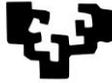


eman ta zabal zazu



Universidad  
del País Vasco

Euskal Herriko  
Unibertsitatea

# **Identificación de variantes genéticas implicadas en el desarrollo de dislexia y falta de atención: Estrategias de caso-control y *loci* de rasgos cuantitativos**

**Mirian Sánchez Morán**

**2017**

Directores:

Ana María Aransay  
Manuel Carreiras



La consecución de esta tesis doctoral ha sido posible gracias a la financiación recibida por el proyecto COEDUCA (Proyecto CONSOLIDER – CSD2008-00048), el Departamento de Industria, Turismo y Comercio del Gobierno Vasco (Programa Eortek 2009-2013) y el Departamento de Innovación Tecnológica de la Diputación de Bizkaia, y ha sido realizada en colaboración entre el CIC bioGUNE y el BCBL.





## Agradecimientos

Por fin se termina este larguísimo viaje! tras años visualizando la meta, la cual cada vez parecía estar más lejos, con un último *sprint*, solo queda un pequeño pasito para cruzarla. Muchas han sido las personas que me han acompañado en este viaje, y a quienes tengo mucho que agradecer, pero sin alargarme demasiado intentaré sintetizar al máximo mis palabras.

En primer lugar me gustaría agradecer a todos esos niños que se sometieron en su día a todas las pruebas psicométricas y que muy generosamente donaron un poquito de saliva, ya que sin ellos no hubiese sido posible realizar este estudio.

A mis directores Ana y Manuel, por haberme guiado en el camino correcto para no desviarme de esa meta.

A mis compañeros de laboratorio por la ayuda con esas miles de extracciones de saliva, cerca de las 800 placas (sí, he hecho la suma) y más de 140 arrays de genotipado que ni me imagino cuanto tiempo me hubieran llevado a mi sola. Igualmente a mis compañeros de “análisis” y al estadístico del COEDUCA por enseñarme a manejar esos archivos de datos infinitos (esos de columnas y filas incontables que no se pueden abrir con un Excel...) así como por la ayuda a la hora de realizar los análisis estadísticos.

Al “koko-team” y a las “Chicas Labayru” (las originales y las nuevas) que han aguantado todas mis quejas (sí, quejas, porque con más o menos razón, todo doctorando se queja).

A mi familia por estar siempre ahí, y especialmente a Yera, la artista de la familia, quien me ha ayudado a diseñar la portada. Ta azkenik zuri, bitxo, holako pazientzia haundixa euki izanagatik, hainbeste urtez helmugara heldu bitartean nere onduan egoteagatik, nire kromosomei buruzko txapak arretaz entzuteagatik eta batez be nigan konfiantza ez galtzeagatik.

Eskerrik asko guztio!





## ABREVIATURAS

<b>ANT</b>	<i>Attention Network Task</i>
<b>APC</b>	<i>Análisis de Componentes Principales</i>
<b>BAF</b>	<i>B Allele Frequency</i>
<b>BRCA</b>	<i>Breast Cancer</i>
<b>C</b>	<i>Comórbido</i>
<b>CD</b>	<i>Trastorno de Conducta</i>
<b>Chr</b>	<i>Cromosoma</i>
<b>CN</b>	<i>Copy Number</i>
<b>CNV</b>	<i>Copy Number Variation</i>
<b>CNVR</b>	<i>Region consenso de CNV</i>
<b>Ct</b>	<i>Cycle Thershold</i>
<b>D</b>	<i>Dislexia</i>
<b>DNA</b>	<i>DesoxiriboNucleic Acid</i>
<b>FA</b>	<i>Falta de Atención</i>
<b>gDNA</b>	<i>Genomic DNA</i>
<b>GWAS</b>	<i>Genome Wide Association Study</i>
<b>GxE</b>	<i>Gene x Environment (interaction)</i>
<b>KEGG</b>	<i>Kyoto Encyclopedia of Genes and Genomes</i>
<b>KO</b>	<i>Knocked out</i>
<b>LD</b>	<i>Linkage disequilibrium</i>
<b>LI</b>	<i>Language Impairment</i>
<b>LRR</b>	<i>logR Ratio</i>
<b>LRR_SD</b>	<i>logR Ratio SD</i>
<b>MAF</b>	<i>Minor Allele Frequency</i>
<b>MMN</b>	<i>MistMatch Negativity</i>
<b>NCBI</b>	<i>National Center for Biotechnology Information</i>
<b>OCDE</b>	<i>Organización para la Cooperación y el Desarrollo Económicos</i>
<b>ODD</b>	<i>Oppositional Defiant Disorder</i>
<b>pb</b>	<i>Pares de bases</i>
<b>PCR</b>	<i>Polimerase Chain Reaction</i>
<b>PISA</b>	<i>Programme for International Student Assessment</i>
<b>qPCR</b>	<i>Quantitative Polimerase Chain Reaction</i>
<b>QTL</b>	<i>Quantitative Trait Locus</i>
<b>R</b>	<i>Repeticiones</i>
<b>RAN</b>	<i>Rapid Automatized Naming</i>

<b>RD</b>	<i>Reading Disability</i>
<b>RNAi</b>	RNA interferente
<b>SD</b>	Desviación estándar
<b>SLI</b>	<i>Speech Language Impairment</i>
<b>SNP</b>	<i>Single Nucleotide Polymorphism</i>
<b>TAE</b>	Tris/Acetato/EDTA
<b>TDHA</b>	Trastorno por Déficit de Atención e Hiperactividad
<b>TEL</b>	Trastorno Específico del Lenguaje
<b>TR</b>	Tiempo de Reacción
<b>UCSC</b>	UC Santa Cruz
<b>UTR</b>	<i>UnTranslated Region</i>
<b>UV</b>	Ultra-Violeta
<b>VNTR</b>	<i>Variable Number of Tandem Repeats</i>

# ÍNDICE

<b>1</b>	<b>INTRODUCCIÓN .....</b>	<b>1</b>
<b>1.1</b>	<b>PANORAMA ACTUAL EN EDUCACIÓN .....</b>	<b>3</b>
<b>1.2</b>	<b>ESTUDIOS GENÉTICOS DE LOS TRASTORNOS COGNITIVOS .....</b>	<b>4</b>
<b>1.3</b>	<b>DISLEXIA .....</b>	<b>6</b>
1.3.1	En busca de rutas causales para explicar la dislexia.....	8
1.3.1.1	La migración neuronal como posible causa principal de la dislexia.....	8
1.3.2	Bases genéticas de la dislexia.....	10
1.3.2.1	DYX1 (15q15-15q21).....	10
1.3.2.2	DYX2 (6p22).....	12
1.3.2.3	DYX3 (2p11-2p16) .....	16
1.3.2.4	DYX5 (3p12-3q13).....	16
1.3.2.5	DYX4 (6q12-q15), DYX6 (18p11.2), DYX7 (11p15) y DYX8 (1p34-1p36) .....	17
1.3.2.6	DYX9 (Xq26-Xq28).....	18
1.3.2.7	FOXP2 (7q31): un gen para el lenguaje y el habla .....	18
1.3.2.8	Análisis de número de copias variable (CNV) en la dislexia.....	19
1.3.3	Estudios de neuroimagen en la dislexia .....	21
<b>1.4</b>	<b>TRASTORNO POR DÉFICIT DE ATENCIÓN E HIPERACTIVIDAD .....</b>	<b>24</b>
1.4.1	Posibles rutas metabólicas afectadas en los individuos con TDAH .....	27
1.4.2	Bases genéticas del TDAH .....	28
1.4.2.1	DRD4 (11p15.5): receptor de la dopamina 4 .....	28
1.4.2.2	DAT1 o SLC6A3 (5p15.3): transportador de la dopamina .....	29
1.4.2.3	COMT (22q11.1-q11.2): catechol-O-metiltransferasa .....	30
1.4.2.4	DBH (9q34): dopamina beta-hidroxilasa.....	31
1.4.2.5	MAOA (Xp11.4-p11.3): monoamina oxidasa A.....	31
1.4.2.6	Otros genes relacionados con el TDAH .....	32
1.4.2.7	Análisis de CNVs en el TDAH .....	33
1.4.3	Estudios de neuroimagen en el TDAH.....	33
<b>1.5</b>	<b>COMORBIDAD ENTRE LA DISLEXIA Y EL TDAH .....</b>	<b>35</b>
<b>1.6</b>	<b>IMPORTANCIA DE LA DEFINICIÓN DEL FENOTIPO EN LOS ESTUDIOS GENÉTICOS DE ASOCIACIÓN EN DISLEXIA Y TDAH .....</b>	<b>36</b>
<b>1.7</b>	<b>INTERACCIÓN ENTRE LOS GENES Y EL AMBIENTE .....</b>	<b>39</b>
<b>1.8</b>	<b>ESTUDIOS DE ASOCIACIÓN DE GENOMA COMPLETO .....</b>	<b>41</b>
<b>2</b>	<b>OBJETIVOS .....</b>	<b>47</b>
<b>3</b>	<b>MATERIALES y MÉTODOS.....</b>	<b>51</b>

<b>3.1</b>	<b>ESTRATEGIAS DE ASOCIACIÓN UTILIZADAS EN ESTE PROYECTO</b> .....	<b>53</b>
<b>3.2</b>	<b>SUJETOS DE ESTUDIO Y CRITERIOS DE INCLUSIÓN</b> .....	<b>54</b>
3.2.1	Pruebas psicométricas y variables registradas para todos los sujetos del presente proyecto .....	54
3.2.1.1	Conciencia Fonológica .....	55
3.2.1.2	Percepción del habla.....	56
3.2.1.3	Fluidez de lectura y velocidad de procesamiento .....	56
3.2.1.4	Decodificación de grafema-fonema.....	57
3.2.1.5	Procesamiento ortográfico .....	58
3.2.1.6	Procesamiento automático/atención selectiva/efecto conflicto .....	59
3.2.1.7	Proceso inhibitorio.....	60
3.2.1.8	Alerta, orientación y efecto conflicto (control ejecutivo) .....	61
3.2.2	Diagnóstico de los sujetos a partir de las variables medidas .....	62
3.2.3	Criterios de inclusión/exclusión de muestras.....	63
<b>3.3</b>	<b>EXTRACCIÓN Y CARACTERIZACIÓN DE DNAs DE SALIVA</b> .....	<b>63</b>
3.3.1	Recogida de muestras de saliva a los participantes .....	63
3.3.1.1	Extracción y cuantificación del DNA .....	64
3.3.2	Caracterización del contenido del DNA proveniente de saliva humana .....	64
3.3.2.1	Cuantificación de DNA Humano .....	64
3.3.2.2	Cuantificación de DNA bacteriano .....	65
3.3.2.3	Cálculo de porcentajes de DNA humano y bacteriano.....	65
<b>3.4</b>	<b>ESTUDIO DE ASOCIACIÓN DE GENES CANDIDATOS</b> .....	<b>66</b>
3.4.1	Selección de las variantes genéticas candidatas .....	66
3.4.2	Estudio de VNTRs .....	67
3.4.2.1	Genotipado de los VNTRs situados en los genes <i>DRD4</i> y <i>DAT1</i> .....	67
3.4.2.2	Clonación de los distintos alelos de VNTRs amplificados .....	67
3.4.2.3	Secuenciación de los productos de PCR clonados .....	68
3.4.2.4	Análisis de asociación de los VNTRs: caso/control .....	68
3.4.3	Estudio de polimorfismos de nucleótido único (SNPs) .....	69
3.4.3.1	Genotipado de los SNPs candidatos seleccionados .....	69
3.4.3.2	Análisis de asociación caso/control.....	69
<b>3.5</b>	<b>ESTUDIO DE ASOCIACIÓN DE GENOMA COMPLETO</b> .....	<b>70</b>
3.5.1	Genotipado de genoma completo: protocolo <i>Infinium</i> .....	70
3.5.2	Análisis de datos de genoma completo: .....	70
3.5.2.1	Análisis conjunto de diferentes versiones de arrays de genotipado.....	70
3.5.2.2	Análisis de reproducibilidad de los arrays de genotipado masivo.....	71
3.5.2.3	Criterios de filtrado de los datos crudos obtenidos mediante los arrays de genotipado masivo.....	71
3.5.2.4	Estructuración y selección de <i>outliers</i> : Análisis de componentes principales.....	72
3.5.2.5	Estudio de asociación: caso/control.....	72
3.5.2.5.1	Análisis de marcadores individuales: SNPs .....	72
3.5.2.5.2	Análisis de haplotipos .....	73

3.5.2.5.3	Análisis de ventanas corredizas ( <i>Sliding Windows</i> ).....	74
3.5.2.5.4	Análisis de interacciones por parejas de marcadores (epistasia).....	74
3.5.2.5.5	Análisis de CNVs .....	75
3.5.2.6	Análisis de enriquecimiento de rutas funcionales relativas a las mutaciones más significativas del estudio.....	77
3.5.2.7	Anotación de las variantes analizadas.....	77
3.5.3	Validación tecnológica de las variantes de mayor relevancia.....	78
3.5.3.1	Selección de las variantes para su validación tecnológica .....	78
3.5.3.2	Genotipado y análisis de replicación de los SNPs seleccionados .....	78
3.5.3.3	Genotipado y análisis de replicación de los CNVs seleccionados .....	78
<b>3.6</b>	<b>ESTUDIO DE ASOCIACION DE RASGOS COGNITIVOS CUANTITATIVOS (QTL)...</b>	<b>80</b>
3.6.1	Análisis de QTLs con los genotipos de los genes candidatos .....	80
3.6.2	Análisis de QTLs con los genotipos de genoma completo .....	81
<b>4</b>	<b>RESULTADOS .....</b>	<b>83</b>
<b>4.1</b>	<b>PUESTA A PUNTO TECNOLÓGICA.....</b>	<b>85</b>
4.1.1	Selección de muestras .....	85
4.1.2	Cuantificación de DNA humano y bacteriano de las muestras provenientes de saliva.....	87
4.1.3	Puesta a punto de la tecnología de arrays de genotipado partiendo de DNA extraído de saliva.....	88
4.1.4	Análisis de reproducibilidad .....	90
<b>4.2</b>	<b>ESTUDIO DE ASOCIACION DE GENES CANDIDATOS .....</b>	<b>91</b>
4.2.1	Asociación de VNTRs con dislexia y falta de atención .....	92
4.2.1.1	Genotipos identificados para el VNTR situado en el gen <i>DRD4</i> .....	93
4.2.1.2	Genotipos identificados para el VNTR situado en el gen <i>DAT1</i> .....	96
4.2.2	Estudio de polimorfismos de nucleótido único (SNPs).....	98
4.2.2.1	Características de los marcadores analizados.....	98
4.2.2.2	Asociación de los marcadores candidatos en la población estudiada..	99
<b>4.3</b>	<b>ESTUDIO DE ASOCIACION DE GENOMA COMPLETO .....</b>	<b>102</b>
4.3.1	Consecuencias del uso de diferentes versiones de arrays de genotipado de genoma completo .....	102
4.3.2	Estudio de asociación caso/control.....	105
4.3.2.1	Detalles de las poblaciones analizadas en genoma completo .....	105
4.3.2.2	Comparación de los resultados obtenidos para los genes candidatos a partir de los datos de genoma completo .....	106
4.3.2.3	Estudio de asociación de genoma completo de polimorfismos de nucleótido único (SNPs) .....	113
4.3.2.3.1	Estudio de asociación de dislexia.....	114
4.3.2.3.2	Estudio de asociación de falta de atención .....	118
4.3.2.3.3	Estudio de asociación con muestras comórbidas .....	121
4.3.2.3.4	Comparación de los genes asociados entre los estudios de dislexia, falta de atención y muestras comórbidas .....	125

4.3.2.3.5	Diferencias en la significancia de los marcadores de mayor asociación en distintos grupos de muestras .....	126
4.3.2.3.6	Enriquecimiento de los marcadores más asociados según su localización génica .....	129
4.3.2.3.7	Validación tecnológica de los marcadores de mayor asociación con los trastornos investigados .....	131
4.3.2.4	Estudio de asociación de genoma completo de CNVs .....	134
4.3.2.4.1	Estudio de asociación de dislexia .....	134
4.3.2.4.2	Estudio de asociación de falta de atención.....	137
4.3.2.4.3	Validación tecnológica de los CNVs de mayor asociación genética.....	137
<b>4.4</b>	<b>ESTUDIO DE ASOCIACION DE QTLs EN LA POBLACIÓN GENERAL.....</b>	<b>139</b>
4.4.1	Estudio de asociación con los genotipos relativos a los genes candidatos seleccionados.....	139
4.4.1.1	Correlaciones entre variables cuantitativas y genotipos de los VNTRs candidatos .....	140
4.4.1.2	Comparativa de las variables fenotípicas cuantitativas con los genotipos de los SNPs candidatos en la población de estudio.....	141
4.4.2	Asociación de valores cuantitativos pertenecientes a rasgos cognitivos en la población general a partir de los datos de genotipado de genoma completo .....	143
4.4.2.1	Rasgos cognitivos relacionados con la lectura.....	143
4.4.2.2	Rasgos cognitivos relacionados con la atención.....	148
<b>4.5</b>	<b>ENRIQUECIMIENTO DE RUTAS FUNCIONALES A PARTIR DE LOS GENES MÁS ASOCIADOS .....</b>	<b>151</b>
4.5.1	Enriquecimiento de rutas funcionales derivado de los estudios de asociación caso/control .....	151
4.5.2	Enriquecimiento de rutas funcionales derivado de los genes más asociados con los rasgos cognitivos a partir de los datos de genotipado de genoma completo .....	157
4.5.2.1	Rutas funcionales asociadas con las variables de lectura .....	158
4.5.2.2	Rutas funcionales asociadas con las variables atencionales .....	160
4.5.2.3	Comparativa de las rutas asociadas entre las variables .....	163
<b>5</b>	<b>DISCUSIÓN .....</b>	<b>167</b>
<b>5.1</b>	<b>LA SALIVA COMO FUENTE DE DNA PARA ESTUDIOS DE GENOMA COMPLETO: LA PRECISA PUESTA A PUNTO DE LA TÉCNICA PARA ESTE TIPO DE MUESTRA .....</b>	<b>169</b>
5.1.1	Observaciones sobre la reproducibilidad del genotipado en las diferentes versiones de arrays <i>Human OmniExpress</i> utilizadas. ....	171
5.1.2	La importancia del estudio de la epistasia: hallazgos de asociaciones más significativas por interacción de SNPs .....	172

<b>5.2</b>	<b>EL ESTUDIO DE GENES CANDIDATOS REVELA ASOCIACIONES DESIGUALES DEPENDIENDO DEL GRUPO DE MUESTRAS CONSIDERADO, ASÍ COMO UNA RELACIÓN CON LOS RASGOS COGNITIVOS DE LA POBLACIÓN GENERAL .....</b>	<b>172</b>
5.2.1	El origen poblacional y el criterio diagnóstico: dos factores clave en los estudios de asociación genética .....	172
5.2.2	La inclusión/exclusión de las muestras comórbidas altera de forma trascendente los resultados de asociación de algunos marcadores .....	174
5.2.3	El gen <i>COMT</i> como posible candidato para la susceptibilidad hacia los problemas de lectura, así como para el estudio de las diferencias cognitivas entre géneros .....	176
5.2.4	Las asociaciones de los alelos de los genes <i>DRD4</i> y <i>DAT1</i> de riesgo a padecer TDAH reportadas en estudios anteriores no se replican en nuestro proyecto.....	177
5.2.5	SNPs ubicados en los genes candidatos revelan asociaciones de interés ....	178
5.2.6	La importancia del diagnóstico certero y homogéneo de las muestras consideradas en análisis genéticos: la utilidad de endofenotipos como alternativa .....	180
5.2.7	Búsqueda de otros posibles genes causales de estos trastornos neurológicos .....	184
<b>5.3</b>	<b>LOS ESTUDIOS DE GWAS DESVELAN SNPS, GENES Y RUTAS FUNCIONALES DE INTERÉS PARA LA DISLEXIA Y LA FALTA DE ATENCIÓN .....</b>	<b>185</b>
5.3.1	Mutaciones en el gen <i>KANK1</i> candidatas a ser causales del desarrollo de dislexia .....	186
5.3.2	Los elementos involucrados en la transmisión sináptica: la causa más observada de trastornos cognitivos .....	188
5.3.3	La adhesión celular y la estructura estrechamente relacionados con la migración neuronal .....	190
5.3.4	Ciertas disfunciones en el sistema inmunitario podrían estar implicadas en el desarrollo de trastornos neurológicos .....	191
5.3.5	La deficiencia de zinc podría tener relación con el desarrollo neurológico defectuoso y los trastornos cognitivos .....	193
5.3.6	¿Podría ser la dislexia una ciliopatía?.....	194
<b>5.4</b>	<b>EL ANÁLISIS DE QTLS MUESTRA REGIONES DE MAYOR ASOCIACIÓN QUE EL ESTUDIO DE CASOS Y CONTROLES .....</b>	<b>195</b>
5.4.1	Dos regiones situadas en los cromosomas 2 y 8 significativamente asociadas al procesamiento automático, la atención selectiva y el efecto conflicto ..	197
<b>5.5</b>	<b>CONSIDERACIONES FINALES.....</b>	<b>198</b>
<b>6</b>	<b>CONCLUSIONES .....</b>	<b>201</b>
<b>7</b>	<b>REFERENCIAS .....</b>	<b>205</b>



# **1 INTRODUCCIÓN**



## 1.1 PANORAMA ACTUAL EN EDUCACIÓN

La lectura es una habilidad fundamental en nuestra sociedad actual como medio de acceso y transmisión de la información, y consecuentemente, para ejercer los derechos como ciudadano o para acceder a un trabajo de calidad. Los avances científicos y tecnológicos realizados en los últimos años han sido numerosos, pero aun así, es difícil entender por qué hay niños que aprenden a leer sin dificultad mientras que otros fallan en esa habilidad tan básica sobre la cual se construye la instrucción de otras destrezas. El informe PISA (*Programme for International Student Assessment*) (<https://www.oecd.org/pisa/>) evalúa el rendimiento escolar en base a las competencias de lectura, matemáticas y ciencias. Los resultados del informe PISA de 2012 indicaron que la puntuación media en comprensión lectora de los estudiantes españoles era 8 puntos inferior al promedio de los países de la OCDE (Organización para la Cooperación y el Desarrollo Económicos), situando a España en los últimos puestos entre todos los países participantes en la evaluación (Figura 1). En el último informe realizado en el 2015 se observa una leve mejoría en la puntuación obtenida en las pruebas de lectura, sin embargo el sistema educativo de España sigue siendo débil, lo cual resulta alarmante. Aspectos como la falta de atención, la falta de control o la dificultad en la lectura contribuyen a esta debilidad ya que las facultades para estas dos habilidades son fundamentales para la educación. Al margen de otras posibles causas que puedan contribuir en estas deficiencias (factores del entorno por ejemplo), dos trastornos relacionados directamente con el aprendizaje, y en consecuencia con el rendimiento escolar, son la dislexia y el trastorno por déficit de atención e hiperactividad (TDAH). Si bien estos dos trastornos pueden derivarse por múltiples orígenes, las alteraciones en el componente genético son un factor importante a tener en cuenta, de modo que su estudio contribuye a la comprensión de los procesos involucrados en la adquisición de estas habilidades y podría proporcionar una guía para un apropiado diagnóstico y clasificación.

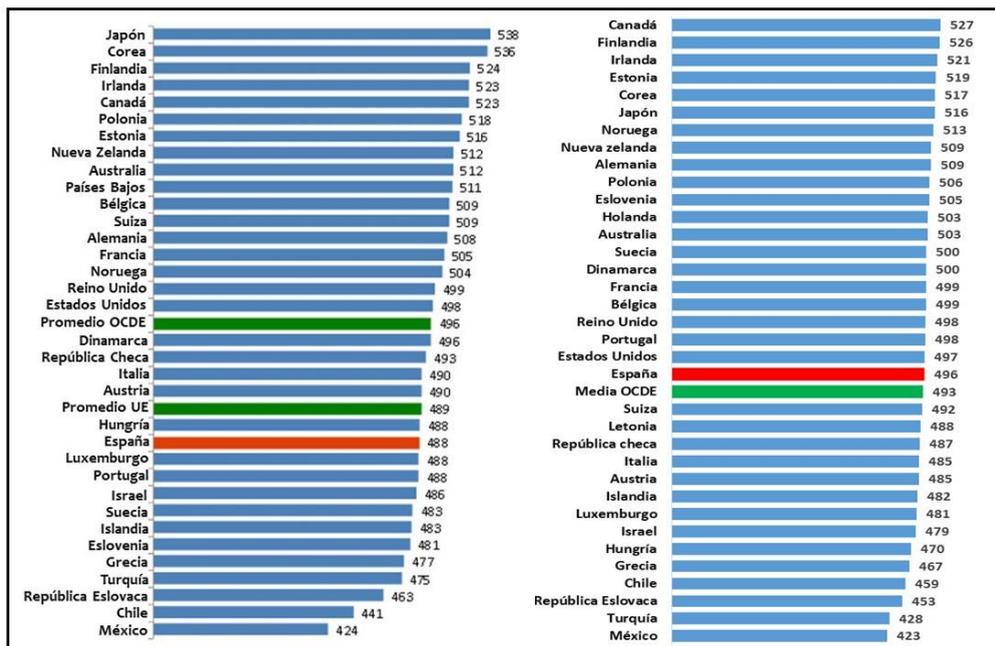


Figura 1.-Resultados en las pruebas de comprensión lectora del PISA-2012 (gráfico de la izquierda) y del PISA-2015 (derecha) en los países participantes de la OCDE.

## 1.2 ESTUDIOS GENÉTICOS DE LOS TRASTORNOS COGNITIVOS

Como se viene demostrado durante varias décadas, muchos de los trastornos y enfermedades humanas tienen una base genética. Sin embargo, es difícil aislar una causa genética en el caso de los trastornos cognitivos o rasgos de la conducta. No obstante, los estudios con gemelos han demostrado que los trastornos cognitivos tienen un alto grado hereditario<sup>1</sup>. Los gemelos que crecen en un mismo entorno se parecerán el uno al otro ya que comparten las mismas condiciones ambientales, sin embargo, en el caso de que la similitud entre gemelos monocigóticos, con una identidad genética completa, sea mayor que la encontrada entre gemelos dicigóticos, los cuales comparten solamente la mitad de su genoma, se les atribuye una influencia a los genes<sup>2</sup>.

La heredabilidad es la proporción de la varianza en un rasgo que puede ser atribuida a la variación genética<sup>3</sup>. Existe una alta heredabilidad entre habilidades cognitivas básicas como la inteligencia o las principales dimensiones de la personalidad, así como en la mayoría de los

trastornos cognitivos del desarrollo. Los neurogenetistas intentan resolver el proceso por el cual surge esta influencia, para lo cual el primer paso es la identificación de los genes en los cuales las variaciones (el genotipo) correlacionan con la diversificación en el dominio cognitivo (el fenotipo)<sup>4</sup>. Para conseguir este objetivo se realizan, entre otros, estudios de asociación de casos y controles, dónde se comparan las frecuencias alélicas o genotípicas de las variantes genéticas entre un grupo de muestras que poseen el rasgo/trastorno y un grupo de muestras que no lo poseen. Estos estudios se pueden abordar desde dos perspectivas: 1) estudio de genes candidatos, el cual se centra en polimorfismos individuales que se sospecha que puedan estar involucrados en la causa del trastorno de interés, 2) GWAS (del inglés *Genome Wide Association Study*, estudio de asociación de genoma completo), el cual consiste en analizar un gran número de polimorfismos a lo largo de todo el genoma para identificar nuevas variantes posiblemente asociadas al fenotipo.

Es fundamental tener en cuenta que ni la conducta ni las funciones cognitivas están directamente codificadas por genes, sino que éstos codifican proteínas, las cuales constituyen el nivel descriptivo más bajo en todo el proceso. Entender bajo qué circunstancias se expresa un gen, qué proteína sintetiza, qué partes de esta proteína se unen a moléculas, cómo influyen todos estos procesos en la célula, y cómo estas cascadas moleculares complejas afectan al desarrollo y funciones cognitivas es el objetivo final de la investigación neurogenética<sup>4</sup>. Por lo tanto, establecer una asociación entre genes y fenotipos cognitivos o conductuales supone una gran dificultad.

Investigaciones previas apuntan a una explicación multifactorial de la influencia genética en la cognición<sup>1</sup>. Además, las mutaciones genéticas relacionadas con este tipo de rasgos están caracterizadas por efectos específicos en diferentes sistemas cuyas inter-relaciones son desconocidas<sup>5</sup>. Si bien se han encontrado numerosos genes que podrían estar relacionados con los procesos cognitivos, no todo se le puede atribuir a los genes. El hecho de que gemelos idénticos no tengan habilidades lingüísticas idénticas, por ejemplo, demuestra que existen factores no-genéticos, tales como factores epigenéticos, ambientales o interacciones entre estos<sup>4</sup>. Las variantes genéticas que afectan a la función cognitiva que se han identificado hasta el momento, son individualmente de efecto pequeño. Igualmente, para que su efecto sea aparente, necesitan la influencia genética de su alrededor, de modo que un alelo concreto puede estar

asociado sólo si co-ocurre con una mutación desventajosa de otro gen o bajo circunstancias ambientales específicas, aunque, en un entorno distinto en cambio, podría ser neutro o incluso ventajoso<sup>1</sup>. Teniendo en cuenta esta heterogeneidad de posibles efectos causales, la noción intuitiva de que una causa singular produce efectos determinantes podría ser insuficiente, y la dicotomía de genes y entorno demasiado simplista como para justificar el desarrollo de los trastornos cognitivos. Por ello, el modelo alternativo de déficit múltiple pasa a ser una explicación más apropiada<sup>4</sup>. Este modelo viene en parte determinado por las ideas de Falconer<sup>6</sup> según las cuales existe un continuo en cuanto a la responsabilidad de causa para un trastorno. Múltiples factores de riesgo podrían afectar a su desarrollo, pero el trastorno sólo es evidente cuando se excede un límite de afecciones.

Es importante tener en cuenta que aunque la mayoría de los efectos genéticos ocurren en interacción con el ambiente, la influencia del entorno o experiencias tienen que pasar primero por rutas sensoriales antes de impactar en el cerebro, y dichos mecanismos sensoriales están diseñados genéticamente. De manera que estas señales externas se convierten en internas y en consecuencia estarán sujetas a las “limitaciones” impuestas por la estructura neuronal. Es decir, cualquier cambio que estas señales pudieran producir en el cerebro estarán determinadas por los mecanismos moleculares que están bajo control genético, por lo que la experiencia sólo puede influir en el organismo mediante el filtro de su propio genoma<sup>4</sup>.

La dislexia y el TDAH son dos trastornos neuro-conductuales con una alta prevalencia en la población<sup>7</sup>. Ambos son considerados anomalías complejas en los que influyen múltiples factores genéticos y ambientales de riesgo<sup>7,8</sup>, y al parecer son muchos los genes implicados en su desarrollo, cada uno contribuyendo con un pequeño efecto a la varianza genética total.

### **1.3 DISLEXIA**

La dislexia es un trastorno del desarrollo caracterizado por dificultades en el aprendizaje de la lectura, aun habiendo tenido unas oportunidades educativas apropiadas, un coeficiente intelectual normal, visión y audición correctas y oportunidades socioculturales adecuadas. Se considera uno de los trastornos neuro-cognitivos más comunes entre los niños<sup>7</sup>. Los síntomas

empiezan en la infancia pero los problemas de lectura suelen persistir en la edad adulta afectando a los logros académicos y/o a la vida social y emocional<sup>9</sup>.

La prevalencia de este trastorno varía entre un 5 y un 10%, pero puede llegar a alcanzar incluso el 17%<sup>10</sup> dependiendo de los criterios establecidos para su diagnóstico o de la lengua nativa de la población<sup>11</sup>, ya que la frecuencia de la dislexia es mayor en las lenguas no-transparentes (como el inglés) que en los idiomas quasi-transparentes (italiano). Una ortografía transparente es aquella en la que existe una correspondencia simple, clara y única entre las letras y los sonidos de la lengua hablada. Asimismo, algunos estudios indican que la prevalencia de la dislexia es mayor en chicos que en chicas<sup>12,13</sup>.

Los individuos disléxicos muestran déficits en procesos cognitivos correlacionados como por ejemplo la memoria verbal, la conciencia fonológica o el procesamiento ortográfico<sup>14</sup>, y algunos investigadores sugieren que la mayor deficiencia ocurre en el procesamiento fonológico, en el nombrado rápido y en el reconocimiento de fonemas<sup>15</sup>. En vista a los diferentes rasgos que pueden estar afectados, existen diversas teorías en lo que concierne a las bases neurológicas de este trastorno. Algunos creen que es un trastorno fonológico, ya que asimilar la correspondencia entre grafema y fonema es un punto esencial para el posterior aprendizaje de la lectura. Otros en cambio, centran sus hipótesis en el procesamiento visual o auditivo<sup>7,10</sup>.

La dislexia es fundamentalmente un déficit familiar y heredable. Los estudios de gemelos han sido cruciales para diferenciar la contribución relativa de los genes y el entorno. Se ha demostrado que la tasa de concordancia entre gemelos disléxicos monocigóticos, los cuales tienen el mismo fondo genético, es del 68% en comparación con un 38% entre los gemelos disléxicos dicigóticos<sup>16</sup>, lo cual indica que los genes contribuyen significativamente en este trastorno cognitivo. También se ha encontrado que la proporción de la varianza en las habilidades para la lectura que se puede explicar mediante los genes es alta, con una heredabilidad de 0.4-0.8<sup>17</sup>. Sin embargo, la búsqueda de estos marcadores genéticos causales relacionados con la dislexia no es tan sencilla, ya que estamos ante un trastorno complejo (causado por diferentes genes y en combinación con el entorno), multifactorial (una combinación de diferentes factores de riesgo) y heterogéneo (el mismo fenotipo puede tener causas distintas en cada individuo). Además hay que considerar que muchos de los factores de riesgo sólo tendrán un tamaño de efecto pequeño o únicamente estarán implicados en algunos casos raros de disléxicos<sup>18</sup>. Aun así,

hay estudios que muestran una fuerte asociación de algunas regiones genéticas con este trastorno<sup>19</sup>. Hasta el momento, se han identificado 9 *loci* (*DYX1-DYX9*)<sup>18</sup> candidatos susceptibles en relación con la dislexia.

### 1.3.1 En busca de rutas causales para explicar la dislexia

Aunque se han propuesto varios genes candidatos asociados con la dislexia (ver sección 1.3.2), todavía son escasas las propuestas sobre las rutas causales. El grupo de Poelman *et al.*<sup>20</sup> propone una ruta molecular cuyo desequilibrio es una posible causa de la dislexia, la cual implica a 10 genes candidatos identificados en la literatura (*ROBO1*, *KIAA0319*, *KIAA0319L*, *S100B*, *DOCK4*, *FMR1*, *DIP2A*, *GTF2I*, *DYX1C1* y *DCDC2*) e introduce 3 nuevos (*SLIT2*, *HMGB1* y *VAPA*). Todos los genes codifican proteínas relacionadas directa o indirectamente con procesos que regulan y modulan los microtúbulos y los filamentos de actina del citoesqueleto celular. La ruta en cuestión, enlaza los genes involucrados en el crecimiento de la neurita y en la migración neuronal, dos procesos neurológicos del desarrollo que utilizan una maquinaria molecular solapada. Sin embargo, no se excluye que puedan existir otros modelos diferentes para conectar dichos genes con la dislexia a un nivel molecular. No es la primera vez que se expone la migración neuronal como la principal causa de la dislexia. Galaburda *et al.*<sup>21</sup> ya mencionan que la función de muchos de los genes involucrados en el desarrollo de la corteza cerebral, incluyendo a *KIAA0319*, *DCDC2*, *DYX1C1* y *ROBO1*, pueden ser los responsables de las malformaciones corticales observadas en cerebros *post-mortem* de pacientes disléxicos<sup>22</sup>, los cuales podrían deberse a defectos en la migración neuronal y en el crecimiento axonal.

#### 1.3.1.1 La migración neuronal como posible causa principal de la dislexia

La neurona es la estructura básica en la composición del cerebro. Durante el desarrollo del sistema nervioso, cada neurona recién nacida debe migrar hasta llegar a su posición definitiva en el cerebro. A este proceso se le denomina migración neuronal, el cual requiere el funcionamiento regulado de una multitud de proteínas que permiten definir el movimiento, la dirección y el lugar dónde debe detenerse cada neurona, para que finalmente éstas alcancen su ubicación final<sup>23</sup>. Entender los mecanismos por los cuales se produce esta migración neuronal es imprescindible para comprender cómo se realizan las conexiones cerebrales. Al mismo tiempo, el estudio de este proceso también ofrece la posibilidad de descifrar posibles causas en los

trastornos del desarrollo neurológico<sup>23,24</sup>. Una migración y organización neuronal apropiada durante el desarrollo es esencial para construir los circuitos sinápticos funcionales en el cerebro, por lo que generar y determinar un número y tipo de neuronas apropiadas es un paso importante para establecer, mantener y modificar dichos circuitos neuronales<sup>24</sup>. De esta forma, alteraciones en el patrón de migración pueden llegar a producir malformaciones cerebrales en los casos más graves o defectos más sutiles de organización neuronal en casos menos severos, pudiendo alterar las conexiones entre las neuronas que nunca llegan hasta su posición final<sup>23</sup>. De manera que cuando el número de neuronas afectadas es grande se pueden producir trastornos neurológicos generalizados. En cambio, si sólo una región cortical es la afectada, estos síntomas serán más restringidos<sup>23</sup>.

Existen tres pasos esenciales y repetitivos en toda migración neuronal. En primer lugar la célula extiende una neurita delantera, ésta se genera como una protusión a partir del cuerpo celular (soma) unida a un paquete de microtúbulos. Al parecer, los microtúbulos estabilizan a la neurita en crecimiento y proveen lo necesario para la extensión de la membrana. Acto seguido se genera un cono de crecimiento que extiende y contrae los filopodios y pliega el avance o colapso de los lamelipodios a medida que va explorando el microentorno. El segundo paso consiste en la translocación del núcleo hacia la neurita destacada, proceso también conocido como nucleokinesis<sup>25</sup>.

Se han identificado dos tipos de migración neuronal en base a su orientación: la radial y la tangencial. En la migración radial las células migran desde la zona dónde han nacido hasta la superficie del cerebro siguiendo el plano radial del tubo neural. De esta manera, las neuronas pasan sobre las capas neuronales creadas anteriormente hasta alcanzar el punto final dónde forman una nueva estructura laminar con un patrón distinto de conectividad. Para ello se ayudan de las células gliales, las cuales sirven como estructura y guía para la migración dirigida. Sin embargo, en la migración tangencial, las células migran ortogonalmente a la dirección de la migración radial. En este proceso no es necesaria la interacción con las células gliales<sup>24,26</sup>. Estas dos rutas de migración no son necesariamente exclusivas y algunas neuronas podrían utilizar ambas para alcanzar su posición final<sup>26</sup>.

### 1.3.2 Bases genéticas de la dislexia

Como ya se ha mencionado anteriormente, pese a la dificultad que conlleva el estudio de trastornos complejos, las investigaciones de los últimos años han identificado 9 *loci* (*DYX1-DYX9*)<sup>18</sup> candidatos que podrían estar vinculados con la dislexia:

#### 1.3.2.1 *DYX1* (15q15-15q21)

Algunas variantes genéticas en el cromosoma 15 fueron posiblemente las primeras que se relacionaron con la habilidad para la lectura<sup>27</sup>. Varios años después, el estudio de Taipale *et al.*<sup>28</sup> indicó el mapeo de un gen denominado *dyslexia susceptibility 1 candidate 1* (*DYX1C1*) en esta región (15q21.3), a partir del punto de ruptura de una translocación (t(2:15)(q11:q21)) hallada anteriormente en una familia finlandesa<sup>29</sup>. La translocación interrumpía el gen *DYX1C1*. En el mismo estudio también describieron una significancia nominal para dos marcadores (-3G>A (rs7343205A) y 1249G>T (rs57809907T)) situados en este gen, así como para el haplotipo formado por ambos.

Los esfuerzos para replicar esta asociación tanto para el fenotipo de dislexia como para rasgos cognitivos y específicamente relacionados con la lectura, han producido resultados dispares (ver detalles de los estudios más relevantes en la Tabla 1) y a veces conflictivos. Por ejemplo, dos estudios han encontrado el mismo marcador asociado pero en la dirección opuesta, es decir, uno de los grupos detectó asociación del alelo de no-riesgo según el estudio anterior<sup>30,31</sup>. Mientras que en el estudio de Taipale *et al.*<sup>28</sup> el alelo A de esa mutación (-3G>A) parecía estar asociado con dislexia en la población finlandesa, en el estudio de Wigg *et al.*<sup>31</sup> fue el alelo G el que se relacionó con este mismo trastorno en la población canadiense. De esta misma forma, el alelo T de la mutación 1249G>T se asoció en el estudio de Taipale *et al.*, mientras en el estudio de Scerri *et al.*, elaborado en base a una población de U.K, fue el alelo G el que se asoció. Se observaron estas mismas discordancias cuando se analizó el haplotipo formado por ambas mutaciones. Otros grupos de investigación también han identificado asociaciones con otros marcadores<sup>32,33</sup>.

Tabla 1. Descripción de los estudios de asociación realizados relativos a los polimorfismos -3G>A y 1249G>T del gen *DYX1C1*. Si=asociación de la variante en el estudio de dislexia. No=no asociación. d.o.= asociación encontrada en dirección opuesta al estudio original. #=asociación únicamente en forma de haplotipo. \*=asociación con rasgos cuantitativos relacionados con la lectura, no en fenotipo dicotómico.

Referencia	Población	-3G>A	1249>T
Taipale <i>et al.</i> 2003 <sup>28</sup>	Finlandia	Si	Si
Scerri <i>et al.</i> 2004 <sup>30</sup>	U.K.	No	Si (d.o.)
Wigg <i>et al.</i> 2004 <sup>31</sup>	Canadá	Si (d.o.)	Si (d.o.)#
BrKanac <i>et al.</i> 2007 <sup>34</sup>	U.S.A.	Si (d.o.)	Si (d.o.)
Marino <i>et al.</i> 2007 <sup>35</sup>	Italia	Si *	Si
Dahdouh <i>et al.</i> 2009 <sup>36</sup>	Alemania	Si#	No
Newbury <i>et al.</i> 2010 <sup>37</sup>	U.K.	No	Si*
Cope <i>et al.</i> 2004 <sup>38</sup>	U.K.	No	No
Meng <i>et al.</i> 2005 <sup>39</sup>	U.S.A.	No	No
Bellini <i>et al.</i> 2005 <sup>40</sup>	Italia	No	No
Saviour <i>et al.</i> 2008 <sup>41</sup>	India	No	No
Bates <i>et al.</i> 2010 <sup>32</sup>	Australia	No	No
Lim <i>et al.</i> 2011 <sup>42</sup>	China	Si	No
Tran <i>et al.</i> 2013 <sup>43</sup>	Meta-análisis	No	No
Becker <i>et al.</i> 2014 <sup>33</sup>	Europa central	No	No
Venkatesh <i>et al.</i> 2014 <sup>44</sup>	India	No	No

Meta-análisis realizados<sup>43,45</sup> indican que no hay evidencias estadísticas para una asociación entre los marcadores -3G>A ó 1249G>T y el riesgo a padecer dislexia, además indican que hay una baja credibilidad de estas asociaciones por la alta heterogeneidad existente entre los estudios.

### **Función molecular del gen *DYX1C1***

Desde su descubrimiento, *DYX1C1* se ha considerado un candidato viable para la etiología de la dislexia, ya que se expresa en diferentes tejidos incluido el cerebro, concretamente en las neuronas corticales y en las células gliales de la materia blanca. Diferentes estudios funcionales han caracterizado la proteína codificada a partir de este gen y, entre sus posibles efectos, se han sugerido los siguientes:

- Tapia *et al.*<sup>46</sup> identificaron factores de transcripción (TFII-I, PARP1 y SFPQ) que podrían regular la expresión de *DYX1C1* de manera alelo dependiente, y modelos animales sugieren que algunos de estos factores tienen funciones en la memoria y en el desarrollo cerebral.
- En el estudio de Massinen *et al.*<sup>47</sup> y posteriormente en el de Tammimies *et al.*<sup>48</sup> se implica funcionalmente a la proteína de *DYX1C1* en la interacción con los receptores  $\alpha$  y  $\beta$  de estrógenos (ESR1 y ESR2), los cuales son importantes para el desarrollo cerebral y procesos neuronales como la migración neuronal y la plasticidad sináptica, incluso podrían estar relacionados con funciones cognitivas como la memoria y el aprendizaje.
- Estudios realizados con ratones en los que se disminuyó la expresión de este gen mediante RNA interferente (RNAi) *in utero* en el neocórtex embrionario, mostraron patrones aberrantes de migración neuronal<sup>49-51</sup>, que eran similares a los observados en los cerebros *post mortem* de disléxicos<sup>22</sup>.
- Varios genes relacionados con la dislexia se encuentran entre los regulados durante la ciliogénesis, lo cual sugiere que los cilios podrían ser causa de una disfunción en la lectura<sup>52</sup>. Además, recientes descubrimientos indican que *DYX1C1* podría estar involucrado en el crecimiento y función de los cilios<sup>52-54</sup>. Aunque la localización principal de la proteína *DYX1C1* es el centrosoma, ésta también se ha observado en el cilio primario de algunas células<sup>52</sup>. El gen ortólogo de *DYX1C1* del pez cebra, el cual se expresa en muchos tejidos ciliados, es esencial para el desarrollo normal de los cilios y su *knockdown* conlleva a múltiples fenotipos relacionados con cililopatías<sup>54</sup>. Por otra parte Tarkar *et al.*<sup>55</sup> reportan que ratones con la delección de los exones 2-4 del gen *DYX1C1* presentan defectos de motilidad ciliar.
- También se ha observado que *DYX1C1* interacciona con LIS1 (una proteína implicada en la lisencefalia) y con *DCDC2* pero no con *KIAA0319*, así como con proteínas componentes del citoesqueleto (*TUBB2B*, *TUBAI* y *Antaxin*)<sup>56</sup>.
- Otros grupos de investigación también han relacionado este gen con la inhibición de la ubiquitina ligasa<sup>57</sup> y como biomarcador para el cáncer colorectal<sup>58</sup>, así como para el cáncer de mama<sup>59,60</sup>.

### 1.3.2.2 *DYX2* (6p22)

El locus *DYX2* es sin duda el más estudiado de todos los loci asociados con dislexia hasta el momento. La primera evidencia de asociación de esta región con la dislexia proviene del estudio

de Cardon *et al.* en 1994<sup>61</sup> y desde entonces muchos han sido los que han identificado esta zona como susceptible para ciertos rasgos relacionados con la lectura<sup>62-64</sup>. Posteriormente se han identificado dos genes en esta región, *DCDC2* (*doublecortin domain containing 2*) (6p22.1) y *KIAA0319* (6p22.3) que se han asociado sucesivamente en numerosos estudios, pero como en el caso de *DYX1C1* los resultados entre ellos no han sido siempre consistentes. En 2004, Deffenbacher *et al.*<sup>65</sup> analizaron 31 marcadores localizados a lo largo de 10 genes de este locus y descubrieron una región enmarcando los genes *VMP*, *KIAA0319*, *TTRAP* y *THEM2*, asociada a rasgos relacionados con la ortografía y la fonología en familias disléxicas de U.S.A.. Partiendo de esta región y con intención de identificar el gen responsable de esta señal, Meng *et al.*<sup>66</sup> seleccionaron 147 marcadores alrededor de un microsatélite previamente identificado (*JA04*) y asociado a una medida de ortografía<sup>62</sup> en esta misma zona. Entre los resultados de su estudio destaca la identificación de una delección de 2445 pares de bases (*BV677278*), localizada en el intrón 2 del gen *DCDC2*, y la asociación de varios marcadores situados en este mismo gen con índices de severidad de dislexia en la población de U.K., por lo que propone el gen *DCDC2* como candidato para la dislexia. El equipo de Schumacher por su parte y en paralelo, asoció el SNP rs793862 así como el haplotipo formado por este marcador junto con rs807701 (ambos ubicados en el gen *DCDC2*) con la dislexia en la población alemana<sup>14</sup>. Ninguno de los dos estudios en cambio, logró relacionar el gen *KIAA0319* con la lectura. Estudios posteriores han intentado replicar estos resultados en diferentes poblaciones, pero las conclusiones logradas no siempre han sido las mismas. En un meta-análisis realizado en base a 4 variantes del gen *DCDC2*, así como de la delección identificada por Meng *et al.*<sup>67</sup>, concluyen que sólo el alelo C del SNP rs807701 está asociado significativamente con el riesgo a padecer dislexia, aunque parece ser que el análisis de sensibilidad indica que los resultados son poco fiables por lo que hay que interpretarlos con precaución.

Por otro lado, los grupos de Francks<sup>68</sup> y Cope<sup>69</sup> también se centraron en la misma región asociada por Deffenbacher<sup>65</sup>. En el primer estudio<sup>68</sup>, identifican un haplotipo de riesgo formado por dos SNPs del gen *KIAA0319* (rs4504469 y rs2038137) y uno del gen *TTRAP* (rs2143340) relacionado con medidas de lectura en familias disléxicas de U.K. y de U.S.A.. Cope *et al.*<sup>69</sup> replican la asociación de los dos marcadores de *KIAA0319* en la población de U.K. (estudio de casos y controles) pero no la del SNP de *TTRAP*. Una de las mayores asociaciones encontradas fue para rs4504469-*KIAA0319*. Un año después, con intención de replicar los resultados obtenidos por estos dos grupos, Harold *et al.*<sup>70</sup> analizan los mismos marcadores (incluyendo algunos nuevos) del

gen *KIAA0319* así como del gen *DCDC2*. Asocian el SNP rs4504469-*KIAA0319* a la dislexia y a diferentes medidas de lectura, y el rs2038137-*KIAA0319* sólo con medidas de lectura. Ni este estudio ni el estudio de Cope<sup>69</sup> obtuvieron resultados significativos para ningún marcador situado en el gen *DCDC2*. Los resultados de asociación obtenidos para el gen *KIAA0319* por los diferentes grupos de investigadores en diversas poblaciones tampoco han sido muy consistentes entre ellos. Sin embargo, del meta-análisis realizado en 2012<sup>45</sup> en base a 5 estudios se dedujo que el alelo T del rs4504469-*KIAA0319* estaba significativamente asociado con el riesgo a padecer dislexia. Diversos estudios han encontrado una interacción entre los genes *DCDC2* y *KIAA0319*, por lo que se sugiere que los dos podrían contribuir independientemente a la dislexia<sup>70-72</sup>. En estudios más recientes<sup>73,74</sup> se asocian ambos genes con la habilidad para la lectura y se proponen tres nuevos genes candidatos: *FAM65B*, *CMAHF* y *S100B*.

En cuanto a los otros genes localizados en esta misma región (*DYX2*), parece ser que el marcador situado en *TTRAP* está en un alto desequilibrio de ligamiento (LD, *linkage disequilibrium*) con el SNP rs9461045 ubicado en *KIAA0319*, el cual afecta a la expresión de *KIAA0319*, pudiendo ser esta la razón de haber encontrado la asociación con *TTRAP*. Por otra parte, la asociación de los genes *THEM* y *VMP* no se ha replicado con respecto a la dislexia<sup>20</sup>.

### **Función molecular del gen DCDC2**

*DCDC2* pertenece a la familia de genes *DCX* (*Doublecortin, DCX-repeat gene family*), caracterizada por su función de empaquetar y estabilizar los microtúbulos neuronales, esenciales para el crecimiento de neuritas y la migración neuronal. Los genes *DCX* son conocidos por su implicación en la lisencefalia, una severa malformación provocada por defectos severos en la migración neuronal<sup>20</sup>. En vista a estas funcionalidades, no es extraño pensar que *DCDC2* esté involucrado en la migración neuronal y varios estudios funcionales parecen haber confirmado estas sospechas. El estudio realizado por Meng *et al.*<sup>66</sup> y posteriormente por Alder *et al.*<sup>75</sup> donde se inhibe la expresión de *DCDC2* (RNAi *in utero*) en ratas, se vio obstaculizada la migración neuronal caracterizada por una reducción progresiva de la distancia recorrida. Estas anomalías eran similares a las encontradas en cerebros humanos *post mortem* de disléxicos<sup>76,77</sup>. En otro estudio sin embargo<sup>77</sup>, no se observaron déficits en la migración neuronal en ratones *DCDC2*-knockout (KO), pero sí tenían dificultades para el aprendizaje, con las tareas de discriminación

visual y con la memoria a largo plazo. Además, *DCDC2* está altamente expresado en las mismas regiones cerebrales activadas durante la lectura tanto en los lectores fluidos como en los disléxicos<sup>66</sup>. A este gen también se le ha atribuido relación con los cilios. Al parecer, una sobreexpresión del gen en las neuronas del hipocampo incrementa la largura media del cilio primario en aproximadamente el doble<sup>78</sup>. En estudios realizados con *Caenorhabditis elegans* también se sugiere que *DCDC2* es importante para la morfología de la neurona y que su sobreexpresión causa una migración neuronal anormal que sólo suele ser observable en las neuronas ciliadas.

### **Función molecular del gen KIAA0319**

Estudios celulares han demostrado que el haplotipo de riesgo asociado con la dislexia, formado por los SNPs rs4504469-rs2038137-rs2143340 en *KIAA0319*, está relacionado con una reducción de la expresión del mismo gen, la cual podría incluso verse reducida hasta más de un 40%<sup>79</sup>. En concreto, en un estudio realizado por este mismo grupo<sup>80</sup>, en el cual se analizaron varias versiones del promotor de *KIAA0319*, se concluyó que el alelo menos frecuente del rs9461045, el cual se encuentra en completo LD con el SNP rs2143340 del haplotipo de riesgo, creaba un sitio de unión para OCT-1 (un factor de transcripción silenciador), razón por la cual se podría explicar la reducción de expresión de este gen, ya que se ha observado que el KO de *OCT-1* hace recuperar el nivel de expresión de *KIAA0319* en el haplotipo de riesgo.

El patrón de expresión de *KIAA0319* en el neocórtex en desarrollo hace pensar que este gen también podría estar involucrado en la migración neuronal, y diversas investigaciones así lo han postulado<sup>75,79,81</sup>. RNAi *in utero* del gen homólogo a *KIAA0319* en ratas resultó en una reducción significativa de la distancia migrada por las neuronas así como en un alargamiento de las dendritas apicales. También se han encontrado heterotopias ventriculares (grupos de neuronas desorganizadas) con neuronas que no migraban adecuadamente. Al corregir este silenciamiento mediante una sobreexpresión del gen *KIAA0319*, la migración neuronal se recuperaba.

Debido al *splicing* alternativo, este gen produce diferentes variantes de la proteína: A (la principal), B y C<sup>82</sup>. La isoforma A contribuye en el plegamiento de las proteínas, la estabilidad, la adhesión celular y la interacción célula-célula, razón por la cual se sugiere que podría estar

involucrada en la interacción entre las neuronas y las fibras gliales durante la migración neuronal<sup>21,83,84</sup>. La isoforma B se ha detectado secretada en el medio extracelular, por ello se hipotetiza que KIAA0319 podría estar involucrado en funciones más generales en las que se incluye la señalización extracelular<sup>83,85,86</sup>.

Además, a parte de los genes *DYX1C1* y *DCDC2*, el estudio bioinformático del grupo de investigación de Ivlieb *et al.*<sup>53</sup> añade el gen *KIAA0319* a la lista de genes con función ciliar.

### 1.3.2.3 *DYX3* (2p11-2p16)

Este locus se mapeó por primera vez en una familia multigeneracional de Noruega la cual presentaba un patrón de herencia dominante de la dislexia<sup>87</sup>. Estudios posteriores también han asociado esta región con dislexia o con varias medidas cuantitativas relacionadas con la lectura en familias disléxicas y en la población general<sup>88-90</sup>. En 2002, Franck *et al.*<sup>88</sup> propusieron los genes *SEMA4F* (perteneciente a una familia de proteínas con función en la orientación axonal), *OTX1* (un factor de transcripción involucrado en la morfogénesis cerebral) y *TACR1* (receptor de la sustancia P de *tachykinin* y expresado en el sistema nervioso) como posibles candidatos a estar asociados con dislexia dentro de este locus, pero no lograron hallar variantes de riesgo relacionadas. En 2007, el grupo de Anthoni<sup>91</sup> identificó los genes *C2ORF3* y *MRPL19*, los cuales muestran un patrón de expresión cerebral en las mismas regiones donde se expresan *DYX1C1*, *DCDC2*, *ROBO1* y *KIAA0319*. También se describió un haplotipo de riesgo que contenía los genes *C2ORF3* y *MRPL19*, y se mostró que los individuos que portaban los alelos de riesgo presentaban una expresión menor de ambos genes en comparación con las muestras que portaban alelos de no-riesgo.

### 1.3.2.4 *DYX5* (3p12-3q13)

La búsqueda de posibles regiones relacionadas con la dislexia a cargo de Nopola-Hemmi *et al.* se realizó en cuatro generaciones de una familia finlandesa con 21 familiares disléxicos, en la que el trastorno se segregaba de forma autosómica dominante. Este estudio localizó la región situada en el cromosoma 3 que se denominó *DYX5*<sup>92</sup>. Este mismo locus ha sido asociado posteriormente con rasgos relacionados con la lectura en otras familias disléxicas así como en la población general<sup>93,94</sup>. Pero el gen responsable de estas asociaciones no se descubrió hasta que

Hannula-Jouppi *et al.* identificaron una translocación cromosómica en un paciente disléxico (t(3;8)(p12;q11)) que interrumpía el gen *ROBO1* (3p12) entre dos de sus promotores<sup>15</sup>. Tras este descubrimiento, este mismo grupo volvió a analizar a la familia finlandesa estudiada por Nopola Hemmi<sup>92</sup> y se percató de que 19 de los 21 familiares afectados poseían un haplotipo de riesgo formado por marcadores de *ROBO1*, el cual provocaba una atenuación de la expresión del gen, por lo que se sugirió que el mecanismo causal podría ser una regulación alterada del mismo. Según concluyeron en su estudio, son necesarias dos copias funcionales de *ROBO1* para adquirir una habilidad normal en la lectura, y una haplo-insuficiencia parcial de *ROBO1* podría predisponer a desarrollar dislexia.

### **Función molecular del gen ROBO1**

*ROBO1* (*Roundabout Guidance Receptor 1*) se expresa abundantemente en el tejido cerebral y su función en el desarrollo del cerebro y en la migración neuronal está bien establecida. Es uno de los 4 genes ortólogos del gen *Roundabout* de *Drosophila*, el cual es crítico para el entrecruzado axónico normal entre las dos mitades del cerebro. La proteína codificada por *ROBO1* actúa como receptor de las proteínas SLIT, las cuales se encargan de dirigir la migración de las neuronas al neocórtex. Interaccionando con netrinas y sus receptores, aparte de regular la dirección, también regulan el ratio del crecimiento axonal/neurita<sup>95</sup>. Estudios en ratones *Robo1*-KO indican que estos animales mueren al nacer y tienen defectos en la migración neuronal entre otros síntomas<sup>96</sup>.

#### **1.3.2.5 DYX4 (6q12-q15), DYX6 (18p11.2), DYX7 (11p15) y DYX8 (1p34-1p36)**

La asociación del locus *DYX4* con dislexia no se ha replicado desde que éste se asoció a la codificación fonológica en pacientes con este trastorno<sup>97</sup>, aunque se ha sugerido una posible relación con el deletreo de palabras irregulares en la población general<sup>90</sup>. En otro estudio Fisher *et al.*<sup>94</sup> identificaron marcadores del cromosoma 18p11.2 asociados a medidas de lectura de palabras en familias disléxicas, que denominaron como locus *DYX6*. El locus *DYX6* también se ha relacionado con medidas de codificación fonológica y ortográfica, así como para la habilidad en la lectura<sup>90,98</sup>. Algunos de los genes que se han propuesto para esta zona son: *MCR5*, *DYM*, *NEDD4L* y *VAPA*<sup>20,99</sup>. Así mismo, Fisher *et al.*<sup>94</sup> también sugirieron una posible relación en el cromosoma 11

(11p15) y la conciencia fonológica, denominando a esta región asociada *DYX7*. Finalmente, el *locus* *DYX8* se nombró gracias a la relación encontrada de una región del cromosoma 1 con medidas cualitativas y cuantitativas de la dislexia<sup>100,101</sup>. Aunque este cromosoma ya había sido implicado anteriormente a través de una translocación (región 1p22) que se cosegregaba con dificultades para la escritura y la lectura<sup>102</sup>, no fue hasta el año 2008 cuando el grupo de Couto propuso el gen *KIAA0319L* como posible candidato dentro de este *locus*<sup>103</sup> al encontrar significancia nominal con uno de sus marcadores y la dislexia en familias canadienses con dificultad para la lectura. Su similitud de secuencia proteica con *KIAA0319* (61%) hace pensar que podría tener funciones similares a ésta y por consiguiente se considera un buen gen candidato.

### 1.3.2.6 *DYX9 (Xq26-Xq28)*

Mediante el estudio de ligamiento a nivel de genoma completo (*genome-wide linkage scan*) realizado en una familia holandesa en la cual 15 de los 29 miembros se diagnosticaron como disléxicos, se identificó un pico de ligamiento en el cromosoma X, Xq27.3<sup>104</sup> que posteriormente se ha replicado en familias francesas<sup>105</sup>. Sin embargo, en el análisis de la secuencia codificadora de los genes que se encuentran en esta región (*FMRI*, *Cxorf1 (TMEM257)*, *DKFZp574M2010*, *KIAA1854 (SLITRK2)*, *Cxorf5l*, *FMR2*, *ASFMRI* y *FMRINB*) no se detectó ninguna mutación, por lo que el gen causante de la dislexia en estas familias particulares sigue sin estar resuelto.

### 1.3.2.7 *FOXP2 (7q31): un gen para el lenguaje y el habla*

Muchos de los individuos con dislexia padecen problemas similares a los del procesamiento del lenguaje oral, razón por la cual no se descarta que la habilidad para la lectura y otros trastornos del lenguaje compartan mecanismos genéticos comunes. El SLI (del inglés *Speech Language Impairment*), en castellano TEL (Trastorno Específico del Lenguaje), se considera una de las formas principales de los trastornos del lenguaje y es altamente comórbido con la dislexia, pudiendo llegar incluso hasta el 55 %<sup>18</sup>. El gen *Forkhead Box P2 (FOXP2)*, definido como uno de los genes responsables del desarrollo del habla y el lenguaje<sup>106</sup> está involucrado en el SLI<sup>107</sup> y una de sus variantes, el rs21533005, se ha asociado con los problemas del procesamiento del lenguaje escrito en disléxicos<sup>108</sup>. Estudios recientes lo han asociado con algunos rasgos de la dislexia<sup>109</sup>.

El gen *FOXP2* fue descubierto a raíz del estudio realizado en 1987 de una familia trigeracional, nombrada KE, con un trastorno del habla muy severo. De los 37 miembros de los que constaba, 15 padecían dispraxia verbal severa, un trastorno de la coordinación causado por una inadecuada transmisión de los mensajes del cerebro a los músculos de la cara. Un estudio de ligamiento de genoma completo en esta familia identificó un pico de asociación que se localizaba en la región cromosómica 7q31<sup>110</sup>. Dos años después, el grupo de Lai localizó una translocación entre los cromosomas 5 y 7 que afectaba al gen *FOXP2* en dos pacientes con un diagnóstico similar al de la familia KE<sup>111</sup>. La secuenciación del mismo reveló una mutación en esta familia (R553H) que estaba presente de forma heterocigota en todos los afectados<sup>112</sup>. Dicha mutación se localiza en el dominio de unión al DNA del gen, por lo que afecta a su capacidad de unión y a su localización nuclear.

### **Función molecular del gen FOXP2**

*FOXP2* actúa como un represor transcripcional mediante la unión directa a las zonas reguladoras de los genes diana. De esta forma podría controlar el crecimiento y la diferenciación de una clase de neuronas destinadas a inervar tejidos involucrados principalmente en la producción del habla<sup>113</sup>. Uno de los genes que está regulado mediante *FOXP2* es *CNTNAP2*, el cual codifica una de las proteínas neurexinas con múltiples funciones de gran importancia en el sistema nervioso central. Variantes de este gen se han asociado con medidas del lenguaje como la repetición de pseudopalabras en muestras con SLI<sup>107</sup> y también en muestras con dislexia<sup>109</sup>. Estudios funcionales realizados con ratones *knock-in* para *Foxp2*, revelan que los individuos homocigotos para la mutación en dicho gen sufrían un desarrollo atrasado con defectos motores resultando finalmente en una muerte temprana. Los heterocigotos en cambio, si bien se desarrollaban normalmente, padecían una alteración de comunicación en las vocalizaciones emitidas<sup>114</sup>.

### **1.3.2.8 Análisis de número de copias variable (CNV) en la dislexia**

Durante las últimas décadas, la investigación de enfermedades psiquiátricas se ha basado en la hipótesis de enfermedad común, variante común (*Common Disease, Common*

*Variant*, CDCV) para explicar la influencia genética de ésta, es decir, que las variantes comunes de baja penetrancia son las mayores contribuidoras de la susceptibilidad genética de las enfermedades comunes. Sin embargo, las asociaciones encontradas hasta el momento solamente explican una proporción pequeña de la heredabilidad, por eso, la hipótesis alternativa de enfermedad común, variante rara (*Common Disease, Rare Variant*, CDRV) está cobrando cada vez más importancia. En este modelo las variantes raras (de frecuencia menor a 5%) tienen un efecto de riesgo grande, pero cada alteración por sí sola resulta causal en una pequeña proporción de afectados<sup>15</sup>. Si bien es cierto que en la mayoría de los pacientes se encuentra una arquitectura genética compleja con muchas variantes de riesgo, también hay casos inusuales donde el patrón de herencia es consistente con una mutación rara en un gen concreto con un gran efecto<sup>18</sup>. Muchas de estas variantes raras suelen consistir en variaciones en número de copia o CNVs (del inglés *Copy Number Variations*). Los CNVs son anomalías estructurales cromosómicas raras que conforman alrededor del 13% del genoma humano. Estas variaciones, las cuales son el resultado de una recombinación o replicación errónea, se han implicado en la etiología de varias condiciones psiquiátricas<sup>16</sup>. Este tipo de polimorfismos operan a una escala mayor originando deleciones, inserciones, duplicaciones y reordenamientos de regiones de DNA desde 100 pares de bases (pb) hasta varios miles de pb. Así los CNVs pueden abarcar regiones cromosómicas que contienen un gran número de genes y afectar de esta forma a un amplio rango de características fenotípicas<sup>1</sup>.

Hasta el momento el número de estudios centrados en el análisis de CNVs en relación a la **dislexia** no son muchos, pero entre ellos se han encontrado varios genes que podrían tener relación con la susceptibilidad a padecer este trastorno. Los primeros hallazgos que relacionan los CNVs con la dislexia vinieron a través del grupo de Girijan<sup>17</sup>, el cual analizó este tipo de polimorfismo en diferentes fenotipos cognitivos, observando que había una disminución significativa en la frecuencia de CNVs largos en los disléxicos en comparación con los demás trastornos. Stefansson *et al.*<sup>118</sup> también relacionaron varios CNVs con diferentes funciones cognitivas. Específicamente observaron que los controles con historia de dislexia y discalculia poseían una deleción localizada en 15q11.2. Sin embargo, los resultados más prometedores los trajo el grupo de Veerappa<sup>119-122</sup>, quien identificó varios casos de disléxicos con duplicaciones y deleciones en Xq21.3 y Yp11.2 tras un escrutinio de los CNVs a nivel de genoma completo en 11 familias disléxicas. También hallaron CNVs que albergaban genes de las protocaderinas en la

misma región del cromosoma X, particularmente el gen *PCDH11X*, el cual está implicado en la comunicación célula-célula, la habilidad verbal, la asimetría cerebral y la plasticidad dendrítica sináptica, por lo que este gen podría considerarse un buen candidato relacionado con la dislexia<sup>119,120</sup>. Este mismo grupo también identificó 5 CNVs *de novo* en familias disléxicas hindúes que contenían concretamente los genes *NEGR1*, *ACCN1*, *DCDC5* y el ya conocido candidato *CNTNAP2*, los cuales están implicados en el aprendizaje, cognición y procesos de la memoria<sup>121</sup>. Además también identificaron una deleción y 5 microduplicaciones en la región 17q21.31, donde se sitúan los genes *KIAA1267*, *LRRC37A*, *ARL17A/B* y *NSFP1*<sup>122</sup>. Los análisis de rutas moleculares sugieren que todos estos genes podrían estar implicados en la transmisión sináptica, la orientación axonal, la adhesión celular y la fusión de membrana celular/vesicular.

### 1.3.3 Estudios de neuroimagen en la dislexia

La habilidad cognitiva general es un rasgo altamente heredable. Aunque todavía no se han identificado genes con un alto impacto en esta capacidad, algunos se han correlacionado reiteradamente con medidas de lectura<sup>123</sup>. Así mismo, la variación de las regiones cerebrales relacionadas con el lenguaje también parecen estar bajo un fuerte control genético<sup>124</sup>. Por esta razón, la asociación directa de los rasgos anatómicos del cerebro, mediante **imágenes estructurales**, puede dar pistas sobre los posibles efectos de las variantes genéticas asociadas. Por otro lado, el estudio de **imágenes funcionales** del cerebro, es decir, el registro de la activación cerebral mientras los sujetos realizan alguna tarea, podría ofrecer una explicación más exacta de los procesos necesarios para dichas tareas, por lo que también se puede analizar la relevancia funcional de ciertas variantes genéticas en la activación cerebral durante la lectura, por ejemplo<sup>108,125</sup>.

Algunos resultados son prometedores para algunos trastornos neurológicos, como la ansiedad, procesos atencionales o la esquizofrenia. Sin embargo, **son pocos los estudios** realizados para la dislexia<sup>108</sup>. Los primeros hallazgos en base a la estructura cerebral de disléxicos, vienen determinados por el grupo de Galaburda<sup>23</sup>, el cual examinó tejido cerebral *post-mortem* de individuos con dislexia. El estudio reveló, aparte de las ya mencionadas anomalías en la migración neuronal, defectos estructurales corticales microscópicos o ectopias corticales. Investigaciones que comparan entre lectores normales y lectores con dificultades han demostrado que **existen**

**diferencias funcionales y estructurales** en regiones del cerebro relacionadas con la lectura<sup>108,125</sup>. Además, varios análisis han encontrado cambios en la volumetría, densidad o estructura de la materia gris y la materia blanca<sup>124,125</sup>, incluso en un caso han identificado una reducción de la materia gris en los sujetos disléxicos<sup>126</sup>. También se han descrito diferencias en los patrones de activación funcional durante la ejecución de tareas fonológicas o de lectura<sup>108,125,127,128</sup>, demostrando una activación menor de las áreas dorsales y ventrales en los disléxicos y, en ocasiones, una sobre-activación de la zona inferior frontal<sup>108</sup>, aunque esto simplemente podría ser el efecto de una estrategia compensatoria de lectura. Finalmente, se ha concluido que los sujetos con dificultades en la lectura utilizan regiones del cerebro y circuitos neuronales diferentes durante las tareas de lectura, sugiriendo que los determinantes de esta dificultad alteran el desarrollo y el funcionamiento del cerebro<sup>125,129</sup>.

En los pocos proyectos que han indagado sobre la relación entre variantes genéticas y la **estructura cerebral** con respecto a la habilidad para la lectura, se han identificado ciertas regiones cerebrales asociadas a determinados polimorfismos genéticos. El grupo de investigación de Meda<sup>124</sup> observó que los sujetos heterocigotos para la delección del intrón 2 del gen *DCDC2*, en comparación con la ausencia de delección, exhibían volúmenes de materia gris significativamente superiores en regiones cerebrales relacionadas con la lectura/lenguaje en sujetos sanos de Estados Unidos, por lo que sugiere que este gen podría ejercer su efecto a través de todo el espectro de la habilidad para la lectura. Scerri *et al.* por su parte<sup>123</sup> también encontraron asociación del alelo G del marcador rs917235 situado en el locus *MRPL19/C2ORF3* con un menor volumen de la materia gris en la población sueca, sugiriendo así mismo la probabilidad de estar asociado a habilidades cognitivas generales. La concentración de la materia gris también se ha visto alterada en pacientes con esquizofrenia en relación a la variación genética de *FOXP2*<sup>130</sup>. El grupo de Darki<sup>131</sup>, en cambio, describió que los SNPs rs3743204-*DYX1C1*, rs793842-*DCDC2* y rs6935076-*KIAA0319* mostraban efectos significativos en la variabilidad normal del volumen de la materia blanca en las regiones temporoparietales izquierdas, la cual se correlaciona con la habilidad para la lectura. Posteriormente este mismo grupo<sup>132</sup> volvió a escanear a 76 de estos mismos sujetos (un total de 3 veces a lo largo de 2 años) de edades entre 6 y 25 años, sin ningún tipo de trastorno neurológico. En este segundo estudio, aparte de replicar los resultados anteriores, descubrieron que el alelo T del marcador rs793842 de *DCDC2* estaba significativamente asociado a un mayor grosor cortical de las regiones temporales y parietales, relacionadas con un menor volumen de materia blanca

durante el desarrollo de la infancia y en concordancia con menores calificaciones lectoras. Concluyen que el volumen de la materia blanca en estas regiones es un buen predictor de la comprensión lectora futura, es decir, 2 años más tarde. *KIAA0319* también parece afectar a la morfometría cerebral<sup>126</sup>, el KO embrionario de *kiaa0319* en ratas no produce cambios en el volumen del córtex ni del hipocampo, pero sí una reducción significativa en el *corpus callosum*. Estudios recientes de asociación de genoma completo<sup>133</sup> han identificado 5 variantes genéticas que influyen en el volumen del putamen, el núcleo caudado, el hipocampo y el volumen intracraneal.

En cuanto a la investigación de las **áreas funcionales cerebrales** con respecto a las variantes genéticas, en 2012 se llevaron a cabo 4 estudios con resultados destacables:

- (i) La ausencia del alelo G del SNP rs12533005 situado en el gen *FOXP2* mostró sobre-activación en ciertas zonas cerebrales involucradas en el procesamiento del lenguaje durante la ejecución de una tarea de lectura en la población alemana<sup>108</sup>.
- (ii) Ciertas variantes de los genes *DCDC2* y *TTRAP* se asociaron con el patrón de activación de varias regiones cerebrales concretas durante la ejecución de tareas de lectura, pero solo variantes situadas en el gen *DCDC2* pasaron la corrección estadística de falsos positivos<sup>125</sup>.
- (iii) Los sujetos disléxicos con el haplotipo raro de expresión débil de *ROBO1* presentan una interacción interaural defectiva en el proceso de cruzamiento axonal en las rutas auditivas, por lo que se sugiere que para un funcionamiento normal de las rutas auditivas se requiere una adecuada expresión del nivel de este gen<sup>134</sup>.
- (iv) Finalmente, el grupo de Pinel<sup>135</sup> al genotipar y escanear 94 adultos sanos durante tareas de escritura y comprensión del lenguaje hablado, identificaron 5 SNPs en *FOXP2* asociados significativamente con la activación en dos regiones frontales del cerebro y un SNP de *THEM2* asociado con la asimetría de la activación en el lóbulo temporal durante tareas de lectura y de fonología.

En vista a los resultados hallados en los diferentes estudios, la utilidad de *imaging-genetics* queda demostrada, y se puede afirmar que las imágenes cerebrales son un buen endofenotipo para condicionar asociaciones genéticas y explorar mecanismos causales así como rutas biológicas que determinen los trastornos neurocognitivos<sup>125</sup>, por lo que el estudio de neuroimágenes podría servir para establecer puentes entre los marcadores genéticos y la conducta<sup>132</sup>. Además, también se podrían identificar los posibles factores que median las

funciones cognitivas y los trastornos del desarrollo<sup>123</sup>. Sin embargo, el número de estudios realizados entre los genes candidatos en relación a dificultades lectoras y la activación de regiones cerebrales durante las tareas de lectura es todavía muy limitado.

#### 1.4 TRASTORNO POR DÉFICIT DE ATENCIÓN E HIPERACTIVIDAD

El trastorno por déficit de atención e hiperactividad (TDAH), descrito por primera vez por el pediatra George Still en 1902, es un trastorno iniciado en la infancia que se caracteriza por una excesiva actividad motora, impulsividad e inatención o falta de concentración<sup>8</sup>. Este patrón de síntomas es un riesgo importante de fracaso escolar. Dependiendo de los síntomas predominantes se divide frecuentemente en tres subtipos: tipo inatento, tipo hiperactivo-impulsivo o tipo combinado<sup>136</sup>. Estos síntomas son crónicos y persisten en la edad adulta en un 30-60 % de los casos<sup>116</sup>. Los niveles de hiperactividad e impulsividad tienden a decrecer mientras que la inatención es más persistente<sup>136</sup>.

Algunas investigaciones indican que existen diferencias etiológicas para la inatención y los síntomas de hiperactividad/impulsividad<sup>137</sup>. Las publicaciones de los últimos 20 años ofrecen evidencias de la utilidad y validez de distinguir estos dos dominios de síntomas que aun estando altamente correlacionados, predicen diferentes deficiencias. Además se acepta que tengan diferentes correlaciones clínicas y neuronales<sup>138-140</sup>. Específicamente, los síntomas de inatención tienden a predecir los problemas académicos, las dificultades conductuales, el ignorar a sus pares y debilidades neurocognitivas, mientras que los síntomas de hiperactividad/impulsividad tienden a predecir la agresión, o el rechazo de los pares, entre otras dificultades<sup>141</sup>. Según argumentan Milich *et al.*<sup>142</sup>, reducir la variabilidad fenotípica del TDAH podría ser de ayuda para incrementar el poder estadístico a la hora de detectar genes candidatos en los estudios moleculares. Sugieren que tener rutas etiológicas únicas es evidencia suficiente para categorizar los subtipos de TDAH como trastornos diferentes y proponen que las investigaciones futuras deberían buscar las causas neurobiológicas de la sintomatología de inatención en ausencia de síntomas de hiperactividad/impulsividad para que las influencias genéticas de estos síntomas se puedan entender mejor. En este estudio nos hemos centrado en niños que padecían síntomas de inatención o falta de atención (FA).

La prevalencia mundial del TDAH varía entre el 2-10%, y como en el caso de la dislexia se observa una mayor frecuencia en chicos con un ratio de chico:chica de 4:1, aunque esta diferencia podría deberse a su mayor facilidad de detección en el grupo de género masculino ya que el trastorno se presenta normalmente con una mayor alteración. El fenotipo del trastorno también parece diferir entre géneros, los chicos tienden a ser más hiperactivos mientras que las chicas normalmente muestran más síntomas de inatención<sup>143</sup>.

Las teorías neurobiológicas del TDAH se han centrado principalmente en dos modelos comunes y existen evidencias que apoyan ambas, por lo que se sugiere que los dos sistemas psicobiológicos están involucrados en el TDAH (para una información más amplia consultar<sup>144</sup>). El primer modelo, denominado **top-down**, se centra en el funcionamiento ejecutivo<sup>145</sup>. Las funciones ejecutivas permiten una organización compleja de la conducta mediante una selección apropiada de los procesos. Barkley<sup>145</sup> sugiere que los síntomas del trastorno están causados por déficits en la respuesta inhibitoria, lo cual acarrearía alteraciones en funciones ejecutivas específicas, consistente con las deficiencias en las tareas que miden estos procesos en los pacientes con TDAH. Según este modelo, la inhibición de respuesta es la habilidad para suprimir una réplica inapropiada a favor de una alternativa más apropiada, por lo que es un paso fundamental en la conducta dirigida a metas y en consecuencia su relación con el TDAH es un factor importante. En el trastorno también se han identificado problemas en otros dominios relacionados con el ejecutivo central como la memoria de trabajo, el cambio y mantenimiento de tarea, la capacidad de planificación, etc.. El otro modelo, denominado **bottom-up**, se centra en la regulación motivacional, de recompensa, y emocional<sup>146</sup>. Estos procesos se relacionan con una regulación pobre del afecto. El TDAH frecuentemente se asocia a problemas con las emociones como pueden ser el enfado o el humor. La regulación emocional reflejaría la habilidad de cambiar o retrasar la respuesta inicial con el objetivo de crear una conducta más adecuada para el “entorno”. Los **modelos integrados** proponen considerar las dos alternativas para entender el patrón clínico de este trastorno complejo, de forma que la regulación emocional está conectada a los sistemas de memoria de trabajo y su disfunción podría estar más relacionada con los síntomas de inatención, mientras que la inhibición emocional estaría asociada a la habilidad de suprimir una reacción para expresar una conducta más adecuada, relacionada con los síntomas hiperactivos e impulsivos<sup>147</sup>.

Existe una cierta controversia acerca de si este trastorno realmente existe o si es un diagnóstico inventado, debate en el que la industria farmacéutica parece jugar un papel importante. Hay quienes opinan que el TDAH no existe y se oponen a la medicación como tratamiento en niños. Otros, en cambio, aseguran que su causalidad biológica está comprobada y defienden la eficacia de la medicación sobre cualquier otra medida terapéutica. Inventado o no, los estudios de familias, de gemelos y de adopción muestran que el TDAH es altamente heredable (con una estimación de heredabilidad de un 75-90%)<sup>116</sup> y se considera un trastorno complejo<sup>8</sup>. Si bien existen estudios que **identifican diferentes factores genéticos y ambientales** asociados con los diferentes subtipos del TDAH<sup>148</sup>, esta evidencia no siempre es consistente, ya que también se han encontrado factores genéticos comunes para ambos<sup>149</sup>. El TDAH se manifiesta como un espectro variante de síntomas y es dudoso que un único gen pueda contribuir a todos ellos<sup>150</sup>. Hasta el momento no se ha encontrado ningún gen que describa más del 5% de la varianza fenotípica del TDAH. Por eso se piensa que muchos genes, cada uno contribuyendo con un pequeño efecto en la varianza genética total, están implicados en este trastorno y son candidatos para poder explicar la neurobiología del mismo. Estudios longitudinales realizados con gemelos indican que estas influencias genéticas también afectan a la continuidad de los síntomas en la edad adolescente<sup>151</sup>.

En 1971, Wender<sup>152</sup> desarrolló una hipótesis sobre la posible base biológica del TDAH la cual dirigía hacia una disfunción en los neurotransmisores catecolaminérgicos (dopamina y norepinefrina), sugiriendo un posible efecto genético. Levy refinó esta hipótesis desarrollando la teoría del déficit de la dopamina<sup>153</sup>. Años después estudios de imágenes cerebrales documentaron anomalías anatómicas en regiones cerebrales específicas coincidiendo con altas densidades de receptores de dopamina. Además estas regiones parecen ser cruciales para ciertos componentes atencionales<sup>154</sup>. Estudios con modelos animales sugieren que las anomalías en la neurotransmisión de la dopamina podrían ser fundamentales en el control motor y otras funciones neuropsicológicas que están afectadas en los pacientes con TDAH<sup>150</sup>. Sin embargo, la evidencia farmacológica es la más sólida en cuanto a la involucración de las rutas dopaminérgicas, ya que la medicación con psicoestimulantes como el metilfenidato o la anfetamina, los cuales inhiben la actividad del transportador de la dopamina y en consecuencia elevan la cantidad de esta molécula en la sinapsis, controlan eficazmente los síntomas del TDAH<sup>150</sup>.

### 1.4.1 Posibles rutas metabólicas afectadas en los individuos con TDAH

Como se ha mencionado anteriormente, la teoría que más aceptación tiene en cuanto a la etiología molecular del TDAH es una disfunción en el sistema dopaminérgico. La dopamina es un neurotransmisor que activa 5 receptores de dopamina: DRD1, DRD2, DRD3, DRD4 y DRD5. Esta molécula se produce en varias zonas cerebrales y se libera cuando existe un estímulo por el cual merece la pena trabajar para conseguir su recompensa. Las rutas que transmiten el neurotransmisor dopamina desde una región del cerebro a otra se denominan rutas dopaminérgicas, y existen 8 diferentes, cada una responsable de transmitir la dopamina de una zona a otra. Dos de estas rutas, la mesolímbica y la nigrostriatal, se han estudiado extensivamente en el TDAH por su potencial papel en la modulación de la respuesta de la conducta.

En una transmisión sináptica se libera la dopamina empaquetada en vesículas de la neurona presináptica y se une al receptor de la dopamina localizado en la membrana postsináptica, lo cual hará que la señal se propague. El nivel de dopamina en una sinapsis está regulado de manera que no se acumule esta molécula en la hendidura sináptica<sup>154</sup>. Para degradar la dopamina existen dos rutas. En la mayoría de las áreas cerebrales, la dopamina se inactiva mediante su recaptación vía el transportador de dopamina DAT1 seguido de su ruptura mediante la enzima monoamina oxidasa (MAOA y MAOB) para convertirse en ácido 3,4-dihidroxifenil. En la corteza cerebral en cambio, hay muy pocas proteínas transportadoras de dopamina por lo que esta molécula se inactiva mediante su recaptación vía el transportador de norepinefrina (NET) y, después, la Catechol-O-Metiltransferasa (COMT) la rompe enzimáticamente convirtiéndola en 3-metoxitiramina. La dopamina que no se destruye de este modo se vuelve a empaquetar en vesículas gracias al transportador vesicular monoamina (VMAT2) para su reutilización (ver Figura 2). Por este motivo los principales genes candidatos y los que más se han estudiado en este campo son *DRD4* y *DAT1* y, en menor medida, *COMT* y *MAOA*.

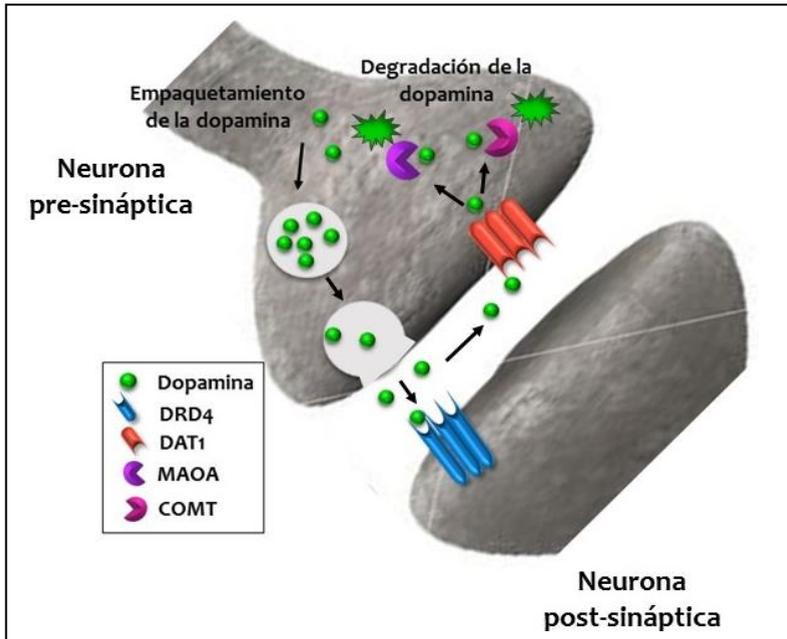


Figura 2 .Representación de la ruta de la dopamina desde una neurona presináptica a una postsináptica

#### 1.4.2 Bases genéticas del TDAH

La causa más estudiada para el TDAH siempre ha sido una disfunción en el sistema dopaminérgico y los genes candidatos para el origen del mismo también han girado en esa dirección:

##### 1.4.2.1 *DRD4* (11p15.5): receptor de la dopamina 4

Desde su descubrimiento, el gen del receptor de la dopamina 4 (*DRD4*) ha sido un gen de gran interés en los estudios de la conducta humana. El polimorfismo que más se ha estudiado y más se ha replicado de este gen es un VNTR (del inglés “*Variable Number of Tandem Repeats*” o número variable de repeticiones en tándem) de 48 pb situado en el exón III. Esta variante por lo tanto, se localiza en una región codificadora y en consecuencia afecta a la estructura proteica del receptor. Los alelos descritos hasta el momento contienen entre 2 y 11 repeticiones de este fragmento y, aunque éstos pueden variar en frecuencia dependiendo del grupo étnico, los alelos de 2, 4 y 7 repeticiones son los más frecuentes globalmente<sup>155</sup>. Estudios iniciales sugieren que el alelo de 7 repeticiones (7R) está asociado al TDAH<sup>156</sup>. Al parecer este alelo reduce la habilidad del

receptor para unirse a la dopamina, llegando a ser esta eficiencia incluso del 50% en comparación con la de los alelos de 2 y 4 repeticiones<sup>157</sup>. Existen numerosos estudios que analizan esta variante y los resultados entre ellos no siempre son consistentes (consultar el artículo<sup>155</sup> para una revisión de los estudios realizados). Algunos autores describen que los portadores del alelo 7R realizan mejor ciertas tareas cognitivas (la atención sostenida y la variabilidad en el tiempo de respuesta) que los no portadores<sup>158</sup>. Otros, en cambio, muestran que los portadores de 7R se caracterizan por un fenotipo extremo pero mantienen intacta la cognición<sup>159</sup>. En diversos meta-análisis<sup>160-162</sup> se ha demostrado una asociación significativa entre este VNTR y el TDAH, dándole importancia a la ruta de la dopamina en su etiología. No obstante, en la investigación de Todd *et al.*, realizada en gemelos<sup>163</sup>, no encuentran ninguna relación entre *DRD4* y el TDAH. Todos los estudios mencionados hasta ahora se enfocaron en el número de repeticiones del VNTR de *DRD4*, sin embargo se sabe que la secuencia repetitiva en sí muestra variación considerable en su composición de bases así como en su longitud<sup>8</sup>. A parte de este VNTR otros investigadores también han analizado una duplicación de 120 pb identificada en este mismo gen<sup>164-166</sup>.

#### 1.4.2.2 ***DAT1* o *SLC6A3* (5p15.3): transportador de la dopamina**

Desde la primera caracterización del gen del transportador de la dopamina (*DAT1* ó *SLC6A3*)<sup>167</sup>, su polimorfismo más estudiado ha sido un VNTR de 40 pb situado en la región 3' no traducida (3'UTR). Se han identificado alelos del VNTR con repeticiones entre 3 y 11, siendo los alelos más frecuentes los de 9 y 10 repeticiones (9R y 10R), aunque las frecuencias de éstos varían entre los diferentes grupos étnicos. En numerosas ocasiones se ha demostrado asociación entre este VNTR y el riesgo a padecer TDAH<sup>167,168</sup>, pero en otras no hay evidencias significativas de este resultado<sup>169-171</sup>. Dado que el polimorfismo no se encuentra en zona codificadora, y por lo tanto en principio no afecta a la secuencia proteica, si podría afectar a la regulación de la expresión de esta proteína en las regiones cerebrales<sup>154</sup>. También, se ha sugerido que el alelo 10R podría influir en la estabilidad del mRNA afectando el nivel de expresión del gen *DAT1*<sup>158</sup>, además de que la longitud de este tipo de polimorfismos podría contribuir a la regulación de la expresión génica<sup>150</sup>. Se ha demostrado que los portadores de 2 copias del alelo de 10R tiene una menor disponibilidad del transportador<sup>172</sup>. Sin embargo, otros estudios han expuesto lo contrario, y observan un mayor potencial efectivo de unión con este alelo en homocigosis<sup>173</sup>, lo cual es compatible con la hipótesis de un exceso de metabolización de la dopamina en el TDAH<sup>8</sup>.

Hay investigadores que han diseccionado el fenotipo de este trastorno y se han centrado en medidas directas de la atención. Por ejemplo, en las investigaciones de Bellgrove *et al.*<sup>174</sup> y de Loo *et al.*<sup>175</sup> se muestra que el alelo 10R del VNTR de *DAT1* confiere riesgo a padecer TDAH, y que los individuos portadores de esta variante en homocigosis poseen menor atención constante en comparación con los portadores del alelo 9R, lo cual se refleja en un mayor número de errores de comisión u omisión en las tareas. Este gen también mostraba relación con la variabilidad del tiempo de respuesta, indicando que los individuos homocigotos para 10R eran deficientes al tratar de mantener la atención durante las tareas.

En varios meta-análisis realizados en base a este gen se ha concluido que el alelo de 480 pb de longitud (10R) estaba asociado significativamente a TDAH<sup>160,162,176</sup>, aunque también se han identificado asociaciones con otros polimorfismos en este mismo gen<sup>177</sup>. En principio parece ser el candidato más fuerte para el trastorno, si bien es cierto que en los meta-análisis mencionados han encontrado muchas evidencias de heterogeneidad entre los estudios realizados con este gen, lo cual sugiere un efecto pequeño de esta variante.

Se insiste en la caracterización apropiada del transportador de la dopamina y su efecto en el TDAH porque el metilfenidato, el tratamiento más efectivo para este trastorno, actúa inhibiendo la recaptación de la dopamina (y la noradrenalina) vía *DAT1*.

#### **1.4.2.3 COMT (22q11.1-q11.2): catechol-O-metiltransferasa**

El gen *catechol-O-metiltransferasa (COMT)* codifica una enzima que participa en la degradación metabólica de la dopamina, norepinefrina y epinefrina, los cuales son neurotransmisores que se proponen como involucrados en la etiología del TDAH<sup>150</sup>.

El polimorfismo más estudiado del gen *COMT* es un SNP funcional (rs4680, G>A) que resulta en un cambio de valina por metionina (Val158Met) en la proteína. El genotipo de Met158 en homocigosis reduce en 3-4 veces la actividad de *COMT*<sup>178</sup>, de manera que cataboliza con menor eficiencia la dopamina. Por el contrario, el genotipo Val158/Val158 incrementa la degradación de la dopamina causando una reducción de su concentración en la sinapsis. Esta variante es la que se cree que podría estar implicada en los déficits del TDAH<sup>179</sup>.

En el estudio de Eisenber *et al.*<sup>180</sup> observaron asociación entre la variante de *COMT* de alta actividad enzimática y el subtipo impulsivo/hiperactivo del trastorno. Esta variante provocaría una menor cantidad de dopamina y norepinefrina en la sinapsis. En el estudio de DeYoung *et al.*<sup>181</sup> en cambio, es la variante *Met158* la que se asocia al trastorno. Sin embargo, otros estudios realizados no han encontrado asociación de este gen con dicho trastorno<sup>182-184</sup>, al igual que en dos meta-análisis realizados<sup>160,179</sup>, ni siquiera después de estratificar los resultados por género o por subtipo de TDAH.

#### 1.4.2.4 **DBH (9q34): dopamina beta-hidroxilasa**

La dopamina beta-hidroxilasa (DBH) es la responsable de la conversión de la dopamina en norepinefrina. Se han identificado varios polimorfismos en este gen que podrían influir en la actividad de su proteína. Uno de ellos es la variación del sitio de restricción *TaqI* dentro del intrón 5<sup>185</sup>. Daly *et al.*<sup>186</sup> fueron de los primeros en mostrar una transmisión preferencial del alelo A1 de este polimorfismo en los individuos con TDAH, resultados que confirmaron Roman *et al.*<sup>187</sup> unos años más tarde. Se ha descrito también que los niños homocigotos para el alelo A2 de esta variante hacen más errores de comisión en el test SART que los niños sin ese alelo, lo cual indica una atención sostenida más pobre<sup>188</sup>.

Tal y como ocurre con los resultados de los otros genes candidatos mencionados, los estudios de asociación del gen *DBH* no muestran consistencia entre ellos<sup>187,189-192</sup>. Otros polimorfismos que también se han estudiado en este mismo gen son el marcador -1021C>T (rs1611115) de la región flanqueada 5<sup>193,194</sup> o el SNP rs110850<sup>195</sup>.

#### 1.4.2.5 **MAOA (Xp11.4-p11.3): monoamina oxidasa A**

La monoamina oxidasa A (MAOA) es una flavoenzima que degrada la dopamina vía deaminación oxidativa, por lo que es un regulador clave para la función normal del cerebro, ya sea por exceso o defecto de actividad, y por lo tanto, su alteración podría ser responsable de trastornos neurológicos. El gen *MAOA* se ha visto asociado entre otros al autismo<sup>196</sup>, a una conducta agresiva<sup>197</sup>, al trastorno afectivo bipolar<sup>198</sup> o depresión<sup>199</sup>. Estudios clínicos muestran que los inhibidores de MAOB son efectivos en el tratamiento del TDAH, por lo que se sugiere que los genes *MAO* podrían estar involucrados en el origen del mismo. Estudios de ligamiento han encontrado asociación de un locus cercano a estos dos genes y este trastorno de atención<sup>200</sup> y,

posteriormente, se ha confirmado el ligamiento entre un VNTR en MAOA y el TDAH<sup>201</sup>. Desde entonces numerosos estudios se han centrado en esta enzima aunque no todos han logrado encontrar significancia entre los polimorfismos analizados y el TDAH<sup>192,201-203</sup>.

#### **1.4.2.6 Otros genes relacionados con el TDAH**

##### **SNAP-25 (20p11.2): proteína asociada al sinaptosoma**

Otros genes, como los serotoninérgicos y proteínas asociadas a sinaptosoma (SNAP-25) también se han expuesto como asociados al TDAH<sup>204-206</sup>. SNAP-25 es un gen funcional candidato porque se localiza en la parte citosólica de la membrana plasmática sináptica y junto con VAP/synaptobrevin y syntaxin forma el complejo proteico SNARE (del inglés "SNAP Soluble NSF Attachment Protein Receptor"), el cual forma parte de la maquinaria molecular requerida para la exocitosis de las vesículas sinápticas durante la neurotransmisión. Variantes en secuencia de SNAP-25 podrían influir en la susceptibilidad del TDAH puesto que podrían afectar a la liberación de la dopamina y otros neurotransmisores<sup>8</sup>.

##### **DRD5 (4p16.3): receptor de la dopamina 5**

Algunos proyectos han investigado la asociación del alelo de 148 pb de un microsatélite dinucleótido situado en el extremo 5' del gen *DRD5* con TDAH<sup>186</sup>, encontrando asociación en 4 meta-análisis<sup>160,176,207,208</sup>. Dicho polimorfismo no se localiza en la región codificadora y su función no es conocida, por lo que podría darse el caso de que éste se heredase junto con otra variante funcional o que influya en la función del gen de una manera aún desconocida<sup>136</sup>.

##### **SLC6A4, SERT, HTT ó 5-HTT (17p11.1-q12): transportador de la serotonina**

En la literatura se describe un polimorfismo de inserción/delección de 44 pb en la región promotora de *SLC6A4* que actúa en la eficiencia transcripcional del gen<sup>8</sup>. El alelo de longitud larga se relaciona con una mayor expresión del transportador<sup>209</sup>, así como con una mayor unión a la membrana y absorción de la serotonina<sup>210</sup>. Diferentes estudios<sup>211,212</sup> han encontrado una frecuencia mayor del alelo de longitud larga en los niños con TDAH y un meta-análisis confirma la asociación de este gen con el trastorno<sup>160</sup>.

### **5HT1B ó HTR1B (6q13): Receptor de la serotonina**

Se sugiere que *5HT1B* está relacionado con TDAH por ser el receptor de la serotonina y porque los ratones que carecen de este receptor muestran hiperactividad, pero este hallazgo sólo ha sido replicado parcialmente y sus resultados son ambiguos<sup>8</sup>. No obstante, el meta-análisis de Gizer et al.<sup>160</sup> muestra evidencias de asociación del gen con este trastorno.

Otros genes que se han sugerido o estudiado en relación con TDAH han sido los receptores de la dopamina 1, 2 y 3 (*DRD1*, *DRD2* y *DRD3*), pero los resultados son inconsistentes como en los casos anteriores<sup>150</sup>.

#### **1.4.2.7 Análisis de CNVs en el TDAH**

Elia et al.<sup>213</sup> observaron que los CNVs raros (poco frecuentes) de una muestra de TDAH estaban enriquecidos con genes conocidos por su importancia en funciones psicológicas y neurológicas como el aprendizaje, la conducta, la transmisión sináptica y el desarrollo del sistema nervioso. Diversos estudios se han centrado en el análisis de CNVs sugiriendo que genes como los receptores metabotrópicos de glutamato (*GRM5*, *GRM7*, *GRM8*, y *GRM1*), *PARK3*, *CHRNA7*, *AUTS2*, *CNTNAP*, *NRXN1* o *NPY* podrían tener relación con el TDAH<sup>214-219</sup>.

#### **1.4.3 Estudios de neuroimagen en el TDAH**

El TDAH, siendo un trastorno neuronal, refleja anomalías en el funcionamiento del sistema nervioso. Por ello, el cerebro de los pacientes con este trastorno se ha estudiado desde un punto de vista estructural y funcional<sup>220</sup>. A partir de los resultados de **imágenes estructurales cerebrales**, se han reportado anomalías en diversas áreas del cerebro como el córtex prefrontal, cerebellum, *striatum*, ganglios basales, *corpus callosum* y en el córtex parietal. En el estudio de Castellanos et al.<sup>220</sup> los pacientes con TDAH tenían un volumen cerebral menor en todas las regiones cerebrales, además de mostrar una reducción del volumen de la materia blanca en comparación con los controles. Sowell et al.<sup>221</sup> localizaron esta reducción del volumen cerebral principalmente en la zona inferior de la corteza prefrontal y la corteza anterior temporal. Mahone et al.<sup>222</sup> observaron un incremento de la materia gris y volúmenes reducidos durante el desarrollo del caudado en los pacientes con TDAH. Estas anomalías se han **relacionado en ocasiones con ciertos genotipos de**

**genes** considerados candidatos para el trastorno. Diferentes estudios han demostrado una asociación entre el *DRD4* y las anomalías en el volumen cerebral de los niños con TDAH. Por ejemplo, se ha descrito que los sujetos portadores del alelo de 7R del VNTR de 48 pb de *DRD4* tenían un volumen del córtex frontal superior y del córtex del cerebelo significativamente menor, en comparación con los sujetos sin este alelo<sup>223</sup>. Otros en cambio no encuentran ninguna relación entre este alelo y medidas anatómicas cerebrales<sup>224</sup>. El VNTR localizado en el gen *DAT1* también ha sido estudiado mediante MRI (del inglés “*Magnetic Resonance Imaging*”). Algunos indican que el alelo de 10R de *DAT1* podría mediar el funcionamiento neuropsicológico e incluso la función cerebral y la estructura cerebral en los niños con TDAH. En concreto, el grupo de Fernández-Jaén *et al.*<sup>225</sup> concluye que los pacientes con dos copias del alelo 10R exhibían significativamente una reducción en el grosor del córtex prefrontal lateral derecho.

**Estudios funcionales** indican que existen anomalías en los individuos con TDAH, particularmente en el córtex prefrontal, el *striatum* y en el córtex parietal<sup>226</sup>. El gen *DRD4* ha sido el más estudiado en cuanto a su funcionalidad neurológica, y diferentes alelos del VNTR de este gen se han relacionado con la actividad funcional cerebral durante la realización de distintas tareas cognitivas. De esta forma, el alelo 5R se ha asociado a la actividad neuronal durante tareas atencionales y la capacidad para la creatividad<sup>227</sup>, el alelo 4R se relaciona con una mayor actividad en el lóbulo temporal derecho durante el procesamiento de estímulos emocionales negativos<sup>228</sup> y el 7R con un incremento de la activación en la mitad izquierda e inferior del giro frontal en la tarea de “respuesta incompatible frente a estímulos”<sup>229</sup>, una mayor activación cerebelar en la tarea de discriminación temporal<sup>229</sup> y una menor activación en varias áreas cerebrales durante las tareas de control inhibitorio (“Go-No-Go”)<sup>230</sup>. Cambios en la activación neuronal relativos a esta última función cognitiva también se han relacionado con genotipos concretos de *DAT1* y *COMT*<sup>231</sup>. Además, hay indicaciones de que *DRD4* interacciona epistáticamente con el gen *COMT* durante la respuesta inhibitoria<sup>232</sup>. También se ha reportado que la alteración de la conectividad de la materia blanca se basa en el genotipo de *COMT*, ya que una ruta que conecta 18 regiones cerebrales (conexiones de materia blanca) estaba significativamente debilitada en los jóvenes con TDAH que portaban el alelo *COMT*-Met158 en comparación con los portadores homocigotos de Val158<sup>233</sup>.

## 1.5 COMORBIDAD ENTRE LA DISLEXIA Y EL TDAH

Tanto la dislexia como el TDAH frecuentemente aparecen acompañados de otros trastornos psiquiátricos o dificultades del aprendizaje. La dislexia, por ejemplo, suele co-ocurrir con SLI (del inglés, *Specific Language Impairment*) o TEL (Trastorno Específico del Lenguaje). Existen evidencias que indican que los problemas del desarrollo en el lenguaje oral y las matemáticas pueden co-ocurrir en los individuos con dislexia. Algunas investigaciones indican que la prevalencia de la discalculia (falta de habilidad para las matemáticas) es mayor en los niños con dislexia. Además se han encontrado correlaciones genéticas entre el lenguaje, las matemáticas y los rasgos de lectura. Según Oliver SP Lavis *et al.*<sup>234</sup> existe un componente genético sustancial en la habilidad de la lectura y de las matemáticas de los niños, y se estima que alrededor de la mitad de la correlación observada entre estos rasgos se deben a efectos genéticos compartidos, lo que indicaría que si una gran proporción de los factores genéticos que afectan estos rasgos son pleiotrópicos, las diferencias individuales en cuanto a estas habilidades vendrían determinadas relativamente por el entorno. Marino *et al.*<sup>235</sup> muestran evidencias significativas entre variantes de *DCDC2* y los “hechos numéricos”, y entre variantes de *DYX1C1* y el “cálculo mental”. Sin embargo ninguno de estos dos genes fue relacionado con fenotipos del lenguaje. En una muestra de 200 niños con discapacidad en el deletreo y en otra población de disléxicos se identificó un marcador situado en el gen *MYO18B* asociado a la habilidad para las matemáticas<sup>236</sup>, sin embargo este mismo marcador no se consiguió replicar en otro estudio subsiguiente<sup>237</sup>.

Asimismo, el TDAH suele ir acompañado del trastorno de oposición desafiante (ODD, del inglés *Oppositional Defiant Disorder*), trastorno de conducta (CD) o autismo. Esta comorbidad es común, ya que alrededor de dos tercios de los niños con TDAH tienen el diagnóstico de otro trastorno psiquiátrico<sup>136</sup>. Establecer las bases genéticas de un trastorno particular bajo este patrón es complicado ya que es imposible discernir a qué síntoma particular se asocia el componente genético. Por ejemplo, el genotipo Val158/Val158 de *COMT* está asociado con los síntomas del trastorno conductual o con una conducta antisocial en los pacientes que tienen TDAH<sup>238,239</sup>, lo cual indicaría que algunas variantes genéticas operan modificando el fenotipo del trastorno, en vez de estar asociadas al trastorno en sí.

Es esencial mencionar que **la dislexia y el TDAH también aparecen juntos con mayor frecuencia de lo esperado**<sup>240</sup>, hipotetizando que podría haber factores genéticos comunes entre ambos. Algunos estudios han demostrado una misma señal de asociación (ligamiento o *linkage*) entre las habilidades para la lectura y los síntomas de inatención o los de hiperactividad/impulsividad<sup>241</sup>, mientras otros predicen una relación más fuerte específicamente entre las dificultades lectoras y los síntomas de falta de atención<sup>242-245</sup>. Los estudios de gemelos y de familias sugieren que este solapamiento es, en parte, debido a un fondo genético compartido<sup>246</sup>. Es más, varios grupos han demostrado que existen variantes de regiones genómicas coincidentes entre estos síndromes. Se ha relacionado el TDAH con la región cromosómica 6p22<sup>241</sup>, la misma región donde se sitúan los genes *DCDC2* y *KIAA0319*, lo que refuerza la posibilidad de la existencia de un gen pleiotrópico en esta zona. También se han relacionado con la dislexia las regiones cromosómicas 4q13.1, 7q36.3 y 16p12.1, regiones previamente implicadas en el TDAH<sup>247</sup>, y el gen *ADRA2A* con la situación comórbida entre ambos trastornos<sup>248</sup>.

La búsqueda de genes de susceptibilidad específicos para la dislexia y el TDAH está siendo complicada no sólo por su complejidad y heterogeneidad, sino también por su co-ocurrencia y la de otros trastornos cognitivos. Esta situación hace todavía más difícil la obtención de resultados fructuosos, particularmente si uno de los trastornos no está diagnosticado. Si se ignora la comorbidad, uno puede concluir erróneamente que alguna variante está asociada a una alteración cuando en realidad está asociada a la condición comórbida<sup>249</sup>. Por lo tanto, por las razones mencionadas, es de vital importancia establecer correctamente los principios diagnósticos de la dislexia y del TDAH con el fin de identificar asociaciones interpretables y reproducibles<sup>3</sup>. Es por ello que la caracterización de genes candidatos para la dislexia en pacientes con TDAH (y viceversa) es una estrategia fundamental.

## **1.6 IMPORTANCIA DE LA DEFINICIÓN DEL FENOTIPO EN LOS ESTUDIOS GENÉTICOS DE ASOCIACIÓN EN DISLEXIA Y TDAH**

Si bien se han encontrado variantes genéticas asociadas a dislexia y TDAH, la consistencia entre los resultados de los distintos estudios es muy escasa. Esta falta de reproducibilidad entre estudios no es una sorpresa, puesto que este tipo de trastornos parece estar relacionado con diversos factores de riesgo, los cuales podrían interactuar entre ellos provocando

manifestaciones múltiples de los síntomas. Por otra parte, diferentes factores de riesgo podrían terminar también produciendo el mismo fenotipo<sup>1</sup>. Bajo este escenario tan heterogéneo, no es fácil establecer claramente unos criterios diagnósticos, y ésta es una de las razones por las cuales está siendo tan difícil la búsqueda de genes individuales de gran efecto en los estudios de asociación (especialmente cuando el trastorno es de carácter complejo).

Desafortunadamente, no existen criterios universales ni pruebas psicométricas generalizados para el diagnóstico de la **dislexia**. Por esta razón cada grupo de investigación utiliza sus propios criterios y establece sus límites de inclusión-exclusión, generalmente dependiendo del idioma y la población estudiada. Algunos investigadores aplican como pauta para el diagnóstico la discrepancia entre la habilidad para la lectura observada y la esperada en base a la inteligencia del individuo. Otros en cambio, lo implementan según la edad cronológica del niño. Obviamente, la falta de medidas uniformes influye directamente en la categorización de muestras y, por supuesto, dificulta la comparación entre estudios de diferentes poblaciones<sup>10</sup>. Por consiguiente, la elección de factores decisivos podría ser un punto crucial. En muchas ocasiones, los diagnósticos de tipo todo-o-nada podrían no ser óptimos para los estudios genéticos ya que este tipo de caracterización no captura la esencia completa del fenotipo<sup>7</sup>. Este tipo de trastornos neurales podrían surgir por varios procesos cognitivos, y consecuentemente se ha sugerido que el identificar genes susceptibles para fenotipos intermedios o alternativos sería una estrategia más fructífera<sup>8,250,251</sup>. Es decir, usar directamente los valores registrados en las pruebas psicométricas para establecer una relación entre los factores moleculares y el trastorno, ya que estos rasgos cuantitativos se vinculan directamente con las dificultades definidas. Estas medidas pueden considerarse **endofenotipos** o índices cuantitativos más fáciles de relacionar con el componente genético<sup>250</sup>. Por las razones mencionadas, existen problemas a la hora de conceptualizar la dislexia como un rasgo dicotómico (caso/control) y se propone que simplemente podría representar el extremo bajo de la cola de la variabilidad normal de la habilidad para la lectura<sup>252</sup>.

En lo que concierne al **TDAH**, el uso de los criterios diagnósticos aceptados internacionalmente ha incrementado la fiabilidad del diagnóstico, resultando incluso en una de las caracterizaciones clínicas mejor validadas. Existen dos tipos de manuales que describen los criterios para definir el TDAH: “*Diagnostic and Statistical Manual of Mental Disorders*” quinta

edición (DSM-V)<sup>253</sup> e “*International Classification of Mental and behabioural Disorders*” décima edición (ICD-10)<sup>254</sup>. En ambos se establece el fenotipo en base a los síntomas identificados en los pacientes<sup>136</sup>, en ocasiones por los síntomas descritos por los padres, en otras por los referidos por los profesores y a veces combinando ambos, la cual parece ser la opción que mejores diagnósticos produce<sup>148</sup>. Aunque la mayoría de los estudios genéticos han utilizado criterios estandarizados para su diagnóstico, algunos también consideran que este trastorno está distribuido de una manera continua en la población<sup>136</sup>. Es más, hace ya mucho tiempo que un estudio de gemelos de gran escala mostró que el TDAH se define mejor como el extremo final del continuo que como un fenotipo categórico<sup>255</sup>. Por lo tanto, en vez de buscar diferencias genéticas entre casos y controles, los grupos extremos de este patrón de continuidad en la distribución, el análisis directo de los índices continuos de severidad en la población general podría ayudar a encontrar genes relacionados con los procesos cognitivos específicos influyentes en cada uno de los trastornos<sup>256,257</sup>. Por esta razón, los investigadores empiezan a extender sus análisis a dichas poblaciones incrementando así el poder estadístico, además de facilitar la detección de factores genéticos con tamaño de efecto individual pequeño<sup>18</sup>.

Un factor muy importante a tener en cuenta es que estamos hablando de trastornos del desarrollo y por lo tanto, los déficits perceptuales que se ven en las primeras fases del mismo podrían resolverse, pero a su vez podría estar afectado el subsiguiente progreso de las funciones cognitivas<sup>2</sup>. La constelación de síntomas que define un trastorno cognitivo podría cambiar a medida que el individuo madura, algunos síntomas podrían mantenerse estables y otros en cambio podrían empeorar o mejorar, por lo que las asociaciones encontradas en un principio en relación a estos síntomas (incluyendo relaciones causales) podrían no ser detectables en años futuros. Como ejemplos se pueden plantear los déficits del procesamiento fonológico y auditivo (este último propuesto en muchas ocasiones como causa del primero), dos síntomas que han demostrado tener muchos **disléxicos**. Otros individuos que padecen dislexia, en cambio, podrían no presentar el déficit auditivo, ya que con la edad este síntoma se podría haber corregido siendo imposible de detectar en el momento de diagnosticar la dislexia. Sin embargo, el déficit auditivo presente en los primeros años de vida, podría haber afectado al procesamiento fonológico, el cual persiste en edad más avanzada<sup>21</sup>. Por otra parte, el perfil sintomatológico y neurofisiológico del **TDAH** cambia al entrar en la adolescencia, los síntomas de inatención permanecen constantes hasta los 9-12 años mientras que los síntomas de hiperactividad/impulsividad si bien no llegan a

normalizarse suelen decrecer<sup>226,258</sup>. También se sugiere que ciertas anomalías en la estructura cerebral asociadas a este trastorno son más prominentes en niños, mientras que otras lo son en adultos<sup>259</sup>. Muchos de los estudios genéticos de la dislexia y del TDAH han adoptado una perspectiva fija de las habilidades cognitivas y conductuales sin tener en cuenta que podría existir un “cambio” en términos de la definición del fenotipo. Para una mejor comprensión de las causas que afectan a este tipo de trastornos, sería imprescindible el considerar las trayectorias del desarrollo que empiezan en la infancia de los niños que están siendo estudiados<sup>18,222,259</sup>.

## 1.7 INTERACCIÓN ENTRE LOS GENES Y EL AMBIENTE

Los estudios realizados con gemelos han sido cruciales a la hora de establecer ciertas atribuciones genéticas. En este tipo de abordaje se efectúa la comparativa de los ratios de concordancia entre gemelos monocigóticos y dicigóticos para **separar los factores genéticos de la contribución del medio ambiente**<sup>10</sup> y concluyen en ambos casos, dislexia y TDAH, que el factor genético interviene significativamente en el trastorno, pero no completamente, lo cual hace probable la hipótesis de una interacción entre los genes y el entorno (GxE, del inglés *Gene versus Environment*). Esta interacción se refiere al hecho de que el ambiente puede modificar la expresión del fondo genético de un individuo, reforzando o debilitando el efecto de los genes en el fenotipo<sup>260</sup>. Según la literatura, la dislexia presenta un tipo de **interacción “bioecológica”**, es decir, el trastorno es más heredable bajo un entorno favorable. Es más, cuanto mayor sea el nivel educativo en una población más probable será que las diferencias individuales en cuanto a la habilidad para la lectura sean atribuibles a las diferencias genéticas entre los niños<sup>3</sup>. Así lo demuestra el estudio de Friend *et al.* en el cual concluyen que la heredabilidad de la dislexia entre gemelos depende del nivel educativo de los padres, donde el efecto de los genes es particularmente evidente en los niños que fracasan en el aprendizaje de la lectura aun teniendo un soporte ambiental óptimo<sup>261</sup>, y el estudio que describe que la calidad del profesorado modera el efecto genético en cuanto al logro en la lectura<sup>262</sup>. Esto es, si los profesores son excelentes, la variabilidad en cuanto al aprendizaje de la lectura parece estar atribuida enormemente a los genes, sin embargo, una enseñanza de peor calidad impide a los niños llegar hasta su mayor potencial.

El TDAH en cambio, sigue una **interacción de “diathesis-stress”** dónde es más heredable bajo un ambiente desfavorable o de “riesgo”, de forma que la vulnerabilidad genética sumada al estrés medioambiental aumentará las probabilidades del trastorno<sup>263</sup>. Ésta es una de las pocas condiciones psiquiátricas donde se han implicado factores ambientales específicos con relativa confianza, entre ellos el peso bajo al nacer, nacimiento prematuro, estrés maternal, adversidad psicológica o el consumo de tabaco maternal prenatal. Polimorfismos en los genes *DAT1*, *DRD4* y *CHRNA4* en interacción con la exposición al tabaco durante el embarazo incrementan el riesgo para el TDAH del tipo combinado <sup>264–266</sup>. También se ha detectado una interacción entre *DAT1* y el consumo de alcohol durante el embarazo<sup>267</sup>. La característica común de muchos de estos factores ambientales de riesgo es su exposición durante un periodo crítico (el embarazo principalmente) por lo que se sugiere que probablemente estén influidos por cambios epigenéticos. La epigenética es un sistema de regulación que controla la expresión de los genes pero sin alterar la secuencia de DNA, y se ha propuesto como mediadora de los efectos del riesgo ambiental en los trastornos psiquiátricos<sup>268</sup>. También se han encontrado interacciones con factores psicológicos, como la interacción de *DRD4* con la auto-culpa de un niño frente a conflictos maritales, lo cual incrementa la susceptibilidad hacia la inatención<sup>269</sup>. Se sugiere que los factores de riesgo bioambientales podrían influir en la manifestación de los síntomas de hiperactividad/impulsividad, mientras que los factores de riesgo psicológicos tendrían una mayor influencia en la dimensión de los síntomas de inatención<sup>270</sup>.

En el estudio de gemelos realizado por el grupo de Rosenberg *et al.*<sup>137</sup> replican la interacción GxE contraria descrita para la dificultad lectora (bioecológica) y para el subtipo inatento de TDAH (“*diathesis-stress*”) con respecto a la educación parental en la misma muestra de participantes. Los autores hipotetizan que sólo los genes específicos para cada trastorno, y no los genes compartidos en la comorbidad, contribuyen en estas interacciones, aunque esta hipótesis se quedó sin confirmar en los análisis realizados en su estudio.

Finalmente, es preciso mencionar que no todos los individuos expuestos a los factores de riesgo ambientales desarrollan el trastorno en cuestión, es decir, los individuos difieren en su respuesta hacia los factores ambientales específicos. De esta misma manera, en la mayoría de los trastornos complejos, no todos los individuos que portan las variantes genéticas de

susceptibilidad desarrollan la enfermedad a no ser que estén expuestos a factores ambientales de riesgo<sup>8</sup>.

## 1.8 ESTUDIOS DE ASOCIACIÓN DE GENOMA COMPLETO

Mediante los estudios de asociación de genes candidatos se han identificado algunos genes que podrían explicar parte de la variabilidad genética que influye en la dislexia y el TDAH pero, como ya se ha mencionado anteriormente, los resultados entre los diferentes estudios no son consistentes y todavía falta explicar gran parte de la variabilidad genética. Por ello, la alternativa de aplicar en estas circunstancias estudios de asociación de genoma completo o GWAS, basados en el genotipado de un set denso de polimorfismos a lo largo de todo el genoma, es una buena opción para identificar nuevas variantes que puedan estar implicadas en los trastornos que nos ocupan<sup>271,272</sup>.

El genoma humano posee una estructura haplotípica, lo cual permite predecir con una alta probabilidad la variación genética de un *locus* frente a otro adyacente<sup>272</sup>. Gracias al proyecto internacional Hapmap<sup>273</sup> se han determinado los patrones de LD a escala global, y este avance ha sido crucial para que los estudios de genoma completo se puedan realizar de una forma eficiente. El éxito en la búsqueda de variantes asociadas a una enfermedad reside en la densidad de marcadores utilizada, es decir, en la proximidad en la que se encuentran los polimorfismos analizados<sup>274</sup>, ya que cuanto más cerca estén dos marcadores, siempre y cuando estos marcadores no estén en LD, mayor será la posibilidad de encontrar la variante asociada. Un millón de SNPs seleccionados aleatoriamente así como unos cuantos cientos de miles seleccionados óptimamente mediante LD deberían ofrecer una cobertura completa del genoma<sup>275</sup>.

Hoy en día existen técnicas de alto-rendimiento que permiten analizar una enorme cantidad de marcadores genéticos en un gran número de muestras de una forma rápida y relativamente barata, plataformas de genotipado que son muy eficientes para la caracterización de SNPs. Sin embargo para el genotipado de variantes estructurales como inserciones, deleciones, inversiones o CNVs, estas mismas plataformas no son del todo óptimas.

Un punto a tener en cuenta es que los GWAS son una aproximación indirecta, es decir, en muchos casos el SNP asociado no es necesariamente el marcador de riesgo funcional,

simplemente puede estar en LD con una región que contiene un gen funcional significativo<sup>3</sup>. También posee limitaciones como un enorme potencial para generar falsos positivos y, en la mayoría de los casos, una falta de poder estadístico para detectar el efecto de variantes raras con frecuencia menor al 1%<sup>275</sup>.

A pesar de estas desventajas, se han realizado múltiples estudios de genoma completo en relación con la **dislexia** o con alteraciones en la **lectura**, en base a los cuales se han sugerido variantes asociadas en diferentes regiones cromosómicas. Los primeros hallazgos vinieron a cargo de **estudios de ligamiento**, en concreto:

- El grupo de Fisher<sup>94</sup> analizó el genoma completo en familias de U.S.A. y U.K.. Aparte de algunos *loci* relacionados con medidas de lectura, ortografía y fonología, los resultados más remarcables fueron la relación de la región 6p21 con la decodificación fonémica en la población de U.K., así como la región 18p21 vinculada a medidas de lectura en ambas poblaciones.
- El *locus* 6p21 también se ha identificado en otro estudio<sup>276</sup> pero no siempre se consigue asociar con rasgos de lectura<sup>277</sup>.
- En estudios posteriores también se han relacionado los *loci* 2p11 (muy cercano al *locus* DYX3), 2q, 2p, 4q13.1, 7q31-32, 7q36.1-q36.2, 7q36.3, 10q, 13q 16p12.1, 17q22 y Xq37.3 con el fenotipo disléxico, con medidas cuantitativas de lectura de palabras o con el nombrado rápido de letras<sup>104,247,277-279</sup>.
- En el cromosoma X, pero en un *locus* diferente al mencionado, en Xq27.3, sugieren el gen FRAXA como posible candidato para la dislexia en familias francesas y refuerzan la implicación del *locus* DYX9 en este trastorno<sup>105</sup>.

Los estudios de asociación de SNPs han sido más numerosos y con resultados prometedores (ver Tabla 2 para un resumen de los estudios):

- En uno de los primeros estudios han asociado SNPs a la habilidad para la lectura situados en los genes *CSE1L*, *SOX5*, *IGSF22*, *CDC42BPA*, *NKAIN2*, *DPF3* y *TIAM1*. Sin embargo, ninguno de estos SNPs aportaba más de un 0,5% de la varianza de este rasgo<sup>280</sup>. Además, ninguno de los 10 marcadores más significativos de este estudio fue replicado en muestras australianas, por lo que sugieren que podrían ser errores de tipo I del estudio original<sup>281</sup>.

- En el estudio realizado por el grupo de Scerri *et al.*<sup>282</sup> en el cual se evalúa la habilidad relativa de la mano en individuos con dislexia, la señal de asociación más significativa la dan dos SNPs, rs9806256 y rs11855415, ambos situados en el gen *PCSK6*, llegando este último incluso a pasar el umbral de significancia establecido para su GWAS ( $p=2E-08$ ). En un meta-análisis posterior se replicó la asociación de este gen pero en otro SNP, rs7182874<sup>283</sup>.
- Dos variantes situadas en el gen *CLSTN2* en 3q23 (rs1365152 y rs2114167) también se han asociado en un GWAS cuyo objetivo era buscar relaciones con medidas de *MistMatch Negativity* (MMN), pero no consiguieron replicar esta asociación en una segunda muestra independiente. Sin embargo, otro de los marcadores, rs4234898 (4q32.1), obtuvo un valor de  $p$  nominal en ambas fases e incluso llegó al umbral de significancia para GWAS en la muestra combinada ( $p=5.14E08$ )<sup>284</sup>.
- En la investigación de Luciano *et al.* sugieren que los genes *ABCC13* y *DAZAP1* son posibles implicados en la habilidad para la lectura, ya que algunas variantes de estos genes muestran asociación con medidas de lectura, de deletreo y de repetición de pseudopalabras<sup>285</sup>. En un estudio reciente de replicación<sup>286</sup>, analizan 4 de los SNPs más significativos de este estudio (rs7187223, rs4839516, rs764255 y rs2192161) en muestras holandesas y asocian nominalmente el marcador rs7187223, el cual se encuentra situado cerca del gen *CDH13*, con medidas de RAN y conciencia fonológica.
- La comorbidad de la dislexia con otros trastornos neurológicos también se ha analizado en estudios de genoma completo. Eicher *et al.*<sup>287</sup> intentan buscar marcadores comunes para la comorbidad entre la dislexia y el LI (del inglés *Language Impairment*), y sugieren los genes *ZNF385D* y *COL4A2* como posibles candidatos. Es más, el estudio de replicación realizado por Carrión-Castillo *et al.*<sup>286</sup> confirma la asociación de un marcador situado en el gen *ZNF385D* con medidas de lectura en la población holandesa. Por otra parte, los marcadores que obtuvieron la mayor asociación entre las muestras disléxicas (sin comorbidad) del estudio de Eicher *et al.*<sup>287</sup> se situaban en los genes *OPA3*, *KIAA0196*, *NUAK1*, *PLEKHA1* y *FAM19A1*.
- En un meta-análisis realizado en base a tres GWAS sobre la lectura y el lenguaje, identifican nuevas asociaciones de los SNPs rs59197085 y rs5995177 localizados en los genes *CCDC136/FLNC* y *RBFox2*, respectivamente. Además, el locus *CCDC136/FLNC* también

mostró asociación con estas medidas en la población general<sup>288</sup>. Sin embargo, en un posterior estudio<sup>286</sup> no replican la asociación del marcador situado en el gen *RBFOX2*.

Tabla 2.-Descripción de los estudios de genoma completo realizados para la dislexia. QTL=Quantitative Trait Loci.

Referencia	Tipo de estudio	Población	Genes con mayor asociación
Meaburn et al. 2007 <sup>280</sup>	QTL	Reino Unido	<i>CSE1L</i> , <i>SOX5</i> , <i>IGSF22</i> , <i>CDC42BPA</i> , <i>NKAIN2</i> , <i>DPF3</i> y <i>TIAM1</i>
Roeske et al. 2011 <sup>284</sup>	QTL	Alemania	<i>CLSTN2</i>
Scerri et al. 2011 <sup>282</sup>	QTL	Reino Unido	<i>PCSK6</i>
Eicher et al. 2013 <sup>287</sup>	Caso/ Control	Reino Unido	<i>OPA3</i> , <i>KIAA0196</i> , <i>NUAK1</i> , <i>PLEKHA1</i> y <i>FAM19A1</i>
Luciano et al. 2013 <sup>285</sup>	Meta-análisis	Reino Unido/ Australia	<i>ABCC13</i> y <i>DAZAP1</i>
Gialluisi et al. 2014 <sup>288</sup>	Meta-análisis	Reino Unido/ Colorado	<i>CCDC136/FLNC</i> y <i>RBFOX2</i>

Las primeras investigaciones de genoma completo en el **TDAH**, se realizaron mediante estudios de ligamiento. De esta forma se identificaron regiones en 5p12, 10q26, 12q23 y 16p13<sup>144</sup>. La mayoría de las veces sin embargo, este tipo de estudios han fallado a la hora de identificar y replicar factores genéticos consistentes. Los estudios de asociación de SNPs han conseguido identificar algunos genes que podrían estar vinculados a la causa del trastorno, aunque en la mayoría de los estudios falta la replicación de los hallazgos en muestras independientes (ver Tabla 3 para detalles de los estudios):

- En el primer GWAS realizado en familias con TDAH, ninguno de los marcadores alcanzó el umbral de significancia establecido para GWAS ( $p < E-08$ ), aunque los SNPs más significativos estaban ubicados en los genes *DCLK1* (organizador del citoesqueleto), *SPCK3* (componente de la matriz extracelular), *CDH13* (proteína de adhesión célula-célula) y dos reguladores de canales de potasio (*KCNIP1* y *KCNIP4*)<sup>289</sup>. Dos años más tarde, el estudio de Lantieri et al.<sup>290</sup> intentó replicar los marcadores de mayor asociación de este estudio y encontró significancia en dos marcadores, rs2939678 y rs12505502, situados en los genes *XKR4* y *FAM190A*, respectivamente.
- Utilizando las mismas muestras que en el estudio de Neale<sup>289</sup>, Lasky-Su implementó un análisis de rasgos cuantitativos de los síntomas de TDAH, e identificó dos marcadores con significancia localizados en los genes *CDH13* (rs6565113) y *GFO1* (rs552655)<sup>291</sup>.

- Es preciso remarcar que un tercer estudio<sup>292</sup> también encontró una tendencia de asociación para el gen *CDH13*, a parte de otros genes encargados de la adhesión celular (*ASTN2*) y reguladores de la plasticidad sináptica (*CTNNA2* y *KALRN*).
- La mayoría de los GWAS realizados posteriormente no han llegado a obtener resultados de asociación significativos<sup>293-299</sup>, aunque en base a sus mejores asociaciones han propuesto varios genes candidatos como son *SLC9A1*, *GRM5*, *CHRNA7* o *GPR139* entre otros. Sin embargo, un reciente GWAS en el que utilizan la misma muestra del primer estudio de Yang *et al.*<sup>298</sup>, pero centrándose esta vez en medidas endofenotípicas, este mismo grupo, ha identificado un SNP (rs11514810) relacionado con la inhibición ejecutiva de los pacientes con TDAH. Este marcador afecta en la expresión del gen *MICALL2* en el cerebro<sup>300</sup>.
- En el meta-análisis<sup>301</sup> basado en los resultados de 4 estudios no se encontró ningún marcador significativamente asociado, pero entre los genes cuyos marcadores habían conseguido un valor de p menor se identificaron *SH3BP5*, *AGBL1*, *NUCB1* y *FURIN*. Sin embargo, en un meta-análisis<sup>302</sup> más reciente en base a datos de 17666 individuos (9 cohortes de Europa, Australia y U.S.A.), encuentran asociación en 3 genes, *LMOD2*, *WASL* y *ASB15*.
- Teniendo en cuenta la heterogeneidad en el diagnóstico de las muestras y la diversidad genética entre distintas poblaciones, dos estudios recientes de gran interés para el presente proyecto son los realizados en población Española, considerando muestras adultas. A pesar de que ninguno de los dos obtiene resultados significativos, el primero señala al gen *SORCS2*<sup>303</sup> como asociado con TDAH y el segundo al gen *FBXO33*, además de sugerir la ubiquitinación (un tipo de modificación post-traduccional de proteínas) como nuevo posible mecanismo causante de este trastorno<sup>304</sup>. Asimismo es de gran importancia un tercer estudio también realizado en sujetos españoles<sup>305</sup> el cual sugiere el gen *PID1* como candidato para el buen funcionamiento atencional de los niños.

Tabla 3.- Descripción de los estudios de genoma completo realizados para el TDAH. QTL=Quantitative Trait Loci.

Referencia	Tipo de estudio	Población	Genes con mayor asociación
Lesch et al.2008 <sup>292</sup>	Caso/control	Alemania	CDH13,ASTN2, CTNNA2 y KALRN
Neale et al. 2008 <sup>289</sup>	Familias	Caucásica	DCLK1,SPCK3, CDH1, KCNIP1 y KCNIP4
Lasky -Su et al.2008 <sup>291</sup>	QTL	Caucásica	CDH13 y GFOD1
Neale et al. 2010 <sup>293</sup>	Caso/control	Caucásica	PRKG1, PPM1H, TCEG1L y FLNC
Mick et al. 2010 <sup>294</sup>	Familias	U.S.A.	EMP2, ELOVL6, UGT1A9, C21orf34, BMPR1B,CCDC46 y SLC9A9
Neale et al. 2010 <sup>301</sup>	Meta-análisis	U.S.A. y Europa	SH3BP5, AGBL1, NUCB1 y FURIN
Hinney et al. 2011 <sup>306</sup>	Caso/control y familias	Alemania	GPC5, BC11A y KIAA1267
Stergiakouli et al.2012 <sup>296</sup>	Caso/control	Caucásica	IL20RA, HOXB1, TRIO, MAGI2 y ATXN2
Ebejer et al. 2013 <sup>297</sup>	QTL	Australia	CPLX2, GPR139, DNMT3B y MAPRE1
Yang et al. 2013 <sup>298</sup>	Caso/control	China	KIAA1486 y ARSB
Alemani et al.2015 <sup>303</sup>	QTL	España	SORCS2
Sánchez-Mora et al. 2015 <sup>304</sup>	Caso/control	España	FBXO33
Zayats et al. 2015 <sup>299</sup>	Caso/control	Noruega	TRIM36 y ZBTB16
Middeldorp et al.2016 <sup>302</sup>	Meta-análisis	Europa, U.S.A. y Australia	LMOD2, WASL y ASB15
Alemaný et al. 2016 <sup>305</sup>	QTL	España	PID1
Yang et al.2017 <sup>300</sup>	QTL	China	MICALL2

Como resumen cabe remarcar que los resultados entre los diferentes GWAS no se solapan y, además, que ninguno de los genes candidatos “clásicos”, tanto de dislexia como de TDAH, aparece entre los marcadores más asociados de estos estudios.

## **2 OBJETIVOS**



**OBJETIVO GENERAL:**

La lectura y la atención son dos habilidades cognitivas básicas para el aprendizaje. Defectos en estos rasgos cognitivos podrían desencadenar trastornos como la dislexia y la falta de atención, dos discapacidades de alta prevalencia en España. Si bien se consideran trastornos complejos con una base genética, las variantes biológicas identificadas en los numerosos estudios realizados hasta el momento no son consistentes ni se replican en diferentes países. Por ello, el **objetivo general** de esta tesis doctoral es **identificar variantes genéticas** asociadas a la lectura y la atención en una población de niños españoles, haciendo especial hincapié en la **dislexia, la falta de atención y la comorbidad entre ambos trastornos**, con el propósito de hipotetizar rutas funcionales causales que den luz sobre posibles métodos para diagnóstico temprano de estos trastornos cognitivos.

**OBJETIVOS ESPECÍFICOS:****Objetivo Técnico:**

1.-Verificación de la adecuación del DNA proveniente de saliva para estudios de caracterización de genoma completo. Teniendo en cuenta la distinta proporción de células humanas y bacterias que hay en la saliva, es un requisito el **ajuste de los protocolos estándar de genotipado** para asegurar su aplicación al DNA extraído de este tipo de muestra.

**Objetivos Científicos:**

1.-**Validar** en la cohorte conseguida las variantes seleccionadas de tipo **SNP y VNTR** localizadas en genes propuestos como **candidatos a la susceptibilidad a padecer dislexia y/o TDAH** en estudios previos.

2.-**Identificación** de nuevos **SNPs, grupos de SNPs o regiones de número de copia variable** asociados por un lado con **la dislexia, la falta de atención y/o la comorbidad** de ambas, y por otro lado **con funciones cognitivas de lectura y atencionales** en la cohorte española estudiada, a través de distintos abordajes que nos permitan la **descripción de rutas funcionales causales de estas afecciones**.



### **3 MATERIALES y MÉTODOS**



### 3.1 ESTRATEGIAS DE ASOCIACIÓN UTILIZADAS EN ESTE PROYECTO

En esta tesis se han llevado a cabo diversos estudios de asociación en base a dos grandes estrategias: **genes candidatos** y **genoma completo (GWAS)**. Asimismo, ambas estrategias a su vez, se han subdividido en dos metodologías de análisis: **casos/controles** (para la dislexia, la falta de atención y la comorbidad entre ambos trastornos) y análisis de **QTLs** (*Quantitative Trait Loci*) en la población general, es decir, independientemente del fenotipo de la muestra. Finalmente partiendo de los resultados obtenidos en el estudio de genoma completo también se ha querido hipotetizar sobre posibles **rutas funcionales** para cada uno de los trastornos así como para las funciones cognitivas analizadas. En la Figura 3 se muestra un esquema general de las estrategias utilizadas en este proyecto.

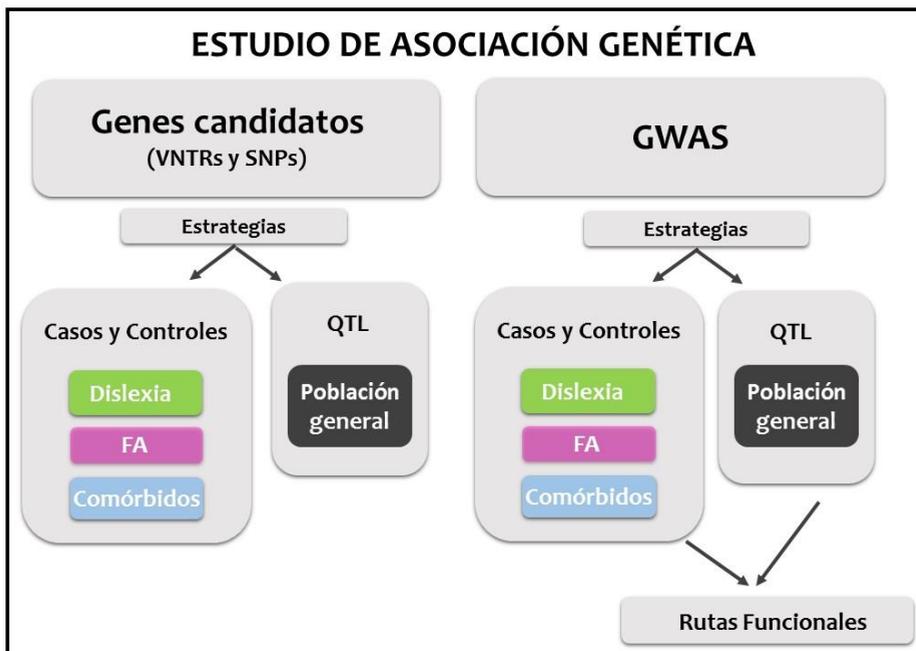


Figura 3.- Esquema general de las estrategias de asociación seguidas en este proyecto.

### 3.2 SUJETOS DE ESTUDIO Y CRITERIOS DE INCLUSIÓN

En la ejecución de un estudio de asociación genética, tanto si es un estudio de genes candidatos como si es de genoma completo, hay un factor sin el cual sería imposible cumplir con los objetivos del proyecto, que es contar con una población bien definida en cuanto al diagnóstico del rasgo que se desea investigar.

#### 3.2.1 Pruebas psicométricas y variables registradas para todos los sujetos del presente proyecto

Tras previo consentimiento informado de los padres o tutores, se reclutaron un total de 5008 muestras de saliva de niños de diversas zonas de España. Los investigadores del consorcio COEDUCA (Proyecto CONSOLIDER – CSD2008-00048) se encargaron de administrar a cada uno de los niños una batería de pruebas psicométricas que permiten evaluar variables relacionadas con la lectura y la atención. En total fueron 174 las variables medidas, las cuales se agruparon en 12 tareas (ver Tabla 4 y descripción de las tareas más adelante), siendo posible inferir de cada una de éstas una medida de un proceso cognitivo diferente.

Tabla 4. Tareas evaluadas y funciones cognitivas asociadas para cada trastorno cognitivo evaluado.

Nº	TAREA	Función cognitiva	Trastorno cognitivo
1	Eliminación del primer fonema	Conciencia fonológica	Dislexia
2	Identificación de fonema inicial en dibujos		
3	Repetición en pseudopalabras		
4	Identificación silábica	Habilidad para la percepción del habla	
5	Nombrado rápido de letras, números, colores y dibujos	Fluidez de lectura y velocidad de procesamiento	
6	Lectura de palabras y pseudopalabras	Decodificación de grafema-fonema	
7	Identificación de la identidad de letras	Procesamiento ortográfico	
8	Identificación de la posición de letras		
9	<i>Stroop</i> numérico	Conflicto e inhibición de respuesta	Falta de atención
10	<i>Stroop</i> verbal		
11	Señal de STOP	Proceso inhibitorio	
12	ANT ( <i>Attentional Network Test</i> )	Alerta, orientación y efecto conflicto (control ejecutivo)	

A continuación se describen las pruebas realizadas para evaluar cada una de estas tareas, para las cuales las variables registradas fueron el número de aciertos y el tiempo necesario para

la ejecución de la misma. Posteriormente se calcularon varios índices a partir de estas dos medidas.

### 3.2.1.1 Conciencia fonológica

Uno de los aspectos de gran importancia para el aprendizaje de la lectura es el conocimiento de las unidades que componen el lenguaje, y para una buena habilidad lectora es indispensable tomar conciencia de la estructura fonológica del lenguaje hablado o de las estructuras del sonido. La conciencia fonológica se refiere a la capacidad de ser consciente de las unidades en que puede dividirse el habla (palabras, sílabas y fonemas). Es un determinante importante a la hora de evaluar el éxito del aprendizaje de la lectura y de la escritura. Para medir esta función se utilizaron los valores obtenidos de las siguientes tareas:

#### Tarea 1: Eliminación del primer fonema

La prueba consiste en presentar auditivamente a los participantes una palabra la cual tienen que reproducir eliminando el primer fonema (ej: de la palabra MINA reproducir INA). Las palabras utilizadas para ello fueron de diversas estructuras de vocales y consonantes.

#### Tarea 2: Identificación de fonema inicial en dibujos

En esta prueba, se muestran dos dibujos y se presenta auditivamente una palabra. Los participantes tienen que elegir el dibujo que empieza con el mismo fonema que la palabra presentada auditivamente (ver Figura 4 como ejemplo del ejercicio).

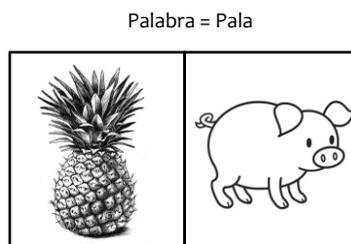


Figura 4 Ejemplo del ejercicio de identificación de fonemas en dibujo. Se presenta auditivamente la palabra “pala” y el participante debería elegir el dibujo de la piña.

El tiempo global empleado para realizar los ejercicios de eliminación del primer fonema e identificación de fonemas en dibujos se definió como “*Tiempo global en identificación de fonema*” (var 1).

### **Tarea 3: Repetición de pseudopalabras**

En esta tarea los participantes escuchan una serie de pseudopalabras cada vez de mayor dificultad y ellos tienen que repetirlas con la mayor rapidez posible.

#### **3.2.1.2 Percepción del habla**

La percepción del habla es el proceso por el cual los sonidos del lenguaje son escuchados, interpretados y entendidos. Su estudio pretende entender cómo las personas reconocen los sonidos del habla y cómo utilizan esta información para entender el lenguaje hablado y para ello se realizó la siguiente prueba:

### **Tarea 4: Identificación silábica**

En este ejercicio se presentan auditivamente un par de sílabas y los participantes tienen que discriminar entre ellas diciendo si son iguales o distintas (ej: pe-pe/pe-be). Mediante esta tarea se evalúa la habilidad para discriminar entre el contraste de consonantes en un contexto de parejas de sílabas. El tiempo global empleado para realizar este ejercicio, para los dos tipos de sílabas (iguales y distintas), se denominó “*Tiempo global en discriminación silábica*” (var 2).

#### **3.2.1.3 Fluidez de lectura y velocidad de procesamiento**

La fluidez de la lectura representa la capacidad para leer de manera correcta, con la entonación, ritmo, fraseo y pausas apropiadas. Para ello es necesario que el lector entienda cómo se relacionan las letras con los sonidos de la lengua, la forma en la que estos sonidos se mezclan para formar palabras y el significado de estas palabras. Un lector fluido no tiene que parar para “descifrar” cada palabra, es decir, puede leer la mayoría de ellas automáticamente, por ello, puede centrar su atención en comprender lo que está leyendo. La fluidez lectora va unida a la velocidad de procesamiento. Ésta, representa la relación entre la ejecución cognitiva y el tiempo

invertido en realizar una tarea, por lo tanto implica realizar con fluidez dichas tareas. Se describe como la capacidad de procesar información de forma automática sin pensar conscientemente en ella, cuanto mayor velocidad de procesamiento, más eficientemente se pensará y se aprenderá. El ejercicio empleado para evaluar dicho proceso fue el siguiente:

### Tarea 5: Nombrado rápido de letras, números, colores y dibujos

Esta tarea también denominada RAN (por sus siglas en inglés de *Rapid Automated Naming*) consiste en entregar unos listados con dibujos, números, letras o con colores (por separado) y el participante tiene que indicar el dibujo, número, letra o color que ve con la mayor rapidez posible (ver Figura 5 como ejemplo). Este tipo de ejercicio está fuertemente correlacionado con medidas de la velocidad de procesamiento, en consecuencia puede predecir la fluidez para la lectura y la habilidad para leer que un niño puede desarrollar en un futuro. Dentro de esta tarea se calculó un tiempo compuesto identificado como “*Tiempo de nombrado rápido compuesto de dibujos y colores*” (var<sub>3</sub>). Para obtener esta medida, se hizo un cálculo compuesto del tiempo empleado para la realización de las pruebas de dibujos y colores.

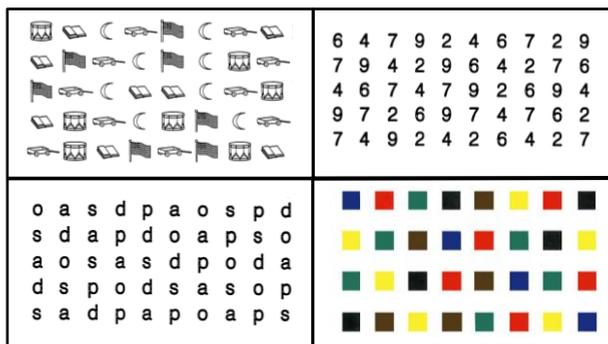


Figura 5. Ejemplo del ejercicio de nombrado rápido.

#### 3.2.1.4 Decodificación de grafema-fonema

En lingüística, un grafema se considera como la unidad mínima de escritura de una lengua, es decir, los grafemas se corresponden con las letras. Un fonema, en cambio, es la articulación mínima de un sonido vocálico o consonántico, son los sonidos del habla que permiten distinguir las palabras de una lengua. El proceso de decodificación de grafema-fonema se refiere al proceso de relacionar las representaciones de las palabras escritas con su representación verbal, es decir,

implica relacionar las letras que componen una palabra con su correspondiente sonido. Para ello es inevitable la conciencia fonológica, ya que es necesario el conocimiento de los sonidos de las palabras para poder relacionarlos con los sonidos de una letra. La tarea considerada para esta función fue:

#### **Tarea 6: Lectura de palabras y pseudopalabras**

A los participantes se les presentan una serie de palabras y pseudopalabras (compuestas por diferentes estructuras de vocales y consonantes) y ellos tienen que leerlas lo más rápidamente posible. Dentro de esta tarea se definió la “*Eficiencia de palabras y pseudopalabras*” (var 4) y ésta se calculó a partir del número de aciertos cometidos y el tiempo requerido para realizar el ejercicio.

#### **3.2.1.5 Procesamiento ortográfico**

El procesamiento ortográfico hace posible el reconocimiento de la palabra de forma fluida y sin esfuerzo. Esta habilidad aparece con posterioridad al procesamiento fonológico, ya que según se van aumentando las representaciones léxicas se favorece la unión entre el deletreo, la pronunciación y el significado de la palabra. En consecuencia los niños se convierten en lectores más eficientes y dependen en menor medida del procesamiento fonológico. Es por ello que el procesamiento ortográfico sirve como indicador del rendimiento lector. Las pruebas realizadas para medir el procesamiento ortográfico fueron:

#### **Tarea 7: Identificación de la identidad de letras**

En esta prueba, al participante se le presentan dos pares de dos letras y éste tiene que decir si son iguales o no. Estas letras pueden variar en su identidad (igual o distinta), en su similitud (letras parecidas o no) y en su representación de caja (mayúscula-minúscula).

#### **Tarea 8: Identificación de la posición de letras**

En esta ocasión, la tarea consiste en mostrar dos pares de cuatro letras y el participante tiene que decir si son iguales o no. En el caso de ser diferentes, estas diferencias pueden ser debidas a una transposición, es decir, con las dos consonantes interiores transpuestas (irfo-ifro) o a un remplazo, con las dos consonantes interiores remplazadas por otras (isto-ifro).

Para definir los “*Aciertos del efecto de transposición*” (var 5), se calculó el número de aciertos de la condición de transposición. Es decir, el número de veces que los participantes respondieron que (irfo-ifro) eran diferentes.

### 3.2.1.6 Procesamiento automático/atención selectiva/efecto conflicto

Existen dos tipos de procesos cognitivos, por un lado el procesamiento controlado, el cual implica un esfuerzo y un consumo de atención, y se produce bajo consciencia. Por otro lado el procesamiento automático, el cual incluye procesos que se aprenden y se hacen de una manera menos consciente o de forma automática. Para evaluar este segundo tipo de procesamiento se le genera al participante una situación de **conflicto**, donde tiene que responder sobre un aspecto (ej.: tamaño físico del número/color de la palabra) ignorando otro de los aspectos presentes (ej.: cantidad numérica real/significado de la palabra), por lo que el sujeto tiene que ser **selectivo**, centrarse únicamente en uno de los estímulos, e inhibir otras respuestas. Mediante estos ejercicios se puede evaluar el procesamiento automático ya que la mente automáticamente determina el significado de la palabra o la cantidad numérica, por otra parte, también sirve como indicadora de la atención selectiva. Para medir esta función cognitiva las tareas utilizadas en el presente proyecto fueron:

#### **Tarea 9: Stroop numérico**

Se exponen dos números y el participante tiene que decidir cuál de los dos es de mayor tamaño físico, sin tener en cuenta su verdadera cantidad numérica. Estos números pueden ser de la misma cantidad numérica (condición neutral) o de diferente cantidad numérica. En este segundo caso, el tamaño físico puede ser congruente con la magnitud numérica o incongruente (ej.: en la pareja de números 1 y 6, si el 1 es de menor tamaño físico que el 6 estaríamos ante una situación congruente, y si fuera al revés ante una situación incongruente) (ver Figura 6 como ejemplo). Este ejercicio permite calcular varios efectos a partir de la resta del número de aciertos o del tiempo empleado entre las tareas congruentes y las tareas incongruentes. De esta forma se calcularon el “*Efecto en aciertos de condiciones incongruente menos congruente (Stroop numérico)*” (var 6) y el “*Efecto en tiempos de condiciones incongruente menos congruente (Stroop numérico)*” (var 7).

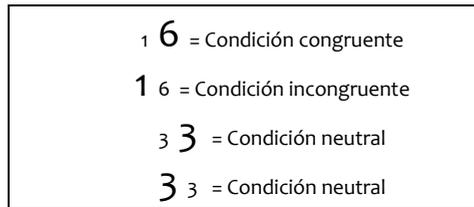


Figura 6. Ejemplo de ejercicio de Stroop numérico

### Tarea 10: Stroop verbal

Se muestran una serie de palabras/símbolos (estímulos) escritas en diferentes colores y el participante tiene que decir de qué color está escrita la palabra. Se utilizan para ello estímulos de tres tipos: palabras que nombran colores, palabras que no nombran colores y grupos de símbolos. Cuando la palabra que nombra un color y el color en el que está escrita es el mismo, estamos ante una situación congruente, en caso contrario ante una situación incongruente. Las palabras que no determinan colores y los grupos de símbolos pertenecen a una situación neutra (ver Figura 7 como ejemplo). Al igual que en la tarea anterior, los efectos denominados “Efecto en aciertos de condiciones incongruente menos congruente (Stroop verbal)” (var 8) y “Efecto en tiempos de condiciones incongruente menos congruente (Stroop verbal)” (var 9) se calcularon a partir de la resta del número de aciertos o del tiempo empleado entre las tareas congruentes y las tareas incongruentes.

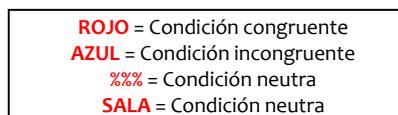


Figura 7.-Ejemplo de ejercicio de Stroop verbal

#### 3.2.1.7 Proceso inhibitorio

La respuesta inhibitoria consiste en la capacidad de inhibir o controlar las respuestas automáticas o impulsivas para dar lugar a respuestas mediadas por la atención. Los ejercicios para determinar dicha función se basan en la habilidad para suprimir una respuesta que ya no es necesaria, concepto que apoya el poseer una conducta flexible en ambientes de constante cambio. Este proceso se considera un importante factor en la función ejecutiva<sup>307</sup> y se midió con esta prueba:

**Tarea 11: Señal de STOP**

En la pantalla del ordenador aparece un avión en una determinada dirección (derecha o izquierda). El participante tiene que presionar una tecla (S o L) dependiendo de la dirección del avión y dejar de presionar la tecla cuando en la pantalla aparece una señal de STOP.

**3.2.1.8 Alerta, orientación y efecto conflicto (control ejecutivo)**

La atención se define como la capacidad de entender y tener en consideración un objetivo. Se utiliza como filtro de los estímulos ambientales, decidiendo y priorizando los estímulos más relevantes por medio de la concentración, de esta forma, se convierte en un mecanismo capaz de controlar de una manera estructurada la actividad humana. También se encarga de controlar y regular los procesos cognitivos. Existen tres mecanismos de atención en el cerebro humano: la alerta, la orientación y la resolución del conflicto. Éstos caracterizaron de la siguiente forma:

**Tarea 12: ANT**

ANT (de las siglas en inglés *Attention Network Test*) es un test que permite evaluar los tres mecanismos de atención mencionados anteriormente. El ejercicio consiste en decir en qué dirección va la flecha central de un conjunto de flechas (derecha o izquierda) (ver Figura 8 como ejemplo). Estas flechas (estímulos) se presentan de una manera congruente (la flecha central va en la misma dirección que las de su alrededor) o de una manera incongruente (la flecha central va en dirección opuesta a las de su alrededor). En el ejercicio se presentan, además de las flechas, señales adicionales (asteriscos) como pista de dónde aparecerán las flechas en la pantalla. Dependiendo de la posición en la que aparezcan estos asteriscos se diferencian: una señal válida (en la misma posición donde aparecerá el estímulo), doble señal (encima y debajo de donde aparecerá el estímulo), señal contraria (en posición contraria a donde aparecerá el estímulo), señal neutra (en la mitad de la pantalla) o sin señal. Al final de la tarea se reporta el número de errores y el tiempo invertido en la misma. A partir de los datos registrados con este ejercicio se calculó el “Índice conflicto” (var 10) mediante la resta del tiempo necesario para todos los ejercicios incongruentes y todos los congruentes (con todas las posibles señales de pista).

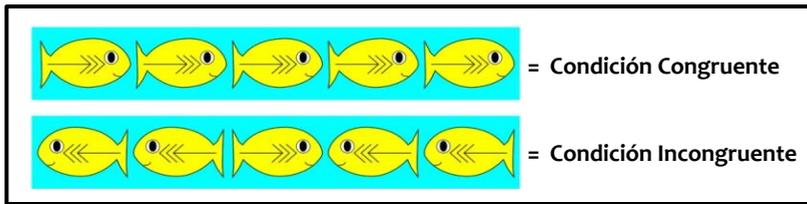


Figura 8. Ejemplo de ejercicio ANT

### 3.2.2 Diagnóstico de los sujetos a partir de las variables medidas

#### Selección de los sujetos disléxicos

Para la selección de los sujetos disléxicos (D) y controles ( $D_{ctr\_dys}$ ) se llevó a cabo un análisis discriminante a partir de un grupo *a priori* diagnosticados como disléxicos ( $N=43$ ) y un grupo de controles con una desviación típica por encima de la media en las pruebas de comprensión de textos ( $N=470$ ). Las variables de la función discriminante que mejor separaban a disléxicos y controles, con una sensibilidad de 91% y una especificidad del 94%, fueron la eficiencia de palabras y pseudopalabras, el RAN de dibujos y colores, el emparejamiento de dibujo-fonema, y la identificación de sílabas. Posteriormente, se aplicó esa función discriminante a todos los datos y se consideraron como disléxicos aquellos participantes situados en los deciles 1 y 2 de la función discriminante y como controles los situados en el decil 5 o por encima de éste. Todos los participantes tenían un coeficiente intelectual no verbal superior a 80.

#### Selección de los sujetos con falta de atención

La selección de los participantes con falta de atención (FA), se llevó a cabo teniendo en cuenta la ejecución de los participantes en las tareas de Stroop verbal, y Stroop numérico (tanto el tiempo de reacción como aciertos) y en la ANT (tiempo de reacción). En concreto se consideró como participantes con falta de atención aquellos que puntuasen en el cuartil 4 del tiempo de reacción (lentos) o bien en el cuartil 1 de aciertos (pocos aciertos), ambos en las tareas de Stroop numérico, Stroop verbal y ANT. Se clasificaron como controles ( $Ctr_{FA}$ ) los participantes situados en los cuartiles 1, 2 y 3 de los tiempos de reacción. Asimismo, los participantes que puntuaron bajo simultáneamente en el tiempo de reacción del Stroop verbal y numérico, se definieron como

controles más estrictos (Ctr\_FA\_2). Todos los participantes tenían un coeficiente intelectual no verbal superior a 80.

### **3.2.3 Criterios de inclusión/exclusión de muestras**

Entre todas las muestras recibidas en el laboratorio se hizo una selección para el estudio. El origen español de las muestras fue un criterio obligatorio por lo que se descartaron aquellas muestras procedentes de niños con: madre o padre de origen no-español, madre o padre sin dato de origen, niño de nacimiento no-español y niño sin dato de nacimiento. También se excluyeron del estudio las muestras de los niños a los que no se les había realizado la batería completa de pruebas psicométricas y las muestras con una calidad de DNA inapropiada para este estudio.

## **3.3 EXTRACCIÓN Y CARACTERIZACIÓN DE DNAs DE SALIVA**

Uno de los objetivos de este proyecto es obtener DNAs de calidad óptima a partir de saliva que nos permitan la correcta interpretación de los resultados, los cuales serían dudosos si los DNAs caracterizados estuvieran degradados.

### **3.3.1 Recogida de muestras de saliva a los participantes**

Cada uno de los tutores legales de los niños participantes en el estudio accedió a ceder una muestra de saliva que fue recogida con el kit de saliva "Oragene-500" (DNA Genotek Inc). Cada individuo depositó 2 ml de saliva en un tubo que se mezcló de inmediato con otros 2 ml de buffer de almacenamiento, el cual inactiva cualquier tipo de microorganismo presente en la muestra y el cual preserva el DNA estable durante años incluso almacenado a temperatura ambiente. Para un mejor rendimiento de las técnicas posteriores, se les pidió a los participantes no comer, beber, fumar o masticar goma de mascar 30 minutos previos a la recogida de su saliva. El CICbioGUNE contaba con el consentimiento del Comité de Ética y Bienestar Animal para la manipulación de este tipo de material en el momento de recibir las muestras de saliva.

### 3.3.1.1 Extracción y cuantificación del DNA

La extracción del DNA del total de la saliva se efectuó mediante el protocolo “*Laboratory protocol for manual purification of DNA from 4,0 ml of Oragene ® DNA/saliva*” siguiendo las indicaciones de la casa Oragene. La concentración total del DNA extraído se midió una primera vez mediante Nanodrop 2000 (Thermo Fisher Scientific) (método de absorbancia a 260 nm) y posteriormente se recuantificó mediante Qubit (Thermo Fisher Scientific, método de fluorimetría). Las muestras que no llegaron a la concentración requerida para la experimentación se precipitaron con etanol para disminuir el volumen de eluido. Todas las muestras se normalizaron de una manera automática mediante el robot epMotion 5070 (Eppendorf) a una concentración determinada y se almacenaron a -80°C. La integridad del DNA se evaluó mediante electroforesis en geles de agarosa al 0,7 %, 1x TAE y el DNA fue visualizado con luz UV previa tinción con bromuro de etidio.

### 3.3.2 Caracterización del contenido del DNA proveniente de saliva humana

#### 3.3.2.1 Cuantificación de DNA Humano

Para cuantificar la proporción de DNA humano presente en cada muestra se realizó una qPCR con sonda TaqMan™ (Thermo Fisher Scientific) diseñada en una región del gen BRCA (*Breast Cancer*), la cual está muy conservada en humanos en cuanto a secuencia y a número de copias. Las secuencias de los cebadores para la amplificación y la sonda TaqMan™ que se utilizaron son los descritos por Covault *et al.* en 2003<sup>308</sup>: BRCA\_F 5'- AAA CAT GTT CCT CCT AAG GTG CTT T-3', BRCA\_R 5'-ATG AAA CCA GAA GTA AGT CCA CCA GT-3', BRCA\_MGB (6FAM-MGB) 5'-ACA CAG CTA GGA CGT-3'. Las condiciones en las que se llevó a cabo la reacción de qPCR fueron: 0.4 µM de ambos cebadores BRCA, 0.25 µM de sonda BRCA, 1x TaqMan™, Universal PCR Master Mix (No AmpErase. UNG (P/N 4324018)) y 20 ng de DNA genómico total, en un volumen final de 10µl. La amplificación se llevó a cabo en un termociclador CFX96-C1000 (Bio-Rad Laboratories, Inc.) durante 49 ciclos de 30 segundos a 95°C y 1 minuto a 60°C. Para poder determinar el porcentaje de DNA humano presente en cada muestra se realizó una curva estándar utilizando para ello estándares artificiales de una mezcla de DNA humano (ABI, Ref. 4312660) y bacteriano (Sigma, Ref. D4889). Estos porcentajes variaron entre el 100 % de DNA humano (0 % de DNA bacteriano) y 0 % de DNA humano (100 % DNA bacteriano). Tanto los estándares de DNA humano:bacteriano

como las muestras se procesaron por triplicado.

### 3.3.2.2 Cuantificación de DNA bacteriano

La cuantificación del porcentaje de DNA bacteriano presente en cada una de las muestras también se llevó a cabo mediante qPCR. En este caso se utilizó EvaGreen como fluoróforo para la medición de fluorescencia durante la amplificación del target establecido. Los cebadores empleados para la PCR cuantitativa fueron los descritos por Muyzer *et al.* en 1993<sup>309</sup> los cuales amplifican una región del gen 16S ribosomal de 193pb, región muy conservada entre las especies bacterianas: BACT\_F1 5'-CCT ACG GGA GGC AGC AG-3' y BACT\_R1 5'- ATT ACC GCG GCT GCT GG-3'.

Las condiciones elegidas finalmente para la cuantificación del DNA bacteriano fueron: 1x mi-Hot Taq Mix (Metabion, mi-E8011), 1μM de cebadores, 1x EvaGreen (Biotium Inc. ref. 31000) y 20 ng de DNA en un volumen total de 15 μl. Las reacciones se mantuvieron durante 3 min a 95°C, seguido de 26 ciclos de 30 segundos a 95°C, 30 segundos a 62,8°C y 1 min a 72°C. La curva de disociación (*melting curve*) consistió en un gradiente de 65°C a 95°C en incrementos de 0,5°C cada 5 segundos (CFX96, C1000 thermal cycler). Para estimar el porcentaje de DNA bacteriano, se utilizaron los mismos estándares empleados en la determinación de la proporción de DNA humano (ver apartado 3.3.3.1). Tanto los estándares DNA humano:bacteriano como las muestras se procesaron por triplicado.

### 3.3.2.3 Cálculo de porcentajes de DNA humano y bacteriano

Los datos obtenidos de las PCRs de cuantificación de DNA humano y bacteriano se analizaron mediante el software “Bio-Rad CFX Manager”. Este programa detecta un valor Ct (*Cycle Thershold*, número de ciclo de la PCR en el que se detecta producto amplificado y que es proporcional a la cantidad inicial de DNA de la muestra) para cada uno de los estándares de distintos porcentajes de DNA humano y bacteriano, y construye con ellos una recta de calibrado. Esta recta se define como una regresión lineal con un determinado coeficiente de correlación ( $r^2$ ) que se representa, mediante la fórmula:  $Y = aX + B$ , donde Y es el Ct medio de la muestra-estándar, a es la pendiente de la recta estándar, X es la proporción de DNA inicial de la muestra y B es una

constante matemática. A partir del Ct medio de cada una de las muestras cuantificadas y empleando la ecuación mencionada, se pudo estimar su porcentaje de DNA humano y bacteriano.

### 3.4 ESTUDIO DE ASOCIACIÓN DE GENES CANDIDATOS

#### 3.4.1 Selección de las variantes genéticas candidatas

Para la validación de la asociación de genes candidatos con dislexia y falta de atención en nuestra población de estudio, se seleccionaron 11 marcadores basándonos en estudios anteriores<sup>28,66,109,167,181,310-312</sup>. En concreto fueron 9 SNPs y 2 repeticiones en tándem de número variable (VNTR, del inglés *Variable Number of Tandem Repeats*) localizados en 9 genes. Siete de los marcadores (5 SNPs y 2 VNTRs) se habían relacionado previamente con TDAH y 4 SNPs con dislexia (ver Tabla 5). La elección del marcador exacto se hizo en base a su previa asociación con el trastorno cognitivo o por su posible efecto funcional de acuerdo a la base de datos FastSNP<sup>313</sup>.

Tabla 5 .Características de los marcadores candidatos seleccionados. Gen=nombre del gen. Chr=cromosoma donde se ubica el gen. Variant\_ID=nombre del marcador. Tipo de variante: SNP o VNTR. Asociado previamente: Sí o No, junto con la referencia donde se ha publicado cada estudio. Trastorno cognitivo asociado=trastorno con el cual se ha asociado el marcador.

Gen	Chr	Variant_ID	Tipo de variante	Asociado previamente	Trastorno cognitivo asociado
KIAA0319	6	rs4504469	SNP	Sj <sup>311</sup>	DISLEXIA
DCDC2	6	rs2274305	SNP	Sj <sup>66</sup>	
DYX1C1	15	rs57809907	SNP	Sj <sup>28</sup>	
FOXP2	7	rs12533005	SNP	Sj <sup>109</sup>	
COMT1	22	rs4680	SNP	Sj <sup>181</sup>	TDAH
DBH	9	rs1611115	SNP	Sj <sup>314</sup>	
DBH	9	rs13306306	SNP	No	
MAOA	9	rs6323	SNP	Sj <sup>202</sup>	
MAOA	X	rs72554632	SNP	No	
DRD4	11	DRD4	VNTR	Sj <sup>310</sup>	
DAT1	5	DAT1	VNTR	Sj <sup>167</sup>	

### 3.4.2 Estudio de VNTRs

#### 3.4.2.1 Genotipado de los VNTRs situados en los genes *DRD4* y *DAT1*

El genotipado del VNTR de 48 pares de bases (pb) situado en el exón III del gen *DRD4* se realizó en un volumen total de reacción de 20  $\mu$ l que contenía: 1x mi-Hot *Taq* Mix (Metabion), 0.2  $\mu$ M de cada cebador (Forward: 5'-GGT CTA CTC GTC CGT GTC CTC CTT-3' y Reverse: 5'-CTC TCC TCC CCG CCC CTC AG-3') y 20 ng DNA genómico. Al ser esta una secuencia rica en CG se añadieron 2  $\mu$ l de DMSO (Dimetil Sulfoxido) 100 % a la mezcla de reacción para facilitar su desnaturalización. Las muestras se procesaron en un termociclador MyCycler (Bio Rad) con las siguientes condiciones: 40 ciclos de 30 seg a 95°C, 30 seg a 62°C y 30 seg a 72°C. Los productos se visualizaron en geles de agarosa del 2% y 1x TAE y los genotipos correspondientes se identificaron en base al marcador de tamaño  $\phi$ x174/*Hinf*I (Biotools) y un marcador compuesto por fragmentos de número de repeticiones específicos generado a la carta en el laboratorio (ver sección 3.4.2.3).

El genotipado de polimorfismo VNTR de 40 pb de la región 3'UTR del gen *DAT1*, se realizó en un volumen total de reacción de 20  $\mu$ l que contenía: 1x mi-Hot *Taq* Mix (Metabion), 0.5  $\mu$ M de cebadores y 20 ng de DNA genómico. Los cebadores fueron diseñados según lo descrito en Vandenberg et al.<sup>315</sup> con una modificación en el cebador reverso para mejorar la amplificación: Forward: 5'TGT GGT GTA GGG AAC GGC CTG AG- 3' y Reverse: 5'-TGG AGG TCA CCG CTC AAG G-3'. La región que contenía el VNTR se amplificó en un termociclador MyCycler (Bio-Rad), durante 40 ciclos de 30 seg a 95°C, 30 seg a 68°C y 30 seg a 72°C. Los productos se visualizaron en geles de agarosa del 2 % y 1x TAE, y sus correspondientes genotipos se identificaron según el tamaño del marcador de peso molecular  $\phi$ x174/*Hinf*I (Biotools) y un marcador compuesto por fragmentos de número de repeticiones específicos creado en el laboratorio (ver sección 3.4.2.3).

#### 3.4.2.2 Clonación de los distintos alelos de VNTRs amplificados

Los productos de PCR que contenían un número distinto de repeticiones de los dos VNTRs estudiados fueron clonados para poder ser secuenciados y evaluar el número exacto de repeticiones que contenían. Para ello se utilizó el kit de clonación "TOPO TA Cloning kit" (Invitrogen). El kit consta del vector "pCR™2.1-TOPO® vector", células competentes DH5 $\alpha$ ™ ("One

Shot® Chemically Competent cells”) y de un plásmido control pUC19. Para la transformación de estas células se utilizó una reacción de ligación con un ratio inserto:vector de 3:1. La discriminación de colonias positivas (aquellas que presentaban el inserto del producto de PCR), se realizó en base al color de la colonia, siendo las colonias blancas positivas y las azules negativas (la diferencia de color se debe al gen *LacZ $\alpha$*  del vector, el cual se rompe al insertarse el fragmento de PCR clonado). De esta forma, se seleccionaron 10 colonias que potencialmente contenían vector con cada fragmento de PCR clonado para verificar su tamaño mediante PCR con los cebadores M13 específicos del vector.

### 3.4.2.3 Secuenciación de los productos de PCR clonados

Con el objetivo de identificar el número exacto de repeticiones de los alelos-VNTR clonados (ver apartado anterior), cada clon se secuenció con tecnología Sanger tanto con los cebadores específicos del vector (M13) como con los cebadores utilizados para la amplificación de dichos VNTRs. La secuenciación se llevó a cabo en la compañía STABVIDA (<http://www.stabvida.com/index.php/es>). El alineamiento y comparación de las secuencias obtenidas se realizó mediante el programa Sequencher™ 4.7 (Gene Codes). Una vez determinados los números de repeticiones de cada fragmento, se amplificaron todos los alelos diferentes con los cebadores específicos para cada uno de los genes y se construyó un marcador de tamaño para cada uno de los VNTRs.

### 3.4.2.4 Análisis de asociación de los VNTRs: caso/control

Las comparativas siempre se realizaron entre un grupo de casos frente a un grupo control, tal y como se especifica en la Tabla 10 de resultados. El análisis de asociación de los VNTRs se realizó de diversas maneras. Por un lado se compararon las frecuencias de cada genotipo frente al resto, por otro, la presencia de cada genotipo frente a la ausencia del mismo y, finalmente, por agrupaciones denominadas “repeticiones largas” frente “repeticiones cortas”. Todos los análisis se realizaron mediante el test exacto de Fisher utilizando la herramienta informática R (versión 2.13.0).

### 3.4.3 Estudio de polimorfismos de nucleótido único (SNPs)

#### 3.4.3.1 Genotipado de los SNPs candidatos seleccionados

El genotipado de las muestras para todos los SNPs seleccionados se realizó utilizando los sistemas de PCR a tiempo real ViiA™ 7 o QuantStudio™ 6 (QS-6) mediante la tecnología TaqMan™ (Thermo Fisher Scientific). Los detalles de las sondas utilizadas así como el número de ciclos de PCR utilizado para cada una de estas reacciones se pueden encontrar en la tabla suplementaria 1. Las reacciones se efectuaron en microplacas de 384 pocillos en un volumen total de 5 µl o 10µl en el caso de *FOXP2* y utilizando 20 ó 40 ng de DNA genómico total, respectivamente. La amplificación se hizo según las indicaciones del fabricante (50-55 ciclos de 15 segundos a 95°C y 90 segundos a 60°C).

#### 3.4.3.2 Análisis de asociación caso/control

##### **Análisis individual de SNPs**

El análisis de asociación de SNPs individuales se realizó mediante la comparación de las frecuencias alélicas entre el grupo de casos y de controles utilizando para ello el test de chi cuadrado y el test de Fisher (ver Tabla 10 de resultados para los detalles de las comparativas realizadas). Por otra parte también se analizaron los datos bajo varios modelos genéticos (genotípico, dominante y recesivo), así como utilizando el test de tendencia de Cochran-Armitage. Todos estos análisis se han realizado con la herramienta PLINK<sup>316</sup>.

##### **Análisis de interacciones por parejas de marcadores (epistasia)**

El análisis de interacciones entre SNPs (dos a dos) se realizó implementando un modelo de regresión logística mediante el programa PLINK.

##### **Análisis de asociación haplotípica partiendo de los datos de genotipado de genoma completo**

El estudio de asociación en este proyecto se abordó de dos maneras diferentes. Por una parte se realizó un análisis de asociación de SNPs o VNTRs en genes candidatos (descrito en esta misma sección) y por otra parte un análisis de asociación de genoma completo (ver sección 3.5). De los datos obtenidos en el GWAS se pudo extraer un gran número de marcadores para 8 de los genes candidatos. Para ello se identificaron las posiciones de comienzo y fin de cada uno de los

genes y mediante PLINK se filtraron estos datos del resto. El análisis de haplotipos se realizó utilizando el programa Haploview<sup>37</sup> según los detalles descritos en el apartado 3.5.2.5.2.

### **3.5 ESTUDIO DE ASOCIACIÓN DE GENOMA COMPLETO**

#### **3.5.1 Genotipado de genoma completo: protocolo *Infinium***

El genotipado de genoma completo se realizó utilizando los arrays *Human Omniexpress* (Illumina Inc): v1.0 y v1.1. El panel completo de estos arrays ofrece la posibilidad de caracterizar 760525 SNPs distribuidos a lo largo de todo el genoma. Este panel se puede abordar para estudiar SNPs individuales y para el análisis de regiones con número de copias variables (CNVs). El protocolo utilizado para este genotipado de genoma completo fue el “*Illumina® Infinium® HD Assay Super protocol*” tal como lo describe Illumina Inc. con la modificación de partir de 400 ng de DNA genómico total en vez de los 200 ng recomendados, ya que el DNA utilizado proviene de saliva (ver detalles en la sección 4.1.3). En este protocolo, en primer lugar se amplifica el genoma completo en una incubación de 20 horas a 24°C con oligos arbitrarios y DNA-polimerasas. El gDNA amplificado se fragmenta enzimáticamente y se diluye en un tampón apto para la hibridación con las sondas que se encuentran inmovilizadas en el array *Human Omniexpress*. Tras la retirada de los fragmentos de DNA no hibridados o hibridados inespecíficamente se extiende una sola base del oligonucleótido que está hibridado con una hebra de DNA, la cual incorpora un marcaje capaz de emitir fluorescencia si se excita a una longitud de onda determinada. El escáner HiScan de Illumina distingue la luz emitida por cada fluoróforo en cada una de las muestras, generando un fichero con datos creados a partir de las fluorescencias registradas. Este fichero lo decodifica el software GenomeStudio (Illumina® GenomeStudio 2010, (c) Illumina, Inc. 2003-2010), transformando los valores de fluorescencias en datos de genotipado.

#### **3.5.2 Análisis de datos de genoma completo:**

##### **3.5.2.1 Análisis conjunto de diferentes versiones de arrays de genotipado**

A causa de la discontinuidad de la versión 1.0 de los arrays *Human Omniexpress* se tuvo que utilizar la siguiente versión (v1.1) para genotipar parte de las muestras de este estudio. Con el fin de analizar conjuntamente las muestras genotipadas con ambas versiones de *Human*

*OmniExpress*, se construyó un fichero con todos los genotipos. Dicha fusión se realizó mediante el programa PLINK utilizando el comando `--merge`. Esta función permite sobreponer dos ficheros en función de los marcadores y de los individuos que contengan, es decir se crea un nuevo fichero con todos los individuos y todos los marcadores que existan en los dos ficheros que se quieren juntar (para el análisis de CNVs no fue posible realizar un análisis conjunto, ver apartado 4.3.1).

### 3.5.2.2 Análisis de reproducibilidad de los arrays de genotipado masivo

Para que una técnica se considere fiable, ésta tiene que ser reproducible, es decir, bajo las mismas condiciones se deben obtener los mismos resultados independientemente de la técnica aplicada. En este estudio para evaluar la reproducibilidad de genotipado de los arrays *Human OmniExpress* se genotipó un mismo DNA en cada una de las placas utilizadas para el estudio de genoma completo (un total de 8 réplicas para la versión 1.0 de los arrays y un total de 11 réplicas para la versión 1.1 de los arrays, repartidas en 8 y 11 chips diferentes). Para medir el grado de concordancia de genotipado entre todas las réplicas se utilizó el comando `--merge --mode 7` en el programa PLINK el cual compara los genotipos de dos ficheros y calcula la similitud entre los marcadores genotipados en ambos. De esta forma, se calculó el grado de concordancia dentro de las réplicas de cada una de las versiones de los arrays y también la concordancia entre las réplicas de ambas versiones.

### 3.5.2.3 Criterios de filtrado de los datos crudos obtenidos mediante los arrays de genotipado masivo

Fueron descartados aquellos SNPs que tenían una codificación discordante entre las dos versiones de los arrays (ver apartado 4.3.1.), así como los marcadores que no llegaban al umbral de *GenTrain Score* de 0.6 considerado por la compañía Illumina (este parámetro es un indicador de la calidad con la que se ha efectuado el *cluster* del marcador). De esta forma, se inspeccionaron visualmente los *clusters* de los marcadores con un valor de *GenTrain Score* menor a 0.6 y se corrigieron o descartaron según fuera necesario.

Además, se excluyeron las muestras con un porcentaje de genotipado menor al 98 %, en base al valor de *Call Rate* indicado en GenomeStudio, y aquellos SNPs que no se genotiparon en más de un 90 % de las muestras, los que no superaron una frecuencia del 5 % en cuanto a su alelo

minoritario (MAF, *Minor Allele Frequency*), así como los que no pasaron el test de Hardy Weinberg con un valor de  $p$  de 0.001.

### 3.5.2.4 Estructuración y selección de *outliers*: Análisis de componentes principales

El análisis de componentes principales (APC) sirve para reducir la dimensionalidad de un conjunto de datos. En el caso de los estudios de genoma completo al tener genotipos de cientos de miles de marcadores, mediante el APC se reduce la información de todos estos marcadores, de manera que se calculan los dos componentes principales (los cuales explican la mayoría de la variabilidad del conjunto de marcadores) y se pueden correlacionar uno contra el otro. Mediante este análisis se puede deducir si los individuos estudiados son un grupo genéticamente homogéneo (todos los individuos se distribuyen en un único grupo) o bien concluir si existe estructuración poblacional dentro de ellos (los individuos se separan en diferentes grupos). El APC se realizó mediante PLINK. Primero se estimó el factor IBD (*Identical By Descent*, idénticos por descendencia) por parejas de individuos, con el objetivo de buscar parejas que fueran demasiado parecidas, es decir, más de lo esperado por azar. Partiendo de esta información se crearon los *clusters* y se calcularon los 4 componentes principales. Correlacionando los dos primeros componentes se obtuvo un gráfico de dimensión reducida del conjunto de datos. Este gráfico se realizó mediante la herramienta informática R. Aquellos puntos que se separan del grupo principal en el gráfico se denominan *outliers*. Para su detección se utilizaron dos técnicas: visualización directa en el gráfico y el paquete de R “*mvoutlier*”.

### 3.5.2.5 Estudio de asociación: caso/control

#### 3.5.2.5.1 Análisis de marcadores individuales: SNPs

El análisis de asociación de SNPs individuales se realizó mediante la comparación de las frecuencias alélicas del grupo de casos y de controles utilizando para ello el test de chi cuadrado y el test de Fisher. Por otra parte también se analizaron los datos bajo varios modelos genéticos (genotípico, dominante y recesivo), así como utilizando el test de tendencia de Cochran-Armitage. Todos los abordajes estadísticos se realizaron con la herramienta PLINK.

Una manera sencilla de revisar visualmente todos estos resultados es mediante un gráfico de manhattan, donde cada punto corresponde a un SNP. En el eje X de este tipo de gráficos se

representa la posición cromosómica de cada uno de los SNPs en pares de bases y, en el eje Y, el valor de significancia de cada una de las asociaciones en escala logarítmica. Cada cromosoma se simboliza con un color distinto de puntos para una mejor interpretación. Estas visualizaciones gráficas se obtuvieron mediante el software Haploview<sup>317</sup>.

### **Análisis de enriquecimiento de los marcadores más asociados según su localización génica**

Con el objetivo de ver si entre los marcadores más asociados existía una proporción mayor o menor (enriquecimiento) de marcadores ubicados en posiciones génicas concretas se realizó el siguiente análisis: 1) se calculó la proporción de cada uno de los tipos de mutaciones en el panel completo de genotipado. Los tipos de mutaciones fueron: intrón, codificador sinónimo, codificador “missense”, codificador “nonsense”, UTR-3’, UTR-5’, splice-3’, splice-5’, cerca del gen 3’ y cerca del gen 5’. Estas proporciones se denominaron “esperadas”. 2) En los 1500 marcadores más asociados de cada comparativa, se calcularon estas mismas proporciones, y así sucesivamente restando 50 marcadores en cada cálculo. Estas proporciones se denominaron “observadas”. 3) Se calculó la proporción entre los valores observados y esperados para cada tipo de mutación. Si este valor es cercano a 1, las proporciones observadas son similares a las esperadas, si el valor es mayor a 1, existen más marcadores con ese tipo de mutación entre los X marcadores más significativos en comparación con lo esperado, y si por el contrario el valor es menor a 1, existen menos marcadores significativos de lo esperado.

#### **3.5.2.5.2 Análisis de haplotipos**

En muchas ocasiones un único marcador no está asociado individualmente al rasgo de interés. En cambio, al analizarse conjuntamente varios marcadores con una posición cromosómica cercana, los valores de asociación cambian. Se denominan haplotipos a una serie de alelos (en un mismo cromosoma y cercanos en cuanto a su posición) que poseen una alta probabilidad de ser heredados conjuntamente, es decir, que están en desequilibrio de ligamiento (LD del inglés *Linkage Disequilibrium*). Para el análisis de haplotipos se utilizó el software Haploview<sup>317</sup>. Este programa se basa en las frecuencias alélicas de los marcadores para estimar cuáles de ellos están en LD e inferir así los denominados bloques haplotípicos. Para la creación de dichos bloques se utilizó el algoritmo de Gabriel *et al.*<sup>318</sup>, el cual viene establecido por defecto.

Este método ignora los marcadores con un MAF menor de 0.05%. Para el resto de los marcadores, calcula un valor de LD o  $D'$  por cada comparación dos a dos, y agrupa estos valores como "fuerte LD", "no concluyente" o "fuerte recombinación". Los marcadores se agrupan en un mismo bloque haplotípico si el 95 % de las comparaciones informativas (excluyendo las "no concluyentes") son del tipo "fuerte LD". Una vez generados los bloques, se calcula la asociación que hay entre casos y controles para cada uno de los bloques haplotípicos mediante un test de chi cuadrado.

### 3.5.2.5.3 Análisis de ventanas corredizas (*Sliding Windows*)

Otra metodología distinta para analizar conjuntos de marcadores es basarse en marcadores contiguos, lo cual se denomina análisis de ventanas corredizas. La diferencia que tiene con respecto al análisis haplotípico es que en este caso los bloques no se establecen en base al LD entre las variantes, si no por su localización adyacente. PLINK permite seleccionar el número de marcadores que contiene cada "ventana" estudiada y compara las frecuencias alélicas del genotipo compuesto entre casos y controles. Para este estudio se seleccionaron "ventanas" de SNPs desde 3 hasta 13 marcadores y se analizaron todos los posibles bloques generados en todos los cromosomas.

### 3.5.2.5.4 Análisis de interacciones por parejas de marcadores (epistasia)

Para el análisis de interacción entre SNPs de los estudios de dislexia, FA y comórbidos se utilizó el software SIXPAC<sup>319</sup>. Este software se basa en el test de contraste de LD de Zhou *et al.*<sup>320</sup>. Según sus autores, la interacción entre dos genes o *loci* crea desequilibrio de ligamiento entre ellos en la población caso, aun sin estar estos dos *loci* ligados. El nivel de LD dependerá de la magnitud de la interacción entre los dos *loci*. SIXPAC utiliza una versión simplificada de este test de contraste y de esta forma ofrece la posibilidad de abarcar todos los SNPs analizados en el genoma. El test compara los SNPs que están en un fuerte LD en casos pero que no están en desequilibrio (o tiene un bajo LD o un LD-reverso) en los controles. Muestra pequeños grupos de casos y resalta, en cada uno de estos, combinaciones de alelos presentes en todos los individuos del grupo. De esta forma, se disminuye drásticamente el número de combinaciones necesarias para analizar todas las parejas de SNPs resultantes de cada array. Cada genotipo (0/1/2 alelos minoritarios) se reduce a dos variables binarias: estado recesivo (rec, 2 alelos minoritarios) o estado dominante (dom, 0 ó 1 alelo minoritario). De modo que para cada pareja de SNPs, se analizan 4 modelos genéticos de interacción dependiendo del estado de cada uno de los dos SNPs

analizados: dom-dom, dom-rec, rec-dom, rec-rec. Para algunos de los análisis de epistasia, en concreto los relativos a las comparativas descritas en la Tabla 10, las cuales se realizaron sólo con las parejas de SNPs más significativas, se utilizó el programa PLINK.

### 3.5.2.5.5 Análisis de CNVs

#### Obtención de ficheros con datos LRR y BAF

En los arrays de genotipado de Illumina se utilizan dos sondas para detectar la presencia de los dos posibles alelos de cada uno de los SNPs. Los alelos medidos por el canal X (“Cy5 dye”, fluorescencia roja) se denominan alelos A mientras que los alelos medidos por el canal Y (“Cy3 dye”, fluorescencia verde) se denominan alelos B. A partir de estas dos medidas, el software de Illumina GenomeStudio calcula los siguientes parámetros: 1) LRR (logR ratio), el cual es una medida relativa al total de la intensidad por lo que también es proporcional al número de copias. 2) BAF (del inglés *B Allele Frequency*, frecuencia del alelo B), la cual es una medida de la contribución relativa del alelo B al total de la señal, por lo que refleja la composición alélica.

El cálculo del valor LRR se obtiene a partir de la siguiente ecuación:

$$\text{Log R ratio: } \log_2(R_{\text{observado}}/R_{\text{esperado}})$$

$R_{\text{observado}}$  es el valor de intensidad de fluorescencia normalizado, el cual se calcula mediante la transformación polar de las dos intensidades medidas para el SNP (una para cada uno de los alelos).

$R_{\text{esperado}}$  es una interpolación del ratio alélico observado con respecto a los *clusters* de genotipo canónico.

Si el valor de LRR es mayor a 0, se considera que existe una ganancia de copias mientras que si es menor a 0 se considera como una delección. En cuanto a la BAF, se calculó mediante el valor *theta* (*allelic intensity ratio*) de la muestra y la posición de *cluster* esperada. Este valor puede ser de 0 (sin ningún alelo B), 0.5 (un alelo B) ó 1 (dos alelos B) en un estado de diploidía. Variaciones en estos valores indican ganancias o pérdidas de segmentos de secuencia de DNA.

Para el análisis de asociación de número de copias variable (CNV) se exportaron los archivos de LRR y BAF del GenomeStudio a partir de los datos de fluorescencia obtenidos de los arrays *Human OmniExpress*.

### **Detección de las regiones de número de copia variable**

Para definir los segmentos de DNA que contenían diferentes números de copias se utilizó el programa informático PennCNV<sup>321</sup>. Este software utiliza un modelo oculto de Markov (HMM) para inferir el número de copias a partir de los datos de genotipado. Los segmentos definidos debían tener un mínimo de 10 SNPs para que fueran considerados, de lo contrario se descartaban. Durante el proceso, el programa calcula para cada una de las muestras la desviación estándar del LRR (LRR SD) como criterio de calidad de la muestra. Se descartaron para su posterior análisis las muestras con un LRR SD > 0.2, para minimizar así el número de falsos positivos.

### **Análisis de asociación caso/control**

Una vez identificadas las regiones de número de copias variable en cada una de las muestras, se utilizó la herramienta informática ParseCNV<sup>322</sup> para realizar un análisis de asociación de dichas zonas entre casos y controles. Este programa lo primero que hace es establecer las regiones CNV consenso (CNVR) a partir de los segmentos de número de copia variable identificados en PennCNV. Este paso es necesario para obtener las regiones exactas que se quieren comparar, ya que PennCNV identifica en cada muestra los segmentos de copia variable, pero éstos pueden no ser exactamente de la misma longitud en todas las muestras. De esta forma, se establece una región consenso entre los CNVs identificados en una misma región basándose en los SNPs de alrededor de similar significancia (casos por un lado y controles por otro) y la frecuencia de la CNVR será la que se compare entre ambos grupos. El estadístico utilizado para la comparativa es un test exacto de Fisher de dos-colas basado en la tabla de contingencia: casos con delección/casos diploides/controles con delección/controles diploides, y lo mismo para las duplicaciones (considerando como delección CN=0, 1 y como duplicación CN=3, 4).

### 3.5.2.6 Análisis de enriquecimiento de rutas funcionales relativas a las mutaciones más significativas del estudio

Otra perspectiva para analizar los datos obtenidos de un genoma completo es centrarse en determinar qué rutas funcionales son las más significativas en vez de analizar todos los marcadores por separado. Para ello existen diferentes programas<sup>323-326</sup> entre los cuales se seleccionó el de GSA-SNP<sup>326</sup>, por su fácil manejo, rapidez y simplificación de resultados. Tras probar diferentes opciones que ofrece el programa y comparar los resultados obtenidos, los parámetros utilizados para los análisis de este proyecto fueron los siguientes: 1) dentro de los SNPs de un mismo gen, utilizar el valor de p del mejor SNP para determinar el valor de significancia que se le dará al gen; 2) establecer el límite superior e inferior de un gen en +/- 20 kb de su posición cromosómica, para el cual se consideraran los SNPs; 3) establecer un número mínimo de 10 genes y un máximo de 200 para definir una ruta funcional. La definición de rutas se realizó de acuerdo a las bases de datos de “Gene Ontology” y “KEGG”.

### 3.5.2.7 Anotación de las variantes analizadas

Para la anotación de los SNPs más significativos de todos los estudios realizados se utilizó el software “SeattleSeq Annotation 137”, basado en la versión de genoma humano NCBI 36/hg18<sup>327</sup> (<http://snp.gs.washington.edu/SeattleSeqAnnotation138/HelpAbout.jsp>). Esta anotación incluye, entre otros campos, los identificadores rs de los SNPs, el nombre de los genes y su número de acceso, la función de la variación (*missense*, *nonsense*...) la posición de la proteína, el cambio aminoacídico y la asociación clínica.

Para anotar la posición génica de los marcadores más asociados se utilizó la herramienta informática SCAN (<http://www.scandb.org/newinterface/about.html>), ya que esta herramienta permite obtener un solo resultado por cada uno de los SNPs en un tiempo razonable para la gran magnitud de datos que se querían anotar. Esta herramienta diferencia entre las siguientes localizaciones que puede tener un SNP dentro de un gen: Intrón, codificador sinónimo, codificador “*missense*”, codificador “*nonsense*”, UTR-3', UTR-5', splice-3', splice-5', cerca del extremo 3' del gen y cerca del extremos 5' del gen.

Para anotar los genes incluidos en los CNVs detectados se utilizó el sistema de anotación del propio programa ParseCNV<sup>322</sup>, el cual está basado en la versión de genoma humano UCSC (BuildHg18).

### **3.5.3 Validación tecnológica de las variantes de mayor relevancia**

#### **3.5.3.1 Selección de las variantes para su validación tecnológica**

Cuando se realiza un estudio de genoma completo es importante contar con una validación tecnológica, es decir, ver si con una tecnología diferente a la utilizada se obtienen los mismos resultados, para evitar así, falsos positivos debidos a la tecnología. Tras analizar los resultados obtenidos, se seleccionaron 48 SNPs y 22 CNVs entre los resultados más significativos.

#### **3.5.3.2 Genotipado y análisis de replicación de los SNPs seleccionados**

Los 48 SNPs elegidos para su replicación, se genotiparon en todas las muestras con diagnóstico (2314 muestras) utilizando la tecnología TaqMan™ (Thermo Fisher Scientific). Las especificaciones de las sondas así como el número de ciclos empleados para cada una de éstas, se encuentran en la tabla suplementaria 1. Para realizar el análisis de replicación se utilizó el mismo método que para calcular la concordancia entre las réplicas del estudio de genoma completo (ver apartado 3.5.2.2). Esta concordancia se calculó para cada uno de los SNPs individualmente y para todos los SNPs en conjunto. Puesto que el número de muestras genotipadas en la validación fue mayor que en el estudio de asociación de genoma completo, se repitieron los tests comparativos de caso/control descritos en el apartado 3.4.3.2 (análisis de SNPs individuales, haplotipos, ventanas corredizas y epistasia).

#### **3.5.3.3 Genotipado y análisis de replicación de los CNVs seleccionados**

Los 22 CNVs seleccionados para su validación, se genotiparon mediante la tecnología TaqMan™ (Thermo Fisher Scientific). Esta tecnología consiste en dos sondas marcadas con fluoróforos y cebadores sin marcar. Una de las sondas (sonda CN) detecta la región del gen o secuencia de interés y está marcada con el fluoróforo FAM-MGB, la otra sonda (sonda de referencia), marcada con VIC-TAMRA, corresponde a la secuencia de una región genómica presente en doble copia en un genoma diploide. Estas dos sondas se cuestionan simultáneamente

en la misma reacción de manera que se registra la fluorescencia emitida por ambos fluoróforos. La sonda de referencia utilizada corresponde al gen *RNAseP* la cual detecta el H1RNA (*Ribonuclease P RNA component H1*) del gen *RPPH1* localizado en el cromosoma 14, citobanda 14q11.2. La posición específica de la sonda es chr.14:20811565 según la versión de genoma Humano NCBI-GR37 y es un amplicón de 87 pb que corresponde al único exón del gen *RPPH1*. Los detalles de las 22 sondas CNs seleccionadas se describen en la Tabla 6:

Tabla 6. Descripción de las sondas CNs seleccionadas para la replicación tecnológica de los CNVs más significativos. Nombre\_sonda=Identificador de la sonda según Thermo Fisher Scientific. Chr=cromosoma. Localización=posición cromosómica de la sonda.

CNV_ID	Nombre sonda	Chr	Gen	Banda citogenética	Localización
CNV1	Hs02189896_cn	Chr1	<i>SCNN1D</i>	1p36.33a	1221589
	Hs07480167_cn	Chr1	<i>GLTPD1</i>	1p36.33a	1262722
	Hs02057266_cn	Chr1	<i>MXRA8</i>	1p36.33a	1290679
CNV2	Hs00980423_cn	Chr11	<i>PTDSS2</i>	11p15.5d	450613
	Hs02600799_cn	Chr11	<i>DRD4</i>	11p15.5d	637563
	Hs00988676_cn	Chr11	<i>EPS8L2</i>	11p15.5c	727639
CNV3	Hs02219266_cn	Chr16	<i>SOX8</i>	16p13.3f	1035900
	Hs00554929_cn	Chr16	<i>BAIAP3</i>	16p13.3e	1391402
	Hs01993673_cn	Chr16	<i>IFT140, TMEM204</i>	16p13.3e	1604907
	Hs00593499_cn	Chr16	<i>MAPK8IP3</i>	16p13.3e	1756421
	Hs00769725_cn	Chr16	<i>MEIOB</i>	16p13.3e	1903092
	Hs00407143_cn	Chr16	<i>ZNF598</i>	16p13.3e	2049800
CNV4	Hs01719675_cn	Chr19	<i>DPF1</i>	19q13.2a	38713054
	Hs02045154_cn	Chr19	<i>KCNK6</i>	19q13.2a	38817513
CNV5	Hs00752842_cn	Chr5	<i>PCDHA</i>	5q31.3a	140186720
	Hs02533973_cn	Chr5	<i>PCDHA</i>	5q31.3a	140201839
CNV6	Hs04347242_cn	Chr8	<i>ARHGAP39</i>	8q24.3h	145767835
	Hs00533492_cn	Chr8	<i>ARHGAP39</i>	8q24.3h	145773336
CNV7	Hs02134493_cn	Chr9	<i>LOC100129722, C9orf173</i>	9q34.3f	140146912
	Hs00295274_cn	Chr9	<i>NELFB (COBRA1)</i>	9q34.3f	140151405
CNV8	Hs05366201_cn	Chr15	<i>GABRB3</i>	15q12b	26854714
	Hs07457448_cn	Chr15	<i>GABRB3</i>	15q12b	26870079

Las reacciones se prepararon según las instrucciones del fabricante en un volumen total de 10 µl, de los cuales 2 µl correspondían a la mezcla de reacción, 0,5 µl a cada una de las sondas,

30 ng de DNA genómico total dispensados en 2  $\mu$ l, y el resto a agua. Las condiciones de amplificación fueron 10 minutos a 95°C seguidos de 40 ciclos de amplificación compuestos por 15 segundos a 95°C y 1 minuto a 60°C. Todas las reacciones se prepararon en placas de 384 pocillos y se amplificaron en el sistema de PCR a tiempo real Quant Studio™ -6 (QS-6). Cada muestra se analizó por triplicado y en cada placa se añadieron 9 muestras negativas (sin DNA) y 2 muestras control (cada una 6 veces) las cuales se incluyeron en todas las placas como controles positivos.

Los resultados obtenidos tras la PCR cuantitativa fueron analizados con el software CopyCaller® Software v2.0 (Thermo Fisher Scientific). En este software se introducen los valores de Ct obtenidos para la sonda de referencia y para la sonda de CN en cada una de las muestras, y el programa calcula un Ct comparativo ( $\Delta\Delta$  CT) determinando de esta forma el número de copias de la región cuestionada.

### **3.6 ESTUDIO DE ASOCIACION DE RASGOS COGNITIVOS CUANTITATIVOS (QTL)**

#### **3.6.1 Análisis de QTLs con los genotipos de los genes candidatos**

En relación a los genes candidatos, la comparativa de genotipos con las variables psicométricas se ha ejecutado con un análisis de covarianza o ANCOVA, el cual se basa en eliminar de la variable dependiente el efecto predicho por una o más covariables mediante regresión lineal múltiple. En nuestro caso, la variable dependiente fue cada una de las 174 medidas de las pruebas realizadas a cada niño, y como covariable se utilizó la edad de cada uno de los niños. De esta forma se pudo eliminar la heterogeneidad causada por la influencia de la edad de los niños en la variable de interés, reduciendo así la diversidad y aumentando en consecuencia la potencia estadística. De esta manera se comparó cada uno de los genotipos correspondientes a los 5 SNPs frente a las variables cuantitativas, así como la presencia/ausencia de cada alelo identificado de los VNTRs. Para estos análisis se han utilizado todas las muestras para las que se disponía de la batería completa de pruebas psicométricas realizada. Quedaron excluidos del análisis aquellos niños con un coeficiente intelectual no verbal inferior a 80.

### 3.6.2 Análisis de QTLs con los genotipos de genoma completo

En cuanto a los datos de genoma completo, se seleccionaron 10 variables de entre la batería completa de pruebas psicométricas realizadas a los participantes, para examinar si existía relación genotipo-fenotipo con alguno de los marcadores estudiados en el panel de genoma completo. En la Tabla 7 se describen las variables seleccionadas (el detalle de cada variable se puede encontrar en la sección 3.2.1).

Tabla 7 Variables de lectura y atencionales seleccionadas para el estudio de QTLs con los genotipos del panel de genoma completo. Se describen las funciones cognitivas asociadas.

Variable		Función cognitiva	Rasgo	
Var1	Tiempo global en identificación de fonema	Conciencia fonológica	Lectura	
Var2	Tiempo global en discriminación silábica	Percepción del habla		
Var3	Tiempo de nombrado rápido compuesto de dibujos y colores	Fluidez de lectura y velocidad de procesamiento		
Var4	Eficiencia de palabras y pseudopalabras	Decodificación de grafema-fonema		
Var5	Aciertos del efecto de transposición	Procesamiento ortográfico		
Var6	Aciertos	Stroop numérico	Atención	
Var7	Tiempo			
Var8	Aciertos			Stroop verbal
Var9	Tiempo			
Var10	Índice conflicto	Alerta, orientación y efecto conflicto (control ejecutivo)		

A diferencia de la asociación caso control descrita en el apartado 3.4.3.2, en este análisis los datos no contienen información de diagnóstico dicotómico, si no variables con medidas cuantitativas. De esta forma, se hizo un análisis de asociación cuantitativo basado en el test asintótico de Wald bajo una distribución t de los datos mediante PLINK. Este análisis se llevó a cabo bajo distintos modelos: lineal, genotípico, dominante y recesivo. PLINK también ofrece la posibilidad de testar la diferencia de asociación de un rasgo cuantitativo entre dos ambientes diferentes o dos grupos en nuestro caso. En dicho análisis simplemente se compara la diferencia entre dos coeficientes de regresión de casos y controles (“*Quantitative trait interaction (GxE)*”).

La visualización detallada de las regiones de interés localizadas en los análisis descritos arriba se efectuó mediante la herramienta LocusZoom v1.1<sup>328</sup>. Esta herramienta expone visualmente la señal de asociación relativa a la posición genómica de los marcadores, además de

presentar el desequilibrio de ligamiento, el patrón de recombinación y la posición de los genes en la región seleccionada.

Asimismo, con los resultados del estudio de asociación lineal, también se realizó un análisis de enriquecimiento de rutas funcionales tal y como se ha descrito en el apartado 3.5.2.6

## **4 RESULTADOS**



## 4.1 PUESTA A PUNTO TECNOLÓGICA

### 4.1.1 Selección de muestras

Como parte del proyecto COEDUCA, el total de muestras de saliva recogidas fue de 5008. Según el origen de nacimiento paterno, materno y del propio niño se excluyeron del análisis un total de 1241 muestras. Así mismo, por su edad adulta (mayor de 16 años) o por no tener realizada la batería completa de pruebas psicométricas se rechazaron otras 328 muestras. El rango de edad de los niños que pasaron todos los criterios de inclusión del estudio fue 6-16 años con un ratio chicos/chicas de 1:0.99. Entre estos sujetos también se excluyeron del estudio 25 muestras porque la calidad del DNA obtenido a partir de su saliva fue insuficiente. En la Tabla 8 se presentan en detalle las cantidades de muestras excluidas según cada criterio:

Tabla 8. Número de muestras descartadas en este estudio y causa de su exclusión.

CAUSA	Nº muestras excluidas
Padre o madre nacido fuera de España	563
Sin dato de origen del padre ni de la madre	289
Sin dato de origen del padre	246
Sin dato de origen de la madre	72
Niño nacido fuera de España	38
Sin dato de origen del niño	33
Adultos o batería de pruebas incompleta	328
Calidad de DNA insuficiente	25

De entre los 3414 individuos que pasaron todos los criterios de inclusión, se definieron los siguientes grupos fenotípicos según los criterios de diagnóstico establecidos:

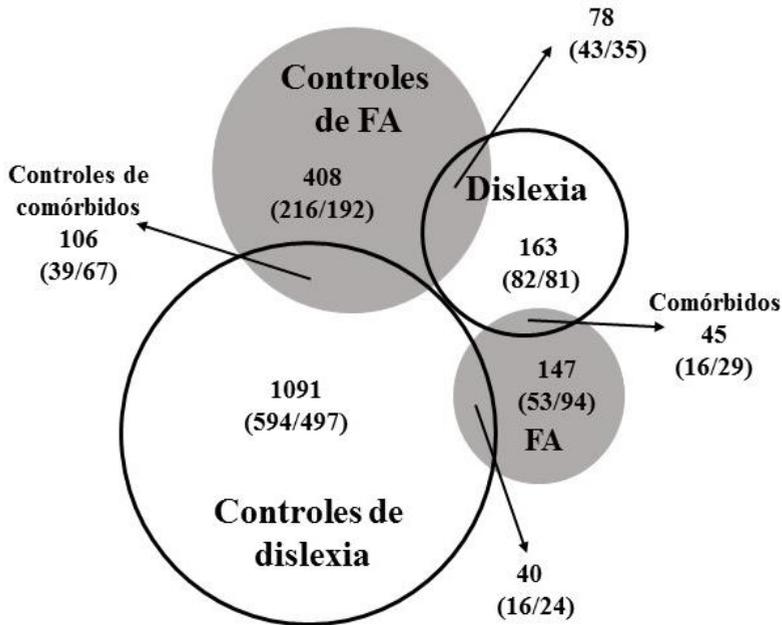


Figura 9. Número total de muestras diagnosticadas para cada trastorno cognitivo estudiado (dislexia y falta de atención –FA–) así como los controles correspondientes. El número de muestras según sexo para cada categoría está detallado entre paréntesis (chicas/chicos).

Del total de muestras que cumplían los criterios de inclusión, se asignó grupo fenotípico a un total de 2078 niños (Figura 9): casos con dislexia ( $D=163+45+78=286$ ), controles para dislexia ( $Ctr_D=1091+106+40=1237$ ), casos con falta de atención ( $FA=147+45+40=232$ ) y controles para falta de atención ( $Ctr_{FA}=408+106+78=592$ ). Tal y como se puede observar, existen muestras que pueden pertenecer a más de un grupo fenotípico. A los 45 individuos que fueron diagnosticados con ambos trastornos, dislexia y falta de atención, se les denominó “comórbidos” (C) y a las 106 muestras que se asignaron como controles tanto para dislexia como para falta de atención se les consideraron “controles para comórbidos” ( $Ctr_c$ ). En la estrategia de genes candidatos realizada en este proyecto, todas las muestras con fenotipo asignado se utilizaron para el estudio de casos y controles. Para el estudio cuantitativo (QTL, del inglés *Quantitative Trait Loci*) poblacional, sin embargo, se utilizaron las 3414 muestras, es decir, todas las muestras que cumplían los criterios de inclusión independientemente del fenotipo.

Según el discriminante utilizado para el diagnóstico de las muestras, hemos observado que los comórbidos obtuvieron los valores más extremos de las variables consideradas (Figura 10). Las muestras representadas en el otro extremo de valores del discriminante corresponden a los controles (línea negra). Interesantemente, la diferencia entre los disléxicos y los controles fue menor que la encontrada entre las muestras con FA y el grupo control.

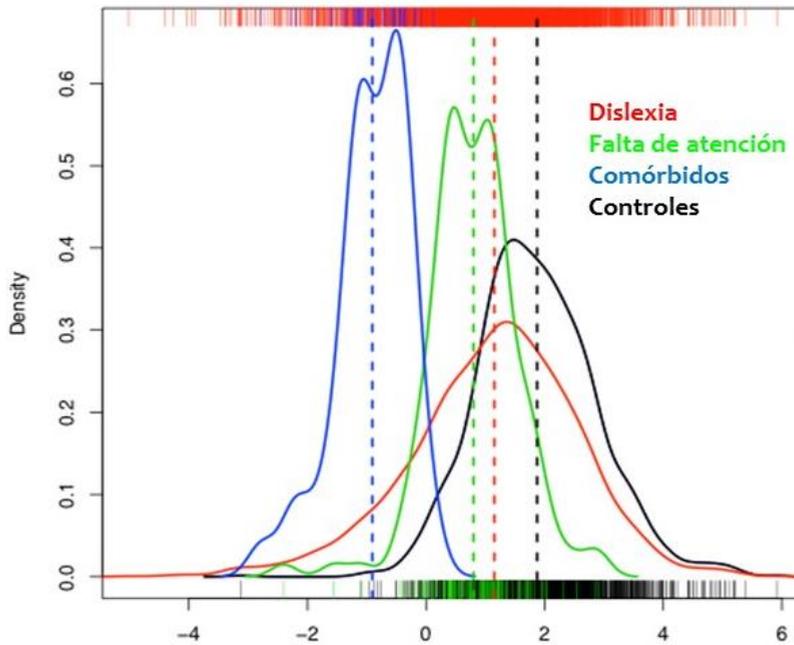


Figura 10. **Distribución de los valores de la función discriminante en los grupos fenotípicos considerados.** Cada histograma se evaluó mediante una estimación de densidad de Kernel. Las muestras comórbidas se consideraron como un grupo fenotípico independiente, por lo que fueron excluidas de los grupos de dislexia y de falta de atención. Las líneas verticales de colores situadas en la parte superior e inferior de la figura muestran los valores individuales del discriminante para las muestras de cada grupo (azul=comórbidos, verde=muestras con FA, rojo=disléxicos y negro=controles). Las líneas discontinuas verticales representan la media de cada una de las distribuciones (azul=comórbidos, verde=muestras con FA, rojo=disléxicos y negro=controles).

#### 4.1.2 Cuantificación de DNA humano y bacteriano de las muestras provenientes de saliva

Puesto que los DNAs que se estudian en este proyecto provienen de saliva, y por lo tanto contienen tanto el DNA genómico de las células humanas como el de bacterias u hongos presentes en la microbiota bucal, antes de comenzar con el genotipado de las muestras, se cuantificó la proporción de DNA humano y DNA bacteriano por qPCR en 30 extractos

provenientes de saliva para establecer los porcentajes de presencia de cada tipo de DNA en nuestra población.

Las rectas patrón utilizadas para cada una de las cuantificaciones, basadas en los resultados de mezclas controladas de DNA con distintos porcentajes de DNA humano y bacteriano comerciales, obtuvieron un índice de correlación ( $r^2$ ) de 0.99. Las proporciones de DNA de origen humano y bacteriano detectadas se resumen en la Tabla 9. El porcentaje de DNA de otro origen se obtuvo restando al total de 100% la proporción cuantificada de los DNAs humano y bacteriano conjunta.

Tabla 9. Proporciones (%) máximas, mínimas, medias y desviación estándar (SD) de DNA humano, bacteriano y de otro origen detectado en los 30 extractos analizados.

Tipo de DNA	Max	Min	Media	SD
Humano	72.8	6.9	26.25	14.28
Bacteriano	4.9	0.5	2.34	1.07
Otro Origen	90.03	13.94	69.6	13.94

#### 4.1.3 Puesta a punto de la tecnología de arrays de genotipado partiendo de DNA extraído de saliva

Como hemos visto en el apartado anterior, la composición del DNA extraído de muestras de saliva no es 100 % humano. Por ello, en este proyecto fue necesario ensayar el genotipado de varias concentraciones de DNA con muestras con diversa proporción de DNA humano/bacteriano. La primera puesta a punto se realizó con los Human660W-quad v1.0 BeadChips, los cuales no se pudieron utilizar en el resto del proyecto porque fueron discontinuados por Illumina Inc.. Para dicha puesta a punto se seleccionaron dos muestras de DNA que tuvieran un % de DNA humano extremo (23.28 y 72.8 %) y se utilizaron 3 concentraciones de partida diferentes a la requerida en el protocolo estándar (200 ng): 250, 300 y 400 ng de DNA total. Si bien cualquier concentración de DNA de partida funcionó satisfactoriamente en base al valor de SNPs genotipados, o *call rate*, el cual fue superior a 0.99 en todos los casos, se seleccionó la concentración de 300 ng ya que su porcentaje de genotipado fue algo superior al resto. De esta forma, se procedió al genotipado de otras 8 muestras con diversos porcentajes de DNA humano partiendo de 300 ng de DNA. El *call rate* obtenido para todas las muestras fue superior a 0.99 excepto para una, independientemente de su % de DNA humano.

Toda esta puesta a punto se repitió con los BeadChips de genotipado de genoma completo de densidad intermedia que estaban disponibles en el momento de comenzar la caracterización de las muestras del estudio, los arrays *Human OmniExpress v1.0* (Illumina Inc.). En este caso, la optimización de la concentración de DNA de partida se hizo en 3 fases. I) En la primera fase se utilizaron 3 concentraciones de DNA de partida (250, 300 y 400 ng) de 4 muestras de distinto porcentaje de DNA humano (7.11/ 12.5/ 24.26/ 38.27 %). II) Para la segunda fase se utilizaron 3 concentraciones de DNA de partida (500, 600 y 700 ng) de 4 muestras de distinto porcentaje de DNA humano (5.33/ 7.11/ 10.33/ 40.74 %). La muestra de 7.11 % de DNA humano fue la misma empleada en la fase 1. III) En la tercera fase de la puesta a punto también se utilizó la muestra de 7.11 % de DNA humano (previamente utilizada en las dos primeras fases) y se repitieron las concentraciones de DNA de partida de 400 y 500 ng, aparte de añadir otras dos intermedias de 430 y 460 ng, para poder determinar si los datos de los tres arrays utilizados eran comparables entre ellos. Los resultados obtenidos una vez normalizados los datos con respecto a esta muestra se detallan en la Figura 11.

El logR ratio (LRR) es un valor proporcional al total de la intensidad de fluorescencia medida con los ensayos Infinium (Illumina Inc.), que refleja el número de copias de cada región cromosómica por cada muestra. Como otro control de calidad durante esta puesta a punto, a partir de los resultados obtenidos se calculó la desviación estándar del LRR para cada una de las muestras, la cual evaluó la calidad de los datos para su posterior análisis de CNVs. Como se puede observar en la Figura 11 los datos obtenidos a partir de las muestras con porcentaje de DNA humano inferior a 10 % no eran de buena calidad para la caracterización de CNVs, independientemente de la concentración de partida que se utilizara en el array (basándonos en el valor de la desviación del LRR). Utilizando una concentración de DNA de partida de 400 ng, la calidad de los datos para el posterior análisis de CNVs resultó adecuada. Por otro lado, el genotipado fue satisfactorio (de acuerdo al valor de *call rate* > 0.99) para cualquier concentración de partida utilizada y con cualquier porcentaje de DNA humano presente en la muestra.

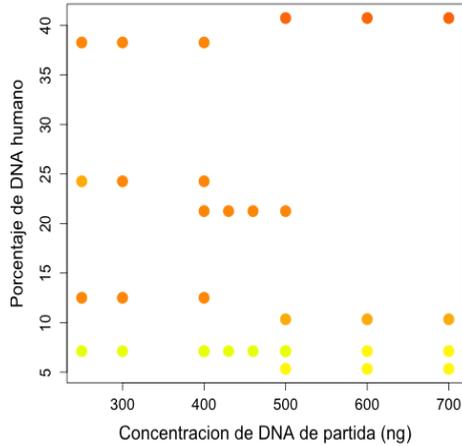


Figura 11. Representación de la desviación estándar del logR ratio (LRR) con respecto a la concentración de partida de DNA utilizada en el chip de genotipado y el porcentaje de DNA humano presente en la muestra. Cada uno de los puntos representa una de las muestras genotipadas. En el eje x se representa la concentración de DNA de partida (ng) utilizada para el genotipado y en el eje Y el porcentaje de DNA humano presente en la muestra. Escala de colores en base a la desviación del LRR de la muestra: tonos amarillos indican un mayor valor de la desviación de LRR, en consecuencia una peor calidad de los datos para el posterior análisis de los CNVs, y los tonos naranjas un valor menor de la desviación del LRR, es decir, una mejor calidad. Resultados normalizados en base a la muestra idéntica utilizada en la misma concentración de DNA de partida para los tres arrays, provenientes de las fases 1, 2 y 3 de la puesta a punto.

#### 4.1.4 Análisis de reproducibilidad

La muestra que contenía un porcentaje de DNA humano de 7.11 % estuvo presente en las 3 primeras pruebas, por lo que se pudo utilizar para calcular la reproducibilidad de resultados entre los distintos chips de *Human OmniExpress v1.0* procesados (ver detalles en la sección 3.5.2.2). De la misma forma, las muestras con porcentaje de DNA humano de 12.5 % y 24.26 % se procesaron en los dos tipos de BeadChips (*Human 660-quad* y *Human OmniExpress v1.0*) por lo que éstas también sirvieron para calcular la reproducibilidad entre los distintos BeadChips Kits. En ambos casos la concordancia fue mayor del 99 %, indicando que tanto entre distintos chips de un mismo kit (*Human OmniExpress v1.0*) como entre los de 2 kits (*Human 660-quad* vs. *Human OmniExpress v1.0*) más del 99 % de los marcadores coincidentes obtuvieron el mismo genotipo.

## 4.2 ESTUDIO DE ASOCIACION DE GENES CANDIDATOS

En el presente estudio, las muestras se diagnosticaron para dislexia y falta de atención. Así mismo, también se seleccionaron muestras control específicas para cada trastorno, de manera que en los estudios de asociación se consideraron los siguientes grupos de acuerdo a su fenotipo:

### Casos

**D:** Muestras diagnosticadas sólo como **disléxicos**.

**FA:** Muestras diagnosticadas sólo con **falta de atención**.

**C:** Muestras diagnosticadas con dislexia y falta de atención (**comórbidos**).

### Controles

**Ctrl\_D:** Muestras caracterizadas como **controles** para el grupo de **disléxicos** (40 de estas muestras también tienen falta de atención).

**Ctrl-P\_D:** Muestras caracterizadas como **controles puros** para el grupo de **disléxicos** (ninguna de estas muestras tiene falta de atención).

**Ctrl\_FA:** Muestras caracterizadas como **controles** para el grupo de **falta de atención** (78 de estas muestras también tienen dislexia).

**Ctrl-P\_FA:** Muestras caracterizadas como **controles puros** para el grupo de **falta de atención** (ninguna de estas muestras tiene dislexia).

**Ctrl\_FA2\*:** Muestras caracterizadas como **controles más estrictos** para el grupo de **falta de atención** (29 de estas muestras también tienen dislexia).

**Ctrl-P\_FA2\*\*:** Muestras caracterizadas como **controles puros más estrictos** para el grupo de **falta de atención** (ninguna de estas muestras tiene dislexia).

**Ctrl\_C:** Muestras caracterizadas como controles validos tanto para el grupo de dislexia como para el grupo de falta de atención (**controles para comórbidos**).

\* Todas las muestras del grupo **Ctrl\_FA2** están incluidas en el grupo **Ctrl\_FA**.

\*\* Todas las muestras del grupo **Ctrl-P\_FA2** están incluidas en el grupo **Ctrl-P\_FA**.

Para el estudio de asociación de casos y controles, teniendo en cuenta los diversos diagnósticos fenotípicos que presentaban las muestras del proyecto, se realizaron diferentes comparativas utilizando distintos grupos de casos y controles. Dichas comparativas se especifican en la Tabla 10.

Tabla 10.- Comparativas de caso/control realizadas y número de muestras consideradas en cada grupo, los cuales están definidos según el recuadro de la página 91.

Casos			Controles	
Grupo	N° muestras		Grupo	N° muestras
D+C	286	vs	Ctr_D	1237
D+C	286	vs	Ctr-P_D	1197
D	241	vs	Ctr-P_D	1197
C	45	vs	Ctr-P_D	1197
C	45	vs	Ctr-P_FA	514
C	45	vs	Ctr-P_FA2	278
C	45	vs	Ctr_c	106
FA+C	232	vs	Ctr_FA	592
FA+C	232	vs	Ctr-P_FA	514
FA+C	232	vs	Ctr_FA2	307
FA+C	232	vs	Ctr-P_FA2	278
FA	187	vs	Ctr-P_FA	514
FA	187	vs	Ctr-P_FA2	278

#### 4.2.1 Asociación de VNTRs con dislexia y falta de atención

Entre las variantes históricamente asociadas con dislexia y/o falta de atención, hay dos variables de número de copia yuxtapuestas (*Variable Number Tandem Repeats* –VNTRs-) en los genes *DRD4* y *DAT1*, las cuales fueron determinadas en este proyecto.

Tras el genotipado de estas variables en varias muestras, se aislaron y clonaron 3 fragmentos correspondientes al VNTR del gen *DRD4* cuya secuenciación verificó la detección de fragmentos de 3, 4 y 7 repeticiones de la región variable de 48 pares de bases (pb). Para el VNTR de *DAT1* fueron 4 los fragmentos aislados y clonados, correspondientes a 7, 9, 10 y 11 repeticiones de la región variable de 40pb. Gracias a los clones obtenidos de estas bandas, se pudieron amplificar cada una de ellas para preparar marcadores de tamaño específico de los alelos presentes en nuestra población para cada VNTR. De esta forma, se establecieron los genotipos de los 2 VNTRs de cada una de las muestras en base a sus perfiles de amplificación resueltos en geles de agarosa y la comparativa de tamaños de los productos de PCR contrastados con los marcadores customizados y los marcadores moleculares comerciales “100bp” y “ $\phi$ x174/HinfI”.

#### 4.2.1.1 Genotipos identificados para el VNTR situado en el gen *DRD4*

Dentro de la población estudiada, para el VNTR del gen *DRD4* se encontraron alelos de 2 a 11 repeticiones, exceptuando la repetición de 10 copias. En la Figura 12 se detallan algunos de los alelos identificados en este estudio. Los alelos predominantes fueron los de 3, 4, 5 y 7 repeticiones (ver Tabla 11-a). Los genotipos más frecuentes se describen en la Tabla 11-b.

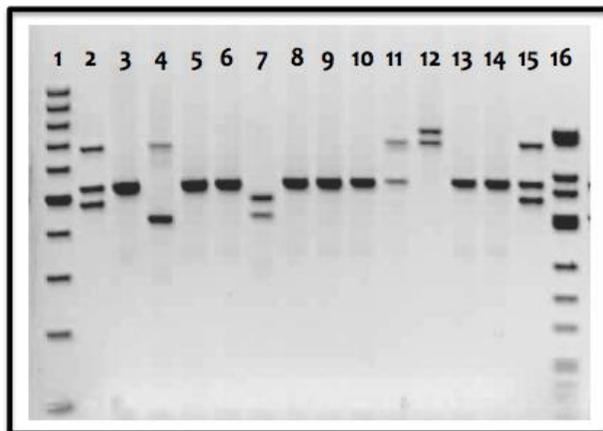


Figura 12. Visualización de productos amplificados del VNTR de 48pb del gen *DRD4* resueltos en gel de agarosa del 2 % y TAE 1x. Las calles 1 y 16 corresponden a marcadores de tamaño molecular de 100bp y  $\phi$ x174/*Hinf*I respectivamente. Las calles 2 y 15 pertenecen al marcador con alelos de 3, 4 y 7 repeticiones (R) elaborado específicamente para este estudio. Las calles 3, 5, 6, 8, 9, 10, 13 y 14 son los resultados de muestras que portan el alelo 04R; la calle 4 corresponde a un heterocigoto con los alelos 02R y 7R; la calle 7 es una muestra con alelos 02R y 03R; la calle 11 con alelos 04R y 07R; y la calle 12 con alelos 7R y 08R.

Tabla 11. a) Alelos y b) genotipos del VNTR situado en el gen *DRD4* encontrados con frecuencia mayor a 0.01 en la población estudiada (3208 muestras). Las variantes que no llegaron a la frecuencia mínima establecida se agrupan bajo la definición "Resto". *NNR* y *NNRZZR*, indican el número de repeticiones de la región variable de 48pb identificada.

a)		b)	
<i>DRD4</i> alelos	Frecuencia	<i>DRD4</i> genotipos	Frecuencia
02R	0.086	02R04R	0.118
03R	0.025	02R07R	0.028
04R	0.679	03R04R	0.035
05R	0.014	04R04R	0.462
07R	0.185	04R05R	0.019
Resto	0.011	04R07R	0.247
		07R07R	0.039
		Resto	0.092

Los resultados de significancia estadística (Test de Fisher) obtenidos de la comparación de presencia/ausencia de cada uno de los alelos y de cada uno de los genotipos observados frente al resto al comparar las poblaciones casos versus controles están descritos en la Tabla 12. En el análisis donde se comparó cada genotipo individual del VNTR del gen *DRD4* frente al resto en casos y controles, el genotipo 04R05R (formado por los alelos de 4 y 5 repeticiones) fue el único que pasó el umbral de significancia de 0.05 en determinadas comparativas, al igual que la presencia del alelo de 5 repeticiones (ver Tabla 12). En el estudio realizado con el grupo de individuos con FA, los valores de significancia fueron menores cuando se eliminaron aquellas muestras con dislexia de dicho grupo (Tabla 12, comparativas FA vs Ctr-P<sub>FA</sub> y FA vs Ctr-P<sub>FA2</sub>). Cabe destacar que en el grupo de comórbidos no se observó ninguna asociación (Tabla 12, C vs Ctr<sub>c</sub>).

Al repetir los análisis sólo con el grupo de chicos, a parte del genotipo 04R05R y la presencia de 05R, también se asoció significativamente el genotipo formado por el alelo de 7 repeticiones en homocigosis (07R07R) en la población diagnosticada para el estudio de dislexia (Tabla 12, comparativas D+C vs Ctr<sub>D</sub> y D vs Ctr-P<sub>D</sub>). En los análisis realizados con el grupo de chicas no se obtuvo ningún valor significativo.

Tabla 12. Resultados de asociación genotípica y de presencia/ausencia alélica del VNTR situado en el gen *DRD4*. E.C.=estudio de comórbidos, Variante=genotipo o alelo analizado, Grupo=grupo de muestras analizadas. Casos/Controles=grupo de muestras utilizadas como casos o controles según la definición del recuadro de la página 91. N° cas.=número de casos totales, N° ctr.=número de controles totales. El número de muestras según sexo para cada categoría está detallado entre paréntesis (chicas/chicos). Sólo se muestran los valores de significancia <0.05.

		Estudio de Dislexia		E.C	Estudio de Falta de Atención			
Casos		D+C	D	C	FA+C	FA+C	FA	FA
Controles		Ctr_D	Ctr-P_D	Ctr_c	Ctr_FA	Ctr_FA2	Ctr-P_FA	Ctr-P_FA2
Variante	Grupo							
04R05R	Todos	0.0296			0.0218	0.0359	0.018	0.017
	Chicos				0.0089	0.031	0.0059	0.0235
	Chicas							
07R07R	Todos							
	Chicos	0.0277	0.014					
	Chicas							
05R	Todos						0.0367	0.0242
	Chicos	0.0206			0.0357	0.031	0.0144	0.0235
	Chicas							
N° cas.		286 (141/145)	241 (125/116)	45 (16/29)	232 (85/147)	232 (85/147)	187 (69/118)	187 (69/118)
N° ctr.		1237 (649/588)	1197 (633/564)	106 (39/67)	592 (298/294)	307 (144/163)	514 (255/259)	278 (125/153)

Por otra parte, al analizar los alelos largos (de 5 a 12 repeticiones), frente a los cortos (de 2 a 4 repeticiones), se observó una diferencia significativa entre las muestras disléxicas y sus controles (D vs Ctr-P<sub>D</sub>), tanto utilizando la población completa como considerando sólo a los chicos (valores de p 0.04367 y 0.04579, respectivamente).

Si bien hay valores estadísticamente significativos mostrando una posible asociación de estos genotipos/alelos con los trastornos cognitivos, es preciso mencionar que el número de muestras con dichas variantes era muy pequeño, sobre todo en los análisis realizados con el grupo de chicos. Sin embargo, en la mayoría de los tests significativos la frecuencia de dichos alelos y genotipos llegó a pasar el valor de 0.01 (ver tabla suplementaria 2 con las frecuencias de cada uno de los genotipos y alelos).

#### 4.2.1.2 Genotipos identificados para el VNTR situado en el gen *DAT1*

Para el VNTR del gen *DAT1* los alelos observados fueron de 3 a 12 repeticiones, exceptuando los alelos de 4 y 5 repeticiones. En la Figura 13 se detallan las bandas observadas de los diferentes fragmentos de repeticiones del VNTR. Las variantes predominantes fueron los alelos de 9 y 10 repeticiones, así como los genotipos formados por ambos (ver Tabla 13).

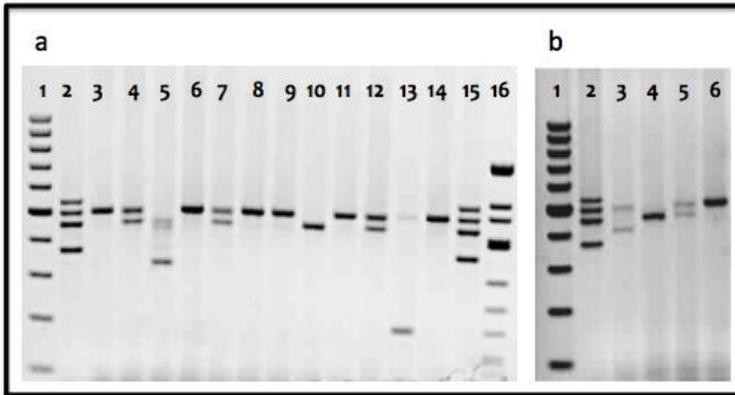


Figura 13. Visualización de productos amplificados del VNTR situado en el gen *DAT1* resueltos en gel de agarosa del 2 % y TAE 1x. a) Las calles 1 y 16 corresponden a marcadores de tamaño molecular de 100bp y al  $\phi$ x174/*Hinf*I, respectivamente. Las calles 2 y 15 son el marcador con alelos de 7, 9, 10 y 11 repeticiones (R) elaborado en el laboratorio. Las calles 3, 6, 8, 9, 11 y 14 tienen muestras homocigotas para el alelo 10R. Las calles 4, 7 y 12 son muestras con los alelos 9R y 10R. La calle 5 con alelos 06R y 09R. La calle 10 con el alelo 09R. La calle 13 heterocigoto con los alelos 03R y 10R. b) La calle 1 corresponde a marcadores de tamaño molecular de 100bp. La calle 2 corresponde al marcador customizado por nosotros. La calle 3 resuelve una muestra con los alelos 08R y 10R. La calle 4 con el alelo 09R. La calle 5 con los alelos 09R y 10R y la calle 6 con el alelo 10R.

Tabla 13. a) Alelos y b) genotipos del VNTR situado en el gen *DAT1* encontrados con frecuencia mayor a 0.01 en la población estudiada (3311 muestras). Las variantes que no llegaron al umbral de frecuencia establecido se agruparon bajo la definición "Resto". Los códigos **NNR** y **NNRZZR** indican el número de repeticiones de la región variable de 48 pb identificada para los alelos y los genotipos, respectivamente.

a)		b)	
<i>DAT1</i> alelos	Frecuencia	<i>DAT1</i> genotipos	Frecuencia
09R	0.354	09R09R	0.128
10R	0.63	09R10R	0.435
Resto	0.016	10R10R	0.403
		Resto	0.03

Los resultados de significancia estadística (Test de Fisher) obtenidos de la comparación de la presencia de cada uno de los alelos y de cada uno de los genotipos frente al resto, están descritos en la Tabla 14. El resultado más relevante se observó en el estudio del grupo de disléxicos, tanto con las muestras comórbidas como sin ellas (Tabla 14, comparativas D+C vs Ctr\_D y D vs Ctr-P\_D), ya que se encontraron dos genotipos asociados dependiendo del grupo de muestras comparado (el genotipo formado por los alelos de 10 y 11 repeticiones y el formado por 9 y 11 repeticiones). La presencia de los alelos 10R y 11R también mostraron valores de significancia en ambos tests dependiendo del grupo de muestras analizado, es decir, mientras que el alelo 10R se asoció únicamente al grupo de chicas, el alelo 11R lo hizo al grupo total así como al formado sólo por chicos. En las comparativas correspondientes al estudio de falta de atención así como en el estudio de comórbidos no se observó ninguna asociación (ver tabla suplementaria 3 con las frecuencias de cada uno de los genotipos y alelos).

Tabla 14. Resultados significativos de asociación genotípica y de presencia/ausencia alélica del VNTR situado en el gen *DAT1*. Variante=genotipo o alelo analizado, Grupo=grupo de muestras analizadas. Casos/Controles=grupo de muestras utilizadas como casos o controles según la definición del recuadro de la página 91. N° cas.=número de casos totales, N° ctr.=número de controles totales. El N de chicos y chicas de cada grupo se muestra entre paréntesis respectivamente. Sólo se muestran los valores de significancia <0.05.

		Estudio de Dislexia	
		D+C	D
		Ctr_D	Ctr-P_D
Variante	Grupo		
10R11R	Todos	0.0323	0.0189
	Chicos	0.0014	0.0008
	Chicas		
09R11R	Todos		
	Chicos		
	Chicas	0.0373	0.0284
10R	Todos		
	Chicos		
	Chicas	0.0338	0.0349
11R	Todos	0.0139	0.017
	Chicos	0.0475	0.07414
	Chicas		
N° cas.		286 (141/145)	241 (125/116)
N° ctr.		1237 (649/588)	1197 (633/564)

## 4.2.2 Estudio de polimorfismos de nucleótido único (SNPs)

### 4.2.2.1 Características de los marcadores analizados

Las frecuencias genotípicas y alélicas de los 7 marcadores de polimorfismos de nucleótido único (en inglés “*Single Nucleotide Polymorphism*” o SNP) encontradas en la población estudiada, según los criterios de inclusión establecidos, están resumidas en las Tabla 15 y Tabla 16. Los marcadores rs13306306 y rs72554632 no superaron la frecuencia del 5% en cuanto a su alelo minoritario por lo que se descartaron del estudio.

Tabla 15. Frecuencias alélicas de los SNPs para la población estudiada. Se especifica el gen, el identificador del marcador (SNP) y el cromosoma (Chr) donde se ubica dicho marcador. A1=nucleótido del alelo mayoritario, A2=nucleótido del alelo minoritario, fq=frecuencia, N=número de muestras caracterizadas.

Gen	SNP	Chr	A1	A2	fq A1	fq A2	N
KIAA0319	rs4504469	6	C	T	0.66	0.34	3334
FOXP2	rs12533005	7	C	G	0.46	0.54	3340
DCDC2	rs2274305	6	C	T	0.61	0.39	3340
DYX1C1	rs57809907	15	A	C	0.1	0.9	3321
DBH	rs1611115	9	C	T	0.78	0.22	3316
COMT	rs4680	22	A	G	0.45	0.55	3338
MAOA	rs6323	X	G	T	0.26	0.74	3339

Tabla 16. Frecuencias genotípicas de los SNPs para la población estudiada. Se especifica el gen, el identificador del marcador (SNP) y el cromosoma (Chr) donde se ubica dicho marcador. Gx=genotipo, fq=frecuencia, N=número de muestras caracterizadas.

Gen	SNP	Chr	G1	G2	G3	fq G1	fq G2	fq G3	N
KIAA0319	rs4504469	6	CC	CT	TT	0.43	0.46	0.12	3334
FOXP2	rs12533005	7	CC	CG	GG	0.22	0.47	0.31	3340
DCDC2	rs2274305	6	CC	CT	TT	0.37	0.49	0.14	3340
DYX1C1	rs57809907	15	AA	AC	CC	0.03	0.15	0.82	3321
DBH	rs1611115	9	CC	CT	TT	0.61	0.34	0.04	3316
COMT	rs4680	22	AA	AG	GG	0.21	0.5	0.3	3338
MAOA	rs6323	X	GG	GT	TT	0.16	0.19	0.65	3339

#### 4.2.2.2 Asociación de los marcadores candidatos en la población estudiada

Los resultados de asociación de los SNPs así como del análisis de interacción dos a dos, están resumidos en las Tabla 17 y Tabla 18, respectivamente. Las señales de asociación (mostradas con el valor de  $p$ ) variaron dependiendo de los grupos de muestras comparados (Tabla 10). El nivel de asociación, estableciendo como mayor significancia un valor menor de  $p$ , fue distinto al analizarse un solo SNP o al analizar la interacción entre cada dos SNPs. Se obtuvo un resultado de gran interés al comprobar que los SNPs de los genes *KIAA0319* y *FOXP2* se asociaban significativamente en las comparativas D+C vs Ctr\_D y D+C vs Ctr-P\_D, pero esta asociación desaparecía cuando se excluían las 45 muestras comórbidas del análisis (D vs Ctr-P\_D), volviendo a manifestarse cuando las muestras comórbidas se analizaron por separado (Tabla 17, C vs Ctr-P\_D/ C vs Ctr-P\_FA/ C vs Ctr-P\_FA2 y C vs Ctr\_C). Se observó un patrón similar al considerar la interacción entre los marcadores de los genes *DCDC2* y *DYX1C1*, los cuales presentaron una asociación significativa en el grupo de falta de atención (Tabla 18, FA+C vs Ctr\_FA/ FA+C vs Ctr-P\_FA y FA+C vs Ctr\_FA2), disminuyendo su significancia al excluir las muestras comórbidas (Tabla 18, FA vs Ctr-P\_FA y FA vs Ctr-P\_FA2) y asociándose de nuevo en los tests de C vs Ctr-P\_FA y C vs Ctr\_C (comórbidos). *DCDC2* también mostró asociación en combinación con *KIAA0319* en el grupo de disléxicos independientemente de la inclusión/exclusión de las muestras comórbidas (Tabla 18, D+C vs Ctr\_D/ D+C vs Ctr-P\_D y D vs Ctr-P\_D). El marcador rs4680, situado en el gen *COMT*, mostró una asociación significativa en el análisis genotípico en las comparativas D+C vs Ctr\_D/ D+C vs Ctr-P\_D y C vs Ctr-P\_D (valores de  $p$  de 0.021, 0.019 y 0.017, respectivamente; ver tabla suplementaria 4). Además, al filtrar el análisis por género, este marcador no apareció asociado a ninguno de los tests realizados, sin embargo, aun no llegando al nivel de significancia, se observó una tendencia interesante en el análisis recesivo del test D vs Ctr-P\_D en el grupo de los chicos (valor de  $p$  0.07, ver tabla suplementaria 5). Para el análisis del marcador rs6323 del gen *MAOA*, situado en el cromosoma X, sólo se utilizó el grupo de chicas. Los resultados mostraron un valor significativo en los análisis recesivos de las comparativas FA+C vs Ctr\_FA/ FA+C vs Ctr-P\_FA / FA+C vs Ctr\_FA2 y FA+C vs Ctr-P\_FA2 (valores de  $p$  0.023, 0.022, 0.014 y 0.033 respectivamente, ver tabla suplementaria 6). En este caso, no pudimos comparar el efecto de las muestras comórbidas con respecto a este marcador ya que las comparativas del modelo recesivo en estos grupos de muestras (comórbidos vs controles) no se llevaron a cabo para *MAOA* puesto que el tamaño de muestra era demasiado pequeño.

Tabla 17. Resultados de asociación alélica de SNPs individuales. Chr=cromosoma, A1=alelo 1, A2=alelo 2. Casos/Controles=grupo de muestras utilizadas como casos o controles según la definición del recuadro de la página 91. N° cas.=número de casos, N° ctr.=número de controles. Sólo se muestran los valores de significancia <0.05.

					Estudio de Dislexia			Estudio de Comórbidos				Estudio de Falta de Atención					
					D+C	D+C	D	C	C	C	C	FA+C	FA+C	FA+C	FA+C	FA	FA
					Contr_D	Contr-P_D	Contr-P_D	Contr-P_D	Contr-P_FA	Contr-P_FA2	Contr_c	Contr_FA	Contr-P_FA	Contr_FA2	Contr-P_FA2	Contr-P_FA	Contr-P_FA2
Chr	SNP	Gen	A1	A2													
6	rs4504469	KIAA0319	T	C	0.01	0.0084		0.0025	0.0032	0.0037	0.01						
7	rs12533005	FOXP2	C	G	0.018	0.0191		0.0079	0.0015	0.0044	0.01						
6	rs2274305	DCDC2	T	C													
15	rs57809907	DYX1C1	A	C													
9	rs1611115	DBH	T	C													
22	rs4680	COMT	A	G													
X	rs6323	MAOA	G	T													
N° cas.					286	286	241	45	45	45	45	232	232	232	232	187	187
N° ctr.					1237	1197	1197	1197	514	278	106	592	514	307	278	514	278

Tabla 18. Resultados de asociación de la interacción de SNPs dos a dos. Gen 1=gen dónde se ubica el SNP 1, Gen 2=gen dónde se ubica el SNP 2. Casos/Controles=grupo de muestras utilizadas como casos o controles según la definición del recuadro de la página 91. \*=fuera del equilibrio de Hardy-Weinberg. Solo se muestran los valores de significancia <0.05. Ver tabla 17 para ubicación de los genes y número de muestras empleadas.

				Estudio de Dislexia			Estudio de Comórbidos				Estudio de Falta de Atención					
Casos				D+C	D+C	D	C	C	C	C	FA+C	FA+C	FA+C	FA+C	FA	FA
Controles				Ctr_D	Ctr-P_D	Ctr-P_D	Ctr-P_D	Ctr-P_FA	Ctr-P_FA2	Ctr_C	Ctr_FA	Ctr-P_FA	Ctr_FA2	Ctr-P_FA2	Ctr-P_FA	Ctr-P_FA2
SNP 1	Gen 1	SNP 2	Gen 2													
rs2274305	DCDC2	rs4504469	KIAA0319	<b>0.0014</b>	<b>0.0014</b>	<b>0.0026</b>										
rs2274305	DCDC2	rs12533005	FOXP2													
rs2274305	DCDC2	rs1611115	DBH													
rs2274305	DCDC2	rs57809907	DYX1C1				<b>0.0151*</b>	<b>0.0237</b>		<b>0.0368</b>	<b>0.0039</b>	<b>0.0056</b>	<b>0.0164</b>		<b>0.0215</b>	
rs2274305	DCDC2	rs4680	COMT													
rs4504469	KIAA0319	rs12533005	FOXP2													
rs4504469	KIAA0319	rs1611115	DBH													
rs4504469	KIAA0319	rs57809907	DYX1C1													
rs4504469	KIAA0319	rs4680	COMT													
rs12533005	FOXP2	rs1611115	DBH													
rs12533005	FOXP2	rs57809907	DYX1C1													
rs12533005	FOXP2	rs4680	COMT													
rs1611115	DBH	rs57809907	DYX1C1													
rs1611115	DBH	rs4680	COMT													
rs57809907	DYX1C1	rs4680	COMT													

### 4.3 ESTUDIO DE ASOCIACION DE GENOMA COMPLETO

#### 4.3.1 Consecuencias del uso de diferentes versiones de arrays de genotipado de genoma completo

El estudio de asociación de genoma completo (GWAS) de este proyecto se realizó en varias etapas, bien dependiendo de las fechas de recepción de muestras o bien de las definiciones de fenotipos. Compañías como Illumina Inc. tienen ofertados muchos arrays para este objetivo, los cuales se actualizan frecuentemente. Debido a esta situación en el presente proyecto tuvimos que utilizar dos versiones distintas de un mismo chip, de forma que las primeras 606 muestras se procesaron con la versión 1.0 del array *Human OmniExpress*, y las siguientes 1124 se genotiparon con la versión 1.1. En ambas versiones, se utilizó un mismo DNA como réplica técnica (un total de 8 réplicas para la versión 1.0 de los arrays y un total de 11 réplicas para la versión 1.1, repartidas en 8 y 11 chips diferentes, respectivamente) con el objetivo de verificar que los resultados eran replicables entre chips de una misma versión y entre chips de distintas versiones. Los resultados obtenidos para esta réplica técnica indicaron una concordancia superior al 99 %, es decir, más del 99 % de los SNPs resultaron en el mismo genotipo con cualquiera de los chips y versión de panel de genotipado.

Al estudiar a fondo qué SNPs se genotiparon diferentemente entre las 2 versiones de BeadChips, observamos que había varios marcadores que estaban decodificados de manera reversa-complementaria para ambas versiones. Es decir, el alelo que en la versión 1.0 estaba codificado con fluorescencia roja, en la versión 1.1 lo estaba con verde, y por lo tanto las frecuencias alélicas resultaron opuestas. Si bien la compañía Illumina niega haber cambiado esta codificación, identificamos claramente algunos de los SNPs codificados de esta forma opuesta (ver Figura 14 con tres ejemplos de este evento). Todos los SNPs decodificados de distinta manera fueron descartados para los análisis. También se inspeccionaron visualmente cada uno de los gráficos de los SNPs con resultado significativo en alguno de los análisis (análisis individual, por grupos o interacción epistática) para verificar si correspondían a alguno con decodificación complementaria o con una clusterización dudosa.

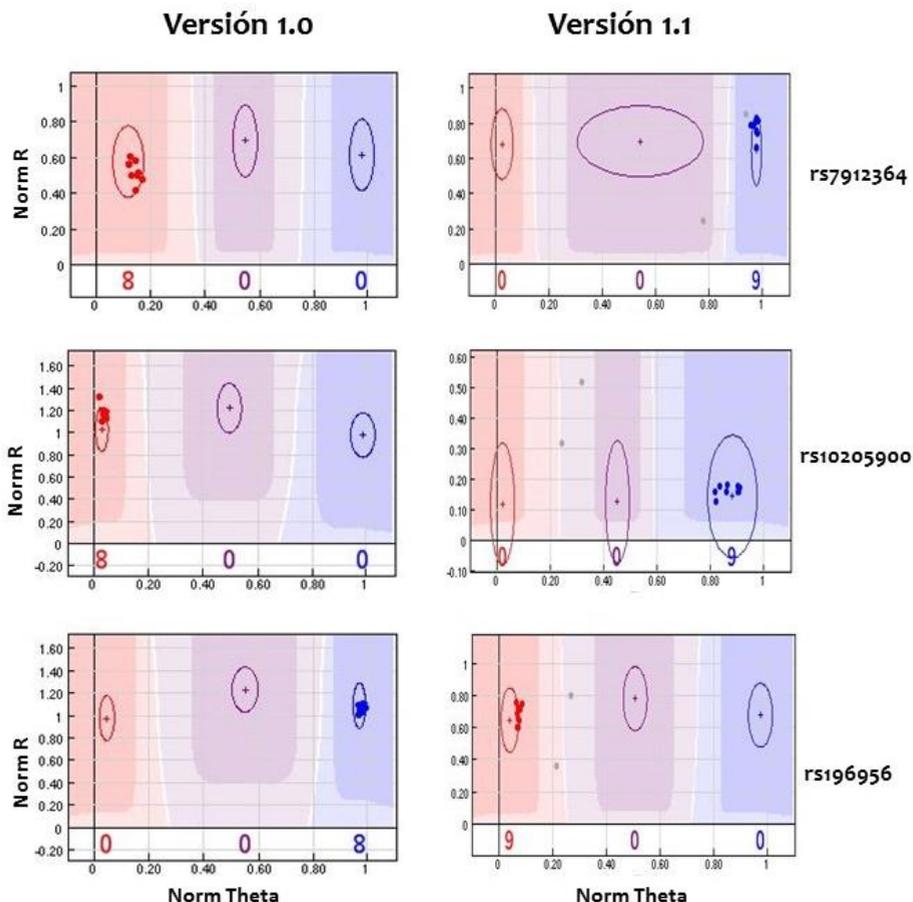


Figura 14. Representaciones gráficas de los clusters genotípicos generados para los SNPs rs7912364, rs10205900 y rs196956, en base a los genotipos obtenidos de 8 y 11 réplicas técnicas con la versión 1.0 y la 1.1 de los arrays *Human OmniExpress*, respectivamente. En rojo muestras con genotipo AA, en morado con AB y en azul con genotipo BB. La intensidad de la señal se muestra en el eje Y (Norm R) y la frecuencia alélica en el eje X (Norm Theta).

Otra de las dificultades derivada del uso de dos versiones de arrays fue la imposibilidad de realizar un análisis conjunto de CNVs debido a la diferencia de calidad de datos obtenida en cada caso. En base a los valores de genotipado de las réplicas técnicas, se evaluaron 3 parámetros para determinar dicha calidad: 1) LRR SD, que corresponde a la variación de la desviación estándar del LRR. Una mayor desviación indica una peor calidad de datos. 2) BAF *drift*, el cual tiene en cuenta los patrones de la frecuencia del alelo B (*B Allele Frequency* o BAF), es decir, si estos valores son cercanos bien a 0, 0.5 ó a 1. La BAF describe la contribución relativa del alelo B al total de la señal de intensidad, por lo que refleja la composición alélica. Un mayor desvío indica una peor calidad de datos. 3) "LRR Waviness Factor" corresponde al patrón que forma el valor de LRR a lo largo de

un cromosoma, el cual puede estar formado por una línea recta en torno a 0, o puede formar diferentes “olas” hacia arriba (en torno a 1) o hacia abajo (en torno a -1). Un valor de waviness cercano a 0 indica una mejor calidad de los datos. Estos resultados se resumen en la Figura 15, donde se puede comprobar que la calidad de los datos fue mejor para las muestras genotipadas en la versión 1.1 en cuanto a los tres parámetros.

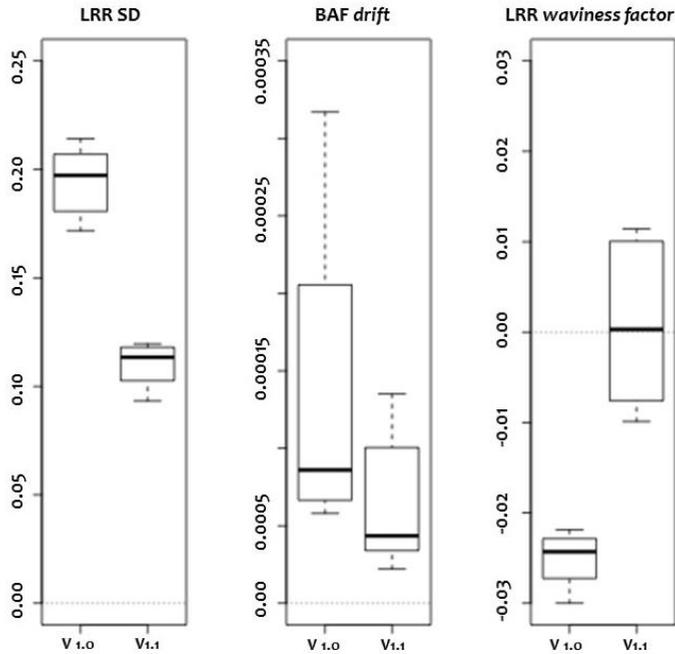


Figura 15. Representación gráfica de los parámetros de calidad calculados para las muestras genotipadas con ambas versiones de arrays de *Human OmniExpress* (v1.0 y v1.1). Gráficos correspondientes a LRR SD, BAF drift y al LRR waviness factor, respectivamente.

Al correlacionar los valores de LRR SD con los del BAF drift (Figura 16-a) y con los valores del LRR waviness factor (Figura 16-b), se pudo observar que las muestras formaban dos grupos diferenciados, correspondientes a las muestras genotipadas con la versión 1.0 y con la versión 1.1. Los valores más cercanos a 0 correspondieron a la versión 1.1 en ambos casos.

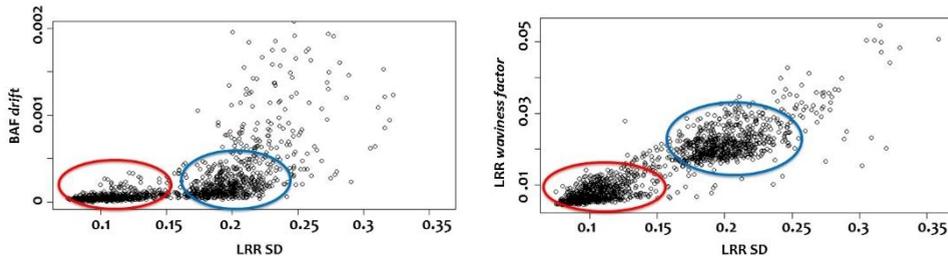


Figura 16. Comparativa de la calidad de genotipado obtenida con las dos versiones de chips *Human OmniExpress* (Illumina Inc.) utilizados en este proyecto. Los puntos dentro de los círculos rojos corresponden a las muestras genotipadas con la versión 1.1 de los arrays y los englobados en los círculos azules, a las genotipadas con la versión 1.0. En el gráfico de la izquierda se muestra la correlación entre la LRR SD y el BAF *drift* y, en el de la derecha, la correlación entre la LRR SD y el LRR *waviness factor* de las muestras genotipadas.

## 4.3.2 Estudio de asociación caso/control

### 4.3.2.1 Detalles de las poblaciones analizadas en genoma completo

Se genotiparon un total de 1730 muestras con los arrays *Human OmniExpress* correspondientes a los grupos que figuran en la Tabla 19. Tras descartar las muestras y los SNPs según los criterios de calidad descritos en los apartados 3.2.3 y 3.5.2.3, los datos analizados finalmente se resumen en esa misma tabla (Tabla 19).

Tabla 19. Resumen de muestras consideradas. Número de muestras genotipadas con cada una de las versiones de Arrays *Human OmniExpress*, así como número de muestras y de marcadores considerados en cada uno de los estudios tras excluir las muestras y marcadores oportunos (número de muestras separado por diagnóstico). Nota: Los 106 controles comórbidos (Ctr\_c) genotipados con la v1.1 están contabilizados en los tres estudios.

ESTUDIO	OmniExpress v1.0		OmniExpress v1.1		N° muestras y SNPs finales considerados		
	Casos	Controles	Casos	Controles	Casos	Controles	N° SNPs
Dislexia	184	390	61	466	238	841	642617
FA	0	0	137	356	136	352	646608
Comórbidos	32	0	13	106	45	102	641651

Con el objetivo de tener una muestra lo más homogénea posible en cuanto a su fondo genético, se descartaron 22, 5 y 4 *outliers* para los estudios de dislexia, falta de atención y comórbidos, respectivamente. La similitud genética entre las muestras restantes fue visualizada mediante la representación de los dos primeros componentes principales (ver Figura 17).

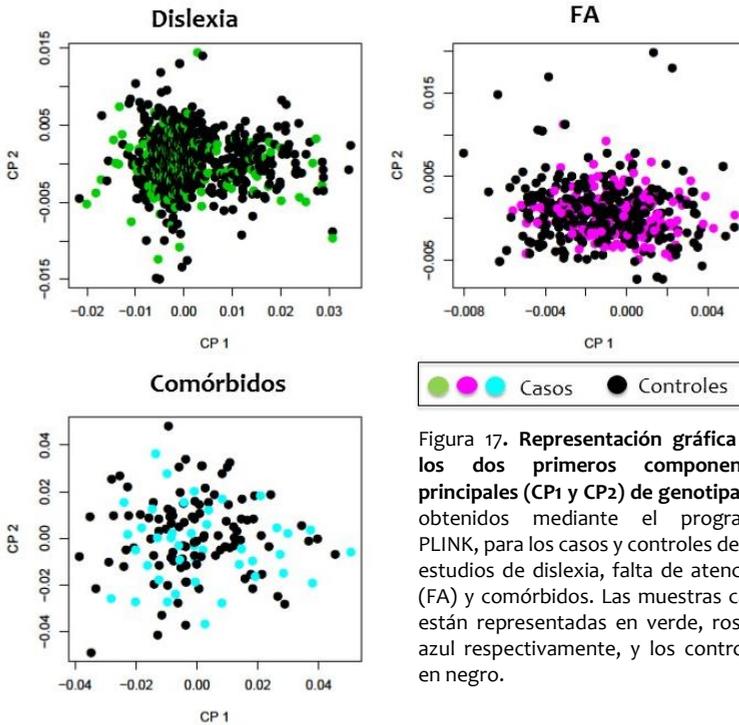


Figura 17. Representación gráfica de los dos primeros componentes principales (CP1 y CP2) de genotipado, obtenidos mediante el programa PLINK, para los casos y controles de los estudios de dislexia, falta de atención (FA) y comórbidos. Las muestras caso están representadas en verde, rosa y azul respectivamente, y los controles en negro.

#### 4.3.2.2 Comparación de los resultados obtenidos para los genes candidatos a partir de los datos de genoma completo

A partir de los genotipos obtenidos mediante los arrays de genoma completo, se estudiaron en detalle los resultados de asociación de todos los SNPs ubicados en 8 de los 9 genes candidatos de este estudio, ya que el panel de genotipado *Human OmniExpress* no contenía ningún marcador del gen *DRD4*.

El grado de desequilibrio de ligamiento (*Linkage Disequilibrium* o LD) entre los marcadores caracterizados para el gen *DCDC2* y el gen *KIAA0319* (ambos genes ubicados en el cromosoma 6) en nuestra población se muestra en la Figura 18. Estos dos genes representan dos bloques de desequilibrio independientes. Además, el grado de LD entre los dos SNPs candidatos en este proyecto es prácticamente nulo (círculo verde de la Figura 18, LD=7).

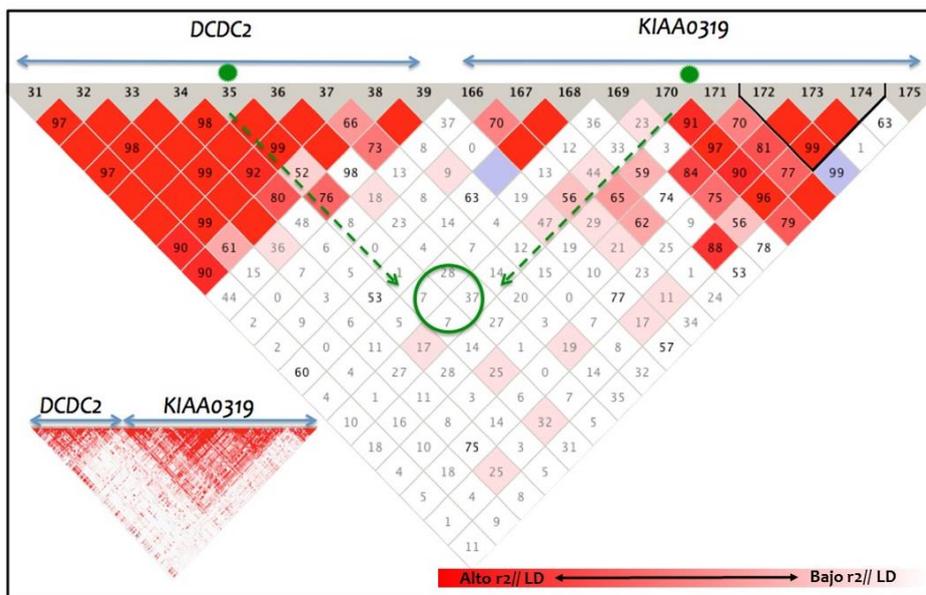


Figura 18. Representación gráfica del patrón de desequilibrio de ligamiento entre los marcadores situados en los genes *DCDC2* y *KIAA0319* estimado para el análisis de dislexia (238 casos vs. 841 controles). En el triángulo más pequeño (abajo a la izquierda) se representan todos los marcadores situados en estos dos genes mientras que en el triángulo principal se han seleccionado solo los marcadores ubicados cerca de los dos SNPs candidatos de este proyecto (números de la franja gris). Los puntos verdes están situados en las posiciones de los SNPs rs2274305 y rs4504469, respectivamente. El valor del grado de desequilibrio entre dos marcadores está indicado por el número que aparece en el recuadro dónde se cruzan los dos marcadores a comparar, el cual se representa también por la intensidad mayor o menor del color rojo de los cuadros, siendo los tonos más intensos los que tienen un mayor grado de LD. El círculo verde enmarca el valor de LD entre estos dos marcadores.

Los resultados de los análisis de asociación de marcadores individuales correspondientes a los modelos alélico y genotípico para cada uno de los estudios (correspondientes a las comparativas D vs Ctr-P<sub>D</sub> (dislexia), FA vs Ctr-P<sub>FA</sub> (FA) y C vs Ctr<sub>C</sub> (comórbidos)) se encuentran detallados en la Figura 19. Tal y como se observa, en el estudio de dislexia hay marcadores de los genes *KIAA0319* y *FOXP2* que muestran asociación según el modelo alélico y el genotípico, incluyendo también SNPs en *DAT1* y *COMT* en este último modelo. En el estudio de falta de atención en cambio, son algunos marcadores ubicados en los genes *DCDC2*, *KIAA0319*, *FOXP2* y *DYX1C1* los que mayor significancia presentan en el modelo alélico y sólo los dos primeros genes pasan el umbral de significancia en el modelo genotípico. En el estudio con individuos comórbidos se identifican un número mayor de marcadores significativos, los cuales se ubican en todos los genes candidatos excepto en *DYX1C1* y *MAOA* en ambos modelos genéticos considerados. (Los

identificadores de los marcadores y los valores de  $p$  obtenidos bajo los distintos modelos genéticos se encuentran en la tabla suplementaria 7).

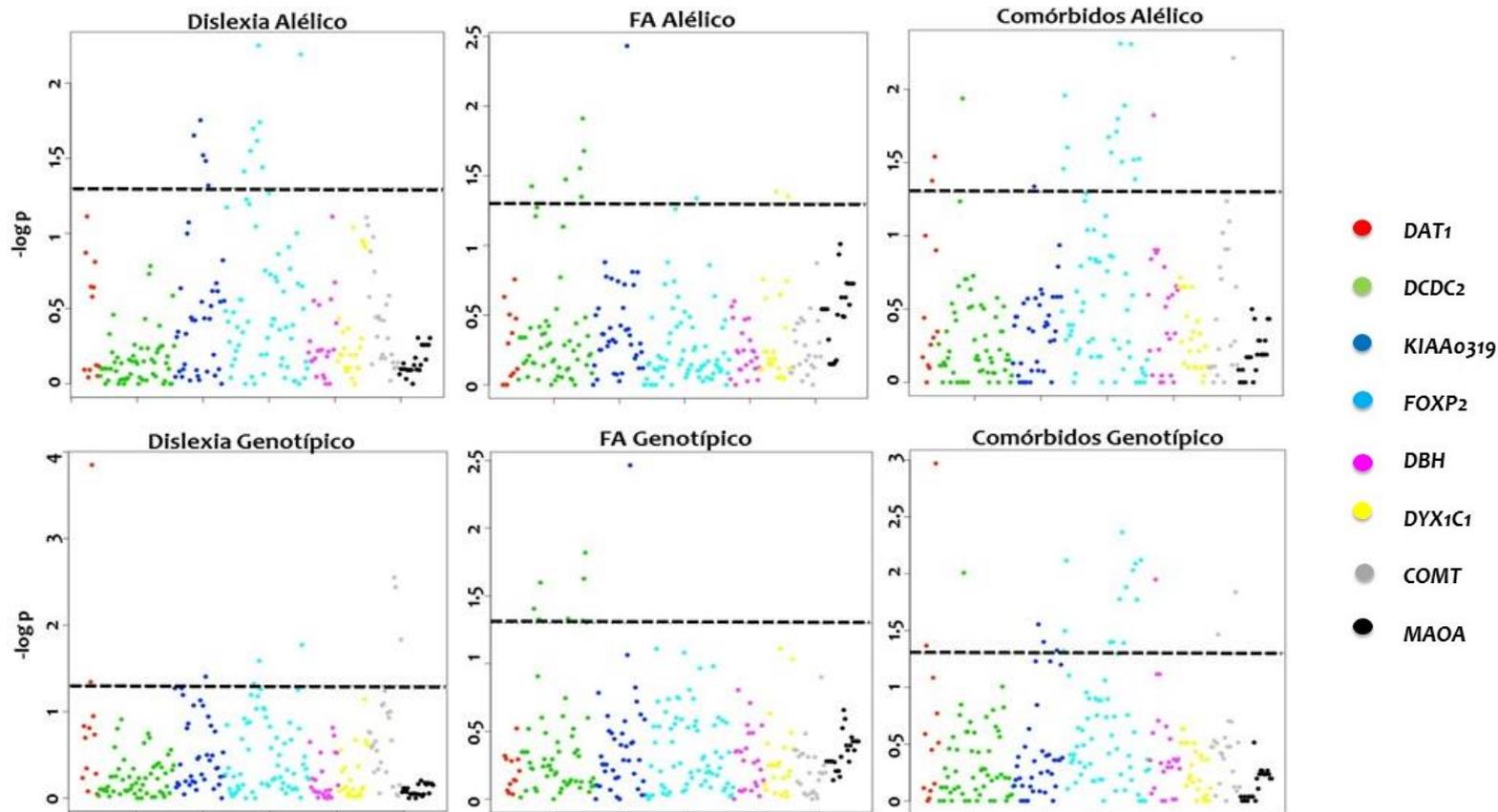


Figura 19. Representación gráfica de los valores de asociación en escala logarítmica ( $\log_p$ ) de todos los marcadores del panel de *Human OmniExpress* (Illumina Inc.) ubicados en los 8 genes candidatos (cada uno con puntos de distinto color). La línea discontinua indica el valor de  $p=0.05$ . Los marcadores están ordenados en el eje X según su posición cromosómica.

En la asociación de **SNPs independientes**, considerando únicamente los tres tests que contenían los individuos con diagnósticos puros (Figura 19), es decir, las comparativas D vs Ctr-P<sub>D</sub> (dislexia), FA vs Ctr-P<sub>FA</sub> (FA) y C vs Ctr<sub>C</sub> (comórbidos), no se observó ningún SNP significativo coincidente entre las tres pruebas según el análisis alélico, genotípico o recesivo. En el análisis de dominancia hubo solamente 3 SNPs (rs464049 ubicado en el gen *DAT1*, y rs1194328 y rs1527153 ubicados en el gen *FOXP2*) con p menor a 0.05 coincidentes entre el test D vs Ctr-P<sub>D</sub> y el test C vs Ctr<sub>C</sub>. Al incluir en el análisis las comparativas D+C vs Ctr-P<sub>D</sub> y FA+C vs Ctr-P<sub>FA</sub>, las cuales consideran las 45 muestras comórbidas tanto dentro de la población disléxica como de la de FA, la significancia de estos SNPs varió sin un patrón claro de interpretar (ver tabla suplementaria 7).

Los resultados del análisis de **haplotipos** en las poblaciones puras (comparativas D vs Ctr-P<sub>D</sub>/ FA vs Ctr-P<sub>FA</sub> y C vs Ctr<sub>C</sub>) se resumen en la Tabla 20. Los haplotipos significativos no coincidieron entre estudios. Curiosamente hubo un haplotipo formado por marcadores ubicados en el gen *DYX1C1* que mostró significancia en el estudio de comórbidos a pesar de que ninguno de los marcadores fue significativo individualmente para la misma comparativa.

Los resultados obtenidos en el análisis de **epistasia** con todos los SNPs ubicados en los 8 genes candidatos están detallados en la Tabla 21, en los que el SNP rs793674 ubicado en *DCDC2* se asociaba con mayor fuerza a las poblaciones de niños disléxicos (tanto en solitario como añadiendo los comórbidos) en interacción con varios SNPs situados en *FOXP2*. Cabe destacar, que en la comparativa FA vs Ctr-P<sub>FA</sub> existe una interacción significativa entre un SNP de *DCDC2* y otro situado en *KIAA0319*, ambos genes previamente relacionados con problemas de lectura, y en el test D vs Ctr-P<sub>D</sub> (dislexia) se observó una asociación entre la interacción de dos SNPs ubicados en el gen *DAT1*, gen previamente relacionado con problemas de atención.

Tabla 20. Descripción de los haplotipos asociados significativamente en los estudios de dislexia, falta de atención y comórbidos a partir de los genotipos obtenidos mediante los arrays de SNPs de genoma completo para los genes candidatos. Haplotipo=nucleótidos específicos que forman el haplotipo. SNPs=identificador de los SNPs que componen el haplotipo. En las columnas referentes a cada uno de los estudios se muestra el valor de p obtenido para el haplotipo correspondiente: D=valores de p para el grupo de disléxicos; FA=valores de p para el grupo de individuos con falta de atención; C=valores de p para el grupo de individuos comórbidos. Sólo se muestran los valores de p <0.05. Los SNPs coincidentes en más de un haplotipo están subrayados.

Gen	Haplotipo	SNPs	D	FA	C
COMT	GGGGGA	rs4646312, rs165656,rs165722,rs4633,rs2239393,rs4680,rs4646316	0.0404		
DBH	GAGGG	rs2797851,rs1611118,rs1108580,rs1108581,rs2873804			0.0113
KIAA0319	CAAGAG	rs4576240, rs17307478,rs4236031,rs6915373,rs9379671,rs7763790	0.0207		
	AACGAG		0.043		
DAT1	GGA	rs464049, rs11737901,rs460000			0.0233
DCDC2	AAAGAAAG	rs11755011, rs1417215, rs1340697, rs1856837, rs11962116, rs1770909, rs1770909, rs793839, rs793836		0.0495	
DCDC2	AAAGAG	rs1417215, rs1340697,rs1856837,rs11962116,rs1770909,rs793839			0.0143
DCDC2	AGA	rs3846829, rs707862, rs707864		0.0278	
DCDC2	AGAAGGAG	rs7765678, rs4052662, rs12192603, rs1535331, rs1277351,rs2753912,rs793674,rs6921810		0.011	
	AGAGAAGG				0.024
DYG1C1	GAGAGCGCAG	rs501231, rs8027135, rs10518810, rs8042509, rs11857139, rs12440680, rs17819108, rs16953240, rs7183937,rs11856786,rs687623,rs11855391		0.0397	
	AAGAGCGGAAG			0.0303	
	GAGAACCGCGGA				0.0088
FOXP2	ACAGG	rs7778560, rs1830282, rs1528090, rs12538008, rs717233,rs12535428	0.0321		
FOXP2	AGAAG	rs1194328, rs11771168, rs1209254, rs12112572, rs12705958, rs1527152	0.0187		
FOXP2	GGGGA	rs1194328, rs11771168, rs1209254, rs12705958, rs1527152		0.0467	
FOXP2	AGG	rs2694933, rs1916986, rs2690833	0.0298		
FOXP2	AAGGAA	rs2690833, rs2245192, rs7785799, rs2140619, rs2694943, rs10249531			0.0242
FOXP2	AAAGG	rs2245192, rs7785799, rs2140619, rs2694943, rs10249531	0.047		
FOXP2	GGAAC	rs6966051, rs17213159, rs1563408, rs2396765, rs12705971	0.0369		

Gen	Haplotipo	SNPs	D	FA	C
FOXP2	GAAG	rs10486026, rs10262192, rs10262103, rs727644		0.0351	
FOXP2	AGCAGGG	rs10486026, rs10262192, rs10262103, rs727644, rs4727799, rs17372842, rs6980093			0.0108
FOXP2	GAA				0.0091
	ACG	rs10269986, rs1229761, rs12705966			0.03
FOXP2	AG				0.0203
	AA	rs1868757, rs13243235			0.0103

Tabla 21. Resultados de asociación de la interacción de SNPs dos a dos, realizado a partir de los datos obtenidos en el GWAS para los SNPs ubicados en los 8 genes candidatos. Gen 1=gen dónde se ubica el SNP 1, Gen 2=gen dónde se ubica el SNP 2 (el cromosoma dónde se localizan los genes se detalla entre paréntesis al lado de cada gen). Casos/Controles=grupo de muestras utilizadas como casos o controles según la definición del recuadro de la página 91. Sólo se muestran los valores de  $p < 1E-04$ .

				Estudio de Dislexia		Estudio de Comórbidos	Estudio de Falta de Atención	
				D+C	D	C	FA+C	FA
				Ctr-P_D	Ctr-P_D	Ctr_c	Ctr-P_FA	Ctr-P_FA
Gen 1	SNP 1	Gen 2	SNP 2					
DCDC2 (6)	rs793674	FOXP2 (7)	rs10262192	4.89E-06	5.43E-06			
DCDC2 (6)	rs793674	FOXP2 (7)	rs10269986	3.14E-05	2.44E-05			
DCDC2 (6)	rs793674	FOXP2 (7)	rs10262103	9.45E-06	2.78E-05			
DAT1 (5)	rs2042449	DAT1 (5)	rs403636		3.99E-05			
DCDC2 (6)	rs793674	FOXP2 (7)	rs10486026	4.95E-05	4.37E-05			
DCDC2 (6)	rs793674	FOXP2 (7)	rs727644	1.68E-05	5.64E-05			
DCDC2 (6)	rs7765678	KIAA0319 (6)	rs807541					8.97E-05

#### 4.3.2.3 Estudio de asociación de genoma completo de polimorfismos de nucleótido único (SNPs)

Con los arrays *Human Omniexpress v1.0* y *v1.1* (Illumina Inc) se genotiparon un total de 1612 muestras (238 disléxicos, 841 controles de dislexia, 136 FA, 352 controles de FA y 45 comórbidos). Los datos de genotipado obtenidos se analizaron para cada uno de los tres estudios (dislexia, falta de atención y comórbidos) de diversas maneras: i) teniendo en cuenta marcadores individuales, ii) haplotipos, iii) ventanas corredizas y iv) epistasia. Debido a la ingente cantidad de análisis realizados es imposible mostrar todos los resultados en este apartado, razón por la que los valores estadísticos más significativos para cada uno de los modelos genéticos analizados y para cada uno de los tres estudios se presentan en las tablas suplementarias 8 a 28).

Entre todos los resultados adquiridos, se seleccionaron los que obtuvieron una  $p < 9.99 \text{ E-}05$  en los casos del análisis de marcadores individuales y de haplotipos, y un valor menor a  $9.99 \text{ E-}07$  en el caso del análisis de ventanas corredizas, a excepción del estudio de comórbidos en el cual el umbral de ventanas corredizas se puso en  $9.99 \text{ E-}06$ . Debido a la diferencia de valores de significancia encontrados en el análisis de epistasia para cada uno de los estudios, el criterio de corte para esta aproximación varió en cada caso. Tras este primer filtrado según el valor de  $p$ , nos centramos tan sólo en aquellos marcadores cuya posición los situaba dentro de algún gen.

En algunas ocasiones al analizar conjuntos de marcadores situados en los mismos genes, se obtuvo un valor de significancia mayor que al analizar los marcadores individuales (la descripción de los marcadores específicos que constituyen los haplotipos o ventanas, así como los nucleótidos específicos que los componen se encuentra en las tablas suplementarias 12, 13, 19, 20, 26 y 27). También se encontraron genes que sólo mostraron asociación en los análisis de conjuntos de SNPs.

La función y características de cada uno de los genes cuyos marcadores pasaron el umbral de asociación establecido se identificaron en las bases de datos Pubmed (<http://www.ncbi.nlm.nih.gov/pubmed>) y STRING (<http://string-db.org/>).

### 4.3.2.3.1 Estudio de asociación de dislexia

#### Identificación de SNPs individuales, haplotipos y ventanas corredizas asociados con dislexia

Los valores de mayor significancia obtenidos en el análisis de marcadores individuales para cualquiera de los modelos genéticos abordados fueron del orden de  $1E-07$  menos para el modelo recesivo que obtuvo un valor del orden de  $1E-06$  (los valores de  $p$  para cada uno de los marcadores más significativos bajo los distintos modelos genéticos se pueden ver en las tablas suplementarias 8 a 11 y los relativos a haplotipos y ventanas corredizas en las tablas suplementarias 12 y 13). El listado de genes asociados mediante alguno de los marcadores situados dentro de los mismos se resume en la Tabla 22.

Tabla 22. Listado de genes con valores de significancia en alguno de sus marcadores según lo descrito en el texto. Resultados correspondientes a la asociación individual de SNPs bajo distintos modelos genéticos, haplotipos y ventanas corredizas con dislexia. Chr=cromosoma, p\_X=valor de  $p$  bajo el modelo genético especificado, pudiendo ser "X": alel=alélico, geno=genotípico, dom=dominante, rec=recesivo, haplo=análisis de haplotipos, VC=ventanas corredizas. Los valores de  $p < 9.9E-07$  están remarcados en negrita. Los genes cuya función está relacionada con el sistema nervioso están representados en negrita y subrayados.

Chr	Gen	p_alel	p_geno	p_dom	p_rec	p_haplo	p_VC
14	ACTN1						<b>8.43E-08</b>
4	AFAP1					9.72E-05	
15	AKAP13	2.14E-05				6.22E-06	
10	ARID5B					6.07E-05	
2	BAZ2B					2.48E-06	
14	C14orf182	1.59E-05	7.72E-05	9.52E-05		3.48E-05	
12	<b>CACNA1C</b>		3.02E-05	5.79E-05			
10	<b>CAMK1D</b>		7.93E-05		2.46E-05		
3	CDCP1				9.61E-05		
18	CELF4	8.97E-05					
7	CFTR		2.08E-05				
3	CHCHD6					8.24E-05	
3	<b>CHL1</b>						<b>1.92E-07</b>
19	<b>CHST8</b>						<b>2.47E-07</b>
18	CLUL1,C18orf56	9.87E-05					
8	<b>CSGALNACT1</b>	2.55E-05	6.11E-05	1.70E-05			
6	DDO	7.87E-06	2.38E-05	2.66E-05		1.68E-05	<b>7.35E-07</b>
2	DIS3L2		8.70E-05	3.97E-05		6.85E-05	
8	<b>DLC1</b>				7.71E-05		
2	DYNC1I2					5.70E-05	
19	ELAVL1				5.91E-05		
22	ELFN2	4.37E-05		9.43E-05			

Chr	Gen	p_alel	p_geno	p_dom	p_rec	p_haplo	p_VC
3	<b><u>EPHA6</u></b>	2.00E-05				1.22E-05	
8	<b><u>EYA1</u></b>	9.47E-05					
6	EYS					9.94E-05	
3	FHIT		9.52E-05	6.25E-05			
5	<b><u>GABRG2</u></b>				5.92E-05		
X	<b><u>GRIA3</u></b>		1.56E-05	5.41E-05			
19	HIF3A						<b>6.95E-07</b>
16	HS3ST4		9.76E-05		7.42E-05		
4	IRF2	1.90E-05	3.56E-06	8.12E-05	1.28E-06	2.51E-05	<b>8.94E-07</b>
9	<b><u>KANK1</u></b>	5.30E-06	3.44E-06	<b>9.07E-07</b>			
22	<b><u>LARGE</u></b>				2.89E-05		
17	LASP1	2.73E-05	8.44E-05	8.27E-05		8.61E-06	<b>8.63E-07</b>
18	<b><u>LDLRAD4</u></b>				7.12E-05		
1	LINGO4					9.59E-05	
8	LOC100287313	4.12E-05		4.82E-05			
20	<b><u>MACROD2</u></b>			8.19E-05			
18	MAPK4				5.37E-05		
12	MUC19		1.56E-05		2.94E-06		
2	OSBPL6	6.18E-05					
13	PDS5B					5.62E-05	
17	PIPOX	<b>1.80E-07</b>	<b>3.20E-07</b>	<b>4.49E-07</b>			
8	<b><u>PLAT</u></b>	5.05E-05		8.19E-05			
10	PLXDC2		7.86E-05				
12	<b><u>PLXNC1</u></b>				6.24E-05		
2	PPP1R1C	5.21E-05		8.97E-05			
19	PPP5C						<b>5.15E-07</b>
13	PRHOXNB					3.04E-05	
5	<b><u>RAB3C</u></b>			7.85E-05			
2	RNF103-CHMP3				5.73E-05		
3	RORC					9.59E-05	
20	<b><u>SDCBP2</u></b>			9.21E-05			
5	<b><u>SEMA5A</u></b>			9.22E-05			
20	SLC13A3				8.00E-05		
2	<b><u>SLC25A12</u></b>					5.70E-05	
20	<b><u>SNPH</u></b>	7.26E-05				6.72E-05	
4	SORBS2					4.48E-05	
4	<b><u>SORCS2</u></b>	3.74E-05	3.88E-05		2.38E-05		
2	SPRED2	7.24E-05		8.11E-05		9.66E-05	
3	<b><u>SRGAP3</u></b>	3.65E-05		4.69E-05		6.01E-05	
13	ST13P4				8.52E-05		
6	SYNE1				9.29E-05		

Chr	Gen	p_alel	p_geno	p_dom	p_rec	p_haplo	p_VC
1	TCEA3		4.07E-05				
1	TDRKH					9.59E-05	
12	TESPA1	2.96E-05	7.83E-05	1.32E-05		3.56E-05	
12	<b><u>TMEM132D</u></b>						<b>5.84E-07</b>
1	TRABD2B		2.96E-05		1.37E-05		
20	TSHZ2		8.57E-05	9.51E-05			
12	WSCD2	9.50E-05					
16	WVOX						<b>8.48E-07</b>

Un 34 % del total de los genes seleccionados por su significancia (los 24 genes resaltados en negrita y subrayados en la Tabla 22) tienen una relación directa con el sistema nervioso o con sus componentes según la base de datos STRING o según su definición de función en Pubmed (desarrollo del sistema nervioso y cerebro, neurogénesis, migración neuronal, axones, transmisión sináptica y transporte de neurotransmisores o relación con alguna enfermedad neurológica). También se encontraron 12 genes relacionados con la estructura del citoesqueleto y la adhesión celular (*EYA1*, *PLNX1*, *KANK1*, *CHL1*, *DLC1*, *SEMA5A*, *SORBS2*, *ACTN1*, *SYNE1*, *CDCP1*, *AFAP1* y *LASP1*), 6 de ellos coincidentes con los relacionados con el sistema nervioso mencionados anteriormente. Además, 4 genes de este listado (*BAZ2B*, *TCEA3*, *RORC* y *LASP1*) pertenecen al grupo de moléculas de unión al zinc.

### **Interacciones de parejas de SNPs asociadas con dislexia (epistasia)**

En la Tabla 23 se detallan las asociaciones epistáticas con dislexia más significativas. Un total de 122 interacciones SNP-SNP pasaron el umbral establecido de  $p < 9.99E-12$  (los detalles de estas interacciones se pueden consultar en la tabla suplementaria 14). De todos los hallazgos, se excluyeron aquellos que contemplaban los dos SNPs con localización fuera de un gen, por su interpretación incierta. De esta forma finalmente nos centramos en 66 interacciones, 23 de éstas se producían entre SNPs situados en dos genes mientras que 43 fueron entre un marcador génico y otro intergénico. En la Tabla 23 se presentan las 23 interacciones entre marcadores incluidos en genes (gen-gen). En 26 de las 66 interacciones citadas, al menos uno de los marcadores está ubicado en un gen con una función relacionada con el sistema neurológico (neurotransmisores, desarrollo del sistema nervioso, regulación sináptica, neurogénesis, axones, etc.) según las bases de datos STRING y Pubmed. Algunos de los genes correspondientes a estas funciones, los que interaccionaban con otro gen, son los escritos en negrita y subrayados en la Tabla 23. Además de

éstos, también se identificaron SNPs en los genes *NRP1*, *GABRB1*, *NPAS3*, *ORPM1*, *DLGAP1* y *CSMD1* interaccionando con marcadores de zonas no codificantes del genoma.

Tabla 23. Resultados más significativos obtenidos a partir del análisis de interacción de los SNPs, dos a dos, para el estudio de dislexia. Chr 1/2=cromosoma en el que se sitúa el SNP. SNP 1/2=identificador del marcador analizado. Gen 1/2=gen dónde se ubica el marcador analizado. Valor de p=valor de significancia del análisis de interacción entre los dos marcadores de cada línea. Los genes cuya función está relacionada con el sistema nervioso están representados en negrita y subrayados.

Chr 1	SNP 1	Gen 1	Chr 2	SNP 2	Gen 2	Valor de p
12	rs12231356	<b><u>TPH2</u></b>	15	rs17235533	<i>FMN1</i>	1.01E-14
2	rs6718486	<i>ZNF385B</i>	20	rs742922	<i>ANGPT4</i>	1.08E-14
12	rs12231356	<b><u>TPH2</u></b>	15	rs11072170	<i>FMN1</i>	1.77E-14
2	rs2438028	<i>SAP130</i>	10	rs6593449	<i>TMEM72</i>	3.37E-14
2	rs11680146	<i>SAP130</i>	10	rs6593449	<i>TMEM72</i>	3.38E-13
1	rs11589363	<i>DESI2</i>	11	rs2697825	<b><u>SPON1</u></b>	3.46E-13
1	rs11589363	<i>DESI2</i>	11	rs2697823	<b><u>SPON1</u></b>	6.70E-13
4	rs16860198	<b><u>GABRB1</u></b>	22	rs9618216	<i>USP18</i>	9.74E-13
5	rs16901583	<i>PDZD2</i>	11	rs58795	<b><u>NTM</u></b>	1.02E-12
1	rs4658683	<i>EFCAB2</i>	18	rs304933	<i>NFATC1</i>	1.13E-12
4	rs10028945	<b><u>GABRB1</u></b>	22	rs9618216	<i>USP18</i>	1.19E-12
3	rs2255532	<i>COL7A1</i>	11	rs4944452	<b><u>DLG2</u></b>	1.33E-12
12	rs11043265	<i>WDR66</i>	15	rs3751586	<b><u>GLDN</u></b>	1.52E-12
2	rs447106	<b><u>TANC1</u></b>	3	rs361052	<i>IQSEC1</i>	2.09E-12
1	rs6702844	<i>ST6GALNAC3</i>	12	rs10505830	<i>PLCZ1</i>	3.91E-12
17	rs767574	<i>ARHGAP44</i>	20	rs6110189	<i>SIRPB2</i>	4.56E-12
9	rs7047127	<b><u>KANK1</u></b>	19	rs4808331	<i>CYP4F8</i>	5.39E-12
4	rs2661665	<i>APBB2</i>	7	rs375981	<b><u>CDHR3</u></b>	6.39E-12
3	rs11710743	<i>ZNF385D</i>	4	rs13110423	<i>LDB2</i>	6.59E-12
2	rs1529289	<b><u>CNTNAP5</u></b>	18	rs1372770	<b><u>DOK6</u></b>	7.81E-12
3	rs11710743	<i>ZNF385D</i>	4	rs283029	<i>LDB2</i>	8.76E-12
2	rs1529289	<b><u>CNTNAP5</u></b>	18	rs7238969	<b><u>DOK6</u></b>	8.92E-12
11	rs7116459	<i>DENND5A</i>	16	rs4290460	<b><u>CDH13</u></b>	9.47E-12

Uno de los resultados más interesantes fue el obtenido para ciertos marcadores del gen *KANK1*, ya que el SNP rs3739592 obtuvo un valor de significancia de  $p=9.07E-07$  en el análisis de marcadores individuales bajo el modelo genético dominante (Tabla 22), y el rs7047127 resultó ser aún más significativo al considerar su interacción epistática con un SNP del gen *CYP4F8* ( $p=5.39E-12$ , Tabla 23).

#### 4.3.2.3.2 Estudio de asociación de falta de atención

##### Identificación de SNPs individuales, haplotipos y ventanas corredizas asociados con falta de atención

En la Tabla 24 se muestran los genes cuyos SNPs han resultados significativamente asociados con falta de atención (FA) bajo los distintos modelos genéticos empleados. El menor valor de significancia obtenido en esta comparativa fue de 2.24E-07 (los valores de p obtenidos para cada uno de los marcadores de mayor significancia bajo los diferentes modelos genéticos empleados se especifican en las tablas suplementarias 15, 16, 17 y 18 y los detalles de los resultados relativos a haplotipos y ventanas corredizas en las tablas suplementarias 19 y 20).

Tabla 24. Listado de genes con valores de significancia en alguno de los marcadores según lo establecido en el texto. Resultados del estudio de asociación de SNPs a lo largo del genoma completo con falta de atención bajo distintos modelos genéticos, haplotipos y ventanas corredizas. Chr=cromosoma, p\_X=valor de p bajo el modelo genético especificado, pudiendo ser "X": alel=alélico, geno=genotípico, dom=dominante, rec=recesivo, haplo=análisis de haplotipos, VC=ventanas corredizas. Los valores de  $p < 9.9E-07$  están remarcados en negrita. Los genes cuya función está relacionada con el sistema nervioso están representados en negrita y subrayados.

Chr	Gen	p_alel	p_geno	p_dom	p_rec	p_haplo	p_VC
3	ADAMTS9			6.38E-05			
2	ANKRD44				4.17E-05		
11	APLNR	6.51E-05					
14	ARG2	7.28E-05	4.16E-05		5.92E-05		
7	<b><u>ATXN7L1</u></b>					9.37E-06	
2	BRE	4.00E-06	8.42E-06	9.59E-06		1.89E-05	<b>2.24E-07</b>
5	<b><u>C5orf64</u></b>	6.04E-05		7.65E-05		2.59E-05	
3	CACNA2D3			5.88E-05			
11	CD44			3.37E-05			
1	CEP350		5.68E-05		3.13E-05		
15	<b><u>CHRNA3</u></b>	9.86E-05					
11	<b><u>CNTN5</u></b>		7.59E-05				
9	COL27A1				9.91E-05		
12	CORO1C					2.68E-05	
2	CPS1	9.64E-05		9.12E-05			
13	CRYL1		9.22E-05		3.33E-05		
8	<b><u>CSMD1</u></b>		8.97E-05				
22	DEPDC5		1.16E-05		6.85E-05		
7	ELMO1			8.92E-05			
7	<b><u>EMID2/COL26A1</u></b>						<b>3.87E-08</b>
2	FARSB					1.96E-05	
18	FHOD3		9.89E-05		9.40E-05	5.70E-05	

Chr	Gen	p_alel	p_geno	p_dom	p_rec	p_haplo	p_VC
12	GLT1D1	7.96E-05		8.89E-05			
2	GREB1		7.47E-05				
16	HS3ST4		3.87E-05				
20	<b>JAG1</b>	2.35E-05	9.99E-05	3.79E-05		8.97E-05	
1	KDM5B	5.19E-05					
16	KIAA1609	1.25E-05	2.89E-05	7.89E-05		7.84E-06	<b>8.36E-07</b>
1	KIF26B	5.10E-05	9.62E-05		9.25E-05	4.53E-05	
9	LOC100506834		9.65E-05				
6	LOC100652953		3.72E-05				
1	MAEL	8.83E-05		6.05E-05		8.74E-05	
2	MOGAT1	9.48E-06	7.40E-05			1.96E-05	
13	NALCN					1.69E-05	
11	NDUFC2- KCTD14		9.50E-05			7.31E-05	
6	<b>NHSL1</b>		5.19E-05				
15	NOP10	8.38E-05					
7	<b>NRCAM</b>	5.03E-05		5.46E-05			
2	<b>NRXN1</b>			6.83E-05			
1	<b>NTNG1</b>		8.61E-05	8.61E-05			
6	<b>PARK2</b>	3.18E-05		7.70E-05		5.06E-05	
21	<b>PCP4</b>	1.85E-06	5.06E-06		2.10E-05		
1	PLD5		5.03E-06				
14	PLEK2						<b>8.90E-07</b>
6	PPP1R14C		7.90E-05				
22	PRR14L		9.17E-05				
3	<b>PTPRG</b>	1.46E-05	8.14E-05	5.96E-05			
23	<b>RPGR</b>		1.73E-05	4.12E-06			
11	<b>SCN3B</b>						<b>2.99E-07</b>
18	SETBP1	2.73E-05	8.31E-05			1.23E-05	
10	SFTPD	2.54E-05	3.03E-05	3.03E-05			
1	SPATA1		7.42E-05	1.90E-05			
12	<b>SSH1</b>	8.31E-05				2.68E-05	
12	SSPN		7.31E-05			7.78E-05	
14	STON2	3.56E-05				1.69E-05	
14	STON2					1.69E-05	
12	<b>TMEM132D</b>				3.30E-05	8.62E-06	
12	TMEM132D					8.62E-06	
9	<b>TNC</b>						<b>9.85E-07</b>
18	TPGS2					5.70E-05	
3	TRIM71	3.13E-05	4.92E-05				
8	<b>UNC5C</b>					7.84E-05	

El 32 % del total de los genes asociados con FA (los 19 genes en negrita y subrayados en la Tabla 24) resultaron tener una función neurológica (desarrollo del sistema nervioso, neurogénesis, axones, etc.) según la base de datos STRING y la descripción detallada en Pubmed. De este set de SNPs también destacaron los ubicados en genes relacionados con el citoesqueleto y adhesión celular (*CD44*, *KIF26B*, *SFTPD*, *SSPN*, *CEP350*, *ELMO1*, *FHOD3*, *PLEK2*, *COROC1C*, *NRCAM*, *NRXN1*, *SSH1*, *TNC* y *PARK2*), así como los relacionados con canales de sodio y calcio (*CACNA2D3*, *NDUF2-KCTD14*, *NALCN*, *CHRNA3* y *SCN3B*). También se han de mencionar 5 genes con funciones relacionadas con el metabolismo de lípidos (*CPS1*, *CRYL1*, *MOGAT1*, *TNC* y *PARK2*). Además, 2 de los SNPs asociados están localizados en genes relacionados con la formación y organización de cilios (*JAG1* y *RPGR*) y 5 con la unión, homeostasis y respuesta a la presencia de moléculas del zinc (*PARK2*, *CPS1*, *ADAMT59*, *KDM5B* y *TRIM71*).

### **Interacciones de parejas de SNPs asociadas con falta de atención (epistasia)**

Los resultados más relevantes del análisis de interacción de parejas de SNPs se encuentran en la Tabla 25. Los valores de p obtenidos en el estudio de FA (valor menor de  $p=1.75E-11$ ) fueron mayores que en el estudio de dislexia (valor menor de  $p=1.97E-15$ ), por lo que, en este caso se consideraron las interacciones con valores de p menores a  $9.99E-10$  (los detalles de las interacciones de mayor significancia pueden consultarse en la tabla suplementaria 21). En total fueron 57 las interacciones consideradas. Como en el caso del estudio anterior, se examinaron sólo aquellas interacciones cuyos SNPs se localizaban dentro de un gen (43 interacciones). Diez de éstas fueron entre marcadores situados en dos genes, mientras que en el resto (33) sólo uno de los marcadores era génico. En la Tabla 25 se muestran las 10 interacciones formadas por marcadores situados en genes. Las casillas en negrita y subrayadas representan genes relacionados con en el sistema nervioso (desarrollo del sistema nervioso, neurogénesis, axones, etc.) según la base de datos STRING y Pubmed. Además de éstos, también se encontraron interacciones entre marcadores de los genes *NRXN3*, *RELN*, *RBFOX1* y *SHISA9* y marcadores intergénicos.

Tabla 25. Resultados más significativos obtenidos a partir del análisis de interacción de los SNPs dos a dos para el estudio de falta de atención. Chr 1/2=cromosoma en el que se sitúa el SNP. SNP 1/2=identificador del marcador analizado. Gen 1/2=gen dónde se ubica el marcador analizado. Valor de p=valor de significancia del análisis de interacción entre los dos marcadores de cada línea. Los genes cuya función está relacionada con el sistema nervioso están representados en negrita y subrayados.

Chr 1	SNP 1	Gen 1	Chr 2	SNP 2	Gen 2	Valor de p
16	rs4534848	<i>HS3ST4</i>	17	rs9916110	<b><u>ASIC2</u></b>	4.57E-11
22	rs9624712	<i>KIAA1671</i>	7	rs7801843	<b><u>SDK1</u></b>	2.18E-10
22	rs9624712	<i>KIAA1671</i>	7	rs6971090	<b><u>SDK1</u></b>	2.55E-10
2	rs10495492	<i>DCDC2C</i>	13	rs9317622	<b><u>PCDH9</u></b>	2.76E-10
7	rs6971911	<i>DGKB</i>	15	rs716218	<i>MEIS2</i>	2.88E-10
3	rs811322	<i>FAIM</i>	17	rs8067393	<i>C17orf101</i>	4.15E-10
5	rs2910260	<b><u>GRIA1</u></b>	20	rs1892318	<i>CDH4</i>	4.92E-10
8	rs16915883	<i>LINC00535</i>	13	rs17678470	<b><u>NALCN</u></b>	5.95E-10
2	rs4927635	<i>SNTG2</i>	21	rs2837108	<i>B3GALT5</i>	7.98E-10
3	rs1545991	<b><u>NLGN1</u></b>	9	rs7036989	<b><u>SH3GL2</u></b>	8.17E-10

Hubo dos genes, *HS3ST4* y *NALCN*, con valores de significancia de 3.87E-05 y 1.69E-05 en el análisis genotípico y de haplotipos respectivamente, que aumentaron su valor de asociación en el análisis de interacción de marcadores por parejas, llegando hasta valores de 4.57E-11 (*HS3ST4* con *ASIC2*) y de 5.95E-10 (*NALCN* con *LINC00535*).

#### 4.3.2.3.3 Estudio de asociación con muestras comórbidas

##### **Identificación de SNPs individuales, haplotipos y ventanas corredizas asociadas a muestras comórbidas**

El valor de mayor asociación en el estudio de las muestras comórbidas fue  $p=8.97E-07$  (los valores de p de mayor significancia obtenidos para cada uno de los marcadores bajo los distintos modelos genéticos empleados se muestran en las tablas suplementarias 22, 23, 24 y 25 y los referentes a haplotipos y ventanas corredizas en las tablas suplementarias 26 y 27). En la Tabla 26 se muestra el valor de p del marcador más significativo situado en cada uno de los genes detallados bajo los distintos modelos genéticos. Sólo se han añadido los valores de significancia menores a 9.99E-05 en todos los casos excepto en el análisis de ventanas corredizas que el umbral se puso en 9.99E-06.

Tabla 26. Valores de significancia obtenidos del estudio de asociación de SNPs individuales bajo distintos modelos genéticos, haplotipos y ventanas corredizas para la población de comórbidos. Chr=cromosoma, p\_X=valor de p bajo el modelo genético especificado, pudiendo ser "X": alel=alélico, geno=genotípico, dom=dominante, rec=recesivo, haplo=análisis de haplotipos, VC=ventanas corredizas. Los valores de  $p < 9.9 \times 10^{-7}$  están remarcados en negrita. Los genes cuya función está relacionada con el sistema nervioso están representados en negrita y subrayados.

Chr	Gen	p_alel	p_geno	p_dom	p_rec	p_haplo	p_VC
7	AGAP3		7.29E-05	3.31E-05			
8	<b><u>ANK1</u></b>		8.35E-05				
2	ANTXR1					6.46E-05	
7	BBS9	8.06E-05					
12	C12orf28	6.14E-05					
5	CDH12	7.45E-05					
7	<b><u>CDK4</u></b>			5.75E-05			
13	<b><u>COL4A1</u></b>					2.37E-05	8.58E-06
6	<b><u>COL9A1</u></b>	1.39E-05					5.12E-06
8	<b><u>CSGALNACT1</u></b>	7.10E-06	6.76E-06	6.76E-06			4.76E-06
8	CSMD1	8.87E-05		9.06E-05		5.91E-05	<b>9.45E-07</b>
7	DNAH11					1.34E-05	
14	<b><u>DPF3</u></b>		2.08E-05				
7	ELMO1					6.49E-05	
4	<b><u>GAK</u></b>		8.72E-05	5.47E-05		4.61E-05	
18	IMPA2	6.22E-05	7.18E-05	2.97E-05			
1	<b><u>KCNN3</u></b>	8.03E-05	4.20E-05	4.20E-05		2.53E-05	4.26E-06
1	LOC100505768		2.75E-05	3.91E-05			
13	LOC100506063		3.48E-05	3.48E-05		2.80E-05	
14	LOC728755	8.68E-05	4.94E-05	4.94E-05			
1	LPPR5					1.80E-05	
2	<b><u>LRRTM4</u></b>			5.29E-05			
1	MARC1	4.35E-05	2.83E-05	2.97E-05			
1	MKKN1					8.92E-06	5.64E-06
1	MOB3C					8.92E-06	5.96E-06
2	MYO3B	1.56E-05	7.16E-05	3.41E-05		1.93E-05	
19	NMRK2		3.93E-05				
16	NOD2					5.54E-05	
11	<b><u>OPCML</u></b>		2.03E-05				
12	PDE3A					1.58E-05	
6	PHACTR1		1.74E-05				
5	S100Z			2.52E-05			
6	SAMD5	6.60E-05	1.01E-05	5.14E-06			
12	SCARB1			3.67E-05			
11	<b><u>SHANK2</u></b>			8.62E-05			
1	SLC2A1					4.31E-05	6.73E-06
1	SLC35F3					7.44E-05	
17	SLC39A11	9.44E-05					

Chr	Gen	p_alel	p_gen	p_dom	p_rec	p_haplo	p_VC
16	SNX20					5.54E-05	
2	SRBD1					6.95E-05	2.58E-06
1	<b><u>SSBP3</u></b>						<b>8.97E-07</b>
15	<b><u>TCF12</u></b>					7.47E-05	
12	TDG	6.22E-05					
21	<b><u>TIAM1</u></b>	9.09E-05					
1	TNNI1		4.86E-05				
6	TRERF1					4.35E-05	
9	TRPM3	4.35E-05					
3	<b><u>ULK4</u></b>					7.71E-05	
3	VGLL4		2.94E-06				
2	ZNF385B	8.01E-05					

Entre los 50 genes con marcadores de mayor significancia listados en la Tabla 26, un 30 % (15 genes, remarcados en negrita y subrayados en la tabla) resultaron tener una función en el sistema nervioso (desarrollo, neurogénesis, neurotransmisores, axones, etc.). Dentro de esta lista también se encuentran genes relacionados con adhesión y citoesqueleto (ANK1, CDH12, CDK4, NMRK2, OPCML, SCARB1, TIAM1, ANTXR1, ELMO1 y ULK4). También hemos de mencionar que 5 de estos genes tienen relación con el transporte o la unión del ion zinc (AGAP3, DPF3, SLC39A11, VGLL4 y ZNF385B), otros 3 genes forman parte de procesos que involucran a los cilios (BBS9, DNAH11 y ULK4) y 4 genes codificantes o reguladores de colágeno (COL9A1, COL4A1, ANTXR1 y SSBP3).

### **Interacciones de parejas de SNPs asociadas a muestras comórbidas (epistasia)**

Los valores de p obtenidos en el análisis de interacción por parejas de marcadores fueron mayores que en los dos estudios anteriores (ver listado de interacciones más significativas en la tabla suplementaria 28). El listado de interacciones consideradas en este estudio se detalla en la Tabla 27.

Tabla 27. Resultados más significativos obtenidos a partir del análisis de interacción de los SNPs dos a dos para el estudio de comórbidos. Chr 1/2=cromosoma en el que se sitúa el SNP. SNP 1/2=identificador del marcador analizado. Gen 1/2=gen dónde se ubica el marcador analizado. Valor de p=valor de significancia del análisis de interacción entre los dos marcadores de cada línea. Los genes cuya función está relacionada con el sistema nervioso están representados en negrita y subrayados.

Chr 1	SNP 1	Gen 1	Chr 2	SNP 2	Gen 2	Valor de p
9	rs7874876	PTPLAD2	10	rs1060093	FRMD4A	1.83E-08
9	rs7874876	PTPLAD2	10	rs17153667	FRMD4A	3.68E-08
4	rs4859880	PARM1	7	rs505532	CPVL	1.73E-07
2	rs17031056	THADA	5	rs4916689	<b><u>GPR98</u></b>	1.88E-07
4	rs4583740	HTRA3	23	rs2998250	<b><u>CASK</u></b>	2.13E-07
3	rs359045	LMCD1-AS1	10	rs7901991	<b><u>RNLS</u></b>	2.42E-07
2	rs17031056	THADA	5	rs4916688	<b><u>GPR98</u></b>	2.66E-07
3	rs986613	<b><u>PTPRG</u></b>	7	rs2373284	<b><u>CNTNAP2</u></b>	2.91E-07
2	rs1022386	RBMS3	17	rs3751957	CBX2	3.24E-07
8	rs10505173	CSMD3	21	rs2824607	CHODL	3.47E-07
2	rs900439	IQCA1	9	rs1144434	GNA14	3.81E-07
8	rs2685320	SCARA5	11	rs1031381	NCAPD3	3.90E-07
2	rs3771033	<b><u>NRP2</u></b>	3	rs7644727	<b><u>SLC9A9</u></b>	5.05E-07
4	rs13110849	GALNTL6	9	rs460154	BRD3	5.16E-07
4	rs2677605	SH3TC1	8	rs9297586	SAMD12	5.55E-07
12	rs2277414	FOXJ2	17	rs4386173	IKBKB	5.68E-07
7	rs10488578	ABCB5	20	rs8117911	RBL1	5.70E-07
2	rs13387277	PARD3B	2	rs6746926	<b><u>PIKFYVE</u></b>	5.76E-07
7	rs7796223	<b><u>CHN2</u></b>	22	rs6007640	MIRLET7BHG	6.61E-07
7	rs3812367	<b><u>CHN2</u></b>	22	rs6007640	MIRLET7BHG	6.61E-07
7	rs3812368	<b><u>CHN2</u></b>	22	rs6007640	MIRLET7BHG	6.61E-07
4	rs6832401	LOC285423	18	rs525374	FHOD3	6.72E-07
5	rs1428439	<b><u>ST8SIA4</u></b>	14	rs10130801	LOC283587	6.89E-07
5	rs1428439	<b><u>ST8SIA4</u></b>	14	rs12884100	LOC283587	6.89E-07
5	rs1428439	<b><u>ST8SIA4</u></b>	14	rs4144361	LOC283587	6.89E-07
2	rs1519634	LOC150568	11	rs7949010	TEAD1	6.93E-07
3	rs1520203	CCDC50	6	rs13209503	ANKRD6	7.18E-07
6	rs1799910	COL11A2	20	rs2064272	PLCB1	7.83E-07
5	rs7736159	GALNT10	11	rs902869	LOC100127946	7.85E-07
6	rs9320552	FRK	16	rs464349	CPNE7	7.90E-07
6	rs1040285	LOC100996634	11	rs17133433	<b><u>CNTN5</u></b>	8.59E-07
10	rs12769066	A1CF	13	rs7337686	<b><u>MYCBP2</u></b>	8.87E-07
10	rs12769066	A1CF	13	rs2329029	<b><u>MYCBP2</u></b>	8.87E-07
1	rs3767498	<b><u>CACNA1S</u></b>	16	rs1520233	<b><u>CDH11</u></b>	9.41E-07
17	rs736866	KIF18B	17	rs2279732	<b><u>SDK2</u></b>	9.81E-07
1	rs9804135	<b><u>CAPN8</u></b>	10	rs734187	<b><u>NRP1</u></b>	9.92E-07

El menor valor de  $p$  identificado en el análisis de epistasia para el estudio de comórbidos fue de  $1.83E-08$ , por lo que se decidió considerar las interacciones con significancia menor a  $9.99E-07$ , es decir, un total de 150 interacciones. Como en los estudios anteriores sólo se analizaron aquellas interacciones cuyos marcadores, al menos uno de ellos, estuviera localizado dentro de algún gen, por lo que el número de interacciones se redujo a 121. Finalmente fueron 36 las interacciones realizadas entre dos marcadores situados ambos en un gen y el resto (85) eran entre un marcador génico y otro intergénico. En la Tabla 27 se presentan los 36 resultados más significativos obtenidos a partir de interacciones entre marcadores ubicados en genes. Los genes escritos en negrita y subrayados corresponden a aquellos con función relacionada con el sistema nervioso según las bases de datos Pubmed y STRING. Entre las interacciones compuestas por marcadores génicos e intergénicos destacan las de los genes *SRRM4*, *SYNPR*, *TSNARE1*, *PDE7B*, *PTPRN2*, *RTN1*, *TRIM9*, *SYNGR1* y *SLIT1*. Ninguno de los genes significativos en el análisis de epistasia coincidió con los asociados al estudiar marcadores individuales, haplotipos o ventanas corredizas.

#### **4.3.2.3.4 Comparación de los genes asociados entre los estudios de dislexia, falta de atención y muestras comórbidas**

Al mirar en detalle los genes cuyos marcadores se han asociado significativamente bajo alguno de los modelos genéticos empleados (alélico, genotípico, dominante, recesivo, haplotipos, ventanas corredizas e interacción epistática) observamos que en rasgos generales los genes asociados en cada estudio (dislexia, FA y comórbidos) son exclusivos, es decir, cada trastorno presenta una serie de genes asociados de forma privativa. Aun así, se han encontrado algunas coincidencias. Los genes *TMEM132D* y *HS3ST4* aparecen asociados tanto con el grupo de disléxicos como con la población que muestra falta de atención. Entre los disléxicos y los comórbidos se encuentran significativamente asociados los genes *CSGALNACT1* y *NRP1*, mostrando este último su asociación tan sólo en modo de interacción epistática en ambos estudios. Curiosamente las interacciones de *NRP1* no son las mismas en los dos estudios, ya que mientras que en el estudio de dislexia interacciona con el marcador rs784500 ubicado en el cromosoma 3 (datos no mostrados en la tabla) en el estudio de comórbidos lo hace con el marcador rs9804135 situado en el gen *CAPN8* del cromosoma 1. Los genes *CSMD1* y *ZNF385B* asociados a la población comórbida también tienen un valor significativo en el estudio de dislexia, pero sólo en su epistasia con otro marcador genético, específicamente con el marcador

rs12529514 situado en el cromosoma 6 en el primer caso (datos no mostrados en la tabla) y con el marcador rs742922 del gen *ANGPT4* del cromosoma 20 en el segundo. También existen coincidencias entre las muestras con falta de atención y los comórbidos: los genes *CSMD1* y *ELMO1* aparecen asociados en ambos estudios, y *CNTN5*, *FHOD3* y *PTPRG* se asocian por si solos en el estudio de falta de atención y en interacción con los comórbidos. En concreto, el gen *CNTN5* interacciona con el marcador rs1040285 ubicado en el cromosoma 6 (*LOC100996634*), *FHOD3* lo hace con el marcador rs6832401 del cromosoma 4 (*LOC285423*) y *PTPRG* con un marcador situado en el gen *CNTNAP2* del cromosoma 7 (rs2373284)(datos no mostrados en la tabla).

#### **4.3.2.3.5 Diferencias en la significancia de los marcadores de mayor asociación en distintos grupos de muestras**

En el subproyecto referente a los genes candidatos (apartado 4.2) se realizaron distintas comparativas de asociación según la definición de fenotipos de las poblaciones (Tabla 10). En este apartado del proyecto, también se realizaron estas mismas comparativas con los datos de genoma completo con el objetivo de identificar posibles diferencias de asociación entre los distintos fenotipos. Los marcadores seleccionados para dicho análisis fueron los que mayor valor de significancia habían alcanzado en los análisis previos (ver tablas suplementarias 8 a 28). Para el análisis de SNPs individuales se escogieron 183 SNPs, los cuales estaban genotipados en todos los grupos comparados (valor de  $p < 9.9E-05$ ), para el estudio de epistasia 176 (valor de  $p$  según el establecido en cada comparativa: en dislexia  $p < 9.99E-12$ , en FA  $p < 9.99E-10$  y en comórbidos  $p < 9.99E-07$ ) y para el análisis de ventanas corredizas y haplotipos se estudiaron 18 bloques (con  $p < 9.9E-07$  y que estuvieran ubicados en algún gen). Los análisis de asociación se realizaron utilizando los mismos filtros y condiciones descritos anteriormente (ver apartado 3.5.2.3), salvo para el análisis de epistasia que se realizó mediante PLINK<sup>316</sup> por su rapidez y simplificación de resultados. Los grupos de muestras utilizados fueron los correspondientes a las comparativas D+C vs Ctr-P\_D (283 casos/841 controles), D vs Ctr-P\_D (238 casos/841 controles), FA+C vs Ctr-P\_FA (181 casos/352 controles), FA vs Ctr-P\_FA (136 casos/352 controles) y C vs Ctr\_C (45 casos/102 controles).

### **Análisis de los marcadores individuales**

Las únicas asociaciones identificadas (valor de  $p < 9.9E-05$ ) fueron en los tests relativos al trastorno de cual se había seleccionado el marcador reanalizado, a excepción de dos marcadores que se asociaron tanto a comorbidad como a falta de atención. El SNP rs1342872 (LOC105373231) obtuvo un valor de  $p=2.8E-05$  bajo el modelo alélico en comórbidos (C vs Ctr\_C), un valor de  $5.92E-05$  en el estudio de falta de atención (FA vs Ctr-P\_FA) y un valor de 0.008116 al juntar las muestras de ambas comparativas (FA+C vs Ctr-P\_FA). Por otra parte, el marcador rs16976956 ubicado en el gen *IMPA2*, obtuvo valores de significancia del orden de  $E-05$  bajo distintos modelos genéticos en el grupo de comórbidos (C vs Ctr\_C), un valor de 0.006007 para falta de atención (FA vs Ctr-P\_FA) bajo el modelo genético dominante, y un valor de  $7,83E-05$  bajo este mismo modelo al añadir las muestras comórbidas (FA+C vs Ctr-P\_FA).

De los 71 marcadores asociados con dislexia (D vs Ctr-P\_D), 54 cambiaron de orden de significancia al incluir las muestras comórbidas (D+C vs Ctr-P\_D). Diez marcadores (rs7682286, rs13120088, rs12649641, rs13115095, rs11726275, rs6918133, rs2252795, rs1058151, rs7236483 y rs7241967) disminuyeron su valor de  $p$  en un orden (aumentando así su significancia de asociación) al incluirse las muestras comórbidas y uno, rs4129455, disminuyó en dos órdenes. Además, 39 marcadores aumentaron en un orden su valor de  $p$  (rs2790661, rs2105233, rs1107901, rs9288078, rs9845704, rs17036126, rs9878908, rs4279078, rs301923, rs11723606, rs9297093, rs9487396, rs2024364, rs16916968, rs7822386, rs7826130, rs10465023, rs12552198, rs10893866, rs10790952, rs3764002, rs9528771, rs4415908, rs10140681, rs8016253, rs927306, rs10134613, rs10220533, rs12147602, rs2355654, rs4900974, rs3825575, rs17623915, rs3865390, rs7231380, rs7407139, rs2071857, rs12557112 y rs5931716) y 4 marcadores (rs7565742, rs10120392, rs3739592 y rs3764714) lo hicieron en dos órdenes de magnitud (disminuyendo así su significancia de asociación).

Centrándonos en las asociaciones con falta de atención, de los 67 marcadores seleccionados (FA vs Ctr-P\_FA), sólo 8 disminuyeron su valor de  $p$  en un orden de magnitud al unir las muestras comórbidas con FA (FA+C vs Ctr-P\_FA) (rs6748782, rs1382381, rs10075443, rs1269662, rs2125391, rs755782 y rs3742855) mientras que 9 marcadores lo aumentaron en dos órdenes de magnitud (rs12143525, rs10920473, rs11127143, rs12342660, rs4964758, rs589346, rs12454120,

rs7248031 y rs2244188) y 38 lo hicieron en un solo orden (rs1041238, rs6701279, rs11127135, rs1968000, rs7568185, rs10203337, rs6547838, rs10495763, rs17041592, rs999497, rs2177250, rs7574494, rs13410347, rs12611436, rs6739216, rs17066305, rs10865564, rs9819574, rs4678297, rs12488404, rs12494312, rs2950220, rs2974452, rs9458413, rs2193241, rs17694897, rs1570856, rs2282624, rs746886, rs7969314, rs11628716, rs8020575, rs7173, rs472054, rs434690, rs1465594, rs2852780 y rs16981324).

### **Análisis de haplotipos y ventanas corredizas**

Tanto los haplotipos como las ventanas corredizas seleccionados fueron en general únicos de cada trastorno, sólo hubo un haplotipo formado por los marcadores rs3820546 y rs3820548 (GG) y situado en el gen *SLC2A1*, con valor de significancia de 2.21E-05 cuando consideramos sólo comórbidos (C vs Ctr\_c), que obtuvo un valor de 5.59E-06 en la comparativa de dislexia más comórbidos (D+C vs Ctr-P\_d). El resto de los haplotipos específicos del estudio de dislexia disminuyeron su valor de p (aumentando así su significancia) en uno o dos órdenes al incluir los comórbidos. Los valores de p para los haplotipos asociados con falta de atención bien se mantuvieron o bien aumentaron en un orden (disminuyendo así su significancia) cuando se consideraban a los comórbidos entre los individuos caso. En cuanto a las ventanas corredizas en cambio, sí hubo bloques que disminuyeron su valor de p, como los ubicados en *CHL1*, *CHST8* y *IRF2* que disminuyeron en un orden en la comparativa de D+C vs Ctr-P\_d con respecto al test de D vs Ctr-P\_d, y el bloque situado en el gen *SCN3B* que también lo hizo en la comparativa de FA+C vs Ctr-P\_FA con respecto al test de FA vs Ctr-P\_FA. El resto bien se mantuvo o bien disminuyó la significancia.

### **Análisis de las interacciones de parejas de marcadores**

Al incluir las 45 muestras comórbidas tanto con las muestras disléxicas (D+C vs Ctr-P\_d) como con las muestras de FA (FA+C vs Ctr-P\_FA) los valores de p se mantuvieron o aumentaron, haciendo estas interacciones menos significativas, excepto para las siguientes parejas de marcadores donde el valor de p disminuyó en el test de D+C vs Ctr-P\_d con respecto a la comparativa de D vs Ctr-P\_d (Tabla 28).

Tabla 28. Interacciones epistásicas que aumentaron su significancia al incluir a las muestras comórbidas en el grupo de casos. Chr 1/2=cromosoma en el que se sitúa el SNP. SNP 1/2=identificador del marcador analizado. Gen 1/2=gen dónde se ubica el marcador analizado. p=valor de significancia del análisis de interacción entre los dos marcadores de cada fila. Casos/Controles=grupo de muestras utilizadas como casos o controles según la definición del recuadro de la página 91.

						Estudio de Dislexia	
						Casos	D
						D+C	D
						Controles	Ctr-P_d
SNP 1	Chr 1	Gen 1	SNP 2	Chr 2	Gen 2	Ctr-P_d	Ctr-P_d
rs6718486	2	ZNF385B	rs742922	20	ANGPT4	7.35E-15	1.08E-14
rs2255532	3	COL7A1	rs4944452	11	DLG2	7.22E-13	1.33E-12
rs11043265	12	WDR66	rs3751586	15	GLDN	2.62E-14	1.52E-12

#### 4.3.2.3.6 Enriquecimiento de los marcadores más asociados según su localización génica

En la Figura 20 se muestra la proporción de marcadores observados frente a los esperados según su ubicación génica (exónico, intrónico, localizado cerca del extremo 3' o 5' del gen, en el UTR-3' o el UTR-5'), entre los 1500 marcadores más asociados ( $p < 0.0023$  para dislexia,  $p < 0.0022$  para FA,  $p < 0.0026$  para comórbidos) según el modelo alélico de los tres GWAS independientes realizados. Como se puede observar, en FA hay un aumento de mutaciones en la región UTR-3' con respecto a lo esperado a medida que se contabilizan los marcadores de mayor valor significativo. Para dislexia en cambio, el aumento de mutaciones significativas ocurren en los marcadores situados cerca del extremo 3' de los genes y en los ubicados en los UTR-5', estos últimos también predominantes entre los marcadores relacionados con comorbidad.

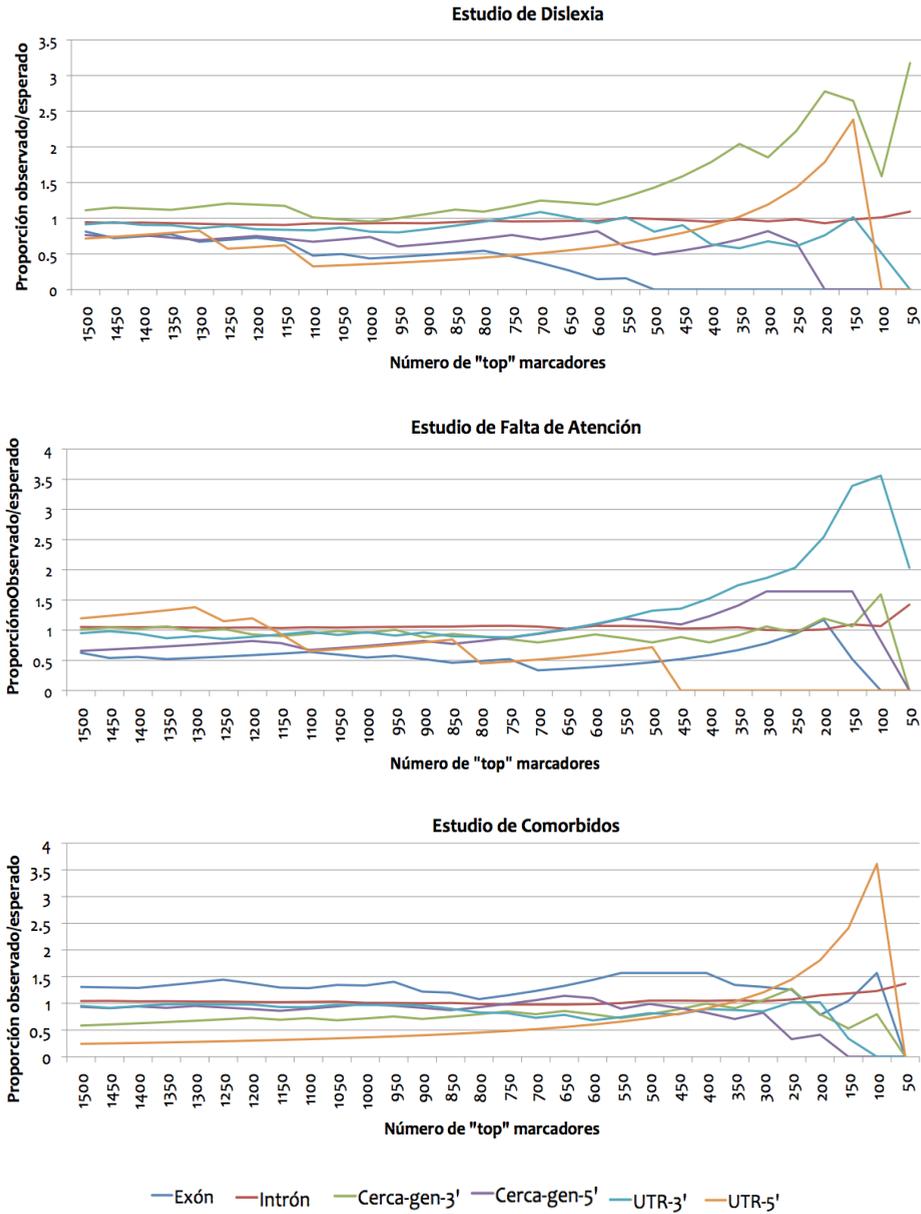


Figura 20. Representación gráfica de la proporción de marcadores observados frente a los esperados considerando los marcadores más significativos ("top" marcadores) con dislexia, falta de atención o comorbidos, según su localización génica.

#### 4.3.2.3.7 Validación tecnológica de los marcadores de mayor asociación con los trastornos investigados

En todo proyecto que considera un número grande de marcadores, se caracterizan un pequeño número de asociaciones mediante otra técnica disponible, con el fin de validar los resultados globales. En esta tesis se seleccionaron 48 marcadores de entre los resultados más significativos descritos, teniendo como criterio que los genes en los que se situaban los SNPs elegidos fueran de interés para el estudio. Concretamente los genes elegidos tenían funciones neuronales o estaban relacionados con la adhesión y la estructura celular:

- 36 SNPs asociados con dislexia de los cuales 3 SNPs pertenecían a 2 interacciones epistáticas (genes *DOK6* y *CNTNAP5*), 14 SNPs de un haplotipo o una ventana corrediza dentro del gen *DDO*, 9 SNPs de un haplotipo y una ventana corrediza dentro del gen *LASP1*, 8 SNPs formaban una ventana corrediza dentro del gen *ACTN1* y 2 SNPs individuales del gen *KANK1*.
- Dentro del estudio de falta de atención se seleccionaron 12 SNPs de la siguiente forma: 2 SNPs que formaban una interacción epistática entre los genes *NLGN1* y *SH3GL2*, 9 SNPs de un haplotipo y ventana corrediza dentro del gen *BRE* y un SNP individual localizado en el gen *PCP4*.

Todos estos marcadores se genotiparon en todas las muestras de la cohorte recogida para este estudio con diagnóstico, de forma que se aumentó así el número de DNAs analizados con respecto a los utilizados en el estudio de genoma completo. Tras aplicar los filtros descritos en el apartado 3.5.2.3, el número de muestras que se consideraron para las validaciones se resume de la siguiente manera: 247 disléxicos, 1421 controles para dislexia, 187 muestras con falta de atención, 520 controles para falta de atención y 45 muestras comórbidas (dentro de los controles se encuentran 112 controles compartidos para ambos trastornos).

La concordancia entre los genotipos obtenidos mediante los arrays *Human OmniExpress* y la tecnología de *TaqMan™* (*Thermo Fisher Scientific*) fue del 97-99 % a excepción de dos de las sondas (las correspondientes a los marcadores rs11127135 y rs9487396), las cuales obtuvieron un valor de concordancia del 87 y 94 % respectivamente. 18 sondas obtuvieron valores entre 97-98 % y el resto (28 sondas) superaron el 99 % de concordancia.

Al haber aumentado el número de muestras genotipadas para estos 48 SNPs, se volvieron a realizar los test comparativos de caso/control descritos en la Tabla 10. Los resultados más significativos del análisis de marcadores individuales se encuentran detallados en las Tabla 29 y Tabla 30:

En general los valores de  $p$  son mayores (menor asociación) que los obtenidos en el estudio de genoma completo. Los valores de significancia de la comparativa de D vs Ctr-P<sub>D</sub> en comparación con la de D+C vs Ctr-P<sub>D</sub> o D+C vs Ctr<sub>D</sub> son menores, es decir, al quitar del grupo de disléxicos aquellas muestras que también presentan falta de atención (Tabla 29). Cabe destacar que el valor de significancia del marcador rs10120392 situado en KANK1 alcanzó el umbral de significancia para los GWASs al aumentar el número de muestras (D vs Ctr-P<sub>D</sub>). Se observa la misma tendencia con los marcadores asociados con FA, al comparar el test de FA+C vs Ctr-P<sub>FA</sub> o FA+C vs Ctr<sub>FA</sub> con el de FA vs Ctr-P<sub>FA</sub> y al comparar el test de FA+C vs Ctr-P<sub>FA2</sub> o FA+C vs Ctr<sub>FA2</sub> con el test de FA vs Ctr-P<sub>FA2</sub>, es decir al quitar del grupo FA las muestras comórbidas (Tabla 30) (en la tabla suplementaria 29 se detallan los valores de  $p$  obtenidos para cada uno de los marcadores replicados bajo los distintos modelos genéticos empleados).

Tabla 29. Resultados obtenidos mediante la tecnología TaqMan™ del estudio de asociación del modelo alélico de marcadores individuales en los distintos grupos de muestras con los marcadores seleccionados del GWAS con dislexia. pGWAS=valor de p obtenido en el estudio de genoma completo de dislexia. Casos/Controles=grupo de muestras utilizadas como casos o controles según la definición del recuadro de la página 91. Sólo se muestran los valores de  $p < 0.01$ . El valor de  $p < 9.9E-08$  se destaca con fuente roja.

			Estudio de Dislexia			Estudio de Comórbidos				Estudio de Falta de Atención					
Casos			D+C	D+C	D	C	C	C	C	FA+C	FA+C	FA+C	FA+C	FA	FA
Controles			Ctr_D	Ctr-P_D	Ctr-P_D	Ctr-P_D	Ctr-P_FA	Ctr-P_FA2	Ctr_C	Ctr_FA	Ctr-P_FA	Ctr_FA2	Ctr-P_FA2	Ctr-P_FA	Ctr-P_FA2
SNP	GEN	pGWAS													
rs9487396	DDO	7.87E-06	5.15E-05	1.59E-05	2.25E-05										
rs10120392	KANK1	8.76E-06	1.28E-06	2.62E-06	4.17E-08										
rs3739592	KANK1	5.30E-06	1.08E-05	2.69E-05	4.18E-07										
rs82977	LASP1	2.73E-05	0.00024	0.00024	0.00017										

Tabla 30. Resultados obtenidos mediante la tecnología TaqMan™ del estudio de asociación del modelo alélico de marcadores individuales en los distintos grupos de muestras con los marcadores seleccionados del GWAS con FA. pGWAS=valor de p obtenido en el estudio de genoma completo de FA. Casos/Controles=grupo de muestras utilizadas como casos o controles según la definición del recuadro de la página 91. Sólo se muestran los valores de  $p < 0.01$ .

			Estudio de Dislexia			Estudio de Comórbidos				Estudio de Falta de Atención					
Casos			D+C	D+C	D	C	C	C	C	FA+C	FA+C	FA+C	FA+C	FA	FA
Controles			Ctr_D	Ctr-P_D	Ctr-P_D	Ctr-P_D	Ctr-P_FA	Ctr-P_FA2	Ctr_C	Ctr_FA	Ctr-P_FA	Ctr_FA2	Ctr-P_FA2	Ctr-P_FA	Ctr-P_FA2
SNP	GEN	pGWAS													
rs4637064	BRE	0.0009								0.00072	0.00103	0.00106	0.00100	0.00075	0.00070
rs11127135	BRE	6.76E-06													
rs12470242	BRE	7.84E-06								4.20E-05	2.47E-05	9.08E-06	9.74E-06	2.10E-05	7.60E-06
rs11127143	BRE	4.00E-06								0.00030	0.00037	0.00060	0.00098	0.00003	0.00012
rs2244188	PCP4	1.85E-06								0.00217	0.00067	0.00014	0.00012	0.00035	0.00006

En cuanto a los haplotipos y ventanas corredizas replicados, si bien los genotipos se correspondían con los obtenidos en los arrays, los resultados de asociación no fueron tan significativos, probablemente por no haber seleccionado todos los marcadores correspondientes al haplotipo o ventana corrediza completo. El patrón de significancia de los tests que consideran FA se parece al observado con los SNPs individuales, es decir, se obtienen valores de p menores en las comparativas de FA vs Ctr-P<sub>FA</sub> o FA vs Ctr-P<sub>FA2</sub> en comparación con los test donde se consideran también las muestras comórbidas dentro de los casos. En cuanto a los relacionados con dislexia en cambio, esta diferencia no fue tan marcada (los marcadores que forman cada haplotipo o ventana corrediza así como los valores de p obtenidos en cada uno de los tests empleados se detallan en la tabla suplementaria 30).

El análisis epistático de los 48 marcadores seleccionados, se realizó mediante PLINK. Los valores de asociación obtenidos fueron mucho mayores (menos significativos) que los obtenidos mediante SIXPAC<sup>319</sup>, por lo que para que los resultados de ambos análisis fueran comparables, el análisis de estos 48 marcadores se realizó con las dos herramientas. La comparación de valores de p obtenidos se detalla en la Tabla 31. Tal y como se puede observar, tanto utilizando SIXPAC como PLINK, las asociaciones obtenidas mediante el genotipado de sondas TaqMan<sup>TM</sup> disminuyen con respecto a las obtenidas en el análisis de genoma completo (ver detalles de los valores de significancia obtenidos en cada uno de los tests en la tabla suplementaria 31).

Tabla 31. Valores de significancia obtenidos del análisis de epistasia con el software SIXPAC y el software PLINK con los genotipos obtenidos mediante arrays de genoma completo (GWAS) y mediante sondas TaqMan<sup>TM</sup> (TAQ). \*Resultados correspondientes al Test D vs Ctr-P<sub>D</sub>. #Resultados correspondientes al Test FA vs Ctr-P<sub>FA</sub>.

Gen 1	SNP 1	Gen 2	SNP 2	GWAS sixpac	GWAS plink	TAQ sixpac	TAQ plink
CNTNAP5	rs1529289	DOK6	rs1372770	7.80E-12	0.4486	1.83E-09 *	0.4137 *
CNTNAP5	rs1529289	DOK6	rs7238969	8.92E-12	0.4014	1.96E-09 *	0.4660 *
NLGN1	rs1545991	SH3GL2	rs7036989	8.17E-10	0.00018	1.54E-06 #	0.0033 #

#### 4.3.2.4 Estudio de asociación de genoma completo de CNVs

##### 4.3.2.4.1 Estudio de asociación de dislexia

Debido a la diferencia de calidad de datos obtenida con las diferentes versiones de arrays *Human OmniExpress* utilizados, para el estudio de dislexia no se pudo hacer un análisis de CNVs

(del inglés, *Copy Number Variations*) con los datos conjuntos de ambas versiones (ver apartado 4.3.1). De manera que, por una parte se analizaron los 184 casos y 390 controles genotipados con la versión 1.0 de los arrays y por otro lado los 61 casos y 466 controles genotipados con la versión 1.1. Los resultados de duplicaciones y deleciones significativas halladas en estas muestras están descritos en la Tabla 32. Con la versión 1.0 de los arrays, fueron 31 las regiones de CNV encontradas que contenían una duplicación y 6 regiones con una deleción, mientras que con los datos de la versión 1.1 sólo se encontró una duplicación (marcada con un asterisco en la Tabla 32). La longitud de dichas regiones se puede estimar a partir de las posiciones genómicas de inicio y fin que figuran en la primera columna de la tabla. Algunas de estas regiones resultaron ser muy largas y contenían una enorme cantidad de genes por lo que en la cuarta columna de la tabla aparecen puntos suspensivos indicando un mayor número de genes a parte de los citados. El número exacto de genes ubicados en las regiones se encuentra entre paréntesis.

Entre los genes duplicados o delecionados se encuentran varios con funciones de interés para el trastorno analizado. Siguiendo en la misma línea que en el análisis de marcadores individuales, grupales e interacción, se encontraron varios genes relacionados con canales iónicos, entre ellos los de calcio (*CACNA1H*), sodio (*SCNN1D*), potasio (*KCNK6*) y cloro (*CLCN7*) y genes con función neuronal: proteínas específicas neuronales (*PTX4*, *CEND1*), neuropeptidos (*NPW*), diferenciación de células gliales (*METRNL*) o relacionados con el neuroblastoma (*DVL1*, *TP73*) entre otros. También hay diversos genes con función en las uniones celulares y estructura del citoesqueleto (*ACRT3C*, *THEM204*, *EPS8L2*, *COL9A3*, *PCDHA1-5*). Hemos de mencionar también el número elevado de genes con secuencias específicas que codifican para estructuras que permiten la interacción con la molécula zinc (*ZNF10*, *TRAF7*, *CASZ1*, *DEAF1*, *DPF1*, *ZNF598*, *ZNF300*, *DPF1*), así como la presencia duplicada de los genes *DRD4* y *DYNGR3*, ambos relacionados con la ruta dopaminérgica. Entre los presentes, también destacan dos genes encargados del tráfico de vesículas (*RAB26*, *MAPK8IP3*) y otros dos relacionados con lípidos (*DGKQ*, *PNPLA2*).

Tabla 32. Duplicaciones y deleciones significativas encontradas en el estudio de dislexia. CNVR=región de número de copias variable. N° SNPs=número de marcadores genotipados dentro de cada CNV. Dist\_gen=distancia entre cada CNV definido y el gen mencionado en la columna "Gen". \*=Duplicación hallada a partir de los datos de genotipado de la versión 1.1 de los arrays.

Duplicaciones					
CNVR(hg18)	N° SNPs	Valor de p	Gen	Dist_gen	
chr1:1201632-1295323	28	1.81E-05	ACAP3...DVL1...UNQ662 (11)	0	
chr7:1462702-1508191	12	0.0004	INTS1 ,KIAA1440, MICALL2	0	
chr5:140165568-140185473	10	0.0005	PCDHA1-5	0	
chr8:145737286-145750506	8	0.0007	KIAA1688	0	
chr17:80012769-80034443	4	0.0008	METRNL	1366758	
chr9:139261086-139279173	12	0.0009	COBRA1,FAM166A,LOC441476	0	
chr22:20098359-20144663	13	0.0010	HIC2	0	
chr14:105660766-106020366	38	0.0016	BC011773...Z49973 (4)	0	
chr10:135023586-135121732	41	0.0017	ECHS1...SPRN (6)	0	
chr21:46934826-46958299	19	0.0023	PRMT2	25535	
chr11:382912-838760	36	0.0024	AK123483...DRD4...TSPAN4 (38)	0	
chr20:62568408-62641575	17	0.0029	C20orf69	163257	
chr16:87862007-87866399	17	0.0029	ANKRD11,AX747440	0	
chr16:818802-2199788	40	0.0039	AK05681, BAIAP3...ZNF598 (72)	0	
chr1:10678843-10743363	19	0.0047	CASZ1	0	
chr2:242712870-242755965	11	0.0062	LOC441309	0	
chr4:871675-965779	29	0.0080	DCKQ...TMEM175 (5)	0	
chr1:3368143-3585041	40	0.0092	ARHGEF16...WDR8 (5)	0	
chr1:2323146-2359280	18	0.0119	PEX10,RER1	0	
chr22:46356554-46519363	14	0.0130	AK093107,AK128136,BC039485	0	
chr11:2859055-2885728	14	0.0181	BWR1B...SLC22A18AS (4)	0	
chr1:6481842-6487031	5	0.0242	PLEKHG5	0	
chr19:629303-657666	14	0.0254	FSTL3,PRSSL1,UNQ674	0	
chr12:133130292-133164690	10	0.0264	ZNF268	840758	
chr12:19372357-19466956	8	0.0276	PLEKHA5	0	
chr20:60880653-60919960	18	0.0318	C20orf20,COL9A3,FLJ00084,OGFR	0	
chr2:239066379-239066379	1	0.0359	AK057376	17691	
chr19:23316947-23364565	4	0.0446	BC038574,ZNF91	0	
chr10:47568296-47568296	1	0.0447	FAM21B,FAM21D	98873	
chr7:27146653-27146807	2	0.0453	HOXA5	714	
chr11:64573589-64606177	11	0.0348	AB231703,CDCA5,NAALADL1	0	
*chr15:24369643-24433948	0	0.0313	GABRB3	0	
Deleciones					
CNVR(hg18)	N° SNPs	Valor de p	Gen	Dist_gen	
chr19:43372386-43513659	1	0.0030	C19orf33,DPF1...YIF1B(8)	0	
chr11:55367889-55424207	17	0.0039	SPRYD5	0	
chr5:150254041-150254620	2	0.0361	ZNF300	0	
chr19:41333284-41333284	15	0.0414	CAPNS1	189	
chr15:34718594-34807851	12	0.0420	C15orf41	0	
chr19:28036744-28036744	0	0.0420	LOC100101266	3898655	

#### 4.3.2.4.2 Estudio de asociación de falta de atención

Todas las muestras del estudio de falta de atención se genotiparon con la versión 1.1 de los arrays, por lo que sólo hizo falta realizar un análisis. Los resultados de las duplicaciones y deleciones significativas se describen en la Tabla 33. En total se detectaron 4 duplicaciones y 3 deleciones que pasaron el umbral de significancia de 0.05.

Tabla 33. Duplicaciones y deleciones significativas encontradas en el estudio de falta de atención. CNVR=región de número de copias variable. N° SNPs=número de marcadores genotipados dentro cada CNV. Dist\_gen=distancia entre cada CNV definido y el gen mencionado en la columna "Gen".

Duplicaciones				
CNVR(hg18)	N° SNPs	Valor de p	Gen	Dist_gen
chr10:47543322-47663696	30	0.0111	FAM21B,FAM21D	73899
chr12:63979105-64108839	13	0.0145	MSRB3	0
chr19:53518747-53518747	1	0.0440	EMP3	1694
chr7:5823221-5888663	16	0.0179	OCM,ZNF815	0
Deleciones				
CNVR(hg18)	N° SNPs	Valor de p	Gen	Dist_gen
chr18:1908848-1972346	33	0.0027	C18orf2	511667
chr15:24357212-24472002	18	0.0027	GABRB3	0
chr4:16880727-168893805	17	0.0222	CR614646	341212

Entre los genes duplicados o delecionados en estos CNVs y de los que se dispone información funcional, destacan el gen *OCM* por codificar una proteína que se une al calcio, los genes *MSRB3* y *ZNF815* por su implicación con el zinc y el gen *GABRB3*, ya que además de codificar una subunidad del canal de cloro, también es un neurotransmisor inhibitor del sistema nervioso.

#### 4.3.2.4.3 Validación tecnológica de los CNVs de mayor asociación genética

Entre las duplicaciones y deleciones encontradas en los estudios de dislexia y falta de atención se eligieron 8 CNVs para replicar utilizando una tecnología diferente a la de los arrays de genoma completo. Los CNVs se eligieron en función del interés funcional de los genes que contenían. Dependiendo de la longitud del CNV identificado en los GWAS, se seleccionaron 2, 3 o 6 sondas por región, para validar dichos resultados. Las 22 sondas ubicadas en los 8 CNVs mencionados (Tabla 6) se genotiparon en 247 muestras disléxicas, 1421 muestras control de dislexia, 187 muestras diagnosticadas con falta de atención, 520 muestras control para la falta de

atención y 45 muestras comórbidas. 112 controles servían como controles para ambos trastornos. Los resultados de 9 de las 22 sondas no llegaron al límite de confianza necesario ( $>0.95$ ) para adjudicarlos como “datos fiables”. Con el resto, si bien se genotiparon de una manera óptima, no se validaron los resultados de número de copias establecido mediante los arrays. En la mayoría de las muestras el número de copias establecido por la tecnología TaqMan™ fue de 2, por lo que entre los resultados obtenidos mediante los datos de genoma completo y los obtenidos por esta técnica alternativa no hubo concordancia. Incluso, en algunos casos, las sondas seleccionadas dentro una misma región de CNV tampoco fueron concordantes entre ellas. Esto se observó con la región situada en el cromosoma 11 (Figura 21), para la que los datos de los arrays de genotipado de genoma completo identificaron una duplicación en la muestra representada en la imagen.

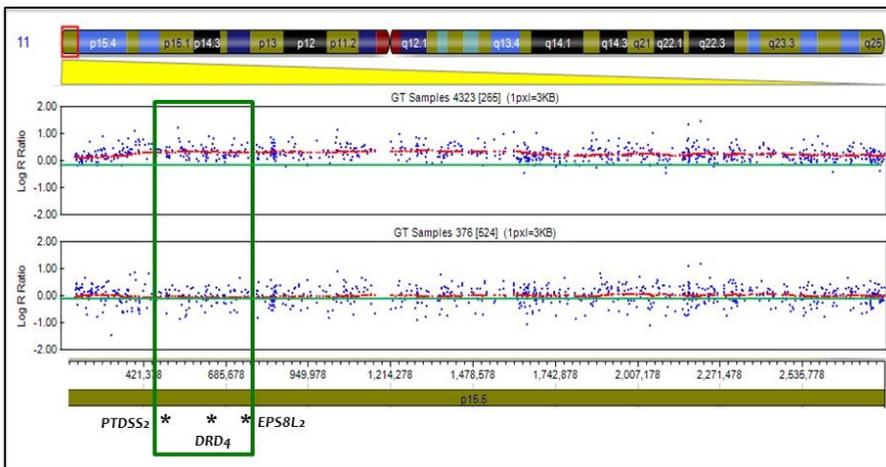


Figura 21. Representación del logR ratio (LRR) de una región de CNV situada en el cromosoma 11 a partir de los datos de fluorescencia obtenidos mediante los arrays de genoma completo. En esta imagen se visualiza tan sólo la región comprendida en el recuadro rojo, en la izquierda de la representación del cromosoma 11 completo. Dentro del recuadro verde se encuentra la región exacta de CNV cuestionada mediante sondas TaqMan™. En la parte inferior de la imagen se pueden observar mediante asteriscos la ubicación de las sondas empleadas así como los genes dónde se ubican. El gráfico de la parte superior corresponde a una muestra con duplicación y el de abajo a una muestra con 2 copias para la región representada, según la definición de segmentos obtenida mediante PennCNV<sup>321</sup>.

La duplicación se puede observar por el aumento en el logR Ratio (por encima del valor 0) con respecto a la muestra de dos copias. Sin embargo, las 3 sondas elegidas para replicar este CNV dieron unos resultados discordantes para esta misma muestra que contenía la duplicación: la sonda situada en el gen *PTDSS2* no identificaba ninguna copia en esta región (delección), la sonda situada en el gen *DRD4* la consideraba de dos copias, mientras que la sonda ubicada en el gen *EPS8L2* la identificaba como de 3 copias.

## 4.4 ESTUDIO DE ASOCIACION DE QTLs EN LA POBLACIÓN GENERAL

### 4.4.1 Estudio de asociación con los genotipos relativos a los genes candidatos seleccionados

Para interpretar los datos de genotipado en base a las características fenotípicas registradas de cada muestra, se procedió a realizar un análisis de locus cuantitativos (QTL) en primer lugar para los SNPs y VNTRs candidatos. Así, se utilizaron los genotipos de 3414 muestras, es decir, todas las muestras de niños con origen español independientemente de su diagnóstico, los cuales se contrastaron con las 174 variables derivadas de las pruebas psicométricas realizadas a todos los individuos. Los resultados más destacables de la relación entre los genotipos de los marcadores candidatos y las 174 variables cuantitativas evaluadas en la población, se encuentran resumidos en la Tabla 34. En dicha tabla se indica con un “SI” la asociación de alguna de las variables del conjunto que componen la tarea cognitiva, es decir, si el correspondiente SNP mostró un valor de  $p$  menor a 0.05 al menos en una de las variables medidas para esa tarea. En los siguientes apartados (4.1.1 y 4.1.2) se describen los detalles de dichas asociaciones.

Tabla 34. Relación de los SNPs y VNTRs analizados con las variables medidas para cada una de las tareas cognitivas estudiadas. Freq=frecuencia, RAN=Rapid Automatic Naming, ANT=Attentional Network Test, SI=relación significativa para alguna variable de la tarea cognitiva (valor de  $p < 0.05$  tras la corrección de falsos positivos Benjamin-Hochberg), vacío=sin relación con ninguna de las variables.

		KIAA0319(rs450446)	FOXP2(rs12533005)	DCDC2(rs2274305)	DYX1C1(rs57809907)	DBH(rs1611115)	COMT(rs4680)	MAOA(rs6323)	DAT1	DRD4
Lectura	Global							SI	SI	
	Palabras									
	Alta freq	SI			SI			SI	SI	
	Baja freq							SI	SI	
	Decodificación pseudopalabras							SI	SI	
Lexicalidad					SI	SI				
Conciencia fonológica				SI		SI		SI		SI
RAN		SI		SI	SI			SI		
Discriminación silábica		SI				SI		SI		SI
STROOP	Verbal		SI		SI				SI	SI
	Númérico	SI	SI		SI			SI	SI	SI
ANT							SI			SI

#### 4.4.1.1 Correlaciones entre variables cuantitativas y genotipos de los VNTRs candidatos

Entre los resultados más remarcables cuando consideramos la presencia/ausencia de los alelos analizados para cada uno de los VNTRs frente a las variables cognitivas, se encontraron los siguientes:

- En el análisis correspondiente a las variables que miden la **habilidad para la lectura** (variables dentro de la tarea de decodificación de grafema-fonema), tanto la presencia del alelo 09R como la del 10R del VNTR situado en el gen *DAT1* (en adelante 09R-*DAT1* y 10R-*DAT1*) se relacionaron con el tiempo de reacción (TR) de lectura de palabras y pseudopalabras. El alelo 10R-*DAT1* también se asoció con el número de aciertos en la lectura de palabras.

- El VNTR situado en el gen *DRD4* en cambio, no resultó asociado en esta tarea y ninguno de los dos VNTRs lo hizo tampoco con el test de nombramiento automático rápido RAN (del inglés *Rapid Automatic Naming*) que evalúa la **fluidez** de la lectura.
- En lo referente a la **conciencia fonológica**, la presencia del alelo 05R del gen *DRD4* (05R-*DRD4*) se asoció al TR en la tarea de dibujos. También se relacionó con el número de aciertos en esta misma tarea así como en la correspondiente a la eliminación del primer fonema. La presencia del alelo de 02R de este mismo gen por otro lado, presentó una asociación aislada con el número de aciertos en la tarea de eliminación del primer fonema.
- Con respecto a la **percepción del habla**, fue el 03R-*DRD4* el que superó el umbral de significancia en relación con el número de aciertos al discriminar entre sílabas diferentes y sílabas iguales, incluso llegando a ser significativo al analizar globalmente estos dos tipos de sílabas. El alelo 07R-*DRD4*, si bien no obtuvo valores significativos individualmente, sí que llegó a ser significativo a nivel global.
- Centrándonos en las variables atencionales, encontramos que ambos genes se asociaron tanto a las tareas de Stroop verbal como a las de Stroop numérico. El alelo 03R-*DRD4* se asoció al número de aciertos en ambas tareas y al TR en la tarea verbal. Sin embargo, el alelo 09R-*DAT1* se relacionó con el TR del Stroop numérico mientras que 10R-*DAT1* lo hizo con el número de respuestas correctas en el Stroop verbal.
- En cuanto a los resultados del test **ANT (Attentional Network Task)**, el alelo 02R-*DRD4* se asoció con el índice de conflicto en errores, calculado como la resta entre el número de errores de las situaciones incongruentes y congruentes. (Los valores de p obtenidos en las variables descritas anteriormente se encuentran en la tabla suplementaria 32).

#### 4.4.1.2 Comparativa de las variables fenotípicas cuantitativas con los genotipos de los SNPs candidatos en la población de estudio

Los resultados de asociación más destacados de la correlación entre los genotipos de los SNPs candidatos y las variables cognitivas se resumen de la siguiente manera:

- Entre las tareas correspondientes a la **habilidad para la lectura**, el marcador rs6323-*MAOA* mostró un patrón de asociación consistente con el tiempo de lectura de palabras y pseudopalabras así como al compuesto de ambas. Además, rs4504469-*KIAA0319* y

rs57809907-DYX1C1 mostraron asociación con los aciertos en la lectura de palabras de alta frecuencia. Este último SNP junto con el rs1611115-DBH fueron también asociados significativamente al efecto de la lexicalidad en aciertos (la diferencia entre los aciertos de palabras y pseudopalabras). En cuanto al resto de genes, aunque si se observó relación con algunas estructuras de palabras específicas, no llegaron a ser significativas a un nivel más global del análisis.

- En lo referente a la **fluidez para la lectura**, evaluada mediante la tarea de RAN, la asociación más llamativa se obtuvo con los marcadores rs4504469-KIAA0319, rs2274305-DCDC2, rs57809907-DYX1C1 y rs6323-MAOA con respecto al número de aciertos en el compuesto formado por las tareas de colores y dibujos. Además, rs4504469-KIAA0319 también mostró asociación con el número de aciertos en estas tareas por separado y rs6323-MAOA únicamente a la tarea de colores.
- En cuanto al proceso cognitivo de la **conciencia fonológica**, los marcadores situados en los genes DCDC2 y DBH se relacionaron significativamente al TR global de las tareas de identificación de fonemas en dibujos (al compuesto de las tareas utilizando dos tipos de distractor diferentes). Por otra parte, rs6323-MAOA se relacionó con el TR de la tarea de eliminación del primer fonema.
- En el análisis de las variables correspondientes a la **percepción del habla**, el marcador rs1611115 situado en el gen DBH y el rs4504469 situado en el gen KIAA0319 se asociaron al número de aciertos en la tarea de discriminación silábica (sílabas iguales o diferentes), mientras que rs6323-MAOA lo hizo al TR únicamente cuando las sílabas eran diferentes, y también al considerar las dos tareas en conjunto.
- Finalmente, en lo referente a las tareas atencionales, los marcadores rs4504469-KIAA0319, rs6323-MAOA, rs57809907-DYX1C1 y rs12533005-FOXP2 se relacionaron con el efecto conflicto calculado por las tareas de Stroop numérico. Específicamente, KIAA0319 y MAOA se correlacionaron con el número de aciertos así como al TR mientras que DYX1C1 se asoció sólo al número de aciertos y FOXP2 solo al tiempo. Además, estos dos últimos genes también fueron significativos en las tareas de Stroop verbal de la misma forma: DYX1C1 con el número de aciertos y FOXP2 con el tiempo.
- Por otro lado, los resultados del test **ANT (Attentional Network Test)** mostraron que el marcador rs4680-COMT se asociaba al índice de conflicto el cual se obtiene mediante la resta del TR de todas las situaciones incongruentes y el TR de todas las congruentes. (Los

valores de p obtenidos en las variables descritas anteriormente se encuentran en la tabla suplementaria 33).

#### **4.4.2 Asociación de valores cuantitativos pertenecientes a rasgos cognitivos en la población general a partir de los datos de genotipado de genoma completo**

El siguiente apartado de resultados está dividido en dos secciones: 1) los resultados QTL obtenidos en base a las 5 **variables de lectura** principales, las cuales se han analizado con las muestras correspondientes al estudio de dislexia, al estudio de FA y al estudio de comórbidos (análisis realizados en cada grupo de muestras por separado y sin tener en cuenta el diagnóstico de caso/control) y 2) los resultados QTL obtenidos en base a las 5 **variables atencionales** seleccionadas, las cuales únicamente se han analizado en las muestras correspondientes al estudio de FA.

Para cada uno de los análisis realizados (modelos genéticos lineal, GxE – donde se compara la diferencia entre dos coeficientes de regresión entre casos y controles-, dominante, recesivo y genotípico) la magnitud de resultados generados ha sido enorme, por lo que se han simplificado en tablas (Tabla 35, Tabla 36 y Tabla 37 para las variables de lectura y Tabla 38 para las variables atencionales), indicando solamente los valores menores de p obtenidos para cada variable y modelo genético empleado, así como el número de marcadores que obtuvieron un valor de  $p < 9.9E-05$ .

##### **4.4.2.1 Rasgos cognitivos relacionados con la lectura**

En la Tabla 35 se incluyen los resultados más significativos para cada una de las 5 variables de lectura analizadas. Los valores de p obtenidos para cada uno de los SNPs mencionados se encuentran en la tabla suplementaria 34.

Tabla 35. Descripción de los resultados de asociación entre genotipos y las 5 variables de lectura utilizando los datos de 1079 muestras (238 disléxicos y 841 controles en conjunto), bajo distintos modelos genéticos (Lineal, GxE=interacción entre genotipo y diagnóstico, Dom=dominante, Rec=recesivo, Geno=genotípico). Var x=identificador de la variable, p\_min=valor de p menor encontrado, N° SNPs=número de marcadores con valor de  $p < 9.9E-05$ . Los valores de  $p < 9.9E-08$  están remarcados en negrita y los valores de  $p < 9.9E-09$  subrayados.

		Población con Dislexia				
		Var 1	Var 2	Var 3	Var 4	Var 5
Lineal	p_min	2.08E-07	1.12E-06	1.73E-07	1.53E-06	3.67E-06
	N° SNPs	101	86	94	64	85
GxE	p_min	5.10E-06	2.49E-07	1.57E-07	3.94E-06	7.96E-06
	N° SNPs	71	85	87	58	59
Dom	p_min	<b>1.66E-08</b>	3.23E-07	<b>8.37E-08</b>	2.41E-06	9.42E-07
	N° SNPs	78	89	89	80	75
Rec	p_min	<b>7.31E-08</b>	5.24E-07	<b>9.58E-13</b>	9.40E-07	2.33E-06
	N° SNPs	149	77	416	75	63
Geno	p_min	<b>3.33E-08</b>	1.11E-06	<b>3.27E-12</b>	3.64E-06	9.89E-07
	N° SNPs	137	82	292	88	83

**Var 1**= “Efecto de transposición en aciertos”: Mide el procesamiento ortográfico.

**Var 2**= “Tiempo de nombrado rápido compuesto de dibujos y colores”: Mide la fluidez de lectura y la velocidad de procesamiento.

**Var 3**= “Tiempo global en discriminación silábica”: Mide la percepción del habla.

**Var 4**=“Tiempo global en identificación de fonema”: Mide la conciencia fonológica.

**Var 5**=“Eficiencia de palabras y pseudopalabras”: Mide la decodificación grafema-fonema.

Las variables 1 y 3 fueron las que mostraron valores de p menores bajo los modelos genéticos dominante, recesivo y genotípico, llegando incluso al umbral de significancia establecido para los estudios de genoma completo ( $E-08$ ). Aun así no se observó ningún conjunto de marcadores destacables para la variable 3. Si bien la variable 1 no llegó a valores del orden de  $E-08$  analizándolo con el modelo lineal, entre las asociaciones más significativas con esta variable, se encontró una región en el cromosoma 11 que se desmarcaba del resto (ver Figura 22).

Fueron 24 los marcadores del cromosoma 11 con  $p < 9.9E-05$  y los genes en los cuales se localizan son: *DLG2*, *CNTN5*, *H2AFX*, *DPGAT1*, *C2CD2L*, *HINFP*, *NRLX1*, *PDZD3*, *CBL* y *MCAM*. Entre estos marcadores, tres codifican una mutación que cambia la pauta de traducción aminoacídica (rs643788, dentro del gen *DPGAT1*, rs4245191, dentro del gen *NRLX1* y rs1815811 dentro del gen *PDZD3*).

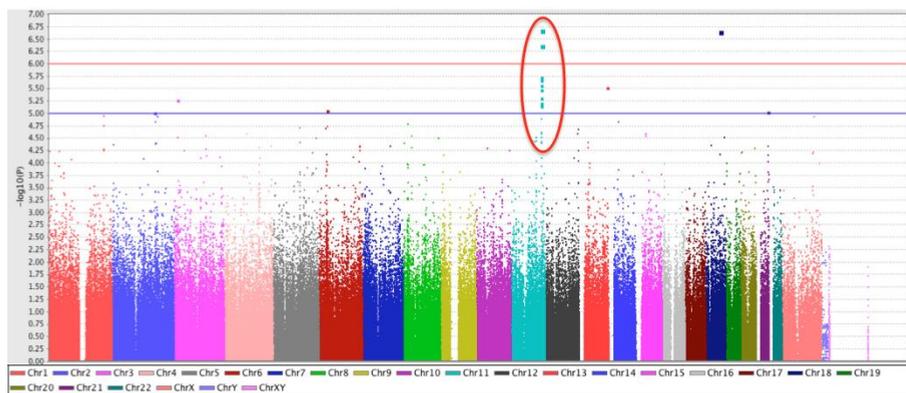


Figura 22. Gráfico de Manhattan obtenido a partir de los valores de asociación de la variable 1 bajo el modelo lineal. En el eje Y se representa el valor de significancia de cada uno de los marcadores en una escala logarítmica. La línea horizontal azul señala un valor de  $p=1E-05$  y la línea horizontal roja un valor de  $p=1E-06$ . Los marcadores están ordenados según su posición cromosómica a lo largo del eje X. La región del cromosoma 11 más significativa se ha resaltado con un círculo rojo.

La variable 2, aunque mostró valores de  $p$  mayores que la variable 1, también demostró estar especialmente asociada con un grupo de marcadores en el cromosoma 13, tanto en el modelo lineal como en el modelo dominante (ver Figura 23).

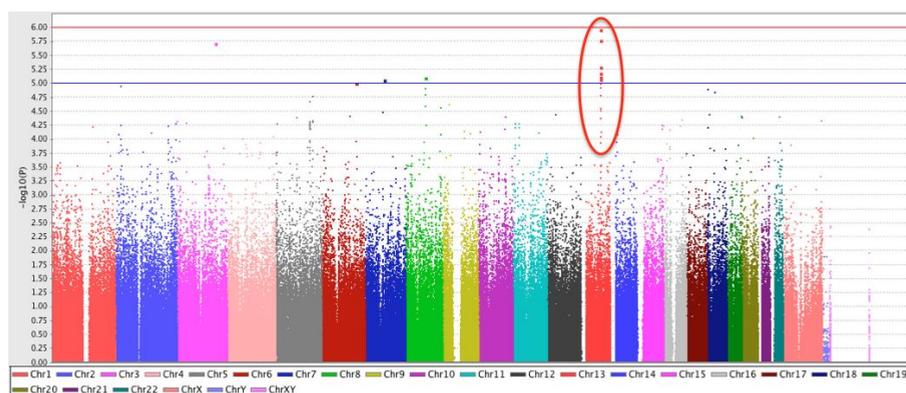


Figura 23. Gráfico de Manhattan obtenido a partir de los valores de asociación de la variable 2 bajo el modelo lineal. En el eje Y se representa el valor de significancia de cada uno de los marcadores en una escala logarítmica. La línea horizontal azul se encuentra a la altura de un valor de  $p=1E-05$  y la línea horizontal roja en un valor de  $p=1E-06$ . Los marcadores están ordenados según su posición cromosómica a lo largo del eje X. La región del cromosoma 13 más asociada es la destacada con un círculo rojo.

El total de marcadores asociados con la variable 2 con valor de  $p < 9.9E-05$  y posicionados en el cromosoma 13 fueron 15 y ninguno de ellos se localizaba dentro de genes.

Cabe mencionar que en este análisis 3 marcadores ubicados en el gen MAOA (rs5906893, rs5906938 y rs6520894) obtuvieron valores de p del orden de E-05 con la variable 3 bajo el modelo genético lineal y que el marcador rs1358278 del gen FOXP2 también obtuvo un valor de significancia de ese mismo orden para la variable 5 bajo el mismo modelo.

La descripción de los resultados de asociación obtenidos para las 5 variables de lectura, pero utilizando esta vez los datos genotípicos de los niños con falta de atención y sus correspondientes controles, se encuentran en la Tabla 36. El listado de valores de p obtenidos para los SNPs citados se encuentra en la tabla suplementaria 35.

Tabla 36. Descripción de los resultados de asociación entre genotipos y las 5 variables de lectura utilizando los datos de 488 muestras (136 muestras con falta de atención y 352 controles en conjunto), bajo distintos modelos genéticos (Lineal, GxE=interacción entre genotipo y diagnóstico, Dom=dominante, Rec=recesivo, Geno=genotípico). Var x=identificador de la variable, p\_min=valor de p menor encontrado, N° SNPs=número de marcadores con valor de p<9.9E-05. Los valores de p<9.9E-09 están subrayados.

		Población con Falta de Atención				
		Var 1	Var 2	Var 3	Var 4	Var 5
Lineal	p_min	2.06E-06	1.85E-06	1.45E-07	9.17E-06	4.72E-06
	N° SNPs	100	50	124	51	67
GxE	p_min	1.03E-07	1.24E-06	<u>7.46E-09</u>	3.35E-06	3.97E-07
	N° SNPs	95	73	170	63	106
Dom	p_min	1.74E-06	4.65E-06	1.02E-07	8.90E-06	3.11E-06
	N° SNPs	80	42	105	46	55
Rec	p_min	4.93E-07	1.83E-06	<u>1.45E-11</u>	6.09E-07	2.16E-06
	N° SNPs	96	47	812	57	54
Geno	p_min	2.16E-06	7.74E-07	<u>3.03E-11</u>	2.72E-06	1.39E-06
	N° SNPs	107	68	625	52	205

**Var 1**= “Efecto de transposición en aciertos”: Mide el procesamiento ortográfico.

**Var 2**= “Tiempo de nombrado rápido compuesto de dibujos y colores”: Mide la fluidez de lectura y la velocidad de procesamiento.

**Var 3**= “Tiempo global en discriminación silábica”: Mide la percepción del habla.

**Var 4**=“Tiempo global en identificación de fonema”: Mide la conciencia fonológica.

**Var 5**=“Eficiencia de palabras y pseudopalabras”: Mide la decodificación grafema-fonema.

La única variable que obtuvo un valor de significancia menor a 9,9E-08 fue la variable 3, que mostró los valores menores de p en los modelos genéticos de interacción (GxE), recesivo y genotípico. Tal y como se puede observar en la Tabla 36, tanto en el modelo genotípico como en el recesivo el número de marcadores con un valor de p< 9.9 E-05 fue elevado.

Al igual que en los dos casos anteriores, también se utilizó la población de muestras que presentaban ambos trastornos (comórbidos) y los controles compartidos para analizar la posible asociación de los marcadores con las variables de lectura en esta población. La descripción de los resultados de dichos análisis se encuentra en la Tabla 37. Los valores de p de los marcadores más asociados para cada una de las variables y modelos genéticos empleados se pueden ver en la tabla suplementaria 36.

Tabla 37. Descripción de los resultados de asociación entre genotipos y las 5 variables de lectura utilizando los datos de 147 muestras (45 comórbidos y 102 controles en conjunto), bajo distintos modelos genéticos (Lineal, GxE=interacción entre genotipo y diagnóstico, Dom=dominante, Rec=recesivo, Geno=genotípico). Var x=identificador de la variable, p\_min=valor de p menor encontrado, N° SNPs=número de marcadores con valor de  $p < 9.9 \times 10^{-5}$ . Los valores de  $p < 9.9 \times 10^{-8}$  están remarcados en negrita y los valores de  $p < 9.9 \times 10^{-9}$  subrayados.

		Población de Comórbidos				
		Var 1	Var 2	Var 3	Var 4	Var 5
Lineal	p_min	<b>7.48E-08</b>	4.66E-07	<b>2.69E-08</b>	2.53E-06	2.49E-06
	N° SNPs	127	53	160	83	57
GxE	p_min	4.11E-07	1.38E-07	<b><u>6.84E-12</u></b>	1.38E-07	1.46E-06
	N° SNPs	150	125	161	135	96
Dom	p_min	2.96E-07	<b>4.86E-08</b>	<b>6.16E-08</b>	6.23E-07	3.86E-06
	N° SNPs	103	69	115	74	55
Rec	p_min	<b>2.63E-08</b>	2.70E-06	<b><u>1.58E-11</u></b>	2.23E-06	1.72E-07
	N° SNPs	126	56	1406	50	33
Geno	p_min	<b>3.01E-08</b>	<b>6.77E-08</b>	<b><u>5.88E-14</u></b>	<b><u>2.20E-10</u></b>	2.88E-07
	N° SNPs	398	194	1246	102	90

**Var 1**= “Efecto de transposición en aciertos”: Mide el procesamiento ortográfico.

**Var 2**= “Tiempo de nombrado rápido compuesto de dibujos y colores”: Mide la fluidez de lectura y velocidad de procesamiento.

**Var 3**= “Tiempo global en discriminación silábica”: Mide la percepción del habla.

**Var 4**=“Tiempo global en identificación de fonema”: Mide la conciencia fonológica.

**Var 5**=“Eficiencia de palabras y pseudopalabras”: Mide la decodificación grafema-fonema.

En este análisis volvió a aparecer la variable 3 con valores de p aún más bajos. Los modelos recesivo y genotípico resultaron tener muchísimos marcadores con valores  $p < 9.9 \times 10^{-5}$  repartidos a lo largo del todo el genoma, por lo que ninguno de ellos resultó destacar sobre el resto. Entre los marcadores más significativos del modelo GxE se encontraron el SNP rs1864952 ubicado en el gen *SLIT3* con un valor de  $p = 6.84 \times 10^{-12}$ , rs1346891 correspondiente al gen *DLG2* con un valor de  $p = 1.37 \times 10^{-11}$ , rs2223065 con significancia de  $2.03 \times 10^{-11}$ , rs10912866 localizado en el gen *CACYBP* y un

valor de  $p=3.65E-10$  y los marcadores rs11806870 y rs12072529 ubicados en el gen *RABGAP1L* con valores del orden de  $p=E-10$ .

Sólo un marcador (rs8055964, cromosoma 16, *LOC105371370*) obtuvo un valor de  $2.2E-10$  con la variable 4, el resto se mantuvo en valores en torno a  $1E-06$ .

#### 4.4.2.2 Rasgos cognitivos relacionados con la atención

Las 5 variables atencionales seleccionadas se analizaron únicamente en la población compuesta por niños con falta de atención y sus controles. Los resultados obtenidos del estudio de asociación de las 5 variables frente a los genotipos del genoma completo de estas muestras se encuentran resumidos en la Tabla 38. Los valores de  $p$  obtenidos para los marcadores más asociados en cada una de las variables y modelo genético empleado se encuentran en la tabla suplementaria 37.

Tabla 38. Descripción de los resultados de asociación entre genotipos y las 5 variables de atención utilizando los datos de 488 muestras (136 muestras con falta de atención y 352 controles en conjunto), bajo distintos modelos genéticos (Lineal, GxE=interacción entre genotipo y diagnóstico, Dom=dominante, Rec=recesivo, Geno=genotípico). Var x=identificador de la variable, p\_min=valor de  $p$  menor encontrado, N° SNPs=número de marcadores con valor de  $p < 9.9E-05$ . Los valores de  $p < 9.9E-08$  están remarcados en negrita y los valores de  $p < 9.9E-09$  subrayados.

		Población con Falta de Atención				
		Var 6	Var 7	Var 8	Var 9	Var 10
Lineal	p_min	<b><u>1.00E-13</u></b>	3.89E-06	4.96E-07	1.16E-06	5.84E-07
	N° SNPs	261	64	100	79	48
GxE	p_min	<b><u>1.30E-25</u></b>	6.24E-07	4.70E-07	1.59E-06	1.14E-06
	N° SNPs	431	101	117	105	111
Dom	p_min	<b><u>7.47E-09</u></b>	4.56E-06	1.04E-06	1.18E-06	2.87E-06
	N° SNPs	166	50	93	76	55
Rec	p_min	<b><u>7.47E-26</u></b>	<b><u>5.96E-09</u></b>	<b><u>4.36E-10</u></b>	4.83E-07	<b><u>7.78E-08</u></b>
	N° SNPs	1837	431	391	58	353
Geno	p_min	<b><u>1.00E-27</u></b>	<b><u>1.70E-08</u></b>	<b><u>1.49E-09</u></b>	1.10E-06	3.29E-07
	N° SNPs	1499	280	274	71	192

**Var 6**= “Efecto en aciertos de condiciones incongruente menos congruente (Stroop numérico)”

**Var 7**= “Efecto en tiempos de condiciones incongruente menos congruente (Stroop numérico)”

**Var 8**= “Efecto en aciertos de condiciones incongruente menos congruente (Stroop verbal)”

**Var 9**= “Efecto en tiempos de condiciones incongruente menos congruente (Stroop verbal)”

Las variables 6, 7, 8 y 9 miden el procesamiento automático, la atención selectiva y el efecto conflicto.

**Var 10**= “Índice conflicto”: Mide la alerta, orientación y efecto conflicto.

La variable 6 obtuvo los valores de  $p$  menores en todos los modelos genéticos empleados. Si bien el valor más significativo del modelo GxE fue de  $1.30E-25$ , sólo dos SNPs fueron los que llegaron hasta esta significancia (rs12121570 del cromosoma 1 y rs274849 del cromosoma 2 y ubicado en el gen *LOC105373661*). Otros dos marcadores (rs6082725 (chr 20) y rs7929179 (chr 11) con localización intergénica) llegaron hasta el orden de  $p=1E-12$  y fueron los terceros más asociados.

Los modelos genotípico y recesivo obtuvieron un número elevado de marcadores significativos con ubicaciones cromosómicas variables por lo que no se pudo establecer ninguna región específica como asociada a esta variable.

Entre los resultados de la asociación lineal, se encontraron dos regiones, una ubicada en el cromosoma 8 y otra en el cromosoma 2, con asociaciones de marcadores claramente destacables con respecto al resto (Figura 24). Específicamente, el marcador rs11989470 (chr 8) obtuvo un valor de  $p=1E-13$  y el SNP rs670302 (chr 2) un valor de  $p=2.33E-08$ . La misma región ubicada en el cromosoma 8 obtuvo con el modelo genotípico valores del orden entre  $1E-27$  y  $1E-25$ .

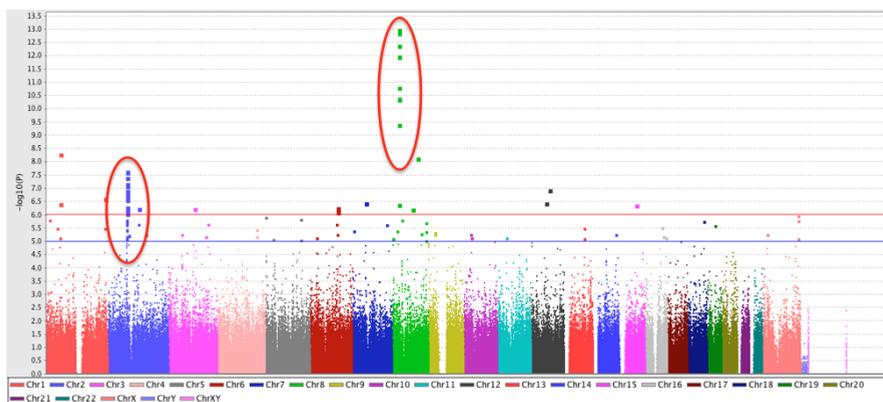


Figura 24. Gráfico de Manhattan obtenido a partir de los valores de asociación de la variable 6 bajo el modelo lineal. En el eje Y se representa el valor de  $p$  de cada uno de los marcadores en una escala logarítmica. La línea horizontal azul señala el valor de  $p=1E-05$  y la línea horizontal roja un valor de  $p=1E-06$ . Los SNPs analizados están ordenados según su posición cromosómica a lo largo del eje X. Las regiones más significativas de los cromosomas 2 y 8 se han destacado con círculos rojos.

Centrándonos únicamente en las dos regiones cromosómicas que han mostrado mayor nivel de asociación con la variable 6, representamos los resultados mediante la herramienta LocusZoom<sup>328</sup> en dos gráficos que muestran el grado de LD entre los marcadores considerados y la anotación de los genes comprendidos en cada región (Figura 25 y Figura 26).

Si bien ninguno de los marcadores más asociados con la variable 6 del cromosoma 8 se ubica dentro de genes, en la Figura 25 se pueden observar los genes que hay en las zonas flanqueantes a dichos marcadores. En cambio la región del cromosoma 2, asociada también con la variable 6, resultó ser mucho más rica en genes (Figura 26).

Para las variables 7 y 8, aun habiendo marcadores asociados muy significativamente no se encontró ninguna región cromosómica destacable.

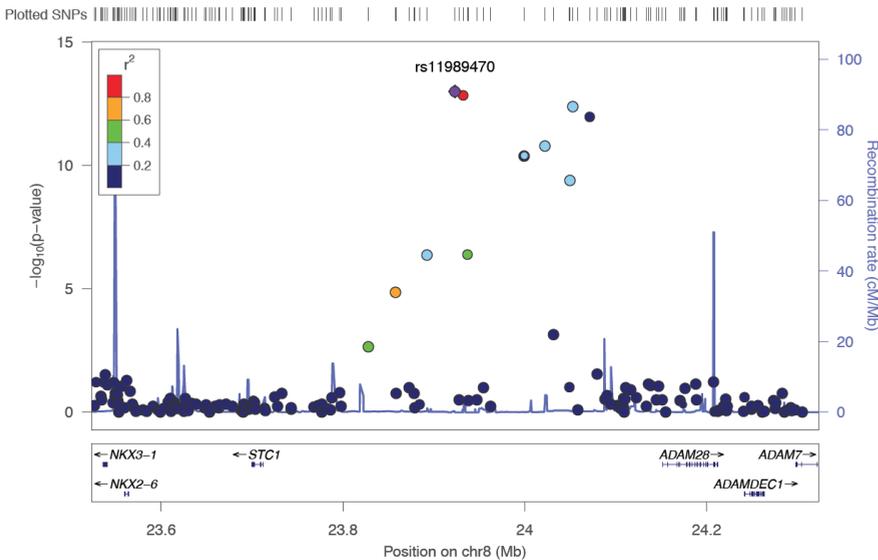


Figura 25. Representación gráfica de los valores de significancia en escala logarítmica (eje Y parte izquierda) del marcador rs11989470 (el más significativo) con la variable 6 y los SNPs de su entorno en la región del cromosoma 8, y visualización del grado de desequilibrio entre los SNPs representado por la escala de color de  $r^2$  (desequilibrio de ligamiento). En la parte inferior de la figura se pueden observar los genes ubicados en esta región. La línea azul representa la tasa de recombinación.

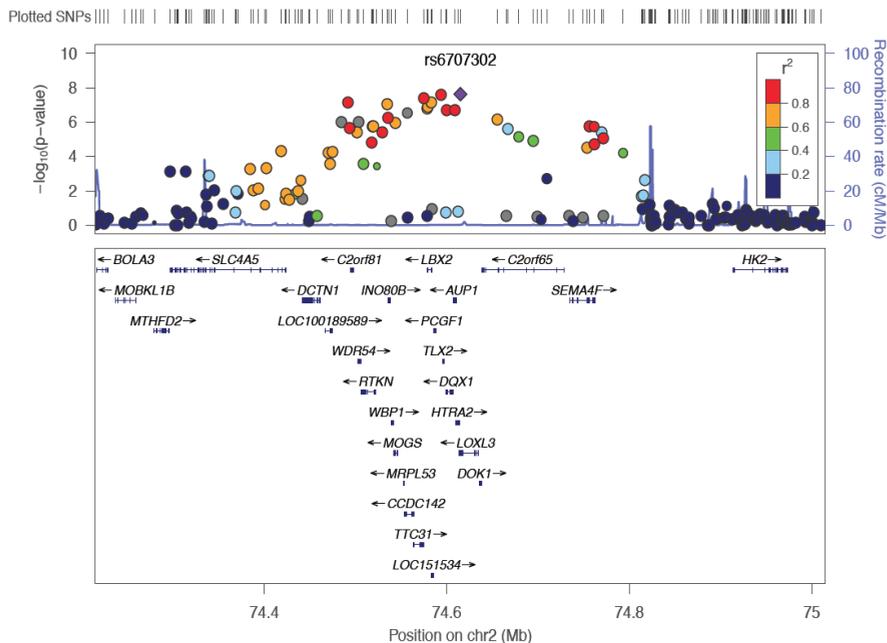


Figura 26. Representación gráfica de los valores de significancia en escala logarítmica (eje Y parte izquierda) del marcador rs6707302 (el más significativo) y los de su entorno en la región del cromosoma 2 y visualización del grado de desequilibrio entre los SNPs representado por la escala de color de  $r^2$ . En la parte inferior de la figura se pueden observar los genes ubicados en esta región. La línea azul representa la tasa de recombinación entre los SNPs considerados.

## 4.5 ENRIQUECIMIENTO DE RUTAS FUNCIONALES A PARTIR DE LOS GENES MÁS ASOCIADOS

### 4.5.1 Enriquecimiento de rutas funcionales derivado de los estudios de asociación caso/control

En este apartado del estudio se realizó un análisis de enriquecimiento de rutas funcionales, considerando para ello la significancia génica hallada a partir de los SNPs bajo el modelo alélico en nuestra población. Este análisis se realizó utilizando dos descripciones de rutas funcionales: GO (Ontología genética, <https://www.ebi.ac.uk/QuickGO/>) y KEGG (<http://www.genome.jp/kegg/>). Puesto que el número de rutas significativas obtenidas fue enorme, estos resultados se agruparon por similitud de función a modo de simplificación, de

forma que se pudieran llevar a cabo comparaciones entre los tres estudios. Para dicha agrupación se han tenido en cuenta tan sólo las rutas funcionales con un valor de asociación de  $p < 9.9 \times 10^{-5}$ . (El listado de las rutas específicas agrupadas en cada uno de estos sets, así como el valor de significancia obtenido para cada una de ellas se encuentran en las tablas suplementarias 38 y 39, columnas E-F-G).

### **Enriquecimiento de rutas funcionales a partir de los genes más asociados con dislexia**

Las agrupaciones derivadas de las rutas con valor de  $p < 9.99 \times 10^{-5}$  identificadas según la base de datos GO (15 rutas en total) y las rutas identificadas según la base de datos KEGG (12 rutas en total), se encuentran detalladas en las Tabla 39 y Tabla 40. Para más detalles sobre las rutas específicas y valores de  $p$  obtenidos en referencia a las bases de datos GO y KEGG ver tablas suplementarias 38 y 39, columna E.

Tabla 39. Agrupaciones realizadas a partir de las rutas funcionales con  $p < 9.99 \times 10^{-5}$  (tras la corrección de Benjamini&Hochbaum por tests múltiples) según la definición de GO para el estudio de dislexia. Se muestran el número de rutas agrupadas en cada caso y el rango de valores de  $p$  de las significancias de enriquecimiento de las rutas que componen cada grupo.

Grupo	Nº rutas	Valor menor de p	Valor mayor de p
Adhesión/estructura celular	4	5.89E-06	6.97E-05
Neurona	4	9.57E-06	3.83E-05
Proteínas G acopladas	1	3.83E-05	3.83E-05
Cerebro (telencéfalo)	1	6.04E-05	6.04E-05
Canales	3	6.97E-05	8.28E-05
Señalización (unión de la calmodulina)	1	8.28E-05	8.28E-05
Proceso catabólico de amino ácidos de la familia del aspartato	1	8.28E-05	8.28E-05

Tabla 40. Rutas funcionales con  $p < 9.99E-05$  (tras la corrección de Benjamini&Hochbaum por tests múltiples) según la definición KEGG para el estudio de dislexia. ARVC= Cardiomiopatía arritmogénica ventricular derecha.

Ruta KEGG	Descripción de la ruta	Valor de p
hsa04020	Ruta de señalización del calcio	2.47E-06
hsa04360	Orientación axonal	2.47E-06
hsa05146	Amoebiasis	2.47E-06
hsa04510	Unión focal	1.27E-05
hsa05412	ARVC	1.27E-05
hsa05414	Cardiomiopatía dilatada	1.46E-05
hsa04970	Secreción salival	1.66E-05
hsa04270	Contracción vascular del músculo liso	1.81E-05
hsa02010	Transportadores ABC	2.49E-05
hsa04514	Moléculas de adhesión (CAMs)	7.07E-05
hsa04530	Uniones estrechas	8.14E-05
hsa05416	Miocarditis viral	8.14E-05

Entre las rutas más significativas según la base de datos GO se encontraron varias relacionadas con la adhesión y estructura celular, así como con neuronas. Según KEGG, en cambio, las relacionadas con la señalización de calcio fueron las más significativas, si bien estos canales también aparecieron entre las más asociadas de GO, seguidas de la ruta de orientación axonal. Entre los resultados de KEGG, destacan también rutas relacionadas con el corazón y adhesión o uniones celulares.

### **Enriquecimiento de rutas metabólicas a partir de los genes más asociadas con falta de atención**

Partiendo de los resultados obtenidos en la asociación alélica, se detectaron un total de 61 rutas enriquecidas con un valor de  $p < 9.9E-05$  según la base de datos GO, las cuales se concentraron en los 14 grupos detallados en la Tabla 41 (el listado de las rutas concretas así como el valor de p obtenido se puede ver en la tabla suplementaria 38, columna F). Fueron 2 los grupos de rutas predominantes: 1) las relacionadas con neuronas, tales como: transmisión y regulación sináptica, despolarización de la membrana, desarrollo de dendritas y reconocimiento de neuronas, y 2) las rutas relacionadas con canales iónicos, tanto los canales activados por ligando como los activados por voltaje, especialmente los canales de calcio de este último grupo, así como las rutas relacionadas con los receptores de glutamato y el transporte iónico. El resto de grupos mencionados en la tabla fueron rutas aisladas si bien obtuvieron valores de p significativos. Según la definición de KEGG en cambio, las rutas predominantes fueron las relacionadas con el corazón

y la ruta de señalización del calcio Tabla 42 (los detalles de estas rutas pueden consultarse en la tabla suplementaria 39, columna F).

Tabla 41. Agrupaciones realizadas a partir de las rutas funcionales con  $p < 9.99E-05$  (tras la corrección de Benjamini&Hochbaum por tests múltiples) según la definición de GO para el estudio de falta de atención. Se muestran el número de rutas agrupadas en cada caso y el rango de valores de p de enriquecimiento de las rutas que componen cada grupo.

Grupo	Nº de rutas	Valor de p menor	Valor de p mayor
Neuronas-dendritas	17	< E-11	3.97E-05
Canales	23	2.60E-11	7.74E-05
Proteínas G acopladas	3	1.58E-05	2.50E-05
Fosfatasa/kinasa (Señalización)	2	1.56E-09	1.56E-09
Cognición	3	5.89E-06	4.12E-05
Músculo esquelético	3	8.82E-06	2.61E-05
Riñón	2	1.58E-05	2.50E-05
Cerebro (telencéfalo)	1	3.93E-05	3.93E-05
Respuesta de alarma	1	2.21E-07	2.21E-07
Conducta locomotora	1	2.82E-07	2.82E-07
Adhesión/estructura	1	5.14E-05	5.14E-05
Regulación del transporte de aminas	1	7.78E-06	7.78E-06
Reconocimiento celular	1	8.37E-05	8.37E-05
Parte de la estructura encapsuladora externa	1	9.82E-05	9.82E-05

Tabla 42. Rutas funcionales con  $p < 9.99E-05$  (tras la corrección de Benjamini&Hochbaum por tests múltiples) según KEGG para el estudio de falta de atención. ARVC= Cardiomiopatía arritmogénica ventricular derecha; HCM=Cardiomiopatía hipertrófica.

Ruta KEGG	Descripción KEGG	Valor de p
hsa05412	ARVC	1.01E-07
hsa05410	HCM	1.94E-05
hsa05414	Cardiomiopatía dilatada	4.22E-05
hsa04020	Ruta de señalización del calcio	9.24E-05

### **Enriquecimiento de rutas funcionales a partir de los genes más asociados a muestras comórbidas**

En total se obtuvieron 64 rutas funcionales enriquecidas con un valor de  $p < 9.9E-05$  en contraste con la base de datos GO y 7 rutas con KEGG (ver detalles de las rutas y sus valores de significancia en las tablas suplementarias 38 y 39, columna G). Las agrupaciones creadas a partir de las rutas de GO y las rutas de KEGG se encuentran detalladas en las Tabla 43 y Tabla 44.

Tabla 43. Agrupaciones realizadas a partir de las rutas funcionales con  $p < 9.99E-05$  (tras la corrección de Benjamini&Hochbaum por tests múltiples) según GO para el estudio de comórbidos. Se muestran el número de rutas agrupadas en cada caso y el rango de valores de p de las significancias de enriquecimiento de las rutas que componen cada grupo.

Grupo	Nº Rutas	Valor de p menor	Valor de p mayor
Proteínas G acopladas	4	1.33E-07	6.83E-05
Función neuronal	15	1.33E-07	7.64E-05
Canales y transporte	16	2.00E-07	6.70E-05
Cognición	3	1.02E-06	1.13E-05
Sistema inmunológico	6	5.55E-06	6.55E-05
Adhesión/estructura	4	6.49E-06	6.83E-05
Proliferación celular	3	1.28E-05	3.02E-05
Cerebro (telencéfalo)	1	1.38E-05	1.38E-05
Reconocimiento celular	1	2.00E-05	2.00E-05
Fosfatas/kinasa (señalización)	3	3.02E-05	8.24E-05
Respuesta a los alcaloides	1	3.06E-05	3.06E-05
Respuesta a los compuestos que contienen purina	1	3.76E-05	3.76E-05
Ciclo del pelo	5	4.69E-05	6.83E-05
Colágeno	1	5.43E-05	5.43E-05

Tabla 44. Rutas funcionales con  $p < 9.99E-05$  (tras la corrección de Benjamini&Hochbaum por tests múltiples) según KEGG para el estudio de comórbidos.

Ruta KEGG	Descripción KEGG	Valor de p
hsa04514	Moléculas de adhesión (CAMs)	2.54E-08
hsa04940	Diabetes Mellitus Tipo I	3.81E-08
hsa05330	Rechazo a aloinjertos	1.09E-07
hsa05332	Rechazo a trasplantes	5.51E-06
hsa04020	Ruta de señalización del calcio	6.18E-06
hsa04510	Adhesión focal	1.95E-05
hsa04360	Orientación axonal	1.95E-05

Al igual que en los estudios anteriores, las funcionalidades más asociadas con las variantes en comórbidos fueron las pertenecientes a canales y transporte celular, con especial presencia de los canales relacionados con calcio y aquellas rutas con funciones neuronales. Las rutas del sistema inmunológico, diabetes o las relacionadas con aloinjertos también obtuvieron valores significativos en este análisis.

### **Comparativa de las rutas funcionales entre estudios**

En cuanto a las rutas funcionales predominantes en cada uno de los estudios, destaca la concordancia de las rutas relacionadas con los canales de iones y en especial la de señalización de calcio, ya que se identifican en los tres tipos de trastornos. También se asocian significativamente las rutas relacionadas con el sistema nervioso y rutas relacionadas con la adhesión y estructura. Si bien existen similitudes entre los tres estudios, también se observan rutas específicas para cada uno de ellos. En las Tabla 45 y Tabla 46 se señalan el número de rutas con valor de  $p < 9.9E-05$  encontradas para cada agrupación (Grupo) en cada una de las comparativas realizadas.

Tabla 45. Número de rutas funcionales con  $p < 9.9E-05$  (tras la corrección de Benjamini&Hochbaum por tests múltiples) según la base de datos GO, agrupadas según la similitud de funciones (Grupo) para cada uno de los estudios realizado. Los números con un asterisco representan las agrupaciones de rutas particulares para la comparativa correspondiente.

<b>Grupo</b>	<b>Dislexia</b>	<b>FA</b>	<b>Comórbidos</b>
Canales y transporte	3	23	16
Función neuronal	4	17	15
Adhesión/estructura	4	1	4
Proteínas G acopladas	1	3	4
Señalización	1	2	3
Cerebro (telencéfalo)	1	1	1
Cognición	0	3	3
Riñón	0	2	1
Proceso catabólico de amino ácidos de la familia del aspartato	<b>1*</b>	0	0
Músculo esquelético	0	<b>3*</b>	0
Respuesta de alarma	0	<b>1*</b>	0
Conducta locomotora	0	<b>1*</b>	0
Regulación del transporte de aminas	0	<b>1*</b>	0
Reconocimiento celular	0	<b>1*</b>	0
Parte de la estructura encapsuladora externa	0	<b>1*</b>	0
Envoltura celular	0	<b>1*</b>	0
Sistema inmunológico	0	0	<b>6*</b>
Ciclo del pelo	0	0	<b>5*</b>
Proliferación celular	0	0	<b>3*</b>
Reconocimiento celular	0	0	<b>1*</b>
Respuesta a los alcaloides	0	0	<b>1*</b>
Respuesta a los compuestos que contienen purina	0	0	<b>1*</b>

Tabla 46. Número de rutas funcionales con  $p < 9.9E-05$  (tras la corrección de Benjamini&Hochbaum por tests múltiples) según la base de datos KEGG, agrupadas según la similitud de funciones (Grupo) para cada uno de los estudios realizado. Los números con un asterisco representan las agrupaciones de rutas particulares para la comparativa correspondiente.

Grupo	Dislexia	FA	Comórbidos
Ruta de señalización de calcio	1	1	1
Adhesión	3	0	2
Orientación axonal	1	0	1
Corazón	3	3	0
Amoebiasis	1*	0	0
Músculo liso	1*	0	0
Secreción salival	1*	0	0
Transportadores ABC	1*	0	0
Injertos	0	0	2*
Diabetes Mellitus Tipo I	0	0	1*

Entre las rutas identificadas mediante GO, se encontraron varias particulares para cada uno de los trastornos, de esta forma, las rutas relacionadas con el músculo esquelético, el riñón, la respuesta de alarma y la conducta locomotora podrían ser más específicas de la falta de atención, mientras que las relacionadas con el sistema inmunológico, el ciclo del cabello y la proliferación celular estarían más vinculadas a la situación comórbida. Centrándonos en las definiciones KEGG en cambio, encontramos rutas como la amoebiasis, el músculo liso, la secreción salivar y los transportadores ABC relacionadas únicamente a la dislexia, mientras que la diabetes y las rutas relacionadas con injertos (estrechamente relacionadas con el sistema inmunológico) se vincularían a la situación comórbida.

#### 4.5.2 Enriquecimiento de rutas funcionales derivado de los genes más asociados con los rasgos cognitivos a partir de los datos de genotipado de genoma completo

Esta sección de resultados se divide de la misma manera en la que se ha dividido la sección 4.2, en dos apartados: por una parte los resultados del análisis de enriquecimiento de rutas en relación con las **variables de lectura** (descritos en detalle únicamente para la población de dislexia) y por otra parte los de las **variables atencionales** (únicamente en la población con FA).

#### 4.5.2.1 Rutas funcionales asociadas con las variables de lectura

Mediante los resultados obtenidos del análisis asintótico de asociación (Lineal) con las variables cuantitativas, se determinó qué genes y, en consecuencia, qué rutas funcionales aparecían enriquecidas según la significancia de los SNPs analizados. En la Tabla 47 se describen las rutas que se muestran enriquecidas por mutaciones de la **población disléxica** asociadas con datos cuantitativos de tests de lectura (variables 1 a 5) de acuerdo con la base de datos GO. En dicha tabla tan sólo se muestran los resultados con valor de  $p < 9.9E-07$ . Los valores de  $p$  obtenidos para cada una de las variables en cada ruta funcional descrita se encuentran en la tabla suplementaria 38, columnas H, I, J, K y L.

Tabla 47. Rutas funcionales más significativas ( $p < 9.9E-07$ , tras la corrección de Benjamini&Hochbaum por tests múltiples) según la base de datos GO de entre las identificadas tras el análisis QTL con las 5 variables de lectura. Var.sig.=Variable de lectura a la que se encuentra asociada la ruta funcional.

Ruta metabólica	Var.sig.
Espina dendrítica	1-2-3-4-5
Espina neuronal	1-2-3-4-5
Parte de la matriz extracelular	1-3-4-5
Actividad del transportador transmembrana del ión de calcio	2-3-5
Actividad de los canales catiónicos dependientes de voltaje	1-2-4
Actividad del transportador transmembrana de cationes inorgánicos divalentes	3-5
Membrana sináptica	3-5
Actividad del factor intercambiador del nucleótido guanina	3-5
Cabeza de la espina dendrítica	2-3
Densidad postsináptica	2-3
Actividad de los canales dependientes de voltaje	2-4
Actividad de los canales iónicos dependientes de voltaje	2-4
Regulación de la actividad del transportador	2-4
Complejo de canales iónicos	3
Unión de la calmodulina	4
Regulación del potencial de membrana	4
Regulación de la actividad fosfolipasa C	4
Regulación de la actividad del transportador transmembrana	4
Desarrollo del telencéfalo	5
Desarrollo del cortex cerebral	5

**Var 1**= “Efecto de transposición en aciertos”: Mide el procesamiento ortográfico.

**Var 2**= “Tiempo de nombrado rápido compuesto de dibujos y colores”. Mide la fluidez de lectura y la velocidad de procesamiento.

**Var 3**= “Tiempo global en discriminación silábica”: Mide la percepción del habla.

**Var 4**=“Tiempo global en identificación de fonema”: Mide la conciencia fonológica.

**Var 5**=“Eficiencia de palabras y pseudopalabras”: Mide la decodificación grafema-fonema.

Dos de las rutas funcionales identificadas, las correspondientes a la espina dendrítica y neuronal, se encontraron enriquecidas en los análisis con las 5 variables de lectura, obteniendo un valor muy significativo ( $p < 2 \text{ E-}11$ ) para las variables 2 y 3. Curiosamente entre los resultados para estas dos mismas variables también se encontró enriquecida la ruta correspondiente a la cabeza de la espina dendrítica. Entre las restantes rutas destacan las relacionadas con canales dependientes de voltaje y con los transportadores transmembrana. Un dato interesante es la relación única de la variable 5 con el desarrollo del cerebro (córtex y telencéfalo).

Al bajar el umbral de significancia hasta  $< 9.99\text{E-}05$ , el número de rutas enriquecidas aumentó hasta 109 (24 de ellas relacionadas con la variable 1, 36 con la variable 2, 39 con la variable 3, 46 con la variable 4 y 51 con la variable 5). Considerando este nuevo umbral las rutas que aparecieron significativamente enriquecidas para todas las variables de lectura, a parte de las correspondientes a la espina dendrítica, fueron las relacionadas con los canales dependientes de voltaje. En cuanto a las rutas asociadas específicamente a una sola variable, se encontró el desarrollo cerebral (córtex, telencéfalo, cerebellum, migración de células cerebrales, capas de materia blanca y gris, robencéfalo) y rutas específicas del cono de crecimiento neural y filopodios, únicamente relacionados con la variable 5. Las rutas encargadas del sistema inmunitario sólo se encontraron asociadas a la variable 4, la regulación de péptidos (secreción y transporte) así como el desarrollo del hígado y una ruta específica de colágeno se relacionaron con la variable 3; la regulación de la movilidad de las células (migración, locomoción, extensión de axón y cono de crecimiento) mostraron asociación con la variable 2 y, finalmente, el transporte de neurotransmisores y la transmisión sináptica GABAérgica obtuvieron valores significativos en la variable 1 (tabla suplementaria 38, columnas H, I, J, K y L).

Al utilizar la definición de las rutas KEGG en cambio, los resultados fueron diferentes. En la Tabla 48 se describen las rutas que obtuvieron un valor de  $p < 9.9\text{E-}07$  para alguna de las 5 variables de lectura analizadas con los genotipos de la **población de dislexia y sus controles**. Al bajar el umbral de significancia hasta  $9.9\text{E-}05$  también se asociaron las rutas de potenciación a largo plazo (hsa04720), la de señalización de GnRH (hsa04912), la de interacción entre receptor y matriz extracelular (hsa04512) y la correspondiente a la enfermedad tiroidea autoinmune (hsa05320). Los valores de  $p$  obtenidos para cada una de estas rutas en cada una de las variables se pueden ver en la tabla suplementaria 39, columnas H, I, J, K y L.

Tabla 48. Rutas funcionales más significativas ( $p < 9.9E-07$ , tras la corrección de Benjamini&Hochbaum por tests múltiples) entre las 5 variables de lectura analizadas según KEGG. Var.sig.=variable de lectura a la cual se encuentra asociada la ruta funcional.

Ruta KEGG	Descripción ruta	Var.sig.
hsa04020	Ruta de señalización del calcio	3-4-5
hsa05412	ARCV (corazón)	2-4
hsa04514	Moléculas de adhesión (CAMs)	1-4
hsa04360	Orientación axonal	2
hsa05414	Cardiomiopatía dilatada	4
hsa04970	Secreción salival	4

**Var 1=** “Efecto de transposición en aciertos”: Mide el procesamiento ortográfico.

**Var 2=** “Tiempo de nombrado rápido compuesto de dibujos y colores”: Mide la fluidez de lectura y la velocidad de procesamiento.

**Var 3=** “Tiempo global en discriminación silábica”: Mide la percepción del habla.

**Var 4=** “Tiempo global en identificación de fonema”: Mide la conciencia fonológica.

**Var 5=** “Eficiencia de palabras y pseudopalabras”: Mide la decodificación grafema-fonema.

Las rutas enriquecidas mediante la definición de KEGG fueron similares a las encontradas en el estudio de caso control de dislexia. Destaca el valor de significancia de  $3.41E-11$  de la ruta de moléculas de adhesión con la variable 1.

A parte de analizar el enriquecimiento de rutas relacionadas con las variables de lectura para la población de dislexia, también se analizaron los datos con la **población FA y sus controles** así como con la **población comórbida y los controles compartidos** (tablas suplementarias 38 y 39, columnas M-V). Ante la imposibilidad de describir con detalle todos los resultados encontrados, se agruparon todas estas rutas según su función para observar qué variables se relacionaban con estos nuevos conjuntos generados, tanto para las variables de lectura como para las variables atencionales, estas últimas descritas en el siguiente apartado. Al final del mismo se encuentra una comparativa de los grupos funcionales encontrados para cada una de las comparativas.

#### 4.5.2.2 Rutas funcionales asociadas con las variables atencionales

Los resultados de asociación de genotipos frente a rasgos cuantitativos atencionales realizados en la **población de FA**, se utilizaron para identificar las rutas funcionales más enriquecidas en dicha correlación. En la Tabla 49 se encuentran las rutas (según GO) que obtuvieron un valor de  $p < 9.9E-07$  para cada una de las 5 variables atencionales estudiadas. Los valores de p obtenidos se encuentran en la tabla suplementaria 38, columnas W-AA).

Tabla 49. Rutas funcionales más significativas ( $p < 9.9E-07$ , tras la corrección de Benjamini&Hochbaum por tests múltiples) entre las 5 variables atencionales analizadas según la descripción de GO. Var.Sig.=variable atencional con la cual se encuentra asociada la ruta funcional.

Ruta	Var.Sig
Espina dendrítica	6-7-9-10
Espina neuronal	6-7-9-10
Membrana sináptica	6-7-9
Membrana postsináptica	6-7
Conducta locomotora	6-7
Regulación del desarrollo de la neurona de proyección	6
Actividad del transportador transmembrana del ion calcio	6
Complejo proteico MHC de clase II	6
Unión de moléculas de adhesión celular	6
Balsas lipídicas	6
Procesamiento antigénico y presentación de antígenos de péptido o polisacárido vía MHC clase II	6
Actividad del transportador transmembrana de cationes divalentes inorgánicos	6
Actividad del receptor transmembrana de proteína fosfatasa	7
Actividad del receptor transmembrana de proteína tirosina fosfatasa	7
Regulación del potencial postsináptico excitatorio de la membrana	9
Parte de la matriz extracelular	9
Organización de la matriz extracelular	9
Organización de la estructura extracelular	9
Actividad de los canales dependientes de ligando	9
Actividad de los canales iónicos dependientes de ligando	9
Complejo de canales iónicos	10
Organización de la sinapsis	10
Complejo de canales catiónicos	10
Célula de adhesión homofílica	10

**Var 6=** “Efecto en tiempos de condiciones incongruente menos congruente (Stroop numérico)”

**Var 7=** “Efecto en tiempos de condiciones incongruente menos congruente (Stroop numérico)”

**Var 8=** “Efecto en tiempos de condiciones incongruente menos congruente (Stroop verbal)”

**Var 9=** “Efecto en tiempos de condiciones incongruente menos congruente (Stroop verbal)”

Las variables 6, 7, 8 y 9 miden el procesamiento automático, la atención selectiva y el efecto conflicto.

**Var 10=** “Índice conflicto”. Mide la alerta, orientación y efecto conflicto.

Entre las rutas más significativas no hubo ninguna asociada a la variable 8. Las correspondientes a la espina dendrítica y neuronal volvieron a estar presentes en la asociación de las variables 6, 7, 9 y 10. Las variables 6 y 7 miden un mismo proceso cognitivo por lo que no asombra que compartan rutas funcionales significativas. Aun así cada una de estas variables obtuvo rutas específicas significativas, destacando entre éstas las relacionadas con el sistema

mayor de histocompatibilidad y la variable 6. Por otra parte, también merece la pena nombrar la ruta relacionada con las balsas lipídicas en el análisis de esta misma variable.

Tal y como se hizo con las variables de lectura, al bajar el umbral de significancia a  $9.9E-05$ , se obtuvieron un total de 87 rutas enriquecidas (29 de las cuales están asociadas con la variable 6, 28 con la variable 7, 6 con la variable 8, 34 con la variable 9 y 45 con la variable 10). Las rutas significativas coincidentes entre las cinco variables atencionales fueron las de la espina dendrítica y neuronal. Se observó mayor coincidencia de rutas entre las variables 6, 7 y 10 que con las restantes variables (8 y 9). Las rutas relacionadas con canales iónicos aumentaron con el umbral más laxo, pero curiosamente ninguna de éstas se asoció a la variable 8, sin embargo se encontraron rutas específicamente enriquecidas para esta variable: el desarrollo de dendritas y la adhesión entre células neuronales. Por otro lado, el transporte de cationes divalentes apareció sólo asociado a la variable 9. La variable 10 obtuvo el mayor número de rutas asociadas y entre las exclusivas se encontraron varias relacionadas con el músculo esquelético, la única ruta relacionada con axones y migración neuronal, así como las encargadas de la transmisión de impulsos nerviosos. Dentro de las variables 6 y 7 no se observaron cambios de gran importancia al utilizar el umbral de significancia de  $p < 9.9E-05$  en comparación con el umbral de  $9.9E-07$ .

Sólo 3 rutas resultaron con significancia menor a  $9.9E-07$  según KEGG: la ruta de señalización del calcio, la de moléculas de adhesión celular y la de orientación axonal, y las 3 resultaron asociarse con la variable 6. Al considerar también las rutas con valores de  $p$  menores a  $9.9E-05$ , fueron 9 las rutas enriquecidas (Tabla 50). Los valores de  $p$  obtenidos para las rutas descritas en cada una de las variables se pueden observar en la tabla suplementaria 39, columnas W-AA.

Tabla 50. Rutas funcionales más significativas ( $p < 9.9 \times 10^{-5}$ , tras la corrección de Benjamini&Hochbaum por tests múltiples) entre las 5 variables atencionales analizadas según la descripción KEGG. Var.Sig.=variable de lectura con la cual se encuentra asociada la ruta funcional.

Ruta KEGG	Descripción ruta	Var.Sig
hsa04020	Ruta de señalización del calcio	6
hsa04514	Moléculas de adhesión (CAMs)	6-7-8
hsa04360	Orientación axonal	6
hsa04540	Uniones gap	6-8
hsa04971	Secreción de ácidos gástricos	6
hsa04971	Secrecion salival	6
hsa05310	Asma	6
hsa05310	Diabetes mellitus tipo I	6
hsa02010	Transportadores ABC	9

**Var 6=** “Efecto en aciertos de condiciones incongruente menos congruente (Stroop numérico)”

**Var 7=** “Efecto en tiempos de condiciones incongruente menos congruente (Stroop numérico)”

**Var 8=** “Efecto en aciertos de condiciones incongruente menos congruente (Stroop verbal)”

**Var 9=** “Efecto en tiempos de condiciones incongruente menos congruente (Stroop verbal)”

Las variables 6, 7, 8 y 9 miden el procesamiento automático, la atención selectiva y el efecto conflicto.

**Var 10=** “Índice conflicto”. Mide la alerta, orientación y efecto conflicto.

#### 4.5.2.3 Comparativa de las rutas asociadas entre las variables

Para poder comparar las rutas funcionales más asociadas ( $p < 9.9 \times 10^{-5}$ , tras la corrección de Benjamini&Hochbaum por tests múltiples) a las variables de lectura y a las variables atencionales, así como las relacionadas con dislexia, falta de atención y al estado comórbido, se agruparon el total de rutas enriquecidas (258) en alguno de los análisis previos en grupos definidos por la similitud de su función. Una vez generados dichos grupos, se contó el número de rutas enriquecidas en cada uno de los análisis. Se han utilizado las 5 variables de lectura en conjunto y las 5 variables atencionales en conjunto para contabilizar las rutas enriquecidas. En la Tabla 51 se muestran los números de rutas resultantes de este análisis. Las rutas que componen cada uno de los grupos, así como el valor de  $p$  obtenido por cada variable en cada una de las rutas ( $p < 9.9 \times 10^{-5}$ ) se describen en la tabla suplementaria 38.

Tabla 51. Número de rutas funcionales enriquecidas para cada set de muestras analizadas en base a la estrategia de QTLs o de casos/controles. E\_D=Población conjunta (casos+controles) del estudio GWAS de dislexia, E\_FA=Población conjunta (casos+controles) del estudio GWAS de falta de atención, E\_C=Población conjunta (casos+controles) del estudio GWAS de comórbidos, Var.atención=variables atencionales, D=dislexia, FA=falta de atención, C=comórbidos.

GRUPO	ESTUDIO DE QTLs				Estudio de Casos vs. Controles		
	Variables de lectura			Var.atención	D	FA	C
	E_D	E_FA	E_C	E_FA			
<b>Total</b>	24	26	28	21	3	23	16
<b>Actividad canales</b>							
Transporte	13	11	10	7	2	7	8
Calcio	6	5	7	3	2	5	4
Potasio	2	2	3	5	0	0	0
Glutamato	0	6	3	1	0	3	0
Receptores	1	5	5	1	0	4	1
<b>Neuronas/sinapsis</b>	16	23	21	17	4	17	15
<b>Adhesión-estructura</b>	9	11	10	10	4	1	4
<b>Señalización</b>							
Fosfatasa/kinasa	2	3	4	2	0	2	3
Otros	3	5	1	6	1	0	0
<b>Cyclic nucleotide</b>	5	15	4	1	0	0	0
GTPasa	3	8	6	2	1	3	4
<b>Migración/locomoción</b>	11	2	2	2	0	1	0
Cerebro	5	1	1	0	1	1	1
Cognición	0	1	3	0	0	3	3
Estímulos	2	1	2	0	0	0	0
Riñón	8	1	2	1	0	2	1
<b>Músculo esquelético</b>	1	4	1	7	0	3	0
<b>S.Inmunológico</b>	4	1	1	3	0	0	6
<b>Proliferación</b>	2	3	0	1	0	0	3
<b>Lípidos</b>	10	0	0	3	0	0	0
<b>Péptido</b>	3	0	1	6	0	0	0
<b>Pelo/muda</b>	0	0	0	0	0	0	5
<b>Respuesta Alarma</b>	0	1	0	0	0	1	0

Los grupos predominantes que aparecen en todos los estudios son los relacionados con canales y neuronas o sinapsis. Las funciones de los canales de calcio y potasio aparecen tanto en las variables de lectura como en las variables atencionales, pero éstos últimos no se observaron en los estudios de caso/control individuales. En cuanto al glutamato, pareció relacionarse principalmente al utilizar muestras con falta de atención. Las rutas encargadas de la señalización molecular, así como las GTPasas y nucleóticos cíclicos, los cuales están estrechamente

---

relacionadas con los canales y la actividad sináptica, también se observaron en la mayoría de las comparativas.

La adhesión y estructura celular son asimismo otras de las principales funciones que se asocian tanto a las variables de lectura como a las atencionales, así como a los tres tipos de estudios realizados.

Como dato de interés, se encontró que dentro de las rutas que afectaban al cerebro únicamente la ruta del desarrollo del telencéfalo coincidía entre estudios, el resto de rutas cerebrales fueron específicamente relacionadas a las variables de lectura al analizarse en las muestras del estudio de dislexia, lo mismo sucedió con la mayoría de las rutas relacionadas con lípidos. Si bien rutas relacionadas con la migración se asociaron a varios de los grupos analizados, su significancia fue predominante en las variables de lectura con el grupo de muestras con dislexia y lo mismo sucedió con las funciones renales.

Por otra parte, las rutas del músculo esquelético fueron predominantes con las variables atencionales. Las rutas relacionadas con el cabello o la muda sólo aparecieron enriquecidas en el estudio de caso/control de comórbidos. Destaca la ruta de respuesta de alarma en el estudio de caso/control de falta de atención, así como en el de QTLs con las variables de lectura al utilizar estas mismas muestras de FA, lo cual no se observa en cambio con las variables atencionales.

Al agrupar las rutas metabólicas asociadas definidas según la base de datos KEGG, se observan algunas similitudes con respecto a las rutas definidas mediante GO. Las rutas relacionadas con la adhesión y estructura celular, las del sistema nervioso, señalización y músculo también aparecen con frecuencia en los distintos análisis realizados. Sin embargo a diferencia de lo anteriormente descrito, aparecen rutas cardiovasculares, rutas relacionadas con diferentes enfermedades así como asociadas al sistema digestivo.



## **5 DISCUSIÓN**



En la presente tesis se ha tratado de identificar variantes genéticas que pudieran estar relacionadas con la lectura y con la atención infantil, haciendo especial hincapié en la dislexia, la falta de atención y en la comorbidad entre ambos trastornos. Para ello se han llevado a cabo estudios de asociación en base a dos grandes estrategias: **genes candidatos** y **genoma completo**. Existen varios genes vinculados a estos trastornos en diversos estudios y, aun observando que los resultados obtenidos no se replican entre las diferentes poblaciones, se ha querido comprobar si en la población española se asociaban. Además, con el propósito de identificar variantes nuevas se ha realizado un estudio de asociación de genoma completo. Asimismo, dada la heterogeneidad sintomatológica dentro de los trastornos estudiados, hemos visto la oportunidad de analizar la variabilidad de ciertos rasgos cognitivos en relación al genotipo. De esta forma, ambas estrategias a su vez, se han subdividido en dos metodologías de análisis: **casos/controles** y análisis de **QTLs** (*Quantitative Trait Loci*). Finalmente partiendo de los resultados obtenidos en el estudio de genoma completo también se ha querido hipotetizar sobre posibles rutas funcionales responsables de cada uno de los trastornos así como para las funciones cognitivas analizadas. Todos estos temas serán discutidos a continuación comenzando con algunas particularidades que se han observado durante el procesamiento de las muestras.

### **5.1 LA SALIVA COMO FUENTE DE DNA PARA ESTUDIOS DE GENOMA COMPLETO: LA PRECISA PUESTA A PUNTO DE LA TÉCNICA PARA ESTE TIPO DE MUESTRA**

Los estudios epidemiológicos de grandes poblaciones requieren la obtención de muestras de DNA de miles de individuos para poder investigar las influencias genéticas así como las respuestas a los posibles tratamientos de la enfermedad. Una alta participación es esencial para tener un número de muestra representativo y poder aumentar así el poder estadístico del estudio. Para incrementar el porcentaje de conformidad de estos individuos es preferible que el método de recogida sea no-invasivo, por ello muchos estudios utilizan células epiteliales bucales como fuente de DNA<sup>329</sup>. Hasta el momento la sangre era la fuente de DNA preferida ya que generalmente se obtienen grandes cantidades y altas calidades de DNA<sup>330</sup>, sin embargo, presenta una serie de limitaciones como una menor participación de voluntarios debido a la invasividad de su recogida, y restricciones de temperatura y tiempo en cuanto a su almacenamiento<sup>331</sup>. En vista a estas limitaciones cada vez son más los estudios que utilizan la saliva como fuente de DNA humano ya que tiene ventajas como no tener que acudir a un centro de salud para poder donar la

muestra<sup>330</sup> (puede enviarse vía correo) y puede almacenarse a temperatura ambiente durante años antes de proceder a su extracción<sup>331</sup>.

Un obstáculo a la hora de utilizar el DNA proveniente de la saliva, a parte de la heterogeneidad con respecto a la cantidad de células bucales de cada individuo, es la **alta variabilidad en cuanto al porcentaje de DNA de origen humano y no-humano** que se obtiene de este tipo de muestra. Estimar la cantidad de DNA genómico humano en las muestras de saliva es de vital importancia ya que podrían contener cantidades variables de DNA bacteriano o de otros orígenes (alimentos u origen fúngico)<sup>332</sup>. Varios estudios realizados al respecto indican que la proporción de DNA humano en una muestra de saliva puede variar entre 11 y el 100 %<sup>330</sup> tal y como se ha demostrado en este estudio, en el que la cantidad media de DNA humano encontrado fue del 26 %. Esto puede considerarse un problema ya que, según determina el estudio de M. Nishita *et al.*<sup>331</sup>, una proporción menor a un 30 % de DNA humano en la muestra puede conllevar un bajo rendimiento de los métodos de genotipado masivo. Hay estudios que revelan que el éxito de genotipado mediante sondas TaqMan™ es mayor al utilizar muestras de DNA provenientes de sangre, y lo asocian a la cantidad de DNA humano presente en la muestra, no a la cantidad total de DNA extraído<sup>333</sup>, aunque no todos opinan lo mismo y apoyan el uso de DNA proveniente de saliva<sup>330,331</sup>.

En el presente proyecto, el grado de genotipado de SNPs (del inglés, *Single Nucleotide Polymorphisms*) mediante arrays (en base al *call rate* de la muestra) fue exitoso en todas las muestras estudiadas independientemente de la cantidad de DNA de partida utilizada, no obstante para el análisis de los CNVs (del inglés, *Copy Number Variations*) las condiciones requeridas fueron algo más exigentes. Por ello, es preciso remarcar **la importancia de establecer exhaustivamente las condiciones óptimas de genotipado** dependiendo y teniendo muy en cuenta los análisis que se quieran realizar posteriormente. En nuestro estudio de puesta a punto no se observó ninguna relación entre el porcentaje de genotipado de una muestra y la de DNA humano que esta contenía, es más, muestras con menores cantidades de DNA humano obtuvieron mejores tasas de genotipado. Sin embargo, para el análisis de los CNVs si parecían influir las proporciones bajas de DNA humano. Así, aunque la compañía Illumina recomienda usar 200 ng de DNA de partida en los chips para el genotipado, se creyó conveniente aumentar dicha cantidad a 400 ng dada la

heterogeneidad de las muestras provenientes de saliva, ya que esto podría ser un factor que afectara a los resultados de esta estrategia.

### 5.1.1 Observaciones sobre la reproducibilidad del genotipado en las diferentes versiones de arrays *Human OmniExpress* utilizadas.

Otro factor de vital importancia en este tipo de tecnologías es que los resultados obtenidos puedan ser reproducibles. En los tres tipos de arrays utilizados en este proyecto la concordancia entre los genotipos obtenidos en cualquiera de ellos superó el 99 %, sin embargo, al analizar más en detalle algunos de sus marcadores, observamos que la codificación para un mismo SNP no era la misma entre las diferentes versiones del array *Human OmniExpress*, lo cual nos llevó a conclusiones preliminares erróneas acerca de las frecuencias observadas para cada alelo. Además la calidad de los datos para su posterior análisis de CNVs fue diferente para cada una de las dos versiones del array utilizadas, **poniendo en duda lo adecuado que pueda resultar mezclar datos de diferentes arrays**. Existen evidencias que indican que cuando los estudios de asociación tienen un número de muestras bajo, la falta de poder estadístico conlleva a resultados contradictorios. Una alternativa para aumentar este número de muestras y cerciorarse de previos resultados inconsistentes es realizar un meta-análisis<sup>334</sup>. Nuestros hallazgos de las diferencias encontradas entre las distintas versiones de arrays de genotipado masivo pueden considerarse un obstáculo a la hora de comparar resultados de diferentes estudios así como a la hora de proceder a fusionar resultados para este tipo de análisis. Por ello es de vital importancia asegurarse de que los datos comparados de diferentes investigaciones tengan exactamente la misma codificación alélica independientemente de la plataforma de genotipado utilizada.

También debe considerarse la reproducibilidad entre distintas técnicas de genotipado. Los marcadores analizados mediante la tecnología TaqMan™ obtuvieron un buen valor de concordancia en su mayoría, sin embargo las sondas para genotipar los CNVs no consiguieron replicar ninguno de los resultados de duplicaciones o deleciones observadas. Por lo tanto los resultados relativos a las diferencias en el número de copias entre los casos y los controles, si bien resultan intrigantes en cuanto al interés funcional de los genes duplicados o deleccionados, han quedado sin confirmar y en consecuencia hemos considerado no discutirlos en la disertación de este trabajo.

### 5.1.2 La importancia del estudio de la epistasia: hallazgos de asociaciones más significativas por interacción de SNPs

En los estudios de asociación de genoma completo la mayoría de las variantes asociadas a las enfermedades tienen un efecto pequeño, por lo que, aun sumando el efecto de todas ellas, explican un bajo porcentaje de la variabilidad genética heredada. Además, al analizar cada *locus* por separado se tiende a implicar genes individuales con las enfermedades sin tener en cuenta la posible interacción entre elementos funcionales de un sistema<sup>319</sup>. **Dos variantes genéticas que individualmente tienen un efecto pequeño para un trastorno podrían ejercer un efecto mucho mayor en combinación** si por ejemplo están involucradas en una misma ruta neural<sup>1</sup>, por lo que los efectos de los diferentes *loci* no se suman independientemente sino interaccionando los unos con los otros. De esta manera, cualquier efecto que un gen singular ejerza sobre el fenotipo podría estar moderado en magnitud y/o en dirección por otra variante genética<sup>335</sup>. Por esta razón, una buena forma de encontrar asociaciones “funcionales” es realizar un análisis de interacciones entre las variantes genotipadas. Analizar todos los SNPs de un genoma completo dos a dos supone tener que superar grandes limitaciones computacionales (memoria y tiempo de ejecución), por ello algunas herramientas informáticas para este tipo de análisis se centran en una selección de marcadores candidatos o analizan solamente los SNPs que están cerca los unos de los otros pudiendo perderse muchas interacciones reales<sup>319</sup>. En este proyecto se ha podido comprobar que efectivamente, **las asociaciones encontradas mediante interacciones epistáticas fueron más significativas que las individuales** incluso para los mismos marcadores. Aun así, la comparativa del uso de diferentes programas informáticos para realizar dichos análisis de interacción produjo resultados diferentes.

## 5.2 EL ESTUDIO DE GENES CANDIDATOS REVELA ASOCIACIONES DESIGUALES DEPENDIENDO DEL GRUPO DE MUESTRAS CONSIDERADO, ASÍ COMO UNA RELACIÓN CON LOS RASGOS COGNITIVOS DE LA POBLACIÓN GENERAL

### 5.2.1 El origen poblacional y el criterio diagnóstico: dos factores clave en los estudios de asociación genética

Un punto esencial que se debe tener en cuenta al tratar de replicar los resultados de los estudios genéticos de asociación es la etnia de las cohortes consideradas, ya que el fondo genético y, consecuentemente, las frecuencias de los polimorfismos analizados son

heterogéneos. Además, no hay que ignorar que de los estudios provenientes del proyecto multipoblacional IMAGE han concluido que los participantes de Israel y España muestran perfiles genéticos diferentes al del resto de los europeos del norte<sup>291</sup>. Valorando el análisis de componentes principales de los tres estudios realizados en la presente tesis doctoral (dislexia, falta de atención (FA) y comórbidos), entre las 1730 muestras genotipadas con arrays de genoma completo se observó que el fondo genético era similar para todas ellas, indicando una homogeneidad en cuanto a la etnia de la población caracterizada. También es importante considerar que los polimorfismos causales de un trastorno en particular no tienen necesariamente porqué ser los mismos en todas las poblaciones<sup>33</sup>, ya que **las frecuencias alélicas de SNPs difieren considerablemente entre las poblaciones humanas**<sup>336-338</sup>, lo cual enfatiza la importancia de la consideración de las etnias en el diseño e interpretación de cualquier estudio de asociación genético.

Otro factor tan importante como el origen poblacional en estos abordajes es el criterio diagnóstico aplicado a las muestras de interés, el cual es particularmente difícil de establecer para ciertos síndromes, como es el caso de los trastornos neurológicos. A pesar de esta dificultad y la complejidad de la búsqueda de marcadores genéticos para rasgos complejos, se han propuesto varios *loci* como biomarcadores potenciales de la dislexia<sup>18</sup> y la FA<sup>8</sup>. Sin embargo, los resultados obtenidos en diferentes estudios son frecuentemente contradictorios y su reproducibilidad es limitada. Los criterios utilizados para describir los grupos categóricos, si bien son cruciales, son heterogéneos, y ésta podría ser una de las razones de las repetidas inconsistencias encontradas por cada grupo de investigación. Este hecho se observa claramente en la asociación significativa del test recesivo del gen MAOA hallada en nuestro proyecto. A parte de apoyar la idea de su posible función en la FA, la señal se acentúa más al utilizar un grupo de controles de FA más estricto, indicando la importancia del riguroso establecimiento de los fenotipos. Lo mismo sucedió con el genotipo 04R05R y la presencia del alelo de 05R del gen DRD4 en los tests referentes a FA. A la hora de definir a los pacientes caso, es incuestionable la necesidad de una buena caracterización, es más, varios artículos indican el impacto de los diferentes grados de severidad del trastorno investigado en los resultados obtenidos. Sin embargo, poco se sabe de las consecuencias de la definición de criterios de inclusión del grupo control<sup>339</sup>. Aunque este campo no se investiga frecuentemente, Smith<sup>340</sup> ha observado que las frecuencias alélicas de la población control podrían ser las que determinan la variabilidad de las asociaciones entre casos y

controles, es más, la fuerza de la asociación podría estar mediada en parte por el método por el cual se ha reclutado al grupo control<sup>341</sup>. Por ello es de importancia crítica la apropiada definición de este grupo en los estudios de asociación, ya que la selección inapropiada de los controles podría haber influido en la inconsistencia de los resultados encontrados hasta el momento<sup>338</sup>. Generalmente en las investigaciones se consideran a las poblaciones control como un solo grupo, sin embargo Fair et al.<sup>342</sup> encuentran varios perfiles cognitivos existentes en los niños con TDAH (Trastorno por Déficit de Atención e Hiperactividad) así como en el grupo de controles, y proponen que diferenciar qué funciones cognitivas son las atípicas en cualquier niño dependerá del contexto que proporciona la población control. Por lo que concluyen que para identificar mecanismos asociados a trastornos mentales, es necesario comparar los individuos con controles bien ajustados al mismo estilo cognitivo. Según nuestros resultados, la estricta caracterización fenotípica de los controles tiene la misma importancia que la definición del mismo trastorno.

### **5.2.2 La inclusión/exclusión de las muestras comórbidas altera de forma trascendente los resultados de asociación de algunos marcadores**

En la mayoría de los proyectos que se centran en trastornos del desarrollo como la dislexia y la falta de atención, estas afecciones se clasifican como grupos clínicos separados, por lo que se consideran independientemente, aunque exista un porcentaje de ellos que comparta las dos condiciones. Sin embargo, recientemente, los estudios con condiciones comórbidas están ganando más relevancia<sup>137,244,343-346</sup>. Un dato importante en referencia a las muestras comórbidas de esta tesis doctoral, es que registraron los valores más extremos del discriminante utilizado para el diagnóstico.

En nuestro proyecto, replicamos los resultados del gen KIAA0319 en las muestras disléxicas y relacionamos el gen FOXP2 con este trastorno. Los distintos valores de significancia obtenidos para los SNPs de estos genes en cada una de las comparaciones apoya la idea de que son las muestras comórbidas las que empujan esas señales de asociación. Esta observación es concordante con el hecho de que los individuos que sufren los dos trastornos son los que presentan una manifestación más severa de los síntomas como se ha visto en investigaciones anteriores<sup>245,347</sup> así como en el presente estudio (Figura 10). Además, se sugiere que la condición

comórbida muestra un perfil más extenso y severo en cuanto a la debilidad neuropsicológica y tiene una mayor carga genética para los dos trastornos, la dislexia y la FA<sup>249</sup>. Según muestran Willcut et al.<sup>245</sup>, la heredabilidad de déficits en lectura y ortografía fueron significativamente mayores cuando los individuos aparte de poseer un trastorno de lectura también cumplían los criterios de TDAH, lo cual apoya nuestros resultados que indican que las muestras comórbidas son las que presentan un valor menor en la función del discriminante. Diversos estudios genéticos relacionados con la lectura coinciden en obtener una mayor asociación cuando las muestras se clasifican según su severidad<sup>14,65,68</sup>.

*DCDC2* y *DYX1C1* son genes potencialmente asociados al desarrollo de la dislexia, pero ninguno de ellos alcanzó un valor de asociación significativo en el presente proyecto. Esta falta de asociación no sorprende, ya que por un lado, se ha contemplado que la relación genética de variantes de *DCDC2* con el fenotipo disléxico difiere dependiendo de la clasificación de subgrupos realizada para este trastorno<sup>348</sup>. En contra a nuestras consideraciones acerca de la homogeneidad de las muestras a la hora de realizar análisis genéticos de asociación, Scerri et al. mantienen la importancia de no excluir a las muestras comórbidas al hacer las agrupaciones. Según su estudio las asociaciones de los marcadores de *DCDC2* con trastornos de lectura se refuerzan si se mantienen las muestras comórbidas de trastorno específico del lenguaje (TEL) y de TDAH<sup>349</sup>, posiblemente por la misma causa que en nuestra población, puesto que las muestras comórbidas son las que presentan un fenotipo neurológico más extremo. Por otro lado, los resultados significativos de *DYX1C1* encontrados por Taipale et al.<sup>28</sup> en una población finlandesa no han sido replicados en ningún otro proyecto<sup>38,40</sup>. En vista de que los resultados de asociación del SNP de *DYX1C1* han sido obtenidos en dirección opuesta para la población estudiada, muchos investigadores creen que habrá algún otro gen causante de dislexia en el locus *DYX1* todavía sin descubrir o que tal vez existan diferentes variables causales en *DYX1C1* entre las diferentes poblaciones. No obstante, ninguno de los marcadores genotipados en los arrays de genoma completo situados en este gen mostró asociación con la población disléxica. También podría haber patrones de desequilibrio de ligamiento (LD) singulares en este gen para cada población<sup>10,20</sup>, o puede que su efecto sólo se presente bajo circunstancias ambientales desfavorables<sup>350</sup>.

### 5.2.3 El gen *COMT* como posible candidato para la susceptibilidad hacia los problemas de lectura, así como para el estudio de las diferencias cognitivas entre géneros

En cuanto a los genes candidatos de TDAH, *DBH*, *COMT* y *MAOA*, su asociación no se replicó en nuestra población de FA mediante el test alélico, aunque encontramos algunas asociaciones en otros análisis realizados (tabla suplementaria 4). Interesantemente, el gen *COMT* apareció asociado en los análisis genotípico y recesivo en las muestras disléxicas. *COMT* se ha relacionado recientemente con tareas de habilidad hacia la lectura (conciencia fonológica y deletreo)<sup>351</sup> y específicamente, con la comprensión lectora tal y como muestran Gringorenko *et al.*<sup>352</sup>, ya que observan diferentes frecuencias de haplotipos de *COMT* en los individuos con y sin dificultades.

La diferencia en la prevalencia de la dislexia entre chicos y chicas está bien documentada en numerosas investigaciones<sup>249</sup> y la neurotransmisión dopaminérgica, la cual está implicada en una multitud de funciones cognitivas, podría estar modulada por el género<sup>353</sup>. Además, diversos estudios de *COMT* en relación al TDAH han descrito efectos específicos dependientes del género para este gen<sup>354,355</sup>. La tendencia hallada en nuestros resultados hace pensar que el gen *COMT* podría estar influyendo de forma diferente según el género en nuestro grupo disléxico (tablas suplementarias 5 y 6). También se observaron efectos más acentuados de asociación en los niños con el genotipo 04R05R y con la presencia del alelo 05R del gen *DRD4* en comparación con las niñas. Según Knopik *et al.* (1998)<sup>356</sup> la correlación genética entre los dos géneros en base a la capacidad de lectura (*reading performance*) es inferior a 1, lo que sugiere bien que existen alelos de riesgo específicos de género o bien, al menos, efectos genéticos específicos para cada género<sup>19</sup>. Esta razón podría contribuir al incremento de ratio 2:1 de la dislexia observada en los chicos. Otras razones podrían ser la diferencia en la plasticidad del desarrollo entre los dos géneros o que la vulnerabilidad a un daño cerebral sea mayor en los chicos, o que exista una diferencia de respuesta ante este daño cerebral entre los dos sexos<sup>21</sup>. Diversos autores muestran también diferencias de género en el TDAH, sin embargo se sabe poco acerca de las causas de estas diferencias. Investigaciones neurobiológicas de años anteriores sostienen que **variantes en el gen *COMT* podrían tener una mayor influencia en los chicos mientras que las de *MAOA* influirían más en las chicas**, y así lo demuestran Biederman *et al.*<sup>357</sup>. La actividad de *COMT* aparentemente también difiere entre géneros, siendo esta actividad mayor en el hígado, eritrocitos y córtex

prefrontal masculino<sup>352</sup>. Por tanto, *COMT* se convierte en un buen candidato para estudiar las posibles diferencias en los procesos cognitivos entre los chicos y las chicas.

#### 5.2.4 Las asociaciones de los alelos de los genes *DRD4* y *DAT1* de riesgo a padecer TDAH reportadas en estudios anteriores no se replican en nuestro proyecto

Si bien el alelo asociado con FA del VNTR de *DRD4* en nuestras muestras es el 05R, el cual anteriormente ya se había relacionado con este trastorno<sup>162</sup>, el alelo de riesgo y el que mayoritariamente aparece asociado a TDAH es el de 07R. Sin embargo, este alelo no se relacionó en nuestra población de FA, aunque en base a los meta-análisis realizados en torno a este alelo, su magnitud de asociación difiere substancialmente entre cohortes<sup>340,358</sup>. Los subtipos de TDAH y la metodología del estudio son las heterogeneidades más estudiadas. Según Sagvolden *et al.*<sup>146</sup>, los síntomas de hiperactividad/impulsividad se relacionan con la hipofunción de la ruta dopaminérgica mesolímbica mientras que los procesos atencionales lo hacen con la ruta mesocortical. El alelo de 07R se ha implicado en ambas rutas, por lo que se esperaría una mayor asociación de éste al tipo combinado de TDAH. Así una proporción elevada de pacientes inatentos entre los casos se relacionaría con una menor asociación de 07R, según lo demuestra Smith<sup>340</sup>. De esta misma forma, el alelo 10R de *DAT1* también se asocia más fuertemente con el tipo combinado de TDAH que con el de inatención<sup>359</sup>. Además, está demostrado que la secuencia repetitiva del VNTR de 48pb de *DRD4* muestra variación considerable en su composición de bases<sup>8</sup>, siendo más probable que esté constituida por una secuencia anormal cuanto más larga sea su longitud<sup>159</sup>. Si estas diferencias en la secuencia afectan a la conducta, es posible que especificar simplemente la longitud del VNTR no sea suficiente para capturar las características fundamentales del gen *DRD4* que pudieran estar asociadas con la manifestación de los síntomas de TDAH<sup>159</sup>. Existen evidencias que indican que haplotipos formados por el alelo de 10R de *DAT1* están asociados a TDAH, como describen Barr *et al.*<sup>360</sup>, quienes encuentran un sesgo en la transmisión del haplotipo formado por el VNTR y dos variantes del intrón 9 de este gen, Galili-Weisstub *et al.*<sup>361</sup> al hallar mayor asociación de un haplotipo formado por el alelo de 10R y un SNP del exón 15, o los resultados de Hawi *et al.*<sup>362</sup>, quienes también sugirieron una asociación haplotípica con el alelo de 10R y el exón 9. Por ello es posible que el alelo 10R esté en LD con una variante de riesgo u otros alelos de *DAT1*, o que la susceptibilidad para el TDAH dependa de la interacción entre este alelo y algún otro polimorfismo funcional<sup>162,334</sup>. Entre los SNPs pertenecientes al gen *DAT1* genotipados en los arrays de genoma

completo, pudimos observar que existe un marcador intrónico (rs46000) que alcanza el valor de significancia ( $p < 0.05$ ) en la población de FA y otros SNPs que se asocian a la población disléxica y comórbida (ver tabla suplementaria 7).

Aun siendo **DRD4** y **DAT1** dos de los principales genes involucrados en la susceptibilidad del TDAH, nuestros análisis mostraron asociación de ambos con la población disléxica. Sin embargo, un estudio anterior en el cual también analizan la posible asociación del alelo de 07R de DRD4 con dislexia no mostró evidencias de significancia, aunque no descartan que otras variantes de DRD4 pudieran estar afectando a la susceptibilidad de este trastorno<sup>363</sup>. El genotipo 10R10R de DAT1 en cambio se ha visto relacionado con la lectura, sugiriendo que este polimorfismo podría influir en un mecanismo neural común tanto para la adquisición de la lectura como para los síntomas de TDAH<sup>364</sup>.

### 5.2.5 SNPs ubicados en los genes candidatos revelan asociaciones de interés

La falta de asociación de algunos de los marcadores no implica necesariamente que no estén relacionados con el trastorno, ya que a veces al analizar el efecto de interacción entre dos SNPs se encuentran nuevas relaciones. Este hecho fue observado entre los SNPs candidatos localizados en los genes *KIAA0319* y *DCDC2*. La distancia genómica entre *DCDC2* y *KIAA0319* es de sólo 130 kb, sin embargo, este intervalo es relativamente rico en recombinaciones. La ausencia de LD entre ellos en la población estudiada apunta hacia una asociación independiente pero cooperativa de estos dos genes para la dislexia. Es muy probable que cualquier asociación detectada en alguno de estos dos genes sea específica del gen implicado y no el reflejo del efecto del otro gen<sup>14</sup>. Además, nuestros resultados son consistentes con los obtenidos por Powers et al.<sup>72</sup>, quienes muestran que el haplotipo de riesgo del gen *DCDC2* interacciona de forma sinérgica con otro ubicado en el gen *KIAA0319*, mostrando unos resultados más extremos al analizarse estos haplotipos en conjunto que cada haplotipo por separado. La interacción entre estos dos genes ha sido descrita anteriormente por otros grupos<sup>70,71,365,366</sup>. *DCDC2* también parece interaccionar con *DYX1C1* en nuestras muestras con FA cuando los dos genes se analizan epistáticamente. Marcadores que previamente se habían asociado a RD (del inglés, *Reading Disability*) se han asociado también a TDAH. Tal y como muestran en Couto et al.<sup>241</sup>, *DCDC2* se asoció a las dos formas de TDAH, a la hiperactividad y a la falta de atención. *DYX1C1* también se ha

analizado en muestras con TDAH, donde varios SNPs situados en este gen así como un haplotipo formado por 6 marcadores se han asociado con los síntomas de falta de atención del trastorno<sup>366,367</sup>.

También hay que considerar que tal vez los marcadores elegidos para este estudio no hayan sido los idóneos para capturar la asociación, ya que otros SNPs de estos mismos genes han obtenido valores con mayor significancia en nuestros GWAS. Entre los resultados más destacados observamos que la mayoría de los SNPs pertenecientes al gen *DCDC2* que llegaron al umbral de significancia lo hicieron al considerar la población de falta de atención, lo que sugiere que podría estar involucrado en este trastorno (tabla suplementaria 7). Tal y como ocurría con los marcadores candidatos seleccionados, también se observó una interacción entre marcadores de los genes *DCDC2* y *KIAA0319*, pero esta vez en la población con falta de atención. Hasta el momento no se han encontrado evidencias de que *KIAA0319* esté relacionado con los procesos atencionales. En concordancia con lo descrito en la literatura, marcadores de los genes *KIAA0319*, *FOXP2* y de *DCDC2* en interacción con *FOXP2* también se asociaron al grupo de muestras disléxicas como era de esperar o al grupo comórbido.

En ocasiones el análisis basado en haplotipos ha revelado evidencias de asociación mayores a las encontradas con cualquier marcador en solitario. Bien sea porque el efecto esté atribuido directamente al haplotipo o bien porque el haplotipo describa más eficientemente la causa de riesgo que cualquier SNP por sí solo<sup>368</sup>. Además, los métodos basados en haplotipos también podrían ser capaces de capturar el efecto de SNPs raros<sup>369</sup>, por lo que con esta estrategia se podría mejorar la identificación de nuevos genes causales. Curiosamente un haplotipo en el gen *COMT* que contenía un marcador en alto LD con nuestro SNP candidato, el cual no obtuvo significancia en ninguno de los tests realizados, se asoció a la población comórbida. Haplotipos en el gen *DCDC2* volvieron a mostrar relación en la población con FA, ofreciendo soporte adicional sobre su posible influencia en el trastorno. Cabe destacar que por lo general cada marcador y cada haplotipo se asociaron a una situación neurológica, indicándonos que es probable que cada uno de los trastornos y la situación comórbida entre ambos dispongan de marcadores específicos que influyan solamente en esa situación.

### 5.2.6 La importancia del diagnóstico certero y homogéneo de las muestras consideradas en análisis genéticos: la utilidad de endofenotipos como alternativa

La falta de uniformidad observada entre los diferentes estudios puede deberse a la heterogeneidad de los criterios de selección de muestras. Investigaciones realizadas con diferentes subgrupos de una misma muestra<sup>65,66,68</sup> apoyan la idea de que **los análisis son sensibles a pequeñas diferencias en la composición de ésta**, especialmente cuando la muestra es de tamaño pequeño<sup>10</sup>. Tal y como hemos observado en nuestros resultados, los comórbidos, aun siendo un subgrupo muy pequeño, pueden alterar el resultado de la asociación dependiendo de su inclusión o exclusión, y asimismo también lo hacen los controles escogidos. La determinación del fenotipo disléxico para los estudios genéticos no está consensuado, por lo que cada investigador utiliza criterios de clasificación arbitrarios, algo que muchas veces es inevitable debido a las diferencias inherentes de los idiomas. La selección de sujetos en base a la codificación fonológica o en base a su habilidad para el deletreo podrían no ser comparables<sup>7</sup>, lo que supone un problema para entender o mezclar datos de diferentes proyectos. Cabe preguntarse si la dislexia se determina mediante un único rasgo o es un grupo de subtipos relacionados con diferentes etiologías, los cuales estarían bajo la influencia de diferentes genes<sup>7</sup>. Esto mismo podría plantearse también para la falta de atención. Podrían existir diferentes etiologías para los mismos síntomas ya que el TDAH no describe simplemente un síndrome, sino estados biológicamente muy heterogéneos y, por lo tanto, agrupar a los sujetos en un único grupo basándose en el fenotipo clínico puede ser limitado a la hora de identificar factores genéticos causantes de dichos fenotipos<sup>144</sup>. Es más, algunos genes podrían estar asociados solamente a ciertas habilidades cognitivas en cuyo caso sería difícil detectar la asociación utilizando grupos categóricos como casos y controles, y esta es la razón por la cual centrar los esfuerzos en buscar relaciones genéticas con rasgos neurológicos que fundamentan cada uno de los trastornos puede producir resultados más exitosos. Todo lo mencionado hace muy dificultosa la comparación de las conclusiones obtenidas hasta el momento. Además, en cada grupo de investigación se analizan distintas variantes genéticas. Esto también podría explicar por qué los resultados que relacionan los genes candidatos de la dislexia o el TDAH con los rasgos cognitivos son muy inconsistentes entre los diferentes estudios<sup>37,68,241</sup>, así como con los obtenidos en este proyecto, los cuales no correlacionan en su totalidad con los descritos en la literatura.

Nuestro estudio muestra algunas relaciones sistemáticas entre procesos cognitivos y genes. En lo que concierne a la **lectura**, diferentes grupos de investigación<sup>32,44,68,257,370</sup> han asociado los genes *DCDC2*, *KIAA0319* y *DYX1C1* con la lectura de palabras regulares, irregulares y pseudopalabras, mientras otros no encuentran esos mismos resultados<sup>37,241</sup>. En nuestra población se replicó la asociación de *KIAA0319* y *DYX1C1* pero solamente en la tarea de lectura de palabras reales, sin embargo el gen que parece tener una relación consistente con esta habilidad es *MAOA*, aunque los resultados que conciernen a este gen hay que considerarlos con precaución, ya que es un gen situado en el cromosoma X y en esta parte del proyecto los chicos y las chicas se han analizado en conjunto. Si bien las deficiencias en la ruta fonológica son una posible explicación para los problemas de lectura, también lo son las deficiencias en la ruta léxica. La implicación de estas rutas estaría relacionada con la lectura de palabras de alta frecuencia (ruta léxica) y de pseudopalabras o de palabras de baja frecuencia (ruta fonológica)<sup>371,372</sup>. En investigaciones previas se ha documentado el **efecto de lexicalidad** en los niños escolares utilizando para ello la tarea de decisión léxica<sup>371-373</sup> y, en nuestro estudio, se han relacionado los polimorfismos de los genes *DYX1C1* y *DBH* con este efecto, particularmente con las respuestas correctas. Diferentes alelos de *DAT1* también se vieron relacionados con esta función.

Es preciso mencionar también que a partir de los resultados analizados con los SNPs de los genes candidatos genotipados en los arrays, se observó una epistasia significativa entre dos marcadores situados en el gen *DAT1* en la población disléxica analizada. La implicación de este gen con la lectura ya se había visto anteriormente. Cornish *et al.*<sup>364</sup> muestran una fuerte y directa relación entre *DAT1* y la lectura, y aunque Shaywitz<sup>374</sup> propone que la alteración de los mecanismos atencionales influye en las dificultades para la lectura, los resultados de Cornish *et al.*<sup>364</sup> eran inexplicables como consecuencia de los síntomas de TDAH, por lo que **se sugiere que el genotipo 10R10R del VNTR en *DAT1* podría ser un posible candidato para explicar las disrupciones en la adquisición de la lectura y los síntomas de inatención del TDAH**. Esta podría ser una de las razones de las inconsistencias halladas entre los estudios de asociación de *DAT1* con los fenotipos neuropsicológicos del TDAH, siendo posible que las asociaciones reflejen en realidad la influencia incontrolada de la habilidad de lectura y la variación en la frecuencia de los genes compartidos entre ambos trastornos<sup>364</sup>.

Otro rasgo cognitivo que está directamente relacionado con el proceso de la lectura es la **fluidez**. La asociación del gen *DCDC2* y la fluidez para la lectura se ha identificado anteriormente<sup>349,375</sup>, y una alteración en *DYX1C1* y varias localizadas en *KIAA0319* también se han asociado previamente con la tarea RAN (del inglés, *Rapid Automated Naming*)<sup>42,370,376</sup>. En este proyecto proponemos añadir a esta lista los marcadores rs4504469-*KIAA0319* y rs6323-*MAOA*, ya que se han asociado al número de respuestas correctas en el conjunto de tareas que consisten en identificar dibujos y colores.

La decodificación fonológica tiene una función central en el desarrollo normal y anormal de la lectura y ha sido repetidamente relacionada con la habilidad para la lectura<sup>377-379</sup>. De hecho, los déficits en la decodificación fonémica, los cuales están estrechamente ligados a la **conciencia fonológica**, se consideran la principal causa de la mayoría de los casos de dificultades en la lectura<sup>380</sup>. Algunos genes candidatos de dislexia se han correlacionado con las tareas que miden este proceso cognitivo, tal como *DYX1C1* y *FOXP2* con la memoria a corto plazo<sup>32,35,44</sup> y con la memoria fonológica<sup>42,109</sup>, o los genes *DCDC2*<sup>65</sup> y *KIAA0319*<sup>65,68,349,376</sup> con la conciencia fonémica. En nuestro estudio, además de replicar la asociación de *DCDC2* con una medida de la conciencia fonológica (el tiempo de reacción global en la tarea de identificación de fonemas en dibujos), también hemos detectado una posible implicación de los genes *DBH* con la misma tarea, *MAOA* con la tarea de delección del fonema inicial y *DRD4* con ambas tareas.

La relación de la **percepción del habla** en las muestras disléxicas se propuso hace tiempo<sup>381</sup>, particularmente con el déficit en el procesamiento automático de la información medida por el componente de *MisMatch Negativity* (MMN). Esta medida, al parecer depende de la memoria a corto plazo y se ha visto reducida en los disléxicos<sup>18</sup>. Se considera una medida objetiva para la habilidad de la discriminación del habla así como un endofenotipo potencial para la dislexia<sup>382</sup>. Además, algunos investigadores han relacionado la **MMN** con variantes genéticas de *SLC2A3*<sup>284</sup> o variantes raras en alto LD en *DCDC2* y una región de *KIAA0319*<sup>383</sup>. En el presente estudio encontramos correlaciones entre el gen candidato para la dislexia *KIAA0319*, así como entre *MAOA*, *DBH* y *DRD4*, todos candidatos para TDAH, con la tarea de **discriminación silábica**. Aunque estas tareas miden procesos diferentes (procesamiento del habla automático vs atento), ambas están relacionadas con el procesamiento del habla.

Debilidades en los **dominios ejecutivos** como en la memoria verbal de trabajo, planteamiento y respuesta inhibitoria, son los rasgos cognitivos más consistentes que podrían explicar parte de los síntomas del TDAH<sup>145,384-387</sup>. En el presente trabajo no se ha replicado la asociación encontrada por Fosella *et al.*<sup>388</sup>, cuyo estudio mostró una relación de MAOA con la atención ejecutiva y de alerta. Sin embargo, la conexión encontrada entre rs4680-COMT y el índice de conflicto (basado en el test ANT) en nuestros resultados, relaciona este gen con el dominio ejecutivo. Varios estudios han asociado recientemente variantes de COMT con dominios cognitivos como la memoria de trabajo<sup>389</sup> o la flexibilidad cognitiva<sup>390</sup>, además de haberse asociado a las tareas de Stroop atencional<sup>391</sup>. Estas tareas atencionales no suelen analizarse generalmente con los genes candidatos de dislexia, por lo que un resultado intrigante obtenido en nuestra cohorte es la relación de los genes KIAA0319, DYX1C1 y FOXP2 con la tarea atencional de Stroop, el cual forma parte de un campo hasta el momento bastante desconocido. No obstante, nuestros resultados indican la posible vinculación de los genes presuntamente relacionados con la lectura y los procesos atencionales.

Sin duda DRD4 y DAT1 son los dos genes clásicamente relacionados con el TDAH, y también nosotros detectamos asociación de diversos alelos de ambos genes con las tareas atencionales medidas a través del test ANT o de la prueba Stroop. Los estudios donde analizan la correlación entre los diferentes alelos de estos genes con respecto a las medidas continuas en la población general no abundan. No obstante, existen evidencias de correlación de los síntomas de hiperactividad<sup>392</sup> e inatención<sup>393</sup> con DAT1 y de ambos síntomas también con DRD4<sup>393</sup>, aunque no todos revelan evidencias significativas<sup>394</sup>. El alelo de 4 repeticiones de DRD4 muestra una peor resolución del conflicto ejecutivo<sup>388</sup> y el de 7 repeticiones una peor memoria de trabajo espacial<sup>395</sup>. Los genes DAT1 y COMT también se han relacionado con esta última medida cognitiva<sup>396</sup>, aunque no todos los investigadores obtienen resultados positivos<sup>397,398</sup>. Sin duda, la tarea cognitiva más consistente ha sido el control inhibitorio. Tanto el alelo 07R de DRD4 como el 10R de DAT1 se han asociado con medidas de inhibición de respuesta<sup>398-400</sup>.

Muchos investigadores<sup>150</sup> se han centrado en la discrepancia entre la presencia o ausencia del alelo de 07R-DRD4, considerado el alelo de riesgo para TDAH. Sorprendentemente, se ha mostrado que los pacientes con el trastorno portadores del alelo de riesgo 07R hacen menos errores de comisión y omisión que los no portadores de este alelo en las pruebas de atención

sostenida<sup>401,402</sup>, y responden más rápido en las tareas de respuesta inhibitoria<sup>403</sup>, aunque estos resultados no siempre han sido consistentes<sup>404</sup>. Estos efectos resultan paradójicos, ya que el alelo de 07R debería conferir riesgo hacia el TDAH, en parte mediante estos procesos cognitivos. Una posible explicación es que los portadores y los no portadores de este alelo tengan subtipos diferentes del trastorno, cada uno caracterizado por diferentes déficits cognitivos<sup>159</sup>. Sin embargo, en las pruebas para medir la función ejecutiva, la orientación y la alerta, los pacientes de TDAH portadores de alelo 07R son más lentos que los pacientes no portadores de este alelo<sup>159</sup>. En cuanto al alelo de 4 repeticiones, los pacientes de TDAH homocigotos 04R04R hacen menos errores de comisión y omisión que los portadores de una única copia de dicho alelo, lo que sugiere que este *locus* tiene un efecto protector<sup>404,405</sup>.

Teniendo todo esto en mente, podemos asumir que aun habiendo encontrado asociación de *FOXP2* y *COMT* con la dislexia en nuestro estudio de casos y controles, estos genes no se relacionan con ninguna variable de lectura en la población general, lo cual sugiere que podrían ejercer su función en los fenotipos más extremos de la habilidad para la lectura, aunque se relacionan con las tareas atencionales. Sin embargo, *KIAA0319*, también asociado con dislexia, parece actuar mediante las variables de discriminación silábica, RAN y lectura de palabras de alta frecuencia, y también se relaciona con la tarea de Stroop numérico. Por otro lado, *DBH* y *MAOA* se correlacionaron con tareas atencionales y de lectura en la población general, pero no mostraron ninguna relación individual con los fenotipos definidos. Los alelos de riesgo de los VNTRs situados en *DRD4* y *DAT1* no se asocian con la falta de atención en nuestra cohorte, si bien varios genotipos están asociados con el grupo disléxico así como con algunas de las tareas cognitivas relacionadas con la lectura y la atención. Estos datos refuerzan la importancia del estudio de medidas cuantitativas aplicadas a la definición de los fenotipos considerados, con el objetivo de descubrir evidencias de causa-consecuencia entre genes y procesos cognitivos respectivamente, en vez de centrarse solamente en los grupos categóricos extrapolados a partir de ellos.

### **5.2.7 Búsqueda de otros posibles genes causales de estos trastornos neurológicos**

La implicación genética en el desarrollo de la dislexia y la falta de atención está universalmente aceptada. Los estudios con gemelos indican que ciertos genes están implicados

en estos trastornos, pero no revelan cuántos ni qué genes. En estos últimos años las investigaciones exponen una serie de *loci* como implicados en la dislexia, predominantemente los mencionados anteriormente, y la migración neuronal como una de las causas principales<sup>21</sup>. Pero existe un problema con esta ruta causal hipotética, ya que es de esperar que una migración neuronal anormal afecte también otros aspectos cognitivos del desarrollo y no sólo al rasgo selectivo de la lectura<sup>1</sup>. De esta forma, se esperaría que estos genes pudieran causar un trastorno del aprendizaje más generalizado o incluso retraso mental y no únicamente dislexia<sup>4</sup>. Además los efectos de los genes candidatos en la migración neuronal no se restringe solamente a áreas corticales específicas relacionadas con la lectura<sup>10</sup>, por lo que en vez de buscar “genes para la dislexia” estaríamos buscando “genes del desarrollo neurológico”, que podrían determinar una gran variedad de condiciones clínicas<sup>247</sup>. Otra posible explicación es que haya otros genes aún sin identificar que interaccionen con los genes responsables de la migración neuronal. En el caso del **TDAH**, si bien **se han encontrado genes que parecen estar fuertemente asociados a este trastorno**, los efectos genéticos de estos no superan el 5 % de la variabilidad fenotípica total del mismo<sup>216</sup>. Es muy probable que todavía falten por descubrir muchos genes de susceptibilidad, siendo muy interesante el verificar si estos implican mecanismos moleculares similares a los encontrados a partir de los candidatos, o nuevas rutas moleculares<sup>18</sup>.

Mediante los estudios de asociación de genoma completo es posible descubrir parte de la influencia genética no observada en las estrategias enfocadas en genes candidatos, así como describir posibles rutas afectadas de las que partir en busca de nuevas variantes genéticas relacionadas con estos trastornos.

### 5.3 LOS ESTUDIOS DE GWAS DESVELAN SNPS, GENES Y RUTAS FUNCIONALES DE INTERÉS PARA LA DISLEXIA Y LA FALTA DE ATENCIÓN

En el presente proyecto, ninguno de los marcadores llegó a pasar el umbral de significancia de  $p < 7.7E-08$ , establecido tras su corrección por el número de tests realizados en este GWAS, bajo ningún modelo genético en ninguno de los estudios de caso/control elaborados. Sin embargo, el hecho de que entre los marcadores más significativamente asociados a los diferentes trastornos hubiera una proporción mayor de la esperada de SNPs ubicados en regiones funcionalmente

importantes como las regiones 3'UTR o 5'UTR (ver Figura 20) aportó confianza con los resultados obtenidos.

No es necesario discutir la importancia de las mutaciones en las regiones exónicas dado que son las responsables más directas de la codificación errónea de proteínas. En referencia a las regiones UTRs (del inglés *UnTranslated Regions*), localizadas en los extremos 5' y 3' de los genes, son las responsables de la regulación de la traducción del mRNA<sup>406,407</sup>. Si bien la mayoría de los trastornos genéticos humanos descritos hasta el momento están causados por mutaciones que afectan a regiones codificantes de proteínas, durante los últimos años, la des-regulación a nivel traduccional (incremento o disminución de la eficiencia de la generación de proteína) se ha definido como un mecanismo causante de ciertas enfermedades o de la progresión de las mismas<sup>407</sup>. Es más, dicha desregulación es tan importante como la regulación transcripcional para un funcionamiento celular adecuado<sup>408</sup>. Los ejemplos mejor estudiados de la des-regulación de la traducción están mediados por las regiones 5'UTR, e involucran mutaciones que alteran el control de iniciación de la misma<sup>406-408</sup>. Por otro lado, las regiones 3'UTR suelen afectar a la traducción indirectamente, es decir, mediante la determinación de la estabilidad del mRNA, aunque variaciones que afectan a las señales de poliadenilación, terminación del codón o en las estructuras secundarias de la región 3'UTR, también podrían estar involucradas en la des-regulación del proceso<sup>407,408</sup>. El listado de enfermedades genéticas causadas por mutaciones que afectan la expresión de mRNA aumenta rápidamente y, debido al gran número de proteínas involucradas en el proceso de traducción, es probable que existan aún más enfermedades relacionadas<sup>409</sup>. Según los datos obtenidos en este proyecto, los individuos con FA exhiben un aumento de mutaciones de alta significancia en la región UTR 3', y los disléxicos y comórbidos en las regiones UTR 5'. Estos resultados nos permiten hipotetizar que estos trastornos cognitivos podrían formar parte del listado de afecciones citado anteriormente.

### **5.3.1 Mutaciones en el gen KANK1 candidatas a ser causales del desarrollo de dislexia**

Al validar los marcadores que resultaron con mayor asociación en nuestros estudios de GWAS mediante la tecnología TaqMan<sup>TM</sup> y aumentando el número de muestras caracterizado, el marcador rs10120392 situado en el gen KANK1 (*kn motif- and ankyrin repeat domain-containing protein 1*), también conocido como ANKRD15, llegó a pasar el umbral de significancia establecido

en nuestro GWAS de dislexia (una vez excluidas las muestras comórbidas) bajo el modelo alélico ( $p=4.17E-08$ ). Este gen se localiza en la posición 9p24 y es miembro de la familia de proteínas KANK. Entre sus funciones destacan (i) la formación del citoesqueleto por la regulación negativa de la polimerización de la actina, (ii) la regulación de la polaridad celular durante la migración dirigida mediante la influencia en los procesos de la reorganización del citoesqueleto<sup>410</sup> y la estabilización de los microtúbulos en las adhesiones focales<sup>411</sup>, (iii) la inhibición de la migración celular vía las rutas de señalización de RHOA y RAC<sup>412,413</sup>, así como la formación de los lamelipodios pero no la de los filopodios<sup>413,414</sup>, (iv) la inhibición de la producción y el crecimiento de las neuritas<sup>413</sup> (v) y también se ha visto que regula positivamente la apoptosis e inhibe el crecimiento celular en los tumores malignos del sistema nervioso<sup>415</sup>. La región 9p24.3-9q34.3, dentro de la cual se ubica KANK1 está enriquecida en deleciones cortas y patógenicas<sup>416</sup>, además esta zona se ha relacionado con diversos carcinomas y otras enfermedades. A parte de su implicación como supresor tumoral del carcinoma renal<sup>417</sup>, se han identificado mutaciones causales en este gen en pacientes con una enfermedad del desarrollo del sistema nervioso denominada *palsy spastic quadriplegic type 2*<sup>418</sup>. Asimismo, mutaciones cercanas a nuestro marcador asociado se han relacionado con la glucosa en plasma<sup>419</sup> y el procesamiento y secreción de la insulina<sup>420</sup>. Diferentes investigaciones también han asociado esta región con el síndrome Afi (caracterizado por retraso mental y autismo)<sup>421</sup>, o con el trastorno obsesivo-compulsivo<sup>422</sup>. En la región flanqueante de KANK1 destacan varios genes que son potencialmente responsables de fenotipos conductuales, incluyéndose FOXD4 (de la misma familia que FOXP2)<sup>106</sup> y DOCK8<sup>423</sup>. Por eso, es razonable especular que KANK1 puede estar involucrado en enfermedades conductuales o relacionadas con el sistema nervioso<sup>413</sup>. Mutaciones en este gen se convierten de esta forma en **buenas candidatas para la susceptibilidad de padecer dislexia**. Asimismo, la mayor significancia encontrada en el grupo de disléxicos puros en comparación con el grupo que también contenía muestras comórbidas podría indicar su especificidad hacia la dislexia ( $p=4.17E-08$  dislexia pura y  $p=2.62E-06$  junto con los comórbidos). Su interacción con el marcador rs7047127 ( $p=5.39E-12$ , Tabla 23) ubicado en el gen CYP4F8 podría **implicar a los ácidos grasos con este trastorno neurológico**. CYP4F8 cataliza numerosas reacciones del metabolismo de fármacos y de la síntesis del colesterol, esteroides y otros lípidos, y también metaboliza ácidos grasos omega-3 (el ácido decosaheptaenoico (DHA) y el ácido eicosapentaenoico). Además, otros genes con relación lipídica también ocupan los primeros puestos de asociación de nuestro estudio disléxico como son: OSBPL6, LDLRAD4 y PLCZ1 (Tabla 22 y Tabla 23). La implicación del metabolismo de los ácidos

grasos y de los fosfolípidos de la membrana celular en los trastornos neurológicos del desarrollo como la dislexia, el TDAH o el autismo ha sido estudiada anteriormente<sup>424</sup>, incluso se propone que su relación con los lípidos podría explicar el grado de comorbidad entre las diferentes condiciones. La deficiencia de ácidos grasos también se ha relacionado con una capacidad menor en la lectura<sup>425</sup> y esta deficiencia está correlacionada con la severidad de los síntomas de la dislexia, particularmente en los dominios visual, auditivo y lingüístico, y asimismo también se relaciona con problemas motores<sup>426</sup>.

Entre los genes más significativos de esta tesis también se han encontrado algunos afines al metabolismo lipídico en el estudio de FA (*MOGAT1*, *DGKB* y *PLD5* Tabla 24 y Tabla 25) y en el de comórbidos (*SCARB1*, *PTPLAD2*, *PLCB1* y *LPPR5* Tabla 26 y Tabla 27), así como rutas implicadas en el metabolismo lipídico enriquecidas en los tres GWAS (ver apartado 4.3). Estudios centrados en determinar el perfil lipídico en pacientes con TDAH han mostrado que ocurren cambios significativos en la proporción de ciertos lípidos en dichos pacientes en comparación con los controles<sup>427</sup>, y que algunos síntomas reportados para TDAH son similares a los observados en las personas con deficiencia de ácidos grasos esenciales<sup>428</sup>. Por ello abundan las investigaciones sobre la ingesta de ciertos ácidos grasos en la dieta para prevenir el TDAH, ya que además de la farmacoterapia como tratamiento, los suplementos nutricionales también podrían tener un gran impacto en la sintomatología de este trastorno<sup>427,429</sup>.

### **5.3.2 Los elementos involucrados en la transmisión sináptica: la causa más observada de trastornos cognitivos**

Cumpliendo con nuestras expectativas, muchas de las rutas enriquecidas a partir de los SNPs más asociados en el presente proyecto pertenecen a la **regulación sináptica** y a la **estructura neuronal** y han aparecido en todas nuestras comparativas, tanto en dislexia, falta de atención y comórbidos como en los resultados derivados de QTLs con variables de lectura y de atención (ver apartado 4.5). La **transmisión sináptica** es un factor esencial para la transferencia de la información por lo que cualquier mal funcionamiento en algún elemento clave así como en la regulación de este proceso (canales de calcio, sodio potasio, estructura de la neurona, etc.) podría estar involucrado en el desarrollo de trastornos cognitivos. En los tres estudios caso/control llevados a cabo, se ha detectado cierta asociación de un número elevado de marcadores ubicados

en genes relacionados con esta función, por lo que discutir la posible implicación de cada uno de ellos resulta complicado. Por ello, tan sólo vamos a resaltar estos hallazgos:

-El marcador rs3764714, del gen *SNPH*, asociado en nuestro estudio de dislexia ( $p_{\text{alel}}=7.26 \text{ E-}05$ , Tabla 22) se encuentra muy cerca del marcador rs3764715 (*SNPH*) asociado débilmente en este caso al TDAH ( $p=0.043$ ) por el equipo de Sánchez-Mora<sup>430</sup>. El gen *SNPH* codifica una proteína que está asociada a la membrana y es responsable de inhibir la formación del complejo SNARE (*Soluble N-ethylmaleimide-sensitive factor Attachment protein REceptor*) mediante la unión a syntaxin-1 (o mediante absorción de la syntaxin-1 libre). El complejo SNARE está formado por syntaxin-1, synaptobrevin/VAMP y SNAP25 y es necesario para el empaquetamiento y la fusión de vesículas.

-Cuatro de los SNPs más significativos en nuestro GWAS de dislexia (rs4944452 (*DLG2*) en interacción con rs2255532 (*COL7A1*),  $p=1.33\text{E-}12$ , Tabla 23; rs7682286 (*SORCS2*),  $p_{\text{alel}}=3.74\text{E-}05$ ; rs13132581 (*SORCS2*),  $p_{\text{geno}}=3.68\text{E-}05$  y rs7682286 (*SORCS2*),  $p_{\text{rec}}=2.38\text{E-}05$  (ver tablas suplementaria 8,9 y 11)) están ubicados en genes (*DLG2* y *SORCS2*) relacionados con la interacción de los sitios postsinápticos. Además, *SORCS2* se expresa fuertemente en el sistema nervioso central (<http://www.ncbi.nlm.nih.gov/pubmed>). Algunos marcadores situados en estos dos genes se han asociado previamente a medidas atencionales en pacientes españoles (rs17147674 (*DLG2*) y rs4689642(*SORCS2*),  $p=5,21\text{E-}06$  y  $p=3.65\text{E-}07$ , respectivamente)<sup>303</sup>.

-El gen *NRXN1* codifica una proteína que se une a las neuroliginas formando el complejo neurexin/neuroligin en las sinapsis del sistema nervioso central, el cual es esencial para una neurotransmisión eficiente y la formación de los contactos sinápticos. Si bien Neale *et al.*<sup>293</sup> no obtuvieron resultados significativos para este gen en su estudio de TDAH, uno de los SNPs (rs17495366) resultó entre los 100 más significativos, lo que aporta confianza a la asociación encontrada en nuestras muestras de falta de atención (rs12617277,  $p_{\text{dom}}=6.83\text{E-}05$ ), tabla suplementaria 17).

-Un marcador del gen *CNTNAP2*, otro de los genes de neurexina, ha resultado entre los más significativos de nuestro estudio comórbido (rs2373284 (*CNTNAP2*) en interacción con rs986613 (*PTPRG*)  $p=2.91\text{E-}07$ , Tabla 27). Este gen se ha asociado previamente a varias condiciones neurológicas del desarrollo como SLI (*Speech Language Disorder*)<sup>107</sup>, autismo<sup>431</sup> y dislexia<sup>109</sup>.

### 5.3.3 La adhesión celular y la estructura estrechamente relacionados con la migración neuronal

Otras rutas predominantes en las asociaciones obtenidas en nuestro proyecto han sido las relacionadas con la estructura y la adhesión celular, tales como la regulación y organización de las uniones celulares o rutas relacionadas con la matriz extracelular. No es de extrañar que estas rutas también predominen como influyentes de estos trastornos cognitivos, ya que la migración neuronal se alcanza mediante el reordenamiento de los componentes del citoesqueleto en respuesta a señales o pistas extracelulares<sup>26,432</sup>. Para ello se requiere un patrón de interacciones coordinadas entre la matriz extracelular, las señales secretadas como guías, las moléculas de adhesión celular, las quinasas de tirosina situadas en la superficie celular y los receptores de señalización de integrinas<sup>24</sup>. Muchos de los SNPs más asociados en las tres comparativas planteadas, es decir, en los GWAS de dislexia, falta de atención y comorbidad, se localizan en genes encargados de la adhesión y señalización molecular. En ese gran listado de genes, aparte del ya mencionado *KANK1*, el único que superó el umbral de significancia establecido para los GWAS en la fase de validación, aparecen otros que ya se habían relacionado en anteriores estudios con alguno de estos trastornos o variables cognitivas afines. El rs2409411 del gen *TIAM1* se asoció ( $p=0.004$ ) a variables de lectura en un estudio con muestras de U.K.<sup>280</sup>. Sin embargo, este resultado no se replicó posteriormente en muestras australianas<sup>281</sup>. Si bien no es la misma variante, en nuestro estudio de comórbidos el rs475142, situado en este mismo gen, también alcanza los primeros puestos de significancia ( $p_{\text{alel}}=9.09\text{E-}05$ , ver tabla suplementaria 22). El gen *PTPRG*, también está representado entre los marcadores de mayor asociación en el meta-análisis genético del TDAH realizado por Neale *et al.*<sup>301</sup>, y en nuestro estudio se refuerza la importancia de este gen con la falta de atención (rs17066305,  $p_{\text{alel}}=1.46\text{E-}05$ , Tabla 24) y su comorbidad con la dislexia (rs986613 (*PTPRG*) en interacción con rs2373284 (*CNTNAP2*)  $p=2.91\text{E-}07$ , Tabla 27).

Otros genes que representan bien la función estructural son las cadherinas puesto que son las principales moléculas de adhesión celular. La cadherina 13 (*CDH13*), cuyas variantes se han analizado en diversos estudios en relación al TDAH, ha obtenido resultados significativos<sup>291</sup> y no significativos<sup>292,433</sup>. Sin embargo, en nuestro GWAS de dislexia obtuvimos un valor de asociación

de  $9.47E-12$  (Tabla 23) en la interacción del SNP rs4290460 de *CDH13* y el rs7116459 de *DENND5A*. Este gen ya se había relacionado anteriormente con la lectura<sup>285,286</sup>. En cuanto a la falta de atención, fue la cadherina 4 (*CDH4*) la única que se ha asociado a este trastorno, esta vez también en interacción con el gen *GRIA1*, el cual codifica un receptor de glutamato y, en consecuencia, es importante en la neurotransmisión del sistema nervioso central (rs2910260 (*GRIA1*) y rs1892318 (*CDH4*),  $p=4.92E-10$ , Tabla 25). En las muestras comórbidas también se identifican SNPs situados en este tipo de moléculas de adhesión, particularmente en los genes *CDH11* y *CDH12*, con valores de asociación interesantes (rs1520233 (*CDH11*) en interacción con rs3767498 (*CACNA1S*),  $p=9.41E-07$ , Tabla 27; rs6884278 (*CHD12*),  $p_{\text{alel}}=7.45E-05$ , tabla suplementaria 22).

### 5.3.4 Ciertas disfunciones en el sistema inmunitario podrían estar implicadas en el desarrollo de trastornos neurológicos

Cabe destacar los resultados más intrigantes en cuanto a las rutas con función inmunitaria enriquecidas de variantes en estos trastornos neurológicos que son las rutas relacionadas con el complejo proteico MHC, señalización mediada por el interferón gamma o presentación de antígenos entre otras. La actividad inmunitaria en el Sistema Nervioso Central (SNC) está restringida principalmente por la barrera hematoencefálica<sup>434</sup>, la cual se sitúa entre los vasos sanguíneos y el SNC e impide la entrada de sustancias tóxicas. Durante tiempo se creyó que la detección de células inmunitarias en el SNC era característico en los pacientes con neuropatologías en los que se hipotetizaba que la barrera hematoencefálica estaría afectada, permitiendo la infiltración de dichas células<sup>435</sup>. Hoy en día en cambio, se acepta la existencia de una colaboración entre el sistema nervioso y el sistema inmunitario, y que dicha interacción es esencial para el funcionamiento normal del cuerpo<sup>436</sup>. Es más, el sistema inmunitario puede ser beneficioso o incluso protector ante una patología del SNC<sup>437</sup>. No se puede ignorar que **un mal funcionamiento inmunitario pueda contribuir potencialmente o incluso sea el factor causal en la etiología de las enfermedades del desarrollo neurológico, cognitivas o psiquiátricas**<sup>437</sup>. Esta interacción inmunitaria podría producirse en la zona de las meninges (membranas que cubren el cerebro) ya que están pobladas densamente por células mieloides (macrófagos, linfocitos T, etc.). Además, Louveau *et al.*<sup>438</sup> han descubierto recientemente que hay vesículas linfáticas capaces de transportar fluidos y células inmunitarias en las meninges de cerebros de ratones, sugiriendo que un mal funcionamiento de estas vesículas podría ser causa de diversos trastornos neurológicos, lo cual podría hacer que se replantea la presunción básica de la neuroinmunología. Este mismo

grupo de investigación<sup>434</sup> ya había definido anteriormente **las áreas meningeales como lugares importantes para la respuesta inmunitaria que ocurre durante el aprendizaje**, y que los linfocitos T ayudan en este proceso. Diversos estudios realizados en ratones muestran que la ejecución de tareas cognitivas produce una acumulación de linfocitos T, productores de IL-4 en las meninges<sup>434</sup>, y su deficiencia está acompañada frecuentemente por defectos cognitivos<sup>439,440</sup> afectando al aprendizaje y a la memoria<sup>435,437</sup>. Además las células inmunitarias también podrían influir en la respuesta ante el estrés psicológico<sup>441</sup> o la neurogénesis<sup>442</sup>. Estos efectos adversos suelen mejorar o revertir mediante la reposición de los linfocitos T<sup>437,442</sup>. En cuanto a su mecanismo de acción, se propone que ante una situación de estrés se liberan mediadores moleculares que dirigen la respuesta inmunitaria. Algunos de estos mediadores podrían ser neuropeptidos o neurotransmisores (acetilcolina, adrenalina, noradrenalina o dopamina)<sup>443</sup>, la dopamina particularmente, tiene función inmunomodulatoria y los linfocitos pueden producirla por sí mismos<sup>435</sup>.

Los argumentos que vinculan la dislexia con el sistema inmunológico vienen de años atrás, cuando en 1991 el grupo de Smith<sup>444</sup> implicó polimorfismos proteicos del cromosoma 6 en las proximidades del complejo MHC con la dislexia. Sin embargo posteriormente, empleando las mismas muestras, no se pudieron conectar las enfermedades inmunológicas con la dislexia<sup>445</sup>. Entre los SNPs de mayor asociación de nuestro estudio, hemos identificado algunos ubicados en genes con función inmunológica como *IRF2* (rs11723606 y rs11726275 bajo diferentes modelos genéticos, como haplotipo y en ventana corrediza, Tabla 22 y tablas suplementarias 8-13) y *TESPA1* (rs2252795 bajo diferentes modelos genéticos y en un haplotipos de 5 marcadores, ver tablas suplementarias 8-10 y 12), ambos asociados a dislexia, o *NOD2* ( $p_{\text{haplo}}=5.54E-05$  incluyendo los marcadores rs4785224, rs5743266, rs2111234, rs13380741, rs11647841, rs2066843 y rs1861757 Tabla 26, y tabla suplementaria 26) y *IKBKB* (rs4386173 (*IKBKB*) en interacción con rs2277414 (*FOXJ2*),  $p=5.68E-07$ , Tabla 27) relacionado con la situación comórbida. En este último grupo también observamos enriquecimiento de las rutas de “rechazo a aloinjertos (hsa05330)” y “rechazo a trasplantes (hsa05332)”, coincidiendo con lo descrito por Alemany et al.<sup>303</sup>, quienes buscaron marcadores a lo largo de todo el genoma que estuvieran asociados a síntomas de falta de atención en adultos españoles con TDAH y analizaron si esas correlaciones estaban enriquecidas en ciertas rutas biológicas.

### 5.3.5 La deficiencia de zinc podría tener relación con el desarrollo neurológico defectuoso y los trastornos cognitivos

Otro dato que nos llamó la atención de nuestros resultados fue la cantidad de **genes relacionados con la unión, transporte y respuesta a la presencia de moléculas de zinc** (ver Tabla 22, Tabla 24 y Tabla 26). Diversas investigaciones han relacionado el zinc o genes relacionados con este elemento y la dislexia. Por ejemplo, se ha mostrado que la concentración de zinc medida en sudor era significativamente menor en los disléxicos<sup>446</sup>. En estudios de asociación, se han propuesto los genes codificantes de “*zinc finger proteins*” (factores de transcripción de DNA) *ZNF280D*, *ZNF385D* y *ZNF804A* como candidatos para la dislexia o trastornos del lenguaje<sup>287,447,448</sup>. Nuestros análisis han relacionado el gen *ZNF385D* con la dislexia (rs11710743 (*ZNF385D*) en interacción con rs13110423 (*LDB2*),  $p=6.59E-12$ , y con rs283029 (*LDB2*),  $p=8.76E-12$ , Tabla 23). Este gen ya se había relacionado previamente con la lectura y con ciertas medidas de vocabulario<sup>287</sup>. En este mismo estudio, *ZNF385D* se asoció con la condición de comorbidad de RD y LI (*Language Impairment*) indicando la posibilidad de una implicación en los procesos comunes entre estos dos trastornos y, recientemente, se ha replicado su relación con la lectura de palabras y la velocidad de procesamiento en la población general holandesa<sup>286</sup>. Además también se ha observado una asociación de *ZNF385D* con el volumen global del cerebro lo cual afectaría a la conectividad y funcionalidad de las regiones cerebrales importantes para la eficiencia fluida del lenguaje verbal y escrito<sup>287</sup>. Asimismo, *ZNF385D* se ha implicado previamente con la esquizofrenia o con el TDAH<sup>20,449</sup>. Por otro lado, aunque no encontramos ninguna asociación significativa de SNPs ubicado en *ZNF804A*, Becker *et al.*<sup>448</sup> asocian este gen con una mejor habilidad cognitiva en base a medidas que constituyen el patrón característico de la dislexia (lectura de palabras y pseudo-palabras, deletreo de palabras, deleción de fonemas y nombrado rápido de letras y números). Teniendo en consideración sus resultados, concluyen que *ZNF804A* podría influir en rasgos cognitivos de importancia para la lectura y el deletreo<sup>448</sup>. En el estudio con muestras de U.K. realizado por Meaburn *et al.*<sup>280</sup> y en la replicación de sus mejores resultados en una segunda muestra Australiana<sup>281</sup> se asociaron marcadores ubicados en el gen *DPF3*, el cual también codifica una “*zinc finger protein*”, con rasgos de lectura. En nuestro proyecto, si bien el SNP no fue el mismo, también obtuvimos asociación de rs8007249 con las muestras comórbidas (valor de  $p_{\text{geno}} = 2.08E-05$ , Tabla 26, y tabla suplementaria 23), añadiendo validez a la posible implicación del gen en este rasgo cognitivo. Marcadores situados en otro de los genes relacionado con el zinc,

ZNF385B, también aparecieron entre los más asociados con nuestra cohorte de dislexia (rs6718486 (ZNF385B) en interacción con rs742922 (ANGPT4),  $p=1.08E-14$ , Tabla 23) así como con comórbidos (rs9541973,  $p_{\text{alel}}=8.01E-05$ , Tabla 26 y tabla suplementaria 22).

Aproximaciones con modelos animales han mostrado que el zinc está involucrado en el control del desarrollo de la neurogénesis y su deficiencia tiene implicaciones en el desarrollo cerebral, la cognición, así como en la propia neurogénesis, ya que este proceso se ve reducido mediante la limitación del número de precursores neuronales de proliferación<sup>450</sup>. Por ejemplo, se ha demostrado que el zinc es significativamente deficiente en los niños con TDAH<sup>451</sup>, aunque en nuestro proyecto no hemos encontrado ningún gen relacionado con él en falta de atención. **Los déficits de aprendizaje causados por una baja concentración de zinc en el organismo se pueden corregir mediante un suplemento de este elemento en la dieta**<sup>452</sup>. Es más, un suplemento de zinc en la dieta, acompañado del tratamiento general, mejora significativamente los síntomas de inatención<sup>453</sup>. Por ello, en los últimos años se está investigando la posible implicación de la dieta en la etiología del TDAH, y los resultados han mostrado que una dieta sana, así como la ingesta de vitaminas y minerales influyen en su desarrollo.

### 5.3.6 ¿Podría ser la dislexia una ciliopatía?

Un gran avance que plantea una nueva posible etiología para los trastornos neurológicos es la caracterización de la función de los cilios. Los defectos en los cilios se asocian con un amplio rango de enfermedades humanas denominadas ciliopatías, las cuales comparten síntomas comunes incluyéndose entre ellos algunos defectos neurocognitivos, sin embargo la relación entre ambos es poco conocida<sup>52,54</sup>. Según nuestros hallazgos con la población de **dislexia, la implicación de los genes relacionados con los cilios queda descartada para este trastorno en su estado puro**. Sin embargo, teniendo en cuenta la comorbidad con la falta de atención, hemos identificado ciertos marcadores ubicados en genes con función ciliar, tales como *BB59* (rs17170286,  $p_{\text{alel}}=8.06E-05$ , Tabla 26 y tabla suplementaria 22), *DNAH11* ( $p=1.34E-05$  para un haplotipo formado por los marcadores rs6979397 y rs2072221, Tabla 26 y la tabla suplementaria 26) y *ULK4* ( $p=7.71E-05$  para un haplotipo formado por los marcadores rs4973889, rs7629556, rs13068312, rs4973978, rs974595 y rs9828875, ver Tabla 26 y la tabla suplementaria 26). Puede que las alteraciones en los cilios sólo estén involucradas en la forma comórbida del trastorno, o puede que sólo afecten a los casos más severos de dislexia. Hasta ahora se desconocía que la

**dislexia** pudiera tener relación con los cilios, sin embargo, el hecho de que los genes candidatos para la etiología de la dislexia (*KIAA0319*, *DCDC2* y *DYX1C1*) tengan posibles implicaciones en las funciones ciliares, tanto en los cilios motiles como con el cilio primario, sugiere que **este trastorno podría considerarse como un nuevo tipo de ciliopatía**. Si bien la alteración de alguno de estos tres genes individualmente resulta en anomalías en la migración neuronal, todavía se desconoce si este fenotipo es una consecuencia de los defectos observados en las estructuras y funciones ciliares. No obstante, la influencia del cilio primario en los patrones cerebrales se ha descrito con anterioridad y varios defectos cognitivos se observan frecuentemente en los síndromes ciliopáticos<sup>54</sup>.

La lista de enfermedades que abarca esta pequeña extensión celular se expande constantemente, con un solapamiento de fenotipos y de causas genéticas entre los diferentes trastornos<sup>454</sup>. El gran rango de fenotipos presentes en las ciliopatías apunta a la posible diversidad de funcionalidad de los cilios. Por una parte, estos parecen ser especialmente importantes en la modulación de las cascadas de señalización subcelular (incluyendo las rutas Shh y Wnt, las cuales son importantes en el desarrollo neuronal) en varias etapas del desarrollo y en la homeostasis adulta<sup>78,454</sup>. Se sugiere que los cilios podrían actuar como foco de señalización capaz de concentrar y amplificar ciertas cascadas, procesos que hasta el momento sólo se le atribuían a la arquitectura neuronal<sup>454</sup>. Por otra parte, también se han visto involucrados en la orientación de la migración y el posicionamiento de las neuronas en el cortex cerebral<sup>455</sup>, en el desarrollo del tubo neuronal<sup>456</sup>, la polaridad celular, la orientación axonal y en la neurogénesis adulta. Esta última función es particularmente interesante ya que el bloqueo de la neurogénesis en el hipocampo podría producir defectos en las funciones cognitivas del aprendizaje y la memoria, en las cuales también se han involucrado los cilios<sup>457,458</sup>.

#### 5.4 EL ANÁLISIS DE QTLS MUESTRA REGIONES DE MAYOR ASOCIACIÓN QUE EL ESTUDIO DE CASOS Y CONTROLES

Como ya se ha mencionado anteriormente, la estrategia de analizar endofenotipos o rasgos cuantitativos en vez del clásico estudio de casos y controles podría aportar resultados más fructíferos, ya que las variables cuantitativas analizadas estarán más cerca de la sintomatología del fenotipo estudiado<sup>8,250,251</sup>. Una buena prueba de ello es la alta significancia de asociación

encontrada en nuestra estrategia de QTLs, tanto de forma global, como considerando únicamente las regiones genéticas fuertemente asociadas a alguna de las variables. El número de marcadores que pasan el umbral de significancia estadística establecido para los GWAS-QTLs es grande, abriendo así un abanico de posibles alteraciones causales que afectan a los rasgos cognitivos analizados. La denominada variable 3 (“Tiempo global en discriminación silábica”), la cual mide la percepción del habla, ha obtenido resultados especialmente significativos en cualquiera de los tres abordajes (GWAS de dislexia, de falta de atención y de comórbidos), indicándonos por un lado que este rasgo cognitivo, si bien se utiliza para medir procesos de lectura, no es específico de los sujetos con dislexia ya que también se ve afectado genéticamente en los demás grupos, o que este atributo está alterado en los dos tipos de trastornos y en la comorbidad de ambos. Por otro lado, el procesamiento ortográfico, medido en la variable 1 (“Efecto de transposición en aciertos”), se vio fuertemente asociado a una región situada en el cromosoma 11 en las muestras con dislexia y sus controles. Entre los genes que se encuentran en dicha región, se ubican 3 genes que por su función neuronal podrían estar relacionados con la habilidad para realizar la tarea cognitiva asociada: el gen *DLG2*, que también se asoció al fenotipo disléxico como se ha mencionado anteriormente, el gen *CNTN5* el cual regula las interacciones celulares durante el desarrollo del sistema nervioso y el gen *NCAM* que codifica una proteína de adhesión neuronal. Lo mismo sucedió con la variable 2 (“Tiempo de nombrado rápido compuesto de dibujos y colores”), la cual se utilizó para evaluar la fluidez de la lectura y la velocidad de procesamiento. Una región intergénica situada en el cromosoma 13 se asoció en la cohorte de dislexia y sus controles.

A partir de los SNPs más asociados con QTLs se identifican las mismas rutas que en la estrategia de casos y controles, es decir, también aparecen rutas relacionadas con canales, con el sistema nervioso, con la adhesión y estructura celular y señalización en general. Gran parte de las rutas funcionales asociadas con las variables 6 (“Efecto en aciertos de condiciones incongruente menos congruente (*Stroop* numérico)”) y 7 (“Efecto en tiempos de condiciones incongruente menos congruente (*Stroop* numérico)”) coinciden en ambas. Sin embargo, las dos regiones significativamente asociadas con la variable 6 no lo están con la 7. Ambas variables evalúan un mismo proceso cognitivo, por lo que estos resultados sugieren una especificidad génica para cada tarea cognitiva dentro de un mismo patrón metabólico alterado.

#### 5.4.1 Dos regiones situadas en los cromosomas 2 y 8 significativamente asociadas al procesamiento automático, la atención selectiva y el efecto conflicto

Uno de los resultados más llamativos y destacables de este proyecto fue la asociación hallada en dos regiones con la variable encargada de medir el procesamiento automático, la atención selectiva y el efecto conflicto (variable 6). Ambas regiones, una situada en el cromosoma 2 (2p13) y otra en el cromosoma 8 (8p21), se relacionaron en la población conjunta de los individuos FA y sus controles. En la región del cromosoma 8 no se localiza ningún gen, aunque no por ello deja de ser un hallazgo importante. Las zonas intergénicas muchas veces se descartan por no ser codificantes, pero ya hace años que se aportan evidencias sobre la utilidad del denominado “DNA-basura” en la investigación médica. Por ejemplo, Lunyak *et al.*<sup>459</sup> describen una secuencia de DNA intergénica crucial para el funcionamiento de un gen situado en otra región alejada. Cuando la causa de una patología se encuentra en zonas reguladoras intergénicas, los pacientes muestran síntomas de tener algún gen particular alterado sin que se observe ninguna mutación en el mismo. Si bien es cierto que en esta tesis debido al gran número de análisis realizados nos hemos centrado en las asociaciones cuyos marcadores se situaban en genes, existe información de indudable relevancia en el DNA intergénico, como las claves que permiten explicar por qué los genes se activan en determinados momentos del desarrollo, o por qué lo hacen en unas células y no en otras. Los marcadores en regiones codificantes están generalmente aceptados como “la respuesta” hacia la patología<sup>460</sup>. Sin embargo, una observación que sorprende es que la mayoría de los *loci* identificados mediante GWAS como asociados a enfermedades están en zonas no-codificantes, los cuales pueden influir en los rasgos biomédicos al afectar a la expresión génica mediante la alteración, por ejemplo, de la eficiencia transcripcional<sup>271,272</sup>. También es preciso remarcar que el reciente estudio de Alemany *et al.*<sup>305</sup> sobre QTLs en población de niños españoles, relaciona entre sus resultados más significativos el rs2320783 con la atención ejecutiva. Este marcador se localiza muy cerca del SNP más significativo de la región del cromosoma 8 descrita en nuestro estudio (rs11989470). Además, un gen de esta misma región, *DOCK5*, pertenece a la familia *DOCK*, la cual está relacionada con el desarrollo cerebral y las enfermedades neurológicas<sup>461</sup>. Otro resultado muy interesante de nuestro proyecto es que en la región flanqueante del gen *KANK1* (candidato de susceptibilidad a padecer dislexia según nuestro GWAS) también se sitúa un gen que pertenece a esta familia, el gen *DOCK8*<sup>423</sup>, previamente asociado a la discapacidad intelectual. Otros genes resaltados por Alemany *et al.*<sup>305</sup>

por la asociación de alguno de sus SNPs, son *NUFIP1*, *MAP2K4*, *PID1*, *EPB41L4A*, *MAG1*, *CLSTN2*, *CTGF* y *CTGF*, y todos coinciden con marcadores asociados significativamente en nuestro estudio de QTLs con variables atencionales.

En nuestro proyecto se identificó también un QTL en una región situada en el cromosoma 2 (2p13), la cual es muy rica en genes aunque, según las funciones adscritas a ellos, no se les puede atribuir una relación directa con procesos cognitivos. Aun así, destacan los genes *DCTN1*<sup>462</sup>, *TLX2*<sup>463</sup> y *SEMA4F*<sup>88</sup> los cuales cumplen funciones en el desarrollo del sistema nervioso y en el transporte axónico. Algo realmente interesante de esta región es que ya se había relacionado anteriormente y en varias ocasiones con la dislexia<sup>88,278</sup>, es más, se trata de la denominada región *DYX3* (2p11-2p16). De esta forma, según los resultados de nuestro estudio existe un solapamiento entre las regiones génicas involucradas en la lectura y en la atención. Por otra parte, diferentes estudios realizados en base al TDAH también han relacionado marcadores y genes en torno a esta región con dicho trastorno. Ejemplo de ello son los estudios de los grupos de Stergiakouli (2012)<sup>296</sup> y Ebejer (2013)<sup>297</sup>, quienes hacen mención especial de los genes *XPO1* (2p15) y *REG1B* (2p12) en relación al TDAH o sus síntomas; el primero a partir de un estudio de casos y controles en población caucásica y el segundo mediante una estrategia de QTLs en australianos.

## 5.5 CONSIDERACIONES FINALES

Tal como se ha descrito en el presente proyecto, la mayor proporción de marcadores que han resultado asociados con algún trastorno cognitivo están en genes distintos a los propuestos hasta el momento como candidatos a conferir alteraciones en la lectura o falta de atención. No obstante, en la estrategia de QTL hemos identificado 3 marcadores ubicados en el gen *MAOA* que obtuvieron valores de asociación moderados con la tarea encargada de evaluar la percepción del habla (variable 3) y dos marcadores situados en el gen *FOXP2* con la habilidad de la decodificación entre grafemas y fonemas. Por lo tanto, este hallazgo no descarta la posible vinculación de estos genes candidatos con las tareas cognitivas estudiadas.

**Hasta el momento, la investigación de las bases genéticas de dislexia y falta de atención, aunque intensa, no ha conseguido determinar variantes genéticas causales “oficiales” de las que partir en busca de soluciones para mejorar estos dos trastornos. La diversidad genética entre las**

poblaciones humanas es un obstáculo para extrapolar los resultados de los distintos estudios, sobre todo en los relacionados con la lectura, ya que al ser un rasgo que va estrechamente relacionado al idioma, los genes o rutas que puedan estar afectadas en este trastorno podrían variar dependiendo de ambos elementos, la población y el lenguaje utilizado por ésta. Se han identificado variantes asociadas a estas dos condiciones neurológicas, sin embargo estas no superan el 5% de la variabilidad fenotípica total del trastorno. Existen varios factores que podrían dar respuesta a este desmotivador resultado. Por una parte, la epístasis o interacción entre genes no se ha explorado lo suficiente hasta el momento y por otra parte, la interacción entre genes y ambiente puede inflar la estimación de la heredabilidad<sup>216</sup>. El concepto de heredabilidad no debería de ser un concepto fijo, ya que la proporción de la varianza total atribuida a los genes es parcialmente dependiente de la varianza en la exposición hacia el factor de riesgo ambiental relevante (el cual es desconocido), y de las características de la población estudiada<sup>19</sup>.

Un impedimento para el avance de las investigaciones neurogenéticas es la escasa comunicación existente entre los dedicados al estudio de las bases genéticas de los rasgos complejos y los investigadores de las funciones cerebrales. Ninguna de las dos disciplinas por separado puede dar una descripción completa de la enfermedad<sup>464</sup>. Es por ello que los abordajes de imágenes funcionales correlacionadas con descripciones genéticas deberían de ser una opción a tener en cuenta, ya que conceden un potencial para esclarecer los efectos de los genes en el desarrollo de las funciones cognitivas<sup>123,125</sup>. Si existen variantes genéticas que afectan al aprendizaje de un niño (en sus capacidades atencionales o lectoras) se necesitaría identificar cómo afectan éstas al desarrollo del cerebro y su función. De esta forma, se podrían desarrollar métodos de intervención para superar el problema, los cuales simplemente podrían consistir en una enseñanza individualizada ajustada a las necesidades de cada niño<sup>465</sup>.



## **6 CONCLUSIONES**



1. En el presente proyecto hemos demostrado que el **DNA que se extrae de saliva humana** es de calidad óptima para llevar a cabo estudios genéticos de alto rendimiento, y que su genotipado requiere una puesta a punto técnica que permita resultados reproducibles, debido al porcentaje variable de DNA humano presente en este tipo de biopsia.
2. Los análisis conjuntos o **meta-análisis de genotipos derivados de múltiples tecnologías**, como por ejemplo de distintas versiones o tipos de arrays y de ensayos de bajo rendimiento tipo TaqMan™, han de plantearse con máximo cuidado, ya que su calidad y concordancia en muchos casos son ambiguas.
3. La asociación de los **SNPs y VNTRs candidatos** seleccionados no se ha validado en la estrategia caso/control de marcador único en nuestra población, pero sí al considerar la epistasia entre pares de SNPs, lo que apoya la causa multigénica de la dislexia y la falta de atención.
4. En esta tesis se ha demostrado que la **selección de criterios de inclusión/exclusión** de la población caso y de la control para un estudio de asociación genética es clave para que los hallazgos sean relevantes y reproducibles.
5. En nuestra población general de niños de origen español, genes candidatos para la dislexia se asocian a variables atencionales, así como genes candidatos para TDAH se asocian a variables relacionadas con la lectura, por lo que nuestro estudio sugiere que podría existir una vinculación entre ambos procesos con **genes o síntomas causales compartidos**.
6. En nuestros GWAS de dislexia, falta de atención y comorbidad, la mayoría de los SNPs con alta significancia de asociación se ubican en **zonas reguladoras génicas y en genes de interés causal** para estos trastornos. Por ello, aunque no hayamos obtenido unos valores de asociación muy significativos, los resultados de este proyecto muestran no ser fruto del azar.

7. Nuestro análisis de enriquecimiento de rutas ha mostrado que los SNPs que han despuntado en cualquiera de las comparativas realizadas pertenecen a genes que codifican proteínas de cascadas funcionales relacionadas con la **transmisión sináptica y la adhesión o estructuración celular**. Gracias a esta metodología también nos es posible proponer ciertos componentes del sistema inmunitario como participantes en el desarrollo de los trastornos estudiados.
  
8. El SNP rs10120392, situado en el gen **KANK1**, ha resultado asociado con nuestra población de dislexia, por lo que consideramos que este polimorfismo y el gen **KANK1** son **candidatos de susceptibilidad** a padecer dicho trastorno.
  
9. El abordaje de asociación de SNPs, o grupos de estos, con **medidas cuantitativas aplicadas a la definición de los fenotipos considerados, denominadas endofenotipos**, ha resultado en hallazgos noveles muy importantes, como por ejemplo la asociación de una región del cromosoma 11 con el procesamiento ortográfico, y de dos regiones situadas en los cromosomas 2 y 8 con el procesamiento automático.

## **7 REFERENCIAS**



- 1 Bishop D V. Genes, cognition, and communication: insights from neurodevelopmental disorders. *Ann N Y Acad Sci* 2009; **1156**: 1–18.
- 2 Bishop DVM. Developmental cognitive genetics: how psychology can inform genetics and vice versa. *Q J Exp Psychol (Hove)* 2006; **59**: 1153–1168.
- 3 Bishop D. The interface between genetic and psychology: lessons from developmental dyslexia. *Proc R Soc London B* 2015; **282**.
- 4 Ramus F. Genes , brain , and cognition : A roadmap for the cognitive scientist. *Cognition* 2006; **101**: 247–269.
- 5 Flint J. The genetic basis of cognition. *Brain*. 1999; **122**: 2015–2031.
- 6 Falconer DS. The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Ann Hum Genet* 1965; **29**: 51–76.
- 7 Fisher SE, DeFries JC. Developmental dyslexia: genetic dissection of a complex cognitive trait. *Nat Rev Neurosci* 2002; **3**: 767–780.
- 8 Thapar A, O'Donovan M, Owen MJ. The genetics of attention deficit hyperactivity disorder. *Hum Mol Genet* 2005; **14 Spec No**: R275-82.
- 9 Eden GF, Moats L. The role of neuroscience in the remediation of students with dyslexia. *Nat Neurosci* 2002; **5 Suppl**: 1080–1084.
- 10 Paracchini S, Scerri T, Monaco AP. The genetic lexicon of dyslexia. *Annu Rev Genomics Hum Genet* 2007; **8**: 57–79.
- 11 Démonet J, Taylor MJ, Chaix Y. Developmental dyslexia. *Lancet* 2004; **363**: 1451–1460.
- 12 Liederman J, Kantrowitz L, Flannery K. Male vulnerability to reading disability is not likely to be a myth: a call for new data. *J Learn Disabil* 2005; **38**: 109–129.
- 13 Rutter M, Caspi A, Fergusson D, Horwood LJ, Goodman R, Maughan B *et al*. Sex Differences in Developmental Reading Disability. *JAMA* 2004; **291**: 2007.
- 14 Schumacher J, Anthoni H, Dahdouh F, König IR, Hillmer AM, Kluck N *et al*. Strong genetic evidence of DCDC2 as a susceptibility gene for dyslexia. *Am J Hum Genet* 2006; **78**: 52–62.
- 15 Hannula-Jouppi K, Kaminen-Ahola N, Taipale M, Eklund R, Nopola-Hemmi J, Kaariainen H *et al*. The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia. *PLoS Genet* 2005; **1**: e50.
- 16 DeFries J, Alarcón M. Genetics of specific reading disability. *Ment Retard Dev Disabil Res Rev* 1996; **2**: 39–47.
- 17 Schumacher J, Hoffmann P, Schmal C, Schulte-Körne G, Nothen MM. Genetics of dyslexia: the evolving landscape. *J Med Genet* 2007; **44**: 289–297.
- 18 Carrion-Castillo A, Franke B, Fisher SE. Molecular genetics of dyslexia: An overview. *Dyslexia*. 2013; **19**: 214–240.
- 19 Williams J, O'Donovan MC. The genetics of developmental dyslexia. *Eur J Hum Genet* 2006; **14**: 681–689.
- 20 Poelmans G, Buitelaar JK, Pauls DL, Franke B. A theoretical molecular network for dyslexia : integrating

- available genetic findings. *Mol Psychiatry* 2011; **16**: 365–382.
- 21 Galaburda AM, Loturco J, Ramus F, Fitch RH, Rosen GD. From genes to behavior in developmental dyslexia. *Nat Neurosci* 2006; **9**: 1213–1217.
- 22 Galaburda AM, Sherman GF, Rosen GD, Aboitiz F, Geschwind N. Developmental dyslexia: Four consecutive patients with cortical anomalies. *Ann Neurol* 1985; **18**: 222–233.
- 23 Marín O. Migración neuronal y enfermedades del desarrollo del sistema nervioso. *SEBBM* 2009; **Noviembre**.
- 24 Ghashghaei HT, Lai C, Anton ES. Neuronal migration in the adult brain: are we there yet? *Nat Rev Neurosci* 2007; **8**: 141–51.
- 25 Tsai LH, Gleeson JG. Nucleokinesis in neuronal migration. *Neuron*. 2005; **46**: 383–388.
- 26 Ayala R, Shu T, Tsai LH. Trekking across the Brain: The Journey of Neuronal Migration. *Cell*. 2007; **128**: 29–43.
- 27 Smith SD, Kimberling WJ, Pennington BF, Lubs H a. Specific reading disability: identification of an inherited form through linkage analysis. *Science (80- )* 1983; **219**: 1345–1347.
- 28 Taipale M, Kaminen N, Nopola-Hemmi J, Haltia T, Myllyluoma B, Lyytinen H et al. A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain. *Proc Natl Acad Sci U S A* 2003; **100**: 11553–11558.
- 29 Nopola-Hemmi J, Taipale M, Haltia T, Lehesjoki AE, Voutilainen A, Kere J. Two translocations of chromosome 15q associated with dyslexia. *J Med Genet* 2000; **37**: 771–775.
- 30 Scerri TS, Fisher SE, Francks C, MacPhie IL, Paracchini S, Richardson AJ et al. Putative functional alleles of DYX1C1 are not associated with dyslexia susceptibility in a large sample of sibling pairs from the UK. *J Med Genet* 2004; **41**: 853–857.
- 31 Wigg KG, Couto JM, Feng Y, Anderson B, Cate-Carter TD, Macciardi F et al. Support for EKN1 as the susceptibility locus for dyslexia on 15q21. 2004.
- 32 Bates TC, Lind PA, Luciano M, Montgomery GW, Martin NG, Wright MJ. Dyslexia and DYX1C1: deficits in reading and spelling associated with a missense mutation. *Mol Psychiatry* 2010; **15**: 1190–1196.
- 33 Becker J, Czamara D, Scerri TS, Ramus F, Csépe V, Talcott JB et al. Genetic analysis of dyslexia candidate genes in the European cross-linguistic NeuroDys cohort. *Eur J Hum Genet* 2014; **22**: 675–80.
- 34 Brkanac Z, Chapman NH, Matsushita MM, Chun L, Nielsen K, Cochrane E et al. Evaluation of candidate genes for DYX1 and DYX2 in families with dyslexia. *Am J Med Genet Part B Neuropsychiatr Genet* 2007; **144**: 556–560.
- 35 Marino C, Citterio A, Giorda R, Facoetti A, Menozzi G, Vanzin L et al. Association of short-term memory with a variant within DYX1C1 in developmental dyslexia. *Genes, Brain Behav* 2007; **6**: 640–646.
- 36 Dahdouh F, Anthoni H, Tapia-Páez I, Peyrard-Janvid M, Schulte-Körne G, Warnke A et al. Further evidence for DYX1C1 as a susceptibility factor for dyslexia. *Psychiatr Genet* 2009; **19**: 59–63.
- 37 Newbury DF, Paracchini S, Scerri TS, Winchester L, Addis L, Richardson AJ et al. Investigation of Dyslexia and SLI Risk Variants in Reading- and Language-Impaired Subjects. *Behav Genet* 2011; **41**: 90–104.
- 38 Cope NA, Hill G, van den Bree M HD, Moskvina V, Green EK, Owen MJ W, O'Donovan J M. No support for

- association between dyslexia susceptibility 1 candidate 1 and developmental dyslexia. *Mol Psychiatry* 2005; **10**: 237–238.
- 39 Meng H, Hager K, Held M, Page GP, Olson RK, Pennington BF et al. TDT-association analysis of EKN1 and dyslexia in a Colorado twin cohort. *Hum Genet* 2005; **118**: 87–90.
- 40 Bellini G, Bravaccio C, Calamoneri F, Donatella Cocuzza M, Fiorillo P, Gagliano A et al. No evidence for association between dyslexia and DYX1C1 functional variants in a group of children and adolescents from Southern Italy. *J Mol Neurosci* 2005; **27**: 311–314.
- 41 Saviour P, Kumar S, Kiran U, Ravuri RR, Rao VR, Ramachandra NB. Allelic variants of DYX1C1 are not associated with dyslexia in India. *Indian J Hum Genet* 2008; **14**: 99–102.
- 42 Lim CK, Ho CS, Chou CH, Waye MM. Association of the rs3743205 variant of DYX1C1 with dyslexia in Chinese children. *Behav Brain Funct* 2011; **7**: 16.
- 43 Tran C, Gagnon F, Wigg KG, Feng Y, Gomez L, Cate-Carter TD et al. A family-based association analysis and meta-analysis of the reading disabilities candidate gene DYX1C1. *Am J Med Genet Part B Neuropsychiatr Genet* 2013; **162**: 146–156.
- 44 Venkatesh SK, Siddaiah A, Padakannaya P, Ramachandra NB. Association of SNPs of DYX1C1 with developmental dyslexia in an Indian population. *Psychiatr Genet* 2014; **24**: 10–20.
- 45 Zou L, Chen W, Shao S, Sun Z, Zhong R, Shi J et al. Genetic variant in KIAA0319, but not in DYX1C1, is associated with risk of dyslexia: An integrated meta-analysis. *Am J Med Genet Part B Neuropsychiatr Genet* 2012; **159 B**: 970–976.
- 46 Tapia-Páez I, Tammimies K, Massinen S, Roy AL, Kere J. The complex of TFII-I, PARP1, and SFPQ proteins regulates the DYX1C1 gene implicated in neuronal migration and dyslexia. *FASEB J* 2008; **22**: 3001–3009.
- 47 Massinen S, Tammimies K, Tapia-Páez I, Matsson H, Hokkanen ME, Söderberg O et al. Functional interaction of DYX1C1 with estrogen receptors suggests involvement of hormonal pathways in dyslexia. *Hum Mol Genet* 2009; **18**: 2802–2812.
- 48 Tammimies K, Tapia-Páez I, Rüegg J, Rosin G, Kere J, Gustafsson J-Å et al. The rs3743205 SNP is important for the regulation of the dyslexia candidate gene DYX1C1 by estrogen receptor  $\beta$  and DNA methylation. *Mol Endocrinol* 2012; **26**: 619–29.
- 49 Wang Y, Paramasivam M, Thomas A, Bai J, Kaminen-Ahola N, Kere J et al. DYX1C1 functions in neuronal migration in developing neocortex. *Neuroscience* 2006; **143**: 515–522.
- 50 Currier TA, Etchegaray MA, Haight JL, Galaburda AM, Rosen GD. The effects of embryonic knockdown of the candidate dyslexia susceptibility gene homologue *Dyx1c1* on the distribution of GABAergic neurons in the cerebral cortex. *Neuroscience* 2011; **172**: 535–546.
- 51 Rosen GD, Bai J, Wang Y, Fiondella CG, Threlkeld SW, Loturco JJ et al. Disruption of neuronal migration by RNAi of *Dyx1c1* results in neocortical and hippocampal malformations. *Cereb Cortex* 2007; **17**: 2562–2572.
- 52 Hoh RA, Stowe TR, Turk E, Stearns T. Transcriptional Program of Ciliated Epithelial Cells Reveals New Cilium and Centrosome Components and Links to Human Disease. *PLoS One* 2012; **7**. doi:10.1371/journal.pone.0052166.

- 53 Ivliev AE, 't Hoen PAC, van Roon-Mom WMC, Peters DJM, Sergeeva MG. Exploring the Transcriptome of Ciliated Cells Using In Silico Dissection of Human Tissues. *PLoS One* 2012; **7**: e35618.
- 54 Chandrasekar G, Vesterlund L, Hultenby K, Tapia-Páez I, Kere J. The Zebrafish Orthologue of the Dyslexia Candidate Gene DYX1C1 Is Essential for Cilia Growth and Function. *PLoS One* 2013; **8**. doi:10.1371/journal.pone.0063123.
- 55 Tarkar A, Loges NT, Slagle CE, Francis R, Dougherty GW, Tamayo J V et al. DYX1C1 is required for axonemal dynein assembly and ciliary motility. *Nat Genet* 2013; **45**: 995–1003.
- 56 Tammimies K, Vitézic M, Matsson H, Le Guyader S, Bürglin TR, Ohman T et al. Molecular networks of DYX1C1 gene show connection to neuronal migration genes and cytoskeletal proteins. *Biol Psychiatry* 2013; **73**: 583–590.
- 57 Hatakeyama S, Matsumoto M, Yada M, Nakayama KI. Interaction of U-box-type ubiquitin-protein ligases (E3s) with molecular chaperones. *Genes to Cells* 2004; **9**: 533–548.
- 58 Kim YJ, Huh JW, Kim DS, Bae MI, Lee JR, Ha HS et al. Molecular characterization of the DYX1C1 gene and its application as a cancer biomarker. *J Cancer Res Clin Oncol* 2009; **135**: 265–270.
- 59 Rosin G, Hannelius U, Lindström L, Hall P, Bergh J, Hartman J et al. The dyslexia candidate gene DYX1C1 is a potential marker of poor survival in breast cancer. *BMC Cancer* 2012; **12**: 79.
- 60 Chen Y, Zhao M, Wang S, Chen J, Wang Y, Cao Q et al. A novel role for DYX1C1, a chaperone protein for both Hsp70 and Hsp90, in breast cancer. *J Cancer Res Clin Oncol* 2009; **135**: 1265–1276.
- 61 Cardon LR, Smith SD, Fulker DW, Kimberling WJ, Pennington BF, DeFries JC. Quantitative trait locus for reading disability on chromosome 6. *Science (80- )* 1994; **266**: 276–279.
- 62 Kaplan DE, Gayan J, Ahn J, Won TW, Pauls D, Olson RK et al. Evidence for linkage and association with reading disability on 6p21.3-22. *Am J Hum Genet* 2002; **70**: 1287–1298.
- 63 Fisher SE, Marlow a J, Lamb J, Maestrini E, Williams DF, Richardson a J et al. A quantitative-trait locus on chromosome 6p influences different aspects of developmental dyslexia. *Am J Hum Genet* 1999; **64**: 146–156.
- 64 Grigorenko EL, Wood FB, Meyer MS, Hart L a, Speed WC, Shuster a et al. Susceptibility loci for distinct components of developmental dyslexia on chromosomes 6 and 15. *Am J Hum Genet* 1997; **60**: 27–39.
- 65 Deffenbacher KE, Kenyon JB, Hoover DM, Olson RK, Pennington BF, DeFries JC et al. Refinement of the 6p21.3 quantitative trait locus influencing dyslexia: linkage and association analyses. *Hum Genet* 2004; **115**: 128–138.
- 66 Meng H, Smith SD, Hager K, Held M, Liu J, Olson RK et al. DCDC2 is associated with reading disability and modulates neuronal development in the brain. *Proc Natl Acad Sci U S A* 2005; **102**: 17053–17058.
- 67 Zhong R, Yang B, Tang H, Zou L, Song R, Zhu LQ et al. Meta-analysis of the association between DCDC2 polymorphisms and risk of dyslexia. *Mol. Neurobiol.* 2013; **47**: 435–442.
- 68 Francks C, Paracchini S, Smith SD, Richardson AJ, Scerri TS, Cardon LR et al. A 77-kilobase region of chromosome 6p22.2 is associated with dyslexia in families from the United Kingdom and from the United States. *Am J Hum Genet* 2004; **75**: 1046–1058.
- 69 Cope N, Harold D, Hill G, Moskvina V, Stevenson J, Holmans P et al. Strong evidence that KIAA0319 on

- chromosome 6p is a susceptibility gene for developmental dyslexia. *Am J Hum Genet* 2005; **76**: 581–591.
- 70 Harold D, Paracchini S, Scerri T, Dennis M, Cope N, Hill G *et al.* Further evidence that the KIAA0319 gene confers susceptibility to developmental dyslexia. *Mol Psychiatry* 2006; **11**: 1085–1091.
- 71 Ludwig KU, Roeske D, Schumacher J, Schulte-Körne G, König IR, Warnke A *et al.* Investigation of interaction between DCDC2 and KIAA0319 in a large German dyslexia sample. *J Neural Transm* 2008; **115**: 1587–1589.
- 72 Powers NR, Eicher JD, Butter F, Kong Y, Miller LL, Ring SM *et al.* Alleles of a polymorphic ETV6 binding site in DCDC2 confer risk of reading and language impairment. *Am J Hum Genet* 2013; **93**: 19–28.
- 73 Matsson H, Huss M, Persson H, Einarsdottir E, Tiraboschi E, Nopola-Hemmi J *et al.* Polymorphisms in DCDC2 and S100B associate with developmental dyslexia. *J Hum Genet* 2015; **60**: 399–401.
- 74 Eicher JD, Powers NR, Miller LL, Mueller KL, Mascheretti S, Marino C *et al.* Characterization of the DYX2 locus on chromosome 6p22 with reading disability, language impairment, and IQ. *Hum Genet* 2014; **133**: 869–881.
- 75 Adler WT, Platt MP, Mehlhorn AJ, Haight JL, Currier TA, Etchegaray MA *et al.* Position of Neocortical Neurons Transfected at Different Gestational Ages with shRNA Targeted against Candidate Dyslexia Susceptibility Genes. *PLoS One* 2013; **8**. doi:10.1371/journal.pone.0065179.
- 76 Burbridge TJ, Wang Y, Volz AJ, Peschansky VJ, Lisann L, Galaburda AM *et al.* Postnatal analysis of the effect of embryonic knockdown and overexpression of candidate dyslexia susceptibility gene homolog Dcdc2 in the rat. *Neuroscience* 2008; **152**: 723–733.
- 77 Gabel LA, Marin I, Loturco JJ, Che A, Murphy C, Manghani M *et al.* Mutation of the dyslexia-associated gene Dcdc2 impairs LTM and visuo-spatial performance in mice. *Genes, Brain Behav* 2011; **10**: 868–875.
- 78 Massinen S, Hokkanen ME, Matsson H, Tammimies K, Tapia-Páez I, Dahlström-Heuser V *et al.* Increased expression of the dyslexia candidate gene DCDC2 affects length and signaling of primary cilia in neurons. *PLoS One* 2011; **6**. doi:10.1371/journal.pone.0020580.
- 79 Paracchini S, Thomas A, Castro S, Lai C, Paramasivam M, Wang Y *et al.* The chromosome 6p22 haplotype associated with dyslexia reduces the expression of KIAA0319, a novel gene involved in neuronal migration. *Hum Mol Genet* 2006; **15**: 1659–1666.
- 80 Dennis MY, Paracchini S, Scerri TS, Prokunina-Olsson L, Knight JC, Wade-Martins R *et al.* A common variant associated with dyslexia reduces expression of the KIAA0319 gene. *PLoS Genet* 2009; **5**: e1000436.
- 81 Peschansky VJ, Burbridge TJ, Volz AJ, Fiondella C, Wissner-Gross Z, Galaburda AM *et al.* The effect of variation in expression of the candidate dyslexia susceptibility gene homolog Kiaa0319 on neuronal migration and dendritic morphology in the rat. *Cereb Cortex* 2010; **20**: 884–897.
- 82 Velayos-Baeza A, Toma C, da Roza S, Paracchini S, Monaco AP. Alternative splicing in the dyslexia-associated gene KIAA0319. *Mamm Genome* 2007; **18**: 627–634.
- 83 Velayos-Baeza A, Toma C, Paracchini S, Monaco AP. The dyslexia-associated gene KIAA0319 encodes highly N- and O-glycosylated plasma membrane and secreted isoforms. *Hum Mol Genet* 2008; **17**: 859–871.
- 84 Gabel LA, Gibson CJ, Gruen JR, LoTurco JJ. Progress towards a cellular neurobiology of reading disability. *Neurobiol. Dis.* 2010; **38**: 173–180.

- 85 Velayos-Baeza A, Levecque C, Kobayashi K, Holloway ZG, Monaco AP. The dyslexia-associated KIAA0319 protein undergoes proteolytic processing with  $\gamma$ -secretase-independent intramembrane cleavage. *J Biol Chem* 2010; **285**: 40148–40162.
- 86 Levecque C, Velayos-Baeza A, Holloway ZG, Monaco AP. The dyslexia-associated protein KIAA0319 interacts with adaptor protein 2 and follows the classical clathrin-mediated endocytosis pathway. *Am J Physiol Cell Physiol* 2009; **297**: C160–C168.
- 87 Fagerheim T, Raeymaekers P, Tonnessen FE, Pedersen M, Tranebjaerg L, Lubs HA. A new gene (DYX3) for dyslexia is located on chromosome 2. *J Med Genet* 1999; **36**: 664–669.
- 88 Francks C, Fisher SE, Olson RK, Pennington BF, Smith SD, DeFries JC et al. Fine mapping of the chromosome 2p12-16 dyslexia susceptibility locus: quantitative association analysis and positional candidate genes SEMA4F and OTX1. *Psychiatr Genet* 2002; **12**: 35–41.
- 89 Kaminen N, Hannula-Jouppi K, Kestila M, Lahermo P, Muller K, Kaaranen M et al. A genome scan for developmental dyslexia confirms linkage to chromosome 2p11 and suggests a new locus on 7q32. *J Med Genet* 2003; **40**: 340–345.
- 90 Bates TC, Luciano M, Castles A, Coltheart M, Wright MJ, Martin NG. Replication of reported linkages for dyslexia and spelling and suggestive evidence for novel regions on chromosomes 4 and 17. *Eur J Hum Genet* 2007; **15**: 194–203.
- 91 Anthoni H, Zucchelli M, Matsson H, Muller-Myhsok B, Fransson I, Schumacher J et al. A locus on 2p12 containing the co-regulated MRPL19 and C2ORF3 genes is associated to dyslexia. *Hum Mol Genet* 2007; **16**: 667–677.
- 92 Nopola-Hemmi J, Myllyluoma B, Haltia T, Taipale M, Ollikainen V, Ahonen T et al. A dominant gene for developmental dyslexia on chromosome 3. *J Med Genet* 2001; **38**: 658–664.
- 93 Bates TC, Luciano M, Medland SE, Montgomery GW, Wright MJ, Martin NG. Genetic variance in a component of the language acquisition device: ROBO1 polymorphisms associated with phonological buffer deficits. *Behav Genet* 2011; **41**: 50–57.
- 94 Fisher SE, Francks C, Marlow AJ, MacPhie IL, Newbury DF, Cardon LR et al. Independent genome-wide scans identify a chromosome 18 quantitative-trait locus influencing dyslexia. *Nat Genet* 2002; **30**: 86–91.
- 95 Rajagopalan S, Nicolas E, Vivancos V, Berger J, Dickson BJ. Crossing the midline: Roles and regulation of Robo receptors. *Neuron* 2000; **28**: 767–777.
- 96 Andrews W, Liapi A, Plachez C, Camurri L, Zhang J, Mori S et al. Robo1 regulates the development of major axon tracts and interneuron migration in the forebrain. *Development* 2006; **133**: 2243–2252.
- 97 Petryshen TL, Kaplan BJ, Hughes ML, Tzenova J, Field LL. Supportive evidence for the DYX3 dyslexia susceptibility gene in Canadian families. *J Med Genet* 2002; **39**: 125–126.
- 98 Seshadri S, DeStefano AL, Au R, Massaro JM, Beiser AS, Kelly-Hayes M et al. Genetic correlates of brain aging on MRI and cognitive test measures: a genome-wide association and linkage analysis in the Framingham Study. *BMC Med Genet* 2007; **8 Suppl 1**: S15.
- 99 Scerri TS, Paracchini S, Morris A, MacPhie IL, Talcott J, Stein J et al. Identification of candidate genes for dyslexia susceptibility on chromosome 18. *PLoS One* 2010; **5**: e13712.

- 100 Tzenova J, Kaplan BJ, Petryshen TL, Field LL. Confirmation of a Dyslexia Susceptibility Locus on Chromosome 1p34-p36 in a set of 100 Canadian Families. *Am J Med Genet B Neuropsychiatr Genet* 2004; **124**: 117–124.
- 101 Grigorenko EL, Wood FB, Meyer MS, Pauls JED, Hart LA, Pauls DL. Linkage Studies Suggest a Possible Locus for Developmental Dyslexia on Chromosome 1p. *Am J Hum Genet* 2001; **129**: 120–129.
- 102 Rabin M, Wen X, Hepburn M, Lubs HA, Feldman E, Duara R. Suggestive linkage of developmental dyslexia to chromosome 1p34-p36. *Lancet* 1993; **342**: 178.
- 103 Couto JM, Other. The KIAA0319-like (KIAA0319L) gene on chromosome 1p34 as a candidate for reading disabilities. *J Neurogenet* 2008; **22**: 295–313.
- 104 de Kovel CG, Hol FA, Heister JG, Willems JJ, Sandkuijl LA, Franke B et al. Genomewide scan identifies susceptibility locus for dyslexia on Xq27 in an extended Dutch family. *J Med Genet* 2004; **41**: 652–657.
- 105 Huc-Chabrolle M, Charon C, Guilmatre A, Vourc'H P, Tripi G, Barthez MA et al. Xq27 FRAXA locus is a strong candidate for dyslexia: Evidence from a genome-wide scan in French families. *Behav Genet* 2013; **43**: 132–140.
- 106 Hakim Arif. FOXP2: A gene for language and speech. *Dhaka Univ J Linguist* 2009; **2**: 173–184.
- 107 Vernes SC, Newbury DF, Abrahams BS, Winchester L, Nicod J, Groszer M et al. A Functional Genetic Link between Distinct Developmental Language Disorders. *N Engl J Med* 2008; **359**: 2337–2345.
- 108 Wilcke A, Jana Burkhardt CL, Alexander M, Wolf C, Quente E, Ahnert P et al. Imaging genetics of FOXP2 in dyslexia. *Eur J Hum Genet* 2012; **20**: 714–714.
- 109 Peter B, Raskind WH, Matsushita M, Lisowski M, Vu T, Berninger VW et al. Replication of CNTNAP2 association with nonword repetition and support for FOXP2 association with timed reading and motor activities in a dyslexia family sample. *J Neurodev Disord* 2011; **3**: 39–49.
- 110 Fisher SE, Vargha-Khadem F, Watkins KE, Monaco AP, Pembrey ME. Localisation of a gene implicated in a severe speech and language disorder. *Nat Genet* 1998; **18**: 168–170.
- 111 Lai CS, Fisher SE, Hurst J a, Levy ER, Hodgson S, Fox M et al. The SPCH1 region on human 7q31: genomic characterization of the critical interval and localization of translocations associated with speech and language disorder. *Am J Hum Genet* 2000; **67**: 357–68.
- 112 Lai CSL, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 2001; **413**: 519–523.
- 113 Kang C, Drayna D. Genetics of Speech and Language Disorders 1. *Annu Rev Genomics Hum Genet* 2011; **12**: 145–164.
- 114 Shu W, Cho JY, Jiang Y, Zhang M, Weisz D, Elder G a et al. Altered ultrasonic vocalization in mice with a disruption in the *Foxp2* gene. *Proc Natl Acad Sci U S A* 2005; **102**: 9643–9648.
- 115 Wang E, Ding YC, Flodman P, Kidd JR, Kidd KK, Grady DL et al. The genetic architecture of selection at the human dopamine receptor D4 (DRD4) gene locus. *Am J Hum Genet* 2004; **74**: 931–944.
- 116 Hawi Z, Cummins TDR, Tong J, Johnson B, Lau R, Samarrai W et al. The molecular genetic architecture of attention deficit hyperactivity disorder. *Mol Psychiatry* 2015; **20**: 289–97.
- 117 Girirajan S, Brkanac Z, Coe BP, Baker C, Vives L, Vu TH et al. Relative Burden of Large CNVs on a Range of

- Neurodevelopmental Phenotypes. *PLoS Genet* 2011; **7**. doi:10.1371/journal.pgen.1002334.
- 118 Stefansson H, Meyer-Lindenberg A, Steinberg S, Magnusdottir B, Morgen K, Arnarsdottir S *et al*. CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature* 2014; **505**: 361–366.
- 119 Veerappa AM, Padakannaya P, Ramachandra NB. Copy number variation-based polymorphism in a new pseudoautosomal region 3 (PAR3) of a human X-chromosome-transposed region (XTR) in the y chromosome. *Funct Integr Genomics* 2013; **13**: 285–293.
- 120 Veerappa AM, Saldanha M, Padakannaya P, Ramachandra NB. Genome-wide copy number scan identifies disruption of PCDH11X in developmental dyslexia. *Am J Med Genet Part B Neuropsychiatr Genet* 2013; **162**: 889–897.
- 121 Veerappa AM, Saldanha M, Padakannaya P, Ramachandra NB. Family-based genome-wide copy number scan identifies five new genes of dyslexia involved in dendritic spinal plasticity. *J Hum Genet* 2013; **58**: 539–47.
- 122 Veerappa AM, Saldanha M, Padakannaya P, Ramachandra NB. Family based genome-wide copy number scan identifies complex rearrangements at 17q21.31 in dyslexics. *Am J Med Genet Part B Neuropsychiatr Genet* 2014; **165**: 572–580.
- 123 Scerri TS, Darki F, Newbury DF, Whitehouse AJO, Peyrard-Janvid M, Matsson H *et al*. The Dyslexia Candidate Locus on 2p12 Is Associated with General Cognitive Ability and White Matter Structure. *PLoS One* 2012; **7**. doi:10.1371/journal.pone.0050321.
- 124 Meda SA, Gelernter J, Gruen JR, Calhoun VD, Meng H, Cope NA *et al*. Polymorphism of DCDC2 reveals differences in cortical morphology of healthy individuals-A preliminary Voxel based morphometry study. *Brain Imaging Behav* 2008; **2**: 21–26.
- 125 Cope N, Eicher JD, Meng H, Gibson CJ, Hager K, Lacadie C *et al*. Variants in the DYX2 locus are associated with altered brain activation in reading-related brain regions in subjects with reading disability. *Neuroimage* 2012; **63**: 148–156.
- 126 Szalkowski CE, Fiondella CF, Truong DT, Rosen GD, LoTurco JJ, Fitch RH. The effects of Kiaa0319 knockdown on cortical and subcortical anatomy in male rats. *Int J Dev Neurosci* 2013; **31**: 116–122.
- 127 Richlan F, Kronbichler M, Wimmer H. Meta-analyzing brain dysfunctions in dyslexic children and adults. *Neuroimage* 2011; **56**: 1735–1742.
- 128 Martin A, Kronbichler M, Richlan F. Dyslexic brain activation abnormalities in deep and shallow orthographies: A meta-analysis of 28 functional neuroimaging studies. *Hum Brain Mapp* 2016; **37**: 2676–2699.
- 129 Martin A, Schurz M, Kronbichler M, Richlan F. Reading in the brain of children and adults: A meta-analysis of 40 functional magnetic resonance imaging studies. *Hum Brain Mapp* 2015; **36**: 1963–1981.
- 130 Španiel F, Horáček J, Tintěra J, Ibrahim I, Novák T, Čermák J *et al*. Genetic variation in FOXP2 alters grey matter concentrations in schizophrenia patients. *Neurosci Lett* 2011; **493**: 131–135.
- 131 Darki F, Peyrard-Janvid M, Matsson H, Kere J, Klingberg T. Three dyslexia susceptibility genes, DYX1C1, DCDC2, and KIAA0319, affect temporo-parietal white matter structure. *Biol Psychiatry* 2012; **72**: 671–676.
- 132 Darki F, Peyrard-Janvid M, Matsson H, Kere J, Klingberg T. DCDC2 Polymorphism Is Associated with Left

- Temporoparietal Gray and White Matter Structures during Development. *J Neurosci* 2014; **34**: 14455–62.
- 133 Hibar D, Stein J, Renteria M, Arias-Vasquez A, Desrivieres S, Jahanshad N et al. Common genetic variants influence human subcortical brain structures. *Nature* 2015.
- 134 Lamminmaki S, Massinen S, Nopola-Hemmi J, Kere J, Hari R. Human ROBO1 Regulates Interaural Interaction in Auditory Pathways. *J Neurosci* 2012; **32**: 966–971.
- 135 Pinel P, Fauchereau F, Moreno A, Barbot A, Lathrop M, Zelenika D et al. Genetic Variants of FOXP2 and KIAA0319 / TTRAP / THEM2 Locus Are Associated with Altered Brain Activation in Distinct Language-Related Regions. *J Neurosci* 2012; **32**: 817–825.
- 136 Stergiakouli E, Thapar A. Fitting the pieces together: current research on the genetic basis of attention-deficit/hyperactivity disorder (ADHD). *Neuropsychiatr Dis Treat* 2010; **6**: 551–60.
- 137 Rosenberg J, Pennington BF, Willcutt EG, Olson RK. Gene by environment interactions influencing reading disability and the inattentive symptom dimension of attention deficit/hyperactivity disorder. *J Child Psychol Psychiatry Allied Discip* 2012; **53**: 243–251.
- 138 Willcutt EG, Nigg JT, Pennington BF, Solanto M V, Rohde L a, Tannock R et al. Validity of DSM-IV attention deficit/hyperactivity disorder symptom dimensions and subtypes. *J Abnorm Psychol* 2012; **121**: 991–1010.
- 139 Fair D a, Nigg JT, Iyer S, Bathula D, Mills KL, Dosenbach NUF et al. Distinct neural signatures detected for ADHD subtypes after controlling for micro-movements in resting state functional connectivity MRI data. *Front Syst Neurosci* 2012; **6**: 80.
- 140 Lahey BB, Pelham WE, Loney J, Lee SS, Willcutt E. Instability of the DSM-IV Subtypes of ADHD from preschool through elementary school. *Arch Gen Psychiatry* 2005; **62**: 896–902.
- 141 Bidwell LC, Willcutt EG, McQueen MB, Defries JC, Olson RK, Smith SD et al. A family based association study of DRD4, DAT1, and 5HTT and continuous traits of attention-deficit hyperactivity disorder. *Behav Genet* 2011; **41**: 165–174.
- 142 Milich R, Balentine AC, Lynam DR. ADHD combined type and ADHD predominantly inattentive type are distinct and unrelated disorders. *Clin Psychol Sci Pract* 2001; **8**: 463–488.
- 143 Cuffe SP, Moore CG, McKeown RE. Prevalence and correlates of ADHD symptoms in the national health interview survey. *J Atten Disord* 2005; **9**: 392–401.
- 144 Matthews M, Nigg JT, Fair DA. Attention deficit hyperactivity disorder. *Curr Top Behav Neurosci* 2014; **16**: 235–66.
- 145 Barkley RA. Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. *Psychol Bull* 1997; **121**: 65–94.
- 146 Sagvolden T, Johansen EB, Aase H, Russell VA. A dynamic developmental theory of attention-deficit/hyperactivity disorder (ADHD) predominantly hyperactive/impulsive and combined subtypes. *Behav Brain Sci* 2005; **28**: 397–419.
- 147 Barkley R a. Advances in the diagnosis and subtyping of attention deficit hyperactivity disorder: what may lie ahead for DSM-V. *Rev Neurol* 2009; **48 Suppl 2**: S101–S106.
- 148 Thapar a, Langley K, O'donovan M, Owen M. Refining the attention deficit hyperactivity disorder phenotype for molecular genetic studies. *Mol Psychiatry* 2006; **11**: 714–720.

- 149 McLoughlin G, Ronald A, Kuntsi J, Asherson P, Plomin R. Genetic support for the dual nature of attention deficit hyperactivity disorder: Substantial genetic overlap between the inattentive and hyperactive-impulsive components. *J Abnorm Child Psychol* 2007; **35**: 999–1008.
- 150 DiMaio S, Grizenko N, Joober R. Dopamine genes and attention-deficit hyperactivity disorder: a review. *J Psychiatry Neurosci* 2003; **28**: 27–38.
- 151 Larsson H, Lichtenstein P, Larsson JO. Genetic contributions to the development of ADHD subtypes from childhood to adolescence. *J Am Acad Child Adolesc Psychiatry* 2006; **45**: 973–981.
- 152 Wender P. Minimal Brain Dysfunction in Children. *John Wiley Sons Inc New York* 1971.
- 153 Levy F. The dopamine theory of attention deficit hyperactivity disorder (ADHD). *Aust N Z J Psychiatry* 1991; **25**: 277–83.
- 154 Swanson J, Posner M, Fusella J, Wasdell M, Sommer T, Fan J. Genes and attention deficit hyperactivity disorder. *Curr Psychiatry Rep* 2001; **3**: 92–100.
- 155 DiMaio S, Grizenko N, Joober R. Dopamine genes and attention-deficit hyperactivity disorder: a review. *J Psychiatry Neurosci* 2003; **28**: 27–38.
- 156 LaHoste GJ, Swanson JM, Wigal SB, Glabe C, Wigal T, King N *et al*. Dopamine D4 receptor gene polymorphism is associated with attention deficit hyperactivity disorder. *Mol Psychiatry* 1996; **1**: 121–124.
- 157 Asghari V, Sanyal S, Buchwaldt S, Paterson A, Jovanovic V, Van Tol HH. Modulation of intracellular cyclic AMP levels by different human dopamine D4 receptor variants. *J Neurochem* 1995; **65**: 1157–65.
- 158 Bellgrove MA, Mattingley JB. Molecular genetics of attention. *Ann N Y Acad Sci* 2008; **1129**: 200–212.
- 159 Swanson J, Oosterlaan J, Murias M, Schuck S, Flodman P, Spence MA *et al*. Attention deficit/hyperactivity disorder children with a 7-repeat allele of the dopamine receptor D4 gene have extreme behavior but normal performance on critical neuropsychological tests of attention. *Proc Natl Acad Sci U S A* 2000; **97**: 4754–4759.
- 160 Gizer IR, Ficks C, Waldman ID. Candidate gene studies of ADHD: a meta-analytic review. *Hum Genet* 2009; **126**: 51–90.
- 161 Faraone S V, Doyle AE, Mick E, Biederman J. Meta-analysis of the association between the 7-repeat allele of the dopamine D(4) receptor gene and attention deficit hyperactivity disorder. *Am J Psychiatry* 2001; **158**: 1052–1057.
- 162 Li D, Sham PC, Owen MJ, He L. Meta-analysis shows significant association between dopamine system genes and attention deficit hyperactivity disorder (ADHD). *Hum Mol Genet* 2006; **15**: 2276–84.
- 163 Todd RD, Neuman RJ, Lobos E a, Jong YJ, Reich W, Heath a C. Lack of association of dopamine D4 receptor gene polymorphisms with ADHD subtypes in a population sample of twins. *Am. J. Med. Genet.* 2001; **105**: 432–8.
- 164 Seaman MI, Fisher JB, Chang F, Kidd KK. Tandem duplication polymorphism upstream of the dopamine D4 receptor gene (DRD4). *Am J Med Genet* 1999; **88**: 705–709.
- 165 McCracken JT, Smalley SL, McGough JJ, Crawford L, Del’Homme M, Cantor RM *et al*. Evidence for linkage of a tandem duplication polymorphism upstream of the dopamine D4 receptor gene (DRD4) with attention deficit hyperactivity disorder (ADHD). *Mol Psychiatry* 2000; **5**: 531–536.

- 166 Barr CL, Feng Y, Wigg KG, Schachar R, Tannock R, Roberts W et al. 5'-untranslated region of the dopamine D4 receptor gene and attention-deficit hyperactivity disorder. *Am J Med Genet* 2001; **105**: 84–90.
- 167 Cook Jr. EH, Stein MA, Krasowski MD, Cox NJ, Olkon DM, Kieffer JE et al. Association of attention-deficit disorder and the dopamine transporter gene. *Am J Hum Genet* 1995; **56**: 993–998.
- 168 Gill M, Daly G, Heron S, Hawi Z, Fitzgerald M. Confirmation of association between attention deficit hyperactivity disorder and a dopamine transporter polymorphism. *Mol Psychiatry* 1997; **2**: 311–313.
- 169 Holmes J, Payton A, Barrett JH, Hever T, Fitzpatrick H, Trumper AL et al. A family-based and case-control association study of the dopamine D4 receptor gene and dopamine transporter gene in attention deficit hyperactivity disorder. *Mol Psychiatry* 2000; **5**: 523–30.
- 170 Todd RD, Jong YJI, Lobos EA, Reich W, Heath AC, Neuman RJ. No association of the dopamine transporter gene 3' VNTR polymorphism with ADHD subtypes in a population sample of twins. *Am J Med Genet - Neuropsychiatr Genet* 2001; **105**: 745–748.
- 171 Aparecida da Silva M, Cordeiro Q, Louzã M, Vallada H. Lack of association between a 3'UTR VNTR polymorphism of dopamine transporter gene (SLC6A3) and ADHD in a Brazilian sample of adult patients. *J Atten Disord* 2011; **15**: 305–309.
- 172 Jacobsen LK, Staley JK, Zoghbi SS, Seibyl JP, Kosten TR, Innis RB et al. Prediction of dopamine transporter binding availability by genotype: A preliminary report. *Am J Psychiatry* 2000; **157**: 1700–1703.
- 173 Heinz A, Goldman D, Jones DW, Palmour R, Hommer D, Gorey JG et al. Genotype influences in vivo dopamine transporter availability in human striatum. *Neuropsychopharmacology* 2000; **22**: 133–139.
- 174 Bellgrove MA, Hawi Z, Kirley A, Gill M, Robertson IH. Dissecting the attention deficit hyperactivity disorder (ADHD) phenotype: Sustained attention, response variability and spatial attentional asymmetries in relation to dopamine transporter (DAT1) genotype. *Neuropsychologia* 2005; **43**: 1847–1857.
- 175 Loo SK, Specter E, Smolen A, Hopfer C, Teale PD, Reite ML. Functional effects of the DAT1 polymorphism on EEG measures in ADHD. *J Am Acad Child Adolesc Psychiatry* 2003; **42**: 986–93.
- 176 Maher BS, Marazita ML, Ferrell RE, Vanyukov MM. Dopamine system genes and attention deficit hyperactivity disorder: a meta-analysis. *Psychiatr Genet* 2002; **12**: 207–15.
- 177 Brookes K, Xu X, Chen W, Zhou K, Neale B, Lowe N et al. The analysis of 51 genes in DSM-IV combined type attention deficit hyperactivity disorder: association signals in DRD4, DAT1 and 16 other genes. *Mol Psychiatry* 2006; **11**: 934–953.
- 178 Lachman HM, Morrow B, Shprintzen R, Veit S, Parsia SS, Faedda G et al. Association of codon 108/158 catechol-O-methyltransferase gene polymorphism with the psychiatric manifestations of velo-cardio-facial syndrome. *Am J Med Genet* 1996; **67**: 468–472.
- 179 Cheuk DKL, Wong V. Meta-analysis of association between a catechol-O-methyltransferase gene polymorphism and attention deficit hyperactivity disorder. *Behav Genet* 2006; **36**: 651–9.
- 180 Eisenberg J, Mei-Tal G, Steinberg A, Tartakovsky E, Zohar A, Gritsenko I et al. Haplotype relative risk study of catechol-O-methyltransferase (COMT) and attention deficit hyperactivity disorder (ADHD):

- Association of the high- enzyme activity val allele with ADHD impulsive-hyperactive phenotype. *Am J Med Genet - Neuropsychiatr Genet* 1999; **88**: 497–502.
- 181 DeYoung CG, Getchell M, Kopolov RA, Yrigollen CM, Haeffel GJ, af Klinteberg B *et al.* Variation in the catechol-O-methyltransferase Val 158 Met polymorphism associated with conduct disorder and ADHD symptoms, among adolescent male delinquents. *Psychiatr Genet* 2010; **20**: 20–24.
- 182 Barr CL, Wigg K, Malone M, Schachar R, Tannock R, Roberts W *et al.* Linkage study of catechol-O-methyltransferase and attention-deficit hyperactivity disorder. *Am J Med Genet* 1999; **88**: 710–713.
- 183 Manor I, Kotler M, Sever Y, Eisenberg J, Cohen H, Ebstein RP *et al.* Failure to replicate an association between the catechol-O-methyltransferase polymorphism and attention deficit hyperactivity disorder in a second, independently recruited Israeli cohort. *Am J Med Genet* 2000; **96**: 858–860.
- 184 Tahir E, Curran S, Yazgan Y, Ozbay F, Cirakoglu B, Asherson PJ. No association between low- and high-activity catecholamine-methyl-transferase (COMT) and attention deficit hyperactivity disorder (ADHD) in a sample of Turkish children. *Am J Med Genet (Neuropsychiatric Genet)* 2000; **96**: 285–8.
- 185 Smith KM, Daly M, Fischer M, Yiannoutsos CT, Bauer L, Barkley R *et al.* Association of the dopamine beta hydroxylase gene with attention deficit hyperactivity disorder: genetic analysis of the Milwaukee longitudinal study. *Am J Med Genet B Neuropsychiatr Genet* 2003; **119B**: 77–85.
- 186 Daly G, Hawi Z, Fitzgerald M, Gill M. Mapping susceptibility loci in attention deficit hyperactivity disorder : preferential transmission of parental alleles at DAT1, DBH and DRD5 to affected children. *Mol Psychiatry* 1999; **4**: 192–196.
- 187 Roman T, Schmitz M, Polanczyk G V, Eizirik M, Rohde LA, Hutz MH. Further Evidence for the Association Between Attention-Deficit / Hyperactivity Disorder and the Dopamine- b -Hydroxylase Gene. *Am J Med Genet B Neuropsychiatr Genet* 2002; **158**: 154–158.
- 188 Bellgrove MA, Hawi Z, Gill M, Robertson IH. The cognitive genetics of attention deficit hyperactivity disorder (ADHD): Sustained attention as a candidate phenotype. *Cortex* 2006; **42**: 838–845.
- 189 Mu K, Daly M, Fischer M, Yiannoutsos CT, Bauer L, Barkley R *et al.* Association of the Dopamine Beta Hydroxylase Gene With Attention Deficit Hyperactivity Disorder : Genetic Analysis of the Milwaukee Longitudinal Study. *Am J Med Genet Part B, Neuropsychiatr Genet* 2003; **119B**: 77–85.
- 190 Nyman ES, Ogdie MN, Loukola A, Varilo T, Taanila A, Hurtig T *et al.* ADHD candidate gene study in a population-based birth cohort: association with DBH and DRD2. *J Am Acad Child Adolesc Psychiatry* 2007; **46**: 1614–1621.
- 191 Wigg K, Zai G, Schachar R, Tannock R, Roberts W, Malone M *et al.* Attention deficit hyperactivity disorder and the gene for dopamine beta-hydroxylase. *Am J Psychiatry* 2002; **159**: 1046–1048.
- 192 Payton A, Holmes J, Barrett JH, Hever T, Fitzpatrick H, Trumper AL *et al.* Examining for association between candidate gene polymorphisms in the dopamine pathway and attention-deficit hyperactivity disorder: A family-based study. *Am J Med Genet - Neuropsychiatr Genet* 2001; **105**: 464–470.
- 193 Bhaduri N, Mukhopadhyay K. Lack of significant association between -1021C??T polymorphism in the dopamine beta hydroxylase gene and attention deficit hyperactivity disorder. *Neurosci Lett* 2006; **402**: 12–16.

- 194 Zhang HB, Wang YF, Li J, Wang B, Yang L. [Association between dopamine beta hydroxylase gene and attention deficit hyperactivity disorder complicated with disruptive behavior disorder]. *Zhonghua Er Ke Za Zhi* 2005; **43**: 26–30.
- 195 Das M, Bhowmik A Das, Bhaduri N, Sarkar K, Ghosh P, Sinha S et al. Role of gene-gene/gene-environment interaction in the etiology of eastern Indian ADHD probands. *Prog Neuro-Psychopharmacology Biol Psychiatry* 2011; **35**: 577–587.
- 196 Cohen IL, Liu X, Lewis MES, Chudley A, Forster-Gibson C, Gonzalez M et al. Autism severity is associated with child and maternal MAOA genotypes. *Clin Genet* 2011; **79**: 355–362.
- 197 Gallardo-Pujol D, Andrés-Pueyo A, Maydeu-Olivares A. MAOA genotype, social exclusion and aggression: An experimental test of a gene-environment interaction. *Genes, Brain Behav* 2013; **12**: 140–145.
- 198 Eslami Amirabadi MR, Rajezi Esfahani S, Davari-Ashtiani R, Khademi M, Emamalizadeh B, Movafagh A et al. Monoamine oxidase a gene polymorphisms and bipolar disorder in Iranian population. *Iran Red Crescent Med J* 2015; **17**: e23095.
- 199 Schulze TG, Muller DJ, Krauss H, Scherk H, Ohlraun S, Sygailo Y V et al. Association between a functional polymorphism in the monoamine oxidase A gene promoter and major depressive disorder. *Am J Med Genet* 2000; **96**: 801–3.
- 200 Jiang S, Xin R, Wu X, Lin S, Qian Y, Ren D et al. Association between attention deficit hyperactivity disorder and the DXS7 locus. *Am J Med Genet - Neuropsychiatr Genet* 2000; **96**: 289–292.
- 201 Jiang S, Xin R, Lin S, Qian Y, Tang G, Wang D et al. Linkage Studies Between Attention-Deficit Hyperactivity Disorder and the Monoamine Oxidase Genes. *Am J Hum Genet* 2001; **105**: 783–788.
- 202 Domschke K, Sheehan K, Lowe N, Kirley A, Mullins C, Sullivan RO et al. Association Analysis of the Monoamine Oxidase A and B Genes With Attention Deficit Hyperactivity Disorder ( ADHD ) in an Irish Sample : Preferential Transmission of the MAO-A 941G Allele to Affected Children. *Am J Med Genet B Neuropsychiatr Genet* 2005; **114**: 110–114.
- 203 Manor I, Tyano S, Mel E, Eisenberg J, Kotler M, Ebstein RP. Family-based and association studies of monoamine oxidase A and attention deficit hyperactivity disorder ( ADHD ): preferential transmission of the long promoter- region repeat and its association with impaired performance on a continuous performance test. *Mol Psychiatry* 2002; **7**: 626–632.
- 204 Sharp SI, McQuillin A, Gurling HMD. Genetics of attention-deficit hyperactivity disorder (ADHD). *Neuropharmacology*. 2009; **57**: 590–600.
- 205 Barr CL, Feng Y, Wigg K, Bloom S, Roberts W, Malone M et al. Identification of DNA variants in the SNAP-25 gene and linkage study of these polymorphisms and attention-deficit hyperactivity disorder. *Mol Psychiatry* 2000; **5**: 405–9.
- 206 Kustanovich V, Merriman B, Mccraken JT, Smalley SL, Nelson SF. Biased paternal transmission of SNAP-25 risk alleles in attention-deficit hyperactivity disorder. *Mol Psychiatry* 2003; **8**: 309–315.
- 207 Lowe N, Kirley A, Hawi Z, Sham P, Wickham H, Kratochvil CJ et al. Joint analysis of the DRD5 marker concludes association with attention-deficit/hyperactivity disorder confined to the predominantly

- inattentive and combined subtypes. *Am J Hum Genet* 2004; **74**: 348–56.
- 208 Li J, Wang Y, Zhou R, Zhang H, Yang L, Wang B *et al.* Association between polymorphisms in serotonin transporter gene and attention deficit hyperactivity disorder in Chinese Han subjects. *Am J Med Genet B Neuropsychiatr Genet* 2007; **144B**: 14–9.
- 209 Lesch KP, Bengel D, Heils a, Sabol SZ, Greenberg BD, Petri S *et al.* Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science* 1996; **274**: 1527–1531.
- 210 Greenberg BD, Tolliver TJ, Huang SJ, Li Q, Bengel D, Murphy DL. Genetic variation in the serotonin transporter promoter region affects serotonin uptake in human blood platelets. *Am J Med Genet* 1999; **88**: 83–87.
- 211 Kent L, Doerry U, Hardy E, Parmar R, Gingell K, Hawi Z *et al.* Evidence that variation at the serotonin transporter gene influences susceptibility to attention deficit hyperactivity disorder (ADHD): analysis and pooled analysis. *Mol Psychiatry* 2002; **7**: 908–12.
- 212 Zoroglu SS, Erdal ME, Alasehirli B, Erdal N, Sivasli E, Tutkun H *et al.* Significance of serotonin transporter gene 5-HTTLPR and variable number of tandem repeat polymorphism in attention deficit hyperactivity disorder. *Neuropsychobiology* 2002; **45**: 176–181.
- 213 Elia J, Gai X, Xie HM, Perin JC, Geiger E, Glessner JT *et al.* Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. *Mol Psychiatry* 2010; **15**: 637–46.
- 214 Elia J, Glessner JT, Wang K, Takahashi N, Shtir CJ, Hadley D *et al.* Genome-wide copy number variation study associates metabotropic glutamate receptor gene networks with attention deficit hyperactivity disorder. *Nat Genet* 2012; **44**: 78–84.
- 215 Jarick I, Volckmar A-L, Pütter C, Pechlivanis S, Nguyen TT, Dauvermann MR *et al.* Genome-wide analysis of rare copy number variations reveals PARK2 as a candidate gene for attention-deficit/hyperactivity disorder. *Mol Psychiatry* 2012; : 1–7.
- 216 Williams NM, Franke B, Mick E, Anney RJL, Freitag CM, Gill M *et al.* Genome-wide analysis of copy number variants in attention deficit hyperactivity disorder: The role of rare variants and duplications at 15q13.3. *Am J Psychiatry* 2012; **169**: 195–204.
- 217 Williams NM, Zaharieva I, Martin A, Langley K, Mantripragada K, Fossdal R *et al.* Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. *Lancet* 2010; **376**: 1401–8.
- 218 Lionel AC, Crosbie J, Barbosa N, Goodale T, Thiruvahindrapuram B, Rickaby J *et al.* Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD. *Sci Transl Med* 2011; **3**: 95ra75.
- 219 Lesch K-P, Selch S, Renner TJ, Jacob C, Nguyen TT, Hahn T *et al.* Genome-wide copy number variation analysis in attention-deficit/hyperactivity disorder: association with neuropeptide Y gene dosage in an extended pedigree. *Mol Psychiatry* 2011; **16**: 491–503.
- 220 Castellanos FX, Lee PP, Sharp W, Jeffries NO, Greenstein DK, Clasen LS *et al.* Developmental Trajectories of Brain Volume Abnormalities in Children and Adolescents With Attention-Deficit/Hyperactivity Disorder.

- JAMA 2002; **288**: 1740.
- 221 Sowell ER, Thompson PM, Welcome SE, Henkenius AL, Toga AW, Peterson BS. Cortical abnormalities in children and adolescents with attention-deficit hyperactivity disorder. *Lancet* 2003; **362**: 1699–1707.
- 222 Mahone EM, Crocetti D, Ranta ME, Gaddis a, Cataldo M, Slifer KJ *et al.* A preliminary neuroimaging study of preschool children with ADHD. *Clin Neuropsychol* 2011; **25**: 1009–28.
- 223 Monuteaux MC, Seidman LJ, Faraone S V., Makris N, Spencer T, Valera E *et al.* A preliminary study of dopamine D4 receptor genotype and structural brain alterations in adults with ADHD. *Am J Med Genet Part B Neuropsychiatr Genet* 2008; **147**: 1436–1441.
- 224 Castellanos FX, Lau E, Tayebi N, Lee P, Long RE, Giedd JN *et al.* Lack of an association between a dopamine-4 receptor polymorphism and attention-deficit/hyperactivity disorder: genetic and brain morphometric analyses. *Mol Psychiatry* 1998; **3**: 431–434.
- 225 Fernández-Jaén A, López-Martín S, Albert J, Fernández-Mayoralas DM, Fernández-Perrone AL, de La Peña MJ *et al.* Cortical thickness differences in the prefrontal cortex in children and adolescents with ADHD in relation to dopamine transporter (DAT1) genotype. *Psychiatry Res - Neuroimaging* 2015; **233**: 409–417.
- 226 Silk T, Genk S, Anderson V, Efron D, Hazell F, Nicholson J *et al.* Developmental brain trajectories in children with ADHD and controls: a longitudinal neuroimaging study. *BMC Psychiatry* 2016; **10.1186/s1**.
- 227 Takeuchi H, Tomita H, Taki Y, Kikuchi Y, Ono C, Yu Z *et al.* Cognitive and neural correlates of the 5-repeat allele of the dopamine D4 receptor gene in a population lacking the 7-repeat allele. *Neuroimage* 2015; **1**.
- 228 Gehricke JG, Swanson JM, Duong S, Nguyen J, Wigal TL, Fallon J *et al.* Increased brain activity to unpleasant stimuli in individuals with the 7R allele of the DRD4 gene. *Psychiatry Res - Neuroimaging* 2015; **231**: 58–63.
- 229 Gilsbach S, Neufang S, Scherag S, Vloet TD, Fink GR, Herpertz-Dahlmann B *et al.* Effects of the DRD4 genotype on neural networks associated with executive functions in children and adolescents. *Dev Cogn Neurosci* 2012; **2**: 417–427.
- 230 Mulligan RC, Kristjansson SD, Reiersen AM, Parra AS, Anokhin AP. Neural correlates of inhibitory control and functional genetic variation in the dopamine D4 receptor gene. *Neuropsychologia* 2014; **62**: 306–318.
- 231 van Rooij D, Hoekstra PJ, Bralten J, Hakobjan M, Oosterlaan J, Franke B *et al.* Influence of DAT1 and COMT variants on neural activation during response inhibition in adolescents with attention-deficit/hyperactivity disorder and healthy controls. *Psychol. Med.* 2015; **45**: 3159–3170.
- 232 Heinzl S, Dresler T, Baehne CG, Heine M, Boreatti-H??mmer A, Jacob CP *et al.* COMT x DRD4 epistasis impacts prefrontal cortex function underlying response control. *Cereb Cortex* 2013; **23**: 1453–1462.
- 233 Hong SB, Zalesky A, Park S, Yang YH, Park MH, Kim B *et al.* COMT genotype affects brain white matter pathways in attention-deficit/hyperactivity disorder. *Hum Brain Mapp* 2015; **36**: 367–377.
- 234 Oliver S.P. D, Band G, Pirinen M, Et. al. The correlation between reading and mathematics ability at age twelve has a substantial genetic component. *Nat Commun* 2014; **5**: 1–6.
- 235 Marino C, Mascheretti S, Riva V, Cattaneo F, Rigoletto C, Rusconi M *et al.* Pleiotropic effects of DCDC2 and DYX1C1 genes on language and mathematics traits in nuclear families of developmental dyslexia.

- Behav Genet 2011; **41**: 67–76.
- 236 Ludwig KU, Sämman P, Alexander M, Becker J, Bruder J, Moll K *et al.* A common variant in myosin-18B contributes to mathematical abilities in children with dyslexia and intraparietal sulcus variability in adults. *Transl Psychiatry* 2013; **3**: e229.
- 237 Pettigrew KA, Fajutrao Valles SF, Moll K, Northstone K, Ring S, Pennell C *et al.* Lack of replication for the myosin-18B association with mathematical ability in independent cohorts. *Genes, Brain Behav* 2015; **14**: 369–376.
- 238 Thapar A, Langley K, Fowler T, Rice F, Turic D, Whittinger N *et al.* Catechol O-methyltransferase gene variant and birth weight predict early-onset antisocial behavior in children with attention-deficit/hyperactivity disorder. *Arch Gen Psychiatry* 2005; **62**: 1275–1278.
- 239 Caspi A, Langley K, Milne B, Moffitt TE, O'Donovan M, Owen MJ *et al.* A replicated molecular genetic basis for subtyping antisocial behavior in children with attention-deficit/hyperactivity disorder. *Arch Gen Psychiatry* 2008; **65**: 203–210.
- 240 Willcutt EG, Pennington BF, Olson RK, Chhabildas N, Hulslander J. Neuropsychological analyses of comorbidity between reading disability and attention deficit hyperactivity disorder: in search of the common deficit. *Dev Neuropsychol* 2005; **27**: 35–78.
- 241 Couto JM, Gomez L, Wigg K, Ickowicz A, Pathare T, Malone M *et al.* Association of Attention-Deficit/Hyperactivity Disorder with a Candidate Region for Reading Disabilities on Chromosome 6p. *Biol Psychiatry* 2009; **66**: 368–375.
- 242 Willcutt EG, Pennington BF, DeFries JC. Twin study of the etiology of comorbidity between reading disability and attention-deficit/hyperactivity disorder. *Am J Med Genet - Neuropsychiatr Genet* 2000; **96**: 293–301.
- 243 Paloyelis Y, Rijdsdijk F, Wood AC, Asherson P, Kuntsi J. The genetic association between ADHD symptoms and reading difficulties: The role of inattentiveness and IQ. *J Abnorm Child Psychol* 2010; **38**: 1083–1095.
- 244 Greven CU, Harlaar N, Dale PS, Plomin R. Genetic Overlap between ADHD Symptoms and Reading is largely Driven by Inattentiveness rather than Hyperactivity-Impulsivity. *J Can Acad Child Adolesc Psychiatry* 2011; **20**: 6–14.
- 245 Willcutt EG, Pennington BF, Olson RK, DeFries JC. Understanding comorbidity: a twin study of reading disability and attention-deficit/hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet* 2007; **144B**: 709–714.
- 246 Loo SK, Fisher SE, Francks C, Ogdie MN, MacPhie IL, Yang M *et al.* Genome-wide scan of reading ability in affected sibling pairs with attention-deficit/hyperactivity disorder: unique and shared genetic effects. *Mol Psychiatry* 2004; **9**: 485–493.
- 247 Field LL, Shumansky K, Ryan J, Truong D, Swiergala E, Kaplan BJ. Dense-map genome scan for dyslexia supports loci at 4q13, 16p12, 17q22; suggests novel locus at 7q36. *Genes, Brain Behav* 2013; **12**: 56–69.
- 248 Stevenson J, Langley K, Pay H, Payton A, Worthington J, Ollier W *et al.* Attention deficit hyperactivity disorder with reading disabilities: preliminary genetic findings on the involvement of the ADRA2A gene. *J Child Psychol Psychiatry Allied Discip* 2005; **46**: 1081–8.

- 249 Boada R, Willcutt EG, Pennington BF. Understanding the comorbidity between Dyslexia and Attention-deficit/Hyperactivity disorder. *Top lang Disord* 2012; **32**: 264–284.
- 250 Gottesman II, Gould TD. The endophenotype concept in psychiatry: Etymology and strategic intentions. *Am. J. Psychiatry*. 2003; **160**: 636–645.
- 251 Flint J, Munafò MR. The endophenotype concept in psychiatric genetics. *Psychol Med* 2007; **37**: 163–80.
- 252 Shaywitz SE, Escobar MD, Shaywitz BA, Fletcher JM, Makuch R. Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *N Engl J Med* 1992; **326**: 145–150.
- 253 American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-V*. 2013 doi:10.1176/appi.books.9780890425596.744053.
- 254 World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders. *Int Classif* 1992; **10**: 1–267.
- 255 Levy F, Hay D a, McStephen M, Wood C, Waldman I. Attention-deficit hyperactivity disorder: a category or a continuum? Genetic analysis of a large-scale twin study. *J Am Acad Child Adolesc Psychiatry* 1997; **36**: 737–744.
- 256 Paracchini S, Steer CD, Buckingham LL, Morris AP, Ring S, Scerri T *et al*. Association of the KIAA0319 dyslexia susceptibility gene with reading skills in the general population. *Am J Psychiatry* 2008; **165**: 1576–1584.
- 257 Lind PA, Luciano M, Wright MJ, Montgomery GW, Martin NG, Bates TC. Dyslexia and DCDC2 : normal variation in reading and spelling is associated with DCDC2 polymorphisms in an Australian population sample. *Eur J Hum Genet EJHG* 2010; **18**: 668–673.
- 258 Thapar A, Langley K, Asherson P, Gill M. Gene-environment interplay in attention-deficit hyperactivity disorder and the importance of a developmental perspective. *Br J Psychiatry* 2007; **190**: 1–3.
- 259 Frodl T, Skokauskas N. Meta-analysis of structural MRI studies in children and adults with attention deficit hyperactivity disorder indicates treatment effects. *Acta Psychiatr Scand* 2012; **125**: 114–126.
- 260 Rutter M, Moffitt TE, Caspi A. Gene-environment interplay and psychopathology: Multiple varieties but real effects. *J. Child Psychol. Psychiatry Allied Discip.* 2006; **47**: 226–261.
- 261 Friend A, DeFries JC, Olson RK. Parental education moderates genetic influences on reading disability. *Psychol Sci a J Am Psychol Soc / APS* 2008; **19**: 1124–1130.
- 262 Taylor J, Roehrig a D, Soden Hensler B, Connor CM, Schatschneider C. Teacher quality moderates the genetic effects on early reading. *Science* 2010; **328**: 512–514.
- 263 Pennington BF, McGrath LM, Rosenberg J, Barnard H, Smith SD, Willcutt EG *et al*. Gene X environment interactions in reading disability and attention-deficit/hyperactivity disorder. *Dev Psychol* 2009; **45**: 77–89.
- 264 Todd RD, Neuman RJ. Gene-environment interactions in the development of combined type ADHD: Evidence for a synapse-based model. *Am J Med Genet Part B Neuropsychiatr Genet* 2007; **144**: 971–975.
- 265 Kahn RS, Khoury J, Nichols WC, Lanphear BP. Role of dopamine transporter genotype and maternal prenatal smoking in childhood hyperactive-impulsive, inattentive, and oppositional behaviors. *J Pediatr* 2003; **143**: 104–110.
- 266 Neuman RJ, Lobos E, Reich W, Henderson CA, Sun LW, Todd RD. Prenatal Smoking Exposure and

- Dopaminergic Genotypes Interact to Cause a Severe ADHD Subtype. *Biol Psychiatry* 2007; **61**: 1320–1328.
- 267 Brookes K-J, Mill J, Guindalini C, Curran S, Xu X, Knight J *et al.* A common haplotype of the dopamine transporter gene associated with attention-deficit/hyperactivity disorder and interacting with maternal use of alcohol during pregnancy. *Arch Gen Psychiatry* 2006; **63**: 74–81.
- 268 Mill J, Petronis A. Pre- and peri-natal environmental risks for attention-deficit hyperactivity disorder (ADHD): The potential role of epigenetic processes in mediating susceptibility. *J. Child Psychol. Psychiatry Allied Discip.* 2008; **49**: 1020–1030.
- 269 Martel MM, Nikolas M, Jernigan K, Friderici K, Waldman I, Nigg JT. The dopamine receptor D4 gene (DRD4) moderates family environmental effects on ADHD. *J Abnorm Child Psychol* 2011; **39**: 1–10.
- 270 Nigg J, Nikolas M, Burt SA. Measured gene-by-environment interaction in relation to attention-deficit/hyperactivity disorder. *J. Am. Acad. Child Adolesc. Psychiatry.* 2010; **49**: 863–873.
- 271 Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 2005; **6**: 95–108.
- 272 Hardy J, Singleton A. Genomewide association studies and human disease. *N Engl J Med* 2009; **360**: 1759–1768.
- 273 International T, Consortium H. The International HapMap Project. *Nature* 2003; **426**: 789–796.
- 274 Martin ER, Lai EH, Gilbert JR, Rogala AR, Afshari AJ, Riley J *et al.* SNPping away at complex diseases: analysis of single-nucleotide polymorphisms around APOE in Alzheimer disease. *Am J Hum Genet* 2000; **67**: 383–94.
- 275 Kruglyak L. The road to genome-wide association studies. *Nat Rev Genet* 2008; **9**: 314–318.
- 276 König I, Schumacher J, Hoffmann P, Kleensang A, Ludwig K, Grimm T *et al.* Mapping for dyslexia and related cognitive trait loci provides strong evidence for further risk genes on chromosome 6p21. *Am J Med Genet Part B Neuropsychiatr Genet* 2011; **156**: 36–43.
- 277 Rubenstein KB, Raskind WH, Berninger VW, Matsushita MM, Wijsman EM. Genome scan for cognitive trait loci of dyslexia: Rapid naming and rapid switching of letters, numbers, and colors. *Am J Med Genet Part B Neuropsychiatr Genet* 2014; **165**: 345–356.
- 278 Kaminen N, Hannula-Jouppi K, Kestilä M, Lahermo P, Muller K, Kaaranen M *et al.* A genome scan for developmental dyslexia confirms linkage to chromosome 2p11 and suggests a new locus on 7q32. *J Med Genet* 2003; **40**: 340–345.
- 279 Igo Jr. RP, Chapman NH, Berninger VW, Matsushita M, Brkanac Z, Rothstein JH *et al.* Genomewide scan for real-word reading subphenotypes of dyslexia: novel chromosome 13 locus and genetic complexity. *Am J Med Genet B Neuropsychiatr Genet* 2006; **141B**: 15–27.
- 280 Meaburn EL, Harlaar N, Craig IW, Schalkwyk LC, Plomin R. Quantitative trait locus association scan of early reading disability and ability using pooled DNA and 100K SNP microarrays in a sample of 5760 children. *Mol Psychiatry* 2007; : 1–12.
- 281 Luciano M, Montgomery GW, Martin NG, Wright MJ, Bates TC. SNP sets and reading ability: testing confirmation of a 10-SNP set in a population sample. *Twin Res Hum Genet* 2011; **14**: 228–32.
- 282 Scerri TS, Brandler WM, Paracchini S, Morris AP, Ring SM, Richardson AJ *et al.* PCSK6 is associated with

- handedness in individuals with dyslexia. *Hum Mol Genet* 2011; **20**: 608–14.
- 283 Brandler WM, Morris AP, Evans DM, Scerri TS, Kemp JP, Timpson NJ *et al.* Common Variants in Left/Right Asymmetry Genes and Pathways Are Associated with Relative Hand Skill. *PLoS Genet* 2013; **9**: doi:10.1371/journal.pgen.1003751.
- 284 Roeske D, Ludwig KU, Neuhoff N, Becker J, Bartling J, Bruder J *et al.* First genome-wide association scan on neurophysiological endophenotypes points to trans-regulation effects on SLC2A3 in dyslexic children. *Mol Psychiatry* 2011; **16**: 97–107.
- 285 Luciano M, Evans DM, Hansell NK, Medland SE, Montgomery GW, Martin NG *et al.* A genome-wide association study for reading and language abilities in two population cohorts. *Genes, Brain Behav* 2013; **12**: 645–652.
- 286 Carrion-Castillo A, van Bergen E, VINO A, van Zuijlen T, de Jong PF, Francks C *et al.* Evaluation of results from genome-wide studies of language and reading in a novel independent dataset. *Genes, Brain Behav* 2016; **15**: 531–41.
- 287 Eicher JD, Powers NR, Miller LL, Akshoomoff N, Amaral DG, Bloss CS *et al.* Genome-wide association study of shared components of reading disability and language impairment. *Genes, Brain Behav* 2013; **12**: 792–801.
- 288 Gialluisi A, Newbury D, Wilcutt E, Olson R, DeFries J, Brandler W *et al.* Genome-wide screening for DNA variants associated with reading and language traits. *Genes, Brain Behav* 2014; **13**: 686–701.
- 289 Neale BM, Lasky-Su J, Anney R, Franke B, Zhou K, Maller JB *et al.* Genome-wide association scan of attention deficit hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet* 2008; **147B**: 1337–1344.
- 290 Lantieri F, Glessner JT, Hakonarson H, Elia J, Devoto M. Analysis of GWAS top hits in ADHD suggests association to two polymorphisms located in genes expressed in the cerebellum. *Am J Med Genet Part B Neuropsychiatr Genet* 2010; **153**: 1127–1133.
- 291 Lasky-Su J, Neale BM, Franke B, Anney R, Zhou K, Maller JB *et al.* Genome-wide association scan of quantitative traits for attention deficit hyperactivity disorder identifies novel associations and confirms candidate gene associations. *Am J Med Genet Part B Neuropsychiatr Genet* 2008; **147**: 1345–1354.
- 292 Lesch KP, Timmesfeld N, Renner TJ, Halperin R, Röser C, Nguyen TT *et al.* Molecular genetics of adult ADHD: Converging evidence from genome-wide association and extended pedigree linkage studies. *J Neural Transm* 2008; **115**: 1573–1585.
- 293 Neale BM, Medland S, Ripke S, Anney R, Asherson P, Buitelaar J *et al.* Case-control genome-wide association study of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 2010; **49**: 906–920.
- 294 Mick E, Todorov A, Smalley S, Hu X, Loo S, Todd RD *et al.* Family-based genome-wide association scan of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 2010; **49**: 898–905.e3.
- 295 Hinney A, Jarick I, Carolin P, Pechlivanis S, Dauvermann MR, Beck S *et al.* Genome-Wide Association Study in German Patients With Attention Deficit / Hyperactivity Disorder. *Am J Med Genet B Neuropsychiatr Genet* 2011; **156**: 888–897.
- 296 Stergiakouli E, Hamshere M, Holmans P, Langley K, Zaharieva I, Hawi Z *et al.* Investigating the

- Contribution of Common Genetic Variants to the Risk and Pathogenesis of ADHD. *Am J Psychiatry* 2011; : 186–194.
- 297 Ebejer JL, Duffy DL, van der Werf J, Wright MJ, Montgomery G, Gillespie N a et al. Genome-wide association study of inattention and hyperactivity-impulsivity measured as quantitative traits. *Twin Res Hum Genet* 2013; **16**: 560–74.
- 298 Yang L, Neale BM, Liu L, Lee SH, Wray NR, Ji N et al. Polygenic transmission and complex neuro developmental network for attention deficit hyperactivity disorder: Genome-wide association study of both common and rare variants. *Am J Med Genet Part B Neuropsychiatr Genet* 2013; **162**: 419–430.
- 299 Zayats T, Athanasiu L, Sonderby I, Djurovic S, Westlye LT, Tamnes CK et al. Genome-wide analysis of attention deficit hyperactivity disorder in Norway. *PLoS One* 2015; **10**. doi:10.1371/journal.pone.0122501.
- 300 Yang L, Chang S, Lu Q, Zhang Y, Wu Z, Sun X et al. A new locus regulating MICALL2 expression was identified for association with executive inhibition in children with attention deficit hyperactivity disorder. *Mol Psychiatry* 2017. doi:10.1038/mp.2017.74.
- 301 Neale BM, Medland SE, Ripke S, Asherson P, Franke B, Lesch K-P et al. Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 2010; **49**: 884–97.
- 302 Middeldorp CM, Hammerschlag AR, Ouwens KG, Groen-Blokhuis MM, St. Pourcain B, Greven CU et al. A Genome-Wide Association Meta-Analysis of Attention-Deficit/Hyperactivity Disorder Symptoms in Population-Based Pediatric Cohorts. *J Am Acad Child Adolesc Psychiatry* 2016; **55**: 896–905.e6.
- 303 Alemany S, Ribasés M, Vilor-Tejedor N, Bustamante M, Sánchez-Mora C, Bosch R et al. New suggestive genetic loci and biological pathways for attention function in adult attention-deficit/hyperactivity disorder. *Am J Med Genet Part B Neuropsychiatr Genet* 2015; **168**: 459–470.
- 304 Sánchez-Mora C, Ramos-Quiroga J a, Bosch R, Corrales M, Garcia-Martínez I, Nogueira M et al. Case-control genome-wide association study of persistent attention-deficit hyperactivity disorder identifies FBXO33 as a novel susceptibility gene for the disorder. *Neuropsychopharmacology* 2015; **40**: 915–26.
- 305 Alemany S, Vilor-Tejedor N, Bustamante M, Pujol J, Macia D, Martinez-Vilavella G et al. A Genome-Wide Association Study of Attention Function in a Population-Based Sample of Children. *PLoS One* 2016; **11**: e0163048.
- 306 Hinney A, Scherag A, Jarick I, Albayrak Ö, Pütter C, Pechlivanis S et al. Genome-wide association study in German patients with attention deficit/hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet* 2011; **156B**: 888–97.
- 307 Verbruggen F, Logan GD. Models of response inhibition in the stop-signal and stop-change paradigms. *Neurosci. Biobehav. Rev.* 2009; **33**: 647–661.
- 308 Covault J, Abreu C, Kranzier H, Oncken C. Quantitative real-time PCR for gene dosage determinations in microdeletion genotypes. *Biotechniques* 2003; **35**: 594–598.
- 309 Muyzer G, de Waal EC, Uitterlinden a G. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Env Microbiol* 1993; **59**: 695–700.

- 310 Tahir E, Yazgan Y, Cirakoglu B, Ozbay F, Waldman I, Asherson PJ. Association and linkage of DRD4 and DRD5 with attention deficit hyperactivity disorder (ADHD) in a sample of Turkish children. *Mol Psychiatry* 2000; **5**: 396–404.
- 311 Luciano M, Lind PA, Duffy DL, Castles A, Wright MJ, Montgomery GW *et al*. A haplotype spanning KIAA0319 and TTRAP is associated with normal variation in reading and spelling ability. *Biol Psychiatry* 2007; **62**: 811–817.
- 312 Kebir O, Tabbane K, Sengupta S, Joobor R. Candidate genes and neuropsychological phenotypes in children with ADHD: review of association studies. *J Psychiatry Neurosci* 2009; **34**: 88–101.
- 313 Yuan HY, Chiou JJ, Tseng WH, Liu CH, Liu CK, Lin YJ *et al*. FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization. *Nucleic Acids Res* 2006; **34**: W635–41.
- 314 Hess C, Reif a, Strobel a, Boreatti-Hümmer a, Heine M, Lesch K-P *et al*. A functional dopamine-beta-hydroxylase gene promoter polymorphism is associated with impulsive personality styles, but not with affective disorders. *J Neural Transm* 2009; **116**: 121–30.
- 315 Vandenberg DJ, Persico AM, Hawkins AL, Griffin CA, Li X, Jabs EW *et al*. Human dopamine transporter gene (DAT1) maps to chromosome 5p15.3 and displays a VNTR. *Genomics* 1992; **14**: 1104–1106.
- 316 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
- 317 Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.
- 318 Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B *et al*. The structure of haplotype blocks in the human genome. *Science (80- )* 2002; **296**: 2225–2229.
- 319 Prabhu S, Pe'er I. Ultrafast genome-wide scan for SNP-SNP interactions in common complex disease. *Genome Res* 2012; **22**: 2230–40.
- 320 Zhao J, Jin L, Xiong M. Test for interaction between two unlinked loci. *Am J Hum Genet* 2006; **79**: 831–845.
- 321 Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF *et al*. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* 2007; **17**: 1665–1674.
- 322 Glessner JT, Li J, Hakonarson H. ParseCNV integrative copy number variation association software with quality tracking. *Nucleic Acids Res* 2013; **41**: e64.
- 323 Zhang K, Cui S, Chang S, Zhang L, Wang J. i-GSEA4GWAS: A web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. *Nucleic Acids Res* 2010; **38**. doi:10.1093/nar/gkq324.
- 324 Zhang K, Chang S, Cui S, Guo L, Zhang L, Wang J. ICSNPathway: Identify candidate causal SNPs and pathways from genome-wide association study by one analytical framework. *Nucleic Acids Res* 2011; **39**. doi:10.1093/nar/gkr391.
- 325 Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM *et al*. A versatile gene-based test for genome-wide association studies. *Am J Hum Genet* 2010; **87**: 139–145.
- 326 Nam D, Kim J, Kim S-Y, Kim S. GSA-SNP: a general approach for gene set analysis of polymorphisms.

- Nucleic Acids Res* 2010; **38**: W749–W754.
- 327 Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C *et al*. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009; **461**: 272–276.
- 328 Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP *et al*. LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics* 2010; **26**: 2336–2337.
- 329 Chartier. Bacterial DNA Content with Oragene® • DNA. *DNA Genotek* 2004; **PD-WP-002**.
- 330 Rylander-Rudqvist T, Hakansson N, Tybring G, Wolk A. Quality and quantity of saliva DNA obtained from the self-administrated oragene method—a pilot study on the cohort of Swedish men. *Cancer Epidemiol Biomarkers Prev* 2006; **15**: 1742–1745.
- 331 Nishita DM, Jack LM, McElroy M, McClure JB, Richards J, Swan GE *et al*. Clinical trial participant characteristics and saliva and DNA metrics. *BMC Med Res Methodol* 2009; **9**: 71.
- 332 Ghannoum MA, Jurevic RJ, Mukherjee PK, Cui F, Sikaroodi M, Naqvi A *et al*. Characterization of the oral fungal microbiome (mycobiome) in healthy individuals. *PLoS Pathog* 2010; **6**: e1000713.
- 333 Philibert RA, Zadorozhnyaya O, Beach SR, Brody GH. Comparison of the genotyping results using DNA obtained from blood and saliva. *Psychiatr Genet* 2008; **18**: 275–281.
- 334 Yang B, Chan RCK, Jing J, Li T, Sham P, Chen RYL. A meta-analysis of association studies between the 10-repeat allele of a VNTR polymorphism in the 3'-UTR of dopamine transporter gene and attention deficit hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet* 2007; **144B**: 541–50.
- 335 Mascheretti S, Bureau A, Trezzi V, Giorda R, Marino C. An assessment of gene-by-gene interactions as a tool to unfold missing heritability in dyslexia. *Hum Genet* 2015. doi:10.1007/s00439-015-1555-4.
- 336 Kang AM, Palmatier MA, Kidd KK. Global variation of a 40-bp VNTR in the 3'-untranslated region of the dopamine transporter gene (SLC6A3). *Biol Psychiatry* 1999; **46**: 151–160.
- 337 Mitchell RJ, Howlett S, Earl L, White NG, McComb J, Schanfield MS *et al*. Distribution of the 3' VNTR polymorphism in the human dopamine transporter gene in world populations. *Hum Biol an Int Rec Res* 2000; **72**: 295–304.
- 338 Judith FC, Kenneth RK, Pakstis AJ, Kidd KK. The world-wide distribution of allele frequencies at the human dopamine D4 receptor locus. *Hum Genet* 1996; **98**: 91–101.
- 339 Kinser PA, Robins JL. Control group design: Enhancing rigor in research of mind-body therapies for depression. *Evidence-based Complement. Altern. Med.* 2013; **2013**. doi:10.1155/2013/140467.
- 340 Smith TF. Meta-analysis of the heterogeneity in association of DRD4 7-repeat allele and AD/HD: stronger association with AD/HD combined type. *Am J Med Genet B Neuropsychiatr Genet* 2010; **153B**: 1189–99.
- 341 Nikolaidis A, Gray JR. ADHD and the DRD4 exon III 7-repeat polymorphism: an international meta-analysis. *Soc Cogn Affect Neurosci* 2010; **5**: 188–93.
- 342 Fair DA, Bathula D, Nikolas MA, Nigg JT. Distinct neuropsychological subgroups in typically developing youth inform heterogeneity in children with ADHD. *Proc Natl Acad Sci* 2012; **109**: 6769–6774.
- 343 Sheikhi AR, Martin N, Hay D, Piek JP. Phenotype refinement for comorbid attention deficit hyperactivity disorder and reading disability. *Am J Med Genet Part B Neuropsychiatr Genet* 2013; **162**: 44–54.
- 344 Czamara D, Tiesler CMT, Kohlböck G, Berdel D, Hoffmann B, Bauer CP *et al*. Children with ADHD

- Symptoms Have a Higher Risk for Reading, Spelling and Math Difficulties in the GINIplus and LISApplus Cohort Studies. *PLoS One* 2013; **8**. doi:10.1371/journal.pone.0063859.
- 345 Gooch D, Snowling MJ, Hulme C. Reaction Time Variability in Children With ADHD Symptoms and/or Dyslexia. *Dev. Neuropsychol.* 2012; **37**: 453–472.
- 346 McGrath LM, Pennington BF, Shanahan MA, Santerre-Lemmon LE, Barnard HD, Willcutt EG *et al.* A multiple deficit model of reading disability and attention-deficit/ hyperactivity disorder: Searching for shared cognitive deficits. *J Child Psychol Psychiatry Allied Discip* 2011; **52**: 547–557.
- 347 Willcutt EG, Betjemann RS, McGrath LM, Chhabildas NA, Olson RK, DeFries JC *et al.* Etiology and neuropsychology of comorbidity between RD and ADHD: The case for multiple-deficit models. *Cortex* 2010; **46**: 1345–1361.
- 348 Wilcke A, Weissfuss J, Kirsten H, Wolfram G, Boltze J, Ahnert P. The role of gene DCDC2 in German dyslexics. *Ann Dyslexia* 2009; **59**: 1–11.
- 349 Scerri TS, Morris AP, Buckingham LL, Newbury DF, Miller LL, Monaco AP *et al.* DCDC2, KIAA0319 and CMIP are associated with reading-related traits. *Biol Psychiatry* 2011; **70**: 237–245.
- 350 Mascheretti S, Bureau A, Battaglia M, Simone D, Quadrelli E, Croteau J *et al.* An assessment of gene-by-environment interactions in developmental dyslexia-related phenotypes. *Genes, Brain Behav* 2013; **12**: 47–55.
- 351 Landi N, Frost SJ, Mencl WE, Preston JL, Jacobsen LK, Lee M *et al.* The COMT Val/Met polymorphism is associated with reading related skills and consistent patterns of functional neural activation. *Dev Sci* 2013; **16**: 13–23.
- 352 Grigorenko EL, Deyoung CG, Getchell M, Haeffel GJ, Klinteberg BAF, Kuposov RA *et al.* Exploring interactive effects of genes and environments in etiology of individual differences in reading comprehension. *Dev Psychopathol* 2007; **19**: 1089–1103.
- 353 Miller DI, Halpern DF. The new science of cognitive sex differences. *Trends Cogn. Sci.* 2014; **18**: 37–45.
- 354 Akutagava-Martins GC, Salatino-Oliveira A, Kieling C, Genro JP, Polanczyk G V., Anselmi L *et al.* COMT and DAT1 genes are associated with hyperactivity and inattention traits in the 1993 Pelotas Birth Cohort: evidence of sex-specific combined effect. *J Psychiatry Neurosci* 2016; **41**: 405–412.
- 355 Gurvich C, Rossell S. Dopamine and cognitive control: Sex-by-genotype interactions influence the capacity to switch attention. *Behav Brain Reserach* 2015; **281**: 96–101.
- 356 Knopik VS, Alarcon M, DeFries JC. Common and specific gender influences on individual differences in reading performance: a twin study. *Pers Individ Dif* 1998; **25**: 269–277.
- 357 Biederman J, Kim JW, Doyle AE, Mick E, Fagerness J, Smoller JW *et al.* Sexually dimorphic effects of four genes (COMT, SLC6A2, MAOA, SLC6A4) in genetic associations of ADHD: a preliminary study. *Am J Med Genet B Neuropsychiatr Genet* 2008; **147B**: 1511–1518.
- 358 Li D, Sham PC, Owen MJ, He L. Meta-analysis shows significant association between dopamine system genes and attention deficit hyperactivity disorder (ADHD). *Hum Mol Genet* 2006; **15**: 2276–2284.
- 359 Waldman ID, Rowe DC, Abramowitz a, Kozel ST, Mohr JH, Sherman SL *et al.* Association and linkage of the dopamine transporter gene and attention-deficit hyperactivity disorder in children: heterogeneity

- owing to diagnostic subtype and severity. *Am J Hum Genet* 1998; **63**: 1767–1776.
- 360 Barr CL, Xu C, Kroft J, Feng Y, Wigg K, Zai G *et al.* Haplotype study of three polymorphisms at the dopamine transporter locus confirm linkage to attention-deficit/hyperactivity disorder. *Biol Psychiatry* 2001; **49**: 333–339.
- 361 Galili-Weisstub E, Levy S, Frisch A, Gross-Tsur V, Michaelovsky E, Kosov A *et al.* Dopamine transporter haplotype and attention-deficit hyperactivity disorder. *Mol Psychiatry* 2005; **10**: 617–8.
- 362 Hawi Z, Lowe N, Kirley a, Gruenhage F, Nöthen M, Greenwood T *et al.* Linkage disequilibrium mapping at DAT1, DRD5 and DBH narrows the search for ADHD susceptibility alleles at these loci. *Mol Psychiatry* 2003; **8**: 299–308.
- 363 Hsiung G-YR, Kaplan BJ, Petryshen TL, Lu S, Field LL. A dyslexia susceptibility locus (DYX7) linked to dopamine D4 receptor (DRD4) region on chromosome 11p15.5. *Am J Med Genet B Neuropsychiatr Genet* 2004; **125B**: 112–119.
- 364 Cornish KM, Savage R, Hocking DR, Hollis CP. Association of the dat1 genotype with inattentive behavior is mediated by reading ability in a general population sample. *Brain Cogn* 2011; **77**: 453–458.
- 365 Shao S, Kong R, Zou L, Zhong R, Lou J, Zhou J *et al.* The Roles of Genes in the Neuronal Migration and Neurite Outgrowth Network in Developmental Dyslexia: Single- and Multiple-Risk Genetic Variants. *Mol Neurobiol* 2016; **56**: 3967–75.
- 366 Mascheretti S, Trezzi V, Giorda R, Boivin M, Plourde V, Vitaro F *et al.* Complex effects of dyslexia risk factors account for ADHD traits: Evidence from two independent samples. *J. Child Psychol. Psychiatry Allied Discip.* 2016. doi:10.1111/jcpp.12612.
- 367 Wigg KG, Couto JM, Feng Y, Crosbie J, Anderson B, Cate-Carter T *et al.* Investigation of the Relationship of Attention Deficit Hyperactivity Disorder to the EKN1 Gene on Chromosome 15q21. *Sci. Stud. Read.* 2005; **9**: 261–283.
- 368 McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008; **9**: 356–69.
- 369 Potkin SG, Turner JA, Guffanti G, Lakatos A, Torri F, Keator DB *et al.* Genome-wide strategies for discovering genetic influences on cognition and cognitive disorders: methodological considerations. *Cogn Neuropsychiatry* 2009; **14**: 391–418.
- 370 Carrion-Castillo A, Maassen B, Franke B, Heister A, Naber M, van der Leij A *et al.* Association analysis of dyslexia candidate genes in a Dutch longitudinal sample. *Eur J Hum Genet* 2017; : 1–9.
- 371 Domínguez A, Cuetos Vega F. Desarrollo de las habilidades de reconocimiento de palabras en niños con distinta competencia lectora. *Cognitiva.* 1992; **4**: 193–208.
- 372 Baquero- Castellanos S. Procesamiento léxico del castellano por parte de niños y adultos. *Forma y Función* 2004; **18**: 45–73.
- 373 Müller O, Baquero Castellano S, Gallo A. El efecto de la lexicalidad en la decisión léxica a lo largo de la primaria. *Forma y Función* 2013; **26**.
- 374 Shaywitz SE, Shaywitz B a, Author A, Department of, Pediatrics YUS of MCSN, Haven CTUS *et al.* Paying

- attention to reading: The neurobiology of reading and dyslexia. *Dev Psychopathol* 2008; **20**: 1329–1349.
- 375 Zhang Y, Li J, Song S, Tardif T, Burmeister M, Villafuerte SM *et al*. Association of DCDC2 polymorphisms with normal variations in reading abilities in a Chinese population. *PLoS One* 2016; **11**.  
doi:10.1371/journal.pone.0153603.
- 376 Lim CK-P, Wong AM-B, Ho CS-H, Waye MM-Y. A common haplotype of KIAA0319 contributes to the phonological awareness skill in Chinese children. *Behav Brain Funct* 2014; **10**: 23.
- 377 Vellutino F, Fletcher JM, Snowling M, Scanlon D. Specific reading disability (dyslexia): what have we learned in the past four decades? *J Child Psychol Psychiatry* 2004; **45**: 2–40.
- 378 Wagner RK. Phonological processing abilities and reading: implications for disabled readers. *J Learn Disabil* 1986; **19**: 623–630.
- 379 Wagner RK, Torgesen JK, Rashotte CA, Hecht S A, Barker T A, Burgess S R *et al*. Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: a 5-year longitudinal study. *Dev Psychol* 1997; **33**: 468–79.
- 380 Wagner RK, Torgesen JK, Rashotte CA. Development of reading-related phonological processing abilities: New evidence of bidirectional causality from a latent variable longitudinal study. *Dev. Psychol.* 1994; **30**: 73–87.
- 381 Schulte-Körne G, Deimel W, Bartling J, Remschmidt H. Auditory processing and dyslexia: evidence for a specific speech processing deficit. 1998.
- 382 Neuhoff N, Bruder J, Bartling J, Warnke A, Remschmidt H, Müller-Myhsok B *et al*. Evidence for the late MMN as a neurophysiological endophenotype for dyslexia. *PLoS One* 2012; **7**.  
doi:10.1371/journal.pone.0034909.
- 383 Czamara. Association of a Rare Variant with Mismatch Negativity in a Region Between KIAA0319 and DCDC2 in Dyslexia. *Behav Genet* 2011; **41**: 110–119.
- 384 Nigg JT. Is ADHD a disinhibitory disorder? *Psychol Bull* 2001; **127**: 571–598.
- 385 Pennington BF, Ozonoff S. Executive Functions and Developmental Psychopathology. *J Child Psychol Psychiatry* 1996; **37**: 51–87.
- 386 Willcutt EG, Doyle AE, Nigg JT, Faraone S V, Pennington BF. Validity of the executive function theory of attention-deficit/hyperactivity disorder: a meta-analytic review. *Biol Psychiatry* 2005; **57**: 1336–1346.
- 387 Willcutt EG, Sonuga-Barke EJS, Nigg JT, Sergeant JA. Recent developments in neuropsychological models of childhood psychiatric disorders. *Adv Biol Psychiatry* 2008; **24**: 195–226.
- 388 Fossella J, Sommer T, Fan J, Wu Y, Swanson JM, Pfaff DW *et al*. Assessing the molecular genetics of attention networks. *BMC Neurosci* 2002; **3**: 14.
- 389 Jin J, Liu L, Gao Q, Chan RCK, Li H, Chen Y *et al*. The divergent impact of COMT Val158Met on executive function in children with and without attention-deficit/hyperactivity disorder. *Genes Brain Behav* 2016; **15**: 271–279.
- 390 Degen C, Zschocke J, Toro P, Sattler C, Wahl H-W, Schönknecht P *et al*. The COMT p.Val158Met polymorphism and cognitive performance in adult development, healthy aging and mild cognitive impairment. *Dement Geriatr Cogn Disord* 2015; **41**: 27–34.

- 391 Schneider KK, Schote AB, Meyer J, Frings C. Genes of the dopaminergic system selectively modulate top-down but not bottom-up attention. *Cogn Affect Behav Neurosci* 2014; **15**: 104–116.
- 392 Mill J, Xu X, Ronald A, Curran S, Price T, Knight J et al. Quantitative trait locus analysis of candidate gene alleles associated with attention deficit hyperactivity disorder (ADHD) in five genes: DRD4, DAT1, DRD5, SNAP-25, and 5HT1B. *Am J Med Genet - Neuropsychiatr Genet* 2005; **133 B**: 68–73.
- 393 Bidwell LC, Willcutt EG, McQueen MB, DeFries JC, Olson RK, Smith SD et al. A family based association study of DRD4, DAT1, and 5HTT and continuous traits of attention-deficit hyperactivity disorder. *Behav Genet* 2011; **41**: 165–74.
- 394 Muglia P, Jain U, Inkster B, Kennedy JL. A quantitative trait locus analysis of the dopamine transporter gene in adults with ADHD. *Neuropsychopharmacology* 2002; **27**: 655–662.
- 395 Froehlich TE, Lanphear BP, Dietrich KN, Cory-Slechta DA, Wang N, Kahn RS. Interactive Effects of a DRD4 Polymorphism, Lead, and Sex on Executive Functions in Children. *Biol Psychiatry* 2007; **62**: 243–249.
- 396 Blanchard MM, Chamberlain SR, Roiser J, Robbins TW, Muller U. Effects of two dopamine-modulating genes (DAT1 9/10 and COMT Val/Met) on n-back working memory performance in healthy volunteers. *Psychol Med* 2011; **41**: 611–618.
- 397 Cornish KM, Wilding JM, Hollis C. Visual search performance in children rated as good or poor attenders: the differential impact of DAT1 genotype, IQ, and chronological age. *Neuropsychology* 2008; **22**: 217–225.
- 398 Cornish KM, Manly T, Savage R, Swanson J, Morisano D, Butler N et al. Association of the dopamine transporter (DAT1) 10/10-repeat genotype with ADHD symptoms and response inhibition in a general population sample. *Mol Psychiatry* 2005; **10**: 686–698.
- 399 Kramer UM, Rojo N, Schule R, Cunillera T, Schols L, Marco-Pallares J et al. ADHD candidate gene (DRD4 exon III) affects inhibitory control in a healthy sample. *BMC Neurosci* 2009; **10**: 150.
- 400 Congdon E, Lesch KP, Canli T. Analysis of DRD4 and DAT polymorphisms and behavioral inhibition in healthy adults: Implications for impulsivity. *Am J Med Genet Part B Neuropsychiatr Genet* 2008; **147**: 27–32.
- 401 Bellgrove MA, Hawi Z, Lowe N, Kirley A, Robertson IH, Gill M. DRD4 gene variants and sustained attention in attention deficit hyperactivity disorder (ADHD): Effects of associated alleles at the VNTR and-521 SNP. *Am J Med Genet Part B-Neuropsychiatric Genet* 2005; **136B**: 81–86.
- 402 Manor I, Tyano S, Eisenberg J, Bachner-Melman R, Kotler M, Ebstein RP. The short DRD4 repeats confer risk to attention deficit hyperactivity disorder in a family-based design and impair performance on a continuous performance test (TOVA). *Mol Psychiatry* 2002; **7**: 790–4.
- 403 Langley K, Marshall L, van den Bree M, Thomas H, Owen M, O'Donovan M et al. Association of the dopamine D4 receptor gene 7-repeat allele with neuropsychological test performance of children with ADHD. *Am J Psychiatry* 2004; **161**: 133–138.
- 404 Kieling C, Roman T, Doyle AE, Hutz MH, Rohde LA. Association between DRD4 Gene and Performance of Children with ADHD in a Test of Sustained Attention. *Biol Psychiatry* 2006; **60**: 1163–1165.
- 405 Kim B, Koo M. Association between Dopamine D4 Receptor Gene Polymorphism and Scores on a Continuous Performance Test in Korean Children with Attention Deficit Hyperactivity Disorder. *Psychiatry Invest* 2009; **6**: 216–221.

- 406 Pickering BM, Willis AE. The implications of structured 5' untranslated regions on translation and disease. *Semin. Cell Dev. Biol.* 2005; **16**: 39–47.
- 407 Cazzola M, Skoda RC. Translational pathophysiology: a novel molecular mechanism of human disease. *Blood* 2000; **95**: 3280–3288.
- 408 Chatterjee S, Pal JK. Role of 5'- and 3'-untranslated regions of mRNAs in human diseases. *Biol Cell* 2009; **101**: 251–262.
- 409 Scheper GC, van der Knaap MS, Proud CG. Translation matters: protein synthesis defects in inherited disease. *Nat Rev Genet* 2007; **8**: 711–723.
- 410 Li C-C, Kuo J-C, Waterman CM, Kiyama R, Moss J, Vaughan M. Effects of brefeldin A-inhibited guanine nucleotide-exchange (BIG) 1 and KANK1 proteins on cell polarity and directed migration during wound healing. *Proc Natl Acad Sci* 2011; **108**: 19228–19233.
- 411 Bouchet BP, Gough RE, Ammon YC, van de Willige D, Post H, Jacquemet G et al. Talin-KANK1 interaction controls the recruitment of cortical microtubule stabilizing complexes to focal adhesions. *Elife* 2016; **5**. doi:10.7554/eLife.18124.
- 412 Kakinuma N, Roy BC, Zhu Y, Wang Y, Kiyama R. Kank regulates RhoA-dependent formation of actin stress fibers and cell migration via 14-3-3 in PI3K-Akt signaling. *J Cell Biol* 2008; **181**: 537–549.
- 413 Kakinuma N, Zhu Y, Wang Y, Roy BC, Kiyama R. Kank proteins: Structure, functions and diseases. *Cell. Mol. Life Sci.* 2009; **66**: 2651–2659.
- 414 Roy BC, Kakinuma N, Kiyama R. Kank attenuates actin remodeling by preventing interaction between IRSp53 and Rac1. *J Cell Biol* 2009; **184**: 253–267.
- 415 Cui Z, Shen Y, Chen KH, Mittal SK, Yang J-Y, Zhang G. KANK1 inhibits cell growth by inducing apoptosis through regulating CXXC5 in human malignant peripheral nerve sheath tumors. *Sci Rep* 2017; **7**.
- 416 JC W, LW M, LP R, A A, R O, FZ. B. Enrichment of small pathogenic deletions at chromosome 9p24.3 and 9q34.3 involving DOCK8, KANK1, EHMT1 genes identified by using high-resolution oligonucleotide-single nucleotide polymorphism array analysis. *Mol Cytogenet* 2016; **9**.
- 417 Sarkar S, Roy BC, Hatano N, Aoyagi T, Gohji K, Kiyama R. A novel ankyrin repeat-containing gene (Kank) located at 9p24 is a growth suppressor of renal cell carcinoma. *J Biol Chem* 2002; **277**: 36585–36591.
- 418 Lerer I, Sagi M, Meiner V, Cohen T, Zlotogora J, Abeliovich D. Deletion of the ANKRD15 gene at 9p24.3 causes parent-of-origin-dependent inheritance of familial cerebral palsy. *Hum Mol Genet* 2005; **14**: 3911–3920.
- 419 Hwang JY, Sim X, Wu Y, Liang J, Tabara Y, Hu C et al. Genome-wide association meta-analysis identifies novel variants associated with fasting plasma glucose in East Asians. *Diabetes* 2015; **64**: 291–298.
- 420 Huyghe JR, Jackson AU, Fogarty MP, Buchkovich ML, Stančáková A, Stringham HM et al. Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nat Genet* 2013; **45**: 197–201.
- 421 Vinci G, Chantot-Bastarad S, El Houate B, Lortat-Jacob S, Brauner R, McElreavey K. Association of deletion 9p, 46,XY gonadal dysgenesis and autistic spectrum disorder. *Mol Hum Reprod* 2007; **13**: 685–689.

- 422 Willour V, Yao Shugart Y, Samuels J, Grados M, Cullen B, Bienvenu O 3rd et al. Replication study supports evidence for linkage to 9p24 in obsessive-compulsive disorder. *Am J Hum Genet* 2004; **75**: 508–513.
- 423 Griggs BL, Ladd S, Saul RA, DuPont BR, Srivastava AK. Deducator of cytokinesis 8 is disrupted in two patients with mental retardation and developmental disabilities. *Genomics* 2008; **91**: 195–202.
- 424 Richardson AJ& RMA. Fatty acid metabolism - a new perspective in neurodevelopmental disorders: ADHD, dyslexia, dyspraxia and autism. *Prostaglandins, Leukot Essent Fat Acids* 2000; **63**,: 1–9.
- 425 Richardson a J, Calvin CM, Clisby C, Schoenheimer DR, Montgomery P, Hall J a et al. Fatty acid deficiency signs predict the severity of reading and related difficulties in dyslexic children. *Prostaglandins Leukot Essent Fatty Acids* 2000; **63**: 69–74.
- 426 Taylor KE, Higgins CJ, Calvin CM, Hall JA, Easton T, McDaid AM et al. Dyslexia in adults is associated with clinical signs of fatty acid deficiency. *Prostaglandins Leukot Essent Fatty Acids* 2000; **63**: 75–78.
- 427 Spahis S, Vanasse M, Belanger SA, Ghadirian P, Grenier E, Levy E. Lipid profile, fatty acid composition and pro- and anti-oxidant status in pediatric patients with attention-deficit/hyperactivity disorder. *Prostaglandins Leukot Essent Fat Acids* 2008.
- 428 Burgess JR, Stevens L, Zhang W, Peck L. Long-chain polyunsaturated fatty acids in children with attention-deficit hyperactivity disorder. *Am J Clin Nutr* 2000; **71**: 327S–30S.
- 429 Cooper RE, Tye C, Kuntsi J, Vassos E, Asherson P. The effect of omega-3 polyunsaturated fatty acid supplementation on emotional dysregulation, oppositional behaviour and conduct problems in ADHD: A systematic review and meta-analysis. *J. Affect. Disord.* 2016; **190**: 474–482.
- 430 Sanchez-Mora C, Cormand B, Ramos-Quiroga J, Hervas A, Bosch R, Palomar G et al. Evaluation of common variants in 16 genes involved in the regulation of neurotransmitter release in ADHD. *Eur Neuropsychopharmacol* 2013; **23**: 426–435.
- 431 Alarcón M, Cantor RM, Liu J, Gilliam TC, Geschwind DH. Evidence for a language quantitative trait locus on chromosome 7q in multiplex autism families. *Am J Hum Genet* 2002; **70**: 60–71.
- 432 Feng Y, Walsh C a. Protein-protein interactions, cytoskeletal regulation and neuronal migration. *Nat Rev Neurosci* 2001; **2**: 408–416.
- 433 Neale BM, Lasky-Su J, Anney R, Franke B, Zhou K, Maller JB et al. Genome-wide association scan of attention deficit hyperactivity disorder. *Am J Med Genet Part B Neuropsychiatr Genet* 2008; **147B**: 1337–1344.
- 434 Derecki NC, Cardani AN, Yang CH, Quinnes KM, Crihfield A, Lynch KR et al. Regulation of learning and memory by meningeal immunity: a key role for IL-4. *J Exp Med* 2010; **207**: 1067–1080.
- 435 Kipnis J, Gadani S, Derecki NC. Pro-cognitive properties of T cells. *Nat Rev Immunol* 2012; **12**: 663–669.
- 436 Marin I, Kipnis J. Learning and memory ... and the immune system. *Learn Mem* 2013; **20**: 601–606.
- 437 Derecki NC, Privman E, Kipnis J. Rett syndrome and other autism spectrum disorders–brain diseases of immune malfunction? *Mol Psychiatry* 2010; **15**: 355–363.
- 438 Louveau A, Smirnov I, Keyes TJ, Eccles JD, Rouhani SJ, Peske JD et al. Structural and functional features of central nervous system lymphatic vessels. *Nature* 2015; **523**: 337–341.
- 439 Brynskikh A, Warren T, Zhu J, Kipnis J. Adaptive immunity affects learning behavior in mice. *Brain Behav*

- Immun* 2008; **22**: 861–869.
- 440 Ron-Harel N, Segev Y, Lewitus GM, Cardon M, Ziv Y, Netanel D et al. Age-dependent spatial memory loss can be partially restored by immune activation. *Rejuvenation Res* 2008; **11**: 903–913.
- 441 Cohen H, Ziv Y, Cardon M, Kaplan Z, Matar MA, Gidron Y et al. Maladaptation to mental stress mitigated by the adaptive immune system via depletion of naturally occurring regulatory CD4+CD25+ cells. *J Neurobiol* 2006; **66**: 552–563.
- 442 Ziv Y, Ron N, Butson A, Landa G, Sudai E, Greenberg N et al. Immune cells contributed to the maintenance of neurogenesis and spatial learning abilities in adulthood. *Nat. Neurosci.* 2007; **1**: 1–8.
- 443 Tracey KJ. Reflex control of immunity. *Nat Rev Immunol* 2009; **9**: 418–428.
- 444 Smith SD, Kimberling WJ, Pennington BF. Screening for multiple genes influencing dyslexia. *Read Writ* 1991; **3**: 285–298.
- 445 Gilger JW, Pennington BF, Harbeck RJ, DeFries JC, Kotzin B, Green P et al. A twin and family study of the association between immune system dysfunction and dyslexia using blood serum immunoassay and survey data. *Brain Cogn* 1998; **36**: 310–333.
- 446 Grant EG. Zinc deficiency in children with dyslexia: concentrations of zinc and other minerals in sweat and hair. *Br Med J* 1988; **296**.
- 447 Buonincontri R, Bache I, Silahatoglu A, Elbro C, Nielsen AM, Ullmann R et al. A cohort of balanced reciprocal translocations associated with dyslexia: identification of two putative candidate genes at DYX1. *Behav Genet* 2011; **41**: 125–133.
- 448 Becker J, Czamara D, Hoffmann P, Landerl K, Blomert L, Brandeis D et al. Evidence for the involvement of ZNF804A in cognitive processes of relevance to reading and spelling. *Transl Psychiatry* 2012; **2**: e136.
- 449 Xu X, Breen G, Chen C-K, Huang Y-S, Wu Y-Y, Asherson P. Association study between a polymorphism at the 3'-untranslated region of CLOCK gene and attention deficit hyperactivity disorder. *Behav Brain Funct* 2010; **6**: 48.
- 450 Levenson CW, Morris D. Zinc and Neurogenesis: Making New Neurons from Development to Adulthood. *Adv Nutr An Int Rev J* 2011; **2**: 96–100.
- 451 Hariri M, Azadbakht L. Magnesium, iron, and zinc supplementation for the treatment of attention deficit hyperactivity disorder: A systematic review on the recent literature. *Int. J. Prev. Med.* 2015; **2015–Sept**. doi:10.4103/2008-7802.164313.
- 452 Grant EG. Developmental dyslexia and zinc deficiency. *www.thelancet.com* 2004; **364**.
- 453 Salehi B, Mohammadbeigi A, Sheykholeslam H, Moshiri E, Dorreh F. Omega-3 and Zinc supplementation as complementary therapies in children with attention-deficit/hyperactivity disorder. *J reasearch pramacy Pract* 2006; **5**: 22–6.
- 454 Lancaster MA, Gleeson GJ. The primary cilium as a cellular signaling center: lessons from disease. *Curr Opin Genet Dev* 2009; **19**: 220–229.
- 455 Higginbotham H, Eom TY, Mariani LE, Bachleda A, Hirt J, Gukassyan V et al. Arl13b in Primary Cilia Regulates the Migration and Placement of Interneurons in the Developing Cerebral Cortex. *Dev Cell* 2012; **23**: 925–938.

- 456 Bay SN, Caspary T. What are those cilia doing in the neural tube? *Cilia* 2012; **1**: 19.
- 457 Lee JH, Gleeson JG. The role of primary cilia in neuronal function. *Neurobiol. Dis.* 2010; **38**: 167–172.
- 458 Amador-Arjona A, Elliott J, Miller A, Ginbey A, Pazour GJ, Enikolopov G et al. Primary cilia regulate proliferation of amplifying progenitors in adult hippocampus: implications for learning and memory. *J Neurosci* 2011; **31**: 9933–44.
- 459 Lunyak V V, Prefontaine GG, Núñez E, Cramer T, Ju B-G, Ohgi KA et al. Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. *Science* 2007; **317**: 248–51.
- 460 Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet* 2003; **33 Suppl**: 228–37.
- 461 Shi L. Dock protein family in brain development and neurological disease. *Commun. Integr. Biol.* 2013; **6**. doi:10.4161/cib.26839.
- 462 Araki E, Tsuboi Y, Daechsel J, Milnerwood A, Vilarino-Guell C, Fujii N et al. A Novel DCTN1 mutation with late-onset parkinsonism and frontotemporal atrophy. *Mov Disord* 2014; **29**: 1201–1204.
- 463 Hu YF, Caron MG, Sieber-blum M. Norepinephrine transport-mediated gene expression in noradrenergic neurogenesis. *BMC Genomics* 2009; **10**: 151.
- 464 Flint J. The genetic basis of cognition. *Brain* 1999; **122 ( Pt 1)**: 2015–2032.
- 465 Bishop DVM. The interface between genetics and psychology: lessons from developmental dyslexia. *Proc Biol Sci* 2015; **282**. doi:10.1098/rspb.2014.3139.