



CENTRO INTERNACIONAL DE ESTUDOS
DE DOUTORAMENTO E AVANZADOS
DA USC (CIEDUS)

TESIS DOCTORAL

**GENÓMICA DE LOS TRASTORNOS
PSIQUIÁTRICOS: ANÁLISIS DE LA
ARQUITECTURA GENÉTICA COMPARTIDA
MEDIANTE MODELOS DE RIESGO POLIGÉNICO**

Xaquín Gurriarán Bas

ESCOLA DE DOUTORAMENTO INTERNACIONAL
PROGRAMA DE DOUTORAMENTO EN MEDICINA MOLECULAR

SANTIAGO DE COMPOSTELA

2019

DECLARACIÓN DEL AUTOR DE LA TESIS

Genómica de los trastornos psiquiátricos: análisis de la arquitectura genética compartida mediante modelos de riesgo poligénico

D./Dña. Xaquín Gurriarán Bas

Presento mi tesis, siguiendo el procedimiento adecuado al Reglamento, y declaro que:

- 1) La tesis abarca los resultados de la elaboración de mi trabajo.*
- 2) En su caso, en la tesis se hace referencia a las colaboraciones que tuvo este trabajo.*
- 3) La tesis es la versión definitiva presentada para su defensa y coincide con la versión enviada en formato electrónico.*
- 4) Confirmo que la tesis no incurre en ningún tipo de plagio de otros autores ni de trabajos presentados por mí para la obtención de otros títulos.*

En Santiago de Compostela, 25 de abril de 2019

Fdo. Xaquín Gurriarán Bas

AUTORIZACIÓN DEL DIRECTOR / TUTOR DE LA TESIS

**Genómica de los trastornos psiquiátricos: análisis de la arquitectura
genética compartida mediante modelos de riesgo poligénico**

D./Dña. Javier Costas Costas

D./Dña. Ángel María Carracedo Álvarez

INFORMA/N:

*Que la presente tesis, corresponde con el trabajo realizado por D/Dña **Xaquín Gurriarán Bas**, bajo mi dirección, y autorizo su presentación, considerando que reúne los requisitos exigidos en el Reglamento de Estudios de Doctorado de la USC, y que como director de ésta no incurriré en las causas de abstención establecidas en Ley 40/2015.*

En Santiago de Compostela., 25 de abril de 2019

Fdo. Javier Costas Costas

Fdo. Ángel María Carracedo Álvarez

El autor declara no tener ningún conflicto de interés en relación con la tesis doctoral.



AGRADECIMIENTOS

En primer lugar, quiero agradecer a mis dos Directores, Javier Costas Costas y Ángel María Carracedo Álvarez, por la oportunidad de haber realizado esta tesis bajo su dirección. Su apoyo, tiempo y rigor científico han servido para realizar el presente trabajo, en un tiempo donde hacer investigación científica no es el camino fácil.

Esta tesis fue realizada con la ayuda de un contrato predoctoral de la Xunta de Galicia (2014).

Fue financiada con los proyectos 2014/075 del Plan Nacional sobre Drogas (Ministerio de Sanidad, Servicios Sociales e Igualdad), ISCIII/PI2014 (Instituto de Salud Carlos III, cofundado por el FEDER), CP11/00163 (Instituto de Salud Carlos III, cofundado por el FEDER) y la Fundación María José Jove.

Además, quiero agradecer a todas las personas pertenecientes a equipos o grupos de investigación/clínicos que participaron en mayor o menor medida en los distintos estudios que conforman esta tesis doctoral. Así mismo, los resultados fueron posibles gracias a la colaboración con las plataformas tecnológicas del Centro Nacional de Genotipado del Instituto de Salud Carlos III (CeGen-ISCIII), la Fundación Pública Galega de Medicina Xenómica (FPGMX) y del Centro de Supercomputación de Galicia (CESGA).

A todas las personas que, desde el primer momento y durante toda mi vida, me enseñaron y apoyaron para no dejar de preguntar y preguntarme “¿Esto que es?”

ÍNDICE

ÍNDICE.....	i
ABREVIATURAS	ix
RESUMEN	xiii
1. INTRODUCCIÓN.....	3
1.1 NATURALEZA Y NOSOLOGÍA DE LOS TRASTORNOS PSIQUIÁTRICOS	3
1.1.1 Introducción	3
1.1.2 Epidemiología psiquiátrica.....	3
1.1.3 La evolución del sistema de clasificación diagnóstico.....	5
1.1.4 Flaquezas del sistema de clasificación diagnóstico.....	8
1.1.4.1 Comorbilidad psiquiátrica	9
1.1.4.2 Fronteras entre normalidad y trastorno.....	10
1.1.4.3 Inestabilidad diagnóstica	11
1.1.4.4 Sistema categórico vs sistema dimensional.....	12
1.2 GENÉTICA Y GENÓMICA PSIQUIÁTRICA	14
1.2.1 Introducción	14
1.2.1.1 Heredabilidad	15
1.2.1.2 Herencia simple vs herencia compleja.....	18
1.2.1.3 Inicios de la investigación en genética psiquiátrica.....	20
1.2.2 Estudios de ligamiento	21
1.2.3 Genes candidato	23
1.2.3.1 Introducción.....	23
1.2.3.2 Bases metodológicas y estadísticas de la asociación...	24

1.2.3.3 Resultados de la estrategia	28
1.2.4 Estudios de asociación de genoma completo.....	32
1.2.4.1 Introducción	32
1.2.4.2 Capturando la variación genética común	34
1.2.4.3 Desequilibrio de ligamiento	35
1.2.4.4 Diseño del estudio	37
1.2.4.5 Imputación de variantes	40
1.2.4.6 Correcciones para test múltiples	42
1.2.4.7 Resultados de la estrategia GWAS	44
1.2.4.8. Variación en el número de copias	51
1.2.4.9 Pleiotropía y el modelo continuo del neurodesarrollo ..	54
1.2.4.10 Heredabilidad perdida	60
1.2.5 Estudios de secuenciación	62
1.2.6 Puntuaciones de riesgo poligénico	68
1.2.6.1 Introducción	68
1.2.6.2 Origen de la predicción de riesgo	69
1.2.6.3 Consideraciones: tamaño muestral, calidad y estándares	70
1.2.6.4 El cálculo de puntuaciones de riesgo poligénico	73
1.2.6.4.1 Selección de variantes	74
1.2.6.4.2 Desequilibrio de ligamiento	74
1.2.6.4.3 Cálculo y evaluación de PRS	76
1.2.6.4.4 Metodologías PRS	78
1.2.6.5 PRS, una aproximación genómica en crecimiento.....	79

1.2.6.5.1 “La variación poligénica común contribuye al riesgo de esquizofrenia y trastorno bipolar” (Purcell et al. 2009)	80
1.2.6.5.2 Análisis del Cross-Disorder Group del PGC	81
1.2.6.6 Correlación genética, otra aproximación en el análisis de solapamiento entre rasgos complejos	83
2. JUSTIFICACIÓN, HIPÓTESIS Y OBJETIVOS.....	87
2.1 JUSTIFICACIÓN E HIPÓTESIS.....	87
2.2 OBJETIVOS	89
3. MATERIALES Y MÉTODOS.....	93
3.1 MUESTRAS	93
3.1.1 Trastorno obsesivo-compulsivo	93
3.1.2 Esquizofrenia.....	94
3.1.3 Trastorno por abuso o dependencia de sustancias.....	95
3.1.4 Trastorno por dependencia de alcohol	96
3.2 GENOTIPADO.....	98
3.3 CONTROL DE CALIDAD DE LOS DATOS GENOTIPADOS	100
3.4 IMPUTACION DE VARIANTES	102
3.5 MODELOS DE RIESGO POLIGÉNICO	102
3.5.1 Muestras descubrimiento	103
3.5.1.1 Muestras descubrimiento del PGC	103
3.5.1.2 Muestras descubrimiento adicionales.....	106
3.5.2 SNPs a incluir en el modelo	107
3.5.3 Generación de los modelos de riesgo poligénico.....	109

3.5.3.1 Estudio 1. Análisis de la arquitectura genética compartida entre el trastorno obsesivo-compulsivo y la esquizofrenia	109
3.5.3.2 Estudio 2. Análisis de la arquitectura genética compartida entre el trastorno por abuso o dependencia de sustancias y los principales trastornos psiquiátricos.	110
3.5.3.3 Estudio 3. Análisis de la capacidad predictiva de las PRS en la modelización combinada con factores predictivos de riesgo conocidos.....	112
3.5.3.4 Estudio 4. Análisis de la capacidad predictiva de modelos poligénicos fundamentados en la selección de variantes en base a criterios biológicos funcionales	113
3.6 ANÁLISIS ESTADÍSTICOS	113
3.6.1 Análisis de regresión.....	113
3.6.2 Estudio 1. Análisis de la arquitectura genética compartida entre el trastorno obsesivo-compulsivo y la esquizofrenia.....	114
3.6.3 Estudio 2. Análisis de la arquitectura genética compartida entre el trastorno por abuso o dependencia de sustancias y los principales trastornos psiquiátricos	115
3.6.4 Estudio 3. Análisis de la capacidad predictiva de las prs en la modelización combinada con factores predictivos de riesgo conocidos	115
3.6.5 Estudio 4. Análisis de la capacidad predictiva de modelos poligénicos fundamentados en la selección de variantes en base a criterios biológicos funcionales	116
3.6.6 Ajustes para contrastes múltiples	116
3.6.7 Correlación	117
3.6.8 Combinación de modelos de riesgo poligénico	117
3.6.9 Evaluación de modelos de predicción	118

4. RESULTADOS	123
4.1 ESTUDIO 1. ANÁLISIS DE LA ARQUITECTURA GENÉTICA COMPARTIDA ENTRE EL TRASTORNO OBSESIVO-COMPULSIVO Y LA ESQUIZOFRENIA	123
4.1.1 Genotipado y control de calidad	124
4.1.2 Análisis principal: esquizofrenia vs trastorno obsesivo- compulsivo	125
4.1.3 Análisis secundario: trastorno obsesivo-compulsivo vs esquizofrenia	126
4.2 ESTUDIO 2. ANÁLISIS DE LA ARQUITECTURA GENÉTICA COMPARTIDA ENTRE EL TRASTORNO POR ABUSO/DEPENDENCIA DE SUSTANCIAS Y LOS PRINCIPALES TRASTORNOS PSIQUIÁTRICOS	128
4.2.1 Genotipado y control de calidad	128
4.2.2 Análisis predictivos	130
4.3 ESTUDIO 3. ANÁLISIS DE LA CAPACIDAD PREDICTIVA DE LAS PRS EN LA MODELIZACIÓN COMBINADA CON FACTORES PREDICTIVOS DE RIESGO CONOCIDOS	141
4.3.1 Genotipado y control de calidad	141
4.3.2 GWAS Fase 1 abuso/dependencia de alcohol.....	145
4.3.3 Combinación de modelos de riesgo	149
4.3.4 PRS vs variables independientes de riesgo conocido.....	150
4.3.5 Evaluación de modelos de predicción con NRI e IDI.....	154
4.3.6 Análisis de correlación	154
4.3.7 Mediación entre PRS y trauma/factores de la personalidad	156
4.4 ESTUDIO 4. ANÁLISIS DE LA CAPACIDAD PREDICTIVA DE MODELOS POLIGÉNICOS FUNDAMENTADOS EN LA	

SELECCIÓN DE VARIANTES EN BASE A CRITERIOS BIOLÓGICOS FUNCIONALES	158
4.4.1 Preparación del archivo diana.....	158
4.4.2 Modelos de riesgo.....	158
4.4.3 Análisis predictivos	161
5. DISCUSIÓN	165
5.1. DISCUSIONES INDIVIDUALES.....	165
5.1.1 Estudio 1. Análisis de la arquitectura genética compartida entre el trastorno obsesivo-compulsivo y la esquizofrenia.....	165
5.1.2 Estudio 2. Análisis de la arquitectura genética compartida entre el trastorno por abuso o dependencia de sustancias y los principales trastornos psiquiátricos	171
5.1.3. Estudio 3. Análisis de la capacidad predictiva de las PRS en la modelización combinada con factores predictivos de riesgo conocidos	177
5.1.4 Estudio 4. Análisis de la capacidad predictiva de modelos poligénicos fundamentados en la selección de variantes en base a criterios biológicos funcionales	181
5.2 DISCUSIÓN GENERAL	186
5.2.1 Arquitectura genética compartida.....	186
5.2.2 Nosología psiquiátrica	191
5.2.2.1 Factor general de la psicopatología: Una visión dimensional	191
5.2.2.2 El DSM y las fronteras nosológicas.....	194
5.2.3 PRS: Mejoras y futuro	196
5.2.4 Traslación a la clínica de las PRS.....	198
5.2.4.1 Traslación en el ámbito de la psiquiatría	199

5.2.4.2 Implementación clínica de las PRS en psiquiatría: Implicaciones éticas, barreras y dificultades	204
6. CONCLUSIONES.....	211
7. BIBLIOGRAFÍA.....	217
8. ANEXO	311





ABREVIATURAS

ADHD-2017	Muestra descubrimiento trastorno déficit de atención e hiperactividad
ADN	Ácido desoxirribonucleico
ALC	Muestra descubrimiento abuso o dependencia de alcohol
ANX	Muestra descubrimiento ansiedad
ASD-2017	Muestra descubrimiento trastornos del espectro autista
AUC-ROC	Área bajo la curva característica operativa del receptor (Area Under Curve - Receiver Operating Characteristic)
AUDIT	Cuestionario de Identificación de los Trastornos debidos al Consumo de Alcohol (Alcohol Use Disorders Identification Test)
BP-CROSS	Muestra descubrimiento trastorno bipolar
CDCV	Enfermedad común — Variante común (Common Disease – Common Variant)
CDRV	Enfermedad común — Variante rara (Common Disease – Rare Variant)
CeGen-ISCI3	Centro Nacional de Genotipado – Instituto de Salud Carlos III
CESGA	Centro de Supercomputación de Galicia
CIE	Clasificación Internacional de Enfermedades
<i>cis</i> -eQTLs	<i>Loci</i> de caracteres cuantitativos de expresión en <i>cis</i> (<i>cis</i> -expression quantitative trait loci)
CNVs	Variación en el número de copias (Copy Number Variants)
CROSS	Muestra descubrimiento Cross-Disorder-PGC
CTQ	Escala de acontecimientos traumáticos en la infancia (Childhood Trauma Questionnaire)
CTQ EA	Subescala CTQ de abuso emocional
CTQ EN	Subescala CTQ de negligencia emocional
CTQ PA	Subescala CTQ de abuso físico
CTQ PN	Subescala CTQ de negligencia física

CTQ	Subescala CTQ de abuso sexual
DSM	Manual diagnóstico y estadístico de los trastornos mentales (Diagnostic and Statistical Manual)
FDR	Tasa de descubrimientos falsos (False Discovery Rate)
FWER	Tasa de error familiar (Family-Wise Error Rate)
GWAS	Estudios de asociación de genoma completo (Genome-wide Associations Studies)
HiTOP	Hierarchical Taxonomy of Psychopathology
IBD	Identidad por descendencia (Identity-By-Descent)
IDIS	Instituto de Investigación Sanitaria de Santiago de Compostela
ISC	International Schizophrenia Consortium
K.SADS-PL	Programa para trastornos afectivos y esquizofrenia en los niños en edad escolar – versión para situación actual y toda la vida (Kids Schedule for affective Disorders and Schizophrenia for School-age Children - Present and Lifetime Version)
LCRs	Repeticiones de bajo número de copias (Low Copy Repeats)
LD	Desequilibrio de ligamiento (Linkage Disequilibrium)
LEAD	Evaluación longitudinal realizada por expertos y utilizando todos los datos disponibles (Longitudinal Expert All Data)
LTE	Escala de acontecimientos vitales estresantes (List of Threatening Experiences)
MDD-CROSS	Muestra descubrimiento trastorno depresivo mayor
MDD2	Muestra descubrimiento trastorno depresivo mayor
MDS	Escalamiento multidimensional (Multidimensional Scaling)
MGS	Molecular Genetics of Schizophrenia
MHC	Complejo principal de histocompatibilidad (Major Histocompatibility Complex)
NAHR	Recombinación homóloga no alélica (Non-Allelic Homologous Recombination)
NEO PI-R	Inventario de Personalidad NEO Revisado
NEO-FFI	Inventario NEO reducido de Cinco Factores
NGS	Secuenciación masiva o de nueva generación (Next-Generation Sequencing)

NIAAA	National Institute on Alcohol Abuse and Alcoholism
NRI	Mejora de reclasificación neta (Net Reclassification Improvement)
nsSNPs	Polimorfismos de un único nucleótido no sinónimos (Non-Synonymous Single Nucleotide Polymorphisms)
OCD	Muestra descubrimiento trastorno obsesivo-compulsivo
OR	Odds Ratio
PGC	Psychiatric Genomics Consortium
PRS	Puntuaciones de riesgo poligénico (Polygenic Risk Scores)
RDoC	Research Domain Criteria
RFLPS	Fragmentos de restricción (Restriction Fragment Length Polymorphisms)
SCID-I	Entrevista Clínica Estructurada para los Trastornos del Eje I del DSM-IV (Structured Clinical Interview for DSM-IV Axis I Disorders - Clinician Version)
SCZ-5100	Muestra diana esquizofrenia
SCZ-CROSS	Muestra descubrimiento esquizofrenia
SCZ2	Muestra descubrimiento esquizofrenia
SNiPA	Single Nucleotide Polymorphisms Annotator
SNPs	Polimorfismos de un único nucleótido (Single Nucleotide Polymorphisms)
SNVs	Variaciones de un único nucleótido <i>de novo</i> (Single Nucleotide Variants)
SRTs	Repeticiones en tándem cortas (Short Tandem Repeats)
SUD	Trastorno por abuso o dependencia de sustancias (Substance Use Disorder)
WES	Secuenciación de exoma (Whole-Exome Sequencing)
WGS	Secuenciación de genoma completo (Whole-Genome Sequencing)
WTCCC	Wellcome Trust Case Control Consortium
xMHC	Complejo principal de histocompatibilidad extendido



RESUMEN

Los trastornos psiquiátricos son heredables, están influidos por el ambiente y constituyen uno de los conjuntos de enfermedades comunes con mayor desconocimiento de su biología. La aproximación diagnóstica se asienta en un sistema que delimita las fronteras entre trastornos y con la variación normal. El sistema, útil y fiable, muestra flaquezas. Estas están representadas por las altas tasas de comorbilidad, la inestabilidad diagnóstica, la ausencia de una nosología dimensional o espectral y la falta de validez biológica.

El desentrañamiento de la arquitectura genética de los trastornos mentales, basado en el estudio de la variación genética común mediante estudios de asociación de genoma completo (GWAS), señala hacia una arquitectura poligénica compuesta por un gran número de variantes de pequeño efecto. Los modelos de riesgo poligénico aprovechan el peso de la variación genética común y emplean los resultados de GWAS con el fin de generar puntuaciones de riesgo individuales en un conjunto muestral independiente. Estos han permitido observar el extenso solapamiento genético entre trastornos y la relevancia del efecto aditivo de las variantes comunes involucradas en masa, mayor que el basado en variación rara o las variantes comunes significativas a nivel genómico.

El presente proyecto de tesis analizó, mediante cuatro estudios, las relaciones poligénicas entre trastornos psiquiátricos: el solapamiento de las arquitecturas, su potencial uso clínico y mecanismos biológicos subyacentes. Con ese fin, un total de 2563 individuos (1245 casos y 1318 controles) fueron genotipados en tres muestras independientes. Estas conformaron las muestras diana en las que se estimaron puntuaciones de riesgo poligénico individuales, empleando los modelos generados con datos de distintos trastornos (principalmente derivados de megaaanálisis caso – control del Psychiatric Genomics Consortium [PGC], con tamaños muestrales variables, con $N_{\text{min.}} - \text{máx.} = 11411 - 173005$).

El primer estudio exploró el riesgo poligénico compartido entre el trastorno obsesivo-compulsivo y la esquizofrenia. El análisis principal incluyó, como muestra genotipada diana, 917 individuos (433 pacientes con trastorno obsesivo-compulsivo y 484 controles) y datos del mayor GWAS de esquizofrenia realizado por el PGC como muestra descubrimiento, base para la generación de modelos. El secundario, empleó los datos de trastorno obsesivo-compulsivo para la generación de un modelo poligénico aplicado en una muestra independiente de esquizofrenia. Ambos análisis confirmaron el solapamiento parcial de la genética de predisposición de los dos trastornos, llegando a explicar un 3,7% de la varianza en la escala fenotípica. Estos resultados representan la primera evidencia descrita de una arquitectura genética solapante entre ambos trastornos, que podría ser parcialmente responsable de su frecuente comorbilidad. Además, los resultados se incorporan al debate nosológico sobre la validez de un subtipo de pacientes esquizo-obsesivos.

El segundo estudio buscó analizar la arquitectura poligénica compartida entre el trastorno por abuso o dependencia de sustancias y los principales trastornos psiquiátricos con datos disponibles (esquizofrenia, trastorno bipolar, trastorno por déficit de atención e hiperactividad, trastorno del espectro autista, trastorno depresivo mayor y trastorno de ansiedad generalizada). Los distintos análisis realizados incluyeron, como muestra genotipada diana, 1176 individuos (572 pacientes con criterios de dependencia de una sustancia y abuso/dependencia de otra/s y 604 controles). Las muestras descubrimiento para la generación de modelos de riesgo fueron ocho y de acceso abierto (en su mayoría aportadas por el PGC): una por cada uno de los seis trastornos psiquiátricos comparados (aproximadamente del mismo tamaño), una conformada por el sumatorio de cinco trastornos analizados (Cross-Disorder-PGC) y otra del mayor GWAS realizado en esquizofrenia. Los resultados indicaron una susceptibilidad compartida en mayor grado con esquizofrenia respecto al resto de trastornos, siendo la única relación significativa. El porcentaje de varianza explicada en la escala fenotípica alcanzó el 3,92%. Este riesgo compartido es independiente de la comorbilidad psiquiátrica de la muestra diana. Los resultados obtenidos con la muestra Cross-Disorder-

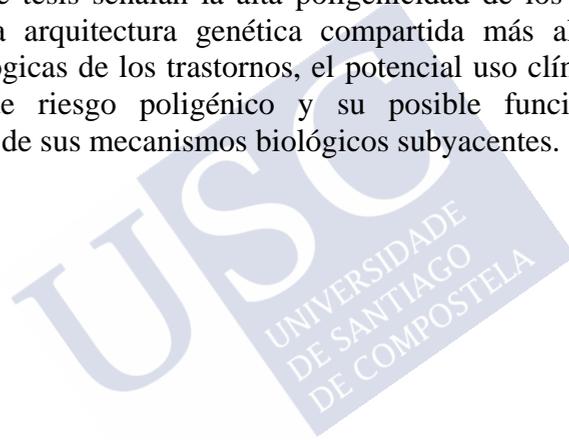
PGC señalan el menor solapamiento del resto de trastornos mentales. En el conjunto de pacientes con trastornos por abuso o dependencia de alcohol, las mujeres mostraron una mayor carga poligénica de predisposición respecto a los hombres, probablemente debido a la mayor presión social contra el uso excesivo de sustancias en mujeres.

El tercer estudio examinó la capacidad predictiva de la modelización combinada de factores predictivos de riesgo conocidos junto a puntuaciones de riesgo poligénico. Con este fin, la muestra genotipada diana de 470 individuos (240 pacientes con trastorno por dependencia de alcohol y 230 controles), incluyó información clínica (comorbilidad psiquiátrica/adictiva, historia clínica familiar, etc.), psicológica (factores de la personalidad), patrón de consumo, experiencias vitales actuales estresantes y traumáticas de la infancia. El modelo poligénico predictivo se basó en la modelización combinada, mediante una suma ponderada, de las puntuaciones de riesgo generadas a partir del mayor GWAS de esquizofrenia realizado por el PGC y a partir de la muestra de controles y casos con trastorno por abuso o dependencia de sustancias empleada en el segundo estudio. La evaluación de la capacidad predictiva conjunta (genética y predictores de riesgo conocidos) señaló el incremento predictivo, no redundante, aportado por las puntuaciones de riesgo poligénico. Este, aunque limitado por el conocimiento parcial de la genética de los trastornos psiquiátricos, evidencia el potencial de la predicción de riesgo y el valor del uso conjunto de fenotipos psiquiátricos relacionados genéticamente. Además, los análisis de este estudio indicaron la relación existente entre la carga poligénica y algunas variables de riesgo conocido, interpretada como una mediación pasiva en el caso de la negligencia emocional traumática sufrida en la infancia y activa en dos factores de la personalidad, neuroticismo y responsabilidad.

Por último, el cuarto estudio buscó explorar la capacidad predictiva de un modelo poligénico asentado en la selección de variantes en base a criterios biológicos. Con este fin, los modelos de riesgo generados se fundamentaron en el conjunto de variantes asociadas a la cascada transcripcional inmediata de respuesta a estrés agudo. Los modelos, generados a partir de datos del mayor GWAS de trastorno depresivo

mayor realizado por el PGC, permitieron la generación de puntuaciones de riesgo poligénico en la muestra diana, conformada por el conjunto de los individuos con trastorno por abuso/dependencia de alcohol del segundo y tercer estudio de la presente tesis (1524 individuos, 721 casos y 803 controles). Los resultados exploratorios señalan la posible, pero débil, relación entre el riesgo a desarrollar alcoholismo y las diferencias individuales en la respuesta transcripcional a estrés. La pequeña capacidad predictiva indica la dificultad de aclarar los mecanismos biológicos en el marco de la poligenicidad extendida.

Los resultados expuestos en los distintos estudios presentados en este proyecto de tesis señalan la alta poligenicidad de los trastornos psiquiátricos, la arquitectura genética compartida más allá de las fronteras nosológicas de los trastornos, el potencial uso clínico de las puntuaciones de riesgo poligénico y su posible función en el esclarecimiento de sus mecanismos biológicos subyacentes.



INTRODUCCIÓN





1. INTRODUCCIÓN

1.1 NATURALEZA Y NOSOLOGÍA DE LOS TRASTORNOS PSIQUIÁTRICOS

1.1.1 Introducción

Los trastornos psiquiátricos representan uno de los mayores retos para la Medicina moderna. Muestran una enorme morbilidad, son extremadamente comunes y sus costes personales y sociales son los mayores a nivel global. Su sintomatología es heterogénea y su naturaleza, a pesar de la considerable investigación hasta la fecha, no ha sido desentrañada. Los trastornos mentales son heredables y constituyen uno de los conjuntos de enfermedades comunes con un mayor desconocimiento de su biología.

Debido al pequeño conocimiento adquirido sobre los mecanismos de los trastornos, la aproximación diagnóstica sigue siendo en gran medida meramente descriptiva y sindrómica. El diagnóstico psiquiátrico actual se asienta en un sistema categórico que diferencia entre variación normal y trastorno, que delimita entidades nosológicas entre sí y se apoya en la detección politética de signos y síntomas que tienden a coocurrir. Este proceso permite a los clínicos elegir tratamientos, estimar resultados y comunicarse con el paciente. El concepto de trastorno psiquiátrico ha evolucionado a lo largo de la Historia y su diagnóstico ha sufrido diversos cambios en su aproximación, consideraciones y naturaleza. La metodología actual es resultado de la búsqueda de un remedio ante la falta de fiabilidad diagnóstica, consolidada en el “Manual diagnóstico y estadístico de los trastornos mentales” (DSM).

1.1.2 Epidemiología psiquiátrica

Décadas de investigación en epidemiología psiquiátrica han permitido llevar a cabo la detección de su carga o prevalencia global,

su mortalidad, sus riesgos y causas ambientales y su heredabilidad estimada:

- La magnitud de la prevalencia global y el impacto de la enfermedad mental han sido estimadas mediante numerosas iniciativas (López & Murray 1998; Eaton *et al.* 2008; Whiteford *et al.* 2013; Steel *et al.* 2014; Vigo *et al.* 2016). Su prevalencia anual ronda el 20% y la vital se acerca al 30% (Steel *et al.* 2014) y muestra diferencias de género, origen geográfico y patología (Steel *et al.* 2014; Vigo *et al.* 2016). Además, el cálculo de su carga, estimada mediante diferentes métodos como “años vividos con discapacidad, es enorme, encontrándose cinco patologías psiquiátricas entre las veinte patologías de mayor carga global (Global Burden of Disease Study 2013 Collaborators 2015).

- Los trastornos psiquiátricos son una de las principales causas de mortalidad global (Eaton *et al.* 2008; Walker *et al.* 2015). Los estudios que tratan de dilucidar esta relación han tendido a reconocer un ligero incremento de la misma (Whiteford *et al.* 2013; Walker *et al.* 2015), pero la dificultad de la estimación reside en dos causas principales: la mayoría de los pacientes con una patología mental fallecen por otras causas médicas o ambientales (Whiteford *et al.* 2013) y están expuestos a una mayor prevalencia de “comportamientos adversos para la salud” que contribuyen a las altas estimaciones de patologías comórbidas (Laurson *et al.* 2014).

- El ambiente juega un papel clave en las patologías mentales por lo que la exposición de un individuo a determinadas condiciones de riesgo puede incrementar la probabilidad a desarrollar el trastorno. Los factores de riesgo incluyen riesgos obstétricos o del embarazo (infecciones, malnutrición o metales pesados), origen migratorio, riesgos perinatales (nacimiento prematuro o complicaciones en el parto), entorno de desarrollo en la infancia (urbanidad, pobreza o maltrato) o la exposición a sustancias en la adolescencia (cannabis o estimulantes) entre otros riesgos conocidos (Tsuang 2000; van Os *et al.* 2010; Vassos *et al.* 2016;

Davis *et al.* 2016; Uher & Zwicker 2017). El exposoma de un individuo está conformado por el conjunto de factores ambientales a los que un individuo está expuesto. Análogo al término genoma, este elemento ofrece importantes oportunidades para la comprensión más profunda de la interacción gen-ambiente (Wild 2005; Caspi & Moffit 2006; Guloksuz *et al.* 2018).

1.1.3 La evolución del sistema de clasificación diagnóstico

El origen del DSM reside en la necesidad de un modelo estandarizado operacional como respuesta a las consecuencias de la Segunda Guerra Mundial en la salud mental de los americanos combatientes. La necesidad de un sistema clasificatorio de los trastornos de los veteranos resultó en el documento que supuso el primer paso cara los modelos de clasificación psiquiátrica actual, el Boletín Médico Técnico número 203 del Ejército de los Estados Unidos publicado en el año 1945 (Artigas-Pallarés & Paula-Pérez 2015; Shorter 2015).

Los siguientes en la evolución de la nosología psiquiátrica moderna se basaron en el descontento de la Asociación Americana de Psiquiatría con el sistema en uso y su encargo, en 1948, de la preparación de una nosología oficial sobre el resto de metodologías (Shorter 2015) que finalizaría en la publicación del DSM-I en 1952. A su vez, la Organización Mundial de la Salud incluyó en 1949 una sección de trastornos mentales en la Clasificación Internacional de Enfermedades (CIE) (Rössler 2013). El psicoanálisis y el conductivismo tuvieron una importante influencia en las dos primeras versiones del DSM, donde el uso del término “reacción” sobre “trastorno” estaba extendido (Artigas-Pallarés & Paula-Pérez 2015). El DSM-II, publicado en 1968, presentaba una arquitectura parecida al DSM-I en los principales diagnósticos. La evolución de la psiquiatría y su búsqueda por una nosología global se apoyó en distintas apreciaciones, como las diferencias en las tasas de diagnóstico de la esquizofrenia entre EE.UU. y el Reino Unido. Otro de los puntos de inflexión fue la publicación del

artículo “Establishment of diagnostic validity in psychiatry: its application to schizophrenia” (Robins & Guze 1970). En este trabajo, los autores propusieron un cambio sobre la tradición dominante basada en el psicoanálisis mediante una aproximación basada en la naturaleza de la enfermedad (Hyman 2010; Zachar & Kendler 2017). Robins y Guze fijaron cinco fases para el desarrollo de una clasificación válida en psiquiatría: descripción clínica precisa, delimitación entre trastornos, el empleo de estudios de seguimiento de pacientes, estudios familiares y estudios de laboratorio (Mullins-Sweatt & Widiger 2009). El trabajo realizado durante esta década, con otros ejemplos importantes (Feighner *et al.* 1972), resaltó las diferencias diagnósticas, señaló la necesidad de estandarizar los diagnósticos de una manera global y llevó al desarrollo de las clasificaciones operacionales (Allardyce *et al.* 2007; Owen 2014).

En 1980 el DSM-III fue publicado y su impacto global fue clave para la psiquiatría moderna al decantarse hacia la orientación biológica. Además, el nuevo manual introdujo un gran número de nuevos diagnósticos y redefinió o modificó en profundidad otros ya existentes (Wilson 1993). Pero su notoriedad no se debe a la arquitectura de los diagnósticos, sino a la visión kraepeliana o neo-kraepeliana de la enfermedad mental (Compton & Guze 1995), al enfatizar la diferenciación de las condiciones basándose en sus síntomas, signos y curso natural. El DSM-I y el DSM-II se limitaban a una descripción de cada diagnóstico pero el DSM-III delineó los signos y síntomas específicos de cada diagnóstico y la aproximación por la cual debían combinarse para su establecimiento (Lilienfeld & Treadway 2016). La revolución que supuso el DSM-III se mantiene hasta hoy en día, ya que las posteriores versiones del DSM mantienen el formato y estructuras básicas desarrolladas en esta importante evolución del manual. El cambio se basó en el inicio del fin del psicoanálisis en el núcleo de la psiquiatría y el acercamiento entre psiquiatría y el resto de la Medicina, al incorporar parte de los postulados de Robins y Guze. La decantación hacia la orientación biológica buscaba esclarecer la etiología y fisiopatología de la enfermedad (Artigas-Pallarés & Paula-Pérez 2015).

Las ediciones subsecuentes (DSM-III-R, DSM-IV, DSM-IV-R y DSM-5) fueron un conjunto de cambios y esperadas mejoras que fueron sucediéndose hasta la publicación de su última versión en 2013. A pesar de las modificaciones incluidas, ninguna se acercó a la magnitud del giro del DSM-III y numerosas críticas fueron surgiendo al comparar su utilidad con su validez (Kendell & Jablensky 2003). El impacto de nuevos diagnósticos fue palpable, volviéndose algunos “virtualmente epidémicos” (Shorter 2015). La evolución del DSM también supuso un efecto multiplicador en el número de diagnósticos posibles, ya que en su primera versión presentó 106 trastornos mentales y evolucionó hasta más de 400 en el DSM-5 (Rössler 2013), modificando en parte los umbrales o fronteras que separaban el trastorno de la variación normal. Este hecho fue ampliamente comentado en la comunidad científica por sus consecuencias en la sanidad global, al poder incrementar el número de diagnósticos (Frances 2016).

Las ventajas del sistema actual se basan en la existencia de criterios operativos que aumentan la fiabilidad del diagnóstico psiquiátrico (Frances 2016). Un punto necesario para las decisiones racionales de tratamiento o para la comunicación entre médico y paciente se basa en la existencia de definiciones fiables compartidas de manera extendida (Hyman 2010). Uno de los mayores éxitos del DSM ha sido disminuir la confusión lingüística internacional, contribuyendo ampliamente a la creación de un lenguaje internacional común (van Heugten-ban der Kloet & van Heugten 2015). A nivel educativo, el sistema diagnóstico del DSM es una herramienta de entrenamiento útil para los clínicos en formación que se extiende también al campo de la epidemiología psiquiátrica, ya que sin sistema diagnóstico no sería posible estimar la prevalencia de los trastornos mentales (Frances 2016). Resulta innegable que el DSM ha facilitado importantes avances terapéuticos, pues la diferenciación de grupos de pacientes medianamente homogéneos facilita el manejo o terapias (Artigas-Pallarés & Paula-Pérez 2015). El DSM-5 es un sistema útil para el clínico, ya que proporciona información sobre la probabilidad de recuperación, recaída, guía las decisiones de tratamiento, indica los perfiles de los síntomas o informa sobre el deterioro del paciente (Jablensky 2016).

El equilibrio entre validez y utilidad sustenta la actual nosología psiquiátrica. Este equilibrio y los dos conceptos que lo conforman no están libres de críticas o puntualizaciones sobre la relación entre ambos conceptos (Kendell & Jablensky 2003; First *et al.* 2004; Mullins-Sweatt & Widiger 2009). Mientras la fiabilidad de los diagnósticos psiquiátricos puede mejorar mediante el uso de criterios diagnósticos como el actual, su validez sigue siendo algo incierto.

1.1.4 Flaquezas del sistema de clasificación diagnóstico

La observación de la fenomenología de los trastornos mentales podría poner en duda la categorización empleada en la actualidad (Artigas-Pallarés & Paula-Pérez 2015). Las críticas al sistema nosológico han sido generadas por las altas tasas de comorbilidad psiquiátrica, los fenotipos intermedios, la inestabilidad diagnóstica, la inespecificidad o las fronteras establecidas entre trastornos o entre variación sana y enferma (Owen 2014; Stephan *et al.* 2016; Kotov *et al.* 2017).

El sistema diagnóstico tradicional considera todos los trastornos mentales como categorías, pero la evidencia genética y epidemiológica señala hacia una estructura de la psicopatología de tipo espectral o continua. La falta de un sistema jerárquico de la sintomatología y de la clasificación de los trastornos se suma a las críticas emitidas sobre el sistema actual (Kendler & First 2010; Smoller 2013; Caspi *et al.* 2014; Kotov *et al.* 2017; Smoller *et al.* 2018). Las llamadas de atención y al debate sobre la estructura nosológica llevan sucediéndose desde hace décadas (Kendell 1989). Éstas ponen en duda la deriva del DSM, que basa su evolución en la redefinición de síndromes, estrategia que podría no ayudar a descubrir las etiologías subyacentes y podría dificultar la incorporación de una base científica que module la nosología (Kendler 1990). La heterogeneidad de síntomas en los diagnósticos basados en la aproximación politética conforma otro de los elementos discutidos (Lilienfeld & Treadway 2017), que complica en parte la búsqueda de agentes etiológicos específicos y grupos homogéneos (Watson 2003).

A continuación se desarrollan algunos de estos puntos críticos de la nosología psiquiátrica actual, que incluyen la comorbilidad, el equilibrio y fronteras entre normalidad y trastorno, la inestabilidad diagnóstica o la crítica a las fronteras nosológicas entre trastornos y la evidencia de una posible estructura espectral.

1.1.4.1 Comorbilidad psiquiátrica

La concurrencia de dos o más diagnósticos psiquiátricos en un mismo individuo es frecuente. En el campo de la salud mental, esta situación clínica es ampliamente conocida y analizada en el transcurso de la evolución y cambios guiados por el DSM en distintas poblaciones, edades y trastornos: individuos de EEUU diagnosticados mediante criterios del DSM-III-R (Kessler *et al.* 1994), pacientes australianos diagnosticados mediante criterios del DSM-IV (Andrews *et al.* 2002), centrado en pacientes adolescentes (Kessler *et al.* 2012), analizando diferencias de género (Ottosen *et al.* 2016) o aproximaciones con población global (Kessler *et al.* 2011). Algunos autores resumen la alta comorbilidad mediante “la regla del 50%”: aproximadamente la mitad de los individuos que cumplen con los criterios diagnósticos para un trastorno cumplen con los criterios de diagnóstico para un segundo que se presenta simultáneamente, la mitad de los individuos con dos trastornos cumplen con los criterios para un tercer trastorno y así sucesivamente (Newman *et al.* 1998).

Los sucesivos análisis han demostrado que, la concurrencia de múltiples diagnósticos psiquiátricos es ahora más frecuente que en el pasado (Maj 2005), pudiendo existir diversas causas que explican la deriva. Los patrones de comorbilidad concurrente o secuencial/acumulativa, pueden surgir de distintos factores al margen de una posible arquitectura de riesgo genético o ambiental compartida. Una de las causas probables de este aumento reside en la propia evolución de los manuales diagnósticos, en los que la fragmentación secuencial de la psicopatología en un gran número de trastornos facilita la proliferación de diagnósticos comórbidos que, en palabras de Jablensky, "desdibujan la distinción entre la verdadera comorbilidad

(conurrencia de trastornos etiológicamente independientes) y la comorbilidad espuria” (Caspi *et al.* 2014; Jablensky 2016). La proliferación de categorías diagnósticas en las clasificaciones psiquiátricas recientes presenta un riesgo importante, ya que podrían realizar divisiones no presentes en la naturaleza que fomentarían los poli-tratamientos no justificados (Maj 2005). Como señala Maj, otra de las causas reside en la consecuencia del uso de entrevistas estandarizadas, que permiten identificar distintos aspectos clínicos que en el pasado no serían detectados tras el diagnóstico principal. Además, la amplia ausencia de reglas jerárquicas en los sistemas de diagnóstico favorece o facilita el diagnóstico comórbido (Maj 2005). Por último, cabe indicar que algunas causas de la comorbilidad psiquiátrica son resultado inherente al curso de la enfermedad psiquiátrica. En la patología mental de individuos jóvenes se puede reconocer un potencial de derivación en diferentes diagnósticos a partir de las primeras manifestaciones de la enfermedad (McGorry *et al.* 2014).

Algunas de las causas de los altos índices de comorbilidad pueden ser también el reflejo de una buena práctica clínica, ya que la consideración de posibles diagnósticos secundarios o el reconocimiento de diferentes estadios o extensión de los trastornos, son básicos en el pronóstico del paciente.

1.1.4.2 Fronteras entre normalidad y trastorno

El análisis de las fronteras en psiquiatría no se reduce a las relaciones entre trastornos, sino que se extiende a las fronteras que separan “patología” de “variación normal” (Jablensky 2016). En la práctica clínica, la asunción de un estado mental patológico en un individuo proviene de la aplicación de criterios operacionales que no reflejan parte de la naturaleza del trastorno, ya que la mayor parte del comportamiento humano está localizado en un continuo (van Os *et al.* 2009). El “todo o nada” de la clasificación categórica de los trastornos psiquiátricos se topa frontalmente con “el continuo” (Rössler 2013).

Esta categorización cuestiona la presencia del llamado fenotipo ampliado (Artigas-Pallarés & Paula-Pérez 2015). Este, suele

presentarse entre familiares de primer grado, y representa ese espacio entre trastorno y normalidad. Alguno de los ejemplos que evidencian esta situación de un continuo dividido artificialmente son los casos de la depresión o la ansiedad, ampliamente reconocidos en diferentes grados de severidad (Owen 2014). Otro ejemplo de la extensión del continuo en el marco de los trastornos psiquiátricos reside en los síntomas psicóticos. Las alucinaciones auditivas y el pensamiento paranoico se producen de forma menguada en el 5-8% de la población sana (van Os & Kapur 2009). Las experiencias del espectro psicótico son más comunes en la población general de lo que se pensaba (van Os *et al.* 2009; Stip & Letourneau 2009; Kelleher *et al.* 2012). La incorporación y análisis de este espectro o variación continua sería beneficioso para investigadores del campo de la salud mental, ya que al interrogar esta frontera se podrían obtener nuevas perspectivas sobre la aparición y desarrollo de los trastornos mentales.

1.1.4.3 Inestabilidad diagnóstica

Otro de los fenómenos que pone en cuestión la categorización de los trastornos psiquiátricos es la estabilidad del diagnóstico clínico. Durante el curso clínico de un trastorno psiquiátrico es bastante común que el paciente pueda entrar o salir fácilmente de un diagnóstico (Artigas-Pallarés & Paula-Pérez 2015). El solapamiento de síntomas entre categorías diagnósticas implica dificultades en la asignación de un diagnóstico en una única categoría específica, lo cual puede facilitar cierta volubilidad del diagnóstico (Craddock & Owen 2005).

Existen distintas causas o factores que implican una mayor inestabilidad en el diagnóstico psiquiátrico: la valoración subjetiva de cada síntoma por parte del paciente o el clínico, cambios sustanciales en las características clínicas a lo largo de la enfermedad, modificación de los criterios diagnósticos en el proceso evolutivo del DSM o del CIE (Fennig *et al.* 1994), la alta frecuencia de individuos diagnosticados con un trastorno en la franja límite del diagnóstico (Artigas-Pallarés & Paula-Pérez 2015) o la propia naturaleza del trastorno, ya que algunas entidades clínicas presentan patrones temporales específicos de

síntomas y otras incluyen diferentes polaridades que requieren tiempo para ser reconocidas entre otras características del curso de los trastornos psiquiátricos (Bromet *et al.* 2011). Algunas de estas características son inherentes al sistema diagnóstico actual y podrían reflejar la debilidad o limitaciones de la nosología actual.

1.1.4.4 Sistema categórico vs sistema dimensional

Los múltiples desafíos, como los anteriormente comentados, sitúan al actual sistema categórico en el punto de mira debido a las dudas generadas de su validez para el avance en la investigación psiquiátrica (Owen 2014). ¿Las evidencias nos señalan hacia una arquitectura de la psicopatología con un perfil de mayor continuidad entre entidades? ¿Es posible llegar a conocer la etiología de los trastornos mentales mediante la investigación basada en la nosología actual? La genética y genómica han proporcionado numerosas evidencias sobre las fronteras en la arquitectura genética entre trastornos, que sobrepasan las establecidas por el sistema nosológico actual al no ser tan agudas (Smoller 2013) (esta situación, tema central de la presente Tesis, será desarrollado en profundidad a lo largo del manuscrito). Desde el punto de vista dimensional o espectral, en la nosología actual los conceptos “externalizantes” e “internalizantes” están ampliamente extendidos. Estos, confirmados por análisis factoriales, ofrecen una visión espectral de trastornos relacionados (Kotov *et al.* 2017).

La reevaluación y publicación del DSM de manera consecutiva trajo la esperanza de incluir mejoras y modificaciones sobre la base de nuevos conocimientos de la etiología y patogénesis, pero su comodidad y fiabilidad en el manejo de pacientes impusieron otro enfoque y evolución (Owen 2014). El enfoque se sostuvo en un proceso de “reificación” (Hyman 2010) que se ha prolongado hasta la última edición. De todas formas, en el DSM-5, se puede observar que la asunción de las limitaciones del modelo está presente, como mínimo de manera parcial o modesta. En el texto, los autores indican “la evidencia científica coloca a muchos de los trastornos, si no la mayoría, en un

espectro con trastornos estrechamente relacionados que tienen síntomas y factores de riesgo genéticos y ambientales compartidos”, revelando también “debemos reconocer que los límites entre trastornos son más porosos de lo que se percibía originalmente”. A pesar de esto, la publicación del DSM-5 mantuvo el sistema categórico, por los mismos motivos de eficiencia que las ediciones previas. Las evidencias obtenidas mediante la búsqueda de biomarcadores, la investigación clínica, de la estructura de la psicopatología y la evolución del DSM indican que el DSM podría ser un obstáculo para el progreso de la investigación psiquiátrica, ya que las entidades nosológicas podrían ser constructos artificiales (Craddock & Owen; Bernstein 2011; Kapur *et al.* 2012).

Ante esta situación, la comunidad científica se ha propuesto la inclusión de nuevas estrategias de diagnóstico. Las distintas opciones varían en su naturaleza y pretensiones, pero las distintas iniciativas o soluciones al estancamiento podrían seguir estos tres principios (Owen 2014): se deben incrementar los estudios que no empleen únicamente categorías diagnósticas como se definen en el DSM o CIE, debería prestarse cada vez más atención a las medidas dimensionales de la psicopatología y se deberían analizar en profundidad posibles fenotipos intermedios.

Algunas de estas iniciativas llevan años desarrollándose y planteándose como alternativa al sistema actual. *Research Domain Criteria* (RDoC) (Insel *et al.* 2010) o *Hierarchical Taxonomy Of Psychopathology* (HiTOP) (Kotov *et al.* 2017) son dos de las más importantes. RDoC es un método de clasificación dimensional de procesos psicológicos potencialmente relevantes en los trastornos psiquiátricos partiendo del conocimiento y avance de parámetros biológicos, que plantea el objetivo de definir las dimensiones básicas de las disfunciones que actúan de manera transversal en los distintos trastornos (Insel 2014). Aunque diversos autores ven un futuro prometedor a este cambio de paradigma como solución a la situación actual (Yee *et al.* 2015) otros critican su carácter simplista y reduccionista (Winberger *et al.* 2015). HiTOP es un modelo dimensional que construye síndromes psicopatológicos a partir de la

covariación observada de síntomas, agrupándolos. Además, combina síndromes concurrentes en espectros caracterizando los fenómenos de forma dimensional que permiten analizar los problemas de inestabilidad diagnóstica y de límites entre trastornos (Kotov *et al.* 2017). Las críticas hacia HiTOP residen en “el énfasis excesivo es la psicopatología transversal” así como en “el descuido de las vías dinámicas de desarrollo y los problemas de diagnóstico diferencial”, elementos “relevantes para el tratamiento y el manejo” (Wittchen & Beesdo-Baum 2018). Ambos modelos se encuentran en desarrollo y en distintas etapas (Kotov *et al.* 2018).

1.2 GENÉTICA Y GENÓMICA PSIQUIÁTRICA

1.2.1 Introducción

Nuestro conocimiento de lo que hoy en día llamamos trastornos psiquiátricos ha cambiado profundamente en los últimos doscientos años. Tras aceptar la naturaleza biológica de los trastornos psiquiátricos surge la siguiente pregunta: ¿Hasta qué punto los factores hereditarios contribuyen a la variación en la susceptibilidad de los trastornos psiquiátricos?

En Biología y Medicina, uno de los debates más extendidos se centra en la relación entre ambiente y herencia genética. El fenotipo de un trastorno es el resultado de factores de predisposición genética, factores ambientales y la interacción del genotipo y el ambiente, ya que la expresión génica está regulada parcialmente por las características ambientales a las que se expone un individuo, determinando el fenotipo detectado. La naturaleza genética de los trastornos psiquiátricos fue observada por los clínicos mucho antes de identificar el ácido desoxirribonucleico (ADN) como molécula responsable de la herencia. En la actualidad, sabemos que la mayor parte de la genética de estos trastornos complejos se debe a la acumulación de un gran número de variantes o *loci* de pequeño efecto que influyen a un número sustancial de trastornos al mismo tiempo. El camino llevado a cabo por los investigadores del campo de la salud mental y la genética ha sido largo

y duro, pero comienza a dar sus frutos en la búsqueda de la etiología escondida a pesar del gran número de desafíos existentes.

1.2.1.1 Heredabilidad

En la actualidad, gracias al abaratamiento de las tecnologías de genotipado y secuenciación masiva, podemos responder a un gran número de cuestiones sobre cómo los factores genéticos contribuyen al desarrollo de los trastornos psiquiátricos examinándolos de forma directa (Sullivan *et al.* 2012; Pasaniuc & Price 2017). Pero el marco bajo el que se ha desarrollado esta tesis es reciente. El primer paso llevado a cabo en la investigación de la heredabilidad y estructura de los trastornos mentales se basó en el reclutamiento de familias, hermanos gemelos y poniendo el foco en individuos adoptados. El fin de esta aproximación era y es la estimación de la heredabilidad de los trastornos y su relación con el ambiente. Cada aproximación tiene sus fortalezas y puntos débiles.

Los estudios familiares investigan el grado de acumulación familiar de un trastorno. La concentración de un trastorno en una familia concreta ocurre cuando el riesgo a sufrir el trastorno es mayor en familiares de un paciente que en controles de la población general. Estos estudios no diferencian entre efectos genéticos y ambientales ya que no tienen en cuenta el ambiente familiar compartido, por lo que las aproximaciones más exactas se basan en gemelos y adopciones. La premisa básica de los estudios de gemelos es que, al heredarse todos o la mitad de los genes (monocigóticos o dicigóticos), puede inferirse la contribución genética al fenotipo. Este tipo de estudios permiten también inferir información de factores ambientales. Por ello, la mejor aproximación para la inferencia de la heredabilidad de un rasgo o un trastorno reside en los estudios de adopción, al diferir los individuos analizados en su ambiente de crianza o al modificar el ambiente paterno (de enfermedad presente a libre de enfermedad o viceversa) (McGuffin *et al.* 2003). Las heredabilidades estimadas se suelen expresar en la escala de susceptibilidad y el empleo de distintos tipos de familiares permite ir analizando en distinto grado la carga ambiental y la genética

de manera conjunta e individual (Lichtenstein *et al.* 2009; Visscher *et al.* 2012). Los resultados de estos estudios son en general muy consistentes y replicables, pero pueden mostrar cierto sesgo a la hora de estimar la heredabilidad a partir de conjuntos de individuos que representan un origen concreto y constreñido (sujetos provenientes de un mismo complejo hospitalario, un pequeño tamaño muestral o sesgo de ascendencia entre las causas) (Kendler *et al.* 1995; Visscher *et al.* 2012; Wray & Gottesman 2012; Kendler 2013). Estos inconvenientes pueden ser evadidos con el empleo de registros nacionales, muy extendidos en los países nórdicos, como los resultados obtenidos a partir de registros suecos de 9 millones de individuos (Lichtenstein *et al.* 2009) o 2,5 millones de individuos en Dinamarca (Wray & Gottesman 2012). Este tipo de aproximaciones también muestran algunas limitaciones, pero pueden ser estimaciones más acertadas o representativas en relación con aproximaciones experimentales basadas en grandes cohortes y proyectos colaborativos con pacientes provenientes de colecciones nacionales y clínicas (Visscher *et al.* 2012; Wray & Gottesman 2012). El grado de heredabilidad de cada patología es diferente, con un abanico del 25-90% aproximadamente (Figura 1) (Kendler 2001; Freitag 2007; Bienvenu *et al.* 2011). Este tipo de estudios puso también el foco sobre un posible solapamiento entre patologías, al detectarse numerosas veces en familias la ocurrencia de trastornos psiquiátricos secundarios respecto al estudiado (Smoller *et al.* 2018).

La variación fenotípica en una población es el resultado de los factores genéticos y/o factores ambientales (Falconer & Mckay 1996). La heredabilidad es un concepto estadístico característico de cada población en la que se hayan estimado las varianzas ambientales y genéticas. Las estimaciones de heredabilidad pueden variar entre poblaciones y en función de los factores no genéticos recogidos e incluidos en la estimación (Tenesa & Haley 2013). La heredabilidad se define como “una proporción de varianzas, concretamente como la proporción de la varianza total en una población para una medida en un momento particular, que es atribuible a la variación de los valores genéticos aditivos o genéticos totales” (Visscher *et al.* 2008). La basada en valores genéticos aditivos es conocida como “heredabilidad en

sentido estricto” y la basada en valores genéticos totales como “heredabilidad en sentido amplio”. Este valor puede variar en el tiempo, debido a la evolución y cambios en el efecto de la influencia ambiental sobre un determinado fenotipo que, consecuentemente, afectará a la contribución relativa de la varianza genética a la varianza fenotípica total (Visscher *et al.* 2008; Wray & Visscher 2008).

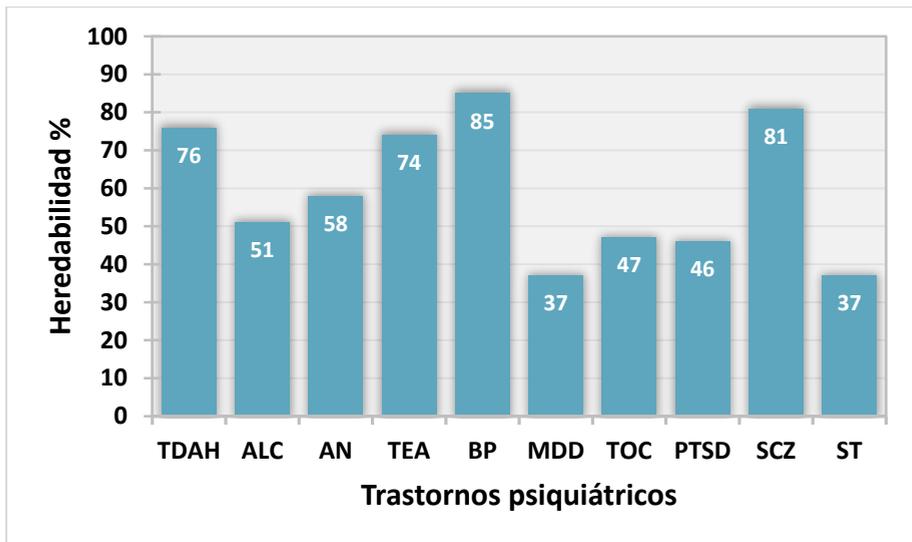


FIGURA 1. HEREDABILIDAD DE LOS TRASTORNOS PSIQUIÁTRICOS. TDAH (trastorno por déficit de atención e hiperactividad), ALC (dependencia de alcohol), AN (anorexia nerviosa), TEA (trastornos del espectro autista), BP (trastorno bipolar), MDD (trastorno depresivo mayor), TOC (trastorno obsesivo-compulsivo), PTSD (trastorno de estrés posttraumático), SCZ (esquizofrenia) y ST (síndrome de Tourette). Datos de Sullivan y Geschwind (Sullivan & Geschwind 2019).

La heredabilidad se calcula mediante el siguiente método (Byers 2008):

$$H^2 = \frac{\sigma_G}{\sigma_P} = \frac{\sigma_G}{\sigma_G + \sigma_A} = \frac{(\sigma_{AD} + \sigma_D + \sigma_E)}{(\sigma_{AD} + \sigma_D + \sigma_E) + (\sigma_{Aesp} + \sigma_{Agen} + \sigma_{GxA})} \cong \frac{\sigma_{AD}}{\sigma_P} = h^2$$

La heredabilidad es la fracción entre la varianza genética (σ_G) y la varianza fenotípica (σ_P). La varianza genética está constituida por el sumatorio de la varianza aditiva (σ_{AD} ; refiere a la desviación fenotípica resultado de la presencia de uno u otro alelo), la varianza de dominancia (σ_D ; refiere a la posible situación de dominancia entre alelos) y la varianza por epistasis o interacción génica (σ_E ; refiere a la interacción entre alelos de distintos *loci*). Debido a que distintos autores han debatido sobre la carga de cada contribución de cada varianza al cálculo de H^2 , la heredabilidad en sentido estricto (h^2) parece el modelo más acertado, donde se considera la varianza aditiva entre la varianza fenotípica (Visscher *et al.* 2008 a). La varianza ambiental (σ_A) está compuesta por la varianza específica de individuo (σ_{Aesp} ; refiere a la varianza menor asociada a factores ambientales individuales), la varianza ambiental general (σ_{Agen} ; refiere a la varianza global asociada a factores ambientales característicos de la población general) y la varianza de interacción genotipo-ambiente ($\sigma_{G \times A}$; refiere a las posibles diferencias en la respuesta al entorno en función del genotipo individual).

1.2.1.2 Herencia simple vs herencia compleja

Las patologías y rasgos humanos pueden presentar distintos patrones de herencia. Estos son debidos a los factores causales involucrados y a la arquitectura genética subyacente. Algunas patologías muestran patrones clásicos de herencia simple o Mendeliana, en los que un único gen es responsable de la patología. En algunos casos el alelo mutado causal es dominante, pero comúnmente es recesivo (Visscher *et al.* 2012). El análisis en familias, con un número suficiente de individuos afectados, puede ser utilizado para determinar la localización de un gen causal y para conocer la dominancia del mismo (Chial 2008). Normalmente hay más de una variante causal, ya que es probable la presencia de diversas mutaciones en distintas variantes que afecten o interfieran con el normal funcionamiento del producto de un mismo gen (Visscher *et al.* 2012). Las variantes de gran efecto son poco frecuentes en las distintas poblaciones, presentan una importante penetrancia y las patologías causadas por las mismas muestran una baja

prevalencia (McCarthy *et al.* 2008; Manolio *et al.* 2009; Robinson *et al.* 2014).

Las patologías comunes, en las que se engloban los trastornos psiquiátricos con la excepción de algunas patologías degenerativas y algunas formas de autismo, no presentan patrones de herencia mendelianos y se consideran patologías complejas (Visscher *et al.* 2012; O'Donovan 2015; Sullivan *et al.* 2018). La complejidad surge cuando la correspondencia entre genotipo y fenotipo no es directa. Esto puede ser debido a que distintos genotipos dan lugar al mismo fenotipo o porque el mismo genotipo da lugar a fenotipos distintos por azar, ambiente o epistasia (Lander & Schork 1994). Cuando la combinación de un genotipo de riesgo y los factores ambientales superan un umbral, el individuo desarrollará la enfermedad o rasgos de la misma (Visscher *et al.* 2012). A pesar de que una pequeña parte de los casos psiquiátricos podrían estar influidos por alelos de herencia simple (O'Donovan 2015), las patologías psiquiátricas muestran una arquitectura poligénica (Sullivan *et al.* 2013a). La penetrancia de las variantes implicadas en el genotipo de riesgo de esta clase de patologías es mucho menor que en el caso de las de herencia mendeliana (McCarthy *et al.* 2008; O'Donovan 2015) pero su efecto reside en el conjunto. De hecho, es común que mutaciones de gran efecto y herencia mendeliana muestren una menor concordancia con el fenotipo esperado o individuos portadores no desarrollen la enfermedad. Esto es debido a una combinación de genotipo y factores ambientales que “modifican” el efecto de la variante de gran tamaño. Por lo tanto, un conjunto de variantes de pequeño efecto junto a unas condiciones ambientales aditivas podrían ser responsables de la ausencia total o parcial del efecto de la variante de gran tamaño (Visscher *et al.* 2012; Mitchel 2012). Otra situación o modelo posible es que la variante de importante efecto precise la acción aditiva de un genotipo de riesgo basado en variantes comunes en un modelo combinado para el desarrollo de la patología, algo probable cuando la variante de mayor efecto sea de penetrancia intermedia o incompleta (Mitchell 2012). Además, existen dos tipos de variantes que podrían estar involucradas en la arquitectura genética de patologías complejas, aquellas con una baja penetrancia y baja frecuencia o raras y de difícil detección, y de forma mucho más inusual e improbable

aquellas variantes comunes de gran penetrancia (McCarthy *et al.* 2008). Hoy en día no cabe duda sobre la poligenicidad generalizada de los trastornos psiquiátricos, pero el camino hasta este punto ha sido largo (Sullivan *et al.* 2018).

1.2.1.3 Inicios de la investigación en genética psiquiátrica

El origen de la genética psiquiátrica actual reside, en primera instancia aunque no de forma directa, en Kraepelin. Su papel en la psiquiatría moderna y su aportación a la nosología (Kendler 2002), la implementación de la práctica clínica en la Alemania de finales del siglo XIX y comienzos del XX con el *German Research Institute for Psychiatry* inaugurado en 1917 marcaron los primeros pasos en la investigación moderna del campo, con Ernst Rüdin como máximo exponente de la rama genética (Propping 2005).

Las leyes de Mendel fueron re-descubiertas en 1900 (de Vries *et al.* 1900) y en pocos años estas fueron reconocidas en humanos. Rüdin fue el primero en proponer el uso sistemático de estudios familiares para confirmar la herencia mendeliana en trastornos psiquiátricos en 1911, publicando el primer estudio familiar en esquizofrenia en 1916. Durante los años 30, Rüdin se unió al nacionalsocialismo y se involucró fuertemente en las doctrinas genético-raciales de la época (Weber 1993). Las ideaciones ultranacionalistas, también extendidas a Kraepelin, se propagaron junto al pensamiento eugenésico de las patologías psiquiátricas en la psiquiatría alemana bajo el desarrollo y crecimiento del nazismo y la segregación racial (Gottesman & Bertelsen 1996). Además del trabajo de Rüdin, otros importantes avances en el estudio de la genética fueron llevados a cabo por Hans Luxenburger, quien en 1928 publicó los resultados del primer estudio sistemático de gemelos en esquizofrenia (Luxenburger 1928). Las aproximaciones mendelianas de estos autores no sirvieron para desentrañar la genética de los trastornos mentales, pero sí para detectar un patrón de segregación más complejo que el esperado y la existencia de un componente ambiental complementario a los factores hereditarios (Luxenburger 1928; Propping 2005; Kendler 2015).

Los trabajos de estos autores fueron seguidos por un auge del enfoque genético en la búsqueda de los factores etiológicos de los trastornos psiquiátricos, la mayoría con aproximaciones mendelianas. Importantes trabajos llevados a cabo por Kallman (Kallman 1938), Slater (Slater 1958), Meehl (Meehl 1962) o Heston (Heston 1966) chocaron de frente con la poligenicidad estructural de la patología, pero asentaron importantes direcciones en la investigación psiquiátrica (Propping 2005; Kendler 2015). Seguidos por la publicación del libro “La Genética de los trastornos Mentales” en 1970 (Slater & Cowie 1970) y el auge los estudios basados en el análisis de registros de adopciones (Propping 2005), la investigación en genética psiquiátrica llegó a la década de los años 70 con dos modelos posibles para la esquizofrenia: un modelo complejo poligénico o un modelo mendeliano (Kendler 2015).

1.2.2 Estudios de ligamiento

Los análisis de ligamiento conformaron la primera metodología de aproximación extendida con el fin de identificar variantes de susceptibilidad en un gran número de trastornos psiquiátricos. Este enfoque tradicional buscaba la localización de genes causales al detectar la asociación entre un marcador polimórfico y el trastorno en una determinada familia (Merikangas & Risch 2003). La distribución de los marcadores a lo largo del genoma permite analizar qué genes están ligados, pudiendo ser varios en función del patrón de recombinación de esa región. Tras el uso extendido y exitoso del posicionamiento clonal para la identificación de genes responsables de trastornos de herencia mendeliana, en el campo de la salud mental se extendió un cierto optimismo. Después de la incertidumbre inicial (Sturt & McGuffin 1985), en la segunda mitad de la década de los años 80 el interés que buscaba replicar ese éxito en los trastornos psiquiátricos se extendió (Burmeister *et al.* 2008)

Inicialmente esta clase de estudios emplearon polimorfismos de longitud de fragmentos de restricción (RFLPS, *Restriction Fragment Length Polymorphisms*) (Cloninger *et al.* 1983) o repeticiones en

tándem cortas (SRTs, *Short Tandem Repeats*). Tras consolidarse y avanzar en el conocimiento de los polimorfismos humanos, éstos se centraron en buscar polimorfismos de un único nucleótido (SNPs, *Single Nucleotide Polymorphisms*) ligados (Merikangas & Risch 2003).

La significación estadística del resultado de un estudio de ligamiento se suele resumir en el valor de una máxima “puntuación lod” (*lod score*). El término “lod” fue acuñado por Barnard para referirse al logaritmo de la *odds ratio* (OR) (Barnard 1949). Es un estimador que permite detectar ligamiento al estimar la frecuencia de recombinación más probable y se calcula de la siguiente forma:

$$\text{Puntuación lod} = \log_{10} \frac{\text{Prob. de un genotipo a una determinada freq. de recombinación } \theta}{\text{Prob. del mismo genotipo con segregación independiente } (\theta = 0,5)}$$

En 1974 se decidió el criterio a satisfacer para considerar significativo el resultado de un análisis de ligamiento: el valor de lod debería ser igual o mayor a 3 (McKusick & Edwards 1975). Los estudios de ligamiento son estudios generadores de hipótesis, ya que no es preciso un conocimiento *a priori* de qué regiones o genes están potencialmente involucrados. Su éxito depende, además de la potencia de análisis (limitada por el tamaño muestral y su grado de homogeneidad) (Altmüller *et al.* 2001), del número de marcadores polimórficos empleados.

Durante los primeros años un importante número de ligamientos de interés fueron detectados en trastorno bipolar o esquizofrenia (Egeland *et al.* 1987; Sherrington *et al.* 1988), pero ninguno de ellos fue replicado (Kennedy *et al.* 1988; Kelsoe *et al.* 1989; Gershon 1991). Debido al creciente escepticismo, muchos autores atribuyeron al diseño del análisis o a la heterogeneidad de las muestras la falta de réplica de resultados de interés (Merikangas & Risch 2003). Ante el aumento del número de estudios y su extensión genómica de análisis, Lander y Kruglyak formalizaron los criterios para determinar un ligamiento significativo a nivel genómico (Lander & Kruglyak 1995) y muy pocos estudios lograron resultados de interés posteriormente replicados. Exceptuando algún resultado de interés en esquizofrenia (Badner & Gershon 2002; Lewis *et al.* 2003) o, por ejemplo, autismo (Yang & Hill 2007), la mayoría de los resultados no han sido consistentes (Owen *et*

al. 2000; Prathikanti & McMahon 2001; Segurado *et al.* 2003; Crow 2007; Yang & Hill 2007). La “era de los análisis de ligamiento” de los trastornos psiquiátricos (1980-2005), no logró identificar ningún *locus* que se replicara inequívocamente en distintas muestras independientes de manera consistente (Burmeister *et al.* 2008).

Años antes del ocaso de los análisis de ligamiento, Rish y Merikangas publicaron un artículo con una gran influencia en el campo de la genética de los trastornos complejos en el que señalaban hacia una arquitectura genética basada en variantes comunes con pequeños efectos (Rish & Merikangas 1996). Ante esta situación, planteaban que los análisis de ligamiento necesitarían miles de muestras para la identificación de *loci* de riesgo. Los años de trabajo de los análisis de ligamiento incorporaron sus resultados de mayor de interés a los análisis de asociación de genes candidatos (Burmeister *et al.* 2008), aproximación ya propuesta por Rish y Merikangas como opción con mayor posibilidad de éxito en patologías complejas.

En la actualidad, los estudios de ligamiento se reducen a un menor número y su utilidad reside en la capacidad de detectar múltiples alelos de riesgo raros. Ante la oleada de estudios de asociación basados en análisis de variantes comunes que será comentada a continuación, los estudios de ligamientos todavía ofrecen algo de luz en la búsqueda de variantes raras que se segregan en linajes concretos. El abaratamiento de las tecnologías de secuenciación masiva con sus múltiples ventajas comienza, por otra parte, a desequilibrar la balanza a la hora de seleccionar una metodología de aproximación.

1.2.3 Genes candidato

1.2.3.1 Introducción

Otra de las aproximaciones desarrolladas en el campo con el fin de dilucidar la etiología de los trastornos mentales son los estudios de asociación de genes candidatos. Ante el lento proceso de los estudios de ligamiento de escaneo de genoma completo, un gran número de estudios se centraron en esta estrategia (Plomin 2005). Además, los

estudios de asociación de genoma completo todavía no habían sido desarrollados, por lo que los estudios se limitaban a analizar un menor número de variantes seleccionadas *a priori* (Smoller 2017), conocidas como variantes asociadas a genes candidatos. Un gen candidato puede ser seleccionado, normalmente, en función de cierto conocimiento previo adquirido mediante estudios de ligamiento, que detectaron regiones cromosómicas de potencial interés, o en base a la evidencia de que el gen en cuestión podría jugar un papel en la fisiopatología del fenotipo (Malhotra & Goldman 1999). Este tipo de análisis se extendió aproximadamente durante una década en psiquiatría, comenzando en 1996 hasta co-existir con los estudios de asociación de genoma completo (GWAS, *Genome-wide Associations Studies*) (Burmeister *et al.* 2008). A diferencia de los estudios de ligamiento, los análisis de asociación buscan la existencia de una asociación simple. Su planteamiento más sencillo consiste en establecer dos conjuntos de individuos, casos y controles, y determinar las frecuencias alélicas y genotípicas de variantes interrogadas en el estudio. En caso de existir una distribución de frecuencias divergente entre ambos grupos, ésta puede ser detectada y calculada con el fin de conocer su riesgo relativo (Zondervan & Cardon 2007; Burmeister *et al.* 2008 Clarke *et al.* 2011).

1.2.3.2 Bases metodológicas y estadísticas de la asociación

Los marcadores frecuentemente analizados en este tipo de estudios son los SNPs. Este marcador genético, que consiste en un único *locus* bialélico con los alelos “A” y “a” con los siguientes genotipos posibles: “AA”, “Aa” y “aa”. La penetrancia de la enfermedad asociada con un genotipo determinado es el riesgo a sufrir la enfermedad en los individuos portadores del genotipo. Los modelos estándar de penetrancia de la enfermedad incluyen modelos multiplicativos, aditivos, recesivos y dominantes (Clarke *et al.* 2011). La asociación debe ser cuantificable mediante algún índice o magnitud que informe del grado de asociación y de su significación estadística. Una medida bastante intuitiva y extendida de la medida de la fuerza de asociación es el riesgo relativo, que compara la penetrancia de la enfermedad entre individuos expuestos a diferentes genotipos. En el caso del

planteamiento más sencillo, caso control, la medida más extendida es el *OR* (razón de dos cocientes). Esta medida informa, en función de la desviación de la unidad, si un alelo está confiriendo riesgo ($OR > 1$) o protección ($OR < 1$) (Clarke *et al.* 2011). La obtención de esta medida se basa en la probabilidad de que el alelo “A” esté presente en el grupo de casos o en el de controles, calculándose de la siguiente forma:

$$OR = \frac{\frac{\text{Casos portadores del alelo A}}{\text{Casos no portadores del alelo A}}}{\frac{\text{Controles portadores del alelo A}}{\text{Controles no portadores del alelo A}}}$$

El valor de esta medida no informa de la significación estadística de la asociación analizada en cuestión. Bajo la hipótesis nula de “no asociación con la enfermedad”, las frecuencias relativas del alelo o genotipo entre el grupo control y el grupo de casos deberían ser las mismas y las diferencias encontradas son debidas al error aleatorio propio de la distribución. La prueba estadística para contrastar la hipótesis nula en los estudios de asociación suele ser la prueba “chi-cuadrado” (χ^2). La hipótesis alternativa sería la existencia de asociación debido a que las diferencias detectadas en las frecuencias no pueden ser explicadas por azar para un nivel de significación α determinado. Otros modelos más complicados son analizados mediante regresión logística, al ser necesario incluir en los análisis covariables de ajuste del rasgo complejo a analizar (Cardon & Bell 2001; Zondervan & Cardon 2007; Clarke *et al.* 2011).

Un resultado significativo en una prueba de asociación rara vez implica que el SNP analizado influya directamente en el riesgo a sufrir el trastorno, enfermedad o rasgo. La asociación puede ser directa, indirecta o espuria.

- Asociación directa: Este tipo de asociación, también conocida como causal, se presenta cuando los alelos del *locus* analizado están relacionados con la etiología de la enfermedad. En esta condición, la vía de actuación será normalmente una ruta biológica de desarrollo de la enfermedad. La asociación causal

de este tipo de variantes no suele ser determinante *per se* para el desarrollo de la enfermedad, por el pequeño efecto del SNP en cuestión.

- Asociación indirecta: Este tipo de asociación, también conocida como no causal, se presenta cuando el alelo en el *locus* analizado está en desequilibrio de ligamiento (LD, *Linkage Disequilibrium*) con el *locus* causal (el concepto genético de LD será desarrollado en profundidad a continuación). Una situación de LD entre los dos *loci* indica que ambos están ligados y segregan de forma conjunta en mayor o menor grado, debido a la existencia de un bloque común para ambos definido por el patrón de recombinación.

- Asociación espuria: Este tipo de asociación puede ocurrir como consecuencia de la calidad de los datos, del muestreo estadístico o de la presencia de estratificación poblacional o mezcla genética (Anderson *et al.* 2010; Clarke *et al.* 2011). La estratificación poblacional se presenta cuando casos y controles son seleccionados de distintas poblaciones de forma no proporcional (Cardon & Palmer 2003; Freedman *et al.* 2004). La asociación, entonces, es una característica de la estructura de la población ya que las frecuencias alélicas pueden diferir entre poblaciones humanas por lo que es necesario su control mediante componentes principales o escalamiento multidimensional (Anderson *et al.* 2010). La mezcla genética se presenta en una población cuando su composición genética ha recibido una aportación de uno o más grupos poblacionales distintos en un pasado reciente (Hellendal *et al.* 2014).

Otro foco de irregularidades en los estudios se basa en su diseño (Malhotra & Goldman 1999; Cardon & Bell 2001; Merikangas & Risch 2003) que si no es realizado con detalle y presenta deficiencias, puede conllevar resultados inconsistentes con las consecuentes interpretaciones erróneas (Hattersley & McCarthy 2005). Los errores derivados de esta situación son los “errores de tipo I” y “tipo II”, muy extendidos en los estudios de asociación de genes candidatos con importantes problemas de replicación de resultados (Sullivan 2003;

NCI-NHGRI Working Group on Replication in Association Studies 2007; Farrell *et al.* 2015):

- Error de tipo I: También denominados falsos positivos, implican el rechazo de la hipótesis nula cuando la “no asociación” es verdadera. Este tipo de errores presentan tres causas principales:

○ La existencia de un sesgo de cara a la publicación de estudios de asociación con resultados positivos frente a negativos, problema inherente a al sistema de publicaciones científicas actuales (Sullivan 2003; Ioannidis *et al.* 2008; Duncan & Keller 2011).

○ El azar. A pesar de incluir un umbral de significación, siendo $\alpha < 0,05$ el más extendido por defecto, es posible obtener resultados espurios por simple probabilidad, ya que las diferencias observadas podrían ser debidas al azar en un 5% de los eventos. A un gran número de pruebas (múltiples comparaciones), aumenta la probabilidad de obtener un resultado significativo por azar. Existen un gran número de métodos de corrección para estas situaciones: Bonferroni, tasa de error familiar (FWER, *Family-wise Error Rate*), permutaciones caso-control o tasa de descubrimientos falsos (FDR, *False Discovery Rate*) (Dunn 1961; Westfall & Young 1993; Benjamini & Hochberg 1995). Los de uso más extendido serán desarrollados en el sub-apartado “Estudios de asociación de genoma completo”.

○ Estratificación poblacional, previamente comentada.

- Error de tipo II: También denominados falsos negativos, implican la aceptación de la hipótesis nula cuando esta es falsa. Este tipo de errores presentan dos causas principales que residen en el diseño del estudio:

○ Insuficiente potencia estadística por tamaño muestral insuficiente. La capacidad de inferir la potencia

estadística de un estudio antes de su realización permite calcular la capacidad de detección de asociaciones de interés (Ioannidis *et al.* 2008; Clarke *et al.* 2011).

o Clasificación incorrecta del fenotipo. La selección de individuos para la conformación de las muestras de casos y controles debe seguir altos estándares de calidad, con el fin de adquirir un conjunto muestral homogéneo. Para un gran número de trastornos complejos o de difícil fenotipo se ha propuesto el uso de endofenotipos. El término, introducido en 1972 (Gottesman & Shields 1972), hace referencia a la división del fenotipo original en síntomas conductuales o cognitivos que conforman fenotipos más estables, con la premisa o suposición de una conexión más directa con la etiología genética por un menor número de variantes implicadas respecto a la enfermedad compleja, con la consiguiente mejora del manejo de los datos genéticos (Gottesman & Gould 2003). Esta estrategia se ha extendido en psiquiatría en las aproximaciones de genes candidatos (Gottesman & Gould 2003; Walters & Owen 2007).

1.2.3.3 Resultados de la estrategia

En psiquiatría, tras asentarse esta aproximación de análisis, el flujo de estudios fue enorme (Sullivan 2017). La limitada comprensión de la base biológica de los trastornos psiquiátricos implicó que casi cualquier gen podría ser un candidato (Smoller 2017). La mayoría de los estudios se centraron en genes codificantes de receptores, transportadores y enzimas involucradas en sistemas de neurotransmisión, y pronto se alcanzó un ritmo de publicación frenético, con más de un artículo diario (Munafò & Flint 2004; Smoller 2017).

La etiología de prácticamente todos los trastornos y síndromes del espectro de la salud mental ha intentado ser desentrañada por los

estudios de genes candidatos. Por lo tanto, el arrollador número de estudios publicados impiden una revisión exhaustiva de los resultados. El tono general de las revisiones de estos análisis y los grandes meta-análisis de resultados previos arrojan una idea extendida: los estudios de genes candidatos generaron un gran interés, pero su éxito fue muy limitado, con enormes problemas de replicación que dieron lugar a una inmensa mayoría de falsos resultados positivos. La perpetua dificultad de los estudios de asociación de genes candidatos en psiquiatría se ha debido principalmente a estudios con potencia insuficiente por su tamaño muestral (Munafò 2006), incluso cuando iniciativas de gran tamaño lo han intentado de manera inconsistente (Hamer 2002). Este frustrante marco se extiende a múltiples patologías fuera del margen de la salud mental (Lohmueller *et al.* 2003), como demostraron algunos de los primeros análisis extensivos de asociaciones de genes candidatos mediante variantes comunes en enfermedades complejas (Hirschhorn *et al.* 2002): de un total de 600 asociaciones positivas la mayoría no eran robustas, ya que solamente 166 asociaciones fueron replicadas tres o más veces y solamente 6 fueron consistentes. Los metaanálisis son herramientas de gran potencia para el análisis conjuntos de múltiples resultados de un conjunto de estudios previos. Además, por los problemas producidos por la estratificación poblacional comentados previamente, los meta-análisis constituyen un esfuerzo extra para la detección de heterogeneidad poblacional o étnica en estudios de asociación (Ioannidis *et al.* 2001; Ioannidis *et al.* 2003). Pero la fortaleza de los meta-análisis reside en la calidad de los estudios llevados a cabo en primer término, por lo que en sí no conforman una estrategia de reemplazo con potencia adecuada (Munafò & Flint 2004). De todas formas, son aproximaciones de gran interés para la reevaluación de resultados previos de genes candidatos al igual que la comparación con resultados obtenidos en GWAS. Son múltiples los ejemplos en la literatura, que a pesar de detectar alguna señal significativa replicada tienden a señalar al desacierto general:

- En esquizofrenia, Sander y colaboradores (Sander *et al.* 2008), tras analizar los 14 genes candidatos más robustos en la literatura que despertaron un gran interés y esfuerzo investigador, no encontraron señales significativas. Farrell y

colaboradores (Farrell *et al.* 2015), realizaron un exhaustivo análisis de 25 variantes de enorme interés mediante meta-análisis, la comparación con estudios de gran tamaño actuales y un recorrido por los estudios de asociación clásicos y sus investigadores de referencia. El estudio concluyó que, sobre la evidencia empírica actual, “la literatura histórica de genes candidatos no produjo una comprensión clara de la base genética de la esquizofrenia” a pesar de marcar los primeros pasos en el terreno de la genética del trastorno.

- En depresión mayor un buen ejemplo se encuentra en Bosker y colaboradores (Bosker *et al.* 2011), donde comparan los resultados del GAIN MDD GWAS con los genes candidatos señalados históricamente como asociados significativamente con la enfermedad. De los 55 genes candidatos, solamente cuatro fueron confirmados.

- En trastorno por déficit de atención e hiperactividad, un buen ejemplo reside en el meta-análisis realizado por Gizer y colaboradores (Gizer *et al.* 2009), donde a pesar de mostrar cierto optimismo ante alguna señal de asociación confirmada, señalaban la falta de homogeneidad y replicabilidad general en las decenas de estudios llevados a cabo.

- En los trastornos de espectro autista, Freitag muestra en su extensiva revisión de la literatura científica el carácter problemático de la falta de replicación extendida. Señalando también los éxitos alcanzados en el campo, algo más numerosos que en otros trastornos debido a la arquitectura del trastorno (Freitag 2007).

- La búsqueda de genes candidatos no se restringió a la búsqueda de variantes asociadas a un único trastorno psiquiátrico o endofenotipo, sino que otros enfoques más amplios y espectrales fueron desarrollados. Algunos de estos análisis se centraron en las fronteras entre trastornos, con el análisis de variantes de efecto pleiotrópico. El surgimiento de señales de asociación en genes similares en distintos trastornos psiquiátricos llamó la atención de numerosos grupos de investigación. El optimismo inicial se apagó por los mismos

problemas de replicación generalizados comentados anteriormente (Smoller 2013).

- Estudios de interacciones gen-ambiente candidatas no han resultado muy fructíferos ante la falta de replicación y al sesgo de publicación (Duncan & Keller 2011).

A pesar de las indicaciones para considerar con cautela los resultados positivos detectados de manera inicial por su potencial origen en la casualidad, sesgo, problemas de análisis o interpretación (Buchanan *et al.* 2006; Ioannidis 2006), los falsos positivos se extendieron en la literatura del campo de la genética psiquiátrica. La justificación más extendida de la falta de replicación se resume en dos argumentos: aceptar los resultados por su marco biológico, a pesar de “la falacia de la plausibilidad biológica”, como garantía *per se* o justificar la divergencia de resultados por una posible arquitectura compleja con el fin de minimizar el peso de las falsas asociaciones positivas (Sullivan 2003). Aunque parte de las justificaciones podrían ser consideradas como ciertas por las evidencias actuales, la heterogeneidad de *locus* o la heterogeneidad etiológica no conforman una causa firme de la falta de éxito de los estudios de asociación de genes candidatos. Diversos autores resaltaron el esfuerzo dedicado, el coste excesivo y los pobres resultados de la investigación basada en genes candidatos, con críticas al diseño y enfoque de una gran parte de los estudios y a la estrategia en general (Johnson *et al.* 2017; Sullivan *et al.* 2018).

Tras diez años desde la publicación del artículo de Risch y Merikangas en 1996, el panorama general del campo de la genética psiquiátrica no mostraba claros caminos abiertos por los estudios de asociación de genes candidatos. Pero el abaratamiento considerable de las tecnologías y los grandes pasos dados por importantes proyectos masivos internacionales abrieron de nuevo la oportunidad de avance (Smoller 2017). La aproximación libre de hipótesis basada en GWAS se estableció como herramienta dominante en la búsqueda conjunta de un gran número de polimorfismos genómicos (Maier *et al.* 2018).

1.2.4 Estudios de asociación de genoma completo

1.2.4.1 Introducción

Los GWAS son una aproximación libre de hipótesis previa que implica el análisis de un amplio conjunto de variantes comunes potencialmente asociadas a un fenotipo de estudio. El propio GWAS es un elemento generador de hipótesis que, a diferencia de la aproximación clásica de asociación basada en genes candidatos, puede inferir información masiva de un gran número de marcadores. Para cada variante del genotipo obtenido se realiza un test de asociación, empleando las mismas herramientas estadísticas ampliadas respecto a los estudios de asociación de genes candidatos. Tras una asociación significativa o de potencial interés, el esfuerzo investigador puede ir dirigido al análisis concreto de una hipótesis biológica o a la generación de modelos para detectar y prevenir una patología.

Los dos primeros GWAS fueron publicados en 2005 y 2006 analizando la genética de la degeneración macular asociada a la edad (Klein *et al.* 2005; Dewan *et al.* 2006). Pero el primer GWAS de gran tamaño y diseño cuidado fue publicado en 2007 (Wellcome Trust Case Control Consortium 2007), analizando 14000 casos de siete patologías y 3000 controles compartidos. La importancia de este estudio en el marco de la investigación genómica en psiquiatría reside en tres puntos: una de las siete patologías analizadas era el trastorno bipolar, confirmó la importancia de importantes tamaños muestrales para la detección de pequeños tamaños de efecto en las variantes analizadas y supuso la primera aproximación estandarizada y masiva con el empleo de chips de genotipado de amplia cobertura (Visscher *et al.* 2012).

Esta nueva estrategia, basada en el genotipado de miles de SNPs que representan una gran parte de la variación común existente, surgió tras la secuenciación de la totalidad del genoma humano (International Human Genome Sequencing Consortium 2001; Venter *et al.* 2001) y el mapeo de un importante número de SNPs distribuidos por todo el genoma. El abaratamiento de las tecnologías de genotipado junto a la oportunidad de examinar variantes de riesgo comunes con alta

resolución revolucionó el campo (Hirschhorn & Daly 2005; Sullivan *et al.* 2018).

La hipótesis “enfermedad común – variante común” (CDCV, *Common Disease – Common Variant*), es la base principal en la que se estableció el interés en el análisis conjunto de un gran número de variantes comunes (Hemminki *et al.* 2008). El origen de esta hipótesis es un conjunto de publicaciones clave (Lander 1996; Risch & Merikangas 1996; Collins *et al.* 1997; Chakravarti 1999; Reich & Lander 2001). CDCV, como hipótesis, plantea que las variaciones genéticas frecuentes en la población general, con una penetrancia relativamente baja, son las contribuyentes principales a la susceptibilidad genética de las patologías comunes (Schork *et al.* 2009).

Los distintos estudios clave no realizaron una predicción del número de *loci* de riesgo involucrados o del tamaño del efecto, sino que señalaban hacia la frecuencia de estas variantes (Lander 1996; Reich & Lander 2001) o puntos esenciales a nivel experimental y estadístico para el análisis de estas variantes (Risch & Merikangas 1996). Algunos autores consideran que Reich y Lander (Reich & Lander 2001), sin defender la hipótesis CDCV por completo, proporcionaron “una perspectiva teórica” o “marco” del espectro alélico y sus frecuencias compatible con un gran número de patologías (Schork *et al.* 2009). Todos estos estudios de referencia tienen en común la capacidad predictiva de la “era de GWAS” antes de la publicación de las iniciativas que abrieron el camino a la investigación genómica y al desarrollo de las tecnologías que la hicieron posible (Visscher *et al.* 2012, *Five years GWAS*). Además del número de variantes potencialmente involucradas en este modelo, algunas de las cuestiones inherentes a la hipótesis CDCV, previas a los resultados de GWAS, son resumidas por Becker (Becker 2004): “¿Los factores genéticos comunes son específicos de una única patología?, ¿cuáles son los efectos de esos alelos involucrados cuando están en distintas combinaciones alélicas o ante distintas influencias ambientales o epigenéticas? o si el riesgo de estas variantes es dado en contextos concretos ¿cuál es su efecto en contextos sin riesgo?”.

Esta hipótesis también recibió diversas críticas desde su planteamiento, como las aportadas por Pritchard (Pritchard 2001). El autor se inclinaba hacia un modelo centrado en la existencia de un conjunto de variantes raras de mayor efecto, justificado en los procesos operativos poblacionales de los que la especie humana es objeto (Schork *et al.* 2009). La hipótesis para patologías comunes basada en variantes raras se denomina “enfermedad común - variante rara” (CDRV, *Common Disease – Rare Variant*). Ésta propone una arquitectura genética basada en múltiples variantes raras que, con su mayor penetrancia, contribuyen a la susceptibilidad genética de las enfermedades comunes. El debate existente entre seguidores de una u otra hipótesis es clave tanto para comprender la naturaleza de las patologías comunes como las estrategias de exploración de las mismas (Schork *et al.* 2009). Ambas hipótesis no son mutuamente exclusivas (Goldstein & Chikhi 2002), pero distintos investigadores tendieron a decantarse por una u otra opción a lo largo de la investigación genómica.

El avance del análisis de variación común en psiquiatría reside, en gran medida, en el esfuerzo colaborativo internacional surgido en el año 2009 con el nombre de *Psychiatric Genomics Consortium* (PGC) (Cichon *et al.* 2009; Craddock *et al.* 2009; Psychiatric GWAS Consortium Steering Committee 2009). Esta iniciativa, todavía activa y en proceso de crecimiento, ha realizado gran parte de las mayores aproximaciones a la genética psiquiátrica común analizando masivamente cientos de miles de individuos de distintas patologías psiquiátricas (Sullivan *et al.* 2018). Las características del PGC y la accesibilidad a sus datos son comentadas en “Materiales y Métodos”.

1.2.4.2 Capturando la variación genética común

Para la realización de un GWAS es necesario conocer la localización y densidad de la variación genética más común en el genoma humano, los SNPs. La publicación de un mapa inicial formado por 1,42 millones de SNPs al unir las iniciativas The SNP Consortium y análisis de variación en los datos obtenidos a partir del International

Human Genome Consortium (Sachidanandam *et al.* 2001) y el establecimiento de un catálogo público estandarizado y ordenado de esta variación (Sherry *et al.* 1999; Sherry *et al.* 2001) denominado dbSNP puso al alcance de los investigadores una “herramienta genética”. Además, el creciente número de iniciativas permitió conocer las diferencias en la variación genética específica de las distintas y principales poblaciones humanas (Bush & Moore 2012). Por último, esta variación capturada necesitaba un mapa haplotípico del genoma humano que informase de la estructura y correlación existente entre la variación genética común recién detectada y publicada con el fin de sortear la información redundante en los estudios de asociación (Reich *et al.* 2001). Esta iniciativa, que comenzó en 2002, caracterizó el LD existente a lo largo del genoma humano (International HapMap Consortium 2005) a lo largo de sus distintas fases, ampliando secuencialmente el número de variantes incluidas y las poblaciones disponibles en esta base de datos de acceso público (International HapMap Consortium 2007; The International HapMap 3 Consortium 2010). La disposición de la estructura haplotípica del genoma humano fue indispensable para el desarrollo de estudios de asociación dirigidos y equilibrados a través de los patrones de LD (Hirschhorn & Daly 2005). Posteriormente, la caracterización haplotípica del genoma humano fue mejorada con distintas iniciativas como el proyecto 1000 Genomas en sus distintas fases (1000 Genomes Project Consortium 2010; 1000 Genomes Project Consortium 2012; 1000 Genomes Project Consortium 2015; Sudmant *et al.* 2015) o el panel de referencia Haplotype Reference Consortium (McCarthy *et al.* 2016) con más de 39 millones de SNPs incluidos.

1.2.4.3 Desequilibrio de ligamiento

El LD puede implicar una asociación no independiente de alelos en dos o más *loci*. Los bloques haplotípicos varían entre poblaciones en función de sus historias demográficas concretas (cuellos de botella poblacionales, mezclas con otras poblaciones, expansiones, mutaciones, selección, etc.). En una familia, los eventos de recombinación, generación tras generación, fracturan segmentos

cromosómicos. Este efecto es amplificado a lo largo de generaciones. En teoría, si este evento recombinatorio ocurriese consecutivamente hasta fragmentar todos los haplotipos presentes que contienen alelos correlacionados, todas las variantes se encontrarían en equilibrio de ligamiento (Bush & Moore 2012). Pero en el genoma humano, las distintas poblaciones muestran bloques no afectados durante la recombinación meiótica con importantes diferencias poblacionales: la población africana muestra una mayor diversidad haplotípica respecto al resto de poblaciones humanas, debido al cuello de botella sufrido tras la expansión del ser humano por el planeta (Tishkoff & Verrelli 2003; Noonan *et al.* 2006). Esta diversidad haplotípica diferencial entre poblaciones implica distintos tamaños de los bloques haplotípicos: menores en poblaciones africanas o descendientes recientes de la población africana respecto a poblaciones europeas o asiáticas (Bush & Moore 2012).

El LD es un parámetro cuantificable para el cual existen diversas formas de estimación y cada una captura distintas características de la asociación no aleatoria (Devlin & Risch 1995; Wang *et al.* 2005; Slatkin 2008). Las dos más empleadas en estudios de asociación son D' y r^2 (Bush & Moore 2012), pero todas están basadas en “ D ”, coeficiente que mide diferencias de frecuencias haplotípicas esperadas y observadas a partir de cada alelo.

- D' es un estimador extendido en genética de poblaciones con escala entre 0 y ± 1 . Cuando D' presenta un valor igual a ± 1 indica un patrón de LD completo, en el que las variantes en el bloque haplotípico segregan siempre juntas, mientras que D' con un valor igual a 0 indica un completo equilibrio de ligamiento. El cálculo de D' se basa en la división de D (coeficiente que mide diferencias de frecuencias de haplotipos observadas y esperadas de cada alelo) entre el valor máximo (cuando $D \geq 0$) o mínimo ($D < 0$) que puede alcanzar D .

- r^2 es un estimador muy extendido en análisis genéticos. Es más restrictivo que D' y está inversamente relacionado con el tamaño muestral requerido en estos análisis. Altos niveles de este estimador indican que dos SNPs transmiten información

similar, ya que un alelo de un SNP está asociado al alelo de otro SNP, indicando el nivel de redundancia estadística entre información inferida de esas variantes. Esta característica permite que solamente la información de un alelo de las distintas variantes a un r^2 tenga que ser capturada o genotipada para inferir información del resto de variantes, englobando la variabilidad haplotípica. La inferencia sistemática se denomina imputación de variantes.

Cuando el patrón poblacional de LD es conocido, solamente es necesario genotipar una fracción del número total de variantes. Estas variantes informativas reciben el nombre de “tag SNPs” y capturan la mayoría de la variación común de cada región (Johnson *et al.* 2001). Los tag SNPs pueden diferir entre poblaciones y su correcta selección implica reducción de costes con baja pérdida de potencia (de Bakker *et al.* 2005).

La asociación de una variante analizada puede reflejar dos situaciones distintas: asociación directa (donde el SNP genotipado es funcional e influye biológicamente al fenotipo estudiado) o asociación indirecta (el tag SNP genotipado no presenta una relación causal con el genotipo pero muestra un importante LD con la variante causal) (Bush & Moore 2012).

1.2.4.4 Diseño del estudio

Los fenotipos aptos para el análisis mediante GWAS responden a dos modelos: categórico (caso-control) o cuantitativo (Hirschhorn & Daly 2005; Bush & Moore 2012). En psiquiatría, la aproximación más extendida o común compara una muestra de individuos no afectados con una muestra de individuos con el trastorno. Esto es debido a que en los trastornos mentales, el modelo nosológico actual clasifica a los individuos de forma binaria (Owen 2014; Artigas-Pallarés & Paula-Pérez 2015). Otro enfoque posible es el que emplea trios, con los progenitores y su descendiente analizados. Sin embargo, esto no implica que la nosología actual impida realizar análisis en base a rasgos cuantitativos. Aunque menos extendidos, los estudios que

emplean información cuantitativa también son desarrollados para múltiples trastornos o rasgos de la personalidad (p.ej. escalas de impulsividad, unidades de alcohol consumido, etc.). Ambas aproximaciones permiten inferir información genética de cualquier rasgo humano, pero en términos generales, el análisis de rasgos cuantitativos en el marco de susceptibilidad continuo suele aportar una clasificación más ajustada (Bush & Moore 2012).

Distintos factores pueden influir en la potencia estadística de los estudios genéticos de enfermedades complejas: aquellos que no dependen del control y diseño del estudio (complejidad de la arquitectura genética del fenotipo, frecuencias y tamaño del efecto de los alelos, la estabilidad del fenotipo y las características genéticas no ajustables de la población de estudio) y aquellos controlables (selección de individuos, tamaño muestral, medidas fenotípicas y genotípicas y métodos para el control de calidad o análisis estadísticos) (Sham & Purcell 2014):

- Tamaño muestral: El número de variantes asociadas significativamente está fuertemente correlacionado con el tamaño de la muestra analizada (Hirschhorn & Daly 2005; Sham & Purcell 2014). Con un conjunto suficiente de casos y una muestra control adecuada, un umbral de significación puede ser superado y, debido a la correlación tamaño-número de variantes detectadas, al incrementar el número de muestras así lo hará el de asociaciones significativas (Visscher *et al.* 2012). El grado de correlación está definido en parte por la naturaleza de la arquitectura genética subyacente (Sham & Purcell 2014). Algunos trastornos psiquiátricos presentan una arquitectura genética que permite la detección de variantes asociadas empleando un menor número de muestras respecto a otros de naturaleza más heterogénea, pero a partir de un tamaño muestral suficiente la relación pasa a ser lineal (Sullivan *et al.* 2018). La muestra requerida para la detección de una variante genética concreta está determinada por la frecuencia alélica y tamaño del efecto de la variante.

- Frecuencia alélica: La frecuencia de un alelo determina la potencia estadística. Una variante rara tiene una menor probabilidad de encontrarse en LD con una variante común incluida en un chip comercial y, además, dos *loci* con grandes diferencias de frecuencia presentan un patrón de LD necesariamente bajo (Sham & Purcell 2014).

- Chip/Array: La tecnología de genotipado empleada debe ajustarse a las características poblacionales, buscando aquellas variantes que permiten incrementar la cobertura en función de los patrones de LD (Sham & Purcell 2014). La selección de un chip eficiente en el análisis del fenotipo deseado permite una mejor inferencia, por lo que distintas consideraciones pueden ser realizadas: ajuste del chip a la población de estudio, ajuste del chip a la patología de estudio (arrays que incluyen variantes específicamente susceptibles de informar sobre la arquitectura genética subyacente) o ajuste a las regiones genómicas de interés (por ejemplo el empleo de un chip exónico o uno con mayor cobertura en regiones cromosómicas concretas). Además, algunas casas comerciales muestran productos configurables, donde la inclusión de variantes concretas de interés para el investigador es posible.

- Fenotipo: Desde una perspectiva estadística, los rasgos cuantitativos mejoran la potencia para detectar un efecto genético (Bush & Moore 2012). La selección de fenotipos extremos (en el extremo de la distribución normal) (Zabaneh *et al.* 2018) o un conjunto específico de individuos en base a la estructura nosológica de la patología (inicio temprano, importante carga familiar, etc.) también son estrategias de interés (Sham & Purcell 2014). Por último, aquellos estudios longitudinales con múltiples medidas del mismo fenotipo en distintos puntos temporales o el empleo conjunto de fenotipos distintos altamente correlacionados genéticamente entre sí, podrían reducir el efecto de las fluctuaciones temporales y distintos errores de medida (Sham & Purcell 2014).

- Población: Además de una selección cuidada de individuos para minimizar la estratificación poblacional (Köhler

& Bickeböllner 2006), el uso de poblaciones aisladas concretas puede ser una ventaja debido a la mayor homogeneidad entre individuos expuestos, además, una menor variabilidad ambiental (Hirschhorn & Daly 2005).

- Control de calidad: Seguir un protocolo de control de calidad permite mejorar la potencia estadística para detectar asociaciones verdaderas. Con parámetros ajustados a las características propias del estudio (Anderson *et al.* 2010) el protocolo debe: excluir individuos altamente emparentados, identificar individuos con sexo discordante, identificar individuos con elevada ratio de genotipos perdidos o heterocigosidad, detectar individuos duplicados, reconocer individuos con ancestralidad discordante, identificar marcadores con una elevada ratio de genotipos perdidos, reconocer marcadores con importante desviación del equilibrio Hardy-Weinberg, identificar variantes con distinta ratio de genotipación entre casos y controles y eliminar aquellas variantes con una frecuencia del alelo menor muy baja.

1.2.4.5 Imputación de variantes

La imputación de genotipos es otra de las vías para mejorar la potencia de un GWAS y su empleabilidad posterior. Este término describe el proceso de predecir genotipos que no han sido directamente evaluados por el chip utilizado en el estudio (Marchini & Howie 2010). En este proceso, la inferencia de genotipos no genotipados en primera instancia se basa en explotar los patrones de LD conocidos y las frecuencias haplotípicas obtenidas a partir de HapMap, 1000 Genomas o Haplotype Reference Consortium entre otros paneles de referencia (Li *et al.* 2009; Das *et al.* 2016). Este proceso puede ser centrado en una región específica, con el fin de lograr un mapeado detallado, o puede ser llevado a cabo a lo largo del genoma (Marchini & Howie 2010).

La imputación de variantes en muestras no emparentadas se realiza en varias fases y sobre varias premisas. Este conjunto de individuos debe presentar un conjunto de variantes tipadas distribuidas por el

genoma y los haplotipos de estos individuos serán semejantes por descendencia. Tras computar los haplotipos presentes en los individuos estudiados para estimar sus frecuencias haplotípicas, el proceso de imputación de variantes comienza con un genotipo parcial que busca ser “completado”. Los distintos algoritmos que tratan de inferir esta información “perdida” presentan diferencias en sus aproximaciones, pero todos ellos buscan en los individuos analizados el mosaico haplotípico basado en el panel de referencia (Marchini & Howie 2010). Existe una gran diversidad de programas informáticos para la imputación de variantes, los cuales presentan distintas metodologías y mejoras o implementaciones. La imputación de genotipos es un proceso con una importante carga computacional que requiere el acceso a supercomputadores y precisa una cantidad de tiempo variable, que depende del número de muestras y el método empleado. Muchos de estos programas buscan reducir las necesidades computacionales sin perder precisión. Algunos de los más novedosos permiten la automatización parcial del proceso mediante el uso de servidores externos, con el consiguiente ahorro temporal y simplificación del procedimiento bioinformático (Das *et al.* 2016). Algunos de los distintos programas de uso más extendido son las distintas versiones de IMPUTE (Marchini *et al.* 2007; Howie *et al.* 2009), Beagle (Browning & Browning 2016; Browning *et al.* 2018) o minimac (Howie *et al.* 2012; Fuchsberger *et al.* 2015; Das *et al.* 2016).

Los genotipos son imputados con incertidumbre, que deberá ser considerada en análisis posteriores, produciéndose una distribución de probabilidad entre genotipos posibles. Esta situación es producida cuando los haplotipos de la muestra estudiada coinciden con varios haplotipos de referencia. Tras la imputación, la calidad de los genotipos imputados debe ser considerada con el fin de evitar aquellos que no superen un umbral de calidad mínimo. La potencia estadística para detectar la potencial asociación de estos SNPs depende de su calidad de imputación (Marchini & Howie 2010; Bush & Moore 2012).

Este genotipo más extenso permite mejorar ciertos aspectos de la predicción genómica en la que se centran los GWAS. Por una parte permite poner el foco detallado en regiones concretas ya que permite

mejorar la resolución genotípica, implicando una mayor probabilidad de detectar la variante causal. Además, los genotipos imputados pueden permitir una mayor potencia de detección de variantes asociadas a un fenotipo concreto. Asimismo, esta mayor resolución genotípica mejora en gran medida dos aproximaciones: meta-análisis de datos de GWAS y cálculos de puntuaciones de riesgo basadas en modelos poligénicos. Estas dos aproximaciones emplean información genotípica de distintas muestras, que pueden mostrar importantes diferencias en su diseño entre estudios afectando al número de variantes incluidas. Al incrementar el número de variantes disponibles la probabilidad de solapamiento genotípico aumenta, permitiendo la inclusión de un mayor número de variantes con el consecuente incremento de información analizada. Por último, cabe indicar que la imputación acertada de variantes raras es más compleja que la imputación de variantes comunes. Las variantes raras presentan una menor probabilidad de encontrarse en alto desequilibrio de ligamiento respecto a variantes comunes, lo cual afecta directamente a la calidad e imputación (debido a su dependencia del LD existente) (Marchini & Howie 2010; Sham & Purcell 2014). Algunos paneles de referencia actuales de alta densidad permiten la imputación de variantes con frecuencias del 0,1-0,5% (McCarthy *et al.* 2016).

1.2.4.6 Correcciones para test múltiples

Al genotipar un gran número de variantes, el nivel de significación estadístico debe corregirse por el número de test realizados. La aproximación general se basa en las mismas premisas y mecanismos de control comentados previamente en el apartado del texto en el que se desarrolló el concepto de asociación. Las dos correcciones más empleadas son:

- Corrección por Bonferroni: Esta aproximación al control por tests múltiples es una de las más sencillas y exigentes, ya que reduce de forma sencilla el umbral de significación acorde al número de tests independientes llevados a cabo en el estudio (Bush & Moore 2012; Sham & Purcell 2014). La corrección de

Bonferroni ajusta el valor de α establecido al número “k” de tests llevados a cabo ($\alpha_{\text{corregido}}=\alpha/k$). Risch y Merikangas propusieron el valor $\alpha=5 \times 10^{-8}$ como el umbral conservador para declarar una variante significativa a nivel genómico (Risch & Merikangas 1996). Este umbral arbitrario responde al cálculo derivado de un estudio con las siguientes características: que incluya variación común en individuos de origen europeo, con 1000 casos y 1000 controles en los que se estiman un millón de test de asociación independiente (Sveinbjornsson *et al.* 2016). Distintas aproximaciones y correcciones a este método han sido propuestas para minimizar la arbitrariedad de esta corrección en base a la incorporación de información de la magnitud del efecto analizado, considerando los bloques de LD y la presencia de *singleton SNPs* (SNPs que solamente informan de su propio genotipo sin contribuir al cálculo de genotipos del resto de la región cromosómica) o métodos basados directamente en las dependencias de los resultados entre pares de alelos (Moskvina & Schmidt 2008; Sveinbjornsson *et al.* 2016). En la actualidad, el uso de 5×10^{-8} como umbral de significación estadística en GWAS está extremadamente extendido, a pesar de su arbitrariedad y el empleo de un mayor número de variantes incluidas en los análisis o poblaciones concretas (Pe'er *et al.* 2008). Es el método más empleado.

- Tasa de descubrimientos falsos (FDR): Esta aproximación al control por múltiples test es una estimación de la proporción de resultados significativos que son falsos positivos. Esta técnica estadística, desarrollada por Benjamini y Hochberg (Benjamini & Hochberg 1995), corrige para este número esperado de descubrimientos falsos, proporcionando una estimación del número de resultados verdaderos entre el conjunto de resultados significativos. El uso de FDR está menos extendido que el de Bonferroni. El cálculo del valor p corregido se basa en realizar una ordenación creciente de los resultados significativos ($p < 0,05$) en función de su significación estadística resultante del test de asociación, asignando un rango a cada uno (p.ej. al resultado más significativo se le asigna el rango 1, al

siguiente el 2, hasta completar todos los rangos). La estimación se basa en la siguiente fórmula:

$$\text{Valor corregido} = \left(\frac{i}{m}\right) \times Q$$

Donde “i” es el rango asignado a cada resultado, “m” el número total de tests ordenados y “Q” la tasa de descubrimientos falsos elegida (porcentaje). Análogamente, el método estadístico cuantil-cuantil captura y muestra gráficamente cuan cerca está la distribución de un conjunto de datos a su distribución “ideal”, comparando los valores de p obtenidos vs los esperados (esta clase de representación gráfica captura la misma información que el método FDR, sin ser un resultado del mismo sino uno análogo) (Sham & Purcell 2014).

1.2.4.7 Resultados de la estrategia GWAS

En psiquiatría, el análisis mediante GWAS ha producido un volumen de resultados de gran tamaño. Tras el primer GWAS en trastorno bipolar (WTCCC 2007), el incremento en el número de estudios ha seguido prácticamente un patrón exponencial, con cientos de estudios publicados anualmente en los últimos años (Horwitz *et al.* 2018). El abaratamiento de la tecnología y el acceso a cada vez mayores colecciones muestrales permiten el avance de esta aproximación. El estudio de decenas de fenotipos, rasgos, conjuntos de síntomas y otras medidas relacionadas con los trastornos psiquiátricos han permitido la detección de cientos de *loci* significativos. Los fenotipos más comúnmente analizados son aquellos determinados por la nosología actual: trastorno depresivo mayor, esquizofrenia, trastorno bipolar, trastorno por déficit de atención e hiperactividad, trastornos del espectro autista, anorexia nerviosa, estrés post-traumático, trastornos por uso o abuso de sustancias, síndrome de Tourette o trastorno obsesivo compulsivo entre otros. Otros fenotipos comúnmente estudiados son aquellos relacionados con los principales rasgos de personalidad, cognitivos, educativos o relacionados con patologías de interés por su coocurrencia o “cercanía” clínica (patologías

neurológicas o autoinmunes entre otras). El análisis de fenotipos basados o relacionados con síntomas o descriptores clínicos concretos (agresividad, impulsividad, manía, neuroimagen, etc.), con la respuesta a fármacos (agranulocitosis, resistencia, farmacocinética, etc.) o subtipos/endofenotipos también ha aportado cierto avance al campo. La comparación entre casos y controles (o entre casos y casos) con una característica clínica diferenciadora presenta ventajas a la hora de detectar susceptibilidades de dominios específicos de la psicopatología. El análisis de fenotipos desde la perspectiva cuantitativa, como se ha comentado previamente, es de gran interés en una nosología regida por un sistema categórico. La combinación de varios fenotipos o síntomas de los mismos también está aportando nuevas evidencias de peso para el campo, así como la comparación entre patrones de resultados de asociación de diferentes diagnósticos.

Estos cientos de iniciativas han seguido un desarrollo desigual, con mayor o menor recorrido y éxito. El campo, al igual que la mayoría de la investigación genómica, ha estado sesgado desde el punto de vista poblacional, con el foco puesto en poblaciones de origen europeo. Además, un gran número de estudios no han logrado asociaciones significativas. Esto puede ser debido a las características de su diseño que afectan a la potencia estadística del estudio. Las iniciativas de mayor peso hasta la fecha son aquellas realizadas mediante consorcios, principalmente el PGC, ya que han permitido alcanzar resultados robustos, replicados y altamente asociados.

A pesar de las diferencias existentes entre fenotipos analizados, el progreso logrado en esquizofrenia representa el potencial de los GWAS en psiquiatría. Por ello, solamente serán presentados en este subapartado parte los resultados obtenidos en los últimos años. La selección de los mismos se basa en su relevancia y capacidad representativa, pero el proceso selectivo empleado es parcialmente arbitrario sin buscar representar la totalidad.

El primer GWAS publicado en esquizofrenia, junto a otras de las primeras iniciativas, estaban basados en el genotipado de conjuntos de muestras (también conocido como *pools*), pocos individuos involucrados y chips limitados a un menor número de variantes con

poca cobertura (Mah *et al.* 2006; Ferreira *et al.* 2008; Shifman *et al.* 2008). Estos estudios no detectaron resultados significativos, pero sí sugirieron evidencias sin peso. Los dos siguientes GWAS emplearon el chip de Affymetrix 500k, logrando una cobertura total del genoma (Lencz *et al.* 2007; Sullivan *et al.* 2008). Pero, para los estándares actuales, el número de individuos involucrados seguía siendo todavía demasiado bajo (178 casos y 144 controles en el trabajo de 2007 y 738 pacientes y 733 controles en el trabajo de Sullivan y colaboradores). Ninguna de las asociaciones fue significativa y solamente se obtuvieron resultados que podrían sugerir ciertas evidencias de interés. Es en 2008 cuando O'Donovan y colaboradores (O'Donovan *et al.* 2008) logran la primera señal significativa. Inicialmente en el estudio se incluyeron 479 pacientes y 2937 controles, obteniendo varias señales que rozaban el nivel de significación genómica. Al analizar conjuntamente esta muestra con otras cohortes a modo de réplica (alcanzando 6666 pacientes y 9897 controles), se obtuvieron todavía resultados más asociados, pero no fue hasta al unir un conjunto de muestras de pacientes con trastorno bipolar cuando se obtuvo la primera señal significativa (en el gen *ZNF804A*).

Tras la insatisfacción general de los primeros estudios y la fundación del PGC como consorcio internacional de referencia, las tornas comenzaron a cambiar (Psychiatric GWAS Consortium Steering Committee 2009). Tras las suposiciones optimistas iniciales, los autores asumieron que los efectos esperados de las variantes comunes eran mucho menores de lo esperado, con la consiguiente dificultad a la hora de lograr detectar asociaciones significativas (Sullivan *et al.* 2018). La confluencia en el PGC de distintas iniciativas internacionales permitió el inicio de la “cascada” de señales significativas (Purcell *et al.* 2009; Shi *et al.* 2009; Stefansson *et al.* 2009). Las variantes asociadas implicaban a los genes *TCF4* y *NRGN* y a la región cromosómica del complejo principal de histocompatibilidad (MHC). El trabajo de Purcell y colaboradores supuso un giro sustancial en el campo, al demostrar por primera vez la arquitectura poligénica en una patología. Esta estrategia de análisis presentada podría ser entendida como la base de esta tesis y de un gran número de iniciativas y estrategias de análisis claves para la predicción de riesgo. Esta estrategia considera el conjunto de variantes

estudiadas, incluyendo aquellas con asociación significativa o sin ella, y será ampliamente desarrollada en el apartado “Puntuaciones de riesgo poligénico”. Este trabajo incluyó 3322 casos y 3587 controles (la réplica incluyó 4692 casos y 15493 controles). Los otros dos, emplearon cifras muestrales semejantes o ligeramente superiores.

En 2011, el consorcio Schizophrenia Psychiatric Genome-Wide Association Study GWAS Consortium (Schizophrenia Psychiatric Genome-Wide Association Study GWAS Consortium 2011), empleó 21856 individuos para sus análisis, así como una muestra independiente para réplica de 29839 individuos. Este estudio detectó 10 variantes asociadas significativamente a 7 *loci*, de los cuales 5 nunca habían sido descritos: *MIR137*, *PCGEM1*, *CSMD1*, *MMP16* y *CNNM2/NT5C2*. La variante más asociada, afectando al *locus MIR137*, sugirió la potencial involucración del microARN 137, implicado en el desarrollo neuronal. Al analizar conjuntamente los datos de esquizofrenia con individuos sanos y con trastorno bipolar, tres *loci* superaron el umbral de significación genómica: *CACNA1C*, *ANKK3* y la región *ITIH3/4*. Este incremento de muestras, además de una mayor detección de variantes asociadas, permitió la replicación de resultados previos indicando la necesidad de incrementar el número de muestras y no prescindir de los GWAS (Sullivan 2012). A su vez, Steinberg y colaboradores (Steinberg *et al.* 2011) extendieron el trabajo previamente comentado de Stefansson (Stefansson *et al.* 2009) meta-analizando distintos conjuntos de datos. Mediante esta aproximación detectaron dos variantes asociadas a tres *loci*: *VRK2*, *CCDC68* y *TCF4*.

En 2011, dos interesantes GWAS fueron llevados a cabo en población china detectando distintos *loci* significativos (Shi *et al.* 2011; Yue *et al.* 2011). La investigación genómica de la esquizofrenia estaba monopolizada por poblaciones europeas, por lo que estos dos trabajos forman parte de los intentos iniciales por diversificar poblacionalmente el conocimiento genético en el trastorno. Además del estudio de la población china, dos GWAS realizados en población japonesa fueron publicados, aportando nuevas evidencias del modelo poligénico subyacente y permitiendo descifrar la diversidad étnica de la esquizofrenia (Ikeda *et al.* 2011; Yamada *et al.* 2011). En esta misma

dirección Lencz y colaboradores (Lencz *et al.* 2013) analizaron una muestra de individuos compuesta por judíos asquenacés y realizaron distintas réplicas y meta-análisis con individuos de distinto origen étnico (incluyendo población africana y asiática). Los autores encontraron un SNP asociado significativamente en la población inicial con efectos consistentes en las cohortes de replicación y extensión.

Ese mismo año, se publica un estudio que describe 13 nuevos *loci* no descritos previamente (Ripke *et al.* 2013). Éstos, forman parte de los 22 descritos por el estudio al realizar un GWAS en múltiples etapas empleando diversas cohortes y réplicas. Este estudio volvía a señalar el peso de una arquitectura poligénica del trastorno, al estimar que 8300 SNPs independientes contribuyen al riesgo a desarrollar esquizofrenia. Los análisis poligénicos detectaron un 32% de varianza explicada en la escala de susceptibilidad.

En 2014 es publicado un estudio que revolucionó el campo (Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014). Al analizar 49 cohortes compuestas por 34241 casos, 45604 controles y 1235 tríos detectó 108 *loci* involucrados significativamente en la enfermedad. Las 128 variantes asociadas hicieron posible, por vez primera, el obtener un marco bastante consistente de la biología subyacente del trastorno. Dentro de estos *loci* asociados, muchos no muestran una clara funcionalidad afectada pero bastantes responden a hipótesis actuales de la etiología o el tratamiento del trastorno. Las evidencias significativas de mayor interés son las siguientes:

- Dianas terapéuticas. Dos asociaciones significativas revelan el peso de las rutas dopaminérgicas y glutamatérgicas. *DRD2* es un gen responsable de la “cognición, recompensa, motivación, aprendizaje y memoria” y es diana de los antipsicóticos de uso más extendido. *GRM3* es un candidato terapéutico histórico. La confirmación de dianas terapéuticas farmacológicas de uso clínico extendido representa un importante hallazgo en el campo de la genómica psiquiátrica y, además, simboliza una de las metas del PGC (Sullivan *et al.* 2018).

- Neurotransmisión glutamatérgica. La hipótesis glutamatérgica es apoyada por un conjunto variado de asociaciones que superaron el umbral de significación a nivel genómico: la subunidad *GRIN2A* del receptor NMDA es un mediador de la plasticidad sináptica. *GRIAI* es una subunidad de AMPA, mediador de la transmisión sináptica rápida que se encuentra involucrado en la organización dendrítica de receptores y transmisión sináptica del hipocampo. *SRR* codifica un catalizador de un agonista esencial y activador de los receptores NMDA, con altos niveles asociados previamente al trastorno. *CLCN3* es un canal de cloruro dependiente de voltaje de cloruro responsable de la sinapsis glutamatérgica en el hipocampo. Además, *GRM3* y *SLC38A7*, podrían estar involucrados en el reciclaje del glutamato en su ciclo molecular.

- Función sináptica y plasticidad: *KCTD13*, *NLGN4X*, *IGSF9B*, *CNTN4*, *MEF2C*, *PTN*, *CNKSRL1*, *PAK6* y *SNAP91* son los *loci* significativos. Algunos de ellos han sido descritos previamente en esquizofrenia, trastorno bipolar o trastornos del espectro autista y parecen estar involucrados en funciones como la regulación de la estructura del citoesqueleto, induciendo la formación presináptica glutamatérgica y GABAérgica, adhesión molecular expresada en interneuronas GABAérgicas, moléculas de adhesión celular, regulando la neurogénesis o la morfogénesis dendrítica entre otras.

- Señalización neuronal de calcio: *CACNA1I*, *RIMS1*, *CACNA1C*, *CACNB2*, *CAMKK2*, *NRGN*, *ATP2A2*. Involucrados en la plasticidad sináptica o la liberación de neurotransmisores, algunos de ellos se sumaron a las evidencias previas que involucran algunos de estos *loci*.

- Canales iónicos neuronales: *KCNB1*, *HCN1*, *CHRNA3*, *CHRNA5* y *CHRNA4*. El primero está relacionado con la excitabilidad neuronal y el potencial de acción. El segundo *locus* es un canal de potasio involucrado en numerosos procesos neuronales. Los tres últimos *loci* son receptores nicotínicos previamente involucrados en fenotipos dependientes de nicotina.

- Neurodesarrollo: Los *loci* significativos *FXR1*, *SATB2*, *PODXL*, *BCL11B*, *TLE1*, *TLE3* y *FAM5B* están de distintas formas involucrados en el neurodesarrollo (proteínas de unión a ARN, asociados al desarrollo neocortical, etc.).

Otro de los resultados de mayor importancia de esta iniciativa son las nuevas evidencias del peso de la estructura poligénica basada en variación común considerada como un conjunto. Los modelos poligénicos llegan a explicar un 18% de la varianza en la escala fenotípica, implicando aproximadamente 10000 SNPs. Las variantes significativas a nivel genómico explican un 3% de esta varianza, mientras que un único marcador explica como mucho el 0,1%. Estos resultados resaltan la oportunidad analítica residente en las estadísticas resumen resultantes del trabajo (Pasaniuc & Price 2016). Este conjunto de datos condensa información estadística de cada variante analizada en el estudio (grado de asociación estadística, alelos y sus frecuencias, efecto, etc.) y pueden ser empleados para distintos objetivos (Visscher *et al.* 2017). La aplicación de modelos poligénicos basados en los resultados de este GWAS podrían ser generados con una mayor exactitud y menor ruido que los obtenidos en trabajos previos.

Tras el trabajo realizado por el PGC, otras iniciativas han sido llevadas a cabo siguiendo algunas de las vías de estudio comentadas en el primer párrafo de este sub-apartado. Por ejemplo, en 2014 es publicado un trabajo enfocado a la genética subyacente de la psicosis en sentido amplio, incluyendo individuos afectados por esquizofrenia, trastorno esquizo-afectivo o trastorno bipolar psicótico (Psychosis Endophenotypes International Consortium & Wellcome Trust Case-Control Consortium 2014). Aunque ninguna variante superó el umbral de significación genómica este tipo de iniciativas pueden aportar nuevos conocimientos de los distintos trastornos. En este mismo sentido, Edwards y colaboradores (Edwards *et al.* 2016) llevaron a cabo un GWAS de escalas de síntomas positivos y negativos. Sin encontrar señales significativas en los SNPs analizados, algunas rutas biológicas sí podrían mostrar indicios en la diferenciación sintomatológica. También han sido ampliadas las evidencias en otras poblaciones como

iniciativas como la llevada a cabo por Li y colaboradores (Li *et al.* 2017) describiendo 30 nuevos *loci* significativos.

Hasta la publicación de la nueva fase de GWAS en esquizofrenia (fase dependiente del PGC y conocida como PGC3), el meta-análisis con un mayor número de muestras implicadas es el publicado en 2018 por Pardiñas y colaboradores (Pardiñas *et al.* 2018), en el que detectaron 50 nuevos *loci* asociados al trastorno de un total de 145 detectados. Además, los autores describen una presión selectiva sobre los genes enriquecidos en señales significativas.

Además de la detección de asociaciones significativas a nivel genómico, los GWAS en psiquiatría y en otros campos han dado lugar a la detección de variaciones en el número de copias (CNVs, *Copy Numer Variants*), cuantificación de la arquitectura mediante distintas aproximaciones, estimaciones de heredabilidad, detección de pleotropía (correlación y modelos de riesgo poligénico), análisis de relaciones causales o mapeados de alta resolución así como reconstruir la historia poblacional del trastorno mediante comparación de poblaciones y poniendo el foco en la detección de procesos selectivos.

1.2.4.8. Variación en el número de copias

Junto a los avances proporcionados por los GWAS, bajo el paraguas de las tecnologías de genotipado mediante chips creció la oportunidad de analizar a gran escala variaciones estructurales del genoma, con un gran peso de las CNVs. Este foco ha conducido a la investigación genómica a un gran número de hallazgos en psiquiatría a partir del año 2008 (Kirov 2015). Aunque diseñados para la detección de SNPs, mediante el análisis de la intensidad de las sondas del chip es posible inferir información sobre esta variedad de variación rara.

Esta clase de variación estructural se refiere a aquellos segmentos de ADN o regiones cromosómicas en las que se han observado, entre dos o más genomas, diferencias en el número de copias (Valsesia *et al.* 2013). Su tamaño suele ser mayor a 1000 pares de bases y no alcanzan tamaños que permitan su detección mediante estudios de cariotipo

clásicos (Kirov 2015). Su detección, aunque posible en gran medida gracias al proceso de genotipado de alta densidad, se complica con la existencia de CNVs de muy pequeño tamaño, donde los métodos de secuenciación permiten alcanzar una resolución máxima y abren nuevas oportunidades (Mills *et al.* 2011). Las CNVs, fuente evidente de diversidad genética humana, presentan dos clases principales: deleciones o duplicaciones. Los mecanismos más comunes mediante los cuales una CNV puede dar lugar a un fenotipo son: cambios en la dosis génica, efectos en regiones reguladoras, la interrupción de genes o el desenmascaramiento de mutaciones recesivas por deleciones (Stankiewicz & Lupski 2010). Este tipo de variación rara tiene un peso heterogéneo entre las distintas patologías psiquiátricas y suelen afectar a genes intolerantes a la pérdida de función.

La mayoría de las CNVs patogénicas son recurrentes, no azarosas, ya que ocurren en las mismas regiones cromosómicas (Kirov 2015). En general, los límites de las CNVs recurrentes se encuentran en regiones con alta homología, normalmente en regiones con secuencias repetitivas como las repeticiones de bajo número de copias (LCRs, *Low Copy Repeats*) (Stankiewicz & Lupski 2002). Estas regiones cromosómicas favorecen que los cromosomas se desalineen durante la recombinación, en lo que se conoce como recombinación homóloga no alélica (NAHR, *Non-Allelic Homologous Recombination*). Este mecanismo es la principal causa de la formación de CNVs en procesos de mitosis y meiosis (Malhotra & Sebat 2012). Además de las CNVs recurrentes, el genoma humano puede presentar CNVs distribuidas de forma más azarosa, “CNVs no recurrentes” (Gu *et al.* 2008). Esta clase de CNVs son producidas por otros mecanismos que, junto a NAHR conforman los principales mecanismos de variación estructural: recombinación no homóloga, retrotransposición mediada por L1 y parada de la horquilla, cambio de la hebra molde (FoSTeS) y NAHR (Lupski 2010; Malhotra & Sebat 2012). La tasa de mutación estimada de CNVs descritas como factor de riesgo para distintas patologías es relativamente alta, encontrándose entre ratios de 1:4000 y 1:25000 en recién nacidos (Lupski 2007).

En psiquiatría, hasta 2008, los únicos indicios de alteraciones estructurales se reservaban a esquizofrenia y autismo. La única CNV descrita como factor de riesgo a desarrollar esquizofrenia era la delección 22q11.2, de 2,3 mega-bases que implicaba la delección de más de 40 genes (Murphy *et al.* 1999). En autismo, la duplicación del *locus* 15q11-13 había sido descrita en algunos casos de autismo idiopático junto a alteraciones en cromosomas sexuales (Gillberg 1998). En 2008, dos artículos marcaron un hito, el confirmar la delección 22q11.2 y el identificar por primera vez tres delecciones (1q21.1, 15q11.2 y 15q13.3) que conforman riesgo a sufrir esquizofrenia al eliminar entre 4 y 11 genes cada una (International Schizophrenia Consortium 2008; Stefansson *et al.* 2008). Tras la detección de estas variantes, estudios posteriores que analizaban la existencia de CNVs, confirmaron el papel de esas variantes de manera casi universal. Este éxito en la replicación de resultados puede ser debido a que todas esas CNVs presentan unas altas tasas de mutación, una presión selectiva operando en contra de las mismas pudiendo explicar las tasas similares en distintas poblaciones y OR grandes que facilitan su detección estadística (Kirov 2015). Un ejemplo reside en el trabajo llevado a cabo por Kushima y colaboradores, en el que al analizar CNVs de riesgo a esquizofrenia en población japonesa, replicaron resultados encontrados en otras poblaciones humanas (Kushima *et al.* 2017). De forma más reciente, otro estudio ha permitido la detección de nuevas CNVs y la confirmación de otras descritas previamente, con un conjunto de 11 CNVs con asociación significativa de riesgo para esquizofrenia (Rees *et al.* 2014). Estas 11 variantes son las de mayor validación en esquizofrenia. Las evidencias que resaltan el importante papel de las CNVs en esquizofrenia son el enriquecimiento y asociación significativa de delecciones y duplicaciones raras en casos en comparación con grupos de individuos sanos y una mayor tasa de CNVs *de novo* en casos respecto a controles (Xu *et al.* 2008; Kirov *et al.* 2012). En términos generales, las CNVs confirmadas como factores de riesgo para esquizofrenia son de gran tamaño y afectan a un importante número de genes (Kirov 2015; Owen & O'Donovan 2017). Algunos de los genes afectados por las CNVs descritas en esquizofrenia están involucrados en la señalización neuronal, en la plasticidad

sináptica o la neurotransmisión gabaérgica (Kirov *et al.* 2012; Pocklington *et al.* 2015; Marshall *et al.* 2017).

Además de en esquizofrenia, distintas CNVs están consolidadas como factor de riesgo también en trastorno por déficit de atención e hiperactividad. Distintos estudios han detectado CNVs con implicaciones biológicas en la cadena de transporte de electrones, en la actividad del transportador transmembrana, proceso catabólico de los derivados de los carbohidratos o la actividad de canales iónicos entre otras (Faraone & Larsson 2018). Algunos análisis han encontrado evidencias de CNVs implicadas en rutas biológicas de regulación del funcionamiento inmune y el estrés oxidativo (Thapar *et al.* 2016). También han sido detectadas CNVs en “rendimiento escolar” (Männik *et al.* 2015) o trastorno bipolar (Green *et al.* 2016). Por último, los trastornos con una mayor implicación de CNVs en su arquitectura genética son los del espectro autista, discapacidad intelectual y otros síndromes del neurodesarrollo como Prader-Willi, Angelman o Williams-Beuren con CNVs de gran penetrancia específicas de este conjunto de trastornos o compartidas con esquizofrenia (Kirov *et al.* 2014). Este solapamiento con esquizofrenia será comentado en mayor profundidad en el siguiente sub-apartado. Algunas de estas CNVs muestran efectos en portadores sanos, ya que déficits cognitivos son apreciables en controles (Steffanson *et al.* 2014).

1.2.4.9 Pleiotropía y el modelo continuo del neurodesarrollo

Uno de los resultados de los estudios de GWAS que han despertado gran interés reside en la detección de *loci* asociados significativamente en distintos rasgos o fenotipos. Además de la variación común, las variantes raras también pueden ser susceptibles de aportar efectos pleiotrópicos. Estos resultados se pueden observar tanto en patologías psiquiátricas como en un enorme número de rasgos humanos (Solovieff *et al.* 2013; Smoller 2013; Owen 2014). En psiquiatría, los ejemplos son numerosos y prácticamente extendidos desde la primera detección de una variante significativa a nivel genómico en esquizofrenia, que mostraba mejora en el grado de

asociación tras la inclusión de casos con trastorno bipolar junto a individuos con esquizofrenia (O'Donovan *et al.* 2008). En algunos casos las mismas variantes muestran asociación con múltiples rasgos y, en otros, distintos marcadores cercanos entre sí o co-localizados muestran asociación entre distintos rasgos. Distinguir qué asociaciones representan efectos realmente compartidos es de gran importancia, debido a los mecanismos de acción implicados (Solovieff *et al.* 2013). Los distintos tipos de pleiotropía detectada en GWAS, resumidos por Solovieff y colaboradores en 2013 y O'Donovan y Owen en 2016 (O'Donovan & Owen 2016), son los siguientes:

- Pleiotropía biológica: una variante genética o gen presenta influencia biológica sobre más de un fenotipo distinto. Esta clase puede responder distintos escenarios: “pleiotropía biológica al nivel de alelo” (la variante causal afecta a ambos fenotipos), “Asociación co-localizante” (donde la variante genética observada en asociación se encuentra en LD con dos variantes causales en el mismo gen que afectan a distintos fenotipos) y “pleiotropía biológica a nivel génico” (dos variantes causales independientes en el mismo gen afectan a diferentes fenotipos).
- Pleiotropía mediada: un fenotipo está, en sí mismo, relacionado causalmente con un segundo fenotipo, por lo que la variante del primer fenotipo está asociada indirectamente con el segundo.
- Pleiotropía espuria o pseudo-pleiotropía: una variante no causal está asociada con múltiples fenotipos. Esta clase puede responder a distintos escenarios: por co-localización (donde dos fenotipos están afectados por dos variantes causales en genes distintos en fuerte LD) o porque la variante causal afecta solamente al primer fenotipo pero el segundo está relacionado con el primero por un mal diseño del estudio, sesgo de publicación o de clasificación del fenotipo.

La hipótesis de verdadera pleiotropía en psiquiatría parece estar presente tanto en CNVs como en variantes comunes. Aunque en el caso

de las CNVs muchas podrían responder a pleiotropía espuria por el gran número de genes que engloban. La detección de variantes de gran tamaño intragénicas asociadas a varios fenotipos psiquiátricos es clave para poder establecer una posible pleiotropía real (O'Donovan & Owen 2016). Este es el caso de del gen *NRXN1*, afectado por la CNV 2p16.3, asociado a autismo, esquizofrenia y discapacidad intelectual (Kirov *et al.* 2008). Además, la acumulación de distintas CNVs de gran tamaño asociadas siempre a un mismo conjunto heterogéneo de fenotipos añaden peso a la hipótesis de una pleiotropía real (Rujescu 2009; Malhotra & Sebat 2012; O'Donovan & Owen 2016). En esta dirección, un importante estudio demostró en 2014 cómo aquellas CNVs reconocidas como factor de riesgo a desarrollar esquizofrenia estaban en realidad más asociadas con otros fenotipos: trastorno del desarrollo, trastornos del espectro autista y malformaciones congénitas (Kirov *et al.* 2014). En el trabajo, los autores estimaban la penetrancia de la delección 22q11.21 y la CNV que afecta al gen *NRXN1*, previamente comentada, con valores de 0,12 y 0,064 respectivamente en esquizofrenia y 0,88 y 0,26 en el caso del conjunto de trastornos asociados al neurodesarrollo o trastornos de inicio temprano. Estos resultados fueron sorprendentes en el campo de la genética psiquiátrica, debido a que el vínculo genético entre estas patologías no había sido descrito con tanto peso en los estudios familiares o de gemelos (Kirov 2015). Otros autores han detectado enriquecimiento en duplicaciones en trastorno por déficit de atención e hiperactividad previamente descritas e implicadas en esquizofrenia y, en menor medida, en trastornos del espectro autista (Thapar *et al.* 2016). En el caso del trastorno bipolar, las CNVs descritas también han sido asociadas a esquizofrenia (Green *et al.* 2016). Este conjunto de resultados, alejan la pleiotropía mediada como causa de esta co-ocurrencia en distintos trastornos (O'Donovan & Owen 2016) y señalan en la dirección en la que la esquizofrenia y un conjunto de trastornos del neurodesarrollo son causados por las mismas variantes raras patogénicas (Owen & O'Donovan 2017).

Este solapamiento de variantes patogénicas de gran efecto entre trastornos infantiles del neurodesarrollo (discapacidad intelectual, trastornos del espectro autista y trastorno por déficit de atención e

hiperactividad) y trastornos psiquiátricos adultos (incluyendo esquizofrenia y trastorno bipolar) puede reflejar un continuo o espectro etiológico y del desarrollo neurológico. La hipótesis del modelo continuo del neurodesarrollo considera un marco en el que estos distintos trastornos “representan una diversa gama de resultados que derivan de un desarrollo cerebral alterado” (Owen & O’Donovan 2017). Esta hipótesis considera los factores de riesgo ambiental y la genética compartida para proponer la existencia de mecanismos patogénicos comunes. En este modelo, los principales síntomas clínicos reflejan la gravedad, el patrón de desarrollo cerebral anormal y las principales características funcionales resultantes (Craddock & Owen 2010; Owen *et al.* 2011; Owen 2012; Owen & O’Donovan 2017).

El origen parcial de este modelo se encuentra en la hipótesis del neurodesarrollo de la esquizofrenia (Weinberger 1987; Murray & Lewis 1988). A pesar de la actual aceptación general del peso del neurodesarrollo en la esquizofrenia (Birnbaum & Weinberger 2017), este trastorno sigue diferenciándose nosológicamente de los otros síndromes infantiles con importante papel del neurodesarrollo. En cambio, son múltiples los factores que evidencian cierto solapamiento. Éstos se resumen en deficiencias cognitivas compartidas, mayor frecuencia en hombres, retraso en el desarrollo, anomalías motoras, comorbilidad significativa de alguno de los mismos y la genética compartida (Craddock & Owen 2010; Owen & O’Donovan 2017). Además, el concepto de espectro o continuo también muestra influencias de las propuestas previas realizadas para trastornos del neurodesarrollo infantil (Lilienfeld *et al.* 1955; Capute & Palmer 1980).

El modelo continuo del neurodesarrollo permite considerar las distintas patologías en un gradiente del déficit del neurodesarrollo (Owen & O’Donovan 2017). Este gradiente considera la edad de inicio, el deterioro cognitivo, la gravedad general del trastorno o la persistencia del deterioro funcional. Basándose en evidencia empíricas, los distintos trastornos psiquiátricos, son ordenados por el grado de decrecimiento del déficit del neurodesarrollo que presentan: discapacidad intelectual, trastornos del espectro autista, trastorno por déficit de atención e hiperactividad, esquizofrenia y trastorno bipolar. Las variantes

patogénicas asociadas a los distintos trastornos permiten ordenarlos en el gradiente propuesto, apoyándose en datos empíricos de penetrancia, solapamiento y carga por individuo (Owen & O'Donovan 2017).

Los resultados significativos a nivel genómico del análisis de variantes comunes mediante GWAS ahondan también en un pleiotropismo extendido. Tras las primeras evidencias obtenidas por O'Donovan y colaboradores (O'Donovan *et al.* 2008), otros GWAS analizaron conjuntamente esquizofrenia, trastorno depresivo mayor y trastorno bipolar evidenciando el papel de los genes *ZNF804A*, *CACNA1C*, *NEK4/ITIH1,3,4* o *NCAN* (Green *et al.* 2010; Schizophrenia Psychiatric Genome-Wide Association Study GWAS Consortium 2011; Sklar *et al.* 2011; Williams *et al.* 2011; Mühleisen *et al.* 2012) (algunos de ellos han sido desarrollados previamente). Uno de los trabajos de mayor peso histórico analizó conjuntamente cinco trastornos psiquiátricos: esquizofrenia, trastorno bipolar, trastornos del espectro autista, trastorno por déficit de atención e hiperactividad y depresión mayor (Cross-Disorder Group of the Psychiatric Genomics Consortium 2013). En este estudio, cuatro *loci* significativos fueron descritos en los análisis de los cinco trastornos combinados: dos en las regiones cromosómicas 3p21 y 10q24 que incluyen un gran número de genes y dos asociadas a dos subunidades de canales de calcio, *CACNA1C* y *CACNB2*. Este tipo de estudios señalan hacia un equilibrio entre la genética específica de trastorno y una serie de variantes potencialmente comunes para el conjunto de los trastornos psiquiátricos, subrayando los posibles mecanismos comunes involucrados en el curso de los trastornos psiquiátricos. En este mismo sentido, otra de las iniciativas del PGC se enfocó en la interpretación funcional de las señales de asociación de GWAS mediante el análisis de posibles rutas moleculares implicadas en los cinco trastornos analizados en la iniciativa Cross-Disorder (Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium 2015). Integrando datos de los grandes GWAS en los distintos trastornos, los autores buscaron mediante un nuevo método robusto la detección de estas rutas compartidas entre trastornos, sin las limitaciones típicas de los estudios clásicos como las producidas por la heterogeneidad alélica o solapamiento parcial de variantes entre conjuntos de datos. Los

resultados previos en canales de calcio fueron significativos nominalmente en este estudio, confirmando de nuevo el posible papel de la actividad de estos canales en las patologías psiquiátricas. Los análisis de rutas biológicas específicas de trastorno mostraron una mayor correlación de las señales detectadas entre esquizofrenia y trastorno bipolar (aunque cada trastorno mostró diferencias en las asociaciones más robustas: metilación de histonas en trastorno bipolar y procesos sinápticos y postsinápticos en esquizofrenia). Al analizar conjuntamente los trastornos psiquiátricos adultos (esquizofrenia, trastorno bipolar y trastorno depresivo mayor) las señales significativas adquirieron una asociación más robusta, permitiendo la descripción de aspectos biológicos que comprenden estos trastornos. Las asociaciones significativas compartidas involucran principalmente rutas asociadas al sistema inmune, sinapsis neuronal o metilación de histonas, pudiendo estar relacionadas con una vulnerabilidad prenatal a distintos efectos ambientales.

Poniendo el foco en mega-análisis como el realizado en esquizofrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014), los resultados de variantes significativas muestran diversos ejemplos descritos en estudios previos en distintos trastornos o rasgos como el caso de los principales trastornos del ánimo (con un gran número de *loci* asociados a trastorno bipolar) o en señales asociadas a distintos fenotipos del consumo de tabaco como la dependencia a nicotina (CHRNA3, CHRNA5, y CHRNA4 son señales no independientes situadas en la citobanda 15q25) (Thorgeirsson *et al.* 2008; Tobacco and Genetics Consortium 2010). Otro ejemplo reside en el mega-análisis realizado en trastorno depresivo mayor (Wray *et al.* 2018), donde al comparar sus resultados con los del mega-análisis de esquizofrenia detectaron seis *loci* compartidos entre los que varios se encuentran en la región del MHC o ligados a TCF4, esencial para un normal desarrollo cerebral. Además, este estudio confirmó el papel de los canales de calcio (*CACNA1E* y *CACNA2D1*), neurotransmisión dopaminérgica (*DRD2*) y glutamatérgica (*GRIK5* y *GRM5*) involucrados en diversos trastornos psiquiátricos. Además, los análisis de rutas biológicas detectaron asociaciones descritas previamente mediante GWAS o análisis de variación exónica rara en esquizofrenia

y autismo (De Rubeis *et al.* 2014; Genovese *et al.* 2016). Estos ejemplos, aunque representan una pequeña fracción del número total de detecciones, permiten observar a grandes rasgos la extensión pleiotrópica en psiquiatría. Además de señales asociadas a distintos fenotipos psiquiátricos, algunos de los resultados significativos en psiquiatría han sido descritos en patologías o rasgos no psiquiátricos. Un ejemplo son los *loci OLFM4* y *NEGR1* asociados al trastorno depresivo mayor, previamente descritos en obesidad o índice de masa corporal (Speliotes *et al.* 2010; Locke *et al.* 2015).

Este conjunto de ejemplos de variantes significativas detectadas por GWAS señalan hacia una biología compartida generalizada (Smoller *et al.* 2018). ¿Qué ocurre con aquellas variantes analizadas en GWAS que no han superado el umbral de significación? En el caso de las variantes comunes asociadas pero no significativas a nivel genómico con un pequeño efecto cada una, parece observarse un efecto pleiotrópico en masa que ahonda más en esta susceptibilidad compartida (Sullivan *et al.* 2012) que será comentado más adelante, en la introducción a los modelos poligénicos.

1.2.4.10 Heredabilidad perdida

Los GWAS han demostrado ser una aproximación de éxito a la genética de los trastornos complejos, identificando un inmenso número de variantes asociadas que las aproximaciones previas no pudieron detectar, pero los cálculos de su contribución a la varianza en la escala de susceptibilidad es limitada (Manolio *et al.* 2009; Wray & Maier 2014). La contribución de cada *locus* significativo a nivel genómico puede ser sumada para la estimación conjunta de la proporción de varianza explicada (h^2_{GWA} ; heredabilidad *Genome-Wide Association*) (Wray & Maier 2014). El problema reside en que el cálculo de h^2_{GWA} en cualquier rasgo o patología compleja refleja una pequeña fracción de la heredabilidad en sentido estricto (h^2) estimada a partir de estudios familiares o de gemelos. Esta ausencia se denominó como “heredabilidad perdida” (Maher 2008; Manolio *et al.* 2009). Inicialmente, esta situación provocó que parte de la comunidad

científica pusiese parcialmente en duda el coste-efectividad de los GWAS (Goldstein 2009), posición que hoy en día ya casi no es sostenida por los éxitos alcanzados y el abaratamiento de la tecnología. La heredabilidad perdida se encuentra en todos los rasgos y patologías estudiadas por GWAS, incluidos los trastornos psiquiátricos.

Además de la h^2_{GWA} , otro método de cálculo empleando la totalidad de variantes incluidas en un chip para GWAS (Yang *et al.* 2010) permitió la aproximación que estima el porcentaje de varianza explicada de forma conjunta por h^2_{GWA} y el resto de variantes que no han superado el umbral de significación genómica, surgiendo el término “heredabilidad basada en SNPs” o “heredabilidad de chip” (h^2_{SNP}). La heredabilidad basada en SNPs es mayor que la basada en variantes significativas a nivel genómico, pero inferior a la calculada a partir de estudios de gemelos ($h^2_{GWA} < h^2_{SNP} < h^2$). El concepto de h^2_{SNP} ha sufrido cierta evolución desde su aparición y las metodologías de cálculo han ido diversificándose. La comparación de las contribuciones individuales de cada variante al riesgo a desarrollar la patología, el contraste de la contribución total de variantes asociadas o analizar la contribución adicional comparada con estudios previos es un procedimiento extendido y útil. Pero existe cierta heterogeneidad entre la distintas metodologías de cálculo (técnicas que han sido validadas, comparadas e incorporadas) (Witte *et al.* 2014; Speed *et al.* 2017; Yang *et al.* 2017; Evans *et al.* 2018). Una estimación acertada de h^2_{SNP} es clave para la resolución del debate de la heredabilidad perdida, ya que indica la potencial utilidad de la predicción basada en SNPs (Speed *et al.* 2017). La dificultad y diferencias en las distintas estimaciones residen en la alta sensibilidad de esta medida a las frecuencias, tamaño del efecto y niveles de LD de las variantes causales. La asunción de resultados debe tener en cuenta las diferencias entre metodologías y las sensibilidades de cada una, con el fin de una correcta interpretación de las medidas obtenidas (Evans *et al.* 2018).

Inicialmente surgieron distintas propuestas para explicar la heredabilidad perdida: presencia de un gran número de variantes comunes de pequeño efecto que todavía no habían sido detectadas, factores epigenéticos, variación genética no aditiva, variantes raras de

gran efecto no incluidas en estudios de GWAS e incluso estimaciones exageradas de la h^2 a partir de estudios de gemelos (Manolio *et al.* 2009; Eichler *et al.* 2010; Zhu *et al.* 2015; Tropf *et al.* 2017). Pero el camino iniciado por los GWAS en el campo de la salud mental acababa de dar solamente los primeros pasos y los esfuerzos colaborativos con los consiguientes aumentos en los tamaños muestrales y nuevas aproximaciones aportaron algo de luz (Visscher *et al.* 2012b “Five years of GWAS discovery.”; Horwitz *et al.* 2017). Los modelos de riesgo poligénico juegan un papel clave en este aspecto, gracias a su capacidad de ampliar el espectro de heredabilidad capturada al no poner únicamente el foco en las variantes asociadas significativamente a nivel genómico. Por último, cabe indicar el papel jugado por los estudios enfocados al análisis de variantes raras. Sus primeros pasos en psiquiatría serán comentados en el siguiente sub-apartado.

1.2.5 Estudios de secuenciación

El uso de tecnologías de secuenciación masiva o de nueva generación (NGS, *Next-Generation Sequencing*) es reciente. Antes de la llegada de las mismas, el escudriñamiento del genoma se basaba en métodos fiables pero con un gasto económico y temporal grande. La secuenciación Sanger (Sanger *et al.* 1977), método más extendido, permitió la caracterización del genoma humano y un gran número de hitos en la historia de la investigación genética. El empleo de NGS comenzó en 2004 y se diversificó a través de distintas metodologías de análisis disponibles en el mercado. Con ellas, el progreso se extendió gracias a distintos factores o ventajas: una mayor velocidad metodológica, mejoras en el rendimiento y longitud de lectura y un abaratamiento masivo del coste por base secuenciada (Liu *et al.* 2012; van Dijk *et al.* 2014). Aunque las tecnologías NGS presentan diversas ventajas, todavía exhiben ciertos inconvenientes como el manejo de secuencias cortas. En este proceso evolutivo tecnológico, las tecnologías han ido evolucionando hasta llegar a las de tercera generación o de lectura larga, que pueden obtener secuencias de mayor calidad, secuenciar de forma completa el transcriptoma o detectar modificaciones epigenéticas (van Dijk *et al.* 2018). El avance de los

GWAS en los primeros años, aunque aportase resultados significativos en ciertas variantes comunes, no satisfacía a una parte de la comunidad investigadora y muchos pusieron el foco en la búsqueda de la variación rara (Cirulli & Goldstein 2010).

En psiquiatría, ante el pequeño éxito de las primeras fases de GWAS, el debate CDCV-CDRV se vio incrementado, con algunos autores alzando la voz por una arquitectura predominante centrada en variación rara y una visión investigadora más enfocada en el análisis de este tipo de variación (McClellan *et al.* 2007; Goldstein 2009). Además de esta discusión sobre la arquitectura genética subyacente, el interés por caracterizar regiones genéticas concretas o de forma masiva también se vio incrementado por el abaratamiento de las tecnologías. Aunque una pequeña parte de las variantes incluidas en chips para el desarrollo de GWAS puede estar reservada para el análisis de variación rara, éstas solamente estarán presentes en pocos individuos o un único individuo del conjunto analizado (Gelernter 2015). El uso de la secuenciación era prometedor, aunque arduo (Geschwind & Flint 2015) debido a la inherente baja frecuencia de las mismas, los costes computacionales y económicos.

En psiquiatría, al igual que en un gran número de patologías comunes, la estrategia de análisis siguió dos vías principales: estudios caso-control y estudios de mutaciones *de novo* en tríos. Ambas estrategias pueden ser llevadas a cabo por distintas aproximaciones: la secuenciación de regiones o genes de interés seleccionadas a través del conocimiento previo adquirido por estudios de GWAS (pudiendo incluir hipótesis de genes candidatos clásicos), la secuenciación de exoma (WES, *Whole-Exome Sequencing*; alcanzando el 1-1,5% del genoma) y la secuenciación de genoma completo (WGS, *Whole-Genome Sequencing*; obteniendo el total de la información genética de un individuo). El éxito de estos análisis en psiquiatría es desigual y dependiente de patología. Aquellos trastornos psiquiátricos con una arquitectura genética híbrida, basada en variantes de mayor efecto y una contribución de variación común, han logrado resultados de interés (Sullivan *et al.* 2018). Por ejemplo, en trastornos del espectro autista, más de 50 genes han sido identificados mediante WES (Sanders *et al.*

2017). La mayoría de las variantes conformadoras de riesgo de estos genes son muy raras o ultra raras con efectos grandes, truncantes de proteínas que pueden indicar las rutas moleculares o biológicas subyacentes de este conjunto de trastornos. WES es, además, una aproximación de estudio con costes más reducidos que la basada en WGS. Las ventajas de WGS en comparación con WES y GWAS son la capacidad de detectar variantes no codificantes de mayor efecto, variantes estructurales y mejorar la detección de resultados de variantes previamente estudiadas en GWAS (SNPs y CNVs inferidas) (Owen *et al.* 2016; Collins *et al.* 2017).

La detección de variantes raras en psiquiatría mediante estas tecnologías ha sido compleja (Smoller *et al.* 2018), pero numerosos ejemplos muestran su potencial en varios trastornos. En trastorno por déficit de atención e hiperactividad, Demontis y colaboradores (Demontis *et al.* 2016) analizaron una pequeña muestra caso-control secuenciando 52 genes candidatos, detectando 22 variantes raras asociadas. En el mismo trastorno, Corominas y colaboradores (Corominas *et al.* 2018) analizaron mediante genotipado o WES a 70 miembros de tres familias distintas con múltiples individuos afectados. Tras seleccionar 24 genes de interés detectados en la muestra multifamiliar, estudiaron en una muestra independiente con genotipado de exoma (N=9365) mediante análisis de conjuntos de genes funcionalmente relacionados (*gene-set*) si el conjunto de genes estaba asociado con trastorno por déficit de atención e hiperactividad, detectando resultados positivos. En el caso del síndrome de Tourette, Willsey y colaboradores (Willsey *et al.* 2017), estudiaron un total de 511 tríos mediante WES detectando cuatro genes con múltiples variantes *de novo* de riesgo y lograron estimar el riesgo conferido por variantes dañinas *de novo* en aproximadamente 400 genes en el 12% de los pacientes. En trastorno bipolar, dos estudios fueron publicados en 2016. El llevado a cabo por Goes y colaboradores (Goes *et al.* 2016) se basó en una primera aproximación mediante WES en 36 pacientes de ocho familias, seguida de un análisis de asociación en tres muestras independientes caso-control (N casos=3541; N controles=4774). Los autores detectaron 84 variantes raras en 82 genes, de las cuales parte fueron descritas previamente en distintos trastornos psiquiátricos

(esquizofrenia, trastornos del espectro autista y síndrome de X frágil). Los resultados obtenidos mediante el meta-análisis de las aproximaciones caso-control, logrados resultados nominalmente significativos en 19 genes. Kataoka y colaboradores (Kataoka *et al.* 2016) detectaron, mediante WES de tríos, 71 mutaciones *de novo* en 79 casos. Además, analizaron conjuntamente los resultados con datos de trastorno esquizoafectivo. Los resultados confirmaron el peso de las mutaciones de pérdida de función y aquellas alteradoras de proteínas y el enriquecimiento de parte de estas variantes en el análisis combinado con el trastorno esquizoafectivo.

Estos resultados, interesantes por la detección de variantes raras que segregan en familias, mutaciones *de novo* y que señalan posibles mecanismos biológicos, son escasos en comparación con los obtenidos en esquizofrenia y trastornos del espectro autista. En estas patologías, el número de individuos analizado es mucho mayor, logrando una mayor caracterización de estas variantes raras. En trastornos del espectro autista, un gran número de estudios realizados mediante WES y secuenciación dirigida a genes candidatos resaltan la importancia de la variación *de novo* alteradoras de proteínas (Iossifov *et al.* 2012; Neale *et al.* 2012; O’Roak *et al.* 2012; Sanders *et al.* 2012; De Rubeis *et al.* 2014; Sanders *et al.* 2015; Stessman *et al.* 2017) en un gran número de CNVs y genes asociados con implicaciones biológicas como la formación sináptica, canales iónicos dependientes de voltaje, histonas, cromatina, regulación transcripcional, etc. El peso de las regiones codificantes, parece ser mayor a las no codificantes en las primeras aproximaciones para la generación de un marco de trabajo en WGS (Werling *et al.* 2018). Además, mediante WGS un estudio ha detectado una media de 73,8 variaciones de un único nucleótido *de novo* (SNVs; *Single Nucleotide Variants*) y 12,6 CNVs *de novo* por caso de trastorno del espectro autista tras analizar más de 5200 individuos (C Yuen *et al.* 2017). Analizando de forma conjunta los trastornos del neurodesarrollo (Deciphering Developmental Disorders Study 2017), se ha detectado mediante WES una alta carga de variantes *de novo* en secuencias codificantes y disruptivas de la función génica.

En esquizofrenia, la contribución de variantes raras detectadas mediante secuenciación es reseñable pero de menor grado que en los trastornos del neurodesarrollo (Smoller *et al.* 2018). En 2014, dos estudios basados en WES (Fromer *et al.* 2014; Purcell *et al.* 2014). Fromer y colaboradores no detectaron en casos un aumento total de mutaciones *de novo* que causasen pérdida de función, pero sí resultados significativos al poner el foco en genes candidatos asociados previamente a esquizofrenia (involucrados en proteínas post-sinápticas glutamatérgicas y asociados a complejos involucrados en el síndrome de X frágil). Las mutaciones *de novo* de pérdida de función sí mostraban una representación significativa cuando se seleccionaban pacientes con un menor éxito escolar (Fromer *et al.* 2014). Purcell y colaboradores detectan resultados en la misma dirección que el estudio comentado previamente y variantes involucradas en canales de iones de calcio. Además, este estudio permitió confirmar la existencia de un componente poligénico conformado por variantes de baja frecuencia, en la misma dirección que los resultados detectados en estudios de GWAS pero representando una pequeña fracción de la varianza en la escala fenotípica explicada por las variantes comunes (Purcell *et al.* 2009; Purcell *et al.* 2014; Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014). Más recientemente, en 2016, los resultados de un análisis de más de 12000 individuos mediante WES en población sueca han sido publicados (Genovese *et al.* 2016). Este estudio, enfocado al análisis de variantes ultra-raras, ha detectado un enriquecimiento de las mismas en casos respecto a controles sin la asociación robusta de un gen concreto, resaltando la arquitectura poligénica del trastorno. Estas variantes estaban concentradas en genes de expresión en cerebro. Al igual que en el caso de Purcell y colaboradores, el porcentaje de varianza explicada es menor que el desentrañado por los modelos poligénicos basados en variantes comunes. Por último, dos de los estudios más recientes reseñables en esquizofrenia, son llevados a cabo por Singh y colaboradores (Singh *et al.* 2016) y Rees y colaboradores (Rees *et al.* 2018). El primero, combinando sus datos de esquizofrenia con los de Purcell y colaboradores, ha permitido confirmar la presencia de variantes de pérdida de función en el gen *SETD1A* tanto en esquizofrenia como al

analizar fenotipos del conjunto de trastornos del neurodesarrollo en muestras secundarias, confirmando los resultados previos en esquizofrenia (Takata *et al.* 2014). Rees y colaboradores han presentado los resultados de la secuenciación dirigida de 187 genes asociados a esquizofrenia en 5207 casos y 4991 controles. Tras meta-analizar los datos con estudios previos y sumar 11319 casos y 15854 controles junto a 1136 tríos, ningún gen estaba asociado significativamente con esquizofrenia. En cambio, sí han obtenido evidencias de la posible implicación de múltiples variantes que afectan a distintos canales de sodio y complejos post-sinápticos mediante análisis de conjuntos de genes funcionalmente relacionados.

Los resultados arrojados por los distintos estudios de secuenciación de variantes raras muestran un patrón parecido al encontrado en CNVs, con un gradiente similar al de la hipótesis del continuo del neurodesarrollo comentado previamente (Owen & O'Donovan 2017). El gradiente en variantes raras secuenciadas muestra una mayor carga generalizada de variantes raras y *de novo* en trastornos del neurodesarrollo respecto a esquizofrenia, una mayor carga de variantes en pacientes esquizofrénicos con peor rendimiento cognitivo, un solapamiento de variantes entre trastornos y un gradiente en la misma dirección del grado de conservación de las variantes (Owen & O'Donovan 2017).

Una pequeña parte de la heredabilidad perdida en los trastornos psiquiátricos puede y es explicada por las variantes raras detectadas mediante estudios de secuenciación (Wray & Maier 2014), pero hasta la fecha los resultados señalan hacia un modelo de enfermedad basado en la hipótesis CDCV (Sullivan *et al.* 2018). Existe una contribución variable de la variación rara a lo largo de los trastornos pero, en la actualidad el peso y potencial de la variación común es punta de lanza de la investigación de la genómica psiquiátrica.

1.2.6 Puntuaciones de riesgo poligénico

1.2.6.1 Introducción

El debate genético sobre cómo la variación genotípica contribuye a la fenotípica es de largo recorrido (Boyle *et al.* 2017). En sus orígenes, el debate enfrentaba a dos grandes grupos: “mendelianos” (liderados por Bateson) vs “biométricos” (liderados por Pearson) (Kendler 2015; Boyle *et al.* 2017). Tras varios intentos de acercamiento, Fisher publicó en 1918 la integración estadística de ambas teorías (Fisher 1918). El modelo, conocido como “modelo infinitesimal”, propone un sistema de herencia de los caracteres compuesto por un gran número de factores mendelianos con efectos aditivos. A medida que el número de genes crece, la contribución de cada uno es más pequeña. Este riesgo genético aditivo presentaría una distribución normal. Aunque el modelo no eliminó el debate, sí creó un puente metodológico entre ambas posiciones científicas (Kendler 2015). En 1965 y en el marco de la genética cuantitativa, Falconer publicó un trabajo estadístico clave, en el que proponía una distribución continua de la causa de las enfermedades (Falconer 1965). El modelo propuesto es conocido como “modelo umbral de predisposición” (*Liability threshold model*). En el marco del “todo o nada” de una enfermedad, el modelo planteado propone la existencia de una graduación subyacente de susceptibilidad integrada tanto por contribución genética (“tendencia innata”) como ambiental (“combinación de circunstancias externas”) (Falconer 1965). El punto de la escala por encima del cual todas las personas se ven afectadas por la enfermedad o trastorno es el reconocido como umbral. La conexión entre psiquiatría y la postura integradora está representada en parte por el trabajo que propuso por primera vez una teoría poligénica en la esquizofrenia (Gottesman & Shields 1967; Kendler 2015). Esta teoría, basada en el modelo poligénico propuesto por Fisher y aplicado a los trastornos por Falconer, es aplicable a todos los trastornos psiquiátricos comunes. En el año 2009, 42 años más tarde de la propuesta realizada por Gottesman y Shields, Purcell y colaboradores (Purcell *et al.* 2009) presentaron la primera aproximación poligénica basada en GWAS en esquizofrenia.

Este trabajo, comentado de forma extendida en un subapartado, supuso el punto de partida para las estrategias basadas en la generación de puntuaciones de riesgo poligénico (PRS, *Polygenic Risk Scores*).

Los GWAS, fundamentados en este modelo (Mitchell 2012), conforman la continuación metodológica al modelo propuesto por Gottesman y Shields. Pero su extensión reside, de forma más completa, en los modelos de riesgo poligénico. Tras observar la poligenicidad basal en las primeras aproximaciones basadas en GWAS, se buscó el desarrollo de una metodología que permitiese capturar y manejar esta información (Wray *et al.* 2014). A diferencia de la estrategia GWAS, el foco de la aproximación y método poligénico no está centrado en la detección de asociaciones altamente significativas, sino que se centra en el empleo del conjunto de variantes que no tienen que haber superado el umbral de significación estadístico aplicado en GWAS.

Las estadísticas resumen obtenidas a partir de GWAS son la base para la generación de modelos de riesgo poligénico. La muestra empleada para la estimación de los efectos individuales y el nivel de asociación de cada variante considerada para el modelo es denominada “descubrimiento”. La muestra empleada para generar PRS individuales (para cada individuo que la conforma) es conocida como muestra “diana”. Esta muestra diana presentará información genotípica individual, con variantes presentes en común con la muestra descubrimiento y estará conformada por un conjunto distinto de individuos. En cada individuo se genera una puntuación de riesgo basada en el conteo de alelos de riesgo presentes y su tamaño del efecto. El fenotipo diana puede diferir del considerado para la conformación del modelo y el conjunto de puntuaciones generadas deberá ser evaluado mediante una aproximación estadística teniendo en cuenta otras covariables (Wray *et al.* 2014).

1.2.6.2 Origen de la predicción de riesgo

La predicción genética ha tenido un papel de gran importancia e importante recorrido en agricultura y ganadería, con estimaciones de valores de reproducción desarrolladas antes que la predicción de riesgo

en humanos. El concepto de riesgo poligénico se estableció en la genética humana clásica y desde ese momento ha sido aplicado en distintas estrategias, como los análisis de segregación de rasgos complejos (Risch 1990), o tras la revolución acercada por los GWAS, en el cálculo de las heredabilidades basadas en gemelos (Visscher *et al.* 2008). Los GWAS, despertaron una nueva vía para la predicción, gracias a su naturaleza de análisis extendido y denso de variantes comunes y al no detectar variantes asociadas a nivel genómico durante sus primeros pasos. Bajo este paraguas nacieron las primeras propuestas de predicción de riesgo empleando esta arquitectura poligénica analizada (Wray *et al.* 2007; Daetwyler *et al.* 2008).

1.2.6.3 Consideraciones: tamaño muestral, calidad y estándares

Una de las piezas clave a considerar en los análisis de cálculo de PRS es el tamaño muestral de la muestra descubrimiento y diana, ya que la primera establece la capacidad predictiva en función de la precisión del cálculo del efecto de cada variante y de la segunda depende la capacidad de detección de una varianza significativa (Wray *et al.* 2014).

Cuando la muestra diana tiene un tamaño suficiente para la detección de varianza explicada, el peso recae en la muestra descubrimiento. Cuanto mayor es el tamaño, el efecto de cada variante se refina reduciendo el ruido general. Uno de los artículos más importantes para la estimación del poder y precisión de las PRS fue publicado por Dudbridge en 2013. El autor estimaba, en distintos escenarios, los valores de tamaño óptimos para la estimación ajustada de PRS, aportaba un método de cálculo de potencia, evaluaba los umbrales de inclusión de marcadores, analizaba la heredabilidad explicada y distintas vías de ponderación de los tamaños de los efectos. El trabajo establecía las condiciones ideales de cálculo en el que, con un tamaño muestral diana de 2000 casos y 2000 controles, permitirían realizar las PRS con acierto. Además, indicaba el menor éxito del sistema de aproximación basado en validaciones cruzadas entre

subconjuntos generados a partir de la muestra original. El autor demostró que los resultados “desalentadores” de estudios previos se deben al bajo número de individuos analizados y la necesidad de grandes tamaños muestrales para una correcta estimación del riesgo genético individual (Dudbridge 2013). Los científicos que trabajan con esta clase de herramientas genómicas suelen valorar en gran medida la posibilidad de inferir potencia de cálculo. Existen diversas metodologías de interés que permiten estimar el poder predictivo (Chatterjee *et al.* 2013; Palla & Dudbridge 2015).

Existe cierto consenso sobre la preparación, procedimientos y filtros para la realización de PRS. Aunque cada aproximación requiere considerar la variabilidad, características y necesidades intrínsecas a cada análisis en su puesta a punto individualizada, una serie de controles son extensibles a la mayoría de abordajes poligénicos. Existen pocas guías o revisiones que asienten las bases y pautas a seguir, pero en la literatura pueden encontrarse varios ejemplos que, desde una perspectiva técnica o práctica, analizan la utilidad y puntos de consenso en este tipo de análisis (Dudbridge 2013; Wray *et al.* 2014; Chatterjee *et al.* 2016; Dudbridge 2016; Anderson *et al.* 2018; Bogdan *et al.* 2018). En términos generales, el poder predictivo y la validez de los análisis PRS dependen de dos factores principales: la calidad y características de las muestras descubrimiento y diana. La implementación de ambos conjuntos de datos puede realizarse siguiendo los altos estándares de calidad establecidos en las estrategias GWAS (Anderson *et al.* 2010) mediante el uso de una herramienta bioinformática. Una de las más extendidas es PLINK (Purcell *et al.* 2007; Chang *et al.* 2015), de código abierto basada en lenguaje C/C++ permite realizar extensivos controles de calidad en datos genómicos, además, la segunda generación del programa presenta importantes mejoras generales en el funcionamiento y compatibilidad. Además de los filtros de calidad clásicos de GWAS aplicados por individuo y por marcador genético en la muestra diana (frecuencia del alelo menor, equilibrio Hardy-Weinberg, porcentaje de información genotípica perdida, calidad de la imputación de variantes, etc.) es de gran utilidad considerar otras cuestiones más específicas de los análisis PRS, siendo críticas varias de ellas:

- Descarga u obtención de la muestra descubrimiento: Las estadísticas resumen obtenidas de páginas/sitios web deben provenir siempre de repositorios oficiales, será necesario comprobar si se presenta más de una versión de los datos de interés con el fin de elegir la que se ajusta a nuestro análisis y será necesario comprobar la integridad de los datos así como cualquier documento secundario con información relevante.

- Ancestralidad: La muestra descubrimiento y la muestra diana deben mostrar mismo origen étnico o deben haber sido seguido guías establecidas para procedimientos trans-ancestrales (Zondervan & Cardon 2007; Kim *et al.* 2018).

- Solapamiento muestral: Se precisa prestar especial interés al origen y conformación de la muestra descubrimiento. Si las muestras descubrimiento y diana presentan individuos en común los resultados de los análisis podrán mostrar una inflación trascendente que distorsione completamente los resultados de interés (Wray *et al.* 2014). Además, si la naturaleza de los datos lo permite, comprobar si existe alto grado de parentesco entre fracciones de ambas muestras, se establece como un control de calidad extra.

- Alelo de efecto: Si la documentación o información asociada a las estadísticas resumen obtenidas como muestra diana no indica qué alelo es el de riesgo puede asumirse incorrectamente su efecto. Cuando no sea posible asegurar la naturaleza del alelo, puede llegar a discernirse a partir de las frecuencias en la muestra descubrimiento/diana y la población de referencia.

- Variantes ambiguas: Cuando muestra descubrimiento y diana no presentan la información necesaria para conocer la cadena en la que las variantes han sido genotipadas y además han sido obtenidas a partir de distintas tecnologías de genotipado, es preferible eliminar aquellas variantes con alelos ambiguos (variantes con las combinaciones A/T o C/G) ante la imposibilidad de conocer si ambas muestras se refieren al mismo alelo o no. La situación puede ser solventada empleando las frecuencias alélicas al igual que en el caso del alelo de efecto,

pero el riesgo potencial y la gran disponibilidad de variantes en la mayoría de los conjuntos de datos empleados en la actualidad permiten su exclusión.

- Variantes duplicadas y nomenclatura empleada: Asegurarse de la presencia de variantes duplicadas es otro elemento de gran importancia. PLINK comprueba de manera automatizada la presencia de incongruencias alélicas, pero otras herramientas bioinformáticas no incluyen el análisis de forma automatizada. Además, es necesario comprobar la nomenclatura empleada en la denominación de variantes. Si la muestra descubrimiento y diana no siguen los mismos estándares de nomenclatura, una parte de la información genética será desestimada en el cálculo de PRS. La aproximación y solución más sencilla se basa en la generación de una nomenclatura basada en la combinación del valor del cromosoma y del par de bases en el que se encuentra cada variante (p. ej. Una variante dominada “rsX” incorporada desde el cromosoma 6 en la posición 334510 puede ser renombrada como “6:334510”).

- Cromosomas sexuales: Al igual que en la mayoría de los GWAS duplicados, los cromosomas sexuales deben ser excluidos para evitar posibles efectos de confusión (Khramtsova *et al.* 2018) siempre que su análisis sea prescindible.

1.2.6.4 El cálculo de puntuaciones de riesgo poligénico

Para proceder al cálculo de PRS es necesario seguir una serie de directrices y pasos. Tras identificar una muestra descubrimiento con datos de estadística resumen de GWAS, junto a la obtención de una muestra diana con datos genotípicos comunes a la muestra descubrimiento y realizar un control de calidad apropiado, el siguiente paso se basa en la generación de modelos de riesgo y su aplicación a nivel individual. Debido a la variedad de herramientas bioinformáticas de cálculo, esta descripción reúne únicamente las características principales de la metodología estándar.

1.2.6.4.1 Selección de variantes

En estudios de asociación, el fin de los análisis es identificar variantes asociadas de manera específica por lo que el umbral en el que un SNP es declarado significativo es un elemento clave. Con este nivel de restricción, las variantes son identificadas con seguridad, disminuyendo en gran medida los falsos positivos (Wray *et al.* 2014). Como se ha comentado anteriormente, los análisis poligénicos permiten la inclusión de un gran número de variantes asociadas en distinto grado al rasgo o patología de análisis sin considerar un umbral estricto a superar. Con el fin de generar y seleccionar el modelo compuesto por un conjunto específico de variantes que mejor clasifique los individuos en la muestra diana, la aproximación metodológica debe considerar que una proporción de las variantes incluidas en el análisis puede incluir ruido, con el consiguiente aumento de los errores estándar. Por lo tanto, el uso de umbrales basados en el valor de p de asociación, permite incluir de forma progresiva las variantes en modelos exigentes en mayor o menor medida. Al emplear estos umbrales, se crean múltiples PRS que representan distintos grados de errores de tipo I/II (Bogdan *et al.* 2018).

Cuando se emplean umbrales para la inclusión/exclusión de variantes basados en el valor de p , el tamaño del efecto de los SNPs excluidos se reduce a cero mientras los incluidos mantienen el tamaño del efecto. El establecimiento de umbrales permite analizar distintos modelos de manera secuencial y ordenada, ya que las variantes son incluidas en una selección que avanza ordenadamente a través del grado de asociación. La comparación de distintos modelos de forma ordenada tiene como fin seleccionar aquel con rendimiento óptimo. Con este fin, el programa informático PRSice (Euesden *et al.* 2015) permite comparar múltiples umbrales generados en un sistema secuencial continuo, con el fin de detectar el umbral óptimo.

1.2.6.4.2 Desequilibrio de ligamiento

Uno de los sesgos más importantes que pueden afectar al resultado de PRS es el LD (Wray *et al.* 2014; Bogdan *et al.* 2018; Martin

et al. 2018). Como se ha desarrollado anteriormente, los SNPs que se encuentran en una misma región genómica tienden a ser heredados de manera conjunta, por lo que aquellas variantes correlacionadas aportarían al modelo efectos no independientes. Esta situación puede reducir en gran medida la precisión de los PRS (Wray *et al.* 2014). Regiones con un alto LD que presentan un gran número de variantes no independientes incluidas en el proceso de genotipado e imputación estarían sobrerrepresentadas en el modelo poligénico en comparación con aquellas regiones con menor LD y un número menor de variantes en la región cromosómica.

Esta falta de uniformidad extendida en el genoma requiere un control específico. El examen clásico para PRS se basa en un podado de los datos estableciendo un valor de r^2 umbral. Existen distintos algoritmos para el manejo de variantes en LD, pero es el podado informativo informativo (*clumping*) uno de los más empleados (Wray *et al.* 2014; Bogdan *et al.* 2018). Los valores recomendados a seguir pueden diferir en función de la población de estudio, la cobertura genotípica tras la imputación, el control de calidad de variantes u otras variables a considerar, como los filtros de frecuencia de alelo menor aplicados a la muestra. Algunas revisiones recomiendan seguir umbrales de r^2 de 0,2 o 0,1 (Wray *et al.* 2014; Bogdan *et al.* 2018) en ventanas de 500 kilobases. En cada uno de los bloques, el algoritmo selecciona en la muestra descubrimiento aquella variante independiente más asociada al rasgo o patología de estudio y de mayor efecto. En la mayoría de los estudios que emplean esta metodología de PRS, es necesario considerar la eliminación completa o el mantenimiento de una única variante de la región extendida del complejo principal de histocompatibilidad (xMHC) debido a los importantes patrones de LD de la región del cromosoma 6. La aplicación del algoritmo de podado informativo informativo puede realizarse con PLINK seleccionando los valores de los parámetros principales.

A pesar de constituir una herramienta clave en el estudio de datos derivados de GWAS, existen ciertas críticas sobre la arbitrariedad en la selección de un valor umbral (Wray *et al.* 2014). Uno de los puntos débiles del algoritmo se basa en la naturaleza de la estrategia, que al

seleccionar arbitrariamente variantes genómicas, puede desequilibrar la ratio de selección entre variantes causales independientes y variantes asociadas con una única variante causal. El algoritmo de podado informativo informativo puede seleccionar varias variantes asociadas por estar correlacionadas con la misma variante causal en detrimento de otras variantes causales independientes con un valor de asociación algo menor (Wray *et al.* 2014). Durante años se han desarrollado otras aproximaciones y algoritmos de selección de variantes, siendo LDpred uno de los métodos más extendidos que prescinde del algoritmo o técnica de podado informativo (Vilhjálmsson *et al.* 2015). Este método infiere el tamaño medio del efecto posterior de cada marcador empleando un panel de referencia externo e información previa de los tamaños del efecto. Las simulaciones llevadas a cabo por los autores logran un incremento de la varianza explicada en la escala fenotípica, pasando del 20,1% al 25,3% con datos de esquizofrenia y del 9,8% al 12% con datos de esclerosis múltiple.

1.2.6.4.3 Cálculo y evaluación de PRS

Una PRS se calcula para cada individuo de la muestra diana como la suma de alelos de riesgo ponderados por su efecto. En estudios de fenotipo binario, caso-control, el tamaño del efecto suele presentarse como el logaritmo natural de la OR calculado u observado en el GWAS descubrimiento. Es importante conocer la escala logarítmica empleada para interpretaciones futuras.

Para cada umbral o modelo empleado, las PRS (S) se computan como:

$$PRS = \sum_{j=1}^m g_j \beta_j$$

Donde “j” es un SNP en “m” SNPs independientes asociados con el fenotipo de interés. “g” es el número de alelos de riesgo para un SNP en particular y β representa la estimación del tamaño del efecto (logOR).

Tras la obtención de las PRS, la validez del modelo debe ser analizada. Las herramientas estadísticas más empleadas se basan en interrogar la capacidad predictiva y de ajuste del modelo en la población diana. La aproximación más común es el análisis de regresión. La asociación entre las PRS y el rasgo de análisis puede ser medida con asociación estándar o métricas de bondad de ajuste, muy extendidas en epidemiología (Wray *et al.* 2014). Estas herramientas estadísticas son la estimación del tamaño del efecto (OR o coeficiente β), varianza fenotípica explicada (R^2), el área bajo la curva característica operativa del receptor (AUC-ROC, *Area Under Curve - Receiver Operating Characteristic*) y el valor de p de asociación de los análisis de regresión. La varianza fenotípica explicada se expresa mediante distintos estimadores de la correlación cuadrada entre el rasgo y las PRS (Heinzl *et al.* 2005). Una de las más utilizadas y extendidas dentro de las pseudo- R^2 en los análisis de regresión logística, cuando se analiza un fenotipo binario, es la pseudo- R^2 de Nagelkerke (Nagelkerke 1991). A pesar de su amplio uso, la pseudo- R^2 de Nagelkerke presenta ciertas limitaciones, como la dependencia que muestra sobre la proporción de casos y controles en la muestra diana. Algunos autores han buscado desarrollar medidas de correlación que consideren este sesgo, pero su uso no está tan extendido (Lee *et al.* 2012; Wray *et al.* 2014).

En el análisis de las PRS es necesario incorporar covariables de ajuste matemático. La edad, sexo, estructura poblacional y otras variables deben ser consideradas en los análisis con el fin de minimizar el efecto de elementos confusores. Estos elementos o artefactos, de no ser ajustados en los modelos matemáticos, podrían llegar a generar resultados espurios de los análisis genéticos (Bogdan *et al.* 2018). Además de estas covariables, otras medidas pueden ser incorporadas a los modelos matemáticos para sus análisis conjuntos con los datos de PRS: covariables informativas de elementos de tipo ambiental o clínico o información genética complementaria.

1.2.6.4.4 Metodologías PRS

La mayoría de los métodos que se presentan a continuación requieren datos en dos clases principales de formatos: genotípicos a nivel individual o estadísticas resumen de GWAS. Algunos métodos emplean solamente datos de una clase y otros las combinan. Aunque permiten un mayor número de opciones analíticas y presentan ciertas ventajas de precisión en cálculo, los datos completos a nivel genotípico individual presentan ciertas complicaciones intrínsecas a su naturaleza:

- El volumen del conjunto de datos puede ser de gran tamaño, con sus consecuentes cargas computacionales asociadas. Los métodos que requieren estadísticas resumen de GWAS presentan tiempos de cálculo reducidos y una menor memoria computacional al presentar procesos de cálculo más eficientes (Ge *et al.* 2015).
- Están sujetos a protecciones estrictas de tipo ético y regulatorio, limitando su acceso con el fin de mantener protegida la privacidad de los individuos. Los métodos que requieren estadísticas resumen presentan limitaciones éticas menores (Pasaniuc & Price 2016).
- Los métodos que requieren estadísticas resumen suelen verse favorecidos de la mayor disponibilidad de estos datos derivados, normalmente, de consorcios internacionales que siguen altos estándares de privacidad y seguridad (Pasaniuc & Price, 2016).

Esta tesis doctoral no busca analizar de manera exhaustiva la totalidad de herramientas bioinformáticas existentes y útiles para el cálculo de PRS y análisis relacionados, sino reflejar algunas de las diferencias y opciones existentes en la literatura, con el fin de mostrar una imagen amplia y diversa de los métodos empleados por la comunidad científica. Algunos de los métodos estadísticos/computacionales más empleados se muestran a continuación:

- PLINK (Purcell *et al.* 2007; Chang *et al.* 2015): Una de las herramientas de uso más extendido por su sencillo y práctico uso. Este método emplea como muestra descubrimiento estadísticas resumen de GWAS.
- GCTA/BLUP/GBLUP (Yang *et al.* 2011; Chen *et al.* 2015): Estos métodos, basados en el modelo lineal generalizado, ajustan modelos lineales mixtos con componentes principales como covariables con el fin de obtener la mejor predicción. Estas metodologías emplean como muestra descubrimiento datos genotípicos a nivel individual.
- LDpred (Vilhjálmsón *et al.* 2015): Comentado previamente, a diferencia de otros métodos, esta herramienta permite inferir el tamaño medio del efecto de cada marcador empleando un panel de referencia externo para inferir patrones de LD y efectos de cada marcador, sin ser necesario el podado informativo de los datos, al indicar un radio de LD. Este método emplea como muestra descubrimiento estadísticas resumen de GWAS.
- BayesR (Erbe *et al.* 2012): Esta aproximación al cálculo de puntuaciones de riesgo individuales presenta una mayor complejidad en su modelo, de tipo jerárquico, se ajusta a todos los SNPs de manera simultánea. El efecto atribuible a SNPs en LD entre ellos es compartido entre todos los SNPs correlacionados. Este método emplea como muestra descubrimiento datos genotípicos a nivel individual.

1.2.6.5 PRS, una aproximación genómica en crecimiento

Podrían emplearse un gran número de ejemplos del avance en el estudio de la arquitectura poligénica de las patologías o rasgos complejos, incluso limitando los ejemplos a aquellos en el campo de psiquiatría/psicología. Dentro de los ejemplos de estimación de riesgo poligénico en el campo de psiquiatría, dos de los hitos más importantes son el estudio de Purcell y colaboradores, que describió por primera vez

con éxito un modelo poligénico, y el realizado por el grupo Cross-Disorder del PGC, analizando la arquitectura genética compartida entre cinco patologías psiquiátricas comunes. Estos dos estudios asentaron las bases y direcciones futuras de una gran parte de la investigación global en genética psiquiátrica, sirviendo como punto de partida para el desarrollo de un gran número de estudios en el campo.

1.2.6.5.1 “La variación poligénica común contribuye al riesgo de esquizofrenia y trastorno bipolar” (Purcell et al. 2009)

El ISC (*International Schizophrenia Consortium*), con Shaun M. Purcell liderando sus análisis, publicó en el año 2009 en *Nature* los resultados de un GWAS que incluía un total de 6909 individuos. Además de los resultados de asociación de los 3322 casos con esquizofrenia de origen europeo y 3587 controles, los autores focalizaron los análisis en un exhaustivo estudio de las variantes genéticas localizadas en el MHC y en el análisis del componente poligénico. Como se ha desarrollado anteriormente, este análisis marcó un hito en el estudio de la variación común ya que pusieron a prueba y confirmaron la teoría clásica de la herencia poligénica. La arquitectura poligénica no había sido descrita hasta la fecha en ninguna patología o rasgo humano a partir de datos de GWAS, por lo que este estudio aportó y asentó las herramientas para el estudio de este componente genético de variación común en el campo de los rasgos complejos humanos.

Estos análisis concluyen que la base poligénica descrita en esquizofrenia está conformada por SNPs comunes, que explican al menos un tercio de la variación en la escala fenotípica, se comparte con el trastorno bipolar y no presenta solapamiento con patologías no psiquiátricas.

Tras dividir la muestra del ISC por sexo, calcularon las PRS en mujeres a partir de los efectos inferidos en la población masculina. Tras detectar un componente poligénico significativo que explicaba un 3% de la varianza fenotípica, los autores analizaron esta arquitectura genética en tres muestras independientes de esquizofrenia (dos

muestras diana del Molecular Genetics of Schizophrenia [MGS]) de origen europeo-americano y africano-americano y una muestra diana derivada de un conjunto de muestras de origen británico (O'Donovan *et al.* 2008). El porcentaje de varianza explicada en la distintas muestras diana fue mayor en el caso de la muestra europea-americana de MGS (máxima $R^2 = 3,2\%$), seguida por la de origen británico (máxima $R^2 = 2,3\%$) y por último, con un una predicción menor, la muestra africana-americana (máxima $R^2 = 0,4\%$).

Los análisis comparativos con otras patologías, se basaron en el empleo de la muestra del ISC como muestra descubrimiento. Además, emplearon en el caso de trastorno bipolar dos muestras independientes (STEP-BD [Sklar *et al.* 2008] y WTCCC [Wellcome Trust Case Control Consortium 2007]) y cinco muestras independientes derivadas del WTCCC (diabetes tipo II, diabetes tipo I, artritis reumatoide, hipertensión y arterioesclerosis coronaria). Ninguna de las cinco patologías no psiquiátricas mostró solapamiento significativo con esquizofrenia, mientras que en las muestras independientes de trastorno bipolar se llegó a explicar entre el 1,4 y el 1,9% de la varianza fenotípica.

El modelo poligénico desarrollado y aplicado por Purcell y colaboradores ha servido como referencia para análisis posteriores, marcando las pautas y estándares a seguir (Wray *et al.* 2014).

1.2.6.5.2 Análisis del Cross-Disorder Group del PGC

El PGC, liderado por Jordan Smoller y Stephan Ripke, publicó en el año 2013 en *The Lancet* los resultados de análisis conjuntos de cinco patologías psiquiátricas: esquizofrenia, trastorno bipolar, trastorno depresivo mayor, trastornos del espectro autista y trastorno por déficit de atención e hiperactividad. Además de estos análisis, el consorcio internacional publicó a partir de esta iniciativa resultados paralelos en *Nature Genetics* en el mismo año, con un estudio liderado por S Hong Lee y Naomi R. Wray.

En el primer análisis comentado, los autores analizaron de forma conjunta los cinco trastornos en 33332 casos y 27888 controles de origen europeo. Además del GWAS general, los análisis se centraron en desentrañar las relaciones poligénicas entre los cinco trastornos y en enfoques de análisis de rutas biológicas específicas. Para ello, los autores realizaron PRS empleando cada patología como muestra descubrimiento y las cuatro restantes como diana. Los resultados principales mostraron un solapamiento notable de la arquitectura poligénica entre trastornos de inicio en la etapa adulta (esquizofrenia, trastorno bipolar y trastorno depresivo mayor), en mayor grado en el caso de esquizofrenia y trastorno bipolar. El solapamiento, aunque menor, también fue detectado mediante PRS entre trastornos del espectro autista y esquizofrenia o trastorno bipolar. En cambio, el trastorno por déficit de atención e hiperactividad no presentó ninguna relación detectable con el resto de patologías. Estos resultados, desde el punto de vista del continuo del neurodesarrollo, muestran el menor peso de la variación común en la relación existente entre los trastornos que siguen el modelo y su gradiente. El estudio publicado en Nature Genetics detectó importantes grados de correlación genética entre distintos trastornos. Las patologías más correlacionadas fueron: esquizofrenia-trastorno bipolar ($r_g=0,68$), esquizofrenia y trastorno depresivo mayor ($r_g=0,43$), trastorno bipolar y trastorno depresivo mayor ($r_g=0,47$) o el trastorno por déficit de atención e hiperactividad y el trastorno depresivo mayor ($r_g=0,32$).

Los hallazgos llevados a cabo por esta iniciativa fueron de gran relevancia, ya que ningún estudio previo había mostrado o examinado el solapamiento genético común entre esas cinco patologías psiquiátricas. La relevancia reside en la demostración del poder predictivo de las PRS, que en este trabajo permite ir más allá de las entidades clínicas básicas poniendo la primera piedra de cara a comprender la coocurrencia de fenotipos clínicos. Los efectos aditivos de variantes comunes independientes abrían una ventana al análisis de la estructura de la psicopatología.

La pleiotropía extendida en los trastornos psiquiátricos detectada en este estudio proporcionó las pistas iniciales sobre una biología

compartida por trastornos. Estos resultados se sumaron a efectos pleiotrópicos detectados previamente en variantes raras de gran efecto (Malhotra & Sebat 2012). Tras la publicación de este trabajo, decenas de análisis han explorado la genética común compartida entre patologías psiquiátricas desde distintas perspectivas: buscando evidencia de etiología o arquitectura genética compartida u optimizando la predicción en la muestra diana empleando patologías distintas.

1.2.6.6 Correlación genética, otra aproximación en el análisis de solapamiento entre rasgos complejos

Otros métodos de interés son aquellos que exploran la arquitectura genética compartida basándose en mediciones de correlación genética. Estas aproximaciones, derivadas de métodos de cálculo de heredabilidad basada en SNPs, permiten estimar el grado de correlación entre parejas de trastornos y, a diferencia de los PRS, no proporcionan predicciones genéticas a nivel individual. Dos rasgos o trastornos están correlacionados genéticamente cuando existe una correlación entre los tamaños y dirección de los efectos de SNPs que afectan a ambos, situación que señala hacia una posible etiología compartida. Por lo tanto, aunque sin apoyarse en el cálculo de puntuaciones de riesgo, el uso de estas metodologías es de gran interés desde la perspectiva de análisis de solapamiento entre rasgos poligénicos (Maier *et al.* 2018).

En síntesis, pueden diferenciarse dos grandes metodologías de aproximación a este análisis genómico: aquellas que precisan datos genotípicos a nivel individual y aquellas que permiten realizar las estimaciones basándose en las estadísticas resumen de GWAS.

Una de las metodologías de estimación de correlación genética más extendida, se basa en cálculos de máxima verosimilitud restringida basados en el genoma (GREML) empleando datos genotípicos a nivel individual. Este método, extendido en el cálculo de heredabilidad basada en SNPs, puede emplearse en el cálculo de correlación genética

al pasar a una forma bivariante. Los análisis bivariantes de GREML dependen de la heredabilidad de SNPs de ambos rasgos a comparar, el valor de correlación genética de la población y el solapamiento de individuos entre muestras (Visscher *et al.* 2014). Esta aproximación, empleada en sus orígenes en la mejora genética de especies de interés comercial, presenta diversos métodos desarrollados para el cálculo en humanos. El más empleado, GCTA (Yang *et al.* 2011), es una herramienta bioinformática bastante empleada en el cálculo de heredabilidad basada en SNPs con un uso extendido al cálculo de correlaciones, junto a otras funciones. Otras herramientas bioinformáticas como BOLT-LMM, logran realizar estimaciones mediante análisis bivariantes GREML en un tiempo menor (Loh *et al.* 2015).

Otra metodología de análisis de correlación genética ampliamente extendida, basada en estadísticas resumen de GWAS, es representada por LDSC (Bulik-Sullivan *et al.* 2015a; Bulik-Sullivan *et al.* 2015b). Uno de los mayores inconvenientes de este método es el aumento de errores estándar respecto a GREML. Su uso está ampliamente extendido y nuevas iniciativas han surgido a raíz de su éxito con el fin de mejorar y facilitar su uso de manera extendida. El mejor ejemplo reside en LD Hub (Zheng *et al.* 2017), una base de datos centralizada de estadísticas resumen de GWAS que permite realizar múltiples análisis de correlación basados en LDSC en una interfaz web, de manera automatizada integrando múltiples conjuntos de datos de GWAS simultáneamente.

**JUSTIFICACIÓN,
HIPÓTESIS Y OBJETIVOS**





2. JUSTIFICACIÓN, HIPÓTESIS Y OBJETIVOS

2.1 JUSTIFICACIÓN E HIPÓTESIS

Los trastornos psiquiátricos presentan, en el marco nosológico actual, una importante comorbilidad extendida. Las relaciones concretas entre trastornos en las que se asienta directamente esta tesis son la comorbilidad entre el trastorno obsesivo-compulsivo y la esquizofrenia y la observada entre el trastorno por abuso o dependencia de sustancias y los trastornos psiquiátricos generales (también denominada “patología dual”).

La prevalencia estimada del trastorno obsesivo-compulsivo en pacientes con un diagnóstico en el espectro de la esquizofrenia es del 12,1% (Achim *et al.* 2011). Además, la presencia de un diagnóstico previo de trastorno obsesivo-compulsivo y la presencia de un diagnóstico del trastorno en parientes cercanos están asociados a un mayor riesgo a desarrollar esquizofrenia (Meier *et al.* 2014; Cederlöf *et al.* 2015).

El trastorno por abuso o dependencia de sustancias muestra una importante heredabilidad estimada, del 40% al 70% dependiendo de la sustancia, con una importante fracción de la susceptibilidad genética compartida entre las distintas adicciones (Kendler *et al.* 2003; Goldman *et al.* 2005; Kendler *et al.* 2007). La comorbilidad de un trastorno por abuso o dependencia de sustancias y los trastornos psiquiátricos generales es muy común y está asociada con un peor pronóstico en los pacientes (Regier *et al.* 1990; Pereiro *et al.* 2013).

En base a lo anteriormente desarrollado, y al gran avance de la genética psiquiátrica en los últimos años, se plantean las siguientes hipótesis:

La arquitectura genética de los distintos trastornos psiquiátricos seguirá una estructura poligénica en el marco del modelo de enfermedad común – variante común. Esta arquitectura, formada por un gran conjunto de variantes de pequeño efecto cada una, podrá ser

empleada para la estimación de riesgo individual a desarrollar un trastorno mediante la generación de puntuaciones de riesgo poligénico.

La generación de modelos de predicción de riesgo, y su capacidad predictiva, dependen del grado de conocimiento del tamaño del efecto y el nivel de asociación de las distintas variantes involucradas en un fenotipo concreto (datos obtenidos mediante grandes GWAS). Estos, aunque reflejan una imagen incompleta de la genética de los distintos trastornos, podrán ser suficientemente útiles en el esclarecimiento de diversas cuestiones etiológicas.

La capacidad predictiva permitirá indagar en las relaciones entre distintos trastornos, posibilitando el análisis del grado de solapamiento genético entre ellos. La alta comorbilidad entre algunos trastornos, así como el solapamiento en distintos aspectos clínicos, podrán responder parcialmente a esta relación genética subyacente.

La predicción de riesgo poligénico podrá ser combinada con variables de riesgo conocidas, con el fin de analizar su comportamiento y utilidad clínica, más allá de las variables conocidas, en modelos conjuntos.

La genética potencialmente compartida entre trastornos podrá incrementar la capacidad predictiva del modelo de riesgo poligénico cuando es modelizada conjuntamente.

Será posible discernir parte de la etiología de los distintos trastornos psiquiátricos mediante la estimación de puntuaciones de riesgo poligénico basadas en modelos fundamentados en mecanismos biológicos funcionales subyacentes.

La mejor comprensión de la poligenicidad y la estimación de riesgo de los trastornos psiquiátricos permitirán avanzar hacia una nosología con mayor validez biológica. Esta, repercutirá positivamente en el diagnóstico, tratamiento y pronóstico de los pacientes

2.2 OBJETIVOS

El objetivo general de los cuatro estudios que conforman la presente tesis doctoral consistió en analizar la utilidad de las puntuaciones de riesgo poligénico en el marco de los trastornos psiquiátricos. Con este fin, se marcaron los siguientes tres objetivos concretos:

- Analizar el solapamiento de la arquitectura genética entre distintos trastornos con diferentes grados de comorbilidad:
 - o Estudiando la susceptibilidad poligénica compartida entre el trastorno obsesivo-compulsivo y la esquizofrenia.
 - o Examinando la susceptibilidad poligénica compartida entre el trastorno por abuso o dependencia de sustancias y los principales trastornos psiquiátricos y las diferencias de carga de susceptibilidad entre sexos.

- Analizar, en una muestra diana de alcoholismo, la capacidad predictiva de la modelización conjunta de factores predictivos de riesgo conocidos junto a puntuaciones de riesgo poligénico, empleando modelos genéticos combinados. Estos últimos, estarán compuestos por puntuaciones poligénicas obtenidas a través de modelos basados en distintos trastornos.

- Analizar la capacidad predictiva de modelos poligénicos fundamentados en la selección de variantes en base a criterios biológicos funcionales, con el fin de explicar la relación entre el estrés y la dependencia al alcohol.



MATERIALES Y MÉTODOS





3. MATERIALES Y MÉTODOS

3.1 MUESTRAS

En la realización de esta tesis doctoral se emplearon cuatro conjuntos muestrales.

3.1.1 Trastorno obsesivo-compulsivo

En el estudio de trastorno obsesivo-compulsivo se incluyeron un total de 917 muestras. El conjunto de pacientes está compuesto por 433 individuos (206 mujeres y 227 hombres) diagnosticados con trastorno obsesivo-compulsivo. El inicio de la patología fue antes de los 15 años de edad en el 38% de los pacientes y antes de los 12 años en el 21% de los individuos. Los diagnósticos fueron llevados a cabo por dos psiquiatras experimentados siguiendo en adultos el (SCID-I, *Structured Clinical Interview for DSM-IV Axis I Disorders - Clinician Version*) (First *et al.* 1997), y, en el grupo pediátrico, el (K.SADS-PL, *Kids Schedule for affective Disorders and Schizophrenia for School-age Children - Present and Lifetime Version*) (Kaufman *et al.* 1996). Los casos fueron reclutados entre los años 2003 y 2012 en la unidad de trastorno obsesivo-compulsivo del Hospital de Bellvitge (L'Hospitalet de Llobregat) y el Departamento de Psiquiatría y Psicología del Niño y el Adolescente del Hospital Clínic (Barcelona). Todos los pacientes son de ascendencia europea y nacidos en España. Los criterios de exclusión para este grupo de individuos son presencia de trastornos psicóticos, dependencia a drogas psicoactivas, discapacidad intelectual y patologías neurológicas u orgánicas severas, exceptuando trastornos por tics.

La muestra control para el estudio de trastorno obsesivo-compulsivo estaba compuesta por 484 individuos (202 mujeres y 282 hombres) del Banco Nacional de ADN Carlos III (Universidad de Salamanca) y sujetos que acuden a centros de atención primaria de Galicia. Todos ellos se encontraban en buen estado de salud, declarando

no sufrir ninguna enfermedad, tras verse sometidos a un breve examen médico y un cuestionario. Se obtuvo consentimiento informado de todos los participantes o de sus tutores legales y la investigación fue realizada acorde a los principios de la Declaración de Helsinki tras la aprobación de los distintos Comités Éticos (Comité Ético del Hospital Universitario de Bellvitge [referencia 198/02], Comité Ético del Hospital Clínic [referencia HCB/2009/4884] y el Comité de Ética en Investigación Clínica De Galicia [referencia 2013/106]).

3.1.2 Esquizofrenia

Esta muestra fue generada en el estudio de Carrera y colaboradores (Carrera *et al.* 2012). Una descripción adicional de los sujetos, la genotipación y el control de calidad aplicado al conjunto muestral se presenta más detalladamente en el manuscrito. Los datos presentados a continuación son el resultado del control de calidad aplicado en el trabajo.

El número total de individuos que conforman esta muestra es 923, divididos en 476 casos y 447 controles e incluyen 5100 SNPs no sinónimos (nsSNPs, *Non-Synonymous Single Nucleotide Polymorphisms*). Los pacientes pertenecen al área sanitaria de Santiago de Compostela, presentan un diagnóstico en esquizofrenia basado en los criterios del DSM-IV a través de entrevistas no estructuradas, su diagnóstico es estable (mínimo 2 años). Los pacientes con trastorno esquizoafectivo fueron excluidos.

Las muestras del grupo control fueron obtenidas de individuos que participaron en donaciones de sangre en el Centro de Transfusión de Galicia (Santiago de Compostela), de origen gallego.

Todos los participantes aportaron por escrito su consentimiento informado para participar en este estudio. La investigación fue realizada acorde a la última versión de la Declaración de Helsinki y fue aprobado por el Comité de Ética en Investigación Clínica De Galicia (referencia 2008/116).

3.1.3 Trastorno por abuso o dependencia de sustancias

En la primera fase del estudio de adicciones, en el que se engloba este conjunto muestral, se incluyeron un total de 1176 individuos: 572 casos y 604 controles (383 mujeres y 793 hombres). La recogida de muestras fue llevada a cabo entre los años 2013 y 2014. Los pacientes son principalmente individuos que acuden a Unidades de Asistencia de Drogodependencias de las principales áreas de salud de Galicia. Los diagnósticos fueron realizados por parte del equipo clínico de estas unidades, todos ellos psiquiatras con amplia trayectoria profesional y experiencia, basándose en todos los datos disponibles utilizando el (LEAD, *Longitudinal Expert All Data*) (Spitzer *et al.* 1983). Los criterios de inclusión se resumen en una edad comprendida entre 18 y 65 años, origen en Galicia (cuatro abuelos nacidos en la comunidad autónoma), criterios de dependencia de una sustancia y abuso/dependencia de otra de las siguientes drogas: alcohol, tabaco, cannabis, cocaína, opiáceos, hipnóticos, estimulantes, alucinógenos y solventes volátiles. Durante el proceso de asistencia a los pacientes se evaluó la posible comorbilidad de trastornos psiquiátricos generales, a través de un examen clínico del estado mental y la revisión exhaustiva de la historia clínica electrónica del paciente. Los criterios de exclusión para este grupo de pacientes se resumen en: presencia de discapacidad intelectual y patologías neurológicas graves.

Las muestras del grupo control, 604 participantes, fueron obtenidas de individuos que participaron en donaciones de sangre en el Centro de Transfusión de Galicia (Santiago de Compostela) y pertenecen a la colección nº 0003122 del Registro Nacional de Biobancos del Instituto de Salud Carlos III. El alcoholismo crónico y el uso de drogas ilícitas, información aportada por los donantes, son causas de exclusión para la donación de sangre. Según datos oficiales del Ministerio de Sanidad, Servicios Sociales e Igualdad del año 2015, la prevalencia del consumo de alcohol alguna vez en la vida en la población (15-64 años de edad) es del 93,5% (EDADES 2015). Por lo tanto, esta muestra está compuesta por individuos que mayoritariamente han estado expuestos a alcohol y no han desarrollado trastorno por abuso o dependencia de sustancias (SUD, *Substance Use Disorder*). Todos los participantes

aportaron por escrito su consentimiento informado para participar en este estudio. La investigación fue realizada acorde a la última versión de la Declaración de Helsinki y fue aprobado por el Comité de Ética en Investigación Clínica De Galicia (referencia 2012/368).

3.1.4 Trastorno por dependencia de alcohol

La segunda fase del estudio incluyó un total de 470 muestras, 240 casos y 230 controles, apareados por sexo y edad. La recogida de muestras fue llevada a cabo en 2015. Las muestras de pacientes fueron recogidas en distintas unidades de asistencia: el 71,67% de la muestra proviene de la Unidad de Conductas Adictivas de Ourense (unidad multidisciplinar dedicada en exclusiva al alcoholismo), el 20,83% del Servicio de Psiquiatría de Santiago de Compostela y el 7,5% restante proviene de varios centros de asistencia del área de Oviedo, Asturias. Los casos son individuos dependientes de alcohol (criterios DSM-IV), diagnosticados por los distintos equipos clínicos. Los criterios de inclusión fueron: mayoría de edad, dependencia de alcohol (criterios CIE-10) y origen gallego (cuatro abuelos gallegos) o asturiano (cuatro abuelos asturianos). Los criterios de exclusión se resumen a continuación: discapacidad intelectual, enfermedad neurológica, deterioro cognitivo y parentesco con muestras de la Fase I. El 67,5% de los casos presentan dependencia de otra sustancia y el 23,75% comorbilidad de otros trastornos psiquiátricos del Eje I. Todos los participantes han aportado por escrito su consentimiento informado para participar en este estudio.

Los individuos que componen el grupo control de la Fase II son sujetos que están expuestos al alcohol pero ninguno ha llegado a mostrar criterios de dependencia de la sustancia. Además del emparejamiento comentado en el párrafo anterior, cabe destacar que estos individuos fueron seleccionados entre adultos que vivían en la misma comunidad que los casos en el momento de la evaluación. El origen de las muestras de Galicia es variado y presenta la siguiente distribución: 70% amigos, vecinos, conocidos y compañeros de trabajo, 20% centros de salud privados y el 10% restante de centros de salud

públicos. Los controles que provienen de Asturias fueron reclutados tras haber pasado consultas y evaluaciones previas por el mismo médico de familia tras haber buscado atención médica con el fin de tratar un evento agudo no grave (resfriado, otitis, etc.). Todas las muestras aportaron su consentimiento informado para este estudio. La investigación fue realizada acorde a la última versión de la declaración de Helsinki y ha sido aprobado por el Comité de Ética en Investigación Clínica De Galicia (referencia 2014/428) y por el Comité de Ética en Investigación Clínica De Asturias (referencia 77/14).

Todos los individuos de esta muestra, añadieron a su muestra biológica información ambiental recogida por los distintos equipos clínicos. Además del sexo y la edad, todos los individuos aportaron información de otras variables: meses/edad de dependencia al alcohol/tabaco, meses/edad de inicio de consumo de alcohol, unidades de alcohol al día durante el último mes, número de cigarrillos al día durante el último mes, estado civil, situación de convivencia en el hogar, situación laboral, años de estudio completados, intentos de suicidio, comorbilidad psiquiátrica CIE-10, comorbilidad adictiva e historia familiar de dependencia de alcohol. También participaron incluyendo distintos cuestionarios psicológicos:

- Cuestionario de patrón de consumo de alcohol: Este cuestionario es una adaptación al castellano del facilitado por el *National Institute on Alcohol Abuse and Alcoholism* (NIAAA, Bethesda, Maryland, EE.UU.), el cual permite conocer, del individuo entrevistado, los hábitos de consumo durante el último año: frecuencia, cantidad y patrón de consumo.

- Escala de acontecimientos vitales estresantes de Brugha (LTE, *The List of Threatening Experiences*): Cuestionario autoadministrado con 12 categorías y cuestiones de las posibles experiencias estresantes o amenazantes que ha sufrido un individuo. Por su estructura y el número de categorías analizadas, permite realizar el análisis con rapidez sin reducir en gran medida el grado de asociación respecto a otros cuestionarios. Los pacientes y controles fueron examinados con

el fin de conocer posible elementos estresantes vitales durante el último año (Brugha *et al.* 1985).

- Escala de acontecimientos traumáticos en la infancia (CTQ, *Childhood Trauma Questionnaire*): Cuestionario autoadministrado con 28 ítems o cuestiones tipo Likert que contempla cinco categorías de exposición a trauma infantil: abuso emocional, abuso físico, abuso sexual, negligencia emocional y negligencia física. Puede ser usado tanto para medir presencia o ausencia de exposición como para medir la severidad de esta exposición (Bernstein & Fink 1988).

- Inventario NEO reducido de Cinco Factores (NEO-FFI): Cuestionario autoadministrado de 60 ítems o cuestiones tipo Likert que miden los cinco grandes dominios o factores de la personalidad: neuroticismo, extraversión, apertura, amabilidad y responsabilidad. Este cuestionario es la versión reducida del Inventario de Personalidad NEO Revisado (NEO PI-R) e incluye 12 ítems de cada factor (Costa & McCrae 1999).

3.2 GENOTIPADO

La tecnología de genotipado de las muestras, así como el manejo de las mismas o el tipo de muestra biológicas presenta diferencias entre los distintos estudios incluidos, por lo que son presentados por separado.

Las muestras donadas por pacientes y controles en el estudio de trastorno obsesivo-compulsivo fueron recogidas en tubos con anticoagulante ácido etilendiaminotetraacético (EDTA) con sangre en su interior. La tecnología de genotipado utilizada fue el chip de Affymetrix (Santa Clara, California, EE.UU.) Axiom Exome Array y el proceso de genotipado fue llevado a cabo en el Centro Nacional de Genotipado (CeGen-ISCIH, nodo de Santiago de Compostela) en la plataforma Gene Titan de Affymetrix. Este panel incluye cerca de 300000 variantes localizadas en regiones codificantes. Las sondas son marcadores de GWAS del *National Human Genome Research Institute Catalog of Published Associations* (agosto 2011) y marcadores

informativos de ancestralidad. Para prevenir y minimizar efectos de lote, todos los chips presentaron aproximadamente el mismo número de casos que de controles. La asignación de variantes (*variant calling*) se realizó con el software Affymetrix Genotyping Console Software utilizando el algoritmo Axiom GT1. Este chip fue utilizado en casos y controles siguiendo las instrucciones del fabricante.

En el caso de los dos conjuntos muestrales de trastornos por uso o abuso de sustancias las muestras biológicas fueron recibidas en el laboratorio de Santiago de Compostela en dos formatos distintos: muestras de sangre en tubos EDTA en la primera y la segunda fase del estudio y saliva únicamente en la segunda fase (kit Oragene•DNA OG-500 [DNA Genotek, Ottawa, Ontario, Canadá]). La extracción del ADN se realizó, en la primera fase, con la plataforma chemagen y el kit chemagenic DNA Blood Kit special (PerkinElmer chemagen Technologie GmbH, Baesweiler, Alemania). En la segunda fase, el ADN genómico fue extraído con el kit de extracción ReliaPrep Blood gDNA Miniprep System (Promega, Madison, Wisconsin, EE.UU.) en el caso de las muestras de sangre. Las muestras de saliva fueron procesadas con el kit prepIT•L2P (DNA Genotek). La cuantificación del ADN se realizó mediante dos métodos distintos: todas las muestras fueron medidas en el laboratorio con el fluorímetro Qubit dsDNA BR Assay Kit de Life Technologies (Eugene, Oregon, EE.UU.) y, parte de las muestras de la segunda fase, con el espectrofotómetro Nanodrop 2000 de Thermo Fisher Scientific (Waltham, Massachusetts, EE.UU.) con el fin de conocer la pureza aproximada de las muestras.

El genotipado utilizado en el estudio de adicciones se basó en la plataforma iScan System de Illumina (Illumina, San Diego, CA, EE.UU.). En ambas fases el chip utilizado fue el BeadChip Illumina Infinium PsychArray (v1.0 en la primera fase y v1.1 en la segunda fase del estudio) y el genotipado fue llevado a cabo en el Centro Nacional de Genotipado (CeGen-ISCI, nodo de Madrid). Este chip ha sido desarrollado por Illumina en colaboración con el Psychiatric Genomics Consortium e incluye más de medio millón de marcadores ampliamente distribuidos por el genoma, de los cuales 50000 están asociados con trastornos psiquiátricos. Para prevenir y minimizar efectos de lote,

todos los chips presentan aproximadamente el mismo número de casos que de controles, con un diseño previo de placa que consideraba el manejo de las muestras en las plataformas robotizadas. Este chip ha sido utilizado en casos y controles siguiendo las instrucciones del fabricante.

3.3 CONTROL DE CALIDAD DE LOS DATOS GENOTIPADOS

El filtrado de genotipos es un procedimiento de gran importancia que, en una gran parte de la bibliografía, no se desarrolla en detalle incluyendo cada paso específico, existiendo cierta pero pequeña variabilidad en los métodos entre distintos estudios. En este trabajo, la mayoría de los parámetros aplicados son compartidos entre los distintos conjuntos de muestras, existiendo ciertas diferencias puntuales. Para reflejar en detalle este proceso, se desarrollan por separado ambos protocolos (trastorno obsesivo-compulsivo y SUD).

En el estudio de trastorno obsesivo-compulsivo el proceso de filtrado de genotipos se realizó de forma exhaustiva con procedimientos estándar de PLINK v1.07 (Purcell *et al.* 2007). PLINK es una herramienta de código abierto C/C++ que permite trabajar con grandes conjuntos de datos. Los SNPs que no superasen alguno de los siguientes filtros fueron eliminados: (i) ratio de genotipación mayor al 95% (*SNP call rate*); (ii) desviaciones significativas del equilibrio Hardy-Weinberg en controles ($P > 1 \times 10^{-3}$); (iii) ratio de genotipación casos vs controles ($P > 1 \times 10^{-3}$; *call rate differences cases vs controls*) y (iv) frecuencia del alelo menor $> 5\%$ (MAF, *minor allele frequency*). Las muestras que presentan alguna de estas condiciones fueron eliminadas con PLINK v1.07: (i) ratio de genotipación menor al 95% (*call rate*); (ii) género discordante entre el registrado en nuestra base de datos y el inferido de los genotipos; (iii) niveles de heterocigosidad que se apartan tres desviaciones estándar de la media; (iv) relaciones de parentesco crípticas, detectadas por valores de la proporción del genoma idéntica por descendencia mayores a 0,05 (IBD, *identity-by-descent*, $\hat{\pi}$ PI-HAT), calculado con variantes filtradas por $MAF > 5\%$. El uso de IBD permite reconocer parejas de individuos que se parecen más de lo que cabría esperar en una muestra al azar. En los casos en los que esta

situación se cumplió, se eliminó la muestra de la pareja que presentaba una menor ratio de genotipación. Por último, se analizaron datos informativos de ancestralidad. Utilizando 3410 marcadores informativos de ancestralidad se buscó identificar aquellos individuos con menos de un 90% genotipos de origen no europeo utilizando Structure v2.3.3 (Pritchard *et al.* 2000) y las muestras de HapMap de origen europeo (CEU), africano (YRI) y asiático (JPT+CHB) como muestras de referencia.

En el estudio de adicciones el proceso de filtrado de genotipos se realizó con PLINK v1.9 (Chang *et al.* 2015). Las variantes que no superasen alguno de los siguientes filtros fueron eliminadas: (i) ratio de genotipación mayor al 95%; (ii) ratio de genotipación casos *vs* controles ($P > 1 \times 10^{-3}$); (iii) sin desviaciones significativas del equilibrio Hardy-Weinberg en controles ($P > 1 \times 10^{-3}$) y (iv) frecuencia del alelo menor $> 1\%$.

Las muestras que presentan alguna de estas condiciones fueron eliminadas: (i) ratio de genotipación menor al 95%; (ii) relaciones de parentesco crípticas entre individuos detectadas mediante IBD, el umbral seleccionado es PI-HAT mayor a 0,15, calculado con variantes filtradas por $MAF > 10\%$. En los casos en los que esta situación se cumplió, se eliminó aquella muestra con menor ratio de genotipación correcta en las variantes que presentan un $MAF > 1\%$ y (iii) se analizó la posible existencia de muestras duplicadas. Para evaluar la estructura de la población, corregir la posible estratificación y detectar muestras atípicas (*outliers*) se calcularon las primeras cinco dimensiones del método de escalamiento multidimensional (MDS, *multidimensional scaling*).

Al incorporarse ambas fases del estudio para análisis combinados se analizó el posible solapamiento de muestras entre ambas fases del estudio y la presencia de relaciones de parentesco no detectadas al seleccionar las muestras ($PI-HAT = 0,125$).

3.4 IMPUTACION DE VARIANTES

La imputación de variantes fue realizada para los datos derivados de las dos fases del estudio de SUD. Al existir diferencias en la metodología entre la primera y la segunda fase, ambos procedimientos se exponen por separado.

En la primera fase, la inferencia de haplotipos (*prephasing*), fue llevada a cabo con SHAPEIT, en su versión v2.r837 (Delaneau *et al.* 2011) con el fin de acelerar el proceso de imputación y reducir el coste de computacional. La imputación fue realizada con IMPUTE2, en su versión v2.3.2 (Howie *et al.* 2), utilizando como datos de referencia la fase 3 b37 del estudio 1000 Genomes siguiendo las indicaciones de uso de IMPUTE2. Tras la imputación, las asignaciones de genotipo (*calls*) con una incertidumbre $>0,1$ fueron tratadas como perdidas, así como las variantes con una puntuación de calidad de imputación (*info*, *imputation quality score*) menor a 0,8. Finalmente, los mismos controles de calidad aplicados a los SNPs genotipados se aplicaron a los SNPs imputados.

En la segunda fase del estudio de SUD, las variantes fueron imputadas utilizando el Michigan Imputation Server (<https://imputationserver.sph.umich.edu/index.html>) (Das *et al.* 2016) utilizando como panel de referencia la versión fase 3 v5 b37 del proyecto 1000 Genomes, SHAPEIT en su versión v2.r837 y Minimac3 (<https://genome.sph.umich.edu/wiki/Minimac>) (Das *et al.* 2016). El servidor lleva a cabo varios pasos en segundo plano de tipo control: de calidad estadística (con el fin de detectar variantes duplicadas, SNPs eliminados, control del MAF etc.) y de validez de los datos introducidos (muestras, cromosomas, SNPs).

3.5 MODELOS DE RIESGO POLIGÉNICO

Las PRS de cada individuo en la muestra diana fueron calculadas como la suma de los efectos por SNP. El efecto de cada variante resulta de la ponderación por el logaritmo neperiano de las OR de la muestra descubrimiento.

3.5.1 Muestras descubrimiento

En este subapartado se resumen las características de las distintas muestras descubrimiento utilizadas en los distintos estudios de esta tesis doctoral para la generación de modelos de riesgo poligénico.

3.5.1.1 Muestras descubrimiento del PGC

Los modelos de riesgo poligénico utilizados en esta tesis doctoral provienen, mayoritariamente, del uso de estadísticas resumen de acceso público. La mayor parte de éstos proviene del PGC. Esta organización, como se ha comentado en la introducción de esta Tesis, es uno de los principales consorcios genómicos del mundo que apoya la ciencia abierta y el intercambio de datos. Esta estrategia responde a una perspectiva “precompetitiva”, considerando la identificación de resultados genómicos como un bien público (Sullivan *et al.* 2018). Para asegurar el progreso en el campo, ofrecen un gran número de estadísticas resumen obtenidas a partir de distintos GWAS llevados a cabo a lo largo de su existencia como organización. La difusión de resultados se realiza a través del sitio web propio, <https://pgc.unc.edu>, un repositorio de acceso controlado asociado al centro University of North Carolina at Chapel Hill School of Medicine. Según indica el sitio web, los datos deben ser utilizados únicamente con fines científicos y/o educativos. Debido a que el PGC incluye numerosos grupos de investigación, se debe esperar siempre cierta heterogeneidad en el origen de las muestras, siendo principalmente de origen europeo (los distintos estudios presentan muestras con un porcentaje de europeos que oscila entre 85% y la totalidad de la muestra). Los recursos utilizados para conformar los distintos modelos de riesgo obtenidos a partir de la iniciativa del PGC se presentan a continuación ordenados por trastorno (la Tabla 1 muestra la información de forma resumida):

TABLA 1. MUESTRAS DESCUBRIMIENTO PGC. El tamaño efectivo de las muestras fue calculado como el doble de la media armónica (parte de las muestras incluyen pseudocontroles o pseudocasos de tríos). Las estimaciones de heredabilidad de las muestras de trastornos individuales de CROSS son las aportadas por Cross-Disorder Group of the PGC 2013, en ANX se aportan las obtenidas por dos métodos distintos, la de MDD2 fue calculada sin los individuos excluidos del conjunto muestral accesible en el repositorio del PGC y la de SCZ2 está basada en un conjunto muestral menor al expuesto aquí, debido a diferentes controles de calidad aplicados. *Los distintos acrónimos responden a las siguientes referencias: ADHD-2017 (Demontis *et al.* 2017), ASD-2017 (Autism Spectrum Disorders Working Group of the PGC 2016), BP-CROSS (Cross-Disorder Group of the PGC 2013), MDD-CROSS (Cross-Disorder Group of the PGC 2013), MDD2 (Wray *et al.* 2018), SCZ-CROSS (Cross-Disorder Group of the PGC 2013), CROSS (Cross-Disorder Group of the PGC 2013), SCZ2 (Schizophrenia Working Group of the PGC 2014) y ANX (Otowa *et al.* 2016).

ACRÓNIMO MUESTRA*	TRASTORNO / RASGO	N CASOS	N CONTROLES	TAMAÑO EFECTIVO	SNPs - h ² (E.S.)
ADHD-2017	Trastorno por déficit de atención e hiperactividad	19099	34194	49017	0,22 (0,01)
ASD-2017	Trastorno del espectro autista	7387	8567	15867	0,18 (0,02)
BP-CROSS	Trastorno bipolar	6990	4820	11411	0,25 (0,012)
MDD-CROSS	Trastorno depresivo mayor	9227	7383	16405	0,21 (0,021)
MDD2	Trastorno depresivo mayor	59851	113154	173005	0,1421 (0,037)
SCZ-CROSS	Esquizofrenia	9379	7736	16957	0,23 (0,008)
CROSS	Cross-Disorder	33332	27888	60736	—
SCZ2	Esquizofrenia	35476	46839	80746	0,274 (0,0007)
ANX	Trastorno de ansiedad generalizada	5690	11620	15279	0,095 (0,037) 0,138 (0,18)

- **Trastorno por déficit de atención e hiperactividad:** Muestra conformada por 19099 casos y 34194 controles, con un tamaño efectivo de 49017 individuos, todas de origen europeo y 8047421 variantes (Demontis *et al.* 2017). El utilizado para esta muestra es **ADHD-2017**.

- **Trastorno del espectro autista:** Muestra constituida por 7387 casos y 8567 controles, con un tamaño efectivo de 15867 individuos. El 85,08% de la muestra es de origen europeo. Incluye 6517325 variantes (Autism Spectrum Disorders Working Group of the PGC 2016). El acrónimo utilizado para esta muestra es **ASD-2017**.

- **Trastorno bipolar:** Muestra conformada por 6690 casos y 4820 controles, con un tamaño efectivo de 11411 individuos. Todas las muestras son de origen europeo. Esta muestra es una muestra secundaria del estudio Cross-Disorder e incluye 1233535 variantes (Cross-Disorder Group of the PGC 2013). El acrónimo utilizado para esta muestra es **BP-CROSS**.

- **Trastorno depresivo mayor:** Dos muestras. La primera de ellas está constituida por 9227 casos y 7383 controles, con un tamaño efectivo de 16405 individuos. Todas las muestras son de origen europeo. Esta muestra es una muestra secundaria del estudio Cross-Disorder e incluye 1232794 variantes (Cross-Disorder Group of the PGC 2013). El acrónimo utilizado para esta muestra es **MDD-CROSS**. La segunda, presenta 59851 casos y 113154 controles, con un tamaño efectivo de 173005 individuos. Todas las muestras son de origen europeo. Incluye 9622763 variantes (Wray *et al.* 2018). El acrónimo utilizado para esta muestra es **MDD2**.

- **Esquizofrenia:** Dos muestras. La primera de ellas está conformada por 9379 casos y 7736 controles, con un tamaño efectivo de 16957 individuos. Todas las muestras son de origen europeo. Esta muestra es una muestra secundaria del estudio Cross-Disorder e incluye 1237959 variantes (Cross-Disorder Group of the PGC 2013). El acrónimo utilizado para esta

muestra es **SCZ-CROSS**. La segunda, conformada por 35476 casos y 46839 controles, con un tamaño efectivo de 80746 individuos. El 96,94% de la muestra es de origen europeo. Incluye 9444230 variantes (Schizophrenia Working Group of the PGC 2014). El acrónimo utilizado para esta muestra es **SCZ2**.

- **Cross-Disorder**: Muestra conformada por 33332 casos y 27888 controles, con un tamaño efectivo de 60736 individuos. Todas las muestras son de origen europeo. Esta muestra es la muestra principal del estudio Cross-Disorder, presenta casos de cinco patologías psiquiátricas distintas (Trastorno por déficit de atención e hiperactividad, trastorno del espectro autista, trastorno bipolar, trastorno depresivo mayor y esquizofrenia) e incluye 1249700 variantes (Cross-Disorder Group of the PGC 2013). El acrónimo utilizado para esta muestra es **CROSS**.

- **Trastorno de ansiedad generalizada**: Muestra formada por 5690 casos y 11620 controles, con un tamaño efectivo de 15279 individuos. Todas las muestras son de origen europeo. Esta muestra fue generada por el Anxiety Neuro Genetics Study Consortium e incluye 6330996 variantes (Otowa *et al.* 2016). El acrónimo utilizado para esta muestra es **ANX**. Aunque este proyecto no está promovido o gestionado en su totalidad por el PGC, el conjunto de datos puede ser descargado desde su repositorio.

3.5.1.2 Muestras descubrimiento adicionales

Otros modelos de riesgo poligénico se basaron en cohortes y estudios no dependientes del PGC. Estas cohortes pertenecen a la colección del grupo de Xenética Psiquiátrica del Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS). Fueron empleadas para generar modelos de riesgo poligénico o para ser utilizados como muestras diana y se resumen a continuación:

- **Trastorno obsesivo-compulsivo:** La muestra de trastorno obsesivo-compulsivo fue utilizada para un análisis secundario, utilizando la estadísticas resumen con el efecto de cada SNP para generar modelos de riesgo. Las estadísticas resumen presenta datos de 38305 variantes. Los individuos que conforman esta muestra son 443 casos y 370 controles, con un tamaño efectivo de 813 muestras. El acrónimo utilizado para esta muestra es **OCD**.

- **Dependencia de alcohol:** Muestra conformada por una parte del total de individuos reclutados para la Fase I. Este subgrupo de individuos está formado por 481 casos de abuso/dependencia de alcohol y 583 controles, un total de 1064 individuos sin relaciones de parentesco con los individuos de la Fase II. Incluye 4257523 variantes y su acrónimo es **ALC**. Fue utilizada para la generación de un modelo de predicción de riesgo genético combinando distintas muestras descubrimiento mediante una estandarización ponderada (desarrollo del método en el apartado 3.6).

3.5.2 SNPs a incluir en el modelo

Las distintas variantes de cada muestra descubrimiento fueron filtradas y seleccionadas siguiendo los siguientes pasos generales: puntuación de calidad de imputación de variantes, eliminación de SNPs en situación de ambigüedad de cadena, podado informativo de variantes correlacionadas (*pruned*). Debido al complejo patrón de desequilibrio de ligamiento del xMHC, todos los análisis se realizaron por duplicado. Los resultados principales son aquellos en los que se ha eliminado el xMHC de la megabase 26 a la 33 del cromosoma 6 en la versión del genoma hg19 (*Genome Reference Consortium Human Build 37 [GRCh37]*) (Bakker *et al.* 2006). Los modelos poligénicos se generaron seleccionando variantes de la muestra descubrimiento empleando distintos umbrales de significación e incluyendo únicamente variantes autosómicas.

Para el cuarto estudio de esta tesis se emplearon modelos de riesgo poligénico fundamentados en la selección de variantes en base a criterios biológicos funcionales. El estudio de Arloth y colaboradores (Arloth *et al.* 2015), en el que se analizan cambios transcripcionales asociados a *loci* de caracteres cuantitativos de expresión en *cis* (*cis*-eQTLs) en respuesta a la activación del receptor de glucocorticoides mediante el agonista dexametasona, fue utilizado como una fuente de variantes asociadas a la respuesta a estrés. Estas variantes son una lista 296 tag *cis*-eSNPs (SNPs de expresión *cis*) no correlacionados asociados a 320 *cis*-eQTLs de respuesta transcripcional mediada por el receptor de glucocorticoides. Los análisis originales fueron realizados en 160 individuos (91 casos y 61 controles) y se publicaron en la revista *Neuron* en el 2015. Debido al solapamiento incompleto entre el conjunto de variantes del manuscrito original y la muestra diana se utilizó un método basado en la búsqueda de variantes en desequilibrio de ligamiento con la del listado original. SNI PA (*Single Nucleotide Polymorphisms Annotator*) fue la herramienta bioinformática online empleada (Arnold *et al.* 2015), permitiendo buscar variantes en LD en ventanas de 500kb. Siguiendo las indicaciones y requerimientos del software se indicaron los siguientes parámetros: versión de ensamblaje del genoma (hg19), datos para el cálculo de LD (1000 genomas fase 3v5), tipo de población (europea), el tipo de variantes incluidas y el valor de r^2 a emplear ($r^2=0,5$).

Además de los filtros y procedimientos desarrollados anteriormente (selección de variantes en función de la puntuación info de imputación de variantes, podado informativo informativo de SNPs, etc.) fue necesario coordinar distintas versiones del genoma, ya que algunas muestras descubrimiento empleadas no presentaban la misma versión de ensamblaje (*genome assembly*). Las muestras diana se encuentran codificadas en la versión de ensamblaje hg19 y aquellas muestras descubrimiento pertenecientes o derivadas de la iniciativa Cross-disorder en hg18 (NCBI36). Esta diferencia entre versiones del genoma humano implica un procedimiento extra. Para realizar el ajuste de la versión hg18 a la hg19 fueron utilizadas dos herramientas bioinformáticas: Lift Genome Annotations (UCSC Genome Browser Utilities, <https://genome.ucsc.edu/cgi-bin/hgLiftOver>) y NCBI

Genome Remapping Service (<https://www.ncbi.nlm.nih.gov/genome/tools/remap>). Ambas herramientas permiten múltiples opciones, siendo utilizado para este trabajo su instrumento para cambio de versión de ensamblaje siguiendo las indicaciones de ambos servicios.

3.5.3 Generación de los modelos de riesgo poligénico

3.5.3.1 Estudio 1. Análisis de la arquitectura genética compartida entre el trastorno obsesivo-compulsivo y la esquizofrenia

- Análisis principal (SCZ2 vs trastorno obsesivo-compulsivo):

Los modelos de riesgo poligénico fueron generados a partir de la muestra SCZ2 siguiendo los siguientes pasos: tras comprobar la versión genómica de los datos y confirmar el alelo de riesgo, todos los SNPs con la combinación de alelos A/T o C/G fueron eliminados para evitar problemas por ambigüedad de cadena y, los SNPs imputados con una puntuación info (calidad de imputación) menor a 0,9, también fueron excluidos de los modelos de riesgo. Se mantuvieron aquellos SNPs presentes en ambas muestras y se controló la presencia de variantes repetidas. Los SNPs correlacionados entre sí por desequilibrio de ligamiento fueron eliminados mediante un proceso de podado informativo en PLINK v1.07, considerando una $r^2=0,2$ y una ventana de 500 kilobases de tamaño. Además, los datos fueron preparados para cumplir los parámetros y formato de PRSice, el software empleado para llevar a cabo las puntuaciones individuales.

Se utilizaron un total de 12 umbrales de significación para la generación de modelos de riesgo: 0,01; 0,05; 0,1; 0,15; 0,2; 0,25; 0,3; 0,35; 0,4; 0,45; 0,5 y 1 (totalidad de las variantes). Debido al patrón de desequilibrio de ligamiento extremadamente complicado del MHC se llevaron a cabo análisis prescindiendo de las variantes situadas en esa región. Los análisis que excluyeron el xMHC eliminaron todas las variantes que se encontraban en la región contenida entre la megabase 26 y la 33 del cromosoma 6 (el rango que delimita el xMHC está marcado por parámetros propios de PRSice).

- Análisis secundario (OCD vs SCZ-5100):

Para generar los modelos de riesgo se llevó a cabo un GWAS incluyendo las variantes con los mismos filtros aplicados en el análisis previo, corrigiendo la posible estratificación poblacional mediante el uso de las 10 primeras dimensiones de MDS como control.

Los modelos de riesgo siguieron unos parámetros muy parecidos a los comentados en detalle en el análisis principal. Se comprobó la versión de ensamblaje del genoma de ambas muestras, se eliminaron variantes ambiguas, se controló la presencia de variantes repetidas, se mantuvieron los SNPs presentes en ambas muestras y se seleccionaron variantes mediante podado informativo para evitar SNPs relacionados entre sí siguiendo los mismos valores que en el análisis principal. Además, los datos fueron preparados para cumplir los parámetros y formato de PRSice, el software empleado para llevar a cabo las PRS.

Se utilizaron 20 umbrales de significación para la generación de los modelos de riesgo: 0,05; 0,1; 0,15; 0,2; 0,25; 0,3; 0,35; 0,4; 0,45; 0,5; 0,55; 0,6; 0,65; 0,7; 0,75; 0,8; 0,85; 0,9; 0,95 y 1. Los análisis incluyeron el xMHC.

3.5.3.2 Estudio 2. Análisis de la arquitectura genética compartida entre el trastorno por abuso o dependencia de sustancias y los principales trastornos psiquiátricos.

- Trastornos psiquiátricos principales vs trastorno por abuso o dependencia de sustancias

En primer lugar se analizó el efecto en SUD de distintos modelos de riesgo poligénico basados en diferentes patologías psiquiátricas con los siguientes conjuntos de muestras descubrimiento: trastorno por déficit de atención e hiperactividad (ADHD-2017), ansiedad (ANX), trastorno del espectro autista (ASD-2017), trastorno bipolar (BP-CROSS), trastorno depresivo mayor (MDD-CROSS) y esquizofrenia (SCZ-CROSS).

Respecto a la preparación de los modelos de riesgo poligénico, los distintos pasos llevados a cabo se resumen a continuación: se confirmó el alelo de riesgo en cada muestra, las variantes con una puntuación info menor a 0,8 fueron excluidas, todos los SNPs con la combinación de alelos A/T o C/G fueron eliminados para evitar problemas por ambigüedad de cadena, se controló la presencia de variantes repetidas y se mantuvieron únicamente aquellas presentes también en la muestra diana. Varias muestras descubrimiento se encontraban en una versión de ensamblaje del genoma anterior (hg18) al de nuestra muestra diana. Por lo tanto fue necesario coordinar las distintas versiones del genoma con el uso de herramientas que permiten convertir datos de anotación. Tras el cambio de versión al deseado (hg19) se realizaron controles de duplicación de variantes, se eligieron las de ensamblaje primario, y se seleccionaron aquellas solapantes con la muestra diana. Las variantes correlacionadas entre sí por LD fueron eliminadas mediante un algoritmo de podado informativo en PLINK v1.9, considerando una $r^2=0,1$ y una ventana de 500kb. Los datos fueron preparados para realizar las PRS siguiendo los parámetros de PLINK v1.9. La región correspondiente al xMHC fue eliminada en todas las pruebas principales (región del cromosoma 6 de la megabase 26 a la 33), manteniéndose en análisis secundarios.

En los análisis iniciales, empleando los principales trastornos psiquiátricos, se utilizaron un total de 6 umbrales de significación para la selección de variantes que conforman los distintos modelos de riesgo poligénico: 0,001; 0,05; 0,1; 0,25; 0,5 y 1. Todas las PRS se llevaron a cabo con la función “—score” de PLINK v1.9.

- SCZ2/CROSS vs trastorno por abuso o dependencia de sustancias

En segundo lugar, analizó la capacidad predictiva de los modelos de riesgo poligénico generados a partir de SCZ2 y CROSS.

La preparación y selección de los modelos de riesgo basados en SCZ2 siguió unos pasos prácticamente similares a los del primer análisis. Se confirmó el alelo de riesgo, filtro $INFO>0,8$, control de ambigüedad de cadena, variantes repetidas y se realizó un podado

informativo de los datos siguiendo los mismos parámetros ($r^2=0,1$ y ventana de 500kb).

La preparación de los modelos de riesgo poligénico de CROSS siguió los mismos pasos y parámetros que en SCZ2, aunque fue necesario el cambio de versión de ensamblaje del genoma, realizado de igual manera que con las muestras de trastornos derivadas de CROSS empleadas en el primer análisis de este estudio.

La región correspondiente al xMHC fue eliminada en todas las pruebas principales (región del cromosoma 6 de la megabase 26 a la 33), manteniéndose en análisis secundarios.

Los datos fueron preparados para realizar las PRS siguiendo los parámetros de PLINK 1.9. Estos análisis presentan la inclusión de un mayor número de umbrales de selección de variantes para generar los modelos. Esto se efectuó con el fin de conocer mejor el efecto en esquizofrenia tras detectarlo con SCZ-CROSS. En este análisis se utilizaron 10: 5×10^{-8} ; 0,001; 0,01; 0,05; 0,1; 0,2; 0,3; 0,5; 0,75 y 1.

Los umbrales empleados con CROSS son los mismos que los empleados en SCZ2.

3.5.3.3 Estudio 3. Análisis de la capacidad predictiva de las PRS en la modelización combinada con factores predictivos de riesgo conocidos

Modelo predictivo combinado, basado en la suma ponderada estandarizada de las PRS generadas en la Fase 2 de SUD a partir de modelos basados en la muestra SCZ2 y ALC.

La preparación y selección de los modelos de riesgo basados en SCZ2 siguió unos pasos similares a los del análisis del Estudio 2. Se confirmó el alelo de riesgo, filtro $INFO > 0,8$, control de ambigüedad de cadena, variantes repetidas y se realizó un podado informativo de los datos siguiendo los mismos parámetros ($r^2=0,1$ y ventana de 500kb).

3.5.3.4 Estudio 4. Análisis de la capacidad predictiva de modelos poligénicos fundamentados en la selección de variantes en base a criterios biológicos funcionales

- Respuesta a estrés vs dependencia de alcohol

En el caso del conjunto de datos basados en respuesta a estrés, los modelos se basaron en la muestra MDD2. Los distintos pasos llevados a cabo antes de la combinación o filtrado con el listado de *cis*-eSNPs basado en el conjunto de variantes de respuesta a estrés (Arloth *et al.* 2015) se presentan a continuación: se confirmó el alelo de riesgo, las variantes con una puntuación info menor a 0,3 fueron excluidas, se controló la presencia de variantes repetidas y, por último, se seleccionaron aquellas variantes en común con la muestra diana de SUD_1524 (muestra caso-control basada en la unión de Fase 1 y Fase 2 de SUD, seleccionando aquellos pacientes con trastorno por abuso o dependencia de alcohol). Se mantuvieron las variantes susceptibles de ambigüedad de cadena ya que, el modelo final, tendría un tamaño relativamente pequeño como para analizar aquellas variantes problemáticas una a una comprobando la distribución de las frecuencias de cada alelo, siendo prioritaria la necesidad de cubrir el mayor número de variantes de la lista de *cis*-eSNP. Los datos fueron preparados para realizar las PRS siguiendo los parámetros de PLINK v1.9. Los umbrales empleados para la generación de modelos se presentan a continuación: 0,05; 0,1; 0,25; 0,5; 0,75 y 1.

3.6 ANÁLISIS ESTADÍSTICOS

3.6.1 Análisis de regresión

Los modelos de riesgo poligénico fueron evaluados en el software R (<https://www.r-project.org/>) mediante regresión logística y regresión lineal. Como covariables, fueron utilizadas el porcentaje de genotipos perdidos, la edad y sexo de los individuos incluidos en cada análisis, las diez primeras dimensiones generadas con escalamiento multidimensional para ajustar la posible estratificación poblacional

(utilizando todas o solamente aquellas significativas, dependiendo del análisis) e información psicológica o ambiental en los análisis en los que esa información es examinada. Los distintos análisis y sus características particulares (covariables, umbrales, ajustes, etc.) se detallan en los siguientes subapartados.

El porcentaje de varianza explicada se estima como el incremento de la pseudo R^2 de Nagelkerke (Nagelkerke 1991), coeficiente de determinación utilizado en regresión logística para medir la correlación entre las variables independientes y la variable categórica, entre el modelo completo y el modelo sin PRS.

Además, mediante el uso de regresión logística se analizaron posibles interacciones entre covariables y distintos subgrupos de muestras fueron analizados por separado. Mediante el uso de regresión lineal fueron analizadas posibles mediaciones entre la genética y otras variables.

3.6.2 Estudio 1. Análisis de la arquitectura genética compartida entre el trastorno obsesivo-compulsivo y la esquizofrenia

- Análisis principal (SCZ2 vs trastorno obsesivo-compulsivo):

En los análisis de regresión logística, las 10 primeras dimensiones de MDS fueron incluidas como covariables informativas de ancestralidad con el fin de ajustar el modelo.

- Análisis secundario (OCD vs SCZ-5100):

En los análisis de regresión logística, las 10 primeras dimensiones de MDS fueron incluidas como covariables informativas de ancestralidad con el fin de ajustar el modelo.

3.6.3 Estudio 2. Análisis de la arquitectura genética compartida entre el trastorno por abuso o dependencia de sustancias y los principales trastornos psiquiátricos

- Trastornos psiquiátricos principales vs trastorno por abuso o dependencia de sustancias

Los análisis de regresión logística utilizados para conocer la validez de los distintos modelos incluyeron las dimensiones de MDS significativas (la segunda y octava de las diez primeras), el sexo, la edad y el porcentaje de genotipos perdidos. Las PRS fueron estandarizadas para presentar una media de 0 y una desviación estándar de 1, con el propósito de obtener resultados más fácilmente interpretables.

- SCZ2/CROSS vs trastorno por abuso o dependencia de sustancias

Se emplearon las mismas covariables con SCZ2 y CROSS que en los análisis iniciales con los trastornos psiquiátricos principales.

3.6.4 Estudio 3. Análisis de la capacidad predictiva de las prs en la modelización combinada con factores predictivos de riesgo conocidos

La suma ponderada estandarizada de PRS (ver subapartado “3.6.8 Combinación de modelos de riesgo poligénico” fueron evaluadas mediante regresión logística, con las covariables de ajuste (edad, sexo y diez primeras dimensiones de MDS) y el resto de posibles covariables basadas en predictores de riesgo conocidos: historia familiar (dicotómica), puntuación global del CTQ, puntuaciones en cada una de las cinco subescalas de CTQ, LTE (dicotómico, presencia/ausencia; continuo: recuentos acumulados) y las puntuaciones para cada uno de los factores de la personalidad del NEO-FFI. Las muestras que mostrasen ausencia de información o datos para alguno de estos predictores de riesgo fueron separadas de cada análisis.

3.6.5 Estudio 4. Análisis de la capacidad predictiva de modelos poligénicos fundamentados en la selección de variantes en base a criterios biológicos funcionales

- Respuesta a estrés vs dependencia de alcohol

Tras la selección de SNPs en base a las variantes de respuesta a estrés y MDD2 para la generación de modelos de riesgo se prepararon aquellas variables de ajuste necesarias. Edad, sexo y las diez primeras dimensiones de MDS fueron las covariables utilizadas. Para el cálculo de MDS, se realizó un podado informativo de datos que redujo a 84661 variantes el conjunto de datos ($r^2=0,1$ y ventana de 500kb).

Por último, se decidió analizar las posibles interacciones entre PRSs basadas en las variantes de respuesta a estrés y trauma infantil. Para ello, se utilizó la segunda fase de SUD como muestra diana, empleando la información traumática infantil recogida para todos los individuos que conforman la muestra. Mediante modelos de regresión logística, ajustados por edad, sexo y diez primeras dimensiones de MDS, se analizó la interacción PRS*CTQ mediante la inclusión en el modelo de todas las interacciones posibles entre la variable CTQ y la variable PRS con el resto de covariables de ajuste.

3.6.6 Ajustes para contrastes múltiples

Método de Bonferroni: Este procedimiento es uno de los más utilizados ya que permite contrastar de forma sencilla y sin asunciones adicionales cada una de las hipótesis individuales involucradas en el estudio. El nivel de significación se calcula dividiendo el valor α (nivel de significación) entre k (número de contrastes).

Permutaciones caso-control/sexo: Este método se basa en la permutación de estatus caso-control o la etiqueta de sexo generando así un conjunto artificial de datos bajo la hipótesis nula. Estos pseudocasos/pseudocontroles y los individuos con la etiqueta sexual permutada permiten realizar un contraste de frecuencias entre pseudogrupos, obteniendo una distribución empírica del estadístico al

azar para ser comparado con el valor real. El principal inconveniente de este método es su alto consumo computacional. Si el número de cálculos es elevado el proceso puede ser lento. En este trabajo se realizaron permutaciones simples de fenotipo y procedimientos más exigentes a nivel computacional, combinando múltiples fenotipos.

Permutaciones empleando SNPsnap (Pers *et al.* 2014): Este método se basa en la generación de modelos de riesgo a partir de conjuntos de variantes con características parecidas a un conjunto original o de referencia. Para la selección de estas variantes de características parecidas se utilizó la herramienta bioinformática online SNPsnap. Las variantes fueron seleccionadas apareando con el conjunto de SNPs la frecuencia de alelo menor, patrones de LD, distancia a genes cercanos u otras regiones genómicas con el fin de generar conjuntos de variantes que permitan generar modelos de riesgo basándose en un conjunto original. Al igual que las permutaciones caso-control/sexo, los resultados de los modelos generados a partir de la búsqueda con SNPsnap permiten realizar un contraste entre modelos, obteniendo una distribución empírica del estadístico al azar con el fin de ser comparada con el valor del modelo original.

3.6.7 Correlación

La correlación entre covariables e información genética fue evaluada mediante el método de Pearson y el de Spearman. Estos métodos fueron ampliamente utilizados con el fin de conocer el grado de relación entre distintas variables.

3.6.8 Combinación de modelos de riesgo poligénico

Este método empleado se basó en la estandarización ponderada presentada en el trabajo de Grove y colaboradores (Grove *et al.* 2019). Este procedimiento busca explotar el solapamiento genético con otros fenotipos para ayudar a mejorar la predicción de un PRS, combinando PRS de un fenotipo con la puntuación previamente internamente

entrenada en SUD. En este caso, las muestras descubrimiento utilizadas son ALC y SCZ2. El modelo de cada fenotipo incluido debe ser aquel que mayor varianza explique en el entrenamiento previo en una muestra independiente. A continuación se muestra la estandarización ponderada utilizada:

$$S_k = \frac{S_{AD} + \sum_{i=1}^k \log(OR_{P_i})S_{P_i} - \mu(S_{AD} + \sum_{i=1}^k \log(OR_{P_i})S_{P_i})}{\sigma(S_{AD} + \sum_{i=1}^k \log(OR_{P_i})S_{P_i})}$$

Donde S_{AD} es la puntuación generada con el modelo SUD, S_{P_i} es la puntuación para el fenotipo P_i , OR_{P_i} es la OR del modelo de regresión logística de S_{P_i} sobre SUD (como factor de ajuste), μ es la media, σ la desviación estándar y k el número de individuos.

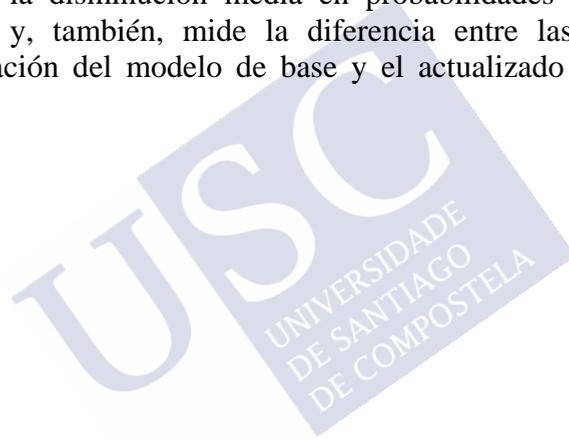
3.6.9 Evaluación de modelos de predicción

Área bajo la curva característica operativa del receptor (AUC-ROC): Índice que puede emplearse para analizar la exactitud global de la prueba, pudiendo interpretarse como la probabilidad de que un modelo poligénico ordenará o puntuará un caso elegido aleatoriamente más alto que a un control. El incremento en esta medida tras la adición del modelo genético es evaluado, siendo 0 el valor mínimo y 1 el máximo. Es el método más utilizado para la evaluación de modelos de predicción de riesgo (Martens *et al.* 2018). Esta medida permite discriminar entre individuos que manifestarán o no el resultado de interés. El incremento de esta medida tras la adición de la variable de PRS se mide como Δ AUC-ROC, evaluando el aumento de la capacidad discriminatoria.

Mejora de Reclasificación Neta (NRI, *Net Reclassification Improvement*): Índice que busca cuantificar cómo de bien un nuevo modelo reclasifica sujetos en comparación con un modelo anterior, analizando cómo la adición de factores de riesgo conduce a una mejor clasificación de riesgos. Es “la suma del porcentaje de reclasificaciones correctas en eventos (pacientes) y no-eventos (controles), siendo la reclasificación los cambios entre categorías de riesgo en NRI categórico

y cambios en el riesgo en NRI de tipo continuo” (Martens *et al.* 2018). En la tesis actual se emplea el método NRI continuo. El NRI mide cambios en clasificaciones de riesgo. La sensibilidad y la especificidad son las proporciones de las correctas reclasificaciones.

Mejora de Discriminación Integrada (IDI, *Integrated Discrimination Improvement*): Este índice es una herramienta para evaluar la capacidad de un marcador para predecir un resultado binario de interés. IDI evalúa la mejora de la diferencia de riesgo entre eventos (pacientes) y no-eventos (controles) de la prueba, evaluando cambios en las diferencias de riesgo. “Analiza la diferencia entre los incrementos medios y la disminución media en probabilidades estimadas entre modelos” y, también, mide la diferencia entre las pendientes de discriminación del modelo de base y el actualizado (Martens *et al.* 2018).





RESULTADOS





4. RESULTADOS

4.1 ESTUDIO 1. ANÁLISIS DE LA ARQUITECTURA GENÉTICA COMPARTIDA ENTRE EL TRASTORNO OBSESIVO-COMPULSIVO Y LA ESQUIZOFRENIA

Siendo el principal objetivo de este estudio el analizar el posible solapamiento de la arquitectura genética entre el trastorno obsesivo-compulsivo y la esquizofrenia, se llevaron a cabo dos análisis distintos (Figura 2):

- Análisis principal: Muestra descubrimiento SCZ2 vs muestra diana, formada por los casos y controles de trastorno obsesivo-compulsivo.
- Análisis secundario: Muestra descubrimiento OCD vs muestra diana SCZ-5100.

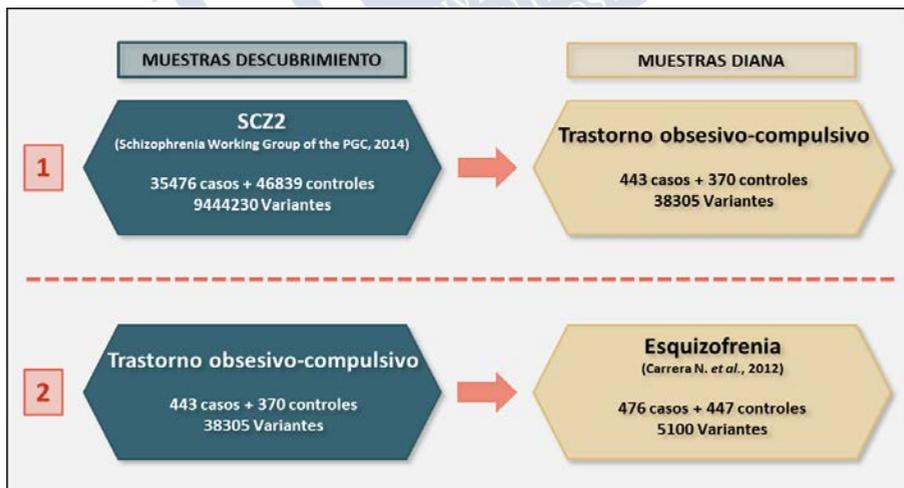


FIGURA 2. ANÁLISIS LLEVADOS A CABO EN EL PRESENTE ESTUDIO.

4.1.1 Genotipado y control de calidad

Tras el proceso de genotipado se obtuvieron datos genotípicos para 295983 SNPs en 484 controles y 433 casos de trastorno obsesivo compulsivo. Tras el control de calidad con los filtros aplicados, el número de variantes y de individuos se redujo a 289339 SNPs y 443 controles y 370 pacientes. El control de la MAF del 5% redujo el número de variantes a 38305. Los detalles de la aplicación del control de calidad son mostrados en la Tabla 2. Los resultados de los análisis de ancestralidad revelan que la muestra es bastante homogénea, siendo eliminados un total de nueve individuos (0,981% de la muestra enviada al servicio de genotipado) al presentar más de un 10% de los genotipos de origen no europeo.

TABLA 2. DETALLES DE LOS CRITERIOS DE FILTRADO EMPLEADOS EN EL CONTROL DE CALIDAD APLICADO EN LOS DATOS DEL TRASTONO OBSESIVO-COMPULSIVO

FILTRO	SNPs	CONTROLES	CASOS
Tamaño inicial	295983	484	433
Ratio de genotipación por SNP >95%	290105	484	433
Desviaciones equilibrio Hardy-Weinberg ($P > 1 \times 10^{-3}$)	289564	484	433
Ratio de genotipación casos vs controles ($P > 1 \times 10^{-3}$)	289339	484	433
MAF 5%	38305	484	433
Ratio de genotipación por muestra >95%	38305	477	423
Sexo discordante	38305	456	412
Niveles de heterocigosidad	38305	450	395
Relaciones crípticas (PI-HAT)	38305	445	377
Ancestralidad	38305	443	370

4.1.2 Análisis principal: esquizofrenia vs trastorno obsesivo-compulsivo

Los resultados del análisis principal mostraron que los modelos de riesgo basados en la muestra SCZ2 del PGC lograron diferenciar significativamente en la muestra diana, y en la dirección esperada, entre pacientes con trastorno obsesivo-compulsivo y controles.

Ambas muestras presentaban en común un total de 16675 variantes. Como se refleja en la Figura 3A el modelo que explicó una mayor varianza en la muestra diana fue el basado en el umbral 0,05. Este modelo empleó 2760 SNPs independientes que explicaron un 3,17% de la varianza con una p de significación de la puntuación de $1,35 \times 10^{-5}$. Asumiendo un OR=1 para el primer quintil, el quinto presentó una OR de 2,44 (95% I.C.=1,52-3,92, $p=1,4 \times 10^{-4}$). Al realizar el análisis sin el xMHC, el porcentaje de varianza explicada mejora pasando a ser del 3,7% ($p=2,3 \times 10^{-6}$), lográndose con el mismo umbral que el resultado comentado anteriormente (Figura 3B).

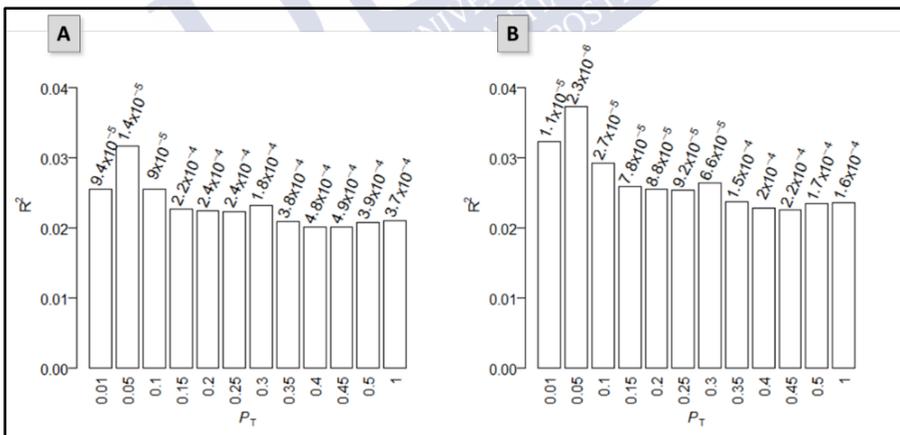


FIGURA 3. RESULTADOS DE LOS ANÁLISIS DE PRS UTILIZANDO SCZ2 COMO MUESTRA DE DESCUBRIMIENTO Y LA MUESTRA DE TRASTORNO OBSESIVO-COMPULSIVO COMO MUESTRA DIANA (ANÁLISIS PRINCIPALES). El eje y representa el porcentaje de varianza explicada en la escala observada. El eje x representa los diferentes umbrales utilizados (P_T). En la parte superior de cada columna de la figura se indica el valor de p de la puntuación. (A) xMHC incluido, (B) xMHC excluido.

Como control interno de los resultados se realizaron 100 permutaciones de fenotipo en la muestra diana de trastorno obsesivo-compulsivo, con el fin de analizar a través de réplicas la resistencia de los resultados. Ninguna de las 100 réplicas generadas con permutaciones aleatorias llegó a explicar una varianza tan amplia como la detectada por el modelo original: la ΔR^2 máxima fue de 1,67%, siendo la media de 0,30% (Figura 4A).

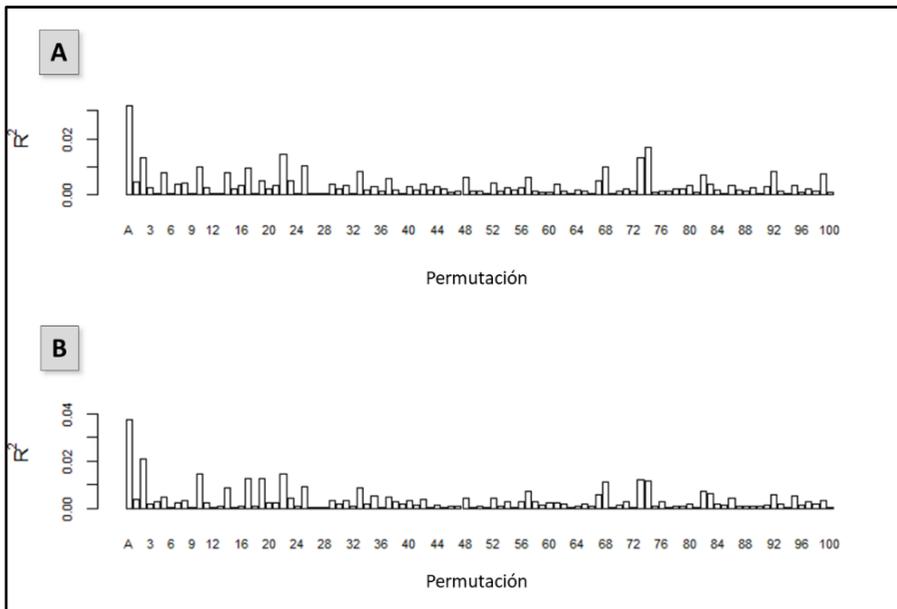


FIGURA 4. 100 PERMUTACIONES CASO-CONTROL. Resultados de las PRS utilizando los datos SCZ2 como muestra descubrimiento y 100 permutaciones al azar de la muestra de trastorno obsesivo-compulsivo como diana. El valor “A” representa los resultados originales. El eje y muestra el porcentaje de varianza explicada en la escala observada. (A) xMHC incluido, (B) xMHC excluido.

4.1.3 Análisis secundario: trastorno obsesivo-compulsivo vs esquizofrenia

Los resultados del análisis secundario mostraron que los modelos de riesgo basados en la muestra OCD lograron diferenciar

significativamente en la muestra diana (SCZ-5100), y en la dirección esperada, entre pacientes de esquizofrenia y controles. Este resultado sirvió para confirmar el resultado principal y añadir mayor firmeza al solapamiento genético detectado en el análisis principal.

A pesar de existir un menor número de variantes entre ambas muestras se detectó el solapamiento entre arquitecturas genéticas de las dos patologías. El efecto más significativo se obtuvo con el umbral 0,25, explicando un 1,16% de la varianza y un nivel de significación de la puntuación $p=0,0044$. Este umbral incluyó 821 SNPs independientes (Figura 5).

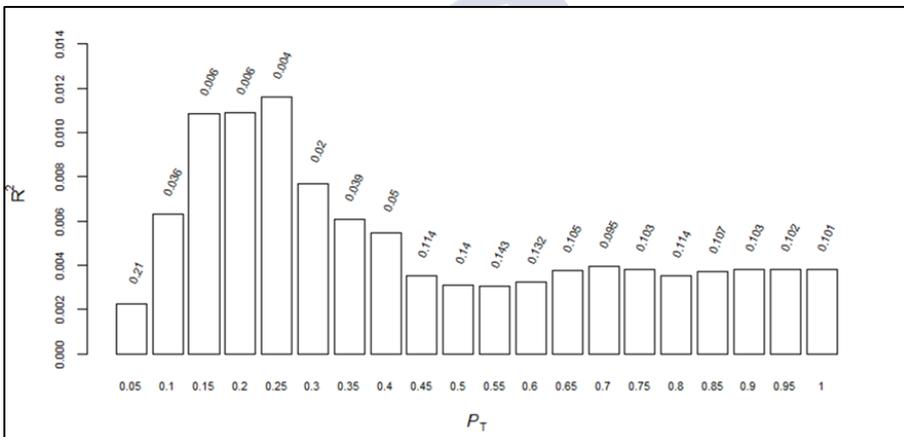


FIGURA 5. RESULTADOS DE LOS ANÁLISIS DE PRS UTILIZANDO OCD COMO MUESTRA DESCUBRIMIENTO Y LA MUESTRA SCZ-5100 COMO MUESTRA DIANA. El eje y representa el porcentaje de varianza explicada en la escala observada. El eje x representa los diferentes umbrales utilizados (P_T). En la parte superior de cada columna de la figura se indica el valor de p de la puntuación.

4.2 ESTUDIO 2. ANÁLISIS DE LA ARQUITECTURA GENÉTICA COMPARTIDA ENTRE EL TRASTORNO POR ABUSO/DEPENDENCIA DE SUSTANCIAS Y LOS PRINCIPALES TRASTORNOS PSIQUIÁTRICOS

Siendo el principal objetivo de este estudio el analizar el posible solapamiento de la arquitectura genética entre los principales trastornos psiquiátricos y el trastorno por abuso o dependencia de sustancias, se llevaron a cabo los siguientes análisis:

- Trastornos psiquiátricos principales vs trastorno por abuso o dependencia de sustancias
- SCZ2/CROSS vs trastorno por abuso o dependencia de sustancias

4.2.1 Genotipado y control de calidad

Tras el proceso de genotipado se obtuvieron datos genotípicos para 571054 variantes en 572 pacientes con trastorno por abuso o dependencia a sustancias (464 hombres y 70 mujeres) y 604 controles (293 hombres y 294 mujeres) (ver Tabla 3 para conocer en detalle el efecto de los filtros del control de calidad). Tras la imputación y el control de calidad el número de variantes e individuos se redujo a 6283934 variantes, 534 casos y 587 controles (1121 individuos), incluyendo un control del alelo menor del 1%, buscando un conjunto de datos sólido. El número medio de sustancias utilizadas por caso fue 3,75 (D.E. 1,51), la mediana fue 4. La Tabla 4 muestra la distribución de los casos por sustancia de uso tras el control de calidad.

Tras el control de calidad de genotipación, el control de la posible estratificación poblacional en los análisis se llevó a cabo mediante MDS. Primero se realizó un podado informativo de los datos ($r^2=0,1$ y ventana de 500kb) pasándose de 6283934 a 127752 variantes, con el fin de realizar el cálculo de MDS de las 10 primeras dimensiones con datos independientes. Después, tras el cálculo de MDS, se analizó mediante regresión logística qué dimensiones de las 10 calculadas eran significativas (el modelo de análisis contempló como variable

categoría el fenotipo, caso-control, y, como covariables, únicamente las dimensiones de MDS con el fin de analizar su correlación). La segunda y octava dimensión fueron significativas (Figura 6).

TABLA 3. DETALLE DE LOS CRITERIOS DE FILTRADO EMPLEADOS EN EL CONTROL DE CALIDAD DE SUD.

FILTRO	SNPs	CONTROLES	CASOS
Tamaño inicial	571054	604	572
Ratio de genotipación por SNP >90%	542324	604	572
Ratio de genotipación por muestra >95%	542324	595	541
Ratio de genotipación por SNP >95%	523339	595	541
Ratio de genotipación casos vs controles ($P > 1 \times 10^{-3}$)	515866	595	541
Desviaciones equilibrio Hardy-Weinberg ($P > 1 \times 10^{-3}$)	515363	595	541
Relaciones críticas (PI-HAT)	515363	587	534
MAF 1%	265827	587	534

TABLA 4. DISTRIBUCIÓN DE LOS CASOS POR SUSTANCIA DE USO TRAS LA APLICACIÓN DE LOS FILTROS Y CONTROLES DE CALIDAD GENOTÍPICOS

SUSTANCIA	N ABUSO (%)	N DEPENDENCIA (%)	N TOTAL (%)
Alcohol	144 (26,97)	359 (67,23)	503 (94,19)
Nicotina	15 (2,81)	475 (88,95)	490 (91,76)
Cocaína	94 (17,60)	219 (41,01)	313 (58,61)
Cannabis	68 (12,73)	217 (40,64)	285 (53,37)

SUSTANCIA	N ABUSO (%)	N DEPENDENCIA (%)	N TOTAL (%)
Opioides	7 (1,31)	218 (40,82)	225 (42,13)
Hipnóticos	50 (9,36)	70 (13,11)	120 (22,47)
Estimulantes	33 (6,18)	8 (1,5)	41 (7,68)
Alucinógenos	26 (4,87)	0 (0,00)	26 (4,87)
Solventes volátiles	0 (0,00)	0 (0,00)	0 (0,00)

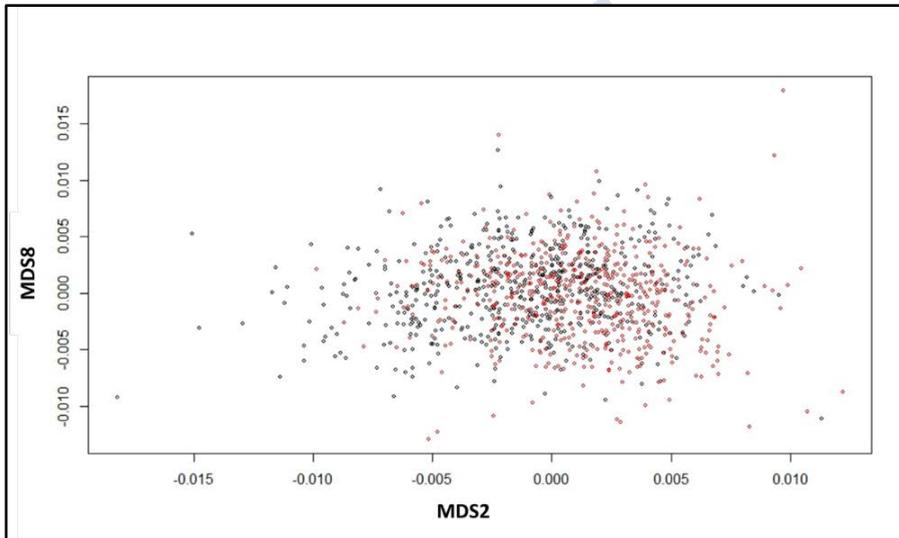


FIGURA 6. CONTROL DE LA ESTRATIFICACIÓN POBLACIONAL MEDIANTE MDS. Representación gráfica de las dos dimensiones de MDS (MDS2 y MDS8) significativas utilizadas para el control de la estratificación poblacional en los análisis del presente estudio.

4.2.2 Análisis predictivos

En primer lugar se analizó el efecto en SUD de distintos modelos de riesgo poligénico basados en diferentes patologías psiquiátricas con los siguientes conjuntos de muestras descubrimiento: trastorno por

déficit de atención e hiperactividad (ADHD-2017), ansiedad (ANX), trastorno del espectro autista (ASD-2017), trastorno bipolar (BP-CROSS), trastorno depresivo mayor (MDD-CROSS) y esquizofrenia (SCZ-CROSS).

Para la corrección del nivel de significación por contrastes múltiples se utilizó un sistema de permutaciones. En total se efectuaron 5000 permutaciones de la etiqueta caso-control de los seis trastornos/modelos utilizados en este análisis. De esta forma se estableció como umbral de significación $p=0,0022$.

La Figura 7 muestra los resultados, sin xMHC (Figura 7A) y con xMHC (Figura 7B) del porcentaje de varianza explicada por cada modelo de riesgo poligénico. Las PRS generadas a partir de los datos SCZ-CROSS son las que explicaron un mayor porcentaje de la varianza entre todos los trastornos empleados (el máximo incremento de la pseudo- $R^2=1,35\%$, umbral=0,25, $p=0,000142$). Cabe indicar que los únicos PRS que se mantienen significativos tras la corrección por permutaciones son los generados con SCZ-CROSS, ya que ADHD-2017 y BP-CROSS, a pesar de presentar resultados significativos a distintos umbrales, no resisten la corrección. Por último, la Figura 8 muestra la alta correlación de los PRS a diferentes umbrales en los distintos trastornos de este primer análisis.

Con el fin de caracterizar aún más la susceptibilidad genética compartida entre las variantes comunes de la esquizofrenia y SUD se llevó a cabo un nuevo análisis con la muestra SCZ2.

Las PRS generadas empleando esta muestra explicaron una mayor variabilidad que los de SCZ-CROSS: incremento máximo de la pseudo- R^2 del 3,92% y $p=2,1 \times 10^{-10}$ con el umbral 0,3 (Figura 9). Un incremento de la desviación estándar en las PRS de SCZ2 estaba asociado con el aumento del 62% de la OR de ser un caso (OR=1,62; intervalo de confianza del 95% 1,40-1,89). Los resultados sin xMHC se representan en la Figura 9 y, aquellos incluyendo la región del xMHC en la Figura 10. Las diferentes PRS están altamente correlacionadas entre sí (Figura 8).

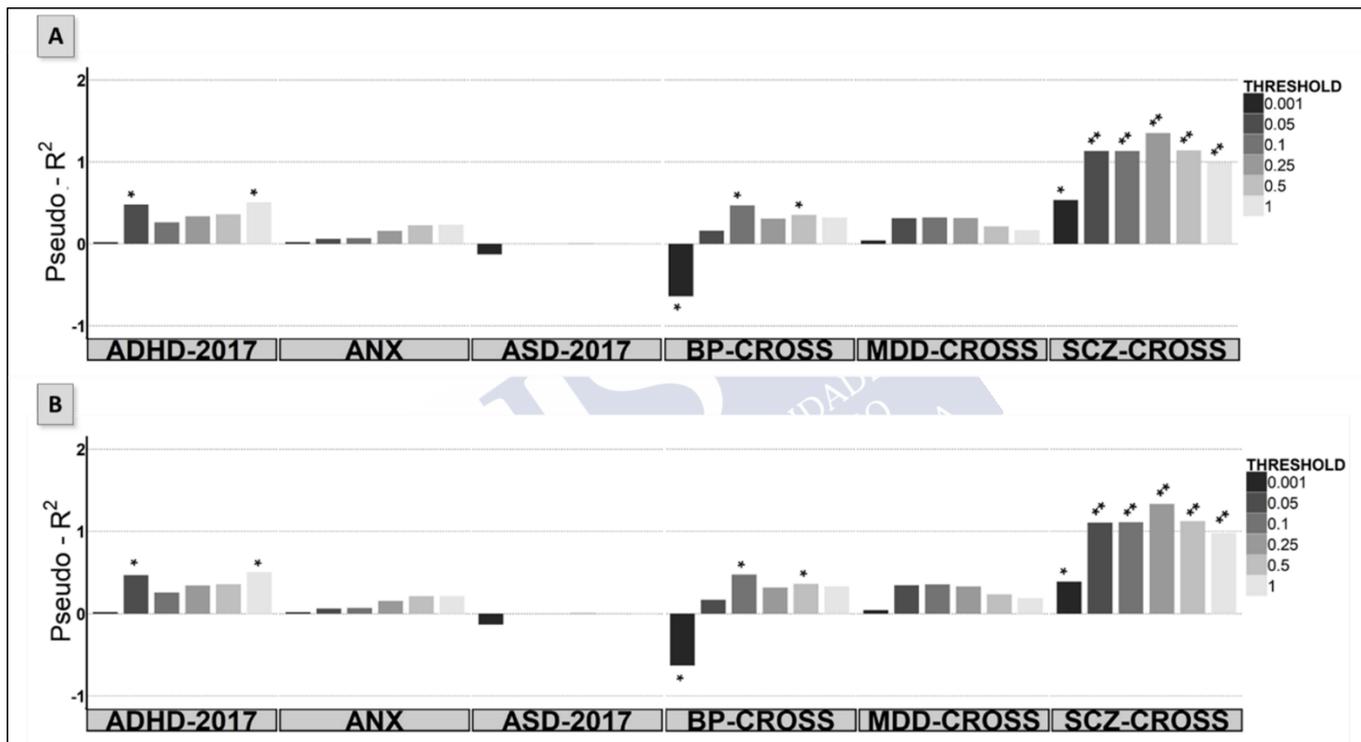


FIGURA 7. PORCENTAJE DE VARIANZA EXPLICADA POR CADA PRS UTILIZANDO SEIS MUESTRAS DESCUBRIMIENTO DISTINTAS. (A) muestra los resultados principales, excluyendo el xMHC; (B) incluyendo el xMHC. La significación de los PRS se indican mediante el uso de “*”: * 0,05<p<0,0022; ** p<0,0022 (corrección para tests múltiples mediante permutaciones). Gráfico creado con el paquete de R ggplot2. El código de colores “threshold” indica el umbral empleado en cada prueba.

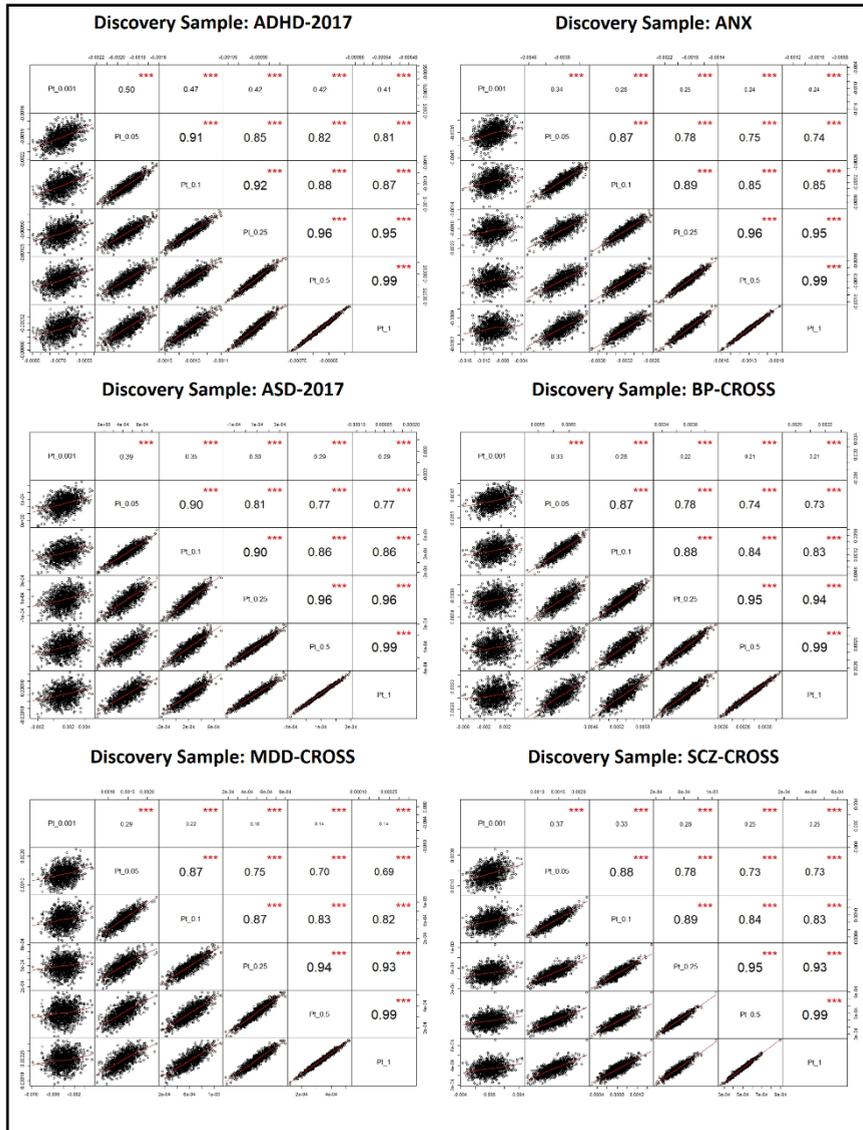


FIGURA 8. CORRELACIÓN ENTRE PRS GENERADOS CON DISTINTOS UMBRALES EN CADA TRASTORNO. Los diferentes umbrales están indicados en la diagonal. Los diagramas de dispersión con línea de tendencia se muestran debajo de la diagonal y, los coeficientes de correlación de Pearson por pares, se muestran por encima de la diagonal. Los niveles de significación de la correlación son: * $0,05 < p < 0,01$; ** $0,01 < p < 0,001$ y *** $p < 0,001$. Gráfico creado con la función `chart.correlation` del paquete de R `PerformanceAnalytics`.

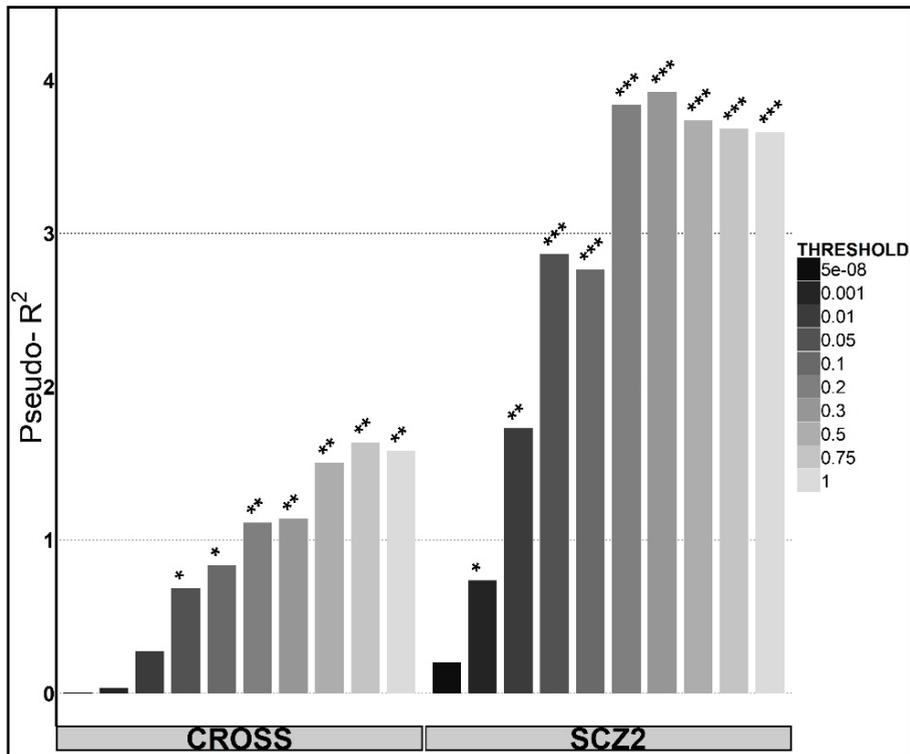


FIGURA 9. PORCENTAJE DE VARIANZA EXPLICADA POR CADA PRS UTILIZANDO CROSS Y SCZ2 COMO MUESTRAS DESCUBRIMIENTO EXCLUYENDO EL xMHC. La significación de las PRS se indica con el siguiente código: * $0,05 < p < 0,0025$ (corrección de Bonferroni para 20 pruebas), ** $0,0025 < p < 1 \times 10^{-5}$; *** $p < 1 \times 10^{-5}$. El código de colores "threshold" indica el umbral empleado en cada prueba. Gráfico creado con el paquete de R ggplot2.

Tras observarse únicamente resultados significativos con el uso de modelos de riesgo poligénico generados a partir de datos de esquizofrenia (SCZ-CROSS) se decidió analizar si esto podría ser debido a una falta de potencia en el resto de muestras descubrimiento del primer análisis. Con el fin de analizar la capacidad predictiva de los modelos generados a partir de una muestra de mayor tamaño y aprovechar sus ventajas, se empleó como muestra descubrimiento los

datos generales de CROSS (conjunto total de individuos sin ser divididos por trastorno).

Las PRS calculadas con los modelos creados a partir de CROSS explicaron un menor porcentaje de la varianza que las de SCZ2: pseudo- $R^2=1,64\%$; $p=2,7 \times 10^{-5}$ con el umbral 0,75 (OR=1,34; intervalo de confianza del 95%=1,17 – 1,54) (Figura 9). Los resultados incluyendo el xMHC se presentan en la Figura 10. Las diferentes PRS están altamente correlacionadas entre sí (Figura 11).

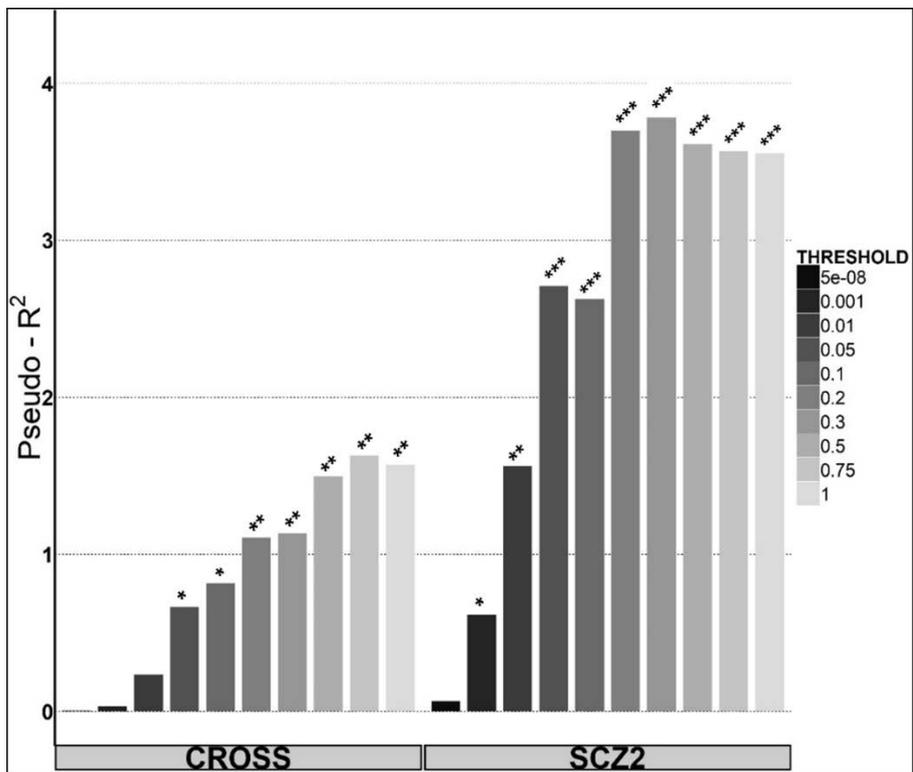


FIGURA 10. PORCENTAJE DE VARIANZA EXPLICADA POR CADA PRS UTILIZANDO CROSS Y SCZ2 COMO MUESTRAS DESCUBRIMIENTO SIN EXCLUIR EL xMHC. La significación de las PRS y el código de colores siguen las mismas pautas que en la Figura 9.

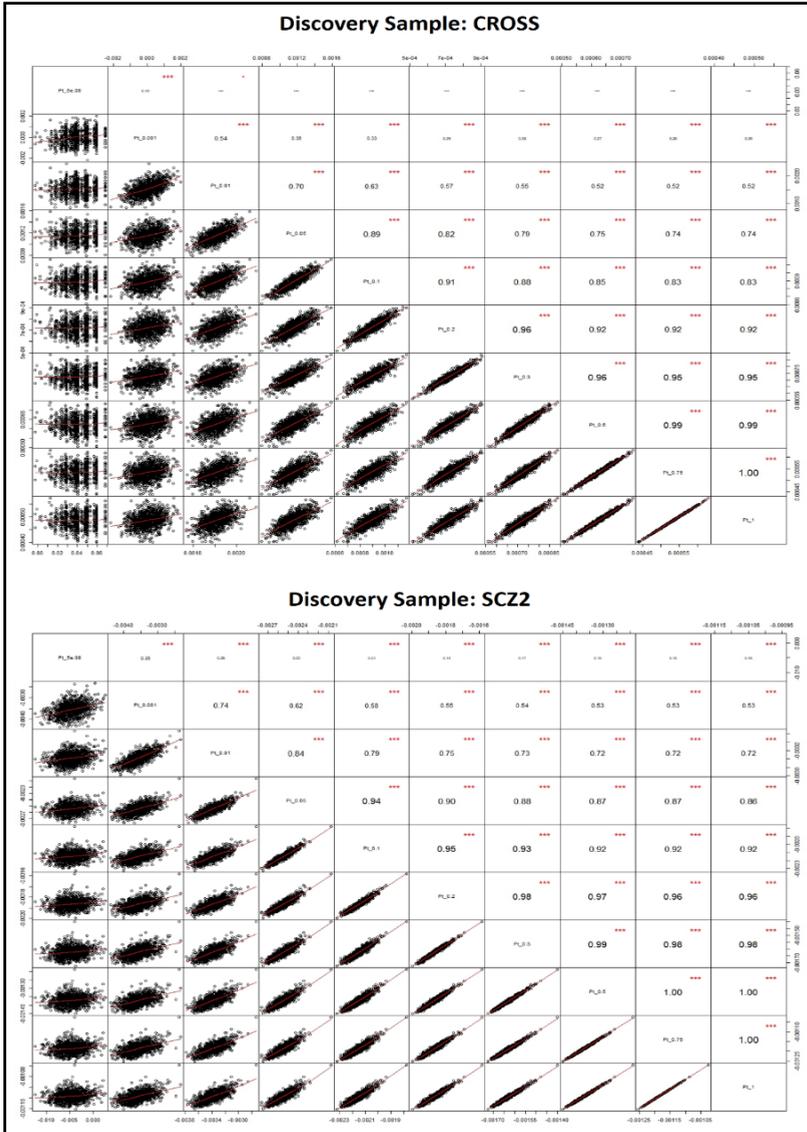


FIGURA 11. CORRELACIÓN ENTRE DISTINTAS PRS GENERADAS CON CROSS Y SCZ2. Los diferentes umbrales están indicados en la diagonal. Los diagramas de dispersión con línea de tendencia se muestran debajo de la diagonal y, los coeficientes de correlación de Pearson por pares, se muestran por encima de la diagonal. Los niveles de significación de la correlación son: * $0,05 < p < 0,01$; ** $0,01 < p < 0,001$ y *** $p < 0,001$. Gráfico creado con la función `chart.correlation` del paquete de R `PerformanceAnalytics`.

La comparación de la distribución de las PRS estandarizadas de SCZ2 con las de los casos generadas con CROSS fue significativa ($t(533)=2,506$; $p=0,01251$ con el umbral 1 y, con el umbral que explicó un mayor porcentaje de la variabilidad en cada una de las muestras descubrimiento, $t(533)=2,461$; $p=0,01393$). Las dos PRS estaban correlacionadas: correlación de Pearson $r^2=0,314$; $p<2,2\times 10^{-16}$ con el umbral 1 y, con el umbral que explicó un mayor porcentaje de la variabilidad en cada una de las muestras descubrimiento, $r^2=0,306$; $p<2,2\times 10^{-6}$.

Para comprobar si los efectos eran parcialmente independientes, se analizaron de forma conjunta incluyendo en los análisis de regresión logística las dos PRS generadas con las dos muestras descubrimiento distintas (SCZ2 y CROSS). La significación de las PRS generadas con SCZ2 pasaron de $2,1\times 10^{-10}$ a $8,4\times 10^{-8}$, el OR (intervalo de confianza del 95%) estandarizado pasó de 1,62 (1,40 – 1,89) a 1,53 (1,31-1,79) tras la inclusión de las PRS generadas con CROSS. La significación de las PRS creadas con CROSS cambio de $2,7\times 10^{-5}$ a $2,1\times 10^{-2}$ (el OR estandarizado [intervalo de confianza del 95%] cambió de 1,34 [1,17 – 1,54] a 1,18 [1,03 – 1,37]) cuando se analizaron conjuntamente en el modelo logístico con las de SCZ2. El porcentaje de variabilidad explicada con el modelo incluyendo las PRS de SCZ2 y CROSS pasó a ser 4,47%.

Como se ha desarrollado en objetivos, uno de los propósitos de este trabajo era analizar las diferencias de carga de susceptibilidad entre sexos, con el fin de replicar los resultados de Reginsson y colaboradores (Reginsson *et al.* 2018). Los autores, detectaron en población islandesa una mayor carga de susceptibilidad a abuso o dependencia de alcohol en mujeres respecto a hombres. Previamente se llevaron a cabo análisis de regresión logística por sexo utilizando SCZ2 como descubrimiento, usando como diana la totalidad de los casos (abuso/dependencia de sustancias), que revelaron una tendencia no significativa en los individuos analizados ($p=0,063$; test de una cola), empleando las PRS del umbral más explicativo en la muestra de los análisis previos. Enfocando el análisis en replicar los resultados del equipo islandés, se restringió el análisis a todas las muestras que cumpliesen el criterio de

abuso o dependencia de alcohol. Estos resultados fueron significativos ($p=0,028$, test de una cola), confirmando el efecto en la dirección estimada. Las PRS en mujeres fueron 2,7 veces mayor que en el conjunto de hombres ($\Delta R^2=9,55\%$ y $3,57\%$ respectivamente) (Figura 12). Mediante la generación de una distribución nula de las diferencias de la pseudo- R^2 explicada en mujeres vs hombres se estimó el valor de p del conjunto de muestras original empleando un umbral de significación de 0,05. Con este fin se realizaron 100000 permutaciones de la etiqueta de sexo en los dos análisis realizados (todos los casos y abuso/dependencia de alcohol) manteniendo las covariables como la edad y dimensiones de MDS. El nivel de significación para el primero estimado por las permutaciones fue $p=0,056$. En el segundo análisis fue significativo, con $p=0,032$.

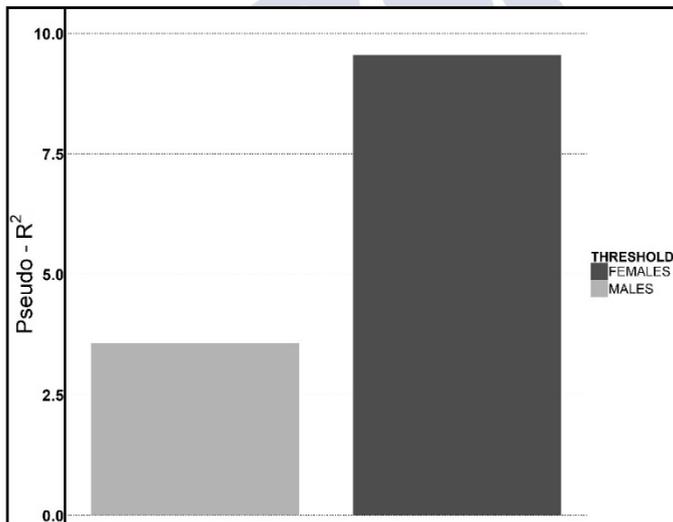


FIGURA 12. PORCENTAJE DE VARIANZA EXPLICADA EN FUNCIÓN DEL SEXO EN EL CONJUNTO DE CASOS DE ABUSO/DEPENDENCIA DE ALCOHOL. El modelo de riesgo fue generado con el umbral más significativo de los análisis previos con SCZ2.

Tras este análisis por sexo, se seleccionó un conjunto específico de muestras diana: controles y aquellos casos que no presentasen

trastornos psiquiátricos comórbidos (el 52,25% de los casos SUD). El objetivo de la generación de este subconjunto de muestras fue esclarecer si los resultados de solapamiento genético detectados eran producidos por las muestras con comorbilidad psiquiátrica. Las patologías comórbidas presentaban la siguiente distribución general: trastornos de la personalidad (21,91%), esquizofrenia (11,61%), trastornos del ánimo (10,67%) y trastornos de ansiedad (4,68%). los resultados se mantuvieron significativos siguiendo un patrón y dirección muy similar al detectado en los análisis con el conjunto total de muestras (Figura 13). El porcentaje de varianza máximo explicado fue de 4,11%, $p=2,8 \times 10^{-8}$ con el umbral 0,2. Al calcular las puntuaciones y llevar a cabo los análisis de regresión para analizar cada modelo sin excluir el xMHC se observaron resultados con patrones muy parecidos a los principales (excluyendo el xMHC) ($\Delta R^2=4,10\%$, $p=3,7 \times 10^{-10}$, con el umbral 0,2).

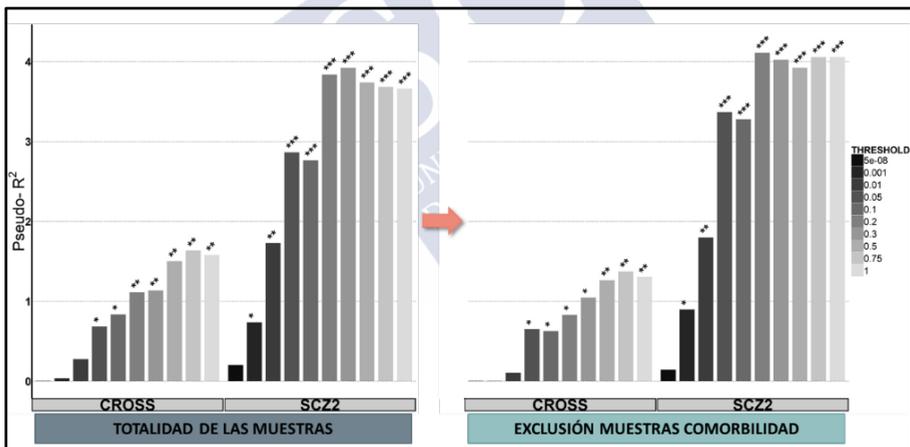


FIGURA 13. COMPARACIÓN DEL PORCENTAJE DE VARIANZA ENTRE LA MUESTRA DIANA ORIGINAL Y LA MUESTRA TRAS EXCLUIR LOS PACIENTES CON PATOLOGÍAS PSIQUIÁTRICAS COMÓRBIDAS. "TOTALIDAD DE LAS MUESTRAS" indica los resultados de la muestra diana original (Figura 9). "EXCLUSIÓN MUESTRAS COMORBILIDAD" indica los resultados de la muestra diana tras separar al 47,75% individuos de la muestra original.

Por último, se analizó el posible impacto del número de sustancias empleadas por cada consumidor. Mediante regresión lineal, empleando como variable dependiente el número de sustancias consumidas, no se encontró asociación con el umbral de significación más significativo de los análisis de SCZ2 principales: $p=0,46$.



4.3 ESTUDIO 3. ANÁLISIS DE LA CAPACIDAD PREDICTIVA DE LAS PRS EN LA MODELIZACIÓN COMBINADA CON FACTORES PREDICTIVOS DE RIESGO CONOCIDOS

Siendo el principal objetivo de este estudio el analizar la capacidad predictiva de la modelización conjunta de factores predictivos de riesgo conocidos junto a PRS, empleando modelos genéticos combinados, se presentan a continuación los resultados empleando como variable genética predictiva la suma ponderada estandarizada de las PRS generadas a partir de modelos fundamentados en ALC y SCZ2 en una muestra diana de dependencia de alcohol (Fase 2 SUD).

4.3.1 Genotipado y control de calidad

Tras el proceso de genotipado se obtuvieron datos genotípicos para 588628 variantes en 240 casos y 225 controles. Tras el control de calidad aplicado siguiendo los parámetros y filtros indicados en “Materiales y métodos” y el proceso de imputación se mantienen 6765662 variantes y 460 individuos (ver Tabla 5 para conocer en detalle el efecto de los filtros aplicados en el control de calidad), de estos, 240 son casos y 220 controles (342 hombres y 123 mujeres). En la Tabla 6 se refleja, de forma extendida, información sociodemográfica general informativa de esta muestra diana: sexo, estado civil, compañía en el hogar, situación laboral y años de estudio completados.

Una de las características principales de este conjunto de muestras de pacientes y controles es la información recogida junto a la información genética capturada. Además de las características sociodemográficas antes comentadas, se recogió información complementaria de la salud mental de los individuos (Tabla 7): historia familiar de dependencia, comorbilidad psiquiátrica, comorbilidad adictiva e intentos de suicidio. El consumo de sustancias ilegales se reduce al 10,83% del conjunto de casos con la siguiente distribución: 6,25% consumía cocaína, 2,5% cannabis, 1,25% opioides y 0,83% sedantes/hipnóticos.

TABLA 5. DETALLE DE LOS CRITERIOS DE FILTRADO EMPLEADOS EN EL CONTROL DE CALIDAD DE SUD FASE 2.

FILTRO	SNPs	CONTROLES	CASOS
Tamaño inicial	588628	225	240
Ratio de genotipación por SNP >90%	587124	225	240
Ratio de genotipación por muestra >95%	587124	225	240
Ratio de genotipación por SNP >95%	585544	225	240
Ratio de genotipación casos vs controles ($P > 1 \times 10^{-3}$)	585454	225	240
Desviaciones equilibrio Hardy-Weinberg ($P > 1 \times 10^{-3}$)	585049	225	240
Relaciones críticas (PI-HAT) (incluyendo SUD Fase 1)	585049	220	240
MAF 1%	253158	220	240

TABLA 6. INFORMACIÓN GENERAL DEL ESTATUS DEMOGRÁFICO DE LOS INDIVIDUOS RECOGIDOS EN ESTE ESTUDIO TRAS EL CONTROL DE CALIDAD APLICADO.

SEXO	CASOS N	CASOS %	CONTROLES N	CONTROLES%
Mujer	64	26,67	58	26,36
Hombre	176	73,33	162	73,64
ESTADO CIVIL	CASOS N	CASOS %	CONTROLES N	CONTROLES %
Casado/a	95	39,58	160	72,73
Soltero/a	62	25,83	27	12,27
Viudo/a	3	1,25	2	0,91
Separado/a	26	10,83	4	1,82
Divorciado/a	37	15,42	12	5,45
Pareja de hecho	16	6,67	15	6,82
NAs	1	0,42	0	0

VIVE CON	CASOS N	CASOS %	CONTROLES N	CONTROLES %
Cónyuge/Pareja	116	48,33	170	77,27
Padres	48	20,00	9	4,09
Hijo/s menores de 18 años	9	3,75	5	2,27
Otros familiares adultos	21	8,75	19	8,64
Otros adultos no familiares	3	1,25	1	0,45
Solo/a	40	16,67	16	7,27
NAs	3	1,25	0	0
SITUACIÓN LABORAL	CASOS N	CASOS %	CONTROLES N	CONTROLES %
Activo completo	81	33,75	159	72,27
Activo parcial	12	5,00	22	10,00
Incapacidad temporal	19	7,92	3	1,36
Pensionista por edad	20	8,33	11	5,00
Pensionista por enfermedad	36	15,00	6	2,73
Parado	60	25,00	9	4,09
Estudiante	0	0	3	1,36
Sus labores	8	3,33	5	2,27
Trabajo informal	0	0	2	0,91
NAs	4	1,67	0	0
AÑOS DE ESTUDIO COMPLETADOS	CASOS N	CASOS %	CONTROLES N	CONTROLES %
Sin estudios	10	4,17	1	0,45
Primarios	147	61,25	27	12,27
Secundarios	60	25,00	116	52,73
Universitarios	23	9,58	74	33,64
NAs	0	0	2	0,91

TABLA 7. INFORMACIÓN GENERAL DE LA SALUD MENTAL DE LOS INDIVIDUOS RECOGIDOS EN ESTE ESTUDIO TRAS EL CONTROL DE CALIDAD APLICADO.

HISTORIA FAMILIAR DEPENDENCIA	CASOS N°	CASOS %	CONTROLES N°	CONTROLES %
Respuesta positiva	129	53,75	14	6,36
Respuesta negativa	111	46,25	203	92,27
NAs	0	0	3	1,36
COMORBILIDAD PSIQUIÁTRICA	CASOS N°	CASOS %	CONTROLES N°	CONTROLES %
Respuesta positiva	80	33,33	0	0
Respuesta negativa	160	66,67	220	100
NAs	0	0	0	0
COMORBILIDAD ADICTIVA	CASOS N°	CASOS %	CONTROLES N°	CONTROLES %
Respuesta positiva (tabaco)	172	71,67	35	15,91
Respuesta positiva (sustancias ilegales)	26	10,83	0	0
Respuesta negativa	42	17,50	185	84,09
NAs	0	0	0	0
INTENTOS DE SUICIDIO	CASOS N°	CASOS %	CONTROLES N°	CONTROLES %
Respuesta positiva	41	17,08	0	0
Respuesta negativa	198	82,50	215	97,73
NAs	1	0,42	5	2,27

Una de las particularidades de este estudio fue el tener que ajustar los datos al empleo de dos versiones distintas del chip de genotipado Illumina Psycharray entre la Fase 1 y la Fase 2: v1.0 y v.1.1. Se analizaron las posibles diferencias entre chips, restringiendo finalmente el número de variantes a aquellos SNPs presentes en ambas muestras, eliminando las variantes que podrían generar complicaciones por ambigüedad de cadena (5753854, MAF>1%).

Al realizar los análisis de control entre Fase 1 y Fase 2 fue necesario analizar con especial detalle el posible solapamiento entre muestras y el control de posibles relaciones de parentesco. La colección general de muestras fue creada gracias al trabajo multicéntrico de un importante número de colaboradores clínicos en un proceso dividido en dos períodos, incrementando levemente la posibilidad de incluir individuos emparentados. Por esto, fue aplicado un control de calidad más exhaustivo, bajando el umbral de PI-HAT a 0,125, analizando clústers de individuos concretos y eliminando aquellas muestras repetidas entre fases. En total, el 2,06% de la muestra fue separada de los análisis por estos nuevos controles (32 individuos de los 1556 totales, ya ajustados por los controles de calidad aplicados previamente).

4.3.2 GWAS Fase 1 abuso/dependencia de alcohol

Se realizó un estudio de asociación de genoma completo de las muestras de abuso/dependencia de alcohol de la Fase 1 con el fin de generar de un modelo poligénico discriminador en la Fase 2.

El número de variantes incluidas en este análisis se redujo a 5753854 (MAF>1% y los controles extra debidos a la diferencia de chip empleado comentados en el apartado previo). Se incluyeron como elementos de ajuste las covariables: sexo, edad y las dimensiones significativas de las diez primeras del análisis de MDS. Un total de 129 SNPs presentaron un nivel de asociación alto ($p > 1 \times 10^{-5}$) (Figura 14). El valor de lambda estimado fue 1,049025 (error estándar= $5,28 \times 10^{-6}$) (Figura 15).

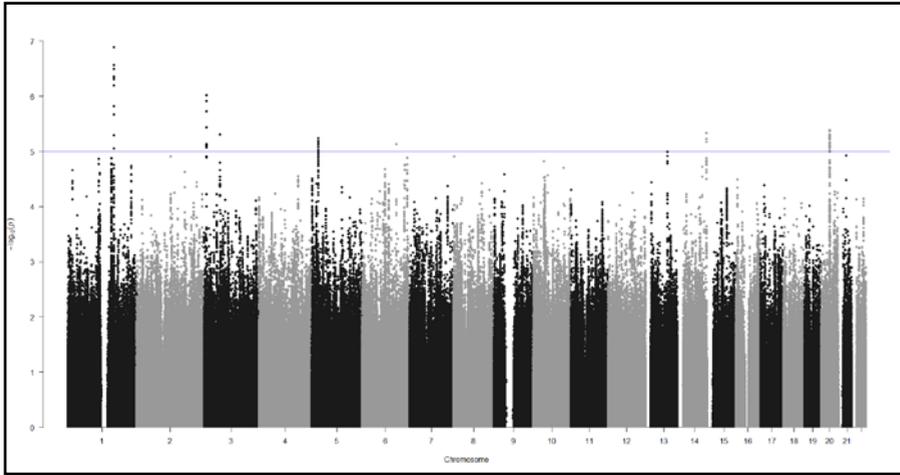


FIGURA 14. GRÁFICO MANHATTAN DEL GWAS DE LA FASE 1 (MUESTRA ALC). 5753854 SNPs fueron incluidos en el análisis. La línea horizontal azul señala el umbral $p=1 \times 10^{-5}$.

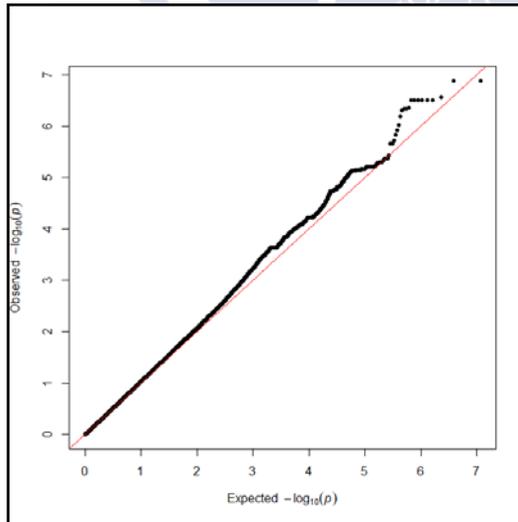


FIGURA 15. GRÁFICO CUANTIL-CUANTIL DEL GWAS DE LA FASE 1 (MUESTRA ALC). Distribución de valores de asociación observados vs esperados del GWAS representado en la Figura 14.

El segundo paso para la generación de un modelo de riesgo capaz de discriminar entre casos y controles en la muestra diana requirió el análisis del efecto y de la distribución de las PRS en SUD. Con el fin de conocer el umbral óptimo más explicativo en SUD se decidió analizar mediante validación cruzada el comportamiento de los modelos en la Fase 1.

El análisis de validación cruzada se basó en la generación aleatoria de 10 subconjuntos de muestras con el 90% de los individuos cada una para conformar distintas muestras descubrimiento, reservando un 10% de los individuos para actuar como muestra diana en cada una de las 10 pruebas.

Para la generación de las muestras descubrimiento que constituyeron los distintos modelos de riesgo fue necesario realizar 10 GWAS distintos. Estos GWAS incluyeron como elementos de ajuste las covariables: sexo, edad y las dimensiones significativas de las diez primeras del análisis de MDS. Los datos empleados para este proceso son los mismos que los resultantes del control de calidad del segundo estudio de esta tesis doctoral. Tras obtener el efecto y grado de asociación de cada variante se realizaron 200 estimaciones de riesgo poligénico (utilizando 20 umbrales por cada validación cruzada) empleando todas las muestras diana. Las PRS generadas en cada muestra diana fueron analizadas mediante regresión logística para conocer el porcentaje de varianza explicada y la validez de cada modelo, empleando el sexo, la edad y las dimensiones significativas, al igual que en el segundo proyecto desarrollado.

Los resultados de los análisis, representados en la Figura 16, mostraron limitaciones a la hora de reconocer un patrón general, ya que algunas pruebas de validación cruzada mostraron un patrón muy diferente entre sí. El porcentaje de varianza máximo explicado fue generado, en cada validación, con umbral muy distinto y una distribución de la ΔR^2 de Nagelkerke similar entre los distintos umbrales de cada validación cruzada (de 0,05 a 1). Finalmente, se eligió el umbral 0,5 para ser empleado en los siguientes análisis. La inconsistencia parcial de los perfiles de riesgo se debe, en parte, al pequeño tamaño muestral de las dianas.

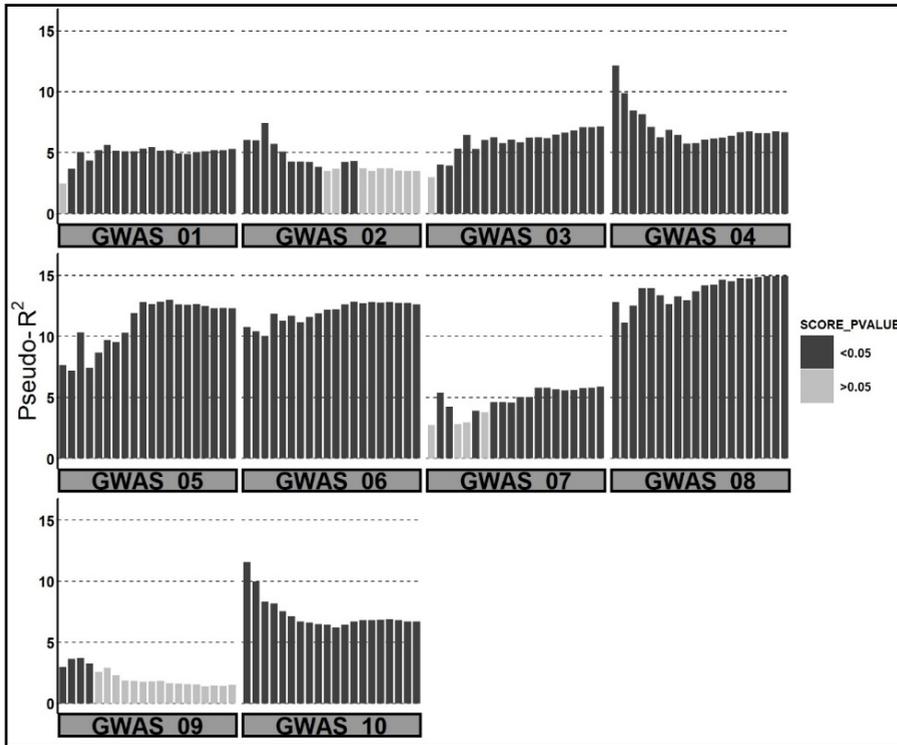


FIGURA 16. VALIDACIÓN CRUZADA EN LA FASE 1 DE SUD (MUESTRA ALC). Se llevaron a cabo 10 GWAS con los que se generaron 20 modelos de riesgo en cada caso incluyendo, cada uno, el 90% de la muestra. Cada modelo de riesgo fue enfrentado a una muestra diana correspondiente, que incluía el 10% restante de la muestra. El código de colores indica qué PRS fueron significativas en los análisis de regresión logística.

Empleando como muestra descubrimiento el GWAS de 1064 individuos de abuso/dependencia de alcohol de la Fase 1 (muestra ALC), seleccionando aquellas variantes bajo el umbral 0,5, se realizó una puntuación de riesgo en los individuos de la Fase 2 de SUD. Empleando sexo, edad y las diez primeras dimensiones de MDS como control de estratificación poblacional. Este PRS explicó el 2,3834% de la varianza en la susceptibilidad de dependencia de alcohol en la Fase 2 (Tabla 8) con el efecto en la dirección esperada.

TABLA 8. RENDIMIENTO DE LAS PRS CALCULADAS EN LA FASE 2. ALC vs Fase 2 SUD (modelo basado en la Fase 1 de SUD vs Fase 2 como muestra diana). PRS combinado (suma ponderada estandarizada de las PRS ALC vs Fase 2 y SCZ2 vs Fase 2 SUD)

TEST	ΔR^2	$\Delta AUC-ROC$	SIGNIFICACIÓN PRS
ALC vs Fase 2 SUD	2,3834	1,5142	0,00261
PRS combinado	3,6664	1,7945	0,000194

4.3.3 Combinación de modelos de riesgo

Siguiendo los resultados previos del segundo estudio de esta Tesis, en los que el modelo más explicativo se basó en la muestra descubrimiento SCZ2, se analizó la combinación de las PRS de estos modelos en combinación con las generadas con el GWAS de 1064 individuos de la Fase 1. Este método empleado se basó en la suma ponderada estandarizada comentada en “Materiales y Métodos”.

Antes de realizar la combinación de modelos de riesgo, se decidió examinar en detalle el umbral o modelo óptimo a emplear sobre la Fase 2, analizándolo sobre la Fase 1 (ALC), con la inclusión de un mayor número de umbrales en la generación de modelos (20) a partir de SCZ2, resultando 0,35 el que mayor varianza explicaba.

Finalmente, las dos PRS fueron sumadas siguiendo el método propuesto por Grove y colaboradores (Grove *et al.* 2019) ponderando las PRS de SCZ2 por el logaritmo OR de la puntuación de riesgo estandarizada en la Fase I (0,4971) (beta-coeficiente de la regresión). Este modelo combinado explicó un 3,6664% de la varianza en dependencia de alcohol en la Fase 2 (Tabla 8). La inclusión de las PRS en un modelo de regresión logística con las variables de control (sexo, edad y diez dimensiones de MDS) explicó un incremento en el área bajo la curva ROC del 1,7945%.

4.3.4 PRS vs variables independientes de riesgo conocido

Como se desarrolló en materiales y métodos, la AUC-ROC fue utilizada para evaluar la mejora de la capacidad discriminativa tras la adición de la PRS, NRI se utilizó para comprobar si se habían producido mejoras en la predicción de riesgo tras la adición de la PRS e IDI se utilizó para comprobar el aumento de las diferencias de riesgo medias previstas entre los casos y los controles tras la adición de las PRS al modelo.

La exclusión de muestras que presentasen incompleta parte de la información recogida a analizar, redujo el número de individuos incluidos en los análisis a un mínimo del 96,74% de la muestra tras el control de calidad y un máximo del 99,57% (ver Tabla 9 para conocer el número de individuos incluidos en cada análisis).

TABLA 9. NÚMERO DE MUESTRAS INCLUIDAS POR ANÁLISIS CON INFORMACIÓN COMPLETA. Nº de muestras información completa / número total tras control de calidad. CTQ EA (abuso emocional), CTQ PA (abuso físico), CTQ SA (abuso sexual), CTQ EN (negligencia emocional) y CTQ PN (negligencia física).

VARIABLE	Nº MUESTRAS
Historia Familiar	457 / 460
CTQ TOTAL	445 / 460
CTQ EA	455 / 460
CTQ PA	458 / 460
CTQ SA	457 / 460
CTQ EN	453 / 460
CTQ PN	455 / 460
LTE acumulativo	457 / 460
LTE dicotómico	457 / 460
Neuroticismo	455 / 460
Extraversión	455 / 460
Apertura	458 / 460
Amabilidad	457 / 460
Responsabilidad	455 / 460

La Tabla 10 muestra el conjunto total de resultados de los análisis. A continuación se enumeran, de forma resumida, los resultados más importantes y extremos (mínimos-máximos) y comparación de modelos (basal, con variable de riesgo [predictor conocido] o con variable de riesgo conocida y PRS).

- Todos los predictores de riesgo empleados como variables de análisis fueron significativos (valor significación mínimo: $p=0,000545$ [variable: CTQ abuso sexual]; valor significación máximo: $p<2 \times 10^{-16}$ [variable: Historia familiar y neuroticismo]) en modelos que incluían cada predictor de riesgo y PRS.

- Las PRS incluidas en los modelos de forma conjunta con las variables predictoras de riesgo fueron significativas en todos los análisis (valor significación mínimo: $p=0,02294$ [variable: Neuroticismo]; valor significación máximo: $p=9,64 \times 10^{-5}$ [variable: Historia familiar]) en modelos que incluían cada predictor de riesgo y PRS.

- Valores de AUC-ROC incluyendo predictor de riesgo conocido y PRS: mínimo 0,744 (variable: LTE dicotómico) y máximo 0,879 (variable: Neuroticismo) en modelos que incluía cada predictor de riesgo y PRS.

- Mejora AUC-ROC modelo basal (no variables predictoras de riesgo) vs modelo con variable de riesgo conocida: mínimo 0,018 (variable: LTE dicotómico) y máximo 0,17 (variable: Neuroticismo).

- Mejora AUC-ROC modelo con variable de riesgo conocida vs variable de riesgo conocida y PRS actuando de manera combinada: mínimo 0,004 (variable: CTQ negligencia emocional y neuroticismo) y máximo 0,019 (variable: LTE dicotómico).

- Ratio de mejora de AUC-ROC “variable de riesgo conocida y PRS actuando de manera combinada” vs “modelo con variable de riesgo conocida”: mínimo 2,35% (variable: Neuroticismo) y máximo 105,56% (variable: LTE dicotómico).

- Valores pseudo- R^2 en el modelo variable de riesgo conocida y PRS actuando de manera combinada: mínimo 0,226 (variable: amabilidad) máximo 0,536 (variable: Neuroticismo).
- Mejora pseudo- R^2 modelo basal vs modelo con variable de riesgo: mínimo 0,036 (variable: LTE dicotómico) y máximo 0,371 (variable: Neuroticismo).
- Mejora pseudo- R^2 modelo con variable de riesgo conocida vs variable de riesgo conocida y PRS actuando de manera combinada: mínimo 0,01 (variable: Neuroticismo) y máximo 0,038 (variable: LTE dicotómico).
- Ratio de mejora de pseudo- R^2 “variable de riesgo conocida y PRS actuando de manera combinada” vs “modelo con variable de riesgo”: mínimo 2,70% (variable: Neuroticismo) y máximo 105,56% (variable: LTE dicotómico).

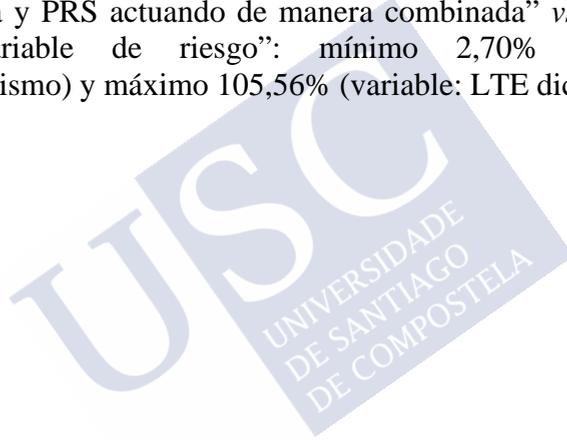


TABLA 10. RESULTADOS PRINCIPALES DE LOS ANÁLISIS DE VARIABLES PREDICTORAS COMBINADAS CON PRS. P variable (valor de p de las distintas variables predictoras de riesgo conocidas en el modelo predictivo combinado con las PRS), P PRS (valor de p de las PRS incluidas en el modelos con las variables predictoras de riesgo conocidas), AUC-ROC H2 (valores de AUC-ROC incluyendo predictor de riesgo conocido + PRS), Δ AUC-ROC H1 (mejora AUC-ROC modelo basal vs modelo con variable predictora conocida), Δ AUC-ROC H2 (mejora AUC-ROC modelo con variable predictora conocida vs variable predictora conocida + PRS), ratio AUC-ROC (H2 vs H1), psR² H2 (valores pseudo-R² en el modelo variable predictora conocida + PRS), Δ psR² H1 (mejora pseudo-R² modelo basal vs modelo con variable predictora conocida), Δ psR² H2 (mejora pseudo-R² modelo con variable predictora conocida vs variable predictora conocida + PRS) y Ratio psR² (H2 vs H1).

Variable	p variable	P PRS	AUC-ROC H2	Δ AUC-ROC H1	Δ AUC-ROC H2	Ratio AUC-ROC	psR ² H2	Δ psR ² H1	Δ psR ² H2	Ratio psR ²
Historia familiar	< 2e-16	9,64E-05	0,862	0,145	0,013	8,97%	0,48	0,294	0,031	10,54%
CTQ TOTAL	7,70E-12	0,004061	0,825	0,111	0,006	5,41%	0,375	0,198	0,018	9,09%
CTQ EA	4,66E-09	0,00101	0,786	0,069	0,013	18,84%	0,305	0,123	0,026	21,14%
CTQ PA	1,80E-05	0,000701	0,765	0,056	0,005	8,93%	0,261	0,078	0,029	37,18%
CTQ SA	0,000545	0,001021	0,747	0,032	0,012	37,50%	0,245	0,064	0,027	42,19%
CTQ EN	8,15E-12	0,00722	0,803	0,093	0,004	4,30%	0,336	0,167	0,016	9,58%
CTQ PN	7,05E-09	0,000749	0,786	0,073	0,007	9,59%	0,305	0,12	0,027	22,50%
LTE acum.	1,07E-09	0,000298	0,784	0,059	0,018	30,51%	0,308	0,12	0,031	25,83%
LTE dicot.	0,00013	0,000132	0,744	0,018	0,019	105,56%	0,231	0,036	0,038	105,56%
Neuroticismo	< 2e-16	0,02294	0,879	0,17	0,004	2,35%	0,536	0,371	0,01	2,70%
Extraversión	2,41E-10	0,000897	0,78	0,065	0,01	15,38%	0,304	0,122	0,026	21,31%
Apertura	3,71E-05	0,000108	0,744	0,023	0,017	73,91%	0,236	0,043	0,038	88,37%
Amabilidad	0,000317	0,000379	0,745	0,024	0,017	70,83%	0,226	0,039	0,032	82,05%
Responsabilidad	6,96E-12	0,00417	0,79	0,078	0,008	10,26%	0,328	0,154	0,019	12,34%

4.3.5 Evaluación de modelos de predicción con NRI e IDI

Los resultados de los análisis de NRI e IDI se muestran en la Tabla 11 y son congruentes con los obtenidos con AUC-ROC. El NRI total tras la inclusión de las PRS en el modelo fue siempre significativo. El resultado más significativo, $6,1 \times 10^{-7}$ se corresponde al modelo que incluye la historia familiar como covariable predictora de riesgo. NRI en casos y controles fue significativo en la mayoría de los análisis, excepto en el caso de “negligencia emocional” sufrida durante la infancia (CTQ EN) en casos y CTQ EN, CTQ SA, neuroticismo, extraversión y responsabilidad en controles.

Los análisis de IDI resultaron en un patrón parecido al detectado por NRI general, siendo el modelo que incluye la historia familiar como covariable predictora el más significativo ($p=1,11 \times 10^{-4}$). El único modelo que no llegó a ser significativo fue aquel que incluyó el neuroticismo ($p=0,0809$).

4.3.6 Análisis de correlación

La Tabla 12 muestra los resultados generales de los análisis de correlación de las PRS generadas y cada una de las variables de riesgo conocidas. Un gran número de variables mostraron cierta correlación significativa. Los resultados de mayor interés se resumen a continuación: en CTQ, el más significativo fue el detectado en negligencia emocional (coeficiente de correlación de Pearson=0,1592, $p=0,0007$); en el caso de LTE e historia familiar no se detectaron correlaciones significativas y, por último, en los análisis con los distintos factores de la personalidad, el resultado más significativo fue el detectado en neuroticismo (coeficiente de correlación de Pearson=0,1634, $p=0,0005$) siendo también de interés lo resultados detectados en la dirección contraria empleando extraversión (coeficiente de correlación de Pearson=-0,1175, $p=0,0122$) y responsabilidad (coeficiente de correlación de Pearson=-0,1472, $p=0,0016$).

TABLA 11. RESULTADOS DE LOS ANÁLISIS MEJORA DE RECLASIFICACIÓN NETA (NRI) Y MEJORA DE DISCRIMINACIÓN INTEGRADA (IDI).

	NRI	Valor de p NRI	Casos NRI	P casos NRI	Controles NRI	P controles NRI	IDI	Valor de p IDI
Hist. familiar	0,455	6,10E-07	2,75E-01	9,37E-06	1,80E-01	7,12E-03	2,92E-02	1,11E-04
TOTAL CTQ	0,352	0,00016	0,207	0,00128	0,146	0,0318	0,0148	0,01213
CTQ EA	0,353	0,00013	0,212	0,00087	0,142	0,03434	0,02064	0,00198
CTQ PA	0,356	1,11E-04	0,21	9,16E-04	0,145	2,92E-02	0,02108	0,00237
CTQ SA	0,308	0,00086	0,19	0,00291	0,118	0,07752	0,02064	0,00254
CTQ EN	0,224	0,0164	0,123	0,0566	0,101	0,1342	0,0111	0,04173
CTQ PN	0,336	0,00028	0,203	0,00142	0,132	0,04804	0,02101	0,00226
LTE	0,387	2,55E-05	0,227	3,26E-04	0,16	1,66E-02	0,02649	0,00034
LTE Dicot.	0,396	1,61E-05	0,227	3,26E-04	0,169	1,12E-02	0,02919	0,00023
Neuroticismo	0,238	0,0106	0,1597	0,0126	0,0783	0,247	0,0076	0,08088
Extraversión	0,335	0,0003	0,205	0,0012	0,13	0,05469	0,01967	0,0035
Apertura	0,374	4,71E-05	0,205	1,20E-03	0,169	1,12E-02	0,02873	0,00031
Amabilidad	0,359	9,47E-05	0,227	3,26E-04	0,132	4,80E-02	0,024984	0,000668
Responsabilidad	0,239	0,0103	0,1464	0,0221	0,0926	0,1717	0,01435	0,01251

TABLA 12. RESULTADOS DE LOS ANÁLISIS DE CORRELACIÓN ENTRE LA VARIABLES PREDICTORA DE RIESGO CONOCIDAS Y LAS PRS.

Variable	Cor. Pearson	P Pearson	Cor. Spearman	P Spearman
Hist. familiar	na	na	0,06810979	0,146
TOTAL CTQ	0,1409586	0,002882	0,1393137	0,00323
CTQ EA	0,09975793	0,03339	0,08400902	0,07342
CTQ PA	0,1016112	0,02969	0,04896651	0,2957
CTQ SA	0,06127747	0,191	0,08178832	0,08071
CTQ EN	0,1592121	0,0006714	0,1734879	0,0002069
CTQ PN	0,1003091	0,03242	0,1071069	0,02232
LTE Acumulativo	0,05947571	0,2044	0,01676314	0,7208
LTE Dicotómico	na	na	-0,01038611	0,8248
Neuroticismo	0,1634332	0,000465	0,1788033	1,26E-04
Extraversión	-0,1174506	0,01217	-0,141035	0,002568
Apertura	-0,01698751	0,7169	-0,01890825	0,6865
Amabilidad	-0,0694404	0,1383	-0,06374706	0,1737
Responsabilidad	-0,1472148	0,00164	-0,169437	0,0002828

4.3.7 Mediación entre PRS y trauma/factores de la personalidad

Por último, decidió analizarse si alguno de los resultados de correlación detectados podrían implicar la mediación del efecto de las PRS por la variable correlacionada, mediante el uso de modelos de regresión lineal, con la variable de riesgo conocida como variable dependiente y las PRS y las variables de control (sexo, edad y 10 dimensiones de MDS) como variables independientes. Además, en una segunda hipótesis, se analizó esta misma mediación potencial con la variable de riesgo conocida “historia familiar” como covariable de ajuste con el fin de conocer si la mediación podría estar realmente causada por un ambiente familiar en la que parte de los individuos de la

estructura familiar presentasen dependencia de alcohol. Los resultados de correlación analizados son aquellos con mayores niveles de correlación.

De las variables predictoras de riesgo conocidas analizadas (negligencia emocional sufrida en la infancia, neuroticismo, extraversión y responsabilidad), se detectan resultados significativos en todas excepto en los análisis con “extraversión” como variable dependiente (mínimo $p=0,0553$). En los otros modelos, la posible mediación fue detectada, manteniéndose significativa tras el ajuste por “historia familiar” (mínimo $p=0,00469$, “neuroticismo” como variable dependiente). Los resultados se muestran en la Tabla 13.

TABLA 13. RESULTADOS DE LOS ANÁLISIS DE MEDIACIÓN. Los resultados precedidos por “H1” responden a la inclusión de las PRS en el modelo de evaluación estadística y “H2” PRS junto a las covariables predictoras conocidas.

Análisis	Valor de p PRS
H1 CTQ_EN	0,0062
H2 CTQ_EN	0,0084
H1 Neuroticismo	0,0047
H2 Neuroticismo	0,0087
H1 Extraversión	0,0553
H2 Extraversión	0,0831
H1 Responsabilidad	0,0066
H2 Responsabilidad	0,0104

4.4 ESTUDIO 4. ANÁLISIS DE LA CAPACIDAD PREDICTIVA DE MODELOS POLIGÉNICOS FUNDAMENTADOS EN LA SELECCIÓN DE VARIANTES EN BASE A CRITERIOS BIOLÓGICOS FUNCIONALES

Siendo el principal objetivo de este estudio analizar la capacidad predictiva de modelos poligénicos fundamentados en la selección de variantes en base a criterios biológicos funcionales, con el fin de explicar la relación entre el estrés y la dependencia al alcohol, se presentan a continuación los resultados de la predicción poligénica basada en modelos generados a partir de MDD2 en una muestra conjunta de la Fase 1 y 2 de SUD.

4.4.1 Preparación del archivo diana

En este estudio se empleó como muestra diana la suma de la Fase 1 de SUD y la Fase 2 de SUD. Las particularidades de trabajar de forma conjunta con estas dos muestras fueron desarrolladas en el Estudio 3, así como los resultados de la aplicación del control de calidad. Por lo tanto, se emplearon los dos conjuntos de muestras tras haber sido aplicado el control de calidad diseñado para análisis conjuntos. La Fase 1 incluyó los 1064 individuos (muestra ALC) y la Fase 2 los 460 individuos con pacientes con dependencia de alcohol.

Tras unir con PLINK v1.9 ambas muestras, se mantuvieron un total de 1524 individuos (721 casos and 803 controles; 1047 hombres y 477 mujeres) y 5499467 SNPs. Esta muestra diana se denomina SUD_1524.

4.4.2 Modelos de riesgo

En primer lugar fue necesario realizar la puesta a punto de un modelo de riesgo basado en la selección de variantes en la muestra descubrimiento MDD2 a partir de datos de respuesta a estrés. A continuación se presentan los distintos pasos de la generación del modelo de riesgo.

La tabla S1 del material suplementario del artículo publicado por Arloth y colaboradores (Arloth *et al.* 2015) incluye el listado de los 320 *cis*-eQTLs de respuesta transcripcional a estrés, con sus *cis*-eSNPs asociados. Tras seleccionar el conjunto único de 296 *cis*-eSNPs, fue necesario realizar un cambio de versión de ensamblaje del genoma siguiendo los procedimientos indicados en “Materiales y Métodos”, ya que el conjunto de datos de variantes de respuesta a estrés se encontraba en versión hg18. Tras el cambio de versión al deseado (hg19) se realizaron controles de duplicación de variantes y se seleccionaron las de ensamblaje primario. Este listado de SNPs fue empleado para la selección de las variantes en las estadísticas resumen de MDD2.

El siguiente paso fue analizar el solapamiento de las variantes filtradas de MDD2 con los 296 *cis*-eSNPs. El solapamiento fue del 66,89%, al detectarse 198 variantes comunes (Tabla S1 ANEXO). Tras ello fue necesario utilizar como estrategia la búsqueda de variantes en LD, empleando la herramienta bioinformática de búsqueda online SNIIPA, para aquellos 98 *cis*-eSNPs no solapantes. Los parámetros seleccionados para la exploración de variantes en LD se resumen a continuación: versión de ensamblaje del genoma, hg19; datos para el cálculo de LD, 1000 Genomas fase 3 v5; variantes, SNPs; y, por último, un umbral de $r^2 > 0,5$. Los resultados incluyen información para 96 de las 98 variantes comprendidas en la búsqueda, ya que 2 SNPs no estaban incluidos en la base de datos de SNIIPA. El número de variantes en desequilibrio de ligamiento ascendió a 3828 SNPs. La media de SNPs en LD por variante original fue de 39,98 (número mínimo de variantes en LD por SNP original: 1; número máximo de variantes en LD por SNP original: 680). El número de variantes en común entre MDD2 filtrado y las 3838 variantes en LD se redujo a un total de 2464 SNPs, informativos de 58 SNPs originales.

La media de SNPs en LD por variante original fue de 42,48 (número mínimo de variantes en LD por SNP original=1; número máximo de variantes en LD por SNP original=535). Con el fin de seleccionar aquellas variantes de mayor valor para el análisis, se seleccionó de cada una de las 58 variantes originales la variante en mayor LD (mayor valor de r^2). Por último, fue necesario analizar

aquellas variantes con alelos de riesgo por ambigüedad de cadena, examinando las frecuencias de los alelos, desechando aquellas variantes ambiguas y seleccionando la siguiente por LD. El conjunto final de 58 SNPs seleccionados presentó una r^2 media respecto a la variantes original de 0,8243 (r^2 mínimo=0,5225; r^2 máximo=1) (Tabla S2 ANEXO). Por lo tanto, el número final de variantes a incluir en el modelo basado en MDD2 y el listado de *cis*-eSNPs, fue de 256 SNPs, cubriendo el 86,49% del conjunto original (Figura 16).

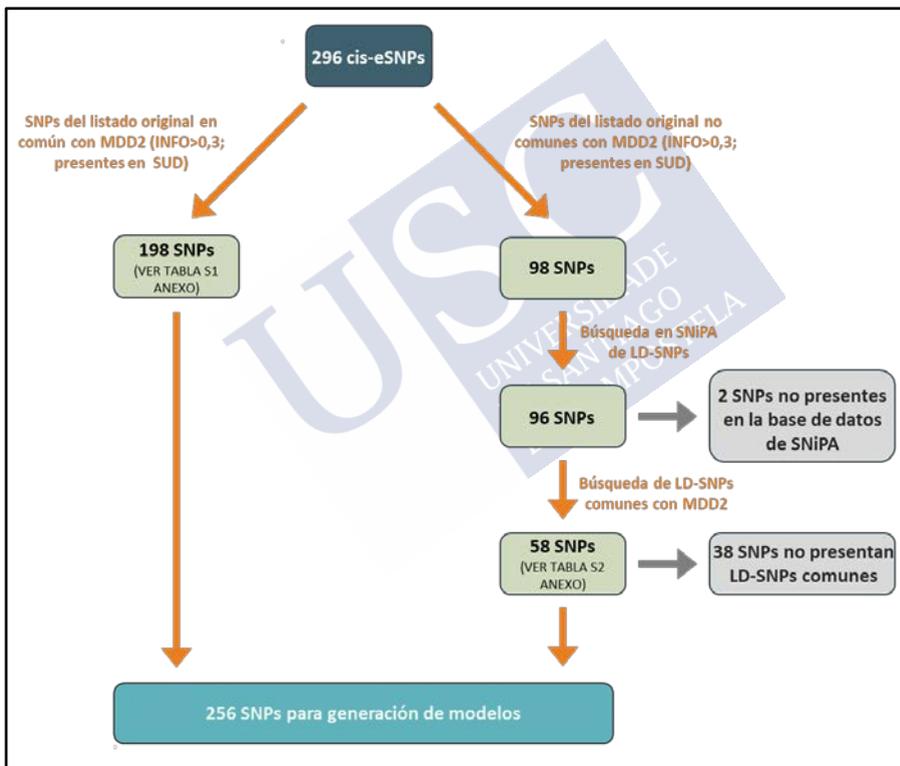


FIGURA 16. DIAGRAMA DE FLUJO DEL PROCESO DE BÚSQUEDA DE SNPs COMUNES A PARTIR DE LA LISTA ORIGINAL DE *cis*-eSNPs ASOCIADOS A LA RESPUESTA A ESTRÉS AGUDO.

4.4.3 Análisis predictivos

Los análisis de regresión logística (Figura 17) incluyendo las PRS generadas en SUD_1524 mostraron un único umbral significativo: incremento pseudo- $r^2=0,2948$; $p=0,0422$. El umbral que explicó el máximo porcentaje de la varianza fue aquel que incluyó los 38 SNPs con un nivel de significación menor a 0,1. El resto de umbrales analizados no superaron el nivel de significación $<0,05$.

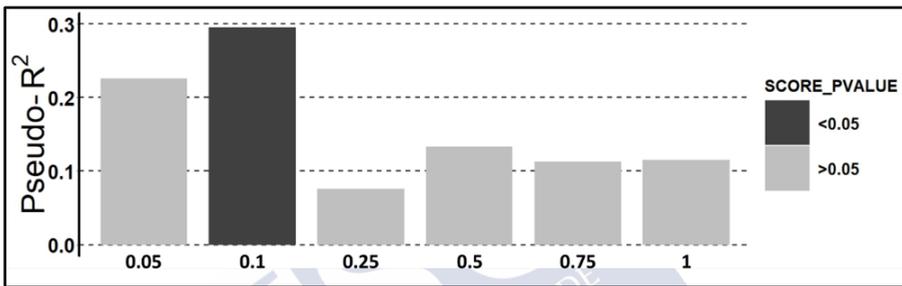


FIGURA 17. PORCENTAJE DE VARIANZA EXPLICADA POR LOS DISTINTOS MODELOS DE RIESGO GENERADOS. El código de colores indica qué modelos empleados fueron significativos.

Como control de los resultados se empleó un sistema de permutaciones basado en la herramienta bioinformática online SNPsnap. Ajustando los parámetros de búsqueda a SNPs con una frecuencia de alelo menor, densidad génica, LD y distancia al sitio de inicio de transcripción del gen más cercano, se generaron 1000 modelos de riesgo poligénico basados en SNPs de características similares o parecidas a los 38 incluidos en el modelo de riesgo significativo. Los resultados del test de permutaciones proyectaron un valor $p=0,039$ (39 de las 1000 permutaciones explicaron una pseudo- R^2 mayor que el valor presentado por el modelo significativo).

Por último, ninguno de los resultados de las interacciones entre PRS basadas en las variantes de respuesta a estrés y los datos de trauma

infantil trauma infantil muestran una interacción significativa entre las PRS y su CTQ (mínimo $p=0,2279$, umbral 0,1).



DISCUSIÓN





5. DISCUSIÓN

5.1. DISCUSIONES INDIVIDUALES

5.1.1 Estudio 1. Análisis de la arquitectura genética compartida entre el trastorno obsesivo-compulsivo y la esquizofrenia

En este primer estudio, se analizó la potencial susceptibilidad compartida entre la esquizofrenia y el trastorno obsesivo-compulsivo. Los resultados muestran, desde una aproximación poligénica, la primera evidencia de arquitectura genética compartida entre ambos trastornos.

En contraste con la mayoría de trastornos psiquiátricos comunes, donde los avances en el estudio de su base genética se basaron en un gran número de GWAS iniciados en el año 2007, en el caso del trastorno obsesivo-compulsivo el número de estudios es limitado. Las primeras indagaciones en la arquitectura poligénica de este trastorno se basaron en las detecciones de riesgo compartido con el síndrome de Tourette (Davis *et al.* 2013; Yu *et al.* 2015). Por lo tanto, en este contexto, los hallazgos de heredabilidad compartida con esquizofrenia detectados en el presente estudio constituyen uno de los primeros pasos importantes dados en el campo de la genética para el entendimiento de la estructura genética de la enfermedad.

El uso de la estadística resumen de SCZ2 (Schizophrenia Working Group of the PGC, 2014) permite la generación de modelos de riesgo robustos que pueden detectar la arquitectura genética compartida. El porcentaje de varianza explicada en estos resultados (3,7% tras excluir las variantes comprendidas en el xMHC) supera al detectado por Yu y colaboradores, en el que empleando la misma estrategia de análisis poligénico, llegaron a explicar un máximo del 3,2% de la varianza en la escala fenotípica (trastorno obsesivo-compulsivo *vs* trastorno obsesivo-compulsivo). En su análisis transversal, analizando la heredabilidad compartida entre el trastorno obsesivo-compulsivo y el síndrome de Tourette, explicaron un 0,6% de la varianza pero con

resultados no significativos (Yu *et al.* 2015). Otra relación de interés, también detectada mediante PRS, es la existente con trastornos del espectro autista. Esta relación es débil pero significativa, ya que solamente explican el 0,11% de la varianza fenotípica en los análisis transversales (Guo *et al.* 2017). La heredabilidad compartida entre anorexia nerviosa y el trastorno obsesivo-compulsivo también ha sido analizada recientemente confirmándose las expectativas generadas por estudios epidemiológicos previos, así como lográndose la identificación de enriquecimiento de patrones de expresión génica celulares o de tejidos y la detección de la correlación genética del fenotipo combinado “anorexia-nerviosa y trastorno obsesivo-compulsivo” con otros trastornos psiquiátricos (Yilmaz *et al.* 2018).

El presente trabajo conforma una base para futuros análisis en el propósito de dilucidar la coocurrencia de estos trastornos. El obtener un resultado en la misma dirección empleando una muestra independiente (SCZ-5100), incluso con menor potencia estadística, y el realizar controles para múltiples test, añaden robustez a los resultados. Estos también señalan que el empleo de variantes genéticas comunes (MAF > 5%) basadas en un chip de genotipado exónico permite analizar la arquitectura poligénica del trastorno obsesivo-compulsivo, siendo necesario plantear futuros análisis de interés empleando variantes que informen de toda la extensión del genoma, sin la restricción de ceñirse a regiones codificantes.

El hecho de analizar el solapamiento genético entre esquizofrenia y el trastorno obsesivo-compulsivo sin incluir variantes de la región del xMHC permitió explicar una mayor varianza fenotípica en el trastorno obsesivo-compulsivo. A pesar de la hipótesis autoinmune y la importancia del MHC en los resultados del GWAS de SCZ2 (Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014), donde el locus más asociado se encontraba en la región codificante inmunitaria, los análisis apuntan con claridad a que la inmunidad no tiene un papel de relevancia en la susceptibilidad compartida. Teniendo en cuenta la hipótesis del papel de la autoinmunidad en un subgrupo de pacientes con trastorno obsesivo-compulsivo, conocido como trastornos neuropsiquiátricos pediátricos

autoinmunes (Davison 2012), y el papel de la inmunidad en la esquizofrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014), los resultados parecen indicar que aunque la inmunidad puede tener un papel relevante en ambos trastornos, esta no es debida a las variantes compartidas en la región cromosómica donde se localiza el MHC.

La relación entre el trastorno obsesivo-compulsivo, o sus síntomas, y la esquizofrenia es un elemento bien conocido en la psiquiatría moderna (Poyurovsky *et al.* 2012) y ha sido analizada durante las últimas décadas. Existe un importante debate sobre la validez de considerar un subtipo de pacientes esquizo-obsesivos en el espectro de la esquizofrenia (Zohar 1997, Poyurovsky *et al.* 2012; Scotti-Muzzi & Saide, 2016). A pesar de que el conjunto esquizo-obsesivo no ha sido incluido en la quinta versión del DSM, la posible existencia de “obsesiones sin percepción de las mismas” en el espectro de la esquizofrenia crea de alguna forma un continuo parcial entre delirios y obsesiones (Catapano *et al.* 2010). En este debate, uno de los puntos de mayor peso se centra en el origen de estas obsesiones en pacientes esquizofrénicos. Los antipsicóticos atípicos han sido descritos como un factor de riesgo del desarrollo de síntomas obsesivo-compulsivos en pacientes esquizofrénicos susceptibles (Lykouras *et al.* 2003). Este factor puede haber influido en el aumento de la prevalencia de estos síntomas en pacientes del espectro de la esquizofrenia (Scotti-Muzzi & Saide 2016). El tratamiento con clozapina fue el primer señalado como inductor de síntomas obsesivos (Baker *et al.* 1992) y con el paso de los años se mantiene como la medicación más asociada al desarrollo de trastorno obsesivo-compulsivo junto a la olanzapina (Poyurovski & Koran 2005; Schirmbeck *et al.* 2011). Además, la duración del tratamiento con clozapina es un factor de riesgo replicado (Fernández-Egea *et al.* 2018), por lo que un seguimiento específico de estos pacientes sería de gran interés. Estos efectos y características deben ser tenidos en cuenta a la hora de aplicar tratamientos, pues la aparición de síntomas podría limitar su uso en pacientes pertenecientes a esa posible entidad nosológica esquizo-obsesiva (Scotti-Muzzi & Saide 2016).

Los pacientes que podrían englobarse dentro de este posible subgrupo esquizo-obsesivo presentan características clínicas distintivas, tales como una edad de inicio de la enfermedad más temprana, déficits cognitivos específicos, un menor grado de percepción de las obsesiones, una severidad global mayor de los síntomas positivos y negativos, un mayor número de suicidios y los síntomas obsesivos-compulsivos son normalmente severos, tendiendo a aparecer durante las etapas prodrómicas de la esquizofrenia (Niendam *et al.* 2009; Poyurovsky *et al.* 2012; Scotti-Muzzi & Saide 2016). De todas formas es necesario considerar la heterogeneidad y limitaciones de los estudios que han ido permitiendo la caracterización de este posible subgrupo de pacientes (Scotti-Muzzi & Saide 2016). Los últimos avances en epidemiología del subgrupo de pacientes y los primeros pasos en la caracterización de la neurobiología de estos individuos se suman a los esfuerzos en genética, donde los resultados de esta tesis aportan las posibles bases a considerar para la generación y uso de modelos de riesgo para la estratificación de pacientes en un subgrupo esquizo-obsesivo. La consideración de un subgrupo de pacientes tendría implicaciones en el pronóstico y tratamiento, ya que las medicaciones para cada trastorno o síntoma podrían afectar a los síntomas del otro (Poyurovsky *et al.* 2012). Futuras mejoras de la estimación del riesgo compartido entre la esquizofrenia y el trastorno obsesivo-compulsivo serán de gran interés. La detección de subgrupos de pacientes permitirá un ajuste de los tratamientos con la consiguiente mejora reflejada en el pronóstico de estos conjuntos de individuos.

Aunque el riesgo de que un paciente con trastorno obsesivo-compulsivo desarrolle síntomas psicóticos es notablemente bajo (1,7%) (Haan *et al.* 2009), al considerar los distintos trastornos del espectro esquizofrenia – trastorno obsesivo-compulsivo, Poyurovsky y Koran (Poyurovsky & Koran 2005) analizaron estos subgrupos dentro del espectro: trastorno obsesivo-compulsivo puro, trastorno obsesivo-compulsivo con baja percepción de la enfermedad, trastorno obsesivo-compulsivo con trastorno esquizotípico de la personalidad, esquizofrenia con síntomas compulsivos, esquizofrenia con trastorno obsesivo-compulsivo y esquizofrenia pura. Los autores encontraron asociaciones heterogéneas con el trastorno obsesivo-compulsivo en el

conjunto de pacientes que presentaban trastorno esquizotípico de la personalidad. El estudio epidemiológico longitudinal empleando registros nacionales de Dinamarca, llevado a cabo por Meier y colaboradores, identificó un aumento del riesgo a desarrollar esquizofrenia en individuos que previamente fueron diagnosticados con trastorno obsesivo-compulsivo (ratio de incidencia=6,90, 95% I.C.=6.25–7.60) (Meier *et al.* 2014). De los 3 millones de individuos seguidos durante 17 años, 30556 desarrollaron esquizofrenia, de los cuales el 2,29% había sido diagnosticado con trastorno obsesivo-compulsivo previamente. Además, este riesgo era aproximadamente dos veces mayor que el de otros trastornos infanto-juveniles como autismo, bulimia nerviosa o trastorno por déficit de atención e hiperactividad. Otro de los resultados interesantes de este estudio longitudinal fue el detectar que un diagnóstico parental de trastorno obsesivo-compulsivo incrementaba la ratio de incidencia de esquizofrenia en la descendencia al 4,31 (95% I.C.=2,72–6,43) y el riesgo asociado al diagnóstico parental de trastorno obsesivo-compulsivo fue mayor que el asociado al diagnóstico parental de cualquier otro trastorno psiquiátrico al margen de esquizofrenia o trastornos del espectro de la esquizofrenia. Fontenelle y colaboradores (Fontenelle *et al.* 2011), al analizar aquellos individuos considerados en riesgo a desarrollar psicosis, detectaron que el 12,1% mostraba síntomas obsesivos-compulsivos y, el 5,1% de los individuos, trastorno obsesivo-compulsivo. Por último, en el estudio llevado a cabo por Zink y colaboradores, al analizar aquellos individuos considerados en riesgo a desarrollar psicosis, el 11% mostraba síntomas obsesivos-compulsivos y el 5% trastorno obsesivo-compulsivo (Zink *et al.* 2014).

Las estimaciones de prevalencia de trastorno obsesivo compulsivo en pacientes de esquizofrenia también son bastante heterogéneas. Las primeras indagaciones (Berrios 1989) estimaron una proporción del 1-3,5%. Estas estimaciones generales han cambiado con el paso de años. Los estudios más recientes han documentado una prevalencia mucho mayor. Un meta-análisis de los últimos años que incluía 3007 muestras de 34 estudios distintos estimó una prevalencia del trastorno obsesivo-compulsivo en pacientes con trastornos del espectro esquizofrénico del 12,1% (95% I.C.=7.0–17.1%) (Achim *et al.* 2011). Estas estimaciones

también han sido empleadas para analizar la presencia de síntomas obsesivo-compulsivos en pacientes. La heterogeneidad en los resultados también está presente en estas medidas, tal y como exponen Scotti-Muzzi y Saide (Scotti-Muzzi & Saidade 2016). Una de las estimaciones más fiables es resumida por Schirmbeck y Zink (Schirmbeck & Zink 2013), indicando que la comorbilidad de síntomas de este espectro en pacientes esquizofrénicos es del 25%.

La principal limitación de este estudio es el tamaño de la muestra de individuos control y pacientes con trastorno obsesivo-compulsivo. Es de gran interés conocer estos efectos de la arquitectura genética compartida empleando futuras muestras de mayor tamaño, con el fin de replicar los resultados detectados por el análisis y mejorar las capacidades predictivas. La consideración de subgrupos de pacientes con rasgos más específicos para analizar como conjunto más homogéneo es una de las estrategias a seguir. Otra limitación de este estudio se basa en el tipo de chip de genotipado empleado, ya que las variantes analizadas eran todas de regiones exónicas. Con el abaratamiento de los costes y desde la perspectiva actual, considerar un análisis de marcadores distribuidos por todo el genoma aportaría una perspectiva de análisis más amplia, con un mayor número de variantes en común entre muestra diana y descubrimiento que permitiría la generación de modelos más ajustados.

5.1.2 Estudio 2. Análisis de la arquitectura genética compartida entre el trastorno por abuso o dependencia de sustancias y los principales trastornos psiquiátricos

Los resultados de este estudio, en el que se emplearon seis trastornos psiquiátricos principales distintos como muestras descubrimiento (esquizofrenia, trastorno bipolar, trastorno por déficit de atención e hiperactividad, trastorno depresivo mayor, trastorno de ansiedad generalizada y autismo), revelaron que la esquizofrenia es la que comparte un mayor riesgo genético con los trastornos por uso o abuso de sustancias. Estos resultados se suman a los primeros avances llevados a cabo en el estudio del solapamiento y estructura genética compartida entre SUD y patologías psiquiátricas generales, que señalaban, mediante modelos de riesgo, la susceptibilidad compartida (Power *et al.* 2014; Carey *et al.* 2016; Andersen *et al.* 2017; Hartz *et al.* 2017; Reginsson *et al.* 2018).

La esquizofrenia es el trastorno que muestra un mayor solapamiento con los trastornos por uso o abuso de sustancias y, además, es el único con resultados significativos tras aplicar las correcciones por comparaciones múltiples. Respecto a los resultados no significativos, obtenidos en los análisis empleando las otras muestras descubrimiento, es necesario reflexionar sobre el sentido de los mismos. Aunque las señales significativas están asociadas a los modelos basados en el riesgo a desarrollar esquizofrenia, las otras patologías también podrían presentar un riesgo compartido no detectado en nuestros análisis. El primer punto que señala en esta dirección se basa en que el porcentaje de varianza explicada en la escala de susceptibilidad fenotípica de SUD, con los modelos generados a partir de la muestra descubrimiento de trastornos psiquiátricos combinados Cross-Disorder, es mayor que la explicada por SCZ-CROSS, muestra incluida en el conjunto Cross-Disorder. Esto indica que en el conjunto de trastornos (muestra Cross-Disorder) existen otros, además de esquizofrenia, con un posible efecto predictor o arquitectura solapante. El segundo punto se basa en que las PRS generadas con Cross-Disorder se mantuvieron significativas tras el ajuste con los modelos analizados basados en SCZ2.

Una de las intenciones de este estudio fue el de minimizar las diferencias en tamaño entre las distintas muestras descubrimiento, empleando nuevas estadísticas resumen de acceso público. Empleando muestras de mayor tamaño (ASD-2017 y ADHD-2017) que las utilizadas por Carey y colaboradores (Carey *et al.* 2016) en sus análisis de solapamiento con trastornos del espectro autista y trastorno por déficit de atención e hiperactividad, con resultados no significativos, tampoco obtuvimos resultados significativos. En este estudio también se ha estudiado por primera vez la posible susceptibilidad compartida entre trastorno de ansiedad generalizada y los trastornos por uso y abuso de sustancias, no logrando detectar una asociación significativa. Por último, no confirmamos las asociaciones entre trastorno depresivo mayor y SUD detectadas por Carey y colaboradores y Andersen y colaboradores (Carey *et al.* 2016; Andersen *et al.* 2017). Esto podría ser por el bajo porcentaje de varianza explicada por los autores que sí detectaron estas asociaciones. De todas formas, estas hipótesis deberían ser cuestionadas de nuevo en el futuro con la aparición de muestras más amplias como MDD2 (Wray *et al.* 2018), posterior a la realización y publicación del presente estudio (Gurriarán *et al.* 2019).

Uno de los resultados de gran interés de este estudio es el porcentaje de varianza explicada por los modelos de SCZ-CROSS, ya que llegaron a explicar un porcentaje de la varianza en la escala fenotípica de SUD mayor a cualquiera detectado empleando la misma muestra descubrimiento contra otras patologías psiquiátricas, con la excepción de trastorno bipolar (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013). El GWAS al que se recurre en los análisis de SCZ-CROSS se basa en un conjunto de muestras que a su vez está incluido en SCZ2. Este conjunto de muestras representa solamente una fracción de SCZ2, ya que el conjunto de casos de SCZ-CROSS representa aproximadamente el 26% de los pacientes incluidos en SCZ2. Este incremento muestral se traduce un mayor poder predictivo de la muestra SCZ2, que llega a explicar un 3,92% de la varianza en nuestra muestra diana frente al 1,35% explicada por SCZ-CROSS. Estos resultados concuerdan con los obtenidos por Hartz y colaboradores, que discutieron sobre el porcentaje de varianza explicada en sus resultados, concluyendo que el solapamiento genético

entre SCZ2 y su muestra de individuos con adicciones era “inusualmente grande”. Este estudio llegó a explicar un 3,7% de la varianza en la escala fenotípica de SUD (Hartz *et al.* 2017). Un resultado en la misma dirección también fue observado en el presente estudio incluido en la Tesis, cuando las muestras con patologías psiquiátricas comórbidas de la muestra diana fueron separadas de los análisis. Este hecho sugiere en gran medida que la esquizofrenia y los trastornos por uso y abuso de sustancias presentan una alta pleiotropía.

Este posible efecto pleiotrópico puede ser de dos clases: biológico o mediado. En el caso de la pleiotropía biológica el modelo se basaría en que las mismas variantes genéticas incrementarían el riesgo a sufrir esquizofrenia y SUD de una forma directa a través del mismo mecanismo biológico. En el caso de pleiotropía mediada, la esquizofrenia incrementarían el riesgo de SUD o viceversa (Solovieff *et al.* 2013). En esta situación, las PRS se relacionan causalmente con un trastorno e indirectamente con el segundo trastorno, mediado por la presencia del primero. El hecho de no poder diferenciar entre tipos de pleiotropía es una limitación del presente estudio y del método empleado, ya que los modelos de riesgo poligénicos no informan de los mecanismos biológicos que podrían encontrarse tras el efecto detectado. A pesar de haber realizado análisis eliminando las muestras comórbidas detectándose efectos de igual magnitud, no se puede descartar la mediación de forma completa, debido a que algunos pacientes presentarían síntomas subclínicos del trastorno. Futuros estudios permitirán identificar los mecanismos neurobiológicos y rutas genéticas subyacentes a este solapamiento genético detectado. Éstos podrían estar basados en aproximaciones como la aleatorización mendeliana o estudios longitudinales. Los estudios de aleatorización mendeliana ofrecen la oportunidad de análisis de la asociación entre SUD y patologías psiquiátricas generales. Existen varios ejemplos en la literatura, como los empleados para investigar la causalidad en la asociación entre cannabis y esquizofrenia (Vaucher *et al.* 2017; Pasman *et al.* 2018).

Una de las características principales de este estudio es que los individuos que constituyen la muestra de pacientes son

policonsumidores. La susceptibilidad compartida entre distintas drogas podría estar basada en mecanismos neurobiológicos centrales y las redes cerebrales, como aquellos que modulan la recompensa, excitación afectiva, control del comportamiento, ansiedad o la respuesta a estrés. Alguno de estos procesos pueden ser comunes a los distintos trastornos psiquiátricos (Buckholtz & Meyer-Lindenberg 2012), describiéndose, mediante estudios de gemelos, una posible genética común (Tsuang *et al.* 1998). Resultados recientes de un mega-análisis del volumen de materia gris en individuos con trastornos de dependencia mostraron evidencias en el mismo sentido: un conjunto de regiones cerebrales presentaron asociación a la dependencia de sustancias en general, indicando un sustrato neural común, y otras específicas al consumo de alcohol y nicotina, logrando clasificar con éxito individuos con esas dependencias y biomarcadores específicos (Mackey *et al.* 2018). En contraste, la susceptibilidad genética específica para las adicciones a una sustancia concreta podría estar más relacionada con el metabolismo de las sustancias. Uno de los ejemplos más claros se basa en la asociación de *ADH1B* y *ALDH2* (alcohol deshidrogenasa y aldehído deshidrogenasa) con la dependencia de alcohol (Edenberg *et al.* 2019). Además, otros indicios se basan en la asociación a la dependencia de alcohol de genes involucrados en el metabolismo de los cuerpos cetónicos, ya que la cetoacidosis es causada por un excesivo consumo de alcohol (Biernacka *et al.* 2013).

Por último, con este estudio replicamos los resultados de Reginsson y colaboradores (Reginsson *et al.* 2018) en los que detectaron diferencias entre sexos en pacientes con trastorno por abuso del alcohol. Sus resultados mostraban que las mujeres presentaban una mayor predisposición o carga genética basada en los PRS fundamentados en esquizofrenia, coincidiendo con resultados previos de estudios familiares en la misma población. Los resultados del presente estudio de la Tesis, en la misma dirección, podrían reflejar que existe una mayor presión social en contra del consumo de alcohol hacia las mujeres respecto a los hombres. Las mujeres, a igual puntuación de riesgo que los hombres, deberían presentar una menor probabilidad para desarrollar el trastorno por abuso o dependencia de alcohol, ya que en nuestra sociedad actual el ambiente es más protector de cara a las

mujeres. Una situación similar también fue detectada por el estudio islandés, ya que describen una clara tendencia al incremento del riesgo genético en fumadores con el paso de los años y el endurecimiento sobre su consumo y una mayor presión social general actual.

Además de la limitación, previamente comentada, basada en la incapacidad a la hora de diferenciar entre el tipo de pleiotropía asociada a los PRS utilizados, este estudio presenta otras limitaciones. Las muestras descubrimiento empleadas podrían presentar trastorno por abuso o dependencia de sustancias, complicando en parte la interpretación de los resultados. El tamaño de la muestra diana, aunque de una magnitud suficiente para detectar el efecto con una significación estadística muy robusta, podría ser objeto de mejora en siguientes aproximaciones de análisis. Otro elemento que añade cierta incertidumbre se basa en el apareamiento clasificado. Este hecho es bastante común entre los trastornos psiquiátricos y SUD (Nordsletten *et al.* 2016). En este trabajo no encontramos asociaciones significativas entre el trastorno por déficit de atención e hiperactividad y SUD a pesar de las mayores correlaciones presentes entre ADHD-2017 y SUD que esquizofrenia y SUD.

La aproximación genómica a la coocurrencia de SUD y trastornos psiquiátricos generales presenta un corto recorrido temporal. Con los primeros resultados significativos detectados y el incremento de iniciativas de peso, el número de resultados que apuntan en la misma dirección que los presentados en esta tesis también crece. Recientemente han sido publicados varios estudios de interés. Se han llevado a cabo predicciones poligénicas en muestras longitudinales en las que se ha detectado que los participantes con un PRS mayor, basado en esquizofrenia, presentaban un mayor uso de sustancias en la etapa adulta emergente (18-25 años) (Travis *et al.* 2018). Aas y colaboradores (Aas *et al.* 2018), aportaron más evidencias del posible solapamiento genético entre esquizofrenia y consumo de cannabis. Dos grandes estudios de 2018 también han aportado información de gran interés a la coocurrencia de trastornos por uso y abuso de sustancias (Sánchez-Roige *et al.* 2018; Walter *et al.* 2018). El primer estudio, empleando medidas cuantitativas del consumo de alcohol mediante el test

Cuestionario de Identificación de los Trastornos debidos al Consumo de Alcohol (AUDIT, *Alcohol Use Disorders Identification Test*), detectó una importante arquitectura genética compartida entre las puntuaciones de AUDIT y otros fenotipos adictivos así como correlaciones genéticas con trastornos psiquiátricos generales. El segundo estudio, representa el mayor GWAS de dependencia de alcohol hasta la fecha. En este trabajo los resultados señalan también en la misma dirección: importantes correlaciones genéticas con otros fenotipos psiquiátricos que podrían indicar una genética común entre fenotipos que se encuentren dentro del espectro de toma de riesgos y bajo control de impulsos.



5.1.3. Estudio 3. Análisis de la capacidad predictiva de las PRS en la modelización combinada con factores predictivos de riesgo conocidos

Los resultados de este estudio muestran la utilidad de las PRS junto a factores de riesgo conocidos en la predicción de dependencia de alcohol. La evaluación de este modelo de riesgo que combina información predictora basada en la predicción genética combinada de la esquizofrenia y el uso y abuso de sustancias se realizó mediante diferentes medidas como el incremento del área bajo la curva ROC, NRI continuo e IDI. El efecto, confirmado en todos los análisis, fue siempre menor que otros factores de riesgo evaluados, con la excepción de LTE dicotómico.

Todas las PRS incluidas en los distintos modelos junto a las variables predictoras de riesgo conocidas se mantuvieron significativas. Es interesante que, en el modelo que incluye la historia familiar, las PRS se mantuviesen significativas. De hecho, la correlación entre la historia familiar y las PRS no fue significativa. Por una parte, la historia familiar puede ser interpretada como un indicador de factores genéticos de distinto efecto, pues informaría de variantes comunes y variantes raras, así como de factores de riesgo ambientales compartidos. Por otra parte, las PRS están basadas en variantes genéticas comunes y podrían añadir especificidad individual. Un resultado similar fue previamente descrito en esquizofrenia empleando registros nórdicos (Lu *et al.* 2018).

Los análisis en los que se incluyeron los valores individuales en la escala de experiencias traumáticas en la infancia como covariable predictora confirman los resultados previos que informaban del importante papel del trauma infantil en el abuso de alcohol (Rothman *et al.* 2008; Keyes *et al.* 2011; Schwandt *et al.* 2013). En la muestra diana, la escala CTQ total mostró un alto nivel de asociación. Examinando las subescalas incluidas en este test, la más explicativa fue la que informaba de la negligencia emocional sufrida por los individuos de la muestra diana. Además, esta variable fue la más correlacionada con las PRS generadas. El efecto de las PRS en la predicción de la escala fenotípica fue muy baja en el modelo que incluyó la subescala de trauma por negligencia emocional. Estos resultados sugieren que el

efecto de las PRS estaría mediada en parte por su efecto, aumentando el riesgo de trauma infantil. De hecho, los análisis de regresión lineal mostraron la asociación significativa entre las PRS y las puntuaciones en la subescala de negligencia emocional, incluso cuando se ajustó el modelo incluyendo la información de historia familiar. Estos resultados parecen ser un claro ejemplo de correlación entre genética y ambiente (rGE) pasiva. En rGE pasiva, la misma persona hereda el riesgo genético de predisposición al trastorno así como el riesgo ambiental. Por lo tanto, los hijos de padres alcohólicos, que heredarán más variantes de riesgo genético, son criados en un ambiente más estresante, con una mayor probabilidad de sufrir experiencias traumáticas infantiles crónicas como la negligencia emocional. Existe un gran número de evidencias sobre el hecho de la heredabilidad de los eventos estresantes (Power *et al.* 2013). De hecho, uno de los diez descubrimientos más replicados en la genética del comportamiento (Plomin *et al.* 2016) es que “la mayoría de medidas del “ambiente” muestran una influencia genética significativa”. Además, Plomin y colaboradores, indican también que “muchas de las asociaciones entre medidas ambientales y los rasgos psicológicos están significativamente mediados por la genética”, algo que parece estar ocurriendo en la relación entre PRS, el trauma infantil y la dependencia al alcohol.

Los análisis en los que se analizaron los factores de la personalidad mostraron una situación general similar. Los factores de la personalidad parecen jugar un papel de peso en la dependencia al alcohol, donde grande metaanálisis de tipo transversales detectaron que altos valores de neuroticismo y bajos de amabilidad y responsabilidad, están asociados con la dependencia al alcohol (Kotov *et al.* 2010). Los estudios longitudinales muestra un patrón distinto, donde una baja responsabilidad y una alta extraversión fueron asociados al alcoholismo como factores de riesgo (Hakulinen *et al.* 2015). La subescala de los rasgos de personalidad que informa de neuroticismo fue el factor de riesgo más asociado con un aumento en el área bajo la curva ROC de todos los analizados en la muestra diana, superando el valor alcanzado por historia familiar (17% vs 14,5%). La mejora aportada por el modelo genético combinado fue muy bajo y los análisis de correlación confirmaron la relación entre ambas variables. Los análisis de

mediación, al igual que los análisis de trauma infantil, detectaron una asociación entre neuroticismo y las PRS. Estos hechos parecen indicar que existe una rGE activa. En esta situación, la genética del individuo le predispone tanto a riesgos ambientales como genéticos de predisposición a dependencia de alcohol. Por lo tanto, el papel de las PRS debe estar mediado por el neuroticismo y la responsabilidad. Estos resultados coinciden de nuevo con los encontrados por Plomin y colaboradores en la genética del comportamiento (Plomin *et al.* 2016).

Los resultados de los análisis incluyendo la información de experiencias traumáticas vitales sufridas durante el último año no mostraron ninguna correlación con las PRS. Además, LTE de tipo dicotómico fue la variable menos explicativa, informando en menor grado que las PRS. Este resultado puede entenderse desde la perspectiva de tratamiento clínico de los pacientes ya que muchos de ellos, a pesar de mostrar un importante alcoholismo crónico, presentan remisión del trastorno. Además, alguno de los elementos medidos por la escala de Brugha incluyen eventos que implican la involucración activa de los pacientes, como el divorcio y, otras, que implican una involucración pasiva, como el fallecimiento de un ser querido. Al emplear esta escala desde una perspectiva acumulativa, aumentó la capacidad predictora de esta variable un 330%.

El modelo basado en la suma ponderada de las PRS generadas con las puntuaciones previamente entrenadas en un conjunto de muestras permitió aprovechar la genética compartida entre trastornos con el fin de mejorar la predicción en alcoholismo. Estas PRS “multifenotipo” permitieron mejorar la predicción, lo cual demuestra que la carga de riesgo individual de un individuo a sufrir dependencia de alcohol se basa en el sumatorio de miles de variantes comunes de predisposición al fenotipo que pueden ser reforzadas mediante la adición de efectos de rasgos correlacionados. El uso de fenotipos combinados en la predicción poligénica es un punto prometedor, existiendo otros ejemplos o tipos de análisis basados en esta premisa, ya que el explotar el solapamiento y correlación presente entre arquitecturas moleculares de trastornos psiquiátricos o fenotipos relacionados mejoraría la

predicción en la escala fenotípica (Maier et al. 2015; Krapohl *et al.* 2018).

Este estudio presenta ciertas limitaciones. La retrospección implícita al cuestionario que analiza la exposición a trauma infantil implica un mayor riesgo de distorsiones. Este hecho ha sido previamente analizado, detectando baja correlación entre el análisis retrospectivo del trauma y el seguimiento longitudinal del mismo. A pesar de esto, ambas medidas del trauma infantil están consistentemente asociadas al riesgo de desarrollo de dependencia de alcohol y los falsos positivos suelen ser raros (Keyes *et al.* 2011). Según los mismos autores, el principal resultado de esta inestabilidad es la infraestimación de la prevalencia de eventos adversos en la infancia, “lo cual podría debilitar las aparentes asociaciones, pero no sesgarlas”. Algo similar ocurre con los factores informativos de rasgos de la personalidad ya que la dependencia al alcohol podría afectar la personalidad del individuo (Hakulinen & Jokela 2018). Por último, indicar que la adición de un mayor número de rasgos o trastornos genéticamente relacionados con la capacidad predictora del riesgo en una muestra diana de alcoholismo podría mejorar la discriminación del modelo. Con la llegada de nuevas estadísticas resumen basadas en estudios cada vez más grandes podrá mejorarse la capacidad predictora de manera sustancial.

5.1.4 Estudio 4. Análisis de la capacidad predictiva de modelos poligénicos fundamentados en la selección de variantes en base a criterios biológicos funcionales

Los resultados de este estudio exploratorio reflejan la limitada capacidad predictiva actual de los modelos de riesgo poligénico fundamentados en la selección de variantes en base a criterios biológicos funcionales. Este modelo basado en la selección de variantes asociadas a la respuesta a estrés agudo señala, a través de los efectos medidos en el mayor GWAS de trastorno depresivo mayor hasta la fecha, la posible relación entre el riesgo a desarrollar abuso o dependencia al alcohol y las diferencias individuales en la cascada transcripcional como respuesta al cortisol liberado en situaciones de estrés.

Arloth y colaboradores observaron en su estudio que las variantes genéticas comunes en elementos amplificadores de largo alcance que alteran la respuesta inmediata del transcriptoma mediada por glucocorticoides están situadas a una distancia media mayor que los eQTLs basales medidos. Además, en sus análisis indican que al menos un 25% de las variantes podrían mostrar una interacción de largo alcance, mediada por interacción directa de la cromatina entre las regiones amplificadoras y los respectivos puntos de inicio transcripcionales. De todas formas, hay que considerar que todos los eQTLs incluidos en este análisis son *cis*-eQTLs, ya que se encuentran a una distancia siempre menor a 1 megabase. En análisis futuros, sería de interés analizar en profundidad la regulación de amplio rango, mediada por *trans*-eQTLs, debido a los indicios detectados del potencial papel de estas variantes para la identificación de genes conductores clave que contribuyen a la enfermedad (Westra *et al.* 2013; Lloyd-Jones *et al.* 2017). Una de las limitaciones de esta propuesta reside en el tamaño del efecto en los *trans*-eQTLs, que al ser menor requiere todavía mayores tamaños muestrales para su detección, pero existen indicios de su potencial (Võsa *et al.* 2018).

La capacidad de predicción de riesgo es uno de los factores limitantes en este estudio. El porcentaje de varianza explicada por el modelo fue limitada, ya que el valor máximo detectado fue del

0,2948%. A pesar de la baja capacidad predictiva, el modelo resistió el control por permutaciones basado en la selección de variantes mediante SNPsnap. Esto es indicativo de que los SNPs del modelo explican más variación que la esperada para igual número de SNPs al azar con las mismas características en relación al conjunto variantes asociadas a la respuesta a estrés. Estos resultados señalan la necesidad de analizar en mayor profundidad la respuesta aguda a estrés y su relación con el alcoholismo. Aunque es bien conocido que el riesgo a sufrir un trastorno por abuso o dependencia de la sustancia se ve exacerbado por factores genéticos y por la exposición a estrés, los mecanismos biológicos funcionales que dirigen la interacción de estos factores se encuentran pobremente analizados. La cascada transcripcional detectada por Arloth y colaboradores puede ser clave para comprender los procesos de señalización y transcripción génica asociados, así como otras interacciones génicas secundarias. El empleo de la genética compartida de depresión mayor y alcoholismo en estos análisis se basa en tres aspectos: los efectos detectados por Arloth y colaboradores en trastorno depresivo mayor en su investigación original, la existencia de un nuevo GWAS que multiplica el número de individuos analizados y la importante correlación genética existente entre el trastorno depresivo mayor y la dependencia de alcohol (Walter *et al.* 2018).

Los análisis que indagaban en la posible interacción entre las distintas PRS y la escala CTQ no encontraron señales significativas. Ninguna de las subescalas de trauma infantil, ni el valor general, mostraron una interacción significativa con las PRS. Estos resultados negativos pueden deberse a la importante limitación existente en el tamaño muestral del conjunto de individuos diana. Al emplear solamente aquellos individuos con información recogida sobre su infancia, el tamaño de la muestra diana se redujo un 70%, con la consiguiente pérdida de potencia estadística. Si los resultados principales mostraron una importante limitación en la detección del efecto, la ausencia de resultados significativos en este análisis secundario era algo esperable.

Una de las características interesantes de este modelo basado en criterios biológicos funcionales es que, aunque la selección de variantes

se basó en análisis de células sanguíneas periféricas, los autores encontraron importantes evidencias de su importancia en circuitos neurales clave en la generación y regulación de la respuesta a situaciones adversas. Las variantes identificadas estaban enriquecidas en amplificadores específicos del cerebro, sugiriendo que su papel puede tener efectos en distintos tejidos, con mayor peso del cerebral. Además, en sus análisis con modelos animales, detectaron cambios diferenciales en la expresión de la cascada en el hipocampo, córtex prefrontal y la amígdala. Por último, empleando genética de neuroimagen (*imaging genetics*) demostraron que el riesgo acumulativo de variantes asociadas a trastorno depresivo mayor basadas en la cascada transcripcional de respuesta a estrés predecía una mayor reactividad generalizada de la amígdala ante estímulos neutrales.

El estudio de las bases biológicas de las adicciones presenta el reto de integrar los resultados de GWAS en el marco de alta especificidad de tejido de los efectos regulatorios y funcionales. Existen bajas correlaciones en la expresión génica entre tejido cerebral y sangre, presentando una importante especificidad de tejido, correlacionada entre las distintas estructuras cerebrales (GTEx Consortium 2015; GTEx Consortium 2017). La detección de variantes de respuesta a estrés con enriquecimiento específico en cerebro y patologías psiquiátricas, así como los resultados en modelos animales o neuroimagen, contrarrestan los problemas por la diferencia transcripcional dependiente de tejido.

El esfuerzo investigador basado en neuroimagen del espectro alcohólico ha detectado numerosas evidencias de una posible función de la amígdala en la relación entre el estrés y el alcoholismo. Estudios estructurales y de análisis de la actividad de esta región cerebral han señalado su rol en la predicción de problemas de consumo y patrones de uso del alcohol (Hariri *et al.* 2009; Nikolova & Hariri 2012; Koob & Volkow 2016; Nikolova *et al.* 2016; Mackey *et al.* 2018). El estrés reciente podría incrementar el riesgo al uso o abuso del alcohol, ya que las experiencias vitales traumáticas son “potentes moduladores” en ambos circuitos neurales (Koob *et al.* 2007; Roozental *et al.* 2009). Artloth y colaboradores detectaron una mayor reactividad de la

amígdala ante imágenes de rostros de naturaleza emocional neutra asociada a una mayor carga de riesgo de las variantes comprendidas en el modelo transcripcional de respuesta a estrés en el marco del trastorno depresivo mayor. En la misma dirección, Nikolova y colaboradores (Nikolova *et al.* 2016) detectaron en consumo problemático de alcohol en dos modelos de respuesta de la amígdala en combinación con otras regiones cerebrales, en el que uno de ellos mostraba una mayor actividad de esta región cerebral ante imágenes de rostros de naturaleza emocional neutra. Este modelo se ha relacionado con el consumo problemático de alcohol a través de la ansiedad y la depresión. La heterogeneidad en los resultados de distintos estudios comentados en este párrafo, podría reflejar distintos mecanismos o ventanas de desarrollo a través de la respuesta y exposición a eventos estresantes asociados al consumo de alcohol problemático. Las diferencias en los tipos de estrés, predisposición individual y otros factores de riesgo confusores deben ser tenidos en cuenta.

Las limitaciones de este estudio se reflejan en la pequeña capacidad predictiva predictiva. El débil efecto inherente a regiones reguladoras como los *cis*-eQTLs, sumado a un tamaño limitado de la muestra diana empleada y al uso de una aproximación transversal (MDD2 – abuso/dependencia de alcohol) puede explicar la frágil señal detectada. Futuras aproximaciones incluyendo datos de dependencia de alcohol (Schizophrenia Working Group of the PGC, 2014; Walters *et al.* 2018) serán de gran interés. También es necesario considerar las limitaciones asociadas a la retrospección en las aproximaciones que incluyen información de trauma infantil, concepto ya desarrollado en la discusión del tercer estudio de esta Tesis. Por último, como señalan Arloth y colaboradores, aunque emplean un agonista selectivo del receptor de glucocorticoides, no se puede descartar la respuesta compartida con otros receptores esteroideos como el de mineralocorticoides, por lo que no se puede excluir una importante contribución de otros receptores a esta cascada regulatoria transcripcional.

El conocimiento de las vías neurobiológicas que llevan a la adicción y al alcoholismo permanece todavía incompleto (Hancock *et*

al. 2018). En este momento, el entendimiento del desarrollo y regulación transcripcional de la respuesta a estrés es limitado, pero ofrece un importante punto de partida para dilucidar las relaciones entre la exposición a estrés y el desarrollo del alcoholismo. Las variantes que cambian la respuesta transcripcional inmediata a estrés mediada por glucocorticoides podrían tener un rol transversal en los trastornos psiquiátricos relacionados con el estrés, por lo que futuras investigaciones en esta misma dirección arrojarán más luz a este posible mecanismo molecular compartido.



5.2 DISCUSIÓN GENERAL

5.2.1 Arquitectura genética compartida

Desde la perspectiva generada gracias a los últimos años de investigación en genómica psiquiátrica, partiendo del artículo señalado de Purcell y colaboradores (Purcell *et al.* 2009), los modelos de riesgo poligénico han jugado un papel clave en el esclarecimiento de la genética común subyacente a distintos trastornos psiquiátricos. El número de estudios, metodologías y diversos enfoques empleados, destacan en la literatura de los últimos años. El rápido crecimiento ha permitido reconocer por primera vez entre los distintos trastornos psiquiátricos la amplia carga poligénica extendida y el solapamiento parcial de la genética de riesgo de un modo casi generalizado.

Las evidencias obtenidas permiten reconocer grupos de trastornos más ampliamente relacionados, como puede ser el caso de esquizofrenia y trastorno bipolar, o diferenciados, como el conjunto de trastornos de inicio en edad temprana (p.ej. trastornos del espectro autista o trastorno por déficit de atención e hiperactividad) y aquellos trastornos de inicio en etapa juvenil o adulta (esquizofrenia o depresión). El porcentaje de varianza explicada en la escala fenotípica y las predicciones de riesgo logradas han ido mejorando en los últimos años, como se ha comentado en la Introducción de esta Tesis, y uno de los mejores ejemplos del avance en el campo reside en el uso de datos derivados del mega-análisis llevado a cabo en esquizofrenia en 2014 (Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014; Bogdan *et al.* 2018). Del mismo modo que en la Introducción de esta tesis se ejemplificó con la esquizofrenia la evolución de los GWAS en psiquiatría, el progreso en el uso de modelos de riesgo poligénico está bien reflejado en el mismo trastorno (Bogdan *et al.* 2018). A fecha de la publicación de esta tesis doctoral, el número de estudios que emplean datos de esquizofrenia desde una perspectiva poligénica o de PRS asciende a varios cientos y el ritmo de publicación es frenético. Estos, utilizando distintos enfoques, se han centrado en estudiar fenotipos relacionados: cognición, la relación con otros trastornos o fenotipos psiquiátricos, la relación con otras patologías no mentales y otros fenotipos clínicos y no clínicos. Algunos de los

ejemplos de fenotipos relacionados se basan en: el estudio de síntomas (Fanous *et al.* 2012; Stepniak *et al.* 2014; Sengupta *et al.* 2017), el curso del trastorno en sus distintas etapas vitales (Stepniak *et al.* 2014; Jones *et al.* 2016), edad de inicio (Ahn *et al.* 2016), historia familiar (Agerbo *et al.* 2015) o respuesta a tratamientos (comentada extensamente en el subapartado “Traslación a la clínica”).

Tras los estudios iniciales (Purcell *et al.* 2009; Cross Disorder Group of the Psychiatric GWAS Consortium 2013), en los que se detectó un solapamiento de la arquitectura poligénica de la esquizofrenia con el trastorno bipolar y el trastorno depresivo mayor, las PRS derivadas de GWAS de esquizofrenia han sido ligadas a un gran número de trastornos/fenotipos psiquiátricos: trastorno de estrés postraumático (Duncan *et al.* 2018), trastorno de ansiedad en adolescentes (Jones *et al.* 2016), capacidades comunicativas (Riglin *et al.* 2017; St. Pourcain *et al.* 2017), trastorno por déficit de atención e hiperactividad junto a otros trastornos durante la etapa infantil (ansiedad, trastorno negativista desafiante, depresión) (Nivard *et al.* 2017), trastorno obsesivo-compulsivo (Costas *et al.* 2016) (primer estudio desarrollado en este proyecto de Tesis), uso y abuso de sustancias (comentadas de forma extendida en el segundo y tercer proyecto de Tesis), suicidio (Laursen *et al.* 2017), neuroticismo (Gale *et al.* 2016) o subfenotipos/endofenotipos en trastorno bipolar y la propia esquizofrenia (Allardyce *et al.* 2018; Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium 2018).

En cognición, algunos de los distintos estudios se centran en cambios cognitivos dependientes de la edad (McIntosh *et al.* 2013), creatividad (Power *et al.* 2015), aprendizaje (Riglin *et al.* 2017), inteligencia (Smeland *et al.* 2019) o rendimiento neurocognitivo (Hatzimanolis *et al.* 2015).

Algunos de los fenotipos clínicos relacionados a nivel fisiológico cerebral están basados en estudios imagen por resonancia magnética funcional (Kauppi *et al.* 2015; Chen *et al.* 2018) o electrofisiología (Roussos *et al.* 2016; Liu *et al.* 2017). Por último, la relación con otras patologías no psiquiátricas, interacciones o fenotipos no clínicos

presenta una gran diversidad de ejemplos: índice de masa corporal (Peterson *et al.* 2011), interacción con complicaciones obstétricas (Ursini *et al.* 2018), psoriasis (Yin *et al.* 2016) o urbanicidad (Paksarian *et al.* 2017).

El porcentaje de varianza explicada varía en gran medida en las distintas relaciones analizadas con segundos fenotipos, con valores relativamente modestos en la mayoría de las aproximaciones (un gran número de aproximaciones no llegan a obtener resultados superiores al 1% y “rara vez” superan el 3%) (Bogdan *et al.* 2018). En general, las PRS de esquizofrenia explican un mayor porcentaje de la varianza en otros trastornos que con características clínicas concretas y otros fenotipos. En este aspecto, en el mismo sentido y como ha sido comentado anteriormente, los trabajos aportados en esta tesis han alcanzado valores relativamente altos al emplear PRS basados en modelos de esquizofrenia, en comparación con el resto de literatura científica. El número de estudios en otros trastornos distintos a esquizofrenia, al igual que en el caso de GWAS, es menor y suelen tener una menor potencia discriminatoria generalizada. Esto es debido a que el mayor conocimiento y base genómica para el desarrollo de PRS reside en el grado de desarrollo de los GWAS, mayor en el caso de esquizofrenia. Pero en general, la visión actual es la de un amplio solapamiento entre trastornos o rasgos históricamente relacionados en el ámbito clínico y algunos hallazgos sorprendentes que desafían a la visión nosológica clásica. De acuerdo con esta afirmación, los resultados obtenidos en los estudios de adicciones de esta tesis indican las relaciones subyacentes con el resto de trastornos psiquiátricos generales, a pesar de su separación histórica.

En el caso de los trastornos psiquiátricos y otras patologías complejas, la heterogeneidad biológica inherente a la propia entidad nosológica que describe un trastorno concreto podría estar eclipsando parte del potencial de los modelos de riesgo poligénico. Esta heterogeneidad podría estar afectando a las señales significativas de GWAS así como al resto de variantes estudiadas, perjudicando al descubrimiento de señales significativas e incrementando el ruido incluido en la estimación de PRS. Además, la configuración de

muestras diana también puede estar influenciada por este aspecto. Sin embargo, es necesario considerar siempre la inesperada complejidad de la arquitectura genética subyacente de los trastornos, que puede implicar un mayor esfuerzo investigador en ciertos casos (Sullivan *et al.* 2018). Al considerar la heterogeneidad clínica y otras variables como la inestabilidad diagnóstica o comorbilidad como elementos presentes en las muestras que configuran modelos de riesgo poligénico surge una pregunta ¿en qué medida los resultados de solapamiento son debidos a estas causas?

Los GWAS de mayor tamaño realizados en psiquiatría se basan en distintos conjuntos de cohortes obtenidas a partir de registros clínicos y colecciones de población general. Los pacientes adheridos a los distintos programas pueden mostrar diferencias sustanciales subyacentes a pesar de un mismo diagnóstico. En función del trastorno, el efecto de estas diferencias puede ser de distinta magnitud y en diferentes direcciones. Por ejemplo, aquellos individuos dependientes de una sustancia que acuden a solicitar tratamiento clínico podrían presentar una mayor carga psicopatológica respecto a aquellos incluidos a partir de colecciones generales. Por lo tanto, estas diferencias pueden dar lugar a resultados que no reflejan exactamente las relaciones de la arquitectura genética subyacente. A la hora de admitir unas evidencias concretas, es necesario sopesar las posibles diferencias en la conformación de las muestras descubrimiento y diana, que aunque definidas por el mismo fenotipo, podrían contener y reflejar distintos orígenes o características clínicas. Existen un gran número de mecanismos que pueden ser controlados en este aspecto que ofrecen oportunidades de mejora en las estimaciones: selección de fenotipos más concretos, selección de fenotipos más extremos, selección de súper-controles o fenotipos intermedios con el fin de aclarar la heterogeneidad del trastorno. Su utilidad se extiende en muchos aspectos genéticos, llegando a jugar un papel relevante en procesos traslacionales a la clínica (Breen *et al.* 2016; Gandal *et al.* 2016).

Recuperando el concepto de arquitectura genética compartida, como ha sido comentado en la introducción, los métodos de análisis de heredabilidad basada en chips de SNPs se han extendido a la estimación

de la correlación de la genética común entre parejas de trastornos psiquiátricos (Smoller 2017). Los primeros pasos de relevancia, basados en la iniciativa Cross-Disorder del PGC, detectaron distintos patrones de correlación entre patologías psiquiátricas (Cross-Disorder Group of the PGC 2013 [b]). Otras aproximaciones han desentrañado la correlación entre los cinco factores de la personalidad y las patologías psiquiátricas (Lo *et al.* 2017), sumándose a otras iniciativas que, con éxito, han ido descifrando las distintas relaciones entre trastornos y rasgos mentales. Para tener una imagen más global de estos distintos grados de correlación, la base de datos LD Hub resume este compendio de relaciones (Zheng *et al.* 2017).

La aproximación llevada a cabo por Brainstorm Consortium (Brainstorm Consortium 2018) es, hasta la fecha, uno de los mejores ejemplos en la cuantificación de la heredabilidad y correlación genética de los trastornos del cerebro. Incorporando un total de 265218 casos y 784643 controles, los investigadores analizaron 25 trastornos del cerebro, incluyendo para cada uno de ellos meta-análisis de amplio tamaño muestral. Los resultados principales se resumen en una importante correlación entre patologías psiquiátricas, mientras que los trastornos neurológicos muestran mayores diferencias entre sí y con los trastornos psiquiátricos (exceptuando la epilepsia). Además, los autores demostraron la importante correlación del neuroticismo, entre los distintos factores de personalidad, con las patologías psiquiátricas y la migraña. Asimismo, se realizaron análisis incluyendo rasgos fenotípicos potencialmente relacionados con las patologías del cerebro (como habilidades cognitivas, rasgos psicológicos, inteligencia, etc.) presentándose solapamiento en distintos grados (en mayor medida en los trastornos psiquiátricos respecto a las patologías neurológicas). Este estudio, añade mayores evidencias de cara a cuestionar las fronteras clínicas de los fenotipos psiquiátricos, ya que “no reflejan distintos procesos patogénicos subyacentes” y sugiere una mayor interconexión etiológica entre patologías psiquiátricas. Estos hechos están relacionados con los patrones de expresión global compartidos entre distintos trastornos psiquiátricos detectados por Gandal y colaboradores. El nivel de compartición de la desregulación transcripcional está relacionado con el solapamiento poligénico entre

trastornos, mostrando “vías de convergencia y especificidad molecular” (Gandal *et al.* 2018).

5.2.2 Nosología psiquiátrica

Los sistemas diagnósticos empleados en la clínica y la investigación psiquiátrica consideran los trastornos mentales desde una postura de categorización, independencia y distinción tal y como ha sido comentado en el primer apartado de la introducción de esta Tesis. Desde la perspectiva clínica y científica hay evidencias e interés en resaltar el marco continuo general, en vez de categórico, que rodea a las distintas patologías (Kotov *et al.* 2017). A pesar de esto, diversos autores coinciden en que la nosología psiquiátrica es útil (Caspi *et al.* 2014; Owen 2014), aunque imperfecta (Sanislow *et al.* 2010). Algunos autores señalan la existencia de “una crisis de confianza en la nosología psiquiátrica” (Caspi & Moffit, 2018). El sistema de clasificación actual debería ser capaz de incorporar los avances del presente y el futuro observados y previstos en la investigación genómica en psiquiatría o neurociencias (Kupfer *et al.* 2011). Los últimos avances nos indican que los efectos genéticos en las patologías psiquiátricas son poligénicos y pleiotrópicos (Insel *et al.* 2010; Robinson *et al.* 2016; Boyle *et al.* 2017; Brainstorm Consortium 2018; Wray *et al.* 2018), poniendo en duda la idea de que los trastornos psiquiátricos sean entidades discretas. A continuación serán discutidos distintos aspectos que atañan al futuro de la nosología psiquiátrica en su avance en paralelo a los resultados aportados por la genética.

5.2.2.1 Factor general de la psicopatología: Una visión dimensional

Ante los resultados obtenidos mediante aproximaciones poligénicas, inevitablemente surgen dos preguntas: ¿es sostenible esta diferenciación definida entre trastornos con las evidencias obtenidas mediante investigación genética? Y ¿podría existir una arquitectura

nosológica integradora y global que mejore la predicción de susceptibilidad o riesgo?

La correlación genética entre trastornos previamente comentada se suma a la correlación existente a otro nivel de la estructura de la psicopatología: del espectro (Caspi *et al.* 2014). La correlación entre entidades clínicas incluidas en el espectro Externalizante y el Internalizante es aproximadamente del 0,5 y, entre Internalizante y “Trastornos del pensamiento” (*Thought Disorders*) la estimación sube al 0,6 (Wright *et al.* 2013). El espectro “Trastornos del pensamiento” es un espectro que puede ser reconocido, según los autores, como un tercer espectro o como subdominio del espectro Internalizante (Keyes *et al.* 2013; Wright *et al.* 2013; Caspi *et al.* 2014). Este espectro, incluiría aquellos trastornos del marco de las experiencias psicóticas, manía y trastorno obsesivo-compulsivo. Dadas estas altas correlaciones resultados basados en investigación con gemelos, Lahey y colaboradores (Lahey *et al.* 2012) sugirieron la posibilidad de la existencia de un factor psicopatológico general junto a los factores Internalizantes y Externalizantes. Este factor psicopatológico general resumiría desde una perspectiva general integradora subyacente la proclividad de los individuos a desarrollar cualquier forma de psicopatología común de la mano de causas únicas de heredabilidad para cada trastorno (Caspi *et al.* 2014). Caspi y colaboradores proponen la existencia de un Factor Psicopatológico general al que denominan “factor p”. Este factor, basado en una idea unidimensional, mejoraría la explicación aportada por la clásica jerarquía de los espectros Internalizante, Externalizante, y “Trastornos del pensamiento” y se situaría sobre ellas. Además, los autores concluyen que el “factor p” es insuficiente para describir la estructura de la psicopatología, ya que las dimensiones Internalizante y Externalizante añaden información más allá de la aportada por el “factor p” (Caspi *et al.* 2014). Esta dimensión une a todos los trastornos e influye sobre la presencia o ausencia de cientos de síntomas psiquiátricos (Caspi & Moffit, 2018). Estos síntomas son agregados en unas docenas de diagnósticos clásicos, que a su vez se organizan en tres dominios generales: Internalizante, Externalizante y “Trastornos del pensamiento”, integrados en el “factor p”. Este factor tendría, según diversos autores, la capacidad de mejorar

el avance en el entendimiento de la base biológica de los trastornos psiquiátricos con importantes implicaciones en la estructura de la psicopatología (Pettersson *et al.* 2015).

El “factor p” es análogo al “factor g”. Este segundo factor se refiere a la estructura común existente en el ámbito de las habilidades cognitivas que, pudiendo ser separadas en distintas categorías, muestran un factor que resume o predice que aquellos individuos que realizan con éxito un test cognitivo tenderán a realizar otro tipo de pruebas con un éxito equivalente (Jensen 1998; Caspi & Moffit 2018).

Por último, cabe indicar que estos autores han logrado comparar las correlaciones genéticas de cuatro métodos diferentes con el fin de examinar su posible contribución o apoyo a la existencia de un “factor p” genético en los principales trastornos psiquiátricos. Los cuatro métodos se basaban en la estimación de un componente principal que apoyase la hipótesis del “factor p” al analizar las correlaciones genéticas entre los trastornos, calculadas mediante información clínica familiar, GCTA y LDSC y una matriz de correlaciones de PRS. Todos los trastornos presentaron un componente principal que informaba del 57, 43, 35, y 22% de la varianza respectivamente a cada una de las cuatro aproximaciones (Selzam *et al.* 2018).

Estas aproximaciones permiten interpretar los resultados del segundo estudio de esta tesis desde otra perspectiva más integradora en lo que a la estructura de la psicopatología se refiere. Esta perspectiva permite hipotetizar sobre el posible valor de las PRS basadas en esquizofrenia como agente informador de esta dimensión psicopatológica superior, ya que existen importantes evidencias de que valores extremos del “factor p” están relacionados con los “Trastornos del pensamiento” (Caspi *et al.* 2014). Estos valores pueden dar lugar a trastornos psicóticos, por lo que validarían el uso de datos de GWAS de esquizofrenia con el fin de generar modelos de riesgo predictores.

5.2.2.2 El DSM y las fronteras nosológicas

Los debates sobre la estructura nosológica y la aparición de modelos predictores que consideran nuevas estructuras nosológicas superiores a la nosología clásica se suman a la información derivada de los grandes saltos dados por la genética psiquiátrica planteando la siguiente cuestión: ¿existen fronteras entre trastornos psiquiátricos y entre el trastorno y la variación normal? Es evidente que el incremento de descubrimientos que evidencian esa falta de fronteras nítidas aviva el debate sobre la validez de la estructura nosológica actual.

El DSM, a pesar de haber avanzado hacia la consideración de entidades nosológicas espectrales en su última versión (American Psychiatric Association, 2013), la estructura nosológica general no se rige por ese esquema (Zachar & Kendler 2017; Smoller *et al.* 2018), como se ha comentado en la introducción. Aunque estas categorías siguen siendo necesarias desde un punto de vista práctico, revisiones futuras deberán incorporar los nuevos conocimientos derivados de la investigación masiva llevada a cabo durante los últimos años, fuertemente apoyados en la iniciativa PGC (Sullivan *et al.* 2018). Estos no se reducen únicamente a los resultados obtenidos mediante el estudio de la variación común, sino que también están reflejados en la variación rara y su papel en la hipótesis del continuo del neurodesarrollo (Owen & O'Donovan 2017). En este sentido, como ha sido comentado en la introducción, aunque la hipótesis del neurodesarrollo ha tenido una influencia en la investigación de la esquizofrenia (Owen *et al.* 2011), los resultados obtenidos a través de la genética psiquiátrica no han sido trasladados a la estructura nosológica desde la perspectiva del continuo del neurodesarrollo. Desde esta perspectiva espectral, el concepto nosológico de los trastornos del neurodesarrollo debería ser ampliado para incluir las psicosis funcionales (Owen & O'Donovan 2017).

El proceso de incorporación de la “nosología genética” (junto a otros biomarcadores) permitirá al DSM evolucionar a través de distintos modelos, como la iteración epistémica. Según las premisas que acompañan a este modelo, se produciría una adquisición reiterada y secuencial de los nuevos datos y evidencias biológicas (Kendler 2009; Kendler 2014; Smoller *et al.* 2018). Esta revisión es un reto sustancial

para la nosología psiquiátrica, poco acostumbrada a la incorporación de biomarcadores. La situación ideal desde esta perspectiva sería la de asegurar una mejora en cada revisión periódica de la nosología, por lo que solamente deberían ser incorporados aquellos elementos que aseguren un mejor rendimiento de los criterios diagnósticos (Kendler 2009). Esta aproximación progresiva/evolutiva debería permitir conocer mejor “la realidad de los trastornos psiquiátricos” y no se comportaría de manera rupturista con la nosología que hoy conocemos. De todas formas, el debate estará presente en cada proceso, ya que las nuevas medidas podrán chocar con criterios descriptivos actuales (Kendler 2014).

Otras opciones podrían basarse en un “giro más radical” en el que se presentaría una nosología con una estructura de la psicopatología muy diferente a la actual (Smoller *et al.* 2018). Esta opción, refundaría la nosología en opciones más espectrales con categorías diagnósticas dimensionales con umbrales bien definidos. En este sentido, los desafíos más importantes que encara el DSM tienen que ver con la visión espectral/dimensional o iniciativas como RDoC (Insel 2013; Zachar & Kendler 2017; Smoller *et al.* 2018). En el caso de RDoC, sus defensores confían obtener conocimiento de mecanismos causales con mayor facilidad a través de la iniciativa. RDoC buscaría emplear categorías o constructos basados en la neurociencia cognitiva y conductual, con la confianza de que esta estructura sería útil tanto en la clínica como en la investigación (Zachar & Kendler 2017).

Independientemente de la estrategia a seguir, la incorporación de informadores de riesgo o diagnósticos con base biológica, no debería constituir un inconveniente. El proceso en sí podría parecerse en gran medida a la agregación de nuevos síntomas o indicios en la nosología clásica.

En este sentido, los modelos de riesgo poligénico representan un sistema de apoyo relativamente fácil de incorporar que, además, han tenido una importante magnitud en lo que al debilitamiento de las fronteras entre patologías psiquiátricas se refiere. Los resultados de esta tesis son un paso más en el camino llevado a cabo en el proceso de aclarar la arquitectura subyacente y su relación con la estructura de la

psicopatología mediante el empleo de la variación genética común. Sin embargo, el proceso a la hora de buscar mejoras en la nosología y su aplicación, deberá buscar más allá de la variación común o la genómica en sí, ya que deberá apoyarse en distintas aproximaciones (neuroimagen, bioquímica o registros de dispositivos electrónicos). La genética de forma aislada podría aportar algunos mecanismos útiles para la nosología, pero esperar de ella un cambio radical podría ser erróneo. Echando la vista atrás, el proceso de avance en genómica psiquiátrica ha sido un proceso gradual en el que no se han presentados cambios radicales que pudieran suponer un cambio de paradigma. Aunque puedan surgir dudas respecto al grado de “especificidad de trastorno” de estas herramientas (p. ej. la pleiotropía extendida en parte de los descubrimientos genéticos), éstas no deberían diferir mucho de las que surgen en el proceso de agregación de un mismo síntoma a varios trastornos.

5.2.3 PRS: Mejoras y futuro

La estrategia poligénica ha llegado a consolidarse en los últimos años, pero todavía queda margen de mejora de estas estimaciones de PRS. Parte del progreso dependerá de la generación de nuevos datos y otra parte se apoyará en mejoras metodológicas o refinamientos que permitirán el esclarecimiento de las relaciones entre trastornos.

La dependencia del poder predictivo de PRS respecto al tamaño muestral de estrategias GWAS indica la importancia de incrementar su tamaño. En este sentido, un gran número de iniciativas globales prometedoras cumplirán un papel de referencia en este progreso. La iniciativa PGC, con la liberación secuencial de las nuevas fases de GWAS crecientes en el número de individuos involucrados permitirá afinar todavía más esta aproximación (Sullivan *et al.* 2018). En este refinamiento y aumento de potencia jugarán un papel decisivo las cohortes de gran tamaño de iniciativas nacionales como el UK Biobank (Bycroft *et al.* 2018) o el programa “All of US” del NIH (Precision Medicine Initiative Working Group 2015). Además, los biobancos nórdicos así como iniciativas centradas en otras poblaciones (BioBank

Japan o China Kadoorie Biobank) arrojarán también algo más de luz. La accesibilidad a los datos obtenidos mediante estas estrategias marcará también el ritmo de mejora, donde estrategias de “ciencia abierta” serán con probabilidad aquellas que más rápido permitirán traducir el progreso en la predicción de riesgo genético. El avance, como se ha comentado antes, es frenético. Algunos de los ejemplos más recientes obtenido gracias a la mejora de las PRS, es la rápida evolución en el esclarecimiento de la arquitectura genética. Esta puede resumirse en la multiplicación de *loci* significativos en trastorno depresivo mayor alcanzando 102 señales (Howard *et al.* 2019), el análisis de 1,2 millones de individuos que contribuyen a nuevos conocimientos sobre la etiología genética del consumo de tabaco y alcohol (Liu *et al.* 2019), los resultados que demuestran el peso de la variación común en trastornos del neurodesarrollo que históricamente han sido considerado monogénicos (Niemi *et al.* 2018), los trabajos e iniciativas centradas en la genómica de neuroimagen como ENIGMA o resultados derivados de UK Biobank (Bearden & Thompson 2017; Elliot *et al.* 2018) o la capacidad de diseccionar la heterogeneidad de los trastornos mediante el uso de PRS (Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium 2018; Grove *et al.* 2019). Los ejemplos se extienden rápidamente durante los últimos años, donde el esfuerzo colaborativo y la bajada de costes están dando sus frutos y el poder acceder a las estadísticas resumen darán un nuevo impulso a los modelos de riesgo poligénico.

Los modelos de riesgo poligénico, aunque han respondido irregularmente en la búsqueda de la heredabilidad perdida/escondida (Wray & Maier 2014), no han cubierto todo el rango esperado. Aunque el de fin esta aproximación no es esclarecer esta parte de la heredabilidad, sí cumplen un papel relevante. La “heredabilidad todavía perdida” puede estar causada por diversos motivos que se relacionan o afectan a la aproximación poligénica: una sobrestimación de la heredabilidad basada en estudios de gemelos y familiares, la existencia de variantes no cubiertas por los SNPs comunes y la heterogeneidad de la enfermedad y las muestras empleadas (Manolio *et al.* 2009; Tenesa & Haley 2013; Wray *et al.* 2014; Wray & Maier 2014; Tropf *et al.* 2017). Estos factores afectan también a otros aspectos y clases de

estudios de los trastornos complejos en general. Desde la perspectiva de crecimiento se espera una mayor capacidad explicativa de las PRS en las escalas de predisposición, pero este crecimiento alcanzará el *plateau*. Con el fin de mejorar el entendimiento de las patologías y rellenar parcialmente el hueco de “heredabilidad todavía perdida”, existen múltiples estrategias. Una de ellas es emplear la correlación genética subyacente a los trastornos psiquiátricos, en los que aproximaciones como MTAG (Turley *et al.* 2018) podrían mejorar las estimaciones del efecto de las variantes analizadas mediante la incorporación de la información genética de otros rasgos. Como se ha comentado a lo largo de esta tesis y como elemento de estudio de la misma, la estimación de PRS puede mejorarse empleando esta genética combinada. En el caso de MTAG, las expectativas teóricas son de incrementar hasta un 25% la capacidad predictiva de las PRS. Otra opción de mejora se basa en el uso de cohortes longitudinales y la “interacción PRS x ambiente”, donde los modelos de riesgo poligénico pueden ser de gran utilidad y serán incluidos en distintas iniciativas (Wray *et al.* 2014; Smoller *et al.* 2018; Sullivan *et al.* 2018).

Por último, para comprender la capacidad predictiva de las PRS y su traslación clínica, un buen ejemplo reside en el trabajo de Khera y colaboradores (Khera *et al.* 2018). En este trabajo, los autores logran generar PRS equivalentes a las estimaciones de riesgo de mutaciones monogénicas de gran efecto. A continuación se desarrollará extensamente el potencial de las PRS en este aspecto.

5.2.4 Traslación a la clínica de las PRS

El objetivo final de la investigación en genética de enfermedades complejas resultados al ámbito clínico (Wray *et al.* 2018). Además, el conocimiento comunes es trasladar sus adquirido puede ser empleado como herramienta de prevención en el ámbito clínico. Algunos de los distintos atractivos de la implementación de los modelos de riesgo poligénico en la clínica residen en la facilidad de cálculo y almacenamiento de datos, la estabilidad predictiva durante la vida del individuo, la posibilidad de permitir la predicción de la enfermedad con

gran antelación al inicio de la misma y el coste-efectividad (Chatterjee *et al.* 2016; Lewis & Vassos 2017). El incremento de señales significativas detectadas por GWAS y los avances proporcionados mediante aproximaciones poligénicas tienen diversos potenciales en el campo de los trastornos mentales. Poniendo el foco en el valor de las PRS en psiquiatría, este sub-apartado busca resaltar el posible papel a jugar por estas herramientas, su comparación con otros campos de la Medicina y los posibles contratiempos y dificultades en su posible traslación.

Torkamani y colaboradores (Torkamani *et al.* 2018) establecieron en su revisión tres grandes categorías que resumen la utilidad de las PRS en la intervención clínica: (1) intervención terapéutica informada por PRS, (2) monitorización de la enfermedad informada por PRS y (3) la planificación vital basada en las PRS. En la primera de las aplicaciones, las PRS podrían jugar un papel en la selección de intervenciones para un manejo de la enfermedad individualizado. Numerosos estudios han mostrado la utilidad de las PRS en el ámbito de la enfermedad coronaria de manera independiente de la historia familiar, identificando individuos de riesgo que recibieron tratamientos con estatina (Mega *et al.* 2015; Torkamani *et al.* 2018). En el caso de monitorización de la enfermedad con el uso de PRS los puntos clave residen en la capacidad de decidir el inicio del monitoreo del paciente o en la interpretación de los resultados de los mismos. Los mayores avances en esta dirección han sido realizados en cáncer de mama o cáncer colorrectal. Otras aplicaciones podrían residir en el rol de interpretación de falsos positivos en controles de cáncer de próstata (Torkamani *et al.* 2018). Por último, la utilidad preventiva general de las estimaciones de riesgo con el fin de evitar actividades u otros factores de riesgo vitales.

5.2.4.1 Traslación en el ámbito de la psiquiatría

En psiquiatría, todavía estamos a varios años de distancia de la aplicación de modelos de riesgo poligénico en la práctica clínica (Zheutlin & Ross 2018). La naturaleza de las patologías psiquiátricas

inherente a su arquitectura genética molecular dificulta la optimización de estrategias de intervención a corto plazo. Los tratamientos farmacológicos conforman parte de las distintas estrategias clínicas iniciales, por lo que la detección de individuos o grupos de pacientes con características de respuesta a fármacos concretas representa uno de los intereses principales. En este marco, varias aproximaciones se enfocaron en analizar la posible capacidad predictiva de las PRS en la respuesta a tratamientos con antipsicóticos (Frank *et al.* 2015; Wimberley *et al.* 2017; Li *et al.* 2018). La detección de pacientes resistentes a antipsicóticos clásicos o de segunda generación es una de las metas a alcanzar en esquizofrenia, ya que aproximadamente un 30-40% de los individuos no responden al tratamiento (Conley & Kelly 2001). Los estudios de Frank y colaboradores (2015) y Wimberley y colaboradores (2017) emplearon modelos de riesgo poligénico para detectar individuos con esta condición, ya que utilizaron registros de terapia con clozapina como información de resistencia a tratamiento. Los primeros detectaron diferencias entre aquellos individuos con resistencia al tratamiento y aquellos que nunca habían recibido tratamiento con clozapina, pero los resultados de los segundos no consiguieron replicar los resultados (el tamaño muestral era ligeramente menor en el estudio de resultados negativos). Li y colaboradores (2018) emplearon modelos de riesgo poligénico en relación con la respuesta clínica a la lurasidona en pacientes con esquizofrenia crónica, con el fin de detectar biomarcadores útiles para predecir la respuesta a antipsicóticos atípicos. Si la detección de individuos que responden al tratamiento mediante predicción de riesgo genético es posible, la ratio existente entre beneficio y riesgo asociado a este tipo de medicaciones podría mejorar.

Al poner el foco en la predicción de la eficacia de antidepresivos, algunos de los estudios más recientes emplean distintas estrategias con este fin (García-González *et al.* 2017; Amare *et al.* 2018; Ward *et al.* 2018). La respuesta positiva a la primera aproximación farmacológica suele presentarse aproximadamente en el 40% de las muestras, por lo que la estratificación de pacientes en función del tratamiento idóneo mediante PRS es una de las metas a perseguir. El estudio de García-González mostró resultados negativos en este sentido tras emplear en

su estudio muestras de siete trabajos enfocados a la farmacogenética. Los resultados fueron negativos tanto al intentar predecir mejora o remisión de síntomas depresivos mediante PRS generadas a partir de dos muestras de respuesta farmacológica, como al analizar la capacidad predictiva de los síntomas de PRS generadas a partir de GWAS de esquizofrenia o trastorno depresivo mayor. Amare y colaboradores estudiaron, en dos cohortes independientes, la predicción poligénica de la respuesta al tratamiento del trastorno depresivo mayor mediante inhibidores selectivos de la re-captación de serotonina. Los modelos poligénicos, basados en los cinco principales factores de la personalidad detectaron resultados significativos al predecir respuestas al tratamiento al emplear datos basados en neuroticismo, apertura o responsabilidad. Estos resultados indican el potencial uso de modelos poligénicos basados en los principales factores de la personalidad “combinados con otra información clínica predictiva”. Por último, Ward y colaboradores, meta-analizando los resultados obtenidos en distintas cohortes con la generación de PRS, buscaron analizar la capacidad predictiva de modelos basados en trastorno depresivo mayor y neuroticismo en el marco de la respuesta a antidepresivos. Los resultados, aunque estadísticamente inconsistentes, mostraron indicios de la probable utilidad de estos modelos.

La posible utilidad de la variación genética común en el ámbito de los tratamientos farmacológicos en psiquiatría no está ceñida únicamente a los modelos de riesgo poligénico, sino que los propios GWAS conforman una de las herramientas de mayor peso a la hora de detectar posibles nuevas dianas farmacológicas (Breen *et al.* 2016; Gandal *et al.* 2016). Desde una perspectiva distinta a la de las PRS y su probable papel en la estratificación farmacológica de pacientes, los GWAS son una de las herramientas con mayor peso a la hora de iniciar la traslación de los resultados genómicos a la farmacología.

La detección poligénica de pacientes “extremos” podría tener importantes implicaciones en el curso de la enfermedad, ya que podrían evitarse recidivas mejorando la evolución del paciente en distintos aspectos clínicos y sociales. Distintos trabajos han analizado el potencial de las PRS en el primer episodio psicótico (Vassos *et al.* 2017;

Sengupta *et al.* 2017; Zhang *et al.* 2019). Vassos y colaboradores lograron detectar diferencias significativas entre aquellos individuos que acabaron desarrollando esquizofrenia de aquellos en los que se desarrollaron otros trastornos psicóticos, pero la precisión en la discriminación fue relativamente baja. Sengupta y colaboradores analizaron dimensiones de síntomas, detectando una correlación positiva entre PRS y la dimensión general de psicopatología pero no con los síntomas positivos o negativos cuantificados por separado. Además, los autores demostraron la importancia de los cambios dinámicos de los síntomas con el paso del tiempo a la hora de afinar la relación entre factores genéticos y fenotipos. Por último, Zhang y colaboradores predijeron una menor respuesta a antipsicóticos en aquellos individuos con una mayor carga de riesgo a sufrir esquizofrenia. El empleo de estos modelos en el contexto de posible aplicación en la atención temprana podría ser clave en las etapas prodrómicas de la esquizofrenia. Además, las PRS generadas podrían llegar a tener aplicaciones a la hora de estimar el grado de cronicidad de la patología y la frecuencia de ingresos hospitalarios (Meier *et al.* 2016).

La extensión de estas herramientas potenciales a otros trastornos psiquiátricos puede ser posible. Por ejemplo, en trastorno bipolar, es probable predecir mayores síntomas psicóticos a partir de modelos basados en esquizofrenia o trastorno bipolar (Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium 2018). Como se ha señalado en la discusión individual del proyecto de trastorno obsesivo-compulsivo, su posible papel estratificador de pacientes con el fin de detectar subgrupos con necesidades y características clínicas concretas podría implicar una potencial utilidad en individuos con el trastorno. La detección de estos pacientes a corta edad podría implicar una importante mejora en el pronóstico y el manejo del paciente.

Futuros estudios deberán emplear aproximaciones combinando información genética común, basada en modelos poligénicos, y variantes raras de gran efecto, como las CNVs (Bergen *et al.* 2019). La relación existente entre estas dos clases de variantes está comenzando

a elucidarse gracias a trabajos que ahondan en análisis combinados en esquizofrenia, trastorno por déficit de atención e hiperactividad y trastornos del espectro autista (Martin *et al.* 2015; Tansey *et al.* 2016; Weiner *et al.* 2017; Bergen *et al.* 2019).

El beneficio de las PRS en la clínica es que pueden sumarse y unirse a algoritmos predictivos más complejos (Suvisari *et al.* 2018). Aunque prometedoras, las PRS requieren un importante desarrollo previo a su implantación (Smoller *et al.* 2018). Una aproximación interesante a la hora de aplicar los modelos de riesgo poligénico en la clínica se basa en la unión de efectos de diferentes factores predictores (Chatterjee *et al.* 2016). Estos modelos combinados, como se ha mostrado en el tercer estudio de esta Tesis, tienen un futuro de gran interés en psiquiatría, dado que la incorporación de distintos tipos de información externa podría mejorar el poder predictivo. La modelización conjunta es importante, ya que la historia familiar es de gran importancia, considerando que las PRS solamente pueden explicar una fracción del riesgo a la enfermedad asociado a la historia familiar. Aun así, su éxito radica en parte en una mayor especificidad de individuo, diferenciando entre individuos de una misma familia. Como Chatterjee y colaboradores indican, la historia familiar suele presentarse como una variable binaria, por lo que la inclusión de historia familiar más extendida o detallada (incluyendo edad de inicio entre otros elementos clínicos informadores) podría mejorar la capacidad predictiva. En esta misma dirección, los datos recogidos en la segunda fase del estudio de adicciones incluyeron un gran número de variables clínicas del individuo afecto y sus familiares. La implementación y proyección de estos modelos es ya una realidad en distintos tipos de cáncer. En cáncer de mama, un estudio (Mavaddat *et al.* 2015), que incluía información genética basada en 77 variantes logró, de manera conjunta con la historia familiar, una importante estratificación de individuos en función del riesgo individual, superior a la lograda empleando únicamente predictor no genético. Esta aproximación fue más efectiva que el criterio básico de estratificación empleado en los mecanismos de cribado mediante mamografías, basado en la edad (Burton *et al.* 2013). Resultados en la misma dirección han sido mostrados en el caso del cáncer colorrectal, en el que un trabajo mostró

un impacto sustancial en la predicción de riesgo combinando modelos de riesgo basados en 27 SNPs, historia familiar y datos derivados de endoscopia. Este trabajo logró mejorar las anteriores recomendaciones clásicas de edad de inicio de tests de cribado (Hsu *et al.* 2015). Aunque con menor capacidad predictiva, también existen ejemplos en psiquiatría (Lu *et al.* 2017). Este estudio ya comentado en la tercera discusión individual de esta Tesis, empleó datos de registros nórdicos para la integración de datos de historia familiar y variantes comunes estableciendo un posible uso futuro de interés en el espectro psicótico.

5.2.4.2 Implementación clínica de las PRS en psiquiatría: Implicaciones éticas, barreras y dificultades

El potencial uso de las PRS en la mejora de los resultados clínicos en pacientes de una amplia variedad de patologías presenta importantes desafíos éticos, clínicos y científicos (Wray *et al.* 2013). Parte de los mismos son exclusivos de esta clase de herramienta genética, pero en su mayoría son comunes a las dificultades asociadas al diagnóstico y consejo genético: implicaciones para miembros de la misma familia, estigma, confidencialidad y privacidad, comunicación e interpretación de los resultados (Hoge & Appelbaum 2012).

Desde el punto de vista genético, las PRS aportan una imagen parcial del riesgo a desarrollar el trastorno. La información cubierta por esta herramienta será de gran utilidad, pero implicará un conocimiento incompleto del riesgo total. Esta situación se ve incrementada en aquellos trastornos psiquiátricos sindrómicos, donde el peso de la variación común capturada por GWAS y constituyente de los modelos poligénicos representa una fracción menor. La ratio coste-beneficio es clave para la implantación y desarrollo de las distintas herramientas genómicas en el ámbito clínico. El debate sobre la estrategia genómica a seguir, PRS *vs* WGS, debería ser aplazado al punto en el que la aproximación de secuencia completa represente un coste-beneficio asumible por los sistemas de salud.

La “medicina poligénica” supondrá un cambio importante en la genética clínica, ya que requerirá una transformación desde una medicina más acostumbrada a trabajar en el campo de trastornos raros (Lewis & Vassos, 2017): el paso de un sistema bivariable a uno continuo. Las PRS en el percentil superior serían indicadoras de riesgo, no conformarían un pronóstico definitivo. El reto en la psiquiatría actual no es la clasificación de casos *vs* controles, sino que está enfocada a la detección de subconjuntos de individuos con el fin de mejorar su trayectoria clínica (Kapur *et al.* 2012). Por lo tanto, es necesario poner el foco en los cambios necesarios en la práctica clínica, en la que serán necesarias nuevas capacidades en la comprensión de métricas de riesgo más dinámicas y complejas. Existen múltiples ejemplos de falsa dicotomía en la relación entre variantes deterministas en enfermedad de tipo familiar y el riesgo poligénico probabilístico (Torkamani *et al.* 2018). Algunos modelos de riesgo poligénico modifican o señalan en la dirección contraria respecto a la historia familiar o variantes genéticas raras de gran efecto en el mismo individuo en cáncer de mama u ovario (Gaudet *et al.* 2013; Kuchenbaecker *et al.* 2017). Además, los sistemas sanitarios deberán adaptarse a nuevos factores éticos y legales en la adopción de nuevos programas de estratificación con información genética. La rápida evolución de la genómica psiquiátrica requiere una rápida respuesta del conjunto de profesionales clínicos y legisladores para la conformación de un escenario adaptado a los nuevos retos bioéticos (Ward *et al.* 2019).

La formación para clínicos y el público general implicaría la apropiada percepción del riesgo así como una mejor interpretación individual y social de los resultados (Chaterjee *et al.* 2016; Lewis & Vassos, 2017; Schork *et al.* 2018; Torkamani *et al.* 2018). Para lograr esta meta, además de la formación, serán necesarias estrategias y herramientas de comunicación de riesgos (Chaterjee *et al.* 2016; Torkamani *et al.* 2018). La interpretación del riesgo y de conceptos estadísticos asociados es una dificultad intrínseca de las PRS a las que deberá adaptarse la comunicación clínico-paciente. En el proceso de mejora de la predicción de riesgo de los modelos de riesgo poligénicos surgirán nuevas necesidades en la relación médico-paciente, como aquellas centradas en la prevención del impacto psicosocial de los

resultados de estas pruebas (Torkamani *et al.* 2018). La mala interpretación de los hallazgos por parte del paciente puede llegar a tener importantes implicaciones a nivel de adherencia al tratamiento, estigmatización y discriminación. La información derivada de la investigación científica que llega a la población general suele implicar importantes simplificaciones y exageraciones con el fin de incrementar la atención del público (Caulfield & Condit 2012). Además, el mayor conocimiento del papel jugado por los factores genéticos en los trastornos mentales está frecuentemente asociado a interpretaciones erróneas deterministas y esencialistas (Kong *et al.* 2017). Esta posible sesgo en la interpretación de los resultados genéticos, puede implicar una peor interpretación con importantes implicaciones a nivel de individuo y su entorno social y familiar, por lo que la transmisión de los resultados deberá considerar esta posible situación. Los resultados genéticos pueden tener, junto a otras “explicaciones biológicas” de la enfermedad, connotaciones negativas en diversos aspectos debido al reduccionismo y las creencias deterministas, afectando incluso a la percepción del clínico (Lebowitz & Ahn 2014). También pueden surgir actitudes de tolerancia, de gran interés siempre que no residan en un sentimiento determinista subyacente, actualmente en boga en distintos contextos científicos con impacto en la sociedad de consecuencias imprevistas (Comfort 2018). Por lo tanto, podría afirmarse que existe cierta dualidad en las consecuencias, con impactos que pueden ser positivos con la correcta interpretación y transmisión.

Esta dualidad “beneficio-riesgo” puede ser observada en otros escenarios, como en el consejo genético en menores tanto en etapa infantil como adolescente. Debido a la herencia compleja de los trastornos psiquiátricos, la información obtenida mediante PRS podría ser útil con el fin de eludir riesgos ambientales conocidos durante esta etapa de desarrollo (además de los beneficios intrínsecos a la atención temprana comentados previamente).

Al mismo tiempo, surgen importantes inquietudes respecto al potencial uso de las PRS en el marco prenatal, donde podrían tener importantes propósitos eugenésicos (Gronowski *et al.* 2014). Este riesgo se ve incrementado en la etapa actual, en la que la

comercialización de tests genéticos dirigidos directamente al consumidor por parte de grandes empresas podría incrementar la intolerancia de cara a los trastornos mentales y rasgos relacionados, ya que parte de estos tests comienzan a estar enfocados al cribado prenatal.

Una de las limitaciones más importantes en la implantación normalizada del uso de modelos poligénicos predictivos en la sanidad global reside en la observación de que actualmente su valor predictivo es mucho mayor en individuos de ascendencia europea en comparación con otras poblaciones humanas. La comunidad científica ha reconocido ampliamente la sobrerrepresentación de participantes de ascendencia europea en la investigación de la genética humana (Rosenberg *et al.* 2010; Bustamante *et al.* 2011; Dalvie *et al.* 2015; Popejoy & Fullerton 2016; Morales *et al.* 2018). Aproximadamente, en la actualidad, el 78% de los participantes en GWAS presentan ascendencia europea, aunque representan el 16% de la población global (Morales *et al.* 2018). La mejora en la representación de individuos no europeos es ligeramente notable, respecto los datos del 2009, donde la representación estaba reducida al 4% de los participantes en GWAS (Bustamante *et al.* 2011). El incremento de diversas poblaciones es un reto que la investigación genética en psiquiatría ha considerado priorizar (CONVERGE Consortium 2015; Duncan *et al.* 2017), ya que existen múltiples evidencias de la mayor capacidad predictiva en europeos respecto a poblaciones no europeas empleando modelos de riesgo basados en grandes GWAS (Li *et al.* 2017; Vassos *et al.* 2017; Wedow *et al.* 2018). La genética de poblaciones prevé que la precisión en la predicción de riesgo genético decaerá con el aumento de la divergencia genética entre la muestra original de GWAS y la diana (Scutari *et al.* 2016, Torkamani *et al.* 2018). Empíricamente, las PRS rinden mucho mejor en poblaciones europeas, ya que su precisión en la predicción se reduce de 2 a 5 veces en poblaciones del este asiático o de origen africano respectivamente (Martin *et al.* 2019). Cuando se emplean estudios basados en poblaciones no europeas, a pesar de representar solamente una fracción del tamaño de los conjuntos europeos, su capacidad predictiva se incrementa: Li y colaboradores (Li *et al.* 2017) mostraron un incremento del 260% en la predicción empleando una muestra con un tamaño del 20% respecto a la muestra de mayor tamaño disponible

(Schizophrenia Working Group of the PGC 2014). Desde un contexto evolutivo, las poblaciones europeas presentan cierto empobrecimiento en la diversidad genética entre la ascendencia europea respecto a la africana, por lo que las características poblacionales también están capturadas en los GWAS. El efecto de la selección, la deriva o los efectos ambientales son causas potenciales en estas diferencias en los resultados en el marco de la multi-etnicidad.

Aunque parte de este problema podría paliarse con mejoras metodológicas, una de las principales soluciones reside, básicamente, en aumentar el esfuerzo inversor en GWAS de poblaciones no europeas (H3Africa Consortium 2014). Esta solución es costosa, lenta y compleja. En psiquiatría, se están realizando esfuerzos para mejorar e incrementar la diversidad genética que además apoyan el intercambio de datos. Iniciativas como Stanley Global Neuropsychiatric Genetics Initiative buscan paliar esta situación (Sullivan *et al.* 2018). Hasta ahora las aproximaciones multiétnicas son escasas, por lo que el avance hasta la fecha es limitado (Duncan *et al.* 2017; Walters *et al.* 2018). Desde el marco de la sanidad española el problema es menor, ya que la estructura étnica de la población presenta una mayor representación de individuos de origen europeo. Pero en el marco global la situación es otra. Para poder desplegar el uso de las predicciones de riesgo poligénico en regiones con alta diversidad o con representación principalmente no europea, es necesario equilibrar, poco a poco, la balanza. El déficit actual debe ser considerado desde la perspectiva de la sanidad global, ya que el cálculo de este desequilibrio es necesario para la estimación de su impacto a corto-medio plazo (Martin *et al.* 2019). Las oportunidades de mejora basadas en las PRS en la sanidad global y la psiquiatría en pacientes de distintas poblaciones deben paliar las disparidades en la salud mundial, no incrementarlas.

CONCLUSIONES





6. CONCLUSIONES

1. Los análisis de riesgo poligénico compartido entre esquizofrenia y trastorno obsesivo-compulsivo representan la primera evidencia descrita de una arquitectura genética solapante entre ambos trastornos. Esta susceptibilidad genética común podría ser parcialmente responsable de la frecuente comorbilidad entre ambos trastornos.
2. El solapamiento poligénico subyacente detectado entre esquizofrenia y trastorno obsesivo-compulsivo se incorpora al debate nosológico sobre la validez de un subtipo de pacientes esquizo-obsesivos. Las puntuaciones de riesgo poligénico podrían ser útiles en la detección y estratificación de pacientes que, siendo diagnosticados con uno de los dos trastornos, podrían pertenecer a un subconjunto clínico más específico y homogéneo.
3. Los análisis de las relaciones de los principales trastornos psiquiátricos con el trastorno por abuso o dependencia de sustancias confirman y amplían las observaciones previas que buscan desentrañar la susceptibilidad genética compartida. De los seis trastornos analizados, el que comparte una predisposición genética significativa es la esquizofrenia. El solapamiento poligénico de los cinco trastornos restantes, aunque significativamente menor, se ve reflejado en los resultados obtenidos mediante el uso de una muestra conjunta que engloba cuatro de ellos.
4. La susceptibilidad genética compartida entre la esquizofrenia y el trastorno por abuso o dependencia de sustancias no es debida a la comorbilidad psiquiátrica en la muestra diana. El nivel de pleiotropía detectado es considerable, encontrándose entre los valores más altos descritos en la literatura científica hasta la fecha.

5. En el conjunto de pacientes con abuso o dependencia al alcohol, las mujeres muestran una mayor carga poligénica de predisposición respecto a los hombres. Estas diferencias podrían reflejar la mayor presión social en contra del consumo excesivo de alcohol en mujeres.
6. La evaluación de la capacidad predictiva de la modelización combinada de factores de riesgo conocidos junto a puntuaciones de riesgo poligénico señala la posible utilidad de la predicción genética en el ámbito de los trastornos mentales. El incremento predictivo aportado por las puntuaciones de riesgo, aunque no redundante, está limitado por el conocimiento parcial de su arquitectura genética. El uso de nuevas muestras descubrimiento, basadas en resultados de nuevos y futuros mega-análisis, permitirá mejoras predictivas.
7. Los resultados predictivos de las puntuaciones de riesgo generadas en la muestra diana a partir de datos combinados de esquizofrenia y alcoholismo, evidencian el valor del uso conjunto de fenotipos psiquiátricos relacionados genéticamente frente a la capacidad predictiva de un modelo basado únicamente en datos de alcoholismo.
8. Los análisis de correlación entre puntuaciones de riesgo poligénico y variables clínicas de riesgo conocidas, acreditan la relación subyacente. La mediación detectada entre variables parece ser pasiva en el caso de la negligencia emocional traumática sufrida en la infancia y activa en el marco de los factores de la personalidad, observada en neuroticismo y responsabilidad.
9. Los resultados exploratorios obtenidos mediante un modelo poligénico basado en la selección de variantes asociadas a la respuesta a estrés, señalan la posible relación entre el riesgo a desarrollar alcoholismo y las diferencias individuales en la cascada transcripcional inmediata como respuesta al estrés agudo. Asimismo, estos resultados indican la dificultad de

esclarecer los mecanismos biológicos funcionales en el marco de la poligenicidad extendida.

- 10.** Los modelos de riesgo poligénico señalan, desde una perspectiva general, hacia una arquitectura compartida entre trastornos psiquiátricos, basada en un importante componente poligénico y con un importante pleiotropismo extendido más allá de las fronteras nosológicas entre trastornos. Estos resultados deberán ser considerados en la evaluación de la estructura nosológica actual y la visión dimensional de la misma.
- 11.** La traslación a la clínica de las puntuaciones de riesgo poligénico debe estar, en principio, enfocada a la estratificación de individuos en grupos más homogéneos en función de su riesgo genético con el fin de mejorar su manejo y pronóstico clínico, no al diagnóstico individual. El progreso en la capacidad predictiva podría tener importantes implicaciones éticas y terapéuticas en el marco de los trastornos mentales.





BIBLIOGRAFÍA





7. BIBLIOGRAFÍA

- 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. A map of human genome variation from population-scale sequencing. *Nature*. 2010 Oct 28;467(7319):1061-73. doi: 10.1038/nature09534. Erratum in: *Nature*. 2011 May 26;473(7348):544. Xue, Yali [added]; Cartwright, Reed A [added]; Altshuler, David L [corrected to Altshuler, David]; Keibel, Andrew [corrected to Keebler, Jonathan]; Koko-Gonzales, Paula [corrected to Kokko-Gonzales, Paula]; Nickerson, Debbie A [corrected to Nickerson, Debo. PubMed PMID: 20981092; PubMed Central PMCID: PMC3042601.
- 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012 Nov 1;491(7422):56-65. doi: 10.1038/nature11632. PubMed PMID: 23128226; PubMed Central PMCID: PMC3498066.
- 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. *Nature*. 2015 Oct 1;526(7571):68-74. doi: 10.1038/nature15393. PubMed PMID: 26432245; PubMed Central PMCID: PMC4750478.
- Aas M, Melle I, Bettella F, Djurovic S, Le Hellard S, Bjella T, Ringen PA, Lagerberg TV, Smeland OB, Agartz I, Andreassen OA, Tesli M. Psychotic patients who used cannabis frequently before illness onset have higher genetic predisposition to schizophrenia than those who did not. *Psychol Med*. 2018 Jan;48(1):43-49. doi: 10.1017/S0033291717001209. Epub 2017 Oct 2. PubMed PMID: 28967348.
- Abbott A. Novartis to shut brain research facility. *Nature*. 2011 Dec 6;480(7376):161-2. doi: 10.1038/480161a. PubMed PMID: 22158218.
- Achim AM, Maziade M, Raymond E, Olivier D, Mérette C, Roy MA. How prevalent are anxiety disorders in schizophrenia? A meta-analysis and critical review on a significant association. *Schizophr Bull*. 2011 Jul;37(4):811-21. doi: 10.1093/schbul/sbp148. Epub 2009 Dec 3. Review. PubMed PMID: 19959704; PubMed Central PMCID: PMC3122284.
- Agerbo E, Sullivan PF, Vilhjálmsson BJ, Pedersen CB, Mors O, Børghlum AD, Hougaard DM, Hollegaard MV, Meier S, Mattheisen M, Ripke S, Wray NR, Mortensen PB. Polygenic Risk Score, Parental Socioeconomic Status, Family

- History of Psychiatric Disorders, and the Risk for Schizophrenia: A Danish Population-Based Study and Meta-analysis. *JAMA Psychiatry*. 2015 Jul;72(7):635-41. doi: 10.1001/jamapsychiatry.2015.0346. PubMed PMID: 25830477.
- Ahn K, An SS, Shugart YY, Rapoport JL. Common polygenic variation and risk for childhood-onset schizophrenia. *Mol Psychiatry*. 2016 Jan;21(1):94-6. doi: 10.1038/mp.2014.158. Epub 2014 Dec 16. PubMed PMID: 25510512.
- Allardyce J, Gaebel W, Zielasek J, van Os J. Deconstructing Psychosis conference February 2006: the validity of schizophrenia and alternative approaches to the classification of psychosis. *Schizophr Bull*. 2007 Jul;33(4):863-7. Epub 2007 Jun 4. PubMed PMID: 17548844; PubMed Central PMCID: PMC2632332.
- Allardyce J, Leonenko G, Hamshere M, Pardiñas AF, Forty L, Knott S, Gordon-Smith K, Porteous DJ, Haywood C, Di Florio A, Jones L, McIntosh AM, Owen MJ, Holmans P, Walters JTR, Craddock N, Jones I, O'Donovan MC, Escott-Price V. Association Between Schizophrenia-Related Polygenic Liability and the Occurrence and Level of Mood-Incongruent Psychotic Symptoms in Bipolar Disorder. *JAMA Psychiatry*. 2018 Jan 1;75(1):28-35. doi: 10.1001/jamapsychiatry.2017.3485. PubMed PMID: 29167880; PubMed Central PMCID: PMC5833541.
- Allen NC, Bagade S, McQueen MB, Ioannidis JP, Kavvoura FK, Khoury MJ, Tanzi RE, Bertram L. Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat Genet*. 2008 Jul;40(7):827-34. doi: 10.1038/ng.171. PubMed PMID: 18583979.
- Altmüller J, Palmer LJ, Fischer G, Scherb H, Wjst M. Genomewide scans of complex human diseases: true linkage is hard to find. *Am J Hum Genet*. 2001 Nov;69(5):936-50. Epub 2001 Sep 14. Erratum in: *Am J Hum Genet* 2001 Dec;69(6):1413. PubMed PMID: 11565063; PubMed Central PMCID: PMC1274370.
- Amare AT, Schubert KO, Tekola-Ayele F, Hsu YH, Sangkuhl K, Jenkins G, Whaley RM, Barman P, Batzler A, Altman RB, Arolt V, Brockmüller J, Chen CH, Domschke K, Hall-Flavin DK, Hong CJ, Illi A, Ji Y, Kampman O, Kinoshita T, Leinonen E, Liou YJ, Mushiroda T, Nonen S, Skime MK, Wang L, Kato M, Liu YL, Praphanphoj V, Stingl JC, Bobo WV, Tsai SJ, Kubo M, Klein TE, Weinshilboum RM, Biernacka JM, Baune BT. Association of the Polygenic Scores for Personality Traits and Response to Selective Serotonin Reuptake Inhibitors in Patients with Major Depressive Disorder. *Front Psychiatry*. 2018 Mar 6;9:65. doi: 10.3389/fpsy.2018.00065. eCollection 2018. PubMed PMID: 29559929; PubMed Central PMCID: PMC5845551.

- Andersen AM, Pietrzak RH, Kranzler HR, Ma L, Zhou H, Liu X, Kramer J, Kuperman S, Edenberg HJ, Nurnberger JI Jr, Rice JP, Tischfield JA, Goate A, Foroud TM, Meyers JL, Porjesz B, Dick DM, Hesselbrock V, Boerwinkle E, Southwick SM, Krystal JH, Weissman MM, Levinson DF, Potash JB, Gelernter J, Han S. Polygenic Scores for Major Depressive Disorder and Risk of Alcohol Dependence. *JAMA Psychiatry*. 2017 Nov 1;74(11):1153-1160. doi: 10.1001/jamapsychiatry.2017.2269. PubMed PMID: 28813562; PubMed Central PMCID: PMC5710224.
- Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nat Protoc*. 2010 Sep;5(9):1564-73. doi: 10.1038/nprot.2010.116. Epub 2010 Aug 26. PubMed PMID: 21085122; PubMed Central PMCID: PMC3025522.
- Andrews G, Slade T, Issakidis C. Deconstructing current comorbidity: data from the Australian National Survey of Mental Health and Well-Being. *Br J Psychiatry*. 2002 Oct;181:306-14. PubMed PMID: 12356657.
- Arloth J, Bogdan R, Weber P, Frishman G, Menke A, Wagner KV, Balsevich G, Schmidt MV, Karbalai N, Czamara D, Altmann A, Trümbach D, Wurst W, Mehta D, Uhr M, Klengel T, Erhardt A, Carey CE, Conley ED; Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium (PGC), Ruepp A, Müller-Myhsok B, Hariri AR, Binder EB; Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium PGC. Genetic Differences in the Immediate Transcriptome Response to Stress Predict Risk-Related Brain Function and Psychiatric Disorders. *Neuron*. 2015 Jun 3;86(5):1189-202. doi: 10.1016/j.neuron.2015.05.034. PubMed PMID: 26050039; PubMed Central PMCID: PMC4490780.
- Arnold M, Raffler J, Pfeufer A, Suhre K, Kastenmüller G. SNIIPA: an interactive, genetic variant-centered annotation browser. *Bioinformatics*. 2015 Apr 15;31(8):1334-6. doi: 10.1093/bioinformatics/btu779. Epub 2014 Nov 26. PubMed PMID: 25431330; PubMed Central PMCID: PMC4393511.
- Artigas-Pallarés J, Paula-Pérez I. [Unresolved issues in the DSM-5]. *Rev Neurol*. 2015 Feb 25;60 Suppl 1:S95-101. Spanish. PubMed PMID: 25726831.
- Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol Autism*. 2017 May 22;8:21. doi: 10.1186/s13229-017-0137-9. eCollection 2017. PubMed PMID: 28540026; PubMed Central PMCID: PMC5441062.
- Baca-Garcia E, Perez-Rodriguez MM, Basurte-Villamor I, Fernandez del Moral AL, Jimenez-Arriero MA, Gonzalez de Rivera JL, Saiz-Ruiz J, Oquendo MA.

- Diagnostic stability of psychiatric disorders in clinical practice. *Br J Psychiatry*. 2007 Mar;190:210-6. PubMed PMID: 17329740.
- Badner JA, Gershon ES. Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. *Mol Psychiatry*. 2002;7(4):405-11. PubMed PMID: 11986984.
- Baker RW, Chengappa KN, Baird JW, Steingard S, Christ MA, Schooler NR. Emergence of obsessive compulsive symptoms during treatment with clozapine. *J Clin Psychiatry*. 1992 Dec;53(12):439-42. PubMed PMID: 1487472.
- Barnard, G. A. Statistical inference. *Journal of the Royal Statistical Society. Series B*. 1949 (Methodological), 11(2), 115-149.
- Bearden CE, Thompson PM. Emerging Global Initiatives in Neurogenetics: The Enhancing Neuroimaging Genetics through Meta-analysis (ENIGMA) Consortium. *Neuron*. 2017 Apr 19;94(2):232-236. doi: 10.1016/j.neuron.2017.03.033. Review. PubMed PMID: 28426957; PubMed Central PMCID: PMC5918136.
- Becker KG. The common variants/multiple disease hypothesis of common complex genetic disorders. *Med Hypotheses*. 2004;62(2):309-17. PubMed PMID: 14962646.
- Benjamini, Y., & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B*. 1995 (Methodological), 289-300.
- Bergen SE, Ploner A, Howrigan D; CNV Analysis Group and the Schizophrenia Working Group of the Psychiatric Genomics Consortium, O'Donovan MC, Smoller JW, Sullivan PF, Sebat J, Neale B, Kendler KS. Joint Contributions of Rare Copy Number Variants and Common SNPs to Risk for Schizophrenia. *Am J Psychiatry*. 2019 Jan 1;176(1):29-35. doi: 10.1176/appi.ajp.2018.17040467. Epub 2018 Nov 5. PubMed PMID: 30392412; PubMed Central PMCID: PMC6408268.
- Bernstein, C. A. Meta-structure in DSM-5 process. *Psychiatric News*. 2011. 46(5), 7.
- Bernstein, D. P., & Fink, L. (1998). *Childhood Trauma Questionnaire: A retrospective self-report (CTQ)*. San Antonio, TX: NCS Pearson.
- Berrios GE. Obsessive-compulsive disorder: its conceptual history in France during the 19th century. *Compr Psychiatry*. 1989 Jul-Aug;30(4):283-95. PubMed PMID: 2667880.
- Bienvenu OJ, Davydow DS, Kendler KS. Psychiatric 'diseases' versus behavioral disorders and degree of genetic influence. *Psychol Med*. 2011 Jan;41(1):33-40.

- doi: 10.1017/S003329171000084X. Epub 2010 May 12. Review. PubMed PMID: 20459884.
- Biernacka JM, Geske J, Jenkins GD, Colby C, Rider DN, Karpyak VM, Choi DS, Fridley BL. Genome-wide gene-set analysis for identification of pathways associated with alcohol dependence. *Int J Neuropsychopharmacol.* 2013 Mar;16(2):271-8. doi: 10.1017/S1461145712000375. Epub 2012 May 2. PubMed PMID: 22717047; PubMed Central PMCID: PMC3854955. 267: Davison K. Autoimmunity in psychiatry. *Br J Psychiatry.* 2012 May;200(5):353-5. doi: 10.1192/bjp.bp.111.104471. PubMed PMID: 22550326.
- Bierut LJ, Agrawal A, Bucholz KK, Doheny KF, Laurie C, Pugh E, Fisher S, Fox L, Howells W, Bertelsen S, Hinrichs AL, Almasy L, Breslau N, Culverhouse RC, Dick DM, Edenberg HJ, Foroud T, Grucza RA, Hatsukami D, Hesselbrock V, Johnson EO, Kramer J, Krueger RF, Kuperman S, Lynskey M, Mann K, Neuman RJ, Nöthen MM, Nurnberger JI Jr, Porjesz B, Ridinger M, Saccone NL, Saccone SF, Schuckit MA, Tischfield JA, Wang JC, Rietschel M, Goate AM, Rice JP; Gene, Environment Association Studies Consortium. A genome-wide association study of alcohol dependence. *Proc Natl Acad Sci U S A.* 2010 Mar 16;107(11):5082-7. doi: 10.1073/pnas.0911109107. Epub 2010 Mar 2. PubMed PMID: 20202923; PubMed Central PMCID: PMC2841942.
- Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Electronic address: douglas.ruderfer@vanderbilt.edu; Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. *Cell.* 2018 Jun 14;173(7):1705-1715.e16. doi: 10.1016/j.cell.2018.05.046. PubMed PMID: 29906448; PubMed Central PMCID: PMC6432650.
- Birnbaum R, Weinberger DR. Genetic insights into the neurodevelopmental origins of schizophrenia. *Nat Rev Neurosci.* 2017 Dec;18(12):727-740. doi: 10.1038/nrn.2017.125. Epub 2017 Oct 26. Review. PubMed PMID: 29070826.
- Bogdan R, Baranger DAA, Agrawal A. Polygenic Risk Scores in Clinical Psychology: Bridging Genomic Risk to Individual Differences. *Annu Rev Clin Psychol.* 2018 May 7;14:119-157. doi: 10.1146/annurev-clinpsy-050817-084847. PubMed PMID: 29579395.
- Bosker FJ, Hartman CA, Nolte IM, Prins BP, Terpstra P, Posthuma D, van Veen T, Willemsen G, DeRijk RH, de Geus EJ, Hoogendijk WJ, Sullivan PF, Penninx BW, Boomsma DI, Snieder H, Nolen WA. Poor replication of candidate genes for major depressive disorder using genome-wide association data. *Mol Psychiatry.* 2011 May;16(5):516-32. doi: 10.1038/mp.2010.38. Epub 2010 Mar 30. PubMed PMID: 20351714.

Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*. 2017 Jun 15;169(7):1177-1186. doi: 10.1016/j.cell.2017.05.038. Review. PubMed PMID: 28622505; PubMed Central PMCID: PMC5536862.

Brainstorm Consortium, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, Duncan L, Escott-Price V, Falcone GJ, Gormley P, Malik R, Patsopoulos NA, Ripke S, Wei Z, Yu D, Lee PH, Turley P, Grenier-Boley B, Chouraki V, Kamatani Y, Berr C, Letenneur L, Hannequin D, Amouyel P, Boland A, Deleuze JF, Duron E, Vardarajan BN, Reitz C, Goate AM, Huentelman MJ, Kambouh MI, Larson EB, Rogaeva E, St George-Hyslop P, Hakonarson H, Kukull WA, Farrer LA, Barnes LL, Beach TG, Demirci FY, Head E, Hulette CM, Jicha GA, Kauwe JSK, Kaye JA, Leverenz JB, Levey AI, Lieberman AP, Pankratz VS, Poon WW, Quinn JF, Saykin AJ, Schneider LS, Smith AG, Sonnen JA, Stern RA, Van Deerlin VM, Van Eldik LJ, Harold D, Russo G, Rubinsztein DC, Bayer A, Tsolaki M, Proitsi P, Fox NC, Hampel H, Owen MJ, Mead S, Passmore P, Morgan K, Nöthen MM, Rossor M, Lupton MK, Hoffmann P, Kornhuber J, Lawlor B, McQuillin A, Al-Chalabi A, Bis JC, Ruiz A, Boada M, Seshadri S, Beiser A, Rice K, van der Lee SJ, De Jager PL, Geschwind DH, Riemenschneider M, Riedel-Heller S, Rotter JI, Ransmayr G, Hyman BT, Cruchaga C, Alegret M, Winsvold B, Palta P, Farh KH, Cuenca-Leon E, Furlotte N, Kurth T, Ligthart L, Terwindt GM, Freilinger T, Ran C, Gordon SD, Borck G, Adams HHH, Lehtimäki T, Wedenoja J, Buring JE, Schürks M, Hrafnisdóttir M, Hottenga JJ, Penninx B, Artto V, Kaunisto M, Vepsäläinen S, Martin NG, Montgomery GW, Kurki MI, Hämäläinen E, Huang H, Huang J, Sandor C, Webber C, Muller-Myhsok B, Schreiber S, Salomaa V, Loehrer E, Göbel H, Macaya A, Pozo-Rosich P, Hansen T, Werge T, Kaprio J, Metspalu A, Kubisch C, Ferrari MD, Belin AC, van den Maagdenberg AMJM, Zwart JA, Boomsma D, Eriksson N, Olesen J, Chasman DI, Nyholt DR, Avbersek A, Baum L, Berkovic S, Bradfield J, Buono R, Catarino CB, Cossette P, De Jonghe P, Depondt C, Dlugos D, Ferraro TN, French J, Hjalgrim H, Jamnadas-Khoda J, Kälviäinen R, Kunz WS, Lerche H, Leu C, Lindhout D, Lo W, Lowenstein D, McCormack M, Møller RS, Molloy A, Ng PW, Oliver K, Privitera M, Radtke R, Ruppert AK, Sander T, Schachter S, Schankin C, Scheffer I, Schoch S, Sisodiya SM, Smith P, Sperling M, Striano P, Surges R, Thomas GN, Visscher F, Whelan CD, Zara F, Heinzen EL, Marson A, Becker F, Stroink H, Zimprich F, Gasser T, Gibbs R, Heutink P, Martinez M, Morris HR, Sharma M, Ryten M, Mok KY, Pulit S, Bevan S, Holliday E, Attia J, Battey T, Boncoraglio G, Thijs V, Chen WM, Mitchell B, Rothwell P, Sharma P, Sudlow C, Vicente A, Markus H, Kourkoulis C, Pera J, Raffeld M, Silliman S, Boraska Perica V, Thornton LM, Huckins LM, William Rayner N, Lewis CM, Gratacos M, Rybakowski F, Keski-Rahkonen A, Raevuori A, Hudson JI, Reichborn-Kjennerud T, Monteleone P, Karwautz A, Mannik K, Baker JH, O'Toole JK, Trace SE, Davis OSP, Helder SG, Ehrlich S, Herpertz-Dahlmann B, Danner UN, van Elburg AA, Clementi M, Forzan M, Docampo E, Lissowska J, Hauser J, Tortorella

A, Maj M, Gonidakis F, Tziouvas K, Papezova H, Yilmaz Z, Wagner G, Cohen-Woods S, Herms S, Julià A, Rabionet R, Dick DM, Ripatti S, Andreassen OA, Espeseth T, Lundervold AJ, Steen VM, Pinto D, Scherer SW, Aschauer H, Schosser A, Alfredsson L, Padyukov L, Halmi KA, Mitchell J, Strober M, Bergen AW, Kaye W, Szatkiewicz JP, Cormand B, Ramos-Quiroga JA, Sánchez-Mora C, Ribasés M, Casas M, Hervas A, Arranz MJ, Haavik J, Zayats T, Johansson S, Williams N, Dempfle A, Rothenberger A, Kuntsi J, Oades RD, Banaschewski T, Franke B, Buitelaar JK, Arias Vasquez A, Doyle AE, Reif A, Lesch KP, Freitag C, Rivero O, Palmason H, Romanos M, Langley K, Rietschel M, Witt SH, Dalsgaard S, Børglum AD, Waldman I, Wilmot B, Molly N, Bau CHD, Crosbie J, Schachar R, Loo SK, McGough JJ, Grevet EH, Medland SE, Robinson E, Weiss LA, Bacchelli E, Bailey A, Bal V, Battaglia A, Betancur C, Bolton P, Cantor R, Celestino-Soper P, Dawson G, De Rubeis S, Duque F, Green A, Klauck SM, Leboyer M, Levitt P, Maestrini E, Mane S, De-Luca DM, Parr J, Regan R, Reichenberg A, Sandin S, Vorstman J, Wassink T, Wijsman E, Cook E, Santangelo S, Delorme R, Rogé B, Magalhaes T, Arking D, Schulze TG, Thompson RC, Strohmaier J, Matthews K, Melle I, Morris D, Blackwood D, McIntosh A, Bergen SE, Schalling M, Jamain S, Maaser A, Fischer SB, Reinbold CS, Fullerton JM, Guzman-Parra J, Mayoral F, Schofield PR, Cichon S, Mühleisen TW, Degenhardt F, Schumacher J, Bauer M, Mitchell PB, Gershon ES, Rice J, Potash JB, Zandi PP, Craddock N, Ferrier IN, Alda M, Rouleau GA, Turecki G, Ophoff R, Pato C, Anjorin A, Stahl E, Leber M, Czerski PM, Cruceanu C, Jones IR, Posthuma D, Andlauer TFM, Forstner AJ, Streit F, Baune BT, Air T, Sinnamon G, Wray NR, MacIntyre DJ, Porteous D, Homuth G, Rivera M, Grove J, Middeldorp CM, Hickie I, Pergadia M, Mehta D, Smit JH, Jansen R, de Geus E, Dunn E, Li QS, Nauck M, Schoevers RA, Beekman AT, Knowles JA, Viktorin A, Arnold P, Barr CL, Bedoya-Berrio G, Bienvenu OJ, Brentani H, Burton C, Camarena B, Cappi C, Cath D, Cavallini M, Cusi D, Darrow S, Denys D, Derks EM, Dietrich A, Fernandez T, Figee M, Freimer N, Gerber G, Grados M, Greenberg E, Hanna GL, Hartmann A, Hirschtritt ME, Hoekstra PJ, Huang A, Huyser C, Illmann C, Jenike M, Kuperman S, Leventhal B, Lochner C, Lyon GJ, Macciardi F, Madruga-Garrido M, Malaty IA, Maras A, McGrath L, Miguel EC, Mir P, Nestadt G, Nicolini H, Okun MS, Pakstis A, Paschou P, Piacentini J, Pittenger C, Plessen K, Ramensky V, Ramos EM, Reus V, Richter MA, Riddle MA, Robertson MM, Roessner V, Rosário M, Samuels JF, Sandor P, Stein DJ, Tsetsos F, Van Nieuwerburgh F, Weatherall S, Wendland JR, Wolanczyk T, Worbe Y, Zai G, Goes FS, McLaughlin N, Nestadt PS, Grabe HJ, Depienne C, Konkashbaev A, Lanzagorta N, Valencia-Duarte A, Bramon E, Buccola N, Cahn W, Cairns M, Chong SA, Cohen D, Crespo-Facorro B, Crowley J, Davidson M, DeLisi L, Dinan T, Donohoe G, Drapeau E, Duan J, Haan L, Hougaard D, Karachanak-Yankova S, Khrunin A, Klovins J, Kučinskis V, Lee Chee Keong J, Limborska S, Loughland C, Lönngqvist J, Maher B, Mattheisen M, McDonald C, Murphy KC, Nenadic I, van Os J, Pantelis C, Pato M, Petryshen T, Quested D, Roussos P, Sanders AR, Schall U, Schwab SG, Sim K, So HC, Stögmann E,

- Subramaniam M, Toncheva D, Waddington J, Walters J, Weiser M, Cheng W, Cloninger R, Curtis D, Gejman PV, Henskens F, Mattingsdal M, Oh SY, Scott R, Webb B, Breen G, Churchhouse C, Bulik CM, Daly M, Dichgans M, Faraone SV, Guerreiro R, Holmans P, Kendler KS, Koeleman B, Mathews CA, Price A, Scharf J, Sklar P, Williams J, Wood NW, Cotsapas C, Palotie A, Smoller JW, Sullivan P, Rosand J, Corvin A, Neale BM, Schott JM, Anney R, Elia J, Grigoriou-Serbanescu M, Edenberg HJ, Murray R. Analysis of shared heritability in common disorders of the brain. *Science*. 2018 Jun 22;360(6395). pii: eaap8757. doi: 10.1126/science.aap8757. PubMed PMID: 29930110; PubMed Central PMCID: PMC6097237.
- Breen G, Li Q, Roth BL, O'Donnell P, Didriksen M, Dolmetsch R, O'Reilly PF, Gaspar HA, Manji H, Huebel C, Kelsoe JR, Malhotra D, Bertolino A, Posthuma D, Sklar P, Kapur S, Sullivan PF, Collier DA, Edenberg HJ. Translating genome-wide association findings into new therapeutics for psychiatry. *Nat Neurosci*. 2016 Oct 26;19(11):1392-1396. doi: 10.1038/nn.4411. Review. PubMed PMID: 27786187; PubMed Central PMCID: PMC5676453.
- Bromet EJ, Kotov R, Fochtmann LJ, Carlson GA, Tanenberg-Karant M, Ruggero C, Chang SW. Diagnostic shifts during the decade following first admission for psychosis. *Am J Psychiatry*. 2011 Nov;168(11):1186-94. doi: 10.1176/appi.ajp.2011.11010048. Epub 2011 Jun 15. PubMed PMID: 21676994; PubMed Central PMCID: PMC3589618.
- Browning BL, Browning SR. Genotype Imputation with Millions of Reference Samples. *Am J Hum Genet*. 2016 Jan 7;98(1):116-26. doi: 10.1016/j.ajhg.2015.11.020. PubMed PMID: 26748515; PubMed Central PMCID: PMC4716681.
- Browning BL, Zhou Y, Browning SR. A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am J Hum Genet*. 2018 Sep 6;103(3):338-348. doi: 10.1016/j.ajhg.2018.07.015. Epub 2018 Aug 9. PubMed PMID: 30100085; PubMed Central PMCID: PMC6128308.
- Brugha T, Bebbington P, Tennant C, Hurry J. The List of Threatening Experiences: a subset of 12 life event categories with considerable long-term contextual threat. *Psychol Med*. 1985 Feb;15(1):189-94. PubMed PMID: 3991833.
- Buchanan AV, Weiss KM, Fullerton SM. Dissecting complex disease: the quest for the Philosopher's Stone? *Int J Epidemiol*. 2006 Jun;35(3):562-71. Epub 2006 Mar 15. Review. PubMed PMID: 16540539.
- Buckholtz JW, Meyer-Lindenberg A. Psychopathology and the human connectome: toward a transdiagnostic model of risk for mental illness. *Neuron*. 2012 Jun 21;74(6):990-1004. doi: 10.1016/j.neuron.2012.06.002. Review. PubMed PMID: 22726830.

- Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J; Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson N, Daly MJ, Price AL, Neale BM. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.* 2015 Mar;47(3):291-5. doi: 10.1038/ng.3211. Epub 2015 Feb 2. PubMed PMID: 25642630; PubMed Central PMCID: PMC4495769.
- Burmeister M, McInnis MG, Zöllner S. Psychiatric genetics: progress amid controversy. *Nat Rev Genet.* 2008 Jul;9(7):527-40. doi: 10.1038/nrg2381. Review. PubMed PMID: 18560438.
- Burton H, Chowdhury S, Dent T, Hall A, Pashayan N, Pharoah P. Public health implications from COGS and potential for risk stratification and screening. *Nat Genet.* 2013 Apr;45(4):349-51. doi: 10.1038/ng.2582. PubMed PMID: 23535723.
- Bush WS, Moore JH. Chapter 11: Genome-wide association studies. *PLoS Comput Biol.* 2012;8(12):e1002822. doi: 10.1371/journal.pcbi.1002822. Epub 2012 Dec 27. Review. PubMed PMID: 23300413; PubMed Central PMCID: PMC3531285.
- Bustamante CD, Burchard EG, De la Vega FM. Genomics for the world. *Nature.* 2011 Jul 13;475(7355):163-5. doi: 10.1038/475163a. PubMed PMID: 21753830; PubMed Central PMCID: PMC3708540.
- Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, Cortes A, Welsh S, Young A, Effingham M, McVean G, Leslie S, Allen N, Donnelly P, Marchini J. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018 Oct;562(7726):203-209. doi: 10.1038/s41586-018-0579-z. Epub 2018 Oct 10. PubMed PMID: 30305743.
- Byers D. Components of phenotypic variance. *Nature Education.* 2008 1(1), 161.
- C Yuen RK, Merico D, Bookman M, L Howe J, Thiruvahindrapuram B, Patel RV, Whitney J, DeFlaux N, Bingham J, Wang Z, Pellicchia G, Buchanan JA, Walker S, Marshall CR, Uddin M, Zarrei M, Deneault E, D'Abate L, Chan AJ, Koyanagi S, Paton T, Pereira SL, Hoang N, Engchuan W, Higginbotham EJ, Ho K, Lamoureux S, Li W, MacDonald JR, Nalpathamkalam T, Sung WW, Tsoi FJ, Wei J, Xu L, Tasse AM, Kirby E, Van Etten W, Twigger S, Roberts W, Drmic I, Jilderda S, Modi BM, Kellam B, Szego M, Cytrynbaum C, Weksberg R, Zwaigenbaum L, Woodbury-Smith M, Brian J, Senman L, Iaboni A, Doyle-Thomas K, Thompson A, Chrysler C, Leef J, Savion-Lemieux T, Smith IM, Liu X, Nicolson R, Seifer V, Fedele A, Cook EH, Dager S, Estes A, Gallagher L, Malow BA, Parr JR, Spence SJ, Vorstman J, Frey BJ, Robinson JT, Strug LJ, Fernandez BA, Elsabbagh M, Carter MT, Hallmayer J, Knoppers BM, Anagnostou E, Szatmari P, Ring RH, Glazer D, Pletcher MT, Scherer SW. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat Neurosci.* 2017 Apr;20(4):602-611. doi: 10.1038/nn.4524.

- Epub 2017 Mar 6. PubMed PMID: 28263302; PubMed Central PMCID: PMC5501701.
- Capute AJ, Palmer FB. A pediatric overview of the spectrum of developmental disabilities. *J Dev Behav Pediatr.* 1980 Jun;1(2):66-9. PubMed PMID: 6164696.
- Cardon LR, Bell JI. Association study designs for complex diseases. *Nat Rev Genet.* 2001 Feb;2(2):91-9. Review. PubMed PMID: 11253062.
- Cardon LR, Palmer LJ. Population stratification and spurious allelic association. *Lancet.* 2003 Feb 15;361(9357):598-604. Review. PubMed PMID: 12598158.
- Carey CE, Agrawal A, Bucholz KK, Hartz SM, Lynskey MT, Nelson EC, Bierut LJ, Bogdan R. Associations between Polygenic Risk for Psychiatric Disorders and Substance Involvement. *Front Genet.* 2016 Aug 15;7:149. doi: 10.3389/fgene.2016.00149. eCollection 2016. PubMed PMID: 27574527; PubMed Central PMCID: PMC4983546.
- Carrera N, Arrojo M, Sanjuán J, Ramos-Ríos R, Paz E, Suárez-Rama JJ, Páramo M, Agra S, Brenlla J, Martínez S, Rivero O, Collier DA, Palotie A, Cichon S, Nöthen MM, Rietschel M, Rujescu D, Stefansson H, Steinberg S, Sigurdsson E, St Clair D, Tosato S, Werge T, Stefansson K, González JC, Valero J, Gutiérrez-Zotes A, Labad A, Martorell L, Vilella E, Carracedo Á, Costas J. Association study of nonsynonymous single nucleotide polymorphisms in schizophrenia. *Biol Psychiatry.* 2012 Jan 15;71(2):169-77. doi: 10.1016/j.biopsych.2011.09.032. Epub 2011 Nov 10. PubMed PMID: 22078303.
- Caspi A, Houts RM, Belsky DW, Goldman-Mellor SJ, Harrington H, Israel S, Meier MH, Ramrakha S, Shalev I, Poulton R, Moffitt TE. The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders? *Clin Psychol Sci.* 2014 Mar;2(2):119-137. PubMed PMID: 25360393; PubMed Central PMCID: PMC4209412.
- Caspi A, Moffitt TE. All for One and One for All: Mental Disorders in One Dimension. *Am J Psychiatry.* 2018 Sep 1;175(9):831-844. doi: 10.1176/appi.ajp.2018.17121383. Epub 2018 Apr 6. PubMed PMID: 29621902; PubMed Central PMCID: PMC6120790.
- Caspi A, Moffitt TE. Gene-environment interactions in psychiatry: joining forces with neuroscience. *Nat Rev Neurosci.* 2006 Jul;7(7):583-90. Review. PubMed PMID: 16791147.
- Catapano F, Perris F, Fabrazzo M, Cioffi V, Giacco D, De Santis V, Maj M. Obsessive-compulsive disorder with poor insight: a three-year prospective study. *Prog Neuropsychopharmacol Biol Psychiatry.* 2010 Mar 17;34(2):323-30. doi: 10.1016/j.pnpbp.2009.12.007. Epub 2009 Dec 16. PubMed PMID: 20015461.

- Caulfield T, Condit C. Science and the sources of hype. *Public Health Genomics*. 2012;15(3-4):209-17. doi: 10.1159/000336533. Epub 2012 Apr 4. PubMed PMID: 22488464.
- Chakravarti A. Population genetics--making sense out of sequence. *Nat Genet*. 1999 Jan;21(1 Suppl):56-60. Review. PubMed PMID: 9915503.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015 Feb 25;4:7. doi: 10.1186/s13742-015-0047-8. eCollection 2015. PubMed PMID: 25722852; PubMed Central PMCID: PMC4342193.
- Chatterjee N, Shi J, García-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat Rev Genet*. 2016 Jul;17(7):392-406. doi: 10.1038/nrg.2016.27. Epub 2016 May 3. Review. PubMed PMID: 27140283; PubMed Central PMCID: PMC6021129.
- Chen CY, Han J, Hunter DJ, Kraft P, Price AL. Explicit Modeling of Ancestry Improves Polygenic Risk Scores and BLUP Prediction. *Genet Epidemiol*. 2015 Sep;39(6):427-38. doi: 10.1002/gepi.21906. Epub 2015 May 21. PubMed PMID: 25995153; PubMed Central PMCID: PMC4734143.
- Chen Q, Ursini G, Romer AL, Knodt AR, Mezeivtch K, Xiao E, Pergola G, Blasi G, Straub RE, Callicott JH, Berman KF, Hariri AR, Bertolino A, Mattay VS, Weinberger DR. Schizophrenia polygenic risk score predicts mnemonic hippocampal activity. *Brain*. 2018 Apr 1;141(4):1218-1228. doi: 10.1093/brain/awy004. PubMed PMID: 29415119; PubMed Central PMCID: PMC5888989.
- Chial H. Mendelian genetics: patterns of inheritance and single-gene disorders. *Nature Education*. 2008, 1(1), 63.
- Choi KW, Chen CY, Stein MB, Klimentidis YC, Wang MJ, Koenen KC, Smoller JW; Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. Assessment of Bidirectional Relationships Between Physical Activity and Depression Among Adults: A 2-Sample Mendelian Randomization Study. *JAMA Psychiatry*. 2019 Jan 23. doi: 10.1001/jamapsychiatry.2018.4175. [Epub ahead of print] PubMed PMID: 30673066; PubMed Central PMCID: PMC6450288.
- Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet*. 2010 Jun;11(6):415-25. doi: 10.1038/nrg2779. Review. PubMed PMID: 20479773.
- Clarke GM, Anderson CA, Pettersson FH, Cardon LR, Morris AP, Zondervan KT. Basic statistical analysis in genetic case-control studies. *Nat Protoc*. 2011

Feb;6(2):121-33. doi: 10.1038/nprot.2010.182. Epub 2011 Feb 3. PubMed PMID: 21293453; PubMed Central PMCID: PMC3154648.

Cloninger CR, Reich T, Yokoyama S. Genetic diversity, genome organization, and investigation of the etiology of psychiatric diseases. *Psychiatr Dev.* 1983 Autumn;1(3):225-46. Review. PubMed PMID: 6369309.

Collins AL, Kim Y, Sklar P; International Schizophrenia Consortium, O'Donovan MC, Sullivan PF. Hypothesis-driven candidate genes for schizophrenia compared to genome-wide association results. *Psychol Med.* 2012 Mar;42(3):607-16. doi: 10.1017/S0033291711001607. Epub 2011 Aug 19. PubMed PMID: 21854684; PubMed Central PMCID: PMC4188923.

Collins FS, Guyer MS, Charkravarti A. Variations on a theme: cataloging human DNA sequence variation. *Science.* 1997 Nov 28;278(5343):1580-1. PubMed PMID: 9411782.

Collins RL, Brand H, Redin CE, Hanscom C, Antolik C, Stone MR, Glessner JT, Mason T, Pregno G, Dorrani N, Mandrile G, Giachino D, Perrin D, Walsh C, Cipicchio M, Costello M, Stortchevoi A, An JY, Currall BB, Seabra CM, Ragavendran A, Margolin L, Martinez-Agosto JA, Lucente D, Levy B, Sanders SJ, Wapner RJ, Quintero-Rivera F, Kloosterman W, Talkowski ME. Defining the diverse spectrum of inversions, complex structural variation, and chromothripsis in the morbid human genome. *Genome Biol.* 2017 Mar 6;18(1):36. doi: 10.1186/s13059-017-1158-6. PubMed PMID: 28260531; PubMed Central PMCID: PMC5338099.

Comfort N. Genetic determinism rides again. *Nature.* 2018 Sep;561(7724):461-463. doi: 10.1038/d41586-018-06784-5. PubMed PMID: 30254357.

Compton WM, Guze SB. The neo-Kraepelinian revolution in psychiatric diagnosis. *Eur Arch Psychiatry Clin Neurosci.* 1995;245(4-5):196-201. Review. PubMed PMID: 7578281.

Conley RR, Kelly DL. Management of treatment resistance in schizophrenia. *Biol Psychiatry.* 2001 Dec 1;50(11):898-911. Review. PubMed PMID: 11743944.

CONVERGE consortium. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature.* 2015 Jul 30;523(7562):588-91. doi: 10.1038/nature14659. Epub 2015 Jul 15. PubMed PMID: 26176920; PubMed Central PMCID: PMC4522619.

Corominas J, Klein M, Zayats T, Rivero O, Ziegler GC, Pauper M, Neveling K, Poelmans G, Jansch C, Svirin E, Geissler J, Weber H, Reif A, Arias Vasquez A, Galesloot TE, Kiemenev LALM, Buitelaar JK, Ramos-Quiroga JA, Cormand B, Ribasés M, Hveem K, Gabrielsen ME, Hoffmann P, Cichon S, Haavik J, Johansson S, Jacob CP, Romanos M, Franke B, Lesch KP. Identification of ADHD

- risk genes in extended pedigrees by combining linkage analysis and whole-exome sequencing. *Mol Psychiatry*. 2018 Aug 16. doi: 10.1038/s41380-018-0210-6. [Epub ahead of print] PubMed PMID: 30116028.
- Coryell W. Diagnostic instability: how much is too much? *Am J Psychiatry*. 2011 Nov;168(11):1136-8. doi: 10.1176/appi.ajp.2011.11081191. PubMed PMID: 22193597.
- Costa,P.T.-McCrae,R.R.(1999): NEO PI-R MANUAL. NEO PI-R, Inventario de Personalidad NEO Revisado. NEO-FFI, Inventario NEO reducido de Cinco Factores. Madrid: TEA Ediciones.
- Costas J, Carrera N, Alonso P, Gurriarán X, Segalàs C, Real E, López-Solà C, Mas S, Gassó P, Domènech L, Morell M, Quintela I, Lázaro L, Menchón JM, Estivill X, Carracedo Á. Exon-focused genome-wide association study of obsessive-compulsive disorder and shared polygenic risk with schizophrenia. *Transl Psychiatry*. 2016 Mar 29;6:e768. doi: 10.1038/tp.2016.34. PubMed PMID: 27023174; PubMed Central PMCID: PMC4872458.
- Craddock N, Owen MJ. The beginning of the end for the Kraepelinian dichotomy. *Br J Psychiatry*. 2005 May;186:364-6. PubMed PMID: 15863738.
- Craddock N, Owen MJ. The Kraepelinian dichotomy - going, going... but still not gone. *Br J Psychiatry*. 2010 Feb;196(2):92-5. doi: 10.1192/bjp.bp.109.073429. PubMed PMID: 20118450; PubMed Central PMCID: PMC2815936.
- Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH, Mowry BJ, Thapar A, Goddard ME, Witte JS, Absher D, Agartz I, Akil H, Amin F, Andreassen OA, Anjorin A, Anney R, Anttila V, Arking DE, Asherson P, Azevedo MH, Backlund L, Badner JA, Bailey AJ, Banaschewski T, Barchas JD, Barnes MR, Barrett TB, Bass N, Battaglia A, Bauer M, Bayés M, Bellivier F, Bergen SE, Berrettini W, Betancur C, Bettecken T, Biederman J, Binder EB, Black DW, Blackwood DH, Bloss CS, Boehnke M, Boomsma DI, Breen G, Breuer R, Bruggeman R, Cormican P, Buccola NG, Buitelaar JK, Bunney WE, Buxbaum JD, Byerley WF, Byrne EM, Caesar S, Cahn W, Cantor RM, Casas M, Chakravarti A, Chambert K, Choudhury K, Cichon S, Cloninger CR, Collier DA, Cook EH, Coon H, Cormand B, Corvin A, Coryell WH, Craig DW, Craig IW, Crosbie J, Cuccaro ML, Curtis D, Czamara D, Datta S, Dawson G, Day R, De Geus EJ, Degenhardt F, Djurovic S, Donohoe GJ, Doyle AE, Duan J, Dudbridge F, Duketis E, Ebstein RP, Edenberg HJ, Elia J, Ennis S, Etain B, Fanous A, Farmer AE, Ferrier IN, Flickinger M, Fombonne E, Foroud T, Frank J, Franke B, Fraser C, Freedman R, Freimer NB, Freitag CM, Friedl M, Frisén L, Gallagher L, Gejman PV, Georgieva L, Gershon ES, Geschwind DH, Giegling I, Gill M, Gordon SD, Gordon-Smith K, Green EK, Greenwood TA, Grice DE, Gross M, Grozeva D, Guan W, Gurling H, De Haan L, Haines JL, Hakonarson H, Hallmayer J, Hamilton SP, Hamshere ML, Hansen

TF, Hartmann AM, Hautzinger M, Heath AC, Henders AK, Herms S, Hickie IB, Hipolito M, Hoefels S, Holmans PA, Holsboer F, Hoogendijk WJ, Hottenga JJ, Hultman CM, Hus V, Ingason A, Ising M, Jamain S, Jones EG, Jones I, Jones L, Tzeng JY, Kähler AK, Kahn RS, Kandaswamy R, Keller MC, Kennedy JL, Kenny E, Kent L, Kim Y, Kirov GK, Klauck SM, Klei L, Knowles JA, Kohli MA, Koller DL, Konte B, Korszun A, Krabbendam L, Krasucki R, Kuntsi J, Kwan P, Landén M, Långström N, Lathrop M, Lawrence J, Lawson WB, Leboyer M, Ledbetter DH, Lee PH, Lencz T, Lesch KP, Levinson DF, Lewis CM, Li J, Lichtenstein P, Lieberman JA, Lin DY, Linszen DH, Liu C, Lohoff FW, Loo SK, Lord C, Lowe JK, Lucae S, MacIntyre DJ, Madden PA, Maestrini E, Magnusson PK, Mahon PB, Maier W, Malhotra AK, Mane SM, Martin CL, Martin NG, Mattheisen M, Matthews K, Mattingsdal M, McCarroll SA, McGhee KA, McGough JJ, McGrath PJ, McGuffin P, McInnis MG, McIntosh A, McKinney R, McLean AW, McMahon FJ, McMahon WM, McQuillin A, Medeiros H, Medland SE, Meier S, Melle I, Meng F, Meyer J, Middeldorp CM, Middleton L, Milanova V, Miranda A, Monaco AP, Montgomery GW, Moran JL, Moreno-De-Luca D, Morken G, Morris DW, Morrow EM, Moskvina V, Muglia P, Mühleisen TW, Muir WJ, Müller-Myhsok B, Murtha M, Myers RM, Myin-Germeys I, Neale MC, Nelson SF, Nievergelt CM, Nikolov I, Nimgaonkar V, Nolen WA, Nöthen MM, Nurnberger JI, Nwulia EA, Nyholt DR, O'Dushlaine C, Oades RD, Olincy A, Oliveira G, Olsen L, Ophoff RA, Osby U, Owen MJ, Palotie A, Parr JR, Paterson AD, Pato CN, Pato MT, Penninx BW, Pergadia ML, Pericak-Vance MA, Pickard BS, Pimm J, Piven J, Posthuma D, Potash JB, Poustka F, Propping P, Puri V, Queded DJ, Quinn EM, Ramos-Quiroga JA, Rasmussen HB, Raychaudhuri S, Rehnström K, Reif A, Ribasés M, Rice JP, Rietschel M, Roeder K, Roeyers H, Rossin L, Rothenberger A, Rouleau G, Ruderfer D, Rujescu D, Sanders AR, Sanders SJ, Santangelo SL, Sergeant JA, Schachar R, Schalling M, Schatzberg AF, Scheftner WA, Schellenberg GD, Scherer SW, Schork NJ, Schulze TG, Schumacher J, Schwarz M, Scolnick E, Scott LJ, Shi J, Shilling PD, Shyn SI, Silverman JM, Slager SL, Smalley SL, Smit JH, Smith EN, Sonuga-Barke EJ, St Clair D, State M, Steffens M, Steinhausen HC, Strauss JS, Strohmaier J, Stroup TS, Sutcliffe JS, Szatmari P, Szelinger S, Thirumalai S, Thompson RC, Todorov AA, Tozzi F, Treutlein J, Uhr M, van den Oord EJ, Van Grootheest G, Van Os J, Vicente AM, Vieland VJ, Vincent JB, Visscher PM, Walsh CA, Wassink TH, Watson SJ, Weissman MM, Werge T, Wienker TF, Wijsman EM, Willemsen G, Williams N, Willsey AJ, Witt SH, Xu W, Young AH, Yu TW, Zammit S, Zandi PP, Zhang P, Zitman FG, Zöllner S, Devlin B, Kelsoe JR, Sklar P, Daly MJ, O'Donovan MC, Craddock N, Sullivan PF, Smoller JW, Kendler KS, Wray NR; International Inflammatory Bowel Disease Genetics Consortium (IBDGC). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet.* 2013 Sep;45(9):984-94. doi: 10.1038/ng.2711. Epub 2013 Aug 11. PubMed PMID: 23933821; PubMed Central PMCID: PMC3800159.

- Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet*. 2013 Apr 20;381(9875):1371-1379. doi: 10.1016/S0140-6736(12)62129-1. Epub 2013 Feb 28. Erratum in: *Lancet*. 2013 Apr 20;381(9875):1360. Erratum in: *Lancet*. 2013 Apr 20;381(9875):1360. PubMed PMID: 23453885; PubMed Central PMCID: PMC3714010.
- Cross-Disorder Phenotype Group of the Psychiatric GWAS Consortium, Craddock N, Kendler K, Neale M, Nurnberger J, Purcell S, Rietschel M, Perlis R, Santangelo SL, Schulze TG, Smoller JW, Thapar A. Dissecting the phenotype in genome-wide association studies of psychiatric illness. *Br J Psychiatry*. 2009 Aug;195(2):97-9. doi: 10.1192/bjp.bp.108.063156. Erratum in: *Br J Psychiatry*. 2009 Oct;195(4):371. Schulze, Thomas [corrected to Schulze, Thomas G]. PubMed PMID: 19648536; PubMed Central PMCID: PMC4739810.
- Crow TJ. How and why genetic linkage has not solved the problem of psychosis: review and hypothesis. *Am J Psychiatry*. 2007 Jan;164(1):13-21. Review. PubMed PMID: 17202538.
- Daetwyler HD, Villanueva B, Woolliams JA. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One*. 2008;3(10):e3395. doi: 10.1371/journal.pone.0003395. Epub 2008 Oct 14. PubMed PMID: 18852893; PubMed Central PMCID: PMC2561058.
- Dalvie S, Koen N, Duncan L, Abbo C, Akena D, Atwoli L, Chiliza B, Donald KA, Kinyanda E, Lochner C, Mall S, Nakasujja N, Newton CR, Ramesar R, Sibeko G, Teferra S, Stein DJ, Koenen KC. Large Scale Genetic Research on Neuropsychiatric Disorders in African Populations is Needed. *EBioMedicine*. 2015 Oct 9;2(10):1259-61. doi: 10.1016/j.ebiom.2015.10.002. eCollection 2015 Oct. PubMed PMID: 26629498; PubMed Central PMCID: PMC4634677.
- Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, Schlessinger D, Stambolian D, Loh PR, Iacono WG, Swaroop A, Scott LJ, Cucca F, Kronenberg F, Boehnke M, Abecasis GR, Fuchsberger C. Next-generation genotype imputation service and methods. *Nat Genet*. 2016 Oct;48(10):1284-1287. doi: 10.1038/ng.3656. Epub 2016 Aug 29. PubMed PMID: 27571263; PubMed Central PMCID: PMC5157836.
- Davis J, Eyre H, Jacka FN, Dodd S, Dean O, McEwen S, Debnath M, McGrath J, Maes M, Amminger P, McGorry PD, Pantelis C, Berk M. A review of vulnerability and risks for schizophrenia: Beyond the two hit hypothesis. *Neurosci Biobehav Rev*. 2016 Jun;65:185-94. doi: 10.1016/j.neubiorev.2016.03.017. Epub 2016 Apr 9. Review. PubMed PMID: 27073049; PubMed Central PMCID: PMC4876729.

- Davis LK, Yu D, Keenan CL, Gamazon ER, Konkashbaev AI, Derks EM, Neale BM, Yang J, Lee SH, Evans P, Barr CL, Bellodi L, Benarroch F, Berrio GB, Bienvenu OJ, Bloch MH, Blom RM, Bruun RD, Budman CL, Camarena B, Campbell D, Cappi C, Cardona Silgado JC, Cath DC, Cavallini MC, Chavira DA, Chouinard S, Conti DV, Cook EH, Coric V, Cullen BA, Deforce D, Delorme R, Dion Y, Edlund CK, Egberts K, Falkai P, Fernandez TV, Gallagher PJ, Garrido H, Geller D, Girard SL, Grabe HJ, Grados MA, Greenberg BD, Gross-Tsur V, Haddad S, Heiman GA, Hemmings SM, Hounie AG, Illmann C, Jankovic J, Jenike MA, Kennedy JL, King RA, Kremeyer B, Kurlan R, Lanzagorta N, Leboyer M, Leckman JF, Lennertz L, Liu C, Lochner C, Lowe TL, Macciardi F, McCracken JT, McGrath LM, Mesa Restrepo SC, Moessner R, Morgan J, Muller H, Murphy DL, Naarden AL, Ochoa WC, Ophoff RA, Osiecki L, Pakstis AJ, Pato MT, Pato CN, Piacentini J, Pittenger C, Pollak Y, Rauch SL, Renner TJ, Reus VI, Richter MA, Riddle MA, Robertson MM, Romero R, Rosàrio MC, Rosenberg D, Rouleau GA, Ruhrmann S, Ruiz-Linares A, Sampaio AS, Samuels J, Sandor P, Sheppard B, Singer HS, Smit JH, Stein DJ, Strengman E, Tischfield JA, Valencia Duarte AV, Vallada H, Van Nieuwerburgh F, Veenstra-Vanderweele J, Walitza S, Wang Y, Wendland JR, Westenberg HG, Shugart YY, Miguel EC, McMahon W, Wagner M, Nicolini H, Posthuma D, Hanna GL, Heutink P, Denys D, Arnold PD, Oostra BA, Nestadt G, Freimer NB, Pauls DL, Wray NR, Stewart SE, Mathews CA, Knowles JA, Cox NJ, Scharf JM. Partitioning the heritability of Tourette syndrome and obsessive compulsive disorder reveals differences in genetic architecture. *PLoS Genet.* 2013 Oct;9(10):e1003864. doi: 10.1371/journal.pgen.1003864. Epub 2013 Oct 24. PubMed PMID: 24204291; PubMed Central PMCID: PMC3812053.
- dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 2001 Jande Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. *Nat Genet.* 2005 Nov;37(11):1217-23. Epub 2005 Oct 23. PubMed PMID: 16244653.
- de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. *Nat Genet.* 2005 Nov;37(11):1217-23. Epub 2005 Oct 23. PubMed PMID: 16244653.
- de Haan L, Dudek-Hodge C, Verhoeven Y, Denys D. Prevalence of psychotic disorders in patients with obsessive-compulsive disorder. *CNS Spectr.* 2009 Aug;14(8):415-7. PubMed PMID: 19890235.
- De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, Kou Y, Liu L, Fromer M, Walker S, Singh T, Klei L, Kosmicki J, Shih-Chen F, Aleksic B, Biscaldi M, Bolton PF, Brownfeld JM, Cai J, Campbell NG, Carracedo A, Chahrour MH, Chiocchetti AG, Coon H, Crawford EL, Curran SR, Dawson G, Duketis E, Fernandez BA, Gallagher L, Geller E, Guter SJ, Hill RS, Ionita-Laza J, Jimenez Gonzalez P, Kilpinen H, Klauck SM, Kolevzon A, Lee I, Lei I, Lei J,

- Lehtimäki T, Lin CF, Ma'ayan A, Marshall CR, McInnes AL, Neale B, Owen MJ, Ozaki N, Parellada M, Parr JR, Purcell S, Puura K, Rajagopalan D, Rehnström K, Reichenberg A, Sabo A, Sachse M, Sanders SJ, Schafer C, Schulte-Rüther M, Skuse D, Stevens C, Szatmari P, Tammimies K, Valladares O, Voran A, Li-San W, Weiss LA, Willsey AJ, Yu TW, Yuen RK; DDD Study; Homozygosity Mapping Collaborative for Autism; UK10K Consortium, Cook EH, Freitag CM, Gill M, Hultman CM, Lehner T, Palotie A, Schellenberg GD, Sklar P, State MW, Sutcliffe JS, Walsh CA, Scherer SW, Zwick ME, Barrett JC, Cutler DJ, Roeder K, Devlin B, Daly MJ, Buxbaum JD. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*. 2014 Nov 13;515(7526):209-15. doi: 10.1038/nature13772. Epub 2014 Oct 29. PubMed PMID: 25363760; PubMed Central PMCID: PMC4402723.
- De Vries H. *The mutation theory* (Vol. 1). 1909. Open Court Publishing Company.
- Deciphering Developmental Disorders Study. Prevalence and architecture of de novo mutations in developmental disorders. *Nature*. 2017 Feb 23;542(7642):433-438. doi: 10.1038/nature21062. Epub 2017 Jan 25. PubMed PMID: 28135719; PubMed Central PMCID: PMC6016744.
- Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods*. 2011 Dec 4;9(2):179-81. doi: 10.1038/nmeth.1785. PubMed PMID: 22138821.
- Demontis D, Lescai F, Børglum A, Glerup S, Østergaard SD, Mors O, Li Q, Liang J, Jiang H, Li Y, Wang J, Lesch KP, Reif A, Buitelaar JK, Franke B. Whole-Exome Sequencing Reveals Increased Burden of Rare Functional and Disruptive Variants in Candidate Risk Genes in Individuals With Persistent Attention-Deficit/Hyperactivity Disorder. *J Am Acad Child Adolesc Psychiatry*. 2016 Jun;55(6):521-3. doi: 10.1016/j.jaac.2016.03.009. PubMed PMID: 27238071.
- Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, Baldursson G, Belliveau R, Bybjerg-Grauholm J, Bækvad-Hansen M, Cerrato F, Chambert K, Churchhouse C, Dumont A, Eriksson N, Gandal M, Goldstein JI, Grasby KL, Grove J, Gudmundsson OO, Hansen CS, Hauberg ME, Hollegaard MV, Howrigan DP, Huang H, Maller JB, Martin AR, Martin NG, Moran J, Pallesen J, Palmer DS, Pedersen CB, Pedersen MG, Poterba T, Poulsen JB, Ripke S, Robinson EB, Satterstrom FK, Stefansson H, Stevens C, Turley P, Walters GB, Won H, Wright MJ; ADHD Working Group of the Psychiatric Genomics Consortium (PGC); Early Lifecourse & Genetic Epidemiology (EAGLE) Consortium; 23andMe Research Team, Andreassen OA, Asherson P, Burton CL, Boomsma DI, Cormand B, Dalsgaard S, Franke B, Gelernter J, Geschwind D, Hakonarson H, Haavik J, Kranzler HR, Kuntsi J, Langlely K, Lesch KP, Middeldorp C, Reif A, Rohde LA, Roussos P, Schachar R, Sklar P, Sonuga-Barke EJS, Sullivan PF, Thapar A, Tung JY, Waldman ID, Medland SE, Stefansson K,

- Nordentoft M, Hougaard DM, Werge T, Mors O, Mortensen PB, Daly MJ, Faraone SV, Børglum AD, Neale BM. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet.* 2019 Jan;51(1):63-75. doi: 10.1038/s41588-018-0269-7. Epub 2018 Nov 26. PubMed PMID: 30478444; PubMed Central PMCID: PMC6481311.
- Devlin B, Risch N. A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics.* 1995 Sep 20;29(2):311-22. PubMed PMID: 8666377.
- Dewan A, Liu M, Hartman S, Zhang SS, Liu DT, Zhao C, Tam PO, Chan WM, Lam DS, Snyder M, Barnstable C, Pang CP, Hoh J. HTRA1 promoter polymorphism in wet age-related macular degeneration. *Science.* 2006 Nov 10;314(5801):989-92. Epub 2006 Oct 19. PubMed PMID: 17053108.
- Dudbridge F. Polygenic Epidemiology. *Genet Epidemiol.* 2016 May;40(4):268-72. doi: 10.1002/gepi.21966. Epub 2016 Apr 7. Review. PubMed PMID: 27061411; PubMed Central PMCID: PMC4982028.
- Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* 2013 Mar;9(3):e1003348. doi: 10.1371/journal.pgen.1003348. Epub 2013 Mar 21. Erratum in: *PLoS Genet.* 2013 Apr;9(4). doi: 10.1371/annotation/b91ba224-10be-409d-93f4-7423d502cba0. PubMed PMID: 23555274; PubMed Central PMCID: PMC3605113.
- Duncan LE, Keller MC. A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *Am J Psychiatry.* 2011 Oct;168(10):1041-9. doi: 10.1176/appi.ajp.2011.11020191. Epub 2011 Sep 2. Review. PubMed PMID: 21890791; PubMed Central PMCID: PMC3222234.
- Duncan LE, Ratanatharathorn A, Aiello AE, Almlil LM, Amstadter AB, Ashley-Koch AE, Baker DG, Beckham JC, Bierut LJ, Bisson J, Bradley B, Chen CY, Dalvie S, Farrer LA, Galea S, Garrett ME, Gelernter JE, Guffanti G, Hauser MA, Johnson EO, Kessler RC, Kimbrel NA, King A, Koen N, Kranzler HR, Logue MW, Maihofer AX, Martin AR, Miller MW, Morey RA, Nugent NR, Rice JP, Ripke S, Roberts AL, Saccone NL, Smoller JW, Stein DJ, Stein MB, Sumner JA, Uddin M, Ursano RJ, Wildman DE, Yehuda R, Zhao H, Daly MJ, Liberzon I, Ressler KJ, Nievergelt CM, Koenen KC. Largest GWAS of PTSD (N=20 070) yields genetic overlap with schizophrenia and sex differences in heritability. *Mol Psychiatry.* 2018 Mar;23(3):666-673. doi: 10.1038/mp.2017.77. Epub 2017 Apr 25. PubMed PMID: 28439101; PubMed Central PMCID: PMC5696105.
- Dunn, O. J. Multiple comparisons among means. *Journal of the American Statistical.* 1961
- Eaton WW, Martins SS, Nestadt G, Bienvenu OJ, Clarke D, Alexandre P. The burden of mental disorders. *Epidemiol Rev.* 2008;30:1-14. doi: 10.1093/epirev/mxn011.

- Epub 2008 Sep 19. Review. PubMed PMID: 18806255; PubMed Central PMCID: PMC2683377.
- Edenberg HJ, Gelernter J, Agrawal A. Genetics of Alcoholism. *Curr Psychiatry Rep.* 2019 Mar 9;21(4):26. doi: 10.1007/s11920-019-1008-1. Review. PubMed PMID: 30852706.
- Edwards AC, Bigdeli TB, Docherty AR, Bacanu S, Lee D, de Candia TR, Moscati A, Thiselton DL, Maher BS, Wormley BK; Molecular Genetics of Schizophrenia Collaboration (MGS), Walsh D, O'Neill FA, Kendler KS, Riley BP, Fanous AH. Meta-analysis of Positive and Negative Symptoms Reveals Schizophrenia Modifier Genes. *Schizophr Bull.* 2016 Mar;42(2):279-87. doi: 10.1093/schbul/sbv119. Epub 2015 Aug 27. PubMed PMID: 26316594; PubMed Central PMCID: PMC4753595.
- Egeland JA, Gerhard DS, Pauls DL, Sussex JN, Kidd KK, Allen CR, Hostetter AM, Housman DE. Bipolar affective disorders linked to DNA markers on chromosome 11. *Nature.* 1987 Feb 26-Mar 4;325(6107):783-7. Review. PubMed PMID: 2881209.
- Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet.* 2010 Jun;11(6):446-50. doi: 10.1038/nrg2809. Review. PubMed PMID: 20479774; PubMed Central PMCID: PMC2942068.
- Elliott LT, Sharp K, Alfaro-Almagro F, Shi S, Miller KL, Douaud G, Marchini J, Smith SM. Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature.* 2018 Oct;562(7726):210-216. doi: 10.1038/s41586-018-0571-7. Epub 2018 Oct 10. PubMed PMID: 30305740.
- Emdin CA, Khera AV, Kathiresan S. Mendelian Randomization. *JAMA.* 2017 Nov 21;318(19):1925-1926. doi: 10.1001/jama.2017.17219. Review. PubMed PMID: 29164242.
- Encuesta sobre Alcohol y Drogas en España (EDADES) 2015-2016 Delegación del Gobierno del Plan Nacional sobre Drogas
- Erbe M, Hayes BJ, Matukumalli LK, Goswami S, Bowman PJ, Reich CM, Mason BA, Goddard ME. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *J Dairy Sci.* 2012 Jul;95(7):4114-29. doi: 10.3168/jds.2011-5019. Erratum in: *J Dairy Sci.* 2014 Oct;97(10):6622. PubMed PMID: 22720968.
- Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinformatics.* 2015 May 1;31(9):1466-8. doi: 10.1093/bioinformatics/btu848. Epub 2014 Dec 29. PubMed PMID: 25550326; PubMed Central PMCID: PMC4410663.

- Evans LM, Tahmasbi R, Vrieze SI, Abecasis GR, Das S, Gazal S, Bjelland DW, de Candia TR; Haplotype Reference Consortium, Goddard ME, Neale BM, Yang J, Visscher PM, Keller MC. Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. *Nat Genet.* 2018 May;50(5):737-745. doi: 10.1038/s41588-018-0108-x. Epub 2018 Apr 26. PubMed PMID: 29700474; PubMed Central PMCID: PMC5934350.
- Falconer DS, Mackay TFC. *Introduction to Quantitative Genetics.* 1996 4th Edition, Addison Wesley Longman, Harlow.
- Falconer, D. S. The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Annals of human genetics.* 1965. 29(1), 51-76.
- Fanous AH, Zhou B, Aggen SH, Bergen SE, Amdur RL, Duan J, Sanders AR, Shi J, Mowry BJ, Olincy A, Amin F, Cloninger CR, Silverman JM, Buccola NG, Byerley WF, Black DW, Freedman R, Dudbridge F, Holmans PA, Ripke S, Gejman PV, Kendler KS, Levinson DF; Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study of clinical dimensions of schizophrenia: polygenic effect on disorganized symptoms. *Am J Psychiatry.* 2012 Dec;169(12):1309-17. doi: 10.1176/appi.ajp.2012.12020218. PubMed PMID: 23212062; PubMed Central PMCID: PMC3646712.
- Faraone SV, Larsson H. Genetics of attention deficit hyperactivity disorder. *Mol Psychiatry.* 2019 Apr;24(4):562-575. doi: 10.1038/s41380-018-0070-0. Epub 2018 Jun 11. Review. PubMed PMID: 29892054.
- Farrell MS, Werge T, Sklar P, Owen MJ, Ophoff RA, O'Donovan MC, Corvin A, Cichon S, Sullivan PF. Evaluating historical candidate genes for schizophrenia. *Mol Psychiatry.* 2015 May;20(5):555-62. doi: 10.1038/mp.2015.16. Epub 2015 Mar 10. Review. PubMed PMID: 25754081; PubMed Central PMCID: PMC4414705.
- Feighner JP, Robins E, Guze SB, Woodruff RA Jr, Winokur G, Munoz R. Diagnostic criteria for use in psychiatric research. *Arch Gen Psychiatry.* 1972 Jan;26(1):57-63. PubMed PMID: 5009428.
- Feinstein AR. THE PRE-THERAPEUTIC CLASSIFICATION OF CO-MORBIDITY IN CHRONIC DISEASE. *J Chronic Dis.* 1970 Dec;23(7):455-68. PubMed PMID: 26309916.
- Fennig S, Kovaszny B, Rich C, Ram R, Pato C, Miller A, Rubinstein J, Carlson G, Schwartz JE, Phelan J, et al. Six-month stability of psychiatric diagnoses in first-admission patients with psychosis. *Am J Psychiatry.* 1994 Aug;151(8):1200-8. PubMed PMID: 8037256.

- Fernandez-Egea E, Worbe Y, Bernardo M, Robbins TW. Distinct risk factors for obsessive and compulsive symptoms in chronic schizophrenia. *Psychol Med.* 2018 Dec;48(16):2668-2675. doi: 10.1017/S003329171800017X. Epub 2018 Feb 19. PubMed PMID: 29455687; PubMed Central PMCID: PMC6236440.
- Ferreira MA, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L, Fan J, Kirov G, Perlis RH, Green EK, Smoller JW, Grozeva D, Stone J, Nikolov I, Chambert K, Hamshere ML, Nimgaonkar VL, Moskvina V, Thase ME, Caesar S, Sachs GS, Franklin J, Gordon-Smith K, Ardlie KG, Gabriel SB, Fraser C, Blumenstiel B, Defelice M, Breen G, Gill M, Morris DW, Elkin A, Muir WJ, McGhee KA, Williamson R, MacIntyre DJ, MacLean AW, St CD, Robinson M, Van Beck M, Pereira AC, Kandaswamy R, McQuillin A, Collier DA, Bass NJ, Young AH, Lawrence J, Ferrier IN, Anjorin A, Farmer A, Curtis D, Scolnick EM, McGuffin P, Daly MJ, Corvin AP, Holmans PA, Blackwood DH, Gurling HM, Owen MJ, Purcell SM, Sklar P, Craddock N; Wellcome Trust Case Control Consortium. Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet.* 2008 Sep;40(9):1056-8. doi: 10.1038/ng.209. PubMed PMID: 18711365; PubMed Central PMCID: PMC2703780
- First, Gibbon, Spitzer, Williams, & Benjamin. *Structured Clinical Interview for DSM-IV (SCID-I/SCID-II)*; 1997
- First MB, Pincus HA, Levine JB, Williams JB, Ustun B, Peele R. Clinical utility as a criterion for revising psychiatric diagnoses. *Am J Psychiatry.* 2004 Jun;161(6):946-54. Review. PubMed PMID: 15169680.
- Fisher, RA. "XV. The correlation between relatives on the supposition of Mendelian inheritance." *Earth and Environmental Science Transactions of the Royal Society of Edinburgh* . 1919 52.2: 399-433.
- Frances A. A report card on the utility of psychiatric diagnosis. *World Psychiatry.* 2016 Feb;15(1):32-3. doi: 10.1002/wps.20285. PubMed PMID: 26833602; PubMed Central PMCID: PMC4780303.
- Frank J, Lang M, Witt SH, Strohmaier J, Rujescu D, Cichon S, Degenhardt F, Nöthen MM, Collier DA, Ripke S, Naber D, Rietschel M. Identification of increased genetic risk scores for schizophrenia in treatment-resistant patients. *Mol Psychiatry.* 2015 Feb;20(2):150-1. doi: 10.1038/mp.2014.56. Epub 2014 Jun 3. PubMed PMID: 24888364; PubMed Central PMCID: PMC4356742.
- Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, Gabriel SB, Topol EJ, Smoller JW, Pato CN, Pato MT, Petryshen TL, Kolonel LN, Lander ES, Sklar P, Henderson B, Hirschhorn JN, Altshuler D. Assessing the impact of population stratification on genetic association studies. *Nat Genet.* 2004 Apr;36(4):388-93. Epub 2004 Mar 28. PubMed PMID: 15052270.

- Freitag CM. The genetics of autistic disorders and its clinical relevance: a review of the literature. *Mol Psychiatry*. 2007 Jan;12(1):2-22. Epub 2006 Oct 10. Review. PubMed PMID: 17033636.
- Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P, Georgieva L, Rees E, Palta P, Ruderfer DM, Carrera N, Humphreys I, Johnson JS, Roussos P, Barker DD, Banks E, Milanova V, Grant SG, Hannon E, Rose SA, Chambert K, Mahajan M, Scolnick EM, Moran JL, Kirov G, Palotie A, McCarroll SA, Holmans P, Sklar P, Owen MJ, Purcell SM, O'Donovan MC. De novo mutations in schizophrenia implicate synaptic networks. *Nature*. 2014 Feb 13;506(7487):179-84. doi: 10.1038/nature12929. Epub 2014 Jan 22. PubMed PMID: 24463507; PubMed Central PMCID: PMC4237002.
- Fuchsberger C, Abecasis GR, Hinds DA. minimac2: faster genotype imputation. *Bioinformatics*. 2015 Mar 1;31(5):782-4. doi: 10.1093/bioinformatics/btu704. Epub 2014 Oct 22. PubMed PMID: 25338720; PubMed Central PMCID: PMC4341061.
- Fusar-Poli P, Cappucciati M, Rutigliano G, Heslin M, Stahl D, Brittenden Z, Caverzasi E, McGuire P, Carpenter WT. Diagnostic Stability of ICD/DSM First Episode Psychosis Diagnoses: Meta-analysis. *Schizophr Bull*. 2016 Nov;42(6):1395-1406. doi: 10.1093/schbul/sbw020. Epub 2016 Mar 15. PubMed PMID: 26980142; PubMed Central PMCID: PMC5049518.
- Gale CR, Hagenaars SP, Davies G, Hill WD, Liewald DC, Cullen B, Penninx BW; International Consortium for Blood Pressure GWAS, CHARGE Consortium Aging and Longevity Group, Boomsma DI, Pell J, McIntosh AM, Smith DJ, Deary IJ, Harris SE. Pleiotropy between neuroticism and physical and mental health: findings from 108 038 men and women in UK Biobank. *Transl Psychiatry*. 2016 Apr 26;6:e791. doi: 10.1038/tp.2016.56. PubMed PMID: 27115122; PubMed Central PMCID: PMC4872414.
- Gandal MJ, Haney JR, Parikshak NN, Leppa V, Ramaswami G, Hartl C, Schork AJ, Appadurai V, Buil A, Werge TM, Liu C, White KP; CommonMind Consortium; PsychENCODE Consortium; iPSYCH-BROAD Working Group, Horvath S, Geschwind DH. Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science*. 2018 Feb 9;359(6376):693-697. doi: 10.1126/science.aad6469. PubMed PMID: 29439242; PubMed Central PMCID: PMC5898828.
- Gandal MJ, Leppa V, Won H, Parikshak NN, Geschwind DH. The road to precision psychiatry: translating genetics into disease mechanisms. *Nat Neurosci*. 2016 Oct 26;19(11):1397-1407. doi: 10.1038/nn.4409. Review. PubMed PMID: 27786179.
- García-González J, Tansey KE, Hauser J, Henigsberg N, Maier W, Mors O, Placentino A, Rietschel M, Souery D, Žagar T, Czernski PM, Jerman B,

- Buttenschön HN, Schulze TG, Zobel A, Farmer A, Aitchison KJ, Craig I, McGuffin P, Giupponi M, Perroud N, Bondolfi G, Evans D, O'Donovan M, Peters TJ, Wendland JR, Lewis G, Kapur S, Perlis R, Arolt V, Domschke K; Major Depressive Disorder Working Group of the Psychiatric Genomic Consortium, Breen G, Curtis C, Sang-Hyuk L, Kan C, Newhouse S, Patel H, Baune BT, Uher R, Lewis CM, Fabbri C. Pharmacogenetics of antidepressant response: A polygenic approach. *Prog Neuropsychopharmacol Biol Psychiatry*. 2017 Apr 3;75:128-134. doi: 10.1016/j.pnpbp.2017.01.011. Epub 2017 Jan 31. PubMed PMID: 28159590.
- Gaudet MM, Kuchenbaecker KB, Vijai J, Klein RJ, Kirchoff T, McGuffog L, Barrowdale D, Dunning AM, Lee A, Dennis J, Healey S, Dicks E, Soucy P, Sinilnikova OM, Pankratz VS, Wang X, Eldridge RC, Tessler DC, Vincent D, Bacot F, Hogervorst FB, Peock S, Stoppa-Lyonnet D; KConFab Investigators, Peterlongo P, Schmutzler RK, Nathanson KL, Piedmonte M, Singer CF, Thomassen M; Ontario Cancer Genetics Network, Hansen Tv, Neuhausen SL, Blanco I, Greene MH, Garber J, Weitzel JN, Andrulis IL, Goldgar DE, D'Andrea E, Caldes T, Nevanlinna H, Osorio A, van Rensburg EJ, Arason A, Rennert G, van den Ouweland AM, van der Hout AH, Kets CM, Aalfs CM, Wijnen JT, Ausems MG; HEBON; EMBRACE, Frost D, Ellis S, Fineberg E, Platte R, Evans DG, Jacobs C, Adlard J, Tischkowitz M, Porteous ME, Damiola F; GEMO Study Collaborators, Golmard L, Barjhoux L, Longy M, Belotti M, Ferrer SF, Mazoyer S, Spurdle AB, Manoukian S, Barile M, Genuardi M, Arnold N, Meindl A, Sutter C, Wappenschmidt B, Domchek SM, Pfeiler G, Friedman E, Jensen UB, Robson M, Shah S, Lazaro C, Mai PL, Benitez J, Southey MC, Schmidt MK, Fasching PA, Peto J, Humphreys MK, Wang Q, Michailidou K, Sawyer EJ, Burwinkel B, Guénel P, Bojesen SE, Milne RL, Brenner H, Lochmann M; GENICA Network, Aittomäki K, Dörk T, Margolin S, Mannermaa A, Lambrechts D, Chang-Claude J, Radice P, Giles GG, Haiman CA, Winqvist R, Devilee P, García-Closas M, Schoof N, Hooning MJ, Cox A, Pharoah PD, Jakubowska A, Orr N, González-Neira A, Pita G, Alonso MR, Hall P, Couch FJ, Simard J, Altshuler D, Easton DF, Chenevix-Trench G, Antoniou AC, Offit K. Identification of a BRCA2-specific modifier locus at 6p24 related to breast cancer risk. *PLoS Genet*. 2013;9(3):e1003173. doi: 10.1371/journal.pgen.1003173. Epub 2013 Mar 27. PubMed PMID: 23544012; PubMed Central PMCID: PMC3609647.
- GB, Xu X; Autism Spectrum Disorder Working Group of the Psychiatric Genomics Consortium; BUPGEN; Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium; 23andMe Research Team, Stefansson K, Geschwind DH, Nordentoft M, Hougaard DM, Werge T, Mors O, Mortensen PB, Neale BM, Daly MJ, Børglum AD. Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet*. 2019 Mar;51(3):431-444. doi: 10.1038/s41588-019-0344-8. Epub 2019 Feb 25. PubMed PMID: 30804558; PubMed Central PMCID: PMC6454898.

- Ge T, Nichols TE, Lee PH, Holmes AJ, Roffman JL, Buckner RL, Sabuncu MR, Smoller JW. Massively expedited genome-wide heritability analysis (MEGHA). *Proc Natl Acad Sci U S A*. 2015 Feb 24;112(8):2479-84. doi: 10.1073/pnas.1415603112. Epub 2015 Feb 9. PubMed PMID: 25675487; PubMed Central PMCID: PMC4345618.
- Gelernter J. Genetics of complex traits in psychiatry. *Biol Psychiatry*. 2015 Jan 1;77(1):36-42. doi: 10.1016/j.biopsych.2014.08.005. Epub 2014 Aug 13. Review. PubMed PMID: 25444161; PubMed Central PMCID: PMC4282183.
- Genovese G, Fromer M, Stahl EA, Ruderfer DM, Chambert K, Landén M, Moran JL, Purcell SM, Sklar P, Sullivan PF, Hultman CM, McCarroll SA. Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat Neurosci*. 2016 Nov;19(11):1433-1441. doi: 10.1038/nn.4402. Epub 2016 Oct 3. PubMed PMID: 27694994; PubMed Central PMCID: PMC5104192.
- Gershon ES. Marker genotyping errors in old data on X-linkage in bipolar illness. *Biol Psychiatry*. 1991 Apr 1;29(7):721-9. PubMed PMID: 1888383.
- Geschwind DH, Flint J. Genetics and genomics of psychiatric disease. *Science*. 2015 Sep 25;349(6255):1489-94. doi: 10.1126/science.aaa8954. Epub 2015 Sep 24. Review. PubMed PMID: 26404826; PubMed Central PMCID: PMC4694563.
- Gillberg C. Chromosomal disorders and autism. *J Autism Dev Disord*. 1998 Oct;28(5):415-25. Review. PubMed PMID: 9813777.
- Girirajan S, Rosenfeld JA, Coe BP, Parikh S, Friedman N, Goldstein A, Filipink RA, McConnell JS, Angle B, Meschino WS, Nezarati MM, Asamoah A, Jackson KE, Gowans GC, Martin JA, Carmany EP, Stockton DW, Schnur RE, Penney LS, Martin DM, Raskin S, Leppig K, Thiess H, Smith R, Aberg E, Niyazov DM, Escobar LF, El-Khechen D, Johnson KD, Lebel RR, Siefkas K, Ball S, Shur N, McGuire M, Brasington CK, Spence JE, Martin LS, Clericuzio C, Ballif BC, Shaffer LG, Eichler EE. Phenotypic heterogeneity of genomic disorders and rare copy-number variants. *N Engl J Med*. 2012 Oct 4;367(14):1321-31. doi: 10.1056/NEJMoa1200395. Epub 2012 Sep 12. Erratum in: *N Engl J Med*. 2012 Dec 13;367(24):2362. PubMed PMID: 22970919; PubMed Central PMCID: PMC3494411.
- Gizer IR, Ficks C, Waldman ID. Candidate gene studies of ADHD: a meta-analytic review. *Hum Genet*. 2009 Jul;126(1):51-90. doi: 10.1007/s00439-009-0694-x. Epub 2009 Jun 9. Review. PubMed PMID: 19506906. 336: Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev*. 2009 Jun;19(3):212-9. doi: 10.1016/j.gde.2009.04.010. Epub 2009 May 28. Review. PubMed PMID: 19481926; PubMed Central PMCID: PMC2914559.

- Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*. 2015 Aug 22;386(9995):743-800. doi: 10.1016/S0140-6736(15)60692-4. Epub 2015 Jun 7. Review. PubMed PMID: 26063472; PubMed Central PMCID: PMC4561509.
- Goes FS, Pirooznia M, Parla JS, Kramer M, Ghiban E, Mavruk S, Chen YC, Monson ET, Willour VL, Karchin R, Flickinger M, Locke AE, Levy SE, Scott LJ, Boehnke M, Stahl E, Moran JL, Hultman CM, Landén M, Purcell SM, Sklar P, Zandi PP, McCombie WR, Potash JB. Exome Sequencing of Familial Bipolar Disorder. *JAMA Psychiatry*. 2016 Jun 1;73(6):590-7. doi: 10.1001/jamapsychiatry.2016.0251. PubMed PMID: 27120077; PubMed Central PMCID: PMC5600716.
- Goldman D, Oroszi G, Ducci F. The genetics of addictions: uncovering the genes. *Nat Rev Genet*. 2005 Jul;6(7):521-32. Review. PubMed PMID: 15995696.
- Goldstein DB, Chikhi L. Human migrations and population structure: what we know and why it matters. *Annu Rev Genomics Hum Genet*. 2002;3:129-52. Epub 2002 Apr 15. Review. PubMed PMID: 12142358.
- Goldstein DB. Common genetic variation and human traits. *N Engl J Med*. 2009 Apr 23;360(17):1696-8. doi: 10.1056/NEJMp0806284. Epub 2009 Apr 15. PubMed PMID: 19369660.
- Gottesman II, Bertelsen A. Legacy of German psychiatric genetics: hindsight is always 20/20. *Am J Med Genet*. 1996 Jul 26;67(4):317-22. Review. PubMed PMID: 8837696.
- Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry*. 2003 Apr;160(4):636-45. Review. PubMed PMID: 12668349.
- Gottesman, I. I., & Shields, J. Schizophrenia and genetics. A twin study vantage point. In *Acad. Aress*, New York, NY. 1972.
- Green EK, Grozeva D, Jones I, Jones L, Kirov G, Caesar S, Gordon-Smith K, Fraser C, Forty L, Russell E, Hamshere ML, Moskvina V, Nikolov I, Farmer A, McGuffin P; Wellcome Trust Case Control Consortium, Holmans PA, Owen MJ, O'Donovan MC, Craddock N. The bipolar disorder risk allele at CACNA1C also confers risk of recurrent major depression and of schizophrenia. *Mol Psychiatry*. 2010 Oct;15(10):1016-22. doi: 10.1038/mp.2009.49. Epub 2009 Jul 21. PubMed PMID: 19621016; PubMed Central PMCID: PMC3011210.
- Green EK, Rees E, Walters JT, Smith KG, Forty L, Grozeva D, Moran JL, Sklar P, Ripke S, Chambert KD, Genovese G, McCarroll SA, Jones I, Jones L, Owen MJ,

- O'Donovan MC, Craddock N, Kirov G. Copy number variation in bipolar disorder. *Mol Psychiatry*. 2016 Jan;21(1):89-93. doi: 10.1038/mp.2014.174. Epub 2015 Jan 6. PubMed PMID: 25560756; PubMed Central PMCID: PMC5038134.
- Gronowski AM, Scott RT Jr, Caplan AL, Nelson LJ. The ethical implications of preimplantation genetic diagnosis. *Clin Chem*. 2014 Jan;60(1):25-8. doi: 10.1373/clinchem.2013.202515. Epub 2013 Oct 2. PubMed PMID: 24089455.
- Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, Pallesen J, Agerbo E, Andreassen OA, Anney R, Awashti S, Belliveau R, Bettella F, Buxbaum JD, Bybjerg-Grauholm J, Bækvad-Hansen M, Cerrato F, Chambert K, Christensen JH, Churchhouse C, Dellenvall K, Demontis D, De Rubeis S, Devlin B, Djurovic S, Dumont AL, Goldstein JI, Hansen CS, Hauberg ME, Hollegaard MV, Hope S, Howrigan DP, Huang H, Hultman CM, Klei L, Maller J, Martin J, Martin AR, Moran JL, Nyegaard M, Nærland T, Palmer DS, Palotie A, Pedersen CB, Pedersen MG, dPoterba T, Poulsen JB, Pourcain BS, Qvist P, Rehnström K, Reichenberg A, Reichert J, Robinson EB, Roeder K, Roussos P, Saemundsen E, Sandin S, Satterstrom FK, Davey Smith G, Stefansson H, Steinberg S, Stevens CR, Sullivan PF, Turley P, Walters
- GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015 May 8;348(6235):648-60. doi: 10.1126/science.1262110. Epub 2015 May 7. PubMed PMID: 25954001; PubMed Central PMCID: PMC4547484.
- GTEx Consortium; Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group; Enhancing GTEx (eGTEx) groups; NIH Common Fund; NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source Site—NDRI; Biospecimen Collection Source Site—RPCI; Biospecimen Core Resource—VARI; Brain Bank Repository—University of Miami Brain Endowment Bank; Leidos Biomedical—Project Management; ELSI Study; Genome Browser Data Integration & Visualization—EBI; Genome Browser Data Integration & Visualization—UCSC Genomics Institute, University of California Santa Cruz; Lead analysts; Laboratory, Data Analysis & Coordinating Center (LDACC); NIH program management; Biospecimen collection; Pathology; eQTL manuscript working group; Battle A, Brown CD, Engelhardt BE, Montgomery SB. Genetic effects on gene expression across human tissues. *Nature*. 2017 Oct 11;550(7675):204-213. doi: 10.1038/nature24277. Erratum in: *Nature*. 2017 Dec 20;.. PubMed PMID: 29022597; PubMed Central PMCID: PMC5776756.
- Gu W, Zhang F, Lupski JR. Mechanisms for human genomic rearrangements. *Pathogenetics*. 2008 Nov 3;1(1):4. doi: 10.1186/1755-8417-1-4. PubMed PMID: 19014668; PubMed Central PMCID: PMC2583991.

- Guloksuz S, van Os J, Rutten BPF. The Exposome Paradigm and the Complexities of Environmental Research in Psychiatry. *JAMA Psychiatry*. 2018 Oct 1;75(10):985-986. doi: 10.1001/jamapsychiatry.2018.1211. PubMed PMID: 29874362.
- Guo W, Samuels JF, Wang Y, Cao H, Ritter M, Nestadt PS, Krasnow J, Greenberg BD, Fyer AJ, McCracken JT, Geller DA, Murphy DL, Knowles JA, Grados MA, Riddle MA, Rasmussen SA, McLaughlin NC, Nurmi EL, Askland KD, Cullen BA, Piacentini J, Pauls DL, Bienvenu OJ, Stewart SE, Goes FS, Maher B, Pulver AE, Valle D, Mattheisen M, Qian J, Nestadt G, Shugart YY. Polygenic risk score and heritability estimates reveals a genetic relationship between ASD and OCD. *Eur Neuropsychopharmacol*. 2017 Jul;27(7):657-666. doi: 10.1016/j.euroneuro.2017.03.011. PubMed PMID: 28641744.
- Gurriarán X, Rodríguez-López J, Flórez G, Pereiro C, Fernández JM, Fariñas E, Estévez V, Arrojo M, Costas J; GenPol Study Group. Relationships between substance abuse/dependence and psychiatric disorders based on polygenic scores. *Genes Brain Behav*. 2019 Mar;18(3):e12504. doi: 10.1111/gbb.12504. Epub 2018 Jul 19. PubMed PMID: 29974660.
- H3Africa Consortium, Rotimi C, Abayomi A, Abimiku A, Adabayeri VM, Adebamowo C, Adebisi E, Ademola AD, Adeyemo A, Adu D, Affolabi D, Agongo G, Ajayi S, Akarolo-Anthony S, Akinyemi R, Akpalu A, Alberts M, Alonso Betancourt O, Alzohairy AM, Ameni G, Amodu O, Anabwani G, Andersen K, Arogundade F, Arulogun O, Asogun D, Bakare R, Balde N, Baniecki ML, Beiswanger C, Benkahla A, Bethke L, Boehnke M, Boima V, Brandful J, Brooks AI, Brosius FC, Brown C, Bucheton B, Burke DT, Burnett BG, Carrington-Lawrence S, Carstens N, Chisi J, Christoffels A, Cooper R, Cordell H, Crowther N, Croxton T, de Vries J, Derr L, Donkor P, Doumbia S, Duncanson A, Ekem I, El Sayed A, Engel ME, Enyaru JC, Everett D, Fadlelmola FM, Fakunle E, Fischbeck KH, Fischer A, Folarin O, Gamielien J, Garry RF, Gaseitsiwe S, Gbadegesin R, Ghansah A, Giovanni M, Goesbeck P, Gomez-Olive FX, Grant DS, Grewal R, Guyer M, Hanchard NA, Happi CT, Hazelhurst S, Hennig BJ, Hertz C, Fowler, Hide W, Hilderbrandt F, Hugo-Hamman C, Ibrahim ME, James R, Jaufeerally-Fakim Y, Jenkins C, Jentsch U, Jiang PP, Joloba M, Jongeneel V, Joubert F, Kader M, Kahn K, Kaleebu P, Kapiga SH, Kassim SK, Kasvosve I, Kayondo J, Keavney B, Kekitiinwa A, Khan SH, Kimmel P, King MC, Kleta R, Koffi M, Kopp J, Kretzler M, Kumuthini J, Kyobe S, Kyobutungi C, Lackland DT, Lacourciere KA, Landouré G, Lawlor R, Lehner T, Lesosky M, Levitt N, Littler K, Lombard Z, Loring JF, Lyantagaye S, Macleod A, Madden EB, Mahomva CR, Makani J, Mamven M, Marape M, Mardon G, Marshall P, Martin DP, Masiga D, Mason R, Mate-Kole M, Matovu E, Mayige M, Mayosi BM, Mbanya JC, McCurdy SA, McCarthy MI, McIlleron H, McLigeyo SO, Merle C, Mocumbi AO, Mondo C, Moran JV, Motala A, Moxey-Mims M, Mpoloka WS, Msefula CL, Mthiyane T, Mulder N, Mulugeta Gh, Mumba D, Musuku J, Nagdee M, Nash O, Ndiaye D, Nguyen AQ, Nicol M, Nkomazana O, Norris S, Nsangi B,

- Nyarko A, Nyirenda M, Obe E, Obiakor R, Oduro A, Ofori-Acquah SF, Ogah O, Ogendo S, Ohene-Frempong K, Ojo A, Olanrewaju T, Oli J, Osafo C, Ouwe Missi Oukem-Boyer O, Ovbiagele B, Owen A, Owolabi MO, Owolabi L, Owusu-Dabo E, Pare G, Parekh R, Patterton HG, Penno MB, Peterson J, Pieper R, Plange-Rhule J, Pollak M, Puzak J, Ramesar RS, Ramsay M, Rasooly R, Reddy S, Sabeti PC, Sagoe K, Salako T, Samassékou O, Sandhu MS, Sankoh O, Sarfo FS, Sarr M, Shaboodien G, Sidibe I, Simo G, Simuunza M, Smeeth L, Sobngwi E, Soodyall H, Sorgho H, Sow Bah O, Srinivasan S, Stein DJ, Susser ES, Swanepoel C, Tangwa G, Tareila A, Tastan Bishop O, Tayo B, Tiffin N, Tinto H, Tobin E, Tollman SM, Traoré M, Treadwell MJ, Troyer J, Tsimako-Johnstone M, Tukei V, Ulasi I, Ulenga N, van Rooyen B, Wachinou AP, Waddy SP, Wade A, Wayengera M, Whitworth J, Wideroff L, Winkler CA, Winnicki S, Wonkam A, Yewondwos M, sen T, Yozwiak N, Zar H. Research capacity. Enabling the genomic revolution in Africa. *Science*. 2014 Jun 20;344(6190):1346-8. doi: 10.1126/science.1251546. PubMed PMID: 24948725; PubMed Central PMCID: PMC4138491.
- Hakulinen C, Elovainio M, Batty GD, Virtanen M, Kivimäki M, Jokela M. Personality and alcohol consumption: Pooled analysis of 72,949 adults from eight cohort studies. *Drug Alcohol Depend*. 2015 Jun 1;151:110-4. doi: 10.1016/j.drugalcdep.2015.03.008. Epub 2015 Mar 18. PubMed PMID: 25823906; PubMed Central PMCID: PMC4447572.
- Hakulinen C, Jokela M. Alcohol use and personality trait change: pooled analysis of six cohort studies. *Psychol Med*. 2019 Jan;49(2):224-231. doi: 10.1017/S0033291718000636. Epub 2018 Mar 15. PubMed PMID: 29540247.
- Hamer D. Genetics. Rethinking behavior genetics. *Science*. 2002 Oct 4;298(5591):71-2. PubMed PMID: 12364769.
- Hancock DB, Markunas CA, Bierut LJ, Johnson EO. Human Genetics of Addiction: New Insights and Future Directions. *Curr Psychiatry Rep*. 2018 Mar 5;20(2):8. doi: 10.1007/s11920-018-0873-3. Review. PubMed PMID: 29504045; PubMed Central PMCID: PMC5983372.
- Hartz SM, Horton AC, Oehlert M, Carey CE, Agrawal A, Bogdan R, Chen LS, Hancock DB, Johnson EO, Pato CN, Pato MT, Rice JP, Bierut LJ. Association Between Substance Use Disorder and Polygenic Liability to Schizophrenia. *Biol Psychiatry*. 2017 Nov 15;82(10):709-715. doi: 10.1016/j.biopsych.2017.04.020. Epub 2017 Jun 6. PubMed PMID: 28739213; PubMed Central PMCID: PMC5643224.
- Hattersley AT, McCarthy MI. What makes a good genetic association study? *Lancet*. 2005 Oct 8;366(9493):1315-23. PubMed PMID: 16214603.
- Hatzimanolis A, Bhatnagar P, Moes A, Wang R, Roussos P, Bitsios P, Stefanis CN, Pulver AE, Arking DE, Smyrnis N, Stefanis NC, Avramopoulos D. Common

- genetic variation and schizophrenia polygenic risk influence neurocognitive performance in young adulthood. *Am J Med Genet B Neuropsychiatr Genet.* 2015 Jul;168B(5):392-401. doi: 10.1002/ajmg.b.32323. Epub 2015 May 12. PubMed PMID: 25963331; PubMed Central PMCID: PMC5008149.
- Hellenthal G, Busby GBJ, Band G, Wilson JF, Capelli C, Falush D, Myers S. A genetic atlas of human admixture history. *Science.* 2014 Feb 14;343(6172):747-751. doi: 10.1126/science.1243518. PubMed PMID: 24531965; PubMed Central PMCID: PMC4209567.
- Hellenthal G, Busby GBJ, Band G, Wilson JF, Capelli C, Falush D, Myers S. A genetic atlas of human admixture history. *Science.* 2014 Feb 14;343(6172):747-751. doi: 10.1126/science.1243518. PubMed PMID: 24531965; PubMed Central PMCID: PMC4209567.
- Helles A, Gillberg IC, Gillberg C, Billstedt E. Asperger syndrome in males over two decades: Quality of life in relation to diagnostic stability and psychiatric comorbidity. *Autism.* 2017 May;21(4):458-469. doi: 10.1177/1362361316650090. Epub 2016 May 26. PubMed PMID: 27233289.
- Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum Mol Genet.* 2018 Aug 1;27(R2):R195-R208. doi: 10.1093/hmg/ddy163. PubMed PMID: 29771313; PubMed Central PMCID: PMC6061876.
- Hemminki K, Försti A, Bermejo JL. The 'common disease-common variant' hypothesis and familial risks. *PLoS One.* 2008 Jun 18;3(6):e2504. doi: 10.1371/journal.pone.0002504. PubMed PMID: 18560565; PubMed Central PMCID: PMC2423486.
- Heslin M, Lomas B, Lappin JM, Donoghue K, Reininghaus U, Onyejiaka A, Croudace T, Jones PB, Murray RM, Fearon P, Dazzan P, Morgan C, Doody GA. Diagnostic change 10 years after a first episode of psychosis. *Psychol Med.* 2015 Oct;45(13):2757-69. doi: 10.1017/S0033291715000720. Epub 2015 May 4. PubMed PMID: 25936425; PubMed Central PMCID: PMC4595854.
- Heston LL. Psychiatric disorders in foster home reared children of schizophrenic mothers. *Br J Psychiatry.* 1966 Aug;112(489):819-25. PubMed PMID: 5966555.
- Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet.* 2005 Feb;6(2):95-108. Review. PubMed PMID: 15716906.
- Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. *Genet Med.* 2002 Mar-Apr;4(2):45-61. Review. PubMed PMID: 11882781.

- Hoge SK, Appelbaum PS. Ethics and neuropsychiatric genetics: a review of major issues. *Int J Neuropsychopharmacol*. 2012 Nov;15(10):1547-57. doi: 10.1017/S1461145711001982. Epub 2012 Jan 25. Review. PubMed PMID: 22272758; PubMed Central PMCID: PMC3359421.
- Horwitz T, Lam K, Chen Y, Xia Y, Liu C. A decade in psychiatric GWAS research. *Mol Psychiatry*. 2019 Mar;24(3):378-389. doi: 10.1038/s41380-018-0055-z. Epub 2018 Jun 25. Review. PubMed PMID: 29942042; PubMed Central PMCID: PMC6372350.
- Howard DM, Adams MJ, Clarke TK, Hafferty JD, Gibson J, Shiralil M, Coleman JRI, Hagenaars SP, Ward J, Wigmore EM, Alloza C, Shen X, Barbu MC, Xu EY, Whalley HC, Marioni RE, Porteous DJ, Davies G, Deary IJ, Hemani G, Berger K, Teismann H, Rawal R, Arolt V, Baune BT, Dannlowski U, Domschke K, Tian C, Hinds DA; 23andMe Research Team; Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, Trzaskowski M, Byrne EM, Ripke S, Smith DJ, Sullivan PF, Wray NR, Breen G, Lewis CM, McIntosh AM. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci*. 2019 Mar;22(3):343-352. doi: 10.1038/s41593-018-0326-7. Epub 2019 Feb 4. PubMed PMID: 30718901.
- Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet*. 2012 Jul 22;44(8):955-9. doi: 10.1038/ng.2354. PubMed PMID: 22820512; PubMed Central PMCID: PMC3696580.
- Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009 Jun;5(6):e1000529. doi: 10.1371/journal.pgen.1000529. Epub 2009 Jun 19. PubMed PMID: 19543373; PubMed Central PMCID: PMC2689936.
- Hsu L, Jeon J, Brenner H, Gruber SB, Schoen RE, Berndt SI, Chan AT, Chang-Claude J, Du M, Gong J, Harrison TA, Hayes RB, Hoffmeister M, Hutter CM, Lin Y, Nishihara R, Ogino S, Prentice RL, Schumacher FR, Seminara D, Slattery ML, Thomas DC, Thornquist M, Newcomb PA, Potter JD, Zheng Y, White E, Peters U; Colorectal Transdisciplinary (CORECT) Study; Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO). A model to determine colorectal cancer risk using common genetic susceptibility loci. *Gastroenterology*. 2015 Jun;148(7):1330-9.e14. doi: 10.1053/j.gastro.2015.02.010. Epub 2015 Feb 13. PubMed PMID: 25683114; PubMed Central PMCID: PMC4446193.
- Hyman SE. The diagnosis of mental disorders: the problem of reification. *Annu Rev Clin Psychol*. 2010;6:155-79. Review. PubMed PMID: 17716032.

- Ikeda M, Aleksic B, Kinoshita Y, Okochi T, Kawashima K, Kushima I, Ito Y, Nakamura Y, Kishi T, Okumura T, Fukuo Y, Williams HJ, Hamshere ML, Ivanov D, Inada T, Suzuki M, Hashimoto R, Ujike H, Takeda M, Craddock N, Kaibuchi K, Owen MJ, Ozaki N, O'Donovan MC, Iwata N. Genome-wide association study of schizophrenia in a Japanese population. *Biol Psychiatry*. 2011 Mar 1;69(5):472-8. doi: 10.1016/j.biopsych.2010.07.010. Epub 2010 Sep 15. PubMed PMID: 20832056.
- Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, Sanislow C, Wang P. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry*. 2010 Jul;167(7):748-51. doi: 10.1176/appi.ajp.2010.09091379. PubMed PMID: 20595427.
- Insel TR. The NIMH Research Domain Criteria (RDoC) Project: precision medicine for psychiatry. *Am J Psychiatry*. 2014 Apr;171(4):395-7. doi: 10.1176/appi.ajp.2014.14020138. PubMed PMID: 24687194.
- International HapMap 3 Consortium, Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, Peltonen L, Dermitzakis E, Bonnen PE, Altshuler DM, Gibbs RA, de Bakker PI, Deloukas P, Gabriel SB, Gwilliam R, Hunt S, Inouye M, Jia X, Palotie A, Parkin M, Whittaker P, Yu F, Chang K, Hawes A, Lewis LR, Ren Y, Wheeler D, Gibbs RA, Muzny DM, Barnes C, Darvishi K, Hurler M, Korn JM, Kristiansson K, Lee C, McCarroll SA, Nemesh J, Dermitzakis E, Keinan A, Montgomery SB, Pollack S, Price AL, Soranzo N, Bonnen PE, Gibbs RA, Gonzaga-Jauregui C, Keinan A, Price AL, Yu F, Anttila V, Brodeur W, Daly MJ, Leslie S, McVean G, Moutsianas L, Nguyen H, Schaffner SF, Zhang Q, Ghorri MJ, McGinnis R, McLaren W, Pollack S, Price AL, Schaffner SF, Takeuchi F, Grossman SR, Shlyakhter I, Hostetter EB, Sabeti PC, Adebamowo CA, Foster MW, Gordon DR, Licinio J, Manca MC, Marshall PA, Matsuda I, Ngare D, Wang VO, Reddy D, Rotimi CN, Royal CD, Sharp RR, Zeng C, Brooks LD, McEwen JE. Integrating common and rare genetic variation in diverse human populations. *Nature*. 2010 Sep 2;467(7311):52-8. doi: 10.1038/nature09298. PubMed PMID: 20811451; PubMed Central PMCID: PMC3173859.
- International HapMap 3 Consortium, Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, Peltonen L, Dermitzakis E, Bonnen PE, Altshuler DM, Gibbs RA, de Bakker PI, Deloukas P, Gabriel SB, Gwilliam R, Hunt S, Inouye M, Jia X, Palotie A, Parkin M, Whittaker P, Yu F, Chang K, Hawes A, Lewis LR, Ren Y, Wheeler D, Gibbs RA, Muzny DM, Barnes C, Darvishi K, Hurler M, Korn JM, Kristiansson K, Lee C, McCarroll SA, Nemesh J, Dermitzakis E, Keinan A, Montgomery SB, Pollack S, Price AL, Soranzo N, Bonnen PE, Gibbs RA, Gonzaga-Jauregui C, Keinan A, Price AL, Yu F, Anttila V, Brodeur W, Daly MJ, Leslie S, McVean G, Moutsianas L, Nguyen H, Schaffner SF, Zhang Q, Ghorri MJ, McGinnis R, McLaren W, Pollack S, Price

AL, Schaffner SF, Takeuchi F, Grossman SR, Shlyakhter I, Hostetter EB, Sabeti PC, Adebamowo CA, Foster MW, Gordon DR, Licinio J, Manca MC, Marshall PA, Matsuda I, Ngare D, Wang VO, Reddy D, Rotimi CN, Royal CD, Sharp RR, Zeng C, Brooks LD, McEwen JE. Integrating common and rare genetic variation in diverse human populations. *Nature*. 2010 Sep 2;467(7311):52-8. doi: 10.1038/nature09298. PubMed PMID: 20811451; PubMed Central PMCID: PMC3173859.

International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Wayne MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallée C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archevêque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J. A second generation human haplotype map of over 3.1 million SNPs. *Nature*. 2007

- Oct 18;449(7164):851-61. PubMed PMID: 17943122; PubMed Central PMCID: PMC2689609.
- International HapMap Consortium. A haplotype map of the human genome. *Nature*. 2005 Oct 27;437(7063):1299-320. PubMed PMID: 16255080; PubMed Central PMCID: PMC1880871.
- International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*. 2008 Sep 11;455(7210):237-41. doi: 10.1038/nature07239. Epub 2008 Jul 30. PubMed PMID: 18668038; PubMed Central PMCID: PMC3912847.
- Ioannidis JP, Boffetta P, Little J, O'Brien TR, Uitterlinden AG, Vineis P, Balding DJ, Chokkalingam A, Dolan SM, Flanders WD, Higgins JP, McCarthy ML, McDermott DH, Page GP, Rebbeck TR, Seminara D, Khoury MJ. Assessment of cumulative evidence on genetic associations: interim guidelines. *Int J Epidemiol*. 2008 Feb;37(1):120-32. Epub 2007 Sep 26. Review. PubMed PMID: 17898028.
- Ioannidis JP, Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis DG. Genetic associations in large versus small studies: an empirical assessment. *Lancet*. 2003 Feb 15;361(9357):567-71. Review. PubMed PMID: 12598142.
- Ioannidis JP. Commentary: grading the credibility of molecular evidence for complex diseases. *Int J Epidemiol*. 2006 Jun;35(3):572-8; discussion 593-6. Epub 2006 Mar 15. Review. PubMed PMID: 16540537.
- Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, Yamrom B, Lee YH, Narzisi G, Leotta A, Kendall J, Grabowska E, Ma B, Marks S, Rodgers L, Stepansky A, Troge J, Andrews P, Bekritsky M, Pradhan K, Ghiban E, Kramer M, Parla J, Demeter R, Fulton LL, Fulton RS, Magrini VJ, Ye K, Darnell JC, Darnell RB, Mardis ER, Wilson RK, Schatz MC, McCombie WR, Wigler M. De novo gene disruptions in children on the autistic spectrum. *Neuron*. 2012 Apr 26;74(2):285-99. doi: 10.1016/j.neuron.2012.04.009. PubMed PMID: 22542183; PubMed Central PMCID: PMC3619976.
- Jablensky A. Psychiatric classifications: validity and utility. *World Psychiatry*. 2016 Feb;15(1):26-31. doi: 10.1002/wps.20284. PubMed PMID: 26833601; PubMed Central PMCID: PMC4780305.
- Jansen PR, Watanabe K, Stringer S, Skene N, Bryois J, Hammerschlag AR, de Leeuw CA, Benjamins JS, Muñoz-Manchado AB, Nagel M, Savage JE, Tiemeier H, White T; 23andMe Research Team, Tung JY, Hinds DA, Vacic V, Wang X, Sullivan PF, van der Sluis S, Polderman TJC, Smit AB, Hjerling-Leffler J, Van Someren EJW, Posthuma D. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat Genet*. 2019

Mar;51(3):394-403. doi: 10.1038/s41588-018-0333-3. Epub 2019 Feb 25. PubMed PMID: 30804565.

Johnson EC, Border R, Melroy-Greif WE, de Leeuw CA, Ehringer MA, Keller MC. No Evidence That Schizophrenia Candidate Genes Are More Associated With Schizophrenia Than Noncandidate Genes. *Biol Psychiatry*. 2017 Nov 15;82(10):702-708. doi: 10.1016/j.biopsych.2017.06.033. Epub 2017 Jul 13. PubMed PMID: 28823710; PubMed Central PMCID: PMC5643230.

Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA. Haplotype tagging for the identification of common disease genes. *Nat Genet*. 2001 Oct;29(2):233-7. PubMed PMID: 11586306.

Jones HJ, Stergiakouli E, Tansey KE, Hubbard L, Heron J, Cannon M, Holmans P, Lewis G, Linden DE, Jones PB, Davey Smith G, O'Donovan MC, Owen MJ, Walters JT, Zammit S. Phenotypic Manifestation of Genetic Risk for Schizophrenia During Adolescence in the General Population. *JAMA Psychiatry*. 2016 Mar;73(3):221-8. doi: 10.1001/jamapsychiatry.2016058. PubMed PMID: 26818099; PubMed Central PMCID: PMC5024747.

Kallman, F. *The Genetics of Schizophrenia*, Locust Valley, New York: J. J. Augustin. Inc., Publishers. 1938.

Kaminsky EB, Kaul V, Paschall J, Church DM, Bunke B, Kunig D, Moreno-De-Luca D, Moreno-De-Luca A, Mulle JG, Warren ST, Richard G, Compton JG, Fuller AE, Gliem TJ, Huang S, Collinson MN, Beal SJ, Ackley T, Pickering DL, Golden DM, Aston E, Whitby H, Shetty S, Rossi MR, Rudd MK, South ST, Brothman AR, Sanger WG, Iyer RK, Crolla JA, Thorland EC, Aradhya S, Ledbetter DH, Martin CL. An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. *Genet Med*. 2011 Sep;13(9):777-84. doi: 10.1097/GIM.0b013e31822c79f9. PubMed PMID: 21844811; PubMed Central PMCID: PMC3661946.

Kapur S, Phillips AG, Insel TR. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol Psychiatry*. 2012 Dec;17(12):1174-9. doi: 10.1038/mp.2012.105. Epub 2012 Aug 7. PubMed PMID: 22869033.

Kataoka M, Matoba N, Sawada T, Kazuno AA, Ishiwata M, Fujii K, Matsuo K, Takata A, Kato T. Exome sequencing for bipolar disorder points to roles of de novo loss-of-function and protein-altering mutations. *Mol Psychiatry*. 2016 Jul;21(7):885-93. doi: 10.1038/mp.2016.69. Epub 2016 May 24. PubMed PMID: 27217147; PubMed Central PMCID: PMC5414074.

- Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, Williamson D, Ryan N. Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry*. 1997 Jul;36(7):980-8. PubMed PMID: 9204677.
- Kauppi K, Westlye LT, Tesli M, Bettella F, Brandt CL, Mattingsdal M, Ueland T, Espeseth T, Agartz I, Melle I, Djurovic S, Andreassen OA. Polygenic risk for schizophrenia associated with working memory-related prefrontal brain activation in patients with schizophrenia and healthy controls. *Schizophr Bull*. 2015 May;41(3):736-43. doi: 10.1093/schbul/sbu152. Epub 2014 Nov 11. PubMed PMID: 25392519; PubMed Central PMCID: PMC4393689.
- Kelleher I, Connor D, Clarke MC, Devlin N, Harley M, Cannon M. Prevalence of psychotic symptoms in childhood and adolescence: a systematic review and meta-analysis of population-based studies. *Psychol Med*. 2012 Sep;42(9):1857-63. doi: 10.1017/S0033291711002960. Epub 2012 Jan 9. Review. PubMed PMID: 22225730.
- Kelsoe JR, Ginns EI, Egeland JA, Gerhard DS, Goldstein AM, Bale SJ, Pauls DL, Long RT, Kidd KK, Conte G, et al. Re-evaluation of the linkage relationship between chromosome 11p loci and the gene for bipolar affective disorder in the Old Order Amish. *Nature*. 1989 Nov 16;342(6247):238-43. PubMed PMID: 2682265.
- Kendell R, Jablensky A. Distinguishing between the validity and utility of psychiatric diagnoses. *Am J Psychiatry*. 2003 Jan;160(1):4-12. PubMed PMID: 12505793.
- Kendell RE. Clinical validity. *Psychol Med*. 1989 Feb;19(1):45-55. Review. PubMed PMID: 2657832.
- Kendler KS, First MB. Alternative futures for the DSM revision process: iteration v. paradigm shift. *Br J Psychiatry*. 2010 Oct;197(4):263-5. doi: 10.1192/bjp.bp.109.076794. PubMed PMID: 20884947.
- Kendler KS, Jacobson KC, Prescott CA, Neale MC. Specificity of genetic and environmental risk factors for use and abuse/dependence of cannabis, cocaine, hallucinogens, sedatives, stimulants, and opiates in male twins. *Am J Psychiatry*. 2003 Apr;160(4):687-95. PubMed PMID: 12668357.
- Kendler KS, Kessler RC, Walters EE, MacLean C, Neale MC, Heath AC, Eaves LJ. Stressful life events, genetic liability, and onset of an episode of major depression in women. *Am J Psychiatry*. 1995 Jun;152(6):833-42. PubMed PMID: 7755111.
- Kendler KS, Myers J, Prescott CA. Specificity of genetic and environmental risk factors for symptoms of cannabis, cocaine, alcohol, caffeine, and nicotine

- dependence. *Arch Gen Psychiatry*. 2007 Nov;64(11):1313-20. PubMed PMID: 17984400.
- Kendler KS. A joint history of the nature of genetic variation and the nature of schizophrenia. *Mol Psychiatry*. 2015 Feb;20(1):77-83. doi: 10.1038/mp.2014.94. Epub 2014 Aug 19. PubMed PMID: 25134695; PubMed Central PMCID: PMC4318712.
- Kendler KS. An historical framework for psychiatric nosology. *Psychol Med*. 2009 Dec;39(12):1935-41. doi: 10.1017/S0033291709005753. Epub 2009 Apr 16. Review. PubMed PMID: 19368761; PubMed Central PMCID: PMC2783473.
- Kendler KS. DSM issues: incorporation of biological tests, avoidance of reification, and an approach to the "box canyon problem". *Am J Psychiatry*. 2014 Dec 1;171(12):1248-50. doi: 10.1176/appi.ajp.2014.14081018. PubMed PMID: 25756765.
- Kendler KS. Hierarchy and heritability: the role of diagnosis and modeling in psychiatric genetics. *Am J Psychiatry*. 2002 Apr;159(4):515-8. PubMed PMID: 11925285.
- Kendler KS. The nosologic validity of paranoia (simple delusional disorder). A review. *Arch Gen Psychiatry*. 1980 Jun;37(6):699-706. PubMed PMID: 7387341.
- Kendler KS. Toward a scientific psychiatric nosology. Strengths and limitations. *Arch Gen Psychiatry*. 1990 Oct;47(10):969-73. PubMed PMID: 2222134.
- Kendler KS. Twin studies of psychiatric illness: an update. *Arch Gen Psychiatry*. 2001 Nov;58(11):1005-14. Review. PubMed PMID: 11695946.
- Kendler KS. What psychiatric genetics has taught us about the nature of psychiatric illness and what is left to learn. *Mol Psychiatry*. 2013 Oct;18(10):1058-66. doi: 10.1038/mp.2013.50. Epub 2013 Apr 30. Review. PubMed PMID: 23628988.
- Kennedy JL, Giuffra LA, Moises HW, Cavalli-Sforza LL, Pakstis AJ, Kidd JR, Castiglione CM, Sjogren B, Wetterberg L, Kidd KK. Evidence against linkage of schizophrenia to markers on chromosome 5 in a northern Swedish pedigree. *Nature*. 1988 Nov 10;336(6195):167-70. PubMed PMID: 2903450.
- Kessler RC, Avenevoli S, Costello EJ, Georgiades K, Green JG, Gruber MJ, He JP, Koretz D, McLaughlin KA, Petukhova M, Sampson NA, Zaslavsky AM, Merikangas KR. Prevalence, persistence, and sociodemographic correlates of DSM-IV disorders in the National Comorbidity Survey Replication Adolescent Supplement. *Arch Gen Psychiatry*. 2012 Apr;69(4):372-80. doi: 10.1001/archgenpsychiatry.2011.160. Epub 2011 Dec 5. PubMed PMID: 22147808; PubMed Central PMCID: PMC3445020.

- Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen HU, Kendler KS. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. *Arch Gen Psychiatry*. 1994 Jan;51(1):8-19. PubMed PMID: 8279933.
- Kessler RC, Ormel J, Petukhova M, McLaughlin KA, Green JG, Russo LJ, Stein DJ, Zaslavsky AM, Aguilar-Gaxiola S, Alonso J, Andrade L, Benjet C, de Girolamo G, de Graaf R, Demyttenaere K, Fayyad J, Haro JM, Hu Cy, Karam A, Lee S, Lepine JP, Matchsinger H, Mihaescu-Pintia C, Posada-Villa J, Sagar R, Ustün TB. Development of lifetime comorbidity in the World Health Organization world mental health surveys. *Arch Gen Psychiatry*. 2011 Jan;68(1):90-100. doi: 10.1001/archgenpsychiatry.2010.180. PubMed PMID: 21199968; PubMed Central PMCID: PMC3057480.
- Keyes KM, Eaton NR, Krueger RF, Skodol AE, Wall MM, Grant B, Siever LJ, Hasin DS. Thought disorder in the meta-structure of psychopathology. *Psychol Med*. 2013 Aug;43(8):1673-83. doi: 10.1017/S0033291712002292. Epub 2012 Nov 21. PubMed PMID: 23171498; PubMed Central PMCID: PMC3767418.
- Keyes KM, Hatzenbuehler ML, Hasin DS. Stressful life experiences, alcohol consumption, and alcohol use disorders: the epidemiologic evidence for four main types of stressors. *Psychopharmacology (Berl)*. 2011 Nov;218(1):1-17. doi: 10.1007/s00213-011-2236-1. Epub 2011 Mar 5. Review. PubMed PMID: 21373787; PubMed Central PMCID: PMC3755727.
- Khandaker GM, Zuber V, Rees JMB, Carvalho L, Mason AM, Foley CN, Gkatzionis A, Jones PB, Burgess S. Shared mechanisms between coronary heart disease and depression: findings from a large UK general population-based cohort. *Mol Psychiatry*. 2019 Mar 19. doi: 10.1038/s41380-019-0395-3. [Epub ahead of print] PubMed PMID: 30886334.
- Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander ES, Lubitz SA, Ellinor PT, Kathiresan S. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018 Sep;50(9):1219-1224. doi: 10.1038/s41588-018-0183-z. Epub 2018 Aug 13. PubMed PMID: 30104762; PubMed Central PMCID: PMC6128408.
- Khramtsova EA, Heldman R, Derks EM, Yu D; Tourette Syndrome/Obsessive-Compulsive Disorder Working Group of the Psychiatric Genomics Consortium, Davis LK, Stranger BE. Sex differences in the genetic architecture of obsessive-compulsive disorder. *Am J Med Genet B Neuropsychiatr Genet*. 2018 Nov 20. doi: 10.1002/ajmg.b.32687. [Epub ahead of print] PubMed PMID: 30456828.
- Kim MS, Patel KP, Teng AK, Berens AJ, Lachance J. Genetic disease risks can be misestimated across global populations. *Genome Biol*. 2018 Nov 14;19(1):179.

doi: 10.1186/s13059-018-1561-7. PubMed PMID: 30424772; PubMed Central PMCID: PMC6234640.

Kirov G, Gumus D, Chen W, Norton N, Georgieva L, Sari M, O'Donovan MC, Erdogan F, Owen MJ, Ropers HH, Ullmann R. Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. *Hum Mol Genet.* 2008 Feb 1;17(3):458-65. Epub 2007 Nov 6. PubMed PMID: 17989066.

Kirov G, Pocklington AJ, Holmans P, Ivanov D, Ikeda M, Ruderfer D, Moran J, Chambert K, Toncheva D, Georgieva L, Grozeva D, Fjodorova M, Wollerton R, Rees E, Nikolov I, van de Lagemaat LN, Bayés A, Fernandez E, Olason PI, Böttcher Y, Komiyama NH, Collins MO, Choudhary J, Stefansson K, Stefansson H, Grant SG, Purcell S, Sklar P, O'Donovan MC, Owen MJ. De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol Psychiatry.* 2012 Feb;17(2):142-53. doi: 10.1038/mp.2011.154. Epub 2011 Nov 15. PubMed PMID: 22083728; PubMed Central PMCID: PMC3603134.

Kirov G, Rees E, Walters JT, Escott-Price V, Georgieva L, Richards AL, Chambert KD, Davies G, Legge SE, Moran JL, McCarroll SA, O'Donovan MC, Owen MJ. The penetrance of copy number variations for schizophrenia and developmental delay. *Biol Psychiatry.* 2014 Mar 1;75(5):378-85. doi: 10.1016/j.biopsych.2013.07.022. Epub 2013 Aug 28. PubMed PMID: 23992924; PubMed Central PMCID: PMC4229045.

Kirov G. CNVs in neuropsychiatric disorders. *Hum Mol Genet.* 2015 Oct 15;24(R1):R45-9. doi: 10.1093/hmg/ddv253. Epub 2015 Jun 30. Review. PubMed PMID: 26130694.

Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J. Complement factor H polymorphism in age-related macular degeneration. *Science.* 2005 Apr 15;308(5720):385-9. Epub 2005 Mar 10. PubMed PMID: 15761122; PubMed Central PMCID: PMC1512523.

Köhler K, Bickeböller H. Case-control association tests correcting for population stratification. *Ann Hum Genet.* 2006 Jan;70(Pt 1):98-115. PubMed PMID: 16441260.

Kong C, Dunn M, Parker M. Psychiatric Genomics and Mental Health Treatment: Setting the Ethical Agenda. *Am J Bioeth.* 2017 Apr;17(4):3-12. doi: 10.1080/15265161.2017.1284915. PubMed PMID: 28328372.

Koob GF, Volkow ND. Neurobiology of addiction: a neurocircuitry analysis. *Lancet Psychiatry.* 2016 Aug;3(8):760-73. doi: 10.1016/S2215-0366(16)00104-8. Review. PubMed PMID: 27475769; PubMed Central PMCID: PMC6135092.

- Koob GF. A role for brain stress systems in addiction. *Neuron*. 2008 Jul 10;59(1):11-34. doi: 10.1016/j.neuron.2008.06.012. Review. PubMed PMID: 18614026; PubMed Central PMCID: PMC2748830.
- Kostick KM, Brannan C, Pereira S, Lázaro-Muñoz G. Psychiatric genetics researchers' views on offering return of results to individual participants. *Am J Med Genet B Neuropsychiatr Genet*. 2018 Oct 25. doi: 10.1002/ajmg.b.32682. [Epub ahead of print] PubMed PMID: 30358063; PubMed Central PMCID: PMC6483893.
- Kotov R, Gamez W, Schmidt F, Watson D. Linking "big" personality traits to anxiety, depressive, and substance use disorders: a meta-analysis. *Psychol Bull*. 2010 Sep;136(5):768-821. doi: 10.1037/a0020327. PubMed PMID: 20804236.
- Kotov R, Krueger RF, Watson D, Achenbach TM, Althoff RR, Bagby RM, Brown TA, Carpenter WT, Caspi A, Clark LA, Eaton NR, Forbes MK, Forbush KT, Goldberg D, Hasin D, Hyman SE, Ivanova MY, Lynam DR, Markon K, Miller JD, Moffitt TE, Morey LC, Mullins-Sweatt SN, Ormel J, Patrick CJ, Regier DA, Rescorla L, Ruggero CJ, Samuel DB, Sellbom M, Simms LJ, Skodol AE, Slade T, South SC, Tackett JL, Waldman ID, Waszczuk MA, Widiger TA, Wright AGC, Zimmerman M. The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *J Abnorm Psychol*. 2017 May;126(4):454-477. doi: 10.1037/abn0000258. Epub 2017 Mar 23. Review. PubMed PMID: 28333488.
- Kotov R, Krueger RF, Watson D. A paradigm shift in psychiatric classification: the Hierarchical Taxonomy Of Psychopathology (HiTOP). *World Psychiatry*. 2018 Feb;17(1):24-25. doi: 10.1002/wps.20478. PubMed PMID: 29352543; PubMed Central PMCID: PMC5775140.
- Krapohl E, Patel H, Newhouse S, Curtis CJ, von Stumm S, Dale PS, Zabaneh D, Breen G, O'Reilly PF, Plomin R. Multi-polygenic score approach to trait prediction. *Mol Psychiatry*. 2018 May;23(5):1368-1374. doi: 10.1038/mp.2017.163. Epub 2017 Aug 8. PubMed PMID: 28785111; PubMed Central PMCID: PMC5681246.
- Kuchenbaecker KB, McGuffog L, Barrowdale D, Lee A, Soucy P, Dennis J, Domchek SM, Robson M, Spurdle AB, Ramus SJ, Mavaddat N, Terry MB, Neuhausen SL, Schmutzler RK, Simard J, Pharoah PDP, Offit K, Couch FJ, Chenevix-Trench G, Easton DF, Antoniou AC. Evaluation of Polygenic Risk Scores for Breast and Ovarian Cancer Risk Prediction in BRCA1 and BRCA2 Mutation Carriers. *J Natl Cancer Inst*. 2017 Jul 1;109(7). doi: 10.1093/jnci/djw302. PubMed PMID: 28376175; PubMed Central PMCID: PMC5408990.
- Kushima I, Aleksic B, Nakatochi M, Shimamura T, Shiino T, Yoshimi A, Kimura H, Takasaki Y, Wang C, Xing J, Ishizuka K, Oya-Ito T, Nakamura Y, Arioka Y, Maeda T, Yamamoto M, Yoshida M, Noma H, Hamada S, Morikawa M, Uno Y,

- Okada T, Iidaka T, Iritani S, Yamamoto T, Miyashita M, Kobori A, Arai M, Itokawa M, Cheng MC, Chuang YA, Chen CH, Suzuki M, Takahashi T, Hashimoto R, Yamamori H, Yasuda Y, Watanabe Y, Nunokawa A, Someya T, Ikeda M, Toyota T, Yoshikawa T, Numata S, Ohmori T, Kunimoto S, Mori D, Iwata N, Ozaki N. High-resolution copy number variation analysis of schizophrenia in Japan. *Mol Psychiatry*. 2017 Mar;22(3):430-440. doi: 10.1038/mp.2016.88. Epub 2016 May 31. PubMed PMID: 27240532.
- Lahey BB, Applegate B, Hakes JK, Zald DH, Hariri AR, Rathouz PJ. Is there a general factor of prevalent psychopathology during adulthood? *J Abnorm Psychol*. 2012 Nov;121(4):971-7. doi: 10.1037/a0028355. Epub 2012 Jul 30. PubMed PMID: 22845652; PubMed Central PMCID: PMC4134439.
- Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet*. 1995 Nov;11(3):241-7. PubMed PMID: 7581446.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann Y, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissole SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L,

- Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kasprzyk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowki J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, Szustakowki J; International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*. 2001 Feb 15;409(6822):860-921. Erratum in: *Nature* 2001 Jun 7;411(6838):720. Szustakowki, J [corrected to Szustakowski, J]. Erratum in: *Nature* 2001 Aug 2;412(6846):565. PubMed PMID: 11237011.
- Lander ES, Schork NJ. Genetic dissection of complex traits. *Science*. 1994 Sep 30;265(5181):2037-48. Review. Erratum in: *Science* 1994 Oct 21;266(5184):353. PubMed PMID: 8091226.
- Lander ES. The new genomics: global views of biology. *Science*. 1996 Oct 25;274(5287):536-9. PubMed PMID: 8928008.
- Laursen TM, Nordentoft M, Mortensen PB. Excess early mortality in schizophrenia. *Annu Rev Clin Psychol*. 2014;10:425-48. doi: 10.1146/annurev-clinpsy-032813-153657. Epub 2013 Dec 2. Review. PubMed PMID: 24313570.
- Laursen TM, Trabjerg BB, Mors O, Børglum AD, Hougaard DM, Mattheisen M, Meier SM, Byrne EM, Mortensen PB, Munk-Olsen T, Agerbo E. Association of the polygenic risk score for schizophrenia with mortality and suicidal behavior - A Danish population-based study. *Schizophr Res*. 2017 Jun;184:122-127. doi: 10.1016/j.schres.2016.12.001. Epub 2016 Dec 6. PubMed PMID: 27939829.
- Lázaro-Muñoz G, Farrell MS, Crowley JJ, Filmyer DM, Shaughnessy RA, Josiassen RC, Sullivan PF. Improved ethical guidance for the return of results from psychiatric genomics research. *Mol Psychiatry*. 2018 Jan;23(1):15-23. doi: 10.1038/mp.2017.228. Epub 2017 Nov 21. PubMed PMID: 29158581; PubMed Central PMCID: PMC5752587.
- Lebowitz MS, Ahn WK. Effects of biological explanations for mental disorders on clinicians' empathy. *Proc Natl Acad Sci U S A*. 2014 Dec 16;111(50):17786-90. doi: 10.1073/pnas.1414058111. Epub 2014 Dec 1. PubMed PMID: 25453068; PubMed Central PMCID: PMC4273344.
- Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, Nguyen-Viet TA, Bowers P, Sidorenko J, Karlsson Linnér R, Fontana MA, Kundu T, Lee C, Li H, Li R, Royer R, Timshel PN, Walters RK, Willoughby EA, Yengo L; 23andMe

- Research Team; COGENT (Cognitive Genomics Consortium); Social Science Genetic Association Consortium, Alver M, Bao Y, Clark DW, Day FR, Furlotte NA, Joshi PK, Kemper KE, Kleinman A, Langenberg C, Mägi R, Trampush JW, Verma SS, Wu Y, Lam M, Zhao JH, Zheng Z, Boardman JD, Campbell H, Freese J, Harris KM, Hayward C, Herd P, Kumari M, Lencz T, Luan J, Malhotra AK, Metspalu A, Milani L, Ong KK, Perry JRB, Porteous DJ, Ritchie MD, Smart MC, Smith BH, Tung JY, Wareham NJ, Wilson JF, Beauchamp JP, Conley DC, Esko T, Lehrer SF, Magnusson PKE, Oskarsson S, Pers TH, Robinson MR, Thom K, Watson C, Chabris CF, Meyer MN, Laibson DI, Yang J, Johannesson M, Koellinger PD, Turley P, Visscher PM, Benjamin DJ, Cesarini D. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet.* 2018 Jul 23;50(8):1112-1121. doi: 10.1038/s41588-018-0147-3. PubMed PMID: 30038396; PubMed Central PMCID: PMC6393768.
- Lee SH, DeCandia TR, Ripke S, Yang J; Schizophrenia Psychiatric Genome-Wide Association Study Consortium (PGC-SCZ); International Schizophrenia Consortium (ISC); Molecular Genetics of Schizophrenia Collaboration (MGS), Sullivan PF, Goddard ME, Keller MC, Visscher PM, Wray NR. Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat Genet.* 2012 Feb 19;44(3):247-50. doi: 10.1038/ng.1108. PubMed PMID: 22344220; PubMed Central PMCID: PMC3327879.
- Lencz T, Guha S, Liu C, Rosenfeld J, Mukherjee S, DeRosse P, John M, Cheng L, Zhang C, Badner JA, Ikeda M, Iwata N, Cichon S, Rietschel M, Nöthen MM, Cheng AT, Hodgkinson C, Yuan Q, Kane JM, Lee AT, Pisanté A, Gregersen PK, Pe'er I, Malhotra AK, Goldman D, Darvasi A. Genome-wide association study implicates NDST3 in schizophrenia and bipolar disorder. *Nat Commun.* 2013;4:2739. doi: 10.1038/ncomms3739. PubMed PMID: 24253340; PubMed Central PMCID: PMC3905728.
- Lencz T, Morgan TV, Athanasiou M, Dain B, Reed CR, Kane JM, Kucherlapati R, Malhotra AK. Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. *Mol Psychiatry.* 2007 Jun;12(6):572-80. Epub 2007 Mar 20. PubMed PMID: 17522711.
- Lewis CM, Levinson DF, Wise LH, DeLisi LE, Straub RE, Hovatta I, Williams NM, Schwab SG, Pulver AE, Faraone SV, Brzustowicz LM, Kaufmann CA, Garver DL, Gurling HM, Lindholm E, Coon H, Moises HW, Byerley W, Shaw SH, Mesen A, Sherrington R, O'Neill FA, Walsh D, Kendler KS, Ekelund J, Paunio T, Lönnqvist J, Peltonen L, O'Donovan MC, Owen MJ, Wildenauer DB, Maier W, Nestadt G, Blouin JL, Antonarakis SE, Mowry BJ, Silverman JM, Crowe RR, Cloninger CR, Tsuang MT, Malaspina D, Harkavy-Friedman JM, Svrakic DM, Bassett AS, Holcomb J, Kalsi G, McQuillin A, Brynjolfsson J, Sigmundsson T, Petursson H, Jazin E, Zoëga T, Helgason T. Genome scan meta-analysis of

- schizophrenia and bipolar disorder, part II: Schizophrenia. *Am J Hum Genet.* 2003 Jul;73(1):34-48. Epub 2003 Jun 11. PubMed PMID: 12802786; PubMed Central PMCID: PMC1180588.
- Lewis CM, Levinson DF, Wise LH, DeLisi LE, Straub RE, Hovatta I, Williams NM, Schwab SG, Pulver AE, Faraone SV, Brzustowicz LM, Kaufmann CA, Garver DL, Gurling HM, Lindholm E, Coon H, Moises HW, Byerley W, Shaw SH, Mesen A, Sherrington R, O'Neill FA, Walsh D, Kendler KS, Ekelund J, Paunio T, Lönnqvist J, Peltonen L, O'Donovan MC, Owen MJ, Wildenauer DB, Maier W, Nestadt G, Blouin JL, Antonarakis SE, Mowry BJ, Silverman JM, Crowe RR, Cloninger CR, Tsuang MT, Malaspina D, Harkavy-Friedman JM, Svrakic DM, Bassett AS, Holcomb J, Kalsi G, McQuillin A, Brynjolfson J, Sigmundsson T, Petursson H, Jazin E, Zoëga T, Helgason T. Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. *Am J Hum Genet.* 2003 Jul;73(1):34-48. Epub 2003 Jun 11. PubMed PMID: 12802786; PubMed Central PMCID: PMC1180588.
- Lewis CM, Vassos E. Prospects for using risk scores in polygenic medicine. *Genome Med.* 2017 Nov 13;9(1):96. doi: 10.1186/s13073-017-0489-y. PubMed PMID: 29132412; PubMed Central PMCID: PMC5683372.
- Li J, Yoshikawa A, Brennan MD, Ramsey TL, Meltzer HY. Genetic predictors of antipsychotic response to lurasidone identified in a genome wide association study and by schizophrenia risk genes. *Schizophr Res.* 2018 Feb;192:194-204. doi: 10.1016/j.schres.2017.04.009. Epub 2017 Apr 19. PubMed PMID: 28431800.
- Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. *Annu Rev Genomics Hum Genet.* 2009;10:387-406. doi: 10.1146/annurev.genom.9.081307.164242. Review. PubMed PMID: 19715440; PubMed Central PMCID: PMC2925172.
- Li Z, Chen J, Yu H, He L, Xu Y, Zhang D, Yi Q, Li C, Li X, Shen J, Song Z, Ji W, Wang M, Zhou J, Chen B, Liu Y, Wang J, Wang P, Yang P, Wang Q, Feng G, Liu B, Sun W, Li B, He G, Li W, Wan C, Xu Q, Li W, Wen Z, Liu K, Huang F, Ji J, Ripke S, Yue W, Sullivan PF, O'Donovan MC, Shi Y. Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. *Nat Genet.* 2017 Nov;49(11):1576-1583. doi: 10.1038/ng.3973. Epub 2017 Oct 9. PubMed PMID: 28991256.
- Lichtenstein P, Yip BH, Björk C, Pawitan Y, Cannon TD, Sullivan PF, Hultman CM. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet.* 2009 Jan 17;373(9659):234-9. doi: 10.1016/S0140-6736(09)60072-6. PubMed PMID: 19150704; PubMed Central PMCID: PMC3879718.
- LILIENFELD AM, PASAMANICK B, ROGERS M. Relationship between pregnancy experience and the development of certain neuropsychiatric disorders

in childhood. *Am J Public Health Nations Health*. 1955 May;45(5 Pt 1):637-43. PubMed PMID: 14361807; PubMed Central PMCID: PMC1623200.

Lilienfeld SO, Treadway MT. Clashing Diagnostic Approaches: DSM-ICD Versus RDoC. *Annu Rev Clin Psychol*. 2016;12:435-63. doi: 10.1146/annurev-clinpsy-021815-093122. Epub 2016 Feb 3. Review. PubMed PMID: 26845519; PubMed Central PMCID: PMC5154554.

Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M. Comparison of next-generation sequencing systems. *J Biomed Biotechnol*. 2012;2012:251364. doi: 10.1155/2012/251364. Epub 2012 Jul 5. Review. PubMed PMID: 22829749; PubMed Central PMCID: PMC3398667.

Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, Datta G, Davila-Velderrain J, McGuire D, Tian C, Zhan X; 23andMe Research Team; HUNT All-In Psychiatry, Choquet H, Docherty AR, Faul JD, Foerster JR, Fritsche LG, Gabrielsen ME, Gordon SD, Haessler J, Hottenga JJ, Huang H, Jang SK, Jansen PR, Ling Y, Mägi R, Matoba N, McMahon G, Mulas A, Orrù V, Palviainen T, Pandit A, Reginsson GW, Skogholt AH, Smith JA, Taylor AE, Turman C, Willemsen G, Young H, Young KA, Zajac GJM, Zhao W, Zhou W, Bjornsdottir G, Boardman JD, Boehnke M, Boomsma DI, Chen C, Cucca F, Davies GE, Eaton CB, Ehringer MA, Esko T, Fiorillo E, Gillespie NA, Gudbjartsson DF, Haller T, Harris KM, Heath AC, Hewitt JK, Hickie IB, Hokanson JE, Hopfer CJ, Hunter DJ, Iacono WG, Johnson EO, Kamatani Y, Kardia SLR, Keller MC, Kellis M, Kooperberg C, Kraft P, Krauter KS, Laakso M, Lind PA, Loukola A, Lutz SM, Madden PAF, Martin NG, McGue M, McQueen MB, Medland SE, Metspalu A, Mohlke KL, Nielsen JB, Okada Y, Peters U, Polderman TJC, Posthuma D, Reiner AP, Rice JP, Rimm E, Rose RJ, Runarsdottir V, Stallings MC, Stančáková A, Stefansson H, Thai KK, Tindle HA, Tyrfingsson T, Wall TL, Weir DR, Weisner C, Whitfield JB, Winsvold BS, Yin J, Zuccolo L, Bierut LJ, Hveem K, Lee JJ, Munafò MR, Saccone NL, Willer CJ, Cornelis MC, David SP, Hinds DA, Jorgenson E, Kaprio J, Stitzel JA, Stefansson K, Thorgeirsson TE, Abecasis G, Liu DJ, Vrieze S. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet*. 2019 Feb;51(2):237-244. doi: 10.1038/s41588-018-0307-5. Epub 2019 Jan 14. PubMed PMID: 30643251; PubMed Central PMCID: PMC6358542.

Liu M, Malone SM, Vaidyanathan U, Keller MC, Abecasis G, McGue M, Iacono WG, Vrieze SI. Psychophysiological endophenotypes to characterize mechanisms of known schizophrenia genetic loci. *Psychol Med*. 2017 Apr;47(6):1116-1125. doi: 10.1017/S0033291716003184. Epub 2016 Dec 20. PubMed PMID: 27995817; PubMed Central PMCID: PMC5352523.

Lloyd-Jones LR, Holloway A, McRae A, Yang J, Small K, Zhao J, Zeng B, Bakshi A, Metspalu A, Dermitzakis M, Gibson G, Spector T, Montgomery G, Esko T,

- Visscher PM, Powell JE. The Genetic Architecture of Gene Expression in Peripheral Blood. *Am J Hum Genet.* 2017 Feb 2;100(2):228-237. doi: 10.1016/j.ajhg.2016.12.008. Epub 2017 Jan 5. Erratum in: *Am J Hum Genet.* 2017 Feb 2;100(2):371. PubMed PMID: 28065468; PubMed Central PMCID: PMC5294670.
- Lo MT, Hinds DA, Tung JY, Franz C, Fan CC, Wang Y, Smeland OB, Schork A, Holland D, Kauppi K, Sanyal N, Escott-Price V, Smith DJ, O'Donovan M, Stefansson H, Bjornsdottir G, Thorgeirsson TE, Stefansson K, McEvoy LK, Dale AM, Andreassen OA, Chen CH. Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. *Nat Genet.* 2017 Jan;49(1):152-156. doi: 10.1038/ng.3736. Epub 2016 Dec 5. PubMed PMID: 27918536; PubMed Central PMCID: PMC5278898.
- Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J, Croteau-Chonka DC, Esko T, Fall T, Ferreira T, Gustafsson S, Kutalik Z, Luan J, Mägi R, Randall JC, Winkler TW, Wood AR, Workalemahu T, Faul JD, Smith JA, Zhao JH, Zhao W, Chen J, Fehrmann R, Hedman ÅK, Karjalainen J, Schmidt EM, Absher D, Amin N, Anderson D, Beekman M, Bolton JL, Bragg-Gresham JL, Buyske S, Demirkan A, Deng G, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Goel A, Gong J, Jackson AU, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Mangino M, Leach IM, Medina-Gomez C, Medland SE, Nalls MA, Palmer CD, Pasko D, Pechlivanis S, Peters MJ, Prokopenko I, Shungin D, Stančáková A, Strawbridge RJ, Sung YJ, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Isaacs A, Albrecht E, Ärnlöv J, Arscott GM, Attwood AP, Bandinelli S, Barrett A, Bas IN, Bellis C, Bennett AJ, Berne C, Blagieva R, Blüher M, Böhringer S, Bonnycastle LL, Böttcher Y, Boyd HA, Bruinenberg M, Caspersen IH, Chen YI, Clarke R, Daw EW, de Craen AJM, Delgado G, Dimitriou M, Doney ASF, Eklund N, Estrada K, Eury E, Folkersen L, Fraser RM, Garcia ME, Geller F, Giedraitis V, Gigante B, Go AS, Golay A, Goodall AH, Gordon SD, Gorski M, Grabe HJ, Grallert H, Grammer TB, Gräßler J, Grönberg H, Groves CJ, Gusto G, Haessler J, Hall P, Haller T, Hallmans G, Hartman CA, Hassinen M, Hayward C, Heard-Costa NL, Helmer Q, Hengstenberg C, Holmen O, Hottenga JJ, James AL, Jeff JM, Johansson Å, Jolley J, Juliusdottir T, Kinnunen L, Koenig W, Koskenvuo M, Kratzer W, Laitinen J, Lamina C, Leander K, Lee NR, Lichtner P, Lind L, Lindström J, Lo KS, Lobbens S, Lorbeer R, Lu Y, Mach F, Magnusson PKE, Mahajan A, McArdle WL, McLachlan S, Menni C, Merger S, Mihailov E, Milani L, Moayyeri A, Monda KL, Morken MA, Mulas A, Müller G, Müller-Nurasyid M, Musk AW, Nagaraja R, Nöthen MM, Nolte IM, Pilz S, Rayner NW, Renstrom F, Rettig R, Ried JS, Ripke S, Robertson NR, Rose LM, Sanna S, Scharnagl H, Scholtens S, Schumacher FR, Scott WR, Seufferlein T, Shi J, Smith AV, Smolonska J, Stanton AV, Steinthorsdottir V, Stirrups K, Stringham HM, Sundström J, Swertz MA, Swift AJ, Syvänen AC, Tan

ST, Tayo BO, Thorand B, Thorleifsson G, Tyrer JP, Uh HW, Vandenput L, Verhulst FC, Vermeulen SH, Verweij N, Vonk JM, Waite LL, Warren HR, Waterworth D, Weedon MN, Wilkens LR, Willenborg C, Wilsgaard T, Wojczynski MK, Wong A, Wright AF, Zhang Q; LifeLines Cohort Study, Brennan EP, Choi M, Dastani Z, Drong AW, Eriksson P, Franco-Cereceda A, Gådin JR, Gharavi AG, Goddard ME, Handsaker RE, Huang J, Karpe F, Kathiresan S, Keildson S, Kiryluk K, Kubo M, Lee JY, Liang L, Lifton RP, Ma B, McCarroll SA, McKnight AJ, Min JL, Moffatt MF, Montgomery GW, Murabito JM, Nicholson G, Nyholt DR, Okada Y, Perry JRB, Dorajoo R, Reinmaa E, Salem RM, Sandholm N, Scott RA, Stolk L, Takahashi A, Tanaka T, van 't Hooft FM, Vinkhuyzen AAE, Westra HJ, Zheng W, Zondervan KT; ADIPOGen Consortium; AGEN-BMI Working Group; CARDIOGRAMplusC4D Consortium; CKDGen Consortium; GLGC; ICBP; MAGIC Investigators; MuTHER Consortium; MIGen Consortium; PAGE Consortium; ReproGen Consortium; GENIE Consortium; International Endogene Consortium, Heath AC, Arveiler D, Bakker SJL, Beilby J, Bergman RN, Blangero J, Bovet P, Campbell H, Caulfield MJ, Cesana G, Chakravarti A, Chasman DI, Chines PS, Collins FS, Crawford DC, Cupples LA, Cusi D, Danesh J, de Faire U, den Ruijter HM, Dominiczak AF, Erbel R, Erdmann J, Eriksson JG, Farrall M, Felix SB, Ferrannini E, Ferrières J, Ford I, Forouhi NG, Forrester T, Franco OH, Gansevoort RT, Gejman PV, Gieger C, Gottesman O, Gudnason V, Gyllenstein U, Hall AS, Harris TB, Hattersley AT, Hicks AA, Hindorf LA, Hingorani AD, Hofman A, Homuth G, Hovingh GK, Humphries SE, Hunt SC, Hyppönen E, Illig T, Jacobs KB, Jarvelin MR, Jöckel KH, Johansen B, Jousilahti P, Jukema JW, Jula AM, Kaprio J, Kastelein JJP, Keinanen-Kiukkaanniemi SM, Kiemenev LA, Knekt P, Kooner JS, Kooperberg C, Kovacs P, Kraja AT, Kumari M, Kuusisto J, Lakka TA, Langenberg C, Marchand LL, Lehtimäki T, Lyssenko V, Männistö S, Marette A, Matise TC, McKenzie CA, McKnight B, Moll FL, Morris AD, Morris AP, Murray JC, Nelis M, Ohlsson C, Oldehinkel AJ, Ong KK, Madden PAF, Pasterkamp G, Peden JF, Peters A, Postma DS, Pramstaller PP, Price JF, Qi L, Raitakari OT, Rankinen T, Rao DC, Rice TK, Ridker PM, Rioux JD, Ritchie MD, Rudan I, Salomaa V, Samani NJ, Saramies J, Sarzynski MA, Schunkert H, Schwarz PEH, Sever P, Shuldiner AR, Sinisalo J, Stolk RP, Strauch K, Tönjes A, Trégouët DA, Tremblay A, Tremoli E, Virtamo J, Vohl MC, Völker U, Waeber G, Willemsen G, Witteman JC, Zillikens MC, Adair LS, Amouyel P, Asselbergs FW, Assimes TL, Bochud M, Boehm BO, Boerwinkle E, Bornstein SR, Bottinger EP, Bouchard C, Cauchi S, Chambers JC, Chanock SJ, Cooper RS, de Bakker PIW, Dedoussis G, Ferrucci L, Franks PW, Froguel P, Groop LC, Haiman CA, Hamsten A, Hui J, Hunter DJ, Hveem K, Kaplan RC, Kivimäki M, Kuh D, Laakso M, Liu Y, Martin NG, März W, Melbye M, Metspalu A, Moebus S, Munroe PB, Njølstad I, Oostra BA, Palmer CNA, Pedersen NL, Perola M, Pérusse L, Peters U, Power C, Quertermous T, Rauramaa R, Rivadeneira F, Saaristo TE, Saleheen D, Sattar N, Schadt EE, Schlessinger D, Slagboom PE, Snieder H, Spector TD, Thorsteinsdottir U, Stumvoll M, Tuomilehto J, Uitterlinden AG, Uusitupa M, van

- der Harst P, Walker M, Wallaschofski H, Wareham NJ, Watkins H, Weir DR, Wichmann HE, Wilson JF, Zanen P, Borecki IB, Deloukas P, Fox CS, Heid IM, O'Connell JR, Strachan DP, Stefansson K, van Duijn CM, Abecasis GR, Franke L, Frayling TM, McCarthy MI, Visscher PM, Scherag A, Willer CJ, Boehnke M, Mohlke KL, Lindgren CM, Beckmann JS, Barroso I, North KE, Ingelsson E, Hirschhorn JN, Loos RJF, Speliotes EK. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015 Feb 12;518(7538):197-206. doi: 10.1038/nature14177. PubMed PMID: 25673413; PubMed Central PMCID: PMC4382211.
- Loh PR, Bhatia G, Gusev A, Finucane HK, Bulik-Sullivan BK, Pollack SJ; Schizophrenia Working Group of Psychiatric Genomics Consortium, de Candia TR, Lee SH, Wray NR, Kendler KS, O'Donovan MC, Neale BM, Patterson N, Price AL. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat Genet*. 2015 Dec;47(12):1385-92. doi: 10.1038/ng.3431. Epub 2015 Nov 2. PubMed PMID: 26523775; PubMed Central PMCID: PMC4666835.
- Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet*. 2003 Feb;33(2):177-82. Epub 2003 Jan 13. PubMed PMID: 12524541.
- Lopez AD, Murray CC. The global burden of disease, 1990-2020. *Nat Med*. 1998 Nov;4(11):1241-3. PubMed PMID: 9809543.
- Lu Y, Pouget JG, Andreassen OA, Djurovic S, Esko T, Hultman CM, Metspalu A, Milani L, Werge T, Sullivan PF. Genetic risk scores and family history as predictors of schizophrenia in Nordic registers. *Psychol Med*. 2018 May;48(7):1201-1208. doi: 10.1017/S0033291717002665. Epub 2017 Sep 25. PubMed PMID: 28942743.
- Lupski JR. Genomic rearrangements and sporadic disease. *Nat Genet*. 2007 Jul;39(7 Suppl):S43-7. Review. PubMed PMID: 17597781.
- Lupski JR. Retrotransposition and structural variation in the human genome. *Cell*. 2010 Jun 25;141(7):1110-2. doi: 10.1016/j.cell.2010.06.014. PubMed PMID: 20602993.
- Luxenburger, H. Vorläufiger bericht über psychiatrische Serienuntersuchungen an zwillingen. *Zeitschrift für die gesamte Neurologie und Psychiatrie* 1928. 116(1), 297-326.
- Lykouras L, Alevizos B, Michalopoulou P, Rabavilas A. Obsessive-compulsive symptoms induced by atypical antipsychotics. A review of the reported cases. *Prog*

Neuropsychopharmacol Biol Psychiatry. 2003 May;27(3):333-46. Review. PubMed PMID: 12691768.

Mackey S, Allgaier N, Charani B, Spechler P, Orr C, Bunn J, Allen NB, Alia-Klein N, Batalla A, Blaine S, Brooks S, Caparelli E, Chye YY, Cousijn J, Dagher A, Desrivieres S, Feldstein-Ewing S, Foxe JJ, Goldstein RZ, Goudriaan AE, Heitzeg MM, Hester R, Hutchison K, Korucuoglu O, Li CR, London E, Lorenzetti V, Luijten M, Martin-Santos R, May A, Momenan R, Morales A, Paulus MP, Pearlson G, Rousseau ME, Salmeron BJ, Schluter R, Schmaal L, Schumann G, Sjoerds Z, Stein DJ, Stein EA, Sinha R, Solowij N, Tapert S, Uhlmann A, Veltman D, van Holst R, Whittle S, Wright MJ, Yücel M, Zhang S, Yurgelun-Todd D, Hibar DP, Jahanshad N, Evans A, Thompson PM, Glahn DC, Conrod P, Garavan H; ENIGMA Addiction Working Group. Mega-Analysis of Gray Matter Volume in Substance Dependence: General and Substance-Specific Regional Effects. *Am J Psychiatry*. 2019 Feb 1;176(2):119-128. doi: 10.1176/appi.ajp.2018.17040415. Epub 2018 Oct 19. PubMed PMID: 30336705; PubMed Central PMCID: PMC6427822.

Mah S, Nelson MR, Delisi LE, Reneland RH, Markward N, James MR, Nyholt DR, Hayward N, Handoko H, Mowry B, Kammerer S, Braun A. Identification of the semaphorin receptor PLXNA2 as a candidate for susceptibility to schizophrenia. *Mol Psychiatry*. 2006 May;11(5):471-8. PubMed PMID: 16402134.

Maher B. Personal genomes: The case of the missing heritability. *Nature*. 2008 Nov 6;456(7218):18-21. doi: 10.1038/456018a. PubMed PMID: 18987709.

Maier R, Moser G, Chen GB, Ripke S; Cross-Disorder Working Group of the Psychiatric Genomics Consortium, Coryell W, Potash JB, Scheftner WA, Shi J, Weissman MM, Hultman CM, Landén M, Levinson DF, Kendler KS, Smoller JW, Wray NR, Lee SH. Joint analysis of psychiatric disorders increases accuracy of risk prediction for schizophrenia, bipolar disorder, and major depressive disorder. *Am J Hum Genet*. 2015 Feb 5;96(2):283-94. doi: 10.1016/j.ajhg.2014.12.006. Epub 2015 Jan 29. PubMed PMID: 25640677; PubMed Central PMCID: PMC4320268.

Maier RM, Zhu Z, Lee SH, Trzaskowski M, Ruderfer DM, Stahl EA, Ripke S, Wray NR, Yang J, Visscher PM, Robinson MR. Improving genetic prediction by leveraging genetic correlations among human diseases and traits. *Nat Commun*. 2018 Mar 7;9(1):989. doi: 10.1038/s41467-017-02769-6. PubMed PMID: 29515099; PubMed Central PMCID: PMC5841449.

Maj M. "Psychiatric comorbidity": an artefact of current diagnostic systems? *Br J Psychiatry*. 2005 Mar;186:182-4. PubMed PMID: 15738496.

- Malhotra AK, Goldman D. Benefits and pitfalls encountered in psychiatric genetic association studies. *Biol Psychiatry*. 1999 Mar 1;45(5):544-50. Review. PubMed PMID: 10088044.
- Malhotra D, Sebat J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell*. 2012 Mar 16;148(6):1223-41. doi: 10.1016/j.cell.2012.02.039. Review. PubMed PMID: 22424231; PubMed Central PMCID: PMC3351385.
- Männik K, Mägi R, Macé A, Cole B, Guyatt AL, Shihab HA, Maillard AM, Alavere H, Kolk A, Reigo A, Mihailov E, Leitsalu L, Ferreira AM, Nõukas M, Teumer A, Salvi E, Cusi D, McGue M, Iacono WG, Gaunt TR, Beckmann JS, Jacquemont S, Kutalik Z, Pankratz N, Timpson N, Metspalu A, Reymond A. Copy number variations and cognitive phenotypes in unselected populations. *JAMA*. 2015 May 26;313(20):2044-54. doi: 10.1001/jama.2015.4845. PubMed PMID: 26010633; PubMed Central PMCID: PMC4684269.
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. *Nature*. 2009 Oct 8;461(7265):747-53. doi: 10.1038/nature08494. Review. PubMed PMID: 19812666; PubMed Central PMCID: PMC2831613.
- Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*. 2007 Jul;39(7):906-13. Epub 2007 Jun 17. PubMed PMID: 17572673.
- Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet*. 2010 Jul;11(7):499-511. doi: 10.1038/nrg2796. Review. PubMed PMID: 20517342.
- Marshall CR, Howrigan DP, Merico D, Thiruvahindrapuram B, Wu W, Greer DS, Antaki D, Shetty A, Holmans PA, Pinto D, Gujral M, Brandler WM, Malhotra D, Wang Z, Fajardo KVF, Maile MS, Ripke S, Agartz I, Albus M, Alexander M, Amin F, Atkins J, Bacanu SA, Belliveau RA Jr, Bergen SE, Bertalan M, Bevilacqua E, Bigdeli TB, Black DW, Bruggeman R, Buccola NG, Buckner RL, Bulik-Sullivan B, Byerley W, Cahn W, Cai G, Cairns MJ, Campion D, Cantor RM, Carr VJ, Carrera N, Catts SV, Chambert KD, Cheng W, Cloninger CR, Cohen D, Cormican P, Craddock N, Crespo-Facorro B, Crowley JJ, Curtis D, Davidson M, Davis KL, Degenhardt F, Del Favero J, DeLisi LE, Dikeos D, Dinan T, Djurovic S, Donohoe G, Drapeau E, Duan J, Dudbridge F, Eichhammer P, Eriksson J, Escott-Price V, Essioux L, Fanous AH, Farh KH, Farrell MS, Frank J, Franke L, Freedman R, Freimer NB, Friedman JI, Forstner AJ, Fromer M, Genovese G, Georgieva L, Gershon ES, Giegling I, Giusti-Rodríguez P, Godard S, Goldstein JI, Gratten J, de Haan L, Hamshere ML, Hansen M, Hansen T,

Haroutunian V, Hartmann AM, Henskens FA, Herms S, Hirschhorn JN, Hoffmann P, Hofman A, Huang H, Ikeda M, Joa I, Kähler AK, Kahn RS, Kalaydjieva L, Karjalainen J, Kavanagh D, Keller MC, Kelly BJ, Kennedy JL, Kim Y, Knowles JA, Konte B, Laurent C, Lee P, Lee SH, Legge SE, Lerer B, Levy DL, Liang KY, Lieberman J, Lönnqvist J, Loughland CM, Magnusson PKE, Maher BS, Maier W, Mallet J, Mattheisen M, Mattingdal M, McCarley RW, McDonald C, McIntosh AM, Meier S, Meijer CJ, Melle I, Meshulam-Gately RI, Metspalu A, Michie PT, Milani L, Milanova V, Mokrab Y, Morris DW, Müller-Myhsok B, Murphy KC, Murray RM, Myin-Germeys I, Nenadic I, Nertney DA, Nestadt G, Nicodemus KK, Nisenbaum L, Nordin A, O'Callaghan E, O'Dushlaine C, Oh SY, Olincy A, Olsen L, O'Neill FA, Van Os J, Pantelis C, Papadimitriou GN, Parkhomenko E, Pato MT, Paunio T; Psychosis Endophenotypes International Consortium, Perkins DO, Pers TH, Pietiläinen O, Pimm J, Pocklington AJ, Powell J, Price A, Pulver AE, Purcell SM, Quested D, Rasmussen HB, Reichenberg A, Reimers MA, Richards AL, Roffman JL, Roussos P, Ruderfer DM, Salomaa V, Sanders AR, Savitz A, Schall U, Schulze TG, Schwab SG, Scolnick EM, Scott RJ, Seidman LJ, Shi J, Silverman JM, Smoller JW, Söderman E, Spencer CCA, Stahl EA, Strengman E, Strohmaier J, Stroup TS, Suvisaari J, Svrakic DM, Szatkiewicz JP, Thirumalai S, Tooney PA, Veijola J, Visscher PM, Waddington J, Walsh D, Webb BT, Weiser M, Wildenauer DB, Williams NM, Williams S, Witt SH, Wolen AR, Wormley BK, Wray NR, Wu JQ, Zai CC, Adolfsson R, Andreassen OA, Blackwood DHR, Bramon E, Buxbaum JD, Cichon S, Collier DA, Corvin A, Daly MJ, Darvasi A, Domenici E, Esko T, Gejman PV, Gill M, Gurling H, Hultman CM, Iwata N, Jablensky AV, Jönsson EG, Kendler KS, Kirov G, Knight J, Levinson DF, Li QS, McCarroll SA, McQuillin A, Moran JL, Mowry BJ, Nöthen MM, Ophoff RA, Owen MJ, Palotie A, Pato CN, Petryshen TL, Posthuma D, Rietschel M, Riley BP, Rujescu D, Sklar P, St Clair D, Walters JTR, Werge T, Sullivan PF, O'Donovan MC, Scherer SW, Neale BM, Sebat J; CNV and Schizophrenia Working Groups of the Psychiatric Genomics Consortium. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet.* 2017 Jan;49(1):27-35. doi: 10.1038/ng.3725. Epub 2016 Nov 21. Erratum in: *Nat Genet.* 2017 Sep 27;49(10):1558. *Nat Genet.* 2017 Mar 30;49(4):651. PubMed PMID: 27869829; PubMed Central PMCID: PMC5737772.

Martens FK, Tonk ECM, Janssens ACJW. Evaluation of polygenic risk models using multiple performance measures: a critical assessment of discordant results. *Genet Med.* 2019 Feb;21(2):391-397. doi: 10.1038/s41436-018-0058-9. Epub 2018 Jun 12. PubMed PMID: 29895851; PubMed Central PMCID: PMC6169739.

Martin AR, Daly MJ, Robinson EB, Hyman SE, Neale BM. Predicting Polygenic Risk of Psychiatric Disorders. *Biol Psychiatry.* 2018 Dec 28. pii: S0006-3223(18)32119-X. doi: 10.1016/j.biopsych.2018.12.015. [Epub ahead of print] Review. PubMed PMID: 30737014.

- Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet.* 2019 Apr;51(4):584-591. doi: 10.1038/s41588-019-0379-x. Epub 2019 Mar 29. Review. PubMed PMID: 30926966.
- Martin J, O'Donovan MC, Thapar A, Langley K, Williams N. The relative contribution of common and rare genetic variants to ADHD. *Transl Psychiatry.* 2015 Feb 10;5:e506. doi: 10.1038/tp.2015.5. PubMed PMID: 25668434; PubMed Central PMCID: PMC4445754.
- Mavaddat N, Pharoah PD, Michailidou K, Tyrer J, Brook MN, Bolla MK, Wang Q, Dennis J, Dunning AM, Shah M, Luben R, Brown J, Bojesen SE, Nordestgaard BG, Nielsen SF, Flyger H, Czene K, Darabi H, Eriksson M, Peto J, Dos-Santos-Silva I, Dudbridge F, Johnson N, Schmidt MK, Broeks A, Verhoef S, Rutgers EJ, Swerdlow A, Ashworth A, Orr N, Schoemaker MJ, Figueroa J, Chanock SJ, Brinton L, Lissowska J, Couch FJ, Olson JE, Vachon C, Pankratz VS, Lambrechts D, Wildiers H, Van Ongeval C, van Limbergen E, Kristensen V, Grenaker Alnæs G, Nord S, Borresen-Dale AL, Nevanlinna H, Muranen TA, Aittomäki K, Blomqvist C, Chang-Claude J, Rudolph A, Seibold P, Flesch-Janys D, Fasching PA, Haeberle L, Ekici AB, Beckmann MW, Burwinkel B, Marme F, Schneeweiss A, Sohn C, Trentham-Dietz A, Newcomb P, Titus L, Egan KM, Hunter DJ, Lindstrom S, Tamimi RM, Kraft P, Rahman N, Turnbull C, Renwick A, Seal S, Li J, Liu J, Humphreys K, Benitez J, Pilar Zamora M, Arias Perez JI, Menéndez P, Jakubowska A, Lubinski J, Jaworska-Bieniek K, Durda K, Bogdanova NV, Antonenkova NN, Dörk T, Anton-Culver H, Neuhausen SL, Ziogas A, Bernstein L, Devilee P, Tollenaar RA, Seynaeve C, van Asperen CJ, Cox A, Cross SS, Reed MW, Khusnutdinova E, Bermisheva M, Prokofyeva D, Takhirova Z, Meindl A, Schmutzler RK, Sutter C, Yang R, Schürmann P, Bremer M, Christiansen H, Park-Simon TW, Hillemanns P, Guénel P, Truong T, Menegaux F, Sanchez M, Radice P, Peterlongo P, Manoukian S, Pensotti V, Hopper JL, Tsimiklis H, Apicella C, Southey MC, Brauch H, Brüning T, Ko YD, Sigurdson AJ, Doody MM, Hamann U, Torres D, Ulmer HU, Försti A, Sawyer EJ, Tomlinson I, Kerin MJ, Miller N, Andrulis IL, Knight JA, Glendon G, Marie Mulligan A, Chenevix-Trench G, Balleine R, Giles GG, Milne RL, McLean C, Lindblom A, Margolin S, Haiman CA, Henderson BE, Schumacher F, Le Marchand L, Eilber U, Wang-Gohrke S, Hooning MJ, Hollestelle A, van den Ouweland AM, Koppert LB, Carpenter J, Clarke C, Scott R, Mannermaa A, Kataja V, Kosma VM, Hartikainen JM, Brenner H, Arndt V, Stegmaier C, Karina Dieffenbach A, Winqvist R, Pylkäs K, Jukkola-Vuorinen A, Grip M, Offit K, Vijai J, Robson M, Rau-Murthy R, Dwek M, Swann R, Annie Perkins K, Goldberg MS, Labrèche F, Dumont M, Eccles DM, Tapper WJ, Rafiq S, John EM, Whittemore AS, Slager S, Yannoukakos D, Toland AE, Yao S, Zheng W, Halverson SL, González-Neira A, Pita G, Rosario Alonso M, Álvarez N, Herrero D, Tessier DC, Vincent D, Bacot F, Luccarini C, Baynes C, Ahmed S, Maranian M, Healey CS, Simard J, Hall P, Easton DF, Garcia-Closas

- M. Prediction of breast cancer risk based on profiling with common genetic variants. *J Natl Cancer Inst.* 2015 Apr 8;107(5). pii: djv036. doi: 10.1093/jnci/djv036. Print 2015 May. PubMed PMID: 25855707; PubMed Central PMCID: PMC4754625.
- McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet.* 2008 May;9(5):356-69. doi: 10.1038/nrg2344. Review. PubMed PMID: 18398418.
- McCarthy S, Das S, Kretschmar W, Delaneau O, Wood AR, Teumer A, Kang HM, Fuchsberger C, Danecek P, Sharp K, Luo Y, Sidore C, Kwong A, Timpson N, Koskinen S, Vrieze S, Scott LJ, Zhang H, Mahajan A, Veldink J, Peters U, Pato C, van Duijn CM, Gillies CE, Gandin I, Mezzavilla M, Gilly A, Cocca M, Traglia M, Angius A, Barrett JC, Boomsma D, Branham K, Breen G, Brummett CM, Busonero F, Campbell H, Chan A, Chen S, Chew E, Collins FS, Corbin LJ, Smith GD, Dedoussis G, Dorr M, Farmaki AE, Ferrucci L, Forer L, Fraser RM, Gabriel S, Levy S, Groop L, Harrison T, Hattersley A, Holmen OL, Hveem K, Kretzler M, Lee JC, McGue M, Meitinger T, Melzer D, Min JL, Mohlke KL, Vincent JB, Nauck M, Nickerson D, Palotie A, Pato M, Pirastu N, McInnis M, Richards JB, Sala C, Salomaa V, Schlessinger D, Schoenherr S, Slagboom PE, Small K, Spector T, Stambolian D, Tuke M, Tuomilehto J, Van den Berg LH, Van Rheenen W, Volker U, Wijmenga C, Toniolo D, Zeggini E, Gasparini P, Sampson MG, Wilson JF, Frayling T, de Bakker PI, Swertz MA, McCarroll S, Kooperberg C, Dekker A, Altshuler D, Willer C, Iacono W, Ripatti S, Soranzo N, Walter K, Swaroop A, Cucca F, Anderson CA, Myers RM, Boehnke M, McCarthy MI, Durbin R; Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet.* 2016 Oct;48(10):1279-83. doi: 10.1038/ng.3643. Epub 2016 Aug 22. PubMed PMID: 27548312; PubMed Central PMCID: PMC5388176.
- McClellan JM, Susser E, King MC. Schizophrenia: a common disease caused by multiple rare alleles. *Br J Psychiatry.* 2007 Mar;190:194-9. Review. PubMed PMID: 17329737.
- McGorry P, Keshavan M, Goldstone S, Amminger P, Allott K, Berk M, Lavoie S, Pantelis C, Yung A, Wood S, Hickie I. Biomarkers and clinical staging in psychiatry. *World Psychiatry.* 2014 Oct;13(3):211-23. doi: 10.1002/wps.20144. PubMed PMID: 25273285; PubMed Central PMCID: PMC4219053.
- McGuffin P, Rijdsdijk F, Andrew M, Sham P, Katz R, Cardno A. The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. *Arch Gen Psychiatry.* 2003 May;60(5):497-502. PubMed PMID: 12742871.
- McIntosh AM, Gow A, Luciano M, Davies G, Liewald DC, Harris SE, Corley J, Hall J, Starr JM, Porteous DJ, Tenesa A, Visscher PM, Deary IJ. Polygenic risk for

- schizophrenia is associated with cognitive change between childhood and old age. *Biol Psychiatry*. 2013 May 15;73(10):938-43. doi: 10.1016/j.biopsych.2013.01.011. Epub 2013 Feb 16. PubMed PMID: 23419543.
- McKusick VA, Edwards JH. Report of the Committee on Unassigned Syntenic Groups and Theoretical Considerations. *Birth Defects Orig Artic Ser*. 1975;11(3):26-8. PubMed PMID: 1203491.
- Meehl, P. E. (1962). Schizophrenia, schizotypy and schizotaxia. *Am Psychol*, 17, 827-838.
- Mega JL, Stitzel NO, Smith JG, Chasman DI, Caulfield M, Devlin JJ, Nordio F, Hyde C, Cannon CP, Sacks F, Poulter N, Sever P, Ridker PM, Braunwald E, Melander O, Kathiresan S, Sabatine MS. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. *Lancet*. 2015 Jun 6;385(9984):2264-2271. doi: 10.1016/S0140-6736(14)61730-X. Epub 2015 Mar 4. PubMed PMID: 25748612; PubMed Central PMCID: PMC4608367.
- Meier SM, Agerbo E, Maier R, Pedersen CB, Lang M, Grove J, Hollegaard MV, Demontis D, Trabjerg BB, Hjorthøj C, Ripke S, Degenhardt F, Nöthen MM, Rujescu D, Maier W; MooDS SCZ Consortium, Werge T, Mors O, Hougaard DM, Børglum AD, Wray NR, Rietschel M, Nordentoft M, Mortensen PB, Mattheisen M. High loading of polygenic risk in cases with chronic schizophrenia. *Mol Psychiatry*. 2016 Jul;21(7):969-74. doi: 10.1038/mp.2015.130. Epub 2015 Sep 1. PubMed PMID: 26324100.
- Meier SM, Petersen L, Pedersen MG, Arendt MC, Nielsen PR, Mattheisen M, Mors O, Mortensen PB. Obsessive-compulsive disorder as a risk factor for schizophrenia: a nationwide study. *JAMA Psychiatry*. 2014 Nov;71(11):1215-21. doi: 10.1001/jamapsychiatry.2014.1011. PubMed PMID: 25188738.
- Merikangas KR, Risch N. Will the genomics revolution revolutionize psychiatry? *Am J Psychiatry*. 2003 Apr;160(4):625-35. PubMed PMID: 12668348.
- Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheetham RK, Chinwalla A, Conrad DF, Fu Y, Grubert F, Hajirasouliha I, Hormozdiari F, Iakoucheva LM, Iqbal Z, Kang S, Kidd JM, Konkel MK, Korn J, Khurana E, Kural D, Lam HY, Leng J, Li R, Li Y, Lin CY, Luo R, Mu XJ, Nemes J, Peckham HE, Rausch T, Scally A, Shi X, Stromberg MP, Stütz AM, Urban AE, Walker JA, Wu J, Zhang Y, Zhang ZD, Batzer MA, Ding L, Marth GT, McVean G, Sebat J, Snyder M, Wang J, Ye K, Eichler EE, Gerstein MB, Hurler ME, Lee C, McCarroll SA, Korb JO; 1000 Genomes Project. Mapping copy number variation by population-scale genome sequencing. *Nature*. 2011 Feb 3;470(7332):59-65. doi: 10.1038/nature09708. PubMed PMID: 21293372; PubMed Central PMCID: PMC3077050.

- Mitchell KJ. What is complex about complex disorders? *Genome Biol.* 2012 Jan 23;13(1):237. doi: 10.1186/gb-2012-13-1-237. PubMed PMID: 22269335; PubMed Central PMCID: PMC3334577.
- Morales J, Welter D, Bowler EH, Cerezo M, Harris LW, McMahon AC, Hall P, Junkins HA, Milano A, Hastings E, Malangone C, Buniello A, Burdett T, Flicek P, Parkinson H, Cunningham F, Hindorff LA, MacArthur JAL. A standardized framework for representation of ancestry data in genomics studies, with application to the NHGRI-EBI GWAS Catalog. *Genome Biol.* 2018 Feb 15;19(1):21. doi: 10.1186/s13059-018-1396-2. PubMed PMID: 29448949; PubMed Central PMCID: PMC5815218.
- Moskvina V, Schmidt KM. Susceptibility of biallelic haplotype and genotype frequencies to genotyping error. *Biometrics.* 2006 Dec;62(4):1116-23. PubMed PMID: 17156286.
- Mühleisen TW, Mattheisen M, Strohmaier J, Degenhardt F, Priebe L, Schultz CC, Breuer R, Meier S, Hoffmann P; GROUP Investigators, Rivandeneira F, Hofman A, Uitterlinden AG, Moebus S, Gieger C, Emeny R, Ladwig KH, Wichmann HE, Schwarz M, Kammerer-Ciernioch J, Schlösser RG, Nenadic I, Sauer H, Mössner R, Maier W, Rujescu D, Lange C, Ophoff RA, Schulze TG, Rietschel M, Nöthen MM, Cichon S. Association between schizophrenia and common variation in neurocan (NCAN), a genetic risk factor for bipolar disorder. *Schizophr Res.* 2012 Jun;138(1):69-73. doi: 10.1016/j.schres.2012.03.007. Epub 2012 Apr 11. PubMed PMID: 22497794.
- Mullins-Sweatt SN, Widiger TA. Clinical utility and DSM-V. *Psychol Assess.* 2009 Sep;21(3):302-12. doi: 10.1037/a0016607. Review. PubMed PMID: 19719343.
- Munafò MR, Flint J. Meta-analysis of genetic association studies. *Trends Genet.* 2004 Sep;20(9):439-44. PubMed PMID: 15313553.
- Munafò MR. Candidate gene studies in the 21st century: meta-analysis, mediation, moderation. *Genes Brain Behav.* 2006;5 Suppl 1:3-8. Review. PubMed PMID: 16417611.
- Murphy KC, Jones LA, Owen MJ. High rates of schizophrenia in adults with velo-cardio-facial syndrome. *Arch Gen Psychiatry.* 1999 Oct;56(10):940-5. PubMed PMID: 10530637.
- Murray RM, Lewis SW. Is schizophrenia a neurodevelopmental disorder? *Br Med J (Clin Res Ed).* 1988 Jan 2;296(6614):63. PubMed PMID: 3122933; PubMed Central PMCID: PMC2544668.
- Nagelkerke, N. J. D. A note on a general definition of the coefficient of determination. *Biometrika.* 1991, 78, 691-693.

- NCI-NHGRI Working Group on Replication in Association Studies, Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, Hirschhorn JN, Abecasis G, Altshuler D, Bailey-Wilson JE, Brooks LD, Cardon LR, Daly M, Donnelly P, Fraumeni JF Jr, Freimer NB, Gerhard DS, Gunter C, Guttmacher AE, Guyer MS, Harris EL, Hoh J, Hoover R, Kong CA, Merikangas KR, Morton CC, Palmer LJ, Phimister EG, Rice JP, Roberts J, Rotimi C, Tucker MA, Vogan KJ, Wacholder S, Wijsman EM, Winn DM, Collins FS. Replicating genotype-phenotype associations. *Nature*. 2007 Jun 7;447(7145):655-60. PubMed PMID: 17554299.
- Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, Lin CF, Stevens C, Wang LS, Makarov V, Polak P, Yoon S, Maguire J, Crawford EL, Campbell NG, Geller ET, Valladares O, Schafer C, Liu H, Zhao T, Cai G, Lihm J, Dannenfels R, Jabado O, Peralta Z, Nagaswamy U, Muzny D, Reid JG, Newsham I, Wu Y, Lewis L, Han Y, Voight BF, Lim E, Rossin E, Kirby A, Flannick J, Fromer M, Shakir K, Fennell T, Garimella K, Banks E, Poplin R, Gabriel S, DePristo M, Wimbish JR, Boone BE, Levy SE, Betancur C, Sunyaev S, Boerwinkle E, Buxbaum JD, Cook EH Jr, Devlin B, Gibbs RA, Roeder K, Schellenberg GD, Sutcliffe JS, Daly MJ. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*. 2012 Apr 4;485(7397):242-5. doi: 10.1038/nature11011. PubMed PMID: 22495311; PubMed Central PMCID: PMC3613847.
- Neale BM, Sham PC. The future of association studies: gene-based analysis and replication. *Am J Hum Genet*. 2004 Sep;75(3):353-62. Epub 2004 Jul 22. PubMed PMID: 15272419; PubMed Central PMCID: PMC1182015.
- Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat Neurosci*. 2015 Feb;18(2):199-209. doi: 10.1038/nn.3922. Epub 2015 Jan 19. Erratum in: *Nat Neurosci*. 2015 Jun;18(6):926. *Nat Neurosci*. 2015 Dec;18(12):1861. PubMed PMID: 25599223; PubMed Central PMCID: PMC4378867.
- Newman DL, Moffitt TE, Caspi A, Silva PA. Comorbid mental disorders: implications for treatment and sample selection. *J Abnorm Psychol*. 1998 May;107(2):305-11. PubMed PMID: 9604559.
- Niemi MEK, Martin HC, Rice DL, Gallone G, Gordon S, Kelemen M, McAloney K, McRae J, Radford EJ, Yu S, Gecz J, Martin NG, Wright CF, Fitzpatrick DR, Firth HV, Hurles ME, Barrett JC. Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. *Nature*. 2018 Oct;562(7726):268-271. doi: 10.1038/s41586-018-0566-4. Epub 2018 Sep 26. PubMed PMID: 30258228.
- Niendam TA, Berzak J, Cannon TD, Bearden CE. Obsessive compulsive symptoms in the psychosis prodrome: correlates of clinical and functional outcome.

- Schizophr Res. 2009 Mar;108(1-3):170-5. doi: 10.1016/j.schres.2008.11.023. Epub 2008 Dec 20. PubMed PMID: 19097751; PubMed Central PMCID: PMC2670932.
- Nikolova YS, Hariri AR. Neural responses to threat and reward interact to predict stress-related problem drinking: A novel protective role of the amygdala. *Biol Mood Anxiety Disord*. 2012 Nov 14;2:19. doi: 10.1186/2045-5380-2-19. PubMed PMID: 23151390; PubMed Central PMCID: PMC3502334.
- Nikolova YS, Knodt AR, Radtke SR, Hariri AR. Divergent responses of the amygdala and ventral striatum predict stress-related problem drinking in young adults: possible differential markers of affective and impulsive pathways of risk for alcohol use disorder. *Mol Psychiatry*. 2016 Mar;21(3):348-56. doi: 10.1038/mp.2015.85. Epub 2015 Jun 30. PubMed PMID: 26122584; PubMed Central PMCID: PMC4696925.
- Nivard MG, Gage SH, Hottenga JJ, van Beijsterveldt CEM, Abdellaoui A, Bartels M, Baselmans BML, Ligthart L, Pourcain BS, Boomsma DI, Munafò MR, Middeldorp CM. Genetic Overlap Between Schizophrenia and Developmental Psychopathology: Longitudinal and Multivariate Polygenic Risk Prediction of Common Psychiatric Traits During Development. *Schizophr Bull*. 2017 Oct 21;43(6):1197-1207. doi: 10.1093/schbul/sbx031. PubMed PMID: 28338919; PubMed Central PMCID: PMC5737694.
- Nolte IM, van der Most PJ, Alizadeh BZ, de Bakker PI, Boezen HM, Bruinenberg M, Franke L, van der Harst P, Navis G, Postma DS, Rots MG, Stolk RP, Swertz MA, Wolfenbutter BH, Wijmenga C, Snieder H. Missing heritability: is the gap closing? An analysis of 32 complex traits in the Lifelines Cohort Study. *Eur J Hum Genet*. 2017 Jun;25(7):877-885. doi: 10.1038/ejhg.2017.50. Epub 2017 Apr 12. PubMed PMID: 28401901; PubMed Central PMCID: PMC5520063.
- Noonan JP, Coop G, Kudaravalli S, Smith D, Krause J, Alessi J, Chen F, Platt D, Pääbo S, Pritchard JK, Rubin EM. Sequencing and analysis of Neanderthal genomic DNA. *Science*. 2006 Nov 17;314(5802):1113-8. PubMed PMID: 17110569; PubMed Central PMCID: PMC2583069.
- Nordsletten AE, Larsson H, Crowley JJ, Almqvist C, Lichtenstein P, Mataix-Cols D. Patterns of Nonrandom Mating Within and Across 11 Major Psychiatric Disorders. *JAMA Psychiatry*. 2016 Apr;73(4):354-61. doi: 10.1001/jamapsychiatry.2015.3192. PubMed PMID: 26913486; PubMed Central PMCID: PMC5082975.
- O'Donovan MC, Craddock N, Norton N, Williams H, Peirce T, Moskvina V, Nikolov I, Hamshere M, Carroll L, Georgieva L, Dwyer S, Holmans P, Marchini JL, Spencer CC, Howie B, Leung HT, Hartmann AM, Möller HJ, Morris DW, Shi Y, Feng G, Hoffmann P, Propping P, Vasilescu C, Maier W, Rietschel M, Zammit

- S, Schumacher J, Quinn EM, Schulze TG, Williams NM, Giegling I, Iwata N, Ikeda M, Darvasi A, Shifman S, He L, Duan J, Sanders AR, Levinson DF, Gejman PV, Cichon S, Nöthen MM, Gill M, Corvin A, Rujescu D, Kirov G, Owen MJ, Buccola NG, Mowry BJ, Freedman R, Amin F, Black DW, Silverman JM, Byerley WF, Cloninger CR; Molecular Genetics of Schizophrenia Collaboration. Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet.* 2008 Sep;40(9):1053-5. doi: 10.1038/ng.201. PubMed PMID: 18677311.
- O'Donovan MC, Owen MJ. The implications of the shared genetics of psychiatric disorders. *Nat Med.* 2016 Nov;22(11):1214-1219. doi: 10.1038/nm.4196. Epub 2016 Oct 26. PubMed PMID: 27783064.
- O'Donovan MC. What have we learned from the Psychiatric Genomics Consortium. *World Psychiatry.* 2015 Oct;14(3):291-3. doi: 10.1002/wps.20270. PubMed PMID: 26407777; PubMed Central PMCID: PMC4592644.
- O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, Levy R, Ko A, Lee C, Smith JD, Turner EH, Stanaway IB, Vernot B, Malig M, Baker C, Reilly B, Akey JM, Borenstein E, Rieder MJ, Nickerson DA, Bernier R, Shendure J, Eichler EE. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature.* 2012 Apr 4;485(7397):246-50. doi: 10.1038/nature10989. PubMed PMID: 22495309; PubMed Central PMCID: PMC3350576.
- Otowa T, Hek K, Lee M, Byrne EM, Mirza SS, Nivard MG, Bigdeli T, Aggen SH, Adkins D, Wolen A, Fanous A, Keller MC, Castelao E, Kutalik Z, der Auwera SV, Homuth G, Nauck M, Teumer A, Milaneschi Y, Hottenga JJ, Direk N, Hofman A, Uitterlinden A, Mulder CL, Henders AK, Medland SE, Gordon S, Heath AC, Madden PA, Pergadia ML, van der Most PJ, Nolte IM, van Oort FV, Hartman CA, Oldehinkel AJ, Preisig M, Grabe HJ, Middeldorp CM, Penninx BW, Boomsma D, Martin NG, Montgomery G, Maher BS, van den Oord EJ, Wray NR, Tiemeier H, Hetttema JM. Meta-analysis of genome-wide association studies of anxiety disorders. *Mol Psychiatry.* 2016 Oct;21(10):1485. doi: 10.1038/mp.2016.11. Epub 2016 Feb 9. PubMed PMID: 26857599.
- Ottosen C, Petersen L, Larsen JT, Dalsgaard S. Gender Differences in Associations Between Attention-Deficit/Hyperactivity Disorder and Substance Use Disorder. *J Am Acad Child Adolesc Psychiatry.* 2016 Mar;55(3):227-34.e4. doi: 10.1016/j.jaac.2015.12.010. Epub 2015 Dec 24. PubMed PMID: 26903256.
- Owen MJ, Cardno AG, O'Donovan MC. Psychiatric genetics: back to the future. *Mol Psychiatry.* 2000 Jan;5(1):22-31. Review. PubMed PMID: 10673765.
- Owen MJ, O'Donovan MC, Thapar A, Craddock N. Neurodevelopmental hypothesis of schizophrenia. *Br J Psychiatry.* 2011 Mar;198(3):173-5. doi:

- 10.1192/bjp.bp.110.084384. PubMed PMID: 21357874; PubMed Central PMCID: PMC3764497.
- Owen MJ, O'Donovan MC. Schizophrenia and the neurodevelopmental continuum:evidence from genomics. *World Psychiatry*. 2017 Oct;16(3):227-235. doi: 10.1002/wps.20440. PubMed PMID: 28941101; PubMed Central PMCID: PMC5608820.
- Owen MJ, Sawa A, Mortensen PB. Schizophrenia. *Lancet*. 2016 Jul 2;388(10039):86-97. doi: 10.1016/S0140-6736(15)01121-6. Epub 2016 Jan 15. Review. PubMed PMID: 26777917; PubMed Central PMCID: PMC4940219.
- Owen MJ. Intellectual disability and major psychiatric disorders: a continuum of neurodevelopmental causality. *Br J Psychiatry*. 2012 Apr;200(4):268-9. doi: 10.1192/bjp.bp.111.105551. PubMed PMID: 22474230.
- Owen MJ. New approaches to psychiatric diagnostic classification. *Neuron*. 2014 Nov 5;84(3):564-71. doi: 10.1016/j.neuron.2014.10.028. Epub 2014 Nov 5. Review. PubMed PMID: 25442935.
- Paksarian D, Trabjerg BB, Merikangas KR, Mors O, Børglum AD, Hougaard DM, McGrath JJ, Pedersen CB, Mortensen PB, Agerbo E. The role of genetic liability in the association of urbanicity at birth and during upbringing with schizophrenia in Denmark. *Psychol Med*. 2018 Jan;48(2):305-314. doi: 10.1017/S0033291717001696. Epub 2017 Jun 29. PubMed PMID: 28659227; PubMed Central PMCID: PMC6361630.
- Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, Legge SE, Bishop S, Cameron D, Hamshere ML, Han J, Hubbard L, Lynham A, Mantripragada K, Rees E, MacCabe JH, McCarroll SA, Baune BT, Breen G, Byrne EM, Dannlowski U, Eley TC, Hayward C, Martin NG, McIntosh AM, Plomin R, Porteous DJ, Wray NR, Caballero A, Geschwind DH, Huckins LM, Ruderfer DM, Santiago E, Sklar P, Stahl EA, Won H, Agerbo E, Als TD, Andreassen OA, Bækvad-Hansen M, Mortensen PB, Pedersen CB, Børglum AD, Bybjerg-Grauholm J, Djurovic S, Durmishi N, Pedersen MG, Golimbet V, Grove J, Hougaard DM, Mattheisen M, Molden E, Mors O, Nordentoft M, Pejovic-Milovancevic M, Sigurdsson E, Silagadze T, Hansen CS, Stefansson K, Stefansson H, Steinberg S, Tosato S, Werge T; GERAD1 Consortium;; CRESTAR Consortium.; Collier DA, Rujescu D, Kirov G, Owen MJ, O'Donovan MC, Walters JTR; GERAD1 Consortium; CRESTAR Consortium; GERAD1 Consortium; CRESTAR Consortium. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet*. 2018 Mar;50(3):381-389. doi: 10.1038/s41588-018-0059-2. Epub 2018 Feb 26. PubMed PMID: 29483656; PubMed Central PMCID: PMC5918692.

- Pasaniuc B, Price AL. Dissecting the genetics of complex traits using summary association statistics. *Nat Rev Genet.* 2017 Feb;18(2):117-127. doi: 10.1038/nrg.2016.142. Epub 2016 Nov 14. Review. PubMed PMID: 27840428; PubMed Central PMCID: PMC5449190.
- Pasman JA, Verweij KJH, Gerring Z, Stringer S, Sanchez-Roige S, Treur JL, Abdellaoui A, Nivard MG, Baselmans BML, Ong JS, Ip HF, van der Zee MD, Bartels M, Day FR, Fontanillas P, Elson SL; 23andMe Research Team, de Wit H, Davis LK, MacKillop J; Substance Use Disorders Working Group of the Psychiatric Genomics Consortium; International Cannabis Consortium, Derringer JL, Branje SJT, Hartman CA, Heath AC, van Lier PAC, Madden PAF, Mägi R, Meeus W, Montgomery GW, Oldehinkel AJ, Pausova Z, Ramos-Quiroga JA, Paus T, Ribases M, Kaprio J, Boks MPM, Bell JT, Spector TD, Gelernter J, Boomsma DI, Martin NG, MacGregor S, Perry JRB, Palmer AA, Posthuma D, Munafò MR, Gillespie NA, Derks EM, Vink JM. GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. *Nat Neurosci.* 2018 Sep;21(9):1161-1170. doi: 10.1038/s41593-018-0206-1. Epub 2018 Aug 27. PubMed PMID: 30150663; PubMed Central PMCID: PMC6386176.
- Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genome-wide association studies of nearly all common variants. *Genet Epidemiol.* 2008 May;32(4):381-5. doi: 10.1002/gepi.20303. PubMed PMID: 18348202.
- Pereiro C, Pino C, Flórez G, Arrojo M, Becoña E; COPSIAD Group. Psychiatric Comorbidity in Patients from the Addictive Disorders Assistance Units of Galicia: The COPSIAD Study. *PLoS One.* 2013 Jun 18;8(6):e66451. doi: 10.1371/journal.pone.0066451. Print 2013. PubMed PMID: 23823135; PubMed Central PMCID: PMC3688915.
- Pers TH, Timshel P, Hirschhorn JN. SNPsnap: a Web-based tool for identification and annotation of matched SNPs. *Bioinformatics.* 2015 Feb 1;31(3):418-20. doi: 10.1093/bioinformatics/btu655. Epub 2014 Oct 13. PubMed PMID: 25316677; PubMed Central PMCID: PMC4308663.
- Peterson RE, Maes HH, Holmans P, Sanders AR, Levinson DF, Shi J, Kendler KS, Gejman PV, Webb BT. Genetic risk sum score comprised of common polygenic variation is associated with body mass index. *Hum Genet.* 2011 Feb;129(2):221-30. doi: 10.1007/s00439-010-0917-1. Epub 2010 Nov 23. PubMed PMID: 21104096; PubMed Central PMCID: PMC3403709.
- Pettersson E, Larsson H, Lichtenstein P. Common psychiatric disorders share the same genetic origin: a multivariate sibling study of the Swedish population. *Mol Psychiatry.* 2016 May;21(5):717-21. doi: 10.1038/mp.2015.116. Epub 2015 Aug 25. PubMed PMID: 26303662.

- Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet.* 2016 Jul;48(7):709-17. doi: 10.1038/ng.3570. Epub 2016 May 16. Erratum in: *Nat Genet.* 2016 Sep 28;48(10):1296. PubMed PMID: 27182965; PubMed Central PMCID: PMC5207801.
- Plomin R, DeFries JC, Knopik VS, Neiderhiser JM. Top 10 Replicated Findings From Behavioral Genetics. *Perspect Psychol Sci.* 2016 Jan;11(1):3-23. doi: 10.1177/1745691615617439. PubMed PMID: 26817721; PubMed Central PMCID: PMC4739500.
- Plomin R. Commentary: missing heritability, polygenic scores, and gene-environment correlation. *J Child Psychol Psychiatry.* 2013 Oct;54(10):1147-9. doi: 10.1111/jcpp.12128. PubMed PMID: 24007418; PubMed Central PMCID: PMC4033839.
- Plomin R. Finding genes in child psychology and psychiatry: when are we going to be there? *J Child Psychol Psychiatry.* 2005 Oct;46(10):1030-8. PubMed PMID: 16178925.
- Pocklington AJ, Rees E, Walters JT, Han J, Kavanagh DH, Chambert KD, Holmans P, Moran JL, McCarroll SA, Kirov G, O'Donovan MC, Owen MJ. Novel Findings from CNVs Implicate Inhibitory and Excitatory Signaling Complexes in Schizophrenia. *Neuron.* 2015 Jun 3;86(5):1203-14. doi: 10.1016/j.neuron.2015.04.022. PubMed PMID: 26050040; PubMed Central PMCID: PMC4460187.
- Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature.* 2016 Oct 13;538(7624):161-164. doi: 10.1038/538161a. PubMed PMID: 27734877; PubMed Central PMCID: PMC5089703.
- Power RA, Steinberg S, Bjornsdottir G, Rietveld CA, Abdellaoui A, Nivard MM, Johannesson M, Galesloot TE, Hottenga JJ, Willemsen G, Cesarini D, Benjamin DJ, Magnusson PK, Ullén F, Tiemeier H, Hofman A, van Rooij FJ, Walters GB, Sigurdsson E, Thorgeirsson TE, Ingason A, Helgason A, Kong A, Kiemeny LA, Koellinger P, Boomsma DI, Gudbjartsson D, Stefansson H, Stefansson K. Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. *Nat Neurosci.* 2015 Jul;18(7):953-5. doi: 10.1038/nn.4040. Epub 2015 Jun 8. PubMed PMID: 26053403.
- Power RA, Verweij KJ, Zuhair M, Montgomery GW, Henders AK, Heath AC, Madden PA, Medland SE, Wray NR, Martin NG. Genetic predisposition to schizophrenia associated with increased use of cannabis. *Mol Psychiatry.* 2014 Nov;19(11):1201-4. doi: 10.1038/mp.2014.51. Epub 2014 Jun 24. PubMed PMID: 24957864; PubMed Central PMCID: PMC4382963.

- Power RA, Wingenbach T, Cohen-Woods S, Uher R, Ng MY, Butler AW, Ising M, Craddock N, Owen MJ, Korszun A, Jones L, Jones I, Gill M, Rice JP, Maier W, Zobel A, Mors O, Placentino A, Rietschel M, Lucae S, Holsboer F, Binder EB, Keers R, Tozzi F, Muglia P, Breen G, Craig IW, Müller-Myhsok B, Kennedy JL, Strauss J, Vincent JB, Lewis CM, Farmer AE, McGuffin P. Estimating the heritability of reporting stressful life events captured by common genetic variants. *Psychol Med*. 2013 Sep;43(9):1965-71. doi: 10.1017/S0033291712002589. Epub 2012 Dec 14. PubMed PMID: 23237013.
- Poyurovsky M, Koran LM. Obsessive-compulsive disorder (OCD) with schizotypy vs. schizophrenia with OCD: diagnostic dilemmas and therapeutic implications. *J Psychiatr Res*. 2005 Jul;39(4):399-408. Epub 2004 Nov 13. Review. PubMed PMID: 15804390.
- Poyurovsky M, Zohar J, Glick I, Koran LM, Weizman R, Tandon R, Weizman A. Obsessive-compulsive symptoms in schizophrenia: implications for future psychiatric classifications. *Compr Psychiatry*. 2012 Jul;53(5):480-3. doi: 10.1016/j.comppsy.2011.08.009. Epub 2011 Oct 27. PubMed PMID: 22036006.
- Prathikanti S, McMahon FJ. Genome scans for susceptibility genes in bipolar affective disorder. *Ann Med*. 2001 May;33(4):257-62. PubMed PMID: 11405547.
- Pritchard JK, Cox NJ. The allelic architecture of human disease genes: common disease-common variant...or not? *Hum Mol Genet*. 2002 Oct 1;11(20):2417-23. Review. PubMed PMID: 12351577.
- Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000 Jun;155(2):945-59. PubMed PMID: 10835412; PubMed Central PMCID: PMC1461096.
- Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet*. 2001 Jul;69(1):124-37. Epub 2001 Jun 12. PubMed PMID: 11404818; PubMed Central PMCID: PMC1226027.
- Propping P. The biography of psychiatric genetics: from early achievements to historical burden, from an anxious society to critical geneticists. *Am J Med Genet B Neuropsychiatr Genet*. 2005 Jul 5;136B(1):2-7. PubMed PMID: 15924298.
- Psychiatric GWAS Consortium Coordinating Committee, Cichon S, Craddock N, Daly M, Faraone SV, Gejman PV, Kelsoe J, Lehner T, Levinson DF, Moran A, Sklar P, Sullivan PF. Genomewide association studies: history, rationale, and prospects for psychiatric disorders. *Am J Psychiatry*. 2009 May;166(5):540-56. doi: 10.1176/appi.ajp.2008.08091354. Epub 2009 Apr 1. Review. PubMed PMID: 19339359; PubMed Central PMCID: PMC3894622.

- Psychiatric GWAS Consortium Steering Committee. A framework for interpreting genome-wide association studies of psychiatric disorders. *Mol Psychiatry*. 2009 Jan;14(1):10-7. doi: 10.1038/mp.2008.126. Epub 2008 Nov 11. Review. PubMed PMID: 19002139.
- Psychosis Endophenotypes International Consortium; Wellcome Trust Case-Control Consortium 2, Bramon E, Pirinen M, Strange A, Lin K, Freeman C, Bellenguez C, Su Z, Band G, Pearson R, Vukcevic D, Langford C, Deloukas P, Hunt S, Gray E, Dronov S, Potter SC, Tashakkori-Ghanbaria A, Edkins S, Bumpstead SJ, Arranz MJ, Bakker S, Bender S, Bruggeman R, Cahn W, Chandler D, Collier DA, Crespo-Facorro B, Dazzan P, de Haan L, Di Forti M, Dragović M, Giegling I, Hall J, Iyegbe C, Jablensky A, Kahn RS, Kalaydjieva L, Kravariti E, Lawrie S, Linszen DH, Mata I, McDonald C, McIntosh A, Myin-Germeys I, Ophoff RA, Pariante CM, Paunio T, Picchioni M; Psychiatric Genomics Consortium, Ripke S, Rujescu D, Sauer H, Shaikh M, Sussmann J, Suvisaari J, Tosato S, Touloupoulou T, Van Os J, Walshe M, Weisbrod M, Whalley H, Wiersma D, Blackwell JM, Brown MA, Casas JP, Corvin A, Duncanson A, Jankowski JA, Markus HS, Mathew CG, Palmer CN, Plomin R, Rautanen A, Sawcer SJ, Trembath RC, Wood NW, Barroso I, Peltonen L, Lewis CM, Murray RM, Donnelly P, Powell J, Spencer CC. A genome-wide association analysis of a broad psychosis phenotype identifies three loci for further investigation. *Biol Psychiatry*. 2014 Mar 1;75(5):386-97. doi: 10.1016/j.biopsych.2013.03.033. Epub 2013 Jul 17. PubMed PMID: 23871474; PubMed Central PMCID: PMC3923972.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007 Sep;81(3):559-75. Epub 2007 Jul 25. PubMed PMID: 17701901; PubMed Central PMCID: PMC1950838.
- Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P, O'Dushlaine C, Chambert K, Bergen SE, Kähler A, Duncan L, Stahl E, Genovese G, Fernández E, Collins MO, Komiyama NH, Choudhary JS, Magnusson PK, Banks E, Shakir K, Garimella K, Fennell T, DePristo M, Grant SG, Haggarty SJ, Gabriel S, Scolnick EM, Lander ES, Hultman CM, Sullivan PF, McCarroll SA, Sklar P. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature*. 2014 Feb 13;506(7487):185-90. doi: 10.1038/nature12975. Epub 2014 Jan 22. PubMed PMID: 24463508; PubMed Central PMCID: PMC4136494.
- Rees E, Carrera N, Morgan J, Hambridge K, Escott-Price V, Pocklington AJ, Richards AL, Pardiñas AF; GROUP Investigators, McDonald C, Donohoe G, Morris DW, Kenny E, Kelleher E, Gill M, Corvin A, Kirov G, Walters JTR, Holmans P, Owen MJ, O'Donovan MC. Targeted Sequencing of 10,198 Samples Confirms Abnormalities in Neuronal Activity and Implicates Voltage-Gated Sodium Channels in Schizophrenia Pathogenesis. *Biol Psychiatry*. 2019 Apr 1;85(7):554-

562. doi: 10.1016/j.biopsych.2018.08.022. Epub 2018 Oct 1. PubMed PMID: 30420267; PubMed Central PMCID: PMC6428681.
- Rees E, Walters JT, Georgieva L, Isles AR, Chambert KD, Richards AL, Mahoney-Davies G, Legge SE, Moran JL, McCarroll SA, O'Donovan MC, Owen MJ, Kirov G. Analysis of copy number variations at 15 schizophrenia-associated loci. *Br J Psychiatry*. 2014 Feb;204(2):108-14. doi: 10.1192/bjp.bp.113.131052. Epub 2013 Dec 5. PubMed PMID: 24311552; PubMed Central PMCID: PMC3909838.
- Regier DA, Farmer ME, Rae DS, Locke BZ, Keith SJ, Judd LL, Goodwin FK. Comorbidity of mental disorders with alcohol and other drug abuse. Results from the Epidemiologic Catchment Area (ECA) Study. *JAMA*. 1990 Nov 21;264(19):2511-8. PubMed PMID: 2232018.
- Reginsson GW, Ingason A, Euesden J, Bjornsdottir G, Olafsson S, Sigurdsson E, Oskarsson H, Tyrfingsson T, Runarsdottir V, Hansdottir I, Steinberg S, Stefansson H, Gudbjartsson DF, Thorgeirsson TE, Stefansson K. Polygenic risk scores for schizophrenia and bipolar disorder associate with addiction. *Addict Biol*. 2018 Jan;23(1):485-492. doi: 10.1111/adb.12496. Epub 2017 Feb 23. PubMed PMID: 28231610; PubMed Central PMCID: PMC5811785.
- Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES. Linkage disequilibrium in the human genome. *Nature*. 2001 May 10;411(6834):199-204. PubMed PMID: 11346797.
- Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends Genet*. 2001 Sep;17(9):502-10. PubMed PMID: 11525833.
- Riglin L, Collishaw S, Richards A, Thapar AK, Maughan B, O'Donovan MC, Thapar A. Schizophrenia risk alleles and neurodevelopmental outcomes in childhood: a population-based cohort study. *Lancet Psychiatry*. 2017 Jan;4(1):57-62. doi: 10.1016/S2215-0366(16)30406-0. Epub 2016 Dec 6. PubMed PMID: 27932233.
- Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kähler AK, Akterin S, Bergen SE, Collins AL, Crowley JJ, Fromer M, Kim Y, Lee SH, Magnusson PK, Sanchez N, Stahl EA, Williams S, Wray NR, Xia K, Bettella F, Borglum AD, Bulik-Sullivan BK, Cormican P, Craddock N, de Leeuw C, Durmishi N, Gill M, Golimbet V, Hamshere ML, Holmans P, Hougaard DM, Kendler KS, Lin K, Morris DW, Mors O, Mortensen PB, Neale BM, O'Neill FA, Owen MJ, Milovancevic MP, Posthuma D, Powell J, Richards AL, Riley BP, Ruderfer D, Rujescu D, Sigurdsson E, Silagadze T, Smit AB, Stefansson H, Steinberg S, Suvisaari J, Tosato S, Verhage M, Walters JT; Multicenter Genetic Studies of Schizophrenia Consortium, Levinson DF, Gejman PV, Kendler KS, Laurent C, Mowry BJ, O'Donovan MC, Owen MJ, Pulver AE, Riley BP, Schwab SG, Wildenauer DB, Dudbridge F, Holmans P, Shi J, Albus M, Alexander M, Campion D, Cohen D, Dikeos D, Duan

- J, Eichhammer P, Godard S, Hansen M, Lerer FB, Liang KY, Maier W, Mallet J, Nertney DA, Nestadt G, Norton N, O'Neill FA, Papadimitriou GN, Ribble R, Sanders AR, Silverman JM, Walsh D, Williams NM, Wormley B; Psychosis Endophenotypes International Consortium, Arranz MJ, Bakker S, Bender S, Bramon E, Collier D, Crespo-Facorro B, Hall J, Iyegbe C, Jablensky A, Kahn RS, Kalaydjieva L, Lawrie S, Lewis CM, Lin K, Linszen DH, Mata I, McIntosh A, Murray RM, Ophoff RA, Powell J, Rujescu D, Van Os J, Walshe M, Weisbrod M, Wiersma D; Wellcome Trust Case Control Consortium 2, Donnelly P, Barroso I, Blackwell JM, Bramon E, Brown MA, Casas JP, Corvin AP, Deloukas P, Duncanson A, Jankowski J, Markus HS, Mathew CG, Palmer CN, Plomin R, Rautanen A, Sawcer SJ, Trembath RC, Viswanathan AC, Wood NW, Spencer CC, Band G, Bellenguez C, Freeman C, Hellenthal G, Giannoulatou E, Pirinen M, Pearson RD, Strange A, Su Z, Vukcevic D, Donnelly P, Langford C, Hunt SE, Edkins S, Gwilliam R, Blackburn H, Bumpstead SJ, Dronov S, Gillman M, Gray E, Hammond N, Jayakumar A, McCann OT, Liddle J, Potter SC, Ravindrarajah R, Ricketts M, Tashakkori-Ghanbaria A, Waller MJ, Weston P, Widaa S, Whittaker P, Barroso I, Deloukas P, Mathew CG, Blackwell JM, Brown MA, Corvin AP, McCarthy MI, Spencer CC, Bramon E, Corvin AP, O'Donovan MC, Stefansson K, Scolnick E, Purcell S, McCarroll SA, Sklar P, Hultman CM, Sullivan PF. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet.* 2013 Oct;45(10):1150-9. doi: 10.1038/ng.2742. Epub 2013 Aug 25. PubMed PMID: 23974872; PubMed Central PMCID: PMC3827979.
- Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science.* 1996 Sep 13;273(5281):1516-7. PubMed PMID: 8801636.
- Robins E, Guze SB. Establishment of diagnostic validity in psychiatric illness: its application to schizophrenia. *Am J Psychiatry.* 1970 Jan;126(7):983-7. PubMed PMID: 5409569.
- Robinson EB, St Pourcain B, Anttila V, Kosmicki JA, Bulik-Sullivan B, Grove J, Maller J, Samocha KE, Sanders SJ, Ripke S, Martin J, Hollegaard MV, Werge T, Hougaard DM; iPSYCH-SSI-Broad Autism Group, Neale BM, Evans DM, Skuse D, Mortensen PB, Børglum AD, Ronald A, Smith GD, Daly MJ. Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat Genet.* 2016 May;48(5):552-5. doi: 10.1038/ng.3529. Epub 2016 Mar 21. PubMed PMID: 26998691; PubMed Central PMCID: PMC4986048.
- Robinson MR, Wray NR, Visscher PM. Explaining additional genetic variation in complex traits. *Trends Genet.* 2014 Apr;30(4):124-32. doi: 10.1016/j.tig.2014.02.003. Epub 2014 Mar 11. Review. PubMed PMID: 24629526; PubMed Central PMCID: PMC4639398.

- Roosendaal B, McEwen BS, Chattarji S. Stress, memory and the amygdala. *Nat Rev Neurosci.* 2009 Jun;10(6):423-33. doi: 10.1038/nrn2651. Review. PubMed PMID: 19469026.
- Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M. Genome-wide association studies in diverse populations. *Nat Rev Genet.* 2010 May;11(5):356-66. doi: 10.1038/nrg2760. Review. PubMed PMID: 20395969; PubMed Central PMCID: PMC3079573.
- Rössler W, Riecher-Rössler A, Angst J, Murray R, Gamma A, Eich D, van Os J, Gross VA. Psychotic experiences in the general population: a twenty-year prospective community study. *Schizophr Res.* 2007 May;92(1-3):1-14. Epub 2007 Mar 23. PubMed PMID: 17363221.
- Rössler W. What is Normal? The Impact of Psychiatric Classification on Mental Health Practice and Research. *Front Public Health.* 2013 Dec 9;1:68. doi: 10.3389/fpubh.2013.00068. eCollection 2013. PubMed PMID: 24350236; PubMed Central PMCID: PMC3859926.
- Rothman EF, Edwards EM, Heeren T, Hingson RW. Adverse childhood experiences predict earlier age of drinking onset: results from a representative US sample of current or former drinkers. *Pediatrics.* 2008 Aug;122(2):e298-304. doi: 10.1542/peds.2007-3412. PubMed PMID: 18676515.
- Roussos P, Giakoumaki SG, Zouraraki C, Fullard JF, Karagiorga VE, Tsapakis EM, Petraki Z, Siever LJ, Lencz T, Malhotra A, Spanaki C, Bitsios P. The Relationship of Common Risk Variants and Polygenic Risk for Schizophrenia to Sensorimotor Gating. *Biol Psychiatry.* 2016 Jun 15;79(12):988-96. doi: 10.1016/j.biopsych.2015.06.019. Epub 2015 Jun 27. PubMed PMID: 26212897.
- Rujescu D, Ingason A, Cichon S, Pietiläinen OP, Barnes MR, Toulopoulou T, Picchioni M, Vassos E, Ettinger U, Bramon E, Murray R, Ruggeri M, Tosato S, Bonetto C, Steinberg S, Sigurdsson E, Sigmundsson T, Petursson H, Gylfason A, Olason PI, Hardarsson G, Jonsdottir GA, Gustafsson O, Fossdal R, Giegling I, Möller HJ, Hartmann AM, Hoffmann P, Crombie C, Fraser G, Walker N, Lonnqvist J, Suvisaari J, Tuulio-Henriksson A, Djurovic S, Melle I, Andreassen OA, Hansen T, Werge T, Kiemenev LA, Franke B, Veltman J, Buizer-Voskamp JE; GROUP Investigators, Sabatti C, Ophoff RA, Rietschel M, Nöthen MM, Stefansson K, Peltonen L, St Clair D, Stefansson H, Collier DA. Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum Mol Genet.* 2009 Mar 1;18(5):988-96. doi: 10.1093/hmg/ddn351. Epub 2008 Oct 22. PubMed PMID: 18945720; PubMed Central PMCID: PMC2695245.
- Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B,

- Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D; International SNP Map Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*. 2001 Feb 15;409(6822):928-33. PubMed PMID: 11237013.
- Sanchez-Roige S, Palmer AA, Fontanillas P, Elson SL; 23andMe Research Team, the Substance Use Disorder Working Group of the Psychiatric Genomics Consortium, Adams MJ, Howard DM, Edenberg HJ, Davies G, Crist RC, Deary IJ, McIntosh AM, Clarke TK. Genome-Wide Association Study Meta-Analysis of the Alcohol Use Disorders Identification Test (AUDIT) in Two Population-Based Cohorts. *Am J Psychiatry*. 2019 Feb 1;176(2):107-118. doi: 10.1176/appi.ajp.2018.18040369. Epub 2018 Oct 19. PubMed PMID: 30336701; PubMed Central PMCID: PMC6365681.
- Sanders AR, Duan J, Levinson DF, Shi J, He D, Hou C, Burrell GJ, Rice JP, Nertney DA, Olincy A, Rozic P, Vinogradov S, Buccola NG, Mowry BJ, Freedman R, Amin F, Black DW, Silverman JM, Byerley WF, Crowe RR, Cloninger CR, Martinez M, Gejman PV. No significant association of 14 candidate genes with schizophrenia in a large European ancestry sample: implications for psychiatric genetics. *Am J Psychiatry*. 2008 Apr;165(4):497-506. doi: 10.1176/appi.ajp.2007.07101573. Epub 2008 Jan 15. Erratum in: *Am J Psychiatry*. 2008 Oct;165(10):1359. PubMed PMID: 18198266.
- Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, Murtha MT, Bal VH, Bishop SL, Dong S, Goldberg AP, Jinlu C, Keaney JF 3rd, Klei L, Mandell JD, Moreno-De-Luca D, Poultney CS, Robinson EB, Smith L, Solli-Nowlan T, Su MY, Teran NA, Walker MF, Werling DM, Beaudet AL, Cantor RM, Fombonne E, Geschwind DH, Grice DE, Lord C, Lowe JK, Mane SM, Martin DM, Morrow EM, Talkowski ME, Sutcliffe JS, Walsh CA, Yu TW; Autism Sequencing Consortium, Ledbetter DH, Martin CL, Cook EH, Buxbaum JD, Daly MJ, Devlin B, Roeder K, State MW. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron*. 2015 Sep 23;87(6):1215-1233. doi: 10.1016/j.neuron.2015.09.016. PubMed PMID: 26402605; PubMed Central PMCID: PMC4624267.
- Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, Ercan-Sencicek AG, DiLullo NM, Parikshak NN, Stein JL, Walker MF, Ober GT, Teran NA, Song Y, El-Fishawy P, Murtha RC, Choi M, Overton JD, Bjornson RD, Carriero NJ, Meyer KA, Bilguvar K, Mane SM, Sestan N, Lifton RP, Günel M, Roeder K, Geschwind DH, Devlin B, State MW. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*. 2012 Apr 4;485(7397):237-41. doi: 10.1038/nature10945. PubMed PMID: 22495306; PubMed Central PMCID: PMC3667984.

- Sanders SJ, Neale BM, Huang H, Werling DM, An JY, Dong S; Whole Genome Sequencing for Psychiatric Disorders (WGSPD), Abecasis G, Arguello PA, Blangero J, Boehnke M, Daly MJ, Egan K, Geschwind DH, Glahn DC, Goldstein DB, Gur RE, Handsaker RE, McCarroll SA, Ophoff RA, Palotie A, Pato CN, Sabatti C, State MW, Willsey AJ, Hyman SE, Addington AM, Lehner T, Freimer NB. Whole genome sequencing in psychiatric disorders: the WGSPD consortium. *Nat Neurosci.* 2017 Dec;20(12):1661-1668. doi: 10.1038/s41593-017-0017-9. Erratum in: *Nat Neurosci.* 2018 Mar 16;:. PubMed PMID: 29184211.
- Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A.* 1977 Dec;74(12):5463-7. PubMed PMID: 271968; PubMed Central PMCID: PMC431765.
- Sanislow CA, Pine DS, Quinn KJ, Kozak MJ, Garvey MA, Heinssen RK, Wang PS, Cuthbert BN. Developing constructs for psychopathology research: research domain criteria. *J Abnorm Psychol.* 2010 Nov;119(4):631-9. doi: 10.1037/a0020909. PubMed PMID: 20939653.
- Schirmbeck F, Esslinger C, Rausch F, Englisch S, Meyer-Lindenberg A, Zink M. Antiserotonergic antipsychotics are associated with obsessive-compulsive symptoms in schizophrenia. *Psychol Med.* 2011 Nov;41(11):2361-73. doi: 10.1017/S0033291711000419. Epub 2011 Apr 5. PubMed PMID: 21466748.
- Schirmbeck F, Zink M. Comorbid obsessive-compulsive symptoms in schizophrenia: contributions of pharmacological and genetic factors. *Front Pharmacol.* 2013 Aug 9;4:99. doi: 10.3389/fphar.2013.00099. eCollection 2013. PubMed PMID: 23950745; PubMed Central PMCID: PMC3738863.
- Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet.* 2011 Sep 18;43(10):969-76. doi: 10.1038/ng.940. PubMed PMID: 21926974; PubMed Central PMCID: PMC3303194.
- Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature.* 2014 Jul 24;511(7510):421-7. doi: 10.1038/nature13595. Epub 2014 Jul 22. PubMed PMID: 25056061; PubMed Central PMCID: PMC4112379.
- Schork AJ, Won H, Appadurai V, Nudel R, Gandal M, Delaneau O, Revsbech Christiansen M, Hougaard DM, Bækved-Hansen M, Bybjerg-Grauholm J, Giørtz Pedersen M, Agerbo E, Bøcker Pedersen C, Neale BM, Daly MJ, Wray NR, Nordentoft M, Mors O, Børglum AD, Bo Mortensen P, Buil A, Thompson WK, Geschwind DH, Werge T. A genome-wide association study of shared risk across psychiatric disorders implicates gene regulation during fetal neurodevelopment. *Nat Neurosci.* 2019 Mar;22(3):353-361. doi: 10.1038/s41593-018-0320-0. Epub 2019 Jan 28. PubMed PMID: 30692689.

- Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev.* 2009 Jun;19(3):212-9. doi:10.1016/j.gde.2009.04.010. Epub 2009 May 28. Review. PubMed PMID: 19481926; PubMed Central PMCID: PMC2914559.
- Schwandt ML, Heilig M, Hommer DW, George DT, Ramchandani VA. Childhood trauma exposure and alcohol dependence severity in adulthood: mediation by emotional abuse severity and neuroticism. *Alcohol Clin Exp Res.* 2013 Jun;37(6):984-92. doi: 10.1111/acer.12053. Epub 2012 Dec 27. PubMed PMID: 23278300; PubMed Central PMCID: PMC3620963.
- Schwartz JE, Fennig S, Tanenberg-Karant M, Carlson G, Craig T, Galambos N, Lavelle J, Bromet EJ. Congruence of diagnoses 2 years after a first-admission diagnosis of psychosis. *Arch Gen Psychiatry.* 2000 Jun;57(6):593-600. PubMed PMID: 10839338.
- Scotti-Muzzi E, Saide OL. Schizo-obsessive spectrum disorders: an update. *CNS Spectr.* 2017 Jun;22(3):258-272. doi: 10.1017/S1092852916000390. Epub 2016 Sep 27. Review. PubMed PMID: 27669819.
- Scutari M, Mackay I, Balding D. Using Genetic Distance to Infer the Accuracy of Genomic Prediction. *PLoS Genet.* 2016 Sep 2;12(9):e1006288. doi: 10.1371/journal.pgen.1006288. eCollection 2016 Sep. PubMed PMID: 27589268; PubMed Central PMCID: PMC5010218.
- Segurado R, Detera-Wadleigh SD, Levinson DF, Lewis CM, Gill M, Nurnberger JI Jr, Craddock N, DePaulo JR, Baron M, Gershon ES, Ekholm J, Cichon S, Turecki G, Claes S, Kelsoe JR, Schofield PR, Badenhop RF, Morissette J, Coon H, Blackwood D, McInnes LA, Foroud T, Edenberg HJ, Reich T, Rice JP, Goate A, McInnis MG, McMahon FJ, Badner JA, Goldin LR, Bennett P, Willour VL, Zandi PP, Liu J, Gilliam C, Joo SH, Berrettini WH, Yoshikawa T, Peltonen L, Lönnqvist J, Nöthen MM, Schumacher J, Windemuth C, Rietschel M, Propping P, Maier W, Alda M, Grof P, Rouleau GA, Del-Favero J, Van Broeckhoven C, Mendlewicz J, Adolfsson R, Spence MA, Luebbert H, Adams LJ, Donald JA, Mitchell PB, Barden N, Shink E, Byerley W, Muir W, Visscher PM, Macgregor S, Gurling H, Kalsi G, McQuillin A, Escamilla MA, Reus VI, Leon P, Freimer NB, Ewald H, Kruse TA, Mors O, Radhakrishna U, Blouin JL, Antonarakis SE, Akarsu N. Genome scan meta-analysis of schizophrenia and bipolar disorder, part III: Bipolar disorder. *Am J Hum Genet.* 2003 Jul;73(1):49-62. Epub 2003 Jun 11. PubMed PMID: 12802785; PubMed Central PMCID: PMC1180589.
- Selzam S, Coleman JRI, Caspi A, Moffitt TE, Plomin R. A polygenic p factor for major psychiatric disorders. *Transl Psychiatry.* 2018 Oct 2;8(1):205. doi: 10.1038/s41398-018-0217-4. PubMed PMID: 30279410; PubMed Central PMCID: PMC6168558.

- Sengupta SM, MacDonald K, Fathalli F, Yim A, Lepage M, Iyer S, Malla A, Joobner R. Polygenic Risk Score associated with specific symptom dimensions in first-episode psychosis. *Schizophr Res.* 2017 Jun;184:116-121. doi: 10.1016/j.schres.2016.11.039. Epub 2016 Dec 2. PubMed PMID: 27916287.
- Serrano M, Moreno M, Bassols J, Moreno-Navarrete JM, Ortega F, Ricart W, Fernández-Real JM. Coxsackie and adenovirus receptor is increased in adipose tissue of obese subjects: a role for adenovirus infection? *J Clin Endocrinol Metab.* 2015 Mar;100(3):1156-63. doi: 10.1210/jc.2014-3791. Epub 2014 Dec 2. PubMed PMID: 25459915.
- Sham PC, Purcell SM. Statistical power and significance testing in large-scale genetic studies. *Nat Rev Genet.* 2014 May;15(5):335-46. doi: 10.1038/nrg3706. Review. PubMed PMID: 24739678.
- Sherrington R, Brynjolfsson J, Petursson H, Potter M, Dudleston K, Barraclough B, Wasmuth J, Dobbs M, Gurling H. Localization of a susceptibility locus for schizophrenia on chromosome 5. *Nature.* 1988 Nov 10;336(6195):164-7. PubMed PMID: 2903449.
- Sherry ST, Ward M, Sirotkin K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res.* 1999 Aug;9(8):677-9. Review. PubMed PMID: 10447503.
- Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 2001 Jan 1;29(1):308-11. PubMed PMID: 11125122; PubMed Central PMCID: PMC29783.
- Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Pe'er I, Dudbridge F, Holmans PA, Whittemore AS, Mowry BJ, Olincy A, Amin F, Cloninger CR, Silverman JM, Buccola NG, Byerley WF, Black DW, Crowe RR, Oksenberg JR, Mirel DB, Kendler KS, Freedman R, Gejman PV. Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature.* 2009 Aug 6;460(7256):753-7. doi: 10.1038/nature08192. Epub 2009 Jul 1. PubMed PMID: 19571809; PubMed Central PMCID: PMC2775422.
- Shi Y, Li Z, Xu Q, Wang T, Li T, Shen J, Zhang F, Chen J, Zhou G, Ji W, Li B, Xu Y, Liu D, Wang P, Yang P, Liu B, Sun W, Wan C, Qin S, He G, Steinberg S, Cichon S, Werge T, Sigurdsson E, Tosato S, Palotie A, Nöthen MM, Rietschel M, Ophoff RA, Collier DA, Rujescu D, Clair DS, Stefansson H, Stefansson K, Ji J, Wang Q, Li W, Zheng L, Zhang H, Feng G, He L. Common variants on 8p12 and 1q24.2 confer risk of schizophrenia. *Nat Genet.* 2011 Oct 30;43(12):1224-7. doi: 10.1038/ng.980. PubMed PMID: 22037555; PubMed Central PMCID: PMC3773910.

- Shifman S, Johannesson M, Bronstein M, Chen SX, Collier DA, Craddock NJ, Kendler KS, Li T, O'Donovan M, O'Neill FA, Owen MJ, Walsh D, Weinberger DR, Sun C, Flint J, Darvasi A. Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women. *PLoS Genet.* 2008 Feb;4(2):e28. doi: 10.1371/journal.pgen.0040028. PubMed PMID: 18282107; PubMed Central PMCID: PMC2242812.
- Shorter E. The history of nosology and the rise of the Diagnostic and Statistical Manual of Mental Disorders. *Dialogues Clin Neurosci.* 2015 Mar;17(1):59-67. Review. PubMed PMID: 25987864; PubMed Central PMCID: PMC4421901.
- Singh T, Kurki MI, Curtis D, Purcell SM, Crooks L, McRae J, Suvisaari J, Chheda H, Blackwood D, Breen G, Pietiläinen O, Gerety SS, Ayub M, Blyth M, Cole T, Collier D, Coomber EL, Craddock N, Daly MJ, Danesh J, DiForti M, Foster A, Freimer NB, Geschwind D, Johnstone M, Joss S, Kirov G, Körkkö J, Kuismin O, Holmans P, Hultman CM, Iyegbe C, Lönngqvist J, Männikkö M, McCarroll SA, McGuffin P, McIntosh AM, McQuillin A, Moilanen JS, Moore C, Murray RM, Newbury-Ecob R, Ouwehand W, Paunio T, Prigmore E, Rees E, Roberts D, Sambrook J, Sklar P, St Clair D, Vejjola J, Walters JT, Williams H; Swedish Schizophrenia Study; INTERVAL Study; DDD Study; UK10 K Consortium, Sullivan PF, Hurler ME, O'Donovan MC, Palotie A, Owen MJ, Barrett JC. Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nat Neurosci.* 2016 Apr;19(4):571-7. doi: 10.1038/nn.4267. Epub 2016 Mar 14. PubMed PMID: 26974950.
- Sklar P, Smoller JW, Fan J, Ferreira MA, Perlis RH, Chambert K, Nimgaonkar VL, McQueen MB, Faraone SV, Kirby A, de Bakker PI, Ogdie MN, Thase ME, Sachs GS, Todd-Brown K, Gabriel SB, Sognez C, Gates C, Blumenstiel B, Defelice M, Ardlie KG, Franklin J, Muir WJ, McGhee KA, MacIntyre DJ, McLean A, VanBeck M, McQuillin A, Bass NJ, Robinson M, Lawrence J, Anjorin A, Curtis D, Scolnick EM, Daly MJ, Blackwood DH, Gurling HM, Purcell SM. Whole-genome association study of bipolar disorder. *Mol Psychiatry.* 2008 Jun;13(6):558-69. doi: 10.1038/sj.mp.4002151. Epub 2008 Mar 4. PubMed PMID: 18317468; PubMed Central PMCID: PMC3777816.
- Sklar P. Linkage analysis in psychiatric disorders: the emerging picture. *Annu Rev Genomics Hum Genet.* 2002;3:371-413. Epub 2002 Apr 15. Review. PubMed PMID: 12142356.
- Slater E, Cowie V. *The Genetics of Mental Disorders* London. United Kingdom: Oxford University Press. 1971.
- Slater E. The monogenic theory of schizophrenia. *Acta Genet Stat Med.* 1958;8(1):50-6. PubMed PMID: 13532420.

- Slatkin M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nat Rev Genet.* 2008 Jun;9(6):477-85. doi: 10.1038/nrg2361. Review. PubMed PMID: 18427557; PubMed Central PMCID: PMC5124487.
- Smeland OB, Bahrami S, Frei O, Shadrin A, O'Connell K, Savage J, Watanabe K, Krull F, Bettella F, Steen NE, Ueland T, Posthuma D, Djurovic S, Dale AM, Andreassen OA. Genome-wide analysis reveals extensive genetic overlap between schizophrenia, bipolar disorder, and intelligence. *Mol Psychiatry.* 2019 Jan 4. doi: 10.1038/s41380-018-0332-x. [Epub ahead of print] PubMed PMID: 30610197.
- Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol.* 2003 Feb;32(1):1-22. PubMed PMID: 12689998.
- Smoller JW, Andreassen OA, Edenberg HJ, Faraone SV, Glatt SJ, Kendler KS. Psychiatric genetics and the structure of psychopathology. *Mol Psychiatry.* 2019 Mar;24(3):409-420. doi: 10.1038/s41380-017-0010-4. Epub 2018 Jan 9. Review. Erratum in: *Mol Psychiatry.* 2018 Mar 14;:. PubMed PMID: 29317742.
- Smoller JW. A Quarter Century of Progress in Psychiatric Genetics. *Harv Rev Psychiatry.* 2017 Nov/Dec;25(6):256-258. doi: 10.1097/HRP.000000000000180. Review. PubMed PMID: 29117019.
- Smoller JW. Disorders and borders: psychiatric genetics and nosology. *Am J Med Genet B Neuropsychiatr Genet.* 2013 Oct;162B(7):559-78. doi: 10.1002/ajmg.b.32174. Review. PubMed PMID: 24132891.
- Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: challenges and strategies. *Nat Rev Genet.* 2013 Jul;14(7):483-95. doi: 10.1038/nrg3461. Epub 2013 Jun 11. Review. PubMed PMID: 23752797; PubMed Central PMCID: PMC4104202.
- Speed D, Cai N; UCLEB Consortium, Johnson MR, Nejentsev S, Balding DJ. Reevaluation of SNP heritability in complex human traits. *Nat Genet.* 2017 Jul;49(7):986-992. doi: 10.1038/ng.3865. Epub 2017 May 22. PubMed PMID: 28530675; PubMed Central PMCID: PMC5493198.
- Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, Lango Allen H, Lindgren CM, Luan J, Mägi R, Randall JC, Vedantam S, Winkler TW, Qi L, Workalemahu T, Heid IM, Steinthorsdottir V, Stringham HM, Weedon MN, Wheeler E, Wood AR, Ferreira T, Weyant RJ, Segrè AV, Estrada K, Liang L, Nemes J, Park JH, Gustafsson S, Kilpeläinen TO, Yang J, Bouatia-Naji N, Esko T, Feitosa MF, Kutalik Z, Mangino M, Raychaudhuri S, Scherag A, Smith AV, Welch R, Zhao JH, Aben KK, Absher DM, Amin N, Dixon AL, Fisher E, Glazer NL, Goddard ME, Heard-Costa NL, Hoesel V, Hottenga JJ, Johansson A, Johnson T, Ketkar S, Lamina C, Li S, Moffatt MF, Myers RH, Narisu N, Perry JR, Peters

MJ, Preuss M, Ripatti S, Rivadeneira F, Sandholt C, Scott LJ, Timpson NJ, Tyrer JP, van Wingerden S, Watanabe RM, White CC, Wiklund F, Barlassina C, Chasman DI, Cooper MN, Jansson JO, Lawrence RW, Pellikka N, Prokopenko I, Shi J, Thiering E, Alavere H, Alibrandi MT, Almgren P, Arnold AM, Aspelund T, Atwood LD, Balkau B, Balmforth AJ, Bennett AJ, Ben-Shlomo Y, Bergman RN, Bergmann S, Biebermann H, Blakemore AI, Boes T, Bonnycastle LL, Bornstein SR, Brown MJ, Buchanan TA, Busonero F, Campbell H, Cappuccio FP, Cavalcanti-Proença C, Chen YD, Chen CM, Chines PS, Clarke R, Coin L, Connell J, Day IN, den Heijer M, Duan J, Ebrahim S, Elliott P, Elosua R, Eiriksdottir G, Erdos MR, Eriksson JG, Facheris MF, Felix SB, Fischer-Posovszky P, Folsom AR, Friedrich N, Freimer NB, Fu M, Gaget S, Gejman PV, Geus EJ, Gieger C, Gjessing AP, Goel A, Goyette P, Grallert H, Grässler J, Greenawalt DM, Groves CJ, Gudnason V, Guiducci C, Hartikainen AL, Hassanali N, Hall AS, Havulinna AS, Hayward C, Heath AC, Hengstenberg C, Hicks AA, Hinney A, Hofman A, Homuth G, Hui J, Igl W, Iribarren C, Isomaa B, Jacobs KB, Jarick I, Jewell E, John U, Jørgensen T, Jousilahti P, Jula A, Kaakinen M, Kajantie E, Kaplan LM, Kathiresan S, Kettunen J, Kinnunen L, Knowles JW, Kolcic I, König IR, Koskinen S, Kovacs P, Kuusisto J, Kraft P, Kvaløy K, Laitinen J, Lantieri O, Lanzani C, Launer LJ, Lecoeur C, Lehtimäki T, Lettre G, Liu J, Lokki ML, Lorentzon M, Luben RN, Ludwig B; MAGIC, Manunta P, Marek D, Marre M, Martin NG, McArdle WL, McCarthy A, McKnight B, Meitinger T, Melander O, Meyre D, Midhjelld K, Montgomery GW, Morken MA, Morris AP, Mulic R, Ngwa JS, Nelis M, Neville MJ, Nyholt DR, O'Donnell CJ, O'Rahilly S, Ong KK, Oostra B, Paré G, Parker AN, Perola M, Pichler I, Pietiläinen KH, Platou CG, Polasek O, Pouta A, Rafelt S, Raitakari O, Rayner NW, Ridderstråle M, Rief W, Ruukonen A, Robertson NR, Rzehak P, Salomaa V, Sanders AR, Sandhu MS, Sanna S, Saramies J, Savolainen MJ, Scherag S, Schipf S, Schreiber S, Schunkert H, Silander K, Sinisalo J, Siscovick DS, Smit JH, Soranzo N, Sovio U, Stephens J, Surakka I, Swift AJ, Tammesoo ML, Tardif JC, Teder-Laving M, Teslovich TM, Thompson JR, Thomson B, Tönjes A, Tuomi T, van Meurs JB, van Ommen GJ, Vatin V, Viikari J, Visvikis-Siest S, Vitart V, Vogel CI, Voight BF, Waite LL, Wallaschofski H, Walters GB, Widen E, Wiegand S, Wild SH, Willemssen G, Witte DR, Wittteman JC, Xu J, Zhang Q, Zgaga L, Ziegler A, Zitting P, Beilby JP, Farooqi IS, Hebebrand J, Huikuri HV, James AL, Kähönen M, Levinson DF, Macciardi F, Nieminen MS, Ohlsson C, Palmer LJ, Ridker PM, Stumvoll M, Beckmann JS, Boeing H, Boerwinkle E, Boomsma DI, Caulfield MJ, Chanock SJ, Collins FS, Cupples LA, Smith GD, Erdmann J, Froguel P, Grönberg H, Gyllenstein U, Hall P, Hansen T, Harris TB, Hattersley AT, Hayes RB, Heinrich J, Hu FB, Hveem K, Illig T, Jarvelin MR, Kaprio J, Karpe F, Khaw KT, Kiemenev LA, Krude H, Laakso M, Lawlor DA, Metspalu A, Munroe PB, Ouwehand WH, Pedersen O, Penninx BW, Peters A, Pramstaller PP, Quertermous T, Reinehr T, Rissanen A, Rudan I, Samani NJ, Schwarz PE, Shuldiner AR, Spector TD, Tuomilehto J, Uda M, Uitterlinden A, Valle TT, Wabitsch M, Waeber G, Wareham NJ, Watkins H; Procardis Consortium, Wilson JF, Wright AF, Zillikens

- MC, Chatterjee N, McCarroll SA, Purcell S, Schadt EE, Visscher PM, Assimes TL, Borecki IB, Deloukas P, Fox CS, Groop LC, Haritunians T, Hunter DJ, Kaplan RC, Mohlke KL, O'Connell JR, Peltonen L, Schlessinger D, Strachan DP, van Duijn CM, Wichmann HE, Frayling TM, Thorsteinsdottir U, Abecasis GR, Barroso I, Boehnke M, Stefansson K, North KE, McCarthy MI, Hirschhorn JN, Ingelsson E, Loos RJ. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet.* 2010 Nov;42(11):937-48. doi: 10.1038/ng.686. Epub 2010 Oct 10. PubMed PMID: 20935630; PubMed Central PMCID: PMC3014648.
- Spitzer RL. Psychiatric diagnosis: are clinicians still necessary? *Compr Psychiatry.* 1983;24(5):399-411.
- St Pourcain B, Robinson EB, Anttila V, Sullivan BB, Maller J, Golding J, Skuse D, Ring S, Evans DM, Zammit S, Fisher SE, Neale BM, Anney RJL, Ripke S, Hollegaard MV, Werge T; iPSYCH-SSI-Broad Autism Group, Ronald A, Grove J, Hougaard DM, Børglum AD, Mortensen PB, Daly MJ, Davey Smith G. ASD and schizophrenia show distinct developmental profiles in common genetic overlap with population-based social communication difficulties. *Mol Psychiatry.* 2018 Feb;23(2):263-270. doi: 10.1038/mp.2016.198. Epub 2017 Jan 3. PubMed PMID: 28044064; PubMed Central PMCID: PMC5382976.
- Stankiewicz P, Lupski JR. Genome architecture, rearrangements and genomic disorders. *Trends Genet.* 2002 Feb;18(2):74-82. Review. PubMed PMID: 11818139.
- Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. *Annu Rev Med.* 2010;61:437-55. doi: 10.1146/annurev-med-100708-204735. Review. PubMed PMID: 20059347.
- Steel Z, Marnane C, Iranpour C, Chey T, Jackson JW, Patel V, Silove D. The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013. *Int J Epidemiol.* 2014 Apr;43(2):476-93. doi: 10.1093/ije/dyu038. Epub 2014 Mar 19. Review. PubMed PMID: 24648481; PubMed Central PMCID: PMC3997379.
- Stefansson H, Meyer-Lindenberg A, Steinberg S, Magnusdottir B, Morgen K, Arnarsdottir S, Bjornsdottir G, Walters GB, Jonsdottir GA, Doyle OM, Tost H, Grimm O, Kristjansdottir S, Snorrason H, Davidsdottir SR, Gudmundsson LJ, Jonsson GF, Stefansdottir B, Helgadóttir I, Haraldsson M, Jonsdottir B, Thygesen JH, Schwarz AJ, Didriksen M, Stensbøl TB, Brammer M, Kapur S, Halldorsson JG, Hreidarsson S, Saemundsen E, Sigurdsson E, Stefansson K. CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature.* 2014 Jan 16;505(7483):361-6. doi: 10.1038/nature12818. Epub 2013 Dec 18. PubMed PMID: 24352232.

Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D, Werge T, Pietiläinen OP, Mors O, Mortensen PB, Sigurdsson E, Gustafsson O, Nyegaard M, Tuulio-Henriksson A, Ingason A, Hansen T, Suvisaari J, Lonnqvist J, Paunio T, Børghlum AD, Hartmann A, Fink-Jensen A, Nordentoft M, Hougaard D, Norgaard-Pedersen B, Böttcher Y, Olesen J, Breuer R, Möller HJ, Giegling I, Rasmussen HB, Timm S, Mattheisen M, Bitter I, Réthelyi JM, Magnusdottir BB, Sigmundsson T, Olason P, Masson G, Gulcher JR, Haraldsson M, Fossdal R, Thorgeirsson TE, Thorsteinsdottir U, Ruggeri M, Tosato S, Franke B, Strengman E, Kiemeny LA; Genetic Risk and Outcome in Psychosis (GROUP), Melle I, Djurovic S, Abramova L, Kaleda V, Sanjuan J, de Frutos R, Bramon E, Vassos E, Fraser G, Ettinger U, Picchioni M, Walker N, Touloupoulou T, Need AC, Ge D, Yoon JL, Shianna KV, Freimer NB, Cantor RM, Murray R, Kong A, Golimbet V, Carracedo A, Arango C, Costas J, Jönsson EG, Terenius L, Agartz I, Petursson H, Nöthen MM, Rietschel M, Matthews PM, Muglia P, Peltonen L, St Clair D, Goldstein DB, Stefansson K, Collier DA. Common variants conferring risk of schizophrenia. *Nature*. 2009 Aug 6;460(7256):744-7. doi: 10.1038/nature08186. Epub 2009 Jul 1. PubMed PMID: 19571808; PubMed Central PMCID: PMC3077530.

Stefansson H, Rujescu D, Cichon S, Pietiläinen OP, Ingason A, Steinberg S, Fossdal R, Sigurdsson E, Sigmundsson T, Buizer-Voskamp JE, Hansen T, Jakobsen KD, Muglia P, Francks C, Matthews PM, Gylfason A, Halldorsson BV, Gudbjartsson D, Thorgeirsson TE, Sigurdsson A, Jonasdottir A, Jonasdottir A, Bjornsson A, Mattiasdottir S, Blondal T, Haraldsson M, Magnusdottir BB, Giegling I, Möller HJ, Hartmann A, Shianna KV, Ge D, Need AC, Crombie C, Fraser G, Walker N, Lonnqvist J, Suvisaari J, Tuulio-Henriksson A, Paunio T, Touloupoulou T, Bramon E, Di Forti M, Murray R, Ruggeri M, Vassos E, Tosato S, Walshe M, Li T, Vasilescu C, Mühleisen TW, Wang AG, Ullum H, Djurovic S, Melle I, Olesen J, Kiemeny LA, Franke B; GROUP, Sabatti C, Freimer NB, Gulcher JR, Thorsteinsdottir U, Kong A, Andreassen OA, Ophoff RA, Georgi A, Rietschel M, Werge T, Petursson H, Goldstein DB, Nöthen MM, Peltonen L, Collier DA, St Clair D, Stefansson K. Large recurrent microdeletions associated with schizophrenia. *Nature*. 2008 Sep 11;455(7210):232-6. doi: 10.1038/nature07229. PubMed PMID: 18668039; PubMed Central PMCID: PMC2687075.

Steinberg S, de Jong S; Irish Schizophrenia Genomics Consortium, Andreassen OA, Werge T, Børghlum AD, Mors O, Mortensen PB, Gustafsson O, Costas J, Pietiläinen OP, Demontis D, Papiol S, Huttenlocher J, Mattheisen M, Breuer R, Vassos E, Giegling I, Fraser G, Walker N, Tuulio-Henriksson A, Suvisaari J, Lonnqvist J, Paunio T, Agartz I, Melle I, Djurovic S, Strengman E; GROUP, Jürgens G, Glenthøj B, Terenius L, Hougaard DM, Ørntoft T, Wiuf C, Didriksen M, Hollegaard MV, Nordentoft M, van Winkel R, Kenis G, Abramova L, Kaleda V, Arrojo M, Sanjuán J, Arango C, Sperling S, Rossner M, Ribolsi M, Magni V, Siracusano A, Christiansen C, Kiemeny LA, Veldink J, van den Berg L, Ingason

- A, Muglia P, Murray R, Nöthen MM, Sigurdsson E, Petursson H, Thorsteinsdottir U, Kong A, Rubino IA, De Hert M, Réthelyi JM, Bitter I, Jönsson EG, Golimbet V, Carracedo A, Ehrenreich H, Craddock N, Owen MJ, O'Donovan MC; Wellcome Trust Case Control Consortium 2, Ruggeri M, Tosato S, Peltonen L, Ophoff RA, Collier DA, St Clair D, Rietschel M, Cichon S, Stefansson H, Rujescu D, Stefansson K. Common variants at VRK2 and TCF4 conferring risk of schizophrenia. *Hum Mol Genet.* 2011 Oct 15;20(20):4076-81. doi: 10.1093/hmg/ddr325. Epub 2011 Jul 26. PubMed PMID: 21791550; PubMed Central PMCID: PMC3298077.
- Stephan KE, Bach DR, Fletcher PC, Flint J, Frank MJ, Friston KJ, Heinz A, Huys QJM, Owen MJ, Binder EB, Dayan P, Johnstone EC, Meyer-Lindenberg A, Montague PR, Schnyder U, Wang XJ, Breakspear M. Charting the landscape of priority problems in psychiatry, part 1: classification and diagnosis. *Lancet Psychiatry.* 2016 Jan;3(1):77-83. doi: 10.1016/S2215-0366(15)00361-2. Epub 2015 Nov 11. Review. PubMed PMID: 26573970.
- Stepniak B, Papiol S, Hammer C, Ramin A, Everts S, Hennig L, Begemann M, Ehrenreich H. Accumulated environmental risk determining age at schizophrenia onset: a deep phenotyping-based study. *Lancet Psychiatry.* 2014 Nov;1(6):444-53. doi: 10.1016/S2215-0366(14)70379-7. Epub 2014 Nov 5. PubMed PMID: 26361199.
- Stessman HA, Xiong B, Coe BP, Wang T, Hoekzema K, Fenckova M, Kvarnung M, Gerds J, Trinh S, Cosemans N, Vives L, Lin J, Turner TN, Santen G, Ruivenkamp C, Kriek M, van Haeringen A, Aten E, Friend K, Liebelt J, Barnett C, Haan E, Shaw M, Geetz J, Anderlid BM, Nordgren A, Lindstrand A, Schwartz C, Kooy RF, Vandeweyer G, Helsmoortel C, Romano C, Alberti A, Vinci M, Avola E, Giusto S, Courchesne E, Pramparo T, Pierce K, Nalabolu S, Amaral DG, Scheffer IE, Delatycki MB, Lockhart PJ, Hormozdiari F, Harich B, Castells-Nobau A, Xia K, Peeters H, Nordenskjöld M, Schenck A, Bernier RA, Eichler EE. Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nat Genet.* 2017 Apr;49(4):515-526. doi: 10.1038/ng.3792. Epub 2017 Feb 13. PubMed PMID: 28191889; PubMed Central PMCID: PMC5374041.
- Stip E, Letourneau G. Psychotic symptoms as a continuum between normality and pathology. *Can J Psychiatry.* 2009 Mar;54(3):140-51. Review. PubMed PMID: 19321018.
- Sturt E, McGuffin P. Can linkage and marker association resolve the genetic aetiology of psychiatric disorders? Review and argument. *Psychol Med.* 1985 Aug;15(3):455-62. PubMed PMID: 3863156.
- Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, Zhang Y, Ye K, Jun G, Fritz MH, Konkel MK, Malhotra A, Stütz AM, Shi X, Casale FP,

- Chen J, Hormozdiari F, Dayama G, Chen K, Malig M, Chaisson MJP, Walter K, Meiers S, Kashin S, Garrison E, Auton A, Lam HYK, Mu XJ, Alkan C, Antaki D, Bae T, Cerveira E, Chines P, Chong Z, Clarke L, Dal E, Ding L, Emery S, Fan X, Gujral M, Kahveci F, Kidd JM, Kong Y, Lameijer EW, McCarthy S, Flicek P, Gibbs RA, Marth G, Mason CE, Menelaou A, Muzny DM, Nelson BJ, Noor A, Parrish NF, Pendleton M, Quitadamo A, Raeder B, Schadt EE, Romanovitch M, Schlattl A, Sebra R, Shabalin AA, Untergasser A, Walker JA, Wang M, Yu F, Zhang C, Zhang J, Zheng-Bradley X, Zhou W, Zichner T, Sebat J, Batzer MA, McCarroll SA; 1000 Genomes Project Consortium, Mills RE, Gerstein MB, Bashir A, Stegle O, Devine SE, Lee C, Eichler EE, Korbelt JO. An integrated map of structural variation in 2,504 human genomes. *Nature*. 2015 Oct 1;526(7571):75-81. doi: 10.1038/nature15394. PubMed PMID: 26432246; PubMed Central PMCID: PMC4617611.
- Sullivan P; 96 Psychiatric Genetics Investigators. Don't give up on GWAS. *Mol Psychiatry*. 2012 Jan;17(1):2-3. doi: 10.1038/mp.2011.94. Epub 2011 Aug 9. PubMed PMID: 21826059; PubMed Central PMCID: PMC4187109.
- Sullivan PF, Agrawal A, Bulik CM, Andreassen OA, Børglum AD, Breen G, Cichon S, Edenberg HJ, Faraone SV, Gelernter J, Mathews CA, Nievergelt CM, Smoller JW, O'Donovan MC; Psychiatric Genomics Consortium. Psychiatric Genomics: An Update and an Agenda. *Am J Psychiatry*. 2018 Jan 1;175(1):15-27. doi: 10.1176/appi.ajp.2017.17030283. Epub 2017 Oct 3. Review. PubMed PMID: 28969442; PubMed Central PMCID: PMC5756100.
- Sullivan PF, Geschwind DH. Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. *Cell*. 2019 Mar 21;177(1):162-183. doi: 10.1016/j.cell.2019.01.015. Review. PubMed PMID: 30901538; PubMed Central PMCID: PMC6432948.
- Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry*. 2003 Dec;60(12):1187-92. PubMed PMID: 14662550.
- Sullivan PF, Lin D, Tzeng JY, van den Oord E, Perkins D, Stroup TS, Wagner M, Lee S, Wright FA, Zou F, Liu W, Downing AM, Lieberman J, Close SL. Genomewide association for schizophrenia in the CATIE study: results of stage 1. *Mol Psychiatry*. 2008 Jun;13(6):570-84. doi: 10.1038/mp.2008.25. Epub 2008 Mar 18. Erratum in: *Mol Psychiatry*. 2009 Dec;14(12):1144. PubMed PMID: 18347602; PubMed Central PMCID: PMC3910086.
- Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry*. 2000 Oct;157(10):1552-62. PubMed PMID: 11007705.

- Sullivan PF. How Good Were Candidate Gene Guesses in Schizophrenia Genetics? *Biol Psychiatry*. 2017 Nov 15;82(10):696-697. doi: 10.1016/j.biopsych.2017.09.004. PubMed PMID: 29031917.
- Suvisaari J, Mantere O, Keinänen J, Mäntylä T, Rikandi E, Lindgren M, Kieseppä T, Raji TT. Is It Possible to Predict the Future in First-Episode Psychosis? *Front Psychiatry*. 2018 Nov 13;9:580. doi: 10.3389/fpsy.2018.00580. eCollection 2018. Review. PubMed PMID: 30483163; PubMed Central PMCID: PMC6243124.
- Sveinbjornsson G, Albrechtsen A, Zink F, Gudjonsson SA, Oddson A, Måsson G, Holm H, Kong A, Thorsteinsdottir U, Sulem P, Gudbjartsson DF, Stefansson K. Weighting sequence variants based on their annotation increases power of whole-genome association studies. *Nat Genet*. 2016 Mar;48(3):314-7. doi: 10.1038/ng.3507. Epub 2016 Feb 8. PubMed PMID: 26854916.
- Takata A, Xu B, Ionita-Laza I, Roos JL, Gogos JA, Karayiorgou M. Loss-of-function variants in schizophrenia risk and SETD1A as a candidate susceptibility gene. *Neuron*. 2014 May 21;82(4):773-80. doi: 10.1016/j.neuron.2014.04.043. PubMed PMID: 24853937; PubMed Central PMCID: PMC4387883.
- Tansey KE, Rees E, Linden DE, Ripke S, Chambert KD, Moran JL, McCarroll SA, Holmans P, Kirov G, Walters J, Owen MJ, O'Donovan MC. Common alleles contribute to schizophrenia in CNV carriers. *Mol Psychiatry*. 2016 Aug;21(8):1085-9. doi: 10.1038/mp.2015.143. Epub 2015 Sep 22. Erratum in: *Mol Psychiatry*. 2016 Aug;21(8):1153. PubMed PMID: 26390827; PubMed Central PMCID: PMC4960448.
- Tenesa A, Haley CS. The heritability of human disease: estimation, uses and abuses. *Nat Rev Genet*. 2013 Feb;14(2):139-49. doi: 10.1038/nrg3377. Review. PubMed PMID: 23329114.
- Thapar A, Martin J, Mick E, Arias Vásquez A, Langley K, Scherer SW, Schachar R, Crosbie J, Williams N, Franke B, Elia J, Glessner J, Hakonarson H, Owen MJ, Faraone SV, O'Donovan MC, Holmans P. Psychiatric gene discoveries shape evidence on ADHD's biology. *Mol Psychiatry*. 2016 Sep;21(9):1202-7. doi: 10.1038/mp.2015.163. Epub 2015 Nov 17. PubMed PMID: 26573769; PubMed Central PMCID: PMC4820035.
- Thorgerirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, Manolescu A, Thorleifsson G, Stefansson H, Ingason A, Stacey SN, Bergthorsson JT, Thorlacius S, Gudmundsson J, Jonsson T, Jakobsdottir M, Saemundsdottir J, Olafsdottir O, Gudmundsson LJ, Bjornsdottir G, Kristjansson K, Skuladottir H, Isaksson HJ, Gudbjartsson T, Jones GT, Mueller T, Gottsäter A, Flex A, Aben KKH, de Vegt F, Mulders PFA, Isla D, Vidal MJ, Asin L, Saez B, Murillo L, Blondal T, Kolbeinsson H, Stefansson JG, Hansdottir I, Runarsdottir V, Pola R, Lindblad B, van Rij AM, Dieplinger B, Haltmayer M, Mayordomo JI, Kiemenev

- LA, Mattheiasson SE, Oskarsson H, Tyrfinngsson T, Gudbjartsson DF, Gulcher JR, Jonsson S, Thorsteinsdottir U, Kong A, Stefansson K. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature*. 2008 Apr 3;452(7187):638-642. doi: 10.1038/nature06846. PubMed PMID: 18385739; PubMed Central PMCID: PMC4539558.
- Tishkoff SA, Verrelli BC. Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annu Rev Genomics Hum Genet*. 2003;4:293-340. Review. PubMed PMID: 14527305.
- Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet*. 2010 May;42(5):441-7. doi: 10.1038/ng.571. Epub 2010 Apr 25. PubMed PMID: 20418890; PubMed Central PMCID: PMC2914600.
- Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet*. 2018 Sep;19(9):581-590. doi: 10.1038/s41576-018-0018-x. Review. PubMed PMID: 29789686.
- Tropf FC, Lee SH, Verweij RM, Stulp G, van der Most PJ, de Vlaming R, Bakshi A, Briley DA, Rahal C, Hellpap R, Iliadou AN, Esko T, Metspalu A, Medland SE, Martin NG, Barban N, Snieder H, Robinson MR, Mills MC. Hidden heritability due to heterogeneity across seven populations. *Nat Hum Behav*. 2017 Oct;1(10):757-765. doi: 10.1038/s41562-017-0195-1. Epub 2017 Sep 11. PubMed PMID: 29051922; PubMed Central PMCID: PMC5642946.
- Tsuang M. Schizophrenia: genes and environment. *Biol Psychiatry*. 2000 Feb 1;47(3):210-20. Review. PubMed PMID: 10682218.
- Tsuang MT, Lyons MJ, Meyer JM, Doyle T, Eisen SA, Goldberg J, True W, Lin N, Toomey R, Eaves L. Co-occurrence of abuse of different drugs in men: the role of drug-specific and shared vulnerabilities. *Arch Gen Psychiatry*. 1998 Nov;55(11):967-72. PubMed PMID: 9819064.
- Tsuang MT, Woolson RF, Winokur G, Crowe RR. Stability of psychiatric diagnosis. Schizophrenia and affective disorders followed up over a 30- to 40-year period. *Arch Gen Psychiatry*. 1981 May;38(5):535-9. PubMed PMID: 7235854.
- Turley P, Walters RK, Maghziyan O, Okbay A, Lee JJ, Fontana MA, Nguyen-Viet TA, Wedow R, Zacher M, Furlotte NA, Magnusson P, Oskarsson S, Johannesson M, Visscher PM, Laibson D, Cesarini D, Neale BM, Benjamin DJ; 23andMe Research Team; Social Science Genetic Association Consortium. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet*. 2018 Feb;50(2):229-237. doi: 10.1038/s41588-017-0009-4. Epub 2018 Jan 1. PubMed PMID: 29292387; PubMed Central PMCID: PMC5805593.

- Uher R, Zwickler A. Etiology in psychiatry: embracing the reality of poly-gene-environmental causation of mental illness. *World Psychiatry*. 2017 Jun;16(2):121-129. doi: 10.1002/wps.20436. PubMed PMID: 28498595; PubMed Central PMCID: PMC5428165.
- Ursini G, Punzi G, Chen Q, Marengo S, Robinson JF, Porcelli A, Hamilton EG, Mitjans M, Maddalena G, Begemann M, Seidel J, Yanamori H, Jaffe AE, Berman KF, Egan MF, Straub RE, Colantuoni C, Blasi G, Hashimoto R, Rujescu D, Ehrenreich H, Bertolino A, Weinberger DR. Convergence of placenta biology and genetic risk for schizophrenia. *Nat Med*. 2018 Jun;24(6):792-801. doi: 10.1038/s41591-018-0021-y. Epub 2018 May 28. PubMed PMID: 29808008.
- Valsesia A, Macé A, Jacquemont S, Beckmann JS, Kutalik Z. The Growing Importance of CNVs: New Insights for Detection and Clinical Interpretation. *Front Genet*. 2013 May 30;4:92. doi: 10.3389/fgene.2013.00092. eCollection 2013. PubMed PMID: 23750167; PubMed Central PMCID: PMC3667386.
- van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. *Trends Genet*. 2014 Sep;30(9):418-26. doi: 10.1016/j.tig.2014.07.001. Epub 2014 Aug 6. Review. PubMed PMID: 25108476.
- van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C. The Third Revolution in Sequencing Technology. *Trends Genet*. 2018 Sep;34(9):666-681. doi: 10.1016/j.tig.2018.05.008. Epub 2018 Jun 22. Review. PubMed PMID: 29941292.
- van Heugten-van der Kloet D, van Heugten T. The classification of psychiatric disorders according to DSM-5 deserves an internationally standardized psychological test battery on symptom level. *Front Psychol*. 2015 Aug 4;6:1108. doi: 10.3389/fpsyg.2015.01108. eCollection 2015. PubMed PMID: 26300808; PubMed Central PMCID: PMC4523712.
- van Os J, Kapur S. Schizophrenia. *Lancet*. 2009 Aug 22;374(9690):635-45. doi: 10.1016/S0140-6736(09)60995-8. Review. PubMed PMID: 19700006.
- van Os J, Kenis G, Rutten BP. The environment and schizophrenia. *Nature*. 2010 Nov 11;468(7321):203-12. doi: 10.1038/nature09563. PubMed PMID: 21068828.
- van Os J, Linscott RJ, Myin-Germeys I, Delespaul P, Krabbendam L. A systematic review and meta-analysis of the psychosis continuum: evidence for a psychosis proneness-persistence-impairment model of psychotic disorder. *Psychol Med*. 2009 Feb;39(2):179-95. doi: 10.1017/S0033291708003814. Epub 2008 Jul 8. Review. PubMed PMID: 18606047.
- Vassos E, Agerbo E, Mors O, Pedersen CB. Urban-rural differences in incidence rates of psychiatric disorders in Denmark. *Br J Psychiatry*. 2016 May;208(5):435-40. doi: 10.1192/bjp.bp.114.161091. Epub 2015 Dec 17. PubMed PMID: 26678865.

- Vassos E, Di Forti M, Coleman J, Iyegbe C, Prata D, Euesden J, O'Reilly P, Curtis C, Koliakou A, Patel H, Newhouse S, Traylor M, Ajnakina O, Mondelli V, Marques TR, Gardner-Sood P, Aitchison KJ, Powell J, Atakan Z, Greenwood KE, Smith S, Ismail K, Pariante C, Gaughran F, Dazzan P, Markus HS, David AS, Lewis CM, Murray RM, Breen G. An Examination of Polygenic Score Risk Prediction in Individuals With First-Episode Psychosis. *Biol Psychiatry*. 2017 Mar 15;81(6):470-477. doi: 10.1016/j.biopsych.2016.06.028. Epub 2016 Aug 6. PubMed PMID: 27765268.
- Vaucher J, Keating BJ, Lasserre AM, Gan W, Lyall DM, Ward J, Smith DJ, Pell JP, Sattar N, Paré G, Holmes MV. Cannabis use and risk of schizophrenia: a Mendelian randomization study. *Mol Psychiatry*. 2018 May;23(5):1287-1292. doi: 10.1038/mp.2016.252. Epub 2017 Jan 24. PubMed PMID: 28115737; PubMed Central PMCID: PMC5984096.
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigó R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L,

- Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X. The sequence of the human genome. *Science*. 2001 Feb 16;291(5507):1304-51. Erratum in: *Science* 2001 Jun 5;292(5523):1838. PubMed PMID: 11181995.
- Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. *Lancet Psychiatry*. 2016 Feb;3(2):171-8. doi: 10.1016/S2215-0366(15)00505-2. Review. PubMed PMID: 26851330.
- Vilhjálmsón BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, Genovese G, Loh PR, Bhatia G, Do R, Hayeck T, Won HH; Schizophrenia Working Group of the Psychiatric Genomics Consortium, Discovery, Biology, and Risk of Inherited Variants in Breast Cancer (DRIVE) study, Kathiresan S, Pato M, Pato C, Tamimi R, Stahl E, Zaitlen N, Pasaniuc B, Belbin G, Kenny EE, Schierup MH, De Jager P, Patsopoulos NA, McCarroll S, Daly M, Purcell S, Chasman D, Neale B, Goddard M, Visscher PM, Kraft P, Patterson N, Price AL. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am J Hum Genet*. 2015 Oct 1;97(4):576-92. doi: 10.1016/j.ajhg.2015.09.001. PubMed PMID: 26430803; PubMed Central PMCID: PMC4596916.
- Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet*. 2012 Jan 13;90(1):7-24. doi: 10.1016/j.ajhg.2011.11.029. Review. PubMed PMID: 22243964; PubMed Central PMCID: PMC3257326.
- Visscher PM, Goddard ME, Derks EM, Wray NR. Evidence-based psychiatric genetics, AKA the false dichotomy between common and rare variant hypotheses. *Mol Psychiatry*. 2012 May;17(5):474-85. doi: 10.1038/mp.2011.65. Epub 2011 Jun 14. Review. PubMed PMID: 21670730.
- Visscher PM, Hemani G, Vinkhuyzen AA, Chen GB, Lee SH, Wray NR, Goddard ME, Yang J. Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLoS Genet*. 2014 Apr 10;10(4):e1004269. doi: 10.1371/journal.pgen.1004269. eCollection 2014 Apr. PubMed PMID: 24721987; PubMed Central PMCID: PMC3983037.
- Visscher PM, Hill WG, Wray NR. Heritability in the genomics era--concepts and misconceptions. *Nat Rev Genet*. 2008 Apr;9(4):255-66. doi: 10.1038/nrg2322. Epub 2008 Mar 4. Review. PubMed PMID: 18319743.

- Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet.* 2017 Jul 6;101(1):5-22. doi: 10.1016/j.ajhg.2017.06.005. Review. PubMed PMID: 28686856; PubMed Central PMCID: PMC5501872.
- Walker ER, McGee RE, Druss BG. Mortality in mental disorders and global disease burden implications: a systematic review and meta-analysis. *JAMA Psychiatry.* 2015 Apr;72(4):334-41. doi: 10.1001/jamapsychiatry.2014.2502. Review. Erratum in: *JAMA Psychiatry.* 2015 Jul;72(7):736. Erratum in: *JAMA Psychiatry.* 2015 Dec;72(12):1259. PubMed PMID: 25671328; PubMed Central PMCID: PMC4461039.
- Walters JT, Owen MJ. Endophenotypes in psychiatric genetics. *Mol Psychiatry.* 2007 Oct;12(10):886-90. Review. PubMed PMID: 17895920.
- Walters RK, Polimanti R, Johnson EC, McClintick JN, Adams MJ, Adkins AE, Aliev F, Bacanu SA, Batzler A, Bertelsen S, Biernacka JM, Bigdeli TB, Chen LS, Clarke TK, Chou YL, Degenhardt F, Docherty AR, Edwards AC, Fontanillas P, Foo JC, Fox L, Frank J, Giegling I, Gordon S, Hack LM, Hartmann AM, Hartz SM, Heilmann-Heimbach S, Herms S, Hodgkinson C, Hoffmann P, Jan Hottenga J, Kennedy MA, Alanne-Kinnunen M, Konte B, Lahti J, Lahti-Pulkkinen M, Lai D, Ligthart L, Loukola A, Maher BS, Mbarek H, McIntosh AM, McQueen MB, Meyers JL, Milaneschi Y, Palviainen T, Pearson JF, Peterson RE, Ripatti S, Ryu E, Saccone NL, Salvatore JE, Sanchez-Roige S, Schwandt M, Sherva R, Streit F, Strohmaier J, Thomas N, Wang JC, Webb BT, Wedow R, Wetherill L, Wills AG; 23andMe Research Team, Boardman JD, Chen D, Choi DS, Copeland WE, Culverhouse RC, Dahmen N, Degenhardt L, Domingue BW, Elson SL, Frye MA, Gäbel W, Hayward C, Ising M, Keyes M, Kiefer F, Kramer J, Kuperman S, Lucae S, Lynskey MT, Maier W, Mann K, Männistö S, Müller-Myhsok B, Murray AD, Nurnberger JI, Palotie A, Preuss U, Rääkkönen K, Reynolds MD, Ridinger M, Scherbaum N, Schuckit MA, Soyka M, Treutlein J, Witt S, Wodarz N, Zill P, Adkins DE, Boden JM, Boomsma DI, Bierut LJ, Brown SA, Bucholz KK, Cichon S, Costello EJ, de Wit H, Diazgranados N, Dick DM, Eriksson JG, Farrer LA, Foroud TM, Gillespie NA, Goate AM, Goldman D, Grucza RA, Hancock DB, Harris KM, Heath AC, Hesselbrock V, Hewitt JK, Hopfer CJ, Horwood J, Iacono W, Johnson EO, Kaprio JA, Karpyak VM, Kendler KS, Kranzler HR, Krauter K, Lichtenstein P, Lind PA, McGue M, MacKillop J, Madden PAF, Maes HH, Magnusson P, Martin NG, Medland SE, Montgomery GW, Nelson EC, Nöthen MM, Palmer AA, Pedersen NL, Penninx BWJH, Porjesz B, Rice JP, Rietschel M, Riley BP, Rose R, Rujescu D, Shen PH, Silberg J, Stallings MC, Tarter RE, Vanyukov MM, Vrieze S, Wall TL, Whitfield JB, Zhao H, Neale BM, Gelernter J, Edenberg HJ, Agrawal A. Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat Neurosci.* 2018 Dec;21(12):1656-1669. doi: 10.1038/s41593-018-0275-1. Epub 2018 Nov 26. PubMed PMID: 30482948; PubMed Central PMCID: PMC6430207.

- Wang WY, Barratt BJ, Clayton DG, Todd JA. Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet.* 2005 Feb;6(2):109-18. Review. PubMed PMID: 15716907.
- Ward ET, Kostick KM, Lázaro-Muñoz G. Integrating Genomics into Psychiatric Practice: Ethical and Legal Challenges for Clinicians. *Harv Rev Psychiatry.* 2019 Jan/Feb;27(1):53-64. doi: 10.1097/HRP.000000000000203. PubMed PMID: 30614887; PubMed Central PMCID: PMC6326091.
- Ward J, Graham N, Strawbridge RJ, Ferguson A, Jenkins G, Chen W, Hodgson K, Frye M, Weinshilboum R, Uher R, Lewis CM, Biernacka J, Smith DJ. Polygenic risk scores for major depressive disorder and neuroticism as predictors of antidepressant response: Meta-analysis of three treatment cohorts. *PLoS One.* 2018 Sep 21;13(9):e0203896. doi: 10.1371/journal.pone.0203896. eCollection 2018. PubMed PMID: 30240446; PubMed Central PMCID: PMC6150505.
- Watson D. Investigating the construct validity of the dissociative taxon: stability analyses of normal and pathological dissociation. *J Abnorm Psychol.* 2003 May;112(2):298-305. PubMed PMID: 12784840.
- Weber MM. Ernst Rüdin. Eine kritische Biographie. 1993
- Weinberg A, Dieterich R, Riesel A. Error-related brain activity in the age of RDoC: A review of the literature. *Int J Psychophysiol.* 2015 Nov;98(2 Pt 2):276-299. doi: 10.1016/j.ijpsycho.2015.02.029. Epub 2015 Mar 4. Review. PubMed PMID: 25746725.
- Weinberger DR. Implications of normal brain development for the pathogenesis of schizophrenia. *Arch Gen Psychiatry.* 1987 Jul;44(7):660-9. PubMed PMID: 3606332.
- Weiner DJ, Wigdor EM, Ripke S, Walters RK, Kosmicki JA, Grove J, Samocha KE, Goldstein JI, Okbay A, Bybjerg-Grauholm J, Werge T, Hougaard DM, Taylor J; iPSYCH-Broad Autism Group; Psychiatric Genomics Consortium Autism Group, Skuse D, Devlin B, Anney R, Sanders SJ, Bishop S, Mortensen PB, Børglum AD, Smith GD, Daly MJ, Robinson EB. Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nat Genet.* 2017 Jul;49(7):978-985. doi: 10.1038/ng.3863. Epub 2017 May 15. PubMed PMID: 28504703; PubMed Central PMCID: PMC5552240.
- Weischenfeldt J, Symmons O, Spitz F, Korbel JO. Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat Rev Genet.* 2013 Feb;14(2):125-38. doi: 10.1038/nrg3373. Review. PubMed PMID: 23329113.
- Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* 2007 Jun

7;447(7145):661-78. PubMed PMID: 17554300; PubMed Central PMCID: PMC2719288.

Werling DM, Brand H, An JY, Stone MR, Zhu L, Glessner JT, Collins RL, Dong S, Layer RM, Markenscoff-Papadimitriou E, Farrell A, Schwartz GB, Wang HZ, Currall BB, Zhao X, Dea J, Duhn C, Erdman CA, Gilson MC, Yadav R, Handsaker RE, Kashin S, Klei L, Mandell JD, Nowakowski TJ, Liu Y, Pochareddy S, Smith L, Walker MF, Waterman MJ, He X, Kriegstein AR, Rubenstein JL, Sestan N, McCarroll SA, Neale BM, Coon H, Willsey AJ, Buxbaum JD, Daly MJ, State MW, Quinlan AR, Marth GT, Roeder K, Devlin B, Talkowski ME, Sanders SJ. An analytical framework for whole-genome sequence association studies and its implications for autism spectrum disorder. *Nat Genet.* 2018 Apr 26;50(5):727-736. doi: 10.1038/s41588-018-0107-y. PubMed PMID: 29700473; PubMed Central PMCID: PMC5961723.

Westfall, P. H., & Young, S. S. Resampling-based multiple testing: Examples and methods for p-value adjustment(Vol. 279). John Wiley & Sons. 1993.

Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, Christiansen MW, Fairfax BP, Schramm K, Powell JE, Zhernakova A, Zhernakova DV, Veldink JH, Van den Berg LH, Karjalainen J, Withoff S, Uitterlinden AG, Hofman A, Rivadeneira F, Hoen PAC', Reinmaa E, Fischer K, Nelis M, Milani L, Melzer D, Ferrucci L, Singleton AB, Hernandez DG, Nalls MA, Homuth G, Nauck M, Radke D, Völker U, Perola M, Salomaa V, Brody J, Suchy-Dacey A, Gharib SA, Enquobahrie DA, Lumley T, Montgomery GW, Makino S, Prokisch H, Herder C, Roden M, Grallert H, Meitinger T, Strauch K, Li Y, Jansen RC, Visscher PM, Knight JC, Psaty BM, Ripatti S, Teumer A, Frayling TM, Metspalu A, van Meurs JBJ, Franke L. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet.* 2013 Oct;45(10):1238-1243. doi: 10.1038/ng.2756. Epub 2013 Sep 8. PubMed PMID: 24013639; PubMed Central PMCID: PMC3991562.

Wetherill L, Agrawal A, Kapoor M, Bertelsen S, Bierut LJ, Brooks A, Dick D, Hesselbrock M, Hesselbrock V, Koller DL, Le N, Nurnberger JI Jr, Salvatore JE, Schuckit M, Tischfield JA, Wang JC, Xuei X, Edenberg HJ, Porjesz B, Bucholz K, Goate AM, Foroud T. Association of substance dependence phenotypes in the COGA sample. *Addict Biol.* 2015 May;20(3):617-27. doi: 10.1111/adb.12153. Epub 2014 May 16. PubMed PMID: 24832863; PubMed Central PMCID: PMC4233207.

Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, Charlson FJ, Norman RE, Flaxman AD, Johns N, Burstein R, Murray CJ, Vos T. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet.* 2013 Nov 9;382(9904):1575-

86. doi: 10.1016/S0140-6736(13)61611-6. Epub 2013 Aug 29. Review. PubMed PMID: 23993280.
- Wild CP. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev.* 2005 Aug;14(8):1847-50. Review. PubMed PMID: 16103423.
- Williams HJ, Norton N, Dwyer S, Moskvina V, Nikolov I, Carroll L, Georgieva L, Williams NM, Morris DW, Quinn EM, Giegling I, Ikeda M, Wood J, Lencz T, Hultman C, Lichtenstein P, Thiselton D, Maher BS; Molecular Genetics of Schizophrenia Collaboration (MGS) International Schizophrenia Consortium (ISC), SGENE-plus, GROUP, Malhotra AK, Riley B, Kendler KS, Gill M, Sullivan P, Sklar P, Purcell S, Nimgaonkar VL, Kirov G, Holmans P, Corvin A, Rujescu D, Craddock N, Owen MJ, O'Donovan MC. Fine mapping of ZNF804A and genome-wide significant evidence for its involvement in schizophrenia and bipolar disorder. *Mol Psychiatry.* 2011 Apr;16(4):429-41. doi: 10.1038/mp.2010.36. Epub 2010 Apr 6. PubMed PMID: 20368704; PubMed Central PMCID: PMC3918934.
- Willsey AJ, Fernandez TV, Yu D, King RA, Dietrich A, Xing J, Sanders SJ, Mandell JD, Huang AY, Richer P, Smith L, Dong S, Samocha KE; Tourette International Collaborative Genetics (TIC Genetics); Tourette Syndrome Association International Consortium for Genetics (TSAICG), Neale BM, Coppola G, Mathews CA, Tischfield JA, Scharf JM, State MW, Heiman GA. De Novo Coding Variants Are Strongly Associated with Tourette Disorder. *Neuron.* 2017 May 3;94(3):486-499.e9. doi: 10.1016/j.neuron.2017.04.024. PubMed PMID: 28472652; PubMed Central PMCID: PMC5769876.
- Wilson M. DSM-III and the transformation of American psychiatry: a history. *Am J Psychiatry.* 1993 Mar;150(3):399-410. Review. PubMed PMID: 8434655.
- Wimberley T, Gasse C, Meier SM, Agerbo E, MacCabe JH, Horsdal HT. Polygenic Risk Score for Schizophrenia and Treatment-Resistant Schizophrenia. *Schizophr Bull.* 2017 Sep 1;43(5):1064-1069. doi: 10.1093/schbul/sbx007. PubMed PMID: 28184875; PubMed Central PMCID: PMC5581885.
- Winham SJ, Biernacka JM. Gene-environment interactions in genome-wide association studies: current approaches and new directions. *J Child Psychol Psychiatry.* 2013 Oct;54(10):1120-34. doi: 10.1111/jcpp.12114. Epub 2013 Jun 28. Review. PubMed PMID: 23808649; PubMed Central PMCID: PMC3829379.
- Wittchen HU, Beesdo-Baum K. "Throwing out the baby with the bathwater"? Conceptual and methodological limitations of the HiTOP approach. *World Psychiatry.* 2018 Oct;17(3):298-299. doi: 10.1002/wps.20561. PubMed PMID: 30192104; PubMed Central PMCID: PMC6127812.

- Witte JS, Visscher PM, Wray NR. The contribution of genetic variants to disease depends on the ruler. *Nat Rev Genet.* 2014 Nov;15(11):765-76. doi: 10.1038/nrg3786. Epub 2014 Sep 16. Review. PubMed PMID: 25223781; PubMed Central PMCID: PMC4412738.
- Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res.* 2007 Oct;17(10):1520-8. Epub 2007 Sep 4. PubMed PMID: 17785532; PubMed Central PMCID: PMC1987352.
- Wray NR, Gottesman II. Using summary data from the danish national registers to estimate heritabilities for schizophrenia, bipolar disorder, and major depressive disorder. *Front Genet.* 2012 Jul 2;3:118. doi: 10.3389/fgene.2012.00118. eCollection 2012. PubMed PMID: 22783273; PubMed Central PMCID: PMC3387670.
- Wray NR, Lee SH, Mehta D, Vinkhuyzen AA, Dudbridge F, Middeldorp CM. Research review: Polygenic methods and their application to psychiatric traits. *J Child Psychol Psychiatry.* 2014 Oct;55(10):1068-87. doi: 10.1111/jcpp.12295. Epub 2014 Aug 1. Review. PubMed PMID: 25132410.
- Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, Adams MJ, Agerbo E, Air TM, Andlauer TMF, Bacanu SA, Bækvad-Hansen M, Beekman AFT, Bigdeli TB, Binder EB, Blackwood DRH, Bryois J, Buttenschøn HN, Bybjerg-Grauholm J, Cai N, Castelao E, Christensen JH, Clarke TK, Coleman JIR, Colodro-Conde L, Couvy-Duchesne B, Craddock N, Crawford GE, Crowley CA, Dashti HS, Davies G, Deary IJ, Degenhardt F, Derks EM, Direk N, Dolan CV, Dunn EC, Eley TC, Eriksson N, Escott-Price V, Kiadeh FHF, Finucane HK, Forstner AJ, Frank J, Gaspar HA, Gill M, Giusti-Rodríguez P, Goes FS, Gordon SD, Grove J, Hall LS, Hannon E, Hansen CS, Hansen TF, Herms S, Hickie IB, Hoffmann P, Homuth G, Horn C, Hottenga JJ, Hougaard DM, Hu M, Hyde CL, Ising M, Jansen R, Jin F, Jorgenson E, Knowles JA, Kohane IS, Kraft J, Kretschmar WW, Krogh J, Kutalik Z, Lane JM, Li Y, Li Y, Lind PA, Liu X, Lu L, MacIntyre DJ, MacKinnon DF, Maier RM, Maier W, Marchini J, Mbarek H, McGrath P, McGuffin P, Medland SE, Mehta D, Middeldorp CM, Mihailov E, Milaneschi Y, Milani L, Mill J, Mondimore FM, Montgomery GW, Mostafavi S, Mullins N, Nauck M, Ng B, Nivard MG, Nyholt DR, O'Reilly PF, Oskarsson H, Owen MJ, Painter JN, Pedersen CB, Pedersen MG, Peterson RE, Pettersson E, Peyrot WJ, Pistis G, Posthuma D, Purcell SM, Quiroz JA, Qvist P, Rice JP, Riley BP, Rivera M, Saeed Mirza S, Saxena R, Schoebers R, Schulte EC, Shen L, Shi J, Shyn SI, Sigurdsson E, Sinnamon GBC, Smit JH, Smith DJ, Stefansson H, Steinberg S, Stockmeier CA, Streit F, Strohmaier J, Tansey KE, Teismann H, Teumer A, Thompson W, Thomson PA, Thorgeirsson TE, Tian C, Traylor M, Treutlein J, Trubetskoy V, Uitterlinden AG, Uebachs D, Van der Auwera S, van Hemert AM, Viktorin A, Visscher PM, Wang Y, Webb BT, Weinsheimer SM,

- Wellmann J, Willemsen G, Witt SH, Wu Y, Xi HS, Yang J, Zhang F; eQTLGen; 23andMe, Arolt V, Baune BT, Berger K, Boomsma DI, Cichon S, Dannlowski U, de Geus ECJ, DePaulo JR, Domenici E, Domschke K, Esko T, Grabe HJ, Hamilton SP, Hayward C, Heath AC, Hinds DA, Kendler KS, Kloiber S, Lewis G, Li QS, Lucae S, Madden PFA, Magnusson PK, Martin NG, McIntosh AM, Metspalu A, Mors O, Mortensen PB, Müller-Myhsok B, Nordentoft M, Nöthen MM, O'Donovan MC, Paciga SA, Pedersen NL, Penninx BWJH, Perlis RH, Porteous DJ, Potash JB, Preisig M, Rietschel M, Schaefer C, Schulze TG, Smoller JW, Stefansson K, Tiemeier H, Uher R, Völzke H, Weissman MM, Werge T, Winslow AR, Lewis CM, Levinson DF, Breen G, Børglum AD, Sullivan PF; Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet.* 2018 May;50(5):668-681. doi: 10.1038/s41588-018-0090-3. Epub 2018 Apr 26. PubMed PMID: 29700475; PubMed Central PMCID: PMC5934326.
- Wray NR, Visscher P (2008). Estimating trait heritability. *Nature education*, 1(1), 29.
- Wray NR, Wijmenga C, Sullivan PF, Yang J, Visscher PM. Common Disease Is More Complex Than Implied by the Core Gene Omnigenic Model. *Cell.* 2018 Jun 14;173(7):1573-1580. doi: 10.1016/j.cell.2018.05.051. Review. PubMed PMID: 29906445.
- Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM. Pitfalls of predicting complex traits from SNPs. *Nat Rev Genet.* 2013 Jul;14(7):507-15. doi: 10.1038/nrg3457. Review. PubMed PMID: 23774735; PubMed Central PMCID: PMC4096801.
- Wray, NR, Maier, R. Genetic basis of complex genetic disease: the contribution of disease heterogeneity to missing heritability. *Current Epidemiology Reports.* 2014, 1(4), 220-227.
- Wright AG, Krueger RF, Hobbs MJ, Markon KE, Eaton NR, Slade T. The structure of psychopathology: toward an expanded quantitative empirical model. *J Abnorm Psychol.* 2013 Feb;122(1):281-94. doi: 10.1037/a0030133. Epub 2012 Oct 15. PubMed PMID: 23067258; PubMed Central PMCID: PMC3570590.
- Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M. Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet.* 2008 Jul;40(7):880-5. doi: 10.1038/ng.162. Epub 2008 May 30. PubMed PMID: 18511947.
- Yamada K, Iwayama Y, Hattori E, Iwamoto K, Toyota T, Ohnishi T, Ohba H, Maekawa M, Kato T, Yoshikawa T. Genome-wide association study of schizophrenia in Japanese population. *PLoS One.* 2011;6(6):e20468. doi:

- 10.1371/journal.pone.0020468. Epub 2011 Jun 6. PubMed PMID: 21674006; PubMed Central PMCID: PMC3108953.
- Yang J, Benjamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet.* 2010 Jul;42(7):565-9. doi: 10.1038/ng.608. Epub 2010 Jun 20. PubMed PMID: 20562875; PubMed Central PMCID: PMC3232052.
- Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011 Jan 7;88(1):76-82. doi: 10.1016/j.ajhg.2010.11.011. Epub 2010 Dec 17. PubMed PMID: 21167468; PubMed Central PMCID: PMC3014363.
- Yang J, Zeng J, Goddard ME, Wray NR, Visscher PM. Concepts, estimation and interpretation of SNP-based heritability. *Nat Genet.* 2017 Aug 30;49(9):1304-1310. doi: 10.1038/ng.3941. PubMed PMID: 28854176.
- Yang MS, Gill M. A review of gene linkage, association and expression studies in autism and an assessment of convergent evidence. *Int J Dev Neurosci.* 2007 Apr;25(2):69-85. Epub 2006 Dec 20. Review. PubMed PMID: 17236739.
- Yee CM, Javitt DC, Miller GA. Replacing DSM Categorical Analyses With Dimensional Analyses in Psychiatry Research: The Research Domain Criteria Initiative. *JAMA Psychiatry.* 2015 Dec;72(12):1159-60. doi: 10.1001/jamapsychiatry.2015.1900. PubMed PMID: 26559005.
- Yilmaz Z, Halvorsen M, Bryois J, Yu D, Thornton LM, Zerwas S, Micali N, Moessner R, Burton CL, Zai G, Erdman L, Kas MJ, Arnold PD, Davis LK, Knowles JA, Breen G, Scharf JM, Nestadt G, Mathews CA, Bulik CM, Mattheisen M, Crowley JJ; Eating Disorders Working Group of the Psychiatric Genomics Consortium, Tourette Syndrome/Obsessive-Compulsive Disorder Working Group of the Psychiatric Genomics Consortium. Examination of the shared genetic basis of anorexia nervosa and obsessive-compulsive disorder. *Mol Psychiatry.* 2018 Aug 7. doi: 10.1038/s41380-018-0115-4. [Epub ahead of print] PubMed PMID: 30087453; PubMed Central PMCID: PMC6367065.
- Yin X, Wineinger NE, Wang K, Yue W, Norgren N, Wang L, Yao W, Jiang X, Wu B, Cui Y, Shen C, Cheng H, Zhou F, Chen G, Zuo X, Zheng X, Fan X, Wang H, Wang L, Lee J, Lam M, Tai ES, Zhang Z, Huang Q, Sun L, Xu J, Yang S, Wilhelmsen KC, Liu J, Schork NJ, Zhang X. Common susceptibility variants are shared between schizophrenia and psoriasis in the Han Chinese population. *J Psychiatry Neurosci.* 2016 Oct;41(6):413-421. PubMed PMID: 27091718; PubMed Central PMCID: PMC5082512.

- Yu D, Mathews CA, Scharf JM, Neale BM, Davis LK, Gamazon ER, Derks EM, Evans P, Edlund CK, Crane J, Fagerness JA, Osiecki L, Gallagher P, Gerber G, Haddad S, Illmann C, McGrath LM, Mayerfeld C, Arepalli S, Barlassina C, Barr CL, Bellodi L, Benarroch F, Berrió GB, Bienvenu OJ, Black DW, Bloch MH, Brentani H, Bruun RD, Budman CL, Camarena B, Campbell DD, Cappi C, Silgado JC, Cavallini MC, Chavira DA, Chouinard S, Cook EH, Cookson MR, Coric V, Cullen B, Cusi D, Delorme R, Denys D, Dion Y, Eapen V, Egberts K, Falkai P, Fernandez T, Fournier E, Garrido H, Geller D, Gilbert DL, Girard SL, Grabe HJ, Grados MA, Greenberg BD, Gross-Tsur V, Grünblatt E, Hardy J, Heiman GA, Hemmings SM, Herrera LD, Hezel DM, Hoekstra PJ, Jankovic J, Kennedy JL, King RA, Konkashbaev AI, Kremeyer B, Kurlan R, Lanzagorta N, Leboyer M, Leckman JF, Lennertz L, Liu C, Lochner C, Lowe TL, Lupoli S, Macciardi F, Maier W, Manunta P, Marconi M, McCracken JT, Mesa Restrepo SC, Moessner R, Moorjani P, Morgan J, Muller H, Murphy DL, Nardden AL, Nurmi E, Ochoa WC, Ophoff RA, Pakstis AJ, Pato MT, Pato CN, Piacentini J, Pittenger C, Pollak Y, Rauch SL, Renner T, Reus VI, Richter MA, Riddle MA, Robertson MM, Romero R, Rosário MC, Rosenberg D, Ruhrmann S, Sabatti C, Salvi E, Sampaio AS, Samuels J, Sandor P, Service SK, Sheppard B, Singer HS, Smit JH, Stein DJ, Strengman E, Tischfield JA, Turiel M, Valencia Duarte AV, Vallada H, Veenstra-VanderWeele J, Walitza S, Wang Y, Weale M, Weiss R, Wendland JR, Westenberg HG, Shugart YY, Hounie AG, Miguel EC, Nicolini H, Wagner M, Ruiz-Linares A, Cath DC, McMahon W, Posthuma D, Oostra BA, Nestadt G, Rouleau GA, Purcell S, Jenike MA, Heutink P, Hanna GL, Conti DV, Arnold PD, Freimer NB, Stewart SE, Knowles JA, Cox NJ, Pauls DL. Cross-disorder genome-wide analyses suggest a complex genetic relationship between Tourette's syndrome and OCD. *Am J Psychiatry*. 2015 Jan;172(1):82-93. doi: 10.1176/appi.ajp.2014.13101306. Epub 2014 Oct 31. PubMed PMID: 25158072; PubMed Central PMCID: PMC4282594.
- Yue WH, Wang HF, Sun LD, Tang FL, Liu ZH, Zhang HX, Li WQ, Zhang YL, Zhang Y, Ma CC, Du B, Wang LF, Ren YQ, Yang YF, Hu XF, Wang Y, Deng W, Tan LW, Tan YL, Chen Q, Xu GM, Yang GG, Zuo XB, Yan H, Ruan YY, Lu TL, Han X, Ma XH, Wang Y, Cai LW, Jin C, Zhang HY, Yan J, Mi WF, Yin XY, Ma WB, Liu Q, Kang L, Sun W, Pan CY, Shuang M, Yang FD, Wang CY, Yang JL, Li KQ, Ma X, Li LJ, Yu X, Li QZ, Huang X, Lv LX, Li T, Zhao GP, Huang W, Zhang XJ, Zhang D. Genome-wide association study identifies a susceptibility locus for schizophrenia in Han Chinese at 11p11.2. *Nat Genet*. 2011 Oct 30;43(12):1228-31. doi: 10.1038/ng.979. PubMed PMID: 22037552.
- Zabaneh D, Krapohl E, Gaspar HA, Curtis C, Lee SH, Patel H, Newhouse S, Wu HM, Simpson MA, Putallaz M, Lubinski D, Plomin R, Breen G. A genome-wide association study for extremely high intelligence. *Mol Psychiatry*. 2018 May;23(5):1226-1232. doi: 10.1038/mp.2017.121. Epub 2017 Jul 4. PubMed PMID: 29731509; PubMed Central PMCID: PMC5987166.

- Zachar P, Kendler KS. The Philosophy of Nosology. *Annu Rev Clin Psychol.* 2017 May 8;13:49-71. doi: 10.1146/annurev-clinpsy-032816-045020. Review. PubMed PMID: 28482691.
- Zhang JP, Robinson D, Yu J, Gallego J, Fleischhacker WW, Kahn RS, Crespo-Facorro B, Vazquez-Bourgon J, Kane JM, Malhotra AK, Lencz T. Schizophrenia Polygenic Risk Score as a Predictor of Antipsychotic Efficacy in First-Episode Psychosis. *Am J Psychiatry.* 2019 Jan 1;176(1):21-28. doi: 10.1176/appi.ajp.2018.17121363. Epub 2018 Nov 5. Erratum in: *Am J Psychiatry.* 2019 Mar 1;176(3):252. PubMed PMID: 30392411; PubMed Central PMCID: PMC6461047.
- Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, Hemani G, Tansey K, Laurin C; Early Genetics and Lifecourse Epidemiology (EAGLE) Eczema Consortium, Pourcain BS, Warrington NM, Finucane HK, Price AL, Bulik-Sullivan BK, Anttila V, Paternoster L, Gaunt TR, Evans DM, Neale BM. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics.* 2017 Jan 15;33(2):272-279. doi: 10.1093/bioinformatics/btw613. Epub 2016 Sep 22. PubMed PMID: 27663502; PubMed Central PMCID: PMC5542030.
- Zheutlin AB, Ross DA. Polygenic Risk Scores: What Are They Good For? *Biol Psychiatry.* 2018 Jun 1;83(11):e51-e53. doi: 10.1016/j.biopsych.2018.04.007. PubMed PMID: 29759133.
- Zhu Z, Bakshi A, Vinkhuyzen AA, Hemani G, Lee SH, Nolte IM, van Vliet-Ostaptchouk JV, Snieder H; LifeLines Cohort Study, Esko T, Milani L, Mägi R, Metspalu A, Hill WG, Weir BS, Goddard ME, Visscher PM, Yang J. Dominance genetic variation contributes little to the missing heritability for human complex traits. *Am J Hum Genet.* 2015 Mar 5;96(3):377-85. doi: 10.1016/j.ajhg.2015.01.001. Epub 2015 Feb 12. PubMed PMID: 25683123; PubMed Central PMCID: PMC4375616.
- Zhu Z, Bakshi A, Vinkhuyzen AA, Hemani G, Lee SH, Nolte IM, van Vliet-Ostaptchouk JV, Snieder H; LifeLines Cohort Study, Esko T, Milani L, Mägi R, Metspalu A, Hill WG, Weir BS, Goddard ME, Visscher PM, Yang J. Dominance genetic variation contributes little to the missing heritability for human complex traits. *Am J Hum Genet.* 2015 Mar 5;96(3):377-85. doi: 10.1016/j.ajhg.2015.01.001. Epub 2015 Feb 12. PubMed PMID: 25683123; PubMed Central PMCID: PMC4375616.
- Zink M, Schirmbeck F, Rausch F, Eifler S, Elkin H, Solojenkina X, Englisch S, Wagner M, Maier W, Lautenschlager M, Heinz A, Gudlowski Y, Janssen B, Gaebel W, Michel TM, Schneider F, Lambert M, Naber D, Juckel G, Krueger-Oezguerdal S, Wobrock T, Hasan A, Riedel M, Müller H, Klosterkötter J,

- Bechdolf A. Obsessive-compulsive symptoms in at-risk mental states for psychosis: associations with clinical impairment and cognitive function. *Acta Psychiatr Scand.* 2014 Sep;130(3):214-26. doi: 10.1111/acps.12258. Epub 2014 Feb 26. PubMed PMID: 24571191.
- Zondervan KT, Cardon LR. Designing candidate gene and genome-wide case-control association studies. *Nat Protoc.* 2007;2(10):2492-501. PubMed PMID: 17947991; PubMed Central PMCID: PMC4180089.





ANEXO





8. ANEXO

Las Tablas S1 y S2 de este anexo presentan las variantes incluidas en los modelos de riesgo poligénico generados en el Estudio 4 de esta Tesis. La Tabla S1 muestra los 198 SNPs comunes al listado original de Arloth y colaboradores (Arloth *et al.* 2015), la muestra descubrimiento MDD2 y la muestra diana SUD_1524. La Tabla S2 presenta los 58 SNPs incluidos en el modelo tras la búsqueda de variantes en LD del listado original (mediante el uso de SNiPA), comunes a MDD2 y SUD_1524.

TABLA S1. 198 SNPs COMUNES AL LISTADO ORIGINAL, MDD2 Y SUD_1524. RSID_PRS (código identificador dbSNP), CHR (cromosoma), BP (par de bases), A1 (alelo de riesgo), A2 (alelo mayor), OR (odds ratio, tamaño del efecto en MDD2), P (valor de p en GWAS MDD2) y originales (código identificador Arloth *et al.* 2015).

RSID_PRS	CHR	BP	A1	A2	OR	P	Originales
rs809972	1	9234363	T	G	1,02194	0,03725	rs809972
rs3766746	1	12031866	T	C	1,0144	0,07909	rs3766746
rs2295281	1	12059412	T	C	1,00833	0,2996	rs2295281
rs77158575	1	20944889	A	G	1,002	0,9129	rs77158575
rs2473263	1	22543364	A	G	0,9983	0,8269	rs2473263
rs7349097	1	44034135	T	C	0,99273	0,3658	rs7349097
rs425181	1	61235684	T	C	0,9982	0,83	rs425181
rs1981294	1	111526739	T	C	1,00612	0,5562	rs1981294
rs7544118	1	112032205	T	C	0,9974	0,797	rs7544118
rs72694971	1	150173801	T	G	0,9993	0,951	rs72694971
rs698915	1	150388318	A	G	1,001	0,9096	rs698915
rs7556661	1	150892009	T	C	1,00995	0,2292	rs7556661
rs11264449	1	156135241	C	G	1,0016	0,8501	rs11264449
rs67917627	1	156962719	T	C	1,00481	0,6397	rs67917627
rs10489832	1	158454356	A	G	1,00461	0,6824	rs10489832
rs717450	1	168552192	A	G	0,99154	0,2908	rs717450

XAQUÍN GURRIARÁN BAS

RSID_PRS	CHR	BP	A1	A2	OR	P	Originales
rs35288741	1	204928155	A	G	1,01501	0,07688	rs35288741
rs4845143	1	207003319	T	G	0,98659	0,1655	rs4845143
rs34764163	1	223034487	A	G	1,0027	0,7673	rs34764163
rs4309551	2	10272673	T	C	0,99481	0,5187	rs4309551
rs843631	2	54547225	A	G	0,99243	0,3404	rs843631
rs2422008	2	63780605	A	C	0,99283	0,3624	rs2422008
rs17849707	2	65298839	A	G	1,01015	0,2907	rs17849707
rs11686934	2	70172116	A	G	0,99461	0,4996	rs11686934
rs3771863	2	75419714	T	C	1,00844	0,4845	rs3771863
rs17031905	2	99070902	A	G	1,01015	0,2618	rs17031905
rs2305160	2	101591304	A	G	0,9993	0,9283	rs2305160
rs6543137	2	103065908	T	G	0,9993	0,943	rs6543137
rs17178720	2	128939799	T	C	1,0023	0,8178	rs17178720
rs736020	2	169964332	T	C	1,00723	0,3665	rs736020
rs16858988	2	171702960	A	G	0,97902	0,02644	rs16858988
rs6433294	2	172369069	T	C	1,00592	0,4781	rs6433294
rs6749185	2	172380608	C	G	0,98827	0,1358	rs6749185
rs6709463	2	203545198	T	C	0,9997	0,9671	rs6709463
rs2938387	3	12439387	T	C	0,99243	0,3532	rs2938387
rs2730355	3	16228958	A	G	1,00975	0,3709	rs2730355
rs12497322	3	39249753	A	G	1,00965	0,2594	rs12497322
rs2371129	3	40248393	T	C	1,02051	0,01412	rs2371129
rs17280306	3	41168084	T	C	0,99183	0,3044	rs17280306
rs113274464	3	43126872	C	G	0,98718	0,1882	rs113274464
rs7622109	3	44402063	A	G	0,97287	0,001004	rs7622109
rs1873625	3	49666964	A	C	0,97795	0,01018	rs1873625
rs17304079	3	50085153	A	G	0,99104	0,3166	rs17304079
rs9873175	3	52059636	T	C	0,98069	0,1416	rs9873175
rs11707455	3	107519777	A	G	1,00773	0,4063	rs11707455
rs72952437	3	112084424	A	G	1,00612	0,5851	rs72952437
rs62262832	3	112841860	A	T	0,9979	0,8511	rs62262832
rs4688030	3	119440690	C	G	0,9976	0,776	rs4688030

RSID_PRS	CHR	BP	A1	A2	OR	P	Originales
rs55776343	3	125749982	A	G	0,97961	0,0658	rs55776343
rs2567342	3	197282497	T	C	0,98305	0,05011	rs2567342
rs10939637	4	9977577	A	G	1,02675	0,007423	rs10939637
rs6531673	4	38847917	T	C	0,9982	0,8165	rs6531673
rs10002500	4	47958679	T	C	1,0199	0,09704	rs10002500
rs7673908	4	56394819	A	G	0,99581	0,6098	rs7673908
rs9994839	4	88393780	T	G	1,01298	0,1217	rs9994839
rs2672027	5	566782	A	C	1,02255	0,02571	rs2672027
rs3853657	5	567246	T	C	0,99501	0,7039	rs3853657
rs325828	5	41032232	T	C	1,0016	0,8781	rs325828
rs66507130	5	50792240	A	G	1,00803	0,368	rs66507130
rs2049490	5	75229137	A	G	0,98946	0,1771	rs2049490
rs1529505	5	76114963	T	C	1,0014	0,8565	rs1529505
rs171803	5	102136499	T	C	1,00501	0,5497	rs171803
rs2420147	5	130561257	T	C	0,99412	0,5668	rs2420147
rs4976450	5	136828322	T	C	0,99452	0,4999	rs4976450
rs10039049	5	150517899	T	C	1,00351	0,6836	rs10039049
rs9329125	5	177401509	A	G	1,00552	0,6222	rs9329125
rs11249756	5	180455372	T	C	1,00391	0,6505	rs11249756
rs9503168	6	2382395	A	G	1,00471	0,5527	rs9503168
rs7755418	6	4018886	A	G	1,01664	0,2558	rs7755418
rs531815	6	10790751	T	G	0,98985	0,2847	rs531815
rs1228529	6	13161138	T	C	0,99541	0,6623	rs1228529
rs115978907	6	31140152	A	C	1,03355	3,07E-05	rs115978907
rs116240820	6	32323529	A	G	1,0002	0,9799	rs116240820
rs116070899	6	32433067	A	G	1,0004	0,9714	rs116070899
rs2282444	6	41762808	C	G	0,9985	0,8704	rs2282444
rs2237250	6	111985081	A	G	1,01298	0,2671	rs2237250
rs6904470	6	132927345	T	C	1	0,9979	rs6904470
rs34771359	7	29565995	A	G	1,00965	0,2462	rs34771359
rs12216600	7	31965969	T	C	0,99243	0,3359	rs12216600
rs639459	7	42951936	T	G	1,00371	0,7586	rs639459

XAQUÍN GURRIARÁN BAS

RSID_PRS	CHR	BP	A1	A2	OR	P	Originales
rs2240516	7	43687296	T	C	0,99213	0,3198	rs2240516
rs17239727	7	43844321	T	C	0,97687	0,01579	rs17239727
rs10234768	7	50182079	T	C	1,03241	0,002649	rs10234768
rs11238359	7	55383887	T	C	1,00391	0,6627	rs11238359
rs2460432	7	65554385	T	C	1,00391	0,6438	rs2460432
rs12705071	7	99682026	T	C	0,99124	0,3132	rs12705071
rs17340646	7	128722514	T	G	1,002	0,8342	rs17340646
rs10487531	7	142162715	T	C	1,0028	0,739	rs10487531
rs2072443	7	150490376	T	C	1,00471	0,5595	rs2072443
rs1532445	8	20131799	A	G	0,98718	0,1204	rs1532445
rs12548608	8	29003000	T	C	1,00642	0,4201	rs12548608
rs12335026	8	82454361	A	G	1,01056	0,3156	rs12335026
rs7027886	9	2798358	A	C	0,97804	0,008065	rs7027886
rs7870685	9	31573188	A	G	1,0142	0,2362	rs7870685
rs10746914	9	75804249	A	C	0,99511	0,5517	rs10746914
rs1408069	9	111219000	A	G	1,02398	0,02448	rs1408069
rs4075428	9	130940121	A	G	0,97258	0,001089	rs4075428
rs13285411	9	130997868	A	C	1,00662	0,6038	rs13285411
rs3761821	9	136085783	T	C	1,00844	0,4203	rs3761821
rs10781518	9	139300547	A	G	1,00884	0,2815	rs10781518
rs10906402	10	6430767	T	C	1,0145	0,2461	rs10906402
rs7089504	10	6516327	T	C	0,98876	0,1725	rs7089504
rs7099954	10	7755732	T	G	0,99571	0,6964	rs7099954
rs158391	10	43143078	T	G	0,99422	0,5536	rs158391
rs77280190	10	45080538	T	C	1,00501	0,6954	rs77280190
rs4253082	10	50717586	T	C	1,00401	0,7069	rs4253082
rs2812500	10	71420172	A	C	1,01826	0,03783	rs2812500
rs17108932	10	89866673	A	G	0,98946	0,2293	rs17108932
rs2234768	10	90749943	T	C	0,98442	0,1907	rs2234768
rs10881678	10	91570905	T	C	0,99461	0,5437	rs10881678
rs3758587	10	98946244	A	G	0,98196	0,04621	rs3758587
rs2296887	10	104005410	T	C	1,01207	0,2763	rs2296887

RSID_PRS	CHR	BP	A1	A2	OR	P	Originales
rs10885031	10	112498330	T	C	1,00955	0,2367	rs10885031
rs9418982	10	135320622	T	C	0,97677	0,06767	rs9418982
rs2568032	11	8861077	A	G	0,99641	0,6772	rs2568032
rs4075678	11	11318964	A	G	1,0016	0,8432	rs4075678
rs10835861	11	32219851	A	C	1,01268	0,1668	rs10835861
rs2956993	11	62406162	T	G	0,9858	0,07727	rs2956993
rs3825073	11	64900270	T	G	1,01106	0,3118	rs3825073
rs11227523	11	66305938	T	C	0,9987	0,8899	rs11227523
rs4147779	11	67804807	A	G	0,99054	0,3134	rs4147779
rs624420	11	68620223	T	C	1,01167	0,3258	rs624420
rs3015983	11	77164904	T	C	1,00702	0,4126	rs3015983
rs10790231	11	117967476	T	C	0,9758	0,1228	rs10790231
rs4442562	11	118845813	A	G	0,98758	0,2153	rs4442562
rs4242902	12	7871436	T	C	0,99561	0,5963	rs4242902
rs11055463	12	7879654	A	G	1,00572	0,518	rs11055463
rs10505733	12	7888555	A	C	1,00662	0,4547	rs10505733
rs10784359	12	40445750	T	C	1,01076	0,2322	rs10784359
rs1859441	12	48423233	T	C	0,99074	0,3921	rs1859441
rs6581076	12	56067226	C	G	1,00371	0,6839	rs6581076
rs7955208	12	113203123	T	C	0,9803	0,07213	rs7955208
rs4767092	12	113975009	A	G	0,99173	0,4077	rs4767092
rs12423255	12	123595064	A	G	1,02265	0,01098	rs12423255
rs9525211	13	114803433	C	G	1,00733	0,4996	rs9525211
rs12891572	14	21738262	T	C	1,00874	0,4625	rs12891572
rs2178779	14	22409571	A	G	1,0005	0,95	rs2178779
rs12432242	14	23283905	T	C	0,98738	0,1146	rs12432242
rs8015121	14	23375251	A	G	0,99064	0,3373	rs8015121
rs8007588	14	25287011	A	G	0,99362	0,4272	rs8007588
rs2749883	14	52515716	A	C	0,99372	0,426	rs2749883
rs12886153	14	56221073	A	G	0,99173	0,4329	rs12886153
rs2359795	14	75912479	T	C	0,99511	0,5367	rs2359795
rs59562633	15	26193577	T	C	1,002	0,822	rs59562633

XAQUÍN GURRIARÁN BAS

RSID_PRS	CHR	BP	A1	A2	OR	P	Originales
rs6493387	15	31312023	T	C	1,0022	0,7809	rs6493387
rs12441390	15	38950560	T	C	0,99213	0,4531	rs12441390
rs4924398	15	40175964	A	G	0,99432	0,493	rs4924398
rs524908	15	44187991	A	C	1,02963	0,002921	rs524908
rs1647990	15	60916622	T	G	1,01664	0,04211	rs1647990
rs35406858	15	75279669	C	G	1,01715	0,1583	rs35406858
rs2269799	15	91796583	T	C	0,99164	0,3293	rs2269799
rs4777959	15	92862216	T	C	1,01847	0,02106	rs4777959
rs12438495	15	99477058	A	G	1,01379	0,08411	rs12438495
rs7194275	16	1481020	T	C	0,98738	0,2528	rs7194275
rs9931197	16	23231918	A	G	0,99025	0,3071	rs9931197
rs1423738	16	26799728	A	T	1,0028	0,7273	rs1423738
rs11644259	16	50371463	T	G	0,9994	0,9517	rs11644259
rs166211	16	65922934	A	T	1,01005	0,2209	rs166211
rs59464952	17	5585519	A	C	0,9994	0,9641	rs59464952
rs3826440	17	7411455	T	C	1,03149	0,0006426	rs3826440
rs75664430	17	8064779	C	G	1,01177	0,201	rs75664430
rs6607302	17	36194230	T	C	0,98413	0,1094	rs6607302
rs4795402	17	38085385	A	C	0,9969	0,7362	rs4795402
rs231478	17	41983627	C	G	0,99035	0,2445	rs231478
rs1352312	17	43376761	T	G	0,99302	0,4908	rs1352312
rs2277628	17	48601381	C	G	1,01684	0,04166	rs2277628
rs36074193	17	48983099	T	C	0,9862	0,1182	rs36074193
rs8082593	17	49509802	T	C	0,99531	0,5818	rs8082593
rs4968392	17	57799204	A	G	0,97814	0,04289	rs4968392
rs1893233	18	11411062	T	C	1,0024	0,8172	rs1893233
rs13090	19	868115	T	C	0,98442	0,1319	rs13090
rs2395891	19	2032148	T	G	1,00431	0,6484	rs2395891
rs12981801	19	2837468	T	C	0,9986	0,8618	rs12981801
rs7256770	19	11689911	A	G	0,99342	0,4348	rs7256770
rs2524379	19	14854034	T	C	0,99392	0,4521	rs2524379
rs91710	19	18141123	A	G	0,99312	0,4186	rs91710

RSID_PRS	CHR	BP	A1	A2	OR	P	Originales
rs55678304	19	18316629	A	C	0,99471	0,523	rs55678304
rs73526909	19	18949229	A	T	0,99322	0,6117	rs73526909
rs877836	19	29762562	A	G	0,99442	0,5069	rs877836
rs73048504	19	36191817	C	G	0,97404	0,01199	rs73048504
rs1559155	19	49632772	T	C	1,00531	0,5221	rs1559155
rs2253693	20	1545468	T	C	1,01369	0,09383	rs2253693
rs2209313	20	1599142	T	C	1,00572	0,5511	rs2209313
rs2027237	20	4177104	T	C	1,01187	0,1526	rs2027237
rs914314	20	24897355	T	C	1,01197	0,3044	rs914314
rs2387976	20	25599560	T	C	1,0007	0,9342	rs2387976
rs1007122	20	31346001	T	C	1,00652	0,4189	rs1007122
rs1041898	20	46561617	A	C	0,99481	0,5248	rs1041898
rs6070412	20	56865274	A	T	0,99452	0,4871	rs6070412
rs2828337	21	24995198	A	G	1,01076	0,2517	rs2828337
rs724208	21	34491984	A	G	1,01633	0,0424	rs724208
rs2329844	21	46086901	A	G	1,00441	0,5823	rs2329844
rs5763241	22	29839168	T	C	0,98787	0,3099	rs5763241
rs42931	22	30960412	T	C	0,99293	0,3984	rs42931
rs5994328	22	31046587	T	C	1,00491	0,5477	rs5994328
rs5756520	22	37508507	A	G	0,99651	0,6661	rs5756520
rs749326	22	38051019	A	G	1,00511	0,7183	rs749326
rs6001675	22	40165159	A	G	1,01349	0,1642	rs6001675

TABLA S2. 58 SNPs EN LD CON LISTADO ORIGINAL, MDD2 Y SUD_1524. RSID_PRS (código identificador dbSNP), CHR (cromosoma), BP (par de bases), A1 (alelo de riesgo), A2 (alelo mayor), OR (odds ratio, tamaño del efecto en MDD2), P (valor de p en GWAS MDD2), originales (código identificador Arloth *et al.* 2015), r^2 y D' (medidas de LD entre SNP seleccionado en SNIIPA y listado original de Arloth *et al.* 2015).

RSID_PRS	CHR	BP	A1	A2	OR	P	Originales	SNIIPA	R^2	D'
rs1280971	1	10978316	A	G	0,99045	0,2745	rs1280984	rs1280971	0,654079	0,973906
rs190043	1	26200602	A	G	1,01106	0,1747	rs213637	rs190043	0,98814	1
rs12727117	1	26726494	C	G	1,00381	0,7136	rs7535902	rs12727117	0,937648	1
rs12145754	1	111463268	T	C	0,9998	0,9831	rs2363536	rs12145754	0,703116	0,875624
rs12134335	1	154738779	A	G	0,9756	0,00863	rs4845391	rs12134335	0,529681	0,99305
rs10800356	1	168327555	T	G	1,00441	0,665	rs12128782	rs10800356	0,853463	0,986072
rs4072407	2	26671544	A	G	1,00864	0,461	rs2280516	rs4072407	0,884799	0,940637
rs12464738	2	62294076	A	G	1,00713	0,3761	rs6545924	rs12464738	0,729076	0,946943
rs6755948	2	73877621	A	G	0,99104	0,2662	rs12620091	rs6755948	0,892702	0,982115
rs10190632	2	86017386	A	G	0,99432	0,4727	rs10180924	rs10190632	0,769238	0,9115
rs335127	2	89100215	T	C	0,9975	0,7665	rs2848122	rs335127	0,921131	1
rs1467130	2	101306319	T	C	0,9976	0,7645	rs11123840	rs1467130	0,580172	-1
rs6755354	2	113826072	T	G	1,01939	0,1282	rs60157471	rs6755354	0,868316	0,94096
rs2680966	2	197596310	A	G	1,00652	0,4806	rs10931765	rs2680966	0,662815	0,820741
rs7625881	3	11550723	A	G	1,0139	0,2259	rs17034661	rs7625881	0,702595	0,847728
rs60206722	3	97964552	A	G	1,00371	0,7625	rs57988909	rs60206722	1	1
rs12636575	3	113474359	A	G	1,00823	0,3162	rs2712353	rs12636575	0,648913	0,9572
rs1949113	3	140901132	C	G	0,98147	0,1587	rs6439982	rs1949113	0,666182	0,925741
rs13145055	4	158775584	A	G	1,01268	0,2278	rs7678870	rs13145055	0,980912	1
rs4705421	5	149566932	T	G	0,98817	0,1488	rs917585	rs4705421	0,99603	1
rs6936881	6	1713692	T	C	0,98393	0,04628	rs12206258	rs6936881	0,747329	-0,995004
rs2875910	6	3830774	C	G	1,00975	0,2315	rs9503750	rs2875910	0,796935	-0,995238
rs116739320	6	32413830	C	G	1,02163	0,008902	rs9268671	rs3763327	0,733204	1
rs9481140	6	111394511	A	G	1,00743	0,6034	rs9320357	rs9481140	0,917914	0,975514
rs532409	7	64763167	A	G	1,00351	0,667	rs7796045	rs532409	0,871236	-0,990925
rs11762191	7	73100487	C	G	1,0016	0,8405	rs3793243	rs11762191	0,811691	0,986703
rs223014	7	73579794	T	C	1,00612	0,4857	rs444297	rs223014	0,837872	0,919441
rs7789389	7	150478821	C	G	0,98758	0,1333	rs17173596	rs7789389	0,970468	1

RSID_PRS	CHR	BP	A1	A2	OR	P	Originales	SNiPA	R ²	D'
rs3924018	8	23151236	T	C	1,00743	0,3965	rs7826635	rs3924018	0,918566	0,985612
rs11013110	10	22996441	T	G	0,98926	0,2061	rs6482235	rs11013110	0,903444	0,966355
rs11186787	10	82501616	A	G	0,9971	0,8432	rs12766521	rs11186787	0,605159	0,88722
rs10902108	11	225256	T	C	1,00803	0,4153	rs3802984	rs10902108	0,764584	0,929075
rs2923097	11	10365044	A	T	1,00371	0,6482	rs1562782	rs2923097	0,983667	1
rs12223076	11	74447549	A	G	1,0002	0,9841	rs12417156	rs12223076	0,789568	0,962518
rs483598	11	118509668	A	G	1,00995	0,2174	rs633683	rs483598	0,776941	0,995411
rs111723516	11	118722127	T	G	1,02296	0,2565	rs538645	rs111723516	0,968056	-1
rs11170172	12	52954844	A	G	1,00391	0,7077	rs12372446	rs11170172	1	1
rs4369555	13	48526833	T	C	0,99283	0,3612	rs9526443	rs4369555	0,889484	0,983389
rs8015364	14	21184482	T	C	1,0025	0,7565	rs12433896	rs8015364	0,757296	0,975737
rs55859043	14	61605415	A	G	1,00662	0,4705	rs17834472	rs55859043	0,988651	1
rs4902683	14	69442848	T	C	0,99571	0,6088	rs4902681	rs4902683	0,611379	0,876642
rs62011562	15	22935275	A	T	1,0141	0,1213	rs74450329	rs62011562	0,893702	0,984254
rs10519978	15	35406579	C	G	0,98649	0,1332	rs4924543	rs10519978	0,985133	-0,995011
rs11642741	16	13888790	T	C	1,01288	0,1168	rs7198922	rs11642741	0,654192	0,994609
rs7204332	16	29668804	C	G	0,99332	0,4075	rs11859842	rs7204332	0,793092	1
rs34876239	16	30498824	T	C	1,0006	0,945	rs12443981	rs34876239	0,77261	0,925736
rs11646602	16	35069526	A	G	0,99442	0,4766	rs11645488	rs11646602	0,765355	0,878385
rs36005190	16	86696224	C	G	1,0002	0,9874	rs300035	rs36005190	0,676963	0,825253
rs1548370	17	5951487	A	G	1,0011	0,8883	rs4985671	rs1548370	0,948045	0,991698
rs939540	17	72805783	T	C	0,98334	0,04112	rs2385067	rs939540	0,522486	-0,841317
rs1054695	19	1274601	T	C	1,00541	0,6366	rs60492890	rs1054695	1	1
rs778811	19	5828895	T	G	1,0007	0,9462	rs1379868	rs778811	0,9197	0,993369
rs34155014	19	41188429	A	T	1,002	0,8226	rs3865451	rs34155014	0,851959	0,932546
rs77284436	19	41959358	A	G	0,99422	0,6181	rs11083620	rs77284436	0,938408	1
rs58532	20	1792434	A	G	0,9995	0,9534	rs200891	rs58532	0,627102	0,965583
rs412123	21	38013082	A	G	0,99045	0,2284	rs417557	rs412123	0,933706	1
rs56288252	21	47077829	T	C	0,98926	0,3064	rs57057834	rs56288252	0,923041	0,96075
rs35269245	22	29632596	T	C	0,98946	0,2536	rs5752890	rs35269245	0,989654	0,994813





