



TESE DE DOUTORAMENTO

**Busca de variantes raras
en xenes asociados á
esquizofrenia**

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA
DEPARTAMENTO DE ANATOMÍA PATOLÓXICA E CIENCIAS FORENSES
FACULTADE DE MEDICINA

José Javier Suárez Rama

2017





O Doutor Angel Carracedo Álvarez, Catedrático de Medicina Legal da Facultade de Medicina da Universidade de Santiago de Compostela, e o Doutor Javier Costas Costas, Investigador do Servizo Galego de Saúde no Instituto de Investigación Sanitaria de Santiago de Compostela.

CERTIFICAN:

Que o presente traballo que leva por título “Busca de variantes raras en xenes asociados á esquizofrenia”, realizado por José Javier Suárez Rama baixo a nosa dirección, foi revisado e está en disposición de ser presentado para optar ó grado de Doutor en Medicina molecular.

E para que así conste, firmamos o presente certificado en Santiago de Compostela, a 20 de Xullo de 2017

Fdo: Dr. Ángel Carracedo Álvarez

Fdo: Dr. Javier Costas Costas



RESUMO

A esquizofrenia é unha enfermidade multifactorial complexa xa que na súa aparición interveñen tanto factores ambientais como xenéticos. Trátase dun trastorno mental grave, crónico e recorrente que xera unha gran carga social e familiar onde a súa prevalenza na poboación é dun 0,72% e na que, segundo estudos de epidemioloxía xenética, se estima unha herdabilidade entre o 65 e o 82%. Os estudos de GWAS (Genome Wide Association Studies), os cales están baseados na hipótese enfermidade común- variante común, conseguiron explicar parte da varianza da susceptibilidade a padecer da doenza, existindo unha gran proporción sen coñecerse. Unha hipótese complementaria á anterior é aquela que sostén que existen múltiples variantes pouco frecuentes na poboación, que teñen un risco asociado moderado e que estarían influíndo no desenvolvemento da esquizofrenia, sen que ningunha delas fora por si soa determinante para a doenza, denominada hipótese enfermidade común-variante rara. Este é o suposto no que se centra principalmente o presente traballo de tese.

Un dos primeiros exemplos que apoiaba a hipótese enfermidade común-variante rara en esquizofrenia era o xene *DISC1*. Un estudo identificaba nunha familia escocesa unha translocación balanceada afectando a este xene e que cosegregaba cos individuos afectados de diferentes enfermidades psiquiátricas, entre as que se atopaba a esquizofrenia. A vista deste resultado, o primeiro estudo desta tese centrábase en estudar variantes codificantes a baixa frecuencia no xene *DISC1* nunha mostra de orixe galego. Para isto secuenciabamos este xene nunha cohorte de casos e controis. A continuación xenotipabamos, no total da mostra dispoñible, as distintas variantes raras e non sinónimas atopadas na fase anterior, máis outras que atopabamos mediante buscas bibliográficas e nas bases de datos. Aínda que unha das variantes acadaba unha significación nominal, esta estratexia non nos permitía encontrar ningunha variante rara individual dunha forma robusta. Utilizando os datos de xenotipado dunha das variantes analizadas xunto con datos doutros estudos publicados anteriormente, levabamos a cabo unha meta- análise que nos permitía identificar unha posible variante funcional de risco que estaba situada na rexión 3'UTR de *DISC1*, podendo estar esta involucrada na regulación da expresión deste xene mediante os miARNs.

No seguinte traballo, utilizando os datos de xenotipado dun GWAS levado a cabo previamente polo noso grupo, estudabamos o papel das variantes comúns, codificantes e de sentido incorrecto no interactoma de *DISC1*. Para isto utilizabamos a estratexia que se basea no estudo dos xenes como un conxunto. Mediante esta metodoloxía identificabamos unha sobre- representación das proteínas que forman parte do interactoma de *DISC1*, sinalando a este conxunto de xenes como un factor a ter en conta no risco a padecer esquizofrenia. Ademais viamos que sete dos xenes incluídos na análise eran os responsables desta sobre- representación, estando estes involucrados na organización neuronal do citoesqueleto e no transporte intracelular ó través dos microtúbulos do citoesqueleto, suxerindo que estes procesos poderían estar afectados en individuos esquizofrénicos.

No estudo III da presente tese, tratabamos de buscar variantes raras cun risco asociado alto en xenes que formaban parte do interactoma de *DISC1* e xenes que previamente foran asociados á esquizofrenia mediante estudos de GWAS ou CNVs. Utilizando a metodoloxía de ultrasecuenciación analizabamos 21 xenes nunha cohorte de 51 *pools* de casos e outras tantas de controis (153 individuos en cada unha das cohortes). Cos datos previos do estudo de GWAS e do estudo de *DISC1*, valorabamos o

funcionamento de distintos paquetes informáticos utilizados para a identificación de variantes en *pools*, concluindo que o *software* VarScan é o que maior especificidade acadaba mantendo, a súa vez, unha alta sensibilidade. Cos datos das variantes atopadas mediante o VarScan viamos que existía un exceso de variantes funcionais a baixa frecuencia ($MAF < 5\%$), o cal podería estar indicando un posible papel deste tipo de variantes no desencadeamento da doenza. Despois de xenotipar, no total da mostra dispoñible, as variantes funcionais exclusivas dos casos, levabamos a cabo un estudo de asociación onde atopabamos unha variante cunha significación nominal presente no xene *AKAP9*. A continuación, mediante unha análise de acumulación de variantes, identificabamos unicamente dous xenes que obtiñan un valor de significación nominal, *ZNF804A* e *NRXNI*, existindo neste último xene un exceso de variantes en controis respecto ós casos.

Por último, e a vista dos resultados obtidos no traballo anterior, decidiamos estudar máis polo miúdo os xenes *ZNF804A* e *NRXNI*. Para isto levabamos a cabo unha busca bibliográfica coa finalidade de atopar estudos onde analizaban estes dous xenes en casos e controis para levar a cabo unha meta- análise. Finalmente centrámonos no xene *NRXNI* xa que foi o único xene onde atopabamos os estudos suficientes para levar a cabo unha meta- análise. A maiores dos datos publicados utilizabamos as variantes atopadas mediante secuenciación deste xene nun total de 384 casos e o mesmo número de controis agrupados en *pools* de 4 mostras, e que posteriormente confirmabamos mediante xenotipado. A continuación agrupabamos as variantes segundo a súa frecuencia e a súa funcionalidade para logo levar a cabo unha meta- análise en cada un dos conxuntos, acadando unicamente valores significativos cando considerabamos unicamente as variantes codificantes de sentido incorrecto cunha frecuencia menor ó 1%, confirmando así a existencia dun exceso deste tipo de variantes en controis. Este resultado é contraditorio respecto ó publicado ata o de agora sobre *NRXNI*, o cal está implicado no balance sináptico inhibitor- excitador, polo que pequenos cambios sutís neste equilibrio poderían dar lugar a un efecto protector contra a doenza.



AGRADECEMENTOS

Quería comezar dándolle as a grazas Ángel Carracedo por poder darme a inmensa oportunidade de forma parte deste grupo de investigación. Sempre lle estarei agradecido por permitirme coñecer o mundo da investigación podendo vivilo, desfrutalo e tamén sufrilo, como o fixen. Envexo a paixón con que vive e difunde a ciencia. Dende aquí quero recoñecerlle o traballo que fai día a día para que o laboratorio siga adiante.

Agradecer, como non, a Xabi Costas por guiarme, tutelarme e aconsellarme dunha forma inapelable durante todos estes anos. Tamén pola paciencia que amosou cando as cousas non ían tamén como quixeramos. Increible o traballo que amosa e desempeña como responsable do grupo de psiquiátricas.

Agradecerlle a Inés todas esas “brincas” que me botaba para que espabilase e chegase a este punto da tese. En parte, grazas a iso estou finalizando este traballo.

Tamén quería destacar a miña primeira “compi” de laboratorio: Noa. Foi a miña “iniciadora” no traballo de laboratorio e sempre me botou unha man no que necesitara nos meus inicios.

Juan, María, Bea, Jorge... Todos vós colaboraches dun xeito ou doutro neste traballo, e axudáchesme a resolver moitos dos problemas que me ían xurdindo. Grazas.

Álex, Kata, Eva, Julio, Xaquín, Peñas, Fátima e outra moita xente tamén, que fan que as horas que se pasan no laboratorio se fagan máis amenas, creando un ambiente de traballo agradable.

Os meus pais, que dicir deles, sempre me aconsellaron dun xeito inmejorable, ademais de amosarme sempre o seu apoio nas decisións que tomase. Grazas a eles son a persoa que son a día de hoxe.

Como non agradecer a miña compañeira eterna de viaxe, a que me escoita e me comprende. A que "sufriu" a tese en primeira persoa e que me aturou nos momentos difíciles deste largo proceso e que me soubo calmar e alentar. Grazas Anti, sen o teu apoio incondicional non sería quen de escribir esta tese.

Por último, destacar a tódalas persoas que colaboraron aportando mostras, así como ós doentes e familiares, parte esencial e primordial para poder levar un estudo de investigación.

Este traballo de púidose levar a cabo grazas o financiamento da Xunta de Galicia mediante o proxecto INCITE08PXIB9101149PR, do Instituto de Saúde Carlos III (ISCIII) a través do proxecto FIS/FEDER 08/1522 e da Fundación Galega de Medicina Xenómica (FPGMX) en colaboración co Servizo Galego de Saúde (SERGAS)



ÍNDICE

ABREVIATURAS.....	xii
1. INTRODUCCIÓN	1
1.1. A ESQUIZOFRENIA	3
1.1.1. Definición, orixe e sintomatoloxía.....	3
1.1.2. Evolución da enfermidade	4
1.1.3. Epidemioloxía.....	6
1.2. XENÉTICA DA ESQUIZOFRENIA	7
1.2.1. Epidemioloxía xenética.....	7
1.2.2. Procura de xenos asociados.....	9
1.2.2.1. Os Estudos de ligamento	10
1.2.2.2. Os Estudos de Asociación	11
1.2.2.2.1. Os Estudos de Asociación de Xenos Candidato.....	14
1.2.2.2.2. Os Estudos de Asociación do Xenoma Completo.....	15
1.2.2.3. Os Estudos de Análises de Conxuntos de Xenos.....	23
1.2.2.4. Os Estudos de Variantes Raras.....	25
1.2.2.4.1. O Xene <i>DISC1</i>	26
1.2.2.4.2. CNVs en Esquizofrenia.....	30
1.2.2.4.3. Estudo de SNVs mediante secuenciación tradicional	38
1.2.2.4.4. Estudo de SNVs mediante secuenciación de nova xeración.....	41
1.3. MÉTODOS DE ANÁLISE DUN CONXUNTO DE XENES.....	49
1.4. MÉTODOS DE ANÁLISE DE VARIANTES RARAS.....	55
2. HIPÓTESES E OBXECTIVOS	60
2.1. HIPÓTESES.....	62
2.2. OBXECTIVOS	62
3. MATERIAL E MÉTODOS.....	64
3.1. MOSTRAS.....	66
3.2. EXTRACCIÓN DO ADN.....	67
3.3. SECUENCIACIÓN DE SANGER	68
3.3.1. Optimización da PCR (Polymerase Chain Reaction)	68
3.3.2. Amplificación do ADN.....	68
3.3.2.1. <i>DISC1</i> (ESTUDO I)	68
3.3.2.2. Confirmación de SNVs identificadas co SOLiD4 System (Estudo III)	71
3.3.3. Secuenciación e análise das secuencias	71

3.4. SECUENCIACIÓN DE NOVA XERACIÓN	72
3.4.1. Secuenciación de nova xeración mediante SOLiD4 System (Estudo III)	72
3.4.1.1. Selección dos xenes	72
3.4.1.2. Preparación das mostrás	73
3.4.1.3. Preparación das bibliotecas	74
3.4.1.4. Secuenciación mediante SOLiD4.....	77
3.4.1.5. Análises bioinformáticas	78
3.4.2. Secuenciación de nova xeración de <i>NRXNI</i> mediante Ion Proton™ System (Estudo IV)	79
3.4.2.1. Preparación das mostrás	79
3.4.2.2. Preparación das bibliotecas	79
3.4.2.3. Secuenciación mediante Ion Proton	82
3.4.2.4. Análises Bioinformáticas.....	82
3.5. XENOTIPADO	83
3.5.1. Selección de variantes a xenotipar.....	83
3.5.2. Sequenom Mass Array® System.....	83
3.6. ANÁLISE DE DATOS	86
3.6.1. Secuenciación de <i>DISC1</i> (Estudo I)	86
3.6.2. Interactoma <i>DISC1</i> (Estudo II).....	86
3.6.3. Secuenciación mediante SOLiD4 System (Estudo III).....	87
3.6.4. Secuenciación de nova xeración de <i>NRXNI</i> mediante Ion Proton™ System (Estudo IV)	88
4. RESULTADOS	90
4.1. ESTUDO I: ANÁLISE DE VARIANTES RARAS NO XENE <i>DISC1</i>	92
4.1.1. Introducción.....	92
4.1.2. Busca bioinformática de posibles SNPs funcionais a baixa frecuencia.....	93
4.1.3. Secuenciación das rexións exónicas de <i>DISC1</i> na mostra descubrimento	93
4.1.4. Estudo de asociación caso-control de SNPs funcionais a baixa frecuencia	94
4.1.5. Meta-análise dos resultados de rs3737597 e E751Q	96
4.2. ESTUDO II: PAPEL DAS PROTEÍNAS QUE INTERACTÚAN CON <i>DISC1</i> NO RISCO A PADECER ESQUIZOFRENIA A PARTIR DE DATOS DUN ESTUDO DE ASOCIACIÓN DO XENOMA COMPLETO DE SNPS CODIFICANTES DE SENTIDO INCORRECTO	99
4.2.1. Introducción.....	99
4.2.2. Análise de conxunto de xenes a partires dos datos de xenotipado previos....	100

4.3. ESTUDO III: RESECUENCIACIÓN E ANÁLISE DE ASOCIACIÓN DE REXIÓNS CODIFICANTES DE 21 XENES CANDIDATO NA SUSCEPTIBILIDADE A ESQUIZOFRENIA	103
4.3.1. Introducción.....	103
4.3.2. Análise de cobertura dos xenes estudados	103
4.3.3. Selección do paquete informático para a detección de variantes.....	106
4.3.4. Detección de variantes	108
4.3.5. Estudo de asociación.....	109
4.4. ESTUDO IV: RESECUENCIACIÓN DO XENE NRXN1.....	113
4.4.1. Introducción.....	113
4.4.2. Análise de cobertura de NRXN1	113
4.4.3. Detección de variantes e confirmación mediante Sequenom	115
4.4.4. Estudo de asociación e meta-análise.....	117
5. DISCUSIÓN.....	120
6. CONCLUSIÓNS	141
7. BIBLIOGRAFÍA.....	145





ABREVIATURAS





ADN: Ácido desoxirribonucleico.
AMPc: Adenosín monofosfato cíclico
ANS: Ansiedade.
ARN: Ácido ribonucleico.
ARNm: ARN mensaxeiro.
aSum: data-adaptative sum Test.
C:Citosina.
Ca: Caso.
CADD: Combined Annotation
Dependent Depletion.
CAST: Cohort Allelic Sums Test.
CEU: Colección de residentes en Utah
con devanceiros do norte e oeste de
Europa doCEPH
CMC: Combined Multivariate and
Collapsing Test.
Co: Control.
CEGEN: Centro Nacional de
Genotipado.
CEPH: Centre d'Etude du
Polymorphisme Humain.
CGH: Comparative genomic
hybridization.
CNV: Copy Number Variation.
DI: Discapacidade Intelectual.
Del: Deleción.
DL: Desequilibrio de Ligamento.
DM: Depresión Maior
Dup: Duplicación.
dNTP: desoxinucleotido trifosfato.

EDTA: Ácido
etilendiaminotetraacético.
ENCODE: Encyclopedia of DNA
Elements.
EP: Epilepsia.
ePCR: Emulsion PCR
EREC: Estimated Regression
Coefficient Test.
ExAC: Exome Aggregation
Consortium.
FDR: False discovery rate
FESD: Functional Element SNPs
Database.
FISH: Fluorescence In Situ
Hybridization.
FPGMX: Fundación Pública Galega de
Medicina Xenómica.
G: Guanina.
GATK: Genome Analysis Toolkit.
GO: Gene Ontology.
GWAS: Genome Wide Association
Studies.
IC: Intervalo de confianza.
ISC: International Schizophrenia
Consortium.
ISP: Ion Sphere™ Particle.
Indel: Inserción/deleción.
Kb: Quilobase.
KBAC: Kernel-based Adaptive Cluster
Test.

KEGG: Kyoto Encyclopedia of Genes and Genomes.

LHD: Lugares de Hipersensibilidade a DNasa I.

LINE1: Long interspersed nuclear elements type 1.

LOF: Loss of function.

MALDI-TOF MS: Matrix-Assisted Laser Desorption/Ionization- Time-Of-Flight Mass Spectrometry.

MAF: Minor Allele Frequency.

Mb: Megabase.

MDS: Multidimensional Scaling.

MgCl₂: Cloruro de Magnesio.

MGS: Molecular Genetics of Schizophrenia.

MHC: Major Histocompatibility Complex.

miRNA: Micro ARN.

μL: Microlitro.

MLPA: Multiplex ligation-dependent probe amplification.

MSIGDB: Molecular Dignatures Database.

NCBI: National Center for Biotechnology Information.

ng: Nanogramos.

NGS: Next Generation Sequencing.

NHBLI ESP: National Heart, Lung, and Blood Institute Exome Sequencing Project.

NS: Non Sinónimas

OR: Odds Ratio.

Pb: Pares de bases.

PCA: Principal Component Analysis.

PCR: Polymerase Chain Reaction.

PGC: Psychiatric Genomics Consortium.

PID: Pathway Interaction Database.

qPCR: quantitative PCR.

RES: reguladores exónicos de splicing

RM-MC: Retraso Mental e Malformacións Conxénitas.

S: Sinónimas.

SAP: Shrimp Alkaline Phosphatase.

SBE: Single Base Extension.

SCZ: Esquizofrenia.

SCZA: Trastorno Esquizoafectivo.

SKAT: Sequence Kernel Association Test.

SKAT-O: SKAT-Optimal Test.

SNP: Single Nucleotide Polymorphism.

SNV: Single Nucleotide Variation.

SOLID: Sequencing by Oligonucleotide Ligation and Detection.

SWG of PGC: Schizophrenia Working Group of the Psychiatric Genomics Consortium.

TB: Trastorno Bipolar

T^a: Temperatura.

Taq: Thermus Aquaticus.

TB: trastorno Bipolar

TD: Trastorno depresivo.

tdPCR: Touchdown PCR.

TDV: Trastorno do Desenvolvimento.

TE: Tris-EDTA.

TEA: Trastorno do Espectro Autista.

THDA: Trastorno de Hiperactividade e Déficit de Atención.

TPH: Trastorno Pitt-Hokins.

TSI: Tuscan in Italy.

TVC: Torrent Variant Caller.

T1000G: The 1000 Genomes Project.

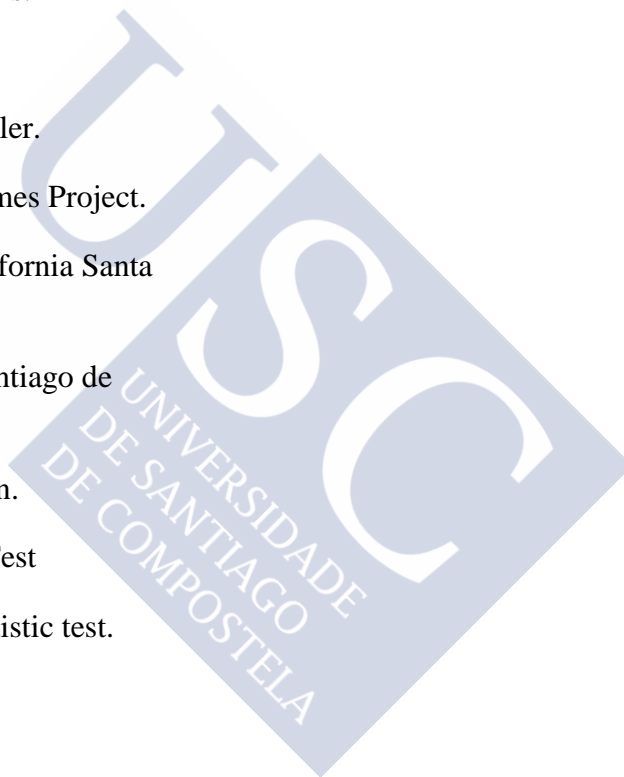
UCSC: University of California Santa Cruz.

USC: Universidade de Santiago de Compostela.

UTR: Untranslated Region.

VT: Variable Threshold Test

WSS: Weighted-Sum Statistic test.







1. INTRODUCCIÓN





1.1. A ESQUIZOFRENIA

1.1.1. Definición, orixe e sintomatoloxía

A esquizofrenia é un trastorno caracterizado por un conxunto de síntomas psicóticos (destacando as alucinacións e delirios), alteración da afectividade, comportamento desorganizado e alteración da percepción da realidade. Os síntomas que presenta un afecto ó longo da súa vida poden variar. Ademais, diferentes pacientes non teñen por que presentar os mesmos síntomas. Todo isto danos unha idea da gran heteroxeneidade etiopatolóxica, sintomatolóxica e do curso do trastorno (Keshavan et al. 2008; Tandon et al. 2008).

En 1857, Morel propoñía por primeira vez a teoría da herdabilidade nos trastornos mentais. Sostiña que estes trastornos son resultado de defectos biolóxicos innatos e a severidade das síndromes mentais aumenta de formar lineal nos descendentes. El acuñaba o termo *démence précoce* no 1852. Este termo non o utilizaba como unha categoría clínica senón que era utilizado simplemente dun xeito descritivo: Nel englobaba a pacientes xoves que presentaban unha demencia prematura (Adityanjee et al. 1999; Ban 2004).

Kraepelin (1856-1926), seguindo a teoría de Morel, foi o primeiro clínico que comezaba a diferenciar as psicoses. Por unha banda distinguía "dementia praecox" para referirse ós casos nos que a aparición da enfermidade tiña lugar na adolescencia ou nos primeiros anos da idade adulta, por presentar tendencia ó deterioro, falta de lucidez mental e demencia. O outro grupo que diferenciaba era a loucura maníaca-depresiva, a cal estaba caracterizada pola súa episodicidade, ausencia de deterioro e por uns resultados máis favorables (Tandon et al. 2009).

Era finalmente Bleuler (1857-1939) quen propoñía substituír "dementia praecox" co termo de esquizofrenia, definindo a enfermidade como un grupo de psicoses caracterizadas por un tipo específico de alteración do pensamento, emocións e relación co mundo exterior. Outra das contribucións destacadas de Bleuler era a creación dunha listaxe de síntomas. Diferenciaba síntomas fundamentais, os cales consideraba característicos da enfermidade e que debían de estar presentes nunha persoa ó longo de toda a vida, e síntomas accesorios que podían ou non estar presentes (Andreasen and Carpenter 1993).

Por último, Schneider (1958) describía unha serie de se síntomas patognomónicos, ós cales chamaba síntomas positivos (Tandon et al. 2009).

Polo tanto, ó longo do século XX era cando se asentaba o concepto de esquizofrenia definíndose este a partir da cronicidade de Kraepelin, os síntomas negativos de Bleuler e os síntomas positivos de Schneider (Tandon et al. 2009). Ata o de agora non se identificou ningún marcador biolóxico que caracterice a enfermidade dunha forma obxectiva, polo que o seu diagnóstico baséase na observación clínica do estado mental do afecto existindo unhas características clínicas xerais da enfermidade, as cales aparecen recollidas na quinta versión do Manual Diagnóstico e Estatístico dos Trastornos Mentais da Asociación Americana de Psiquiatría (DMS-5) (Braff et al. 2013) e na décima revisión da Clasificación Internacional de Enfermidades da Organización Mundial da Saúde (ICD-10; World Health Organization, 1992).

A esquizofrenia é un trastorno mental grave, crónico e cunha expresión moi variada na que os seus principais síntomas clasifícanse en:

Introdución

Os **síntomas positivos**: son exacerbacións de fenómenos normais. Inclúen delirios, alucinacións e outro tipo de distorsións da realidade. Poden ocorrer distintos tipos de delirios e alucinacións afectando a calquera dos 5 sentidos e ademais en diferente grao de persistencia ou sistematización. Este tipo de síntomas son os que mellor resposta teñen fronte a medicación antipsicótica (Tandon et al. 2009).

Os **síntomas negativos**: Implican a redución ou perda dun rango de afectividade e da función conativa. Aquí inclúese incapacidade fronte a experiencias afectivas e expresión, abulia (falta de motivación), falta de empatía, empobrecemento da linguaxe, anhedonia (incapacidade para experimentar pracer), falta de iniciativa e conduta social reducida (Tandon et al. 2009).

Os **síntomas cognitivos** son a diminución da capacidade para inferir e interpretar os sentimentos doutros, a tendencia a atribuír emocións negativas a outras persoas e a distorsión da atención, do procesamento da información, alteración do procesamento da fala e a fluidez verbal.

Os **síntomas de desorganización do pensamento e comportamento**: Refírese á fragmentación da natureza lóxica e progresiva do pensamento normal, podendo variar en gravidade dende a circunstancialidade e tanxencialidade relativamente leve ata a incoherencia máis severa e mestura de palabras. Inclúe tanto o descarrilamento e neoloxismo (trastorno do pensamento formal positivo) como a pobreza do contido do pensamento (trastorno do pensamento formal negativo).

Os **síntomas psicomotores**: Os doentes habitualmente manifestan anomalías tanto na extensión como na natureza da actividade psicomotora. As desordes da actividade psicomotora abranguen dende movementos de postura illados e simples ou manierismos ata patróns de movemento, como os observados en varios estados catatónicos (Tandon et al. 2009).

Síntomas da conduta como hostilidade, impulsividade, illamento e conduta antisocial (Tandon et al. 2009).

1.1.2. Evolución da enfermidade

A esquizofrenia é un trastorno mental grave, crónico e recorrente que xera unha gran carga social e familiar, tendo unha prevalencia na poboación de 0,72% (McGrath et al. 2008). O trastorno soe ter unha traxectoria secuencial que empeza coa **fase premórbida** na que os individuos que van a padecer a enfermidade sofren problemas cognitivos, emocionais e do desenvolvemento do comportamento, acompañados por discapacidade en funcións sociais e académicas. Entre as anormalidades resaltan o retraso no desenvolvemento motor, disfunción da atención, déficit na linguaxe receptiva, malos resultados académicos, illamento social e aplanamento emocional (Schenkel and Silverstein 2004).

A seguinte etapa denomínase **fase prodrómica** e corresponde á fase anterior á aparición da doenza. Está caracterizada por síntomas positivos atenuados (síntomas básicos) e funcións decadentes, xunto cunha serie de signos clínicos: déficits cognitivos, síntomas negativos, síntomas do humor e un descenso da función (Cornblatt et al. 1999). Os síntomas cognitivos, negativos e depresivos aparecen, de media, sobre 5 anos antes do primeiro contacto clínico e a discapacidade social aproximadamente de 1 a 3 anos antes. Por outra banda, os síntomas positivos soen manifestarse un ano antes do primeiro contacto clínico (Tandon et al. 2009).

A aparición do primeiro episodio psicótico marca o inicio da **fase psicótica**, que é cando realmente aparece o trastorno. O inicio desta é difícil de estimar debido ó proceso continuo da evolución da enfermidade dende a fase prodrómica á psicose e tamén debido as distintas definicións existentes na actualidade desta fase de inicio, entre as que destacan: a aparición dos primeiros signos de trastornos mentais, primeiros síntomas positivos, primeiras evidencias de disfunción social, primeiro contacto clínico ou primeira hospitalización (Beiser et al. 1993; Hafner 1998). Segundo o DSM-5 considérase que a enfermidade iníciase se o individuo presenta, ó longo dun mes, polo menos dous destes cinco síntomas: alucinacións, delirios, comportamento ou fala desorganizada ou síntomas negativos. Para a confirmación do diagnóstico é necesario que se manifeste, polo menos, un dos tres primeiros. Nesta fase acentúanse os síntomas positivos e os negativos fanse máis prominentes a medida que o trastorno vai evolucionando. A esquizofrenia tipicamente maniféstase na adolescencia ou no inicio da idade adulta, aínda que nas mulleres debuta, de media, de 5 a 7 anos máis tarde (Hafner 1998). Relacionado con isto dicir que a distribución da idade de aparición nos homes é unimodal mentres que nas mulleres é bimodal, amosando un pico nos dous xéneros entre os 18 e 30 anos e un segundo pico nas mulleres entre os 45 e 49 anos (Fig.1) (Hafner 1998; Hafner 2003).

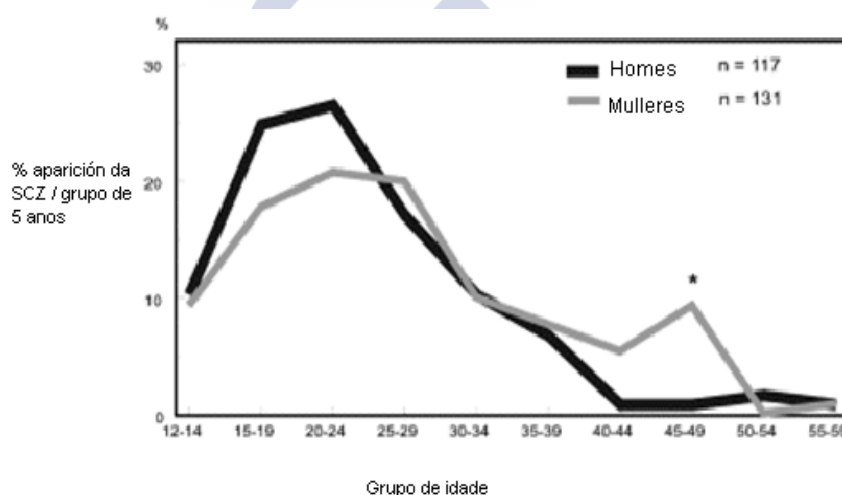


Figura 1. Distribución da idade de aparición da esquizofrenia. Adaptada de (Hafner 2003).

Por último, o trastorno entra nunha **fase estable** ou **de meseta** onde os síntomas positivos tenden a ser menos severos e os síntomas negativos e o déficit cognitivo son cada vez máis predominantes (Tandon et al. 2009). Esta fase está caracterizada ou por unha remisión do trastorno ou por unha cronicidade, aínda que poden existir exacerbacións posteriores.

Aínda que clasicamente divídese nestas fases, a evolución da esquizofrenia varía amplamente entre os pacientes e a separación destas fases moitas veces é imprecisa debido a gran heteroxeneidade do curso do trastorno. Ademais os síntomas prodrómicos e premórbidos non teñen valor para dar unha predición do diagnóstico xa que aparecen nunha gran variedade de alteracións do comportamento en individuos que posteriormente non padecen a enfermidade (Tandon et al. 2009).

1.1.3. Epidemioloxía

A esquizofrenia é un dos trastornos psiquiátricos máis debilitantes tendo grandes efectos sobre os pacientes e as súas familias, afectando a calidade de vida individual e provocando unha gran carga familiar. Os individuos afectados presentan unha elevada taxa de mortalidade, de feito os esquizofrénicos teñen un risco de morte de dúas a tres veces maior que a poboación xeral (McGrath et al. 2008), provocando que a esperanza de vida dun individuo afecto estea reducida entre 15 e 20 anos. Gran parte desta mortalidade é debida á elevada taxa de suicidio dos afectados comparado á poboación xeral, de feito un terzo dos individuos con esquizofrenia intenta suicidarse unha ou máis veces e o 5% dos que padecen o trastorno morren de suicidio. Por outro lado, sobre un 10% desta mortalidade é explicada por accidentes (Saha et al. 2007). Outra das explicacións desta elevada taxa é o gran número de afectados que padecen enfermidades cardiovasculares, sendo a maior causa de morte entre os esquizofrénicos (Olfson et al. 2015).

A esquizofrenia é un trastorno complexo e multifactorial, xa que a súa aparición débese a un conxunto de factores xenéticos e ambientais. Entre estes últimos, existe unha gran variedade de factores implicados na etioloxía da esquizofrenia. Por exemplo, infeccións maternas (influenza, rubéola, toxoplasmore, etc.) ou unha mala nutrición durante o primeiro ou o inicio do segundo trimestre de embarazo están ligadas a un aumento na susceptibilidade a desenvolver esquizofrenia (Penner and Brown 2007). Relacionado con isto, as complicacións obstétricas e perinatais como a hipoxia, dano cerebral no feto e a diabetes materna son outros dos riscos asociados (Brown 2011). Outro dos factores que pode influenciar é a idade paterna avanzada, chegando a provocar o dobre de risco a padecer esquizofrenia (Wohl and Gorwood 2007).

A estación de nacemento tamén inflúe no risco do trastorno, estando a etapa de finais de inverno e principios da primavera asociada a ter entre un 5-10% maior probabilidade de padecer esquizofrenia (Davies et al. 2003). Este efecto da estación de nacemento parece ser que se acentúa co aumento da latitude e a severidade da estación invernal (Tandon et al. 2008). O sexo é outro dos factores que inciden sobre a esquizofrenia sendo máis frecuente nos homes que nas mulleres (ratio 1,4:1) (Aleman et al. 2003; McGrath et al. 2004). O feito de ser inmigrante tamén modifica o risco a sufrir esquizofrenia, de xeito que hai unha maior incidencia e prevalencia da esquizofrenia entre as persoas inmigrantes (McGrath et al. 2008). Existen factores sociais (discriminación, racismo, illamento social, etc.), biolóxicos (falta de vitamina D) e mecanismos epixenéticos que poderían explicar a relación entre a esquizofrenia e a condición de inmigrante (Tandon et al. 2008).

Varios estudos amosaban que a susceptibilidade a padecer esquizofrenia é maior entre as persoas que viven nunha cidade respecto das que habitan nun ámbito rural (Marcelis et al. 1999; March et al. 2008). Isto parece estar relacionado con algúns factores que están presentes en maior grao nas urbes, tales como un maior consumo de cannabis e abuso doutras substancias, maior grao de estrés e relacións sociais, toxinas ambientais... Aínda que todos estes factores foron ligados independentemente ó trastorno, ningún deles explica con firmeza estas diferenzas entre o binomio urbe-campo (Tandon et al. 2008).

A esquizofrenia está ligada a unha fertilidade menor comparada á poboación xeral, acadando unha taxa fertilidade do 39% (Bundy et al. 2011). A explicación disto é que os afectados teñen problemas de integración social e soidade característica facendo difícil

atopar parella coa consecuenta menor probabilidade de ter fillos (Haukka et al. 2003; Bundy et al. 2011).

Todos estes aspectos comentados anteriormente foron ligados algunha vez á esquizofrenia, provocando unha maior susceptibilidade a desenvolver o trastorno, pero a relevancia de todos eles aínda non está clara, de feito ningún destes factores é necesario ou suficiente para causar a esquizofrenia, nin ningún ten tanta influencia na etioloxía como os factores xenéticos.

1.2. XENÉTICA DA ESQUIZOFRENIA

1.2.1. Epidemioloxía xenética

A importancia relativa dos factores ambientais e xenéticos varía moito en función da enfermidade a estudar. Para determinar estes factores habitualmente empréganse estudos de epidemioloxía xenética. Esta é unha disciplina que trata de estudar os factores xenéticos que provocan unha determinada enfermidade, ademais da interacción destes cos factores ambientais. Este tipo de estudos utilízanse para estudar doenzas como a esquizofrenia, xa que se trata dun trastorno multifactorial complexo no que os factores xenéticos teñen unha gran relevancia, e mediante os cales se estimaba unha elevada herdabilidade (proporción de varianza fenotípica dun determinado trazo nunha poboación explicada por factores xenéticos). Ata non fai moito, o valor da herdabilidade que se manexaba máis habitualmente roldaba o 82%, valor estimado mediante unha meta-análise de estudos publicados en xemelgos (Sullivan et al. 2003). A raíz dun traballo publicado máis recentemente por Lichtenstein e cols. no que avaliaban a influencia ambiental e xenética na susceptibilidade á esquizofrenia e trastorno bipolar e no que estimaban que a herdabilidade da esquizofrenia está entorno ó 65% (Lichtenstein et al. 2009), empezouse a utilizar como referencia o rango comprendido entre estes dous valores (65%-82%).

A herdabilidade non proporciona información sobre a influencia de factores xenéticos nun único individuo, nin dá unha estima de cantos xenes poden estar relacionados coa enfermidade. A herdabilidade, normalmente, calcúlase comparando a variación de trazos en individuos moi relacionados, como é o caso de estudos de agregación familiar, adopción e de xemelgos, nos que se demostraba que o risco a padecer esquizofrenia aumenta entre familiares de individuos afectos (Fig. 2). Os estudos de xemelgos permítenos calcular a taxa de concordancia entre irmáns (probabilidade de que un xemelgo padeza a mesma enfermidade que o seu irmán afecto) estimándose que o risco de padecer esquizofrenia en xemelgos monozigóticos dun individuo afecto aumenta ao redor do 50% e en xemelgos dizigóticos dun 17%. Por outro lado, este incremento no risco a padecer enfermidade alcanza un valor dun 12% de media entre familiares de primeiro grao e un 5% entre familiares de segundo grao (Riley and Kendler 2006).

Introdución

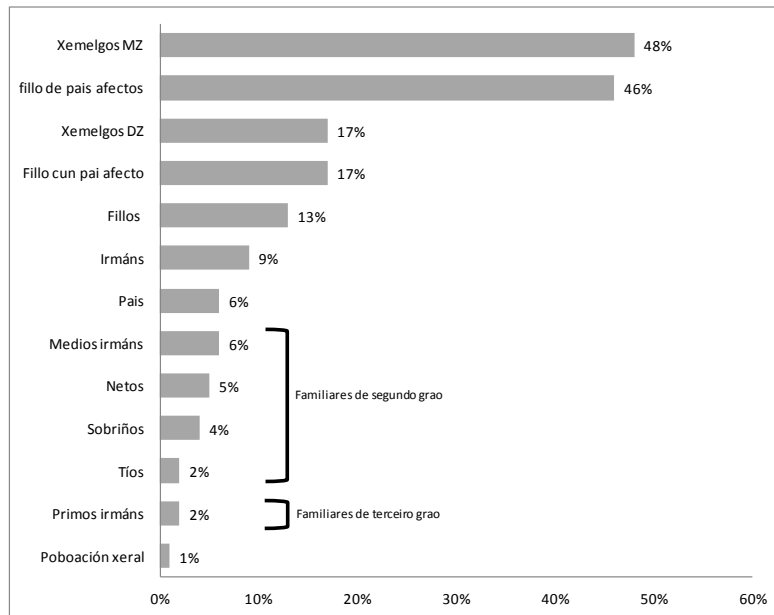


Figura 2. Risco de padecer esquizofrenia segundo o grao de parentesco cun afecto. Adaptación de (Riley and Kendler 2006).

Por outra banda, os estudos de adopción permítenos diferenciar, de forma máis clara do que o fan os estudos de xemelgos, o risco ambiental do risco xenético a padecer unha determinada enfermidade (Ingraham and Kety 2000). Este tipo de estudos permitía ver que o risco a sufrir esquizofrenia estaba relacionado coa parte biolóxica máis que co feito de que os individuos se desenvolvan en ambientes distintos; como así se vía en estudos no que o risco era similar para descendentes de nais con esquizofrenia se a descendencia era criada polos pais biolóxicos (afectos) ou por uns pais adoptivos (non afectos) (Higgins 1976). Ademais amosaban que descendentes de nais non afectas non tiñan un risco incrementado cando son criados por pais adoptivos psicóticos (Wender et al. 1974).

A partir de estudos de agregación familiar pódese concluír que, ó igual que outras enfermidades comúns, a esquizofrenia ten un modo de transmisión complexo e compatible co modelo polixénico (Gottesman and Shields 1967). A complexidade da esquizofrenia non só se explica por factores xenéticos aditivos, é dicir, múltiples alelos de risco independentes, senón tamén por efectos epistáticos (interacción xene-xene). Aínda que a esquizofrenia é altamente herdable, existe unha contribución significativa de factores ambientais, os cales poderían modificar a función cerebral. Alteracións epixenéticas, tanto a nivel de metilación do ADN como a nivel da modificación da cromatina, poden modificar o potencial transcricional dunha célula e mediar a influencia xenética, do desenvolvemento e ambiental no risco a padecer esquizofrenia. Ademais disto, a interacción entre xene-ambiente aumenta a susceptibilidade, non sendo ningún dos factores comentados capaces de orixinar a enfermidade por si mesmo.

1.2.2. Procura de xenes asociados

Está ben establecido que a esquizofrenia ten unha gran compoñente xenética, pero a natureza da súa arquitectura xenética non se coñece. Esta última refírese ó número de loci xenómicos que contribúen ó risco, á distribución da frecuencia dos alelos de risco en cada loci, ó tamaño do efecto e a interacción dos alelos dentro e entre os xenes. Segundo os datos dispoñibles, a arquitectura xenética da esquizofrenia é diversa e inclúe loci ó través do espectro alélico: moitas variantes comúns de efecto sutil como son os SNPs (en inglés, *Single Nucleotide Polimorphisms*), é dicir, polimorfismos dun único nucleótido, que son substitucións dunha única base; as CNVs (en inglés *Copy Number Variants*) que son variantes en número de copia que están presentes a baixa frecuencia na poboación e altamente penetrantes; e as variacións puntuais raras (SNVs, do inglés *Single Nucleotide Variants*) (Fig.3).

Ó longo do tempo foron empregándose distintas estratexias co fin de intentar clarear esta arquitectura xenética. Estes estudos réxense fundamentalmente por dúas hipóteses principais: a hipótese enfermidade común-variante común e a hipótese enfermidade común-variante rara. A primeira postula que a distinta susceptibilidade individual a padecer unha enfermidade complexa é por mor da presenza de múltiples alelos frecuentes na poboación que causan un risco moi baixo de xeito individual (Lander 1996). A presenza destas variantes de elevada frecuencia é por dúas razóns: a primeira é debido a que as variantes que teñen un risco asociado baixo non son eliminadas pola selección purificadora, polo que a deriva xénica fai que aumente a súa frecuencia na poboación; ademais as enfermidades complexas que se manifestan en idades tardías a selección purificadora é menos eficaz a hora de eliminar os alelos que aumenten o risco a padecer ditas doenzas, polo que, outra vez, a actuación da deriva xénica provoca o aumento da súa frecuencia. A segunda razón é que este tipo de variantes poderían verse favorecidas mediante a selección balanceadora xa que poderían conferir algunha vantaxe en estado heterocigoto ou a través do seu efecto nalgún outro trazo (pleiotropismo). Na literatura hai numerosos exemplos que sosteñen esta hipótese. O máis claro, e que a súa vez inspirou esta hipótese, é o do alelo $\epsilon 4$ da *APOE*, que confire unha maior susceptibilidade a padecer doenza de Alzheimer, e a súa frecuencia oscila entre o 10-20 % en poboacións europeas (Saunders et al. 1993). Outro exemplo é a variante atopada a frecuencias elevadas no xene *MTHFR*, a cal altera a función da encima que codifica, provocando a existencia de niveis elevados de homocisteína en plasma, que a súa vez era identificado como un factor de risco para enfermidades coronarias, cerebrovasculares e vasculares periféricas (Frosst et al. 1995).

Por outra banda, a hipótese enfermidade común-variante rara sostén que as variantes que aumentan o risco a padecer unha patoloxía están a moi baixa frecuencia na poboación e o risco asociado a cada unha delas é máis elevado. Sería unha situación similar á que se da en enfermidades mendelianas ou simples (aquelas que son causadas por mutacións nun único xene, e ademais este xene ten gran efecto na enfermidade). A hipótese postula que as variantes de susceptibilidade serían mutacións novas, e habería un equilibrio mutación/selección que impediría que aumentasen a súa frecuencia na poboación. No caso de enfermidades complexas, nas que hai máis dunha variante implicada, o efecto de cada mutación non sería tan severo como o das mutacións responsables das enfermidades mendelianas (Pritchard 2001).

Introducción

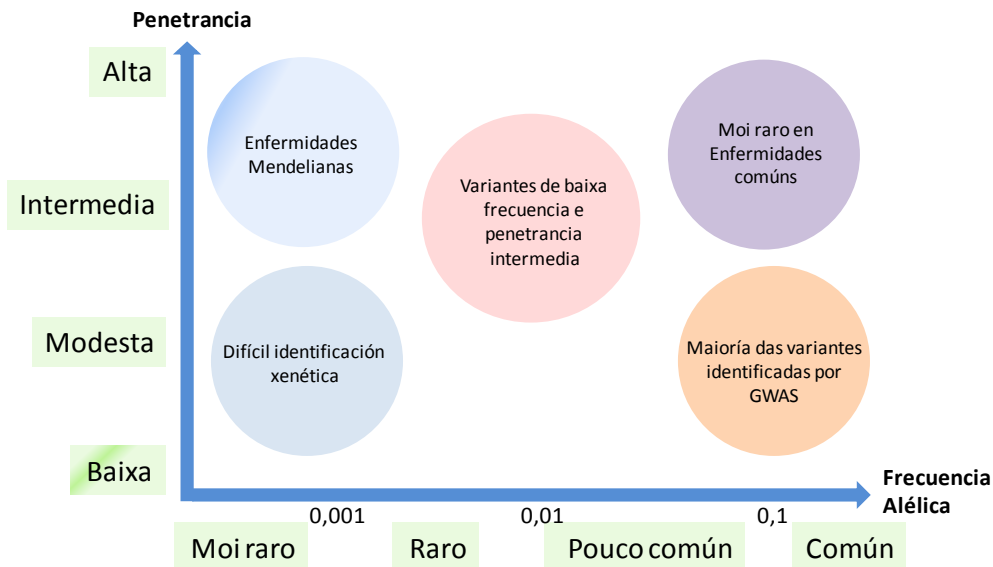


Figura 3. Arquitectura xenética de enfermidades comúns como, por exemplo, a esquizofrenia. Figura adaptada de (McCarthy et al. 2008).

Estes dous modelos non son mutuamente excluíntes, xa que se coñece que variantes raras deletéreas están presentes en xenes que tamén conteñen variantes comúns con efectos modestos no risco a padecer unha enfermidade. Un exemplo disto é o traballo levado a cabo por Steingberg e cols. no 2011, no que, ademais de confirmar a asociación previa do SNP rs134470 coa esquizofrenia e que estaba situado no xene *ZNF804A*, atopaban dúas CNVs raras de efecto moderado que afectaban a este xene alcanzando valores significativos (Steinberg et al. 2011b). Outro exemplo claro sería o estudo onde atopaban unha asociación significativa entre o SNP rs4583255 e a esquizofrenia, o cal está situado no xene *TAOK2*, estando este xene a súa vez localizado no locus 16p11.2, rexión afectada por unha CNV rara asociada previamente á esquizofrenia (Steinberg et al. 2014). Polo tanto, en enfermidades xenéticas complexas como a esquizofrenia, un subconxunto de casos poden ser causados por mutacións raras cun efecto grande, mentres que outro subconxunto podería ser resultado da interacción de múltiple variantes comúns cun efecto pequeno. Destacar que tamén a combinación de ámbolos dous tipos de variantes poderían dar lugar a doenza (Schreiber et al. 2013).

1.2.2.1. Os Estudos de ligamento

Ó longo da pasada década a metodoloxía de estudo máis empregada para tratar de atopar a base xenética dunha enfermidade hereditaria eran os estudos de ligamento. Mediante este tipo de estudos o que se pretende é identificar un determinado marcador molecular que cosegrega cunha enfermidade nunha xenealoxía que presenta varios individuos afectados. Os estudos de ligamento tiveron un gran éxito en identificar loci responsables de enfermidades mendelianas. Por outra banda, estes estudos tiveron moi pouca eficacia á hora de identificar marcadores moleculares ligados a enfermidades comúns xa que non existe un único locus responsable da patoloxía e o efecto asociado a cada un destes loci é baixo ou moderado. Outra das limitacións deste tipo de estudos é que para que sexan efectivos necesítanse un gran número de familias para identificar os xenes implicados, cousa que é difícil de conseguir no caso da esquizofrenia xa que a maioría dos casos son esporádicos. Outra desvantaxe desta estratexia é que é necesario asumir un modelo de herdanza concreto, cousa que nas enfermidades comúns como a esquizofrenia é imposible determinar.

Debido ás limitacións desta estratexia, os resultados atopados non foron os esperados, xa que ningún deles alcanzaba o nivel de significación xenómico necesario nin foron replicados en mostras independentes (Owen et al. 2005). Hai exemplos de estudos de ligamento nos que se atopaban rexións cromosómicas asociadas á esquizofrenia, con valores de significación menores a 0,05, como eran as rexións 6p24–22, 1q21–22 e 13q32–34 (Straub et al. 1995; Blouin et al. 1998; Brzustowicz et al. 2000) pero ningunha delas era replicada, indicando que, probablemente, representen falsos positivos (Owen et al. 2004). Todos estes resultados espurios, como xa comentamos anteriormente, posiblemente sexan debidos á combinación do pequeno efecto xenético atribuído ós marcadores, a un tamaño de mostras inadecuado ou ó uso dun mapa de marcadores cunha densidade insuficiente para conseguir toda a información xenética.

Ó longo de todos estes anos fóronse obtendo e acumulando datos dos estudos de ligamento, o que permitía levar a cabo varias meta-análises (Badner and Gershon 2002; Lewis et al. 2003; Ng et al. 2009). Na meta-análise feita por Badner e Gershon atopaban loci que apoiaban a existencia de xenes de susceptibilidade que estaban situados nos cromosomas 8p, 13q e 22q (Badner and Gershon 2002). No caso do traballo feito por Lewis e cols. atopaban varias rexións ligadas : 2p-q, 5q, 3p, 11q, 6p, 1p-q, 22q11–12, 8p 20p, e 14pter-q132q (Lewis et al. 2003). Curiosamente só dúas rexións eran coincidentes nestes traballos. Na meta-análise máis recente utilizaban 32 estudos de ligamento contabilizando máis de sete mil casos, no cal amosaban varias rexións ligadas á esquizofrenia, sendo a que está situada na rexión 2q (119-152 Mb) a que maior evidencia demostraba, seguida de 5q (142-168 Mb) e 2q (103-134 Mb) (Ng et al. 2009). Este estudo amosaba máis ligamentos nominalmente significativos dos que se esperaría por azar. A causa do exceso de sinais de ligamento podería ser debido a existencia de rexións xenómicas que estiveran altamente enriquecidas por variantes comúns de efecto moderado ou de variantes cun efecto moito maior pero pouco frecuentes (Williams et al. 2009).

Tanto os estudos de ligamento como as súas meta-análises identifican rexións amplas (que poden abarcar gran cantidade de xenes) compartidas entre individuos afectos, as cales poden ser posibles rexións a estudar máis detalladamente en estudos posteriores. De feito, o mapeo máis fino destas rexións permitía a identificación de xenes candidato posicionais. Para algúns destes xenes, como se comenta no seguinte apartado, atopábanse resultados positivos mediante os estudos de asociación.

1.2.2.2. Os Estudos de Asociación

Os estudos de asociación permiten, a diferenza dos estudos de ligamento, identificar rexións moito máis pequenas e que abarcan un único xene ou uns poucos. Con este tipo de estudos o que se pretende é comparar as frecuencias de marcadores entre individuos afectos e controis non emparentados. O seu obxectivo é detectar alelos que son máis comúns (ou menos) nos casos que na poboación xeral, e están fundamentados na hipótese de enfermidade común-variante común anteriormente comentada. A potencia estatística que presentan os estudos de asociación é moito maior que os estudos de ligamento cando o alelo asociado é común e ademais o risco que confire é baixo (Risch and Merikangas 1996), polo que é a metodoloxía apropiada para estudar enfermidades comúns como a esquizofrenia. Outra vantaxe que presentan é que a facilidade para conseguir mostras é bastante maior respecto ós estudos de ligamento.

Introdución

A principal limitación desta estratexia é que é ineficaz fronte a heteroxeneidade alélica, é dicir, a existencia de varios alelos presentes no mesmo xene asociados á patoloxía.

Posto que a recombinación xenética ó longo das xeracións rompe as asociacións existentes entre marcadores contiguos, para os estudos de asociación é necesario utilizar marcadores xenéticos moi abundantes ó longo do xenoma. Esta característica cumpren os SNPs. Estas variantes ocorren, aproximadamente, cada 300 pares de bases ó longo do xenoma. Ademais existe unha gran cantidade de SNPs que son comúns dentro dunha poboación, é dicir, máis do 5% da poboación é portadora dunha variación específica nun locus determinado. Todo isto fai que estas variacións sexan fáciles de estudar mediante os estudos de asociación.

No xenoma humano existen distintos tipos de SNPs. Un deles son os SNPs non sinónimos que provocan un cambio na secuencia aminoacídica da proteína codificada polo xene, podendo así, afectar a actividade desta e a súa vez a susceptibilidade a padecer unha determinada patoloxía. Este tipo de SNPs ou ben poden dar lugar a substitucións dun aminoácido por outro na proteína codificada podendo afectar a actividade da proteína, sobre todo se o cambio ten lugar entre dous aminoácidos de propiedades fisicoquímicas moi distintas; ou ben poden dar lugar a un codón de parada que provoca o remate prematuro da tradución dando lugar a unha proteína truncada. Variantes deste estilo son as que se estudan no **estudos de asociación directos** (Hirschhorn and Daly 2005). Nestes estudos o que se fai é comprobar se existe unha distribución distinta entre casos e controis dunha variante funcional dun xene que estivese implicada na variación na susceptibilidade. Outro tipo de variantes que tamén poden ser obxecto de estudo mediante esta estratexia, pero de forma pouco habitual, son os SNPs intrónicos ou os sinónimos (substitucións na rexión codificante pero que non provocan cambios aminoacídicos). Os SNPs intrónicos poden afectar á expresión do xene por estar situadas en rexións promotoras de unión de factores de transcrición, ó procesamento do ARN mensaxeiro afectando ó corte de intróns e empalme de exóns ou a súa tradución, etc. Outro tipo de SNPs que hai que considerar son os que están sitos en rexións non codificantes, xa que estas variantes poden afectar a lugares de unión de factores de transcrición en promotores, zonas altamente conservadas ou rexións con modificacións epixenéticas, as cales tamén poden ser funcionais (Roussos et al. 2014).

O principal problema da estratexia directa é a necesidade de coñecemento previo das variantes funcionais dos xenes a estudar. Isto resólvese usando o **método indirecto**, que asume que non se poden coñecer todas estas variantes funcionais previamente, polo que trata de identificar unha asociación entre un xene e unha patoloxía vendo se existen diferenzas significativas entre casos e controis para os SNPs marcadores, os cales poden ser os causantes directos da variación na susceptibilidade ou pode que se encontren en desequilibrio de ligamento (DL) cos causais (Hirschhorn and Daly 2005). Defínese o DL como a correlación entre dúas variantes próximas de tal xeito que, os alelos destes marcadores veciños (observados no mesmo cromosoma) están asociados nunha poboación con máis frecuencia que a esperada en función das súas frecuencias individuais (Hirschhorn and Daly 2005). O DL xurde coa aparición dunha nova mutación que, mentres non se produza recombinación, estará en DL co resto das variantes do cromosoma, formando así un novo haplotipo (combinación de variantes alélicas dun cromosoma). O DL pode variar dentro dunha mesma poboación ou entre poboacións como consecuencia da variabilidade rexional nos patróns de recombinación, deriva xenética, selección natural, conversión xénica, idade da mutación, diversidade étnica e mestura recente da poboación, taxa de mutación ou ós patróns de emparellamento na poboación (Carlson et al. 2004).

Para medir o DL entre dous marcadores veciños existen varias medidas (Weiss and Clark 2002). Unha delas é o parámetro D, que mide a diferenza entre as frecuencias haplotípicas observadas e esperadas a partir das frecuencias individuais de cada alelo. Calcúlase do seguinte xeito:

$$D = \text{frec}(A \cap B) - \text{frec}(A) \cdot \text{frec}(B)$$

Onde A e B son os alelos de maior frecuencia de dous SNPs e a e b os menores. O problema deste parámetro é que depende da frecuencia dos distintos alelos, non sendo útil para comparar o DL entre rexións distintas.

O problema anterior soluciónase utilizando D' que se calcula axustando o valor de D en función do máximo ou o mínimo valor que pode alcanzar ás frecuencias alélicas concretas (Dmax):

$$D' = D/D_{\text{max}}$$

Outra forma de calcular o DL é r^2 :

$$r^2 = D^2 / [\text{frec}(A) \cdot \text{frec}(a) \cdot \text{frec}(B) \cdot \text{frec}(b)]$$

Ó igual que D', r^2 toma valores comprendidos 0 e 1. A diferenza é que D' acada o valor de 1 se o alelo de menor frecuencia dun SNP está sempre asociado a un dos dous alelos do outro SNP, mentres que r^2 só será 1 se cada un dos alelos dun SNP está asociado a un único alelo do outro SNP, é dicir, as frecuencias do alelo menor dos dous SNPs son idénticas e ademais atópanse no mesmo haplotipo. Polo tanto, r^2 é unha medida máis restritiva que D'.

O efecto que causa cada un dos SNPs asociados é medido mediante o *odds ratio* (OR) que é a razón de dous cocientes: número de pacientes portadores de alelo de risco (a) entre o número de afectados non portadores (c); e o número de controis portadores do alelo de risco (b) entre o número de controis non portadores (d) ($OR = (a/c)/(b/d)$). Se o valor de *odds ratio* é menor que 1 indica que o alelo é de protección.

A potencia estatística dun estudo indica a probabilidade de que a hipótese nula sexa rexeitada cando a hipótese alternativa é verdadeira, é dicir, a probabilidade de non cometer un erro tipo II (aceptar a hipótese nula de non asociación sendo falsa). O cálculo da potencia estatística dun estudo permite ver se o deseño do estudo é o adecuado para poder atopar unha determinada asociación. Gran parte dos erros tipo II son debidos a que o tamaño mostral é insuficiente; aumentar o número de mostras nun estudo é a forma máis común de aumentar a potencia e, como consecuencia, reducir este tipo de erros. Outros parámetros que tamén se utilizan para a estima da potencia son a frecuencia e a magnitude do efecto atribuído a un alelo, o cal é medido polo OR, comentado anteriormente. O nivel de significación (α) tamén é necesario para o cálculo da potencia. Este parámetro indica o valor por debaixo do cal se considera que as diferenzas observadas non son debidas ó azar. O nivel de significación dun SNP ven dado polo valor-P que indica a probabilidade de obter un resultado polo menos tan extremo como o que se obtivo supoñendo que a hipótese nula fose certa. Hai que ter en conta que este umbral de significación vólvese máis restritivo cando no estudo se levan a cabo múltiples comparacións. O método máis habitual para corrixir isto é o método de Bonferroni, no que o nivel de significación calcúlase dividindo o valor α entre o número de probas. Este método asume que as distintas probas son independentes, cousa que non pasa case nunca nos estudos de asociación, pola presenza do DL entre SNPs, causando así erros de tipo II. Unha alternativa é o uso de permutacións. Mediante este método permútanse as etiquetas de casos e controis na mostra, xerando así pseudocasos e

Introdución

pseudocontros, facéndose a continuación o contraste de frecuencias entre ambos pseudogrupos, e así para cada unha das réplicas, obténdose unha distribución empírica do estatístico ó chou para logo comparala co valor real.

1.2.2.2.1. Os Estudos de Asociación de Xenos Candidato

Posto que, a diferenza dos estudos de ligamento, as asociacións non se estenden ó longo de grandes segmentos xenómicos, para os estudos de asociación é necesario estudar centos de miles de marcadores. Isto, ata o desenvolvemento das novas tecnoloxías de xenotipado non era posible levalo a cabo, polo que os primeiros estudos de asociación centrábanse en analizar xenos candidato, sendo así o número de marcadores a estudar moito menor. Nun principio a selección destes xenos facíase en función das súas características funcionais (**xenos candidato funcionais**), é dicir, xenos que, segundo coñecementos moleculares previos, poderían estar implicados en rutas metabólicas relacionadas coa enfermidade, levar a cabo unha función biolóxica clave na patoloxía, expresarse en rexións implicadas no trastorno, etc.; dito doutro xeito, as proteínas que codificaban estes xenos poderían desenvolver un papel importante na patofisioloxía da enfermidade. No caso da esquizofrenia existe o problema de que se coñece pouco sobre a súa patoxénese ou sobre a súa etioloxía, polo que a selección dos xenos era difícil (Williams et al. 2009). Existen numerosas publicacións con este tipo de estudos pero ningunha delas conseguía resultados consistentes (Farrell et al. 2015).

A alternativa a estes estudos era analizar xenos situados en rexións que tiñan unha elevada probabilidade de estar relacionados na etioloxía da enfermidade, como poden ser xenos situados en rexións cromosómicas previamente asociadas con estudos de ligamento ou xenos que están alterados por aberracións cromosómicas implicadas na patoloxía. Son os chamados **xenos candidato posicionais**. Estes estudos, a pesares da non confirmación nas réplicas das asociacións atopadas cos estudos de ligamento, atopaban varios xenos asociados coa esquizofrenia, entre os que destacan *Neuregulin 1 (NRG1)* (Stefansson et al. 2002), *Dystrobrevin-binding protein 1 (DTNBPI)* (Straub et al. 2002) e *D-amino acid oxidase activator protein (DAOA)* (Detera-Wadleigh and McMahon 2006).

A pesares do descubrimento dalgunhas asociacións positivas, ningún dos xenos candidato estudados acadaba un nivel de significación convincente para asocialos a esquizofrenia. A falta de asociacións positivas neste tipo de estudos pode ser consecuencia do uso de tamaños de mostras moi pequenos, polo que os traballos non tiñan a suficiente potencia estatística para detectar efectos de pequeno tamaño, algo típico dos alelos comúns de enfermidades complexas (Williams et al. 2009).

No 2008 publicábase un artigo que facía referencia a unha base de datos pública e de acceso libre chamada “SchizophreniaGene” (<http://www.szgene.org/>). Aquí recollíanse tódolos estudos de asociación xenéticos sobre a esquizofrenia publicados, incluíndo o número de mostras analizadas, orixe e datos xenotípicos. Ademais, levaban a cabo e incluían os resultados das meta-análises de tódolos polimorfismos dos que se dispoñían de, polo menos, datos de catro mostras caso-control independentes. Atopaban 24 variantes que alcanzaban un valor significativo nos xenos: *APOE*, *COMT*, *DAO*, *DRD1*, *DRD2*, *DRD4*, *DTNBPI*, *GABRB2*, *GRIN2B*, *HP*, *IL1B*, *MTHFR*, *PLXNA2*, *SLC6A4*, *TP53* e *TPH*. Estes xenos tiñan un OR medio de 1,23. Sen embargo, se facían unha selección seguindo uns criterios propostos (Ioannidis et al. 2008) para avaliar a evidencia acumulativa dos estudos de asociación, só 4 destes xenos (*DRD1*, *DTNBPI*, *MTHFR* e *TPH*) amosaban unha elevada credibilidade epidemiolóxica (Allen et al. 2008).

1.2.2.2.2. Os Estudos de Asociación do Xenoma Completo

Posteriormente xurdían os estudos de asociación do xenoma completo (en inglés *Genome Wide Association Studies*, GWAS), nos cales se xenotipaban unha gran cantidade de SNPs situados ó longo do xenoma para avaliar a variación xenética común e cun efecto asociado baixo, coa finalidade de comprobar se algunha destas variantes actuaba como factor de risco dunha enfermidade. Os GWAS están baseados na hipótese enfermidade común-variante común que, como se comentou anteriormente, sostén que as variantes xenéticas comúns, presentes na poboación xeral, causan un risco moi baixo de xeito individual a desenvolver unha determinada enfermidade. Estes son posibles realizalos grazas ós avances nas tecnoloxías de xenotipado, ó coñecemento da secuencia completa do xenoma humano, ó rexistro de millóns de SNPs en base de datos públicas e ó proxecto internacional HapMap (International HapMap 2003). Este último permítenos coñecer os patróns de DL ó longo do xenoma e, por tanto, caracterizar a diversidade haplotípica existente nas distintas poboacións humanas (Hirschhorn and Daly 2005). Os GWAS tiveron gran aceptación debido a que xuntan as vantaxes dos estudos de asociación (permiten detectar pequenos efectos) e os de ligamento (non requiren un coñecemento específico da patoxénese) (Williams et al. 2009). Por outra banda, estes estudos necesitan un gran número de participantes para conseguir unha potencia estatística suficiente para atopar variantes con risco relativo baixo e cunha significación a nivel xenómico. De feito, o valor-P proposto para estes estudos é de 5×10^{-8} , resultante de dividir un valor-P de 0,05 por un millón de test independentes segundo a corrección de Bonferroni (Risch and Merikangas 1996).

A necesidade de grandes tamaños de mostras para alcanzar suficiente potencia estatística para detectar alelos comúns de pouco efecto mediante os GWAS (Táboa 1, páxina 21) fixo que se crearan diferentes consorcios, entre os que se atopan o *International Schizophrenia Consortium* (ISC), *Molecular Genetics of Schizophrenia* (MGS), SGENE e o *Schizophrenia Working Group of the Psychiatric Genomics Consortium* (SWG of PGC). A parte de permitir tamaños de mostras grandes, algúns destes consorcios permiten o acceso público dos datos resumo obtidos, co cal poden ser utilizados en posteriores análises. Un dos problemas asociados a estes consorcios é a diferente orixe das mostras recompiladas, facendo que un dos puntos críticos destes estudos sexa a estratificación poboacional. Esta fai referencia a posibilidade de obter resultados espurios debido as diferenzas ancestrais alélicas entre as sub-poboacións do estudo, no lugar de estar asociados ó trazo estudado (Bergen and Petryshen 2012). Para evitar este problema deberíase ter especial atención no emparellamento entre os casos e os controis segundo á súa orixe xeográfica. Aparte disto, hai métodos que permiten detectar e corrixir o problema da estratificación poboacional a partires dos datos de xenotipado obtidos. As dúas primeiras estratexias utilizadas eran o control xenómico (Devlin et al. 2001) e a asociación estruturada (Pritchard and Donnelly 2001). O control xenómico calcula o valor polo cal se incrementa o estatístico empregado para identificar a asociación como consecuencia da estratificación; mentres que a asociación estruturada intenta identificar cantas sub-poboacións existen na mostra, de xeito que asigna cada un dos individuos de dita mostra a unha sub-poboación antes de contrastar as frecuencias dos marcadores. Outros dous métodos que existen actualmente para avaliar a estratificación poboacional é a Análise de Componentes Principais (das súas siglas en inglés, PCA) e o Escalado Multidimensional (das súas siglas en inglés, MDS). A PCA identifica os componentes principais e úsaos como covariables na análise de asociación. Esta metodoloxía podería non axustar adecuadamente para estratificación poboacional se esta é debida á presenza de varias sub-poboacións discretas, xa que a PCA usa os

Introdución

vectores propios identificados como covariables continuas. Ademais, se hai *outliers*, os resultados baseados neste tipo de axuste poderían ser erróneos (Price et al. 2006). Por outro lado o MDS axusta para estratificación poboacional debido tanto a estrutura poboacional discreta como continua; ademais de funcionar ben tanto para GWAS de gran tamaño como para estudos de pequeno tamaño. Presenta o mesmo problema que o PCA cando hai presenza de *outliers* (Li and Yu 2008).

Os primeiros GWAS publicados sobre a esquizofrenia (Táboa 1) estaban baseados no xenotipado de *pools* de mostras de ADN (Mah et al. 2006; Shifman et al. 2008; Kirov et al. 2009c). Mah e cols. xenotipaban inicialmente 320 esquizofrénicos e 325 controis, nos que atopaban o SNP rs752016 cun $P=0,006$ situado no xene *PLXNA2*. Este resultado era confirmado nunha mostra posterior que comprendía 531 casos de diferente orixe e 605 controis. Este xene pertence á familia dos receptores transmembrana semaforin que está implicado na guía axonal durante o desenvolvemento podendo modular a plasticidade neuronal. Por outra banda, Shifman e cols. atopaban evidencias bastante fortes dunha asociación específica entre *RELN* e mulleres con esquizofrenia nunha poboación Asquenazi xudía ($P=2.9 \times 10^{-5}$), despois de analizar unha mostra inicial de 660 casos e 2271 controis e posteriores réplicas nun total de 2274 casos e 4401 controis. Este xene codifica unha proteína que é unha serín proteasa que actúa mediante un número de rutas metabólicas mediadas por receptores en neuronas ademais de desenvolver un papel clave na corticoxénese (Quattrocchi et al. 2002). Mutacións neste xene eran asociadas cunha forma autosómica recesiva de lisencefalia (trastorno pouco común da formación do cerebro que se caracteriza por microcefalia e axiria, que é a ausencia das dobreces normais dos cerebro). Por último, Kirov e cols. xenotipaban unha *pool* de 605 controis, unha segunda *pool* de 574 casos de esquizofrenia e unha terceira que comprendía os pais dos afectos. Unha vez analizados, o SNP rs1106478, situado no xene *CCDC60*, acadaba o valor máis significativo ($P=1,2 \times 10^{-6}$). Outro dos xenes que destacaban era o *RBP1*, que contiña o SNP (rs893703) co terceiro valor-P máis significativo ($P=0,00016$), o cal fora previamente implicado na patoxénese da esquizofrenia. Destacar que nestes tres estudos non se alcanzaban niveis significativos de asociación a nivel xenómico coa esquizofrenia aínda que si suxerían algunha evidencia bastante forte.

O primeiro estudo publicado sobre esquizofrenia (Táboa 1) no que se analizaban mostras individuais era o levado a cabo por Lencz e cols. Nel xenotipaban unha mostra composta por 178 afectos e 144 controis no que atopaban un SNP rs4129148 que alcanzaba un $P=3,7 \times 10^{-7}$ e que estaba situado nunha rexión próxima ó xene *CSF2RA* (Lencz et al. 2007). Ó ano seguinte, O'Donovan e cols. identificaban un sinal de asociación do SNP (rs1344706) sito no xene *ZNF804A* nunha mostra inicial de 479 esquizofrénicos e 2937 controis e nunha mostra réplica total de 6829 esquizofrénicos e 9897 controis. Unha vez que combinaban esta mostra con mostras de pacientes con trastorno bipolar a asociación aumentaba en dúas ordes de magnitude, acadando un nivel de significación xenómico ($P=9,96 \times 10^{-9}$, OR=1,12). Este xene está localizado no cromosoma 2 e codifica unha proteína non caracterizada e sen función coñecida pero que contén dominios de unión ó ADN e a ións de zinc segundo predicións *in silico*, o cal suxire que pode ter un posible papel como regulador da expresión xénica (O'Donovan et al. 2008).

Dende os primeiros estudos de asociación de xenoma completo ata a actualidade identificábanse gran cantidade de loci asociados á esquizofrenia, sendo o máis consistente de todos eles a rexión que abarca o Complexo Principal de Histocompatibilidade (en inglés *Major Histocompatibility Complex*, MHC). Esta rexión

asociábase coa esquizofrenia en tres artigos distintos publicados ó longo 2009 ademais da meta-análise feita destes tres traballos (International Schizophrenia et al. 2009; Shi et al. 2009; Stefansson et al. 2009). O locus está sito no cromosoma 6 e abarca 6 millóns de pares de bases o que fai que, xunto co elevado desequilibrio de ligamento e a súa complexidade xenómica estrutural, sexa difícil determinar si un ou múltiples xenes están involucrados no trastorno. Aínda así, recentemente publicábase un artigo no que indicaban que parte da asociación do MHC coa esquizofrenia era debido a xenes que forman parte do compoñente 4 do complemento (C4) (Sekar et al. 2016). As principais hipóteses que apoian a relación do MHC coa esquizofrenia son as evidencias que amosan que existe relación entre a exposición a infeccións na época prenatal ou nos primeiros anos de vida (Brown 2006) e o feito de atopar unha elevada taxa de enfermidades autoinmunes e inflamatorias en individuos afectados (Eaton et al. 2006). Entre os xenes en DL situados neste locus, existen varios que non están relacionados coa inmunidade. De feito, hai xenes que desempeñan un papel na regulación (*KDM5C*) e reparación do ADN (*DNMT1*), no mantemento da estrutura da proteína (*HMGN4*), na regulación da transcrición (*ABT1*, *ZNF322A* e *ZNF184*), na sinalización do receptor acoplado a proteína G (*FKSG83*) e no complexo poro nuclear (*POM121L2*). Ademais existen moitos outros nos que a súa función no é ben coñecida (Shi et al. 2009).

O Consorcio Internacional de Esquizofrenia chegaba a obter valores significativos no MHC despois de xenotipar, na fase inicial, 3322 individuos europeos con esquizofrenia e 3587 controis e na fase de réplica 4692 casos e 15493 controis. Neste artigo tamén avaliaban a contribución dos SNPs en masa no risco a padecer a esquizofrenia. Cos resultados do GWAS obtidos na mostra descubrimento, xeraban un modelo polixénico mediante a asignación dun valor de significación e un efecto do alelo de risco (OR) a cada un dos SNPs. A continuación calculaban o risco polixénico de cada un dos individuos nunha mostra diana, sumando o número de alelos de risco e ponderando polo seu efecto (como o logaritmo da OR) para cada un dos SNPs que alcanzaba un valor de significación por debaixo dun umbral de significación dado. Deste xeito detectaban a existencia de diferenzas significativas entre os casos e os controis, que acadaban valores máis significativos se consideraban umbrais de significación menos estritos na mostra descubrimento. Ademais amosaban que a compoñente polixénica tamén contribúe no risco a padecer trastorno bipolar pero non nas distintas enfermidades non psiquiátricas (enfermidade coronaria, enfermidade de Crohn, hipertensión, artrite reumatoide, diabete tipo I e tipo II) (International Schizophrenia et al. 2009).

No traballo de Stefansson e cols., no que xenotipaban unha mostra inicial de 2,663 esquizofrénicos e 13,498 controis e unha réplica posterior de 10,282 casos e 21,093 controis. Destacar que parte das mostras que utilizabamos para levar a cabo o traballo de tese que posteriormente describirei, estaban incluídas neste estudo e no traballo levado a cabo por Steinberg e cols. (Steinberg et al. 2011a). A parte de atopar a asociación co MHC, identificaban dous SNPs con valores significativos a nivel xenómico. Un deles era o rs12807809, que a cada un $P=2,4 \times 10^{-9}$ e un $OR=1,15$ (Stefansson et al. 2009). Este polimorfismo estaba situado a 3457 bases augas arriba do xene *Neurogranin* (*NRGN*). Este xene exprésase exclusivamente no cerebro, concretamente nas espiñas dendríticas, sendo a expresión controlada por hormonas tiroideas. Codifica un substrato proteína quinasa postsináptico que une Calmodulina (CaM) en ausencia de calcio; exprésase de forma abundante en rexións importantes do cerebro para funcións cognitivas, e está especialmente enriquecido en neuronas piramidais CA1 no hipocampo. A principal función deste xene pode ser a de actuar

Introducción

como reservorio de CaM, regulando a súa dispoñibilidade nos compartimentos postsinápticos (Huang et al. 2007b). O outro dos SNPs (rs9960767) que alcanzaba valores significativos ($P = 4,1 \times 10^{-9}$, $OR=1,23$) está situado no intrón 4 do xene *TCF4*. Este xene é membro dunha familia hélice-bucle-hélice básica de factores de transcrición que se expresa altamente en varios tecidos adultos e no sistema nervioso central embrionario tendo un papel importante no desenvolvemento do sistema nervioso. Mutacións neste xene causan a síndrome Pitt-Hopkins, que é un trastorno do desenvolvemento autosómico dominante caracterizado por retraso motor severo e retraso mental, microcefalia, epilepsia e dismorfismo facial (Pitt and Hopkins 1978).

Shi e cols. inicialmente analizaban unha mostra de 3967 esquizofrénicos e 3626 controis. Unha vez xenotipadas tódalas mostras, ningún dos SNPs alcanzaba valores significativos, polo que levaban a cabo unha meta-análise engadindo as súas mostras, 5327 casos e 16424 controis. Deste xeito, 7 SNPs, altamente ligados, obtiñan valores significativos, estando todos eles situados no MHC (cromosoma 6p22.1) (Shi et al. 2009).

No 2011, o *Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium* publicaba un artigo onde analizaban máis de vinte e unha mil mostras na primeira fase e case trinta mil na fase de réplica e no que atopaban 10 SNPs, situados en 7 loci, que alcanzaban valores significativos a nivel xenómico. Cinco dos anteriores non foran previamente descubertos (1p21.3, 2q32.3, 8p23.2, 8q21.3 e 10q24.32-q24.33) e dous xa foran asociados previamente: o locus 6p21.32-p22.1 que corresponde ó MHC e 18q21.2, no cal está situado o xene *TCF4*. O locus que acadaba o nivel de significación máis alto ($1,59 \times 10^{-11}$) era o SNP rs1625579 situado na rexión 1p21.3. Este SNP estaba dentro dun intrón pertencente ó xene *MIR137*. Este xene expresa un microARN implicado na regulación da neuroxénese adulta e na maduración neuronal. Curiosamente outros loci que acadaban valores significativos conteñen xenes (*TCF4*, *CACNA1C*, *CSMD1* e *C10orf26*) que son preditos por ser diana do *MIR137*, suxerindo que unha mala regulación mediada por este xene podería dar lugar a esquizofrenia. Neste estudo usaban casos de distintas enfermidades para intentar aumentar a potencia estatística e tamén para identificar factores de risco xenético compartido xa que parece ser que o risco xenético é altamente pleiotrópico, é dicir, un xene ou un alelo pode afectar a múltiples trazos fenotípicos nun mesmo individuo. Analizaban conxuntamente individuos esquizofrénicos e afectados con trastorno bipolar, no que só tres xenes alcanzaban valores significativos a nivel xenómico (*CACNA1C*, *ANKK1* e *ITIH3-ITIH4*) (Schizophrenia Psychiatric Genome-Wide Association Study 2011).

Neste mesmo ano, Steinberg e cols. levaban a cabo un estudo que era a continuación do feito no 2009 por Stefansson e cols. (Steinberg et al. 2011a). Aquí testaban tódolos SNPs cun valor-P inferior a 1×10^{-4} nunha mostra de 9246 casos e 22356 controis iniciais, e aqueles SNPs que alcanzasen un valor significativo a nivel xenómico despois dunha análise conxunta, eran investigados en 1014 casos e 1144 controis. Mediante esta metodoloxía confirmaban 3 loci previamente asociados (a rexión do MHC e os xenes *NRGN* e *TCF4*). Ademais atopaban novas variantes con significación xenómica: o SNP rs2312147 ($OR = 1,09$, $P = 1,9 \times 10^{-9}$) situado augas arriba de *VRK2* e o SNP rs4309482 que está situado na rexión 18q21.2, entre *CCDC68* e *TCF4* ($OR = 1,09$, $P = 7,8 \times 10^{-9}$). Este último SNP está situado a 400 kb do SNP rs9960767, o cal era previamente asociado coa esquizofrenia por estes mesmos investigadores, pero entre ámbolos dous existe unha correlación moi débil, amosando que *TCF4* é un claro exemplo dun locus que presenta varias variantes comúns de risco. O xene *VRK2* está altamente expresado, cunha marcada expresión en células altamente

proliferativas. Propúxose que xoga un papel importante no mantemento da correcta arquitectura nuclear e que está relacionado coa prevención da apoptose (Steinberg et al. 2011a).

Outro GWAS publicado este ano era o feito por Yue e cols. Estes incluían na súa mostra inicial 746 individuos con esquizofrenia e 1599 controis, todos eles descendentes da poboación chinesa Han. Unha vez feito o xenotipado, seleccionaban 46 SNPs con valor-P menor a 1×10^{-5} e cunha MAF (frecuencia do alelo menor) $>5\%$ para replicar nunha segunda mostra de 4027 casos e 5063 controis. Destes SNPs, unicamente 6 amosaban unha asociación significativa nesta mostra, estando situados na rexión 6p21-p22.1 e 11p11.2. No locus 6p21-p22.1 identificábanse tres xenes (*NKAPL*, *ZKSCAN4* e *PGBDI*) que non teñen función coñecida. Por outro lado, na rexión 11p11.2 estaba localizado o xene *TSPAN18*, que codifica un membro dunha gran familia de tetraspaninas, que son proteínas relacionadas con diversos procesos celulares, aínda que a que codifica este xene ten unha función descoñecida (Yue et al. 2011).

Shi e cols. tamén levaban a cabo un GWAS usando unha mostra Chinesa Han que constaba de 3750 individuos con esquizofrenia e 6468 controis sans. Unha vez feito o xenotipado nesta mostra inicial, seleccionaban os SNPs cos valores máis significativos para analizalos nunha mostra réplica de 4383 casos e 4539 controis de poboación chinesa Han tamén. Aquí atopaban dous loci que alcanzaban valores significativos: 8p12 (rs16887244, $P = 1,27 \times 10^{-10}$) e 1q24.2 (rs10489202, $P = 9,50 \times 10^{-9}$) (Shi et al. 2011).

No 2012, Carrera e cols. publicaban un traballo onde analizaban 5100 SNPs non sinónimos comúns cunha frecuencia superior ó 5% distribuídos ó longo do xenoma, utilizando como mostra descubrimento 447 controis e 476 doentes esquizofrénicos de orixe galega (a maior parte dos individuos desta mostra son os que procesamos e analizamos nesta tese) (Carrera et al. 2012). Posteriormente utilizaban unha mostra réplica que consistía en 4069 casos e 15128 controis de orixe europea. Neste estudo atopaban un SNP, rs13107325, que alcanzaba valores significativos ($P = 2,7 \times 10^{-6}$, OR= 1,32) despois da corrección por Bonferroni na mostra combinada. Este SNP está localizado no xene *SLC39A8*, que pertence a unha familia de transportadores de ións metálicos. Concretamente actúa no transporte de manganeso e zinc, o cal xoga un papel importante no desenvolvemento e funcionamento de moitos órganos, podendo ser tóxico a concentracións moi elevadas. Este xene exprésase altamente no cerebro e na placenta, polo que podería ser que estivese involucrado na homeostase de metais no desenvolvemento embrionario, actuando na barreira hematoplacentaria, así como na barreira hematoencefálica. Neste estudo tamén se confirmaba a existencia do risco polixénico. Mediante validacións cruzadas identificaban unha puntuación polixénica xerada a partir de SNPs con significación individual inferior a 0,1, obtendo unha distribución significativamente distinta entre os casos e os controis.

No 2013, Ripke e cols. publicaban un estudo do xenoma completo no que inicialmente analizaban unha mostra composta de 5001 casos e 6243 controis suecos atopando 312 SNPs que alcanzaban valores significativos a nivel xenómico, e que estaban repartidos entre a rexión do MHC e o cromosoma 2. A continuación utilizaban os datos obtidos deste xenotipado inicial e os obtidos polo PGC (*Psychiatric Genomics Consortium*) no 2011 nunha mostra de 8832 casos 12067 controis, para levar a cabo unha meta-análise. Mediante esta identificaban 168 rexións xenómicas asociadas, que posteriormente intentaban replicar nunha mostra independente de 7413 esquizofrénicos e 19762 controis e 581 tríos. Finalmente vían que 22 loci alcanzaban valores significativos a nivel xenómico, 13 dos cales eran novos e un xa fora asociado ó

Introdución

trastorno bipolar previamente. Entre estes loci destacaban aqueles que afectaban ós canles de calcio, asociados anteriormente á esquizofrenia e trastorno bipolar noutros traballos. Utilizando estes datos intentaban clarificar parte da arquitectura xenética da esquizofrenia, chegando a conclusión de que unha cantidade elevada de SNPs comúns (entre 6300 e 10200) contribuían a etioloxía da esquizofrenia, explicando conxuntamente, polo menos, un 32% da varianza na susceptibilidade (Ripke et al. 2013).

Un dos estudos de asociación do xenoma completo máis recente era o que levaba a cabo o PGC. Neste estudo analizaban na fase de descubrimento 34214 esquizofrénicos, 45604 controis e 1235 tríos. Unha vez xenotipadas todas estas mostras seleccionaban un subconxunto de SNPs independentes ($r^2 < 0,1$ dentro dunha fiestra de 500 Kb) e cun $P < 1 \times 10^{-6}$, que posteriormente estudaban nunha mostra réplica formada por 1513 casos e 66236 controis. Na análise final conxunta, 128 SNPs sobrepasaban o valor de significación a nivel xenómico, os cales estaban situados en 108 loci independentes dos que 83 nunca foran asociados previamente (Schizophrenia Working Group of the Psychiatric Genomics 2014). Ademais vían que as asociacións estaban enriquecidas en xenes expresados no cerebro, así como para tecidos relacionados coa inmunidade. Resaltar que neste artigo atopaban evidencias de asociación coa transmisión dopaminérxica, concretamente o xene asociado era *DRD2*, o cal é a diana de tódolos fármacos antipsicóticos efectivos, amosando a capacidade dos GWAS de identificar posibles dianas dos fármacos. Tamén vían que existía unha asociación cos xenes *GRI1*, *GRIN2A* e *GRM3*, implicando á ruta glutaminérxica. Ademais amosaban unha asociación con xenes que codifican sub-unidades dos canles de calcio (*CACNA1C*, *CACNB2* e *CACNA1I*), aumentando as evidencias da implicación desta familia de proteínas na susceptibilidade a padecer esquizofrenia e outros trastornos psiquiátricos. Estas últimas asociacións apoiaban ás hipóteses previas que sostían que a homeostase neuronal do calcio e a alteración da vía do glutamato poderían ser vías claves no desenvolvemento da esquizofrenia. Isto amosa que o número de novos loci de susceptibilidade situados en determinadas rutas biolóxicas segue en aumento, co cal a tendencia será a analizar estes loci como un conxunto en lugar de individualmente, mediante os distintos tests de análise de conxunto de xenes (comentados no apartado 1.3). Outro dos aspectos que destacaban os autores é o solapamento que atopaban entre as variantes comúns e raras, suxerindo que estas variantes son complementarias en vez de ser antagónicas. Por último, confirmaban, dunha forma robusta, que as puntuacións dos perfís de risco calculados a partir dunha mostra descubrimento, permitían predicir o estatus de caso ou control nunha mostra independente.

Co fin de estudar a relación xenética existente entre distintos trastornos psiquiátricos, no 2013 publicábase un artigo no que facían unha análise combinada de varias enfermidades psiquiátricas (esquizofrenia, trastorno bipolar, depresión maior, trastorno do espectro autista e trastorno de hiperactividade e déficit de atención). En total, a mostra estaba formada por 33332 casos e 27888 controis (nos que estaban incluídos psuedocontrois formados polos alelos non transmitidos no caso dos tríos). Cos datos desta mostra, atopaban tres loci asociados coa totalidade dos trastornos (*ITIH3*, *AS3MT* e *CACNB2*) ademais de confirmar o descubrimento anteriormente comentado do xene *CACNA1C* en esquizofrenia e trastorno bipolar. Por outra banda, utilizando puntuacións de risco, atopaban un solapamento do risco polixénico entre os tres trastornos adultos (esquizofrenia, trastorno depresivo e trastorno bipolar). Aínda que en menor medida, tamén existía un risco xenético solapado entre autismo, esquizofrenia e trastorno bipolar. Ademais, non detectaban un solapamento polixénico significativo

entre trastorno depresivo e autismo ou entre trastorno de hiperactividade e calquera outro trastorno. Por último, facían unha análise de rutas metabólicas onde atopaban un enriquecemento en tódalas doenzas dun conxunto de xenes relacionados coa actividade de canles de calcio asociados coa catálise da difusión de ións calcio ó través dunha canle transmembrana (Cross-Disorder Group of the Psychiatric Genomics 2013).

Autor	Ano	Mostra Inicial		Mostra réplica		Xene/ Rexión	SNPs	Valor-P
		Casos	Controis	Casos	Controis			
Math ^a	2006	320	325	531	605	<i>PLXNA2</i>	rs752016	6,00E-03
Lencz	2007	178	144	NA	NA	<i>CSF2RA</i>	rs4129148	3,70E-07
Shifman ^a	2008	660	2271	2274	4401	<i>RELN</i>	rs7341475	9,00E-07
Kirov ^a	2009	574 tríos	605	NA	NA	<i>CCDC60</i>	rs11064768	1,00E-06
						<i>RBP1</i>	rs893703	1,60E-04
O'Donovan	2008	479	2,937	6829	9897	<i>ZNF804A</i>	rs1344706	2,00E-07
Stefansson	2009	2,663	13,498	10282	21093	<i>SPA17 -NRGN</i>	rs12807809	2,40E-09
						<i>TCF4</i>	rs9960767	4,00E-09
						<i>MHC (HIST1H2BJ)</i>	rs6913660	1,10E-09
						<i>MHC (PRSS16)</i>	rs13219354	1,30E-10
						<i>MHC (PRSS16)</i>	rs6932590	1,40E-12
						<i>MHC (PGBD1)</i>	rs13211507	8,30E-11
						<i>MHC (NOTCH4)</i>	rs3131296	2,30E-10
ISC	2009	3,322	3587	4692	15493	<i>MHC</i>	rs13194053	1,00E-08
Shi	2009	3967	3626	5327	16424	<i>MHC</i>	rs13194053	1,00E-08
PGC	2011	9394	12462	8442	21397	<i>TCF4</i>	rs17512836	1,00E-06
						<i>MIR137HG</i>	rs1625579	2,00E-11
						<i>CSMD1</i>	rs10503253	2,00E-08
Steinberg	2011	9246	22356	1014	1144	<i>TCF4</i>	rs4309482	7,80E-09
						<i>VRK2</i>	rs2312147	1,90E-09
Yue	2011	746	1599	4027	5063	<i>NKAPL</i>	rs1635	6,91E-12
						<i>ZKSCAN4</i>	rs1233710	4,76E-11
						<i>PGBD1</i>	rs2142731	5,14E-10
						<i>TSPAN18</i>	rs11038167	1,09E-11
Shi	2011	3750	6468	4383	4539	<i>LSM1</i>	rs16887244	1,27E-10
						<i>BRP44</i>	rs10489202	9,50E-09
Carrera	2012	447	476	4069	15128	<i>SLC39A8</i>	rs13107325	2,70E-06
Ripke ^b	2013	13833	18310	7413	20343	<i>MHC</i>	rs114002140	9,00E-14
						<i>DPYD, MIR137</i>	rs1198588	2,00E-12
						<i>CACNA1C</i>	rs1006737	5,00E-12
						<i>CACNB2</i>	rs17691888	1,00E-10
PGC ^a	2014	34214 1235 tríos	45604	1513	66236	<i>CACNA1C</i>	rs2007044	3,00E-18
						<i>CACNB2</i>	rs7893279	2,00E-12
						<i>DRD2</i>	rs2514218	3,00E-11

Táboa 1. Principais resultados dos GWAS publicados sobre esquizofrenia. MHC= Complexo Principal de Histocompatibilidade.

^a Estudos nos que se analizaron *pools* de mostras.

^b SNPs situados nos loci destacados no traballo do total dos loci significativos.

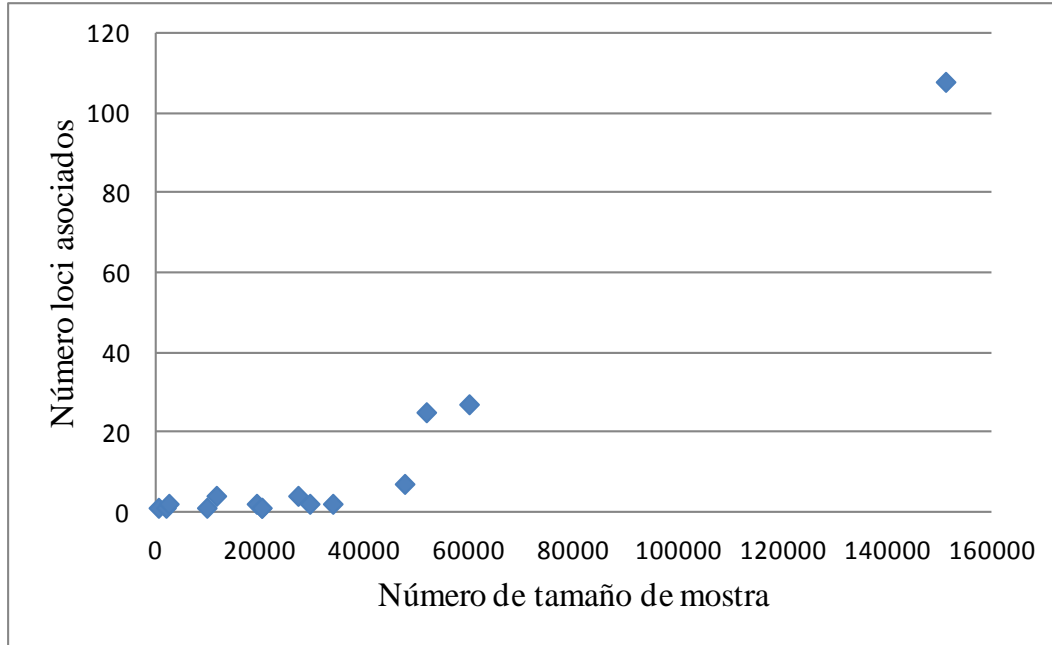


Figura 4. Relación do tamaño de mostra dos GWAS publicados en esquizofrenia co número de loci asociados. O número de mostras corresponde á suma de casos e controis das distintas fases dos traballos considerados.

Os GWAS publicados ata o momento detectaban variantes comúns con riscos relativos baixos, ORs entre 1,1 e 1,5. Este tipo de estudos permitían confirmar a existencia dunha compoñente polixénica (Ripke et al. 2013; Loh et al. 2015) previamente suxerida (Gottesman and Shields 1967), chegando a estimar que existen entre 6300 e 10200 SNPs contribuíndo ó risco de padecer a esquizofrenia e explican, en conxunto, polo menos un terzo da varianza en susceptibilidade, é dicir, case a metade da herdabilidade (International Schizophrenia et al. 2009; Lee et al. 2012c; Ripke et al. 2013); sendo ademais, a contribución destas variantes comúns na herdabilidade dunha orde de magnitude superior que as variantes raras (Purcell et al. 2014; Szatkiewicz et al. 2014). Tamén estimaban que máis do 71% de rexións xenómicas de 1 Mb de tamaño contiñan unha ou máis dunha variante que confire risco a padecer esquizofrenia, observando tamén un enriquecemento significativo da herdabilidade en rexións ricas en G-C e en SNPs con frecuencias elevadas (Loh et al. 2015). A medida que o tamaño de mostra dos GWAS aumenta, tamén o fai a potencia estatística, permitindo atopar maior número de loci asociados (Fig. 4), de feito hai un tamaño de mostra umbral a partir do cal o número de loci asociados crece moi rapidamente, ademais de que aínda non se chegou a unha fase de estabilidade. Isto demostra que esta metodoloxía a medida que aumenta o número de mostras a analizar segue amosando a súa capacidade de aportar información útil sobre a doenza. Neste senso traballa o consorcio PGC, cuxo obxectivo será, mediante a súa terceira fase (PGC3), analizar cen mil casos (sesenta e cinco mil ó longo do 2016), o cal lles permitiría identificar, segundo a súas estimas, o dobre de loci dos atopados ata o momento. Un maior tamaño de mostra unido a unha maior contribución de mostras non europeas permitirá unha maior resolución e uns intervalos de asociación máis estreitos. Ademais axudaría a determinar que variantes teñen maior probabilidade de ser relevantes funcionalmente. Isto facilitaría cumprir co obxectivo principal desta fase, é dicir, identificar a información biolóxica, terapéutica e clínica a partir da crecente fonte de información xenética dispoñible (Corvin and Sullivan 2016).

1.2.2.3. Os Estudos de Análises de Conxuntos de Xenes

Un método que pode facilitar a detección da variación polixénica común que contribúe á esquizofrenia son as análises que se centran en conxuntos de xenes relacionados funcionalmente, xa que as enfermidades complexas poderían ser o resultado da acumulación do efecto de variantes dentro dun determinado conxunto de xenes. Este método permite obter información adicional dos resultados dos GWAS. Isto é porque a maioría dos GWAS céntranse só nos 20-50 SNPs máis significativos e os seus xenes máis próximos pero presta pouca atención ó resto, representando os primeiros só unha pequena fracción das variantes que contribúen a enfermidade complexa, quedando o resto da variación escondida entre o ruído estatístico. Este tipo de estudos permiten integrar os resultados dun GWAS e os xenes dunha ruta biolóxica coñecida. Deste xeito comprobamos se a ruta está asociada cun determinado trastorno, polo que poderían complementar os métodos de SNPs máis significativos para interpretar datos de GWAS en enfermidades complexas permitindo así, diferenciar os sinais reais dos loci que causan ruído (Wang et al. 2007). Esta metodoloxía permite traducir os sinais dos GWAS a un nivel de comprensión bioquímica e/ou biolóxica e pode proporcionar éxito nas réplicas en presenza de heteroxeneidade alélica ou de locus (Cantor et al. 2010). O principal inconveniente deste método é que depende dunha ruta biolóxica previamente coñecida e trátase como unha única unidade, non podendo detectar pequenas porcións da ruta ou outra nova combinación dos xenes os cales poderían estar asociados á enfermidade.

Para a análise destas rutas biolóxicas existen numerosas metodoloxías que posteriormente tratarei máis ó fondo no apartado 1.3. Todos estes procedementos teñen os súas vantaxes e os seus inconvenientes, polo que non existe un consenso sobre cal deles é o idóneo.

Nos estudos publicados ata o momento atopábanse varias rutas metabólicas asociadas á esquizofrenia: rutas relacionadas co metabolismo do glutamato, procesos de apoptose, inflamación e co sistema inmune (Jia et al. 2010), rutas relacionadas co sistema inmune e moléculas de adhesión celular (esta ruta era a única replicada) (O'Dushlaine et al. 2011; Jia et al. 2012), adhesión focal, guía axónica, rutas de sinalización de calcio, interacción neuroactiva ligando receptor (Weng et al. 2011), rutas relacionadas coa actividade das canles específicas de substratos (Zhang et al. 2011) ou procesos de biosíntese de estrógenos (Lee et al. 2013) (Táboa 2).

Recentemente publicábase un estudo de análise de rutas metabólicas no que trataban os datos resumo dos estudos do PGC da fase I que abarcaba máis de sesenta mil individuos afectados de distintas doenzas psiquiátricas (esquizofrenia, trastorno depresivo, trastorno bipolar, trastorno do espectro autista e trastorno de déficit en atención e hiperactividade). Neste estudo utilizaban distintas bases de datos para a recompilación das rutas metabólicas a estudar (GO, KEGG, *Panther*, *Reactome* e *Target Scan*). A continuación, utilizando os datos resumo dos GWAS e mediante cinco métodos de análises de rutas metabólicas (FORGE, INRICH, SET-SCR, MAGENTA e ALIGATOR) calculaban un valor-P para cada ruta metabólica, para posteriormente clasificar en función do valor medio dos 5 métodos. Isto facíanlo para cada unha das doenzas. Por último, comparaban estes valores cos obtidos mediante simulación. Os resultados amosaban un enriquecemento significativo nominal da ruta metabólica involucrada na activade das canles de calcio con tódalas doenzas, descrita no traballo do *Cross-Disorder Group of the Psychiatric Genomics* previamente xa comentado. Tamén demostraban que os procesos relacionados coa sinapse e a post-sinapse pode que estean

Introdución

relacionados coa etioloxía da esquizofrenia. Por outra banda, acadaban unha asociación significativa entre a esquizofrenia, trastorno bipolar e trastorno depresivo e os procesos de metilación de histonas, rutas metabólicas relacionadas coa sinalización neuronal e sistema inmune (Network and Pathway Analysis Subgroup of Psychiatric Genomics 2015).

Autor	Ano	Mostra Total		Tests Utilizados	Definición Rutas metabólicas	Ruta Metabólica Asociada
		Casos	Controis			
Jia	2010	1351 SCZ	1378	GSEA	MSigDB	-Metabolismo Glutamato
				Test Hiperxeométrico		-Sinalización TGF-beta
				Método de Fisher		-Sinalización TNFR1 -Ruta metabólica TOB1
Jia	2012	3129 SCZ	3587	gamGWAS	KEGG	-Lupus eritematoso sistémico
		1351 SCZ	1378			-Rexeito de alotransplantes -Enfermidade transplante vs hóspede -Inmunodeficiencia primaria -Ruta metabólica CAMs
O'Dushlaine	2011	3322 SCZ	3587	SNP ratio test	KEGG	-Ruta metabólica CAMs
		1351 SCZ	1378			
Weng	2011	1351 SCZ	1378	SNP Set Enrichment Analysis	KEGG	-Adhesión focal
		921 SCZ	954			-Guía axónica -Rutas de sinalización de calcio -Interacción neuroactiva ligando receptor
Zhang	2011	1351 SCZ	1378	i-GSEA	BioCarta	-Actividade dos canles específicos de substratos
		1195 SCZ	954		KEGG	
		900 SCZ	877		GO	
		1001 TB	1034			
		362 TB	671			
Lee	2013	1351SCZ	1378	ICSNPPathway	BioCarta	-Procesos de biosíntese de estróxenos
					KEGG	
					GO	
					MSigDB	
NPAS of PGC ^a	2015	9379 SCZ	7736	FORGE	KEGG	-Procesos relacionados coa sinapse e post-sinapse
		9227 TD	7383	SET SCREEN TEST	GO	-Procesos de metilación de histonas
		6990 TB	4820	INRICH	Panther	-Sinalización neuronal
		2787 THDA	2635	ALIGATOR	Reactome	-Sistema inmune
		4949 ASD		MAGENTA	TargetScan	
			OMIM			

Táboa 2. Resultados dos estudos de análises de rutas metabólicas en esquizofrenia. SCZ=Esquizofrenia; TEA=Trastorno do Espectro Autista; TD= Trastorno depresivo; THDA=Trastorno de Hiperactividade e Déficit de Atención

^aThe Network and Pathway Analysis Subgroup of the Psychiatric Genomics Consortium.

1.2.2.4. Os Estudos de Variantes Raras

Como xa se comentou anteriormente, os GWAS feitos ata o momento permitían concluir que cada un dos SNPs identificados contribúen dun xeito moi pequeno a aumentar risco a padecer esquizofrenia, aínda que colectivamente explican case o 50% da varianza da susceptibilidade a padecer a doenza. Isto suxire que a variación xenética común explica a maior parte da herdabilidade da esquizofrenia, pero queda aínda unha proporción sen explicar. Esta última defínese como a herdabilidade perdida (Manolio et al. 2009). En xenética, defínese a proporción da herdabilidade dun trazo explicada por un conxunto de variantes xenéticas coñecidas como: $\pi_{\text{explicada}} = h^2_{\text{coñecido}} / h^2_{\text{total}}$; onde o numerador é a proporción de varianza explicada mediante o efecto aditivo das variantes coñecidas e o denominador é a proporción da varianza fenotípica atribuíble ó efecto aditivo de tódalas variantes, incluíndo aquelas que aínda non se descubriron. O numerador pode ser calculado directamente dos efectos medidos das variantes, mentres que o denominador debe ser inferido a partir de datos poboacionais (Zuk et al. 2012). A explicación máis xeneralizada para a herdabilidade perdida radica no numerador desta ecuación, é dicir, nas variantes adicionais que aínda non foron descubertas. A herdabilidade total (Fig. 5), é dicir, a proporción da varianza fenotípica atribuíble ós factores xenéticos aditivos, pode ser dividida en dúas partes: A herdabilidade debido a variantes coñecidas, ou o que é o mesmo, a varianza atribuída a variantes que alcanzan valores significativos mediante estudos de GWAS; e a herdabilidade perdida, que a súa vez está subdividida en: herdabilidade escondida e herdabilidade aínda perdida. A primeira fai referencia a varianza que podería ser capturada, nun último caso, mediante as variantes analizadas nun GWAS e que descendería a medida que o tamaño mostral aumenta; a segunda é a herdabilidade que non é capturada polos GWAS, é dicir, non descenderá co aumento do tamaño mostral, pero si o fará a medida que a varianza xenética sexa capturada, como por exemplo, mediante as variantes raras (Witte et al. 2014).

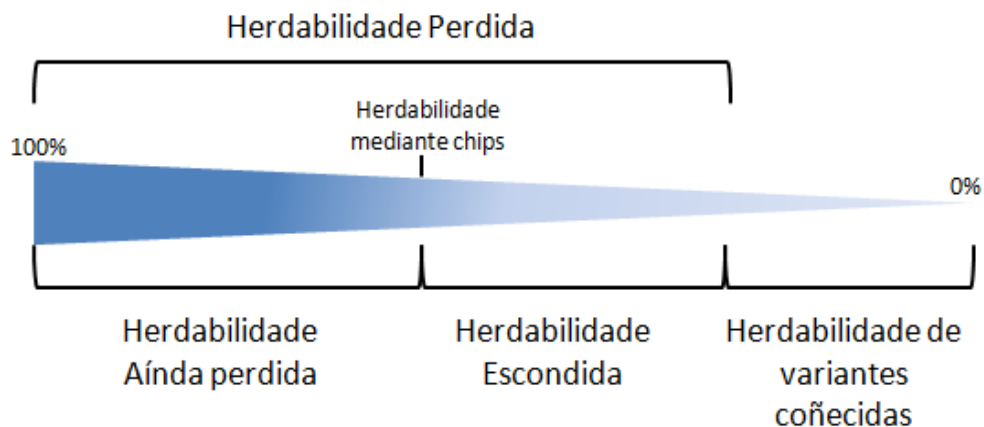


Figura 5. Diferentes compoñentes da herdabilidade explicada. Figurada adaptada de (Witte et al. 2014)

En contraposición á existencia de variantes comúns de predisposición a padecer enfermidades complexas xurde a hipótese de enfermidade común-variante rara. Esta sostén que existen múltiples variantes pouco frecuentes na poboación, que teñen un risco asociado moderado e que estarían influíndo no desenvolvemento da esquizofrenia, sen que ningunha delas fora por si soa determinante para a doenza (Pritchard 2001; Manolio et al. 2009). Esta hipótese cada día está collendo máis forza para explicar a existencia da susceptibilidade xenética a esquizofrenia. Para apoiar esta posibilidade

Introdución

está o feito de que esta doenza se manifeste ó comezo da idade adulta e estea ligada a unha fertilidade reducida provocando que a selección purificadora reduza a frecuencia das variantes alélicas de susceptibilidade. De xeito que, variantes asociadas a un risco elevado estarían sometidas a selección negativa manténdose a frecuencias baixas na poboación ou que, co paso do tempo, acabarían sendo eliminadas (Rees et al. 2011). Por outra banda, estarían as mutacións lixeiramente deletéreas nas que a selección non sería tan forte, polo que esta non permitiría que chegasen a ser comúns pero tampouco as eliminaría totalmente da poboación, podendo segregarse na poboación a baixa frecuencia e contribuíndo a susceptibilidade a padecer determinadas enfermidades comúns (Kryukov et al. 2007; Barreiro et al. 2008; Gorlov et al. 2008). Por último, coñécese ben que a idade paterna no momento do nacemento incrementa o risco a padecer esquizofrenia, feito que se interpretou como reflexo dunha acumulación de mutacións na liña xerminal ó longo da vida paterna (McClellan et al. 2007; Kong et al. 2012).

1.2.2.4.1. O Xene *DISC1*

Un dos primeiros estudos apoiando o modelo de enfermidade común-variante rara atopaba, nunha familia escocesa, unha asociación dunha alteración cromosómica balanceada con varios trastornos mentais (esquizofrenia, trastorno esquizoafectivo, trastorno bipolar e depresión maior) (Fig. 6) (St Clair et al. 1990; Blackwood et al. 2001). Esta alteración consistía nunha translocación cromosómica balanceada recíproca (1;11) (q42;q14.3) afectando ós xenes *Disrupted in Schizophrenia 1 (DISC1)* e *Disrupted in Schizophrenia 2 (DISC2)*. *DISC2* dá lugar a unha molécula de ARN non codificante e transcríbese no sentido contrario á *DISC1*.

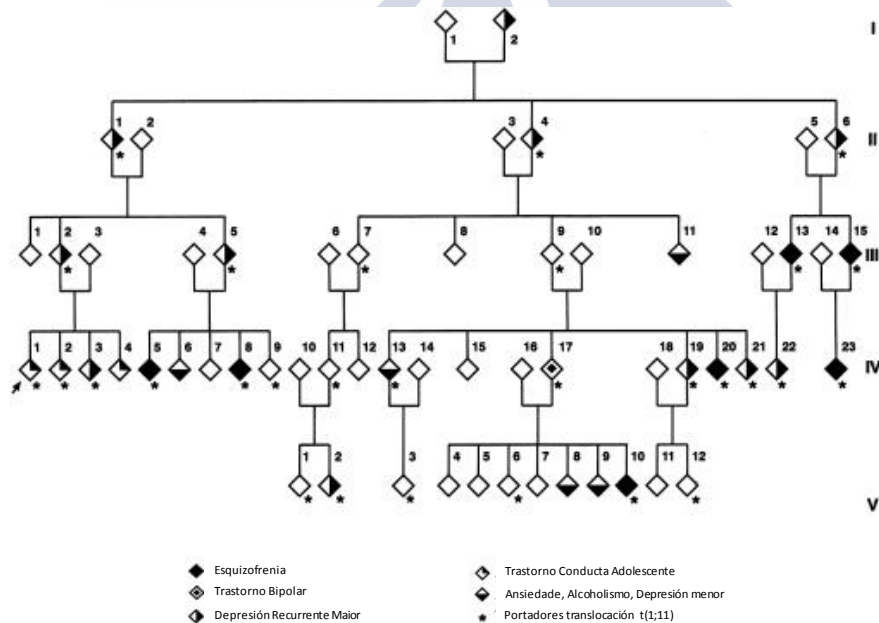


Figura 6. Familia coa translocación (1;11)(q42;q14.3). Adaptada de Blackwood, 2001.

O xene *DISC1* é un xene longo que consta de 13 exóns. En concreto engloba 415 kb de ADN xenómico, dando lugar a un transcrito completo de 7,5 kb (Millar et al. 2000), aínda que se identificaban 4 isoformas principais (*long*, L; *long variant*, Lv; *short*, s; e *extra short*, Es) (Hennah et al. 2006). Exprésase en múltiples localizacións subcelulares, como as mitocondrias, gránulos citoplasmáticos, núcleo, centrosoma e os filamentos de actina (Hennah et al. 2006).

O transcrito completo de *DISC1* dá lugar a unha proteína do mesmo nome que consta de 854 aminoácidos e cunha masa molecular aproximada de 100 kDa (Fig. 7). A rexión N-terminal forma un ou máis dominios de cabeza globular, mentres que a C-terminal contén rexións con potencial para formar espirais enroladas. A rexión N-terminal contén un pequeno motivo moi conservado enriquecido en serina e fenilalaninas e un sinal de localización nuclear bipartido (SLN) (Ma et al. 2002; Taylor et al. 2003; Bord et al. 2006); por outra banda, a rexión C-terminal contén tres motivos de cremalleiras de leucinas (Hennah et al. 2006). Os aminoácidos da posición 1 a 358 de *DISC1*, que corresponde co dominio da “cabeza”, son importantes para a localización da proteína no núcleo (Millar et al. 2005a); a parte central da proteína (entre os residuos 400 e 500) identificábase como un dominio de asociación libre (Kamiya et al. 2005); e a parte da cola da rexión C-terminal é importante para a localización de *DISC1* no citoesqueleto, no núcleo e no centrosoma (Morris et al. 2003; Kamiya et al. 2005; Millar et al. 2005a; Malavasi et al. 2012).

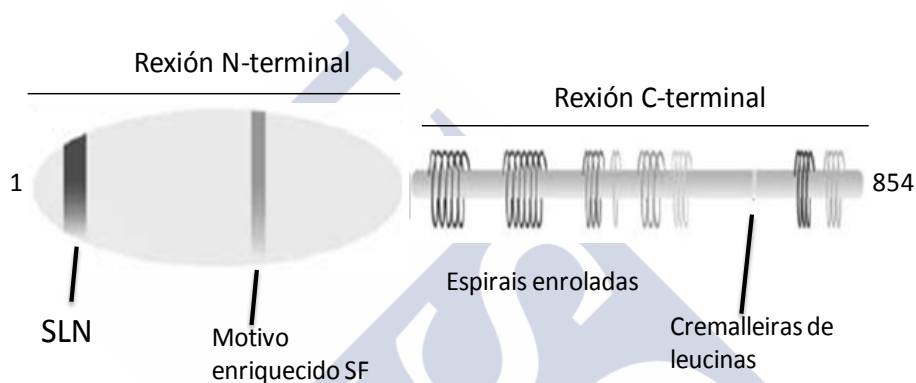


Figura 7. Estrutura esquemática de *DISC1*. Adaptado de (Chubb et al. 2008).

A expresión de *DISC1* é elevada durante o desenvolvemento do sistema nervioso central e descende gradualmente ó longo da vida (Nakata et al. 2009). Aínda que inicialmente se pensaba que *DISC1* se expresaba unicamente en neuronas, existen traballos que amosan que este xene exprésase tamén en múltiples tipos de células gliais, de feito, *DISC1* estaba implicada en funcións celulares de oligodendrocitos e astrocitos (Ma et al. 2013).

Un paso para intentar comprender as funcións nas que está implicado *DISC1* era a descrición do seu interactoma, é dicir, identificar tódalas interaccións moleculares (incluíndo interaccións directas proteína-proteína, interaccións a nivel metabólico ou a nivel de regulación da expresión) nas que está implicada *DISC1*. Esta proteína actúa principalmente como proteína de andamiaxe, unindo máis de 200 proteínas (Fig. 8), moitas das cales foron identificadas mediante experimentos de dobre híbrido de fermento. Utilizando esta técnica Camargo e cols. no 2007, publicaban un traballo no que se representaba unha rede de interaccións de *DISC1* con 127 proteínas e un total 158 interaccións. De feito, este estudo revelaba que *DISC1* probablemente participase na formación e maduración da sinapse, sendo potencialmente unha proteína clave neste proceso. Ademais, amosaba que *DISC1* está relacionado coa estabilidade e organización do citoesqueleto, transporte intracelular e co ciclo e división celular (Camargo et al. 2007).

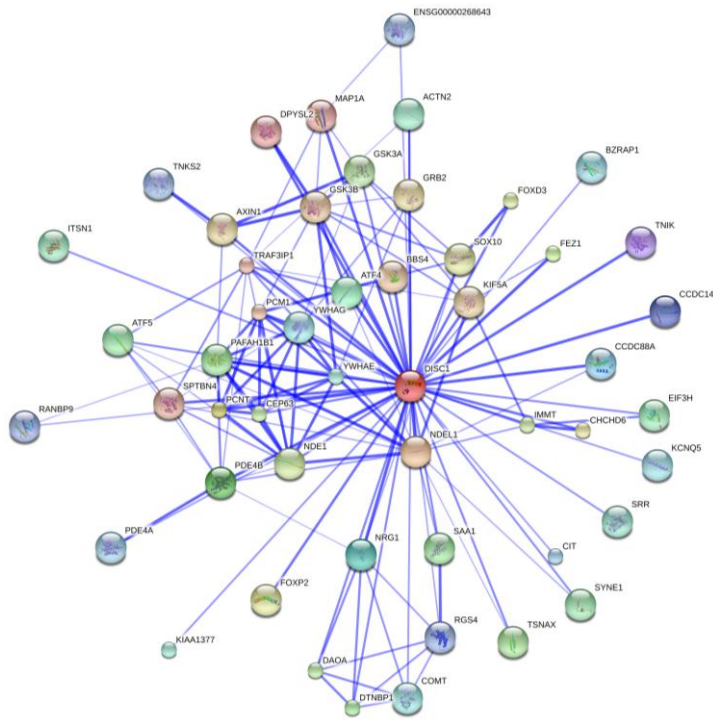


Figura 8. Esquema dalgunha das interaccións de *DISC1*, amosando a este como un xene central de diferentes funcións.

Outra das funcións na que podería estar implicado *DISC1* é na regulación da neuroxénese adulta e embrionaria, posto que este xene exprésase en grandes cantidades en zonas ventriculares ou subventriculares do córtex embrionario, onde se localizan as células proxenitoras neuronais (Mao et al. 2009). Neste estudo suxeríase que *DISC1* desenvolve unha función de regulación na vía de sinalización Wnt-GSK3β/β-catenina, a cal é necesaria para a proliferación celular. Un dos xenes cos que interactúa *DISC1* é o xene *DIXDC1* regulando ambos o proceso de sinalización mediante a via Wnt-GSK3β/β-catenina e a proliferación de proxenitores neurais corticais (Singh et al. 2010). Outros xenes que interactúan con *DISC1* son *LIS1* e *NDE1* (ou o seu parálogo *NDEL1*), os cales están directamente involucrados na regulación da neuroxénese (Burdick et al. 2008; Bradshaw et al. 2009). O complexo que forman este dous xenes regulan moitas funcións relacionadas co citoesqueleto ó través do seu efecto sobre a proteína motora dineína (Lam et al. 2010).

DISC1 tamén podería ser clave na migración neuronal (Young-Pearse et al. 2010; Steinecke et al. 2012). Tamén se demostraba que os xenes comentados anteriormente, *LIS1* e *NDEL1*, están ligados a migración neuronal (Sasaki et al. 2000). O xene *DIXDC1*, está implicado na migración neuronal durante o desenvolvemento do córtex, aínda que, a diferenza do comentado no parágrafo anterior, isto é independente da vía Wnt-GSK3β/β-catenina (Singh et al. 2010).

DISC1 é requirida para a integración neuronal (incorporación de novas neuronas que xorden a partir de células nai neuronais adultas no circuíto excitador), de feito, suxeríase que esta proteína actúa como un regulador, controlando a sincronización da integración neuronal ó través das súas diferentes etapas: morfoxénese neuronal, migración, adquisición da excitabilidade intrínseca e a formación da sinapse (Duan et al. 2007). *KIAA1212* e *AKT*, son outros dous xenes que interactúan con *DISC1*. Os tres forman parte dunha ruta metabólica que regula a maduración neuronal (Kim et al. 2009). Hai evidencias de que a sinalización do AMPc é outra ruta metabólica que

modula a maduración neuronal. *DISC1* está relacionado con esta vía, posto que esta proteína está relacionada directamente coa proteína *PDE4B*, a cal hidroliza o AMPc. Polo tanto, *DISC1* pode regular a sinalización do AMPc mediante a modulación da actividade de *PDE4B* (Millar et al. 2005b). Neste artigo amosaban que o xene *PDE4B* é un factor de risco en esquizofrenia xa que atopaban unha translocación (t(1;16)) afectando a este xene e ó xene *CDH8*, nunha familia escocesa cun individuo afecto con esquizofrenia e outro diagnosticado con trastorno psicótico.

Por outra banda, existen numerosos estudos que mostran que *DISC1* está localizada nas sinapses desempeñando múltiples papeis. Algún destes estudos amosan que *DISC1* está localizada nas espiñas dendríticas, concretamente na densidade postsináptica (Kirkpatrick et al. 2006; Bradshaw et al. 2008; Hayashi-Takagi et al. 2010; Wang et al. 2011b; Paspalas et al. 2013). Ademais, publicábase un estudo no que se identificaba un complexo no que se demostraba que *DISC1* xoga un papel clave na modulación do tamaño e densidade das espiñas dendríticas. En resumo, *DISC1* disocia a interacción entre *KAL-7* e *RAC1* reducindo a actividade deste último, o que produce un efecto na densidade e tamaño das dendritas xa que *RAC1* é un modulador destas espiñas (Hayashi-Takagi et al. 2010). Tamén hai evidencias de que *DISC1* modula a composición sináptica glutamatérxica mediante a interacción coa quinasa *TNIK* na densidade postsináptica (Wang et al. 2011b). A fosforilación e expresión desta proteína son dependentes da actividade do receptor do glutamato polo que *TNIK* podería estar relacionado coa tradución do sinal sináptico. Ademais das localizacións anteriores, *DISC1* tamén está localizada en lugares perisinápticos e extrasinápticos (Paspalas et al. 2013). Demostrábase que *DISC1* ten funcións presinápticas, xa que a sobreexpresión deste xene aumenta a probabilidade de liberación de glutamato na sinapse, mentres que o silenciamento de *DISC1* ten un efecto oposto (Maher and LoTurco 2012). Diferentes experimentos demostraban que a expresión de *DISC1* é dependente da actividade sináptica (Namba et al. 2011; Ramsey et al. 2011). Ma e cols. amosaban que esta proteína pode modular a transmisión sináptica por medio da interacción coa encima racemasa serina (Ma et al. 2013).

Hai varias evidencias que demostran que *DISC1* está directamente relacionado coa regulación transcricional nuclear, xa que a proteína que codifica este xene está parcialmente distribuída no núcleo, onde ten lugar a transcripción activa (Sawamura et al. 2005; Malavasi et al. 2012). Nestes artigos tamén se comprobaba que *DISC1* interactúa con dous factores de transcripción, *ATF4* e *AFT5*, regulando a súa actividade transcricional.

Demostrábase que *DISC1* está localizada no centrosoma en liñas celulares cultivadas e en neuronas primarias (Kamiya et al. 2005; Bradshaw et al. 2008; Kamiya et al. 2008), sendo importante para o correcto funcionamento deste orgánulo (Shimizu et al. 2008). Outros traballos vían que *DISC1* centrosomal está relacionada coa ancoraxe e recrutamento de varias proteínas esenciais no centrosoma. Unha delas é a proteína codificada polo xene *PCMI*, a cal a súa vez é necesaria para a localización doutras proteínas centrosomais (Kamiya et al. 2008) e para a correcta organización do conxunto de microtúbulos (Dammermann and Merdes 2002). *DISC1* centrosomal tamén é necesaria para a ancoraxe do complexo motor de microtúbulos dineína (Kamiya et al. 2005).

A presenza de *DISC1* en neuronas estriadas e células NIH3T3 parece indicar que este xene está relacionado coa formación (cilioxénese) e funcións da estrutura dos cilios primarios, tales como o tráfico e modulación funcional de receptores D2 ciliares (Marley and von Zastrow 2010).

Introdución

Outros dos orgánulos nos que está presente DISC1 é a mitocondria, concretamente localizábase tanto en lugares inespecíficos da membrana mitocondrial (Ramsey et al. 2011; Paspalas et al. 2013) como na membrana mitocondrial externa (Park et al. 2010). DISC1 forma parte dun complexo multiproteico que une as membranas das cristas á membrana mitocondrial interna e promove a ensamblaxe do complexo da cadea respiratoria (Hoppins et al. 2011). Demostrábase que DISC1 modula o transporte e dinámica da mitocondria en dendritas e axóns neuronais (Park et al. 2010; Norkett et al. 2016).

Por último, DISC1 desenvolve un rol directo no transporte intracelular baseado en microtúbulos tanto no transporte axonal anterógrado, no que interactúa con distintas proteínas (GRB2, LIS1, NDEL1, FEZ1 e 14-3-3Epsilon) (Miyoshi et al. 2003; Shinoda et al. 2007; Taya et al. 2007), como no transporte intracelular retrógrado (Kamiya et al. 2005). Esta función era detectada tanto en mitocondrias (comentado anteriormente) como nas vesículas sinápticas (Flores et al. 2011).

Tódolos estudos sobre as funcións de DISC1 revelaban moito sobre os procesos que están afectados cando a proteína non funciona correctamente, pero aínda sigue sendo un reto intentar comprender como as rutas metabólicas nas que está implicado este xene poden contribuír ós trastornos psiquiátricos.

1.2.2.4.2. CNVs en Esquizofrenia

Outro tipo de alteracións cromosómicas que apoian a hipótese enfermidade común-variante rara son as CNVs que están presentes a baixa frecuencia na poboación. Este termo define unha secuencia de ADN de polo menos 1 kb de tamaño que pode variar no número de copias en comparación cun xenoma de referencia e son o resultado dunha duplicación ou dunha deleción do material cromosómico (Williams et al. 2009). Unha CNV pode ser unha estrutura simple, como unha duplicación en tándem, ou pode abarcar ganancias ou perdas complexas de secuencias homólogas en múltiples lugares do xenoma, e a diferenza doutro tipo de variacións estruturais (inversións, translocacións recíprocas, etc.) poden ser non balanceadas, dando lugar a diferenzas no tamaño do xenoma entre individuos e alterando o equilibrio biolóxico nun locus determinado nun xenoma diploide.

Unha CNV pode manifestar o efecto fenotípico a través dunha gran variedade de mecanismos estruturais e reguladores: (1) Unha CNV pode dar lugar a un cambio de dose dun ou máis xenes nunha rexión, podendo exercer un efecto na expresión dos xenes podendo causar un maior risco a padecer unha enfermidade (Hurles et al. 2008). Posto que un cambio na dose causada por unha duplicación é menos acusada que a causada por unha deleción, suponse que as duplicacións teñen, normalmente, un efecto fenotípico menor que as delecións; (2) unha CNV pode alterar a expresión hemicigótica dunha mutación recesiva ou exacerbar o efecto dunha variante funcional nun segmento cromosómico non delecionado (Lupski and Stankiewicz 2005); (3) estas variacións poden alterar a expresión xénica dun xene situado fora da CNV, podendo ter un efecto posicional ou alterar unha trans-regulación dun xene (Lupski and Stankiewicz 2005).

Son variacións amplamente distribuídas ó longo do xenoma humano, representando unha das principais fontes de variación xenética (Iafrate et al. 2004; Sebat et al. 2004). Os mecanismos de formación deste tipo de variacións son inferidos normalmente a partir da información da secuencia nos sitios de rotura ou unión dos reordenamentos, e son principalmente 4 tipos: recombinación homóloga non alélica, que ocorre mediante un cruce entre dous lugares no xenoma que comparten rexións de secuencia homóloga,

dando como resultado unha deleción e unha duplicación recíproca das secuencias implicadas; o segundo tipo é a unión de extremos non homólogos, que ten lugar como resultado de erros nos mecanismos de reparación dos extremos rotos da dobre cadea de ADN, tendo lugar habitualmente en elementos repetitivos; o terceiro é a parada na forcada de replicación e troco do molde o cal é un mecanismo de reordenación xenómica baseado na replicación e que está inducido por erros durante o proceso de replicación do ADN; o cuarto e último mecanismo é a retrotrasposición, que é levada a cabo por elementos nucleares intercalados largos tipo 1 (LINE1), que son a única clase activa de retrotrasposóns en humanos, e que se poden traspoñer a novas rexións xenómicas mediante un mecanismo de transcrición reversa (Malhotra and Sebat 2012).

A primeira CNV confirmada como factor de risco para a esquizofrenia era a deleción situada na rexión 22q11.2 (Karayiorgou et al. 1995; Murphy et al. 1999). Este locus está afectado por dúas delecións máis frecuentes cun tamaño de aproximadamente 2,3 Mb e de 1,5 Mb, as cales están formadas mediante recombinación homóloga non alélica (Sharp et al. 2005). Esta asociábase previamente ó síndrome Velocardiofacial (VCFS), tamén coñecido como Síndrome DiGeorge ou Síndrome de Shprintzen (Shprintzen et al. 1978; Driscoll et al. 1992). A deleción ten unha prevalencia de 1 cada 4000 nacementos (Tezenas Du Montcel et al. 1996; Goodship et al. 1998), aparecendo como *de novo* no 72-92% dos portadores (Driscoll et al. 1993; Ryan et al. 1997).

Os afectos que presentan esta deleción teñen un fenotipo moi variable incluíndo características como insuficiencia velofarinxeal, defectos conxénitos cardíacos, dificultades na aprendizaxe, inmunodeficiencia, estatura baixa e hipocalcemia. Moitos afectos presentan retraso mental ademais de alteracións cognitivas e de comportamento, e aproximadamente un 30% dos portadores da deleción desenvolven esquizofrenia (Murphy et al. 1999).

Posteriormente ós estudos anteriormente comentados, nos que trataban de identificar as características clínicas dos portadores da deleción, levábanse a cabo estudos caso-control nos que analizaban un gran número de mostras coa finalidade de avaliar o risco de padecer esquizofrenia nos portadores desta deleción. Deste xeito, víase que esta variación aumenta o risco a padecer esta doenza (Levinson et al. 2011; Vacic et al. 2011; Szatkiewicz et al. 2014). Pola contra, un estudo publicado no 2014 amosaba que unha duplicación nesta rexión ten efecto protector contra a esquizofrenia (Rees et al. 2014a). Dependendo do tamaño da deleción, a rexión afectada inclúe entre 35 e 60 xenes coñecidos (Edelmann et al. 1999; Shaikh et al. 2000), estando moitos deles expresados no cerebro (Maynard et al. 2003). Aínda que non se coñece cales destes xenes están relacionados coa esquizofrenia, varios foron propostos como xenes candidato funcionais: *COMT*, *PRODH*, *DGCR8*, *ZDHHC8* e *GNBL1* (Tam et al. 2009).

Ó longo de todos estes anos utilizáronse varias técnicas para a detección das CNVs; ó principio, o principal método para a detección das CNVs era a hibridación fluorescente *in situ* (FISH, en inglés *Fluorescence In Situ Hybridization*), a cal non tiña suficiente resolución para identificar a maior parte das CNVs; logo foron xurdindo outras novas metodoloxías moleculares como: a reacción en cadea da polimerasa cuantitativa (qPCR, das súas siglas en inglés), que permite detectar CNVs que afectan a un único exón, pero que, aínda que a súa capacidade é moi limitada para detectar múltiples dianas nun único ensaio, é útil para a validación de resultados atopados con técnicas de xenoma completo (Ceulemans et al. 2012); a hibridación xenómica comparada (CGH, das súas siglas en inglés) que permite a detección de CNVs mediante a comparación do ADN de mostras problema cunha mostra de referencia; e a técnica de amplificación múltiple dependente de ligazón (MLPA, das súas siglas en inglés) que se

basea nun ensaio con múltiples sondas (hibridación das sondas, ligazón, PCR e separación dos fragmentos nun secuenciador capilar), cada unha das cales é específica dunha secuencia diferente de ADN para medir o seu número de copia (Stuppia et al. 2012). Nos últimos anos houbo grandes avances no desenvolvemento das tecnoloxías de *microarrays* de ADN de alto rendemento, facendo que, hoxe en día, a maioría dos estudos de CNVs en esquizofrenia estean baseados na análise de datos obtidos de *microarrays* de SNPs. O suposto é que a maior parte do xenoma ten dúas copias, e a intensidade do sinal dos marcadores do *array* ó longo do cromosoma é comparado entre os dous alelos dun marcador. Debido a que a intensidade de cada alelo e cada marcador varía, é necesario utilizar un grupo control para corrixir os datos xa que o comportamento de cada marcador é distinto. Cando a intensidade do sinal cambia dunha forma clara e continua durante un intervalo xenómico razoable, pódese inferir un cambio no número de copia. O normal é que estas CNVs inferidas necesiten ser verificadas experimentalmente mediante outras técnicas. Esta metodoloxía permite a identificación de CNVs distribuídos ó longo de todo o xenoma.

Estes avances nas tecnoloxías permitían a identificación dun gran número de CNVs raras que causan un risco elevado a padecer esquizofrenia (Táboa 3 e Fig. 9):

Deleción e duplicación en 1q21

A deleción 1q21.1 era das primeiras CNVs asociadas á esquizofrenia. Abrangue un tamaño de 1,35 Mb e ten un risco asociado (OR) de 8,35. Inicialmente, identificábase nunha mostra de 4718 esquizofrénicos e 41201 controis de distintos países (Stefansson et al. 2008). Esta asociación replicábase nun estudo levado a cabo polo Consorcio Internacional de Esquizofrenia no mesmo ano utilizando unha mostra de ancestralidade similar e no traballo levado a cabo por Rees e cols. (International Schizophrenia 2008; Rees et al. 2014c). Unha duplicación recíproca tamén era asociada coa esquizofrenia cun OR entre 3,45, suxerindo que calquera alteración na dose dos xenes que abrangue esta rexión pode conferir risco a padecer a enfermidade (Levinson et al. 2011; Rees et al. 2014c).

Deleción 2p16.3

Esta deleción, cuxo tamaño é variable pero inferior a 1 Mb, abrangue unha rexión que afecta total ou parcialmente ó xene *NRXN1*. A primeira vez que esta CNV se asociaba á esquizofrenia era no 2008. Neste ano Kirov e cols. levaban a cabo un estudo no que utilizaban *arrays* de CGH para analizar 93 casos de esquizofrenia e 392 controis, nos que atopaban unha deleción nun único caso que afectaba a rexión promotora e o primeiro exón de *NRXN1* (Kirov et al. 2008). Pouco despois deste traballo, Walsh e cols., utilizando a mesma metodoloxía que o anterior, identificaban unha deleción que afectaba ós exóns do extremo 3´ deste mesmo xene, en dous xemelgos idénticos con diagnóstico de esquizofrenia, entre 233 casos e 268 controis (Walsh et al. 2008). Un ano máis tarde, Rujescu e cols., utilizando a tecnoloxía de *microarrays*, analizaban o xene *NRXN1* en 2977 esquizofrénicos e 33746 controis. Así identificaban 7 CNVs nos casos e 5 en controis afectando ós exóns deste xene alcanzando valores significativos ($P = 0,0027$, $OR = 8,97$, $95\% CI = 1,8-51,9$) (Rujescu et al. 2009). No 2009 levábase a cabo unha meta-análise que apoiaba a asociación desta CNV coa esquizofrenia, alcanzando un $P = 0,000037$ e un OR de 7,44 ($95\% IC = 3,22-17,18$) (Kirov et al. 2009b). Por último, nun estudo máis recente, analizaban unha mostra de 6882 esquizofrénicos e 6316 controis co fin de determinar a contribución das distintas CNVs asociadas

previamente á esquizofrenia, entre as cales estaban as que afectaban a *NRXN1*. Estes identificaban 11 casos que portaban unha CNV alterando este xene. Con estes datos e cos previamente publicados, facían unha meta-análise confirmando a asociación (Rees et al. 2014c).

Un dos aspectos que cabe resaltar desta CNV é que, a diferenza das outras que afectan a diferentes xenes, é das poucas CNVs que está asociada consistentemente á esquizofrenia e que ademais abrangue un único xene, amosando que o efecto desta CNV é causado por este xene (Kirov et al. 2009b). Tamén destacar que esta CNV non é recorrente, xa que se identificaban CNVs de diferente tamaño e localización afectando ó xene (rexión promotora e primeiro exón, extremo 3' do xene, etc.). Isto é debido a que estas variantes non están flanqueadas por rexións repetitivas, típicas noutros reordenamentos cromosómicos que dan lugar a CNVs recorrentes mediante recombinación homóloga non alélica (Rujescu et al. 2009). Por outra banda, identificábase unha asociación significativa de CNVs que afectaban a *NRXN1* con pacientes con trastorno do espectro autista (Kim et al. 2008; Glessner et al. 2009), sendo isto unha evidencia máis da idea da existencia dunha etioloxía e patoxénese común a ambas doenzas. Tamén suxería que estas CNVs alterando *NRXN1* poderían ter consecuencias no desenvolvemento e función sináptica, sendo un exemplo da hipótese do neurodesenvolvemento en esquizofrenia. Esta hipótese é un modelo explicativo da esquizofrenia e postula que a doenza ten o súa orixe nas alteracións producidas durante o desenvolvemento do sistema nervioso e do cerebro, aínda que esta se manifeste na idade adulta (Weinberger 1987).

NRXN1 codifica unha proteína de adhesión celular presináptica, a cal se une a proteínas postsinápticas chamadas neuregulinas xogando un papel importante na formación, mantemento e liberación de neurotransmisores na sinapse (Sudhof 2008). O xene está involucrado na diferenciación, maduración, estabilización e plasticidade das sinapses inhibitorias e excitadoras (Bang and Owczarek 2013).

Deleción 3q29

O papel desta CNV era reportado por primeira vez por Mulle e cols. (Mulle et al. 2010), sendo posteriormente confirmada (Levinson et al. 2011; Rees et al. 2014c). Esta deleción ten un tamaño de 1,6 Mb e abarca 22 xenes que codifican para proteínas, destacando entre eles os xenes *PAK2*, *DLG1* e *FBXO45*, sendo propostos como xenes candidato para a esquizofrenia (Mulle 2015). O risco asociado é dos máis elevados alcanzando un OR de 41,1 nunha recente meta-análise (Mulle 2015), acadando nesta un nivel de significación global moi forte ($P=5,8 \times 10^{-8}$).

Duplicación 7q36.3

Esta microduplicación de tamaño variable abrangue unha rexión de 362 kb e está localizada augas arriba e ó longo da rexión codificante do xene *VIPR2*. Era asociada por primeira vez á esquizofrenia por Vacic e cols. no 2011 (Vacic et al. 2011) e posteriormente replicada por Levinson e cols. (Levinson et al. 2011) aínda que, nun traballo posterior, onde analizaban 6882 esquizofrénicos e 6316 controis, non alcanzaba valores significativos (Rees et al. 2014c). *VIPR2* codifica para proteína VPAC2, que é un receptor acoplado a proteína G, a cal se activa por un ligando iniciando unha cascada de sinalización neuronal (Piggins 2011). Este xene exprésase no cerebro, no núcleo supraquiasmático, hipocampo, amígdala e hipotálamo (Sheward et al. 1995), estando implicado na regulación da transmisión sináptica no hipocampo (Yang et al. 2009) e

Introdución

promove a proliferación de células proxenitoras neurais no xiro dentado (Zaben et al. 2009).

Duplicación 7q11.23

A primeira vez que se publicaba a asociación entre esta duplicación e a esquizofrenia era no 2014 (Mulle et al. 2014), atopando aquí un risco asociado de 10,78 tras facer unha meta-análise de máis de catorce mil casos e entorno a vinte e oito mil controis. Este valor aumentaba a 11,35 tras engadir as CNVs atopadas no traballo de Rees (Rees et al. 2014c). Esta rexión, ademais de estar previamente asociada ó autismo (Van der Aa et al. 2009; Sanders et al. 2011), presenta unha delección recíproca que causa a Síndrome de Williams-Beuren (Tassabehji 2003).

Delección 7q34–36.1

Esta CNV afecta ó xene *CNTNAP2* e era identificada no 2008 por Friedman e cols. Estes publicaban un estudo no que atopaban 3 individuos afectados de esquizofrenia e/ou epilepsia que presentaban este reordenamento cromosómico non sendo atopado en 512 controis (Friedman et al. 2008). *CNTNAP2* está asociado ao Síndrome de Pitt Hopkins (Zweier et al. 2009) ó igual que *TCF4*, o cal tamén está asociado á esquizofrenia (Stefansson et al. 2009; Steinberg et al. 2011a). Destacar que *TCF4* codifica un factor de transcrición que se une no promotor de *CNTNAP2* para regular a súa expresión (Forrest et al. 2012).

Este xene da lugar a proteína Caspr2 (Proteína 2 asociada a contactina), a cal pertence á superfamilia de neurexinas, un grupo de proteínas transmembrana que media a interacción célula-célula no sistema nervioso. Esta proteína exprésase principalmente no sistema nervioso, na rexión iuxtaparanodal dos axóns de neuronas mielinizadas e nos oligodendrocitos (Poliak et al. 1999). Esta proteína é esencial para a localización de canles de voltaxe activados por K^+ na rexión iuxtaparanodal dos axóns, o cal podería indicar que esta proteína esta implicada na estabilización da condución e na axuda a manter o potencial de repouso intermodal (Vabnick et al. 1999).

CNVs en 15q

Neste locus atopábanse varias asociacións positivas coa esquizofrenia. No ano 2008 publicábanse as primeiras deleccións como factor de risco da enfermidade: a delección 15q11.2 e a delección 15q.13.3, tendo a primeira delas tamaño de 470 kb e un OR =2,73; e a segunda un tamaño de 1,58 Mb e un OR=11,54 (International Schizophrenia 2008; Stefansson et al. 2008). Estas dúas CNVs confirmábanse posteriormente como loci de susceptibilidade para a esquizofrenia (Kirov et al. 2009a; Levinson et al. 2011; Melhem et al. 2011; Grozeva et al. 2012; Rees et al. 2014c).

Neste mesmo cromosoma identificábase unha duplicación de orixe materna no locus 15q11-q13, a cal tamén está asociada ao síndrome de Prader-Willi, como un factor de risco en varios individuos esquizofrénicos e esquizoafectivos (Ingason et al. 2011a). Trala análise combinada de aproximadamente 15 mil casos estimábase un OR asociado a esta duplicación de 13,2 (Rees et al. 2014c). Resaltar que esta duplicación está considerada como un dos factores xenéticos máis comúns para o trastorno do espectro autista aparecendo cunha frecuencia de 1 cada 500 casos (Moreno-De-Luca et al. 2013).

CNVs en 16p

O cromosoma 16 presenta varias CNVs asociadas á esquizofrenia. Unha delas é a duplicación no locus 16p11.2, que abarca 600 kb que contén 28 xenes (McCarthy et al. 2009), alcanzando un risco asociado superior a 11 (Rees et al. 2014c). Tanto duplicacións como deleccións afectando esta mesma rexión tamén están asociadas ó trastorno do espectro autista (Weiss et al. 2008). A segunda das CNVs está situada nunha rexión máis distal da anteriormente comentada, tamén na rexión cromosómica 16p11.2, e era identificada inicialmente en dúas cohortes independentes de xudeus Ashkenazis e búlgaros, sendo replicada nunha mostra maior de Estados Unidos, Europa e Xapón, acadando un OR = 6,25 (Guha et al. 2013), confirmándose tamén posteriormente cun OR de 3,39 (Rees et al. 2014c). Outra das CNVs que afecta a este locus está situada en 16p12.1 e corresponde a unha delección de 480 kb; esta asociación era descuberta no 2014 nunha mostra de 7000 casos (Rees et al. 2014b) e replicada neste mesmo traballo nunha cohorte de 15 mil casos. A delección abrangue 7 xenes e está asociada a un risco de 2,72.

O cromosoma 16 contén unha duplicación situada na rexión xenómica 16p13.11 que alcanzaba, inicialmente, un valor significativo coa esquizofrenia despois de analizar arredor de 4300 casos e aproximadamente trinta e cinco mil controis (Ingason et al. 2011b). A duplicación que afecta a esta rexión ten un tamaño de 1,5 Mb e está fortemente asociada á esquizofrenia cun OR de 7,27. Esta rexión contén ós xenes *NTANI*, que está relacionado coa memoria, e ó xene *NDE1* que interactúa con *DISC1* estando ambos involucrados na neuroxénese (como xa comentamos anteriormente), e que posteriormente se asociaba coa esquizofrenia mediante o descubrimento dunha variante rara neste xene (Kimura et al. 2015). Hai outros estudos que apoiaban a asociación desta CNV coa esquizofrenia alcanzando OR de 2,3 (Kirov et al. 2009a; Ikeda et al. 2010; Rees et al. 2014c). Destacar que duplicacións e deleccións nesta rexión cromosómica están asociadas con autismo e con retraso mental (Ullmann et al. 2007; Hannes et al. 2009).

Delección 17p12 e duplicación/delección 17q12

O primeiro traballo que relacionaba ó locus 17p12 coa esquizofrenia era un estudo de ligamento en familias con individuos afectos (Park et al. 2004). Posteriormente Kirov e cols. atopaban unha asociación dunha delección de 1,3 Mb coa esquizofrenia despois de analizar 5300 casos e 39 mil controis e cun OR = 10 (Kirov et al. 2009a). Por outra banda, a delección 17q12, era observada entre os casos de esquizofrenia pero non estaba presente nos case 52500 controis estudados (Moreno-De-Luca et al. 2010). Esta delección tiña un tamaño de 1,4 Mb e un OR estimado de 4,49. Estas dúas deleccións comentadas eran replicadas no traballo de Rees e cols. (Rees et al. 2014c). Moi recentemente identificábase nesta mesma rexión unha duplicación con evidencias de estar asociada á esquizofrenia cun $P = 0,018$ e un OR de 4,16 (Szatkiewicz et al. 2014).

Introdución

Locus	Tipo	Frec SCZ (%)	Frec. CO (%)	OR	IC 95%	valor-P	Outros trastornos asociadas
1q21	Del	0,17	0,02	8,35	(4,65–14,99)	4,1E-13	RM-MC, TEA, THDA
1q21	Dup	0,13	0,027	3,45	(1,92–6,20)	9,90E-05	RM-MC, TEA, THDA, TB
2p16.3	Del	0,18	0,02	9,01	(4,44–18,29)	1,60E-11	TEA, RM-MC
3q29	Del	0,082	0,0014	41,1	(5,6–1953,6)	5,80E-08	RM-MC, DI, TEA
7q11.23	Dup	0,066	0,0058	11,35	(2,58–49,93)	6,90E-05	TEA, RM-MC, THDA, ANS
7q36.3 ^a	Dup	0,11	0,069	1,54	(0,77–3,09)	2,70E-01	
7q11.23 ^b	Del	NA	NA	NA	NA	NA	EP, SPH
15q11.2	Del	0,59	0,28	2,15	(1,71–2,68)	2,50E-10	RM-MC, THDA, TOC, EP
15q11-q13	Dup	0,083	0,0063	13,2	(3,72–46,77)	5,60E-06	RM-MC, TEA
15q.13.3	Del	0,14	0,019	7,52	(3,98–14,19)	4,00E-10	RM-MC, THDA, TEA, EP
16p11.2	Dup	0,35	0,03	11,52	(6,86–19,34)	2,90E-24	RM-MC, TEA, TB, THDA, ANS, DM
16p11.2	Del distal	0,063	0,018	3,39	(1,21–9,52)	1,70E-02	TEA, RM-MC
16p12.1	Del	0,15	0,057	2,72	(1,48–5,02)	1,60E-03	RM-MC, EP
16p13.11	Dup	0,31	0,13	2,3	(1,57–3,36)	5,70E-05	RM-MC, THDA
17p12	Del	0,094	0,026	3,62	(1,73–7,57)	1,20E-03	
17q12	Del	0,036	0,0054	6,64	(1,78–24,72)	7,20E-03	RM-MC, TEA
17q12	Dup	0,072	0,017	4,16	(1,28-Inf)	1,80E-02	RM-MC, TEA
22q11.2	Del	0,29	0	NA	(28,27-Inf)	4,4E-40	RM-MC, TEA, THDA, ANS, DM

Táboa 3. Resumo das CNVs asociadas á esquizofrenia. Datos extraídos de (Friedman et al. 2008; Rees et al. 2014c; Mulle 2015). SCZ= Esquizofrenia; Co= Control; Del=Delección; Dup=Duplicación; RM-MC=Retraso Mental e Malformacións Conxénitas; TEA=Trastorno do Espectro Autista; DI=Discapacidade Intelectual; ANS= Ansiedade; EP=Epilepsia; THDA=Trastorno de Hiperactividade e Déficit de Atención; DM=Depresión Maior, TB=Trastorno Bipolar, TPH=Trastorno Pitt-Hokins.

^a Esta CNV (7q36.3 dup) asociábase significativamente (P=0,006) coa esquizofrenia en estudos anteriores (Levinson et al. 2011; Vacic et al. 2011).

^b Esta CNV asociábase á esquizofrenia pero non dispoñemos de datos de frecuencia, risco asociado e valor de significación (Friedman et al. 2008).

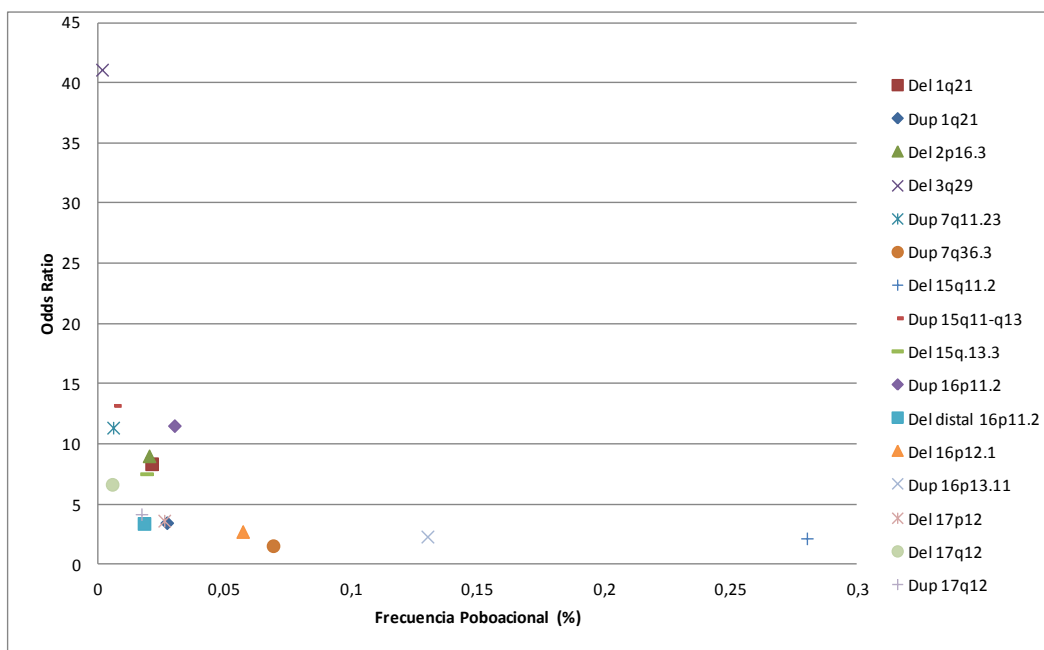


Figura 9. Representación gráfica da frecuencia poboacional das principais CNVs asociadas á esquizofrenia fronte ó risco asociado de cada unha.

Unha das características das CNVs en esquizofrenia é a preponderancia de casos *de novo*. Estimábase que as CNVs *de novo* son 8 veces máis frecuentes en pacientes esquizofrénicos esporádicos comparado con controis, mentres que os casos só teñen 1,5 veces máis posibilidades de albergar CNVs herdadas que controis (Xu et al. 2008). Os riscos asociados (OR) a cada unha das CNVs son moi variables, sendo estes máis elevados cos asociados a loci comúns (Levinson et al. 2011). No 2011, Rees e cols. publicaban un traballo no que trataban de analizar o efecto na poboación xeral deste tipo de CNVs de gran tamaño, onde estimaban a taxa de aparición e presión selectiva; para isto centrábanse en CNVs que provocasen un risco a padecer esquizofrenia e atopaban que estas CNVs estaban baixo unha gran presión selectiva, alcanzando valores de $s = 0,12-0,88$, suxerindo que trala súa aparición, estas CNVs perduran poucas xeracións (non máis de 5 xeracións). En canto a taxa de mutación vían que era alta en comparación coa taxa de mutacións puntuais, alcanzando valores entre 1:3500 e 1:300000 (Rees et al. 2011), polo que esta elevada taxa é o que fai que as CNVs asociadas a esquizofrenia sexan recorrentes. Anteriormente e de forma similar, Itsara e cols. estimaban que as CNVs cun tamaño maior a 500 kb sufrían unha forte selección purificadora alcanzando valores de $s=0,16$ e unha taxa de mutación de 0,0065 por xenoma e por transmisión (Itsara et al. 2010). Por outra banda, a frecuencia destas variantes é moi baixa na poboación xeral, estimándose un frecuencia menor a 0,001. Todo isto fai que exista un equilibrio mutación-selección sobre as CNVs de risco, de xeito que canto maior risco cause unha CNV sobre un determinado trastorno psiquiátrico a súa frecuencia será menor (Fig. 9). Por último, resaltar que as CNVs, aínda que o tamaño do efecto é significativamente máis elevado que as variantes identificadas mediante GWAS, amosan unha penetrancia reducida e unha expresividade variable, tendendo a predispoñer, ademais da esquizofrenia, a un rango amplo de trastornos do neurodesenvolvemento como a epilepsia, autismo, discapacidade intelectual, etc.

1.2.2.4.3. Estudo de SNVs mediante secuenciación tradicional

Xunto ás CNVs, outro tipo de alteracións que apoian a hipótese enfermidade común-variante rara son as SNVs. Os *microarrays* que se usan para levar a cabo os GWAS só estudan as variantes incluídas no deseño; algún destes *microarrays* inclúen no seu deseño variantes a moi baixa frecuencia, pero hai outras que só aparecen en moi poucos individuos ou incluso nun único individuo por ser *de novo*, non podendo ser analizados por estes métodos. Para poder analizar este tipo de variantes é preciso recorrer a secuenciación. Os primeiros estudos trataban de secuenciar e analizar rexións ou xenes candidato en tamaños de mostras bastante reducidos mediante secuenciación de Sanger (Táboa 4). Por exemplo, no ano 2008 Song et al. publicaban un estudo no que trataban de buscar variantes raras en *DISC1* mediante a secuenciación dos exóns e *splice junctions* do xene *DISC1* en 288 casos e 288 controis. Neste traballo identificaban cinco variantes moi raras de risco elevado en 6 doentes e que non estaban presentes en controis ($P=0,015$), considerándoas ultra-raras despois de comprobar que ningunha delas aparecía en dez mil controis. En total atopaban 22 casos e 8 controis que presentaban algunha variante rara (Song et al. 2008).

No 2009, outros autores resecuenciaban o xene *VGLUT1* en 376 pacientes esquizofrénicos e 368 controis dunha poboación de Taiwan atopando 16 variantes raras en casos e ningunha en controis, vendo que estas variantes estaban sobrerrepresentadas en casos respecto dos controis ($P=1,5 \times 10^{-5}$) (Shen et al. 2009). Neste mesmo ano, publicábase un artigo no que facían un *screening* de variantes raras utilizando HRMA (*High Resolution Melting Analysis*) das partes exónicas do xene *DTNPB1* en 669 casos e 710 controis, non alcanzando asociación significativa con ningunha das variantes raras atopadas. Ademais de comprobar que non existía un enriquecemento de variantes raras en casos respecto dos controis (Dwyer et al. 2010a).

O ano seguinte Shen e cols. publicaban un estudo no que atopaban un exceso significativo ($P=2,3 \times 10^{-5}$) de variantes raras en casos ($N=375$) con respecto ós controis ($N=366$) no xene *VGLUT2* nunha poboación de Taiwan unha vez que resecuenciaban as rexións codificantes e as rexións promotoras deste xene (Shen et al. 2010).

Nun artigo publicado no 2011 resecuenciaban 111 xenes ligados ó cromosoma X en 143 doentes esquizofrénicos e 142 de autistas. Como controis utilizaban 277 cromosomas X de individuos sans. Identificaban máis de 200 variantes non sinónimas atopando un exceso de variantes raras prexudiciais ($P=0,0004$), suxerindo que algunha destas podería ser unha variante causal. Ademais facían unha análise para comprobar os xenes que amosaban un exceso de variantes raras non sinónimas. Entre os casos con esquizofrenia, obtíñan diferenzas significativas no xene *SLITRK2* ($P=0,0006$); tamén alcanzaban diferenzas significativas o xene *SEPT6* ($P=0,004$), *KCND1* e *MECP2* (ambos cun valor de $P=0,004$). Por outra banda, entre os casos con autismo, só o xene *CACNA1F* ($P=0,0011$) amosaba un exceso significativo de variantes raras (Piton et al. 2011). Neste mesmo ano, Sun e cols. secuenciaban tódolos exóns e as rexións promotoras do xene *NLGN2* en 584 esquizofrénicos e 549 controis de Taiwan atopando unha tendencia a acumulación de variantes raras pero sen alcanzar valores significativos ($P=0,07$) (Sun et al. 2011). Outro traballo que saía a luz neste ano era o levado a cabo por Green e cols., que secuenciaban o exón 11 do xene *DISC1* en 506 casos e 1211 controis e no que atopaban 6 variantes, 4 das cales só aparecían en esquizoafectivos e vían que existía un exceso significativo de variantes raras en casos respecto a controis ($P=0,008$) (Green et al. 2011). Mühleisen e cols. secuenciaban o xene *NRXN1* en 94

esquizofrénicos e 94 controis atopando 16 variantes raras, existindo maior número nos casos pero sen conseguir valores significativos (Muhleisen et al. 2011).

Gauthier e cols. no 2011 publicaban un traballo onde secuenciaban *NRXN1*, *NRXN2* e *NRXN3* en 142 autistas, 143 esquizofrénicos e 94 afectos de discapacidade intelectual. Estes atopaban dúas variantes *de novo* que truncaban o xene, unha delas situada no xene *NRXN2* nun paciente con trastorno do espectro autista e outra nun esquizofrénico en *NRXN1* (Gauthier et al. 2011).

No 2012, Kushima e cols. publicaban un traballo no que secuenciaban as rexións exónicas dos xenes *KALRN* e *EPHB1*, usando un método baseado en *microarrays* de ADN; despois dunha fase inicial de secuenciación e dúas posteriores de xenotipado, atopaban evidencias de mutacións raras codificantes de sentido incorrecto en *KALRN*, podendo ser este xene un factor de risco para a esquizofrenia (Kushima et al. 2012). Neste ano, Shen e cols. publicaban un experimento onde secuenciaban as rexións exónicas e as rexións promotoras do xene *NRGN* en 346 esquizofrénicos e 345 controis atopando 5 variantes raras en casos e ningunha en controis, pero ningunha delas alcanzaba valores significativos (Shen et al. 2012).

Li e cols., no 2013, publicaban un estudo despois de secuenciar as rexións exónicas do xene *DLGAP3* en 215 esquizofrénicos e 215 controis, no que non atopaban ningunha asociación significativa con ningunha das mutacións identificadas (Li et al. 2013b). Un ano máis tarde saía publicado un artigo onde secuenciaban os xenes *MIR137* e *MIR2682*, xunto coas súas rexións reguladoras en 2610 esquizofrénicos e 2611 controis, vendo que existía unha diferenza significativa na carga de variantes raras en rexións promotoras e *enhancers* entre os casos e os controis (Duan et al. 2014). Kimura e cols. publicaban neste ano un artigo onde trataban de buscar variantes raras no xene *NDE1* en 433 esquizofrénicos e 145 mostras con trastornos no desenvolvemento, para logo facer unha análise nunha mostra maior que incluía casos de esquizofrenia, trastorno bipolar e controis, alcanzando valores significativos para a variante S214F en esquizofrenia e para a R234C en trastorno bipolar (Kimura et al. 2015).

Recentemente Luo e cols. publicaban un traballo que constaba de dúas fases. Na inicial secuenciaban as rexións promotoras e UTR (en inglés, *untranslated region*) do xene *DISC1*, en 100 afectos de esquizofrenia. Mediante este proceso identificaban 17 SNPs que eran xenotipados nunha mostra independente de 1447 pacientes e 1154 controis. Ningún dos SNPs analizados acadaba valores significativos, aínda que dous haplotipos marcadores alcanzaban unha asociación nominal para protección ($P=0,0476$) (Luo et al. 2016).

Introdución

Autor	Ano	Xene/s	Nº Ca	Nº Co	Resultado	Valor-P ^a
Song	2008	<i>DISC1</i>	288	288	Exceso de SNVs raras en casos	1,50E-02
Shen	2009	<i>VGLUT1</i>	376	368	Exceso de SNVs raras en casos	1,50E-05
Dwyer	2009	<i>DTNPB1</i>	669	710	Non existen diferenzas significativas na acumulación de SNVs raras	-
Shen	2010	<i>VGLUT2</i>	375	366	Exceso de SNVs raras en casos	2,30E-05
Piton	2011	111 xenes ligados ó Crom.X	143 SCZ /142TEA	277	Exceso de SNVs raras prexudiciais en casos	4,00E-04
					<i>SLITRK2</i> (SCZ)	6,00E-04
					<i>SEPT6</i> (SCZ)	4,00E-03
					<i>KCND1</i> (SCZ)	4,00E-03
					<i>MECP2</i> (SCZ)	4,00E-03
					<i>CACNA1F</i> (TEA)	1,10E-03
Sun	2011	<i>NLGN2</i>	584	549	Tendencia acumulación SNVs raras	7,00E-02
Green ^b	2011	<i>DISC1</i>	506	1211	Exceso de SNVs raras en casos	8,00E-03
Mühleisen	2011	<i>NRXN1</i>	94	94	Non existen diferenzas significativas na acumulación de SNVs raras	-
Gauthier	2011	<i>NRXN1, NRXN2, NRXN3</i>	142 TEA /143 SCZ /94 DI		Dúas variantes <i>de novo</i> truncando <i>NRXN1</i> e <i>NRXN2</i> nun SCZ en un TEA, respectivamente	-
Kushima ^c	2012	<i>KALRN, EPHB1</i>	320	-	Asociación significativa en <i>KALRN</i>	4,80E-02
Shen	2012	<i>NRGN</i>	346	345	Non existen diferenzas significativas entre as SNVs atopadas	-
Li	2013	<i>DLGAP3</i>	215	215	Non existen diferenzas significativas entre as SNVs atopadas	-
Duan	2014	<i>MIR137, MIR2682</i>	2610	2611	Exceso de SNVs raras en rexións promotoras e <i>enhancers</i> casos	2,10E-03
Kimura ^d	2014	<i>NDE1</i>	433 SCZ /145 TDV	-	S214F (SCZ)	3,90E-03
					R234C (TB)	3,20E-03
Lou ^e	2016	<i>DISC1</i>	100	-	2 haplotipos marcadores	4,76E-02

Táboa 4. Resultados dos estudos de variantes raras mediante secuenciación tradicional. SCZ= Esquizofrenia, TDV= Trastorno do Desenvolvemento, TEA= Trastorno do Espectro Autista.

^a Test exacto de Fisher

^b Green e cols. secuenciaban unicamente o exón 11 do xene.

^c Na segunda fase do estudo analizan 729 casos e 562 controis e na terceira fase 1511 casos e 1517 controis.

^d Na segunda fase analizan 3554 SCZ, 1041 TB e 4746 controis.

^e Na segunda fase analizan 1447 SCZ e 1154 controis.

Aínda que a gran maioría dos resultados dos traballos publicados usando a secuenciación Sanger amosan unha ausencia de variantes raras asociadas dunha forma consistente coa esquizofrenia, principalmente por carecer de potencia dabondo para atopar variantes raras de risco moderado, esta estratexia é unha boa aproximación para buscar variantes raras. Este tipo de estudos son un paso previo ós estudos de

secuenciación de nova xeración comentados a continuación (NGS, das súas siglas en inglés).

1.2.2.4.4. Estudo de SNVs mediante secuenciación de nova xeración.

A vista dos recentes traballos publicados sobre a esquizofrenia, queda demostrado que a contribución da variación común na doenza é significativa, aínda que esta non explica toda a herdabilidade estimada para o trastorno. Por outra banda, os estudos de CNVs explicados anteriormente amosan que as variantes raras contribúen a explicar parte desta herdabilidade. O seguinte paso sería comprobar se as variantes puntuais raras contribúen neste aspecto, xa que os GWAS levados a cabo ata este momento non permiten analizar esta parte do espectro. Recentes avances na NGS permitiron aumentar o rendemento, á vez que descendía o custo destes procesos, podendo así analizar maior número de mostras e rexións do xenoma. Isto unido ó desenvolvemento de modelos estatísticos que incorporaran múltiples variantes de susceptibilidade, aspecto que se comentará nun apartado posterior (apartado 1.4), facilita a análise do papel das variantes puntuais (SNVs) raras na susceptibilidade a padecer a esquizofrenia.

O desenvolvemento da metodoloxía de NGS permitiu a secuenciación de rexións diana ou secuenciación dirixida xa que é a metodoloxía máis alcanzable economicamente. Grazas a esta podemos estudar rexións do xenoma que foron asociadas previamente a un fenotipo mediante GWAS, por exemplo, e que poden albergar variantes raras. As rexións son elixidas segundo unha hipótese inicial, polo que é necesario coñecemento previo.

A medida que esta metodoloxía foi evolucionando o prezo foi descendendo polo que se comezaban a levar a cabo estudos de secuenciación dirixida pero que implicaban todo o exoma. Estes estudos teñen por obxectivo secuenciar o 1% -2% do xenoma que codifica para proteínas. Aínda que o exoma ocupa unha porcentaxe moi pequena do xenoma, estimábase que pode albergar ata o 85% das variantes que causan un gran efecto sobre os trazos relacionados cunha enfermidade, polo que esta é unha estratexia para capturar rexións que, en principio, poderían ser máis susceptibles de albergar variantes causais (Choi et al. 2009). Os métodos de captura do exoma divídense en dúas categorías principais: métodos baseados en solución e métodos baseados en *arrays*. Na primeira categoría as mostras de ADN son fragmentadas e utilízanse sondas marcadas con biotina para que se hibriden as rexións de interese do xenoma; a continuación engádense bolas magnéticas con estraptavidina para que se unan as sondas biotilizadas e despois de facer varios lavados para eliminar as rexións que non interesan, utilízase unha PCR para amplificar a rexión diana. Os métodos baseados en *arrays* son similares ós anteriores pero diferéncianse en que as sondas están unidas a un *microarray* de alta densidade (Warr et al. 2015). Normalmente, mediante a secuenciación do exoma obtéñense elevadas profundidades de lectura (a profundidade de secuenciación fai referencia ó número medio de lecturas que cubre cada unha das bases a estudar). Esta cobertura non é homoxénea ó longo do exoma, principalmente causado polo contido de GC, de xeito que as rexións con elevado contido en GC tenden a ter peor éxito na captura e, polo tanto, unha profundidade de lectura menor. Aínda así, unha profundidade media de 60X-80X no exoma pode alcanzar unha alta probabilidade de cobertura >20X nunha fracción grande da rexión codificante. Debido a que a tecnoloxía de enriquecemento das dianas non é perfecta, a secuenciación dos exomas tamén produce algunhas lecturas de secuenciación fora da rexión diana (Lee et al. 2014). A principal limitación desta metodoloxía é que captura só a variación xenética no exoma,

cando se demostrou que rexións non codificantes poderían xogar tamén un importante papel en enfermidades complexas. Por exemplo, a maioría dos loci de GWAS están sitos nestas rexións, ademais de que os recentes resultados do proxecto ENCODE suxiren que moitas rexións non codificantes poderían ter unha función biolóxica importante.

O seguinte paso nesta sucesión sería a secuenciación do xenoma completo a baixa profundidade, xa que é unha alternativa máis custo-efectiva en relación á secuenciación profunda, que se comenta a continuación. De feito que, no lugar de secuenciar un individuo a unha profundidade de 30X, poderíase secuenciar máis individuos con menor cobertura por base cun custo similar (por exemplo, 7 ou 8 individuos a 4X). Esta estratexia depende de métodos baseados no desequilibrio de ligamento que aproveitan a información á través dos individuos para mellorar a calidade da detección de variantes e os xenotipos estimados (Li et al. 2011). Estudos iniciais de simulación amosaban que a secuenciación a baixa profundidade para mostras grandes podería ser máis potente que secuenciación profunda en mostras menores, tanto para a detección de variantes como para estudos posteriores de asociación (Lee et al. 2014).

Nun futuro, grazas á progresiva redución dos custos destas tecnoloxías poderanse levar a cabo estudos de secuenciación do xenoma en profundidade analizando gran número de mostras. Esta estratexia é a máis informativa xa que se secuencia todo o xenoma con coberturas elevadas. Sen embargo, a combinación da secuenciación do xenoma completo a gran escala e os deseños epidemiolóxicos clásicos, como os estudos caso-control ou estudos de cohortes, actualmente son imposibles de levar a cabo debido ó elevado custo.

Con respecto a esquizofrenia existen diversos estudos publicados, os cales son comentados e subdivididos en función da estratexia utilizada:

- **Estudos de secuenciación dirixida de xenes candidato mediante NGS**

Dende a aparición da secuenciación de nova xeración foron xurdindo varios artigos nos que se secuenciaban diferentes xenes candidato en tamaños de mostra bastante reducidos. Aínda así algún dos resultados obtidos (Táboa 5) permitía elucidar o papel das variantes raras na etiopatoloxía da esquizofrenia.

No 2011 publicábase o primeiro artigo que utilizaba a estratexia de secuenciación dirixida de xenes candidato sobre esquizofrenia. Moens e cols. secuenciaban o xene *DISC1* e dez xenes máis que interactúan con el (*ATF5*, *Grb2*, *FEZ1*, *LIS-1*, *PDE4B*, *NDE1*, *NDEL1*, *TRAF3IP1*, *YWHAE* e *ZNF365*), coa finalidade de ver a contribución destes xenes na susceptibilidade á esquizofrenia. Para isto secuenciaban estes once xenes en dúas *pools* de casos e outras dúas de controis, comprendendo 80 esquizofrénicos e 80 individuos non afectos que pertencían a unha poboación illada do norte de Suecia. Aquí identificaban 50 variantes que eran validadas e que, posteriormente, eran xenotipadas en 486 casos e 514 controis. Mediante unha análise de carga de mutacións atopaban que variantes raras non sinónimas ($MAF < 0,01$) son máis frecuentes en pacientes que en controis ($P = 0,0018$), amosando evidencias de que tanto *DISC1* como os xenes que interactúan con el están involucrados en trastornos psiquiátricos (Moens et al. 2011).

No 2013 Crowley e cols. levaban a cabo a secuenciación Sanger de dez xenes candidato (*COMT*, *DAOA*, *DISC1*, *DRD2*, *DRD3*, *DTNBP1*, *HTR2A*, *NRG1*, *SLC6A3* e *SLC6A4*) en 727 esquizofrénicos e 733 controis. Despois de facer unha selección e

posterior confirmación mediante técnicas de secuenciación de nova xeración das variantes, xenotipaban 92 SNVs en 2191 casos e 2659 controis. Ningunha variante nin ningún dos xenes analizados alcanzaban valores significativos (Crowley et al. 2013). Kenny e cols. atopaban un exceso de variantes que provocaban unha perda de función ($P=0,0007$), unha vez que secuenciaban as rexións exónicas de 215 xenes nunha mostra combinada de 237 casos de esquizofrenia e 147 afectos do trastorno do espectro autista, ademais de 287 controis. Este mesmo exceso detectábase na categoría de xenes de proteínas que interactúan con Neurexina e Neurolixina ($P=0,05$) (Kenny et al. 2013).

Outro dos traballos nos que utilizaban esta metodoloxía era feito por Thomson e cols. Estes secuenciaban o xene *DISC1* nunha mostra de 889 controis e 653 casos diagnosticados de esquizofrenia (240), trastorno bipolar (221) e trastorno depresivo maior recorrente (192). A única asociación significativa que atopaban era entre o trastorno depresivo e o SNP rs16856199; aínda que non era replicada nunha segunda mostra, alcanzaba valores significativos ($P=0,0058$) despois dunha análise conxunta (Thomson et al. 2014). Noutro dos traballos publicados neste ano secuenciaban 101 xenes previamente relacionados coa esquizofrenia nunha mostra de aproximadamente 1000 casos e 1000 controis que incluía unha cohorte descubrimento e unha de réplica. Aquí vían que os casos presentaban case dúas veces máis alelos sen sentido que os controis, aínda que ningún dos xenes amosaba diferenzas significativas entre os dous grupos na presenza de variantes funcionais (Hu et al. 2014).

Girard e cols. publicaban un traballo onde secuenciaban en total 102 xenes, 62 dos cales eran seleccionados por presentar algunha variante *de novo* identificada en estudos previos (Awadalla et al. 2010; Girard et al. 2011; Xu et al. 2011). Os restantes interactuaban con algún dos xenes anteriores. A secuenciación era feita en 235 esquizofrénicos e 233 controis. Unha vez analizados os datos obtidos vían que existía un enriquecemento de variantes privadas sen sentido entre os doentes. Tamén facían un test de carga dividindo os xenes por grupos e vían que existían diferenzas significativas entre casos e controis. Por último, facían unha análise de carga por cada un dos xenes incluídos e atopaban que os xenes *EP300*, *MAPK14* e *SHANK3* amosaban diferenzas significativas entre os casos e os controis (Girard et al. 2015).

No 2016 Ishizuka e cols. levaban a cabo a secuenciación dirixida do xene *PCDH15* en 370 esquizofrénicos e 192 individuos afectos con autismo de orixe xaponés. Como resultado disto identificaban 16 SNVs, 6 das cales eran seleccionadas para xenotipar nunha mostra maior de 2096 casos (1174 esquizofrénicos e 382 TEA) e 1917 controis. Unha vez analizados os datos vían que 4 delas tiñan unha $MAF < 0,0005$ (ultra-raras) e que podían aumentar o risco a padecer a enfermidade, pero ningunha das SNVs raras alcanzaba valores significativos (Ishizuka et al. 2016).

Introducción

Autor	Xene/s	Mostra inicial		Mostra Réplica		Resultado
		Nº Ca	Nº Co	Nº Ca	Nº Co	
Moens (2011)	<i>DISC1, ATF, Grb2, FEZ1, LIS-1, PDE4B, NDE1, NDELI, TRAF3IP1, YWHAE e ZNF365</i>	80	80	486	514	Exceso de SNV raras nonsin. en casos
Crowley (2013)	<i>COMT, DAOA, DISC1, DRD2, DRD3, DTNBP1, HTR2A, NRG1, SLC6A3 e SLC6A4</i>	727	733	2191	2659	Ningún xene alcanzaba valores significativos
Kenny (2013)	215 xenes	237 SCZ, 147 TEA	287	-	-	Exceso de SNVs LOF en casos Exceso de SNVs LOF na categoría de xenes de proteínas que interactúan con <i>NRXN1</i> e <i>NLGN</i>
Thomson (2014)	<i>DISC1</i>	240 SCZ, 221 TB, 192 DM	889	1802 DM	4017	rs16856199 ^a
Hu (2014)	101 xenes	525	619	455	336	Exceso de SNVs sen sentido en casos
Girard (2015)	102 xenes	235	233	-	-	Enriquecemento de SNVs privadas sen sentido entre os doentes Exceso significativo de variantes raras nos xenes EP300, MAPK14 e SHANK3
Ishizuka (2016)	<i>PCDH15</i>	370 SCZ, 192 TEA	-	1174 SCZ, 382 TEA	1197	Ningunha SNV alcanzaba valores significativos

Táboa 5. Resumo dos resultados atopados mediante secuenciación de nova xeración de xenes diana. DM=Depresión Maior; TEA=Trastorno do Espectro Autista; SCZ=Esquizofrenia; TB=Trastorno Bipolar; LOF= *Loss of function*.

^a Este SNP apareceu como significativo en DM (Depresión Maior).

• Estudos de secuenciación de exoma caso-control mediante NGS

No 2012 publicábase un artigo onde utilizaban a secuenciación do exoma ou xenoma de 166 esquizofrénicos ou trastorno esquizoafectivo e 307 controis atopando 337312 variantes, aínda que ningunha delas alcanzaba valores significativos despois da corrección por Bonferroni ($P < 1,5 \times 10^{-7}$). Das variantes atopadas, seleccionábanse aquelas que alcanzaban un $P < 0,005$ e as que eran exclusivas de casos. En total eran máis de 5000 variantes que se xenotipaban nunha mostra maior, composta de 2617 casos e 1800 controis. Os resultados amosaban que ningunha delas acadaba valores significativos utilizando un valor de significación de $1,5 \times 10^{-7}$. Por último, destacar que segundo o estudo de potencia que facían, este traballo permitiríalles identificar variantes cunha frecuencia comprendida entre 1% e 5% e cun risco relativo entre 2 e 6. Polo tanto, a vista dos seus resultados e o cálculo de potencia descartaban a existencia de variantes moderadamente raras que tivesen un risco relativo moderado que estivesen asociadas á esquizofrenia (Need et al. 2012).

Un ano máis tarde publicábase un traballo onde analizaban 5 familias cun total de 24 individuos afectos (18 de esquizofrenia, 4 de depresión e 2 de psicose non especificada) e 17 individuos sans. Estas mostras eran analizadas utilizando un *chip* feito a medida para buscar CNVs en rexións recorrentes de reordenamento con

asociacións coñecidas a doenzas, non identificando ningunha. A continuación seleccionaban dous ou tres esquizofrénicos con menos relación de parentesco de cada familia facendo un total de 12 suxeitos para secuenciar o exoma de cada un deles. Así atopaban variantes raras que provocaban unha alteración na proteína en tres xenes que estaban relacionadas co receptor de glutamato NMDAR, sinalando que esta podería ser unha ruta metabólica clave na patofisioloxía da esquizofrenia (Timms et al. 2013).

No 2014, Balan e cols. publicaban un experimento no que secuenciaban o exoma de dous individuos, un dos cales estaba diagnosticado de esquizofrenia e outro que non padecía ningunha enfermidade psicótica. Ámbolos dous eran portadores da deleción 22q11.2. Os resultados amosaban variantes xenéticas exclusivas do caso e que estaban relacionadas con funcións neuronais (Balan et al. 2014).

Purcell e cols. no 2014, secuenciaban o exoma de 2536 esquizofrénicos e 2543 controis. Os resultados amosaban que existía unha carga polixénica debido a mutacións que afectaban a proteínas e que tiñan unha frecuencia do alelo menor inferior ó 0,5% ou *singletons* (alelos que estaban presentes unha vez en individuos heterocigotos) distribuídas ó longo de 2456 xenes pertencentes a distintas categorías previamente implicadas na doenza. Tamén vian que ningún dos xenes seleccionados individualmente amosaban un exceso significativo de mutacións raras en casos, pero si algún dos conxuntos seleccionados. Curiosamente outros dos resultados obtidos concordaban cun estudo sobre tríos publicado neste mesmo ano (Fromer et al. 2014), xa que atopaban un enriquecemento de mutacións daniñas raras nos complexos ARC (*Activity-Regulated Cytoskeleton-associated protein*) e NMDAR (*N-methyl-D-aspartate receptor*), e nas dianas de FMRP (*fragile X mental retardation protein*). Ademais este tipo de variantes estaban enriquecidas en xenes que codifican para proteínas de canles de calcio dependente da voltaxe. Neste artigo atopaban unha asociación nominal (non resistía a corrección de múltiples tests) nun conxunto de xenes candidato (2507) de autismo e discapacidade intelectual para variantes que alteraban a proteína e variantes non sinónimas que eran preditas como patoxénicas polos 5 preditores utilizados (Purcell et al. 2014). Aínda que este era o primeiro traballo de NGS que demostraba a existencia dunha compoñente polixénica debido a variantes raras, observábase que a contribución deste tipo de variantes era menor á que provocaban as variantes comúns (Schizophrenia Working Group of the Psychiatric Genomics 2014).

Moi recentemente Singh e cols. publicaban un traballo onde secuenciaban inicialmente o exoma de 1745 esquizofrénicos e 6789 controis. A estes engadínanlle os datos de secuenciación dun estudo previo (Purcell et al. 2014), co fin de aumentar a potencia do estudo, facendo un total de 4264 casos e 9343 controis. Con estes datos confirmaban o enriquecemento de variantes raras de perda de función (LOF das súas siglas en inglés, que son variantes que alteran xenes que codifican proteínas e que teñen unha alta probabilidade de ser deletéreas e son de gran interese na etioloxía dunha doenza) nun conxunto de xenes candidato (2507) de autismo e discapacidade intelectual descrito por Purcell e cols. (Purcell et al. 2014). Coa finalidade de atopar algún xene de risco, a maiores incluían datos de 1077 tríos de esquizofrenia de diferentes estudos previos. Con todos estes datos e tendo en conta só variantes LOF raras ($MAF < 0,1\%$) o xene que alcanzaba o valor máis significativo ($P = 3,3 \times 10^{-9}$) era *SETD1A*, o cal foi asociado á esquizofrenia por Takata e cols. no 2014 (artigo comentado posteriormente) (Takata et al. 2014). Este xene codifica unha metiltransferasa que cataliza a metilación dos residuos de lisina na histona H3. Este descubrimento unido ás evidencias de variantes comúns previas, suxería que unha mala regulación epixenética, especialmente

Introdución

no ruta metabólica da metilación de histonas H3K4, podería ser un mecanismo importante na patoxénese da esquizofrenia (Singh et al. 2016).

Autor	Mostra inicial		Mostra Réplica		Resultado
	Nº Ca	Nº Co	Nº Ca	Nº Co	
Need (2012)	166	307	2167	1800	Ningunha SNV alcanzaba valores significativos
Timms (2013) ^a	24	17	-	-	SNVs en 3 xenes relacionados co receptor de glutamato NMDAR
Purcell (2014)	2536	2543	-	-	Existencia dun compoñente polixénico debido a variantes raras Enriquecemento de mutacións daniñas raras nos complexos ARC e NMDAR, e nas dianas de FMRP
Balan (2014)	1	1	-	-	Existencia de SNVs exclusivas nos casos
Singh (2016)	4264, 1077 tríos	9343	-	-	Enriquecemento de variantes LOF raras nun conxunto de xenes ^b Asociación significativa (P=3,3 x 10 ⁻⁹) de <i>SETD1A</i> coa <i>SCZ</i>

Táboa 6. Resumo dos resultados atopados mediante secuenciación de nova xeración en estudos caso-control. LOF= Variante de perda de función; SCZ=Esquizofrenia

^a Trátase dun estudo familiar no que a mostra de casos estaba composta de 18 esquizofrénicos, 4 afectos con depresión maior e 2 con psicose non especificada.

^b O conxunto de xenes era descrito por Purcell e cols. (Purcell et al. 2014).

• Estudos de secuenciación de exoma en tríos mediante NGS

O estudo de tríos (é dicir, un individuo afecto e os seus pais) permite determinar se as variantes son herdadas ou *de novo* (mutacións que xorden esporadicamente nos gametos dos pais ou nas primeiras etapas do desenvolvemento embrionario), de xeito que unha variante *de novo* é máis probable que sexa patoxénica se ámbolos pais non están afectos. Esta estratexia era utilizada con éxito noutras doenzas, como a discapacidade intelectual (Vissers et al. 2010) e o trastorno do espectro autista (O'Roak et al. 2011). A hipótese subxacente a estes estudos é que o conxunto dos casos esporádicos probablemente terán unha acumulación de variantes raras *de novo*, o cal é reflexo dunha elevada taxa de mutación (Schreiber et al. 2013). Ó longo destes anos foron xurdindo traballos (Táboa 7) que utilizaban estes principios co fin de identificar variantes *de novo* asociadas a esquizofrenia.

O primeiro estudo utilizando esta estratexia era levado a cabo por Girard e cols. no cal secuenciaban o exoma de 14 esquizofrénicos e os seus respectivos pais, e no que atopaban 15 variantes *de novo* en 8 dos esquizofrénicos. A taxa de mutación observada neste estudo excedía a taxa de mutación na liña xerminal predita ata o momento, a cal variaba de $1,1 \times 10^{-8}$ a $3,8 \times 10^{-8}$ por nucleótido e por xeración (Girard et al. 2011).

Seguindo unha estratexia similar, aparecía un traballo no que tamén secuenciaban o exoma de 53 casos esporádicos, 22 controis e os seus respectivos pais; aquí atopaban 40 mutacións *de novo* en 27 casos, e que tiñan, segundo predicións *in silico*, un efecto funcional en 40 xenes. Vían que había un exceso de variantes *de novo* non sinónimas nos casos, o cal apoiaba a hipótese de que mutacións *de novo* xogan un papel importante no risco a padecer esquizofrenia (Xu et al. 2011).

Ó ano seguinte Xu e cols. publicaban un artigo onde analizaban un total 795 exomas de 231 tríos de esquizofrenia así como 34 tríos non afectos, observando un exceso de variantes nucleotídicas non sinónimas en casos. Ademais vían que existía unha maior prevalencia en casos de variantes que afectaban a xenes respecto ós controis. Mediante este estudo tamén amosaban que había unha alta contribución das

mutacións en xenes con maior expresión nas primeiras etapas da vida fetal (Xu et al. 2012).

Co fin de avaliar a contribución de variantes *de novo* á esquizofrenia, Guipponi e cols. secuenciaban o exoma de 53 individuos con esquizofrenia esporádica e os seus pais non afectados. Vían que o número de variantes *de novo* por individuo afecto era de 0,88 dando lugar a unha taxa de mutación exónica puntual de $1,7 \times 10^{-8}$ por nucleótido por xeración, vendo que o 34% dos seus casos esporádicos portaban algunha variante *de novo* predita por ser prexudicial. Ademais proporcionaban un listado de 18 posibles xenes candidato para a esquizofrenia (Guipponi et al. 2014).

Takata e cols. utilizando os datos do traballo de secuenciación de Xu e cols. (Xu et al. 2012), no que secuenciaban 795 exomas de 231 tríos de esquizofrenia así como 34 tríos non afectados, analizaban as variantes LOF *de novo*. Concretamente centrábanse nas indels (insercións e deleccións de pequeno tamaño). Estes autores vían que existía un exceso de indels LOF *de novo* en esquizofrénicos aínda que non alcanzaba valores significativos. Por outro banda, amosaban unha correlación entre o número de variantes LOF *de novo* e pacientes con historia de dificultades de aprendizaxe na adolescencia ($P=0,0006$). Tamén atopaban dúas indels LOF no xene *SETD1A*, sendo isto moi pouco probable por azar ($P=0,035$). Este xene era asociado á esquizofrenia nun traballo de Singh e cols. que se comentou anteriormente (Singh et al. 2016). Utilizando os seus datos e os de estudos previos de exoma en tríos seleccionaban un total de 62 xenes que eran intolerantes á variación e que presentaban variantes LOF, levaban a cabo unha análise de conxunto de xenes, alcanzando o valor máis significativo a función da regulación da cromatina. Por último vían, mediante unha análise de patróns de transmisión, que as variantes LOF eran máis probables de ser transmitidas a individuos afectados que a controis (Takata et al. 2014).

Fromer e cols. levaban a cabo o maior estudo de mutacións *de novo* en esquizofrenia feito ata o momento. Nel secuenciaban o exoma completo de 534 esquizofrénicos, 89 pacientes afectados de trastorno esquizoafectivo e os pais de cada un deles. Os resultados atopados non evidenciaban un aumento total na taxa de mutacións *de novo* que causasen perda da función. Un resultado similar era atopado cando se analizaban as variantes non sinónimas. Isto suxería que as mutacións *de novo* xogan un papel menor do que era indicado en estudos previos na susceptibilidade a padecer esquizofrenia. Tamén estudaban varios conxuntos de xenes que foran previamente asociados a esquizofrenia (Kirov et al. 2012; Timms et al. 2013; Schizophrenia Working Group of the Psychiatric Genomics 2014). Entre estes atopaban un enriquecemento significativo de mutacións *de novo* non sinónimas nos xenes que codifican proteínas post-sinápticas glutamatérxicas que comprende a proteína asociada ó citoesqueleto regulada por actividade (ARC) e o complexo receptor N-metil-D-Aspartato (NMDAR). Ademais atopaban un enriquecemento de mutacións en xenes que codifican para proteínas que interactúan con complexos que modulan a resistencia sináptica, tales como proteínas que regulan a dinámica dos filamentos de actina e aqueles cuxos ARN mensaxeiros son diana de proteínas do retraso mental do X fráxil (FMRP). Por outra banda, facían unha análise sen unha hipótese previa para intentar atopar un mecanismo relacionado coa enfermidade non descuberto ata ese momento baseándose nas anotacións de Gene Ontology (GO). Só unha categoría alcanzaba valores significativos ($P_{\text{corrixido}}=0,032$) despois da corrección para tódalas categorías de GO, e que correspondía cos xenes relacionados coa ensamblaxe dos paquetes de filamentos de actina. Outro dos aspectos que destacaban neste traballo era que trataban de ver se existía unha etioloxía xenética compartida entre a esquizofrenia, o trastorno do

Introdución

espectro autista e discapacidade intelectual. Para isto comprobaban se existía solapamento entre os xenes afectados por mutacións *de novo* entre estas tres doenzas, comparando os datos obtidos deste estudo con outros publicados anteriormente. Facendo isto vían que efectivamente xenes afectados por mutacións *de novo* en esquizofrenia tamén o estaban nas outras dúas enfermidades, evidenciando unha etiopatoxenia común entre os trastornos (Fromer et al. 2014).

Recentemente Takata e cols. utilizando bases de datos públicas de secuenciación de exomas de tríos analizaban o papel de mutacións sinónimas *de novo* en diferentes tipos de funcións reguladoras de xenes. En total, incluían 1043 autistas, 1021 esquizofrénicos e 731 controis. Analizando este tipo de variacións vían que, por un lado, aquelas que estaban próximas ós lugares de *splicing* e que modificaban a reguladores exónicos de *splicing* estaban significativamente enriquecidas en individuos autistas. Utilizando unha análise de conxunto de xenes atopaban un enriquecemento destas variantes en autismo en xenes relacionados coas canles e co transporte de calcio, e coa localización e función neuronal. Por outro lado, atopaban que aquelas variantes situadas no cortex prefrontal derivadas de lugares de hipersensibilidade a DNasa I estaban asociadas á esquizofrenia. Por último, avaliaban aqueles xenes que estaban afectados por variantes prexudiciais *de novo* e por mutacións sinónimas potencialmente funcionais. En canto ó autismo, o xene que alcanzaba os valores máis significativos era *RAB2A* ($P= 4,48 \times 10^{-6}$), e no caso da esquizofrenia destacaban ó xene *SETD1A* ($P= 1,79 \times 10^{-6}$) que previamente xa era asociado a esta doenza (Takata et al. 2016).



Autor	Nº Tríos		Resultado
	Ca	Co	
Girard (2011)	14	-	Exceso de mutacións de novo
Xu (2011)	53	22	Exceso de mutacións de novo non sinónimas nos casos
Xu (2012)	231	34	Exceso de variantes non sinónimas. Contribución de mutacións en xenes con maior expresión nas primeiras etapas da vida fetal
Guipponi(2014)	53		Mutación exónica puntual de $1,7 \times 10^{-8}$ por nucleótido por xeración Listado de 18 posibles xenes candidato
Takata (2014)	231	34	Exceso non significativo de indels LOF <i>de novo</i> en esquizofrénicos Dúas indels LOF no xene <i>SETD1A</i> Enriquecemento significativo no xenes relacionados coa "función da regulación da cromatina" Maior probabilidade transmisión de LOF ós casos
Fromer (2014)	534 SCZ, 89 SCZA	-	Taxa de mutacións <i>de novo</i> disruptivas non aumentada Enriquecemento de mutacións de novo en ARC e NMDAR Enriquecemento en xenes da regulación dos filamentos de actina e xenes con ARNm son diana de FMRP Etioloxía compartida entre SCZ, TEA e DI
Takata (2016)	1043 TEA, 1021 SCZ	731	Mutacións de novo sinónimas situadas próximas ós lugares de <i>splicing</i> e que modifican reguladores exónicos de <i>splicing</i> asociadas a TEA Mutacións de novo sinónimas situadas no cortex prefrontal e derivadas de lugares de Hipersensibilidade a DNasa I asociadas á SCZ Enriquecemento en TEA en xenes relacionados cos canles e transporte de Ca, e coa localización e función neuronal Xene <i>SETD1A</i> asociado con SCZ Xene <i>RAB2A</i> asociado con TEA

Táboa 7. Resumo dos resultados atopados mediante secuenciación de nova xeración en estudos de tríos. SCZ= Esquizofrenia; SCZA= Trastorno Esquizoafectivo., DI= Discapacidade Intelectual.

Os estudos de secuenciación de nova xeración publicados ata este momento amosan unha falta de potencia estatística para implicar xenes específicos e mutacións raras na esquizofrenia. Sen embargo, permítenos concluír que as variantes raras, xunto cas comúns, xogan un papel importante na susceptibilidade a padecer esquizofrenia aínda que a contribución relativa das dúas clases de variantes non está nada clara e permanece pendente de estudos de secuenciación dun tamaño moito maior ademais de ver a contribución daquelas variantes que están sitas fora do exoma.

1.3. MÉTODOS DE ANÁLISE DUN CONXUNTO DE XENES

Como xa se comentou anteriormente, os GWAS teñen por obxectivo identificar os mecanismos etiopatoxénicos das enfermidades complexas, as cales, a miúdo, son resultado dunha interacción de múltiples factores de riscos xenéticos e ambientais. Esta metodoloxía permite a identificación de variantes que explican só, individualmente ou en combinación con outras, unha pequena proporción da herdabilidade da esquizofrenia. Unha posible explicación a isto é que o gran número de polimorfismos xenéticos estudados nun GWAS e a elevada cantidade de tests levados a cabo provocan que asociacións débiles pero reais se perdan debido ó axuste por comparacións múltiples (Wang et al. 2011a). Unha das técnicas máis efectivas para solucionar este problema é o

Introdución

uso de análises baseadas en conxuntos de xenes xa que ó tratar ó grupo de xenes como un todo consegue unha redución da dimensión. Este tipo de test xorde da idea de que o risco xenético das enfermidades complexas é principalmente debido a múltiples xenes cun efecto pequeno ou moderado que actúan dunha forma modular, máis que por mutacións de xenes individuais. Un exemplo tipo de enfermidades complexas que pode ser estudada mediante esta metodoloxía é a esquizofrenia, xa que debido a súa natureza altamente polixénica pode extrapolarse que a unidade que confire o risco a padecela podería ser unha ruta metabólica, no canto de ser un único xene (Sullivan 2012). Unha das consideracións a ter en conta cando levas a cabo este tipo de análise é a heteroxeneidade de locus, é dicir, alelos de loci distintos causan unha enfermidade en poboacións distintas, a cal fai que aumente a dificultade de replicar un marcador cunha determinada enfermidade. Xa que os métodos baseados en conxunto de xenes utilizan información de distintos loci nunha unidade funcional pode producir resultados máis robustos e estables que a análise de xenes individuais.

Para levar a cabo este tipo de estudos é necesario tomar varias decisións. A primeira é elixir a estratexia a seguir para a elección do conxunto de xenes a estudar. Normalmente hai dúas estratexias:

- A primeira delas céntrase en conxuntos de xenes concretos. Estes defínense segundo unha hipótese a priori referente ó conxunto de xenes que probablemente estea involucrado na doenza. No noso traballo que comentaremos posteriormente a hipótese que barallamos é comprobar se o interactoma de *DISC1* xoga algún papel no risco a padecer a esquizofrenia.(Cantor et al. 2010).
- A segunda estratexia consiste en analizar tódolos grupos de xenes que están recollidos nas diferentes bases de datos. Debido a que este método utiliza tódolos datos xenómicos dispoñibles, permite identificar relacións inesperadas entre o conxunto de xenes e a doenza. Por outra banda, este método é computacionalmente moi intenso, xa que require estritas correccións por múltiples comparacións, problema que se minimiza considerablemente cando se estudan conxunto de xenes concretos ó reducirse a dimensionalidade (Ramanan et al. 2012).

O seguinte paso, e non menos importante, é seleccionar a base de datos máis axeitada para delimitar os xenes que forman o conxunto a estudar. Hai numerosas bases de datos que se utilizan para definir os grupos de xenes:

- ✓ "The Pathway Commons database" (Cerami et al. 2011), que é unha colección de datos de rutas biolóxicas públicas de diferentes organismos, que inclúe reaccións bioquímicas, ensamblaxes complexos, eventos de transporte e catálise, interaccións físicas nas que están implicadas proteínas, ADN, ARN, pequenas moléculas e complexos. Todo isto está recollido a partir de diferentes fontes e cunha linguaxe común e un formato estándar.
- ✓ "The Kyoto Encyclopedia of Genes and Genomes (KEGG) " (Kanehisa and Goto 2000) recolle rutas biolóxicas de referencia (mosaicos de varios organismos) e anotacións específicas de organismos, ademais de mapas de rutas biolóxicas ligados a xenes intimamente relacionados.
- ✓ Reactome (Croft et al. 2014) é unha base de datos de rutas biolóxicas que son recollidas mediante referencias cruzadas a PubMed, GO e outras fontes. Pode incluír datos de expresión e outros datos no mapa da ruta biolóxica.
- ✓ "The Gene Ontology" (Ashburner et al. 2000), é a maior base de datos cuxa estrutura é xerárquica (coméntase en detalle máis abaixo).

- ✓ Metacore (<http://thomsonreuters.com/metacore>) é unha base de datos que recolle rutas biolóxicas de enfermidades.
- ✓ Biocarta (www.biocarta.com) é unha base de datos que está dirixida mediante a entrada de usuarios con revisións expertas dalgunhas rutas biolóxicas. Proporciona interacción nas rutas biolóxicas en procesos celulares humanos.
- ✓ Molecular Signatures Database (MSigDB) (<http://software.broadinstitute.org/gsea/msigdb>), é unha base de datos que compila conxuntos de xenes de GO e KEGG, ademais de outras fontes como bandas citoxenéticas e rexións cromosómicas, conxunto de xenes descritos na literatura, motivos reguladores cis e xenes asociados e co-expresados en cancro.
- ✓ Pathway Interaction Database (PID) (Subramanian et al. 2005), contén un amplo rango de rutas celulares, centrándose especialmente na sinalización do cancro. Permite xerar un mapa de interaccións a partir dunha lista de xenes.
- ✓ Ingenuity (www.ingenuity.com) é unha gran colección de rutas metabólicas canónicas, sendo os mapas das rutas de elevada calidade.
- ✓ STRING (Jensen et al. 2009) recolle interaccións proteína-proteína incluíndo tanto interaccións físicas como funcionais. Integra información de numerosas fontes, incluíndo repositorios experimentais, métodos de predición funcional e coleccións de textos públicos, polo que actúa como unha base de datos que mapea tódalas evidencias de interaccións nun conxunto común de xenomas e proteínas.
- ✓ DAVID (Huang et al. 2007a) caracterízase por aumentar e integrar anotacións de outras bases de datos.
- ✓ PANTHER (Thomas et al. 2003) é unha base de datos que fai unha predición da función dunha proteína a partir dos datos de secuencia e de conservación.

As definicións que máis se utilizan para limitar un conxunto de xenes son:

- Rutas biolóxicas: Trátase dun conxunto de xenes (ou os seus produtos) que xuntos levan a cabo unha función biolóxica específica. Unha ruta biolóxica debe incluír información sobre as entidades moleculares ás que afecta, así como a interacción entre estas. Recentemente propúxose a existencia de 5 tipos de rutas biolóxicas: **ruta biolóxica molecular** que fai referencia a accións bioquímicas nunha molécula ou composto; **ruta biolóxica celular** que abarca a regulación de procesos celulares máis globais, como por exemplo apoptose ou división celular; outros dous tipos son a **ruta biolóxica de enfermidades** e a **rutas biolóxica de intervencións**, as cales describen procesos de enfermidades as primeiras e respostas a determinadas intervencións, tipo resposta a medicamentos, as segundas. Estes dous últimos tipos refírense a colección de xenes previamente asociados a un fenotipo e non están baseados en coñecementos biolóxicos. Finalmente está a **ruta biolóxica de sistemas ou órganos** que fai referencia a funcións fisiolóxicas de orde superior (Ramanan et al. 2012). No campo das enfermidades neuropsiquiátricas, como é o caso da esquizofrenia, podería ser importante definir manualmente unha ruta biolóxica utilizando o coñecemento experto, xa que as bases de datos públicas poderían non ter anotados adecuadamente as rutas biolóxicas relacionadas coa función neuronal (Wang et al. 2010a).

Un dos problemas das rutas biolóxicas é que representan modelos de procesos biolóxicos hipotéticos, non existindo unhas regras estandarizadas para construílos

polo que estes modelos poden variar en función da fonte de datos utilizada. Isto pode provocar diferenzas á hora de interpretar e comparar os resultados (Mitrea et al. 2013). Por outra banda, outro dos problemas é a natureza xerárquica das rutas biolóxicas que pode implicar que algunhas destas rutas poidan representar un subconxunto dunha ruta biolóxica de maior tamaño (Mooney et al. 2014).

- **Redes biolóxicas:** Estas redes non describen unha función ou proceso biolóxico que se leva a cabo nun contexto biolóxico específico, senón que describen a relación biolóxica (interacción predita ou observada) entre múltiples xenes ou os seus produtos. Un exemplo son as redes de interacción proteína-proteína (Mooney et al. 2014). A principal diferenza entre as rutas e as redes biolóxicas son que as rutas inclúen regulación direccional e bifurcacións, pero están dirixidas cara un resultado final. Por outra banda, os elementos nunha ruta están conectados mecanicamente, mentres que os elementos dunha rede están conectados ó través de relacións compartidas que poderían non levar a cabo unha acción. Polo tanto, as redes non están dirixidas dende un punto inicial a un resultado final. As redes poden estar divididas en: **sub-redes** nas que tódolos elementos están conectados a un nodo central, e **módulos** que amosan unha gran densidade de conexións (Ramanan et al. 2012). Estas redes de interacción son construídas utilizando diferentes técnicas (cromatografía de afinidade, inmunoprecipitación, *cross-linking*, etc.). Entre todas elas destaca a técnica de dobre híbrido de fermento que se basea na actividade transcricional para medir a interacción proteína-proteína. Depende da natureza modular dos activadores transcricionais específicos de lugar, os cales constan dun dominio de unión ó ADN e un dominio de activación da transcrición. O primeiro serve como diana do activador para un xene específico que será expresado e o segundo contacta con outras proteínas da maquinaria transcricional para que a transcrición teña lugar. O sistema de dobre híbrido está baseado na observación de que os dous dominios do activador non necesitan ligarse covalentemente e poden unirse mediante a interacción das dúas proteínas. A aplicación deste sistema require a construción de dous híbridos: Un dominio de unión ó ADN fusionado a unha proteína e un dominio de activación de transcrición fusionado a outra proteína. Estes dous híbridos son expresados nunha célula que conteña un ou máis xenes de interese. Se os híbridos interactúan, eles crearán un activador funcional mediante a unión do dominio de activación co dominio de unión ó ADN, podendo ser detectado mediante a expresión do xene de interese (Phizicky and Fields 1995). Un exemplo no que se utilizaba esta técnica para identificar unha rede de interaccións era o interactoma de *DISCI* publicado por Camargo e cols. (Camargo et al. 2007).
- **Gene Ontology:** Unha das formas máis comúns para definir o conxunto de xenes a estudar é a base de datos de Gene Ontology (Ashburner et al. 2000), a cal intenta describir as funcións xénicas usando tres categorías xerárquicas: funcións moleculares, procesos biolóxicos e compoñentes celulares. Aínda que determinados xenes estean asignados a unha categoría particular de Gene Ontology non significa que se coñezan as relacións ou interaccións entre estes xenes. Ó igual que as rutas metabólicas, un dos problemas desta categoría é a súa natureza xerárquica que pode producir un solapamento significativo entre categorías provocando un solapamento dos xenes e unha información redundante entre os conxuntos de xenes, o cal ten as súas implicacións tanto a hora de analizar os datos como na interpretación dos resultados (Mooney et al. 2014). A gran vantaxe de Gene Ontology é que produce un vocabulario controlado e dinámico para describir o papel dos xenes e os seus produtos en calquera tipo de eucariota (Ashburner et al. 2000).

Cabe resaltar que os resultados obtidos dun estudo usando un conxunto de xenes probablemente estean nesgados en función da base de datos a utilizar, de xeito que os resultados obtidos reflictan as diferenzas na información do conxunto de xenes incluída na fonte a utilizar e a maneira na que a información está organizada (Elbers et al. 2009). As recomendacións para intentar paliar isto é utilizar diferentes bases de datos para definir un único conxunto de xenes e incorporar un algoritmo para decidir se o conxunto de xenes está asociado cando os resultados son inconsistentes ó través das bases de datos (Cantor et al. 2010).

Unha vez elixida a base de datos e delimitada o conxunto de xenes, compre asignar os SNPs dos GWAS a un xene específico que pertence a un dos grupos elixidos. O método máis simple para mapear un SNP a un xene é mediante a súa localización, é dicir, asignalo a un xene que está a unha distancia específica ou simplemente ó xene máis próximo. Outro método de asignación de SNPs é aquel que ten en conta o DL, de xeito que se consideran tódolos SNPs que están en DL co xene en cuestión. Con esta estratexia o que se pretende é cubrir tódolos marcadores que están ligados á variante causal dentro do mesmo bloque de DL. A principal desvantaxe deste segundo método é que tamén inclúe SNPs irrelevantes, facendo que dilúan a forza do sinal do conxunto de xenes e tamén aumente drasticamente a carga computacional, especialmente cando un ruta biolóxica está formada por un gran número de xenes (Wang et al. 2011a). Ademais usar o DL para mapear os SNPs pode crear problemas cando un SNP está correlacionado con múltiples xenes. Se estes xenes están no mesmo conxunto o SNP pode ser contado varias veces, inflando os valores de asociación (Sedeno-Cortes and Pavlidis 2014).

A continuación hai que elixir o método de análise a usar. Ó longo de todos estes anos foron xurdindo gran variedade de métodos (Táboa 8), os cales se poden clasificar de diferente xeito. Un deles é segundo a estratexia para ter en conta a asociación a nivel de xene:

- Métodos dun paso: Aquí tódolos SNPs do conxunto de xenes son usados para calcular un nivel de asociación sen ter en conta o efecto a nivel de xene. Este tipo de métodos non son os adecuados cando un conxunto de xenes contén un número elevado de SNPs (máis SNPs que mostran)(Mooney et al. 2014).
- Métodos de dous pasos: O primeiro paso neste método é calcular o estatístico a nivel de xene, para logo agregalos todos para calcular un único estatístico para o conxunto de xenes. A forma máis simple é calcular o efecto de tódolos SNPs a nivel do xene, e logo seleccionar para cada un dos xenes o SNP co valor-P menor entre os SNPs mapeados dentro destes xenes. As principais desvantaxes deste método son: se un xene presenta varios sinais de asociación, agás un, o resto son desestimadas; ademais a puntuación de asociación dun xene depende da densidade e do tamaño do xene, definíndose o primeiro como o número de SNPs que hai nun xene e o segundo como o DL que existe entre estes SNPs. Isto fai que os xenes máis longos teñan máis probabilidade de albergar os valores de significación máis baixos. Este método pode inflar a taxa de erro tipo I para un conxunto de xenes que conteñan maior número de xenes longos. Este mesmo efecto ocorre con xenes de densidade elevada (Wang et al. 2011a; de Leeuw et al. 2016). Outra forma sería usar varios SNPs dentro dun mesmo xene para calcular o efecto a nivel xénico. Para isto hai dúas maneiras: Combinar os valores-P individuais de cada SNP ou o modelado de multi-SNP. No primeiro tódolos SNPs dun xene son individualmente testados para asociación e logo os tests estatísticos obtidos son combinados para obter medidas de asociación a

Introdución

nivel xénico. No segundo, o efecto a nivel de xene é calculado mediante un modelado conxunto de múltiples SNPs dentro do mesmo xene usando métodos de regresión múltiple. A principal vantaxe deste método respecto ó anterior é que se pode coñecer a contribución individual de cada xene na asociación do conxunto xénico. Isto é útil cando un conxunto de xenes está formado por un gran número xa que non tódolos xenes contribúen de igual forma na asociación (Mooney et al. 2014).

Por outra banda, segundo a hipótese nula, pódense clasificar como:

- Métodos competitivos: Estes métodos están baseados en identificar primeiro os SNPs (ou xenes) que están asociados cun trazo, para a continuación avaliar se os SNPs asociados significativamente tenden a agruparse nun conxunto de xenes predefinido. Estes métodos son competitivos xa que comparan o sinal da asociación entre os xenes nun conxunto co sinal de asociación de tódolos xenes fora dese conxunto. A hipótese nula que testan é que o nivel de asociación dentro do conxunto de xenes é igual ó nivel de asociación fora deste conxunto. Unha limitación destes métodos é a dicotimización dos resultados de asociación dos SNPs en significativo ou non significativo en base a un umbral de significación predefinido, o cal non ten en conta a forza da asociación (Fridley and Biernacka 2011).
- Métodos autónomos: Consideran só o efecto dentro dun conxunto de xenes de interese. O que trata de ver é se o conxunto de xenes está asociado co trazo a estudar, sendo a hipótese nula a de non asociación. A diferenza do anterior, non necesita datos de xenes distintos ós que pertencen o grupo analizado, nin é necesario un umbral de significación predefinido.

Por último, outro tipo de métodos son os métodos baseados en topoloxía: Utilízanse para analizar conxuntos de xenes derivados de fontes de datos que describen interaccións entre os xenes, como son as redes metabólicas. Estes métodos incorporan gráficos topolóxicos no procedemento para a puntuación do conxunto de datos. O único que necesitan estes métodos é unha lista da xenes e os valores de significación para cada un dos xenes. Teñen en conta non só o efecto individual de cada xene, senón tamén a relación entre os xenes e a importancia relativa de cada xene individual dentro do conxunto. Aquí non tódolos xenes dentro do conxunto son tratados do mesmo xeito, xa que, a priori, ós xenes cun papel máis crucial no proceso dáselle máis peso á hora de testar a asociación. Debido a estas diferenzas de peso non é posible comparar directamente os resultados deste tipo de análises cos comentados anteriormente (Mooney et al. 2014).

Un problema común a tódolos tipos de tests de análise de conxunto de xenes é a estratificación poboacional, xa que este efecto tenderá a varianza na distribución dos tests estatísticos, reducindo potencialmente a potencia para detectar o efecto real do conxunto de xenes. Ademais pode aumentar a taxa de erro tipo I se a estratificación infla desproporcionadamente a asociación dos xenes no conxunto de xenes. Normalmente este problema resólvese aplicando a corrección de estratificación estándar para os SNPs ou para os valores-P dos xenes utilizados do conxunto de xenes antes de levar a cabo a análise do conxunto de xenes (de Leeuw et al. 2016).

Tipo de método	Descrición	Exemplos
Métodos dun paso	Tódolos SNPs do conxunto de xenes son usados para calcular o nivel de asociación	SNP ratio test , INRICH
Métodos de dous pasos	Primeiro calcúlase o estatístico a nivel de xene, e logo agréganse para calcular un único estatístico	BGSAsnp, VEGAS
Métodos competitivos	Permuta a etiqueta dos xenes para cada unha das rutas biolóxicas e compara o conxunto de xenes da ruta biolóxica con outro conxunto que non pertence a esta	GenGen, ALIGATOR
Métodos autónomos	Permuta a etiqueta (fenotipo, por exemplo) e compara tódolos xenes dunha ruta biolóxica consigo mesmo (non considera o resto)	Plink set-based test, GSEA-SNP
Métodos topolóxicos	Utiliza a topoloxía da rede biolóxica para calcular o estatístico a nivel de xene. Asigna maior peso ós xenes máis relevantes	dm GWAS, Metacore

Táboa 8. Clasificación dos métodos e algún dos programas dispoñibles. Algún dos exemplos indicados poden estar clasificados dentro dalgunha das outras categorías.

Debido a gran variedade de métodos existentes para a análise dun conxunto de xenes, hoxe en día non hai un consenso sobre cal é a mellor maneira de resumir o efecto a nivel de SNP nun grupo de xenes, nin tampouco cal é a mellor forma de avaliar o nivel de significación estatística do efecto a nivel de conxunto de xenes. Ademais con esta gran variedade de metodoloxías resulta complicado comparar os resultados entre os estudos. Para intentar solucionar tódolos problemas comentados ó longo desta sección propuxéronse unha serie de retos e metas que mellorarán esta metodoloxía. Un deles sería desenvolver un método para examinar tódalas posibles rutas biolóxicas implicadas nun trazo, de xeito que un método non nesgado podería identificar a ruta biolóxica asociada. Este tipo de métodos son intensos computacionalmente polo que é todo un reto biolóxico. Outra mellora que provocaría un aumento na potencia estatística sería aquela na que un método lle dese máis peso a unha ruta metabólica que contivese varios SNPs asociados independentes. Xa que os alelos de risco para unha doenza complexa poden ser comúns ou raros, os métodos de análise de conxunto de xenes deberían estar deseñados para capturar ambos. Ademais, xa que dous alelos asociados nunha ruta metabólica poderían amosar epistase, os métodos deberían poder reflectir o seu efecto (Cantor et al. 2010). Estes métodos poden utilizar tanto datos xenotípicos crús para cada individuo ou unha lista dos valores-P relacionando un SNP ou un xene a unha doenza. Os primeiros poden incluír covariables e corrixir para o DL ó través de permutacións; polo contra, os que utilizan os valores-P teñen a gran vantaxe de que a distribución destes valores están moi accesibles a través doutros investigadores e poden ser xerados aplicando covariables. Iso si, é necesario corrixir para o DL baseado en poboacións de referencia (Ramanan et al. 2012). A pesares de todo isto, este tipo de metodoloxía ten un gran potencial para aumentar o coñecemento de determinadas enfermidades a partir dos datos dos GWAS.

1.4. MÉTODOS DE ANÁLISE DE VARIANTES RARAS

Os primeiros tests utilizados para a análise de variantes raras foron os tests de marcadores únicos, onde cada variante individual dun xene era testada para comprobar a asociación cunha doenza, utilizando tests estatísticos estándar univariáveis como o test X^2 , o test exacto de Fisher ou o test de tendencia Cochran Armitage, sendo a taxa de erro por familia (probabilidade de cometer falsos positivos ou erro tipo I) controlada mediante corrección por múltiples comparacións (Bonferroni ou permutacións) (Li and Leal 2008).

Introdución

Unha das alternativas ós tests individuais para estudar unha posible asociación é usar métodos multivariados. Estes combinan a información ó través das variantes e analizan simultaneamente as distintas variantes presentes. Os tests máis usados son o Método de Fisher, os Tests T^2 de Hotelling e a regresión loxística multivariable. Todos necesitan múltiples graos de liberdade, o cal reduce a potencia do test, especialmente cando só hai un único sinal forte na rexión (Asimit and Zeggini 2010). Outro problema con estes tests é que son sensibles ás frecuencias alélicas; por exemplo, un estudo demostra que o test T^2 de Hotelling vese amplamente afectado pola frecuencia do alelo menor, ademais de provocar unha redución da potencia cando o número de alelos causais raros aumenta (Li and Leal 2008).

Por outra banda, o desenvolvemento das tecnoloxías de secuenciación masiva permítenos avaliar maior número de variantes de frecuencia baixa e investigar o seu papel nas enfermidades complexas. Un dos retos neste campo fai referencia ós tests estatísticos necesarios para tratar estas variantes, xa que os tests de asociación clásicos teñen pouca potencia estatística, agás que o risco asociado á variante sexa elevado ou se analice unha mostra de gran tamaño. Esta escasa potencia dos tests dunha única variante cando se estudan variantes raras é debido a varias razóns: a primeira é que moi poucos individuos portan alelos raros nunha posición variante única polo que o sinal de asociación é baixo debido á baixa frecuencia. A segunda é que en presenza de heteroxeneidade alélica nun locus en distintos individuos afectados, hai unha redución da potencia posto que varios individuos contribuirán ó sinal en diferentes variantes no locus (Asimit and Zeggini 2010). A terceira e última é que as variantes raras son moito máis abundantes que as comúns e están moi pouco correlacionadas, provocando unha severa penalización a hora de corrixir por múltiples tests (Li et al. 2013a).

Outro dos aspectos que se ven afectados ó analizar variantes raras con tests simples é a taxa de erro tipo I. Debido a baixa frecuencia das variantes raras, a escaseza de datos pode facer que os resultados sexan erróneos para tamaños de mostra pequenos (Li et al. 2013a).

Para tratar de paliar estes problemas, desenvolvéronse métodos estatísticos especialmente configurados para a análise de asociación de variantes raras para aumentar a potencia. Son os chamados tests de agregación. Estes métodos agrupan variantes situadas nunha rexión de interese, que pode ser un xene, unha agrupación de xenes (por exemplo, varios xenes dunha mesma ruta metabólica), ou incluso nunha rexión arbitraria do xenoma definida mediante as posicións en pares de bases nun cromosoma. O que tratan é de testar o efecto da asociación do grupo das variantes como un todo, no lugar de testar cada variante de forma individual.

A principal vantaxe destes métodos é que producen unha redución da dimensión provocada pola agregación das múltiples variantes raras nunha única unidade de análise. Hoxe en día hai unha gran variedade de tests de agregación de variantes que se agrupan nas seguintes categorías:

1) Tests de Carga

O obxectivo destes tests é comprobar se hai un exceso de variantes raras en casos ou en controis. A metodoloxía que seguen é agregar a información de múltiples variantes nunha única puntuación xenética e comprobar se esta puntuación está asociada ó trazo a estudar. O método máis simple compara portadores de alelos raros nos casos fronte ós controis. Outra estratexia é contar o número de variantes raras que porta cada individuo nunha rexión determinada. Para plasmar o diferente

efecto xenético que probablemente teñan as variantes raras xorden os tests que utilizan a suma estatística ponderada, na que as variantes máis raras teñen máis peso, de xeito que os alelos menos frecuentes contribúen máis ó test estatístico (Li et al. 2013a).

Un test que está dentro desta categoría é o test CAST (*Cohort Allelic Sums Test*) que asume que a presenza de calquera variante rara aumenta o risco e analiza as diferenzas de número de mutacións nun grupo de variantes entre casos e controis mediante a codificación dos individuos en función do número de mutacións (Morgenthaler and Thilly 2007). O test CMC (*Combined Multivariate and Collapsing Test*) divide os marcadores en subgrupos en base a un criterio determinado (frecuencia alélica), para a continuación colapsar as variantes dentro de cada grupo, e avaliar o efecto de cada conxunto mediante un test T^2 de Hotelling (Li and Leal 2008). Outro utilizado é o test WSS (*Weighted-Sum Statistic*) que crea unha puntuación de xenotipo composto para cada un dos individuos. Esta puntuación é a suma dos alelos alternativos ponderados en función da frecuencia das variantes no grupo dos controis. Usa o test *Wilcoxon* de suma de rangos nas puntuacións dos xenotipos entre os grupos fenotípicos e utiliza permutacións para obter os valores-P (Madsen and Browning 2009). Mediante estas estratexias o que se consegue é reducir os graos de liberdade debido a que se agregan múltiples variantes raras nunha única unidade de análise e como consecuencia, prodúcese un aumento do sinal e unha redución da dimensionalidade (Li et al. 2013a). Os tests de carga asumen que tódalas variantes no conxunto a estudar son causais e están asociadas na mesma dirección co trazo a estudar. A inclusión dalgunha variante non causal provoca a redución drástica da potencia.

2) Tests de carga adaptativos

Para intentar paliar as limitacións dos tests de carga anteriormente comentados, desenvolvéronse varios métodos que son máis robustos cando existen variantes nulas, de risco ou de protección. A súa robustez é debida a que non require tantas suposicións sobre a arquitectura xenética que está por detrás de cada locus xenético. A maioría de tests están baseados en procedementos de dous pasos, e o feito de que algúns requiran o cálculo de coeficientes de regresión de variantes individuais no primeiro paso é, a miúdo, difícil e inestable para variantes raras. Ademais requiren que se fagan permutacións para o cálculo dos valores-P que, normalmente é computacionalmente custoso.

Un exemplo é o test aSum (*data-adaptative sum Test*) que primeiro estima a dirección do efecto de cada variante nun modelo marxinal para logo levar a cabo o test de carga coas direccións estimadas, ponderando a variante como $w=-1$ ou $w=+1$ (Han and Pan 2010). Por outra banda, o test EREC (*Estimated Regression Coefficient Test*) calcula un coeficiente de regresión para cada unha das variantes e utilízao para ponderar (Lin and Tang 2011). O Test VT (*Variable Threshold*) selecciona un umbral de frecuencia óptimo para un test de carga de variantes raras e estima os valores-P analiticamente ou por permutacións (Price et al. 2010). O test KBAC (*Kernel-based Adaptive Cluster Test*) combina a clasificación de variantes de risco e non-risco co test de asociación mediante o uso dunha ponderación que reflexa o tamaño aparente do efecto de cada variante individual (Liu and Leal 2010).

3) Tests de compoñentes da varianza

Estes métodos calculan a asociación avaliando a distribución do efecto xenético dun grupo de variantes. A gran vantaxe destes tests é que teñen en conta tanto variantes de risco como de protección, independentemente da dirección do efecto. Un exemplo é o test C-alpha, que asume como hipótese nula que a distribución de cada variante entre casos e controis segue unha distribución binomial. Este test detecta a desviación do número de variantes observadas nos casos e nos controis respecto á distribución binomial esperada. Unha vantaxe deste test é que permite incorporar no cálculo do test estatístico un peso a cada unha das variantes analizadas. Esta ponderación pode ser feita en función da frecuencia alélica de cada variante ou segundo as predicións bioinformáticas de funcionalidade (Neale et al. 2011). Outro test, que soluciona o problema da existencia de variantes que teñen un efecto en direccións opostas, é o Test SKAT (*Sequence Kernel Association Test*). O SKAT é unha forma xeneralizada do C-alpha. A diferenza do test C-alpha, o SKAT permite incluír covariables, permite a análise de trazos cualitativos e cuantitativos e o manexo de modelos xenéticos complexos, tales como o efecto epistático (Wu et al. 2011).

4) Test ómnibus que combina test de carga e test de compoñentes da varianza

Este tipo de test propúxose para combinar métodos que son poderosos para variantes co mesmo efecto e métodos que son robustos cando están presentes variantes non causais ou variantes con efecto oposto. Un exemplo é o test SKAT-O (*SKAT-Optimal Test*) o cal combina de forma lineal un test de carga e un test SKAT nun único método; este test permite o cálculo dos valores-P dunha forma analítica e cunha elevada precisión (Lee et al. 2012a). Outra opción é o método de Fisher que combina valores-P obtidos mediante SKAT e o test de carga e utiliza as permutacións para avaliar a significación do test (Derkach et al. 2013).

Posto que o funcionamento de cada un destes métodos depende da arquitectura xenética subxacente das enfermidades complexas, é difícil conseguir un test que sexa óptimo para cada un dos escenarios. A medida que se vaian clarexando novos aspectos sobre a arquitectura xenética das enfermidades complexas poderemos incorporalos a estes tests de asociación para aumentar a potencia e priorizar variantes para os estudos de replicación e para análises funcionais.



2. HIPÓTESES E OBXECTIVOS





2.1. HIPÓTESES

A susceptibilidade xenética a padecer esquizofrenia pode ser provocada por variantes que están a unha frecuencia baixa na poboación axustándose ó modelo de variante rara-enfermidade común.

Segundo o equilibrio mutación-selección, as variantes que maior efecto causan sobre a esquizofrenia son variantes pouco frecuentes na poboación.

Xa que os xenes non actúan dunha forma illada, é de esperar que as variantes implicadas na susceptibilidade a padecer esquizofrenia converxan en redes de interacción, polo que a acumulación de variantes comúns de sentido incorrecto no interactoma de *DISC1* poderían aumentar a o risco a padecer esquizofrenia.

Os xenes asociados á esquizofrenia identificados mediante estudos de GWAS ou de CNVs que afectan a un único xene poderían presentar tanto variantes de susceptibilidade con frecuencias comúns e pouco efecto como variantes raras cun risco elevado.

O risco de padecer esquizofrenia poder ser debido a acumulación de variantes raras nun único xene ou ben nun conxunto de xenes que pertencen a unha vía concreta.

2.2. OBXECTIVOS

No primeiro dos traballos presentados centrabámonos no xene *DISC1* coa finalidade de analizar SNPs non sinónimos cunha frecuencia menor ó 5% que causen, por si mesmos, un aumento a padecer esquizofrenia na poboación galega. Outro dos obxectivos era comprobar se, nesta mesma poboación, existía unha acumulación de variantes raras non sinónimas que fose significativamente maior en casos respecto ós controis en *DISC1*.

No segundo estudo presentado re-analizabamos datos dun estudo de asociación previo para ver se había unha sobre-representación do interactoma de *DISC1* co fin de ver se este complexo proteico podería estar implicado na patofisioloxía da esquizofrenia. Centrabámonos en variantes comúns de sentido incorrecto e intentabamos ver se esta vía presentaba unha acumulación de variantes de predisposición á enfermidade.

No terceiro traballo analizabamos as variantes identificadas en varios xenes asociados previamente á esquizofrenia mediante estudos de GWAS e CNVs, para ver se existía unha acumulación de variantes raras que causasen un incremento na predisposición a esquizofrenia considerablemente maior que as distintas variantes comúns de susceptibilidade descritas. Ademais avaliabamos o funcionamento de varios *softwares* de identificación de variantes utilizando os datos previos e os obtidos da secuenciación dos xenes.

Por último, intentabamos confirmar a acumulación de variantes raras codificantes de sentido incorrecto en controis no xene *NRXN1* mediante unha meta-análise, utilizando os datos obtidos da secuenciación de novas mostras e de buscas bibliográficas.



3. MATERIAL E MÉTODOS





3.1. MOSTRAS

A colección de mostras utilizada neste estudo consistía nun total de 1157 individuos. Desta mostra 580 eran pacientes diagnosticados con esquizofrenia por psiquiatras experimentados seguindo os criterios do DSM-IV. As mostras dos pacientes procedían dos distintos centros hospitalarios da área de Santiago de Compostela e todos tiñan orixe galega. Todos estes pacientes amosaban unha diagnóstico estable polo menos durante dous anos. Entre os afectos non existía ningunha relación de parentesco. Tódolos pacientes con trastorno esquizoafectivo ou esquizofreniforme eran excluídos. Dispoñiamos dun consentimento informado por escrito de tódolos doentes incluídos que nos permitía engadilos a este estudo.

Por outra banda, as restantes mostras que formaban a colección (577) son controis, os cales eran anonimizados e cuxa extracción de sangue foi feita no Centro de Transfusión de Sangue de Galicia. Co fin de evitar duplicados das mostras de sangue dos controis, estas eran recollidas nun espazo de tempo curto. Ademais estes donantes non recibían ningún tipo de remuneración, de xeito que era pouco probable que esta mostra estivera enriquecida en grupos con desvantaxes sociais, os cales poderían ter unha maior taxa de individuos afectos.

Os traballos de investigación facíanse baixo a última versión da declaración de Helsinki. Ademais tódolos estudos eran aprobados polo Comité Ético de Investigación Clínica de Galicia (CEIC), o cal é un organismo independente e acreditado pola Consellería de Sanidade.

No caso do estudo de *DISC1* (Estudo I) secuenciabamos mediante Sanger 47 casos e o mesmo número de controis. A media de idade das mostras secuenciadas era de 47 ± 11 anos, tanto para casos como para controis. O número de varóns entre os pacientes era do 79%. Os casos seleccionados para a secuenciación elixíanse tanto pola precocidade da aparición dos primeiros síntomas da enfermidade así como pola presenza de historia familiar. As idades as cales aparecían os primeiros síntomas dos casos estaban situadas entre os 14 e os 22 anos, cunha media de 18 anos. Os 47 controis seleccionados para este estudo concordaban cos esquizofrénicos en idade e sexo.

Para o traballo do interactoma de *DISC1* utilizabamos os xenotipos de 476 pacientes, dos cales o 68% eran varóns, e 447 controis, sendo o 49% varóns. Todo eles foron previamente xenotipados utilizando a plataforma de Affymetrix usando o chip *Affymetrix GeneChip® Human (Panels 1 & 2) 20K cSNP Kit*. Despois de aplicar ós SNPs os correspondentes filtros de calidade indicados no traballo de Carrera e cols. (Carrera et al. 2012), utilizábamnos 5100 *SNPs* de sentido incorrecto comúns para ver a relación críptica, é dicir, comprobar se existía algunha relación de parentesco entre as mostras, mediante cálculos de identidade por estado utilizando o *software* PLINK (Purcell et al. 2007). Ademais analizabamos 130 marcadores informativos de ancestralidade utilizados par a detección da estrutura poboacional empregando a ferramenta Structure v2.3.1 (Pritchard et al. 2000). Isto permitíanos desbotar 37 casos e 72 controis.

Para o estudo de secuenciación de nova xeración mediante o SOLiD4 (Estudo III) elixiamos 306 mostras da colección completa (153 casos de esquizofrenia e 153 controis). A selección do 70% dos casos facíamola en base ós datos do porcentaxe do éxito de xenotipado que se levaba a cabo utilizando o chip *Affymetrix GeneChip® Human (Panels 1 & 2) 20K cSNP Kit* no traballo de Carrera et al. (2012), como indicativo da calidade do ADN. O restante 30% seleccionábámolo entre aquelas mostras

que tiñan unha porcentaxe de éxito de xenotipado maior á media e que, ademais, presentaban unha aparición da enfermidade temperán e/ou historia familiar con psicose. Seguíamos estes dous últimos criterios co fin de enriquecer a mostra en fenotipos extremos. No conxunto final de casos había 62 mostras con presenza de historia familiar e unha idade media de aparición do trastorno de $23,4 \pm 7,22$, ademais o 70% eran varóns. En canto a mostra control, seleccionabamos aqueles controis que concordaban cos casos na porcentaxe do éxito de xenotipado, deste xeito tratábamos de evitar un nesgo metodolóxico debido ás diferenzas de calidade do ADN, sendo o 49% destes, homes.

Para o estudo do xene *NRXN1* (Estudo IV), escolliamos para secuenciar 384 pacientes esquizofrénicos e 384 controis que non eran previamente secuenciados co SOLiD4 no estudo previo, sendo o 67% e 49% varóns en casos e controis respectivamente. Os casos tiñan unha idade media de 55 anos e os controis 47.

	Nº Total Mostras		Nº homes		Nº mulleres		% homes		% mulleres		Idade Media	
	Ca	Co	Ca	Co	Ca	Co	Ca	Co	Ca	Co	Ca	Co
Estudo I	47	47	37	37	10	10	79	79	21	21	47 ±11	47 ±11
Estudo II	476	447	326	220	150	227	68	49	32	51	57 ±15	47 ±11
Estudo III	153	153	107	75	46	78	70	49	30	51	53 ±14	46 ±14
Estudo IV	384	384	258	190	126	194	67	49	33	51	55 ±16	47 ±11

Táboa 9. Resumo dos datos das mostras utilizadas en cada un dos estudos. A idade dos casos e controis era calculada segundo o ano 2015.

3.2. EXTRACCIÓN DO ADN

As extraccións do ADN facíamolas a partir de sangue total. As mostras, tanto de casos como de controis, eran recibidas en forma de sangue fresco sendo conxeladas e almacenadas a -80°C en tubos con anticoagulantes (EDTA). Para facelas extraccións utilizábamos o robot *Chemagic Magnetic Separation Module I* seguindo as instrucións do protocolo que indicaba Chemagen (Baesweiler, Alemaña) (Biopolymer Technologie Aktiengesellschaft <http://www.chemagen.com/>). Partíase dun volume inicial de sangue de 5 ml e facíamos roldas de extracción de 12 mostras. Unha vez extraído o ADN, este era cuantificado usando o espectrofotómetro Nanodrop, obtendo tanto valores cuantitativos, en forma de concentración, como cualitativos a través dos cocientes dos valores de absorbancia OD (260/280) e OD (260/230), permitíndonos valorar a pureza do ADN extraído.

3.3. SECUENCIACIÓN DE SANGER

3.3.1. Optimización da PCR (Polymerase Chain Reaction)

Os deseños dos cebadores para levar a cabo as PCRs facíámolos co programa Primer3Plus (Untergasser et al. 2007). A partir da temperatura de desnaturalización (T_m), calculada con este programa, levábamnos a cabo as distintas PCRs. O *kit*, o termociclador e as condicións da reaccións utilizadas eran as mesmas que se usaban para a secuenciación (descritas máis abaixo).

O rango de temperatura utilizado na fase da unión do cebador para a posta a punto da PCR variaba en función dos cebadores utilizados. O valor máis baixo deste rango era a do cebador coa T_m menor e a partir desta aumentabamos 5 °C, polo tanto probabamos a amplificar unha única mostra a seis temperaturas diferentes. No caso do exón 3 variabamos a concentración final de $MgCl_2$ con fin conseguir unha correcta optimización da PCR.

Os resultados destas amplificacións comprobábanse cargando as mostras no Agilent 2100 Bioanalyzer seguindo os pasos sinalados no protocolo que Agilent Technologies (Santa Clara, California) indicaba. Os resultados destes reflectían o tamaño do amplicón de interese e a concentración do ADN en función das distintas temperaturas utilizadas. A temperatura que seleccionabamos como T_m para a PCR da secuenciación era a que mostraba a maior concentración e o tamaño de amplicón esperado.

3.3.2. Amplificación do ADN

3.3.2.1. DISC1 (ESTUDO I)

Neste estudo secuenciabamos tódolos exóns do xene *DISC1*, concretamente os 13 exóns da isoforma Lv (NM_001012957.1), o exón 3 da isoforma Es (NM_001012958.1), o exón 10 da isoforma S (NM_001012959.1) e o exón 11 da isoforma L (NM_018662.2). Isto era porque estes exóns destas isoformas son os únicos que son diferentes ós da Isoforma Lv. En total secuenciabamos 8509 bases codificantes.

Para levar a cabo a PCR utilizabamos o *kit* de Promega (Wisconsin, USA), *Gotaq® Flexi ADN Polymerase*. Os exóns de *DISC1* eran amplificados nunha reacción cun volume final de 25 μ L (Táboa 10), empregando un termociclador *Veriti 96 wells* de Applied Biosystems (Carlsbad, California) utilizando unhas condicións determinadas para cada un dos amplicóns (Táboa 11) Os cebadores usados para esta amplificación (Táboa 12) eran normalizados á concentración requirida.

Reactivos	Concentración inicial	Concentración final	V (μL)/ mostra
Mestura de nucleótidos para PCR (dNTPs)	5 mM	0,2 mM	1 μL
5X Colorless Gotaq® Flexi buffer	5X	1X	5 μL
Solución MgCl ² ^a	25 mM	2 mM	2 μL
Cebador <i>forward</i>	10 μM	0,6 μM	1,5 μL
Cebador <i>reverse</i>	10 μM	0,6 μM	1,5 μL
Gotaq® ADN polimerasa	5 u/μL	0,025 u/μL	0,125 μL
ADN	25 ng/μL	2 ng/μL	2 μL

Táboa 10. Reactivos e volumes utilizados para a PCR.

^aNo caso do exón 3 de *DISC1* a concentración final da solución de MgCl² era de 3 mM.

Paso	Temperatura	Tempo	Número de ciclos
Desnaturalización Inicial	94 °C	2 minutos	1 ciclo
Desnaturalización	94 °C	30 seg.	30 ciclos
Anelamento	58-65 °C ^a	45 seg.	
Extensión	72 °C	1 minuto	
Extensión final	72 °C	7 minutos	1 ciclo
Final	4 °C	Indefinido	1 ciclo

Táboa 11. Condicións do termociclador para PCR.

^a As temperaturas neste paso variaban en función do exón a secuenciar (Táboa 12).

Material e Métodos

Exón	Cebador	Secuencia	Tamaño produto	Tm
1	F	GGAAGGAGCAGGAGGCAG	207	62 °C
	R	ACCTACGCGACTTCCTGG		
2.1 ¹	F	TTCTCCAGATGCAGTTCCAG	620	64 °C
	R	GCACTGTGAGAGCCAGGAG		
2.2 ^a	F	GATGGAGCAAGGCGTGTC	614	65 °C
	R	CTGCCAGTGCTAATCTGAG		
3	F	AAAAGAGAATGAAGAAGTTCAGTTTTG	248	59 °C
	R	AAAGATGGAAAGAAAATTGGGAC		
3a ^b	F	CTGTTAGTCCCAGTCAAGGATG	194	62 °C
	R	TCATGATCTACGCTGTGTGG		
4	F	CACCGGGGTTATCTATTTTGC	280	65 °C
	R	TGTGGGAGCTGAGAGGTAGG		
5	F	GGCCACTTGCTGGAGTTTC	329	64 °C
	R	GCAACAGAGCAAGACCCTG		
6	F	CAGAGCAGTTTGCCATGAG	353	59 °C
	R	CCAATGAACAGGTCAAAGAGG		
7	F	CAGTGTCCTAACTGTAGTGGTATTG	191	61 °C
	R	AAAATCAAGTCAAACGAAACTTC		
8	F	ATCTCTGACCTGGCTGTCC	265	62 °C
	R	TTCATCGAGGACAAAACACG		
9	F	TCTCCATGTGTGTGGATGC	323	62 °C
	R	ATTCTGCCTGGCCACAAAAC		
10	F	TCAATCCTTTGGCTTTGAGC	305	62 °C
	R	CACTGATGGGGTTCAAAGG		
10a ^c	F	GGCTCCAGGCACTTTCG	214	60 °C
	R	ACATCATGTCTCTTCAATGCC		
11	F	CAAGCTATCGACTTGGTATGATG	405	62 °C
	R	TCTGTCACCTCCAGCACAG		
11a ^d	F	GCCAGGTAGACAAGCTATCG	405	58 °C
	R	GTCCCAAAGCACCCCTTAG		
12	F	ACTGAGGTGCAGAACTGGG	317	64 °C
	R	TGTTGCCATCTTCTGAGGC		
13	F	AGTACCCATCTGCTTCCAGC	325	64 °C
	R	CATCTTAGCTATGGCGGGG		

Táboa 12. Secuencia dos cebadores (*forward* (F) e *reverse* (R)) utilizados, xunto coa Temperatura de desnaturalización e o tamaño do exón amplificado.

^a O exón 2 debido a súa gran lonxitude era fragmentado en dúas partes (2.1 e 2.2)

^b Exón 3 da isoforma Es.

^c Exón 10 da isoforma S.

^d Exón 11 da isoforma L.

Por problemas de inespecificidade á hora de amplificar o exón 1 cunha PCR convencional, levabamos a cabo unha PCR *Touchdown* (PCRtd) (Táboa 13), xa que esta permitía aumentar o rendimento, a sensibilidade e a especificidade sen a necesidade de prolongar a optimización nin de redeseñar os cebadores.

	Paso	Temperatura	Tempo	Número de ciclos
Fase 1	Desnaturalización Inicial	95 °C	3 minutos	1 ciclo
	Desnaturalización	95°C	30 seg.	12 ciclos
	Anelamento	62+10 °C ^a	45 seg.	
	Extensión	72 °C	1 minuto	
Fase 2	Desnaturalización	95 °C	30	25 ciclos
	Anelamento	62 °C	45 seg.	
	Extensión	72 °C	60 seg.	
Fase 3	Extensión final	72 °C	7 minutos	1 ciclo
	Final	4 °C	Indefinido	1 ciclo

Táboa 13. Programa do termociclador para levar a cabo a PCRtd.

^a Para o anelamento partíase da T_m aumentada 10 °C e en cada ciclo diminúese un grado.

3.3.2.2. Confirmación de SNVs identificadas co SOLiD4 System (Estudo III)

Tres das variantes atopadas con esta tecnoloxía intentabámolas confirmar usando secuenciación tradicional (Táboa14), xa que a ferramenta *Assay Design v4* de Sequenom non era quen de deseñar os cebadores. Unha destas variantes era P208A que estaba situada no xene *NRXN1*. As outras dúas (L252F e S721C) estaban situadas no xene *DISC1*.

SNV	Xene	Exón	Cebador	Secuencia	Tamaño produto (pb)	T _m (°C)
L252F	<i>DISC1</i>	2.2	F	GATGGAGCAAGGCGTGTC	614	65
			R	CTGCCAGTGCTAATCTGAG		
S721C	<i>DISC1</i>	13	F	AGTACCCATCTGCTTCCAGC	325	64
			R	CATCTTAGCTATGGCGGGG		
P208A	<i>NRXN1</i>	1	F	TTACCTTGGCTGCAGTCCTT	223	64
			R	GTGACGTGAGGGTCAACTCC		

Táboa 14. Cebadores e T_m utilizados nas PCRs para a confirmación das SNVs.

3.3.3. Secuenciación e análise das secuencias

O produto da PCR comprobabámolo mediante unha electroforese nun xel de agarosa ó 1% vendo se o exón amplificado era o correcto. A continuación purificabamos o produto da amplificación utilizando o *kit ExoSAP-IT®*, o cal contiña *Exonucleasa I* e *Shrimp Alkaline Phosphatase* diluídos nun tampón especialmente formulado. O volume final da reacción era de 3 µL (0,85µL de ExoSAP-IT® e 2,15µL de produto de PCR). A reacción era introducida nun termociclador (*Veriti 96 wells* de **Applied Biosystems** (Carlsbad, California)) para levar a cabo dous tipos de incubación

Material e Métodos

(Táboa 15); a primeira para degradar os cebadores e os nucleótidos sobrantes, e a segunda para inactivar a encima ExoSAP-IT.

Paso	Temperatura	Tempo	Número de ciclos
1º Incubación	37°C	15 minutos	1 ciclo
2ª incubación	80 °C	15 minutos	1 ciclo
Final	4 °C	Indefinido	1 ciclo

Táboa 15. Condicións do termociclador usadas para a purificación coa ExoSAP-IT®.

Unha vez purificados, secuenciabamos os distintos exóns utilizando o *kit ABI Prims® BygDye® terminator v3.1 Cycle sequencing*. Para esta reacción utilizabamos o termociclador *Veriti 96 wells fast Thermal Cycler* de **Applied Biosystems** (Carlsbad, California) (Táboa 16).

Paso	Temperatura	Tempo	Número de ciclos
Fase I	96 °C	1 minuto	1 ciclo
Fase II	96 °C	10 seg.	25 ciclos
Fase III	50 °C	5 seg.	
Fase IV	60 °C	1:15 minuto	
Final	4 °C	Indefinido	1 ciclo

Táboa 16. Condicións do termociclador usadas para a reacción de secuenciación.

Para a lectura da secuencia das mostras usabamos o *3730xl ADN Analyzer* de Applied Biosystems. Os datos das secuencias eran importados como ficheiros ABI ó *Software Seq Scape® Version 2.5* (Applied Biosystems). Os resultados da lectura das secuencias eran aliñados coas secuencias de referencia, as cales eran obtidas a partir de *Ensembl Genome Browser* (<http://www.ensembl.org/index.html>). Todas as variantes atopadas eran revisadas visualmente ademais de ser confirmadas por medio dunha secuenciación bidireccional.

3.4. SECUENCIACIÓN DE NOVA XERACIÓN

3.4.1. Secuenciación de nova xeración mediante SOLiD4 System (Estudo III)

3.4.1.1. Selección dos xenes

Neste estudo seleccionabamos 21 xenes para a súa secuenciación. Cinco deles (*ZNF804A*, *TCF4*, *NRGN*, *VRK2* e *RELN*) eran seleccionados baseándonos na asociación positiva atopada en GWAS previos sobre esquizofrenia (O'Donovan et al. 2008; Shifman et al. 2008; Stefansson et al. 2009; Steinberg et al. 2011a). *SLC39A8* era seleccionado por presentar unha asociación significativa nun GWAS previo utilizando SNPs codificantes de sentido incorrecto levado a cabo polo noso grupo (Carrera et al. 2012). O xene *SLC39A3* seleccionabámolo pola hipótese xerada a partir dos resultados obtidos do GWAs do noso grupo anteriormente citado, o cal postulaba que a homeostase dos metais no cerebro podería estar involucrada na patoloxía da

esquizofrenia, así como noutros trastornos psiquiátricos (Baum et al. 2008; Ollila et al. 2009). *CNTNAP2* e *NRXN1* seleccionábanse pola súa asociación coa esquizofrenia baseada en datos de CNVs (Friedman et al. 2008; Vrijenhoek et al. 2008). A maiores escolliamos un conxunto de xenes formados por *DISC1* e por varias proteínas que interactuaban con este. *MACF1*, *KIF3A*, *DST* e *AKAP9* eran seleccionados como consecuencia dos resultados da análise de asociación dun conxunto de xenes levada a cabo no estudo II presentado nesta tese (Costas et al. 2013). Outros xenes que formaban parte deste conxunto eran *FEZ1*, *NDE1*, *NDEL1*, *PCMI*, *PDE4B*, *SPTAN1* e *TNIK* que seleccionabamos pola súa relevancia segundo a bibliografía (Millar et al. 2003; Brandon and Sawa 2011; Kamiya et al. 2012).

3.4.1.2. Preparación das mostras

As mostras seleccionadas para este proxecto eran purificadas con glicóxeno e acetato de amonio para evitar que calquera tipo de impureza afectara ós procesos posteriores. Tódalas mostras eran cuantificadas empregando o Nanodrop, o cal nos permitía, ademais, ver a súa pureza mediante as razóns OD (260/280) e OD (260/230). Considerabamos unha mostra presentaba unha boa calidade se o valor da OD (260/280) estaba entre 1,8-2 e o da OD (260/230) estaba entre 1,8-2,2. Ademais, as mostras non debían estar fragmentadas así que cada unha delas era corrida nun xel de agarosa ó 0,8% con marcadores de alto peso molecular para ver a súa integridade. Se as mostras cumprían estas dúas premisas entón cuantificábanse mediante fluorimetría utilizando como fluorocromo Picogreen e como lector Qubit (Invitrogen, Carlsbad, CA, USA). Cada mostra medíase por triplicado e a continuación normalizábase a 50 ng/ μ L en función da media das 3 medidas anteriores, nun volume de 45 μ L. Unha vez feita esta, levabamos a cabo unha nova cuantificación con Qubit para comprobar que a concentración fose a correcta. A continuación mesturabamos as mostras para facer *pools*, que estaban formadas por tres mostras nas que o volume final era de 120 μ L e a concentración de 50 ng/ μ L. As mostras que formaban unha *pool* tiñan igual ou similar porcentaxe de éxito de xenotipado, valor que era obtido a partir dos datos previos do xenotipado das mostras co *chip* de Affymetrix GeneChip® Human (Panels 1 & 2) 20K cSNP Kit (Carrera et al. 2012). En total faciamos 51 *pools* de casos e 51 de controis (Fig.10).

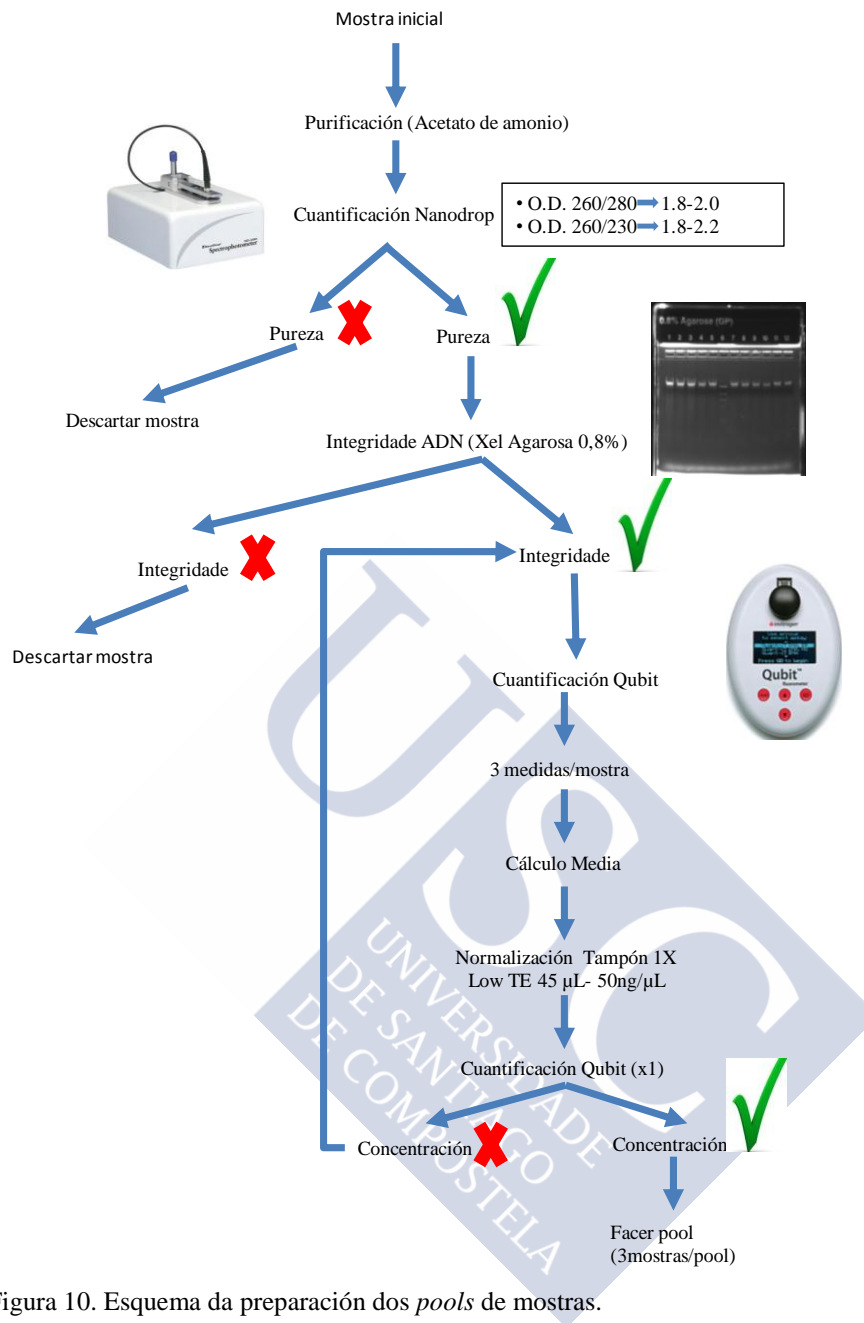


Figura 10. Esquema da preparación dos *pools* de mostrás.

3.4.1.3. Preparación das bibliotecas

Para a preparación das bibliotecas utilizabamos o *kit SureSelect Custom Target Enrichment System* (Agilent Technologies, Santa Clara CA, USA) (Fig.11). Este sistema permite capturar as rexións de interese mediante hibridación en solución utilizando sondas de ARN biotinizadas dunha lonxitude de 120 pares de bases. As sondas eran deseñadas de xeito que cada secuencia fora capturada como mínimo por dúas sondas. Para isto usabamos a ferramenta *online* eArray (<https://earray.chem.agilent.com/earray/>). Neste caso, o tamaño da rexión enriquecida era de aproximadamente 200 kb. A tecnoloxía a utilizar era a secuenciación co SOLiD 4 System (Applied Biosystems, Carlsbad, CA, USA) que permitía a secuenciación masiva e en paralelo de fragmentos de ADN mediante a secuenciación por ligación co método *Pair-End Sequencing* 50 + 35 bp.

A preparación das bibliotecas comezaba cunha fragmentación mediante sonicación do ADN co fin de conseguir un tamaño de fragmento comprendido entre 150 e 180 pb. Logo reparabamos os extremos do ADN para posteriormente engadir uns adaptadores a cada un deles. Estes adaptadores eran secuencias universais complementarias ós cebadores utilizados posteriormente na amplificación das bibliotecas e durante a secuenciación. Ademais, estes adaptadores levaban incorporada unha secuencia específica que actuaba como código de barras de cada pool de mostras, de xeito que, trala preparación das bibliotecas, podíamos mesturar e secuenciar varias mostras a un tempo. Unha vez engadidos os adaptadores seleccionabamos o fragmento de ADN co tamaño necesario, utilizando unha electroforese nun xel de agarosa. Unha vez amplificado o ADN faciamos a captura das rexións de interese por hibridación das sondas de ARN descritas previamente utilizando o *kit* anteriormente comentado (Fig.11). Trala hibridación, as rexións de interese eran capturadas utilizando bolas magnéticas marcadas con estreptavidina.

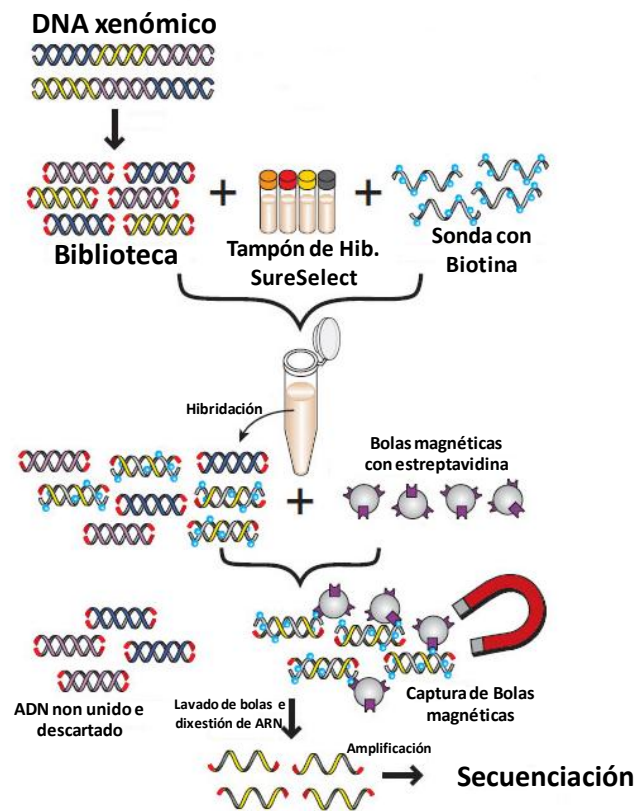


Figura 11. Esquema do sistema de captura co *kit SureSelect Custom Target Enrichment System*.

Unha vez construídas as bibliotecas e enriquecidas as rexións de interese, estas eran amplificadas mediante unha PCR en emulsión (ePCR) previa a secuenciación (Fig.12). Este tipo de PCR consta dunha fase acuosa, formada polos reactivos necesarios para a PCR, e unha oleosa na que, tras unha axitación, fórmanse gran cantidade de pequenas gotas que actúan como microrreactores que conteñen os reactivos necesarios para que a PCR teña lugar, producíndose unha amplificación en paralelo. Un dos cebadores utilizados na ePCR, complementario a un dos adaptadores introducidos durante a preparación das bibliotecas, está unido á superficie dunhas bolas magnéticas, de xeito que a PCR só teña lugar se nun destes microrreactores coinciden un fragmento de ADN e unha bola magnética (Fig.12), xerándose así unha bola recuberta por moitas copias dun único fragmento.

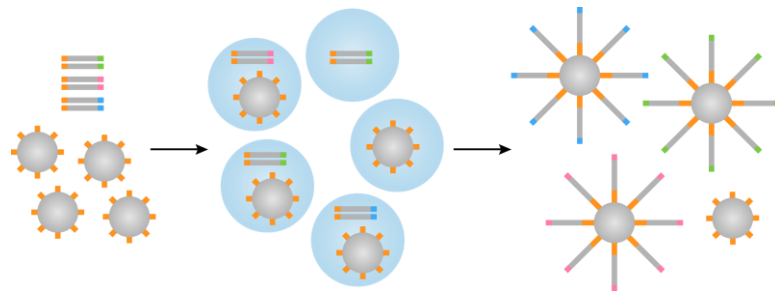


Figura 12. Esquema dunha ePCR

Trala amplificación, cuantificabamos os produtos da ePCR e faciamos un enriquecemento dos mesmos co fin de eliminar todas aquelas bolas magnéticas que non contivesen produto amplificado. Para isto, primeiro tiñamos que romper a emulsión utilizando butanol, de xeito que conseguíamos separar a fase oleosa da fase acuosa que contiña os produtos da PCR. A continuación, utilizando unhas bolas de enriquecemento e mediante unha centrifugación en gradiente de glicerol, conseguíamos separar as bolas co produto amplificado daquelas que non o tiñan. Isto era debido a que estas bolas de enriquecemento contiñan un cebador complementario ó cebador do extremo libre dos produtos amplificados, de xeito que unha destas bolas uníase a varias bolas co produto amplificado formándose un aglomerado. Isto permitíanos separar as bolas co produto amplificado do resto xa que se situaban no sobrenadante do glicerol unha vez centrifugado (Fig.13).

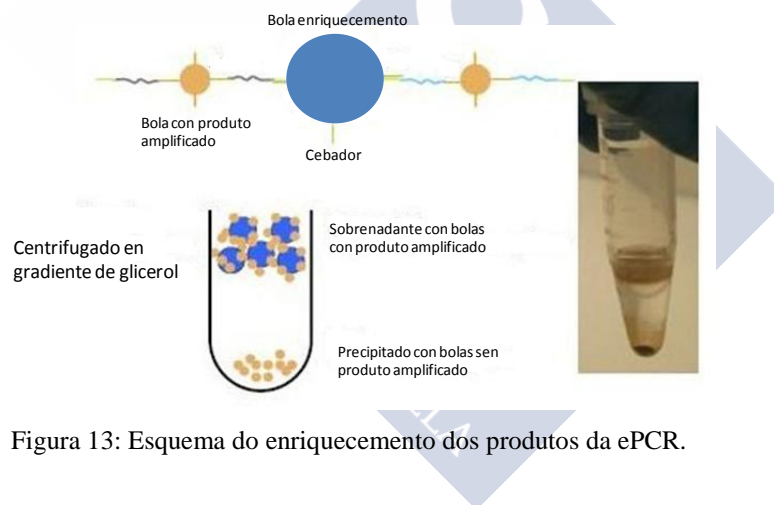


Figura 13: Esquema do enriquecemento dos produtos da ePCR.

3.4.1.4. Secuenciación mediante SOLiD4

Tras depositar estas bolas magnéticas cos fragmentos amplificados sobre un portaobxectos, procedíase a levar cabo a secuenciación por ligazón no SOLiD4. Dado o tamaño das bibliotecas e a capacidade do SOLiD4 secuenciabamos varias mostras en paralelo usando códigos de barras, polo que o portaobxecto a utilizar era o que dispón dunha configuración en octeto. Debido ó número de mostras secuenciadas só utilizabamos 7 cuadrantes do portaobxecto, contendo cada un unha mestura de 15 *pools*, os cales eran incorporados ás mostras tralo enriquecemento mediante a captura por hibridación (Fig.14). Para evitar calquera tipo de nesgo á hora da secuenciación, cada cuadrante do octeto estaba formado por 7 *pools* de casos e 8 *pools* de controis ou 8 de casos e 7 de controis.

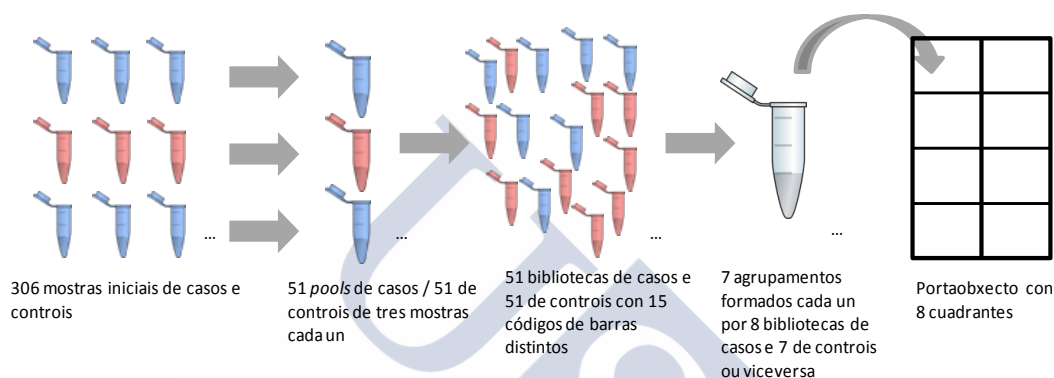
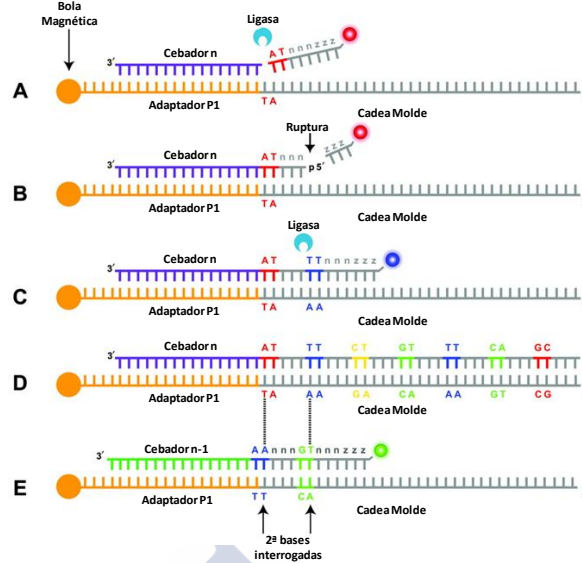


Figura 14. Esquema da preparación das *pools* e deposición no portaobxecto.

Como se comentou previamente, a secuenciación co SOLiD4 trátase dunha secuenciación por ligazón co método *Pair-End Sequencing* 50 + 35 bp, o que permite obter lecturas de 50 pb de lonxitude no sentido *forward* e lecturas de 35 pb no sentido *reverse*. Para isto utilízase ADN ligasa para ir incorporando sondas a un cebador (Fig. 15A). En primeiro lugar un cebador universal, complementario a un adaptador incorporado durante a preparación das bibliotecas, serve de ancoraxe ás sondas que se van ir unindo por medio de reaccións sucesivas de ligazón. Unha mestura de sondas de 8 pb de lonxitude, marcadas no extremo 5' cun fluorocromo, conteñen tódalas combinacións posibles de bases nas posicións 1 a 5, as cales interrogan a secuencia molde. Só se une ó cebador universal á sonda que contivese as 5 primeiras bases complementarias á secuencia molde. O fluorocromo do extremo 5' está determinado polas dúas primeiras bases da sonda. Unha vez unida a sonda capturase unha imaxe e elimínase o extremo das bases 6 a 8 xunto co fluorocromo, para que así teña lugar o seguinte ciclo de ligazón no que se unirá o seguinte segmento tres posicións máis alá do que o fixo o anterior. Así o que se consegue é determinar a secuencia de forma salteada. Isto repítese ó longo de toda a rexión a secuenciar. Para completar a secuencia é necesario o uso de 5 combinacións distintas de cebadores que defiren entre si no punto de unión á cadea molde (cebador n, cebador n-1, ..., cebador n-4) (Fig. 15B). Isto permite combinar a información dos distintos cebadores e inferir a secuencia da rexión a analizar, ademais de que cada base está interrogada dúas veces, o que permite diminuír a taxa de erro.

A)



B)

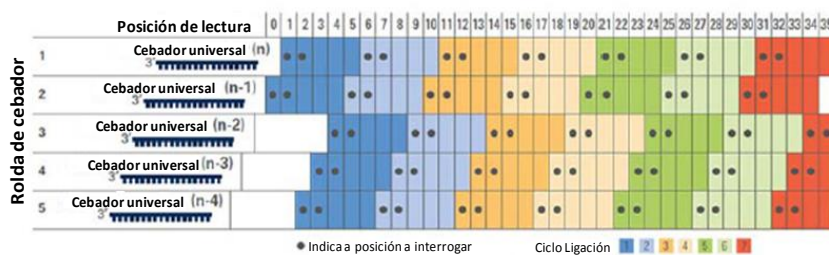


Figura 15. Esquema de secuenciación co SOLiD4. A) Figura da secuenciación por ligación. B) Esquema das 5 posibles combinacións de cebadores necesarios para completar a secuenciación

3.4.1.5. Análises bioinformáticas

Unha vez obtidas as lecturas da secuencias co SOLiD4, estas eran mapeadas contra o xenoma humano de referencia (hg19, GRCh37), mediante un algoritmo contido dentro do paquete informático BioScope (Applied Biosystems) e usando os parámetros por defecto, desta forma obtiñamos os aliñamentos en formato .BAM.

Unha vez obtidos os ficheiros dos aliñamentos utilizabamos 3 paquetes informáticos para a identificación das SNVs:

- **BioScope** (*Applied Biosystems*) por medio da ferramenta diBayes/small INDEL, modificando dous dos parámetros: "Call.stringency" que por defecto estaba a nivel Medium e nos elevámolo a nivel High (require que haxa lecturas do alelo alternativo nas dúas cadeas); e a opción "Het.min.allele.ratio" que tiña un valor por defecto de 0,15 pero que nos diminuíamos polo feito de utilizar *pools* de 3 mostras fixándoo a 0,025.
- **GATK** (DePristo et al. 2011): Esta ferramenta partía do mapeado e aliñado feito co BioScope. O primeiro paso era eliminar os duplicados mediante o Picard Tools. A continuación faciamos un realiñamento local e un recalibrado da

calidade de mapeo e por último, identificabamos as variantes coa ferramenta *GATK Unified Genotyper*, obténdoas nun ficheiro VCF.

- **VarScan** (Koboldt et al. 2009; Koboldt et al. 2012): Esta ferramenta é específica para traballar con *pools*. Parte de ficheiros con formato .PILEUP, os cales xerabamos a partir dos ficheiros .BAM utilizando o programa Galaxy (Giardine et al. 2005; Goecks et al. 2010). O único parámetro que modificábase para a identificación de SNVs nesta ferramenta era o umbral de frecuencia mínima do alelo alternativo, fixándoo nun 5%.

A anotación inicial das SNVs facíamola con Annovar (Wang et al. 2010b), aínda que tamén utilizabamos a ferramenta Emsembl Variant Predictor (Kircher et al., 2014) para obter a anotación funcional mediante a ferramenta CADD (*Combined Annotation Dependent Depletion*) da Universidade de Washington (<http://cadd.gs.washington.edu/score>).

3.4.2. Secuenciación de nova xeración de *NRXN1* mediante Ion Proton™ System (Estudo IV)

3.4.2.1. Preparación das mostras

Cada unha das mostras seleccionadas para este proxecto era cuantificada por triplicado de forma independente mediante fluorimetría utilizando como fluorocromo Picogreen e como lector Qubit (Invitrogen, Carlsbad, CA, USA) para posteriormente normalizarse a 50 ng/μl en función do valor medio de cuantificación. A continuación preparabamos *pools* de 4 mostras, sendo o número final 96 *pools* de casos e o mesmo número de *pools* de controis, mesturando a mesma cantidade de ADN de cada unha das mostras. A concentración final de cada *pool* era de 10 ng/μl e un volume de 30 μl.

3.4.2.2. Preparación das bibliotecas

Para a preparación das bibliotecas usabamos o *kit Ion AmpliSeq™ DNA Library Preparation* (Life Technologies) partindo dunha cantidade inicial de ADN de 10 ng. A diferenza da secuenciación levada a cabo no SOLiD4, neste caso as rexións de interese eran capturadas mediante unha PCR, para a cal usabamos cebadores para amplificación das rexións exónicas do xene *NRXN1*. Concretamente, amplificábanse mediante dúas PCRs, unha delas cunha *pool* de 39 amplicóns e a outra con 38. Os cebadores da captura eran deseñados utilizando o *software Ion AmpliSeq™ Designer* (Life Technologies) (Fig. 16). O tamaño da rexión capturada era de 10,7 kb que incluía os exóns e as rexións flanqueantes, capturando un 98,95% do total da rexión de interese.

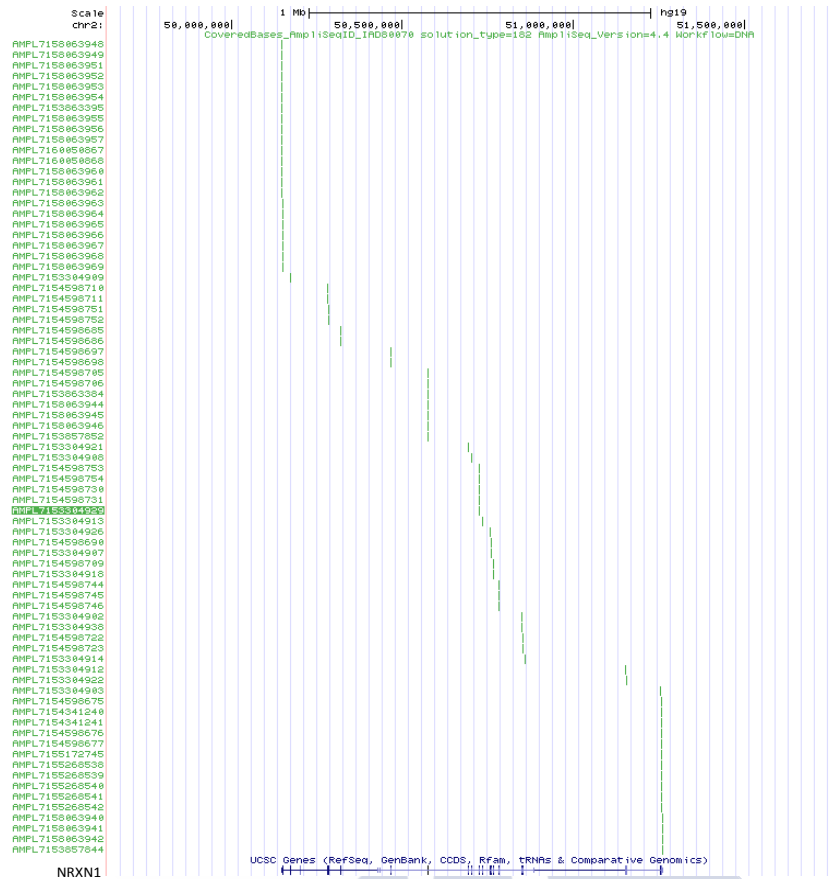


Figura 16. Imaxe do *Genome Browser* da UCSC onde se representa o deseño dos 77 amplicóns para a captura das rexións codificantes do xene *NRXN1*.

Despois da PCR levábase a cabo unha dixestión parcial dos cebadores, para evitar que fosen secuenciados, e a fosforilación dos extremos. A continuación, mediante unha ligación, engadíamose ós amplicóns uns adaptadores que contiñan uns códigos de barras que nos permitían diferenciar as mostras. Neste caso usábase 96 códigos de barras (Fig.17). Logo purificábase as bibliotecas coas bolas magnéticas AMPure® XP. Unha vez purificadas as bibliotecas, cada mostra era cuantificada mediante unha qPCR (PCR cuantitativa) co *kit Ion Library Quantitation* sen necesidade dunha amplificación adicional. Logo, en función dos valores obtidos, normalizábase de forma equimolar as mostras obtendo dúas *pools*, estando cada unha delas formada por 96 mostras.

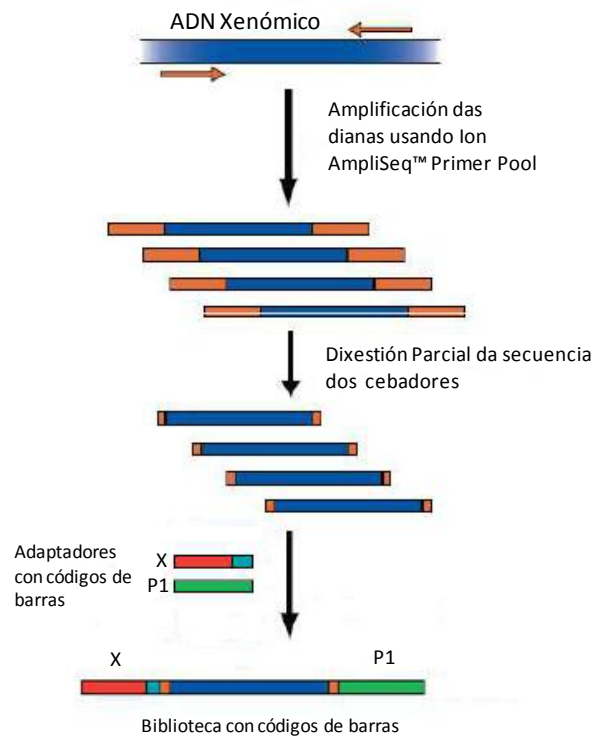


Figura 17. Esquema da preparación de bibliotecas co *kit Ion AmpliSeq™ DNA Library Preparation*.

Unha vez mesturadas as dúas *pools* utilizabamos o Ion Chef para levar a cabo:

- Unha amplificación clonal mediante unha ePCR (explicada anteriormente).
- Un enriquecemento das mostras coa finalidade de quedarnos unicamente coas esferas que conteñen ADN, para o que utilizabamos as *Ion Sphere™ Particle Enrichment* (Fig.18).
- A carga das mostra enriquecidas en dous *chips* (*Ion PI™ Chip*).

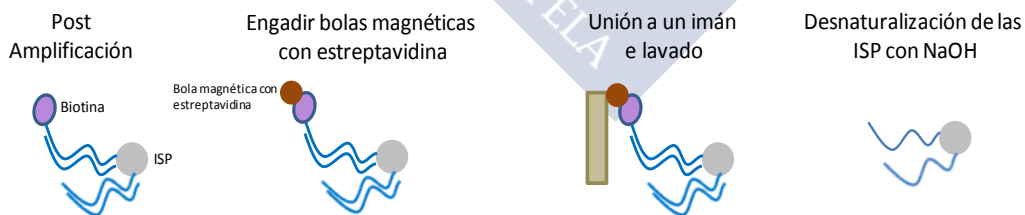


Figura 18. Esquema do enriquecemento coas *Ion Sphere™ Particle* (ISP).

3.4.2.3. Secuenciación mediante Ion Proton

As bibliotecas enriquecidas e cargadas nos *chips* eran secuenciadas ca tecnoloxía baseada en semiconductores na plataforma *Ion Proton* de Ion Torrent. O fundamento desta tecnoloxía consiste na detección de ións de hidróxeno liberados durante a polimerización do ADN. Para isto os 4 nucleótidos proporcionáanse de forma secuencial durante a carreira en fluxos (Fig.19A). Cando o nucleótido proporcionado é complementario a base situada a continuación do cebador de secuenciación, este incorpórase mediante unha polimerasa. Isto provoca un aumento na lonxitude da secuencia do cebador nunha base (ou máis se se trata dun tramo homopolímero) o que provoca a hidrólise dun nucleótido trifosfato, e como consecuencia, a liberación dun pirofosfato e dun protón por cada un dos nucleótidos incorporados. A liberación do protón da lugar a súa vez a un cambio no pH da solución, que é proporcional ó número de nucleótidos incorporados durante un fluxo (Fig.19B). Esta variación é detectada por un sensor situado no fondo de cada pozo, converténdoa en voltaxe e sendo esta última dixitalizada (Fig. 19C).

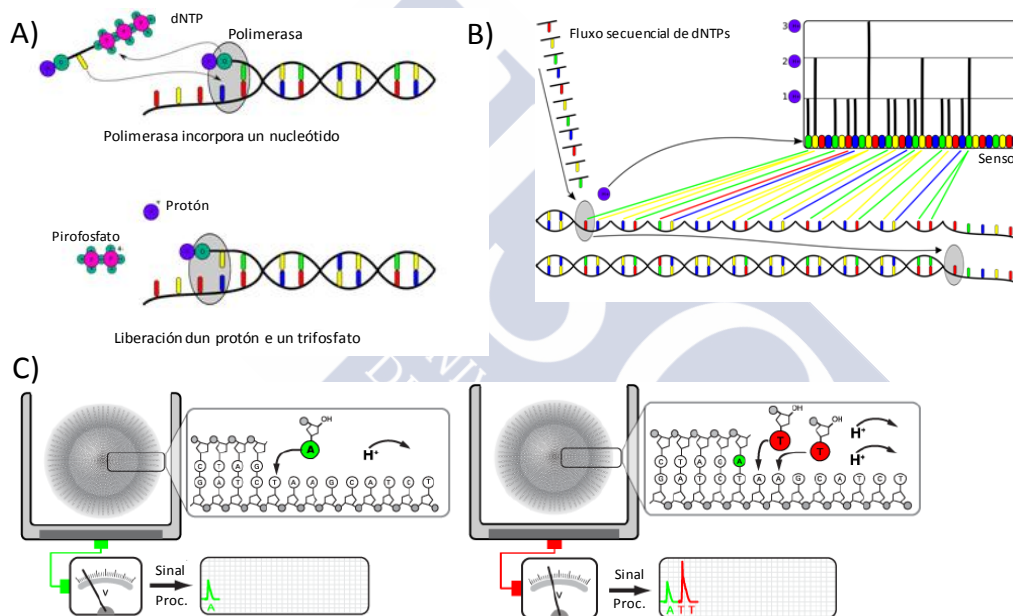


Figura 19 .Esquema de secuenciación co Ion Proton.

3.4.2.4. Análises Bioinformáticas

As lecturas xeradas eran analizadas no *Torrent Suite Server*, onde o sinal producido durante a secuenciación era transformado en bases. Logo aplicabamos uns controis de calidade que consistían nun recorte e nun filtro das lecturas. O obxectivo do primeiro era eliminar tódalas aquelas asignación de bases erradas no extremo 3´da lectura. Pola contra, o filtrado trataba de eliminar: tódalas lecturas policlonais (é dicir, aquelas que proveñen de ISPs con máis dunha biblioteca), lecturas cun sinal baixo ou irrecoñecible, e lecturas de dímeros de cebadores. Unha vez pasados estes controis, as lecturas eran mapeadas contra o xenoma de referencia. Por último, mediante o *Torrent Variant Caller* (TVC) procedíamos a detección de variantes, para o cal utilizabamos os parámetros para a detección de variantes somáticas xa que está opción utilizaba uns filtros moi laxos co fin de non perder ningunha variante, xa que posteriormente cada unha delas ía ser

confirmada mediante Sequenom. Finalmente as SNVs eran anotadas utilizando Annovar (Wang et al. 2010b).

3.5. XENOTIPADO

3.5.1. Selección de variantes a xenotipar

Esta tecnoloxía non só nos permitía confirmar as SNVs atopadas mediante secuenciación de nova xeración senón que tamén podíamos estimar as frecuencias alélicas e xenotípicas, no total da nosa mostra, das variantes atopadas cos distintos métodos de secuenciación. En cada un dos estudos utilizabamos distintos criterios de selección de variantes:

- Estudo I: Para este estudo escollíamos tódalas variantes de sentido incorrecto que tiveran unha frecuencia inferior ó 5%. Ademais intentabamos validar posibles SNPs funcionais con frecuencia inferior ó 5% en mostras europeas de dbSNP e que seleccionabamos usando *Functional Element SNPs Database* (FESD) (Kang et al. 2005) e Patrocles (Clop et al. 2006; Hiard et al. 2009). A primeira utilizabámola para explorar os posibles lugares de unión a factores transcrición, illas CpG, lugares de *splicing* e sitios poly(A). Por outra banda, usando Patrocles podíamos ver se existían posibles lugares diana do miARN en rexións 3'UTR dos distintos transcritos de *DISC1*. Para reducir a taxa de falsos positivos da predición de Patrocles, reducimos a busca a aqueles posibles lugares diana preditos polas dúas fontes usadas por Patrocles (Lewis et al. 2005; Xie et al. 2005).
- Estudo III: Neste caso seleccionabamos aquelas variantes identificadas co SOLiD4 que eran non sinónimas, que estiveran presentes exclusivamente en casos e que eran preditas como prexudiciais por preditores *in silico*.
- Estudo IV: Escollíamos tódalas SNVs non sinónimas identificadas con Ion Proton e que estiveran a unha frecuencia menor ó 1%.

3.5.2. Sequenom Mass Array® System

O xenotipado das mostras facíámolo utilizando espectrometría de masas (MALDI-TOF MS, *Matrix Assited Laser Desorption Ionization-Time of Flight Mass Spectrometry*) coa tecnoloxía MassArray de Sequenom (San Diego, CA), que determina o alelo do SNP segundo o peso molecular dos produtos formados na reacción de minisequenciación. Para iso emprega a química *iPLEX® Gold*, na que a reacción de discriminación alélica é unha extensión de base única (SBE, *Single Base Extension*). Esta reacción baséase na capacidade da ADN polimerasa de incorporar, a partires dun cebador, os nucleótidos complementarios a unha secuencia de ADN molde. O cebador é deseñado de tal xeito que o extremo 3' é complementario á base inmediatamente anterior ó SNP. A base polimórfica é determinada pola adición pola polimerasa dun dNTP (desoxinucleótidos trifosfato) complementario á base que queríamos determinar. Para esta fase as mostras de ADN normalizábanse a unha concentración de 20 ng/μl nun volume final de 40 μl.

Antes de empezar coa parte experimental da tecnoloxía, era necesario facer o deseño do ensaio para o xenotipado. Para isto utilizabamos a ferramenta *Assay Design*

v4, que agrupa as variantes xenéticas en *multiplex*, é dicir, reaccións de xenotipado múltiples nunha única reacción. Ademais deseña a parella de cebadores para a amplificación inicial mediante a PCR de cada unha das variantes e o cebador para a reacción de extensión SBE.

O proceso propiamente dito do xenotipado de Sequenom iPLEX comeza cunha reacción de amplificación na que as mostras sométense a unha PCR multiplex para amplificar as rexións de interese, utilizando unha mestura de cebadores. Trala PCR, lévase a cabo unha purificación encimática coa encima SAP (*Shrimp Alkaline Phosphatase*), coa finalidade de eliminar os dNTPs que non se incorporan, así evítase que se unan ás cadeas de ADN nas seguintes reaccións. A encima SAP é unha encima fosfatasa que modificaba o grupo -OH do grupo fosfórico dos dNTPs impedindo que se unan á cadea de ADN.

O seguinte paso é a extensión que consiste nunha reacción de discriminación alélica por medio dunha reacción SBE. Para isto utilízanse nucleótidos terminadores modificados, de xeito que se incrementa a diferenza de masas entre os catro tipos de terminadores (nucleótidos terminadores de Adenina, Citosina, Guanina e Timina), que é o que se utiliza para a súa posterior identificación. Desta forma conséguese que a masa dos terminadores variase entre 16 e 79,9 Da. Antes de facer a lectura, os produtos da reacción de extensión purifícanse engadindo unha resina (*Clean resin*) co fin de eliminar os sales para que non interfiran na lectura.

Os produtos da reacción de extensión purificados dispénsanse nun *Spectro Chip* para a súa análise coa tecnoloxía MALDI-TOF. A continuación introdúcese o *chip* no Autoflex, que é un espectrómetro de masas MALDI-TOF. Dentro do espectrómetro un láser volatiliza as mostras do *chip*, que voan unha por unha no interior dunha columna de baleiro. Cada molécula de ADN voa de xeito independente, e o tempo que tarda en chegar ó final da columna de baleiro é proporcional a súa masa. Na columna de baleiro xérase un campo eléctrico que acelera as moléculas de ADN que se detectan a medida que van chegando á parte superior, onde se atopa o detector. Dado que o tempo de voo na columna é proporcional á masa, e que cada produto de extensión ten unha masa diferente, este sistema consegue discriminar cada un dos alelos dos polimorfismos analizados no ensaio. A diferenza mínima de masa que pode discriminar a espectrometría de masas MALDI-TOF é de 15 Da, e o rango de masa dos produtos de extensión de tódolos polimorfismos vai de 5000 a 8500 Da. Para ver a relación masa/carga de cada unha das variantes analizadas utilízase o *software Spectro Acquire*, o cal xera un espectro de masas. Cada un dos espectros revisábanse manualmente utilizando o *software TyperAnalyzer v4* (Fig.20).

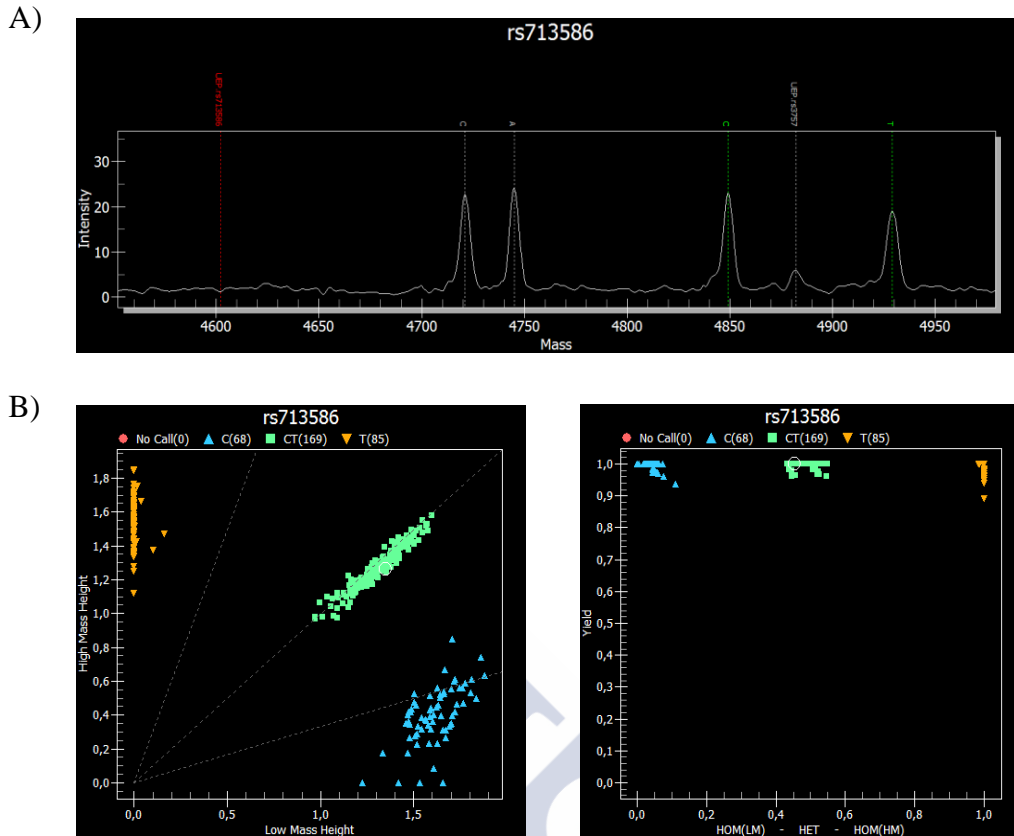


Figura 20. Exemplo das imaxes do *TyperAnalyzer v4* obtidas do xenotipado dun SNP. En A) aparece representado o espectro de masas dun SNP onde se pode diferenciar a posición do cebador (vermello) e a dos dous alelos do SNP C/T (verde). En B) represéntase a asignación de xenotipos para a variante onde cada punto representa unha mostra, sendo os laranxas homocigotos para un alelo, os verdes os heterocigotos e os azuis homocigotos para o outro alelo.

A cada unha das mostras e cada un dos SNPs aplicábase uns controis de calidade. Para isto revisábase cada un dos *clusters* obtidos (Fig. 20 B) e eliminábanse aquelas variantes cunha baixa porcentaxe de xenotipado (<90%) ou que non cumprían o equilibrio Hardy-Weinberg. En canto as mostras, eliminábanse aquelas que amosaban unha baixa porcentaxe de xenotipado (<90%) ou que tiñan unha peor calidade. A modo de control interno, en cada un dos *chips* analizábase incorporábase un trío de mostras Coriell (NIGMS *Human Genetics Cell Repository*, *Coriell Institute for Medical Research*, Camden, New Jersey, EEUU) para o que existían datos publicados e cos que se podían comparar os resultados obtidos e revisar a correcta transmisión de alelos de pais a fillos, co fin de comprobar a eficacia do xenotipado.

O xenotipado das mostras era levado a cabo no Centro Nacional de Genotipado (CEGEN), no nodo de Santiago de Compostela.

3.6. ANÁLISE DE DATOS

3.6.1. Secuenciación de DISC1 (Estudo I)

As frecuencias e o risco das distintas variables atopadas neste caso eran comparadas utilizando o Test Exacto de Fisher dunha única cola co programa WINPEPI (Abramson 2004). Aplicábase a corrección de test múltiples mediante o método Bonferroni tendo en conta as variantes que aparecían na nosa mostra.

Utilizabamos o test de tendencia de Cochran-Armitage para comprobar se existía unha tendencia a acumular variantes en casos respecto ós controis. Para facer o cálculo deste test utilizabamos o mesmo programa que para o anterior.

Utilizabamos o programa PS (Power and Sample Size Calculations) (Dupont and Plummer 1998) para facer o cálculo da frecuencia e do risco asociado a unha variante que eramos capaces de identificar utilizando o noso tamaño de mostra e fixando unha potencia do 80%.

A meta-análise era feita para aqueles SNPs dos que dispoñiamos datos de máis de tres poboacións. Para o seu cálculo usabamos o modelo de efecto fixo de Mantel-Haenszel e o modelo de efecto ó chou de DerSimonian-Laird. O modelo fixo é mais potente en ausencia de heteroxeneidade mentres que o modelo aleatorio é máis robusto cando existe heteroxeneidade. A heteroxeneidade entre os estudos era cuantificada co estatístico I^2 . Tódalas análises levábase a cabo co programa R utilizando o paquete meta 1.6.

As secuencias de *DISC1* das distintas especies utilizadas para facer o aliñamento eran obtidas de Ensembl Genome Browser (<http://www.ensembl.org/index.html>). Para levar a cabo o aliñamento utilizabamos o programa CLUSTALX2 versión 2.0 (Thompson et al. 1997). Os ficheiros que contiñan os resultados deste aliñamento eran abertos co programa GENEDOC (Nicholas et al. 1997) co fin de comprobar o grao de conservación das posicións que ocupaban as distintas variables estudadas.

Para predicir se as substitucións aminoacídicas que afectaban a función da proteína utilizabamos os programas Polyphen2 (Sunyaev et al. 2001) (<http://genetics.bwh.harvard.edu/pph>) e SIFT (Ng and Henikoff 2003) (<http://sift.jcvi.org/>). Polyphen2 é unha ferramenta que fai predicións do posible impacto dunha substitución dun aminoácido nunha proteína baseándose nas características da secuencia e na información estrutural e filoxenética que a caracterizan. Por outra banda, SIFT basea as súas predicións no grao de conservación do residuo aminoacídico cando se aliña con secuencias estreitamente relacionadas.

3.6.2. Interactoma *DISC1* (Estudo II)

Para a selección dun dos grupos de xenes a analizar neste estudo utilizabamos STRING, unha base de datos de interaccións proteína-proteína, tendo en conta experimentos e outras bases de datos (Jensen et al. 2009); a base de datos Entrez GENE do NCBI (Maglott et al. 2005); e buscas bibliográficas. Para o segundo grupo, elixiamos aqueles xenes que interactuaban con polo menos dous xenes adicionais do interactoma de *DISC1* descrito por Camargo e cols. (Camargo et al. 2007).

Unha vez feitos os dous grupos de xenes, analizábase o conxunto de xenes usando o *software* GenGen (Wang et al. 2007). Para isto, o *software* calcula un test de tendencia para cada un dos SNPs e ordena os xenes nunha lista en función do valor-P

máis significativo alcanzado en cada un dos xenes. A continuación estima unha puntuación de enriquecemento baseada no estatístico da suma acumulada de Kolmogorov-Smirnov ponderado. Por último, calcula a significación da puntuación do enriquecemento facendo 5000 permutacións da etiqueta caso e control. Deste xeito, preserva a estrutura do desequilibrio de ligamento e a densidade de SNPs de cada un dos xenes.

Coa fin de confirmar os resultados atopados co procedemento anterior, analizabamos o conxunto de xenes con PLINK 1.07 (Purcell et al. 2007). Para o cálculo do estatístico facíamos a media dos valores de X^2 do test de tendencia de Cochran-Armitage. Para esta análise fixabamos o valor de r^2 en 0,5 como un criterio de inclusión dun SNP no test, e tamén o número de SNPs (10) a incluír neste estatístico. O cálculo da significación facíámolo levando a cabo 10000 permutacións da etiqueta caso-control.

3.6.3. Secuenciación mediante SOLiD4 System (Estudo III)

Para este traballo utilizabamos distintos tipos de tests estatísticos. O primeiro que usabamos era o test de carga WSS que pondera por frecuencias en controis (Madsen and Browning 2009). Calcula a puntuación de cada individuo a partires destes valores ponderados e a suma da puntuación de casos, o cal considera como estatístico. Outro dos tests aplicados era o C-Alpha (Neale et al. 2011) que asume como hipótese nula que a distribución de cada variante entre casos e controis considerada é unha distribución binomial. No caso de existir variantes de risco e/ou protección, habería unha sobredispersión da distribución. Isto podía calcularse sobre todas as variantes e estimarse a súa significación mediante este test. As variantes que aparecían unha soa vez eran agrupadas como se fose unha única variante considerándoas como se tiveran a mesma dirección de efecto. Este test aplicábase a cada unha das SNVs dúas veces, ponderando ou non pola puntuación de PolyPhen2 (versión 2.2.2).

O nivel de significación de ambos tests calculabámolo mediante permutacións (Fig.21). Para este procedemento, seleccionabamos tódalas SNVs presentes en 15 ou menos *pools*, logo asignabamos cada variante a cada unha das 306 mostras, para despois dividir as mostras en 153 casos e 153 controis ó chou, xuntando os casos e os controis en grupos de tres. Este procedemento facíámolo 1000 veces calculando o estatístico nestas *pools* ficticias e comparándoas coas reais, obtendo así o valor de significación de cada unha das SNVs. Todo isto facíase utilizando un *script* en R que creabamos.

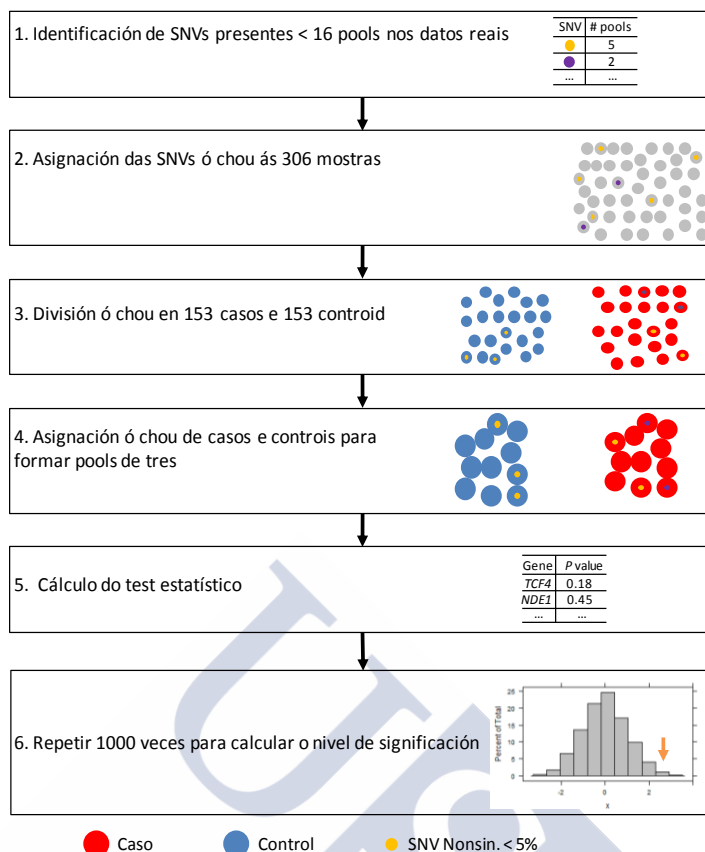


Figura 21: Esquema dos pasos a seguir para calcular o nivel de significación dos test estatísticos.

Aquí tamén aplicabamos o test exacto de Fisher dunha cola para comparar a distribución de portadores de SNVs que aparecen unha única vez na nosa mostra entre afectados e non afectados, asumindo que estas SNVs tiñan unha maior probabilidade de ser prexudiciais. Isto era feito con WINPEPI (Abramson 2004). Resaltar que deste test eran excluídas as SNVs que aparecían cunha frecuencia maior a 0,5 en mostras europeas-americanas no *4300 NHLBI GO Exome Sequencing Project European* (<http://evs.gs.washington.edu/EVS/>).

Ó igual que o estudo I, utilizabamos o programa CLUSTALX2 versión 2.0 para levar a cabo o aliñamento (Thompson et al. 1997).

3.6.4. Secuenciación de nova xeración de *NRXN1* mediante Ion Proton™ System (Estudo IV)

Neste traballo inicialmente aplicabamos os mesmos tests estatísticos utilizados no estudo III, é dicir, tanto o test de carga WSS como o test C-Alpha.

A continuación levabamos a cabo unha meta-análise para a cal faciamos unha busca sistemática dos artigos publicados nos que se secuenciaba *NRXN1* en individuos diagnosticados con esquizofrenia e en controis, para obter os datos das variantes atopadas. Logo usabamos o paquete de R meta 3.5-0 para aplicar o test de Mantel-Haenszel. Xa que a heteroxeneidade non era salientable, considerabamos un modelo de efecto fixo. A heteroxeneidade era cuantificada mediante o parámetro I^2 , usando o estatístico Q para ver o nivel de significación.



4. RESULTADOS





4.1. ESTUDO I: ANÁLISE DE VARIANTES RARAS NO XENE *DISC1*

4.1.1. Introducción

Datos de xenética de poboacións suxiren que as variantes raras xogan un papel importante na susceptibilidade a enfermidades complexas (Kryukov et al. 2007; Barreiro et al. 2008; Gorlov et al. 2008). Baseándose en comparacións de datos de resecuenciación de humano e chimpancé, Kryukov e cols. (2007) amosaban que unha gran parte de mutacións *de novo* codificantes de sentido incorrecto que son medianamente deletéreas estaban suxeitas a selección negativa débil e que poderían chegar a alcanzar unha frecuencia elevada. Estes autores asumían que a maioría deste tipo de alelos deletéreos podían conferir risco a unha enfermidade, polo que concluían que numerosos SNPs raros e funcionais podían considerarse como a principal contribución na susceptibilidade a padecer unha enfermidade común (Kryukov et al. 2007). A esta mesma conclusión tamén chegaban diferentes investigadores despois de analizar a distribución da frecuencia relativa de SNPs codificantes de sentido incorrecto dependendo da súa probabilidade de alterar a función da proteína (Barreiro et al. 2008; Gorlov et al. 2008). Neste senso tamén se confirmaba un exceso de SNPs non sinónimos fronte a sinónimos cunha frecuencia entre 2-4% despois de resecuenciar 200 exomas humanos (Li et al. 2010).

No ano 2000, Millar e cols. identificaban unha variante estrutural que estaba relacionada coa esquizofrenia, que consistía nunha translocación balanceada (1;11) (q42.1;q14.3) detectada nunha familia escocesa, a cal cosegregaba con varios trastornos mentais, como a esquizofrenia, trastorno esquizoafectivo, depresión maior recorrente e trastornos emocionais e da conduta (St Clair et al. 1990). Esta translocación afecta a un xene que codifica unha proteína localizada no cromosoma 1, chamada *DISC1* (Disrupted In Schizophrenia 1) (Millar et al. 2000). Despois da identificación de *DISC1*, facíanse numerosos traballos para tratar de ver o seu papel a nivel molecular. Este xene presenta múltiples transcritos (Nakata et al. 2009) e exprésase en diferentes tecidos (Millar et al. 2000). Ó longo destes anos realizábanse numerosos estudos que intentaban analizar os patróns de interacción de *DISC1*, amosando que é unha proteína central que está relacionada con numerosos procesos tales como: estabilización de microtúbulos, transporte de vesículas mediada por quinesinas, extensión de neuritas ou na regulación da proliferación dos proxenitores neuronais (Camargo et al. 2007; Chubb et al. 2008; Mao et al. 2009).

En canto ós estudos xenéticos feitos sobre *DISC1*, resaltar en primeiro lugar unha meta-análise feita con 50 SNPs comúns en poboación europea utilizando tamaños de mostras que ían dende máis de 3500 ata uns 10000 individuos, no que atopaban un SNP común de susceptibilidade cun risco asociado de 1,27 (Schumacher et al. 2009). Posto que este SNP está situado nun intrón, era probable que fose un *proxy* dalgunha variante funcional indeterminada. Por outro lado, Song e cols. trataban de buscar variantes codificantes de sentido incorrecto ultra-raras, definidas como aquelas detectadas mediante secuenciación e/ou método *multiplex* de polimorfismo de conformación de cadea simple nunha cohorte de 288 pacientes esquizofrénicos e 288 controis pero que non se atopaban en 10000 controis adicionais. Este traballo revelaba un exceso

significativo de variantes ultra raras codificantes de sentido incorrecto nos casos fronte ós controis (Song et al. 2008).

Neste traballo intentabamos investigar o papel de variantes posiblemente funcionais de frecuencia baixa no xene *DISC1* na susceptibilidade á esquizofrenia. Primeiro, utilizabamos ferramentas bioinformáticas, buscas bibliográficas e secuenciación de exóns nun subconxunto de mostras para identificar SNPs posiblemente funcionais a baixa frecuencia en *DISC1*. A continuación levabamos a cabo un estudo de asociación das variantes identificadas. Por último, faciamos unha meta-análise para conseguir conclusións máis robustas para aqueles SNPs dos que dispoñiamos de datos suficientes.

4.1.2. Busca bioinformática de posibles SNPs funcionais a baixa frecuencia

Para esta busca utilizabamos FESD e Patrocles. Co primeiro viamos que non existía ningún SNP validado a frecuencia menor ó 5% en mostras europeas en dbSNP e que ademais afectase ós lugares de unión de factores de transcrición, illas CpG, lugares de *splicing* ou lugares poly(A). Por outra banda, utilizando Patrocles detectabamos un SNP que posiblemente afectase ó funcionamento dun microARN que regula a *DISC1*, rs11122396. O alelo menor deste SNP alteraba un posible lugar diana dos MIR135A1 e MIR135B. Este SNP estaba en completo desequilibrio de ligamento ($r^2=1$) co rs3737597 nas mostras europeas en HapMap, é dicir, na poboación CEU (colección de residentes en Utah con devanceiros do norte e oeste de Europa do *Centre d'Etude du Polymorphisme Humain*, CEPH) e TSI (en inglés, *Tuscan in Italy*); e cun $r^2 < 0,90$ nas diferentes poboacións asiáticas (Frazer et al. 2007). Ámbolos dous SNPs estaban situados na rexión 3'UTR da isoforma L e Lv separados por 2,4 kb. Seleccionabamos o SNP rs3737597 para xenotipar nas nosas mostras xa que este polimorfismo era anteriormente xenotipado por outros grupos ademais de ser un *proxy* perfecto do SNP rs11122396.

4.1.3. Secuenciación das rexións exónicas de *DISC1* na mostra descubrimento

Nesta fase do estudo secuenciabamos os diferentes exóns de *DISC1* en 47 casos e 47 controis (todos eles de orixe galega e non emparentados) co fin de buscar variantes non sinónimas a baixa frecuencia (<5%). En total eran secuenciadas 8509 pb por cada mostra, nas que ían incluídas as partes codificantes da isoforma Lv de *DISC1* máis as partes das secuencias exónicas das outras isoformas (L, Es e S) que eran distintas á Lv.

Aquí identificabamos 12 variantes (Táboa 17), das cales 7 eran non sinónimas, sendo 4 SNPs codificantes de sentido incorrecto e cunha frecuencia baixa (< 5%). Estes últimos estaban presentes en tres casos e tres controis. Resaltar que a variante C565F non fora identificada previamente.

Resultados

Variable ^a	Exón ^b	Posición cromosómica ^c	Tipo SNP	Casos (N=47)			Controis (N=47)		
				2211	21	22	11	21	22
A220A	2	231.830.164	S	47	0	0	46	1	0
rs3738401	2	231.830.295	NS	23	19	5	18	23	6
rs3738402	5	231.903.010	S	39	8	0	43	4	0
rs28930675	5	231.902.975	NS	47	0	0	46	1	0
rs78792190	5	231.902.912	NS	46	1	0	47	0	0
rs2492367	6	231.906.589	S	36	10	1	37	10	0
C565F	8	231.935.858	NS	47	0	0	46	1	0
rs6675281	9	231.954.101	NS	38	6	3	34	12	1
rs12133766	9	231.954.145	S	41	6	0	38	8	1
rs821616	11	232.144.598	NS	27	15	5	24	19	4
E751Q	11	232.144.739	NS	45	2	0	46	1	0
P734P	11	232.144.690	S	46	1	0	47	0	0

Táboa 17. Variantes atopadas mediante secuenciación. En negra están as variantes non sinónimas cunha MAF<5%.

^a Fai referencia ó transcrito L (Secuencia de referencia do NCBI: NM_018662.2).

^b A posición cromosómica está baseada na versión do xenoma de referencia humano de febreiro do 2009 (GRCh37/hg19).

^c SNP sinónimo (S) ou non sinónimo (NS).

4.1.4. Estudo de asociación caso-control de SNPs funcionais a baixa frecuencia

Nesta parte do estudo xenotipabamos, mediante a tecnoloxía de Sequenom iPLEX MassArray, as catro variables que atopabamos na fase anterior que eran variantes codificantes de sentido incorrecto e cunha frecuencia baixa (< 5%) (rs28930675 (T453M), rs78792190 (P432L), E751Q e C565F) xunto con outras descritas na bibliografía e que tamén cumprían as características mencionadas anteriormente: R37W, R418H, S90L, T603I (Song et al. 2008), ademais do SNP rs3737597 localizado na rexión 3'UTR do xene e identificado mediante Patrocles (Figura 22). Nesta fase xenotipabamos un total de 499 casos e 533 controis que, como no paso anterior, eran de orixe galega e non estaban emparentados (Táboa 18).

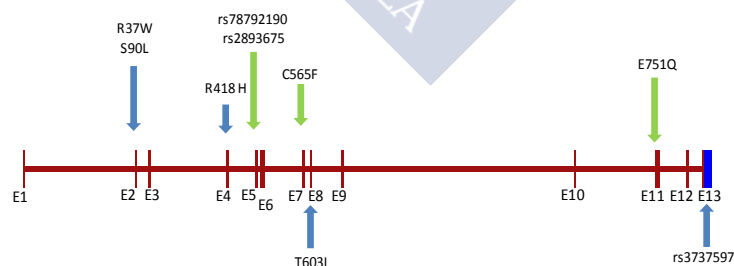


Figura 22. Localización exónica das variantes xenotipadas mediante *Sequenom MassArray*; as frechas verdes indican as variables que foron atopadas mediante secuenciación e as frechas azuis son as que están descritas na bibliografía e que foron incluídas na análise.

Primeiro destacar que como control de calidade, entre as mostras xenotipadas, incluíamos os casos e os controis que tamén foran secuenciadas mediante Sanger. Viamos que existía unha concordancia do 100% entre as variantes das mostras atopadas

mediante secuenciación e os resultados obtidos destas mesmas mostras mediante o xenotipado con Sequenom.

De tódalas variables analizadas, tres (T603I, R37W e S90L) eran monomórficas para a nosas mostras. O resto das variantes detectabámolos en heterocigose (Táboa 18). De todas estas só a variante E751Q presentaba unha asociación positiva coa enfermidade, alcanzando un valor-P igual a 0,027 e cun OR de 2,72 (Intervalo de confianza do 95%, 0,99-8,63) despois de aplicar un test exacto de Fisher dunha cola, aínda que non alcanzaba un nivel significativo cando se aplicaba a corrección de múltiples tests.

Os alelos raros constituían o 5,32% dos alelos totais para cada posición secuenciada dos casos (53 de 998) e o 5,44% dos alelos totais dos controis (58 de 1066), non existindo diferenzas significativas entre ámbolos dous grupos (Test χ^2 , P=0,90).

Variante ID	Posición Cromosómica	Rexión	Alelos ^a	Frec. Casos ^b (%)	Frec. Controis ^b (%)	Valor-P ^c
C565F	231.935.858	Exón 8	G/A	0/998 (0)	2/1066 (0,19)	0,267
E751Q	232.144.739	Exón 11	C/T	15/998 (1,50)	6/1066 (0,56)	0,027
rs78792190	231.902.912	Exón 5	C/T	5/998 (0,50)	8/1066 (0,75)	0,332
R418H	231.885.807	Exón 4	G/T	1/998 (0,20)	2/1066 (0,56)	0,525
rs28930675	231.902.975	Exón 5	G/C	2/998 (0,20)	6/1066 (0,56)	0,166
rs3737597	232.172.838	3'UTR	C/T	30/998 (3,01)	34/1066 (3,19)	0,455

Táboa18. Resultados do estudo de asociación caso-control.

^a Referencia a cadea positiva. O alelo en primeiro lugar é o alelo común

^b Frecuencia do alelo menor.

^c Valor-P do Test Exacto de Fisher alélico dunha cola.

Resaltar que había tres casos nos que se atopaban dúas variables por individuo, cousa que non ocorría nos controis. Dous casos tiñan as mesmas variables (E751Q e P432L) e o outro unha combinación da variable E751Q e a situada na rexión 3'UTR (rs3737597). Con estes datos levabamos a cabo un test de tendencia de Cochran-Armitage para comprobar se existía un maior risco a padecer a enfermidade a medida que se ían acumulando variantes raras. Con este test non se alcanzaban valores significativos (Táboa 19).

	0 variantes	1 variante	2 variantes	Valor-P
Casos	446	47	3	0,922
Controis	475	58	0	

Táboa 19. Test de tendencia de Cochran-Armitage.

Resultados

Tamén aplicabamos o test exacto de Fisher dunha cola para comprobar se existían diferenzas significativas na acumulación de variantes raras en casos respecto a controis. Por medio desta análise víamos que non existían diferenzas significativas (Táboa 20).

	Presenza	Non presenza	Valor-P ^a
Casos	50 (11,2%)	446 (88,8%)	0,376
Controis	58 (12,2%)	475 (87,8%)	

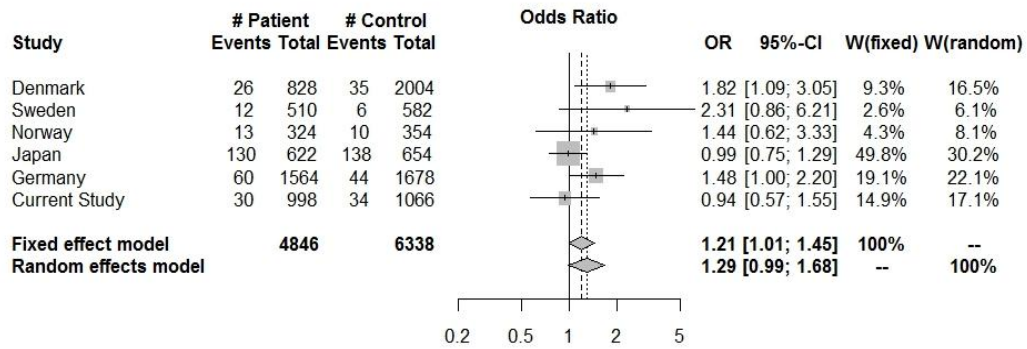
Táboa 20. Acumulación de variantes raras en casos respecto ós controis.

^a Test exacto de Fisher dunha cola.

4.1.5. Meta-análise dos resultados de rs3737597 e E751Q

Nesta parte primeiro levábamos a cabo unha meta-análise cos datos obtidos do xenotipado do SNP rs3737597 na nosa mostra e cos datos de tres traballos previos nos que analizaban este SNP en mostras caso-control. Saetre e cols. no 2008 xenotipaban este SNP en tres poboacións escandinavas diferentes, Zhang e cols. no 2005 analizábano nunha mostra xaponesa, e Schumacher e cols. no seu estudo do 2009 presentaban datos imputados para este SNP nunha mostra alemá (Zhang et al. 2005; Saetre et al. 2008; Schumacher et al. 2009). Polo tanto, levabamos a cabo unha meta-análise de seis poboacións a nivel alélico, nun total de 2423 casos e 3169 controis (Fig. 23A). Os resultados amosaban unha asociación marxinal. Segundo o modelo fixo obtiñamos un $P = 0,04$, cun $OR = 1,21$ (95% IC = 1,01-1,45); e segundo o modelo ó chou alcanzaba un valor-P de 0,06 ($OR = 1,29$, 95% IC = 0,99-1,68). Resaltar que a heteroxeneidade entre os estudos era elevada cun valor de $I^2 = 40,7\%$ (95% CI 0,0-76,5%). O SNP rs3737597 estaba a baixa frecuencia nas mostras europeas (1%-4%), mentres que acadaba unha frecuencia alta en asiáticos (entorno ó 20%). A vista destes datos repetiamos a meta-análise despois de eliminar os datos das mostras xaponesas, para así intentar reducir a heteroxeneidade (Fig. 23B). A mostra para esta segunda meta-análise estaba formada finalmente por 2112 esquizofrénicos e 2842 controis. Como era de esperar a heteroxeneidade reducíase dunha forma considerable a uns valores de $I^2 = 11,4\%$ (95% CI 0,0-81,6%). Por outra banda, a asociación era máis significativa, cuns valores para o modelo fixo de $P = 0,0047$ ($OR = 1,42$, 95% CI=1,11-1,82) e para o modelo ó chou de $P = 0,0084$ e un $OR = 1,43$ (95% CI = 1,10-1,87).

A)



B)

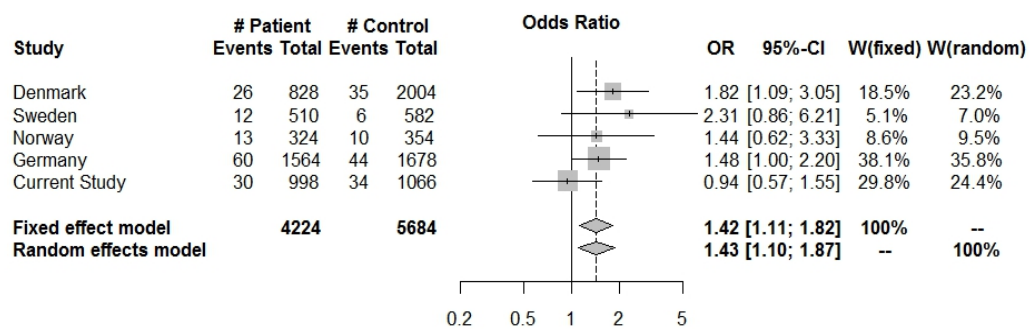


Figura 23. Representación gráfica da meta-análise de rs373597. En A) estaban incluídos os datos de tódalas poboacións dispoñibles. Na figura B) excluíamos a poboación xaponesa. Os ORs para cada poboación individual están representados como cadrados cun tamaño proporcional ó seu peso (estimas do modelo ó chou). As liñas horizontais representan os intervalos de confianza do 95% dos ORs. Os resultados conxuntos das poboacións europeas e de todas as poboacións están representados polos rombos.

Resultados

Por outra parte, levabamos a cabo unha meta-análise a nivel alélico utilizando os datos da variante E751Q do noso traballo xunto cos datos doutros tres estudos onde identificaban esta variante (Song et al. 2008; Moens et al. 2011; Thomson et al. 2014) (Fig. 24). Para isto analizabamos un total de 3000 alelos de casos e 4372 alelos de controis vendo que a variante, segundo o modelo fixo, non acadaba unha asociación significativa ($P = 0,0905$). Comentar que non existía evidencia de heteroxeneidade entre os diferentes estudos ($I^2 = 5,55\%$, cun valor-P do test $Q = 0,136$).

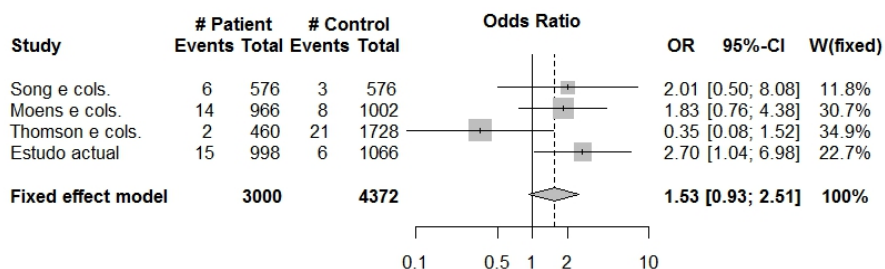


Figura 24. Meta-análise da variante E751Q. Os OR (en escala logarítmica) para os estudos individuais están representados polas caixas en gris cuxo tamaño é proporcional ó peso de cada traballo na meta-análise. As liñas horizontais representan os intervalos de confianza ó 95%. O resultado conxunto e o seu intervalo de confianza do 95% está representado polo rombo

4.2. ESTUDO II: PAPEL DAS PROTEÍNAS QUE INTERACTÚAN CON *DISC1* NO RISCO A PADECER ESQUIZOFRENIA A PARTIR DE DATOS DUN ESTUDO DE ASOCIACIÓN DO XENOMA COMPLETO DE SNPs CODIFICANTES DE SENTIDO INCORRECTO

4.2.1. Introducción

Ó longo destes anos publicábanse varios GWAS sobre esquizofrenia nos que se identificaban variantes comúns que alcanzaban valores de significación moi restritivos (International Schizophrenia et al. 2009; Stefansson et al. 2009; Steinberg et al. 2011a; Ripke et al. 2013). Estes tamén revelaban a existencia de moitos SNPs de susceptibilidade pero que estaban lonxe de alcanzar valores de significación estatística (International Schizophrenia et al. 2009; Lee et al. 2012b). A análise baseada nun conxunto de xenes propúxose como unha maneira de obter información adicional dos resultados dos GWAS (Wang et al. 2007; Cantor et al. 2010).

Dende o seu descubrimento, *DISC1* foi un xene moi estudado para intentar ver a súa relación cos fenotipos psiquiátricos. Estes traballos amosaban que se trataba dun xene central de varias redes, é dicir, que era un nodo altamente conectado na súa rede de interacción proteína-proteína, estando involucrado en procesos como: organización de microtúbulos, transporte de vesículas mediadas por quinesinas, extensión de neuritas ou regulación da proliferación de proxenitores neurais (Millar et al. 2003; Miyoshi et al. 2003; Morris et al. 2003; Camargo et al. 2007; Chubb et al. 2008; Brandon et al. 2009).

Varias das proteínas que interactuaban con *DISC1* podían estar asociadas co risco a padecer esquizofrenia. Un exemplo era a proteína PDE4B, afectada por unha translocación balanceada nun paciente con esquizofrenia e nun familiar deste diagnosticado cunha enfermidade psiquiátrica crónica (Millar et al. 2005b). Ademais o SNP rs910694 situado neste xene, estaba asociado coa esquizofrenia segundo unha meta-análise (Allen et al. 2008). Por outro lado, tamén había evidencias de epistase entre *DISC1* e xenes que interactuaban con el (Burdick et al. 2008; Nicodemus et al. 2010; Andreasen et al. 2011). Sen embargo, pouco se coñecía da relación entre a esquizofrenia e os xenes que interaccionaban con *DISC1*. Algúns modelos animais destes xenes amosaban comportamentos e fenotipos neuroanatómicos que estaban relacionados coa esquizofrenia (Ikeda et al. 2008; Sakae et al. 2008; Youn et al. 2009; Carlisle et al. 2011), os cales eran similares ós ratos mutantes de *DISC1* (Brandon and Sawa 2011).

Previamente o noso grupo levaba a cabo un estudo de asociación de xenoma completo centrado en SNPs codificantes de sentido incorrecto para intentar identificar SNPs específicos asociados que aumentasen o risco a sufrir esquizofrenia. Así identificabamos ó xene *SLC39A8* como un novo xene de susceptibilidade a esquizofrenia (Carrera et al. 2012). No traballo comentado a continuación, partiamos da hipótese de que diferentes alteracións na rede de interacción de *DISC1* poderían aumentar a susceptibilidade individual a desenvolver esquizofrenia; polo que analizabamos os datos do traballo previo para comprobar se existía algunha sobre-representación do conxunto de xenes do interactoma de *DISC1* na parte alta da lista de xenes ordenados por nivel de significación. Para este fin utilizabamos o método *Gene Set Enrichment Analysis*, baseado en estudos previos de expresión xénica, e que estaba especialmente adaptado para estudos de GWAS nos que se analizan SNPs (Wang et al. 2007).

4.2.2. Análise de conxunto de xenes a partires dos datos de xenotipado previos

Neste traballo analizabamos dous conxuntos de xenes. O primeiro deles estaba composto por tódolos xenes dos cales dispoñiamos de datos de xenotipado e que ademais interactuaban con *DISC1* segundo STRING, Entrez GENE ou buscas bibliográficas (interactoma total de *DISC1*). O outro estaba formado por aqueles xenes que interactuaban con, alomenos, dous xenes adicionais do interactoma de *DISC1* descrito por Camargo e cols. (Camargo et al. 2007), ó cal chamabamos "interactoma central de *DISC1*".

Os datos dos xenotipos a analizar consistía, despois de aplicar os correspondentes controis de calidade, en 5100 SNPs codificantes de sentido incorrecto con frecuencias maiores ó 5% que estaban localizados en 3751 xenes. De todos estes, vimos que só 21 xenes pertencían ó interactoma de *DISC1* (Táboa 21). O valor-P máis significativo correspondía ó SNP rs587404, situado no xene *MACF1* (valor-P do test de tendencia= 0,0123, cunha FDR (do inglés, *False discovery rate*) $q= 0,71$, ocupando a posición 85 no rango de significación). Utilizando o *software* GenGen detectabamos un enriquecemento significativo dos xenes do interactoma de *DISC1* na parte superior da lista de xenes ordenados por nivel de significación ($P= 0,0158$), calculado mediante 5000 permutacións da etiqueta do fenotipo. A análise do "interactoma central de *DISC1*", baseado só en datos de 8 xenes, confirmaba a significación deste conxunto ($P = 0,021$). Dous dos oito xenes do "interactoma central de *DISC1*" non estaban incluídos no "interactoma total de *DISC1*", mentres que 13 dos xenes do "interactoma total de *DISC1*" non estaban no "interactoma central de *DISC1*" (Táboa 21).



Xene ^a	Fonte	Mellor SNP	Valor-P test de tendencia	Rango de SNPs ^b	Valor-P PGC ^c
<i>MACF1</i> ^d	STRING	rs587404	0,01226	85	0,039
<i>UTRN</i>	STRING	rs1534443	0,02349	148	0,18
<i>DST</i> ^d	STRING	rs11758339	0,02388	151	0,83
<i>DISC1</i> ^d	STRING, Entrez GENE	rs821616	0,04879	303	n.d.
<i>KIF3A</i>	STRING	rs3798130	0,05682	352	0,373
<i>SYNE1</i>	STRING, Entrez GENE	rs998147	0,05763	358	0,075
<i>AKAP9</i> ^d	STRING	rs6964587	0,06354	386	0,484
<i>PCMI</i>	(Kamiya e cols., 2008)	rs208753	0,1574	883	0,15
<i>ZNF365</i>	STRING	rs7076156	0,2826	1496	0,063
<i>DMD</i>	STRING	rs1800280	0,347	1796	n.d.
<i>KIAA1267</i> ^e	(Camargo e cols., 2007)	rs7220988	0,3508	1816	0,04
<i>PPP4R1</i>	STRING	rs2306134	0,3616	1885	0,137
<i>NUP160</i>	STRING	rs3816605	0,4192	2193	0,07
<i>SPARCL1</i>	STRING	rs1130643	0,4504	2351	0,969
<i>C14orf135</i> ^d	STRING	rs167437	0,4614	2388	n.d.
<i>FAM190B</i> ^e	(Camargo e cols., 2007)	rs11557865	0,5743	2966	0,464
<i>CEP63</i> ^d	STRING, Entrez GENE	rs9868985	0,6022	3096	0,012
<i>CCDC141</i>	STRING, Entrez GENE	rs17362588	0,7121	3619	0,101
<i>COL4A1</i>	STRING, Entrez GENE	rs3742207	0,8205	4174	0,622
<i>PCNT</i>	STRING, Entrez GENE	rs2073380	0,865	4427	n.d.
<i>FEZ1</i>	Entrez GENE	rs597570	0,8733	4480	n.d.

Táboa 21. Lista dos xenes do interactoma de *DISC1* que estaban presentes no conxunto dos datos.

^a Os xenes máis extremos están indicados en negra.

^b Rango ordenado dos SNPs entre os 5100 SNPs baseados no test de tendencia de asociación.

^c PGC: Valores-P obtidos do *Schizophrenia Psychiatric GWAS Consortium* (Ripke et al., 2011). n.d., non dispoñible.

^d Xenes presentes tanto no "interactoma central de *DISC1*" como no do "interactoma total de *DISC1*".

^e Xenes presentes no "interactoma central de *DISC1*" pero ausentes do "interactoma total de *DISC1*" debido a que non teñen unha interacción directa con *DISC1*.

Resultados

A continuación levabamos a cabo unha análise dos xenes máis extremos (*leading edge*) e identificabamos 7 xenes que estaban situados antes ou no punto de puntuación máxima de enriquecemento (Fig. 25). Entre estes xenes só había catro que presentaban polo menos un SNP cun valor-P nominal que fose menor a 0,05 (Táboa 21).

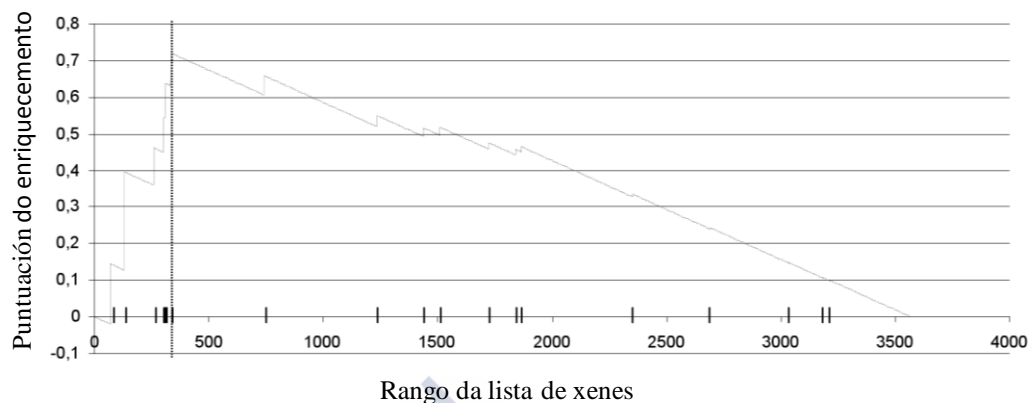


Figura 25. Análise dos xenes máis extremos do conxunto de xenes do interactoma de *DISC1*. As barras verticais do eixo das abscisas representan a posición dos distintos xenes do interactoma de *DISC1* no rango da lista de xenes clasificados segundo o seu mellor valor-P de asociación de cada un dos xenes. A liña punteada vertical indica a localización da puntuación de enriquecemento máximo. Os sete xenes que están a esquerda desta liña son os xenes máis extremos.

Co fin de confirmar os resultados anteriores, faciamos unha análise utilizando un test baseado nun conxunto de xenes mediante PLINK. Para facer este test fixabamos un $r^2=0,5$ ademais de elixir o número de SNPs a incluír (10). Para calcular o nivel de significación utilizabamos permutacións da etiqueta caso-control, concretamente dez mil. Con estes parámetros atopabamos uns resultados similares, cuns valores-P de 0,055 e 0,025 para o conxunto do “interactoma de *DISC1* total” e para o “interactoma central de *DISC1*”, respectivamente.

Por último viamos que o *PGC* de *Schizophrenia* dispoñía de datos para 16 dos 21 SNPs (Táboa 21). Tres destes SNPs alcanzaban un nivel nominal de significación, incluíndo o SNP máis significativo dos nosos datos. Outros tres SNPs presentaban un valor-P menor de 0,1 e catro SNPs presentaban valores de significación menores de 0,2. A comparación da distribución destes valores-P coa distribución dos valores-P dos SNPs do *Schizophrenia PGC* agrupados por desequilibrio de ligamento (N=113774, $r=0,25$ en fiestras de 500 kb) amosaba unha tendencia a sobre-representación do interactoma de *DISC1* na parte superior dos valores-P ordenados (Test de Kolmogorov-Smirnoff, $P=0,078$), aínda que o feito de non poder acceder a datos de xenotipos individuais impedíanos facer unha análise máis detallada.

4.3. ESTUDO III: RESECUENCIACIÓN E ANÁLISE DE ASOCIACIÓN DE REXIÓN CODIFICANTES DE 21 XENES CANDIDATO NA SUSCEPTIBILIDADE A ESQUIZOFRENIA

4.3.1. Introducción

A esquizofrenia é unha enfermidade complexa que presenta unha elevada herdabilidade, a cal foi estimada entre 65 e 82 % (Sullivan et al. 2003; Lichtenstein et al. 2009). Mediante GWAS identificábanse varios SNPs asociados coa esquizofrenia dunha forma fiable, aínda que cun efecto moi baixo (International Schizophrenia et al. 2009; Stefansson et al. 2009; Steinberg et al. 2011a; Ripke et al. 2013). Modelos estatísticos suxerían que variantes comúns explicaban polo menos un terzo da herdabilidade (International Schizophrenia et al. 2009; Lee et al. 2012b; Ripke et al. 2013). Por outra banda, propoñíase que as variantes raras poderían contribuír dunha forma considerable na herdabilidade inexplicada (van Dongen and Boomsma 2013). Un tipo de variantes que son exemplos claros da existencia de variantes raras que presentan un efecto moderado son as CNVs (Rees et al. 2014c). Sen embargo, en xeral, os intentos para identificar SNVs raras asociadas á esquizofrenia non tiñan gran éxito (Need et al. 2012; Crowley et al. 2013; Kenny et al. 2013; Hu et al. 2014; Purcell et al. 2014), aínda que tamén se atopaba algún resultado prometedor (Kushima et al. 2012; Kimura et al. 2015).

Tendo en conta a limitada potencia para identificar asociacións con variantes raras, a análise do efecto acumulativo das diferentes variantes raras dun xene ou dun conxunto de xenes podería ser necesaria para identificar xenes de risco. Por exemplo, no 2011, Moens e cols. estudaban o xene *DISC1* (Moens et al. 2011). Os autores secuenciaban este xene e 10 máis que interactuaban con el, en 80 pacientes esquizofrénicos e en 80 controis, usando para isto 4 *pools* de 40 mostras. A continuación xenotipaban tódalas variantes non sinónimas cunha frecuencia menor ó 1% en 406 doentes adicionais e 434 controis, chegando a identificar un exceso de variantes raras nos casos. Máis tarde, Purcell e cols. secuenciaban o exoma de 2536 esquizofrénicos e 2543 controis. Aínda que non identificaban ningunha SNV nin ningún xene con asociación significativa despois de aplicar unha corrección por múltiples tests, atopaban un enriquecemento significativo de mutacións raras que alteraban o xene en varios conxuntos de xenes candidato como xenes de canles de calcio dependentes de voltaxe e xenes da densidade post-sináptica (Purcell et al. 2014).

4.3.2. Análise de cobertura dos xenes estudados

Neste traballo estudabamos e secuenciabamos mediante o SOLID4 System as rexións exónicas de 21 xenes relacionados coa esquizofrenia mediante estudos de GWAS, CNVs e do interactoma de *DISC1* (Apartado 3.4.1.1 Selección dos xenes, páxina 72). En total secuenciabamos 124,5 kb de secuencia codificante por mostra. Secuenciabamos 153 casos e 153 controis agrupados en *pools* de tres mostras, comprendendo 51 *pools* de casos e outras tantas de controis. A súa vez estas *pools* eran agrupadas para a súa secuenciación en *pools* de 15 *pools* (ver Fig. 21 na páxina 88).

Mediante unha análise visual da cobertura de cada un dos xenes secuenciados utilizando o IGV decatámonos que un dos xenes, *NRGN*, o cal codifica unha pequena proteína de 78 aminoácidos, fora erroneamente capturado xa que presentaba moi pouca cobertura e ademais as lecturas sempre estaban na mesma cadea, de xeito que non había

Resultados

solapamento entre as lecturas nos dous sentidos (Fig.26). Á vista destes datos non considerabamos os resultados obtidos deste xene e non o tiñamos en conta nas análises posteriores.



Figura 26. Imaxe do IGV do xene *NRGN* onde se podía ver os problemas atopados na captura.

No primeiro paso deste traballo analizabamos os datos de cobertura. De media alcanzabamos unha cobertura de 376X (Táboa 22). Un 95% das bases obxecto de estudo de cada unha das *pools* estaba lida, polo menos, 100 veces, e o 88% presentaba, polo menos, 150 lecturas. Con esta profundidade de lectura amosabamos que o noso método de captura era eficaz (Fig. 27).

	Cobertura Media	Proporción de rexión de interese cuberta					
		25x	50x	75x	100x	125x	150x
Mostra Total	376,253	0,988	0,981	0,971	0,952	0,924	0,884
Casos	375,710	0,988	0,981	0,971	0,953	0,924	0,881
Controis	376,829	0,988	0,980	0,970	0,952	0,925	0,887
Octeto_1	360,407	0,988	0,981	0,971	0,952	0,925	0,885
Octeto_2	343,152	0,988	0,978	0,965	0,940	0,902	0,849
Octeto_3	322,342	0,987	0,979	0,966	0,942	0,902	0,846
Octeto_4	406,468	0,989	0,983	0,975	0,961	0,938	0,902
Octeto_5	414,488	0,989	0,982	0,975	0,962	0,941	0,911
Octeto_6	425,474	0,989	0,983	0,976	0,962	0,940	0,912
Octeto_7	361,441	0,987	0,979	0,968	0,948	0,921	0,881

Táboa 22. Datos da cobertura media e da porcentaxe da rexión de interese cuberta do total da mostra, dos casos, dos controis e por octeto.

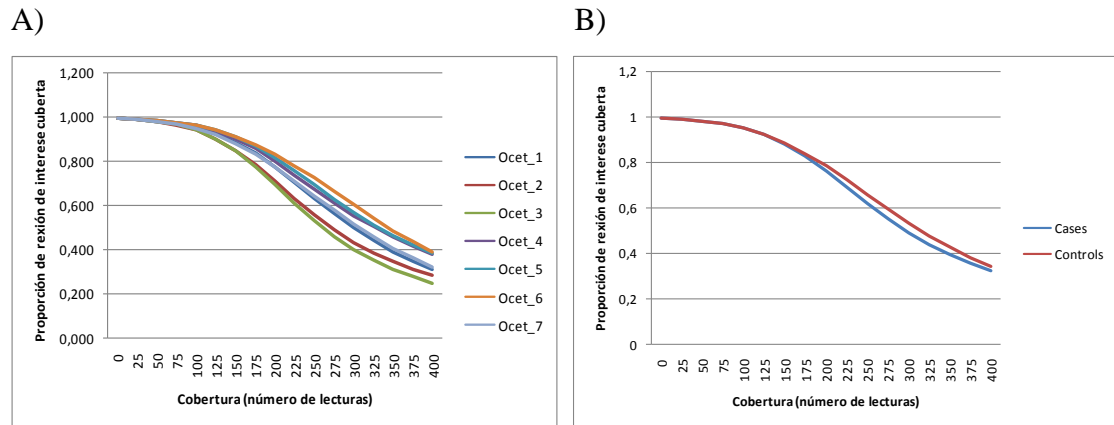


Figura 27. Porcentaxe de rexión de interese cuberta diferenciando octetos (A) e casos e controis (B)

A continuación comprobábase que non existían diferenzas significativas entre a cobertura dos casos e dos controis que provocasen un nesgo na identificación de variantes. Isto facíamolo mediante o test de Mann-Whitney que alcanzaba un valor- $P=0,96$, de xeito que podíamos comparar ámbalas mostras (Fig. 28).

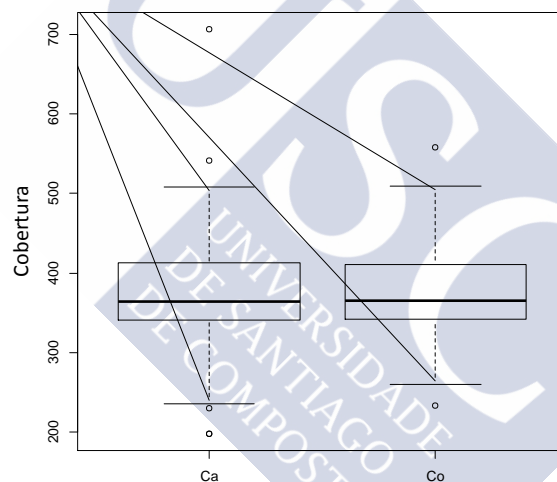


Figura 28. Diagrama de caixas e bigotes no que se comparaba a distribución de coberturas de casos fronte a controis.

Resultados

Ademais vimos que existía unha alta reproducibilidade de cobertura entre as mostras (Fig. 29).

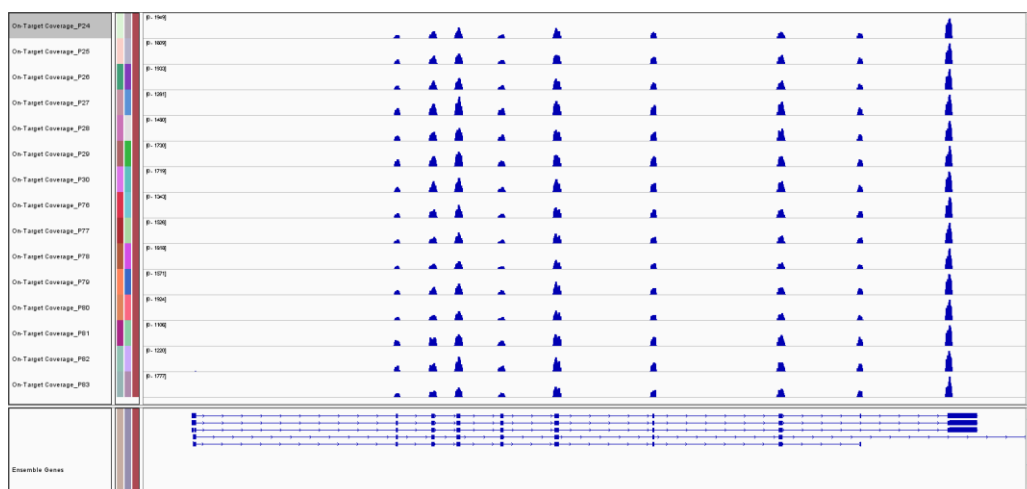


Figura 29. Imaxe do IGV na que se amosan distintas mostras (cada fila é unha mostra) e a cobertura das partes exónicas do xene *KIF3A*.

4.3.3. Selección do paquete informático para a detección de variantes

Para a identificación das variantes inicialmente utilizabamos tres paquetes informáticos: BioScope, GATK e VarScan (Apartado 3.4.1.4 Análises Bioinformáticas, páxinas 79-80). Primeiro faciamos unha avaliación destes tres paquetes para valorar a súa sensibilidade e a súa especificidade. Para isto dispoñiamos de datos de xenotipado de SNPs acadados mediante a plataforma Affymetrix das mostras que secuenciabamos neste traballo, obtidos do estudo de Carrera e cols. 2012, ademais dos datos da secuenciación de Sanger do xene *DISC1*, descrito anteriormente no Estudo I.

Usando estes datos estimabamos o número de variantes identificadas polos algoritmos pero que non eran reais (falsos positivos) baseándonos en 842 posicións nas que as tres mostras que formaban a *pool* eran homocigotas para o alelo de referencia e 134 posicións nas que as tres mostras dunha mesma *pool* eran homocigotas para o alelo alternativo. Logo calculabamos o número de variantes reais que non eran capaces de identificar os paquetes informáticos (falsos negativos) para a cal utilizabamos 580 posicións cun heterocigoto e dous homocigotos para o alelo de referencia na mesma *pool*, e 260 posicións cun heterocigoto e dous homocigotos para o alelo alternativo.

O principal obxectivo desta análise era reducir ó máximo a taxa de falsos positivos, intentando conservar unha sensibilidade considerable. Isto principalmente era debido a que os falsos negativos daban lugar a unha redución da potencia dos tests acumulativos, pero unha baixa especificidade podía dar lugar a conclusións cualitativas erróneas.

Utilizando tanto BioScope como GATK alcanzabamos unha elevada especificidade aínda que a sensibilidade era mellorable (Táboa 23), polo que optabamos por utilizar unha combinación das dúas ferramentas de dous xeitos. A primeira opción consideraba unha variante certa se era identificada por un ou outro paquete informático, mentres que a segunda consideraba unha variante como real se era identificada por ámbolos dous paquetes informáticos. Esta última opción acadaba valores moi elevados de sensibilidade e especificade, pero aínda así valorabamos utilizar o *software* VarScan, xa que se deseñaba especificamente para traballar con *pools*, co fin de aumentar a

especificidade e a sensibilidade. Usabamos VarScan modificando algún dos seus parámetros: o umbral de frecuencia mínimo para identificar unha variante e a presenza do alelo nas dúas cadeas. Primeiro utilizabamos un valor de 2,5% de umbral de frecuencia mínimo obtendo unha taxa de falsos positivos de 0,0218 e unha de falsos negativos 0,0179; mentres que se usabamos un umbral do 5% a taxa de falsos positivos e falsos negativos eran de 0,003 e 0,0214 respectivamente (Táboa 23). Con este último valor, a taxa de falsos positivos era considerablemente menor que co anterior, aínda que aumentaba a taxa de falsos negativos, pero non dunha forma considerable, polo que decidiamos utilizar o umbral do 5% para as posteriores análises.

Estudando a fondo os falsos positivos, víamos que se considerabamos como unha variante a presenza de ambos alelos nas dúas cadeas máis dunha vez o valor da sensibilidade diminuía mentres que o da especificidade mantíñase. Polo que decidíamos incluír un novo filtro segundo o cal considerabamos unha variante como verdadeira se non existía nesgo entre as lecturas dos dous alelos das dúas cadeas (aplicando Test de Fisher considerando un valor- $P < 0,0001$), pero só aplicado cando a frecuencia dun alelo é inferior a 1/6 dentro da pool, xa que viamos que a unha frecuencia inferior a 1/6 a taxa de falsos positivos era elevada. Deste xeito conseguíamos unha especificidade moi alta (0,998) e unha sensibilidade de 0,8929. A vista destes resultados e segundo o obxectivo inicial da redución dos falsos positivos, decidíamos utilizar o VarScan cos filtros comentados anteriormente para a identificación das variantes, xa que aínda que especificidade era un pouco menor (0,998) respecto á cando usabamos Bioscope e GATK (0,999), a sensibilidade era moito maior usando VarScan (0,8929 fronte a 0,8095 de GATK e VarScan).

	BioScope	GATK	BioScope ou GATK ^a	BioScope e GATK ^b	VarScan				Frec. $\geq 5\%$ Strand2=2 Frec. $< 0,166$ $p < 0,0001^f$
					Frec. $\geq 2,5\%$ Strand2=0	Frec. $\geq 2,5\%$ Strand2=2 ^c	Frec. $\geq 5\%$ Strand2=0 ^d	Frec. $\geq 5\%$ Strand2=2 ^e	
Sensibilidade	0,8786	0,8702	0,9488	0,8095	0,9821	0,9476	0,9786	0,9476	0,8929
Especificidade	0,9919	0,9975	0,9904	0,999	0,9782	0,9803	0,997	0,997	0,998
Taxa FN	0,1214	0,1298	0,0512	0,1905	0,0179	0,0524	0,0214	0,0524	0,1071
Taxa FP	0,0081	0,0025	0,0096	0,001	0,0218	0,0197	0,003	0,003	0,002

Táboa 23. Taxa de falsos positivos (FP) e falsos negativos (FN) segundo os distintos paquetes informáticos e filtros utilizados. En negra están indicados os valores do criterio seleccionado.

^a Considerabamos unha variante se era identificada por un ou outro paquete informático.

^b Considerabamos unha variante se era identificada por ámbolos dous paquetes informáticos.

^c Considerabamos unha variante se aparecen os dous alelos nas dúas cadeas e se ten unha frecuencia mínima do alelo alternativo maior ou igual ó 2,5%.

^d Fixabamos un umbral de frecuencia mínima do alelo alternativo igual ou maior ó 5%.

^e Considerabamos unha variante se aparecían os dous alelos nas dúas cadeas e se tiñan unha frecuencia mínima do alelo alternativo maior ou igual ó 5%.

^f Considerabamos unha variante se aparecían os dous alelos nas dúas cadeas, se tiñan unha de frecuencia mínima do alelo alternativo maior ou igual ó 5% e se non existía nesgo entre as lecturas das dúas cadeas dos dous alelos (aplicando Test de Fisher considerando un valor- $P < 0,0001$) pero só aplicado cando a frecuencia dun alelo era inferior a 1/6.

Resultados

Unha vez feita a escolma do paquete informático e as condicións deste a utilizar, faciamos comprobacións do funcionamento de VarScan. A primeira delas consistía na comparación das frecuencias observadas do alelo alternativo utilizando este *software* coas frecuencias esperadas obtidas mediante os datos de xenotipado e secuenciación de Sanger dos que dispoñiamos. Deste xeito obtiñamos unha correlación elevada entre ambas frecuencias ($R^2 = 0,9844$) (Fig. 30).

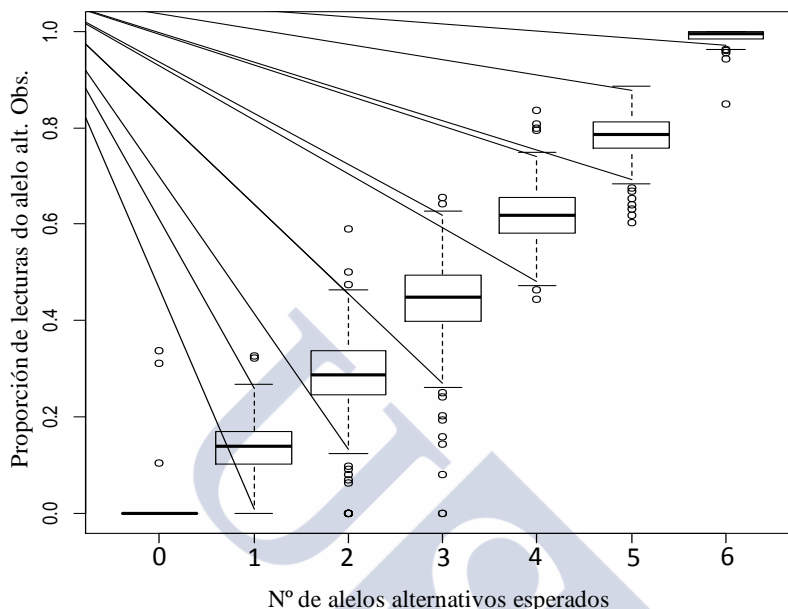


Figura 30. Diagrama de caixas e bigotes no que se comparaba a frecuencia dos alelos alternativos observados mediante a secuenciación co Solid e o número dos alelos alternativos esperados.

Outra das comprobacións era ver a eficacia de VarScan á hora de detectar SNVs raras funcionais, xa que estas estaban enriquecidas en erros de secuenciación (MacArthur and Tyler-Smith 2010). Para isto xenotipabamos, mediante Sequenom, 57 SNVs codificantes de sentido incorrecto presentes en tres ou menos *pools*. Estas SNVs aparecían en 67 de 71 *pools* totais, polo que obtiñamos así unha taxa de falsos positivos dun 5%.

4.3.4. Detección de variantes

Utilizando o VarScan coas condicións comentadas anteriormente, identificabamos 510 posibles SNVs, das cales unha era unha variante sen sentido, 294 codificantes de sentido incorrecto (58%) e 215 sinónimas (42%). A SNV sen sentido (rs148806955), que era identificada nunha *pool* de casos e nunha de controis, estaba recollida nas base de datos públicas e está presente no xene *PCMI* (Táboa 24. Ver material suplementario).

Das variantes identificadas, 118 eran posible ou probablemente prexudiciais segundo PolyPhen2. Asumindo que cada unha das variantes identificadas pertencía a unha única mostra das que forman a *pool*, vimos que había 108 SNVs que tiñan unha MAF menor ou igual ó 5%, é dicir, estaban presentes en menos de 16 *pools*. Con estes datos facíamos unha comparación da relación de variantes non sinónimas fronte a sinónimas a diferentes frecuencias. Vimos que existía un exceso de variantes non sinónimas cunha MAF menor ó 1%, ou o que é o mesmo, que estaban presentes nunha, dúas ou tres *pools*; estas diferenzas tamén as detectabamos entre as variantes que

detectabamos nun rango comprendido entre 4 e 15 *pools*, que era equivalente a que tiveran unha frecuencia do alelo menor entre o 1% e o 5%. Aplicando un test de tendencia de Conchran-Armitage obtiñamos un $P=0,0052$ (Fig.31).

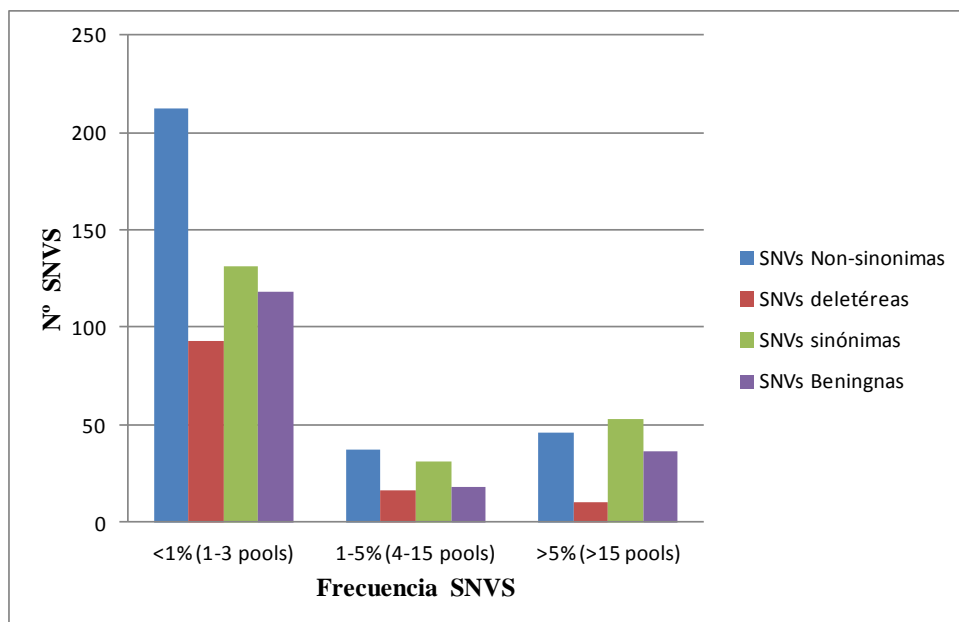


Figura 31. Distribución das SNVs por frecuencia e por predición bioinformática da funcionalidade segundo PolyPhen2. Definíamos as SNVs deletéreas como aquelas que eran consideradas como probable ou posiblemente prexudiciais segundo PolyPhen2.

A continuación separabamos as variantes non sinónimas segundo eran preditas por ser prexudiciais ou benignas e repetimos as análises. Primeiro comparabamos as variantes prexudiciais fronte a SNVs sinónimas e obtiñamos un $P = 2,2 \times 10^{-4}$ mediante un test de tendencia de Conchran-Armitage. Por outra banda, con este mesmo tipo de test obtiñamos un $P= 0,184$ se analizabamos a relación entre as variantes sinónimas e benignas. En función destes datos podíamos concluír que o exceso era causado polas variantes funcionais. A maiores atopabamos que este efecto non só era causado por variantes cunha $MAF < 1\%$, xa que comparando as SNVs prexudiciais fronte a sinónimas e ademais que estiveran presentes entre 4 e 15 *pools* coas que estaban en máis de 15 *pools*, viamos que existía un exceso significativo das variantes funcionais cunha frecuencia menor a pesares de que o número de SNVs era moi baixo ($P = 0,024$ aplicando o test exacto de Fisher). Polo tanto, a vista destes resultados, para levar a cabo o estudo de asociación centrábamonos en analizar as variantes que estaban presentes en menos de 16 *pools*.

4.3.5. Estudo de asociación

Entre as variantes atopadas, 33 eran non sinónimas, preditas como prexudiciais e exclusivas de casos, as cales eran xenotipadas en toda a nosa mostra. Destacar que dúas delas non se podían xenotipar xa que cando deseñabamos as *plexes*, a ferramenta Assay Design v4 non era capaz de incluílas no deseño. Nove destas variantes estaban presentes en mostras adicionais, entre as que se atopaban dúas que aparecían en dous casos a maiores pero ausentes en controis. Ademais, unha SNV que identificabamos en dous casos e en ningún control na mostra que secuenciabamos, non aparecía nas mostras restantes (Táboa 25). Destacar que estas tres variantes non se atopaban nin entre os 107 individuos de poboación ibérica de España que formaban parte da fase 3 de base datos

Resultados

1000 Genomes Project (McVean et al. 2012) , nin entre os 75 controis sans españois incluídos na da base de datos *CIBERER Exome browser* deseñada polo Centro de Investigación Biomédica en Rede de Enfermidades Raras Español (<http://www.ciberer.es/bier/exome-server>). No caso da variante que estaba no xene *AKAP9*, tamén podíamos confirmar a súa ausencia en 58 mostras españolas adicionais que formában parte dunha base de datos interna da Fundación Pública Galega de Medicina Xenómica, nas cales se buscaban mutacións en xenes involucrados en trastornos arritmoxénicos do corazón mediante secuenciación de nova xeración utilizando un panel que incluía *AKAP9* (Os pacientes que presentaban síndrome de QT largo eran excluídos antes da análise, xa que mutacións no xenes *AKAP9* estaban implicadas nesta síndrome). Resaltar que esta SNV alcanzaba un valor de significación nominal, cun $P = 0,031$ despois de aplicar un test exacto de Fisher.

Chr:posición ^a	Xene	dbSNP 141	Predición PolyPhen2	Frec. Secuenciación ^b	Frec. Xenotipado	T1000G ^c (N = 5008)	NHLBI ESP ^d (N = 6503)
Chr1:66458661G>C	<i>PDE4B</i>	rs115074556	Posiblemente prexudicial	2/153 CA, 0/153 CO	0/419 CA, 0/398 CO	17	50
Chr2:185803657T>G	<i>ZNF804A</i>	rs199741095	Probablemente prexudicial	1/153 CA, 0/153 CO	1/419 CA, 0/398 CO	0	2
Chr7:91631849A>G	<i>AKAP9</i>	Ausente	Posiblemente prexudicial	1/153 CA, 0/153 CO	3/419 CA, 0/398 CO	0	0

Táboa 25. SNVs que aparecían en máis dun caso e ausentes nos controis.

^a Baseado na versión do xenoma GRCh37/hg19.

^b CA=pacientes de esquizofrenia; CO= controis.

^c Mostras con variantes no Proxecto 1000 Xenomas.

^d Mostras con variantes no *NHLBI GO Exome Sequencing*.

Unha vez que viamos que os tests sobre variantes individuais non alcanzaban valores significativos, aplicabamos catro tests de análises de acumulación de variantes en cada xene individual, no conxunto total de xenes e nos xenes que formaban parte do interactoma de *DISC1*. *NRXN1* presentaba 10 SNVs raras, 8 delas aparecían unha única vez en controis, unha aparecía unha única vez en casos, e a restante , rs56086732, estaba presente en catro *pools* de casos e dous de controis (Táboa 26). Con estas variantes atopabamos que este xene acadaba o resultado máis significativo despois de levar a cabo o test acumulativo C-Alpha cun $P = 0,028$ (Táboa 27).

Chr:Posición ^a	dbSNP135	Frec. CA ^b	Frec. CO ^b	Predicción Polyphen2 ^c	Puntuación Polyphen2
chr2:50699553C>T		0/51	1/51	D	0,998
chr2:50765421T>C		0/51	1/51	D	0,999
chr2:50765649T>G		0/51	1/51	D	1
chr2:51254790G>C	rs200171245	0/51	1/51	P	0,793
chr2:51254843T>C	rs200792504	0/51	1/51	P	0,564
chr2:50463941G>T		0/51	1/51	B	0,088
chr2:50765589T>C	rs200074974	0/51	1/51	B	0,067
chr2:50780081A>C		0/51	1/51	B	0,141
chr2:51255090G>A	rs199784029	1/51	0/51	B	0,029
chr2:50765412G>T	rs56086732	4/51	2/51	P	0,799

Táboa 26. Listaxe das variantes que atopabamos en *NRXN1*.

^a Baseado na versión do xenoma GRCh37/hg19.

^b CA, pacientes de esquizofrenia; CO, controis.

^c Predicción de PolyPhen2. B= Benigna; P= Posiblemente prexudicial; D= Probablemente prexudicial.

Outros dous dos tests que levabamos a cabo amosaban dúas asociacións nominais. A primeira delas alcanzaba un $P=0,031$ para o xene *NRXN1* despois de aplicar o test exacto de Fisher tendo en conta aquelas variantes que aparecían só nunha única mostra (*singletons*); e a segunda era no xene *ZNF804A* cun $P= 0,038$ unha vez feito o test C-Alpha ponderado (para esta ponderación utilizabamos as puntuacións obtidas coas predicións feitas con PolyPhen2), baseado nun total de 11 variantes raras non sinónimas. Aplicando estes mesmos tipos de análises ó conxunto de xenes do interactoma de *DISC1* e a tódolos xenes xuntos, ningún acadaba un valor significativo, de feito, o valor-P máis baixo alcanzábase cando analizabamos tódolos xenes conxuntamente aplicando o test C-Alpha ponderado sendo o $P= 0,40$.

Resultados

	Non- sinónimas raras					Singletons ^a
	Nº SNVs	Valor-P C_Alpha	Valor-P C_Alpha Ponder.	Valor- P WSS	Valor-P WSS_Polyp.	Test Fisher ^b
<i>MACF1</i>	50	0,408	0,453	0,716	0,63	-
<i>PDE4B</i>	2	0,499	0,496	0,239	0,237	1
<i>DISC1</i>	7	0,204	0,291	0,201	0,413	0,243
<i>NRXN1</i>	10	0,028	0,051	0,968	0,945	0,031
<i>VRK2</i>	2	1	1	0,873	0,767	1
<i>ZNF804A</i>	11	0,124	0,038	0,244	0,191	0,118
<i>TNIK</i>	2	0,597	0,597	0,622	0,462	1
<i>SLC39A8</i>	2	0,432	0,432	0,824	0,886	-
<i>KIF3A</i>	1	-	-	-	-	-
<i>DST</i>	72	0,724	0,351	0,703	0,841	-
<i>AKAP9</i>	21	0,295	0,831	0,663	0,941	-
<i>RELN</i>	19	0,447	0,429	0,762	0,556	-
<i>CNTNAP2</i>	8	0,531	0,343	0,198	0,204	-
<i>PCMI</i>	14	0,372	0,956	0,253	0,617	1
<i>SPTAN1</i>	11	1	1	0,626	0,756	1
<i>FEZ1</i>	3	1	1	0,707	0,787	1
<i>NDE1</i>	6	0,327	0,464	0,07	0,763	1
<i>NDEL1</i>	4	1	1	0,336	0,19	1
<i>TCF4</i>	2	1	1	0,751	0,751	1
<i>SCL39A3</i>	2	0,494	0,494	0,253	0,253	0,495
Interactoma ^c	193	0,542	0,582	0,601	0,929	-
Total xenes	249	0,512	0,398	0,668	0,955	-

Táboa 27. Valores de significación dos diferentes tests de carga aplicados. En negraña están indicados os valores que alcanzaban un resultado significativo.

^a Variantes que só aparecían nunha única mostra.

^b Test exacto de Fisher de dúas colas.

^c Grupo de xenes que interactuaban con *DISC1*.

4.4. ESTUDO IV: RESECUENCIACIÓN DO XENE NRXN1

4.4.1. Introducción

NRXN1 é un xene que codifica para unha proteína implicada na adhesión sináptica neuronal. Está localizada na rexión presináptica e interacciona con neuroliquinas (NLGN) postsinápticas nas sinapses inhibitoras e excitadoras no cerebro (Sudhof 2008). A primeira vez que este xene se relacionaba coa esquizofrenia foi no 2008, cando Kirov e cols. atopaban nun doente unha microdeleción que afectaba a rexión promotora e ó primeiro exón deste xene (Kirov et al. 2008). Estudos posteriores reforzaban este descubrimento (Walsh et al. 2008; Kirov et al. 2009b; Rujescu et al. 2009; Rees et al. 2014c). Unha das características máis destacables destas variantes é que son CNVs non recorrentes, ademais de ser unha das poucas que afectan a un único xene, polo que *NRXN1* está implicado claramente no risco a padecer esquizofrenia.

Contraditoriamente ó publicado ata o de agora, no Estudo III atopabamos que variantes raras codificantes de sentido incorrecto en *NRXN1* eran máis frecuentes en controis. Coa finalidade de intentar analizar en maior detalle os resultados obtidos, secuenciabamos as rexións exónicas de *NRXN1* na totalidade de mostras restantes que formaban parte da nosa colección, é dicir, incluíamos neste estudo aquelas mostras que non foran secuenciadas mediante o SOLID4 System. Das variantes atopadas mediante a secuenciación seleccionabamos aquelas que eran codificantes de sentido incorrecto cunha frecuencia inferior ó 1% e intentábamolas confirmar mediante Sequenom. Por último, facíamos unha busca bibliográfica na procura de estudos adicionais que analizasen *NRXN1* en mostras diagnosticadas con esquizofrenia e en controis para así conseguir suficientes datos para poder levar a cabo unha meta-análise.

4.4.2. Análise de cobertura de NRXN1

Neste estudo secuenciabamos, mediante o Ion Proton, os exóns do xene *NRXN1* coa finalidade de identificar e analizar a variación rara deste xene. Para isto analizabamos 384 doentes de esquizofrenia e 384 controis, en 96 *pools* de casos e outros tantos de controis, é dicir, cada *pool* constaba de 4 mostras. A rexión secuenciada era de 10,7 kb por mostra, a cal estaba cuberta por un total de 77 amplicóns.

Antes de iniciar a identificación e análise das variantes, facíamos unha análise da cobertura do xene secuenciado. Mediante esta metodoloxía de secuenciación obtiñamos unha cobertura media de 8501X (Táboa 28). Mediante a comparación entre a cobertura de casos e controis viamos que non existían diferenzas significativas entre ámbolos dous grupos de mostras, de feito obtiñamos un $P = 0,95$ despois de aplicar o test de Mann-Whitney (Fig.32 e Fig.33).

Resultados

	Cobertura media	Proporción da rexión diana cuberta					
		25x	50x	75x	100x	125x	150x
Mostra total	8501	0,9853	0,9700	0,9573	0,9472	0,9387	0,9324
Casos	8522	0,9877	0,9728	0,9609	0,9512	0,9426	0,9356
Controis	8201	0,9827	0,9672	0,9537	0,9432	0,9348	0,9293

Táboa 28. Datos da cobertura media e da porcentaxe da rexión de interese cuberta do total da mostra, dos casos e dos controis.

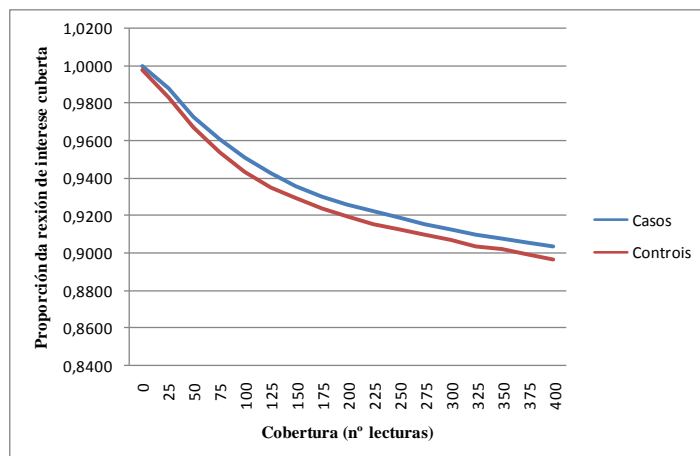


Figura 32. Porcentaxe de rexión de interese cuberta comparando os casos cos controis.

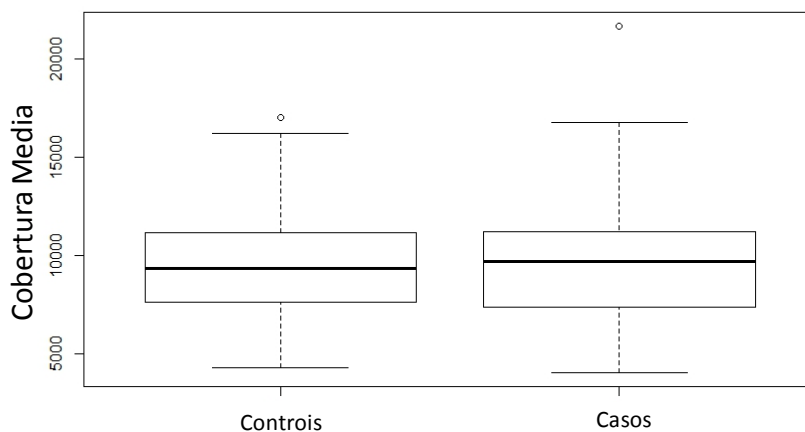


Figura 33. Diagrama de caixas e bigotes no que se comparaba a distribución de coberturas medias dos casos fronte ós controis.

Se comparábase os casos e os controis por amplicón (Fig. 34) vimos que si que existían diferenzas na cobertura entre os distintos amplicóns pero seguían os mesmos patróns tanto nos casos como nos controis. De xeito que se enfrontabamos a cobertura media dentro de cada amplicón dos casos fronte ós controis obtínhamos un factor de correlación de R^2 (cadrado do coeficiente de correlación de Pearson) = 0,998 (Fig. 35).

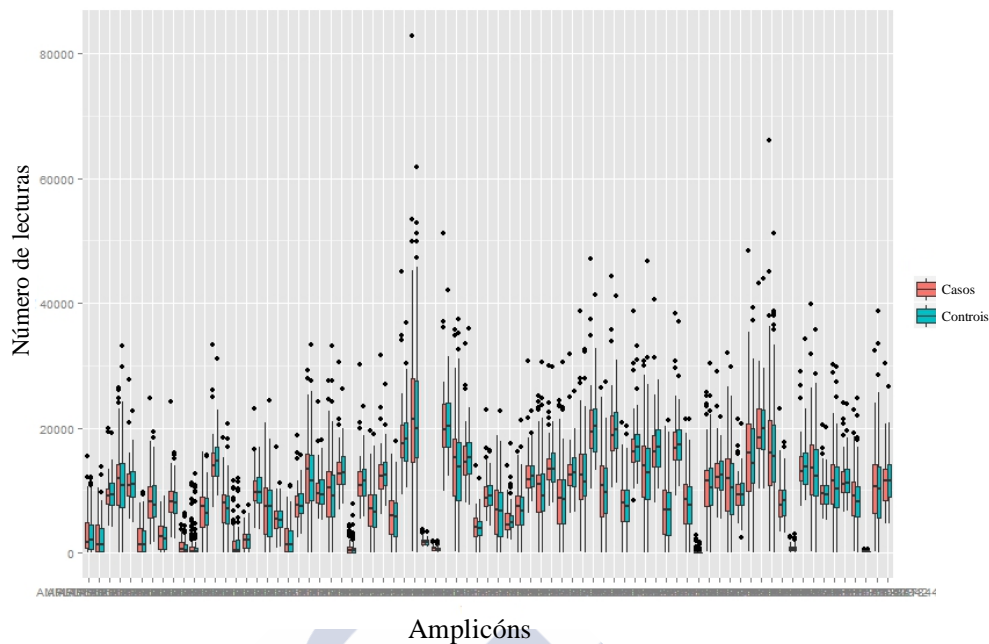


Figura 34. Diagrama de caixas e bigotes no que se comparaba a cobertura de casos e controis nos 77 amplicóns secuenciados.

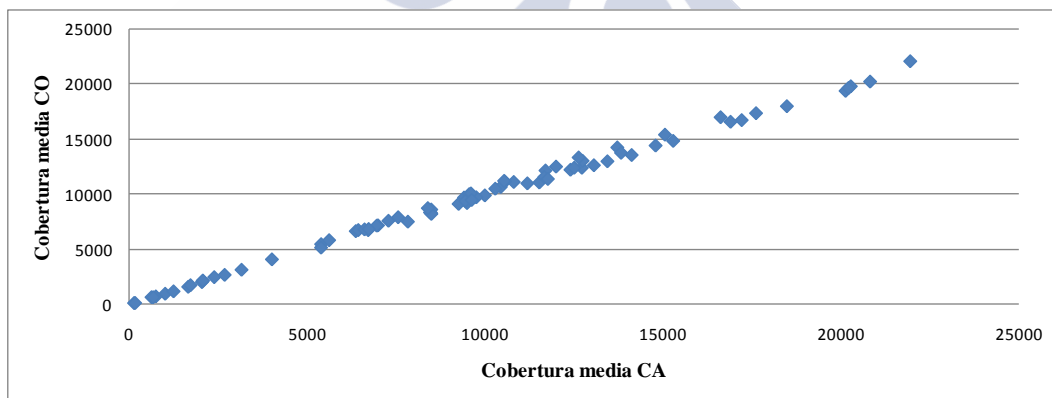


Figura 35. Correlación que existía entre a cobertura dos casos fronte ós controis para cada un dos 77 amplicóns.

4.4.3. Detección de variantes e confirmación mediante Sequenom

Para a identificación das SNVs utilizabamos *Torrent Variant Caller*, o cal nos permitía detectar 59 variantes, das cales 37 eran variantes non sinónimas sendo unha delas unha SNV que provocaba unha perda dun codón de parada. As restantes 22 eran variantes sinónimas. Como no anterior estudo, para ver as predicións funcionais das variantes codificantes de sentido incorrecto utilizabamos PolyPhen2, obtendo que 12 destas SNVs eran benignas, 7 eran posiblemente prexudiciais e 17 eran probablemente prexudiciais. Destacar que 10 destas variantes identificábamolas no estudo anterior.

De tódalas variantes atopadas, seleccionabamos a sen sentido mailas non sinónimas que estaban a unha frecuencia inferior ó 1%, é dicir, que estaban presentes en 15 ou menos pools, para confirmar mediante Sequenom. En total

Resultados

eran 32 SNVs que estaban presentes en 63 *pools*. Unha vez feito o deseño co *Assay Design v4*, estas SNVs eran agrupadas en tres *plexes* para o seu procesado. Das variantes escollidas 17 eran confirmadas (53%) sendo as restantes falsos positivos (47%) (Táboa 29). Esta elevada taxa de falsos positivos era esperada xa que seleccionabamos uns filtros pouco esixentes cando utilizabamos o *Torrent Variant Caller* para a identificación de variantes, co fin de identificar tódalas variantes existentes.

Chr:Posición ^a	Confir. ^b	dbSNP138	Frec. CA ^c	Frec. CO ^c	Predición Polyphen2 ^d	Puntuación Polyphen2	T1000G ^e	ESP6500 ^f	ExAC ^g
chr2:50149084A>G	n		0/384	1/384	S				
chr2:50149311C>A	n		0/384	1/384	D	1			
chr2:50149369C>T	s	rs200604893	1/384	0/384	B	0,087		0,0002	0,0001
chr2:50280432G>A	n		2/384	0/384	P	0,492			
chr2:50463928A>T	s		1/384	0/384	D	0,999			
chr2:50574086A>G	n		0/384	1/384	B	0			
chr2:50699488G>C	n		0/384	1/384	P	0,781			
chr2:50723114A>C	n		4/384	5/384	D	0,981			
chr2:50724562A>C	n		1/384	0/384	D	0,949			
chr2:50724569C>T	n		1/384	1/384	D	0,993			
chr2:50724753A>G	s		0/384	1/384	D	1			
chr2:50724765G>A	s		1/384	1/384	D	1			
chr2:50724817G>A	s	rs199784139	1/384	1/384	D	0,3845	0,0002	0,001	0,0006
chr2:50733693G>T	s		0/384	1/384	D	1			
chr2:50758365C>G	n		0/384	2/384	P	0,818			
chr2:50758515C>T	s		0/384	1/384	B	0,138			
chr2:50765405C>T	s		0/384	2/384	B	0,354			
chr2:50765538C>T	n	rs201731350	1/384	0/384	B	0,011			
chr2:50779731G>C	n		2/384	2/384	D	0,999			
chr2:50780005C>A	n		4/384	6/384	D	1			
chr2:50847195G>A	s	rs78540316	1/384	3/384	D	0,999	0,0014	0,0028	0,0037
chr2:50848364C>T	n		0/384	1/384	D	0,98			
chr2:50850745C>T	n		0/384	1/384	D	0,992			
chr2:51149819G>A	s		1/384	0/384	D	0,987			
chr2:51253562T>C	s	rs144049982	1/384	0/384	B	0,001	0,0012	0,0011	0,0008
chr2:51254756G>T	n		1/384	0/384	B	0,106			
chr2:51254843T>C	s	rs200792504	1/384	0/384	P	0,597	0,0002	0,0009	0,0007
chr2:51255001C>A	s		0/384	2/384	D	0,917			
chr2:51255090G>A	s	rs199784029	0/384	1/384	B	0,087	0,0014	0,0014	0,0008
chr2:51255110G>C	s	rs200184823	1/384	0/384	D	0,999	0,0002	0,0002	0,0004
chr2:51255210G>A	s		2/384	0/384	P	0,899			
chr2:51255329C>T	s		1/384	0/384	P	0,837			

Táboa 29. Táboa das variantes non sinónimas que se xenotipaban para a súa confirmación. S=Stoploss.

^a Baseado na versión do xenoma GRCh37/hg19.

^b Confirmación mediante Sequenom. s=Si; n=Non.

^c CA, pacientes de esquizofrenia; CO, controis.

^d Predición de PolyPhen2. B=Benigna; P= Posiblemente prexudicial; D=Probablemente prexudicial.

^e Frecuencia das variantes no *1000 Genomes Project*.

^f Frecuencia das variantes no *NHLBI GO Exome Sequencing*.

^g Frecuencia das variantes no *Exome Aggregation Consortium*.

4.4.4. Estudo de asociación e meta-análise

Coas variantes confirmadas anteriormente aplicabamos os mesmos tests de carga que no Estudo III. En total eran 17 SNVs presentes en 25 mostras, das cales 13 eran controis e 12 casos (Táboa 29). Ningún dos tests aplicados acadaba valores significativos (Táboa 30).

	Calpha	Calpha_w	WSS	WSS_PolyPhen2	Nº SNVS
<i>NRXNI</i>	0,397	0,408	0,536	0,513	17

Táboa 30. Valores de significación dos diferentes tests de carga aplicados.

A vista dos resultados do Estudo III, no que atopabamos un exceso de variantes raras codificantes de sentido incorrecto no xene *NRXNI*, levabamos a cabo unha meta-análise na que incluíamos os datos do presente estudo, os do traballo anterior ademais de outros estudos publicados e que atopabamos despois de facer unha revisión bibliográfica. Mediante esta busca identificabamos sete artigos nos que se secuenciaban os exóns deste xene en doentes esquizofrénicos e en controis. Dous destes artigos non dispoñían dos datos dos que nós precisabamos (Need et al. 2012; Kenny et al. 2013), pero os outros cinco artigos si. No primeiro deles, Hu e cols. amosaban o número de variantes codificantes de sentido incorrecto ou que causaban unha perda de función a unha frecuencia inferior ó 5% nunha mostra descubrimento e nunha mostra réplica (Hu et al. 2014). Por outra banda, no artigo de Purcell e cols. podiamos obter os datos de variantes non sinónimas a unha MAF < 0,1% identificadas como prexudiciais mediante cinco métodos distintos de predicións bioinformáticas (Purcell et al. 2014). Girard e cols. recollían tódalas variantes que identificaban despois de secuenciar 40 xenes, entre os que se atopaba *NRXNI* (Girard et al. 2015). Por último, Muhleisen e cols. por un lado, e Gauthier e cols. por outro, presentaban nos seus respectivos traballos todas as SNVs non sinónimas que foran capaces de identificar (Gauthier et al. 2011; Muhleisen et al. 2011). Polo tanto, para esta fase incluíamos os datos dispoñibles dos traballos de Hu e cols. e os de Purcell e cols., os datos das variantes codificantes de sentido incorrecto preditas como prexudiciais e que ademais estiveran a unha frecuencia menor ó 1% dos outros estudos, e aquelas identificadas nos dous traballos levados a cabo por nós e que cumprían estas condicións. Utilizando todos estes datos faciamos a meta-análise empregando distintos criterios de frecuencia (variantes *singleton*, é dicir, unha variante que aparece nun único individuo, ou variantes cunha MAF <1%) e/ou funcionalidade (Táboa 32).

Resultados

Autor (ano)	Nº Ca	Nº Co	Miss. Singl	Miss. Singl. Prex.	Miss. MAF < 1%	Miss MAF < 1% Prex.	NS Sing	NS Singl Prex./LOF	NS MAF<1%	NS MAF < 1% Prex./LOF
Gauthier e cols.(2011)	143	190	3/2	1/1	6/3	4/2	4/2	2/1	7/3	5/2
Mühleisen e cols.(2011)	94	94	2/0	0/0	3/3	1/1	2/0	0/0	3/3	1/1
Hu e cols._Inicial(2013)	513	611	10/25	10/25	10/25	10/25	10/25	10/25	10/25	10/25
Hu e cols._Réplica(2013)	409	305	19/16	19/16	19/16	19/16	19/16	19/16	19/16	19/16
Purcell e cols.(2014)	2536	2543	23/34	23/34	23/34	23/34	23/34	23/34	23/34	23/34
Estudo III(2014)	153	153	1/8	0/5	1/8	0/5	1/8	0/5	1/8	0/5
Girard e cols.(2015)	235	233	6/2	4/1	6/4	4/1	7/2	5/1	7/4	5/1
Estudo IV(2015)	384	384	7/4	5/2	12/13	10/9	7/4	5/2	12/13	10/9
Valor-P			0,0906	0,0473	0,0469	0,0678	0,1303	0,073	0,0689	0,0994

Táboa 32. Resultado da meta-análise segundo os criterios de frecuencia do alelo menor (MAF) ou funcionalidade. En negriña están indicados os valores significativos. Singl=*Singleton* .LOF=variantes de perda de función (das súas siglas en inglés, *Loss Of Function*); NS =Non-sinónimas; Prex.=Prexudicial. Miss=Variantes codificantes de sentido incorrecto.

De tódalas posibles combinacións o valor máis significativo acadábase cando considerábase as SNVs codificantes de sentido incorrecto cunha frecuencia do alelo menor inferior ó 1% alcanzando un $P = 0,0469$ cun $OR = 0,74$ e un intervalo de confianza do 95% = (0,55-0,99). Ademais non existía evidencia de heteroxeneidade entre os diferentes estudos ($I^2 = 27,8\%$, cun valor-P do test $Q = 0,20$), analizando un total de 4467 casos e 4513 controis (Fig. 36). Outra das combinacións que acadaba un valor significativo era aquela na que só se tiñan en conta as SNVs codificantes de sentido incorrecto, prexudiciais e que aparecían nun único individuo cun $P = 0,0473$ (Táboa 30)

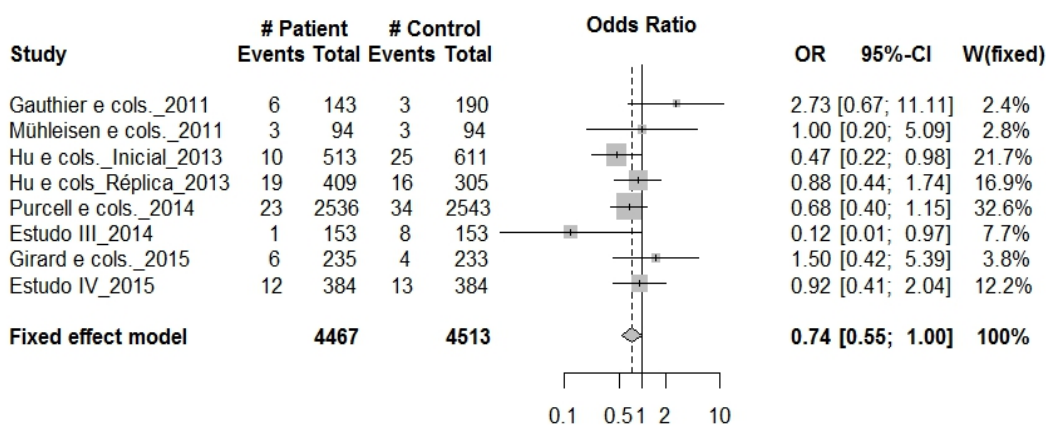


Figura 36. Meta-análise dos estudos de secuenciación de *NRXN1* feitos sobre mostras de esquizofrenia e controis. Os OR (en escala logarítmica) para os estudos individuais están representados polas caixas en gris cuxo tamaño é proporcional ó peso de cada traballo na meta-análise. As liñas horizontais representan os intervalos de confianza ó 95%. O resultado conxunto e o seu intervalo de confianza do 95% está representado polo rombo.



5. DISCUSIÓN





Discusión

Un dos grandes desafíos da xenética na actualidade é a busca de variantes xenéticas que causen unha maior susceptibilidade a padecer enfermidades complexas. As enfermidades psiquiátricas están englobadas dentro deste grupo de trastornos, sendo a esquizofrenia un exemplo claro. Esta é unha enfermidade con importantes alteracións nas diferentes funcións sociais, polo que é probable que os xenes responsables do seu desenvolvemento estean implicados en rutas cerebrais relacionadas co comportamento. Un destes posibles xenes é *DISC1* (*Disrupted in Schizophrenia 1*), xa que codifica unha proteína que parece estar implicada no desenvolvemento do sistema nervioso central. Como xa se comentaba na introdución, este xene parece que xoga un papel importante na regulación da neuroxénese tanto en embrións como en adultos (Mao et al. 2009), ademais de estar relacionado na formación e maduración da sinapse, na integración e migración neuronal (Miyoshi et al. 2003; Ozeki et al. 2003; Kamiya et al. 2005; Camargo et al. 2007; Duan et al. 2007; Brandon et al. 2009).

Ó longo de todos estes anos hai evidencias xenéticas de que este xene podería estar involucrado no desencadeamento da esquizofrenia. De feito, atopábanse tanto variacións estruturais ultra-raras de risco elevado (como a translocación recíproca alterando a *DISC1* nunha familia escocesa afectada por varios trastornos mentais entre os que estaba a esquizofrenia (St Clair et al. 1990; Millar et al. 2000; Blackwood et al. 2001)); como variantes comúns de baixo risco (Saetre et al. 2008; Schumacher et al. 2009)

A vista destes resultados, *DISC1* é un bo xene candidato a estudar para intentar atopar variantes que causen unha maior probabilidade de sufrir esquizofrenia. No primeiro estudo presentado neste traballo de tese (Estudo I) tiñamos como obxectivo principal, guiándonos pola hipótese comentada na introdución enfermidade común - variante rara, buscar variantes non sinónimas que presentasen unha frecuencia inferior ó 5% e que aumentasen o risco a padecer esquizofrenia por si mesmas.

Centrámonos nas variantes non sinónimas posto que existen evidencias de que este tipo de variantes teñen maior probabilidade de ser deletéreas. Unha destas evidencias foi aportada por un traballo publicado por Kryukov e cols. Nel os autores combinaban e analizaban datos de mutacións causantes de enfermidades mendelianas en humanos, datos de diverxencia entre humanos e chimpancé e datos de variación xenética na poboación humana moderna, mediante os cales chegaban a conclusión de que máis dun 50 % das mutacións codificantes *de novo* e non sinónimas tiñan efectos deletéreos, mentres que o 70% dos SNPs codificantes de sentido incorrecto eran lixeiramente deletéreos (Kryukov et al. 2007). Noutro estudo posterior no que analizaban os resultados obtidos de tódolos GWAS publicados ata o momento vían que, entre os SNPs asociados significativamente a unha determinada enfermidade ou trazo, os únicos que estaban sobre-representados na mostra eran os que ocupaban unha posición non sinónima, polo que este tipo de variantes eran os que tiñan maior probabilidade de estar asociadas a unha enfermidade (Hindorff et al. 2009).

Ademais de non sinónimas, tratábase de buscar variantes con frecuencia moi baixa (<5%) como responsables da enfermidade. Isto baseábase nun estudo feito por Gorlov e cols., no que utilizando datos do proxecto HapMap, ENCODE e o proxecto SeattleSNPs vían que existía unha gran proporción de SNPs funcionais entre os SNPs cunha MAF menor ó 5% (representaban máis do 40% do total dos SNPs). Ademais amosaban que había unha relación inversa entre a MAF e a proporción de SNPs non sinónimos que daban lugar a unha alteración na función da proteína segundo predicións *in silico*. Polo tanto, concluían que sería máis probable atopar unha variante causal entre aquelas que tivesen unha frecuencia baixa (Gorlov et al. 2008). A esta mesma

conclusión chegaban Barreiro e cols. despois de analizar o grao de diferenciación poboacional en case tres millóns de SNPs da fase dúas do proxecto HapMap (Barreiro et al. 2008). Evidencias sobre a importancia das variantes raras no desencadeamento dunha enfermidade complexa tamén foron aportadas por Kryukov e cols. no artigo comentado anteriormente (Kryukov et al. 2007). Un estudo posterior ós comentados confirmaba isto despois de secuenciar o exoma de 200 individuos daneses. Concretamente atopaban un exceso de máis 1,8 veces de SNPs non sinónimos fronte a sinónimos a unha frecuencia entre o 2% e o 5% (Li et al. 2010). Todas estas evidencias indicaban que a existencia desta maior proporción de variantes que alteran a proteína a baixa frecuencia na poboación era debido ó efecto da selección purificadora.

Segundo a análise da potencia estatística previa, o Estudo I permitiría identificar variantes raras cun risco moderado. Tendo en conta que a probabilidade de que unha variante cunha frecuencia x fose observada nunha mostra formada por K individuos é $1 - (1-x)^{2k}$, o noso tamaño de mostra na fase de secuenciación tiña un 75%-90% de probabilidade de atopar unha variante neutral cunha frecuencia de 1,5% ou 2,5%, respectivamente. Se unha variante confire un risco a padecer unha doenza, terá maior frecuencia nos casos, aumentando así a potencia para atopala. Por outro lado, co noso tamaño de mostra para o estudo de asociación, tiñamos un 80% de potencia para detectar unha asociación con variantes raras cunha frecuencia ó 5% e cun $OR > 2$ (Fig. 37) aplicando un test alélico exacto de Fisher cun nivel de significación $\alpha = 5\%$. Asumindo unha corrección de Bonferroni para 5 tests, este valor varía entre 1,8 e 4,5 a unha frecuencia de 5% ou 0,5%, respectivamente. Para aumentar a posibilidade de atopar variantes de risco, nos limitabámonos a buscar alelos posiblemente funcionais. Ademais de analizar as variantes a nivel funcional tamén faciamos unha análise como un conxunto. Por último, co fin de aumentar a evidencia faciamos unha meta-análise na que se incluía varios estudos atopados mediante buscas bibliográficas.

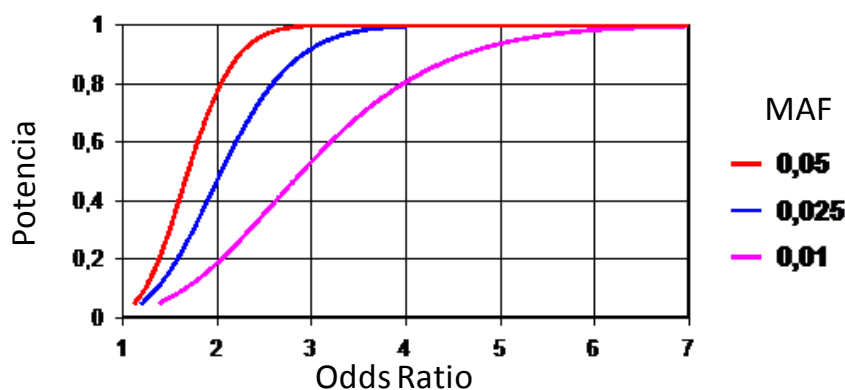


Figura 37. Análise da potencia estatística do estudo I. Cada liña representa a potencia acadada en función da OR e da frecuencia da variante a analizar.

Na primeira parte do Estudo I secuenciabamos tódolos exóns do xene *DISC1* en 47 casos e 47 controis nos que identificabamos seis posibles variantes funcionais. De tódalas variantes atopadas nesta primeira fase de secuenciación só catro eran variantes non sinónimas a unha frecuencia inferior ó 5%. Dúas delas estaban recollidas na base de datos de dbSNP (*build* 131) (rs28930675 e rs78792190); en canto as dúas restantes, a C565F era a primeira vez que se identificaba e a outra (E751Q) xa fora atopada anteriormente nun traballo publicado no 2008 por Song e cols. (Song et al. 2008). Cabe mencionar que neste traballo tamén identificaban as variantes rs28930675 e

Discusión

rs78792190, que os autores nomeaban como T453M e P423L respectivamente, e que nos tamén identificabamos na nosa mostra descubrimento (Táboa 17).

Na segunda fase do Estudo I xenotipabamos as 4 variables anteriores máis outras que estaban descritas na bibliografía, en 499 casos e 533 controis. Concretamente incluíamos variables ultra-raras (é dicir, cada variable só aparecía nun caso pero en ningún control) e non sinónimas descritas no traballo de Song e cols. Neste estudo analizaban inicialmente *DISC1* en 288 doentes esquizofrénicos e 288 controis nos que atopaban 8 variantes presentes exclusivamente nos casos. De todas estas, había 6 que estaban ausentes nos dez mil controis analizados na segunda fase deste traballo pasando a consideralas variantes ultra-raras. Este tipo de variantes amosaban un risco atribuíble combinado dun 2% (Song et al. 2008). Vendo isto decidíamos xenotipar estas variantes na nosa mostra, aínda que só podíamos incluír 4 delas (T603I, R37W, R418H e S90L), debido a problemas no deseño dos cebadores nas dúas restantes. A maiores incluíamos o SNP rs3737597 que estaba situado na rexión 3'UTR do xene *DISC1*, polo que podía chegar a ser funcional xa que variacións nesta rexión poderían alterar a expresión do xene. Outra das razóns para incluír este SNP era que ten unha frecuencia baixa (MAF= 0,04 en poboación CEU), ademais de estar asociado positivamente á doenza en estudos anteriores (Hennah et al. 2003; Saetre et al. 2008).

De tódalas variantes analizadas só unha alcanzaba valores significativos (E751Q, $P= 0,027$), aínda que deixaba de selo se aplicabamos a corrección de Bonferroni para múltiples tests. Esta variante estaba localizada na rexión codificante da isoforma L, a cal difire da isoforma Lv no uso dun sitio donante de *splicing* alternativo no exón 11, dando lugar a unha extensión de 22 aminoácidos adicionais no final do exón (Taylor et al. 2003). A variante E751Q era un destes aminoácidos adicionais. Como esperabamos, segundo a nosa hipótese inicial de que as variantes a baixa frecuencia eran variantes de risco suxeitas a selección purificadora, o aminoácido de referencia nesta posición estaba conservado en mamíferos placentarios, xa que a isoforma L só existe nestes (Fig. 38), e probablemente en peixes (Taylor et al. 2003). Sen embargo, non estaba claro se a similitude entre mamíferos e peixes representaba conservación por homoloxía ou realmente se debía a unha evolución converxente, xa que só a parte distal do exón 11 da isoforma L amosaba homoloxía coa secuencia de peixes.

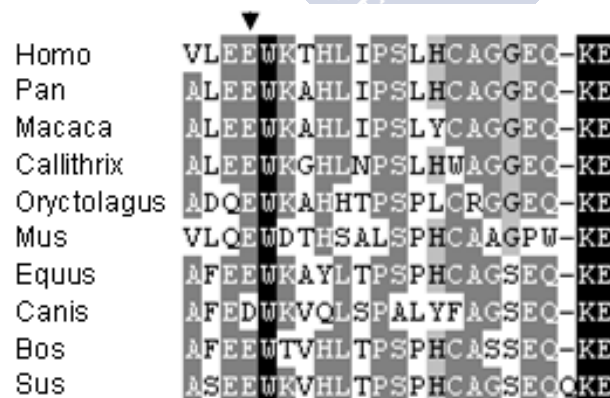


Figura 38. Alíñamento dos 22 aminoácidos exclusivos do exón 11 da isoforma L de *DISC1* en mamíferos placentarios. A frecha indica a posición da variante E751Q.

Por outro lado, predicións bioinformáticas utilizando SIFT e PolyPhen2 suxerían que este E751Q podería alterar a funcionalidade da proteína (Ng and Henikoff 2003;

Adzhubei et al. 2010). Dentro da proteína, esta variante está localizada na parte C-terminal. Esta rexión da proteína está relacionada coa localización de *DISC1* no citoesqueleto e no centrosoma (Morris et al. 2003; Kamiya et al. 2005; Millar et al. 2005a). Hai estudos que inferían que tanto o centrosoma como os microtúbulos, os cales forman parte do citoesqueleto, están implicados na proliferación de células proxenitoras, na migración e diferenciación neuronal (Feng and Walsh 2001; Badano et al. 2005), polo que unha alteración nesta parte da proteína podería dar lugar a problemas nestes procesos celulares.

Tamén comprobabamos se existía un exceso de portadores de variantes posiblemente funcionais raras nos casos. Na nosa mostra non atopabamos un exceso de variante raras nos casos, pero se aplicabamos este mesmo test ós datos de Song e cols. (2008), o cal non foi feito polos autores, vimos que existía un exceso de variantes nos casos. De feito, había 22 portadores entre os esquizofrénicos e 8 entre os controis, alcanzando un $P = 0,007$ despois de facer un test exacto de Fisher dunha cola, cun $OR = 2,89$ e cun intervalo de confianza do 95% de 1,21-7,64. Cabía destacar que neste artigo identificaban dúas variantes (R418H e P432L) que describían como ultra-raras xa que non aparecían en ningún dos mil controis que analizaban, pero que si aparecían nos controis que nós analizabamos. Este feito reflectía o problema da elección do umbral de frecuencia para definir variantes raras, un problema destacado no contexto da secuenciación de nova xeración.

Facendo unha busca bibliográfica identificabamos tres artigos dos que podíamos extraer os datos necesarios dunha variante rara posiblemente funcional que tamén era xenotipada nas nosas mostras e que nos permitía poder levar a cabo unha meta-análise. Esta variante era o SNP rs3737597. Este SNP non amosaba asociación coa esquizofrenia na nosa mostra, aínda que unha vez feita a meta-análise vimos que existía unha asociación positiva, polo menos en poboación europea (Fig. 23). Con tódolos datos recompilados, pertencentes a seis poboacións, vimos que a variante presentaba 271 alelos de risco entre os 2423 esquizofrénicos, estando 141 destes alelos presentes en 2112 doentes de orixe europea. Isto indicaba que o SNP estaba dentro do rango considerado como "cantidade moderada de evidencia" segundo as pautas suxeridas por Ioannidis e cols., as cales clasificaban as variantes en función da súa frecuencia na mostra a estudar. Segundo isto, unha variante considerábase como que amosaba unha "evidencia a gran escala" se presentaba máis de mil alelos na mostra; as que presentaban entre cen e mil alelos eran clasificadas como que amosaban "cantidade moderada de evidencia" (este é o noso caso); por último, se o alelo menor aparece menos de 100 veces considerábase variante con "con pouca evidencia" (Ioannidis et al. 2008).

O alelo máis común deste SNP era infra-transmitido a esquizofrénicos en familias finesas (Hennah et al. 2003), resultado que era confirmado nun estudo caso-control de tres mostras adicionais escandinavas (Saetre et al. 2008). No caso dun estudo sobre unha mostra xermana, vían que existía unha tendencia á asociación no mesmo senso ($P_{alélico} = 0,067$), aínda que os SNPs eran imputados en lugar de ser xenotipados (Schumacher et al. 2009). En cambio noutro artigo non atopaban asociación deste SNP coa esquizofrenia nunha mostra de orixe asiática onde o alelo menor alcanzaba unha frecuencia do 21%. Estes resultados enfrontados entre as poboacións europeas e asiáticas podían ser debidos as diferenzas existentes no desequilibrio de ligamento, xa que o SNP rs3737597 podería ser unha variante en desequilibrio de ligamento cunha variante funcional descoñecida en vez de ser a variante funcional relacionada coa esquizofrenia. Sen embargo, rs11122396, un SNP predito para alterar a unión dun miARN na rexión 3'UTR de *DISC1*, presenta valores elevados de desequilibrio de

ligamento co SNP rs3737597 en poboacións asiáticas ($r^2=0,90$) e europeas ($r^2=0,96$). Polo tanto, esta explicación non era válida se o SNP funcional responsable da asociación era rs11122396. A principal limitación desta meta-análise era a posible existencia de sesgo nas publicacións, xa que normalmente só se publican traballos con tamaños de mostra pequenos que obtiveran resultados positivos quedando sen publicar aqueles que teñen tamaños de mostra pequenos e resultados negativos. Isto non podía ser avaliado debido ó número reducido de estudos.

Do mesmo xeito que o SNP anterior, identificabamos a variante E751Q en varios estudos permitíndonos levar a cabo unha meta-análise. Con estes datos non podíamos confirmar unha asociación desta variante coa esquizofrenia. Esta variante atopábase nun estudo cunha frecuencia maior en individuos caucásicos afectados por esquizofrenia (6 de 576 alelos, 1,04%) ou trastorno bipolar (16 de 1008 alelos, 1,59%) que en controis (10 de 1152, 0,87%) (Song et al. 2008; Song et al. 2010), aínda que as diferenzas non acadaban valores significativos. Neste mesmo ano Moens e cols. publicaban un traballo onde, utilizando secuenciación de nova xeración, analizaban *DISC1* xunto con outros dez xenes que interactúan con el. Secuenciaban as rexións codificantes destes xenes en 486 doentes esquizofrénicos e 514 controis. Neste estudo atopaban a variante E751Q cunha MAF nos casos de 1,45% e 0,80% en controis sendo as diferenzas entre ambas cohortes non significativas. Resaltar que neste estudo esta variante acadaba un valor de OR próximo a 2, similar ó que atopaban Song e cols. (Song et al. 2008; Moens et al. 2011). No 2014 Thomson e cols. secuenciaban o xene *DISC1* identificando E751Q en 2 alelos de 460 doentes esquizofrénicos (0,4 %) e en 21 alelos de 1728 controis (1,2%) (Thomson et al. 2014). Outro dos traballos onde identificaban a E751Q era levado a cabo por Green e cols. Nel vían que esta variante aparecía en 23 alelos de 2422 controis (0,95%) e en 11 alelos de 1012 (1,09%) individuos esquizoafectivos. Os autores definían ós doentes esquizoafectivos como: individuos diagnosticados con esquizofrenia que tivesen alomenos un episodio de manía ó longo da súa vida; individuos con trastorno bipolar I con características psicóticas e incongruencias do estado de ánimo alomenos na metade dos episodios; e individuos que cumprían os *Research Diagnostic Criteria* (RDC) para o síndrome de alienación parental. Os datos das variantes que atopaban estaban publicados conxuntamente polo que non podíamos coñecer as variantes presentes unicamente en doentes esquizofrénicos, de xeito que non podíamos incluír os datos na meta-análise (Green et al. 2011).

Deste primeiro estudo podíamos concluír que só atopabamos unha asociación débil da variante E751Q aínda que non podíamos confirmala mediante a meta-análise. Debido ás limitacións de potencia estatística polo tamaño de mostra utilizado non podemos desbotar a idea da implicación de variantes codificantes de baixa frecuencia en *DISC1* no desencadeamento da esquizofrenia. As recentes melloras na tecnoloxía e nas ferramentas bioinformáticas permiten levar a cabo a secuenciación masiva do exoma completo ou dos exóns de determinados xenes candidato relacionados coa esquizofrenia (Estudo III), permitindo facer unha busca máis ampla de variantes raras que aumenten o risco a padecer a enfermidade. Por outra banda, este primeiro estudo tamén permitía identificar, trala meta-análise feita, unha posible variante funcional de risco na rexión 3'UTR de *DISC1*. Esta variante podería estar involucrada na regulación da expresión deste xene mediante os miARNs.

Este primeiro estudo non nos permitía identificar unha asociación dunha variante individual dunha forma robusta, polo que unha alternativa era analizar os xenes como un conxunto. Por iso levamos a cabo o Estudo II da tese. Nel utilizabamos datos de 5100 SNPs codificantes de sentido incorrecto dun traballo previo (Carrera et al. 2012)

co fin de ver se existía unha sobre-representación das proteínas que interactuaban con DISC1 entre os resultados máis destacados dunha lista de xenes ordenados segundo o valor de significación. Para isto utilizabamos unha análise baseada nun conxunto de xenes. Neste estudo, efectivamente identificabamos esta sobre-representación, amosando a importancia deste conxunto de xenes no desencadeamento da enfermidade.

As características do deseño do noso traballo permitíanos detectar un enriquecemento do conxunto de xenes do interactoma de DISC1, sendo un dos poucos GWAS que, de xeito indirecto, amosaba evidencias dunha asociación positiva entre *DISC1* e a esquizofrenia. Outro dos traballos que detectaban unha asociación entre este xene e a doenza usando datos dun GWAS era o levado a cabo por Jia e cols. Estes utilizaban un método para detectar redes de interacción proteína-proteína que localmente maximizaba a proporción de xenes con valores-P baixos en dous conxuntos de datos de GWAS. Deste xeito identificaban 205 xenes cun efecto conxunto significativamente asociados coa esquizofrenia. Entre estes xenes estaba *DISC1*, ademais de outros 7 que interactuaban directamente con el, incluíndo *SYNE1*, *KIF3A*, *MACF1*, e *UTRN*, os cales identificabamos na nosa análise como xenes máis significativos, liderando a asociación entre a doenza e o interactoma de DISC1, e ós cales chamamos xenes máis extremos (Jia et al. 2012). Estes xenes máis extremos do interactoma de DISC1 estaban relacionados coa organización e dinámica do citoesqueleto neuronal, afectando a procesos como a estabilidade do citoesqueleto ou ó transporte intracelular ó través dos microtúbulos do citoesqueleto, o cal estaba sobre-representado entre as proteínas que interactuaban con DISC1 (Camargo et al. 2007).

UTRN únese á actina do citoesqueleto, ligándoa á matriz extracelular, suxerindo un papel no soporte estrutural das membranas neuronais (Perronet and Vaillend 2010). AKAP9 é necesaria para a ensamblaxe dos microtúbulos nas membranas de Golgi (Rivero et al. 2009), e probablemente está involucrada no acoplamento citoesquelético de receptores de NMDA (Lin et al. 1998). KIF3A é un membro do complexo quinesina II, un complexo proteico motor relacionado co transporte anterógrado dos orgánulos membranosos ós axóns (Kondo et al. 1994; Hirokawa et al. 2009). DST é un factor de entrecruzamento de microtúbulos responsable da elevada estabilidade dos microtúbulos axonais, os cales son necesarios para o transporte de largas distancias (Yang et al. 1999). Unha alteración da súa asociación coa dinactina daba lugar a problemas no transporte axonal retrógrado (Sonnenberg and Liem 2007). MACF1 é tamén un factor de entrecruzamento de microtúbulos intimamente relacionado con DST. Ratos nos que se anulaba DST amosaban que existía unha compensación mediante MACF1 no sistema nervioso central, indicativo dalgunha redundancia funcional (Leung et al. 2002). DISC1 está involucrado no transporte ó longo dos microtúbulos ata a parte distal dos axóns, actuando como un receptor de carga no complexo quinesina-I (Shinoda et al. 2007; Taya et al. 2007). Por último, SYNE 1 forma parte do complexo LINC (do inglés, *linker of nucleoskeleton and cytoskeleton*), envolto na ancoraxe da cuberta á actina do citoesqueleto, xogando un papel na neuroxénese (Zhang et al. 2009). Un fragmento de SYNE únese á sub-unidade KIF3B do complexo quinesina II (Fan and Beck 2004).

Destacar que unha meta-análise de GWAS de trastorno bipolar amosaba unha asociación con *SYNE1* (Sklar et al. 2011), o cal tamén era confirmado e asociado en depresión maior (Green et al. 2013). Xu e cols. publicaban un artigo onde levaban a cabo un GWAS en 950 individuos con trastorno bipolar e 950 controis atopando evidencias da implicación de SYNE1 no desencadeamento do trastorno bipolar (Xu et al. 2014). No 2016, publicábase un traballo no que se analizaban 244 SNPs localizados en 34 xenes candidato en 100 individuos diagnosticados con esquizofrenia ou trastorno

bipolar e 96 controis. Un dos SNPs que acadaba un valor significativo estaba situado no xene *SYNE1* (Gasso et al. 2016). Máis recentemente, Sharp e cols. identificaban variantes non sinónimas raras despois de analizar do exón 14 ó 22 de *SYNE1* en 937 doentes diagnosticados con trastorno bipolar. Estas estaban presentes en nove doentes que son heterocigotos compostos indicando que estas variantes poderían estar contribuíndo na arquitectura xenética da doenza (Sharp et al. 2017). Ademais, mediante a secuenciación do exoma de pacientes con autismo esporádico identificaban mutacións *de novo* de sentido incorrecto preditas con consecuencias funcionais en *SYNE1* (O'Roak et al. 2011) e unha mutación *de novo* alterando *DST* (Iossifov et al. 2012). No 2013, Yu e cols. secuenciaban o exoma de familias consanguíneas con varios membros afectos con trastorno do espectro autista atopando unha mutación de sentido incorrecto posiblemente funcional en *SYNE1* que cosegregaba coa enfermidade (Yu et al. 2013). Schuurs-Hoeijmakers e cols. secuenciaban o exoma de 19 familias con individuos afectos con discapacidade intelectual atopando mutacións patoxénicas en varios xenes, entre os cales se atopaba *SYNE1* (Schuurs-Hoeijmakers et al. 2013). Tendo en conta que os factores xenéticos de susceptibilidade de trastornos psiquiátricos a miúdo son compartidos por diferentes enfermidades psiquiátricas (Gratacos et al. 2009; Malhotra and Sebat 2012; Cross-Disorder Group of the Psychiatric Genomics et al. 2013), entre os que está a translocación orixinal de *DISC1* (St Clair et al. 1990), todos estes descubrimentos daban un apoio adicional ó noso resultado principal.

Tres dos sete xenes máis extremos do interactoma de *DISC1* (*DST*, *MACF1* e *SYNE1*) codifican proteínas que se unen a *DTNBP1* (Benson et al. 2001; Camargo et al. 2007; Guo et al. 2009). Destacar que os primeiros traballos publicados amosaban que *DTNBP1* formaba un complexo con *UTRN* no cerebro, outro dos xenes máis extremos identificados na nosa análise (Benson et al. 2001), pero un estudo posterior non era capaz de replicar este resultado (Nazarian et al. 2006). *DTNBP1* era considerado como outro posible xene de susceptibilidade á esquizofrenia. Straub e cols. identificaban por primeira vez este xene como un xene de susceptibilidade a doenza. Para chegar a este resultado primeiro levaban a cabo un estudo de ligamento onde identificaban un locus ligado á enfermidade e que posteriormente estudaban mediante unha análise de asociación na que se incluían 270 familias. Así atopaban varios SNPs que, independentemente ou en combinación, aumentaban o risco a padecer esquizofrenia (Straub et al. 2002). Anos máis tarde, Allen e cols. levaban a cabo meta-análises daqueles polimorfismos dos que se dispoñía de datos de, polo menos, catro mostras caso-control independentes. Atopaban 24 variantes que acadaban un valor significativo que estaban presentes en varios xenes, entre os que se atopaba *DTNBP1* (Allen et al. 2008). *DTNBP1* está envolto na organización do citoesqueleto e no transporte dende o corpo da célula neuronal ás terminais sinápticas como unha subunidade do complexo BLOC-1 (Kubota et al. 2009; Mead et al. 2010; Larimore et al. 2011).

Destacar que a detección do enriquecemento detectado no interactoma de *DISC1* foi debido a que neste estudo utilizabamos unha hipótese dirixida a un conxunto de xenes, a cal era unha estratexia recente para a análise de datos de GWAS como un conxunto, co fin de evitar unha carga excesiva de corrección por múltiples probas, algo que era intrínseco ó método de hipótese libre (Askland et al. 2011). Outra característica importante era que nos compilabamos manualmente o conxunto de xenes candidato, tendo a vantaxe da existencia dun gran número de estudos sobre os xenes que interactúan con *DISC1*. Isto é importante para enfermidades neuropsiquiátricas xa que as bases de datos de conxunto de xenes e de rutas metabólicas poderían non ter ben anotadas estas rutas para as funcións neuronais (Wang et al. 2010a). Aínda que os xenes

exactos que forman parte do interactoma de DISC1 poderían ser materia de debate, a detección da sobre-representación significativa deste conxunto de xenes utilizando dous criterios distintos para a selección dos xenes amosaba que o noso resultado era robusto para o criterio de selección do conxunto de xenes. Por outra banda, o uso de SNPs codificantes de sentido incorrecto implicaba que podíamos asignar directamente un SNP a un xene, polo que se evitaban criterios arbitrarios á hora de asignar un SNP interxénico a un xene. Por último, mentres outros métodos para testar sobre-representación dun conxunto de xenes consideraban un umbral arbitrario de significación, o método de Wang e cols. tiña en conta tódolos valores-P. Considerando que non hai un consenso claro sobre cal é o mellor método para a análise dun conxunto de xenes, aconsellábase o uso de diferentes métodos para ver a consistencia dos resultados obtidos (Wang et al. 2010a). Seguindo esta apreciación nos utilizabamos PLINK para facer unha análise adicional do conxunto de xenes (Purcell et al. 2007). A diferenza de GenGen, con PLINK é necesario elixir varios parámetros que poderían ter un efecto considerable no resultado final. Utilizando parámetros intermedios á hora de elixir os SNPs que se teñen en conta para o cálculo do estatístico do conxunto de xenes obtiñamos resultados acordes ó GenGen, aumentando así a evidencia da asociación do interactoma de DISC1 coa esquizofrenia.

Un dos principais resultados dos GWAS de esquizofrenia era a existencia dun importante compoñente polixénico común, cuantificado mediante unha puntuación polixénica, normalmente calculada como a suma do número de alelos de susceptibilidade de cada un dos SNPs ponderado polo seu OR (International Schizophrenia et al. 2009; Carrera et al. 2012). Esta puntuación polixénica explicaba unha maior proporción da varianza poboacional na susceptibilidade a padecer esquizofrenia se estivese baseada nun maior conxunto de SNPs no canto de só ter en conta os SNPs máis significativos. De feito, Purcell e cols. atopaban que a puntuación polixénica baseada en tódolos SNPs cun valor-P menor de 0,5 no seu GWAS explicaban a maior proporción da varianza na susceptibilidade a padecer esquizofrenia nunha mostra caso-control independente (International Schizophrenia et al. 2009). Isto suxería que moitas das variantes comúns de susceptibilidade non alcanzaban os valores de significación, aínda que os seus valores-P foran lixeiramente menores do esperado por azar. De acordo con isto, identificabamos unha tendencia á sobre-representación dos SNPs do interactoma de DISC1 analizados neste segundo traballo na parte alta da lista ordenada dos valores-P dos datos do PGC (Ripke et al. 2011).

Unha das limitacións deste traballo era que a asociación de cada xene era resumida mediante o mellor valor-P, polo que este método non tiña en conta a posible heteroxeneidade alélica (Cantor et al. 2010). Ademais, os nosos datos de xenotipado incluían poucos xenes do interactoma de DISC1. Por outra banda, non dispoñíamos de datos adicionais para levar a cabo unha réplica. Para rematar, debemos ter en conta que a análise dun conxunto de xenes dun estudo de asociación permite, a priori, identificar os posibles mecanismos biolóxicos implicados na etiopatoloxía dun trastorno, pero para iso é necesario estandarizar un método para a análise dos datos, cousa que a día de hoxe aínda non se conseguiu.

A pesares destas limitacións, no estudo II atopabamos unha asociación significativa entre un subconxunto de xenes do interactoma de DISC1 e a esquizofrenia, estando de acordo con achados previos (Jia et al. 2012). Este resultado podería ser relevante, sobre todo se temos en conta o suposto de que a esquizofrenia podería considerarse como unha enfermidade de rutas biolóxicas (Sullivan 2012).

Discusión

Os dous primeiros estudos comentados anteriormente poderían considerarse como unha fase previa do Estudo III, xa que no Estudo I faciamos unha primeira aproximación para a busca de variantes raras individuais asociadas á esquizofrenia, e no Estudo II amosabamos a existencia dunha asociación co interactoma de DISC1 mediante o estudo de variantes comúns. Polo tanto, o Estudo III é unha continuación dos anteriores xa que nel tratábamos de buscar variantes de baixa frecuencia cun efecto elevado en rexións codificantes de 9 xenes que formaban parte do interactoma de DISC1, ademais de noutros 11 xenes de susceptibilidade atopados mediante estudos de CNVs ou GWAS. Neste terceiro traballo da tese centrámonos no estudo de variantes a baixa frecuencia seguindo a hipótese de variante rara-enfermidade común, que sostíña que as variantes de risco presentes nesta rexión estarían sometidas á selección purificadora non alcanzando unha frecuencia elevada (McClellan et al. 2007; Manolio et al. 2009).

Este estudo consistía nunha fase inicial na que secuenciábamos, mediante a plataforma de ultrasecuenciación Solid4 System, as rexións codificantes (en total, 124,5 kb) de 21 xenes seleccionados por estar asociados a esquizofrenia mediante estudos de CNVs e GWAS. Analizábamos 153 pacientes diagnosticados con esquizofrenia e 153 controis, todos eles de orixe galega, agrupados en *pools* de tres mostras. A continuación, na segunda fase, xenotipábamos, nas mostras restantes dispoñibles (419 casos e 398 controis), as variantes posiblemente prexudiciais que aparecían unicamente nos casos. Antes de centrarnos nas variantes atopadas, faciamos unha análise da cobertura das rexións secuenciadas. Mediante esta viamos que o xene *NRGN* presentaba moi pouca cobertura e ademais as lecturas sempre estaban na mesma cadea, de xeito que non había solapamento entre as lecturas nos dous sentidos (Fig.25), polo que decidíamos descartalo das análises posteriores. Tamén faciamos unha comparación entre as lecturas dos controis e dos casos, vendo que non existía un nesgo significativo entre ámbalas dúas cohortes, polo que podíamos comparar dunha forma fiable as variantes atopadas entre ámbolos dous grupos de mostras. Por último, xa que dispoñíamos dos datos de secuenciación de Sanger e de xenotipado do Estudo I e Estudo II respectivamente, levamos a cabo unha valoración da eficacia na identificación de variantes de tres *softwares* (BioScope, GATK e VarScan). Partindo da idea de que a meta era conseguir unha alta especificidade mantendo a sensibilidade elevada, viamos que o VarScan cos parámetros especificados na sección RESULTADOS era o paquete informático que mellor se atíña ó noso obxectivo. VarScan é un *software* especificamente deseñado para a identificación de SNVs en *pools*. Isto unido a que as *pools* que utilizábamos estaban formadas por moi poucas mostras, permitíanos reducir a existencia de falsos positivos, un dos principais problemas da secuenciación de nova xeración.

O primeiro resultado que obtiñamos na análise das SNVs era un exceso de variantes funcionais a baixa frecuencia ($MAF < 5\%$) nos xenes das mostras secuenciadas, os cales probablemente estean involucrados na susceptibilidade a esquizofrenia. No noso traballo considerábamos como variantes funcionais aquelas que eran preditas por PolyPhen2 como probablemente ou posiblemente prexudiciais, as cales agrupábamos nun único grupo denominándoas como variantes deletéreas. A vista deste resultados, as seguintes análises estaban centradas neste tipo de variantes.

As variantes codificantes raras xogan un papel importante no desencadeamento de moitas doenzas e contribúen a herdabilidade perdida dos GWAS, as cales son estudadas en maior profundidade grazas ós estudos de secuenciación de nova xeración. Estes estudos permiten o descubrimento de variantes non sinónimas, das cales, na súa gran maioría, coñécese pouco sobre a súa funcionalidade, sendo un reto conseguir o

desenvolvemento de estratexias que axuden a predicir a probabilidade da patoxenicidade dunha variante. Na actualidade existen varias ferramentas bioinformáticas que poden facelo. Unha delas é a que nos utilizabamos: Polyphen2. Esta basea as súas predicións utilizando as características da secuencia e a información estrutural e filoxenética que caracterizan a substitución, pero existen outras moitas ferramentas como SIFT (Ng and Henikoff 2003), PROVEAN (Choi and Chan 2015), MutationAssessor (Reva et al. 2011), MutationTaster (Schwarz et al. 2014), LRT (Chun and Fay 2009), GERP (Cooper et al. 2010), SiPhy (Garber et al. 2009), phastCons (Felsenstein and Churchill 1996), etc. Estes algoritmos son útiles por si mesmos, pero as súas predicións as veces non son similares xa que cada un utiliza un algoritmo diferente aínda que parte da información que utilizan para as predicións é común. Aínda así non hai unha decisión unánime sobre se o convinte é utilizar unha ou varias destas ferramentas para priorizar as variantes nos diferentes estudos e así minimizar os erros nas predicións.

No estudo de asociación de variantes individuais mencionar que ningunha delas alcanzaba valores significativos pero cabía destacar a SNV K873R situada no xene *AKAP9*. Estaba presente en 4 dos 572 casos e en ningún dos 791 controis, acadando un valor de significación nominal ($P=0,031$). Esta variante aparecía cunha frecuencia moi baixa en poboación latina ($1,7 \times 10^{-4}$) e do sur de Asia ($6,1 \times 10^{-5}$) segundo a base de datos do *Exome Aggregation Consortium* (ExAC) aínda que estaba ausente en poboación europea. Ademais non aparecía no resto das bases de datos consultadas, incluíndo as 5008 mostras do *The 1000 Genomes Project phase III* e as 6503 mostras pertencentes ó *NHBLI Exome Sequencing Project*. A frecuencia dunha variante na poboación pode ser considerada como un criterio adicional na interpretación funcional desta. De feito, segundo as recomendacións do Colexio Americano de Xenética e Xenómica Médica e a Asociación de Patoloxía Molecular, cando estudas un trastorno e unha variante presenta unha frecuencia alélica maior a frecuencia poboacional esperada pode considerarse unha evidencia bastante forte para clasificar unha variante como benigna (Richards et al. 2015). Sen embargo, esta recomendación non da un valor detallado para determinar a frecuencia alélica esperada para considerar unha variante como patoxénica (Fig. 39). A publicación do ExAC, base de datos formada por datos da secuenciación de exomas de 60706 individuos, podería axudar a determinar un umbral de frecuencia a partir do cal unha variante pode ser considerada patoxénica. De feito, ExAC proporciona a maior base de datos que hai na actualidade para estimar a frecuencia alélica de variantes xenéticas codificantes, proporcionando un filtro importante para a análise de variantes patoxénicas.

Discusión

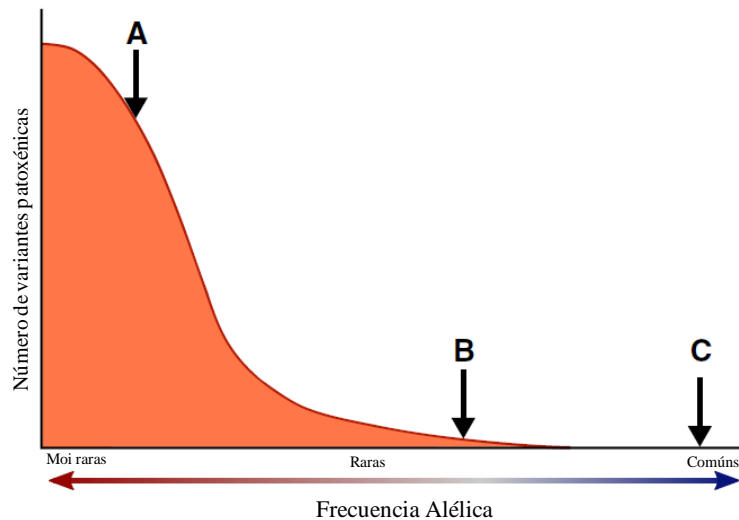


Figura 39 Distribución das variantes patoxénicas en función da súa frecuencia alélica. A, B e C representan tres supostos distintos. En A a variante descoñecida ten unha frecuencia onde se atopan as variantes patoxénicas coñecidas. No escenario B, a variante descoñecida ten unha frecuencia acorde con variantes patoxénicas coñecidas, pero a frecuencia desta é máis elevada que a maioría das mutación patoxénicas. Por último, en C a variante descoñecida ten unha frecuencia fora da distribución da frecuencia de variantes patoxénicas coñecidas. Figura adaptada de Kobayashi e cols. (Kobayashi et al. 2017).

O aminoácido de referencia situado na posición da variante K873R está ben conservado ó longo da evolución agás en peixes (Fig. 38), indicando a importancia desta posición no funcionamento da proteína implicando que un cambio neste lugar podería alterar a función da proteína.

Hcmo	:	LQEEYACLLKVKDDLEDSK
Macaca	:	LQEEYACLLKVKDDLEDSK
Mus	:	LQREYTCLLKIRDDLEATQ
Canis	:	LQEDYACLLKVKTDLEDSK
Opossum	:	LR EEYACLLKVKADLEDSK
Gallus	:	LQKECESLLKAKNVLQENM
Anolis	:	LQEKYASLLKVKLDLEESR
Danio	:	LKEEYTCIANAKLQLEERL

Figura 38. Aliñamento dos aminoácidos situados entorno a variante K873R ó longo dos distintos niveis filoxenéticos. A frecha indica a posición da variante K873R da proteína que codifica o xene *AKAP9*.

As predicións bioinformáticas feitas con PolyPhen2 sobre a variante K873R sinalaban que se trataba dunha mutación posiblemente prexudicial. *AKAP9* actúa como unha proteína de andamiaxe postsináptica no cerebro, no cal forma un complexo co receptor NMDA xunto con outros compoñentes da ruta metabólica da sinalización do AMPc, tales como a proteína quinasa A (PKA), adenilil ciclase ou as fosfodiesterasas PKA (Sanderson and Dell'Acqua 2011). K873R está situada nun lugar de unión a adenilil ciclase 2 (*ADCY2*). Esta unión da lugar á inhibición de *ADCY2* no cerebro

(Piggott et al. 2008). Isto probablemente actúa como un mecanismo de modulación da plasticidade sináptica mediante a regulación de receptores NMDA (Sanderson and Dell'Acqua 2011; Edwards et al. 2012). No ano 2014 Muhleisen e cols. identificaban *ADCY2* como un xene de risco do trastorno bipolar. Para chegar a este achado levaban a cabo un GWAS onde analizaban máis de 24 mil mostras, incluíndo controis e doentes diagnosticados con trastorno bipolar. En total, identificaban 56 SNPs que sobrepasaban o umbral de significación xenómica, os cales estaban situados en 5 rexións xenómicas, estando tres delas previamente asociadas á doenza (*ANK3*, *ODZ4* e *TRANK1*) ademais do locus 5p15.31 (onde se situaba *ADCY2*) e a rexión entre *MIR2113* e *POU3F2* (Muhleisen et al. 2014). Máis recentemente, van Hulzen e cols. levaban a cabo unha meta-análise na que se incluían datos de xenotipado de 4609 individuos con trastorno de hiperactividade e déficit de atención e 9650 doentes de trastorno bipolar, ademais de 21363 controis. Deste xeito, identificaban unha asociación significativa a nivel xenómico no xene *ADCY2*, pero só cando se incluían datos de casos con trastorno bipolar na que a enfermidade debutase en idades temperáns (van Hulzen et al. 2016). Cabe a pena destacar isto debido a existencia do solapamento xenético entre o trastorno bipolar e a esquizofrenia (International Schizophrenia et al. 2009; Lichtenstein et al. 2009). Tamén mencionar que no 2014 publicábase un estudo no que secuenciaban o exoma de 623 tríos de esquizofrenia no que identificaban unha mutación *de novo* de sentido incorrecto no xene *AKAP9* nun dos doentes (Fromer et al. 2014).

A principal limitación deste traballo é a pouca potencia que tiñamos para a detección de variantes de risco a nivel individual. Intentabamos superar este problema utilizando diferentes tests estatísticos que nos permitían estudar o papel acumulativo das variantes raras nun determinado locus. Neste caso utilizabamos o test de carga WSS (do inglés, *Weighted Sum Statistic*), o cal nos permite ponderar as variantes, por un lado pola súa frecuencia, asumindo que as variantes máis raras teñen máis probabilidade de ser funcionais; e por outro lado pola súa predición de impacto na proteína utilizando as puntuacións obtidas mediante PolyPhen2 (Price et al. 2006; Madsen and Browning 2009). Ademais deste test, utilizabamos un test de compoñentes da varianza, o test C-Alpha, que, a diferenza dos tests de carga, considera a existencia tanto de alelos de risco, como protectores ou neutrais (Lin and Tang 2011; Neale et al. 2011). O cálculo do nivel de significación facíámolo mediante permutacións seguindo o noso deseño experimental (Fig. 21). As variantes eran agrupadas e analizadas en función do xene ó que pertencían, atopando que unicamente dous dos xenes analizados, *NRXN1* e *ZNF804A*, acadaban valores significativos nominais. Ademais considerabamos *DISC1* e as proteínas que interactuaban con el como un conxunto de xenes a estudar. A pesares de analizar máis de 94 kb de secuencia por cada unha das mostras, non eramos capaces de identificar novas evidencias da implicación do interactoma de *DISC1* na esquizofrenia.

En resumo, este terceiro traballo da tese, no que tratamos de identificar variantes raras en 21 xenes candidato de susceptibilidade á esquizofrenia, amosaba a existencia de posibles variantes de risco en *AKAP9* ademais dun papel acumulativo das variantes a baixa frecuencia e de sentido incorrecto en *ZNF804A* e *NRXN1*, actuando neste último como posibles factores de protección contra esquizofrenia.

Os resultados máis salientables do Estudo III eran os que obtiñamos cando faciamos a análise conxunta das variantes, polo que decidiamos centrarnos nos xenes *ZNF804A* e *NRXN1* e analizalos nunha mostra de maior tamaño. Para isto, primeiro faciamos unha busca bibliográfica intentando atopar datos de traballos nos que secuenciaban estes xenes, en individuos esquizofrénicos e en controis, co fin de

confirmar estes resultados mediante unha meta-análise. En canto a *ZNF804A*, atopabamos tres estudos nos que analizaban este xene (Dwyer et al. 2010b; Williams et al. 2011; Purcell et al. 2014), pero ningún dos tres dispoñía dos datos necesarios para facer a meta-análise. Pola contra, viamos que había sete estudos que secuenciaban *NRXN1*. Dous deles (Need et al. 2012; Kenny et al. 2013) non dispoñían dos datos que precisabamos, pero os outros cinco si (Gauthier et al. 2011; Muhleisen et al. 2011; Hu et al. 2014; Purcell et al. 2014; Girard et al. 2015).

Unha vez compilada toda esta información, decidiamos levar a cabo o Estudo IV. Nel intentabamos confirmar o exceso de variantes raras codificantes de sentido incorrecto en controis en *NRXN1*, xa que era o único xene do cal atopabamos os datos suficientes para levar a cabo unha meta-análise. Para isto secuenciabamos, utilizando a tecnoloxía Ion Proton™ System, as rexións exónicas (10,7 kb) deste xene en 384 doentes de esquizofrenia e o mesmo número de controis, agrupándoos en *pools* de catro mostras. Ó igual que o traballo anterior, analizabamos os datos de cobertura e viamos que non había ningún nesgo significativo entre os casos e os controis, de xeito que podiamos comparar as SNVs atopadas en ámbalas dúas cohortes cun elevado nivel de fidelidade. Secuenciando este xene atopabamos unha variante sen sentido e 31 codificantes de sentido incorrecto cunha frecuencia inferior ó 1%. Todas estas variantes xenotipabámolas con Sequenom no total da nosa mostra confirmando 17. Con estas SNVs identificadas xunto coas publicadas nos 5 estudos comentados anteriormente, podiamos confirmar a existencia dun exceso, en controis, de variantes codificantes de sentido incorrecto cunha frecuencia menor ó 1% no xene *NRXN1* analizando un total de 4467 casos e 4513 controis. Este exceso acadaba un $P = 0,0469$ cun $OR = 0,74$ e un intervalo de confianza do 95% = (0,55-0,99). Destacar que non existía heteroxeneidade entre os diferentes estudos incluídos ($I^2 = 27,8\%$, cun valor-P do test Q = 0,20). Resaltar que este grupo de variantes era o que acadaba un valor de significación máis baixo, pero se agrupabamos as variantes codificantes de sentido incorrecto, que aparecían nun único individuo e que se clasificaban como prexudiciais segundo PolyPhen2, conseguíamos tamén un valor significativo ($P = 0,0473$). Ademais probabamos outras combinacións (variantes codificantes de sentido incorrecto, variantes codificantes de sentido incorrecto, prexudiciais e cunha frecuencia menor ó 1%; variantes non sinónimas que aparecían nun único individuo; variantes non sinónimas que aparecían nun único individuo, prexudiciais ou que provocasen unha perda de función; variantes non sinónimas cunha frecuencia menor ó 1%; e variantes non sinónimas cunha frecuencia menor ó 1% prexudiciais ou que provocasen unha perda de función) pero ningunha delas acadaba valores significativos.

Dous dos traballos incluídos na meta-análise presentaban datos resumo e non seguían exactamente os mesmos criterios de selección de variantes que nos utilizabamos no noso traballo. Un deles era o traballo de Purcell e cols. onde secuenciaban o exoma de mais de cinco mil mostras, incluíndo casos e controis. Neste artigo clasificaban as variantes en tres grupos: variantes disruptivas (aquelas que eran sen sentido, que alteraban un lugar de *splicing* esencial ou que alteraban a pauta de lectura), variantes non sinónimas estritas (eran as variantes disruptivas máis as preditas por ser prexudiciais polos cinco algoritmos que utilizaban para a predición funcional) e variantes non sinónimas amplas (variantes disruptivas máis aquelas que eran preditas como prexudiciais por alomenos un dos 5 preditores utilizados). Os autores só publicaban as variantes non sinónimas estritas e cunha frecuencia menor ó 0,1% (Purcell et al. 2014). O segundo artigo era o de Hu e cols. no que levaban a cabo un traballo de dúas fases. Na primeira parte do estudo secuenciaban 101 xenes en 525

casos e 619 controis e na fase de réplica analizaban 455 pacientes esquizofrénicos e 336 controis (Hu et al. 2014). Neste traballo os autores non publicaban as frecuencias nin as variantes individuais senón que publicaban os datos resumo das deleccións e insercións que modificaban a pauta de lectura, as variantes sen sentido, variantes non sinónimas e as variantes de *splicing*. Debido ó gran número de mostrás analizadas, incluíamos estes dous artigos na meta-análise aínda que non seguían os mesmos criterios de clasificación de variantes que nos empregabamos. De aí que decidimos levar a cabo varias combinacións de variantes, polo que esta análise podería considerarse como exploratoria.

NRXN1 é unha molécula de adhesión celular presente principalmente en terminais presinápticas inhibitorias GABAérxicas e excitadoras glutamatérxicas (Bang and Owczarek 2013). NRXN1 únese a proteínas neuronais transmembrana con repeticións enriquecidas en leucina (LRRTM), concretamente a catro membros desta familia de LRRTM, amosando todos estes actividade sinaptóxénica, localizándose en sinapses excitadoras e promovendo a diferenciación postsináptica (Linhoff et al. 2009). NRXN1 tamén se une a cerebellin (Cbln). Esta proteína xoga un papel esencial na formación e mantemento da sinapse excitadora entre as células granulares e as células de Purkinje en cerebelo de rato. De feito, para mediar a formación da sinapse in vivo no cerebelo, Cbln1 únese simultaneamente ó receptor $\delta 2$ de glutamato postsináptico e a NRXN presináptica (Lee et al. 2012d). NRXN1 únese ás neurolixinas (NLGN), as cales son as moléculas que interactúan cas neurexinas que máis se teñen estudado na densidade postsináptica, formando un complexo dependente de calcio nas sinapses do sistema nervioso central. A súa vez este complexo únese a proteínas tirosinas fosfatas tipo receptor mediante o seu dominio extracelular, de xeito que forman un complexo que regula a formación da sinapse (Lim et al. 2009). NRXN1 tamén se une á calcio/calmodulina proteína quinasa controlando o mecanismo de liberación das vesículas sinápticas (Biederer and Sudhof 2001). A súa afinidade para unirse a outras moléculas está controlada por un proceso dinámico que supón cambios das diferentes variantes de *splicing* na terminal presináptica (Bang and Owczarek 2013). Ademais, NRXN1 xoga un papel importante na regulación do equilibrio inhibitor/excitador nos circuítos neuronais, o cal podería estar danado nos trastornos do neurodesenvolvemento como a esquizofrenia (Rubenstein and Merzenich 2003; Insel 2010; Bang and Owczarek 2013).

Segundo GTE_x (www.gtexportal.org), NRXN1 exprésase principalmente no cerebro, sendo o cortex prefrontal e o hemisferio cerebelar as rexións onde se atopa un maior nivel de expresión. Concretamente, identificábanse altos niveis deste xene durante os períodos críticos do desenvolvemento sináptico estando enriquecido no córtex prefrontal durante as últimas estadias xestacionais e principios da vida postnatal, especificamente durante o segundo trimestre ata aproximadamente os tres anos de idade postnatal (Jenkins et al. 2016).

O feito de que variantes raras en *NRXN1* puideran ser factores protectores contra a esquizofrenia é sorprendente xa que as CNVs que alteran este xene están asociadas con discapacidade intelectual, autismo e esquizofrenia, aínda que a penetrancia sexa incompleta (Autism Genome Project et al. 2007; Kirov et al. 2008). Mentres que estas alteracións poderían alterar o equilibrio excitador/inhibidor aumentando considerablemente o risco a padecer esquizofrenia, mutacións menos prexudiciais de sentido incorrecto poderían dar lugar a cambios sutís neste equilibrio. Estes cambios poderían optimizar o fondo xénético de susceptibilidade a padecer o trastorno nalgúns individuos, polo que poderían ser protectores contra a esquizofrenia. Este achado

podería ser relevante xa que se estudos posteriores chegaran a confirmalo, este sería unha diana alternativa para o tratamento da esquizofrenia mediante a terapia farmacolóxica do equilibrio excitador/inhibidor, a través da modulación selectiva da interacción NRXN-NLGN (Sudhof 2008; Bang and Owczarek 2013).

A complexidade dos efectos de mutacións específicas nas moléculas de adhesión celular sináptica podería ser exemplificado mediante as NLGNs. No 2013, Bang e Owczarek, estudaban os fenotipos electrofisiolóxicos de dúas mutacións de sentido incorrecto asociadas con autismo en ratos co *NLGN3 Knock-in*. As variantes eran R451C e a R704C; esta última identificada en *NLGN4* pero era testada en *NLGN3* debido a súa baixa conservación e expresión en *NLGN4* en ratos. Estas mutacións tiñan un efecto oposto, aumentando ou descendendo, respectivamente, a transmisión sináptica excitadora mediada polo receptor AMPA no hipocampo. Polo tanto, mutacións de risco en trastornos do neurodesenvolvemento en moléculas de adhesión celular sináptica poderían actuar ó través de cambios no equilibrio inhibitor/excitador modificándoo cara un lado ou cara outro (Bang and Owczarek 2013).

Neste senso, recentemente Jenkins e cols. identificaban que os niveis de expresión de *NRXN1* están alterados nun transcrito dunha maneira específica no contexto de trastornos neuropsiquiátricos. De feito amosaban que a expresión de NRXN1- α e de NRXN1- β estaba lixeiramente alterada, pero dunha forma significativa, no córtex prefrontal. Concretamente vían que a expresión NRXN1- α estaba modificada en doentes con trastorno bipolar e a de NRXN1- β en pacientes con esquizofrenia (Jenkins et al. 2016). Isto é unha evidencia máis do efecto pleiotrópico deste xene. O estudo do efecto que as variantes xenéticas provocan sobre diferentes trastornos é unha estratexia bastante habitual no eido da enfermidades psiquiátricas. De feito, existen varios estudos que amosan que hai CNVs que inflúen no risco a padecer esquizofrenia, pero que tamén o fan sobre discapacidade intelectual, trastorno do espectro autista, retraso mental e o trastorno do déficit de atención e hiperactividade (Kirov et al. 2014). A maioría destas CNVs pleiotrópicas identificadas abarcan varios xenes, polo que non se pode excluír que haxa un pseudo-pleiotropismo (é dicir, dous fenotipos están influenciados por diferentes xenes que están próximos), no cal os xenes que están afectados polas CNVs estean afectando a distintos fenotipos (O'Donovan and Owen 2016). O caso da deleción 2p16.3 é distinto xa que esta CNV afecta unicamente ó xene *NRXN1*, implicando directamente a este xene nas distintas doenzas ás que se asocia, incluíndo esquizofrenia, trastorno do espectro autista e o retraso mental e trastorno no desenvolvemento (Marshall et al. 2008; Glessner et al. 2009; Ching et al. 2010; Gregor et al. 2011).

Un efecto pleiotrópico a nivel de variante foi descuberto recentemente no xene *SETD1A*. Singh e cols. publicaban un artigo onde trataban de analizar variantes de perda de función (LOF) raras (MAF<0,1%), ás que definían como aquelas que alteraban a correcta tradución do xene, introducían un codón de parada prematuro, alteraban a pauta de lectura ou impedían o *splicing* normal. En total, atopaban dez variantes en case 8000 individuos esquizofrénicos pero ningunha delas estaba presente nos máis do 13000 controis. Entre estas variantes LOF había unha que estaba implicada significativamente no risco a padecer esquizofrenia pero ó mesmo tempo tamén estaba asociada a padecer discapacidade individual e retraso no desenvolvemento (Singh et al. 2016). Resaltar que estes autores eran os primeiros que atopaban unha variante rara individual asociada significativamente a esquizofrenia.

Recentemente publicábase o maior estudo de secuenciación sobre esquizofrenia, onde analizaban máis de once mil casos e controis de Suecia (Genovese et al. 2016). O estudo centrábase na análise de variantes ultra-raras, definidas como SNVs que

aparecían unha única vez na cohorte estudada. A pesares de analizar un gran tamaño de mostra, non podían implicar, dunha forma robusta, ningún xene específico. O que atopaban era que existía un enriquecemento, nos casos comparado ós controis, de variantes ultra-raras preditas por ser prexudiciais ou alterar a proteína, reforzando a idea de que a esquizofrenia é unha enfermidade polixénica, e de que non só se debe a variantes comúns senón tamén a variantes raras. De feito, facían unha estima da varianza na susceptibilidade á esquizofrenia que explicaban as variantes ultra-raras obtendo un valor próximo ó 0,5%. Isto representaba un valor de, alomenos, unha orde de magnitude inferior ó calculado previamente para as variantes comúns (Schizophrenia Working Group of the Psychiatric Genomics 2014). Aínda que as variantes raras atopadas ata o de agora explican pouco a nivel poboacional, deberían de terse en conta xa que estas poderían ter un maior efecto a nivel individual ou incluso a nivel dunha familia portadora. Ademais, o feito de que as variantes raras teñen consecuencias máis severas que as variantes comúns, fai posible que estudar estas mutacións a nivel molecular en modelos celulares e animais sexa máis alcanzable e sinxelo. No campo dos estudos de ultrasecuenciación para a identificación tanto de variantes raras que causen un efecto sobre a esquizofrenia como de xenes específicos asociados a esta, será necesario estudar tamaños de mostras máis grandes, xa que o tamaño de mostras utilizados ata o de agora non é o suficiente para acadar a potencia estatística necesaria para identificar variantes de risco cun efecto baixo.

O estudo e caracterización da arquitectura xenética da esquizofrenia levou á conclusión de que se trata dunha doenza altamente polixénica na que interveñen tanto variantes de risco comúns como raras, isto propiciou o desenvolvemento de novos métodos estatísticos para intentar estudar esta característica. Entre eles está a puntuación do risco polixénico que ten por obxectivo predicir o risco a sufrir unha enfermidade usando evidencias de asociación a partir de variantes que non pasan o umbral de significación xenómica. En resumo, esta metodoloxía, a partires dos datos dos alelos de risco e as magnitudes do efecto obtidos dun GWAS nunha mostra descubrimento, calcula unha puntuación de perfil de risco xenómico nunha mostra independente diana, usando aqueles SNPs da mostra descubrimento que están por debaixo dun umbral fixado. Esta puntuación de perfil de risco xenómico é calculado para cada individuo da mostra diana sumando o número de alelos de risco ponderado polo seu efecto (OR) na mostra descubrimento. Esta puntuación do perfil avalíase mediante unha regresión do fenotipo da mostra diana tendo en conta outras covariables coñecidas. Destacar que os fenotipos da mostra descubrimento e mostra diana poden ser distintos, polo que este método permite análises de fenotipos cruzados (Wray et al. 2014). A primeira vez que se xeraba unha puntuación de perfil de risco xenómico en esquizofrenia foi no 2009, no que utilizaban datos do ISC (*International Schizophrenia Consortium*) como mostra descubrimento e cohorte do MGS (*Molecular Genetics of Schizophrenia*) como mostra diana. Así obtiñan un valor da varianza explicada dun 3,2% (International Schizophrenia et al. 2009). Utilizando un tamaño de mostra maior, o *Schizophrenia Working Group of the Psychiatric Genomics Consortium* estimaban un valor da varianza maior ó calculado ata o de agora, aínda que seguía lonxe da estimada para a doenza. Baixo o suposto dunha prevalenza do 1% e a independencia dos SNPs analizados, as estimacións da varianza aumentaban ata un 7% (Schizophrenia Working Group of the Psychiatric Genomics 2014). A medida que os tamaños de mostra aumenten e os descritores fenotípicos sexan mellores, tanto en termos de síntomas clínicos como de factores de risco non xenéticos, as análises de risco polixénico aportarán nova información para poder entender mellor as enfermidades complexas (Wray et al. 2014).

Discusión

Por outra banda, os estudos de secuenciación centrados na parte exónica dos xenes non cumpriron as expectativas iniciais, non podendo identificar a varianza xenética causada pola variación rara que se esperaba. O paso seguinte serían os estudos de xenomas completos ou outros nos que se resecuencian rexións reguladoras de xenes implicados na esquizofrenia mediante GWAS ou CNVs, co fin de estudar a variación das rexións non codificantes, e concretamente aquela que afecta as rexións reguladoras. Neste senso, os últimos GWAS publicados nos que se estudaba a variación común amosaban que as principais asociacións eran debidas a variantes situadas en rexións non codificantes e reguladoras da expresión xénica (Richards et al. 2012; Roussos et al. 2014; Schizophrenia Working Group of the Psychiatric Genomics 2014). Richards e cols. atopaban que variantes asociadas á esquizofrenia polos GWAS publicados naquel momento estaban enriquecidos en eQTLs (do inglés, *expression quantitative trait loci*) presentes en rexións promotoras e *enhancers* (Richards et al. 2012). Por outra banda, Roussos e cols. vían que existía un enriquecemento significativo das variantes asociadas á esquizofrenia nun GWAS publicado no 2013 (Ripke et al. 2013) en elementos reguladores da expresión en *cis*. Ademais atopaban un solapamento significativo entre SNPs que afectaban á expresión e estes elementos reguladores da expresión (Roussos et al. 2014).

Un exemplo da análise de variantes raras situadas en rexións reguladoras era levado a cabo por Duan e cols. no que secuenciaban 6,8 kb entorno ó xene *MIR137*, o cal estaba asociado á esquizofrenia mediante GWAS. Isto facíano en 2610 esquizofrénicos e 2611 controis. Os resultados deste traballo amosaban que existía unha acumulación de variantes raras nas rexións promotoras e *enhancers* deste xene. Ademais este estudo permitía identificar unha variante concreta que aparecía en 11 casos e en ningún control amosando, despois de xenotipala nunha mostra de maior tamaño, unha asociación coa esquizofrenia (Duan et al. 2014).

No 2015, o noso grupo publicaba un artigo que seguía esta mesma liña. Nel analizábanse 74 kb de secuencia reguladora de 29 rexións promotoras na nosa mostra, en total 516 casos e 516 controis agrupados en 83 pools. As rexións reguladoras analizadas pertencían a 23 xenes implicados en esquizofrenia mediante estudos de CNVs e de GWAS. Con estes datos amosábase que existía unha maior proporción significativa en casos respecto ós controis de variantes raras preditas por ter un papel na regulación. Tamén se vía que a proporción de variantes raras en rexións reprimidas da cromatina no estado represor Polycomb en tecidos relevantes era maior nos casos que nos controis. Estes dous resultados combinados daba lugar a identificación dun exceso de portadores de variantes raras cunha predición na regulación localizadas en rexións reprimidas da cromatina no estado represor Polycomb en tecidos relevantes nos casos fronte ós controis. Destacar que estas SNVs están localizadas exclusivamente en dous xenes *HNF1B* e *LHX1*, que pertencen a rexión 17q12. Estes datos suxerían que mecanismos epixenéticos específicos, como a remodelación da cromatina mediante modificación de histonas durante as primeiras etapas do neurodesenvolvemento poderían estar danados nun subconxunto de pacientes con esquizofrenia (Gonzalez-Penas et al. 2016).

A interpretación das variantes identificadas mediante estes estudos de rexións reguladoras non sería posible sen a axuda dos diversos proxectos e consorcios que permiten coñecer con maior profundidade a natureza das rexións reguladoras ó longo de todo o xenoma, a interpretación funcional das variantes presentes nestas e as diferenzas de expresión nas distintas liñas celulares. Un deles é o Proxecto ENCODE, o cal permite identificar e construír un catálogo de variantes funcionais ó longo de todo o

xenoma, proporcionando nova información sobre a organización e regulación dos xenes e o xenoma, e unha fonte ampla de anotacións funcionais (Encode Project Consortium 2012). O proxecto Roadmap describe unha análise integrada de 111 epixenomas con información sobre metilación, transcrición, modificación de histonas e de lugares de hipersensibilidade a DNAsas (Roadmap Epigenomics Consortium 2015). Por outra banda, PsychENCODE permite obter anotacións adicionais a nivel de cerebro humano xa que o obxectivo deste proxecto é caracterizar con maior profundidade as funcións reguladoras no cerebro. Para isto dispoñían de datos epixenómicos de individuos controis e afectos con diferentes enfermidades psiquiátricas (esquizofrenia, trastorno bipolar e trastorno do espectro autista). Con estes datos facían un estudo comparativo dos perfís de expresión espazo-temporal ó longo do desenvolvemento do cerebro humano, de xeito que permitirán coñecer máis polo miúdo as variantes asociadas ás enfermidades psiquiátricas (PsychEncode Consortium 2015). Por último comentar tamén o Brainspan, o cal se centra en estudar a distribución da expresión xénica ó longo das diferentes etapas do neurodesenvolvemento nas distintas áreas do cerebro (Hawrylycz et al. 2012).





6. CONCLUSIONES





1. O estudo da secuenciación das rexións codificantes do xene *DISC1* amosaba que o SNP rs3737597, situado na rexión 3'UTR deste xene, podería ser unha posible variante que aumente o risco a padecer esquizofrenia, xa que esta variante podería estar implicada na regulación da expresión de *DISC1* ó través de miARNs.
2. Aínda que a variante E751Q situada no xene *DISC1* presentaba unha asociación débil coa esquizofrenia, non podiamos confirmar esta asociación mediante a meta-análise, polo que parece que as variantes raras individuais neste xene non aumentaban por si mesmas o risco a padecer esquizofrenia na nosa mostra.
3. Utilizando variantes comúns procedentes dun estudo de GWAS previo analizabamos o papel destes SNPs no interactoma de *DISC1* atopando unha asociación significativa entre este conxunto de xenes e a esquizofrenia, suxerindo a importancia desta rede na susceptibilidade a padecer esquizofrenia, de xeito que se amosaban novas evidencias da implicación do interactoma de *DISC1* na etioloxía da enfermidade.
4. No estudo do interactoma de *DISC1* tamén identificabamos 7 xenes que eran os responsables desta sobre-representación, os cales están involucrados na organización do citoesqueleto neuronal e no transporte intracelular a través dos microtúbulos do citoesqueleto, suxerindo que estes procesos poderían estar danados en pacientes esquizofrénicos.
5. Comparando os datos previos de xenotipado e secuenciación de Sanger cos datos obtidos mediante a secuenciación de nova xeración no Estudo III podíamos concluír que VarScan era o paquete informático que acadaba unha maior sensibilidade á vez que mantiña unha alta especificidade.
6. Neste traballo tamén identificabamos a variante K873R situada no xene *AKAP9*, a cal é unha SNV posiblemente prexudicial aparecendo só nos casos e que alcanzaba unha significación nominal. Isto indica que alteracións en xenes implicados na modulación da plasticidade sináptica poderían estar implicados na etioloxía da esquizofrenia.
7. No Estudo III atopabamos un exceso de variantes raras de sentido incorrecto no xene *NRXN1* nos controis respecto ós casos, o cal confirmabamos mediante unha meta-análise feita no Estudo IV despois de incluír datos da secuenciación de *NRXN1* no total da mostra e de datos obtidos mediante unha busca bibliográfica. Isto podería estar indicando que pequenos cambios na regulación do equilibrio excitador/inhibidor nos circuitos neuronais poderían ser factores de protección fronte á esquizofrenia.



7. BIBLIOGRAFÍA





- Abramson JH. 2004. WINPEPI (PEPI-for-Windows): computer programs for epidemiologists. *Epidemiol Perspect Innov* **1**(1): 6.
- Adityanjee, Aderibigbe YA, Theodoridis D, Vieweg VR. 1999. Dementia praecox to schizophrenia: the first 100 years. *Psychiatry Clin Neurosci* **53**(4): 437-448.
- Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. *Nat Methods* **7**(4): 248-249.
- Aleman A, Kahn RS, Selten JP. 2003. Sex differences in the risk of schizophrenia: evidence from meta-analysis. *Arch Gen Psychiatry* **60**(6): 565-571.
- Allen NC, Bagade S, McQueen MB, Ioannidis JP, Kavvoura FK, Khoury MJ, Tanzi RE, Bertram L. 2008. Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat Genet* **40**(7): 827-834.
- Andreasen NC, Carpenter WT, Jr. 1993. Diagnosis and classification of schizophrenia. *Schizophr Bull* **19**(2): 199-214.
- Andreasen NC, Wilcox MA, Ho BC, Epping E, Ziebell S, Zeien E, Weiss B, Wassink T. 2011. Statistical epistasis and progressive brain change in schizophrenia: an approach for examining the relationships between multiple genes. *Mol Psychiatry*.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**(1): 25-29.
- Asimit J, Zeggini E. 2010. Rare variant association analysis methods for complex traits. *Annu Rev Genet* **44**: 293-308.
- Askland K, Read C, O'Connell C, Moore JH. 2011. Ion channels and schizophrenia: a gene set-based analytic approach to GWAS data for biological hypothesis testing. *Hum Genet* **131**(3): 373-391.
- Autism Genome Project C Szatmari P Paterson AD Zwaigenbaum L Roberts W Brian J Liu XQ Vincent JB Skaug JL Thompson AP et al. 2007. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* **39**(3): 319-328.
- Awadalla P, Gauthier J, Myers RA, Casals F, Hamdan FF, Griffing AR, Cote M, Henrion E, Spiegelman D, Tarabeux J et al. 2010. Direct measure of the de novo mutation rate in autism and schizophrenia cohorts. *Am J Hum Genet* **87**(3): 316-324.
- Badano JL, Teslovich TM, Katsanis N. 2005. The centrosome in human genetic disease. *Nat Rev Genet* **6**(3): 194-205.
- Badner JA, Gershon ES. 2002. Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. *Mol Psychiatry* **7**(4): 405-411.
- Balan S, Iwayama Y, Toyota T, Toyoshima M, Maekawa M, Yoshikawa T. 2014. 22q11.2 deletion carriers and schizophrenia-associated novel variants. *Br J Psychiatry* **204**: 398-399.
- Ban TA. 2004. Neuropsychopharmacology and the genetics of schizophrenia: a history of the diagnosis of schizophrenia. *Prog Neuropsychopharmacol Biol Psychiatry* **28**(5): 753-762.
- Bang ML, Owczarek S. 2013. A matter of balance: role of neurexin and neuroligin at the synapse. *Neurochem Res* **38**(6): 1174-1189.
- Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L. 2008. Natural selection has driven population differentiation in modern humans. *Nat Genet* **40**(3): 340-345.

Bibliografía

- Baum AE, Hamshere M, Green E, Cichon S, Rietschel M, Noethen MM, Craddock N, McMahon FJ. 2008. Meta-analysis of two genome-wide association studies of bipolar disorder reveals important points of agreement. *Mol Psychiatry* **13**(5): 466-467.
- Beiser M, Erickson D, Fleming JA, Iacono WG. 1993. Establishing the onset of psychotic illness. *Am J Psychiatry* **150**(9): 1349-1354.
- Benson MA, Newey SE, Martin-Rendon E, Hawkes R, Blake DJ. 2001. Dysbindin, a novel coiled-coil-containing protein that interacts with the dystrobrevins in muscle and brain. *J Biol Chem* **276**(26): 24232-24241.
- Bergen SE, Petryshen TL. 2012. Genome-wide association studies of schizophrenia: does bigger lead to better results? *Curr Opin Psychiatry* **25**(2): 76-82.
- Biederer T, Sudhof TC. 2001. CASK and protein 4.1 support F-actin nucleation on neuroligins. *J Biol Chem* **276**(51): 47869-47876.
- Blackwood DH, Fordyce A, Walker MT, St Clair DM, Porteous DJ, Muir WJ. 2001. Schizophrenia and affective disorders-- cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: clinical and P300 findings in a family. *Am J Hum Genet* **69**(2): 428-433.
- Blouin JL, Dombroski BA, Nath SK, Lasseter VK, Wolyniec PS, Nestadt G, Thornquist M, Ullrich G, McGrath J, Kasch L et al. 1998. Schizophrenia susceptibility loci on chromosomes 13q32 and 8p21. *Nat Genet* **20**(1): 70-73.
- Bord L, Wheeler J, Paek M, Saleh M, Lyons-Warren A, Ross CA, Sawamura N, Sawa A. 2006. Primate disrupted-in-schizophrenia-1 (DISC1): high divergence of a gene for major mental illnesses in recent evolutionary history. *Neurosci Res* **56**(3): 286-293.
- Bradshaw NJ, Christie S, Soares DC, Carlyle BC, Porteous DJ, Millar JK. 2009. NDE1 and NDEL1: multimerisation, alternate splicing and DISC1 interaction. *Neurosci Lett* **449**(3): 228-233.
- Bradshaw NJ, Ogawa F, Antolin-Fontes B, Chubb JE, Carlyle BC, Christie S, Claessens A, Porteous DJ, Millar JK. 2008. DISC1, PDE4B, and NDE1 at the centrosome and synapse. *Biochem Biophys Res Commun* **377**(4): 1091-1096.
- Braff DL, Ryan J, Rissling AJ, Carpenter WT. 2013. Lack of use in the literature from the last 20 years supports dropping traditional schizophrenia subtypes from DSM-5 and ICD-11. *Schizophr Bull* **39**(4): 751-753.
- Brandon NJ, Millar JK, Korth C, Sive H, Singh KK, Sawa A. 2009. Understanding the role of DISC1 in psychiatric disease and during normal development. *J Neurosci* **29**(41): 12768-12775.
- Brandon NJ, Sawa A. 2011. Linking neurodevelopmental and synaptic theories of mental illness through DISC1. *Nat Rev Neurosci* **12**(12): 707-722.
- Brown AS. 2006. Prenatal infection as a risk factor for schizophrenia. *Schizophr Bull* **32**(2): 200-202.
- . 2011. The environment and susceptibility to schizophrenia. *Progress in neurobiology* **93**(1): 23-58.
- Brzustowicz LM, Hodgkinson KA, Chow EW, Honer WG, Bassett AS. 2000. Location of a major susceptibility locus for familial schizophrenia on chromosome 1q21-q22. *Science* **288**(5466): 678-682.
- Bundy H, Stahl D, MacCabe JH. 2011. A systematic review and meta-analysis of the fertility of patients with schizophrenia and their unaffected relatives. *Acta Psychiatr Scand* **123**(2): 98-106.

- Burdick KE, Kamiya A, Hodgkinson CA, Lencz T, DeRosse P, Ishizuka K, Elashvili S, Arai H, Goldman D, Sawa A et al. 2008. Elucidating the relationship between DISC1, NDEL1 and NDE1 and the risk for schizophrenia: evidence of epistasis and competitive binding. *Hum Mol Genet* **17**(16): 2462-2473.
- Camargo LM, Collura V, Rain JC, Mizuguchi K, Hermjakob H, Kerrien S, Bonnert TP, Whiting PJ, Brandon NJ. 2007. Disrupted in Schizophrenia 1 Interactome: evidence for the close connectivity of risk genes and a potential synaptic basis for schizophrenia. *Mol Psychiatry* **12**(1): 74-86.
- Cantor RM, Lange K, Sinsheimer JS. 2010. Prioritizing GWAS results: A review of statistical methods and recommendations for their application. *Am J Hum Genet* **86**(1): 6-22.
- Carlisle HJ, Luong TN, Medina-Marino A, Schenker L, Khorosheva E, Indersmitten T, Gunapala KM, Steele AD, O'Dell TJ, Patterson PH et al. 2011. Deletion of densin-180 results in abnormal behaviors associated with mental illness and reduces mGluR5 and DISC1 in the postsynaptic density fraction. *J Neurosci* **31**(45): 16194-16207.
- Carlson CS, Eberle MA, Kruglyak L, Nickerson DA. 2004. Mapping complex disease loci in whole-genome association studies. *Nature* **429**(6990): 446-452.
- Carrera N, Arrojo M, Sanjuan J, Ramos-Rios R, Paz E, Suarez-Rama JJ, Paramo M, Agra S, Brenlla J, Martinez S et al. 2012. Association study of nonsynonymous single nucleotide polymorphisms in schizophrenia. *Biol Psychiatry* **71**(2): 169-177.
- Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur O, Anwar N, Schultz N, Bader GD, Sander C. 2011. Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res* **39**(Database issue): D685-690.
- Ceulemans S, van der Ven K, Del-Favero J. 2012. Targeted screening and validation of copy number variations. *Methods Mol Biol* **838**: 311-328.
- Clop A, Marcq F, Takeda H, Pirottin D, Tordoir X, Bibe B, Bouix J, Caiment F, Elsen JM, Eychenne F et al. 2006. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nat Genet* **38**(7): 813-818.
- Consortium EP. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**(7414): 57-74.
- Consortium P, Akbarian S, Liu C, Knowles JA, Vaccarino FM, Farnham PJ, Crawford GE, Jaffe AE, Pinto D, Dracheva S et al. 2015. The PsychENCODE project. *Nat Neurosci* **18**(12): 1707-1712.
- Cooper GM, Goode DL, Ng SB, Sidow A, Bamshad MJ, Shendure J, Nickerson DA. 2010. Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. *Nat Methods* **7**(4): 250-251.
- Cornblatt B, Obuchowski M, Roberts S, Pollack S, Erlenmeyer-Kimling L. 1999. Cognitive and behavioral precursors of schizophrenia. *Dev Psychopathol* **11**(3): 487-508.
- Corvin A, Sullivan PF. 2016. What Next in Schizophrenia Genetics for the Psychiatric Genomics Consortium? *Schizophr Bull* **42**(3): 538-541.
- Costas J, Suarez-Rama JJ, Carrera N, Paz E, Paramo M, Agra S, Brenlla J, Ramos-Rios R, Arrojo M. 2013. Role of DISC1 interacting proteins in schizophrenia risk from genome-wide analysis of missense SNPs. *Ann Hum Genet* **77**(6): 504-512.
- Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, Caudy M, Garapati P, Gillespie M, Kamdar MR et al. 2014. The Reactome pathway knowledgebase. *Nucleic Acids Res* **42**(Database issue): D472-477.

Bibliografia

- Cross-Disorder Group of the Psychiatric Genomics C. 2013. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**(9875): 1371-1379.
- Cross-Disorder Group of the Psychiatric Genomics C Lee SH Ripke S Neale BM Faraone SV Purcell SM Perlis RH Mowry BJ Thapar A Goddard ME et al. 2013. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* **45**(9): 984-994.
- Crowley JJ, Hilliard CE, Kim Y, Morgan MB, Lewis LR, Muzny DM, Hawes AC, Sabo A, Wheeler DA, Lieberman JA et al. 2013. Deep resequencing and association analysis of schizophrenia candidate genes. *Mol Psychiatry* **18**(2): 138-140.
- Ching MS, Shen Y, Tan WH, Jeste SS, Morrow EM, Chen X, Mukaddes NM, Yoo SY, Hanson E, Hundley R et al. 2010. Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders. *Am J Med Genet B Neuropsychiatr Genet* **153B**(4): 937-947.
- Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloglu A, Ozen S, Sanjad S et al. 2009. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A* **106**(45): 19096-19101.
- Choi Y, Chan AP. 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31**(16): 2745-2747.
- Chubb JE, Bradshaw NJ, Soares DC, Porteous DJ, Millar JK. 2008. The DISC locus in psychiatric illness. *Mol Psychiatry* **13**(1): 36-64.
- Chun S, Fay JC. 2009. Identification of deleterious mutations within three human genomes. *Genome Res* **19**(9): 1553-1561.
- Dammermann A, Merdes A. 2002. Assembly of centrosomal proteins and microtubule organization depends on PCM-1. *J Cell Biol* **159**(2): 255-266.
- Davies G, Welham J, Chant D, Torrey EF, McGrath J. 2003. A systematic review and meta-analysis of Northern Hemisphere season of birth studies in schizophrenia. *Schizophr Bull* **29**(3): 587-593.
- de Leeuw CA, Neale BM, Heskes T, Posthuma D. 2016. The statistical properties of gene-set analysis. *Nat Rev Genet*.
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**(5): 491-498.
- Derkach A, Lawless JF, Sun L. 2013. Robust and powerful tests for rare variants using Fisher's method to combine evidence of association from two or more complementary tests. *Genet Epidemiol* **37**(1): 110-121.
- Detera-Wadleigh SD, McMahon FJ. 2006. G72/G30 in schizophrenia and bipolar disorder: review and meta-analysis. *Biol Psychiatry* **60**(2): 106-114.
- Devlin B, Roeder K, Wasserman L. 2001. Genomic control, a new approach to genetic-based association studies. *Theor Popul Biol* **60**(3): 155-166.
- Driscoll DA, Salvin J, Sellinger B, Budarf ML, McDonald-McGinn DM, Zackai EH, Emanuel BS. 1993. Prevalence of 22q11 microdeletions in DiGeorge and velocardiofacial syndromes: implications for genetic counselling and prenatal diagnosis. *J Med Genet* **30**(10): 813-817.
- Driscoll DA, Spinner NB, Budarf ML, McDonald-McGinn DM, Zackai EH, Goldberg RB, Shprintzen RJ, Saal HM, Zonana J, Jones MC et al. 1992. Deletions and microdeletions of 22q11.2 in velo-cardio-facial syndrome. *Am J Med Genet* **44**(2): 261-268.

- Duan J, Shi J, Fiorentino A, Leites C, Chen X, Moy W, Chen J, Alexandrov BS, Usheva A, He D et al. 2014. A rare functional noncoding variant at the GWAS-implicated MIR137/MIR2682 locus might confer risk to schizophrenia and bipolar disorder. *Am J Hum Genet* **95**(6): 744-753.
- Duan X, Chang JH, Ge S, Faulkner RL, Kim JY, Kitabatake Y, Liu XB, Yang CH, Jordan JD, Ma DK et al. 2007. Disrupted-In-Schizophrenia 1 regulates integration of newly generated neurons in the adult brain. *Cell* **130**(6): 1146-1158.
- Dupont WD, Plummer WD, Jr. 1998. Power and sample size calculations for studies involving linear regression. *Controlled clinical trials* **19**(6): 589-601.
- Dwyer S, Carroll L, Mantripragada KK, Owen MJ, O'Donovan MC, Williams NM. 2010a. Mutation screening of the DTNBP1 exonic sequence in 669 schizophrenics and 710 controls using high-resolution melting analysis. *Am J Med Genet B Neuropsychiatr Genet* **153B**(3): 766-774.
- Dwyer S, Williams H, Holmans P, Moskvina V, Craddock N, Owen MJ, O'Donovan MC. 2010b. No evidence that rare coding variants in ZNF804A confer risk of schizophrenia. *Am J Med Genet B Neuropsychiatr Genet* **153B**(8): 1411-1416.
- Eaton WW, Byrne M, Ewald H, Mors O, Chen CY, Agerbo E, Mortensen PB. 2006. Association of schizophrenia and autoimmune diseases: linkage of Danish national registers. *Am J Psychiatry* **163**(3): 521-528.
- Edelmann L, Pandita RK, Morrow BE. 1999. Low-copy repeats mediate the common 3-Mb deletion in patients with velo-cardio-facial syndrome. *Am J Hum Genet* **64**(4): 1076-1086.
- Edwards HV, Christian F, Baillie GS. 2012. cAMP: novel concepts in compartmentalised signalling. *Seminars in cell & developmental biology* **23**(2): 181-190.
- Elbers CC, van Eijk KR, Franke L, Mulder F, van der Schouw YT, Wijmenga C, Onland-Moret NC. 2009. Using genome-wide pathway analysis to unravel the etiology of complex diseases. *Genet Epidemiol* **33**(5): 419-431.
- Fan J, Beck KA. 2004. A role for the spectrin superfamily member Syne-1 and kinesin II in cytokinesis. *J Cell Sci* **117**(Pt 4): 619-629.
- Farrell MS, Werge T, Sklar P, Owen MJ, Ophoff RA, O'Donovan MC, Corvin A, Cichon S, Sullivan PF. 2015. Evaluating historical candidate genes for schizophrenia. *Mol Psychiatry* **20**(5): 555-562.
- Felsenstein J, Churchill GA. 1996. A Hidden Markov Model approach to variation among sites in rate of evolution. *Mol Biol Evol* **13**(1): 93-104.
- Feng Y, Walsh CA. 2001. Protein-protein interactions, cytoskeletal regulation and neuronal migration. *Nat Rev Neurosci* **2**(6): 408-416.
- Flores R, 3rd, Hirota Y, Armstrong B, Sawa A, Tomoda T. 2011. DISC1 regulates synaptic vesicle transport via a lithium-sensitive pathway. *Neurosci Res* **71**(1): 71-77.
- Forrest M, Chapman RM, Doyle AM, Tinsley CL, Waite A, Blake DJ. 2012. Functional analysis of TCF4 missense mutations that cause Pitt-Hopkins syndrome. *Hum Mutat* **33**(12): 1676-1686.
- Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM et al. 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**(7164): 851-861.
- Fridley BL, Biernacka JM. 2011. Gene set analysis of SNP data: benefits, challenges, and future directions. *Eur J Hum Genet* **19**(8): 837-843.

Bibliografía

- Friedman JI, Vrijenhoek T, Markx S, Janssen IM, van der Vliet WA, Faas BH, Knoers NV, Cahn W, Kahn RS, Edelmann L et al. 2008. CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy. *Mol Psychiatry* **13**(3): 261-266.
- Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P, Georgieva L, Rees E, Palta P, Ruderfer DM et al. 2014. De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506**(7487): 179-184.
- Frosst P, Blom HJ, Milos R, Goyette P, Sheppard CA, Matthews RG, Boers GJ, den Heijer M, Kluijtmans LA, van den Heuvel LP et al. 1995. A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nat Genet* **10**(1): 111-113.
- Garber M, Guttman M, Clamp M, Zody MC, Friedman N, Xie X. 2009. Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics* **25**(12): i54-62.
- Gasso P, Sanchez-Gistau V, Mas S, Sugranyes G, Rodriguez N, Boloc D, de la Serna E, Romero S, Moreno D, Moreno C et al. 2016. Association of CACNA1C and SYNE1 in offspring of patients with psychiatric disorders. *Psychiatry Res* **245**: 427-435.
- Gauthier J, Siddiqui TJ, Huashan P, Yokomaku D, Hamdan FF, Champagne N, Lapointe M, Spiegelman D, Noreau A, Lafreniere RG et al. 2011. Truncating mutations in NRXN2 and NRXN1 in autism spectrum disorders and schizophrenia. *Hum Genet* **130**(4): 563-573.
- Genovese G, Fromer M, Stahl EA, Ruderfer DM, Chambert K, Landen M, Moran JL, Purcell SM, Sklar P, Sullivan PF et al. 2016. Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat Neurosci* **19**(11): 1433-1441.
- Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J et al. 2005. Galaxy: a platform for interactive large-scale genome analysis. *Genome Res* **15**(10): 1451-1455.
- Girard SL, Dion PA, Bourassa CV, Geoffroy S, Lachance-Touchette P, Barhdadi A, Langlois M, Joober R, Krebs MO, Dube MP et al. 2015. Mutation burden of rare variants in schizophrenia candidate genes. *PLoS One* **10**(6): e0128988.
- Girard SL, Gauthier J, Noreau A, Xiong L, Zhou S, Jouan L, Dionne-Laporte A, Spiegelman D, Henrion E, Diallo O et al. 2011. Increased exonic de novo mutation rate in individuals with schizophrenia. *Nat Genet* **43**(9): 860-863.
- Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S, Zhang H, Estes A, Brune CW, Bradfield JP et al. 2009. Autism genome-wide copy number variation reveals ubiquitous and neuronal genes. *Nature* **459**(7246): 569-573.
- Goecks J, Nekrutenko A, Taylor J, Galaxy T. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* **11**(8): R86.
- Gonzalez-Penas J, Amigo J, Santome L, Sobrino B, Brenlla J, Agra S, Paz E, Paramo M, Carracedo A, Arrojo M et al. 2016. Targeted resequencing of regulatory regions at schizophrenia risk loci: Role of rare functional variants at chromatin repressive states. *Schizophr Res* **174**(1-3): 10-16.
- Goodship J, Cross I, LiLing J, Wren C. 1998. A population study of chromosome 22q11 deletions in infancy. *Arch Dis Child* **79**(4): 348-351.
- Gorlov IP, Gorlova OY, Sunyaev SR, Spitz MR, Amos CI. 2008. Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. *Am J Hum Genet* **82**(1): 100-112.

- Gottesman, II, Shields J. 1967. A polygenic theory of schizophrenia. *Proc Natl Acad Sci U S A* **58**(1): 199-205.
- Gratacos M, Costas J, de Cid R, Bayes M, Gonzalez JR, Baca-Garcia E, de Diego Y, Fernandez-Aranda F, Fernandez-Piqueras J, Guitart M et al. 2009. Identification of new putative susceptibility genes for several psychiatric disorders by association analysis of regulatory and non-synonymous SNPs of 306 genes involved in neurotransmission and neurodevelopment. *Am J Med Genet B Neuropsychiatr Genet* **150B**(6): 808-816.
- Green EK, Grozeva D, Sims R, Raybould R, Forty L, Gordon-Smith K, Russell E, St Clair D, Young AH, Ferrier IN et al. 2011. DISC1 exon 11 rare variants found more commonly in schizoaffective spectrum cases than controls. *Am J Med Genet B Neuropsychiatr Genet* **156B**(4): 490-492.
- Green EK, Hamshere M, Forty L, Gordon-Smith K, Fraser C, Russell E, Grozeva D, Kirov G, Holmans P, Moran JL et al. 2013. Replication of bipolar disorder susceptibility alleles and identification of two novel genome-wide significant associations in a new bipolar disorder case-control sample. *Mol Psychiatry* **18**(12): 1302-1307.
- Gregor A, Albrecht B, Bader I, Bijlsma EK, Ekici AB, Engels H, Hackmann K, Horn D, Hoyer J, Klapecki J et al. 2011. Expanding the clinical spectrum associated with defects in CNTNAP2 and NRXN1. *BMC Med Genet* **12**: 106.
- Grozeva D, Conrad DF, Barnes CP, Hurles M, Owen MJ, O'Donovan MC, Craddock N, Kirov G, Wtccc. 2012. Independent estimation of the frequency of rare CNVs in the UK population confirms their role in schizophrenia. *Schizophr Res* **135**(1-3): 1-7.
- Guha S, Rees E, Darvasi A, Ivanov D, Ikeda M, Bergen SE, Magnusson PK, Cormican P, Morris D, Gill M et al. 2013. Implication of a rare deletion at distal 16p11.2 in schizophrenia. *JAMA psychiatry* **70**(3): 253-260.
- Guipponi M, Santoni FA, Setola V, Gehrig C, Rotharmel M, Cuenca M, Guillin O, Dikeos D, Georgantopoulos G, Papadimitriou G et al. 2014. Exome sequencing in 53 sporadic cases of schizophrenia identifies 18 putative candidate genes. *PLoS One* **9**(11): e112745.
- Guo AY, Sun J, Riley BP, Thiselton DL, Kendler KS, Zhao Z. 2009. The dystrobrevin-binding protein 1 gene: features and networks. *Mol Psychiatry* **14**(1): 18-29.
- Hafner H. 1998. Onset and course of the first schizophrenic episode. *The Kaohsiung journal of medical sciences* **14**(7): 413-431.
- . 2003. Gender differences in schizophrenia. *Psychoneuroendocrinology* **28 Suppl 2**: 17-54.
- Han F, Pan W. 2010. A data-adaptive sum test for disease association with multiple common or rare variants. *Hum Hered* **70**(1): 42-54.
- Hannes FD, Sharp AJ, Mefford HC, de Ravel T, Ruivenkamp CA, Breuning MH, Fryns JP, Devriendt K, Van Buggenhout G, Vogels A et al. 2009. Recurrent reciprocal deletions and duplications of 16p13.11: the deletion is a risk factor for MR/MCA while the duplication may be a rare benign variant. *J Med Genet* **46**(4): 223-232.
- Haukka J, Suvisaari J, Lonnqvist J. 2003. Fertility of patients with schizophrenia, their siblings, and the general population: a cohort study from 1950 to 1959 in Finland. *Am J Psychiatry* **160**(3): 460-463.
- Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, van de Lagemaat LN, Smith KA, Ebbert A, Riley ZL et al. 2012. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**(7416): 391-399.
- Hayashi-Takagi A, Takaki M, Graziane N, Seshadri S, Murdoch H, Dunlop AJ, Makino Y, Seshadri AJ, Ishizuka K, Srivastava DP et al. 2010. Disrupted-in-Schizophrenia 1

Bibliografía

- (DISC1) regulates spines of the glutamate synapse via Rac1. *Nat Neurosci* **13**(3): 327-332.
- Hennah W, Thomson P, Peltonen L, Porteous D. 2006. Genes and schizophrenia: beyond schizophrenia: the role of DISC1 in major mental illness. *Schizophr Bull* **32**(3): 409-416.
- Hennah W, Varilo T, Kestila M, Paunio T, Arajärvi R, Haukka J, Parker A, Martin R, Levitzky S, Partonen T et al. 2003. Haplotype transmission analysis provides evidence of association for DISC1 to schizophrenia and suggests sex-dependent effects. *Hum Mol Genet* **12**(23): 3151-3159.
- Hiard S, Charlier C, Coppieters W, Georges M, Baurain D. 2009. Patrocles: a database of polymorphic miRNA-mediated gene regulation in vertebrates. *Nucleic Acids Res* **38**(Database issue): D640-651.
- Higgins J. 1976. Effects of child rearing by schizophrenic mothers: a follow-up. *J Psychiatr Res* **13**(1): 1-9.
- Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106**(23): 9362-9367.
- Hirokawa N, Noda Y, Tanaka Y, Niwa S. 2009. Kinesin superfamily motor proteins and intracellular transport. *Nat Rev Mol Cell Biol* **10**(10): 682-696.
- Hirschhorn JN, Daly MJ. 2005. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* **6**(2): 95-108.
- Hoppins S, Collins SR, Cassidy-Stone A, Hummel E, Devay RM, Lackner LL, Westermann B, Schuldiner M, Weissman JS, Nunnari J. 2011. A mitochondrial-focused genetic interaction map reveals a scaffold-like complex required for inner membrane organization in mitochondria. *J Cell Biol* **195**(2): 323-340.
- Hu X, Zhang B, Liu W, Paciga S, He W, Lanz TA, Kleiman R, Dougherty B, Hall SK, McIntosh AM et al. 2014. A survey of rare coding variants in candidate genes in schizophrenia by deep sequencing. *Mol Psychiatry* **19**(8): 857-858.
- Huang DW, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, Guo Y, Stephens R, Baseler MW, Lane HC et al. 2007a. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res* **35**(Web Server issue): W169-175.
- Huang FL, Huang KP, Boucheron C. 2007b. Long-term enrichment enhances the cognitive behavior of the aging neurogranin null mice without affecting their hippocampal LTP. *Learning & memory* **14**(8): 512-519.
- Hurles ME, Dermitzakis ET, Tyler-Smith C. 2008. The functional impact of structural variation in humans. *Trends Genet* **24**(5): 238-245.
- Iafraite AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. 2004. Detection of large-scale variation in the human genome. *Nat Genet* **36**(9): 949-951.
- Ikeda M, Aleksic B, Kirov G, Kinoshita Y, Yamanouchi Y, Kitajima T, Kawashima K, Okochi T, Kishi T, Zaharieva I et al. 2010. Copy number variation in schizophrenia in the Japanese population. *Biol Psychiatry* **67**(3): 283-286.
- Ikeda M, Hikita T, Taya S, Uraguchi-Asaki J, Toyo-oka K, Wynshaw-Boris A, Ujike H, Inada T, Takao K, Miyakawa T et al. 2008. Identification of YWHAE, a gene encoding 14-3-3epsilon, as a possible susceptibility gene for schizophrenia. *Hum Mol Genet* **17**(20): 3212-3222.

- Ingason A, Kirov G, Giegling I, Hansen T, Isles AR, Jakobsen KD, Kristinsson KT, le Roux L, Gustafsson O, Craddock N et al. 2011a. Maternally derived microduplications at 15q11-q13: implication of imprinted genes in psychotic illness. *Am J Psychiatry* **168**(4): 408-417.
- Ingason A, Rujescu D, Cichon S, Sigurdsson E, Sigmundsson T, Pietilainen OP, Buizer-Voskamp JE, Strengman E, Francks C, Muglia P et al. 2011b. Copy number variations of chromosome 16p13.1 region associated with schizophrenia. *Mol Psychiatry* **16**(1): 17-25.
- Ingraham LJ, Kety SS. 2000. Adoption studies of schizophrenia. *Am J Med Genet* **97**(1): 18-22.
- Insel TR. 2010. Rethinking schizophrenia. *Nature* **468**(7321): 187-193.
- International HapMap C. 2003. The International HapMap Project. *Nature* **426**(6968): 789-796.
- International Schizophrenia C. 2008. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**(7210): 237-241.
- International Schizophrenia C, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P. 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**(7256): 748-752.
- Ioannidis JP, Boffetta P, Little J, O'Brien TR, Uitterlinden AG, Vineis P, Balding DJ, Chokkalingam A, Dolan SM, Flanders WD et al. 2008. Assessment of cumulative evidence on genetic associations: interim guidelines. *Int J Epidemiol* **37**(1): 120-132.
- Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, Yamrom B, Lee YH, Narzisi G, Leotta A et al. 2012. De novo gene disruptions in children on the autistic spectrum. *Neuron* **74**(2): 285-299.
- Ishizuka K, Kimura H, Wang C, Xing J, Kushima I, Arioka Y, Oya-Ito T, Uno Y, Okada T, Mori D et al. 2016. Investigation of Rare Single-Nucleotide PCDH15 Variants in Schizophrenia and Autism Spectrum Disorders. *PLoS One* **11**(4): e0153224.
- Itsara A, Wu H, Smith JD, Nickerson DA, Romieu I, London SJ, Eichler EE. 2010. De novo rates and selection of large copy number variation. *Genome Res* **20**(11): 1469-1481.
- Jenkins AK, Paterson C, Wang Y, Hyde TM, Kleinman JE, Law AJ. 2016. Neurexin 1 (NRXN1) splice isoform expression during human neocortical development and aging. *Mol Psychiatry* **21**(5): 701-706.
- Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M et al. 2009. STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* **37**(Database issue): D412-416.
- Jia P, Wang L, Fanous AH, Chen X, Kendler KS, International Schizophrenia C, Zhao Z. 2012. A bias-reducing pathway enrichment analysis of genome-wide association data confirmed association of the MHC region with schizophrenia. *J Med Genet* **49**(2): 96-103.
- Jia P, Wang L, Meltzer HY, Zhao Z. 2010. Common variants conferring risk of schizophrenia: a pathway analysis of GWAS data. *Schizophr Res* **122**(1-3): 38-42.
- Kamiya A, Kubo K, Tomoda T, Takaki M, Youn R, Ozeki Y, Sawamura N, Park U, Kudo C, Okawa M et al. 2005. A schizophrenia-associated mutation of DISC1 perturbs cerebral cortex development. *Nat Cell Biol* **7**(12): 1167-1178.
- Kamiya A, Sedlak TW, Pletnikov MV. 2012. DISC1 Pathway in Brain Development: Exploring Therapeutic Targets for Major Psychiatric Disorders. *Front Psychiatry* **3**: 25.
- Kamiya A, Tan PL, Kubo K, Engelhard C, Ishizuka K, Kubo A, Tsukita S, Pulver AE, Nakajima K, Cascella NG et al. 2008. Recruitment of PCM1 to the centrosome by the

Bibliografía

- cooperative action of DISC1 and BBS4: a candidate for psychiatric illnesses. *Arch Gen Psychiatry* **65**(9): 996-1006.
- Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**(1): 27-30.
- Kang HJ, Choi KO, Kim BD, Kim S, Kim YJ. 2005. FESD: a Functional Element SNPs Database in human. *Nucleic Acids Res* **33**(Database issue): D518-522.
- Karayiorgou M, Morris MA, Morrow B, Shprintzen RJ, Goldberg R, Borrow J, Gos A, Nestadt G, Wolynec PS, Lasseter VK et al. 1995. Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. *Proc Natl Acad Sci U S A* **92**(17): 7612-7616.
- Kenny EM, Cormican P, Furlong S, Heron E, Kenny G, Fahey C, Kelleher E, Ennis S, Tropea D, Anney R et al. 2013. Excess of rare novel loss-of-function variants in synaptic genes in schizophrenia and autism spectrum disorders. *Mol Psychiatry*.
- Keshavan MS, Tandon R, Boutros NN, Nasrallah HA. 2008. Schizophrenia, "just the facts": what we know in 2008 Part 3: neurobiology. *Schizophr Res* **106**(2-3): 89-107.
- Kim HG, Kishikawa S, Higgins AW, Seong IS, Donovan DJ, Shen Y, Lally E, Weiss LA, Najm J, Kutsche K et al. 2008. Disruption of neurexin 1 associated with autism spectrum disorder. *Am J Hum Genet* **82**(1): 199-207.
- Kim JY, Duan X, Liu CY, Jang MH, Guo JU, Pow-anpongkul N, Kang E, Song H, Ming GL. 2009. DISC1 regulates new neuron development in the adult brain via modulation of AKT-mTOR signaling through KIAA1212. *Neuron* **63**(6): 761-773.
- Kimura H, Tsuboi D, Wang C, Kushima I, Koide T, Ikeda M, Iwayama Y, Toyota T, Yamamoto N, Kunimoto S et al. 2015. Identification of Rare, Single-Nucleotide Mutations in NDE1 and Their Contributions to Schizophrenia Susceptibility. *Schizophr Bull* **41**(3): 744-753.
- Kirkpatrick B, Xu L, Cascella N, Ozeki Y, Sawa A, Roberts RC. 2006. DISC1 immunoreactivity at the light and ultrastructural level in the human neocortex. *J Comp Neurol* **497**(3): 436-450.
- Kirov G, Grozeva D, Norton N, Ivanov D, Mantripragada KK, Holmans P, International Schizophrenia C, Wellcome Trust Case Control C, Craddock N, Owen MJ et al. 2009a. Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Hum Mol Genet* **18**(8): 1497-1503.
- Kirov G, Gumus D, Chen W, Norton N, Georgieva L, Sari M, O'Donovan MC, Erdogan F, Owen MJ, Ropers HH et al. 2008. Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. *Hum Mol Genet* **17**(3): 458-465.
- Kirov G, Pocklington AJ, Holmans P, Ivanov D, Ikeda M, Ruderfer D, Moran J, Chambert K, Toncheva D, Georgieva L et al. 2012. De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol Psychiatry* **17**(2): 142-153.
- Kirov G, Rees E, Walters JT, Escott-Price V, Georgieva L, Richards AL, Chambert KD, Davies G, Legge SE, Moran JL et al. 2014. The penetrance of copy number variations for schizophrenia and developmental delay. *Biol Psychiatry* **75**(5): 378-385.
- Kirov G, Rujescu D, Ingason A, Collier DA, O'Donovan MC, Owen MJ. 2009b. Neurexin 1 (NRXN1) deletions in schizophrenia. *Schizophr Bull* **35**(5): 851-854.
- Kirov G, Zaharieva I, Georgieva L, Moskvina V, Nikolov I, Cichon S, Hillmer A, Toncheva D, Owen MJ, O'Donovan MC. 2009c. A genome-wide association study in 574 schizophrenia trios using DNA pooling. *Mol Psychiatry* **14**(8): 796-803.

- Kobayashi Y, Yang S, Nykamp K, Garcia J, Lincoln SE, Topper SE. 2017. Pathogenic variant burden in the ExAC database: an empirical approach to evaluating population data for clinical variant interpretation. *Genome medicine* **9**(1): 13.
- Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, Weinstock GM, Wilson RK, Ding L. 2009. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* **25**(17): 2283-2285.
- Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* **22**(3): 568-576.
- Kondo S, Sato-Yoshitake R, Noda Y, Aizawa H, Nakata T, Matsuura Y, Hirokawa N. 1994. KIF3A is a new microtubule-based anterograde motor in the nerve axon. *J Cell Biol* **125**(5): 1095-1107.
- Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson A, Jonasdottir A, Jonasdottir A et al. 2012. Rate of de novo mutations and the importance of father's age to disease risk. *Nature* **488**(7412): 471-475.
- Kryukov GV, Pennacchio LA, Sunyaev SR. 2007. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am J Hum Genet* **80**(4): 727-739.
- Kubota K, Kumamoto N, Matsuzaki S, Hashimoto R, Hattori T, Okuda H, Takamura H, Takeda M, Katayama T, Tohyama M. 2009. Dysbindin engages in c-Jun N-terminal kinase activity and cytoskeletal organization. *Biochem Biophys Res Commun* **379**(2): 191-195.
- Kushima I, Nakamura Y, Aleksic B, Ikeda M, Ito Y, Shiino T, Okochi T, Fukuo Y, Ujike H, Suzuki M et al. 2012. Resequencing and association analysis of the KALRN and EPHB1 genes and their contribution to schizophrenia susceptibility. *Schizophr Bull* **38**(3): 552-560.
- Lam C, Vergnolle MA, Thorpe L, Woodman PG, Allan VJ. 2010. Functional interplay between LIS1, NDE1 and NDEL1 in dynein-dependent organelle positioning. *J Cell Sci* **123**(Pt 2): 202-212.
- Lander ES. 1996. The new genomics: global views of biology. *Science* **274**(5287): 536-539.
- Larimore J, Tornieri K, Ryder PV, Gokhale A, Zlatic SA, Craige B, Lee JD, Talbot K, Pare JF, Smith Y et al. 2011. The schizophrenia susceptibility factor dysbindin and its associated complex sort cargoes from cell bodies to the synapse. *Mol Biol Cell* **22**(24): 4854-4867.
- Lee S, Abecasis GR, Boehnke M, Lin X. 2014. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* **95**(1): 5-23.
- Lee S, Wu MC, Lin X. 2012a. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* **13**(4): 762-775.
- Lee SH, DeCandia TR, Ripke S, Yang J, Schizophrenia Psychiatric Genome-Wide Association Study C, International Schizophrenia C, Molecular Genetics of Schizophrenia C, Sullivan PF, Goddard ME, Keller MC et al. 2012b. Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat Genet* **44**(3): 247-250.
- Lee SH, Decandia TR, Ripke S, Yang J, Sullivan PF, Goddard ME, Keller MC, Visscher PM, Wray NR. 2012c. Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat Genet* **44**(3): 247-250.
- Lee SJ, Uemura T, Yoshida T, Mishina M. 2012d. GluRdelta2 assembles four neurexins into trans-synaptic triad to trigger synapse formation. *J Neurosci* **32**(13): 4688-4701.

Bibliografía

- Lee YH, Kim JH, Song GG. 2013. Pathway analysis of a genome-wide association study in schizophrenia. *Gene* **525**(1): 107-115.
- Lencz T, Morgan TV, Athanasiou M, Dain B, Reed CR, Kane JM, Kucherlapati R, Malhotra AK. 2007. Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. *Mol Psychiatry* **12**(6): 572-580.
- Leung CL, Green KJ, Liem RK. 2002. Plakins: a family of versatile cytolinker proteins. *Trends Cell Biol* **12**(1): 37-45.
- Levinson DF, Duan J, Oh S, Wang K, Sanders AR, Shi J, Zhang N, Mowry BJ, Olincy A, Amin F et al. 2011. Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. *Am J Psychiatry* **168**(3): 302-316.
- Lewis BP, Burge CB, Bartel DP. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**(1): 15-20.
- Lewis CM, Levinson DF, Wise LH, DeLisi LE, Straub RE, Hovatta I, Williams NM, Schwab SG, Pulver AE, Faraone SV et al. 2003. Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. *Am J Hum Genet* **73**(1): 34-48.
- Li B, Leal SM. 2008. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet* **83**(3): 311-321.
- Li B, Liu DJ, Leal SM. 2013a. Identifying rare variants associated with complex traits via sequencing. *Curr Protoc Hum Genet* **Chapter 1**: Unit 1 26.
- Li JM, Lu CL, Cheng MC, Luu SU, Hsu SH, Chen CH. 2013b. Exonic resequencing of the DLGAP3 gene as a candidate gene for schizophrenia. *Psychiatry Res* **208**(1): 84-87.
- Li Q, Yu K. 2008. Improved correction for population stratification in genome-wide association studies by identifying hidden population structures. *Genet Epidemiol* **32**(3): 215-226.
- Li Y, Sidore C, Kang HM, Boehnke M, Abecasis GR. 2011. Low-coverage sequencing: implications for design of complex trait association studies. *Genome Res* **21**(6): 940-951.
- Li Y, Vinckenbosch N, Tian G, Huerta-Sanchez E, Jiang T, Jiang H, Albrechtsen A, Andersen G, Cao H, Korneliussen T et al. 2010. Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nat Genet* **42**(11): 969-972.
- Lichtenstein P, Yip BH, Bjork C, Pawitan Y, Cannon TD, Sullivan PF, Hultman CM. 2009. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* **373**(9659): 234-239.
- Lim SH, Kwon SK, Lee MK, Moon J, Jeong DG, Park E, Kim SJ, Park BC, Lee SC, Ryu SE et al. 2009. Synapse formation regulated by protein tyrosine phosphatase receptor T through interaction with cell adhesion molecules and Fyn. *EMBO J* **28**(22): 3564-3578.
- Lin DY, Tang ZZ. 2011. A general framework for detecting disease associations with rare variants in sequencing studies. *Am J Hum Genet* **89**(3): 354-367.
- Lin JW, Wyszynski M, Madhavan R, Sealock R, Kim JU, Sheng M. 1998. Yotiao, a novel protein of neuromuscular junction and brain that interacts with specific splice variants of NMDA receptor subunit NR1. *J Neurosci* **18**(6): 2017-2027.
- Linhoff MW, Lauren J, Cassidy RM, Dobie FA, Takahashi H, Nygaard HB, Airaksinen MS, Strittmatter SM, Craig AM. 2009. An unbiased expression screen for synaptogenic proteins identifies the LRRTM protein family as synaptic organizers. *Neuron* **61**(5): 734-749.

- Liu DJ, Leal SM. 2010. A novel adaptive method for the analysis of next-generation sequencing data to detect complex trait associations with rare variants due to gene main effects and interactions. *PLoS Genet* **6**(10): e1001156.
- Loh PR, Bhatia G, Gusev A, Finucane HK, Bulik-Sullivan BK, Pollack SJ, Schizophrenia Working Group of Psychiatric Genomics C, de Candia TR, Lee SH, Wray NR et al. 2015. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat Genet* **47**(12): 1385-1392.
- Luo X, Jin C, Zhou Z, Zhang F, Yuan J, Liu X, Cheng Z. 2016. Association study of DISC1 genetic variants with the risk of schizophrenia. *Psychiatr Genet*.
- Lupski JR, Stankiewicz P. 2005. Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. *PLoS Genet* **1**(6): e49.
- Ma L, Liu Y, Ky B, Shughrue PJ, Austin CP, Morris JA. 2002. Cloning and characterization of Disc1, the mouse ortholog of DISC1 (Disrupted-in-Schizophrenia 1). *Genomics* **80**(6): 662-672.
- Ma TM, Abazyan S, Abazyan B, Nomura J, Yang C, Seshadri S, Sawa A, Snyder SH, Pletnikov MV. 2013. Pathogenic disruption of DISC1-serine racemase binding elicits schizophrenia-like behavior via D-serine depletion. *Mol Psychiatry* **18**(5): 557-567.
- MacArthur DG, Tyler-Smith C. 2010. Loss-of-function variants in the genomes of healthy humans. *Hum Mol Genet* **19**(R2): R125-130.
- Madsen BE, Browning SR. 2009. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet* **5**(2): e1000384.
- Maglott D, Ostell J, Pruitt KD, Tatusova T. 2005. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res* **33**(Database issue): D54-58.
- Mah S, Nelson MR, Delisi LE, Reneland RH, Markward N, James MR, Nyholt DR, Hayward N, Handoko H, Mowry B et al. 2006. Identification of the semaphorin receptor PLXNA2 as a candidate for susceptibility to schizophrenia. *Mol Psychiatry* **11**(5): 471-478.
- Maher BJ, LoTurco JJ. 2012. Disrupted-in-schizophrenia (DISC1) functions presynaptically at glutamatergic synapses. *PLoS One* **7**(3): e34053.
- Malavasi EL, Ogawa F, Porteous DJ, Millar JK. 2012. DISC1 variants 37W and 607F disrupt its nuclear targeting and regulatory role in ATF4-mediated transcription. *Hum Mol Genet* **21**(12): 2779-2792.
- Malhotra D, Sebat J. 2012. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell* **148**(6): 1223-1241.
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A et al. 2009. Finding the missing heritability of complex diseases. *Nature* **461**(7265): 747-753.
- Mao Y, Ge X, Frank CL, Madison JM, Koehler AN, Doud MK, Tassa C, Berry EM, Soda T, Singh KK et al. 2009. Disrupted in schizophrenia 1 regulates neuronal progenitor proliferation via modulation of GSK3beta/beta-catenin signaling. *Cell* **136**(6): 1017-1031.
- Marcelis M, Takei N, van Os J. 1999. Urbanization and risk for schizophrenia: does the effect operate before or around the time of illness onset? *Psychol Med* **29**(5): 1197-1203.
- March D, Hatch SL, Morgan C, Kirkbride JB, Bresnahan M, Fearon P, Susser E. 2008. Psychosis and place. *Epidemiol Rev* **30**: 84-100.
- Marley A, von Zastrow M. 2010. DISC1 regulates primary cilia that display specific dopamine receptors. *PLoS One* **5**(5): e10902.

Bibliografía

- Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, Shago M, Moessner R, Pinto D, Ren Y et al. 2008. Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* **82**(2): 477-488.
- Maynard TM, Haskell GT, Peters AZ, Sikich L, Lieberman JA, LaMantia AS. 2003. A comprehensive analysis of 22q11 gene expression in the developing and adult brain. *Proc Natl Acad Sci U S A* **100**(24): 14433-14438.
- McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* **9**(5): 356-369.
- McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S, Perkins DO, Dickel DE, Kusenda M, Krastoshevsky O et al. 2009. Microduplications of 16p11.2 are associated with schizophrenia. *Nat Genet* **41**(11): 1223-1227.
- McClellan JM, Susser E, King MC. 2007. Schizophrenia: a common disease caused by multiple rare alleles. *Br J Psychiatry* **190**: 194-199.
- McGrath J, Saha S, Chant D, Welham J. 2008. Schizophrenia: a concise overview of incidence, prevalence, and mortality. *Epidemiol Rev* **30**: 67-76.
- McGrath J, Saha S, Welham J, El Saadi O, MacCauley C, Chant D. 2004. A systematic review of the incidence of schizophrenia: the distribution of rates and the influence of sex, urbanicity, migrant status and methodology. *BMC Med* **2**: 13.
- McVean GA, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**(7422): 56-65.
- Mead CL, Kuzyk MA, Moradian A, Wilson GM, Holt RA, Morin GB. 2010. Cytosolic protein interactions of the schizophrenia susceptibility gene dysbindin. *J Neurochem* **113**(6): 1491-1503.
- Melhem N, Middleton F, McFadden K, Klei L, Faraone SV, Vinogradov S, Tiobech J, Yano V, Kuartei S, Roeder K et al. 2011. Copy number variants for schizophrenia and related psychotic disorders in Oceanic Palau: risk and transmission in extended pedigrees. *Biol Psychiatry* **70**(12): 1115-1121.
- Millar JK, Christie S, Porteous DJ. 2003. Yeast two-hybrid screens implicate DISC1 in brain development and function. *Biochem Biophys Res Commun* **311**(4): 1019-1025.
- Millar JK, James R, Christie S, Porteous DJ. 2005a. Disrupted in schizophrenia 1 (DISC1): subcellular targeting and induction of ring mitochondria. *Mol Cell Neurosci* **30**(4): 477-484.
- Millar JK, Pickard BS, Mackie S, James R, Christie S, Buchanan SR, Malloy MP, Chubb JE, Huston E, Baillie GS et al. 2005b. DISC1 and PDE4B are interacting genetic factors in schizophrenia that regulate cAMP signaling. *Science* **310**(5751): 1187-1191.
- Millar JK, Wilson-Annan JC, Anderson S, Christie S, Taylor MS, Semple CA, Devon RS, St Clair DM, Muir WJ, Blackwood DH et al. 2000. Disruption of two novel genes by a translocation co-segregating with schizophrenia. *Hum Mol Genet* **9**(9): 1415-1423.
- Mitreá C, Taghavi Z, Bokanizad B, Hanoudi S, Tagett R, Donato M, Voichita C, Draghici S. 2013. Methods and approaches in the topology-based analysis of biological pathways. *Frontiers in physiology* **4**: 278.
- Miyoshi K, Honda A, Baba K, Taniguchi M, Oono K, Fujita T, Kuroda S, Katayama T, Tohyama M. 2003. Disrupted-In-Schizophrenia 1, a candidate gene for schizophrenia, participates in neurite outgrowth. *Mol Psychiatry* **8**(7): 685-694.

- Moens LN, De Rijk P, Reumers J, Van den Bossche MJ, Glasse W, De Zutter S, Lenaerts AS, Nordin A, Nilsson LG, Medina Castillo I et al. 2011. Sequencing of DISC1 pathway genes reveals increased burden of rare missense variants in schizophrenia patients from a northern Swedish population. *PLoS One* **6**(8): e23450.
- Mooney MA, Nigg JT, McWeeney SK, Wilmot B. 2014. Functional and genomic context in pathway analysis of GWAS data. *Trends Genet* **30**(9): 390-400.
- Moreno-De-Luca D, Consortium S, Mulle JG, Simons Simplex Collection Genetics C, Kaminsky EB, Sanders SJ, GeneStar, Myers SM, Adam MP, Pakula AT et al. 2010. Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia. *Am J Hum Genet* **87**(5): 618-630.
- Moreno-De-Luca D, Sanders SJ, Willsey AJ, Mulle JG, Lowe JK, Geschwind DH, State MW, Martin CL, Ledbetter DH. 2013. Using large clinical data sets to infer pathogenicity for rare copy number variants in autism cohorts. *Mol Psychiatry* **18**(10): 1090-1095.
- Morgenthaler S, Thilly WG. 2007. A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). *Mutat Res* **615**(1-2): 28-56.
- Morris JA, Kandpal G, Ma L, Austin CP. 2003. DISC1 (Disrupted-In-Schizophrenia 1) is a centrosome-associated protein that interacts with MAP1A, MIPT3, ATF4/5 and NUDEL: regulation and loss of interaction with mutation. *Hum Mol Genet* **12**(13): 1591-1608.
- Muhleisen TW, Basmanav FB, Forstner AJ, Mattheisen M, Priebe L, Herms S, Breuer R, Moebus S, Nenadic I, Sauer H et al. 2011. Resequencing and follow-up of neurexin 1 (NRXN1) in schizophrenia patients. *Schizophr Res* **127**(1-3): 35-40.
- Muhleisen TW, Leber M, Schulze TG, Strohmaier J, Degenhardt F, Treutlein J, Mattheisen M, Forstner AJ, Schumacher J, Breuer R et al. 2014. Genome-wide association study reveals two new risk loci for bipolar disorder. *Nature communications* **5**: 3339.
- Mulle JG. 2015. The 3q29 deletion confers >40-fold increase in risk for schizophrenia. *Mol Psychiatry* **20**(9): 1028-1029.
- Mulle JG, Dodd AF, McGrath JA, Wolyniec PS, Mitchell AA, Shetty AC, Sobreira NL, Valle D, Rudd MK, Satten G et al. 2010. Microdeletions of 3q29 confer high risk for schizophrenia. *Am J Hum Genet* **87**(2): 229-236.
- Mulle JG, Pulver AE, McGrath JA, Wolyniec PS, Dodd AF, Cutler DJ, Sebat J, Malhotra D, Nestadt G, Conrad DF et al. 2014. Reciprocal duplication of the Williams-Beuren syndrome deletion on chromosome 7q11.23 is associated with schizophrenia. *Biol Psychiatry* **75**(5): 371-377.
- Murphy KC, Jones LA, Owen MJ. 1999. High rates of schizophrenia in adults with velo-cardio-facial syndrome. *Arch Gen Psychiatry* **56**(10): 940-945.
- Nakata K, Lipska BK, Hyde TM, Ye T, Newburn EN, Morita Y, Vakkalanka R, Barenboim M, Sei Y, Weinberger DR et al. 2009. DISC1 splice variants are upregulated in schizophrenia and associated with risk polymorphisms. *Proc Natl Acad Sci U S A* **106**(37): 15873-15878.
- Namba T, Ming GL, Song H, Waga C, Enomoto A, Kaibuchi K, Kohsaka S, Uchino S. 2011. NMDA receptor regulates migration of newly generated neurons in the adult hippocampus via Disrupted-In-Schizophrenia 1 (DISC1). *J Neurochem* **118**(1): 34-44.
- Nazarian R, Starcevic M, Spencer MJ, Dell'Angelica EC. 2006. Reinvestigation of the dysbindin subunit of BLOC-1 (biogenesis of lysosome-related organelles complex-1) as a dystrobrevin-binding protein. *Biochem J* **395**(3): 587-598.

Bibliografía

- Neale BM, Rivas MA, Voight BF, Altshuler D, Devlin B, Orho-Melander M, Kathiresan S, Purcell SM, Roeder K, Daly MJ. 2011. Testing for an unusual distribution of rare variants. *PLoS Genet* **7**(3): e1001322.
- Need AC, McEvoy JP, Gennarelli M, Heinzen EL, Ge D, Maia JM, Shianna KV, He M, Cirulli ET, Gumbs CE et al. 2012. Exome sequencing followed by large-scale genotyping suggests a limited role for moderately rare risk factors of strong effect in schizophrenia. *Am J Hum Genet* **91**(2): 303-312.
- Network, Pathway Analysis Subgroup of Psychiatric Genomics C. 2015. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat Neurosci* **18**(2): 199-209.
- Ng MY, Levinson DF, Faraone SV, Suarez BK, DeLisi LE, Arinami T, Riley B, Paunio T, Pulver AE, Irmansyah et al. 2009. Meta-analysis of 32 genome-wide linkage studies of schizophrenia. *Mol Psychiatry* **14**(8): 774-785.
- Ng PC, Henikoff S. 2003. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* **31**(13): 3812-3814.
- Nicodemus KK, Callicott JH, Higier RG, Luna A, Nixon DC, Lipska BK, Vakkalanka R, Giegling I, Rujescu D, St Clair D et al. 2010. Evidence of statistical epistasis between DISC1, CIT and NDEL1 impacting risk for schizophrenia: biological validation with functional neuroimaging. *Hum Genet* **127**(4): 441-452.
- Nicholas KB, Jr. NHB, Deerfield DW. 1997. GeneDoc: Analysis and Visualization of Genetic Variation. *EMBNEWNEWS*.
- Norkett R, Modi S, Birsa N, Atkin TA, Ivankovic D, Pathania M, Trossbach SV, Korth C, Hirst WD, Kittler JT. 2016. DISC1-dependent Regulation of Mitochondrial Dynamics Controls the Morphogenesis of Complex Neuronal Dendrites. *J Biol Chem* **291**(2): 613-629.
- O'Donovan MC, Craddock N, Norton N, Williams H, Peirce T, Moskvina V, Nikolov I, Hamshere M, Carroll L, Georgieva L et al. 2008. Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet* **40**(9): 1053-1055.
- O'Donovan MC, Owen MJ. 2016. The implications of the shared genetics of psychiatric disorders. *Nat Med* **22**(11): 1214-1219.
- O'Dushlaine C, Kenny E, Heron E, Donohoe G, Gill M, Morris D, International Schizophrenia C, Corvin A. 2011. Molecular pathways involved in neuronal cell adhesion and membrane scaffolding contribute to schizophrenia and bipolar disorder susceptibility. *Mol Psychiatry* **16**(3): 286-292.
- O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, Karakoc E, Mackenzie AP, Ng SB, Baker C et al. 2011. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat Genet* **43**(6): 585-589.
- Olfson M, Gerhard T, Huang C, Crystal S, Stroup TS. 2015. Premature Mortality Among Adults With Schizophrenia in the United States. *JAMA psychiatry* **72**(12): 1172-1181.
- Ollila HM, Soronen P, Silander K, Palo OM, Kieseppa T, Kaunisto MA, Lonnqvist J, Peltonen L, Partonen T, Paunio T. 2009. Findings from bipolar disorder genome-wide association studies replicate in a Finnish bipolar family-cohort. *Mol Psychiatry* **14**(4): 351-353.
- Owen MJ, Craddock N, O'Donovan MC. 2005. Schizophrenia: genes at last? *Trends Genet* **21**(9): 518-525.
- Owen MJ, Williams NM, O'Donovan MC. 2004. The molecular genetics of schizophrenia: new findings promise new insights. *Mol Psychiatry* **9**(1): 14-27.

- Ozeki Y, Tomoda T, Kleiderlein J, Kamiya A, Bord L, Fujii K, Okawa M, Yamada N, Hatten ME, Snyder SH et al. 2003. Disrupted-in-Schizophrenia-1 (DISC-1): mutant truncation prevents binding to NudE-like (NUDEL) and inhibits neurite outgrowth. *Proc Natl Acad Sci U S A* **100**(1): 289-294.
- Park N, Juo SH, Cheng R, Liu J, Loth JE, Lilliston B, Nee J, Grunn A, Kanyas K, Lerer B et al. 2004. Linkage analysis of psychosis in bipolar pedigrees suggests novel putative loci for bipolar disorder and shared susceptibility with schizophrenia. *Mol Psychiatry* **9**(12): 1091-1099.
- Park YU, Jeong J, Lee H, Mun JY, Kim JH, Lee JS, Nguyen MD, Han SS, Suh PG, Park SK. 2010. Disrupted-in-schizophrenia 1 (DISC1) plays essential roles in mitochondria in collaboration with Mitofilin. *Proc Natl Acad Sci U S A* **107**(41): 17785-17790.
- Paspalas CD, Wang M, Arnsten AF. 2013. Constellation of HCN channels and cAMP regulating proteins in dendritic spines of the primate prefrontal cortex: potential substrate for working memory deficits in schizophrenia. *Cereb Cortex* **23**(7): 1643-1654.
- Penner JD, Brown AS. 2007. Prenatal infectious and nutritional factors and risk of adult schizophrenia. *Expert review of neurotherapeutics* **7**(7): 797-805.
- Perronnet C, Vaillend C. 2010. Dystrophins, utrophins, and associated scaffolding complexes: role in mammalian brain and implications for therapeutic strategies. *J Biomed Biotechnol* **2010**: 849426.
- Phizicky EM, Fields S. 1995. Protein-protein interactions: methods for detection and analysis. *Microbiological reviews* **59**(1): 94-123.
- Piggins HD. 2011. Schizophrenia: Zooming in on a gene. *Nature* **471**(7339): 455-456.
- Piggott LA, Bauman AL, Scott JD, Dessauer CW. 2008. The A-kinase anchoring protein Yotiao binds and regulates adenylyl cyclase in brain. *Proc Natl Acad Sci U S A* **105**(37): 13835-13840.
- Piton A, Gauthier J, Hamdan FF, Lafreniere RG, Yang Y, Henrion E, Laurent S, Noreau A, Thibodeau P, Karemera L et al. 2011. Systematic resequencing of X-chromosome synaptic genes in autism spectrum disorder and schizophrenia. *Mol Psychiatry* **16**(8): 867-880.
- Pitt D, Hopkins I. 1978. A syndrome of mental retardation, wide mouth and intermittent overbreathing. *Australian paediatric journal* **14**(3): 182-184.
- Poliak S, Gollan L, Martinez R, Custer A, Einheber S, Salzer JL, Trimmer JS, Shrager P, Peles E. 1999. Caspr2, a new member of the neuexin superfamily, is localized at the juxtaparanodes of myelinated axons and associates with K⁺ channels. *Neuron* **24**(4): 1037-1047.
- Price AL, Kryukov GV, de Bakker PI, Purcell SM, Staples J, Wei LJ, Sunyaev SR. 2010. Pooled association tests for rare variants in exon-resequencing studies. *Am J Hum Genet* **86**(6): 832-838.
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**(8): 904-909.
- Pritchard JK. 2001. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* **69**(1): 124-137.
- Pritchard JK, Donnelly P. 2001. Case-control studies of association in structured or admixed populations. *Theor Popul Biol* **60**(3): 227-237.

Bibliografía

- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**(2): 945-959.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**(3): 559-575.
- Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P, O'Dushlaine C, Chambert K, Bergen SE, Kahler A et al. 2014. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**(7487): 185-190.
- Quattrocchi CC, Wannenes F, Persico AM, Ciafre SA, D'Arcangelo G, Farace MG, Keller F. 2002. Reelin is a serine protease of the extracellular matrix. *J Biol Chem* **277**(1): 303-309.
- Ramanan VK, Shen L, Moore JH, Saykin AJ. 2012. Pathway analysis of genomic data: concepts, methods, and prospects for future development. *Trends Genet* **28**(7): 323-332.
- Ramsey AJ, Milenkovic M, Oliveira AF, Escobedo-Lozoya Y, Seshadri S, Salahpour A, Sawa A, Yasuda R, Caron MG. 2011. Impaired NMDA receptor transmission alters striatal synapses and DISC1 protein in an age-dependent manner. *Proc Natl Acad Sci U S A* **108**(14): 5795-5800.
- Rees E, Kirov G, Sanders A, Walters JT, Chambert KD, Shi J, Szatkiewicz J, O'Dushlaine C, Richards AL, Green EK et al. 2014a. Evidence that duplications of 22q11.2 protect against schizophrenia. *Mol Psychiatry* **19**(1): 37-40.
- Rees E, Moskvina V, Owen MJ, O'Donovan MC, Kirov G. 2011. De novo rates and selection of schizophrenia-associated copy number variants. *Biol Psychiatry* **70**(12): 1109-1114.
- Rees E, Walters JT, Chambert KD, O'Dushlaine C, Szatkiewicz J, Richards AL, Georgieva L, Mahoney-Davies G, Legge SE, Moran JL et al. 2014b. CNV analysis in a large schizophrenia sample implicates deletions at 16p12.1 and SLC1A1 and duplications at 1p36.33 and CGNL1. *Hum Mol Genet* **23**(6): 1669-1676.
- Rees E, Walters JT, Georgieva L, Isles AR, Chambert KD, Richards AL, Mahoney-Davies G, Legge SE, Moran JL, McCarroll SA et al. 2014c. Analysis of copy number variations at 15 schizophrenia-associated loci. *Br J Psychiatry* **204**(2): 108-114.
- Reva B, Antipin Y, Sander C. 2011. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res* **39**(17): e118.
- Richards AL, Jones L, Moskvina V, Kirov G, Gejman PV, Levinson DF, Sanders AR, Molecular Genetics of Schizophrenia C, International Schizophrenia C, Purcell S et al. 2012. Schizophrenia susceptibility alleles are enriched for alleles that affect gene expression in adult human brain. *Mol Psychiatry* **17**(2): 193-201.
- Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E et al. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**(5): 405-424.
- Riley B, Kendler KS. 2006. Molecular genetic studies of schizophrenia. *Eur J Hum Genet* **14**(6): 669-680.
- Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kahler AK, Akterin S, Bergen SE, Collins AL, Crowley JJ, Fromer M et al. 2013. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet* **45**(10): 1150-1159.
- Ripke S, Sanders AR, Kendler KS, Levinson DF, Sklar P, Holmans PA, Lin DY, Duan J, Ophoff RA, Andreassen OA et al. 2011. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet* **43**(10): 969-976.

- Risch N, Merikangas K. 1996. The future of genetic studies of complex human diseases. *Science* **273**(5281): 1516-1517.
- Rivero S, Cardenas J, Bornens M, Rios RM. 2009. Microtubule nucleation at the cis-side of the Golgi apparatus requires AKAP450 and GM130. *EMBO J* **28**(8): 1016-1028.
- Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* **518**(7539): 317-330.
- Roussos P, Mitchell AC, Voloudakis G, Fullard JF, Pothula VM, Tsang J, Stahl EA, Georgakopoulos A, Ruderfer DM, Charney A et al. 2014. A role for noncoding variation in schizophrenia. *Cell reports* **9**(4): 1417-1429.
- Rubenstein JL, Merzenich MM. 2003. Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes Brain Behav* **2**(5): 255-267.
- Rujescu D, Ingason A, Cichon S, Pietilainen OP, Barnes MR, Toulopoulou T, Picchioni M, Vassos E, Ettinger U, Bramon E et al. 2009. Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum Mol Genet* **18**(5): 988-996.
- Ryan AK, Goodship JA, Wilson DI, Philip N, Levy A, Seidel H, Schuffenhauer S, Oechsler H, Belohradsky B, Prieur M et al. 1997. Spectrum of clinical features associated with interstitial chromosome 22q11 deletions: a European collaborative study. *J Med Genet* **34**(10): 798-804.
- Saetre P, Agartz I, De Franciscis A, Lundmark P, Djurovic S, Kahler A, Andreassen OA, Jakobsen KD, Rasmussen HB, Werge T et al. 2008. Association between a disrupted-in-schizophrenia 1 (DISC1) single nucleotide polymorphism and schizophrenia in a combined Scandinavian case-control sample. *Schizophr Res* **106**(2-3): 237-241.
- Saha S, Chant D, McGrath J. 2007. A systematic review of mortality in schizophrenia: is the differential mortality gap worsening over time? *Arch Gen Psychiatry* **64**(10): 1123-1131.
- Sakae N, Yamasaki N, Kitaichi K, Fukuda T, Yamada M, Yoshikawa H, Hiranita T, Tatsumi Y, Kira J, Yamamoto T et al. 2008. Mice lacking the schizophrenia-associated protein FEZ1 manifest hyperactivity and enhanced responsiveness to psychostimulants. *Hum Mol Genet* **17**(20): 3191-3203.
- Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, Chu SH, Moreau MP, Gupta AR, Thomson SA et al. 2011. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* **70**(5): 863-885.
- Sanderson JL, Dell'Acqua ML. 2011. AKAP signaling complexes in regulation of excitatory synaptic plasticity. *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry* **17**(3): 321-336.
- Sasaki S, Shionoya A, Ishida M, Gambello MJ, Yingling J, Wynshaw-Boris A, Hirotsune S. 2000. A LIS1/NUDEL/cytoplasmic dynein heavy chain complex in the developing and adult nervous system. *Neuron* **28**(3): 681-696.
- Saunders AM, Strittmatter WJ, Schmechel D, George-Hyslop PH, Pericak-Vance MA, Joo SH, Rosi BL, Gusella JF, Crapper-MacLachlan DR, Alberts MJ et al. 1993. Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology* **43**(8): 1467-1472.
- Sawamura N, Sawamura-Yamamoto T, Ozeki Y, Ross CA, Sawa A. 2005. A form of DISC1 enriched in nucleus: altered subcellular distribution in orbitofrontal cortex in psychosis and substance/alcohol abuse. *Proc Natl Acad Sci U S A* **102**(4): 1187-1192.

Bibliografía

- Schenkel LS, Silverstein SM. 2004. Dimensions of premorbid functioning in schizophrenia: a review of neuromotor, cognitive, social, and behavioral domains. *Genetic, social, and general psychology monographs* **130**(3): 241-270.
- Schizophrenia Psychiatric Genome-Wide Association Study C. 2011. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet* **43**(10): 969-976.
- Schizophrenia Working Group of the Psychiatric Genomics C. 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**(7510): 421-427.
- Schreiber M, Dorschner M, Tsuang D. 2013. Next-generation sequencing in schizophrenia and other neuropsychiatric disorders. *Am J Med Genet B Neuropsychiatr Genet* **162B**(7): 671-678.
- Schumacher J, Laje G, Abou Jamra R, Becker T, Muhleisen TW, Vasilescu C, Mattheisen M, Herms S, Hoffmann P, Hillmer AM et al. 2009. The DISC locus and schizophrenia: evidence from an association study in a central European sample and from a meta-analysis across different European populations. *Hum Mol Genet* **18**(14): 2719-2727.
- Schuurs-Hoeijmakers JH, Vulto-van Silfhout AT, Vissers LE, van de V, II, van Bon BW, de Ligt J, Gilissen C, Hehir-Kwa JY, Neveling K, del Rosario M et al. 2013. Identification of pathogenic gene variants in small families with intellectually disabled siblings by exome sequencing. *J Med Genet* **50**(12): 802-811.
- Schwarz JM, Cooper DN, Schuelke M, Seelow D. 2014. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* **11**(4): 361-362.
- Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M et al. 2004. Large-scale copy number polymorphism in the human genome. *Science* **305**(5683): 525-528.
- Sedeno-Cortes AE, Pavlidis P. 2014. Pitfalls in the application of gene-set analysis to genetics studies. *Trends Genet* **30**(12): 513-514.
- Sekar A, Bialas AR, de Rivera H, Davis A, Hammond TR, Kamitaki N, Tooley K, Presumey J, Baum M, Van Doren V et al. 2016. Schizophrenia risk from complex variation of complement component 4. *Nature* **530**(7589): 177-183.
- Shaikh TH, Kurahashi H, Saitta SC, O'Hare AM, Hu P, Roe BA, Driscoll DA, McDonald-McGinn DM, Zackai EH, Budarf ML et al. 2000. Chromosome 22-specific low copy repeats and the 22q11.2 deletion syndrome: genomic organization and deletion endpoint analysis. *Hum Mol Genet* **9**(4): 489-501.
- Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Segraves R et al. 2005. Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet* **77**(1): 78-88.
- Sharp SI, Lange J, Kandaswamy R, Daher M, Anjorin A, Bass NJ, McQuillin A. 2017. Identification of rare nonsynonymous variants in SYNE1/CPG2 in bipolar affective disorder. *Psychiatr Genet*.
- Shen YC, Liao DL, Chen JY, Wang YC, Lai IC, Liou YJ, Chen YJ, Luu SU, Chen CH. 2009. Resequencing and association study of vesicular glutamate transporter 1 gene (VGLUT1) with schizophrenia. *Schizophr Res* **115**(2-3): 254-260.
- Shen YC, Liao DL, Lu CL, Chen JY, Liou YJ, Chen TT, Chen CH. 2010. Resequencing of the vesicular glutamate transporter 2 gene (VGLUT2) reveals some rare genetic variants that may increase the genetic burden in schizophrenia. *Schizophr Res* **121**(1-3): 179-186.
- Shen YC, Tsai HM, Cheng MC, Hsu SH, Chen SF, Chen CH. 2012. Genetic and functional analysis of the gene encoding neurogranin in schizophrenia. *Schizophr Res* **137**(1-3): 7-13.

- Sheward WJ, Lutz EM, Harmar AJ. 1995. The distribution of vasoactive intestinal peptide2 receptor messenger RNA in the rat brain and pituitary gland as assessed by in situ hybridization. *Neuroscience* **67**(2): 409-418.
- Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Pe'er I, Dudbridge F, Holmans PA, Whitemore AS, Mowry BJ et al. 2009. Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* **460**(7256): 753-757.
- Shi Y, Li Z, Xu Q, Wang T, Li T, Shen J, Zhang F, Chen J, Zhou G, Ji W et al. 2011. Common variants on 8p12 and 1q24.2 confer risk of schizophrenia. *Nat Genet* **43**(12): 1224-1227.
- Shifman S, Johannesson M, Bronstein M, Chen SX, Collier DA, Craddock NJ, Kendler KS, Li T, O'Donovan M, O'Neill FA et al. 2008. Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women. *PLoS Genet* **4**(2): e28.
- Shimizu S, Matsuzaki S, Hattori T, Kumamoto N, Miyoshi K, Katayama T, Tohyama M. 2008. DISC1-kendrin interaction is involved in centrosomal microtubule network formation. *Biochem Biophys Res Commun* **377**(4): 1051-1056.
- Shinoda T, Taya S, Tsuboi D, Hikita T, Matsuzawa R, Kuroda S, Iwamatsu A, Kaibuchi K. 2007. DISC1 regulates neurotrophin-induced axon elongation via interaction with Grb2. *J Neurosci* **27**(1): 4-14.
- Shprintzen RJ, Goldberg RB, Lewin ML, Sidoti EJ, Berkman MD, Argamaso RV, Young D. 1978. A new syndrome involving cleft palate, cardiac anomalies, typical facies, and learning disabilities: velo-cardio-facial syndrome. *The Cleft palate journal* **15**(1): 56-62.
- Singh KK, Ge X, Mao Y, Drane L, Meletis K, Samuels BA, Tsai LH. 2010. Dixdc1 is a critical regulator of DISC1 and embryonic cortical development. *Neuron* **67**(1): 33-48.
- Singh T, Kurki MI, Curtis D, Purcell SM, Crooks L, McRae J, Suvisaari J, Chheda H, Blackwood D, Breen G et al. 2016. Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nat Neurosci* **19**(4): 571-577.
- Sklar P, Ripke S, Scott LJ, Andreassen OA, Cichon S, Craddock N, Edenberg HJ, Nurnberger JI, Jr, Rietschel M, Blackwood D et al. 2011. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet* **43**(10): 977-983.
- Song W, Li W, Feng J, Heston LL, Scaringe WA, Sommer SS. 2008. Identification of high risk DISC1 structural variants with a 2% attributable risk for schizophrenia. *Biochem Biophys Res Commun* **367**(3): 700-706.
- Song W, Li W, Noltner K, Yan J, Green E, Grozeva D, Jones IR, Craddock N, Longmate J, Feng J et al. 2010. Identification of high risk DISC1 protein structural variants in patients with bipolar spectrum disorder. *Neurosci Lett*.
- Sonnenberg A, Liem RK. 2007. Plakins in development and disease. *Exp Cell Res* **313**(10): 2189-2203.
- St Clair D, Blackwood D, Muir W, Carothers A, Walker M, Spowart G, Gosden C, Evans HJ. 1990. Association within a family of a balanced autosomal translocation with major mental illness. *Lancet* **336**(8706): 13-16.
- Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D, Werge T, Pietilainen OP, Mors O, Mortensen PB et al. 2009. Common variants conferring risk of schizophrenia. *Nature* **460**(7256): 744-747.

Bibliografía

- Stefansson H, Rujescu D, Cichon S, Pietilainen OP, Ingason A, Steinberg S, Fossdal R, Sigurdsson E, Sigmundsson T, Buizer-Voskamp JE et al. 2008. Large recurrent microdeletions associated with schizophrenia. *Nature* **455**(7210): 232-236.
- Stefansson H, Sigurdsson E, Steinthorsdottir V, Bjornsdottir S, Sigmundsson T, Ghosh S, Brynjolfsson J, Gunnarsdottir S, Ivarsson O, Chou TT et al. 2002. Neuregulin 1 and susceptibility to schizophrenia. *Am J Hum Genet* **71**(4): 877-892.
- Steinberg S, de Jong S, Andreassen OA, Werge T, Borglum AD, Mors O, Mortensen PB, Gustafsson O, Costas J, Pietilainen OP et al. 2011a. Common variants at VRK2 and TCF4 conferring risk of schizophrenia. *Hum Mol Genet* **20**(20): 4076-4081.
- Steinberg S de Jong S Mattheisen M Costas J Demontis D Jamain S Pietilainen OP Lin K Papiol S Huttenlocher J et al. 2014. Common variant at 16p11.2 conferring risk of psychosis. *Mol Psychiatry* **19**(1): 108-114.
- Steinberg S, Mors O, Borglum AD, Gustafsson O, Werge T, Mortensen PB, Andreassen OA, Sigurdsson E, Thorgeirsson TE, Bottcher Y et al. 2011b. Expanding the range of ZNF804A variants conferring risk of psychosis. *Mol Psychiatry* **16**(1): 59-66.
- Steinecke A, Gampe C, Valkova C, Kaether C, Bolz J. 2012. Disrupted-in-Schizophrenia 1 (DISC1) is necessary for the correct migration of cortical interneurons. *J Neurosci* **32**(2): 738-745.
- Straub RE, Jiang Y, MacLean CJ, Ma Y, Webb BT, Myakishev MV, Harris-Kerr C, Wormley B, Sadek H, Kadambi B et al. 2002. Genetic variation in the 6p22.3 gene DTNBP1, the human ortholog of the mouse dysbindin gene, is associated with schizophrenia. *Am J Hum Genet* **71**(2): 337-348.
- Straub RE, MacLean CJ, O'Neill FA, Burke J, Murphy B, Duke F, Shinkwin R, Webb BT, Zhang J, Walsh D et al. 1995. A potential vulnerability locus for schizophrenia on chromosome 6p24-22: evidence for genetic heterogeneity. *Nat Genet* **11**(3): 287-293.
- Stuppia L, Antonucci I, Palka G, Gatta V. 2012. Use of the MLPA assay in the molecular diagnosis of gene copy number alterations in human genetic diseases. *Int J Mol Sci* **13**(3): 3245-3276.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES et al. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**(43): 15545-15550.
- Sudhof TC. 2008. Neuroligins and neurexins link synaptic function to cognitive disease. *Nature* **455**(7215): 903-911.
- Sullivan PF. 2012. Puzzling over schizophrenia: Schizophrenia as a pathway disease. *Nat Med* **18**(2): 210-211.
- Sullivan PF, Kendler KS, Neale MC. 2003. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry* **60**(12): 1187-1192.
- Sun C, Cheng MC, Qin R, Liao DL, Chen TT, Koong FJ, Chen G, Chen CH. 2011. Identification and functional characterization of rare mutations of the neuroligin-2 gene (NLGN2) associated with schizophrenia. *Hum Mol Genet* **20**(15): 3042-3051.
- Sunyaev S, Ramensky V, Koch I, Lathe W, 3rd, Kondrashov AS, Bork P. 2001. Prediction of deleterious human alleles. *Hum Mol Genet* **10**(6): 591-597.
- Szatkiewicz JP, O'Dushlaine C, Chen G, Chambert K, Moran JL, Neale BM, Fromer M, Ruderfer D, Akterin S, Bergen SE et al. 2014. Copy number variation in schizophrenia in Sweden. *Mol Psychiatry* **19**(7): 762-773.

- Takata A, Ionita-Laza I, Gogos JA, Xu B, Karayiorgou M. 2016. De Novo Synonymous Mutations in Regulatory Elements Contribute to the Genetic Etiology of Autism and Schizophrenia. *Neuron* **89**(5): 940-947.
- Takata A, Xu B, Ionita-Laza I, Roos JL, Gogos JA, Karayiorgou M. 2014. Loss-of-function variants in schizophrenia risk and SETD1A as a candidate susceptibility gene. *Neuron* **82**(4): 773-780.
- Tam GW, Redon R, Carter NP, Grant SG. 2009. The role of DNA copy number variation in schizophrenia. *Biol Psychiatry* **66**(11): 1005-1012.
- Tandon R, Keshavan MS, Nasrallah HA. 2008. Schizophrenia, "just the facts" what we know in 2008. 2. Epidemiology and etiology. *Schizophr Res* **102**(1-3): 1-18.
- Tandon R, Nasrallah HA, Keshavan MS. 2009. Schizophrenia, "just the facts" 4. Clinical features and conceptualization. *Schizophr Res* **110**(1-3): 1-23.
- Tassabehji M. 2003. Williams-Beuren syndrome: a challenge for genotype-phenotype correlations. *Hum Mol Genet* **12 Spec No 2**: R229-237.
- Taya S, Shinoda T, Tsuboi D, Asaki J, Nagai K, Hikita T, Kuroda S, Kuroda K, Shimizu M, Hirotsune S et al. 2007. DISC1 regulates the transport of the NUDEL/LIS1/14-3-3epsilon complex through kinesin-1. *J Neurosci* **27**(1): 15-26.
- Taylor MS, Devon RS, Millar JK, Porteous DJ. 2003. Evolutionary constraints on the Disrupted in Schizophrenia locus. *Genomics* **81**(1): 67-77.
- Tezenas Du Montcel S, Mendizabai H, Ayme S, Levy A, Philip N. 1996. Prevalence of 22q11 microdeletion. *J Med Genet* **33**(8): 719.
- Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A. 2003. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res* **13**(9): 2129-2141.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**(24): 4876-4882.
- Thomson PA, Parla JS, McRae AF, Kramer M, Ramakrishnan K, Yao J, Soares DC, McCarthy S, Morris SW, Cardone L et al. 2014. 708 Common and 2010 rare DISC1 locus variants identified in 1542 subjects: analysis for association with psychiatric disorder and cognitive traits. *Mol Psychiatry* **19**(6): 668-675.
- Timms AE, Dorschner MO, Wechsler J, Choi KY, Kirkwood R, Girirajan S, Baker C, Eichler EE, Korvatska O, Roche KW et al. 2013. Support for the N-methyl-D-aspartate receptor hypofunction hypothesis of schizophrenia from exome sequencing in multiplex families. *JAMA psychiatry* **70**(6): 582-590.
- Ullmann R, Turner G, Kirchoff M, Chen W, Tonge B, Rosenberg C, Field M, Vianna-Morgante AM, Christie L, Krepischi-Santos AC et al. 2007. Array CGH identifies reciprocal 16p13.1 duplications and deletions that predispose to autism and/or mental retardation. *Hum Mutat* **28**(7): 674-682.
- Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* **35**(Web Server issue): W71-74.
- Vabnick I, Trimmer JS, Schwarz TL, Levinson SR, Risal D, Shrager P. 1999. Dynamic potassium channel distributions during axonal development prevent aberrant firing patterns. *J Neurosci* **19**(2): 747-758.
- Vacic V, McCarthy S, Malhotra D, Murray F, Chou HH, Peoples A, Makarov V, Yoon S, Bhandari A, Corominas R et al. 2011. Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia. *Nature* **471**(7339): 499-503.

Bibliografía

- Van der Aa N, Rooms L, Vandeweyer G, van den Ende J, Reyniers E, Fichera M, Romano C, Delle Chiaie B, Mortier G, Menten B et al. 2009. Fourteen new cases contribute to the characterization of the 7q11.23 microduplication syndrome. *Eur J Med Genet* **52**(2-3): 94-100.
- van Dongen J, Boomsma DI. 2013. The evolutionary paradox and the missing heritability of schizophrenia. *Am J Med Genet B Neuropsychiatr Genet* **162B**(2): 122-136.
- van Hulzen KJ, Scholz CJ, Franke B, Ripke S, Klein M, McQuillin A, Sonuga-Barke EJ, Group PAW, Kelsoe JR, Landen M et al. 2016. Genetic Overlap Between Attention-Deficit/Hyperactivity Disorder and Bipolar Disorder: Evidence From Genome-wide Association Study Meta-analysis. *Biol Psychiatry*.
- Visser LE, de Ligt J, Gilissen C, Janssen I, Stehouwer M, de Vries P, van Lier B, Arts P, Wieskamp N, del Rosario M et al. 2010. A de novo paradigm for mental retardation. *Nat Genet* **42**(12): 1109-1112.
- Vrijenhoek T, Buizer-Voskamp JE, van der Stelt I, Strengman E, Genetic R, Outcome in Psychosis C, Sabatti C, Geurts van Kessel A, Brunner HG, Ophoff RA et al. 2008. Recurrent CNVs disrupt three candidate genes in schizophrenia patients. *Am J Hum Genet* **83**(4): 504-510.
- Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, Nord AS, Kusenda M, Malhotra D, Bhandari A et al. 2008. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* **320**(5875): 539-543.
- Wang K, Li M, Bucan M. 2007. Pathway-based approaches for analysis of genomewide association studies. *Am J Hum Genet* **81**(6): 1278-1283.
- Wang K, Li M, Hakonarson H. 2010a. Analysing biological pathways in genome-wide association studies. *Nat Rev Genet* **11**(12): 843-854.
- . 2010b. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**(16): e164.
- Wang L, Jia P, Wolfinger RD, Chen X, Zhao Z. 2011a. Gene set analysis of genome-wide association studies: methodological issues and perspectives. *Genomics* **98**(1): 1-8.
- Wang Q, Charych EI, Pulito VL, Lee JB, Graziane NM, Crozier RA, Revilla-Sanchez R, Kelly MP, Dunlop AJ, Murdoch H et al. 2011b. The psychiatric disease risk factors DISC1 and TNIK interact to regulate synapse composition and function. *Mol Psychiatry* **16**(10): 1006-1023.
- Warr A, Robert C, Hume D, Archibald A, Deeb N, Watson M. 2015. Exome Sequencing: Current and Future Perspectives. *G3* **5**(8): 1543-1550.
- Weinberger DR. 1987. Implications of normal brain development for the pathogenesis of schizophrenia. *Arch Gen Psychiatry* **44**(7): 660-669.
- Weiss KM, Clark AG. 2002. Linkage disequilibrium and the mapping of complex human traits. *Trends Genet* **18**(1): 19-24.
- Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, Saemundsen E, Stefansson H, Ferreira MA, Green T et al. 2008. Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* **358**(7): 667-675.
- Wender PH, Rosenthal D, Kety SS, Schulsinger F, Welner J. 1974. Crossfostering. A research strategy for clarifying the role of genetic and experiential factors in the etiology of schizophrenia. *Arch Gen Psychiatry* **30**(1): 121-128.

- Weng L, Macciardi F, Subramanian A, Guffanti G, Potkin SG, Yu Z, Xie X. 2011. SNP-based pathway enrichment analysis for genome-wide association studies. *BMC Bioinformatics* **12**: 99.
- Williams HJ, Norton N, Dwyer S, Moskvina V, Nikolov I, Carroll L, Georgieva L, Williams NM, Morris DW, Quinn EM et al. 2011. Fine mapping of ZNF804A and genome-wide significant evidence for its involvement in schizophrenia and bipolar disorder. *Mol Psychiatry* **16**(4): 429-441.
- Williams HJ, Owen MJ, O'Donovan MC. 2009. Schizophrenia genetics: new insights from new approaches. *Br Med Bull* **91**: 61-74.
- Witte JS, Visscher PM, Wray NR. 2014. The contribution of genetic variants to disease depends on the ruler. *Nat Rev Genet* **15**(11): 765-776.
- Wohl M, Gorwood P. 2007. Paternal ages below or above 35 years old are associated with a different risk of schizophrenia in the offspring. *Eur Psychiatry* **22**(1): 22-26.
- Wray NR, Lee SH, Mehta D, Vinkhuyzen AA, Dudbridge F, Middeldorp CM. 2014. Research review: Polygenic methods and their application to psychiatric traits. *J Child Psychol Psychiatry* **55**(10): 1068-1087.
- Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. 2011. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* **89**(1): 82-93.
- Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, Lander ES, Kellis M. 2005. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* **434**(7031): 338-345.
- Xu B, Ionita-Laza I, Roos JL, Boone B, Woodruff S, Sun Y, Levy S, Gogos JA, Karayiorgou M. 2012. De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nat Genet* **44**(12): 1365-1369.
- Xu B, Roos JL, Dexeheimer P, Boone B, Plummer B, Levy S, Gogos JA, Karayiorgou M. 2011. Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nat Genet* **43**(9): 864-868.
- Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M. 2008. Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet* **40**(7): 880-885.
- Xu W, Cohen-Woods S, Chen Q, Noor A, Knight J, Hosang G, Parikh SV, De Luca V, Tozzi F, Muglia P et al. 2014. Genome-wide association study of bipolar disorder in Canadian and UK populations corroborates disease loci including SYNE1 and CSMD1. *BMC Med Genet* **15**: 2.
- Yang K, Trepanier CH, Li H, Beazely MA, Lerner EA, Jackson MF, MacDonald JF. 2009. Vasoactive intestinal peptide acts via multiple signal pathways to regulate hippocampal NMDA receptors and synaptic transmission. *Hippocampus* **19**(9): 779-789.
- Yang Y, Bauer C, Strasser G, Wollman R, Julien JP, Fuchs E. 1999. Integrators of the cytoskeleton that stabilize microtubules. *Cell* **98**(2): 229-238.
- Youn YH, Pramparo T, Hirotsune S, Wynshaw-Boris A. 2009. Distinct dose-dependent cortical neuronal migration and neurite extension defects in *Lis1* and *Ndel1* mutant mice. *J Neurosci* **29**(49): 15520-15530.
- Young-Pearse TL, Suth S, Luth ES, Sawa A, Selkoe DJ. 2010. Biochemical and functional interaction of disrupted-in-schizophrenia 1 and amyloid precursor protein regulates neuronal migration during mammalian cortical development. *J Neurosci* **30**(31): 10431-10440.

Bibliografía

- Yu TW, Chahrour MH, Coulter ME, Jiralerspong S, Okamura-Ikeda K, Ataman B, Schmitz-Abe K, Harmin DA, Adli M, Malik AN et al. 2013. Using whole-exome sequencing to identify inherited causes of autism. *Neuron* **77**(2): 259-273.
- Yue WH, Wang HF, Sun LD, Tang FL, Liu ZH, Zhang HX, Li WQ, Zhang YL, Zhang Y, Ma CC et al. 2011. Genome-wide association study identifies a susceptibility locus for schizophrenia in Han Chinese at 11p11.2. *Nat Genet* **43**(12): 1228-1231.
- Zaben M, Sheward WJ, Shtaya A, Abbosh C, Harmar AJ, Pringle AK, Gray WP. 2009. The neurotransmitter VIP expands the pool of symmetrically dividing postnatal dentate gyrus precursors via VPAC2 receptors or directs them toward a neuronal fate via VPAC1 receptors. *Stem Cells* **27**(10): 2539-2551.
- Zhang, ZHANG, ZHANG, WANG. 2011. Pathway-based analysis for genome-wide association studies of schizophrenia to provide new insight in schizophrenia study. *Chinese Science Bulletin* **56**(32): 4.
- Zhang X, Lei K, Yuan X, Wu X, Zhuang Y, Xu T, Xu R, Han M. 2009. SUN1/2 and Syne/Nesprin-1/2 complexes connect centrosome to the nucleus during neurogenesis and neuronal migration in mice. *Neuron* **64**(2): 173-187.
- Zhang X, Tochigi M, Ohashi J, Maeda K, Kato T, Okazaki Y, Kato N, Tokunaga K, Sawa A, Sasaki T. 2005. Association study of the DISC1/TRAX locus with schizophrenia in a Japanese population. *Schizophr Res* **79**(2-3): 175-180.
- Zuk O, Hechter E, Sunyaev SR, Lander ES. 2012. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A* **109**(4): 1193-1198.
- Zweier C, de Jong EK, Zweier M, Orrico A, Ousager LB, Collins AL, Bijlsma EK, Oortveld MA, Ekici AB, Reis A et al. 2009. CNTNAP2 and NRXN1 are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic protein in Drosophila. *Am J Hum Genet* **85**(5): 655-666.

