



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

Estudio genético en esquizofrenia:
análisis de variantes funcionales para la
identificación de factores de predisposición

Memoria para optar al grado de Doctor
presentada por:

Noa Carrera Cachaza

Enero 2012



El Doctor Ángel Carracedo Álvarez, Catedrático de Medicina Legal de la Facultad de Medicina de la Universidad de Santiago de Compostela, y el Doctor Javier Costas Costas, Investigador de la Fundación Pública Galega de Medicina Xenómica,

CERTIFICAN:

Que el presente trabajo que lleva por título “Estudio genético en esquizofrenia: análisis de variantes funcionales para la identificación de factores de predisposición”, realizado por Noa Carrera Cachaza bajo nuestra dirección, ha sido revisado y está en disposición de ser presentado para optar al grado de Doctor en Biología.

Y para que así conste, firmamos el presente certificado en Santiago de Compostela, a 27 de diciembre de 2011

Fdo: Dr. Ángel Carracedo

Álvarez Fdo: Dr. Javier Costas Costas

Agradecimientos

Me gustaría, en primer lugar, mostrar mi agradecimiento a Ángel Carracedo por haberme dado la gran oportunidad de formar parte de su grupo de investigación y agradecerle el esfuerzo y la energía que emplea a diario para impulsarlo.

Me gustaría también agradecer a Xabi, mi tutor, su dirección durante estos años de trabajo e intensa formación, y su paciencia. Espero haber incorporado en alguna medida su buen hacer y del rigor con el que siempre aborda sus investigaciones.

Gracias a todas las personas que siempre están dispuestas a ayudarme y que también están contribuyendo en gran medida a mi formación; de forma especial a mis compañeros y amigos del grupo de Psiquiátricas, del CeGen y del grupo de Xenómica, que a lo largo de estos más de cinco años me han acompañado y ayudado, tanto personalmente como en el desarrollo de esta investigación. Con ellos he compartido muchísimo, dentro y fuera del laboratorio. Gracias también a todas las personas que colaboran aportando muestras. Sin su trabajo el nuestro no sería posible.

Por último, gracias a mi familia y a mis amigos por la ayuda y el apoyo cariñoso que siempre me dan y que para mí es tan importante, en especial a Mauris, mi *compañerito*.

Esta tesis fue realizada con la ayuda de una beca de tercer ciclo y un contrato predoctoral del programa María Barbeito de la Xunta de Galicia y estuvo financiado a través de proyectos de investigación por el Instituto de Salud Carlos Tercero (FIS 02/0018, PI052036 y PI081522) y la Xunta de Galicia (INCITE08PXIB9101149PR y REGENPSI).

A mi padre, a mi madre y a Pepe

Índice

Introducción	5
1. La esquizofrenia. Aspectos clínicos	3
1.1 Síntomas y diagnóstico. Problemas asociados	3
1.2 Curso de la enfermedad	6
1.3 Aspectos epidemiológicos	7
1.4 Tratamiento	8
1.5 Hipótesis sobre el origen y el desarrollo de la enfermedad	10
1.5.1 Sistemas de neurotransmisores	10
1.5.2 Hipótesis del neurodesarrollo	11
1.6 Neurología	12
1.6.1 Neuroanatomía	12
1.6.2 Conectividad	13
1.6.3 Alteraciones neurofisiológicas y neurocognitivas	14
2. Genética de la esquizofrenia	15
2.1 Epidemiología genética	15
2.2 Diseño de un estudio genético	17
2.3 Estudios de ligamiento	18
2.4 Estudios de asociación	20
2.4.1 Hipótesis enfermedad común / variante común	20
2.4.1.1 Estudios con genes candidatos	21
2.4.1.2 Estudios de genoma completo	25
2.4.2 Hipótesis enfermedad común / variante rara	32

3. Papel de la selección natural en el espectro alélico de la esquizofrenia	38
3.1 Esquizofrenia, ¿paradoja evolutiva?	38
3.2 Esquizofrenia y selección natural	39
Hipótesis y Objetivos	43
Métodos, resultados y discusión	47
Estudio 1. Esquizofrenia, DRD3 y selección natural	49
Estudio 2. Modelo de susceptibilidad ancestral en esquizofrenia: <i>MAOB</i>	59
Estudio 3. Modelo de pleiotropismo antagónico en esquizofrenia: <i>DAOA</i> , <i>PPPIRIB</i> , y <i>APOLI</i> .	67
Estudio 4. Selección natural reciente en genes candidatos funcionales	73
Estudio 5. Análisis de SNPs no sinónimos comunes	91
Estudio 6. Análisis de SNPs no sinónimos de baja frecuencia en genes candidatos.	103
Discusión general	117
Conclusiones	143
Bibliografía	147

Anexo. Se adjunta un disco compacto con el siguiente material suplementario de la tesis y los artículos presentados.

ESTUDIO 4. Supplementary Table 1

ESTUDIO 5. Supplement 1

ESTUDIO 6. Supplementary Figure 1

ESTUDIO 6. Supplementary Table 1

ESTUDIO 6. Supplementary Table 2

ESTUDIO 6. Supplementary Table 3

TESIS Tabla Suplementaria 1

Introducción

1. La esquizofrenia. Aspectos clínicos

1.1 Síntomas y diagnóstico. Problemas asociados

El concepto de esquizofrenia como enfermedad es bastante reciente, de finales del siglo XIX, y hace referencia a personas que presentan un conjunto muy variado de síntomas entre los que predominan una alteración de la afectividad, del comportamiento social, de la percepción de la realidad, del flujo y asociación de los pensamientos y aparición de brotes psicóticos. El fenotipo englobado bajo esta entidad clínica es muy heterogéneo: los síntomas que sufre un mismo paciente pueden variar a lo largo de toda su vida y varían también entre diferentes individuos, de modo que dos pacientes diagnosticados como esquizofrénicos pueden no tener ningún síntoma en común. Además, estos síntomas pueden presentarse también en otros trastornos mentales considerados diferentes tales como el trastorno esquizotípico, el trastorno bipolar o el autismo.

Desde su definición, el diagnóstico de la enfermedad se ha basado únicamente en observaciones clínicas del estado mental del afectado ya que, aun habiendo sido extensamente estudiada, todavía no se ha encontrado su origen ni ningún marcador biológico que caracterice particularmente a los pacientes, por lo que no se han podido desarrollar pruebas clínicas objetivas. Esto, unido a la enorme heterogeneidad fenotípica, da lugar a que aunque desde su descripción la enfermedad haya constituido una entidad clínica única, su definición y límites hayan variado, y varían, en función de los criterios empleados para su diagnóstico. En este sentido, para tratar de unificar los criterios y lograr un diagnóstico homogéneo se han ido elaborando manuales para el empleo universal, aplicables por todos los psiquiatras del mundo, que tratan de

recoger los síntomas que mejor definen a la enfermedad. Estos manuales se van actualizando y mejorando a medida que pasan los años. En este momento los criterios de referencia que se emplean con más frecuencia son los definidos en la versión cuarta del Manual Diagnóstico y Estadístico de los Trastornos Mentales de la Sociedad Americana de Psiquiatría (DMS-IV; American Psychiatric Association Diagnostic and Statistical Manual of Mental Disorders, 2002), y la décima revisión de la Clasificación Internacional de Enfermedades de la Organización Mundial de la Salud (ICD-10; World Health Organization, 1992), y el diagnóstico se basa en una exploración psicopatológica, una entrevista semi-estructurada del paciente y si es posible de su familia y en su historia clínica.

La definición actual de esquizofrenia recogida en estos manuales deriva principalmente de las aportaciones de tres autores: la cronicidad de Kraepelin (persistencia de algunos de los síntomas en el tiempo), los síntomas negativos y afectivos que destacó Bleuer, y los síntomas positivos de Schneiderian (Andreasen, 2007). A continuación se describen algunos de los síntomas que aparecen en los pacientes (Tandom 2008):

Síntomas positivos: se refieren a aquellos que son una exacerbación de elementos psíquicos normales, o bien elementos que no existen en otras personas. Constituyen lo que en general se conoce como “psicosis”: percepción errónea de la realidad, alucinaciones, ideas delirantes (persecución, sustitución de la personalidad, transmisión de pensamientos...) y otras distorsiones con un convencimiento psicótico de que “su” realidad es la de todos. Son los síntomas más llamativos, pero no caracterizan particularmente a los pacientes esquizofrénicos ya que no siempre están presentes y muchos de ellos pueden aparecer en otras patologías, por ejemplo en la demencia de diversos orígenes, en el trastorno bipolar o en el alcoholismo (Carpenter, 2011).

Síntomas negativos: aquellos relacionados con rasgos de personalidad presentes en personas sanas pero ausentes o alterados en pacientes: aplanamiento afectivo, falta de empatía, abulia, retraimientos emocional y social, apatía, pasividad, falta de espontaneidad, anhedonia, alogia...

Síntomas cognitivos: relacionados con déficits de atención y de memoria a corto plazo (memoria de trabajo, un tipo de memoria a corto plazo que permite manejar temporalmente información necesaria para realizar una tarea), déficit en el procesamiento de la información, alteración en las funciones ejecutivas (procesos mentales que permiten responder y adaptarse de modo apropiado al entorno), trastorno asociativo, dificultad para elaborar nuevas ideas, dificultad de aprendizaje, problemas para concentrarse, priorizar, ordenar... Estos síntomas sí caracterizan de modo más general a las personas esquizofrénicas ya que, a diferencia de otros síntomas, están generalmente presentes en todos los pacientes y durante toda su vida y suelen aparecer ya alrededor de los 6-14 años (Carpenter, 2011).

Otros síntomas importantes son conducta y discurso desorganizados, pensamiento desordenado, limitación en la fluidez y en la productividad del discurso y del pensamiento, desesperanza, ideas e intentos suicidas, estados catatónicos... Síntomas menos obvios son el aislamiento social (evitar el contacto con familiares, amigos y otras personas) o la conducta hostil o impulsiva.

Los manuales para el diagnóstico diferencian además subtipos de la enfermedad en base a los síntomas que predominan. Así, el DSM-IV reconoce cinco tipos de esquizofrenia: paranoide, desorganizado, catatónico, indiferenciado y residual, y el ICD-10 contempla nueve clases: paranoide, hebefrénica, catatónica, indiferenciada, post-esquizofrenia, residual, simple y las denominadas otras esquizofrenias y esquizofrenia no especificada. Sin embargo, dada la riqueza sintomática de cada paciente y el hecho de que los síntomas varíen a lo largo de su vida, es difícil asignar a cada enfermo una única categoría y un mismo individuo puede pasar por diferentes “tipos de esquizofrenia” en el curso de su enfermedad. En relación a esto, se ha sugerido que en la próxima actualización del DSM se eliminen estas categorías (web: <http://www.dsm5.org>).

Esta variabilidad sintomática y el diferente grado en que los síntomas se manifiestan en cada persona hacen que no se pueda hablar de un cuadro típico de esquizofrenia, lo que dificulta su diagnóstico precoz y posterior tratamiento. Además, como hemos comentado, la esquizofrenia comparte rasgos clínicos con un amplio rango de enfermedades mentales, como el trastorno bipolar, el trastorno esquizoafectivo o el autismo (Siever y Davis, 2004). Esto hace que, en algunos casos, resulte difícil establecer un diagnóstico específico con certeza y sea necesario recurrir a más de una categoría diagnóstica para describir el fenotipo de una persona (Krueger, 1999). Otro aspecto importante es que existe inestabilidad diagnóstica de los trastornos psiquiátricos sobre todo en etapas tempranas de la enfermedad; por ejemplo, se ha estimado que en torno al 25% de los pacientes diagnosticados de esquizofrenia cambian el diagnóstico a otro, entre ellos, desorden esquizoafectivo (Haahr et al., 2008; Salvatore et al., 2009).

Este aparente caos sintomático, secundado desde el comienzo de los estudios genéticos de la enfermedad por la identificación de factores genéticos de vulnerabilidad compartidos entre diferentes categorías diagnósticas mentales, plantea un problema a la hora de entender las enfermedades mentales como categorías biológicas diferenciadas, con causas, síntomas y consecuencias específicas, y existe un sentimiento general, sobre todo desde el punto de vista de los que abordan la enfermedad desde la genética, de que la actual división categórica de estos trastornos debería ser reconsiderada. Así, se ha sugerido que sería más adecuado entender las enfermedades mentales como un continuo, con solapamiento entre ellas, tanto a nivel fenotípico como en origen (con causas comunes a varias de estas enfermedades y otras específicas de cada una). El continuo no estaría restringido a síndromes psicóticos o psiquiátricos (Gejman et al., 2011; Owen et al., 2011). Este punto de vista es controvertido. Una opinión contraria a éste puede verse, por ejemplo, en Lawrie et al

(Lawrie et al., 2010), artículo en el que se considera que la hipótesis del continuo para las enfermedades mentales no tiene base científica suficiente y que sería difícilmente incorporable desde el punto de vista de la práctica clínica.

1.2 Curso de la enfermedad

La enfermedad suele manifestarse de manera clara en la adolescencia o inicio de la edad adulta, cuando aparece el primer síntoma psicótico (Messias et al., 2007), pero esto puede ir precedido por meses o incluso años de deterioro cognitivo y alteraciones afectivas y sociales menos evidentes, que pueden ser casi imperceptibles (Carpenter, 2011; Hafner et al., 1992). De hecho, diversos autores entienden que la esquizofrenia se desarrolla progresivamente en tres etapas psicopatológicas que se corresponden con distintas fases clínicas: etapa de desarrollo neuronal, que incluiría la fase premórbida, durante la gestación e infancia; etapa neuroplástica, que incluiría la fase prodrómica (que precede a la enfermedad, que anuncian la llegada de la enfermedad), a partir de la pubertad, y la fase de manifestación de la enfermedad y deterioro posterior; y etapa neuroprogresiva, que sería una fase de mayor deterioro y consolidación crónica (Lieberman et al., 2001). De acuerdo con esto, estudios en individuos de alto riesgo sugieren que múltiples factores premórbidos relacionados con complicaciones pre y perinatales, o disfunciones cognitivas, motoras o sociales durante la infancia se asocian con la posterior aparición de la enfermedad (Lieberman, 2006). Asimismo, se observa un mantenimiento o refuerzo de estas disfunciones en la fase prodrómica, en la que pueden aparecer además síntomas positivos atenuados (ilusiones, pensamientos mágicos, supersticiones), síntomas cognitivos (dificultad para la concentración), alteraciones del estado de ánimo (ansiedad, disforia, irritabilidad), aislamiento social o comportamientos obsesivos que interrumpen la vida normal del individuo (Addington et al., 2007). De todos modos, estos síntomas premórbidos y prodrómicos no dejan de ser muy inespecíficos, pues aparecen también en un amplio rango de alteraciones del comportamiento en personas que posteriormente no desarrollan esquizofrenia, por lo que, aunque se consideren, no tienen valor predictivo o diagnóstico.

Existen diferencias entre hombres y mujeres en cuanto a la manifestación de la enfermedad (Hafner, 2003; Salem y Kring, 1998). Por ejemplo, en los hombres la distribución de la edad en la que la enfermedad se muestra de manera clara presenta un único pico entre los 15 y 25 años y disminuye progresivamente (Hafner, 2003), mientras que en las mujeres los síntomas psicóticos suelen aparecer más tarde y la edad de aparición sigue una distribución bimodal, con un segundo pico agudo entre los 45 y 50 años, coincidiendo con la menopausia (Leung y Chue, 2000; Messias et al., 2007; Riecher-Rossler y Hafner, 2000). Otra diferencia es que los hombres con frecuencia sufren síntomas más severos que las mujeres (Leung y Chue, 2000).

La evolución de la enfermedad en cada paciente es variable. Una proporción de las personas que sufren un brote psicótico se recuperan y no vuelven a sufrir otra crisis, pudiendo llevar una vida normal. Otras personas tienen varios episodios de psicosis en el transcurso de su vida, con remisión parcial o total, y entre dichos períodos es posible que puedan llevar una vida más o menos normal. En el caso más grave los pacientes evolucionan hacia una esquizofrenia "crónica", que se manifiesta de forma continua o recurrente, haciendo que las personas permanezcan absolutamente incapacitadas por la enfermedad, viviendo un estado psicótico continuo y dando lugar a una enorme carga familiar (Walker et al., 2004). Muchos de estos pacientes sufren un deterioro progresivo en sus síntomas asociado a la resistencia al tratamiento.

En general, la enfermedad se asocia a una marcada reducción en las funciones sociales y laborales, con dificultad para mantener conductas motivadas y dirigidas hacia metas concretas, por lo que resulta una patología muy incapacitante. Existe una tasa muy elevada de desempleo entre las personas esquizofrénicas que implica una dependencia económica que aumenta el estrés de estas personas y empeora su integración social (Bartley, 1994).

1.3 Aspectos epidemiológicos

La esquizofrenia es una enfermedad presente en todas las poblaciones y culturas del mundo que afecta aproximadamente al 0,72% de la población general, aunque esta cifra varía en función de diferentes factores (McGrath et al., 2008). Por ejemplo, la enfermedad es más frecuente en hombres que en mujeres (1,4:1) (McGrath et al., 2008). Otro factor que influye es la latitud, pues se ha visto que países de mayor latitud muestran una mayor prevalencia de la enfermedad (Saha et al., 2006); dentro de un mismo país varía además en función de diferentes factores psicosociales como la condición de emigrante, pertenecer a un grupo minoritario o a clases sociales desfavorecidas, crecimiento en un entorno urbano, traumas infantiles, el abuso de drogas como el cannabis, o el estado civil, predominando el porcentaje de solteros y separados (Cantor-Graae y Selten, 2005; Fearon et al., 2006). Algunos autores han sugerido que esta diferente prevalencia podría responder a que una exposición a estos factores durante determinados periodos sensibles podría tener un impacto en el desarrollo del cerebro "social" del individuo, incrementando el estrés y precipitando la aparición de la enfermedad en una persona vulnerable (van Os et al., 2010). Sin embargo, no en todos los casos se puede definir en qué sentido se establece la relación causa efecto, en caso de que exista, ya que la situación desestructurada podría ser provocada por el propio trastorno (Cooper, 2005; Saha et al., 2005; Sullivan, 2005).

Otros factores observados que parecen aumentar el riesgo a desarrollar la enfermedad son complicaciones severas durante la gestación, el parto y

complicaciones perinatales, entre las que se encuentran una nutrición insuficiente, infección materna durante el embarazo, hipoxia, daño cerebral en el feto, diabetes materna, estación en la que se nace (relacionado con infecciones prenatales con incidencia diferente en cada época del año: gripe, rubéola, toxoplasmosis) (Brown, 2011) y edad avanzada del padre (Torrey et al., 2009). Como se verá más adelante, estas observaciones han contribuido a la formulación de la hipótesis del neurodesarrollo como origen de la enfermedad.

Es posible que el efecto de estos factores ambientales modifique el fenotipo a través de la regulación de la expresión génica, es decir, a través de mecanismos epigenéticos, aunque por el momento, debido a dificultades metodológicas, el papel de la epigenética no se ha confirmado (Kaminsky et al., 2009; Roth et al., 2009).

De todos modos, el efecto individual de estos factores de riesgo es moderado y en ningún caso determinante y ninguno de ellos predispone tanto como tener un familiar de primer grado con la enfermedad, lo que sitúa a los factores genéticos como causa principal de esquizofrenia.

La tasa de mortalidad de individuos esquizofrénicos es entre 2 y 3 veces mayor que la de la población general (McGrath et al., 2008) causando más muertes prematuras que otras enfermedades como el cáncer o la muerte física (van Os y Kapur, 2009). Este incremento se debe, entre otros factores, a una elevada tasa de suicidio: el 5% de los pacientes se suicidan, sobre todo hombres, en las primeras fases de la enfermedad (Palmer et al., 2005). Otras causas que contribuyen a aumentar la mortalidad son, por ejemplo, complicaciones cardiovasculares en edades más avanzadas, dieta pobre, poco ejercicio, obesidad o el tabaco (el 80% de los pacientes son fumadores, un porcentaje entre dos y tres veces mayor que en la población general) (Batel, 2000; Saha et al., 2007). Muchas de estas complicaciones están relacionadas con el propio tratamiento.

La enfermedad está asociada también a una fertilidad reducida (con una tasa de fertilidad del 39,95%), especialmente en hombres, que responde a la falta de integración social y a la tendencia a la soledad característica de las personas esquizofrénicas que hace que, en general, tengan dificultad para encontrar pareja y en consecuencia no tengan descendencia (Bundy et al., 2011).

1.4 Tratamiento

Las primeras moléculas con efecto antipsicótico se descubrieron en los años 60 de forma accidental, al administrar un fármaco que se empleaba en cirugía como anestésico a un paciente esquizofrénico en estado maníaco y ver que se reducían sus síntomas positivos. Este descubrimiento supuso una revolución en la psiquiatría y una gran ayuda para las personas enfermas que hasta ese momento no contaban con ningún

tratamiento y generalmente vivían estigmatizadas, recluidas en manicomios y apartadas de la sociedad (Tandon et al., 2010). Este tipo de fármacos, actualmente llamados antipsicóticos típicos o de primera generación y entre los que se encuentran la clorpromazina y el haloperidol, ejercen su acción bloqueando los receptores de dopamina D_2 , y esta acción está estrechamente relacionada con la respuesta antipsicótica, limitando la aparición de síntomas positivos y de desorganización. Estas primeras moléculas presentaban una serie de problemas ya que no eran efectivas en todos los pacientes, no eran efectivas frente a otro tipo de síntomas como los negativos o cognitivos (asociados a un alto grado de incapacidad en los pacientes) y además generaban importantes efectos secundarios, como los efectos extrapiramidales, entre ellos acatisia, parkinsonismo, distonía o discinesia tardía.

Para intentar resolver estos problemas se desarrollaron nuevos antipsicóticos. El primero de ellos fue la clozapina, que apareció a finales de los años 60 (Tandon et al., 2010). El fármaco era muy eficaz contra los síntomas positivos y no causaba efectos extrapiramidales, pero generaba otros efectos secundarios muy graves, como la agranulocitosis, por lo que su uso fue limitado (Tandon et al., 2010). A pesar de esto, el hecho de que no provocase efectos extrapiramidales generó un gran optimismo en cuanto a las posibilidades de mejora del tratamiento de la psicosis y motivó la síntesis de nuevos fármacos con perfil farmacodinámico similar al de la clozapina, que bloqueaba los receptores serotoninérgicos $5-HT_{2A}$. De esta forma surgieron a partir de los años 90 nuevos antipsicóticos, atípicos o de segunda generación.

Así, actualmente existen unos 51 antipsicóticos de primera generación y 13 de segunda generación. Todos ellos ejercen su acción principal mediante el bloqueo de los receptores D_2 de la dopamina y esta acción se asocia directamente con la respuesta antipsicótica. La mayor parte de ellos intervienen además en otros sistemas de neurotransmisión como el serotoninérgico, adrenérgico o muscarínico. El estudio CATIE, en el que se siguió durante más de 18 meses a 1.493 pacientes (Lieberman et al., 2005) demostró que, a pesar de lo que se pensaba inicialmente, no existen grandes diferencias en cuanto a eficacia entre los tratamientos antipsicóticos de primera y segunda generación. En todo caso, para un mismo grado de reducción de síntomas positivos, los antipsicóticos de segunda generación pueden presentar un menor riesgo de aparición de efectos extrapiramidales -asociado a su característico bloqueo del receptor serotoninérgico- y un mayor riesgo de aparición de efectos metabólicos adversos, aunque la diferencia no es significativa y existe una gran heterogeneidad respecto a estos efectos en cada clase (Tandon et al., 2010). En general, ningún antipsicótico es muy efectivo frente a síntomas negativos o cognitivos, y todos están asociados a la aparición de efectos secundarios importantes a diferentes niveles, extrapiramidal, cardiovascular, hematológico, genito-urinario, disfuncionalidad sexual, endocrino o metabólico, entre otros, diferenciándose cada uno de los 64 fármacos en su propensión a causar dichos efectos (Tandon et al., 2010). Todos estos problemas provocan que muchos pacientes no sigan de manera adecuada la medicación o la abandonen de manera brusca, aumentando el riesgo de una crisis psicótica aguda (Lieberman et al., 2005).

El tratamiento actual de un paciente suele consistir en una combinación de antipsicóticos y de otro tipo de fármacos, pero cada paciente necesita una combinación concreta a una dosis determinada y cada uno desarrolla efectos secundarios variables. Esto hace que la dinámica para conseguir un tratamiento eficaz para cada paciente resulte un trabajo muy laborioso, basado en el “ensayo y error”, que supone hacer un seguimiento exhaustivo de cada paciente y modificar la medicación y la dosis a medida que evoluciona la enfermedad. De todos modos, este tratamiento no cura la enfermedad, por lo que ésta puede alargarse durante toda la vida del paciente.

Además del tratamiento químico es muy importante la terapia psico-social para favorecer un mejor pronóstico de la enfermedad, reducir el número de recaídas y ayudar a las personas a convivir con su enfermedad y entenderla (van Os y Kapur, 2009). Las terapias de rehabilitación -individualizadas y en grupo-, los grupos de autoayuda, la educación familiar y la inserción laboral con remuneración ayudan enormemente a los pacientes a sobrellevar su enfermedad y evitan su estigmatización.

En cualquier caso existe una necesidad inmediata de encontrar nuevos tratamientos, pero sin entender las causas y la biología de la enfermedad, la capacidad para desarrollar tratamientos efectivos o de tomar medidas de prevención es muy limitada, por lo que cualquier avance en la biología molecular de la enfermedad permitiría mejorar la eficacia de los tratamientos.

1.5 Hipótesis sobre el origen y el desarrollo de la enfermedad

1.5.1 Sistemas de neurotransmisores

Las primeras hipótesis científicas que intentaron concretar el origen de la enfermedad se basaban en el mecanismo de acción de los antipsicóticos o de otras drogas cuyo consumo daba lugar a síntomas observados en pacientes. Una de las primeras hipótesis formuladas fue la hipótesis dopaminérgica, que surgió al ver que la eficacia de los fármacos antipsicóticos de primera generación era proporcional a su capacidad para bloquear los receptores de la dopamina. Esta teoría sugería que la enfermedad podría estar causada por una hiperactividad dopaminérgica generalizada (Seeman y Lee, 1975). Posteriormente, esta hipótesis se reformuló en base a nuevos hallazgos y actualmente se considera que existe un desequilibrio en el sistema dopaminérgico en determinadas regiones del cerebro: habría un nivel elevado de dopamina a nivel subcortical mediada por los receptores D_2 de la dopamina, que sería el responsable de los síntomas positivos de la enfermedad, y niveles más bajos en otras regiones, como la corteza prefrontal, relacionado con los receptores D_1 , que darían

lugar a síntomas negativos (Pani, 2002). Otro hecho que respalda la hipótesis es que el consumo de drogas psicotrópicas, como las anfetaminas o la cocaína, que aumentan la liberación de dopamina, da lugar a cuadros psicóticos similares a los de la esquizofrenia en personas sanas.

El estudio del mecanismo de acción de la clozapina y de otros antipsicóticos de segunda generación mostró que estas moléculas tenían afinidad por receptores de la serotonina. Esta observación hizo pensar que la alteración de este sistema podría intervenir en la aparición de la enfermedad, formulándose la hipótesis serotoninérgica. Un hecho que respalda esta hipótesis es que el consumo de drogas que intervienen en este sistema, como el LSD (dietilamida del ácido lisérgico), un agonista del receptor de la serotonina, da lugar a alucinaciones y cuadros psicóticos (Geyer y Vollenweider, 2008).

Por último, la hipótesis hipoglutamatergica surgió directamente del estudio del mecanismo de acción de diferentes drogas que administradas a bajas dosis en personas normales daban lugar a la aparición de síntomas concretos observados en pacientes esquizofrénicos, como síntomas negativos, discapacidad cognitiva y alteraciones psicológicas (Lahti et al., 2001). Estas drogas son antagonistas del receptor N-metil-D-aspartato (NMDA) del glutamato, lo que llevó a postular en un primer momento que la hipofunción del receptor podría relacionarse con ciertos síntomas de la enfermedad (Coyle, 2006). Posteriormente se perfiló la hipótesis y actualmente se sugiere que podría existir un desequilibrio del sistema glutamatérgico, sobre todo en regiones frontales, que daría lugar en última instancia a la alteración del sistema dopaminérgico. Se ha observado, además, que agentes moduladores del receptor NMDA reducen en alguna medida síntomas cognitivos de la enfermedad (Insel, 2010).

Estas hipótesis no son excluyentes, de hecho hay evidencias que las relacionan. Así, se ha demostrado que los receptores del glutamato y la dopamina están funcionalmente asociados mediante complejos, particularmente mediante proteínas estructurales (Yao et al., 2008; Zhang et al., 2007a). También se ha visto la formación de complejos funcionales entre los receptores de la serotonina y el receptor metabotrópico del glutamato en la corteza del cerebro, y que la desregulación de este complejo puede estar relacionado con las alteraciones de procesos corticales presentes en pacientes esquizofrénicos (Fribourg et al., 2011; Gonzalez-Maeso et al., 2008).

1.5.2 Hipótesis del neurodesarrollo

La existencia de síntomas prodrómicos asociados a la posterior aparición de la enfermedad (síntomas cognitivos, negativos...), unido a la aparente ausencia de cambios neuropatológicos en las personas enfermas (Weinberger y McClure, 2002) dieron lugar al desarrollo de la hipótesis del neurodesarrollo (Murray y Lewis, 1987; Weinberger et al., 1986). Ésta postula que la esquizofrenia sería el resultado de

alteraciones tempranas en el desarrollo del cerebro ya durante la gestación; es decir, que los factores responsables de la aparición posterior de la enfermedad actuarían durante el neurodesarrollo, pero se manifestarían clínicamente entre una y tres décadas después del nacimiento. Estas alteraciones darían lugar a fallos en el “cerebro social” de las personas y alteraciones cognitivas (Owen et al., 2011). De hecho, estudios en primates mostraron que lesiones cerebrales neonatales daban lugar a alteraciones en el comportamiento a una edad avanzada (Owen et al., 2011). Bajo esta hipótesis se puede comprender la observación de que complicaciones obstétricas como las citadas anteriormente -infecciones maternas, hipoxia, diabetes materna, etc.- sean factores ambientales de riesgo que se relacionan con la aparición de la enfermedad.

La razón por la que la enfermedad se manifiesta de forma dramática tras la adolescencia se desconoce, pero algunos autores sugieren que podría estar relacionada con un fallo en la maduración normal que ocurre típicamente en la adolescencia tardía, particularmente la mielinización de la materia blanca y la poda sináptica cortical; este fallo podría estar provocado por un desarrollo anormal de sistemas neuronales críticos en el cerebro importantes para esta remodelación (Meyer-Lindenberg, 2010)

La hipótesis del neurodesarrollo, extensiva a otros trastornos psiquiátricos, ha sido respaldada por estudios epidemiológicos, de neurodesarrollo y de neuroimagen, y ha sido el paradigma dominante de la esquizofrenia en las últimas décadas (Owen et al., 2011).

1.6 Neurología

Desde la sospecha de que las enfermedades mentales tienen una base neurobiológica, se han intentado identificar rasgos neuropatológicos presentes en estas enfermedades. Entre las numerosas observaciones descritas en este campo podemos destacar las que tratamos a continuación.

1.6.1 Neuroanatomía

Estrategias de neuroimagen estructural han permitido observar rasgos anatómicos presentes en individuos de alto riesgo, en pacientes esquizofrénicos en su primer episodio psicótico y en pacientes crónicos, entre los que destacan reducciones en el volumen total del cerebro y de la materia gris (localizadas principalmente en el lóbulo temporal medial y superior, hipocampo, corteza prefrontal y tálamo), dilatación ventricular (lateral y tercero) y alteración de la materia blanca (Meyer-Lindenberg, 2010; Owen et al., 2011).

Se ha descrito, además, que estos cambios macroscópicos se corresponden con cambios a nivel microscópico. En la corteza prefrontal las neuronas piramidales -la mayor fuente de neurotransmisión cortical de excitación- son más pequeñas y están empaquetadas de manera más densa, lo que indica una reducción en el neuropilo (terminales axónicos, espinas dendríticas y células gliales) que ocupa el espacio entre neuronas. Se ha sugerido que esto puede ser debido a una poda sináptica excesiva durante la pubertad (Meyer-Lindenberg, 2010). Además, en esta región las poblaciones de interneuronas están reducidas. En el hipocampo, el soma de estas neuronas piramidales es también más pequeño y las espinas dendríticas están asimismo reducidas. También se ha descrito disminución en el número de neuronas a nivel del tálamo. En la corteza prefrontal la dopamina ejerce una función reguladora muy importante, por lo que estas reducciones podrían alterar funciones ejecutivas dependientes del nivel óptimo de la señalización dopaminérgica (Meyer-Lindenberg, 2010; Selemon y Goldman-Rakic, 1999). Al no haberse observado gliosis, estas modificaciones parecen responder a alteraciones durante el desarrollo, más que a un proceso neurodegenerativo (Iritani, 2007; Weinberger, 1987).

De las modificaciones anatómicas descritas, la reducción en el volumen del hipocampo y de la materia gris cortical se han observado también en familiares de pacientes, lo que indica un posible componente genético (Boos et al., 2007; Goldman et al., 2008).

De todos modos, las alteraciones observadas aparecen en un porcentaje importante de personas sanas, no están presentes en todos los pacientes y muchas de ellas no son exclusivas de esquizofrenia sino que aparecen también en otros trastornos mentales, lo que hace que tengan un valor diagnóstico escaso. Además, se han descrito numerosos factores que implican cambios en el cerebro maduro, entre ellos, el abuso de alcohol (Taki et al., 2006), la nicotina (Brody et al., 2004), la obesidad (Ward et al., 2005) y el tratamiento con antipsicóticos (Meyer-Lindenberg, A., 2010), por lo que es difícil asumir que todos los cambios observados sean debidos a la enfermedad y no a otros factores. Por otra parte, tal y como indica la hipótesis del neurodesarrollo, el momento en el que se diagnostica la enfermedad no se corresponde con el comienzo de la misma, sino que ésta ya se habría originado durante el desarrollo, evolucionando a lo largo de la vida del individuo; por tanto, las alteraciones observadas en los pacientes reflejarían una condición compleja que podría estar en un estado demasiado avanzado en su trayectoria como para que nos indique su causa (Meyer-Lindenberg, 2010).

1.6.2 Conectividad

Otro tipo de aproximación consiste en acercarse al cerebro *in vivo* a través de estudios de neuroimagen funcional, técnicas empleadas para investigar el funcionamiento de regiones determinadas del cerebro, tanto durante el descanso como

en respuesta a diferentes estímulos o tareas, y tratar de identificar circuitos concretos alterados (Keshavan et al., 2008).

Se han identificado circuitos anormales sobre todo en la corteza dorsolateral prefrontal en pacientes esquizofrénicos, y en general, se observa una conexión alterada entre distintas áreas del cerebro en pacientes, familiares e individuos de alto riesgo (Meyer-Lindenberg, 2010).

1.6.3 Alteraciones neurofisiológicas y neurocognitivas

Se ha observado que existen rasgos neurofisiológicos y neurocognitivos alterados de manera característica en individuos esquizofrénicos y, en muchos casos, en familiares. Entre ellos podemos destacar: déficit en la memoria a corto plazo, memoria verbal o episódica, inhibición o atenuación en la amplitud de potenciales evocados ante estímulos repetidos, inhibición del prepulso del reflejo de la respuesta de sobresalto ante un estímulo acústico brusco -un paciente esquizofrénico responde igual al primer y segundo estímulo; una persona normal sólo al primero, después ya está preparada-, estructura del sueño alterada -que incluye un acortamiento de la fase REM (del inglés, *Rapid Eye Movement*)- o movimientos sacádicos de los ojos, entre otros (Bilder et al., 2011; Bramon et al., 2005; Glahn et al., 2003).

Se ha visto que muchos de estos rasgos son heredables y se ha estimado su heredabilidad en torno al 50% (Greenwood et al., 2007; Sabb et al., 2008). En relación a esto, se ha observado que muchos de los fenotipos cognitivos están significativamente correlacionados con esquizofrenia y que la mayor parte de esta covarianza (del 72 al 92%) se debe a factores genéticos compartidos (Owens et al., 2011; Touloupoulou et al., 2010; Touloupoulou et al., 2007).

En respuesta a la enorme heterogeneidad fenotípica observada en los pacientes esquizofrénicos y teniendo en cuenta el solapamiento sintomatológico observado con otros trastornos mentales, se ha propuesto abordar el estudio de la enfermedad desde el análisis de fenotipos más simples, denominados endofenotipos (Gottesman y Gould, 2003). A estos endofenotipos se le supone una arquitectura genética más simple que la del propio trastorno, lo que podría facilitar la identificación de factores genéticos relacionados con la enfermedad. Los endofenotipos formarían parte del trastorno, estarían presentes en las personas enfermas, pero también en familiares cercanos en una proporción más elevada que en la población general. Muchos de los fenotipos neurofisiológicos y neurocognitivos comentados con anterioridad, como el déficit en la memoria a corto plazo o inhibición de potenciales evocados, se emplean en ocasiones como endofenotipos de esquizofrenia.

2. Genética de la esquizofrenia

2.1 Epidemiología genética

Para llevar a cabo un buen estudio genético es necesario que exista una entidad fenotípica válida y homogénea y que haya evidencias de la existencia de un componente genético hereditario que explique un porcentaje elevado de este fenotipo. Como ya se ha comentado anteriormente, el fenotipo englobado bajo el diagnóstico de esquizofrenia se caracteriza por una gran heterogeneidad, que afectará a la identificación de causas comunes.

Los estudios de epidemiología clásica se emplearon para determinar la contribución relativa del componente genético en el origen de la enfermedad. Este tipo de estudios, que incluyen estudios de agregación de la patología en familias, estudios de concordancia en gemelos homocigóticos frente a heterocigóticos y estudios de adopciones, analizan el grado de concordancia o parecido de un carácter entre personas que comparten porcentajes elevados de su genoma y su combinación permite dividir la varianza fenotípica observada en componentes atribuibles a diferentes causas, como el componente genético y el componente ambiental, y calcular la heredabilidad de un carácter, en nuestro caso, el fenotipo de la esquizofrenia.

Los estudios de agregación familiar en esquizofrenia ponen de manifiesto una mayor incidencia del trastorno en el grupo familiar que en la población general y que el riesgo a padecer esquizofrenia es mayor cuanto más porcentaje de genoma se comparte con una persona que la sufre (figura 1) Además, entre familiares de personas esquizofrénicas existe una elevada incidencia de otro tipo de trastornos psiquiátricos, como trastorno esquizotípico o trastorno bipolar (Riley y Kendler, 2006).

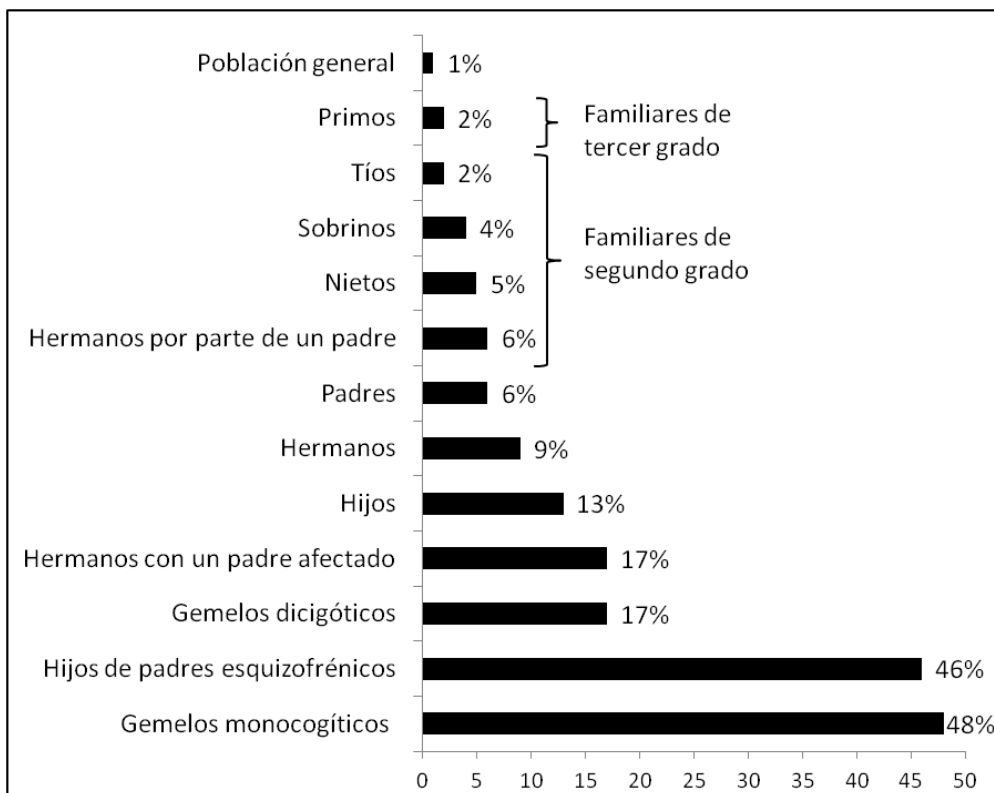


Figura 1. Riesgo de padecer esquizofrenia según el grado de parentesco con una persona que ya la padece. Gráfica adaptada de Riley y Kendler 2006

Los estudios con gemelos se basan en la comparación de la tasa de concordancia de la enfermedad- la probabilidad de que un gemelo desarrolle la enfermedad si el otro ya la padece- en gemelos monocigóticos (MZ) frente a dicigóticos (DZ). Estos estudios muestran que la concordancia en gemelos MZ es de un 48%, mientras que en gemelos DZ se sitúa alrededor de un 17% (Riley y Kendler, 2006). Estos datos sugieren que existe un componente genético responsable de la aparición de la enfermedad, pero que también deben existir otros factores diferentes necesarios para que la enfermedad se desarrolle (Petronis, 2004). De hecho, se ha observado que el riesgo de desarrollar esquizofrenia o patologías asociadas es similar para los hijos de los gemelos MZ aun en los casos en los que sólo uno de los gemelos haya desarrollado la enfermedad (Kringlen y Cramer, 1989).

Frente a los estudios con gemelos, los estudios de adopción permiten separar el efecto de los factores genéticos de los del ambiente familiar (Ingraham y Kety, 2000). Este tipo de estudios en esquizofrenia mostraron que la tasa de aparición de la enfermedad en niños adoptados con antecedentes de psicosis en su familia biológica y no en su familia adoptiva era mayor que en niños adoptados sin antecedentes de psicosis en su familia biológica, lo que indica que el riesgo a padecer la enfermedad

“viaja” con la familia biológica, por lo que los factores ambientales tendrían un papel modulador, potenciador o silenciador, de variantes de susceptibilidad genética.

Estos estudios han permitido confirmar que la esquizofrenia es claramente una enfermedad genética y su heredabilidad se ha estimado en torno al 80% (Cardno et al., 1999; Sullivan et al., 2003), lo que la convierte en una de las patologías complejas con mayor heredabilidad (Uher, 2009). De todos modos, a diferencia de organismos modelo, donde se pueden realizar los cruces necesarios bajo condiciones controladas, los cálculos de heredabilidad en humanos son complicados, por lo que hay que interpretarlos con cierta prudencia. Por ejemplo, un estudio más reciente basado en registros multigeneracionales en familias suecas ha dado lugar a una estimación de la heredabilidad de la esquizofrenia y el trastorno bipolar de 64% y 59% respectivamente (Lichtenstein et al., 2009), es decir, considerablemente más baja que las anteriores estimas. En todo caso se trata de una enfermedad con una alta heredabilidad.

El patrón de transmisión de la enfermedad observado en estos estudios encaja con el modelo de enfermedad multifactorial compleja. Según este modelo, su aparición provendría de la coincidencia en una misma persona de una combinación de factores de predisposición genéticos, ambientales y tal vez de interacción entre ellos, sin que la presencia aislada de cada uno de estos factores sea suficiente para determinar la aparición del trastorno y sin que ninguno de los factores sea imprescindible para su desarrollo. Esto hace que cada variante de susceptibilidad individualmente tenga, en principio, un poder diagnóstico o predictivo muy limitado (Kraft y Hunter, 2009).

Además, cada individuo podría enfermar a causa de una combinación particular de estos factores (Owen et al., 2005). A diferencia de otras enfermedades complejas, hasta el momento no se ha identificado ninguna familia en la que la enfermedad segregue de modo mendeliano, por lo que parece poco probable que existan formas mendelianas de esquizofrenia.

A pesar de mostrar agregación familiar, la mayor parte de los casos de esquizofrenia en la población general son esporádicos, es decir, no tienen un familiar de primer, segundo o tercer grado afectado. La aparente dicotomía entre el número de casos esporádicos y los que se dan con historia familiar de afectados se explica porque en una enfermedad que tiene una prevalencia poblacional del 1% y una heredabilidad de entre el 60 y el 80%, como es el caso, hay una probabilidad de aparición de casos esporádicos mayor del 60% (Yang et al., 2010).

2.2 Diseño de un estudio genético

Una vez que se confirma la existencia de un componente genético de una enfermedad, lo siguiente es identificar variantes genéticas concretas que puedan estar

relacionadas con la causa de esa enfermedad. Los humanos compartimos en torno al 99,9% de la secuencia genética. El 0,1% restante constituye la base de la variabilidad humana y, dentro de ella, la diferente susceptibilidad a padecer un trastorno. Así pues, el estudio genético de una enfermedad heredable pretende identificar qué parte de esta variabilidad es la que contribuye a su aparición.

Existen diferentes estrategias para la identificación de estos factores genéticos de susceptibilidad, algunas de las cuales se comentarán en esta introducción. El éxito de cada una de ellas depende enormemente de la arquitectura genética o espectro alélico de la enfermedad, que se refiere a: el número de alelos de riesgo que contribuyen a su desarrollo -que puede incluir múltiples alelos de un mismo gen o de genes diferentes-, la frecuencia en la población de cada uno de los alelos de riesgo, el efecto de cada alelo de riesgo y el modo en que los alelos actúan en conjunto -de modo aditivo o con interacción-. En el caso de la esquizofrenia, y en general de las enfermedades complejas, esta información es desconocida -los datos obtenidos de los estudios de epidemiología genética clásica comentados son compatibles con más de un modelo de enfermedad (Uher, 2009) y los estudios teóricos de genética de poblaciones tampoco permitieron obtener una conclusión sobre esto (Pritchard, 2001; Reich y Lander, 2001; Wray y Visscher, 2010)-, por lo que no está claro qué estrategia de análisis podría ser la más adecuada.

2.3 Estudios de ligamiento

Una de las primeras estrategias empleadas para la búsqueda de variantes de susceptibilidad a padecer esquizofrenia fueron los estudios de ligamiento en familias (Risch, 1990b), que identifican regiones cromosómicas con alta probabilidad de albergar variantes implicadas en la aparición de un trastorno. A grandes rasgos, en un estudio de ligamiento se evalúan marcadores espaciados a lo largo del genoma en los miembros de una o varias familias con más de un individuo afecto y se investiga si alguno de los alelos de los marcadores cosegrega con la enfermedad a través de las genealogías.

Esta estrategia es eficaz si el riesgo de la variante o variantes causales es elevado y si existen genes de efecto mayor en los que se localizan las variantes de susceptibilidad. Esto hace que los estudios de ligamiento pierdan sensibilidad si existe heterogeneidad de locus (existencia de varios loci implicados en la enfermedad), pero que sean robustos frente a la heterogeneidad alélica (existencia en el mismo gen de distintos alelos que causan susceptibilidad a la patología). Otra ventaja de estos estudios es que se analizan marcadores situados a lo largo del genoma, lo que resulta muy útil en enfermedades con etiología desconocida, como la esquizofrenia. Las regiones identificadas en estos estudios suelen ser largas y puede incluir varias

decenas de genes, por lo que una vez localizada una señal de ligamiento queda el trabajo de rastrear la zona para identificar la localización exacta de la variante causal responsable de la señal.

La aplicación de esta estrategia en esquizofrenia dio lugar a la identificación de numerosos loci asociados al trastorno en familias particulares (MacIntyre et al., 2003; Muir et al., 2006; Williams et al., 2009), pero ninguno de ellos se reprodujo de manera significativa en posteriores estudios con otras familias (Ng et al., 2009; Uher, 2009). A pesar de esto, la estrategia supuso un avance en cuanto a la descripción de la arquitectura genética de la enfermedad, ya que permitió descartar la posibilidad de que las variantes de riesgo de todos los individuos estuviesen restringidas a pocos genes de efecto mayor. Además evidenciaron que, tal y como apuntaban predicciones teóricas (Risch, 1990a), es poco probable que existan loci de susceptibilidad con un efecto muy elevado sobre el fenotipo.

Se han hecho varios meta-análisis de los estudios de ligamiento en esquizofrenia (Badner y Gershon, 2002; Lewis et al., 2003; Ng et al., 2009). Esta aproximación, que analiza conjuntamente resultados de varios trabajos, permite amplificar señales de asociación pequeñas pero comunes a varios estudios que individualmente pueden no destacar, mientras que señales que son fuertes en un único estudio se debilitan. El meta-análisis más reciente y completo incluye 32 estudios de ligamiento en esquizofrenia con 3.255 familias y un total de 7.413 casos de la enfermedad o trastornos relacionados y localiza la mayor evidencia de ligamiento en la región 2q, seguida de las regiones 5q (142-168) y 2q (103-134); otra señal aparece en 8q cuando el meta-análisis se restringe a estudios de muestras europeas. Señales menos significativas aparecen en 1, 2q, 3q, 4q, 5q, 8p y 10q (Ng et al., 2009). Otros meta-análisis identificaron señales de ligamiento en 8p, 13q y 22q (Badner y Gershon, 2002) o en 2p12-22.1 (Lewis et al., 2003). El exceso de señales de ligamiento puede reflejar la existencia de regiones genómicas enriquecidas en variantes comunes de efecto moderado o en variantes raras de efecto mayor (Williams et al., 2009).

El mapeado fino de estas regiones detectadas por ligamiento permitió la identificación de genes candidatos posicionales y, para muchos de ellos, estudios posteriores encontraron también resultados positivos de asociación. Entre estos genes podemos destacar a *NRG1*, *DTNBPI*, *DAOA*, *PPP3CC* o *COMT* (Kirov et al., 2005; Owen et al., 2005; Riley y Kendler, 2006). Además, los resultados de estos estudios de ligamiento sugieren la existencia de susceptibilidad genética compartida con el trastorno bipolar (Collier y Li, 2003).

2.4 Estudios de asociación

2.4.1 Hipótesis enfermedad común / variante común

El hecho de que las enfermedades complejas fuesen comunes en la población hizo pensar que la razón de esta alta prevalencia podría ser que los factores genéticos de predisposición estuviesen presentes con relativa frecuencia en las poblaciones humanas. En base a esta idea se formuló la hipótesis enfermedad común / variante común (Lander, 1996).

En las enfermedades que aparecen a una edad temprana y se asocian a una fertilidad reducida, como la esquizofrenia, en caso de que existan variantes comunes de susceptibilidad, es de esperar que confieran un riesgo pequeño, ya que variantes con un efecto elevado sobre el fenotipo serían eliminadas de la población mediante selección purificadora y no llegarían a ser comunes. Estudios teóricos demostraron que para el estudio de variantes que predisponen a padecer una enfermedad de manera moderada o baja, los estudios de asociación en poblaciones constituían una estrategia más potente que los estudios de ligamiento (Risch y Merikangas, 1996). La culminación del Proyecto Genoma Humano, que supuso un aumento en el conocimiento de variantes a lo largo del genoma y el hecho de que la recogida de muestras para este tipo de estudios sea mucho más simple que para un estudio de ligamiento en familias, provocaron que la comunidad científica centrara el estudio genético de las enfermedades complejas en la búsqueda de variantes de susceptibilidad comunes en la población mediante estudios de asociación caso-control, relegando los estudios de ligamiento o de variantes raras a un segundo plano.

En líneas generales, en un estudio de asociación caso-control se pretende la identificación de variantes alélicas que están presentes en individuos con un fenotipo concreto a una frecuencia significativamente diferente que en individuos sin ese fenotipo. Para determinar si la frecuencia es significativamente diferente entre ambos grupos se emplea una prueba estadística, como la χ^2 o la regresión logística, en la que la hipótesis nula es la igualdad de frecuencias alélicas o genotípicas. El grupo de casos y controles deben pertenecer a la misma población para evitar problemas de estratificación y no estar emparentados. Si se observa que la diferencia de frecuencias es significativa se concluye que existe asociación estadística en la muestra, pero para confirmar su relación con la enfermedad y descartar que el resultado se deba al azar es necesario obtener el mismo resultado en una o varias muestras independientes (Chanock et al., 2007). Los estudios de asociación, a diferencia de los estudios de ligamiento, pierden sensibilidad si existe heterogeneidad alélica, pero son robustos frente a la heterogeneidad de locus. Para cuantificar la fuerza de una asociación, es decir, el efecto que tiene la variante sobre el fenotipo, se calcula la Odds Ratio (OR, riesgo conferido por la variante): sea a el número de casos que presentan la variante, b

el número de controles que presentan la variante, c el número de casos que no presentan la variante y d el número de controles que no presentan la variante, la OR para la variante sería:

$$OR = \frac{\frac{a}{c}}{\frac{b}{d}}$$

Si el valor OR es menor que 1, la variante conferiría protección; si es mayor que 1, la variante incrementaría el riesgo a sufrir la enfermedad.

2.4.1.1 Estudios con genes candidatos

Los estudios tradicionales de asociación se centran en el análisis de uno o varios genes candidatos, por lo que resulta necesario tener una hipótesis de trabajo que guíe la selección. Un criterio empleado para esta selección es el de la funcionalidad: genes que por su función, lugar de expresión o ruta metabólica en la que participan, están relacionados con procesos que podrían ser relevantes para la enfermedad; serían genes candidatos funcionales. En el caso de la esquizofrenia, por el momento no se conocen vías moleculares o procesos fisiopatológicos específicos de la enfermedad que dirijan de manera clara esta selección. Las diferentes hipótesis formuladas en relación al origen de la enfermedad centran la selección principalmente en genes que participan en diferentes sistemas de neurotransmisores, como el dopaminérgico, serotoninérgico o glutamatérgico, o en genes implicados en el desarrollo del cerebro o en la transmisión sináptica; sin embargo, teniendo en cuenta que aproximadamente el 85% de los genes conocidos se expresan en el cerebro y podrían intervenir en estas funciones (Kang et al., 2011), la lista de genes candidatos funcionales es inmensa.

Una alternativa al criterio anterior consiste en seleccionar genes que por su localización en el genoma tienen elevada probabilidad de estar implicados en el desarrollo de la enfermedad, por ejemplo, genes localizados en regiones identificadas mediante estudios de ligamiento o afectados por aberraciones cromosómicas asociadas a la enfermedad; serían genes candidatos posicionales (Owen et al., 2005; Sullivan, 2005).

Los polimorfismos empleados de manera habitual en los estudios de asociación son los SNPs (del inglés, *Single Nucleotide Polymorphism*), debido a su abundancia en el genoma y a la facilidad tecnológica para su detección. Estos polimorfismos son cambios nucleotídicos de una única base que implican el reemplazo de un nucleótido por otro, o la delección o inserción de un nucleótido. Cada SNP marca una posición única en el genoma y se localizan a lo largo de toda la secuencia de ADN, pero su

distribución no es homogénea: existen diferencias en cuanto a su densidad entre regiones exónicas, intrónicas e intergénicas. La mayor parte de estos polimorfismos son neutros -sin repercusión sobre funciones o sobre el fenotipo- pero en algunos casos dan lugar a cambios funcionales que pueden incrementar el riesgo a padecer enfermedades complejas. Además del estudio de SNPs individuales, es interesante también el análisis de haplotipos (combinación de variantes alélicas en un cromosoma).

En la secuencia de cada gen puede haber multitud de SNPs por lo que, una vez seleccionados los genes, se decidirá qué polimorfismos se incluirán en el estudio. Esta selección puede basarse en el “método directo” (Risch, 2000), en el que la selección está guiada por la hipótesis de causalidad de los polimorfismos, es decir, se analizan únicamente SNPs (u otro tipo de polimorfismo) con alta probabilidad de ser funcionales y por tanto de estar implicados en la etiología de las enfermedades. Existen herramientas bioinformáticas *in silico* que permiten estimar la probabilidad de que un SNP sea funcional. Dentro de un gen, la clase de SNP con mayor probabilidad de ser funcional es la de los SNPs no sinónimos: sustituciones nucleotídicas que ocurren en la región codificante del gen y, o bien provocan el reemplazamiento de un aminoácido por otro en la proteína codificada, con posible efecto en su actividad (especialmente si el cambio aminoacídico es entre dos aminoácidos de propiedades físico-químicas muy diferentes), o bien dan lugar a un codón de parada, provocando la terminación prematura de la traducción y la aparición de una proteína truncada. Otras clases de SNPs pueden alterar el funcionamiento de un gen, incluso SNPs aparentemente neutros, como los SNPs intrónicos o los SNPs sinónimos (sustituciones en la región codificante que no conllevan un cambio aminoacídico): pueden alterar el empaquetamiento del gen, su expresión (por ejemplo, SNPs localizados en regiones promotoras de unión de factores de transcripción, en las que el cambio nucleotídico puede interrumpir o crear nuevos sitios de unión, variando la afinidad entre las dos moléculas), el procesamiento del ARN mensajero (afectando el corte de intrones y empalme de exones) o su exportación al citoplasma (afectando a su estabilidad, por ejemplo) (Capon et al., 2004). El problema en estos casos es que la funcionalidad del cambio puede no ser tan evidente como la de un SNP no sinónimo, por lo que su interpretación biológica es más difícil (Hindorff et al., 2009). Por tanto, este método presenta la limitación de que no siempre se conocen todas las variantes funcionales de un gen.

El método alternativo es el indirecto, o mapeo por desequilibrio de ligamiento en el que se analizan variantes sin considerar si son o no funcionales. El desequilibrio de ligamiento se define como la presencia de dos alelos en el mismo cromosoma a una frecuencia significativamente diferente de la esperada por sus frecuencias individuales. Este fenómeno hace que si un SNP incrementa el riesgo a padecer una enfermedad, habrá una asociación estadística en la población entre la enfermedad y ese SNP (asociación directa), y entre la enfermedad y SNPs en desequilibrio de ligamiento con el causal (asociación indirecta debido al desequilibrio de ligamiento). La estrategia

indirecta asume que de entre todos los polimorfismos presentes en cada gen no se puede predecir con seguridad cuáles son funcionales; por tanto, se genotipan varios SNPs a lo largo del gen con la intención de que alguno de ellos sea el causal o de que al menos alguno de los genotipados esté en desequilibrio de ligamiento con él y dé lugar a una señal de asociación. La señal de asociación indirecta será más leve, tanto más cuanto menor sea el desequilibrio de ligamiento entre las dos variantes.

Esta correlación estructural de desequilibrio de ligamiento es compleja y varía a lo largo del genoma en función de diferentes parámetros como la tasa de recombinación, tasa de mutación, selección, migración, conversión génica o deriva genética (Tishkoff y Verrelli, 2003; Weiss y Clark, 2002; Zondervan y Cardon, 2004). No todos los SNPs son igualmente informativos, por lo que para que la estrategia resulte eficaz es esencial una buena selección de SNPs marcadores de la región (o tagSNPs), para lo cual se necesitan conocer las relaciones de desequilibrio de ligamiento de la región en la población que se está estudiando. En respuesta a esta necesidad, en el año 2002 surgió el Proyecto Internacional HapMap, cuyo objetivo era caracterizar los patrones de desequilibrio de ligamiento a lo largo del genoma humano en distintas poblaciones y hacer que esta información estuviese disponible de forma gratuita para el dominio público para permitir abordar de manera sistemática la búsqueda de variantes comunes de susceptibilidad a padecer enfermedades complejas. Este proyecto supuso un salto cualitativo en el avance de los estudios de asociación (Altshuler et al., 2010; The International HapMap Consortium, 2005; Frazer et al., 2007). Existen diferentes medidas para estimar el grado de desequilibrio de ligamiento entre dos variantes (Devlin y Risch, 1995). Las más comunes son D' y r^2 , que se calculan:

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

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

En las que A y B son los alelos de mayor frecuencia de dos SNPs, a y b los menores, $D = \text{frec}(A \cap B) - \text{frec}(A) \cdot \text{frec}(B)$ y D_{\max} el máximo valor que puede alcanzar D en función de las frecuencias alélicas concretas.

Ambas medidas tienen valores de 0 a 1, aunque su interpretación es distinta: si el alelo de menor frecuencia de un SNP está siempre asociado a uno de los dos alelos del otro SNP, el valor de D' será de 1. r^2 sólo será 1 si cada uno de los alelos de un SNP está asociado a un único alelo del otro SNP (mismas frecuencias). Por tanto, r^2 es una medida más restrictiva que D' .

Siguiendo esta estrategia se han analizado miles de variantes comunes de unos 1.000 genes para investigar su posible implicación en la etiología de la esquizofrenia (Allen, N.C., 2008). Estos estudios dieron lugar a multitud de asociaciones positivas en las muestras iniciales, pero cuando se intentaba reproducir el resultado en muestras

independientes las asociaciones desaparecían o se encontraba asociación positiva pero con diferentes alelos de la misma variante o incluso con la misma variante pero en dirección contraria. Un ejemplo de esto es el caso del gen de la disbindina (*DTNBPI*), para el que se encontraron asociaciones con cada uno de los 5 haplotipos del mismo bloque haplotípico (Mutsuddi et al., 2006). La falta de reproducibilidad ha sido la norma en los estudios de asociación con genes candidatos y ninguna de las asociaciones descritas puede considerarse consistentemente establecida.

La dificultad para encontrar factores de riesgo comunes muy posiblemente esté relacionada con que el fenotipo englobado bajo la definición actual de esquizofrenia probablemente no se corresponda con un grupo homogéneo desde el punto de vista biológico. Además de esto, existen diversos factores o situaciones que pueden dificultar la reproducibilidad de un resultado positivo real (error tipo II) o provocar la aparición de asociaciones falsas (error tipo I) (Cardon y Bell, 2001).

Un mal diseño experimental en un intento de reproducción de un resultado podría dar lugar a un error tipo II. Por ejemplo, falta de potencia estadística por tamaño insuficiente de la muestra con la que se pretende reproducir la asociación; el tamaño muestral necesario para identificar una asociación depende en gran medida del riesgo asociado a la variante causal y su frecuencia: cuanto menor sean éstos, mayor tamaño de muestra se necesitará para alcanzar una probabilidad *a priori* suficiente para detectar la asociación. Este fenómeno puede estar agravado por el hecho de que los primeros resultados positivos tienden a sobreestimar el riesgo (deducido al comparar la OR de un primer estudio de asociación de una variante causal con la OR del meta-análisis de estudios posteriores) (Ioannidis, 2008). Otro factor que podría dificultar la reproducción de una asociación verdadera sería que la estructura de desequilibrio de ligamiento de la región asociada de la población inicial fuese diferente de la que existe en la segunda población; esto es un problema relacionado con la estrategia de mapeo por desequilibrio de ligamiento y es esperable que ocurra en alguna ocasión; indicaría que la variante genotipada en el estudio inicial no era la variante causal. Por último, otra posibilidad sería que el riesgo asociado a la variante en la población inicial dependiese de una interacción con otro factor, ambiental o genético (epistasia) y que esta interacción no se diese en la nueva población considerada en el análisis (porque no esté presente el segundo factor o porque está a una frecuencia más baja).

Una proporción de las asociaciones descritas podrían ser en realidad falsas (error tipo I). Al realizar estudios en los que se analizan múltiples polimorfismos se espera que por azar un porcentaje de las asociaciones encontradas sean falsas. Para minimizar esto es necesario adaptar el nivel de significación al número de pruebas realizadas en cada estudio. Este fenómeno puede verse incrementado al estudiar genes candidatos funcionales sin una base biológica sólida. Otro factor que propiciaría la aparición de asociaciones falsas sería la estratificación poblacional (origen étnico mayoritario en casos diferente al de controles) que podría dar lugar a que un rasgo genético característico de una etnia pueda interpretarse como un factor de riesgo identificable

en un estudio de asociación. Por último, errores en el genotipado o el sesgo en las publicaciones (tendencia a publicar resultados positivos independientemente de la calidad de los mismos) podrían incrementar el número errores de tipo I (Ioannidis, 2008).

Una iniciativa para ordenar la gran cantidad de estudios genéticos de asociación publicadas en relación a la esquizofrenia y resaltar los más relevantes fue la creación de “SchizophreniaGene”, una base de datos pública, de acceso libre y actualizada regularmente (Allen et al., 2008). En esta base de datos se muestran los datos genotípicos de las muestras y su origen, se hacen meta-análisis de todos los polimorfismos con datos suficientes y se ordenan los genes por nivel de evidencia en su implicación en la etiología de la enfermedad. Actualmente los genes identificados mediante estudios de asociación de genoma completo (descritos en el siguiente apartado) ocupan los primeros puestos. La implicación del resto de genes sigue siendo controvertida.

2.4.1.2 Estudios de genoma completo

El escaso éxito de los estudios de asociación con genes candidatos en la identificación de variantes comunes de susceptibilidad a padecer enfermedades complejas, el avance en las tecnologías de genotipado a gran escala y el aumento en el número de polimorfismos conocidos, propiciaron la aparición de los estudios de asociación del genoma completo, o GWAS (del inglés, *Genome-wide Association Studies*) (Hirschhorn y Daly, 2005). El objetivo de estos estudios es analizar sistemáticamente la mayor parte de la variabilidad común de cada muestra a partir del genotipado de una proporción de SNPs (estrategia de mapeo por desequilibrio de ligamiento), de modo que la búsqueda de variantes de susceptibilidad no dependa de una hipótesis previa ni de los genes candidatos que se elijan, sino que sean los propios resultados los generadores de hipótesis. Esto permite identificar genes cuya relación con la enfermedad se desconocía y no se sospechaba, que mediante la estrategia de genes candidatos funcionales sería poco probable descubrir (Hirschhorn y Daly, 2005). Esto resulta particularmente útil para el estudio genético de las enfermedades psiquiátricas, para las que todavía no se conoce su patofisiología. Para ello se emplean paneles que permiten el genotipado de más de un millón de SNPs marcadores localizados a lo largo del genoma simultáneamente en una misma muestra.

El éxito de los estudios de asociación del genoma completo se basa en la hipótesis enfermedad común / variante común, ya que la capacidad para identificar asociaciones con variantes poco frecuentes es escasa, aun estando la variante incluida en el panel de marcadores, a no ser que ésta confiriese un riesgo muy elevado y el tamaño muestral fuese suficiente para alcanzar potencia estadística.

Debido al elevado número de comparaciones realizadas en cada estudio, para evitar falsos positivos por pruebas múltiples se ha establecido un umbral general de

significación genómica de para considerar robusta una asociación, $p < 5 \times 10^{-8}$, que responde al número total de SNPs interrogados en cada estudio de genoma completo (Dudbridge y Gusnanto, 2008). Debido a esta corrección y al bajo riesgo asociado a cada variante de susceptibilidad ($< 1,3$ en esquizofrenia), se necesitan muestras de entre 2.000 y 3.000 casos y el mismo número de controles para alcanzar una probabilidad *a priori* del 80% de identificar asociaciones verdaderas. Estudios con tamaños muestrales /inferiores tendrán escasa potencia estadística a no ser que el riesgo de la variante sea elevado.

Los paneles de genotipado incluyen SNPs localizados a lo largo del genoma, incluyendo regiones de funcionalidad desconocida, como regiones intergénicas o intrónicas; en los casos en los que se encuentra asociación con SNPs de estas regiones, resulta difícil, por el momento, entender la asociación desde un punto de vista biológico.

Los primeros estudios de genoma completo en esquizofrenia encontraron asociaciones (Kirov et al., 2009b; Lencz et al., 2007; Need et al., 2009; O'Donovan et al., 2008; Shifman et al., 2008; Sullivan et al., 2008), pero debido al limitado tamaño de muestra empleado sólo una alcanzó significación genómica, la asociación identificada con *ZNF804A*, al unir las muestras de esquizofrenia a muestras de trastorno bipolar (O'Donovan et al., 2008).

En el año 2009 se publicaron otros tres GWAS (ISC, 2009; Shi et al., 2009; Stefansson et al., 2009) y un meta-análisis de los tres (Shi et al., 2009) en los que se incrementó el tamaño de la muestra empleada y se logró la identificación de asociaciones con significación genómica que constituyeron las evidencias más robustas de asociación de variantes comunes con la enfermedad hasta ese momento. Además, se descubrió la existencia de múltiples de variantes de susceptibilidad asociadas a un riesgo casi indetectable (componente poligénico) (ISC, 2009).

Desde entonces se han publicado nuevos estudios de genoma completo (Alkelai et al., 2011a; Alkelai et al., 2011b; Athanasiu et al., 2010; Chen et al., 2011; Ikeda et al., 2011; Ma et al., 2011; Rietschel et al., 2011; Steinberg et al., 2011a; Yamada et al., 2011). El *Psychiatric Genome-Wide Association Study Consortium* ha publicado recientemente un “mega” meta-análisis en el que se analizaron de manera conjunta datos de varios de estos estudios de genoma completo, con un total de 29.839 individuos de origen europeo (Ripke et al., 2011). Paralelo a este estudio, el mismo consorcio hizo otro mega meta-análisis con los estudios de genoma completo de trastorno bipolar (Sklar et al., 2011) y se realizó un análisis en el que se combinaron las muestras de los dos trastornos, de lo que se obtuvieron nuevas evidencias de la vulnerabilidad genética compartida entre ambos trastornos.

A continuación se describen algunos genes con asociación consistente a esquizofrenia.

Tabla 1. Principales resultados de los estudios de genoma completo ($p < 5 \times 10^{-8}$) en esquizofrenia en poblaciones de origen europeo.

Primer autor y año	Muestra I N ca / N co	Muestra II ¹ N ca / N co	Región	Gen	menor valor p	OR
O'Donovan, 2008 ²	479/2937	6666/9897	2q32.1	<i>ZNF804A</i>	2×10^{-7}	1,12
Purcell, 2009	3322/3587	4692 / 15493	6p22.1	Región MHC	$9,54 \times 10^{-9}$	0,82
Stefansson, 2009	2663/13498	10282 / 21093	6p22.1	Región MHC (<i>PRSS16</i>)	$1,4 \times 10^{-12}$	1,16
			6p21.32	Región MHC (<i>NOTCH</i>)	$2,3 \times 10^{-10}$	1,19
			11q24.2	<i>NRGN</i>	$2,4 \times 10^{-9}$	1,15
			18q21.2	<i>TCF4</i>	$4,1 \times 10^{-9}$	1,23
Shi, 2009	2681/2653	5327 / 16424	6p22.1	Región MHC (<i>PRSS16</i>)	$9,54 \times 10^{-9}$	1,28
			6p21.32	<i>HLA-DQA1</i>	$6,88 \times 10^{-8}$	0,90
Psychiatric GWAs Consortium, 2011	9394 / 12462	8442 / 21397	6p22.33	<i>TRIM26</i>	$2,18 \times 10^{-12}$	1,15
			1p21.3	<i>MIR137</i>	$1,59 \times 10^{-11}$	1,12
			10q24.32	<i>CNNM2</i>	$1,82 \times 10^{-9}$	1,10
			8p23.2	<i>CSMD1</i>	$4,14 \times 10^{-8}$	1,11
			10q24.33	<i>NT5C2</i>	$1,11 \times 10^{-8}$	1,15
			18q21.2	<i>CCDC68 - TCF4</i>	$2,6 \times 10^{-10}$	1,09
			2q32.3	<i>RPS17P8 - GLULP6</i>	$4,65 \times 10^{-8}$	1,2
Steinberg, 2011	7946 / 19036	10260 / 23500	2p15.1	<i>VRK2</i>	$1,9 \times 10^{-9}$	1,09
			18q21.2	<i>CCDC68 - TCF4</i>	$7,8 \times 10^{-9}$	1,09

¹Muestra multicéntrica empleada en la fase de confirmación de los resultados de la Muestra I

²La asociación alcanza significación genómica al combinar las muestras de esquizofrenia con muestra de trastorno bipolar

ZNF804A

O'Donovan y colaboradores identificaron una señal de asociación con una variante intrónica de este gen, el SNP rs1344706, en muestras de pacientes esquizofrénicos (valor p de meta-análisis = $1,95 \times 10^{-7}$; OR = 1,12) (O'Donovan et al., 2008). Al combinar estas muestras con muestras de pacientes con trastorno bipolar la señal de asociación se incrementó en dos órdenes de magnitud, alcanzando significación genómica ($p = 9,96 \times 10^{-9}$; OR = 1,12). Desde entonces la asociación de esta variante a esquizofrenia y trastorno bipolar ha sido consistentemente reproducida en estudios posteriores (Riley et al., 2010; Steinberg et al., 2011b; Williams et al., 2011; Zhang et al., 2011b). Esslinger y colaboradores (Esslinger et al., 2009) investigaron el efecto de esta variante en la actividad de diferentes regiones cerebrales durante procesos cognitivos en 115 personas sanas mediante estudios genéticos de neuroimagen, que permiten asociar variantes genéticas con funciones y estructuras concretas del cerebro. Mediante esta estrategia encontraron que el alelo de riesgo daba lugar a una reducción en la conectividad entre la corteza prefrontal dorsolateral dentro de y entre cada hemisferio, mostrando un patrón similar al que se ha descrito en pacientes (Meyer-Lindenberg et al., 2005); además vieron que el alelo de riesgo incrementaba la conectividad entre la amígdala y diferentes áreas reguladoras límbicas y prefrontales. Esta misma variante se ha asociado con diferencias en habilidades cognitivas (como memoria episódica o a corto plazo) en pacientes esquizofrénicos, no en controles (Walters et al., 2010).

Región del Complejo Principal de Histocompatibilidad (Major Histocompatibility Complex, MHC)

La idea de que la región MHC podría estar relacionada con el desarrollo de esquizofrenia es ya antigua (Worden et al., 1976) y algunos estudios identificaron asociaciones con esta región, no reproducidas posteriormente (Lewis et al., 2003). Recientemente, los estudios de genoma completo han permitido confirmar esta asociación y actualmente el locus 6p22.1, situado en esta región, es una de las asociaciones más robustas (ISC, 2009; Ripke et al., 2011; Shi et al., 2009; Stefansson et al., 2009). La mayor parte de las señales identificadas en la región corresponden a SNPs localizados en regiones intergénicas, pero no se puede descartar que provengan de alguna variante localizada en algún gen cercano. Esta región se caracteriza por un elevado grado de desequilibrio de ligamiento, lo que dificulta enormemente la identificación de la variante o las variantes causales responsables de la señal de asociación. De todos modos, toda la región está muy conservada por lo que es posible que las regiones intergénicas tengan alguna función por el momento desconocida (Kleinjan y van Heyningen, 2005).

La región MHC incluye numerosos genes implicados en diferentes funciones entre las que predominan las relacionadas con el sistema inmune. Si la señal proviene

de alguno de estos genes, su implicación en la enfermedad podría explicar la observación de que nacer en invierno -época más propicia para infecciones- incrementa el riesgo de padecer esquizofrenia.

Otra explicación podría estar relacionada con el siguiente hecho. Recientemente se ha visto que, insospechadamente, algunas moléculas que originalmente fueron identificadas en el sistema inmune también tienen funciones, hasta el momento desconocidas, en el sistema nervioso central. Estas moléculas, que participan en el sistema inmune innato (por ejemplo, citoquinas pro-inflamatorias y componentes de la cascada del complemento) o adaptativo (por ejemplo, componentes del MHC), se expresan en el sistema nervioso y juegan un papel regulador importante, no inmunológico, en funciones normales del cerebro durante el desarrollo y en la edad adulta, participando por ejemplo en la formación y posterior plasticidad sináptica (Boulangier, 2009). Entre otras funciones, parecen ser reguladoras clave de circuitos glutamatérgicos (Fourgeaud y Boulangier, 2010).

En esta región existen también genes no relacionados con el sistema inmune. Por ejemplo, el gen *NOTCH4*, asociado previamente a la enfermedad y con función vinculada al neurodesarrollo (Wang et al., 2006).

TCF4

En este gen se ha identificado una variante intrónica asociada a esquizofrenia, el SNP rs9960767 (Ripke et al., 2011; Stefansson et al., 2009). Dado que no hay mucho desequilibrio de ligamiento en torno a la señal, parece que la señal viene del propio gen. A través de la medida de la inhibición del prepulso de la respuesta al sobresalto, se ha visto que el alelo de riesgo de esta variante se asocia a una reducción en la sincronización senso-motora (Lennertz et al., 2011). Recientemente, Steinberg y colaboradores identificaron otro polimorfismo cercano a *TCF4*, localizado a 400kb de la variante intrónica, con una fuerte señal de asociación ($p = 7,8 \times 10^{-9}$) independiente de la anterior. *TCF4* parece un buen candidato a intervenir en la etiología de la esquizofrenia ya que es un factor de transcripción neuronal esencial para el desarrollo del cerebro, específico de la neurogénesis (Flora et al., 2007; Gulacsi y Anderson, 2008). Además, mutaciones y deleciones en este gen causan el Síndrome de Pitt-Hopkins, un trastorno del neurodesarrollo caracterizado por síntomas como el retraso mental y motor severos, que incluyen ausencia de habla, microcefalia, dimorfismo facial y epilepsia (Blake et al., 2010; Brockschmidt et al., 2007). Se ha descrito también una mutación en este gen que causa retraso mental sin estar asociado a otros rasgos del síndrome Pitt-Hopkins (Kalscheuer et al., 2008). Es interesante destacar que deleciones y mutaciones en *CNTNAP2* y *NRXN1*, otros genes asociados a esquizofrenia, al espectro autista y a epilepsia (Friedman et al., 2008; Gejman et al., 2011), dan lugar a fenotipos similares al síndrome de Pitt-Hopkins (Zweier et al., 2009) lo que parece indicar mecanismos moleculares compartidos entre estos trastornos.

NRGN

Se han identificado dos asociaciones independientes correspondientes a los SNPs rs12807809 y rs548181 en el entorno de este gen (Ripke et al., 2011; Stefansson et al., 2009). *NRGN* codifica una proteína quinasa postsináptica con un papel importante en la función y plasticidad sináptica, participando en la ruta de señalización calcio-calmodulina; en esta ruta también interviene el receptor NMDA de glutamato, por lo que su implicación en la etiología de la enfermedad podría estar relacionada con la hipótesis glutamatérgica (Wang et al., 2008; Zhong et al., 2009). Se ha visto que *NRGN* juega un papel en la potenciación a largo plazo, uno de los mecanismos celulares principales que subyace al aprendizaje y la memoria (Harrison y Weinberger, 2005; Huang et al., 2006; Zhabotinsky et al., 2006). Un estudio reciente sugiere que los diferentes alelos del SNP rs12807809 se relacionan con funcionamientos neuronales diferenciados en áreas específicas del cerebro relacionadas con la memoria episódica en personas sanas (Krug et al., 2011).

VRK2

Steinberg y colaboradores (Steinberg et al., 2011a) recientemente publicaron un trabajo en el que ampliaron su GWAS previo (Stefansson et al., 2009), analizando en muestras adicionales variantes que en este trabajo inicial obtuvieron un valor de asociación de $p < 1 \times 10^{-4}$. El resultado de esto fue, además de la reproducción de las asociaciones ya detectadas, la identificación de dos nuevas señales de asociación con significación genómica (tabla 1). Una de ellas se sitúa en el locus 2p15.1, el SNP rs2312147, localizado a 50kb del gen *VRK2* ($p = 1,9 \times 10^{-9}$; OR = 1,09). Este gen codifica una serina/treonina quinasa del grupo de las caseína quinasa I (Manning et al., 2002). Se ha sugerido que este gen podría participar en el mantenimiento de la arquitectura nuclear (Nichols et al., 2006) y en la prevención de la apoptosis (Li et al., 2006). *VRK2* interacciona con otra proteína, JIP1, que en células neuronales actúa como un factor anti-apoptótico en respuesta al estrés (Dong et al., 2005) y participa en el desarrollo axonal (Dajas-Bailador et al., 2008).

En el último mega meta-análisis llevado a cabo por el *Psychiatric GWAS Consortium* (Ripke et al., 2011) se identificaron 5 nuevos loci de susceptibilidad. De éstos, la señal más fuerte corresponde al SNP rs1625579, localizado en un intrón del gen *MIR137* que codifica un microARN regulador del desarrollo y maduración neuronal y de la neurogénesis adulta (Silber et al., 2008; Smrt et al., 2010; Szulwach et al., 2010).

En este estudio del *Psychiatric GWAs Consortium* se combinaron las muestras de esquizofrenia con muestras de trastorno bipolar -empleadas en un mega meta-análisis paralelo del mismo consorcio en trastorno bipolar (Sklar et al., 2011)- y se encontró asociación a ambos trastornos con significación genómica en tres loci:

CACNA1C

Varios estudios han identificado señales de asociación entre este gen y diferentes trastornos psiquiátricos, incluida la esquizofrenia (Curtis et al., 2011; Ferreira et al., 2008; Green et al., 2010; Nyegaard et al., 2010). La asociación identificada por el *Psychiatric GWAs Consortium* fue con el SNP rs10994359 ($p = 7,01 \times 10^{-9}$; OR = 1,11). Otra variante de este gen, el SNP rs1006737, se ha asociado con alteraciones en imágenes de resonancia magnética estructural y funcional (Bigos et al., 2010; Erk et al., 2010; Kempton et al., 2009; Krug et al., 2010). *CACNA1C* codifica la subunidad alfa del canal de calcio tipo L dependiente de voltaje. Los canales de calcio tipo L participan en la regulación génica de aspectos de la plasticidad neuronal y pueden tener efecto directo en la transcripción (Gomez-Ospina et al., 2006).

ANKK1

La señal de asociación identificada en este gen proviene del SNP rs10994359 ($p = 2,45 \times 10^{-8}$; OR=1,22). Al igual que *CACNA1C*, este gen ya se había asociado previamente a trastorno bipolar (Ferreira et al., 2008). *ANKK1* codifica una proteína adaptadora, ankyrina 3, identificada en los segmentos iniciales de los axones. Se ha descrito su participación en la regulación del ensamblaje de canales de sodio voltaje-dependientes (Kordeli et al., 1995; Poliak y Peles, 2003).

ITIH3-ITIH4

La tercera asociación corresponde a la región que contiene a *ITIH3-ITIH4*, codificadores de inhibidores inter- α (globulina) H3 y H4 (rs2239547; $p = 7,8 \times 10^{-9}$ OR=1,12). En esta región ya se habían descrito asociaciones con trastorno bipolar (Scott et al., 2009).

Componente poligénico

Además de la identificación de variantes comunes con significación genómica, los estudios de genoma completo han permitido descubrir que parte de la predisposición a padecer esquizofrenia y otras enfermedades complejas está originada por el efecto global de múltiples de variantes comunes de riesgo muy bajo presentes en un individuo, lo que se conoce como componente poligénico (ISC, 2009; Ripke et al.,

2011). El componente poligénico ya se había propuesto anteriormente (Gottesman y Shields, 1967), pero hasta ese momento no existían pruebas empíricas.

Así pues, los resultados acumulados hasta la fecha indican que individualmente cada variante identificada incrementa débilmente el riesgo a padecer la enfermedad, por lo que cada una tiene bajo valor diagnóstico y explica un porcentaje muy pequeño de la heredabilidad atribuida a esquizofrenia, lo cual no quiere decir que cada asociación encontrada no pueda ser útil para la comprensión de mecanismos presentes en el desarrollo de la enfermedad o para generar nuevas hipótesis sobre su origen.

2.4.2 Hipótesis enfermedad común / variante rara

La hipótesis alternativa, o complementaria, a la existencia de variantes comunes de predisposición a padecer enfermedades complejas es la existencia de variantes raras de susceptibilidad, como propone la hipótesis enfermedad común / variante rara (Manolio et al., 2009; Pritchard, 2001). Según esta hipótesis, mutaciones de baja frecuencia, raras o incluso exclusivas de familias o casos concretos, serían las responsables de la diferente susceptibilidad a padecer enfermedades complejas. Dado que las enfermedades comunes se consideran de origen multifactorial, en general, el efecto de este tipo de mutaciones no sería determinante sobre el fenotipo, como ocurre en las enfermedades monogénicas.

Desde los inicios de la investigación genética de la enfermedad, y en general de las enfermedades complejas, ha existido un intenso debate en relación a la contribución relativa de alelos raros o comunes en la susceptibilidad genética. En el caso de la esquizofrenia existen diferentes argumentos a favor de la hipótesis de las variantes raras (Keller y Miller, 2006; McClellan et al., 2007; Uher, 2009). El principal es que la enfermedad se manifiesta en un momento temprano de la edad reproductiva y está asociada a una fertilidad reducida, por lo que se puede decir que el fenotipo está sometido a selección negativa. En este contexto, variantes asociadas a un riesgo elevado sufrirían esta selección y tendrían pocas posibilidades de alcanzar frecuencia elevada en la población, manteniéndose a muy baja frecuencia o desapareciendo en poco tiempo -por lo que en general serían variantes *de novo* o de origen reciente- (Rees et al., 2011). En caso de mutaciones ligeramente deletéreas la selección no sería tan fuerte, de modo que ésta impediría que llegasen a ser comunes en la población pero no llegaría a eliminarlas totalmente, pudiendo segregarse en la población a frecuencia baja y contribuir a la susceptibilidad a padecer enfermedades comunes en algunas personas (Gorlov et al., 2008; Kryukov et al., 2007).

A pesar de que por el momento la arquitectura genética de la esquizofrenia continúa sin conocerse totalmente, existe una lista en aumento de variantes raras que

incrementan el riesgo a sufrir la enfermedad. Entre ellas podemos destacar las siguientes:

DISC1

St Clair y colaboradores identificaron una traslocación recíproca balanceada (1;11) (q42;q14) presente en una familia escocesa que co-segregaba con esquizofrenia y otras patologías psiquiátricas (St Clair et al., 1990). El pedigrí incluía 21 individuos con esquizofrenia, trastorno bipolar, depresión recurrente y “trastorno conductual”, de los cuales 16 presentaban la traslocación. La traslocación afectaba a dos genes hasta ese momento desconocidos, a los que se les dio el nombre de *Disrupted in Schizophrenia 1 y 2 (DISC1 y DISC2)* (Blackwood et al., 2001; Millar et al., 2000). Estudios posteriores reprodujeron la asociación del locus 1q42 con esquizofrenia y otros trastornos psiquiátricos como el trastorno bipolar en familias de Finlandia (Ekelund et al., 2004), Escocia (Macgregor et al., 2004), el Reino Unido e Irlanda (Hamshere et al., 2005).

Se ha visto que personas portadoras de esta traslocación, tanto individuos sanos como pacientes esquizofrénicos, muestran una reducción en la amplitud y latencia de potenciales evocados (Blackwood et al., 2001), consistente con un defecto cognitivo subyacente. Esto sugiere que *DISC1* podría mediar este rasgo cognitivo característico de personas esquizofrénicas. Además, se han descrito asociaciones entre diversos polimorfismos de este gen y rasgos típicos de personas esquizofrénicas, como la disminución en el tamaño de la materia gris (Di Giorgio et al., 2008) o diferentes endofenotipos relacionados con habilidades cognitivas (Carless et al., 2011).

DISC1 es una *hub protein* (proteína altamente conectada que funciona como un “centro de redes”) que juega un papel muy importante en multitud de procesos, entre ellos, el funcionamiento normal de los microtúbulos (Kamiya et al., 2005), el crecimiento de dendritas y axones (Hattori et al., 2010; Kamiya et al., 2005; Miyoshi et al., 2003), la adhesión celular (Hattori et al., 2010) y la migración neuronal (Kamiya et al., 2005; Ozeki et al., 2003), y se ha visto que él y su interactoma están implicados en diversas funciones críticas cerebrales como el desarrollo y la maduración de la sinapsis (Camargo et al., 2007) o el desarrollo de la corteza cerebral (Ishizuka et al., 2011; Raznahan et al., 2010). Así pues, su implicación en el desarrollo de enfermedades psiquiátricas es compatible con la hipótesis del neurodesarrollo. *DISC2* da lugar a una molécula de ARN no codificante (Millar et al., 2000).

Se ha identificado otra traslocación en pacientes emparentados con trastornos psicóticos crónicos que afecta a *PDE4B*, un gen que interactúa con *DISC1* (Millar et al., 2005). En concreto, se ha visto que las dos proteínas codificadas por estos genes interactúan en la ruta del AMPc (adenosina fosfodiesterasa cíclica), asociada a procesos de aprendizaje y memoria, y se ha sugerido que variaciones en *DISC1* y/o

PDE4B podrían alterar esta interacción y afectar al catabolismo mitocondrial del AMPc, con repercusión sobre los citados procesos cerebrales (Millar et al., 2005).

Microdeleciones en el locus 22q11.2

Otra asociación clásica de la enfermedad con una alteración estructural rara es la microdelección 22q11.2 (Karayiorgou et al., 1995). Esta delección da lugar al síndrome velocardiofacial, o síndrome de DiGeorge, y aparece en 1 de cada 2.000 – 4.000 nacimientos (Botto et al., 2003; Kobrynski y Sullivan, 2007; Robin y Shprintzen, 2005). La delección se origina *de novo* en el 72-92% de los portadores (Driscoll et al., 1993; Ryan et al., 1997). El fenotipo asociado a la delección es muy variable, pero entre las manifestaciones más comunes están anomalías cardiovasculares y craneofaciales, inmunodeficiencia, estatura baja e hipocalcemia (Robin y Shprintzen, 2005). Además, muchas de las personas sufren retraso mental y alteraciones cognitivas y de comportamiento y un 30% de los portadores de la delección desarrollan psicosis, la mayor parte en forma de esquizofrenia (aproximadamente el 80% de éstos) (Murphy et al., 1999). Estos casos se corresponden aproximadamente con el 0,3% del total de casos de esquizofrenia y la delección se asocia a un riesgo superior a 20 (tabla 2) (Stefansson et al., 2008), por lo que se trata del mayor factor genético conocido de riesgo a padecer la enfermedad, aunque no es específico de la misma. En los individuos portadores de la microdelección no hay ninguna diferencia anatómica significativa entre los que desarrollan esquizofrenia y aquellos que no la desarrollan. Se ha sugerido que la gran heterogeneidad fenotípica asociada a la delección puede estar relacionada con la presencia de otras variantes genéticas que contribuyen a la variabilidad del fenotipo cognitivo y esquizofrénico (Karayiorgou et al., 2010).

Por el momento no se conoce el mecanismo por el que la delección incrementa el riesgo a desarrollar esquizofrenia. Una de las hipótesis es que la delección dé lugar al desenmascaramiento de una o más mutaciones deletéreas recesivas, o que provoque una infra-expresión de un número significativo de genes sensibles a la dosis, que puede tener un efecto acumulativo. Dependiendo del tamaño de la delección la región afectada incluye entre 60 y 35 genes conocidos (Edelmann et al., 1999; Shaikh et al., 2000), muchos de los cuales se expresan en el cerebro (Maynard et al., 2003). Entre ellos, algunos que podrían intervenir en el mecanismo son *COMT*, *PRODH*, *DGCR8*, *ZDHHC8*, *GNBIL*.

CNVs

Los avances en tecnologías como la hibridación genómica comparada (CGH del inglés, *Comparative Genomic Hybridization*) han permitido descubrir un tipo de variación estructural en una abundancia hasta el momento desconocida. A este tipo de

variación se le ha llamado *Copy Number Variations* (CNVs), y son variaciones en el número de copias de un segmento de ADN. Esta variación puede consistir en la duplicación o deleción del segmento, por lo que a diferencia de otro tipo de variaciones estructurales, como las inversiones o translocaciones recíprocas, pueden ser no balanceadas, dando lugar a diferencias en el tamaño del genoma entre individuos, perturbando así el equilibrio biológico normal de un genoma diploide en un locus determinado. Las CNVs suelen concentrarse en regiones del genoma con características determinadas que propician su aparición (Iafrate et al., 2004). Estudios recientes revelaron que 11.700 CNVs solapan con unos 1.000 genes conocidos (Conrad et al., 2009), de modo que esta importante fuente de variabilidad podría ser un mecanismo clave para la evolución humana, a través de la duplicación génica, triplicación o reorganización de exones. De hecho, se ha visto que entre los genes afectados por CNVs, existe un enriquecimiento de genes relacionados con el metabolismo de xenobióticos o receptores olfativos, procesos que se han relacionado con procesos adaptativos (Feuk et al., 2006). Además se pensó que este tipo de variación podría contribuir de manera significativa a la susceptibilidad de padecer enfermedades (Lupski, 2009), y como el tamaño de las variaciones oscila entre 1 kilobase y varias megabases (Iafrate et al., 2004; Redon et al., 2006; Sebat et al., 2004), su efecto fenotípico podría ser mayor que el de un SNP. Estas CNVs podrían dar lugar a manifestaciones fenotípicas a través de multitud de mecanismos, entre ellos, alteración de la expresión y regulación génicas o expresión de alelos recesivos desmascarados con una deleción (Zhang et al., 2009). Se ha propuesto, además, que el efecto de una CNV puede estar mediado por procesos de interacción con otra CNV u otro tipo de variante, que sería un segundo *hit*, que podría actuar como modificador genético. Esta teoría del *two hit* (o *multiple hits*) se ha propuesto para explicar casos de autismo, existiendo ya evidencias experimentales de la misma (Girirajan y Eichler, 2010; Girirajan et al., 2010).

Así pues, desde su identificación ha crecido el interés por comprender de qué modo estas variaciones afectan a la variabilidad fenotípica y a la aparición de enfermedades y ha motivado el desarrollo de técnicas más efectivas para detectarlas y determinar de manera precisa su posición y región exacta en el genoma para delimitar la región afectada por cada CNVs (McCarroll, 2010). En la primera generación de estudios de CNVs se empleaban paneles de CGH. Posteriormente se emplearon los paneles de genotipado de SNPs para su detección.

En esquizofrenia, el estudio de este tipo de variantes ha permitido detectar la existencia de un incremento significativo de CNVs largos (> 100 Kb) y raros (<1%) en pacientes, lo que sugiere que muchas de ellas podrían estar relacionadas con la enfermedad (tabla 2) (Gejman et al., 2011; Glessner et al., 2010; Ikeda et al., 2010; ISC, 2008; Kirov et al., 2009a; Kirov et al., 2008; Levinson et al., 2011; Magri et al., 2010; Mitchell y Porteous, 2011; Mulle et al., 2010; Need et al., 2009; Shi et al., 2008; Stefansson et al., 2008; Tam et al., 2010; Vrijenhoek et al., 2008; Walsh et al., 2008; Xu et al., 2008). Una proporción elevada de estos CNVs se originan *de novo*. La

deleción 22q11.2, identificada ya muchos años antes del empleo de estos paneles, es una variación de éstas.

Recientemente, Levinson y colaboradores analizaron un total de 3945 casos y 3611 controles y confirmaron las asociaciones ya descritas en relación a las deleciones en los loci 1q21.1, 15q13.3 y 22q11.21, deleciones que interrumpen exones del gen NRXN1 (2p16.3) y duplicaciones en el locus 16p11.2 (Levinson et al., 2011). Además, identificaron dos nuevas CNVs: la deleción 3q29 y duplicaciones exónicas en VIPR2 (7q36.3). El estudio muestra que los pacientes esquizofrénicos presentan un incremento significativo en el número de deleciones que afectan a genes, pero no de duplicaciones. El OR asociado a estas CNVs es elevado y oscila entre 2 y más de 30 (Levinson et al., 2011). Salvo en el caso de la deleción 22q11, el resto de CNVs se han identificado también en personas sanas. Muchas de estas CNVs se han asociado también con otros trastornos del neurodesarrollo, como retraso mental, retraso en el desarrollo, autismo o epilepsia (Gejman et al., 2011)

Tabla 2. Principales CNVs asociadas significativamente a esquizofrenia

	1q21.1 deleciones	2p16.3 Deleciones (NRXN1)	3q29 deleciones	7q36.2 duplicaciones (VIPR2)
Casos (%) ¹	0,17	0,2	0,08	0,12
Controles (%) ¹	0,02	0,02	0,002	0,01
Valor p ¹	7×10^{-13}	$3,3 \times 10^{-14}$	$3,9 \times 10^{-8}$	$1,4 \times 10^{-10}$
OR ²	9,5	7,5	3,7	4
	15q13.3 deleciones	16p11.2 duplicaciones	22q11.21 deleciones	
Casos (%)	0,14	0,15	0,29	
Controles (%)	0,02	0,004	0	
Valor p ¹	$4,3 \times 10^{-11}$	$7,7 \times 10^{-18}$	$1,6 \times 10^{-37}$	
OR ²	12,1	9,5	20,3	

¹ Información extraída de Gejman, 2011

² Información extraída de Levinson, 2011

Varios estudios han analizado el efecto en la población general de este tipo de CNVs largas, estimando su tasa de aparición y presión selectiva. Así, Itsara y colaboradores observaron que, en general, las CNVs de tamaño superior a 500kb están sometidas a una fuerte selección purificadora ($s=0,16$), esto es, los portadores de estas

CNVs tienen una reducción del 16% en su fertilidad. Además, observaron que la tasa de mutación en población general de estas CNVs largas se sitúa en torno a 0,0065 por genoma y por transmisión (Itsara et al., 2010).

Rees y colaboradores (Rees et al., 2011) hicieron un estudio similar centrado en CNVs que se hubiesen asociado a un incremento en el riesgo a padecer esquizofrenia y encontraron una elevada presión selectiva sobre ellas ($s=0,12-0,88$), lo que sugiere que tras la aparición de una de estas CNVs en un individuo, su persistencia en la población se limita a unas pocas generaciones. También estimaron que la tasa de aparición de estas CNVs es alta, entre 1:3.500 y 1:30.000, y vieron que cuanto más rara es una CNV, mayor es su coeficiente de selección. Otros estudios indican que la tasa de aparición de nuevas CNVs es mayor en pacientes esquizofrénicos o autistas que en controles (Itsara et al., 2010; Xu et al., 2008). Así pues, estas CNVs que individualmente se asocian significativamente a esquizofrenia son mutaciones recurrentes, y el incremento en CNVs totales observado en pacientes esquizofrénicos respondería a una elevada tasa de aparición.

Mutaciones puntuales

Otro tipo de variantes raras son las mutaciones puntuales. Hasta hace poco su búsqueda se restringía a la secuenciación de genes o regiones candidatas en un número pequeño de individuos. Sin embargo, teniendo en cuenta el sustrato poligénico de la enfermedad, la probabilidad de que las personas analizadas en cada estudio tuviesen mutaciones precisamente en el gen o genes analizados es mínima, por lo que la estrategia resulta poco eficaz y por el momento no se puede estimar de forma precisa hasta qué punto este tipo de variantes contribuyen a la etiología de la enfermedad. Esta situación está cambiando gracias a la aparición de la resecuenciación a gran escala o ultrasecuenciación. El uso generalizado de este tipo de tecnologías, que posibilita el análisis del exoma (parte codificante del genoma) o genoma completo de un mismo individuo, supondrá un avance extraordinario en la investigación de este tipo de variantes raras ya que permitirá sistematizar su búsqueda (Cirulli y Goldstein, 2010).

Por el momento, la resecuenciación de algunos genes candidatos ha dado lugar a la identificación de mutaciones raras en *DISC1* (Song et al., 2008) o *ABCA13* (Dwyer et al., 2011), entre otros. Piton y colaboradores llevaron a cabo la secuenciación de 111 genes ligados al cromosoma X en 143 pacientes esquizofrénicos, 142 autistas y 277 individuos sanos e identificaron más de 200 variantes no sinónimas y un exceso de variantes con efecto deletéreo (Piton et al., 2011), por lo que es posible que muchas de ellas sean variantes de predisposición a padecer enfermedades psiquiátricas.

3. Papel de la selección natural en el espectro alélico de la esquizofrenia

3.1 Esquizofrenia, ¿paradoja evolutiva?

La combinación de elevada prevalencia y edad temprana de aparición diferencia a los trastornos mentales de la mayor parte de las enfermedades (Uher, 2009). En el caso de la esquizofrenia, su persistencia en las poblaciones a elevada prevalencia aun a pesar de la desventaja reproductiva observada en pacientes y su elevada heredabilidad se ha considerado una paradoja evolutiva, ya que se espera que variantes asociadas a una pérdida de fertilidad sean eliminadas mediante selección purificadora del “pool” genético y, por tanto, estas variantes y los fenotipos asociados deberían ser raros (Huxley et al., 1964). Sin embargo, la enfermedad continúa siendo común en las poblaciones humanas.

Esta aparente paradoja ha motivado intensos debates (Crow, 1997; Crow, 2008; Horrobin, 1998; Keller y Miller, 2006). Algunos autores han propuesto que es posible que aunque la esquizofrenia esté sometida a presión negativa, las mutaciones de susceptibilidad podrían conferir cierta ventaja en personas que no la desarrollan (Huxley et al., 1964; Jarvik y Deckard, 1977). Por ejemplo, Nettle propuso que las mutaciones de predisposición podrían estar relacionadas con una mayor imaginación, creatividad verbal y artística, cognición esquizotípica, pensamiento divergente o determinadas conductas que hayan estado favorecidas durante la evolución del linaje humano (Nettle y Clegg, 2006). De hecho, Karlsson y colaboradores (Karksson, 1970) observaron que en población Islandesa, familiares de individuos con trastornos psicóticos gozaban de un potencial creativo excepcional. Otros investigadores han sugerido que las ventajas podrían estar asociadas a rasgos ajenos al cerebro, como resistencia a toxinas o a diferentes patógenos (Carter y Watts, 1971). Otros autores

sitúan a la enfermedad como algo inherente a la propia línea evolutiva humana. Por ejemplo, Crow propuso que la vulnerabilidad genética a padecer esquizofrenia puede haber surgido como consecuencia de la evolución de rasgos característicos de los humanos (Crow, 1997); en concreto, que la evolución cerebral que ha dado lugar al lenguaje ha tenido como efecto colateral la aparición de trastornos cognitivos, como la esquizofrenia.

Todas estas hipótesis señalan que los genes que predisponen a padecer esquizofrenia están entre los que han sufrido evolución reciente.

3.2 Esquizofrenia y selección natural

De acuerdo con la hipótesis neutralista de evolución molecular (Kimura, 1979), en general, la mayor parte de los polimorfismos existentes en una población en un momento dado son neutros desde el punto de vista de la selección natural: por un lado, aquellos alelos que confieren alguna ventaja selectiva, si no son eliminados por deriva genética en las primeras generaciones, serán seleccionados de manera positiva e incrementarán rápidamente su frecuencia hasta su fijación en la población, contribuyendo durante poco tiempo al polimorfismo de una especie; por otro lado, alelos deletéreos asociados a una pérdida de fertilidad serían eliminados o se mantendrían a frecuencia baja mediante selección purificadora. De este modo, la mayor parte de la variabilidad común presente en una población en un momento determinado estaría constituida por alelos neutros -que típicamente varían su frecuencia muy lentamente por acción de la deriva genética- y por la eventual escasa proporción de polimorfismos favorecidos en ese momento que estén en vías de fijación.

Como se comentó anteriormente, la esquizofrenia está asociada a una elevada heredabilidad y a una desventaja reproductiva, por lo que se puede decir que está sometida a selección negativa. Siendo esto así, y de acuerdo con la hipótesis neutralista de evolución molecular, es de esperar que aquellos alelos de susceptibilidad con efecto elevado o moderado sobre el fenotipo estén expuestos a esta presión selectiva, de modo que no podrán llegar a ser comunes y segregarán en la población a baja frecuencia o incluso desaparecerán; por otra parte, variantes que incrementen mínimamente el riesgo a padecer la enfermedad (con efecto mínimo sobre el fenotipo) evolucionarán casi como polimorfismos neutros, pudiendo llegar a ser comunes en la población.

Sin embargo, se han descrito algunos escenarios evolutivos en los que variantes que incrementan el riesgo a padecer la enfermedad moderadamente podrían ser comunes en la población.

La frecuencia de una variante en la población en un momento determinado es la consecuencia de su pasado evolutivo, y éste puede estar modulado por multitud de factores, que pueden cambiar a medida que pasa el tiempo (Falconer, 1996). Siendo esto así, uno de los escenarios evolutivos es el que propone el modelo de susceptibilidad ancestral (Di Rienzo, 2006; Di Rienzo y Hudson, 2005; Wright et al., 2003): alelos ancestrales que en condiciones pasadas eran neutros o estaban favorecidos por selección, podrían estar mal adaptados a nuevos ambientes, dando lugar a un incremento en la susceptibilidad a padecer determinadas enfermedades comunes. La historia evolutiva reciente de la especie humana se ha desarrollado en un escenario de incesantes cambios. Es conocido que nuestro entorno es radicalmente diferente al que vivieron poblaciones humanas ancestrales. Procesos como la salida de África y la colonización de nuevos territorios, la aparición de la agricultura, cambios en el clima y en el estilo de vida o las grandes migraciones de poblaciones humanas - que implicaron la propagación de graves enfermedades infecciosas-, generaron numerosos cambios en la dirección de la presión selectiva. En este contexto, variantes implicadas en funciones biológicas que interaccionasen con estos factores cambiantes serían sensibles a la consecuente presión evolutiva y, alelos que antes funcionaban bien, que podrían ser neutros o incluso ventajosos, pasarían a tener un efecto deletéreo, empezando a ser activamente reemplazados por variantes nuevas (derivadas) mejor adaptadas. Estas variantes seleccionadas terminarían fijándose en la población, si no hay cambios posteriores en la dirección de la selección, pero hasta que esto no ocurre, la variante ancestral sigue segregando en la población junto con la variante derivada. Lo que propone el modelo es que serían estas variantes ancestrales mal adaptadas las que contribuirían a la susceptibilidad de enfermedades comunes.

Existen diferentes ejemplos que ilustran el impacto que estos cambios de presión selectiva pueden tener. Patologías como la diabetes tipo II o los problemas vasculares podrán explicarse con la hipótesis del alelo frugal (Neel, 1962): alelos favorecidos en ambientes pasados por incrementar la eficacia de la utilización y el almacenamiento de la energía, en la actualidad podrían abocar a las citadas patologías. En relación con esto, se ha detectado que el alelo ancestral $\epsilon 4$ de la *APOE*, relacionado con el metabolismo lipídico, confiere una mayor susceptibilidad a padecer Alzheimer y enfermedad cardiovascular (Fullerton et al., 2000). Asimismo, se ha visto que variantes ancestrales del gen del Angiotensinógeno incrementan el riesgo a padecer hipertensión (Wu et al., 2004) y se ha sugerido que la variante derivada de uno de estos polimorfismos ha alcanzado elevada frecuencia en la población debido a la acción de la selección natural reciente (Nakajima et al., 2004). Esto fundamenta la hipótesis de “retención de sodio” (Gleibermann, L, 1973), según la cual, poblaciones ancestrales de África ecuatorial adaptadas a condiciones extremas de calor y humedad, con alelos especializados en la retención de agua y sales, experimentarían problemas de hipertensión al colonizar ambientes más fríos y secos.

Otro escenario evolutivo es el que propone el modelo de pleiotropismo (fenómeno por el que un mismo gen es responsable de efectos fenotípicos o caracteres distintos y no relacionados) antagónico (Wright et al., 2003). Según este modelo,

variantes genéticas que incrementan el riesgo a sufrir una enfermedad asociada a fertilidad reducida podrían segregarse a elevada frecuencia en la población durante un período de tiempo si éstas están implicadas también en la aparición de otro fenotipo con éxito reproductivo. La presión selectiva sobre una mutación concreta está determinada por la combinación del efecto de la selección sobre todos los fenotipos en los que interviene y es la presión resultante la que determina la frecuencia de la variante en la población (Zhang y Hill, 2003) Falconer, 1996). Así, según este modelo, alelos que incrementan el riesgo a padecer esquizofrenia podrían sin embargo incrementar su frecuencia y ser comunes si además son o han sido objeto de selección natural reciente por participar en la etiología de otros fenotipos con éxito reproductivo.

Teniendo en cuenta ambos escenarios y nuestra historia evolutiva reciente, es posible que muchos loci del genoma hayan sufrido eventos de selección natural reciente, de modo que existan numerosos alelos funcionalmente diferentes a frecuencias intermedias, entre los que podrían estar aquellos implicados en la aparición de las enfermedades complejas y, en concreto, de la esquizofrenia.

La búsqueda de confirmación de estos modelos necesita la identificación de estos fenómenos de selección. Cuando un alelo aumenta en frecuencia por selección puede dejar una huella o patrón de diversidad característico en el genoma (Sabeti et al., 2006). Gracias a la secuenciación del genoma humano y el de otras especies, al proyecto HapMap y al desarrollo de pruebas específicas, estas huellas genéticas pueden ser detectadas; así, de más antigua a más reciente, las señales y la edad del evento selectivo serían: elevada proporción de cambios funcionales (millones de años), reducción en la diversidad genética (<250.000 años), frecuencia elevada de alelos derivados (<80.000 años), diferencias entre poblaciones (de 50.000 a 75.000 años) y longitud de haplotipos (<30.000) (Sabeti et al., 2006). Dado que los modelos evolutivos descritos, susceptibilidad ancestral y pleiotropismo antagónico, hacen referencia a procesos de selección natural reciente, la estrategia ideal sería la que se basa en la longitud de los haplotipos. Esta estrategia detecta eventos de selección reciente mediante la localización de “barridos incompletos de selección”, detectados por la presencia de un nivel de desequilibrio de ligamiento en torno a una mutación significativamente mayor de lo observado en otros haplotipos del genoma de su misma frecuencia, tras corregir por las diferentes tasas de recombinación a lo largo del genoma (Sabeti et al., 2002; Sabeti et al., 2006; Voight et al., 2006).

Se han llevado a cabo numerosos estudios en los que se analiza todo el genoma humano para buscar señales de selección natural (Barreiro et al., 2008; The International HapMap Consortium, 2005; Costas et al., 2005; Di Rienzo, 2006; Hawks et al., 2007; Sabeti et al., 2006). Si se clasifican los genes con señal positiva de selección, entre las categorías con mayor número de genes están: la que engloba a genes relacionados con el comportamiento, la que abarca a genes relacionados con la anatomía del cerebro y la que incluye genes de interacción con patógenos (Vallender y Lahn, 2004). Es decir, genes que podrían ser relevantes en esquizofrenia.

En un intento de comprobar si genes candidatos a participar en la etiología de la enfermedad presentan este tipo de señales de selección, Crespi y colaboradores analizaron 76 genes, encontrando señales de selección positiva en 28 de ellos (Crespi et al., 2007). Estas señales de selección corresponden tanto a eventos de selección recientes, ocurridos en los últimos miles de años de la evolución humana, como eventos de selección “histórica”, mucho más antigua en el linaje humano. Por otra parte, estudios neurológicos muestran que entre las áreas cerebrales que en pacientes esquizofrénicos están desreguladas se encuentran las regiones con mayor evidencia de haber sufrido cambios evolutivos diferenciales en el linaje humano (Brune, 2004; Burns, 2006). Además, se ha observado que una proporción de los genes implicados en procesos metabólicos del cerebro, alterados en pacientes esquizofrénicos, pueden haber sufrido procesos de adaptación específicos del linaje humano (Khaltovich et al., 2008).

Estas observaciones sugieren que la selección natural podría tener un papel importante en la evolución de algunos genes de susceptibilidad a padecer esquizofrenia, por lo que el análisis de variantes comunes sometidas a selección natural reciente podría dar alguna pista acerca de la etiología de la enfermedad. Esta hipótesis puede evaluarse buscando señales de selección y buscando asociación al fenotipo esquizofrénico en los mismos polimorfismos.

Teniendo en cuenta que la mayor parte de la variabilidad común es fenotípicamente neutra, centrar un estudio en variantes detectadas por la selección natural sería una forma de enriquecer el estudio en variantes funcionalmente diferentes.

Hipótesis y Objetivos

Hipótesis

La vulnerabilidad genética a padecer esquizofrenia podría estar determinada por variantes alélicas en genes relacionados con la transmisión y plasticidad sináptica y el desarrollo del sistema nervioso central.

Polimorfismos sometidos a selección positiva constituyen variantes funcional y fenotípicamente diferentes, por lo que podrían contribuir a la susceptibilidad genética a padecer enfermedades

Genes implicados en la etiología de la esquizofrenia pueden haber participado en eventos recientes de adaptación de la especie humana, por lo que es posible que la frecuencia de una parte de los polimorfismos actuales en estos genes sea el resultado de la acción de la selección natural.

Los SNPs no sinónimos alteran la secuencia aminoacídica de las proteínas, por lo que *a priori* presentan una mayor probabilidad de ser funcionales y, por tanto, de estar implicados en la aparición de enfermedades.

Esta probabilidad será mayor cuanto menor sea la frecuencia de un polimorfismo, por lo que una proporción elevada de los SNPs no sinónimos presentes a baja frecuencia en la población será moderadamente deletérea, contribuyendo a incrementar la susceptibilidad genética a padecer enfermedades complejas, de acuerdo con el modelo de mutación-selección poligénico.

El impacto de SNPs no sinónimos en la estructura y función de las proteínas puede predecirse mediante herramientas bioinformáticas, empleando algoritmos basados en el grado de conservación de la secuencia del genoma y en características físicas de los aminoácidos.

Objetivos

El objetivo central del trabajo de investigación que aquí se presenta fue la identificación de factores genéticos de predisposición a padecer esquizofrenia mediante el análisis de variantes con alta probabilidad *a priori* de ser funcionales. Atendiendo al modo en que hemos abordado esta búsqueda, el conjunto de los seis trabajos que presentamos a continuación pueden ser divididos en dos partes.

En la primera parte examinamos variantes que hubiesen sido objeto de selección natural reciente y, por tanto, funcionalmente diferentes. Para ello nos planteamos los siguientes objetivos concretos:

- 1) buscar polimorfismos sometidos a selección natural reciente en genes candidatos.
- 2) evaluar su implicación en el desarrollo de la enfermedad mediante estudios de asociación caso-control.

En la segunda parte investigamos sistemáticamente el papel de SNPs no sinónimos, tanto comunes como raros, en la etiología de la enfermedad. Seguimos diferentes estrategias de análisis para lograr los siguientes objetivos concretos:

- 3) identificar posibles SNPs no sinónimos, raros o comunes, que por sí mismos causen un incremento considerable en el riesgo a padecer esquizofrenia.
- 4) determinar si la acumulación de variantes comunes de predisposición se asocia a un incremento en el riesgo a padecer la enfermedad.
- 5) determinar si la acumulación de variantes raras con alta probabilidad de ser funcionales en genes candidatos se asocia a un mayor riesgo de padecer el trastorno.

Métodos, resultados y discusión

Estudio 1. Esquizofrenia, *DRD3* y selección natural

A common haplotype of *DRD3* affected by recent positive selection is associated with protection from schizophrenia

Javier Costas · Noa Carrera · Eduardo Domínguez · Elisabet Vilella · Lourdes Martorell · Joaquín Valero · Alfonso Gutiérrez-Zotes · Antonio Labad · Ángel Carracedo

Received: 24 July 2008 / Accepted: 28 October 2008 / Published online: 6 November 2008
© Springer-Verlag 2008

Abstract The number and frequency of susceptibility alleles at loci associated to most psychiatric disorders is largely unknown, in spite of its relevance for the design of studies aiming to find these alleles. Both, common polymorphisms and rare mutations may contribute to the genetic susceptibility to complex psychiatric disorders, being the relative relevance of each type of variation currently under debate. Here, we confirmed the existence of a common protective haplotype against schizophrenia at the dopamine D_3 receptor (*DRD3*) gene, by replication and pooled analysis with previous data (Mantel–Haenszel χ^2 P value = 0.00227; OR = 0.79, 95% CI 0.68–0.92, based on 794 cases and 1,078 controls from three independent

populations of European origin). This protective haplotype is at very low frequency in Sub-Saharan Africans (median 0.06) and at intermediate frequencies in other populations (median 0.25). We also revealed, by examining the patterns of linkage disequilibrium around this gene, that the protective haplotype has reached high frequency in non-African populations due to selection acting, most probably, on a linked functional polymorphism, the non-synonymous single nucleotide polymorphism Ser9Gly (rs6280), also at *DRD3*. Thus, this finding shows that the natural selection may play a role in the existence of common alleles conferring different susceptibility to schizophrenia.

Introduction

Schizophrenia is a common mental disorder, with a lifetime prevalence of 0.72% (Saha et al. 2005). Genetic factors clearly play an important role in its etiology, with heritability estimates of 81% (Sullivan et al. 2003). It often starts in young adulthood and persists during the lifetime of the individual, being associated to decreased fertility (Haukka et al. 2003) and increased mortality (Brown 1997). Therefore, the existence of common susceptibility alleles is considered an evolutionary paradox (Huxley et al. 1964). Evolutionary processes have often been proposed to solve it (Brune 2004; Crespi et al. 2007; Keller and Miller 2006).

One of the most studied genes in relation to schizophrenia susceptibility is *DRD3*, an obvious candidate taking into account that it is an important target for several anti-psychotic drugs (Sokoloff et al. 1990). Most of the work was centered on the non-synonymous SNP Ser9Gly (dbSNP ID: rs6280). Several meta-analyses on this polymorphism have been published in the last years, without any clear conclusion. While some of them suggest a

Electronic supplementary material The online version of this article (doi:10.1007/s00439-008-0584-7) contains supplementary material, which is available to authorized users.

J. Costas (✉) · N. Carrera · Á. Carracedo
Fundación Pública Galega de Medicina Xenómica (FPGMX)-SERGAS, University of Santiago Hospital Complex, Edif. Consultas Planta-2, 15706 Santiago de Compostela, Spain
e-mail: javier.costas@usc.es

E. Domínguez
Applied Pharmacology to Membrane Biology and Signal Transduction Mechanisms Research Group (BIOFARMA), Departament de Farmacology, University of Santiago de Compostela, Santiago de Compostela, Spain

E. Vilella · L. Martorell · J. Valero · A. Gutiérrez-Zotes · A. Labad
Faculty of Medicine, Psychiatric Hospital Institut Pere Mata, University Rovira i Virgili, Reus, Spain

Á. Carracedo
Genomic Medicine Group, Institute of Legal Medicine, CEGEN and CIBERER, University of Santiago de Compostela, Santiago de Compostela, Spain

positive but very modest association with schizophrenia (Lohmueller et al. 2003), the most recent ones failed to find any significant association (Jonsson et al. 2004; Ma et al. 2008).

Recently, our group (Dominguez et al. 2007) and others (Sand et al. 2008; Talkowski et al. 2006) identified common haplotypes of *DRD3* associated to schizophrenia in three independent populations of European origin, by extensive linkage disequilibrium (LD) mapping using different sets of tagSNPs. Our previous study, carried out in Galician population (NW Spain), revealed that the association was mainly due to the over-representation of several low-frequency haplotypes in schizophrenic individuals at the expense of a single common haplotype, more frequent in controls. Thus, this common haplotype may be regarded as protective against schizophrenia (Dominguez et al. 2007). In this work, we further explore the existence of this common haplotype.

Materials and methods

Subjects

The replication study consisted of 273 cases aged 18–60 years (mean 39.9, SD 10.8) from the Hospital Psiquiàtric Universitari Institut Pere Mata (Reus, Catalonia, NE Spain). All patients had a consensus diagnosis of schizophrenia, excluding schizoaffective disorder according to DSM-IV criteria, conducted by two independent, trained, senior psychiatrists whose diagnostic concordance was found to be high ($K > 0.8$). Most of the patients (88.2%) were institutionalized. According to the DSM-IV criteria, paranoid schizophrenia was the most frequent subtype (66.7%).

The control group (512 samples, aged 18–60 years, mean 38.8, SD 11.1) was recruited from primary care centers in the same geographical area. Screening for psychiatric

diseases was performed according to the 28-scaled global health questionnaire (GHQ). Only unrelated individuals with no personal history of psychiatric disease (clinically documented) and a GHQ score < 7 were included. The study complied with the guidelines of the Medical Ethical Committee, and informed consent was obtained from all participants. To limit genetic heterogeneity, we excluded participants from ethnic minorities such as gypsies or those who were born, or whose parents were born, in continents other than Europe. Significant population stratification has been previously discarded by the “structured association” method (Pritchard et al. 2000) using 25 unlinked SNPs (Roig et al. 2007).

We also genotyped the samples of the Centre d’Etude du Polymorphisme Humain (CEPH) human diversity panel, representing 1,064 individuals from 51 populations around the world (Cann et al. 2002), considering only the standardized subset of 952 samples, after exclusion of duplicated, closely related, and atypical individuals (Rosenberg 2006).

Genotyping

The SNPs defining the common protective haplotype in our previous work (Dominguez et al. 2007) (rs7631540, rs1486012, rs2134655, and rs963468) as well as those additional SNPs defining the haplotype region associated to schizophrenia in the work of Talkowski et al. (2006) (rs10934254, rs324030, and rs324029) were genotyped using the MassArray technology from Sequenom (San Diego, CA, USA), as previously described (Costas et al. 2005) (Table 1). As a quality control, we tested for any departure of Hardy–Weinberg equilibrium (HWE) in both cases and controls, using Haploview (Barrett et al. 2005). In addition, one sample from the CEPH human diversity panel was genotyped 24 times, without any discrepant result. Individual SNP data, irrelevant in the present work but useful for meta-analysis, are shown as Supplementary Table.

Table 1 SNPs genotyped at the present work

SNP ID	Strand	Allele 1/2	Position ^a	Gene location	MA in CEU ^b
rs7631540	+	C/T	115313209	3′	2
rs1486012	+	A/T	115322112	3′	1
rs10934254	+	A/G	115324324	3′	Na
rs2134655	+	T/C	115340891	Intron 5	1
rs963468	+	A/G	115345577	Intron 4	1
rs324030	+	C/G	115364131	Intron 2	2
rs324029	+	A/G	115364313	Intron 2	1

^a NCBI build 36.1

^b Minor allele (i.e., allele at the lower frequency) in the CEU (CEPH samples of European ancestry) samples from HapMap. Na, not available, SNP not genotyped in HapMap samples

Statistical analysis

Haplotype estimation and comparison of overall haplotype frequencies between cases and controls were done with the software Unphased v2.4 (Dudbridge 2003). Haplotypes at frequencies lower than 5% in both groups were clustered together prior to the likelihood ratio test. Comparison of the frequencies of the common haplotype versus the other haplotypes was carried out by χ^2 test on 2×2 contingency tables. The Mantel–Haenszel χ^2 was employed to combine samples from the current and previous data, using WINPEPI (Abramson 2004). The PHASE 2.1 program (Stephens et al. 2001) was used to estimate the probability of haplotype inference in the case of the rare haplotype cited in our previous work (Dominguez et al. 2007). A priori power calculation of the replication sample was carried out using the PS Program (Dupont and Plummer 1997).

Long range haplotype (LRH) test

In order to investigate if recent adaptive selection has played a role in the allelic spectrum of *DRD3* in European samples, we applied the LRH test around this locus to phased data from the CEPH (Utah residents with ancestry from northern and western Europe) samples (CEU) of HapMap phase II (International HapMap Consortium 2007), as implemented in the Sweep program (Sabeti et al. 2002). This test is based on the identification of specific haplotypes with a pattern of LD surrounding them higher than expected based on their frequencies. We analyzed all the core haplotypes around *DRD3*, defined by the haplotype block method of Gabriel et al. (2002) as implemented in Sweep (default parameters). For each haplotype at each block (i.e., the core haplotype) we calculated the extended haplotype homozygosity (EHH), defined as the probability that two randomly chosen chromosomes carrying the core haplotype of interest are identical for the entire interval from the core region to a specified distance. We also measured the relative EHH (rEHH), i.e. the ratio of the EHH on the tested core haplotype compared with the EHH of the grouped set of core haplotypes at the region not including the core haplotype tested (Sabeti et al. 2002). By this way, the influence of different recombination rates along the genome in EHH is adjusted using the other haplotypes of the same region. Finally, we consider the significance of the findings by comparison of rEHH values to all the core haplotypes of chromosome 3 at similar frequency (20 bins, intervals of 5%) and at three different distances from core, 0.1 cM, 0.2 cM and marker $H = 0.04$ on both sides of the core. This last distance, the recommended one in the Sweep documentation (<http://www.broad.mit.edu/mpg/sweep/documentation.html>), allows for comparison of core haplotypes matching the observed amount of recombination.

Results

Replication study

The SNPs defining the protective haplotype in our previous work were genotyped in new subjects from Catalonia. All SNPs were consistent with HWE ($P > 0.05$). Haplotype frequencies in the new samples are shown in Table 2. Although the frequencies do not differ significantly between cases and controls, there is a trend for association (overall P value = 0.085). The rare haplotype of susceptibility, rs7631540_2—rs1486012_1, found previously in 8 of 260 Galician schizophrenic patients (1 of them homozygous for this haplotype) and 0 of 354 Galician controls (Dominguez et al. 2007), was almost absent from the new sample, only two controls presenting this haplotype (posterior probability estimate of the haplotype = 100%). The putative protective haplotype (rs7631540_2—rs1486012_2—rs2134655_1—rs963468_2) is present in excess in controls, although it does not reach significance (22.9 vs. 25.6%, χ^2 test, $P = 0.23$).

Pooled analysis of the LD mapping results of *DRD3* in schizophrenia in different studies

We genotyped the tagSNPs defining the associated haplotypes from Pittsburgh (Sand et al. 2008; Talkowski et al. 2006) (rs10934254, rs324030, and rs324029, in addition to rs2134655) in the Catalonian samples to allow for comparisons to our own findings. The common haplotype of protection identified by us was always associated to the haplotype rs10934254_1—rs2134655_1—rs324030_1—rs324029_2 from Pittsburgh, and vice versa. Thus, this haplotype corresponds to the putative protective haplotype. The same correspondence between both haplotypes may be inferred from HapMap data, although SNP rs10934254 is not genotyped (Table 3). The haplotype rs10934254_1—rs2134655_1—rs324030_1—rs324029_2 is also present at higher frequency in control samples from Pittsburgh (26 vs. 31%, Table 3 of Talkowski et al. 2006, see also Sand et al. 2008).

Table 2 Haplotype frequencies in the replication sample^a

Haplotype ^b	No. of cases	Freq cases	No. of controls	Freq controls
1121	231	0.426	413.6	0.412
2212	124	0.229	255.6	0.256
2222	113	0.208	213.1	0.212
1222	43	0.079	73.1	0.073
1122	31	0.057	39.9	0.040

^a Only haplotypes at frequencies higher than 0.05 in any group are shown

^b SNPs are in this order: rs7631540, rs1486012, rs2134655, rs963468

Table 3 Haplotype frequencies in the HapMap phase II CEU samples

Haplotype ^a	No. of chromosomes ^b	Frequency ^b
112112 -Ser	44	0.367
221212 -Ser ^c	32	0.267
222221 -Gly	24	0.200
122212 -Ser	6	0.050
122221 -Gly	6	0.050
112221 -Gly	5	0.042
111212 -Ser	1	0.008
222212 -Ser	1	0.008
112112 -Gly	1	0.008

^a SNPs ordered as follow: rs7631540, rs1486012, rs2134655, rs963468, rs324030, rs324029, rs6280. SNP rs10934254 has not been genotyped by the International HapMap Consortium. SNPs identifying the protective haplotype in Dominguez et al. (2007) are in boldface, and those in Talkowski et al. (2006) are italicized. The Gly/Ser alleles at the non-synonymous SNP rs6280 are also shown

^b Haplotype frequencies taken from the phased data of HapMap

^c Protective haplotype

Allele 1 at SNP rs2134655 is a perfect proxy from the protective haplotype in the Pittsburgh sample. This allele is present in 176 of 654 alleles (26.9%) in the case group and in 168 of 522 alleles (32.2%) in the control group (χ^2 test, $P = 0.048$) (V.L. Nimgaonkar, personal communication). The pooled analysis of the three case-control data (Galician, Catalonian, and Pittsburgh samples) clearly supports the existence of the protective haplotype (Mantel-Haenszel χ^2 P value = 0.00227; OR = 0.79, 95% CI 0.68–0.92, total sample size: 794 cases, 1,078 controls) (Fig. 1). Sequencing of the seven exons of *DRD3* and six non-coding regions highly conserved between mammals in ten individuals homozygous for the haplotype failed to find any putative functional variant strongly linked to this haplotype.

The role of recent positive selection on the allelic spectrum of *DRD3*

Applying the LRH test to the HapMap CEU samples, we identified one core haplotype with empirical P values lower than 0.05 at the three different distances at the region upstream the core (P values of 0.024, 0.005 and 0.008 at -0.1 cM, -0.2 cM and marker $H = 0.04$, respectively). As shown in Fig. 2, this core region is composed of three haplotypes at frequencies higher than 1%, and the haplotype at 65% frequency presents an EHH considerably higher than expected. The use of different core regions adding additional SNPs at each side does not change the results of the LRH test. Interestingly, this core haplotype includes the non-synonymous SNP rs6280. The ancestral allele, shared between humans and other primates, is Gly (C at the DNA level). The derived allele, Ser, is almost always associated

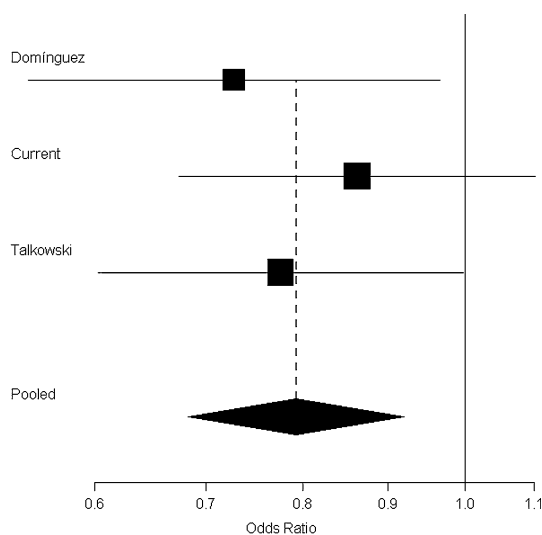


Fig. 1 Graphical representation of allelic metaanalysis for the putative protective haplotype and schizophrenia. ORs (logarithmic scale) for single studies are shown as boxes with size proportional to the weight of the study in the metaanalysis. Horizontal lines represent 95% CI of ORs. The pooled result (fixed model) and its confidence intervals are represented by a diamond

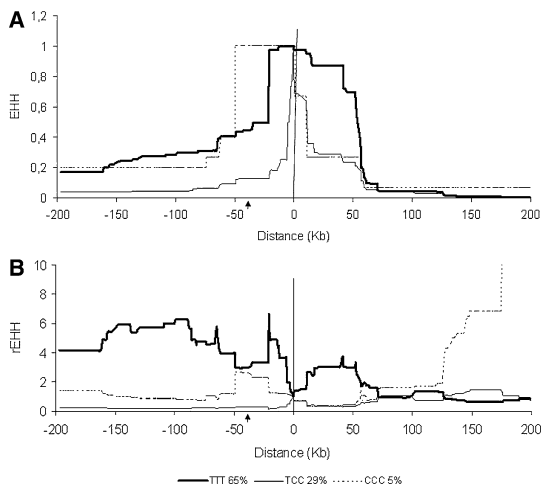


Fig. 2 Extended Haplotype Homozygosity (EHH) (a) and relative EHH (b) plots of the core haplotype within *DRD3* that appears as significant in the LRH test at several different distances in the CEU population from HapMap. The SNP at the core are ordered as follows: rs3732783, rs6280, rs324026. Numbers after the haplotype represent the frequency in the HapMap CEU samples. Location of SNP rs2134655, proxy of the protective haplotype in this population, is shown by a vertical arrow

to the haplotype at 65% frequency. There are just one common haplotype in 40 kb around rs6280 bearing the Ser allele, and five other haplotypes that appeared just once

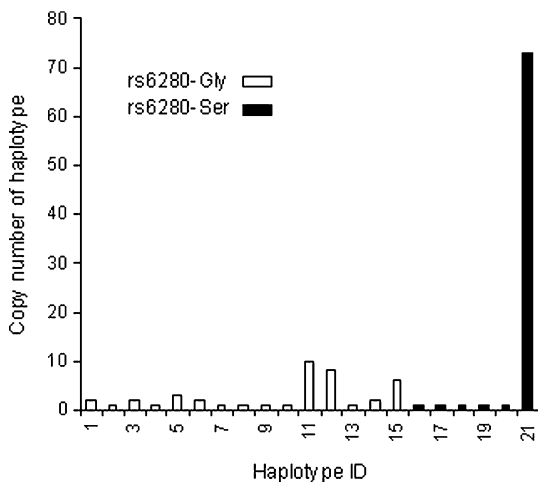


Fig. 3 Frequency distribution of all the haplotypes defined by the SNPs present in 40 kb around rs6280, in CEU population from HapMap. The different haplotypes at the region are shown on the X axis

(Fig. 3). By contrast, there are 15 alleles bearing the Gly variant.

While the derived haplotype is also at high frequency in the Asian populations from HapMap (63 and 76% in Han Chinese from Beijing and Japanese from Tokyo, respectively), it is the allele at lower frequency in Yoruba from Ibadan (the African population genotyped by HapMap), accounting for 12% of the chromosomes. Frequency data from different populations worldwide, taken from the Allele Frequency Database (Rajeevan et al. 2003) (ALFRED, UID for rs6280: SI000672P), confirm this pattern, i.e., in general, the ancestral allele is the commonest in African

populations, while the derived allele is the commonest in the other populations (detailed information on the different populations may be retrieved from ALFRED).

Frequency of the common haplotype of protection in population samples worldwide

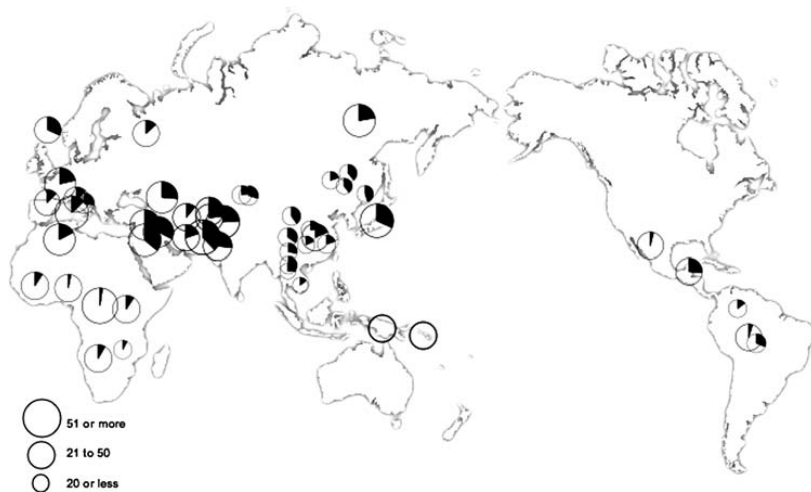
Next, we wanted to know if the variant subject to adaptive selection may influence the frequency of the protective haplotype. As shown in Table 3, this seems to be the case. The rs2134655_1 allele, proxy of the protective haplotype in the HapMap CEU data, and located 32.6 Kb upstream of rs6280, is always associated with the Ser allele at rs6280 ($D' = 1$). Interestingly, the allele 1 at rs2134655 is the derived one. Thus, the protective haplotype has reached its high frequency in Europeans “hitchhiked” with the selected variant.

Finally, we determined the worldwide frequency of the protective haplotype by genotyping SNPs rs7631540, rs1486012, rs2134655, and rs963468, in the samples of the CEPH human diversity panel, representing 51 populations worldwide. The protective haplotype seems to be absent from Oceania, being at low frequency in Sub-Saharan Africans (median 0.06; ranging from 0.02 in Biaka Pygmies to 0.09 in Mandenka), and at intermediate frequencies in other populations (median 0.25; ranging from 0.04 in Americans Pima to 0.45 in Hazhen from East Asia), in agreement with the expectancy taking into account the frequency of the rs6280 alleles around the world (Fig. 4).

Discussion

The present work shows evidence for the existence of a common haplotype of protection against schizophrenia at

Fig. 4 Worldwide frequency of the *DRD3* protective haplotype, rs7631540 T-rs1486012 T-rs2134655 A-rs963468 G, shown in black. Sample size indicates number of chromosomes



the *DRD3* gene. Thus, our combined analysis of 794 cases and 1,078 controls from three independent populations of European origin supports this association (P value = 0.00227), although further studies are needed to totally discard type I errors. The lack of association in the Catalonian sample alone is not surprising, as we have a power of 0.76 to replicate our previous result. Interestingly, this association is not due to a high effect in just one sample, as the removal of any one of the three samples still gives rise to significant results (P values of 0.019, 0.0033, and 0.026 after removal of Pittsburgh, Catalonian and Galician samples, respectively).

Our results also show that a haplotype containing the derived allele Ser at rs6280 has been subject to recent positive selection, at least in European populations, leading to its fast increase in frequency. This is strongly suggested by the unusual pattern of haplotype diversity unveiled in the comparison of haplotypes carrying the Ser versus Gly allele at rs6280 (Fig. 3). The LRH test confirms the exceptional pattern of LD around this SNP (Fig. 2). Thus, the most frequent haplotype at this region, identified by the Ser allele, presents an extreme value of rHH in comparison with empirical distribution of chromosome 3 in the CEU population from HapMap. We must take into account that this comparison is a conservative approach, as other regions on chromosome 3 may also be subject to positive selection. Remarkably, experiments using human embryonic kidney (HEK) 293-transfected cells have shown possible effects of SNP rs6280 on gene function, the Gly variant having higher dopamine affinity and increased signaling responses (Jeanneteau et al. 2006). Thus, SNP rs6280 is most probably the actual polymorphism subject to selection at this haplotype. The selective force acting on *DRD3* is totally unknown, as it is involved in several phenotypic traits that might have experienced recent positive selection, such as cognition or salt retention (Asico et al. 1998; Bombin et al. 2008; Laszy et al. 2005). Interestingly, another dopamine receptor, *DRD4*, has been also subject to recent positive selection (Ding et al. 2002; Wang et al. 2004).

Finally, we have shown that the protective haplotype is linked to the haplotype subject to recent selection and, therefore, it has reached intermediate frequencies by hitchhiking. This recent positive selection seems not to be present in Sub-Saharan African populations. As a consequence, the frequency of the protective haplotype varies considerably among different populations around the world (Fig. 4). Therefore, the impact of this haplotype in protection from schizophrenia, assuming the same OR in all populations, may be quite different at different geographic regions. The low frequency of the haplotype in African populations represents a difficulty for replication studies in these populations.

Our finding may be considered of special interest in the context of the current debate related to the relative importance of the common disease/common variant versus the common disease/rare variant hypotheses in schizophrenia (Craddock et al. 2007; Keller and Miller 2006; McClellan et al. 2007; Walsh et al. 2008), because it constitutes one of the few examples of putative schizophrenia candidate genes with alleles affecting susceptibility that have reached high frequencies due to the action of natural selection. The other putative cases, *GABRB2* and *MAOB*, seem to fit the ancestral susceptibility hypothesis, which proposed that some alleles may be currently disease susceptibility variants due to a miss-adaptation between the ancestral allele and the new environment (Carrera et al. 2008; Di Rienzo and Hudson 2005; Lo et al. 2007). Thus, the analysis of functional candidate genes subject to recent natural selection might render new common susceptibility (or protective) alleles in future research.

Acknowledgments This work was supported by grants PI052036 to JC and PI050842 to EV from Ministerio de Sanidad y Consumo, Spain, and PIGDIT04PXC20802 to AC from Xunta de Galicia. NC was supported by a fellowship from Xunta de Galicia. We thank Dr. V. L. Nimgaonkar (University of Pittsburgh) for sharing the exact numbers of each genotype in their data from the Pittsburgh samples, and to Dr. R. Tarrío (University of Santiago de Compostela) for critical reading of an early version of the manuscript. Genotyping was performed at the Centro Nacional de genotipado (CeGen).

Conflict of interest statement The authors reported no biomedical financial interests or potential conflicts of interest.

References

- Abramson JH (2004) WINPEPI (PEPI-for-Windows): computer programs for epidemiologists. *Epidemiol Perspect Innov* 1:6
- Asico LD, Ladines C, Fuchs S, Accili D, Carey RM, Semeraro C, Pochiari F, Felder RA, Eisner GM, Jose PA (1998) Disruption of the dopamine D3 receptor gene produces renin-dependent hypertension. *J Clin Invest* 102:493–498
- Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265
- Bombin I, Arango C, Mayoral M, Castro-Fornieles J, Gonzalez-Pinto A, Gonzalez-Gomez C, Moreno D, Parellada M, Baeza I, Graell M et al (2008) *DRD3*, but not *COMT* or *DRD2*, genotype affects executive functions in healthy and first-episode psychosis adolescents. *Am J Med Genet B Neuropsychiatr Genet* 147B(6):873–879
- Brown S (1997) Excess mortality of schizophrenia. A metaanalysis. *Br J Psychiatry* 171:502–508
- Brune M (2004) Schizophrenia-an evolutionary enigma? *Neurosci Biobehav Rev* 28:41–53
- Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, Bodmer J, Bodmer WF, Bonne-Tamir B, Cambon-Thomsen A et al (2002) A human genome diversity cell line panel. *Science* 296:261–262
- Carrera N, Sanjuán J, Moltó MD, Carracedo A, Costas J (2008) Recent adaptive selection at *MAOB* and ancestral susceptibility to schizophrenia. *Am J Med Genet B Neuropsychiatr Genet* (in press). doi:10.1002/ajmg.b.30823

- Costas J, Torres M, Cristobo I, Phillips C, Carracedo A (2005) Relative efficiency of the linkage disequilibrium mapping approach in detecting candidate genes for schizophrenia in different European populations. *Genomics* 86:280–286
- Craddock N, O'Donovan MC, Owen MJ (2007) Phenotypic and genetic complexity of psychosis. Invited commentary on: Schizophrenia: a common disease caused by multiple rare alleles. *Br J Psychiatry* 190:200–203
- Crespi B, Summers K, Dorus S (2007) Adaptive evolution of genes underlying schizophrenia. *Proc Biol Sci* 274:2801–2810
- Di Rienzo A, Hudson RR (2005) An evolutionary framework for common diseases: the ancestral-susceptibility model. *Trends Genet* 21:596–601
- Ding YC, Chi HC, Grady DL, Morishima A, Kidd JR, Kidd KK, Flodman P, Spence MA, Schuck S, Swanson JM et al (2002) Evidence of positive selection acting at the human dopamine receptor D4 gene locus. *Proc Natl Acad Sci USA* 99:309–314
- Dominguez E, Loza MI, Padin F, Gesteira A, Paz E, Paramo M, Brenlla J, Pumar E, Iglesias F, Cibeira A et al (2007) Extensive linkage disequilibrium mapping at HTR2A and DRD3 for schizophrenia susceptibility genes in the Galician population. *Schizophr Res* 90:123–129
- Dudbridge F (2003) Pedigree disequilibrium tests for multilocus haplotypes. *Genet Epidemiol* 25:115–121
- Dupont WD, Plummer WD (1997) PS power and sample size program available for free on the Internet. *Control Clin Trials* 18:274
- Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M et al (2002) The structure of haplotype blocks in the human genome. *Science* 296:2225–2229
- 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* 160:460–463
- Huxley J, Mayr E, Osmond H, Hoffer A (1964) Schizophrenia as a genetic morphism. *Nature* 204:220–221
- International HapMap Consortium (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449:851–861
- Jeanneteau F, Funalot B, Jankovic J, Deng H, Lagarde JP, Lucotte G, Sokoloff P (2006) A functional variant of the dopamine D3 receptor is associated with risk and age-at-onset of essential tremor. *Proc Natl Acad Sci USA* 103:10753–10758
- Jonsson EG, Kaiser R, Brockmoller J, Nimgaonkar VL, Crocq MA (2004) Meta-analysis of the dopamine D3 receptor gene (DRD3) Ser9Gly variant and schizophrenia. *Psychiatr Genet* 14:9–12
- Keller MC, Miller G (2006) Resolving the paradox of common, harmful, heritable mental disorders: which evolutionary genetic models work best? *Behav Brain Sci* 29:385–404 Discussion 405–452
- Laszky J, Laszlovszky I, Gyertyán I (2005) Dopamine D3 receptor antagonists improve the learning performance in memory-impaired rats. *Psychopharmacology (Berl)* 179:567–575
- Lo W-S, Xu ZW, Yu ZL, Pun FW, Ng S-K, Chen J, Tong KL, Zhao C, Xu X, Tsang SY et al (2007) Positive selection within the schizophrenia-associated GABAA receptor b2 gene. *PLoS ONE* 2:e462
- Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 33:177–182
- Ma G, He Z, Fang W, Tang W, Huang K, Li Z, He G, Xu Y, Feng G, Zheng T et al (2008) The Ser9Gly polymorphism of the dopamine D3 receptor gene and risk of schizophrenia: an association study and a large meta-analysis. *Schizophr Res* 101:26–35
- McClellan JM, Susser E, King MC (2007) Schizophrenia: a common disease caused by multiple rare alleles. *Br J Psychiatry* 190:194–199
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Rajeevan H, Osier MV, Cheung KH, Deng H, Druskin L, Heinzen R, Kidd JR, Stein S, Pakstis AJ, Tosches NP et al (2003) ALFRED: the ALElle FREquency Database. Update. *Nucleic Acids Res* 31:270–271
- Roig B, Virgos C, Franco N, Martorell L, Valero J, Costas J, Carracedo A, Labad A, Vilella E (2007) The discoidin domain receptor 1 as a novel susceptibility gene for schizophrenia. *Mol Psychiatry* 12:833–841
- Rosenberg NA (2006) Standardized subsets of the HGDP-CEPH Human Genome Diversity Cell Line Panel, accounting for atypical and duplicated samples and pairs of close relatives. *Ann Hum Genet* 70:841–847
- Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ et al (2002) Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419:832–837
- Saha S, Chant D, Welham J, McGrath J (2005) A systematic review of the prevalence of schizophrenia. *PLoS Med* 2:e141
- Sand PG, Langguth B, Prikryl R, Kucerova H, Ceskova E (2008) A putative DRD3 schizophrenia risk haplotype deconstructed. *Biol Psychiatry* 63(3):e21 Author reply e23–25
- Sokoloff P, Giros B, Martres M-P, Bouthenet M-L, Schwartz J-C (1990) Molecular cloning and characterization of a novel dopamine receptor (D-3) as a target for neuroleptics. *Nature* 347:146–151
- Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68:978–989
- Sullivan PF, Kendler KS, Neale MC (2003) Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry* 60:1187–1192
- Talkowski ME, Mansour H, Chowdari KV, Wood J, Butler A, Varma PG, Prasad S, Senwal P, Bhatia T, Deshpande S et al (2006) Novel, replicated associations between dopamine D3 receptor gene polymorphisms and schizophrenia in two independent samples. *Biol Psychiatry* 60:570–577
- 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 E, Ding YC, Flodman P, Kidd JR, Kidd KK, Grady DL, Ryder OA, Spence MA, Swanson JM, Moyzis RK (2004) The genetic architecture of selection at the human dopamine receptor D4 (DRD4) gene locus. *Am J Hum Genet* 74:931–944

**Estudio 2. Modelo de susceptibilidad ancestral en
esquizofrenia: *MAOB***

Recent Adaptive Selection at *MAOB* and Ancestral Susceptibility to Schizophrenia

Noa Carrera,¹ Julio Sanjuán,² María Dolores Moltó,³ Ángel Carracedo,^{1,4} and Javier Costas^{1*}

¹Fundación Pública Galega de Medicina Xenómica (FPGMX), University of Santiago Hospital Complex, Santiago de Compostela, Spain

²Psychiatric Unit, Valencia Medical School, CIBERSAM, Valencia, Spain

³Faculty of Biology, Genetic Department, Valencia University, Valencia, Spain

⁴Genomic Medicine Group, Institute of Legal Medicine, University of Santiago de Compostela, Centro Nacional de Genotipado (CeGen), CIBERER, Santiago de Compostela, Spain

Received 6 March 2008; Accepted 28 May 2008

The ancestral susceptibility hypothesis has been proposed to explain the existence of susceptibility alleles to common diseases. Some ancestral alleles, reflecting ancient adaptations, may be poorly adapted to the more contemporary environmental conditions giving rise to an increased risk to suffer some common disorders. In order to test this hypothesis in schizophrenia, we focused on the monoamine oxidase B gene (*MAOB*). This gene is involved in deamination of several monoamines, including both xenobiotic amines present in several foods, as well as neurotransmitters such as dopamine. In addition, preliminary analysis based on phase I HapMap data suggested that recent natural selection has acted on this locus. We further explored the existence of this recent positive selection using a test based on extension of linkage disequilibrium (LD) to large distance at the specific selected haplotype taking data from HapMap phase II, and searched for association of the ancestral haplotypes with schizophrenia in a sample of 532 schizophrenic patients and 597 controls from Spain. Our analysis suggested the existence of a haplotype of *MAOB* subject to recent selection. In agreement with the ancestral susceptibility hypothesis, the ancestral haplotypes were significantly over-represented in patients ($P=0.047$). These haplotypes conferred an increased risk to schizophrenia, restricted to males ($P=0.024$, OR = 1.41, 95% CI 1.01–1.90). Thus, pending on replication studies, *MAOB* seems to fit the ancestral susceptibility model, validating a new strategy to search for common schizophrenia susceptibility genes by focusing in those functional candidate genes subject to recent positive selection. © 2008 Wiley-Liss, Inc.

Key words: dopaminergic hypothesis; case–control studies; monoamine oxidase; natural selection; common disease–common variant hypothesis

INTRODUCTION

Schizophrenia is a common mental disorder, with a lifetime prevalence of 0.72% [Saha et al., 2005]. It often starts in young adulthood, and is associated with reduced fertility [Haukka et al., 2003], increased rates of suicide [Saha et al., 2007], and higher rates

How to Cite this Article:

Carrera N, Sanjuán J, Moltó MD, Carracedo Á, Costas J. 2009. Recent Adaptive Selection at *MAOB* and Ancestral Susceptibility to Schizophrenia.

Am J Med Genet Part B 150B:369–374.

of mortality before adulthood in offspring of schizophrenic mothers [Suvisaari et al., 2007]. Genetic epidemiology studies revealed heritability values as high as 80% in schizophrenia [Cardno et al., 1999]. Thus, the existence of common alleles of disease susceptibility may be considered an evolutionary paradox, firstly discussed in 1964, but still unresolved nowadays [Huxley et al., 1964; Keller and Miller, 2006].

Recently, the ancestral susceptibility model has been proposed to explain the involvement of common susceptibility alleles in common diseases [Di Rienzo and Hudson, 2005; Di Rienzo, 2006]. This evolutionary model, that may help to resolve the schizophrenia paradox, proposes that ancestral alleles adapted to past environments may be poorly adapted to the new one associated to a shift in lifestyle of human populations. Therefore, they might give rise to an increased susceptibility to common diseases. Currently, there are few examples fitting this model, mainly genes involved in energy metabolism or salt homeostasis, in agreement with the “thirsty genotype” and the “sodium-sensitivity” hypotheses [Di Rienzo, 2006].

Grant sponsor: Ministerio de Sanidad y Consumo, Spain; Grant numbers: PI052036, FIS 02/0018.

*Correspondence to:

Javier Costas, Fundación Pública Galega de Medicina Xenómica-SERGAS, Hospital Clínico Universitario de Santiago, Edif. Consultas Planta-2, E-15706 Santiago de Compostela, Spain. E-mail: bfcostas@usc.es

Published online 13 June 2008 in Wiley InterScience

(www.interscience.wiley.com)

DOI 10.1002/ajmg.b.30823

Over the last years huge amount of single nucleotide polymorphism (SNP) data have been generated along the human genome in different populations. It allowed whole-genome scans in order to find genomic regions subject to recent adaptive selection [Costas et al., 2005; The International HapMap Consortium, 2005; Voight et al., 2006; Wang et al., 2006]. These scans are based on the identification of signatures of incomplete selective sweeps, such as common haplotypes with long range linkage disequilibrium (LD). Such haplotypes are indicative of fast increase in frequency precluding that recombination breaks down the long haplotype [Sabeti et al., 2002, 2006]. These regions subject to incomplete selective sweeps, thus bearing functionally distinct alleles, are the ideal substrate for the ancestral susceptibility model.

One of these whole-genome scans for fingerprints of recent positive selection, based on the HapMap phase I data, identified monoamine oxidase B gene (*MAOB*) as one of the genes with the strongest signal of recent selective sweep [Table 10 of The International HapMap Consortium, 2005]. Monoamine oxidases are enzymes that catalyze the oxidative deamination of several biogenic amines, including neurotransmitters such as serotonin, norepinephrine, and dopamine, or the neuromodulator phenylethylamine [Shih et al., 1999]. The two genes coding monoamine oxidases within the human genome, *MAOA* and *MAOB*, show different affinities for these substrates. *MAOB* has high affinity for phenylethylamine and dopamine. Thus, it may be considered as a functional candidate for schizophrenia susceptibility from the viewpoint of the well-established dopaminergic hypothesis [Winterer and Weinberger, 2004; Iversen and Iversen, 2007; Talkowski et al., 2007]. Nevertheless, *MAOB* has been poorly studied in regard to its putative role in schizophrenia risk [Talkowski et al., 2007]. Thus, Coron et al. [1996] analyzed one intronic SNP in *MAOB* in 110 cases and 87 controls with negative results. Wei and Hemmings [1999] studied one microsatellite at *MAOA* and another at *MAOB* in 140 schizophrenic males and 91 controls and they did not find association with schizophrenia in allelic tests, although an specific haplotype combining both markers seems to be overrepresented in cases (OR = 4.05; 95% CI 1.15–14.26; Fisher's $P = 0.011$). Finally, Hoogendoorn et al. [2005] analyzed two microsatellites at *MAOB* in 282 schizophrenic samples and 582 controls and did not find significant differences in frequency. These three studies were performed on samples from European origin. Taking into account these facts, we have chosen *MAOB* to test the ancestral susceptibility model in schizophrenia.

MATERIALS AND METHODS

Samples

Two independent case–control cohorts were available for the present study. The first one consists of 258 unrelated cases (64.0% males) and 305 unrelated controls (56.7% males) from Galicia (NW Spain). The second one is constituted by 274 unrelated patients (65.7% males) and 292 unrelated controls (61.0% males) from Valencia (Central East Spain).

The cases were diagnosed as schizophrenic patients according to DSM-IV criteria by experimented psychiatrists. The control samples were from blood donors as well as hospital staff. Both patients and

controls were of Spanish family origin. All samples were included in the study after participants signed written informed consent forms. The study was approved by the appropriate Ethic Committees.

Assuming the frequencies of HapMap data, the sample size reaches 80% power to detect a susceptibility allele conferring a risk higher than 1.47 or 1.53, in allelic tests of female and male samples, respectively.

Genotyping

SNP rs5905512 was genotyped by mass spectrometry using the Sequenom MassArray technology following manufacture's instructions. The Galician samples were genotyped twice, with a reproducibility rate of 99.8%. Three putative males appeared as heterozygotes and were excluded from the study. The SNP was in Hardy–Weinberg equilibrium ($P > 0.05$) in both samples.

Long Range Haplotype (LRH) Test

The LRH test is specifically designed to detect common haplotypes showing an LD pattern around them higher than expected by neutrality, taking into account the haplotype frequency. Basically, haplotype frequency is proportional to its age, and then, more common haplotypes have been subject to recombination for a longer period, leading to a break down of LD. Nevertheless, haplotypes subject to recent positive selection increase their frequency very fast, reducing the chance that recombination can destroy LD.

LRH tests were performed on the phased haplotype data from the HapMap CEU (CEPH, Utah residents with European ancestry) sample phase II, using the Sweep program [Sabeti et al., 2002]. We first identified core haplotypes at *MAOB*, that is, each one of the haplotypes present at each haplotype block. We calculated the extended haplotype homozygosity (EHH) for each core haplotype, that is, the probability that two randomly chosen chromosomes carrying the core haplotype of interest are identical for the entire interval from the core region to a specified distance [Sabeti et al., 2002]. Finally, we calculated the relative EHH (rEHH), that is, the ratio of the EHH on the tested core haplotype compared with the EHH of the grouped set of core haplotypes at the region not including the core haplotype tested [Sabeti et al., 2002]. By this way, the influence of different recombination rates along the genome in EHH is adjusted using the other haplotypes of the same region. We measured the rEHH value of the putative selected haplotype at three different genetic distances (0.1, 0.2, and 0.3 cM) and compared it with the distribution of rEHH values of all core haplotypes at similar frequencies (0.50–0.55) along the X chromosome.

Statistical Analysis

Deviation of Hardy–Weinberg (HW) equilibrium was tested in control females using Pearson's chi-squared tests. Allelic frequencies between cases and controls were compared using the Mantel–Haenszel test, as implemented in WINPEPI [Abramson, 2004], to consider samples from two distinct geographic regions. As *MAOB* is located on the X chromosome, the comparisons were done independently for each sex.

RESULTS

According to the International HapMap Consortium, the *MAOB* selected haplotype is centered at position 42,763,073 (NCBI Build 34) on the X chromosome and has a frequency of 0.53 [The International HapMap Consortium, 2005]. Figure 1 shows the location of this haplotype on *MAOB* gene structure, as well as the different haplotypes identified in this region with their frequencies. The selected haplotype spans from intron I to intron III of *MAOB*. There is one SNP, rs5905512, within intron I that is a perfect proxy of this haplotype, at least in the CEU population of HapMap. The derived allele, allele A, is the one linked to the selected haplotype. This result was based on the phase I data of HapMap.

The analysis of HapMap phase II data, with higher density of SNPs, showed *MAOB* out of the 200 more significant results for recent positive selection [The International HapMap Consortium, 2007]. Therefore, we used the new data from HapMap to search for any signature of selection on this haplotype. As shown in Figure 2A, the haplotype marked by the rs5905512 A allele shows an unusually high rEHH at long distances upstream, taking into account its frequency. Comparison of the distribution of rEHH values with haplotypes at similar frequencies, from 0.5 to 0.55, along the entire X chromosome revealed that the *MAOB* haplotype is an outlier of the empirical distribution, being at the 97.1th, 96.9th, and 95.2th percentile ranks at 0.1, 0.2, and 0.3 cM, respectively (Fig. 2B).

In order to test the ancestral susceptibility hypothesis, SNP rs5905512, perfect marker of the selected haplotype, was genotyped in a sample of 532 schizophrenic patients and 597 controls from two different Spanish autonomous regions, Galicia and Valencia. We specifically examined the prediction of the hypothesis, that is, there may be an excess of the ancestral, non-selected allele in patients. The results revealed that this was the case, the ancestral allele was significantly over-represented in the schizophrenic sample (one-tailed Fisher's exact test for stratified data

$P=0.047$). Subsequently, we performed allelic test of association by sex, comprising 343 male patients and 351 male controls, and 186 female patients and 246 female controls (Table I). There were a significant association in the male samples (P value = 0.024), indicating an increased susceptibility to schizophrenia in the presence of the non-selected variant, conferring an OR of 1.41 (95% CI 1.01–1.90). This association was not detected in the female samples (OR = 1.04, 95% CI 0.79–1.36).

DISCUSSION

The ancestral susceptibility model has been proposed as a way to explain the existence of susceptibility alleles to common diseases that reached high frequency, that is, became polymorphic, in populations [Di Rienzo and Hudson, 2005; Di Rienzo, 2006]. To fit this model, a locus has to be subject to very recent natural selection, such as the selected variant is still unfixated and segregating with ancestral variants; and the old variants have to increase susceptibility to a common disorder. Here, we show that *MAOB* conforms to this model.

Thus, this gene is one of the clearest examples of incomplete selective sweep in data from the HapMap phase I [The International HapMap Consortium, 2005]. The derived variant of the SNP rs5905512, the A allele, is a perfect proxy of the selected haplotype in CEU. Its frequency varies from 0.11 in YRI (Yoruba in Ibadan, Nigeria) to 0.80 in JPT (Japanese in Tokyo). When we analyzed the HapMap phase II data, this haplotype appeared at one extreme of the empirical rEHH distribution of haplotypes at similar frequencies on the X chromosome (Fig. 2). Thus, although the evidence of selection was not as strong as suggested from phase I, the rEHH of the haplotype is above the 95th percentile rank of the empirical distribution at several different distances. Taking into account recent work suggesting the widespread nature of adaptive selection in recent evolutionary history, associated to the colonization of

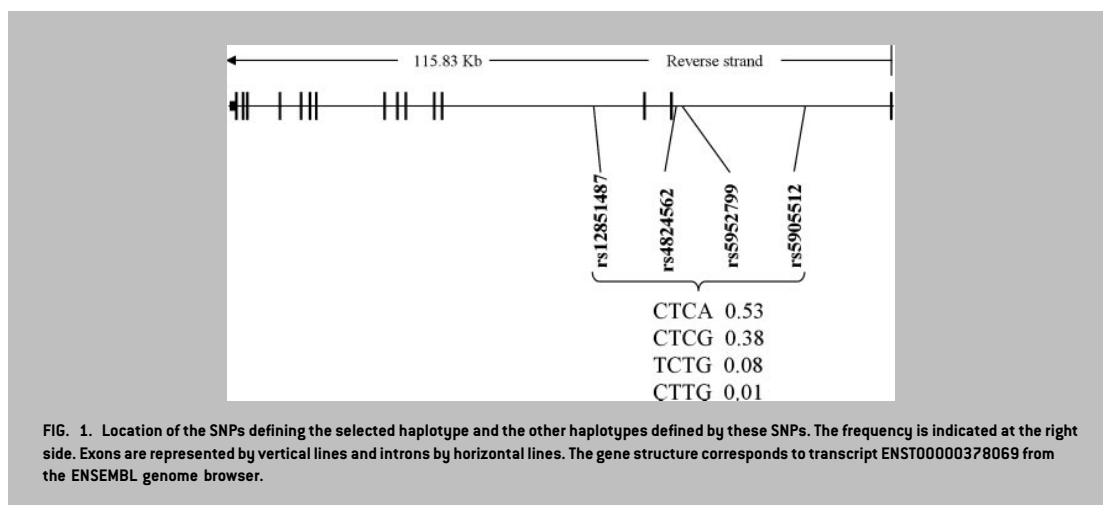


FIG. 1. Location of the SNPs defining the selected haplotype and the other haplotypes defined by these SNPs. The frequency is indicated at the right side. Exons are represented by vertical lines and introns by horizontal lines. The gene structure corresponds to transcript ENST00000378069 from the ENSEMBL genome browser.

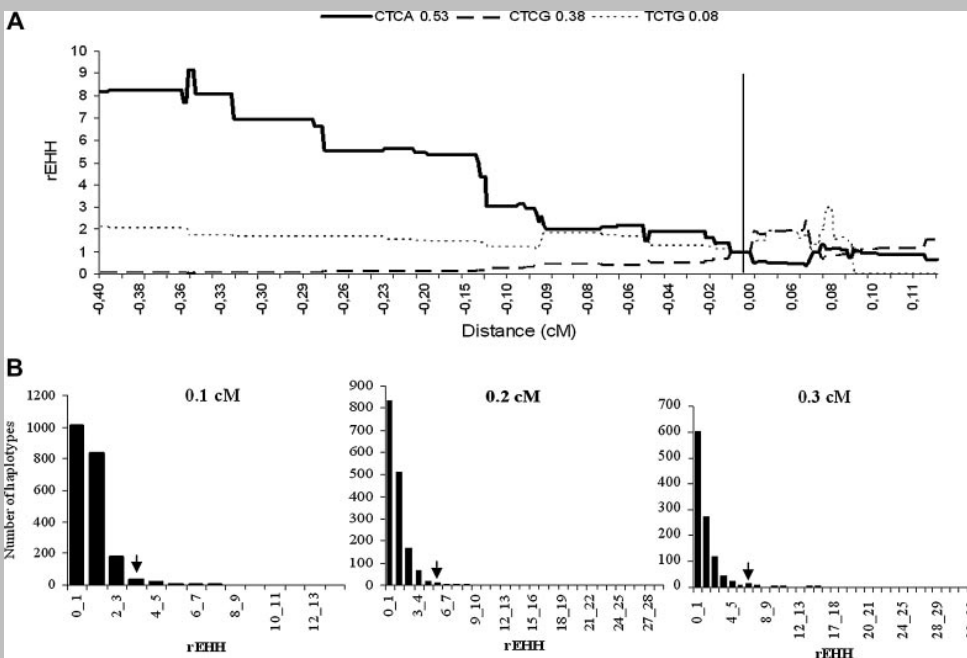


FIG. 2. Results of the LRH test to detect recent adaptive selection on the X chromosome. **A:** Variation of rEHH with distance from the core haplotypes at the *MAOB* selected region. Only haplotypes at frequency higher than 5% are shown. **B:** Distribution of rEHH at different distances from all the core haplotypes on the X chromosome at frequencies 0.50–0.55. Arrows indicate the range of the selected haplotype CTCA at *MAOB*.

novel environments during the expansion of humans out of Africa, *MAOB* seems to remain as a valid candidate of a recently selected locus [Hawks et al., 2007]. Furthermore, *MAOB* catalyses the oxidative deamination of both biogenic monoamines, mainly dopamine, as well as xenobiotic amines, present in a large variety of foods and beverages [Strolin Benedetti et al., 2007]. Thus, it is

involved in traits exposed to major environmental changes, such as cultural behavior or dietary adaptation. Genes involved in these traits are very prone to be subject to recent selection [Vallender and Lahn, 2004]. In fact, the other monoamine oxidase of the human genome, *MAOA*, has also been subject to recent positive selection [Gilad et al., 2002]. As these two genes are very close in the genome,

TABLE I. Case–Control Association Analysis

	Cases [%]		Controls [%]		P value ^a	OR (95% CI) ^a
	A	G	A	G		
Males (N = 694^b)						
Santiago	64 (39.3)	99 (60.7)	86 (49.7)	87 (50.3)		
Valencia	88 (49.0)	92 (51.0)	99 (56.0)	79 (44.0)		
Total	152 (44.3)	191 (55.7)	185 (52.7)	166 (47.3)	0.024	1.41 (1.01–1.90)
Females (N = 864^b)						
Santiago	80 (43.0)	106 (57.0)	130 (49.2)	134 (50.8)		
Valencia	102 (55.0)	84 (45.0)	114 (50.0)	114 (50.0)		
Total	182 (48.9)	190 (51.1)	244 (49.6)	248 (50.4)	0.800	1.04 (0.79–1.36)

^aMantel–Haenszel test.
^bNumber of alleles.

it may be possible that the detected selective sweeps represent just one event of adaptation. The selective pressures acting on *MAOB* did not necessarily target the disease phenotype per se, but rather phenotypes related to normal processes, such as response to xenobiotics, social cognition or creativity. Schizophrenia susceptibility at the locus may be a by-product of this selection [Crespi et al., 2007].

In addition, our case-control association study revealed that ancestral variants of *MAOB*, defined by the G allele at rs5905512, are over-represented in schizophrenic patients. Concretely, the ancestral haplotypes conferred an increased risk to schizophrenia restricted to males (Table I). *MAOB* expression differs by sex [Snell et al., 2002], and is regulated by ovarian steroid hormones [Gundlah et al., 2002]. There are considerable differences by gender in relation to schizophrenia onset, symptoms, course, or treatment, among others, and it has been suggested that these differences may be mediated by estrogens [Salem and Kring, 1998; Halbreich and Kahn, 2003]. It has been reported that estradiol may modulate dopaminergic transmission at various levels, such as release, metabolism, receptors, or transporter, in experimental animals [Cyr et al., 2000]. Our result suggests that *MAOB* might be one of several factors involved in differences in schizophrenia between sexes. Interestingly, another gene involved in dopamine catabolism and whose expression is regulated by estrogen, *COMT*, also seems to confer susceptibility to schizophrenia restricted to one sex [Xie et al., 1999; Shifman et al., 2002].

Our finding may be considered of special interest in the context of the current debate common disease/common variant versus common disease/rare variant in schizophrenia, and by extension, in other common psychiatric diseases [Keller and Miller, 2006; Craddock et al., 2007; Crow, 2007; McClellan et al., 2007a,b]. It may be possible that most biologically functional candidate genes do not present common variants differing in functionality. In fact, both the results of the first whole-genome scan in schizophrenia as well as a recent study of a large European sample using the more generally accepted schizophrenia susceptibility genes, suggested that common susceptibility variants may be very infrequent [Lencz et al., 2007; Sanders et al., 2008]. Here, we follow a new approach to search for common risk alleles, in agreement to the ancestral susceptibility hypothesis [Di Rienzo and Hudson, 2005; Di Rienzo, 2006]. This approach, suggested previously by Crespi et al. [2007], is based on the a priori identification of candidate genes that present functionally different common variants, detected by the signature of very recent adaptive selection. Focusing on this specific subset of candidate genes, instead of all putative candidates, and performing just one test per locus based on functionally different variants, is expected to increase the chance to find common variants of susceptibility to schizophrenia. In this context, it is interesting to highlight that other schizophrenia candidate genes have been subject to recent adaptive selection, such as *GRM3*, *DTNBP1*, *SLC6A4*, or *GABRB2*, among others [Ding et al., 2002; Voight et al., 2006; Crespi et al., 2007; Lo et al., 2007]. It has been suggested that *GABRB2* may also fit the ancestral susceptibility model, although the results of association analyses are somewhat controversial [Lo et al., 2007].

Our study presents several limitations. The probability of an outlier in rEHH distribution being a selected gene or a neutral one

depends on several unknown parameters such as the strength of selection and the proportion of all loci subject to selection [Kelley et al., 2006; Sabeti et al., 2006]. Thus, we cannot calculate a precise posterior probability of selection. Although the use of empirical distributions, rather than theoretical models, is a clear advantage, convincing proof of recent selection will require an understanding of biological function as well as the identification of the functional change at the DNA sequence [Sabeti et al., 2006]. In addition, our sample size has limited power to detect very small genetic effects. Thus, we must consider the possibility that both the negative finding in our female sample as well as the positive finding in our male sample may be due to chance. Further studies are needed to confirm the association of the *MAOB* ancestral variants with schizophrenia risk, as well as to investigate the ancestral susceptibility model in other schizophrenia candidate genes.

ACKNOWLEDGMENTS

This work was supported by grants PI052036 to JC and FIS 02/0018 from Ministerio de Sanidad y Consumo, Spain. We are grateful to the staff of the Centro de Transfusión de Galicia, Unidade de Saúde Mental do Hospital Provincial de Santiago de Compostela, Hospital Psiquiátrico de Conxo and Complexo Hospitalario de Ourense, Dr. Ana Vega (FPGMX), and Dr. Eduardo Domínguez (USC) for providing samples. We thank the patients for their participation. NC was supported by a fellowship from Xunta de Galicia.

REFERENCES

- Abramson JH. 2004. WINPEPI (PEPI-for-Windows): Computer programs for epidemiologists. *Epidemiol Perspect Innov* 1(1):6.
- Cardno AG, Marshall EJ, Coid B, Macdonald AM, Ribchester TR, Davies NJ, Venturi P, Jones LA, Lewis SW, Sham PC, et al. 1999. Heritability estimates for psychotic disorders: The Maudsley twin psychosis series. *Arch Gen Psychiatry* 56(2):162–168.
- Coron B, Campion D, Thibaut F, Dollfus S, Preterre P, Langlois S, Vasse T, Moreau V, Martin C, Charbonnier F, et al. 1996. Association study between schizophrenia and monoamine oxidase A and B DNA polymorphisms. *Psychiatry Res* 62(3):221–226.
- Costas J, Salas A, Phillips C, Carracedo A. 2005. Human genome-wide screen of haplotype-like blocks of reduced diversity. *Gene* 349:219–225.
- Craddock N, O'Donovan MC, Owen MJ. 2007. Phenotypic and genetic complexity of psychosis. Invited commentary on . Schizophrenia: A common disease caused by multiple rare alleles. *Br J Psychiatry* 190: 200–203.
- Crespi B, Summers K, Dorus S. 2007. Adaptive evolution of genes underlying schizophrenia. *Proc Biol Sci* 274(1627):2801–2810.
- Crow T. 2007. Genetic hypotheses for schizophrenia. *Br J Psychiatry* 191:180; author reply 180–1.
- Cyr M, Calon F, Morissette M, Grandbois M, Di Paolo T, Callier S. 2000. Drugs with estrogen-like potency and brain activity: Potential therapeutic application for the CNS. *Curr Pharm Des* 6(12):1287–1312.
- Di Rienzo A. 2006. Population genetics models of common diseases. *Curr Opin Genet Dev* 16(6):630–636.

- Di Rienzo A, Hudson RR. 2005. An evolutionary framework for common diseases: The ancestral-susceptibility model. *Trends Genet* 21(11):596–601.
- Ding YC, Chi HC, Grady DL, Morishima A, Kidd JR, Kidd KK, Flodman P, Spence MA, Schuck S, Swanson JM, et al. 2002. Evidence of positive selection acting at the human dopamine receptor D4 gene locus. *Proc Natl Acad Sci USA* 99(1):309–314.
- Gilad Y, Rosenberg S, Przeworski M, Lancet D, Skorecki K. 2002. Evidence for positive selection and population structure at the human MAO-A gene. *Proc Natl Acad Sci USA* 99(2):862–867.
- Gundlach C, Lu NZ, Bethea CL. 2002. Ovarian steroid regulation of monoamine oxidase-A and -B mRNAs in the macaque dorsal raphe and hypothalamic nuclei. *Psychopharmacology (Berl)* 160(3):271–282.
- Halbreich U, Kahn LS. 2003. Hormonal aspects of schizophrenias: An overview. *Psychoneuroendocrinology* 28(Suppl 2):1–16.
- 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.
- Hawks J, Wang ET, Cochran GM, Harpending HC, Moyzis RK. 2007. Recent acceleration of human adaptive evolution. *Proc Natl Acad Sci USA* 104(52):20753–20758.
- Hoogendoorn ML, Bakker SC, Schnack HG, Selten JP, Otten HG, Verduijn W, van der Heijden FM, Pearson PL, Kahn RS, Sinke RJ. 2005. No association between 12 dopaminergic genes and schizophrenia in a large Dutch sample. *Am J Med Genet B Neuropsychiatr Genet* 134(1):6–9.
- Huxley J, Mayr E, Osmond H, Hoffer A. 1964. Schizophrenia as a Genetic Morphism. *Nature* 204:220–221.
- Iversen SD, Iversen LL. 2007. Dopamine: 50 years in perspective. *Trends Neurosci* 30(5):188–193.
- Keller MC, Miller G. 2006. Resolving the paradox of common, harmful, heritable mental disorders: Which evolutionary genetic models work best? *Behav Brain Sci* 29(4):385–404; discussion 405–52.
- Kelley JL, Madeoy J, Calhoun JC, Swanson W, Akey JM. 2006. Genomic signatures of positive selection in humans and the limits of outlier approaches. *Genome Res* 16(8):980–989.
- Lencz T, Morgan TV, Athanasiou M, Dain B, Reed CR, Kane JM, Kucherpapati R, Malhotra AK. 2007. Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. *Mol Psychiatry* 12(6):572–580.
- Lo WS, Xu Z, Yu Z, Pun FW, Ng SK, Chen J, Tong KL, Zhao C, Xu X, Tsang SY, et al. 2007. Positive selection within the Schizophrenia-associated GABA(A) receptor beta2 gene. *PLoS ONE* 2(5):e462.
- McClellan JM, Susser E, King MC. 2007a. Schizophrenia: A common disease caused by multiple rare alleles. *Br J Psychiatry* 190:194–199.
- McClellan JM, Susser E, King MC. 2007b. Authors' reply to: Genetic hypotheses for schizophrenia. *Br J Psychiatry* 191:180–181.
- Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, et al. 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419(6909):832–837.
- Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varylly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES. 2006. Positive natural selection in the human lineage. *Science* 312(5780):1614–1620.
- Saha S, Chant D, Welham J, McGrath J. 2005. A systematic review of the prevalence of schizophrenia. *PLoS Med* 2(5):e141.
- 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.
- Salem JE, Kring AM. 1998. The role of gender differences in the reduction of etiologic heterogeneity in schizophrenia. *Clin Psychol Rev* 18(7):795–819.
- Sanders AR, Duan J, Levinson DF, Shi J, He D, Hou C, Burrell GJ, Rice JP, Nertney DA, Olincy A, et al. 2008. No significant association of 14 candidate genes with schizophrenia in a large European ancestry sample: Implications for psychiatric genetics. *Am J Psychiatry* 165:497–506.
- Shifman S, Bronstein M, Sternfeld M, Pisante-Shalom A, Lev-Lehman E, Weizman A, Reznik I, Spivak B, Grisaru N, Karp L, et al. 2002. A highly significant association between a COMT haplotype and schizophrenia. *Am J Hum Genet* 71(6):1296–1302.
- Shih JC, Chen K, Ridd MJ. 1999. Monoamine oxidase: From genes to behavior. *Annu Rev Neurosci* 22:197–217.
- Snell LD, Glanz J, Tabakoff B. 2002. Relationships between effects of smoking, gender, and alcohol dependence on platelet monoamine oxidase-B: Activity, affinity labeling, and protein measurements. *Alcohol Clin Exp Res* 26(7):1105–1113.
- Strolin Benedetti M, Tipton KF, Whomsley R. 2007. Amine oxidases and monoxygenases in the in vivo metabolism of xenobiotic amines in humans: Has the involvement of amine oxidases been neglected? *Fundam Clin Pharmacol* 21(5):467–480.
- Suvisaari J, Hakkinen L, Haukka J, Lonnqvist J. 2007. Mortality in offspring of mothers with psychotic disorder. *Psychol Med*: 1–8 (in press). DOI: 10.1017/S0033291707002383.
- Talkowski ME, Bamne M, Mansour H, Nimgaonkar VL. 2007. Dopamine genes and schizophrenia: Case closed or evidence pending? *Schizophr Bull* 33(5):1071–1081.
- The International HapMap Consortium. 2005. A haplotype map of the human genome. *Nature* 437(7063):1299–1320.
- The International HapMap Consortium. 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449(7164):851–861.
- Vallender EJ, Lahn BT. 2004. Positive selection on the human genome. *Hum Mol Genet* 13(SpecNo 2):R245–R254.
- Voight BF, Kudravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. *PLoS Biol* 4(3):e72.
- Wang ET, Kodama G, Baldi P, Moyzis RK. 2006. Global landscape of recent inferred Darwinian selection for Homo sapiens. *Proc Natl Acad Sci USA* 103(1):135–140.
- Wei J, Hemmings GP. 1999. A study of linkage disequilibrium between polymorphic loci for monoamine oxidases A and B in schizophrenia. *Psychiatr Genet* 9(4):177–181.
- Winterer G, Weinberger DR. 2004. Genes, dopamine and cortical signal-to-noise ratio in schizophrenia. *Trends Neurosci* 27(11):683–690.
- Xie T, Ho SL, Ramsden D. 1999. Characterization and implications of estrogenic down-regulation of human catechol-O-methyltransferase gene transcription. *Mol Pharmacol* 56(1):31–38.

**Estudio 3. Modelo de pleiotropismo antagónico en
esquizofrenia: *DAOA*, *PPP1R1B*, y *APOL1*.**



Contents lists available at ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

Testing the antagonistic pleiotropy model of schizophrenia susceptibility by analysis of *DAOA*, *PPP1R1B*, and *APOL1* genes

Noa Carrera^a, Manuel Arrojo^b, Eduardo Paz^b, Ramón Ramos-Ríos^b, Santiago Agra^b, Mario Páramo^b, Julio Brenlla^b, Javier Costas^{a,*}

^a Fundación Pública Galega de Medicina Xenómica, Santiago de Compostela, Spain

^b Servicio de Psiquiatría, Complejo Hospitalario Universitario de Santiago, Santiago de Compostela, Spain

ARTICLE INFO

Article history:

Received 25 November 2008

Received in revised form 16 April 2009

Accepted 26 May 2009

Keywords:

Association study

Natural selection

Single-nucleotide polymorphism

G72

DARPP-32

ABSTRACT

Schizophrenia is a common disease associated with reduced fertility. Therefore, the existence of common susceptibility alleles not removed by natural selection may be considered an evolutionary paradox. The antagonistic pleiotropy model, proposed to explain this paradox, states that an allele may be common because of its overall selective advantage, in spite of deleterious effects on specific traits. Recent work on *DAOA*, *PPP1R1B*, and *APOL1* suggests that these genes present common alleles associated to increase risk of schizophrenia but conferring an overall selective advantage, related to better cognitive performance (*DAOA* and *PPP1R1B*) or protection against pathogens (*APOL1*). To test if these genes fit the antagonistic pleiotropy model, we searched for recent natural selection at these loci applying the long-range haplotype test on data from the HapMap Project; and performed case-control association analysis in a well-powered sample, including 301 schizophrenic patients and 604 controls from Spain. For *DAOA* and *PPP1R1B*, we genotyped the Single-nucleotide polymorphisms (SNPs) needed to replicate previous associations, while for *APOL1*, we genotyped 15 tagSNPs, and seven putative functional SNPs. We did not detect evidence of recent natural selection. Furthermore, we did not find significant associations. Thus, these genes do not fit the antagonistic pleiotropy model.

© 2009 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Schizophrenia is a common, heritable disease, associated with reduced fertility (Haukka et al., 2003). Therefore, the existence of common susceptibility alleles, not yet eliminated by purifying selection (i.e., selection against alleles with deleterious effects), may be considered an evolutionary paradox (Huxley et al., 1964; Keller and Miller, 2006).

Specific evolutionary models may explain this paradox. The ancestral susceptibility model considers that alleles fitted to ancestral environments may be mis-adapted in the current human environment, being the basis for disease susceptibility (Di Rienzo and Hudson, 2005). This model is beginning to be tested in schizophrenia, with promising results (Carrera et al., 2009). The antagonistic pleiotropy model proposes that an allele may be present at high frequency

because of its overall selective advantage, although it may negatively affect some specific traits. For instance, some alleles related to increased risk to schizophrenia might also be associated to creativity (Crespi et al., 2007; Pearson and Folley, 2008).

At least three loci from literature are good candidates to test the antagonistic pleiotropy hypothesis in schizophrenia, as previously suggested (Costas et al., 2005; Meyer-Lindenberg et al., 2007; Opgen-Rhein et al., 2008). Thus, Opgen-Rhein et al. (2008) described a haplotype ~80 kb upstream of the coding region of D-amino acid oxidase activator (*DAOA*) associated with better semantic fluency, one aspect of cognitive performance, and with schizophrenia susceptibility using 178 cases and 144 healthy controls of European ancestry. Other analyses at *DAOA* did not include this distant region, with the exception of the first association report where the region, poorly covered, was not associated (Chumakov et al., 2002).

In a similar way, a common haplotype of protein phosphatase 1, regulatory (inhibitor) subunit 1B (*PPP1R1B*, also known as *DARPP-32*) seems to be associated with more efficient intrastriatal processing, revealed by neuroimaging and a wide range of cognitive tasks in healthy subjects, as well as with schizophrenia susceptibility in a family-based association analysis (Meyer-Lindenberg et al., 2007).

* Corresponding author. Fundación Pública Galega de Medicina Xenómica-SERGAS, Hospital Clínico Universitario de Santiago, Edif. Consultas Planta -2, E-15706 Santiago de Compostela, Spain. Tel.: +34 981951490; fax: +34 981951473.

E-mail address: bfcostas@usc.es (J. Costas).

Interestingly, *PPP1R1B* interacts with discoidin domain receptor tyrosine kinase 1 (*DDR1*), another putative schizophrenia-associated gene (Hansen et al., 2006; Roig et al., 2007).

Finally, the candidacy of *APOL1* is based on different type of evidences. The cluster of apolipoprotein L (*APOL*) genes is subject to recent adaptive selection (Costas et al., 2005; The International HapMap Consortium, 2007). Furthermore, *APOL1* is the lytic factor involved in defense against trypanosome infections (Vanhamme et al., 2003) and is upregulated in schizophrenia, suggesting that it may be a susceptibility gene (Mimmack et al., 2002). The first study to test this possibility evaluated 143 single-nucleotide polymorphisms (SNPs) in an Irish sample within the *APOL* cluster and did not find evidence of association (McGhee et al., 2005). However, the study had limited power to detect odds ratio (OR) <2, as they used pooled DNA from 219 cases and 231 controls. Another study, analyzing 377 families of different ethnicity and 130 tagSNPs, also failed to identify association at individual SNPs, although they did not perform haplotype analysis of *APOL1* (Takahashi et al., 2008).

In this work, we tested the antagonistic pleiotropy model by a combination of two steps, identification of variants of these genes subject to recent selection, and detection of susceptibility alleles by case-control association analysis.

2. Methods

2.1. Samples

The patient group comprised 301 samples (66.78% males; mean age 52.18 ± 15.89 years) being treated by the Galician Mental Health Service in the area of Santiago de Compostela, (Galicia, NW Spain), meeting Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV) criteria for schizophrenia. Patients with mental retardation or comorbid neurological diseases were excluded. Written informed consent was obtained. The research was carried out in accordance with the Declaration of Helsinki and was approved by the appropriate Ethics Committee. The control population consists of 604 unscreened controls (51.6% males; mean age 37.46 ± 12.41 years) obtained from blood donors at the Centro de Transfusión de Galicia (Santiago de Compostela). Taking into account the distribution of age-at-onset of schizophrenia, it is possible that a very small proportion of controls (considerably lower than the ~1% lifetime prevalence) will develop schizophrenia later in life, but this is irrelevant to the analysis. This sample size reaches at least 80% power to identify ORs >1.5, assuming the expected frequencies, while the OR for the *DAOA* haplotype estimated from the published data is higher than 1.7 (Opgen-Rhein et al., 2008). All samples are Spanish of Caucasian origin.

2.2. Genotyping

TagSNPs to cover the *APOL1* coding region as well as 5 kb at each edge were selected using linkage disequilibrium (LD) data from HapMap CEU (CEPH, Utah residents with European ancestry) samples and the tagger option of Haploview (Barrett et al., 2005), with a criterion for tagSNP of $r^2 > 0.8$ and minor allele frequency (MAF) >0.05. We also included six non-synonymous SNPs at *APOL1*, and several SNPs at *PPP1R1B* and *DAOA* that allow the identification of the haplotypes previously reported to be associated with schizophrenia and cognitive traits. SNPs were genotyped using the iPLEX MassArray technology from Sequenom (San Diego, CA, USA), following the manufacturer's instructions. SNPs that departed from the Hardy-Weinberg equilibrium, with more than 25% of missing genotypes, or MAF <0.01, were eliminated from the study. As a genotyping quality control, one CEPH trio (CEPH pedigree 1340, samples NA07019, NA 07029, and NA07062) was genotyped for all the SNPs. There were no Mendelian inconsistencies and the genotypes agree with those present in HapMap data.

2.3. Test for recent natural selection

The long-range haplotype (LRH) test implemented in the Sweep program (Sabeti et al., 2002) was used to detect alleles subject to recent positive selection, as described elsewhere (Carrera et al., 2009). The test is based on the fact that positively selected haplotypes present a LD surrounding them higher than expected by chance, as their rapid increase in frequency limits the action of recombination. We applied the LRH test along the whole genome to data from HapMap CEU or YRI (Yoruba from Ibadan, Nigeria) samples, as specifically indicated. Briefly, for each haplotype at each block of the human genome (i.e., the core haplotype) as defined by the program, we estimated the extended haplotype homozygosity (EHH), a measure of the long-range LD around a haplotype. Specifically, EHH is the probability that two randomly chosen chromosomes carrying the core haplotype of interest are identical for the entire interval from the core region to a specified distance (in our case, 0.1 and 0.2 cM). Then, we calculated the relative extended haplotype homozygosity (rEHH) as a measure of EHH corrected by the recombination at the locus. Specifically, rEHH is the ratio of the EHH on the tested core haplotype compared with the EHH of the grouped set of core haplotypes at the

region not including the core haplotype tested. Then, we compared the rEHH of the haplotypes of interest with the empirical distribution of rEHH of haplotypes with similar frequency along the human genome (20 bins, intervals of 5%). By this way, haplotypes with LD higher than expected under neutral evolution can be detected as outliers of the distribution.

2.4. Statistical analysis

Hardy-Weinberg (HW) equilibrium was tested using Haploview (Barrett et al., 2005). This program was also used to infer the frequency of the haplotypes and perform case-control analyses by 2×2 (allelic) χ^2 tests. In addition, a sliding-window haplotype analysis (window size of 2 SNPs) was performed on the tagSNPs of *APOL1* using the maximum likelihood method implemented in Unphased (Dudbridge, 2008).

3. Results

3.1. Searching for recent positive selection

The LRH test did not detect any evidence of recent natural selection at the haplotypes of *DAOA* and *PPP1R1B* putatively associated to schizophrenia and cognitive performance (Fig. 1). The LD around these haplotypes presents normal levels for haplotypes at similar frequencies.

In the case of *APOL1*, recent natural selection has been described for sub-Saharan African populations (The International HapMap Consortium, 2007). Application of the LRH test to the YRI (individuals with African ancestry) population from HapMap allowed us the

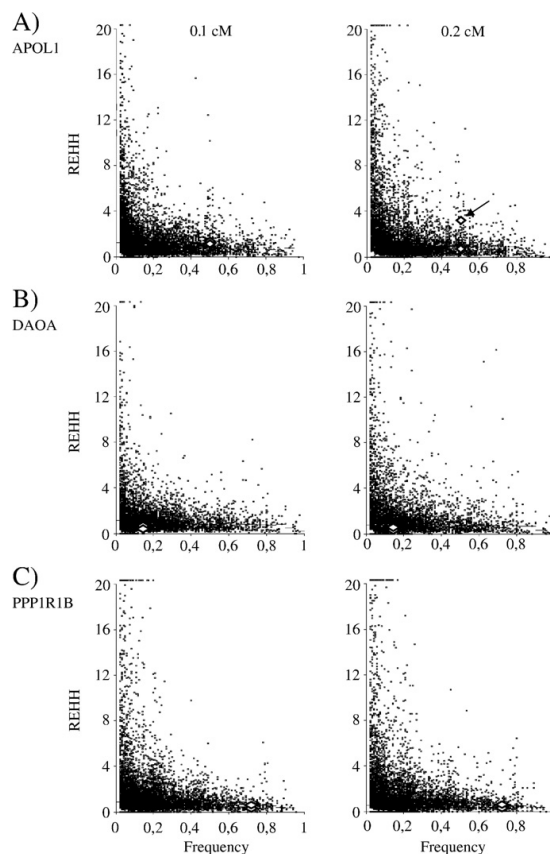


Fig. 1. Results of the LRH test at two different distances. The haplotypes under analysis for *APOL1* (A), *DAOA* (B), and *PPP1R1B* (C) are represented by diamonds. An arrow indicates the position of the unique haplotype with a P value lower than 0.05.

Table 1
Genotyping results.

Gene	SNP_ID	HW <i>P</i> val ^a	%Geno ^b	MA ^c	MAF ^d
<i>APOL1</i>	rs4280560	0.415	100	C	0.499
<i>APOL1</i>	rs2227089	0.196	96.9	T	0.381
<i>APOL1</i>	rs4821472	0.000	100	C	0.122
<i>APOL1</i>	rs5995271	0.000	75.4	T	0.164
<i>APOL1</i>	rs6000218	0.000	74.8	A	0.487
<i>APOL1</i>	rs5756115	1.000	75.1	G	0.082
<i>APOL1</i>	rs9610467	0.166	96.5	A	0.138
<i>APOL1</i>	rs136147	0.161	100	T	0.451
<i>APOL1</i>	rs136148	0.032	97.8	C	0.380
<i>APOL1</i>	rs4820224	0.807	97.7	A	0.149
<i>APOL1</i>	rs713753	1.000	99.2	C	0.385
<i>APOL1</i>	rs2239785 ^e	0.724	99.9	G	0.214
<i>APOL1</i>	rs136175 ^e	0.805	97.8	G	0.207
<i>APOL1</i>	rs136176 ^e	0.739	99.9	G	0.206
<i>APOL1</i>	rs16996616 ^e	1.000	98	A	0.005
<i>APOL1</i>	rs41297245 ^e	0.645	99.7	A	0.050
<i>APOL1</i>	rs3886200	0.555	99.1	T	0.477
<i>APOL1</i>	rs2012928	0.598	99.6	A	0.148
<i>APOL1</i>	rs136179	0.327	97.2	T	0.354
<i>APOL1</i>	rs41311348 ^f	1.000	96.9	–	0.000
<i>APOL1</i>	rs4821475	0.0088	99	T	0.239
<i>APOL1</i>	rs60910145 ^e	1.000	99.7	–	0.000
<i>PPP1R1B</i>	rs4795390	0.397	100	G	0.185
<i>PPP1R1B</i>	rs879606	0.407	99.9	A	0.188
<i>PPP1R1B</i>	rs11651497	0.941	97.9	T	0.238
<i>PPP1R1B</i>	rs907094	0.805	99.5	C	0.243
<i>PPP1R1B</i>	rs3764353	0.804	100	A	0.242
<i>PPP1R1B</i>	rs3764352	0.870	99.4	G	0.244
<i>DAOA</i>	rs1570709	1.000	98.4	C	0.168
<i>DAOA</i>	rs9586843	0.676	99.2	T	0.181
<i>DAOA</i>	rs7324448	0.534	96.7	T	0.228
<i>DAOA</i>	rs1575633	0.928	97.4	A	0.253
<i>DAOA</i>	rs7329966	0.361	99.7	G	0.460

^a *P* value for Hardy–Weinberg equilibrium.^b Percentage of samples successfully genotyped at each SNP.^c Minor allele (genomic + strand).^d Minor allele frequency.^e Coding non-synonymous SNP.^f Coding frameshift SNP.

identification of a recently selected haplotype at *APOL1*, consisting of 5347 bp, from SNP rs4280560 to SNP rs2227089. We also applied the test to the CEU samples (individuals with European ancestry) from HapMap. The haplotype selected in Africans presents a high value of rEHH just at one distance, 0.2 cM (empirical *P* value = 0.0298). There are no other haplotypes at *APOL1* showing clearer evidence of recent positive selection (Fig. 1).

3.2. Association analysis

Four of the SNPs departed from HW equilibrium ($P < 0.01$), two SNPs were monomorphic and one additional SNP has a frequency lower than 0.01 (Table 1). These SNPs were removed from the analysis. In the case of *PPP1R1B* and *DAOA*, we specifically tested association with the haplotypes putatively involved in schizophrenia risk and cognitive performance. For *APOL1*, we analyzed the haplotype subject to selection in Africans as well as the tagSNPs, as a proxy for common variants, both individually and as haplotypes of two consecutive markers (sliding-windows analysis). There is no evidence of association in our samples (Table 2).

4. Discussion

The allelic spectrum (i.e., number and frequency of different alleles) of loci involved in susceptibility to complex diseases, shaped by evolution, is critical for the design of studies aimed to find susceptibility alleles. Thus, the use of an evolutionary genetics framework may facilitate the identification of these genes. Blekman et al. (2008) have shown that while Mendelian-disease genes are under widespread

purifying selection, the class of genes that affect complex disease risk shows little signs of evolutionary conservation, possibly because this category includes targets of both purifying and positive selection. Crespi et al. (2007) have found a significant increase in genes subject to recent positive selection among those putatively associated to schizophrenia in comparison to a set of neuronal genes, although this result may be considered preliminary due to the lack of genes firmly established as schizophrenia susceptibility genes.

The antagonistic pleiotropy model might partially explain this positive selection. To fit this model, one allele must confer an overall selective advantage as well as susceptibility to schizophrenia. Thus, as a first step to test this hypothesis, we searched for recent positive selection on the haplotypes putatively related to both improved cognitive performance and schizophrenia risk at the *DAOA* and *PPP1R1B* genes (Meyer-Lindenberg et al., 2007; Opgen-Rhein et al., 2008). We failed to identify any sign of recent selection at these haplotypes. In addition, the haplotype at *APOL1* selected in the African population of HapMap shows less clear evidence of selection in Europeans. However, the absence of these signs must not be considered a definitive proof against the antagonistic pleiotropy model, if we take into account that the identification of haplotypes subject to recent natural selection is a difficult task, with an unknown proportion of both false-positive and false-negative results (Kelley et al., 2006; Sabeti et al., 2006; Voight et al., 2006). Furthermore, although antagonistic pleiotropy does not lead to a stable polymorphism, the conflicting fitness effects on different traits could lead to a longer persistence of the different alleles at intermediate frequencies (Keller and Miller, 2006). Therefore, there is more chance for recombination to dilute the mark of positive selection.

We also tried to replicate the former associations in the case of the haplotypes of *DAOA* and *PPP1R1B*, as well as to analyzed putative association with *APOL1*. The lack of association strongly suggests that these haplotypes, putatively related to better cognitive performance or pathogen resistance, do not fit the antagonistic pleiotropy model. Although antagonistic pleiotropy cannot be ruled out as an explanation for the existence of some proportion of susceptibility alleles, there

Table 2
Case-control association results.

Marker	Hap/allele	% Cases (n)	% Controls (n)	Marker <i>P</i> value	Sliding-windows <i>P</i> value ^a
<i>DAOA</i>	CCCCG ^b	16.1 (282)	14.8 (602)	0.4785	
<i>APOL1</i> haplot	CGCTGA ^c	73.9 (297)	76.2 (596)	0.291	
<i>APOL1</i> haplot	CA ^d	47.1 (301)	48.6 (604)	0.5643	
rs4280560	T	51.3 (301)	49.3 (604)	0.4057	0.7373
rs2227089	T	38.7 (279)	37.9 (598)	0.7379	0.7081
rs5756115	G	8.2 (214)	8.2 (466)	0.9885	0.3869
rs9610467	A	14.7 (289)	13.4 (584)	0.4416	0.4937
rs136147	G	55.5 (301)	53.1 (604)	0.3475	0.6103
rs136148	C	39.3 (298)	37.3 (587)	0.4235	0.7163
rs4820224	A	15.3 (298)	14.7 (586)	0.7406	0.7645
rs713753	C	39.7 (301)	37.9 (597)	0.4481	0.8696
rs2239785	A	22.1 (301)	21.1 (603)	0.6142	0.1846
rs136175	G	21.8 (298)	20.1 (587)	0.4012	0.2451
rs136176	G	21.6 (301)	20.1 (603)	0.4487	0.2902
rs41297245	A	6.3 (301)	4.4 (601)	0.0816	0.1717
rs3886200	C	54.5 (297)	51.2 (600)	0.1775	0.3689
rs2012928	A	15.3 (301)	14.5 (600)	0.6588	0.5203
rs136179	T	37.1 (298)	34.5 (582)	0.2908	

^a *P* value of haplotypic tests using a sliding-windows approach with a window size of two SNPs. The *P* value of the haplotype is located on the row of the first SNP.^b Haplotype associated with better semantic fluency and schizophrenia susceptibility. Order of SNPs: rs1570709, rs9586843, rs7324448, rs1575633, rs7329966.^c Haplotype associated with more efficient intratrial processing and schizophrenia susceptibility. Order of SNPs: rs4795390, rs879606, rs11651497, rs907094, rs3764353, rs3764352.^d Selected haplotype in the YRI population from HapMap. Order of SNPs: rs4280560, rs2227089.

is no experimental evidence to date. On the contrary, there is an increasing evidence of the important role of rare variants, probably under mutation-selection balance, in the genetic susceptibility to schizophrenia (International Schizophrenia Consortium, 2008; Stefansson et al., 2008; Walsh et al., 2008), in general agreement with evolutionary genetics (Keller and Miller, 2006; Kryukov et al., 2007).

Acknowledgments

This study was supported by grant PI052036 from Ministerio de Sanidad y Consumo, Spain, to J.C. We thank the patients for their participation, and the Centro de Transfusión de Galicia for its collaboration. Genotyping was done at the Centro Nacional de Genotipado, node USC, Fundación Genoma España. NC was supported by a fellowship from Xunta de Galicia. We would like to thank the healthcare professionals of Complejo Hospitalario de Santiago who assisted in the recruitment of study participants.

References

- Barrett, J.C., Fry, B., Maller, J., Daly, M.J., 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263–265.
- Blekhnman, R., Man, O., Herrmann, L., Boyko, A.R., Indap, A., Kosiol, C., Bustamante, C.D., Teshima, K.M., Przeworski, M., 2008. Natural selection on genes that underlie human disease susceptibility. *Current Biology* 18, 883–889.
- Carrera, N., Sanjuán, J., Moltó, M.D., Carracedo, A., Costas, J., 2009. Recent adaptive selection at MAOB and ancestral susceptibility to schizophrenia. *American Journal of Medical Genetics Series B Neuro-psychiatric Genetics* 150B, 369–374.
- Chumakov, I., Blumenfeld, M., Guerassimov, O., Cavarec, L., Palicio, M., Abderrahim, H., Bougueleret, L., Barry, C., Tanaka, H., La Rosa, P., Puech, A., Tahri, N., Cohen-Akenine, A., Delabrosse, S., Lissarrague, S., Picard, F.P., Maurice, K., Essioux, L., Millasseau, P., Grel, P., Debailleu, V., Simon, A.M., Caterina, D., Dufaure, I., Malekzadeh, K., Belova, M., Luan, J.J., Bouillot, M., Sambucy, J.L., Primas, G., Saumier, M., Boubkiri, N., Martin-Saumier, S., Nasroune, M., Peixoto, H., Delaye, A., Pinchot, V., Bastucci, M., Guillou, S., Chevillon, M., Sainz-Fuertes, R., Meguenni, S., Aurich-Costa, J., Cherif, D., Gimalac, A., Van Duijn, C., Gauvreau, D., Ouellette, G., Fortier, J., Raelson, J., Sherbatic, T., Riazanskaia, N., Rogae, E., Raeymaekers, P., Aerssens, J., Konings, F., Luyten, W., Macciardi, F., Sham, P.C., Straub, R.E., Weinberger, D.R., Cohen, N., Cohen, D., 2002. Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America* 99, 13675–13680.
- Costas, J., Salas, A., Phillips, C., Carracedo, A., 2005. Human genome-wide screen of haplotype-like blocks of reduced diversity. *Gene* 349, 219–225.
- Crespi, B., Summers, K., Dorus, S., 2007. Adaptive evolution of genes underlying schizophrenia. *Proceedings. Biological Sciences: The Royal Society*, vol. 274, pp. 2801–2810.
- Di Rienzo, A., Hudson, R.R., 2005. An evolutionary framework for common diseases: the ancestral-susceptibility model. *Trends in Genetics* 21, 596–601.
- Dudbridge, F., 2008. Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. *Human Heredity* 66, 87–98.
- Hansen, C., Greengard, P., Nairn, A.C., Andersson, T., Vogel, W.F., 2006. Phosphorylation of DARPP-32 regulates breast cancer cell migration downstream of the receptor tyrosine kinase DDR1. *Experimental Cell Research* 312, 4011–4018.
- 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. *American Journal of Psychiatry* 160, 460–463.
- Huxley, J., Mayr, E., Osmond, H., Hoffer, A., 1964. Schizophrenia as a genetic morphism. *Nature* 204, 220–221.
- International Schizophrenia Consortium, 2008. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 455, 237–241.
- Keller, M.C., Miller, G., 2006. Resolving the paradox of common, harmful, heritable mental disorders: which evolutionary genetic models work best? *The Behavioral and Brain Sciences* 29, 385–404 discussion 405–452.
- Kelley, J.L., Madeoy, J., Calhoun, J.C., Swanson, W., Akey, J.M., 2006. Genomic signatures of positive selection in humans and the limits of outlier approaches. *Genome Research* 16, 980–989.
- Kryukov, G.V., Pennacchio, L.A., Sunyaev, S.R., 2007. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *American Journal of Human Genetics* 80, 727–739.
- McGhee, K.A., Morris, D.W., Schwaiger, S., Nangle, J.M., Donohoe, G., Clarke, S., Meagher, D., Quinn, J., Scully, P., Waddington, J.L., Gill, M., Corvin, A., 2005. Investigation of the apolipoprotein-L (APOL) gene family and schizophrenia using a novel DNA pooling strategy for public database SNPs. *Schizophrenia Research* 76, 231–238.
- Meyer-Lindenberg, A., Straub, R.E., Lipska, B.K., Verchinski, B.A., Goldberg, T., Callicott, J.H., Egan, M.F., Huffaker, S.S., Mattay, V.S., Kolachana, B., Kleinman, J.E., Weinberger, D.R., 2007. Genetic evidence implicating DARPP-32 in human frontostriatal structure, function, and cognition. *Journal of Clinical Investigation* 117, 672–682.
- Mimmack, M.L., Ryan, M., Baba, H., Navarro-Ruiz, J., Iritani, S., Faull, R.L., McKenna, P.J., Jones, P.B., Arai, H., Starkey, M., Emson, P.C., Bahn, S., 2002. Gene expression analysis in schizophrenia: reproducible up-regulation of several members of the apolipoprotein L family located in a high-susceptibility locus for schizophrenia on chromosome 22. *Proceedings of the National Academy of Sciences of the United States of America* 99, 4680–4685.
- Opgen-Rhein, C., Lencz, T., Burdick, K.E., Neuhaus, A.H., DeRosse, P., Goldberg, T.E., Malhotra, A.K., 2008. Genetic variation in the DAOA gene complex: impact on susceptibility for schizophrenia and on cognitive performance. *Schizophrenia Research* 103, 169–177.
- Pearlson, G.D., Folley, B.S., 2008. Schizophrenia, psychiatric genetics, and Darwinian psychiatry: an evolutionary framework. *Schizophrenia Bulletin* 34, 722–733.
- Roig, B., Virgos, C., Franco, N., Martorell, L., Valero, J., Costas, J., Carracedo, A., Labad, A., Vilella, E., 2007. The discoidin domain receptor 1 as a novel susceptibility gene for schizophrenia. *Molecular Psychiatry* 12, 833–841.
- Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., Ackerman, H.C., Campbell, S.J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., Lander, E.S., 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419, 832–837.
- Sabeti, P.C., Schaffner, S.F., Fry, B., Lohmueller, J., Varily, P., Shamovsky, O., Palma, A., Mikkelsen, T.S., Altshuler, D., Lander, E.S., 2006. Positive natural selection in the human lineage. *Science* 312, 1614–1620.
- Stefansson, H., Rujescu, D., Cichon, S., Pietiläinen, O.P., Ingason, A., Steinberg, S., Fossdal, R., Sigurdsson, E., Sigmundsson, T., Buizer-Voskamp, J.E., Hansen, T., Jakobson, K.D., Muglia, P., Francks, C., Matthews, P.M., Gylfason, A., Halldorsson, B.V., Gudbjartsson, D., Thorgerirsson, T.E., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., Björnsson, A., Mattiasdottir, S., Blondal, T., Haraldsson, M., Magnusdottir, B.B., Giegling, I., Möller, H.J., Hartmann, A., Shianna, K.V., Ge, D., Need, A.C., Crombie, C., Fraser, G., Walker, N., Lonnqvist, J., Suvisaari, J., Tuulio-Henriksson, A., Paunio, T., Toupoulou, T., Bramon, E., Di Forti, M., Murray, R., Ruggeri, M., Vassos, E., Tosato, S., Walshe, M., Li, T., Vasilescu, C., Mühleisen, T.W., Wang, A.G., Ullum, H., Djurovic, S., Melle, I., Olesen, J., Kiemene, L.A., Franke, B., Sabatti, C., Freimer, N.B., Gulcher, J.R., Thorsteinsdottir, U., Kong, A., Andreassen, O.A., Ophoff, R.A., Georgi, A., Rietschel, M., Werge, T., Petursson, H., Goldstein, D.B., Nöthen, M.M., Peltonen, L., Collier, D.A., St Clair, D., Stefansson, K., Kahn, R.S., Linszen, D.H., van Os, J., Wiersma, D., Bruggeman, R., Cahn, W., de Haan, L., Krabbendam, L., Myin-Germeys, I., Genetic Risk and Outcome in Psychosis (group), 2008. Large recurrent microdeletions associated with schizophrenia. *Nature* 455, 232–236.
- Takahashi, S., Cui, Y.H., Han, Y.H., Fagerness, J.A., Galloway, B., Shen, Y.C., Kojima, T., Uchiyama, M., Faraone, S.V., Tsuang, M.T., 2008. Association of SNPs and haplotypes in APOL1 2 and 4 with schizophrenia. *Schizophrenia Research* 104, 153–164.
- The International HapMap Consortium, 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449, 851–861.
- Vanhamme, L., Paturiaux-Hanocq, F., Poelvoorde, P., Nolan, D.P., Lins, L., Van Den Abbeele, J., Pays, A., Tebabi, P., Van Xong, H., Jacquet, A., Moguilevsky, N., Dieu, M., Kane, J.P., De Baetselier, P., Brasseur, R., Pays, E., 2003. Apolipoprotein L-1 is the trypanosome lytic factor of human serum. *Nature* 422, 83–87.
- Voight, B.F., Kudravalli, S., Wen, X., Pritchard, J.K., 2006. A map of recent positive selection in the human genome. *PLoS Biology* 4 (3), e721.
- Walsh, T., McClellan, J.M., McCarthy, S.E., Addington, A.M., Pierce, S.B., Cooper, G.M., Nord, A.S., Kusenda, M., Malhotra, D., Bhandari, A., Stray, S.M., Rippey, C.F., Roccanova, P., Makarov, V., Lakshmi, B., Findling, R.L., Sikich, L., Stromberg, T., Merriman, B., Gogtay, N., Butler, P., Eckstrand, K., Noory, L., Gochman, P., Long, R., Chen, Z., Davis, S., Baker, C., Eichler, E.E., Meltzer, P.S., Nelson, S.F., Singleton, A.B., Lee, M.K., Rapoport, J.L., King, M.C., Sebat, J., 2008. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 320, 539–543.

**Estudio 4. Selección natural reciente en genes candidatos
funcionales.**

Identification of schizophrenia susceptibility alleles subject to recent adaptive selection

Noa Carrera^{1,2}, Manuel Arrojo³, Ramón Ramos-Ríos³, Eduardo Paz³, Santiago Agra³,
Mario Páramo³, Julio Brenlla³, Ángel Carracedo^{1,2,4}, Javier Costas¹

¹Fundación Pública Galega de Medicina Xenómica-SERGAS, Hospital Clínico
Universitario, Santiago de Compostela, Spain

²Dpto. Anatomía Patolóxica e Ciencias Forenses, Universidade de Santiago de
Compostela, Spain

³Servicio de Psiquiatría-SERGAS, Complexo Hospitalario Universitario, Santiago de
Compostela, Spain

⁴CIBERER, Centro Nacional de Genotipado, Universidade de Santiago de Compostela,
Spain

Corresponding author:

Javier Costas

Fundación Pública Galega de Medicina Xenómica-SERGAS, Hospital Clínico

Universitario de Santiago, Edif. Consultas Planta -2

E-15706 Santiago de Compostela, Spain

Phone: +34-981951490

Fax: +34-981951473

E-mail: javier.costas@usc.es

Abstract

Background

The existence of common alleles of susceptibility to complex diseases of reduced fertility, such as schizophrenia, may be considered an evolutionary paradox, except if their effect is almost negligible. Two main evolutionary models have been proposed to explain the existence of common variants of susceptibility to complex diseases, the ancestral susceptibility model and the antagonistic pleiotropy model. Recent adaptive selection plays a key role in both models, suggesting that genes subject to incomplete selective sweeps have greater probability to present common variants of susceptibility to complex diseases of reduced fertility.

Results

In order to test the role of recent adaptive selection on the existence of common variants of susceptibility to schizophrenia, a list of 1694 genes that are functional candidates to be involved in schizophrenia etiopathology were compiled as the first step. Application of the long range haplotype test to the CEU samples from HapMap phase II led to the identification of strong evidence of incomplete selective sweeps at 12 of the 1694 genes. The 12 selected haplotypes were analyzed by a case - control association study comprising 391 schizophrenic cases and 659 controls from Spain. One gene, *PTPN13*, presented a haplotype subject to recent natural selection associated with protection against schizophrenia (trend test $P = 0.002$, corrected $P = 0.024$).

Conclusions

This study used a novel strategy based on evolutionary biology to identify common susceptibility variants to schizophrenia, a complex disease associated to reduced fertility. This strategy led to the identification of common variants of susceptibility to schizophrenia at one gene, *PTPN13*, although the result is pending of replication in larger samples. Taking into account the limits in the identification of footprints of recent selection as well as in the choice of putative functional candidate genes related to schizophrenia, our results suggest that a less stringent threshold to consider a gene as subject to recent natural selection, use of different tests to identify selection and/or analysis of additional genes may lead to the identification of other common variants of susceptibility to schizophrenia.

1. Introduction

Schizophrenia is a chronic severe psychiatric syndrome which affects people from all over the world, with a lifetime prevalence of 0.72% [1]. It often starts in young adulthood [2] and is associated with reduced fertility [3, 4] and high rates of suicide, generally near onset of the illness [5]. It is a highly heritable complex trait, which results from a combination of genetic and environment predisposing factors [6]. Its heritability was estimated as 65-81% [7, 8].

The elevated prevalence of schizophrenia despite its high heritability, early onset and reproductive disadvantage is considered an evolutionary paradox [9, 10]. According to the evolutionary theory, it is expected that genetic variants associated with loss of fitness are pruned out from the genetic pool through purifying selection and as a consequence, these variants and the related phenotypes should be rare. Therefore, how common variants conferring risk to schizophrenia circumvent this negative pressure remains an enigma. An obvious explanation is that each individual common variant escapes the action of natural selection because of its near negligible effect, being neutral from a selective perspective. Two additional evolutionary models have been proposed to explain the existence of common variants with non negligible risk.

The ancestral susceptibility model [11] states that some genetic variants that were advantageous or neutral in ancestral populations, might now be miss-adapted to the current human lifestyle and environment, increasing susceptibility to certain complex diseases. There are some clear examples in the literature fitting the ancestral susceptibility model in complex diseases [12]. Regarding schizophrenia, our group has described a haplotype at *MAOB* subject to recent positive selection while ancestral haplotypes confer an increased risk to schizophrenia, in agreement with the ancestral susceptibility model [13].

The antagonistic pleiotropy model [14] proposes that genetic variants that increase susceptibility to a fitness-reducing disease could reach high frequency in populations if they are also associated with a trait that confers a selective advantage. For instance, it has been suggested that some alleles related to susceptibility to schizophrenia may have undergone recent positive selection associated with aspects of creativity or language [15, 16]. Our group has tested three genes that were candidate to fit this model [17-19], but with negative results [20].

The consideration of these models would be useful to guide the search for common susceptibility variants, focusing on those variants subject to recent positive selection. In addition, this strategy is supposed to enrich the study in functionally different alleles, as they are non neutral, and so this increases the probability to find pathogenic variants.

The progress of dense maps of human genetic variations has allowed whole-genome scans to detect the genetic footprint of recent positive selection. There are several tests for these scans. One of these tests is the long-range haplotype (LRH) test [21], which identifies events of incomplete selective sweeps. The test is relied on the relationship between an allele's frequency and the extent of linkage disequilibrium (LD) surrounding it. Under neutral evolution, haplotype frequency is proportional to its age, and then, more common haplotypes have been subject to recombination for a longer period,

leading to a breakdown of LD. Nevertheless, frequency of haplotypes subject to recent positive selection may increase rapidly enough to reduce change that recombination destroys the association with nearby polymorphisms, and therefore such haplotypes will have an extension of LD around them larger than expected under neutral evolution according to their frequency. Assuming that most part of the human variability has evolved in a neutral mode [22], outliers of the empirical distribution of a statistic which measure the LD around common haplotypes, controlled for frequency and local variation in recombination rates, may be considered as candidates for selection.

The aim of this study was to identify common susceptibility alleles to schizophrenia based on evolutionary models. To this goal, first we searched for evidence of recent adaptive selection on a large list of functional candidate genes, and then we tested for association of the common haplotypes showing signals of recent selection with schizophrenia.

2. Methods

2.1- Samples

The patient group consists of 391 individuals (67.26% males; mean age 50.97 ± 15.39 years) diagnosed as schizophrenic patients according to the DSM-IV criteria and treated by the Galician Mental Health Service in the area of Santiago de Compostela (Galicia, NW Spain). Patients with mental retardation or comorbid neurological diseases were excluded. Written informed consent was obtained. The control sample consists of 659 unscreened individuals (51.6% males; mean age 37.46 ± 12.41 years) obtained from blood donors at the Centro de Transfusión de Galicia (Santiago de Compostela). The research was carried out in accordance with the Declaration of Helsinki and was approved by the appropriate Ethic Committee. All samples are Spanish of Caucasian origin.

2.2- Construction of a set of putative functional candidate genes

Different sources were used to compile a list of putative functional candidate genes for schizophrenia susceptibility: 1) genes from the SchizophreniaGene database [23]; 2) genes under the Gene Ontology terms "nervous system development" and "synaptic transmission" [24]; 3) genes under the Panther terms "neurogenesis" and "synaptic transmission" [25]; 4) genes from SYNAPSEdb [26]; 5) genes from the DISC1 interactome [27]; 6) other previously compiled lists from bibliography, based on functionality or differential expression [28-31]. The list was fully assembled prior to our analyses.

2.3- Test for recent positive selection

The LRH test was performed across the genome on phased genotype data from CEU samples (CEPH, Utah residents with European ancestry) of the phase II HapMap, as implemented in the Sweep program [21]. Firstly, for each haplotype at each haplotype block of the human genome defined by the program (i.e., the core haplotype), we

estimated the extension of LD around it by calculating the extended haplotype homozygosity (EHH), i.e, the probability that two randomly chosen chromosomes carrying the core haplotype of interest are identical by descent for the entire interval from the core region to a specified distance [21]. Instead of analyzing EHH at a single point, we measured the EHH value at four genetic distances, ± 0.1 and 0.2 cM to identify more robust evidence of selection. Following, we calculated the relative EHH (rEHH), defined as the ratio of the EHH on the tested core haplotype compared with the EHH of the grouped set of core haplotypes at the region, excluding the core haplotype tested [21]. In this manner, EHH value is corrected by the influence of different recombination rates along the genome using the other haplotypes of the block. Finally, we compared the rEHH value of the haplotypes of interest with the empirical distribution of rEHH values of all haplotypes with similar frequency of the human genome (20 bins of frequency, intervals of 5%). In that way, haplotypes potentially subject to recent positive selection are detected as outliers of the empirical distribution of rEHH values.

To increase confidence in the value of the test, we consider only core haplotypes at frequencies higher than 10% and with at least 3 of the four measures of EHH, corresponding with the four genetic distances, higher than 0.1. A haplotype was considered putatively subject to recent natural selection if the P value of the LRH test was lower than 0.01 at least at 3 of the four genetic distances.

2.4- SNP selection and genotyping

Appropriate tag SNPs to distinguish the previously identified selected haplotypes were chosen using LD data from HapMap CEU samples. SNPs were genotyped by mass spectrometry (MALDI-TOF) using the Sequenom MassArray iPLEX genotyping system (San Diego, CA, USA), following manufacturer's instructions. Genotyping was done at the Santiago de Compostela node of the Spanish National Genotyping Center (CEGEN).

2.5- Statistical Analysis

For those selected haplotypes tagged by more than one SNP, inference of haplotype phase from the genotype data was performed using PHASE version 2.1 [32]. Deviations from Hardy-Weinberg (HW) equilibrium were calculated using the Haploview program [33]. Case-control analyses were performed in each gene by Cochran-Armitage test for trend implemented in WINPEPI v10.5 [34].

3. Results

Our compilation of putative functional candidate genes consisted of 1749 genes, approximately 7% of the total genes (Supplementary Table 1). Using the LRH test in data from the HapMap CEU samples, we identified strong signals of selection on specific haplotypes in twelve of those genes (Table 1). Several of them have been previously shown to present signs of recent adaptive selection.

In order to test for association of these recently selected haplotypes with schizophrenia, we genotyped 17 SNPs tagging the selected haplotypes in our case-control sample. One SNP, the one tagging TUBB haplotype, departed from HW equilibrium ($P = 0.001$). Nevertheless, inspection of the genotypic clusters did not show any apparent error explaining this departure. Therefore, as departure for HW equilibrium may be due to current natural selection, the SNP was not excluded for the analysis.

As shown in Table 2, the selected haplotype of *PTPN13* was significantly over-represented in controls (Trend test $P = 0.0020$; OR heterozygous = 0.66, 95% CI, 0.49 - 0.89; OR homozygous for the selected haplotype = 0.48, 95%CI, 0.20 - 1.14). This result remains significant after Bonferroni's correction for 12 tests (corrected $P = 0.024$). The core haplotype includes the non-synonymous SNP rs2230600, in perfect LD with the genotyped SNP rs17420966. The derived allele of rs2230600 is always associated with the selected haplotype (Figure 1). Another haplotype in the synaptic transmission gene *SNTG1* is nominally associated ($P = 0.046$), but did not resist multiple testing correction.

4. Discussion

In this work, we used an original strategy based on evolutionary biology to find common variants of susceptibility to schizophrenia focusing on haplotypes subject to recent natural selection in a list of putative functional candidate genes. Although the list is far from being complete, due to the low knowledge of both the molecular biology basis of schizophrenia as well as the function of many human genes, it represents more than 7% of the human genes, comprising genes with plausible functional link to schizophrenia. Using this strategy, we identified a common haplotype of protection against schizophrenia at *PTPN13*; an association that remains significant after Bonferroni's correction. *PTPN13* seems to be over-expressed in prefrontal cortex of schizophrenic patients [30].

PTPN13 is a large intracellular protein with a modular structure consisting of diverse protein-protein interaction modules (Figure 2A) [35]: a KIND domain of unknown function, a FERM domain, typical of proteins that act as plasma membrane-cytoskeleton linkers, five different PDZ domains involved in protein-protein interactions, and one protein tyrosine phosphatase domain, which is located at the C-terminus [35]. PDZ domains are often found in tandem arrays, as in *PTPN13*, in which the PDZ domain-containing protein serves as a central scaffolding protein facilitating the assembly of a multiplicity of different proteins [36]. The haplotype identified in our study is located at the core of the protein, and includes four PDZ domains, from 2 to 5. The missense SNP rs2230600 is located at the third PDZ domain (Figure 2). Nevertheless, we must take into account that in case of two or more PDZ domains connected in tandem, the target binding properties of each domain may be affected by the other domains [36].

PTPN13 is expressed in several tissues and is involved in multiple functions. There are several proteins known to interact with PDZ domains of *PTPN13* [35]. The role of

PTPN13 in brain development and function may be inferred by their protein partners. Thus, PTPN13 binds the neurotrophins receptor NGFR through the second PDZ domain [37], probably controlling neuronal cell death during development [38]. Ephrin B interacts with the fourth PDZ domain and serves also as a substrate for the protein tyrosine phosphatase domain. PTPN13 colocalized with Ephrin B in membrane lipid rafts of cultured cortical neurons [39]. Ephrin B has important roles in processes such as axon guidance, synaptogenesis, synaptic plasticity and proliferation of neural stem cells [39]. PRK2 interacts with the third PDZ domain of PTPN13 [40]. PRK2 is a cytosolic serine/threonine kinase involved in the modulation of the cytoskeleton [41]. Interestingly, it has been found as a constituent of synaptosomes of human cerebral cortex [42].

Thus the concrete combination of variants integrating the selected haplotype of PTPN13, that include several missense SNPs (Figure 1), would implicate some subtle effects on protein binding that would reduce the risk of developing schizophrenia. Nevertheless, the result is pending of replication analysis to discard or confirm the association of *PTPN13* with schizophrenia susceptibility.

Until recently, the study of natural selection in humans has largely been restricted to comparing individual candidate genes to theoretical expectations. The advent of genome-wide sequence and polymorphism data has allowed the use of empirical distributions of DNA sequence variation across the human genome to identify signals of positive selection. However, as we do not know the proportion of the genome that has undergone positive selection during evolution, we cannot define precise thresholds to calculate the probability of selection of an outlier of the rEHH distributions. Therefore, in addition to evidence of selection by a specific test, the declaration of selection at a locus should be enhanced by other functional evidence, such as finding an association with a specific phenotype [43]. The functional evidence supporting selection at PTPN13 would come from the association with schizophrenia revealed in the present work, as well as from association with squamous cell carcinoma of head and neck, as reported previously [44]. The G allele at rs2230600, corresponding to the selected haplotype, is associated with an increased risk of squamous cell carcinoma of head and neck.

Recently, it has been argued that events of natural selection in which a beneficial mutation arises and rapidly fixes in the population were not the dominant mode of human adaptation. Other types of selection, such as polygenic selection or selection on standing variation, may be more probable [45]. If this is the case, a proportion of signals identified in our scan would correspond to other modes of selection. This type of adaptive selection may be more difficult to be detected than typical selective sweeps. Therefore, the role of recent adaptive selection on the existence of common alleles of schizophrenia susceptibility may be higher than appreciated. Currently, there are a few examples of genes showing common variants of susceptibility due to recent natural selection, such as MAOB [13], GABRB2 [46] or DRD3 [47, 48]. In addition, balancing selection has been suggested to explain the existence of polymorphisms at two schizophrenia associated loci, the Val158Met SNP at *COMT* [49, 50] and the variants at the major histocompatibility complex [51-54].

One limitation of our study is the sample size, which has little power to detect small genetic effects. In this regard, the filters applied to the LRH test may result quite strict, and it would be that many of the candidate genes have softer signals of selection. Nevertheless, this restriction responds to a need in limiting the number of tests performed in the study to avoid a very stringent correction for multiple tests.

In summary, although we analyzed only a subset of genes involved in schizophrenia risk and most probably detected only some of the genes subject to recent natural selection from this subset, this strategy led to the identification of a common haplotype involved in schizophrenia susceptibility, pending on replication in additional samples. Reduction of the threshold to consider a gene as subject to recent natural selection, use of different tests to identify selection and/or analysis of additional genes may lead to the identification of other common variants of susceptibility.

5. References

1. McGrath J, Saha S, Chant D, Welham J: **Schizophrenia: a concise overview of incidence, prevalence, and mortality.** *Epidemiol Rev* 2008, **30**:67-76.
2. Messias EL, Chen CY, Eaton WW: **Epidemiology of schizophrenia: review of findings and myths.** *Psychiatr Clin North Am* 2007, **30**(3):323-338.
3. Haukka J, Suvisaari J, Lonnqvist J: **Fertility of patients with schizophrenia, their siblings, and the general population: a cohort study from 1950 to 1959 in Finland.** *Am J Psychiatry* 2003, **160**(3):460-463.
4. Bundy H, Stahl D, MacCabe JH: **A systematic review and meta-analysis of the fertility of patients with schizophrenia and their unaffected relatives.** *Acta Psychiatr Scand* 2011, **123**(2):98-106.
5. Palmer BA, Pankratz VS, Bostwick JM: **The lifetime risk of suicide in schizophrenia: a reexamination.** *Arch Gen Psychiatry* 2005, **62**(3):247-253.
6. Gejman PV, Sanders AR, Kendler KS: **Genetics of Schizophrenia: New Findings and Challenges.** *Annu Rev Genomics Hum Genet* 2011.
7. Lichtenstein P, Yip BH, Bjork C, Pawitan Y, Cannon TD, Sullivan PF, Hultman CM: **Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study.** *Lancet* 2009, **373**:234-239.
8. Sullivan PF, Kendler KS, Neale MC: **Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies.** *Arch Gen Psychiatry* 2003, **60**(12):1187-1192.
9. Huxley J, Mayr E, Osmond H, Hoffer A: **Schizophrenia as a Genetic Morphism.** *Nature* 1964, **204**:220-221.
10. Keller MC, Miller G: **Resolving the paradox of common, harmful, heritable mental disorders: which evolutionary genetic models work best?** *Behav Brain Sci* 2006, **29**(4):385-404; discussion 405-352.
11. Di Rienzo A, Hudson RR: **An evolutionary framework for common diseases: the ancestral-susceptibility model.** *Trends Genet* 2005, **21**(11):596-601.
12. Di Rienzo A: **Population genetics models of common diseases.** *Curr Opin Genet Dev* 2006, **16**(6):630-636.

13. Carrera N, Sanjuan J, Molto MD, Carracedo A, Costas J: **Recent adaptive selection at MAOB and ancestral susceptibility to schizophrenia.** *Am J Med Genet B Neuropsychiatr Genet* 2009, **150B**(3):369-374.
14. Wright A, Charlesworth B, Rudan I, Carothers A, Campbell H: **A polygenic basis for late-onset disease.** *Trends Genet* 2003, **19**(2):97-106.
15. Crespi B, Summers K, Dorus S: **Adaptive evolution of genes underlying schizophrenia.** *Proc Biol Sci* 2007, **274**(1627):2801-2810.
16. Crow TJ: **Is schizophrenia the price that Homo sapiens pays for language?** *Schizophr Res* 1997, **28**(2-3):127-141.
17. Costas J, Salas A, Phillips C, Carracedo A: **Human genome-wide screen of haplotype-like blocks of reduced diversity.** *Gene* 2005, **349**:219-225.
18. Meyer-Lindenberg A, Straub RE, Lipska BK, Verchinski BA, Goldberg T, Callicott JH, Egan MF, Huffaker SS, Mattay VS, Kolachana B *et al*: **Genetic evidence implicating DARPP-32 in human frontostriatal structure, function, and cognition.** *J Clin Invest* 2007, **117**(3):672-682.
19. Opgen-Rhein C, Lencz T, Burdick KE, Neuhaus AH, DeRosse P, Goldberg TE, Malhotra AK: **Genetic variation in the DAOA gene complex: impact on susceptibility for schizophrenia and on cognitive performance.** *Schizophr Res* 2008, **103**(1-3):169-177.
20. Carrera N, Arrojo M, Paz E, Ramos-Rios R, Agra S, Paramo M, Brenlla J, Costas J: **Testing the antagonistic pleiotropy model of schizophrenia susceptibility by analysis of DAOA, PPP1R1B, and APOL1 genes.** *Psychiatry Res* 2011, **179**(2):126-129.
21. Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ *et al*: **Detecting recent positive selection in the human genome from haplotype structure.** *Nature* 2002, **419**(6909):832-837.
22. Kimura M: **The neutral theory of molecular evolution.** *Sci Am* 1979, **241**(5):98-100, 102, 108 passim.
23. Allen NC, Bagade S, McQueen MB, Ioannidis JP, Kavvoura FK, Khoury MJ, Tanzi RE, Bertram L: **Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database.** *Nat Genet* 2008, **40**:827-834.
24. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al*: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25**(1):25-29.
25. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A: **PANTHER: a library of protein families and subfamilies indexed by function.** *Genome Res* 2003, **13**(9):2129-2141.
26. Zhang W, Zhang Y, Zheng H, Zhang C, Xiong W, Olyarchuk JG, Walker M, Xu W, Zhao M, Zhao S *et al*: **SynDB: a Synapse protein DataBase based on synapse ontology.** *Nucleic Acids Res* 2007, **35**(Database issue):D737-741.
27. Camargo LM, Collura V, Rain JC, Mizuguchi K, Hermjakob H, Kerrien S, Bonnert TP, Whiting PJ, Brandon NJ: **Disrupted in Schizophrenia 1 Interactome: evidence for the close connectivity of risk genes and a potential synaptic basis for schizophrenia.** *Mol Psychiatry* 2007, **12**:74-86.
28. Moises HW, Zoega T, Gottesman, II: **The glial growth factors deficiency and synaptic destabilization hypothesis of schizophrenia.** *BMC Psychiatry* 2002, **2**:8.

29. Jungerius BJ, Hoogendoorn ML, Bakker SC, Van't Slot R, Bardoel AF, Ophoff RA, Wijmenga C, Kahn RS, Sinke RJ: **An association screen of myelin-related genes implicates the chromosome 22q11 PIK4CA gene in schizophrenia.** *Mol Psychiatry* 2008, **13**(11):1060-1068.
30. Arion D, Unger T, Lewis DA, Levitt P, Mirnics K: **Molecular evidence for increased expression of genes related to immune and chaperone function in the prefrontal cortex in schizophrenia.** *Biol Psychiatry* 2007, **62**(7):711-721.
31. 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*: **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* 2009, **150B**(6):808-816.
32. Stephens M, Smith NJ, Donnelly P: **A new statistical method for haplotype reconstruction from population data.** *Am J Hum Genet* 2001, **68**(4):978-989.
33. Barrett JC, Fry B, Maller J, Daly MJ: **Haploview: analysis and visualization of LD and haplotype maps.** *Bioinformatics* 2005, **21**(2):263-265.
34. Abramson JH: **WINPEPI (PEPI-for-Windows): computer programs for epidemiologists.** *Epidemiol Perspect Innov* 2004, **1**:6.
35. Erdmann KS: **The protein tyrosine phosphatase PTP-Basophil/Basophil-like. Interacting proteins and molecular functions.** *Eur J Biochem* 2003, **270**(24):4789-4798.
36. Feng W, Zhang M: **Organization and dynamics of PDZ-domain-related supramodules in the postsynaptic density.** *Nat Rev Neurosci* 2009, **10**(2):87-99.
37. Irie S, Hachiya T, Rabizadeh S, Maruyama W, Mukai J, Li Y, Reed JC, Bredesen DE, Sato TA: **Functional interaction of Fas-associated phosphatase-1 (FAP-1) with p75(NTR) and their effect on NF-kappaB activation.** *FEBS Lett* 1999, **460**(2):191-198.
38. Dechant G, Barde YA: **The neurotrophin receptor p75(NTR): novel functions and implications for diseases of the nervous system.** *Nat Neurosci* 2002, **5**(11):1131-1136.
39. Palmer A, Zimmer M, Erdmann KS, Eulenburg V, Porthin A, Heumann R, Deutsch U, Klein R: **EphrinB phosphorylation and reverse signaling: regulation by Src kinases and PTP-BL phosphatase.** *Mol Cell* 2002, **9**(4):725-737.
40. Gross C, Heumann R, Erdmann KS: **The protein kinase C-related kinase PRK2 interacts with the protein tyrosine phosphatase PTP-BL via a novel PDZ domain binding motif.** *FEBS Lett* 2001, **496**(2-3):101-104.
41. Vincent S, Settleman J: **The PRK2 kinase is a potential effector target of both Rho and Rac GTPases and regulates actin cytoskeletal organization.** *Mol Cell Biol* 1997, **17**(4):2247-2256.
42. DeGiorgis JA, Jaffe H, Moreira JE, Carlotti CG, Jr., Leite JP, Pant HC, Dosemeci A: **Phosphoproteomic analysis of synaptosomes from human cerebral cortex.** *J Proteome Res* 2005, **4**(2):306-315.
43. Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Vavilili P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES: **Positive natural selection in the human lineage.** *Science* 2006, **312**(5780):1614-1620.

44. Niu J, Huang YJ, Wang LE, Sturgis EM, Wei Q: **Genetic polymorphisms in the PTPN13 gene and risk of squamous cell carcinoma of head and neck.** *Carcinogenesis* 2009, **30**(12):2053-2058.
45. Pritchard JK, Di Rienzo A: **Adaptation - not by sweeps alone.** *Nat Rev Genet* 2010, **11**(10):665-667.
46. Lo WS, Xu Z, Yu Z, Pun FW, Ng SK, Chen J, Tong KL, Zhao C, Xu X, Tsang SY *et al*: **Positive selection within the Schizophrenia-associated GABA(A) receptor beta(2) gene.** *PLoS One* 2007, **2**(5):e462.
47. Costas J, Carrera N, Dominguez E, Vilella E, Martorell L, Valero J, Gutierrez-Zotes A, Labad A, Carracedo A: **A common haplotype of DRD3 affected by recent positive selection is associated with protection from schizophrenia.** *Hum Genet* 2009, **124**(6):607-613.
48. Nunokawa A, Watanabe Y, Kaneko N, Sugai T, Yazaki S, Arinami T, Ujike H, Inada T, Iwata N, Kunugi H *et al*: **The dopamine D3 receptor (DRD3) gene and risk of schizophrenia: case-control studies and an updated meta-analysis.** *Schizophr Res* 2011, **116**(1):61-67.
49. Goldman D, Weinberger DR, Malhotra AK, Goldberg TE: **The role of COMT Val158Met in cognition.** *Biol Psychiatry* 2009, **65**(1):e1-2; author reply e3-4.
50. Costas J, Sanjuan J, Ramos-Rios R, Paz E, Agra S, Ivorra JL, Paramo M, Brenlla J, Arrojo M: **Heterozygosity at catechol-O-methyltransferase Val158Met and schizophrenia: new data and meta-analysis.** *J Psychiatr Res*, **45**(1):7-14.
51. Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Pe'er I, Dudbridge F, Holmans PA, Whitemore AS, Mowry BJ *et al*: **Common variants on chromosome 6p22.1 are associated with schizophrenia.** *Nature* 2009, **460**(7256):753-757.
52. Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D, Werge T, Pietilainen OP, Mors O, Mortensen PB *et al*: **Common variants conferring risk of schizophrenia.** *Nature* 2009, **460**:744-747.
53. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P: **Common polygenic variation contributes to risk of schizophrenia and bipolar disorder.** *Nature* 2009, **460**(7256):748-752.
54. Hughes AL, Nei M: **Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection.** *Nature* 1988, **335**(6186):167-170.
55. Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L: **Natural selection has driven population differentiation in modern humans.** *Nat Genet* 2008, **40**(3):340-345.
56. Simonson TS, Yang Y, Huff CD, Yun H, Qin G, Witherspoon DJ, Bai Z, Lorenzo FR, Xing J, Jorde LB *et al*: **Genetic evidence for high-altitude adaptation in Tibet.** *Science* 2010, **329**(5987):72-75.
57. Voight BF, Kudaravalli S, Wen X, Pritchard JK: **A map of recent positive selection in the human genome.** *PLoS Biol* 2006, **4**(3):e72.
58. Williamson SH, Hubisz MJ, Clark AG, Payseur BA, Bustamante CD, Nielsen R: **Localizing recent adaptive evolution in the human genome.** *PLoS Genet* 2007, **3**(6):e90.

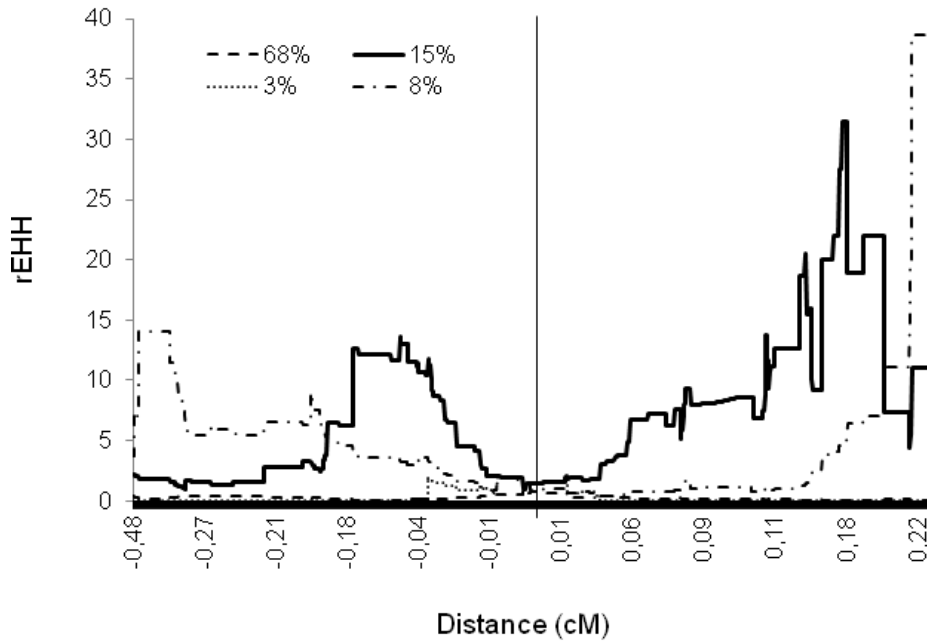


Figure 1. Variation of rEHH with distance from the core haplotypes at the *PTPN13* Selected haplotype. Only haplotypes at frequency higher than 1% are shown.

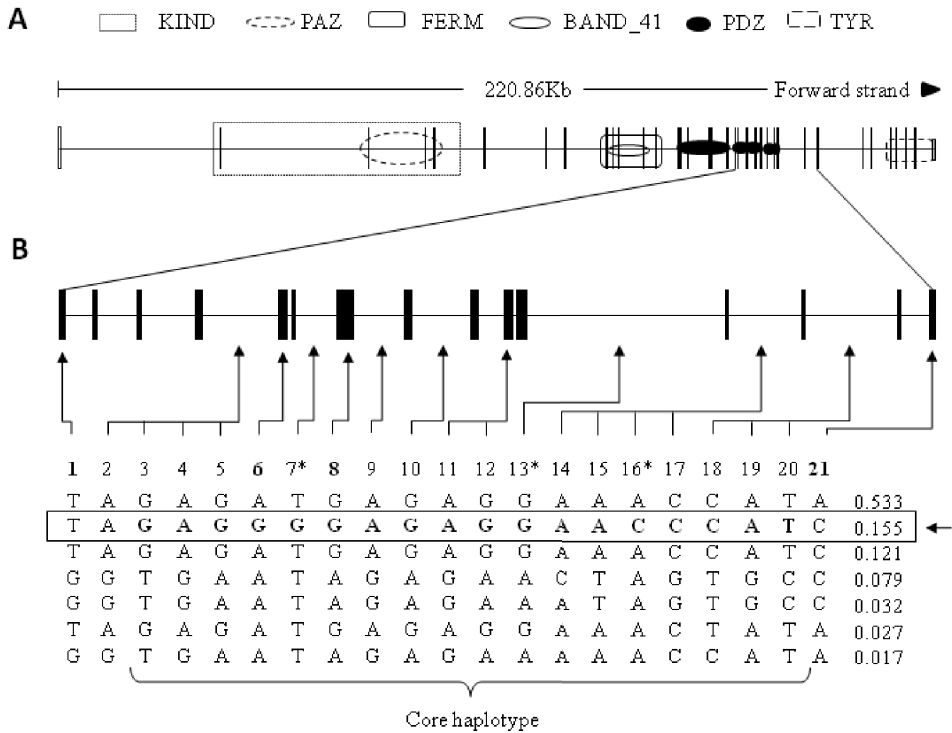


Figure 2. A. Structure of *PTPN13*, corresponding to the transcript ENST00000436978 from the ENSEMBL genome browser. Exons are represented by vertical rectangles, introns by horizontal lines, and functional domains as specified in the figure. The selected haplotype is located on the core of the protein, and spans four of the five PDZ domains. **B.** Location of the SNPs defining the selected haplotype. Note that the core haplotype defined by the Sweep Program comprises from SNP 3 to 20, twenty SNPs altogether, adding monomorphic SNPs not represented in the figure. This is by default the maximum number of SNPs allowed in each haplotype block defined by the program, but inspection of the region reveals that the selected haplotype spans three more SNPs, as shown in figure. Sequence of the selected haplotype (marked by an arrow) as well as the other haplotypes defined by the SNPs is indicated, with each frequency at the right side (only haplotypes at frequency higher than 0.01 are shown). Non synonymous SNPs are in bold type. Tag SNPs defining the selected haplotype are marked by *. The numbers of SNPs correspond as follows: 1: rs10033029; 2: rs11097118; 3: rs12512858; 4: rs11097119; 5: rs12498681; 6: rs2230600; 7: rs17420966; 8: rs12500797; 9: rs12509560; 10: rs11097120; 11: rs12507034; 12: rs12499471; 13: rs13148078; 14: rs9993838; 15: rs12499477; 16: rs17012072; 17: rs13118632; 18: rs9990536; 19: rs10013114; 20: rs10002511; 21: rs989902.

Table 1. Genes with evidence of recent selective sweep in our analysis

Gene	List	Prior ^a	Selected Haplotype			L-RH test <i>P</i> -values				
			Chr	Start SNP	End SNP	Fr. ^b	-0.2	-0.1	0.1	0.2
<i>PTPN13</i>	Bibliography	5 (YRI)	4	rs12512858	rs10002511	0.15	0.17091	0.00303	0.00339	0.00217
<i>CDH10</i>	Panther	5 (CEU)	5	rs6452213	rs16893479	0.47	0.00337	0.02546	0.00093	0.00597
<i>IL3</i>	GO, SZGene		5	rs6596053	rs2069616	0.47	na ^c	0.00026	0.00019	0.00066
<i>GRIK2</i>	GO, Panther, SNPfun, SZgene	1	6	rs9399731	rs12333116	0.47	0.00878	0.00713	0.00180	0.05776
<i>TUBB</i>	DISC1 interact	5 (CEU)	6	rs6929626	rs9968913	0.16	0.00002	0.00303	0.00160	0.00289
<i>SDK1</i>	panther	1, 2	7	rs6976818	rs12333854	0.10	na ^c	0.00871	0.00788	0.00609
<i>SNTG1</i>	panther	3, 4, 5 (CEU, ASN)	8	rs16915079	rs16915106	0.72	0.02576	0.00018	0.00015	0.00037
<i>UNC5D</i>	Panther	1	8	rs7839379	rs11992734	0.10	0.00670	0.00973	0.00285	0.01190
<i>SNCG</i>	SZGene		10	rs4520507	rs760112	0.29	0.00959	0.01544	0.00492	0.00846
<i>CNTN5</i>	Panther, SNPfun		11	rs7931411	rs1373048	0.30	0.01069	0.00776	0.00179	0.00435
<i>NELLI</i>	GO, Panther	2	11	rs12365800	rs11025762	0.73	0.00266	0.00157	0.00277	0.03651
<i>INPP5J</i>	Bibliography, SZGene	5 (YRI)	22	rs2232178	rs4820946	0.17	0.02351	0.00998	0.00232	0.00069

^aPrior evidence of selection; 1. Barreiro et al 2007 [55] 2. Simonson et al. 2010, [56]; 3. Voight et al. 2006 [57], 4. Willansom et al. 2007 [58]; 5. Haplotter HapMap phase II Gene P (in brackets, population with signal of selection).

^bFrequency in HapMap CEU (CEPH, Utah residents with European ancestry).

^cna, not applicable, EHH for the core haplotype reached the minimum value at a genetic distance shorter than -0.2 cM

Table 2. Association analysis

Gen	TagSNPs	Sample size		Core Haplotype Frequency		P-value ¹	OR (95% CI)	
		Cases	Controls	Cases	Controls		het genotype ²	hom genotype ³
PTPN13	rs17420966-rs13148078-rs17012072	386	659	0,12	0,17	0.0020	0.66 (0.49 - 0.89)	0.48 (0.20 - 1.14)
SNTG1	rs1483637	385	643	0,71	0,67	0.046	0.90 (0.57 - 1.41)	1.26 (0.81 - 1.95)
SNCG	rs760112	378	650	0,26	0,23	0.070	1.13 (0.87 - 1.48)	1.86 (1.00 - 3.44)
NELL1	rs11025762	378	650	0,78	0,81	0.16	0.96 (0.50 - 1.85)	0.80 (0.42 - 1.51)
CDH10	rs3903387	382	631	0,49	0,51	0.28	0.87 (0.64 - 1.18)	0.82 (0.57 - 1.18)
IL3	rs2069616	386	659	0,47	0,49	0.33	1.01 (0.75 - 1.36)	0.83 (0.57 - 1.19)
CNTN5	rs11221762	381	643	0,31	0,30	0.37	1.15 (0.88 - 1.51)	1.11 (0.70 - 1.76)
SDK1	rs2190251	374	646	0,05	0,06	0.39	0.86 (0.57 - 1.29)	0.56 (0.06 - 5.45)
PIB5PA	rs4820944-rs5753481	389	659	0,26	0,25	0.44	1.16 (0.89 - 1.51)	1.03 (0.60 - 1.74)
TUBB	rs1059612	375	649	0,06	0,05	0.57	1.13 (0.73 - 1.75)	1.17 (0.33 - 4.17)
GRIK2	rs9399731-rs9404147	390	659	0,43	0,44	0.82	1.10 (0.83 - 1.47)	0.92 (0.64 - 1.33)
UNC5D	rs1436613	373	644	0,10	0,10	0.97	1.06 (0.76 - 1.47)	0.50 (0.10 - 2.40)

¹ Cochran-Armitage trend test² OR of heterozygous (1 selected haplotype) referred to the homozygote with 0 selected haplotypes³ OR of homozygote of 2 selected haplotypes referred to the homozygote of 0 selected haplotypes

Estudio 5. Análisis de SNPs no sinónimos comunes

Association Study of Nonsynonymous Single Nucleotide Polymorphisms in Schizophrenia

Noa Carrera, Manuel Arrojo, Julio Sanjuán, Ramón Ramos-Ríos, Eduardo Paz, Jose J. Suárez-Rama, Mario Páramo, Santiago Agra, Julio Brenlla, Silvia Martínez, Olga Rivero, David A. Collier, Aarno Palotie, Sven Cichon, Markus M. Nöthen, Marcella Rietschel, Dan Rujescu, Hreinn Stefansson, Stacy Steinberg, Engilbert Sigurdsson, David St. Clair, Sarah Tosato, Thomas Werge, Kari Stefansson, Jose Carlos González, Joaquín Valero, Alfonso Gutiérrez-Zotes, Antonio Labad, Lourdes Martorell, Elisabet Vilella, Ángel Carracedo, and Javier Costas

Background: Genome-wide association studies using several hundred thousand anonymous markers present limited statistical power. Alternatively, association studies restricted to common nonsynonymous single nucleotide polymorphisms (nsSNPs) have the advantage of strongly reducing the multiple testing problem, while increasing the probability of testing functional single nucleotide polymorphisms (SNPs).

Methods: We performed a case-control association study of common nsSNPs in Galician (northwest Spain) samples using the Affymetrix GeneChip Human 20k cSNP Kit, followed by a replication study of the more promising results. After quality control procedures, the discovery sample consisted of 5100 nsSNPs at minor allele frequency >5% analyzed in 476 schizophrenia patients and 447 control subjects. The replication sample consisted of 4069 cases and 15,128 control subjects of European origin. We also performed multilocus analysis, using aggregated scores of nsSNPs at liberal significance thresholds and cross-validation procedures.

Results: The 5 independent nsSNPs with false discovery rate $q \leq .25$, as well as 13 additional nsSNPs at $p < .01$ and located in functional candidate genes, were genotyped in the replication samples. One SNP, rs13107325, located at the metal ions transporter gene *SLC39A8*, reached significance in the combined sample after Bonferroni correction (trend test, $p = 2.7 \times 10^{-6}$, allelic odds ratio = 1.32). This SNP presents minor allele frequency of 5% to 10% in many European populations but is rare outside Europe. We also confirmed the polygenic component of susceptibility.

Conclusions: Taking into account that another metal ions transporter gene, *SLC39A3*, is associated to bipolar disorder, our findings reveal a role for brain metal homeostasis in psychosis.

Key Words: Metal brain homeostasis, metal ion transporters, psychosis, *SLC39A8*, whole-genome association, ZIP8

Schizophrenia is a complex disorder with high heritability (1,2). Recent genome-wide association studies (GWAS) involving several thousand samples identified a few loci strongly associated with schizophrenia, at the major histocompatibility complex, and the genes *NRGN* and *TCF4* (3–5). Each one of the loci presents a very low effect (odds ratio [OR] < 1.3). The International Schizophrenia Consortium (ISC) used a test based on the aggregated score of all single nucleotide polymorphisms (SNPs) at liberal significance thresholds to provide some evidence for the existence of common polygenic variation contributing to risk of schizophrenia

(3). This result has been confirmed in a Japanese population (6). Another gene initially identified in GWAS of 479 cases and 2937 control subjects (7), *ZNF804A*, has been confirmed as a schizophrenia susceptibility gene in follow-up studies (8,9). In addition, rare variants may also play an important role, as revealed by copy number variant analysis (10–13).

Genome-wide association studies are based on linkage disequilibrium (LD), using several hundred thousand marker SNPs, most of them with very low a priori probability to be involved in disease. This generates an important problem of multiple testing, requiring a very low p value to declare significance. The analysis of GWAS results revealed that nonsynonymous SNPs (nsSNPs) (SNPs that change the sequence of the protein encoded by a gene) are the

From the Fundación Pública Galega de Medicina Xenómica-SERGAS (NC, JJS-R, AC, JC), Hospital Clínico Universitario; Dpto. Anatomía Patológica e Ciencias Forenses (NC, AC), and CIBERER (AC), Centro Nacional de Genotipado, Universidade de Santiago de Compostela; and Servicio de Psiquiatría-SERGAS (MA, RR-R, EP, MP, SA, JB, SM), Complexo Hospitalario Universitario, Santiago de Compostela; CIBERSAM (JS, OR), Universitat de Valencia; and Hospital Clínico Valencia (JCG), Valencia; and Hospital Psiquiàtric Universitari Institut Pere Mata (JV, AG-Z, AL, LM, EV), IISPV, Rovira i Virgili University, Reus, Spain; Social, Genetic and Developmental Psychiatry Centre (DAC), Institute of Psychiatry, King's College, London, and Department of Mental Health (DSC), University of Aberdeen, Royal Cornhill Hospital, Aberdeen, United Kingdom; Institute for Molecular Medicine Finland FIMM and Department of Medical Genetics (AP), University of Helsinki, Finland, and Wellcome Trust Sanger Institute, Cambridge, United Kingdom; Institute of Neuroscience and Medicine (SC), Structural and Functional Organization of the Brain Genomic Imaging Research Center, Juelich; Institute of Human Genetics and Department of Genomics (SC, MMN), Life & Brain Center, University of Bonn, and German Center for Neurodegenerative Disorders, Bonn; Department of Genetic Epidemiology in Psychiatry (MR), Central Institute of Mental Health, University of Heidelberg, Heidelberg; and Division of Molecular and Clinical Neurobiology (DR), Department of Psychiatry, Ludwig-Maximilians-University, Munich, Germany; deCODE Genetics (HS, SS, KS); Department of Psychiatry (ES), National University Hospital; and University of Iceland (KS), Faculty of Medicine, Reykjavik, Iceland; Section of Psychiatry and Clinical Psychology (ST), University of Verona, Verona, Italy; and Institute of Biological Psychiatry (TW), Mental Health Centre Sct. Hans, Copenhagen University Hospital, Copenhagen, Denmark. Address correspondence to Javier Costas, Ph.D., Hospital Clínico Universitario de Santiago, Fundación Pública Galega de Medicina Xenómica-SERGAS, Edif.

Consultas Planta-2, Santiago de Compostela E-15706, Spain; E-mail: javier.costas@usc.es.

Received Sep 27, 2010; revised Aug 19, 2011; accepted Sep 6, 2011.

0006-3223/\$36.00
doi:10.1016/j.biopsych.2011.09.032

BIOL PSYCHIATRY 2012;71:169–177
© 2012 Society of Biological Psychiatry

functional class with higher probability to be associated with disease (14). Nonsynonymous SNPs present lower LD than intronic or intergenic SNPs, being more likely singletons, i.e., SNPs that only tag themselves (15). Around 35% of nsSNPs in HapMap phase II Utah residents with Northern and Western European ancestry from the CEPH collection (CEU) samples are singletons at $r^2 = 1$ (15). Therefore, if they are not included in GWAS panels, there may be a lack of power to detect association with these nsSNPs using tagging SNPs. For instance, only 45% of the common nsSNPs (minor allele frequency > 5%) from the Illumina (San Diego, California) Human NS-12 BeadChip panel (a panel specifically designed to capture 11,649 nsSNPs) are tagged at $r^2 = 1$ by markers on the Affymetrix (Santa Clara, California) 500K GWAS panel (16). Furthermore, nsSNP panels are expected to be more useful for testing individual profiles based on aggregated SNP scores because of the enrichment in functional variants. In agreement with this, the ISC found a stronger polygenic effect for SNPs within annotated genes (3).

Here, we describe our case-control association study of schizophrenia using the Affymetrix GeneChip Human 20k cSNP Kit, specifically designed to genotype ~14,000 nsSNPs, including all the validated nsSNPs present in the database of single nucleotide polymorphism (dbSNP) b124 (National Center for Biotechnology Information [NCBI], Bethesda, Maryland) that could be accurately genotyped. We used two strategies of analysis based on individual nsSNPs as well as nsSNPs in multilocus data.

Methods and Materials

Discovery Sample

The cases sample was constituted by 509 schizophrenia patients attending the Galician Public Health Services of the Santiago de Compostela area. All the cases met the DSM-IV criteria for schizophrenia, as determined by consensus of the treating psychiatrist and a member of the research team, both senior clinical psychiatrists, after making a complete review of all the psychiatric records, which usually included many years of treatment. The DSM-IV diagnosis was then confirmed by the psychiatrist of the research team by means of an unstructured interview. All patients presented a diagnostic stability of at least 2 years. Patients with schizoaffective disorders were excluded. The mean age was 52.3 ± 15.6 years old (range: 21–89) and 67% were male patients. All the cases gave their written informed consent to participate in this study, which was approved by the Ethical Committee of Clinical Research from Galicia, Spain. All patients were of European origin from Galicia (North-west Spain).

The control group consisted of 519 blood donors at the Galician Transfusion Center at Santiago de Compostela. They were of Galician origin and were not specifically screened to exclude psychiatric disorders. The mean age was 43.11 ± 10.9 years old (range: 20–67) and 48% were male participants. Donors were not financially remunerated, making it unlikely that socially disadvantaged groups, which might have higher rates of schizophrenia, were overrepresented. Therefore, taking into account the prevalence of schizophrenia, the impact of using unscreened control subjects may be negligible.

Replication Samples

Two case-control samples from Valencia (East Spain) and Reus (Northeast Spain) were genotyped for replication purposes. The Valencia sample consisted of 438 cases (66% male cases) and 419 control subjects (63% male subjects). Schizophrenic patients were both inpatients and outpatients from the Valencia University Hospital, meeting the DSM-IV criteria for schizophrenia diagnosis as

determined by consensus of two senior clinical psychiatrists. All patients had a minimum of 1 year evolution of the illness. The control subjects were recruited from the same geographical area and were screened with a semistructured interview, including questions regarding psychological or psychiatric treatment to discard personal history of psychiatric disorder. All the participants gave their written informed consent to enter in this study. The study was approved by the Ethical Committee of the Medicine Faculty, University of Valencia. All the samples were of Spaniard origin.

The Reus samples consisted of 494 cases (62% male cases) and 614 control subjects (52% male subjects), all born in Spain and of European origin. The cases were both inpatients and outpatients from the Hospital Psiquiàtric Universitari Institut Pere Mata (Reus, Spain), meeting the DSM-IV criteria for schizophrenia diagnosis, evaluated as previously described (17). Control subjects from the same geographical area were recruited from primary care centers and were screened to discard psychiatric disease using the 28-scale global health questionnaire (more details in [17]). Informed consent was obtained from all participants and the study was approved by the corresponding Medical Ethical Committee.

Additional samples were from the SGENE-plus Consortium (<http://www.sgene.eu/>). There were 3137 schizophrenic patients and 14,095 control subjects from England, Finland (Kuusamo and general population), Germany (Munich and Bonn), Iceland, Italy, Scotland, and Denmark (described in [5]). The Denmark samples, comprising 541 schizophrenic patients and 474 control subjects, were included in the SGENE-plus collection after publication of the GWAS (5). Patients were recruited to the Danish Psychiatric Biobank from the psychiatric departments of the six hospitals in the Copenhagen region and were clinically diagnosed with schizophrenia according to ICD-10 (F20) without ever having received a diagnosis of mania or bipolar illness (F30 or F31). An experienced research and consultant psychiatrist verified the high reliability of the clinical diagnoses using the Operational Criteria Checklist for Psychotic and Affective Illness (18). The control subjects were recruited through the Danish Blood Donor Corps in the Copenhagen area; apparent behavioral abnormality was an exclusion criterion.

Genotyping

Genotyping was done using the Affymetrix GeneChip Human 20k cSNP Kit following manufacturer's instructions. This kit uses the molecular inversion probe technology (19) to genotype two panels that contain all the validated nsSNPs in dbSNP124 that could be accurately genotyped, 164 ancestry informative markers, additional nonvalidated nsSNPs, and several internal controls and duplicates. Mapping of this set of SNPs in the version 56 of ENSEMBL (<http://www.ensembl.org>) variation identified 14,135 nsSNPs. As a technical validation, the array includes 77 SNPs in duplicate. In addition, seven samples were genotyped twice. The concordance rate based on these genotypes was 99.17% (171,302 of 172,725 comparisons). Figure S1 (Supplement 1) shows the cluster plots for the most significant nsSNPs. Sequenom (San Diego, California) MassArray iPLEX genotyping system was used to genotype the 18 nsSNPs in the replication phase in Valencia and Reus samples, following manufacturer's instructions.

SGENE-plus samples were genotyped using different platforms and genome-wide association arrays, as described previously (5). Denmark samples were genotyped using Illumina 370K or 610K arrays. Proxy SNPs at high LD were used when the interesting nsSNPs were not genotyped in SGENE-plus samples.

Genotyping Quality Control Procedures

Five exclusion criteria for SNPs were applied as genotyping quality control: 1) SNPs if there were no missense SNPs according to

ENSEMBL variation version 56, 2) minor allele frequency lower than 5%, 3) genotyping call rate lower than 95%, 4) significant difference in call rates between cases and control subjects ($p < .001$), and 5) departure from Hardy-Weinberg equilibrium in control samples ($p < .001$).

Five filters were applied to the samples, removing those with 1) less than 90% European ancestry (see Population Stratification), 2) inconsistencies between the expected sex and that based on genotypes, 3) cryptic relatedness detected by identity by state values higher than .2, 4) call rate lower than 95%, and 5) heterozygosity levels that departed three standard deviations from the mean. The exact order of application of SNP and sample quality control filters is shown in Table S1 (Supplement 1). Finally, for the relevant associations, genotype cluster plots were reviewed to check genotype calling (Figure S1 in Supplement 1).

Population Stratification

The percentage of European ancestry in each sample was estimated running Structure v2.3.1 (<http://pritch.bsd.uchicago.edu/structure.html>) (20). Samples were removed if they showed less than 90% European ancestry as determined using the 130 ancestry informative markers of the panel that passed quality controls and were also genotyped in HapMap samples. HapMap CEU, YRI (Yoruba in Ibadan, Nigeria), and CHB (Han Chinese in Beijing, China)/JPT (Japanese in Tokyo) individuals were used as reference samples and the program was run under the admixture model with 100,000 replications for the burnin period and 100,000 replications after burnin for parameter estimations (Figure S2 in Supplement 1).

The genomic inflation factor λ (i.e., the median value of all association tests divided by the expected value under the null hypothesis of no association) (21) was calculated for the primary analysis and compared with the value for a logistic regression test, including as covariates the first four quantitative indices of ancestry based on a multidimensional scaling analysis, as implemented in PLINK v1.07 (<http://pngu.mgh.harvard.edu/purcell/plink/>) (22). For comparison purposes, we also calculated λ_{1000} , i.e., the inflation factor expected in a study of 1000 cases and 1000 control subjects.

Statistical Analysis

The primary analysis was a Cochran-Armitage trend test, as implemented in PLINK (22). To correct for multiple tests, false discovery rate (FDR) was calculated in our primary tests on the discovery sample using Q-Value software (<http://genomics.princeton.edu/storeylab/qvalue/>) (23). We established a threshold of $q \leq .25$ (that is, 25% of the tests declared significant will be false discoveries) for inclusion of an nsSNP in the replication phase. In addition, we included other nsSNPs with $p < .01$, based on their putative function (Supplement 1). Taking into account that joint analysis of data from the different phases almost always increases the power to detect association, despite more stringent significance levels (24), the results of the discovery and replication samples were combined using extended Mantel-Haenszel test for trend, considering 12 strata. Heterogeneity among samples was measured by I^2 (25). Statistical analyses were done with WinPEPI (<http://www.brixtonhealth.com/pepi4windows.html>) (26). Forest plots were drawn with the R package meta (<http://www.r-project.org/index.html>). Multiple tests correction was done by Bonferroni method considering the total number of nsSNPs passing the quality control filters.

Multilocus Analysis

A tenfold cross-validation procedure was applied to the discovery sample to detect polygenic effects, following a strategy similar to that of the ISC (3). Basically, case-control samples were divided

into 10 subsets of equal size. Nine subsets were used to generate p values and ORs for the different nsSNPs using allelic tests. Nonsynonymous SNPs with p values under a predetermined threshold were used to generate a score for the remaining one tenth of the samples, as the testing subset. The score for each individual, calculated with PLINK, was the sum across nsSNPs of the number of associated alleles (0, 1, or 2) at that nsSNP multiplied by the score for that SNP, $\log(\text{OR})$, averaging the score on those nsSNPs with missing genotyping data. To limit the effect of missing genotyping data, we matched cases and control subjects for this value before the tenfold partition. This procedure was repeated 10 times, using each of the 10 subsets as the testing subset. Scores were compared between cases and control subjects with a logistic regression model, including subset and percentage of missing genotypes as covariates. The multilocus analysis was based on a pruned set of nsSNPs, after removing one for each pair of nsSNP at $r^2 > .8$. We also considered a subset of nsSNPs based on its functional relevance. We included all the genes present in SzGene (<http://www.szgene.org/>) (27), a database of case-control association studies in schizophrenia, as well as those genes with a Gene Ontology (<http://www.geneontology.org/>) (28) associated term related to synaptic transmission and/or neurodevelopment, after manual review of all the terms associated with the genes in the final set. To quantify the value of the scoring nsSNPs to discriminate cases and control subjects, we calculated the area under the curve (AUC) of the receiver operating characteristic curve. The difference in the Nagelkerke pseudo R^2 between a model including scores and covariates and the model including only covariates was calculated as a measure of the percentage of variance explained by the scoring alleles, as in ISC (3). All the calculations were done with R package Design v2.3.0 (<http://biostat.mc.vanderbilt.edu/s/Design>).

Results

Quality Control Procedures

As shown in Table S1 (Supplement 1), the final set after application of quality control filters consists of 5100 common nsSNPs genotyped in 447 control subjects and 476 cases. The mean genotyping call rate for nsSNP in the final set was 98.91% (SD 1.16%), while the mean genotyping call rate for samples was 98.91% (SD .89%), without significant differences (t test, p value = .47) between cases (98.89%) and control subjects (98.93%). The information captured by these 5100 nsSNPs is partially redundant, as there are some nsSNPs in the same gene with high LD values. For instance, 276 and 387 nsSNPs present $r^2 > .95$ or $r^2 > .90$ with other nsSNPs of the set of 5100, respectively.

Association Analysis at nsSNPs in the Discovery Sample

We used the Cochran-Armitage trend test as the primary analysis to test for association in the discovery sample. Figure 1A shows the quantile-quantile plot for the analysis. The inflation factor λ was estimated as 1.035 ($\lambda_{1000} = 1.068$), a low value that more likely does not reflect stratification, as it remains almost identical ($\lambda = 1.034$, $\lambda_{1000} = 1.067$) after correcting by the four main components of ancestry using multidimensional scaling (Figure 1B). Figure S3 (Supplement 1) shows the p values by genome location. Three hundred eight nsSNPs present p values lower than .05, 70 lower than .01, and 8 lower than .001 (Table S2 in Supplement 1). Seven nsSNPs at five genes present a FDR $\leq .25$ (Table 1). None of the markers reached significance after Bonferroni correction. The top result is SNP rs6856 at the *BSC2* gene. According to the ENSEMBL database, this SNP is synonymous in the main transcripts of *BSC2* but is nonsynonymous in the transcript ENST00000278893. Nevertheless, visual in-

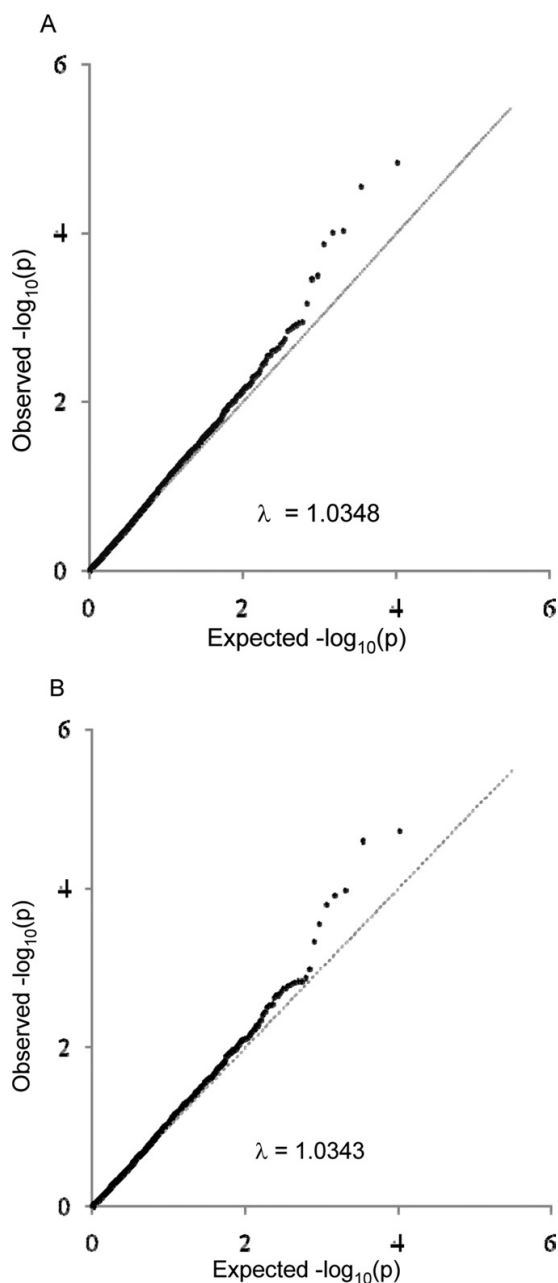


Figure 1. Quantile-quantile plots and λ estimates for the main analysis using Cochran-Armitage trend test (A) and the alternative analysis after correction by the four main components of ancestry using multidimensional scaling (B).

spection of the genotype intensity clustering plot revealed suboptimal assignment of genotypes (Figure S1 in Supplement 1). Thus, the significance of this finding may be overestimated. As there was

no evidence of association in the replication samples (see below), additional experiments to determine the correct genotype were not done. Among the top results, there were several SNP in tight LD in *MKI67* (a gene involved in cellular proliferation) and rs17101193 in *NRG3*, the homologous of the schizophrenia risk gene *NRG1*. Other interesting genes in relation to schizophrenia risk within the top 25 results include *MRPL40* located at the commonly deleted region in the velocardiofacial syndrome; *INSC*, involved in nervous system development; or *BCHE*, involved in cognition (Table 1 and Supplement 1).

Comparison with previous GWAS results supports the role of *ZNF804A* (Table 2). The two nsSNPs at *ZNF804A* with $p < .05$ are in LD and in the expected phase within them ($r^2 = .81$ in our control samples) and with rs1344706, the intronic SNP associated in the previous GWAS ($r^2 = .31$ and $.45$ with rs12476147 and rs1366842, respectively, in HapMap samples from Tuscan, Italy).

Replication in Independent Samples and Combined Analysis

Eighteen nsSNPs, including the most significant nsSNPs ($q \leq .25$) as well as those nsSNPs with $p < .01$ and suggestive of involvement in schizophrenia after bibliographic review (Table 1 and Supplement 1) were genotyped in two additional case-control samples from Spain. Additional data were obtained from the SGENE-plus collection. Single nucleotide polymorphism rs13107325 at *SLC39A8* (also known as ZIP8), reaches significance in the meta-analysis after Bonferroni correction for 5100 tests ($p = 2.7 \times 10^{-6}$) (Table 3; Table S3 in Supplement 1 for detailed results). As shown in Figure S4 (Supplement 1), genotype cluster plots are of high quality, indicating that the association is not due to genotyping artifacts. There is little heterogeneity among samples ($I^2 = 16.7\%$). After exclusion of the discovery sample, rs13107325 presented a p value of .00036, significant after Bonferroni correction for 18 tests. Heterogeneity between samples was minimal ($I^2 = 0\%$).

We also performed allelic meta-analysis to discard overestimation of the effects by chance in the trend test due to the low frequency of homozygous for the risk allele. Allelic meta-analysis confirmed the significance of the results ($p = 2.3 \times 10^{-6}$; allelic OR = 1.32; 95% confidence interval, 1.18–1.49; $I^2 = 15.8\%$). The association was again significant after exclusion of the discovery sample ($p = .00033$, OR = 1.26, 95% confidence interval, 1.11–1.43; $I^2 = 0\%$). Figure 2 shows the effects of rs13107325 at the genotype and allele levels in the different populations.

Multilocus Analysis

Using a tenfold cross-validation method, we generated a list of nsSNPs and associated ORs at a significance threshold of $p < .05$ in each training test. This list was used to generate a score in each testing test, based on the sum of log (OR). As shown in Table 4, there is a significant difference in the score of cases and control subjects ($p = .0068$) in the expected direction (i.e., cases scored higher than control subjects), although the explanatory power is limited (AUC = .545, Nagelkerke pseudo $R^2 = .011$). Repeating the same procedure using a more liberal threshold ($p < .1$) increased the discrimination between cases and control subjects ($p = .0002$, AUC = .567, Nagelkerke pseudo $R^2 = .021$).

As it is possible that most of the scoring alleles are not true schizophrenia susceptibility alleles, we repeated the analysis considering only that subset of nsSNPs located in genes with a function putatively involved in schizophrenia susceptibility. This reduced the number of nsSNPs to approximately one tenth. Although the differences in scores are still significant, the scores did not improve the explanatory power in comparison with the whole set of genes (Table 4).

Table 1. Results of the Trend Test for Association for the Most Relevant SNPs in Phase I

Rank ^a	SNP ^b (Minor Allele) ^c	HWE <i>p</i> ^d	Missing Rate			Allelic Frequency				OR (95% CI) ^g		Gene
			Cases	Controls	Missing <i>p</i> ^e	Cases	Controls	Trend <i>p</i> ^f	FDR <i>q</i>	Het Genotype	Hom Genotype	
1	rs6856 [C] ^h	.90	.04	.02	.012	.16	.24	1.5E-05	.07	.64 (.48–.85)	.23 (.10–.55)	<i>BSCL2</i>
2	rs997983 [C]	.52	.00	.00	.484	.07	.13	2.8E-05	.07	.56 (.39–.79)	.09 (.01–.73)	<i>MKI67</i>
3	rs4750685 [A]	1.00	.00	.00	1.000	.07	.12	9.4E-5	.12	.57 (.41–.81)	.06 (.00–1.15) ⁱ	<i>MKI67</i>
4	rs17101193 [A]	1.00	.03	.01	.077	.04	.08	9.9E-05	.12	.47 (.31–.72)	.12 (.01–2.42) ⁱ	<i>NRG3</i>
5	rs7918199 [T]	.83	.00	.01	.359	.07	.13	1.3E-04	.13	.56 (.40–.79)	.14 (.02–1.16)	<i>MKI67</i>
6	rs13107325 [T]	.34	.01	.01	.767	.13	.08	3.2E-04	.25	1.84 (1.31–2.60)	2.12 (.63–7.10)	<i>SLC39A8</i>
7	rs2276314 [G]	.03	.01	.01	1.000	.21	.28	3.5E-04	.25	.58 (.44–.76)	.65 (.36–1.19)	<i>C18orf21</i>
10	rs964307 [G]	.14	.00	.00	1.000	.24	.30	.0012	.51	.63 (.48–.83)	.62 (.37–1.06)	<i>PKHD1L1</i>
11	rs7575 [A]	.03	.02	.02	1.000	.09	.14	.0012	.51	.58 (.42–.79)	.83 (.17–4.15)	<i>MRPL40</i>
19	rs17507577 [A]	.02	.00	.01	.438	.07	.12	.0023	.57	.69 (.48–.99)	.22 (.06–.77)	<i>INSC</i>
21	rs1803274 [A] ^h	.04	.05	.01	.001	.17	.22	.0025	.57	.60 (.45–.80)	.82 (.38–1.75)	<i>BCHE</i>
32	rs160277 [A]	.840	.00	.01	.359	.44	.38	.0045	.66	1.43 (1.07–1.91)	1.67 (1.12–2.48)	<i>VCAN</i>
33	rs540825 [T]	.868	.00	.01	.113	.23	.17	.0046	.66	1.50 (1.13–1.99)	1.55 (.77–3.12)	<i>OPRM1</i>
35	rs2303291 [T]	.887	.00	.00	.234	.27	.21	.0048	.66	1.31 (.99–1.72)	2.04 (1.14–3.64)	<i>ITSN2</i>
36	rs10282493 [A]	.192	.00	.00	1.000	.16	.21	.0050	.66	.67 (.50–.88)	.64 (.29–1.44)	<i>ICA1</i>
44	rs7281684 [T]	.564	.01	.02	.317	.13	.09	.0064	.69	1.28 (.92–1.78)	7.02 (1.58–31.11)	<i>ABCG1</i>
53	rs1105168 [G]	.047	.00	.00	1.000	.44	.50	.0074	.69	.71 (.52–.97)	.60 (.41–.88)	<i>FAT2</i>
56	rs12476147 [A]	.610	.00	.00	.484	.43	.37	.0080	.69	1.14 (.86–1.52)	1.76 (1.19–2.62)	<i>ZNF804A</i>
58	rs1801582 [C]	.765	.01	.00	.062	.15	.20	.0083	.69	.60 (.44–.81)	.87 (.45–1.69)	<i>PARK2</i>
60	rs7603997 [T]	.057	.00	.00	.234	.36	.30	.0084	.69	1.22 (.93–1.60)	1.91 (1.18–3.07)	<i>ITSN2</i>

CI, confidence interval; FDR, false discovery rate; Het, heterozygous; Hom, homozygous; HWE, Hardy-Weinberg equilibrium; OR, odds ratio; SNP, single nucleotide polymorphism.

^aRank of SNP based on significance of the association.

^bAll the SNPs were included in the replication phase except rs997983 and rs4750685 due to high linkage disequilibrium with rs7918199, the SNP at *MKI67* with higher chance to work properly in the Sequenom assay.

^cReferred to the plus strand.

^dHardy-Weinberg equilibrium *p* value.

^eFisher's exact test for comparison of missing genotypes between cases and control subjects.

^fCochran-Armitage test for trend.

^gOR referred to homozygous for the major allele.

^hVisual inspection of genotype clusters showed suboptimal assignment of genotypes for this SNP. Therefore, the significance of this SNP may be overestimated.

ⁱAn increment of .5 was added to each cell in 2 × 2 tables to calculate approximate ORs, as there were zero cases homozygous for the minor allele.

Discussion

Our main result is the identification of one nsSNP, rs13107325, which reached significance after Bonferroni correction in the analysis of the whole set, comprising 4545 schizophrenia patients and 15,575 control subjects. Taking into account that there is considerable LD among different nsSNPs, this correction is highly conservative. Several facts suggest that this SNP is the functional SNP responsible for the association. First, the common allele at rs13107325 encodes an alanine that is well conserved in mammals and birds (29,30) (Figure S5 in Supplement 1). Second, the SNP gives rise to a substitution of a hydrophobic amino acid, alanine, to hy-

drophilic polar amino acid, threonine. Third, this SNP is at low LD with the rest of SNPs in HapMap samples ($r^2 < .5$). Finally, rs13107325 has been recently associated with two related phenotypes in genome-wide association studies, obesity and plasma high-density lipoprotein cholesterol concentrations (31,32).

The minor allele was present at a frequency of 7% in our control subjects from Spain and at similar frequency in the other European samples with the exception of Nordic countries, which showed lower frequencies (Figure S6 in Supplement 1). The SNP is almost absent from non-European populations of the HapMap project (2/482 alleles). Therefore, this nsSNP may be considered a rare

Table 2. nsSNPs at $p < .05$ in Loci Associated with Schizophrenia in Previous GWAS

Locus	Tested nsSNPs	Significant nsSNPs ^a	Trend <i>p</i> ^b	OR (95% CI)
<i>ZNF804A</i>	3	rs12476147 (T) rs1366842 (G)	.0079 .0090	1.29 (1.07–1.56) 1.28 (1.07–1.55)
MHC ^c	52	rs1265754 (T)	.0227	.62 (.42–.92)
<i>NRGN</i>	0	–	–	–
<i>TCFA</i>	0	–	–	–

CI, confidence interval; GWAS, genome-wide association studies; MHC, major histocompatibility complex; nsSNP, nonsynonymous single nucleotide polymorphism; OR, odds ratio.

^aMinor allele, referred to the plus strand.

^bCochran-Armitage test for trend.

^cDue to the complex linkage disequilibrium relationships within the MHC region, we considered the region from gene *SLC17A1* to *HLA-DQA1* instead of specific genes.

Table 3. Association Results at the Different Phases and the Meta-Analysis of All Samples

SNP (Gene)	Discovery Sample, Santiago 476 Cases/447 Control Subjects		Replication Samples, Spain 932 Cases/1033 Control Subjects		Replication SGENE-Plus 3137 Cases/14,095 Control Subjects		All Samples, 4545 Cases/15,575 Control Subjects	
	p Value ^a	OR Het	p Value ^b	OR Hom ^d	p Value ^b	OR Het ^c	p Value ^b	OR Het (95% CI) ^c
rs13107325 (SLC39A8)	3.2E-04	1.84	2.12	3.12	.013	1.11	4.13	1.25 (1.10–1.41)
rs7918199 (MKI67) ^e	1.3E-04	.56	.14	.93	.082	.90	.75	.87 (78–97)
rs964307 (PKHD1L1)	.0012	.63	.62	.89	.034	.96	.83	.94 (87–1.02)
rs2276314 (C18orf21)	3.5E-04	.58	.65	1.02	.21	.96	.88	.92 (85–1.00)
rs1105168 (FAT2)	.0074	.71	.60	.88	.41	.98	.94	.94 (86–1.03)
rs17101193 (NRG3) ^f	9.9E-05	.47	.15	1.35	.018	1.15	2.37	1.11 (98–1.25)
rs12476147 (ZNF804A) ^g	.0080	1.14	1.76	.99	.77	1.04	1.01	1.08 (98–1.18)
rs7281684 (ABCG1) ^h	.0064	1.28	7.02	.77	.76	1.05	.83	1.07 (98–1.17)
rs7575 (MRPL40)	.0012	.58	.83	.94	.97	1.03	.83	.96 (87–1.06)
rs7603997 (ITSN2)	.0084	1.22	1.91	.90	.42	.94	1.14	.96 (88–1.04)
rs540825 (OPRM1) ⁱ	.0046	1.50	1.55	.88	.64	1.01	1.06	1.02 (94–1.11)
rs1803274 (BCHE)	.0025	.60	.82	.97	.12	1.11	1.02	1.01 (93–1.11)
rs17507577 (INSC) ^j	.0023	.69	.22	1.04	.66	1.03	1.00	1.00 (90–1.11)
rs10282493 (ICA1) ^k	.0050	.67	.64	.35	.96	.98	1.03	.97 (89–1.05)
rs1801582 (PARK2) ^l	.0083	.60	.87	1.10	.92	1.02	.95	.98 (89–1.07)
rs6856 (BSCL2) ^m	1.5E-05	.64	.23	1.13	.78	1.00	1.05	.99 (91–1.08)
rs2303291 (ITSN2)	.0048	1.31	2.04	.96	.88	.95	1.12	.98 (90–1.06)
rs160277 (VCAN)	.0045	1.43	1.67	.93	.75	.92	1.02	.96 (88–1.05)

CEU, Utah residents with Northern and Western European ancestry from the CEPH collection; CI, confidence interval; Het, heterozygous; Hom, homozygous; OR, odds ratio; SNP, single nucleotide polymorphism.

^aCochran-Armitage test for trend.

^bExtended Mantel-Haenszel test for trend (two-tailed).

^cOR of heterozygous for the major allele.

^dOR of homozygous for the minor allele referred to homozygous for the major allele.

^eProxy SNP in SGENE-plus; rs4750685 ($r^2 = 1.00$ in HapMap CEU samples).

^fProxy SNP in SGENE-plus; rs7069185 ($r^2 = 1.00$ in HapMap CEU samples).

^gProxy SNP in SGENE-plus; rs1366842 ($r^2 = 1.00$ in HapMap CEU samples).

^hProxy SNP in SGENE-plus; rs3787986 ($r^2 = .89$ in HapMap CEU samples).

ⁱProxy SNP in SGENE-plus; rs558948 ($r^2 = .90$ in HapMap CEU samples).

^jProxy SNP in SGENE-plus; rs1945627 ($r^2 = .85$ in HapMap CEU samples).

^kProxy SNP in SGENE-plus; rs2058519 ($r^2 = .64$ in HapMap CEU samples).

^lProxy SNP in SGENE-plus; rs11964284 ($r^2 = 1.00$ in The 1000 Genomes CEU samples).

^mProxy SNP in SGENE-plus; rs549312 ($r^2 = .71$ in The 1000 Genomes CEU samples).

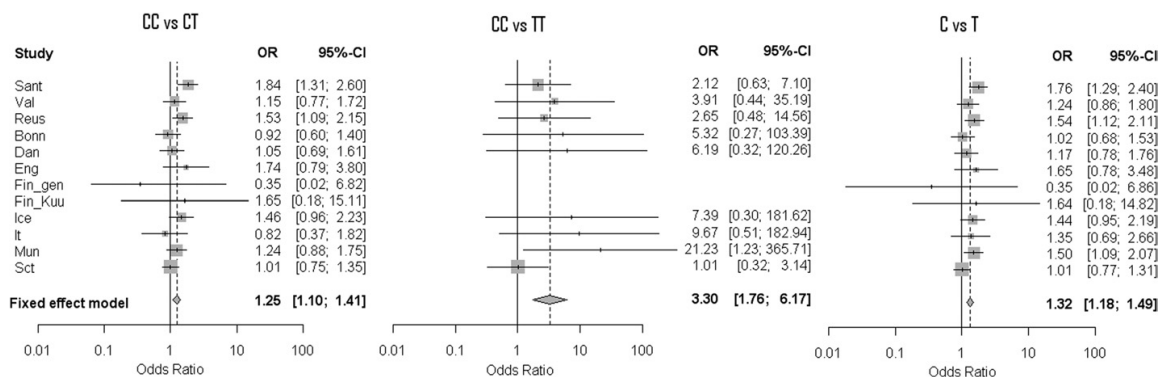


Figure 2. Forest plot of rs13107325 at genotype and allelic levels. The size of the squares reflects relative weights of samples within each plot but are at different scale at the different plots. An increment of .5 was added to each cell in 2×2 tables with zero frequencies in one cell to calculate approximate odds ratios. CI, confidence interval; OR, odds ratio.

variant of small effect in most populations, a type of variant very hard to identify by genetic tools. Its increase in frequency in some European populations, probably by genetic drift, allowed its identification.

SLC39A8 is one of the 14 human members of the family 39 of solute carrier transporters (*SLC39*), the metal ion transporters. It acts in the transport of the essential metals manganese (Mn^{2+}) and zinc (Zn^{2+}) (33,34), which play important roles in the development and functioning of many organs, including the brain, although may be neurotoxic at high concentrations (35,36).

Approximately 20% of Zn in the brain is located in synaptic vesicles, instead of being protein-bound Zn (37), playing a role in the neuromodulation of synaptic transmission and plasticity in limbic structures and cerebral cortex (36,37). Manganese is a cofactor of several enzymes, such as the antioxidant enzyme superoxide dismutase, the main detoxifier of reactive oxygen species in the mitochondria, or glutamine synthetase, involved in the conversion of glutamate to glutamine in astrocytes. High levels of blood Mn increase neuropsychiatric symptoms in people at risk of alcohol use disorders (38). Excessive environmental exposure to Mn is neurotoxic, leading to parkinsonian as well as psychiatric symptoms such as depression, agitation, or hallucinations (39). *SLC39A8* also acts in the transport of the nonessential neurotoxic metal cadmium (34,40).

Among other processes, *SLC39A8* is involved in uptake of these cations from blood in vascular endothelial cells into different tissues and organs (34,40). First, identified in monocytes after innate immune activation, *SLC39A8* is expressed in many tissues including

placenta (at high levels) and brain (at low levels) (29,33). Therefore, *SLC39A8* may be involved in metal homeostasis in the developing embryo, acting at the blood-placental barrier, as well as in the brain, acting at the blood-brain barrier (39,41).

Interestingly, a *de novo* deletion affecting *SLC39A8* and 10 additional genes has been discovered in a patient with nonsyndromic mental retardation (42). Although control subjects present some deletions affecting some of these genes, this large deletion is absent from control subjects and there is not deletion in control subjects affecting *SLC39A8*. Thus, according to the workflow for decision making proposed by the authors, the copy number variant may be considered as related to mental retardation.

Other members of the *SLC39* family may be involved in psychiatric phenotypes. *SLC39A3* has been associated with bipolar disorder, as revealed in meta-analysis of two GWAS ($p = 5 \times 10^{-6}$) (43) and subsequent replication study (44). Although far from genome-wide significance, *SLC39A11* was one of the most significant genes in a GWAS study of major depressive disorder (45). In this regard, the 22nd more significant SNP in our study, rs2478568 ($p = .0025$, FDR = .57), is also located in a member of the *SLC39* family, *SLC39A12* (Table S2 in Supplement 1). Functional redundancy among the different members has been suggested (46). Therefore, metal brain homeostasis may be an important factor in susceptibility to different related psychiatric diseases.

Our multilocus analysis agrees with the existence of multiple loci of susceptibility, as reported previously (3,6). We did not attempt direct comparison with previous results because different sample sizes may affect considerably the quality of the polygenic score.

Table 4. Aggregated nsSNPs Score Analysis

Set	Threshold	Mean <i>n</i> nsSNPs (Range)	p^a	AUC ^b	R^2 Difference ^c
All SNPs	$p < .05$	508 (486–528)	.0068	.545	.011
All SNPs	$p < .1$	976 (948–998)	.0002	.567	.021
GO + SzGene	$p < .05$	53 (44–64)	.0186	.542	.009
GO + SzGene	$p < .1$	93 (80–106)	.0073	.545	.011

AUC, area under the curve; GO, Gene Ontology; nsSNP, nonsynonymous single nucleotide polymorphism; SNP, single nucleotide polymorphism.

^a p value of logistic regression of disease state on score, including percentage of missing genotypes and subset of the cross-validation as covariates.

^bAUC of receiver operating characteristic curve.

^cDifference in Nagelkerke's pseudo R^2 between a model including scores and covariates and the model including only covariates.

Instead, we compared the results of the aggregated nsSNPs score analysis using all genes versus those using a subset of functional candidate genes, representing approximately 10% of the total. If most nsSNPs contributing to the score were, in fact, not related to schizophrenia, better results would be expected, limiting the analysis to functional candidate genes. Nevertheless, a more significant and explanatory result was obtained when all genes were included in the analysis, suggesting that multiple unexpected genes may be related to schizophrenia susceptibility.

The main limit of our work is the reduced sample size in the discovery sample, although the sample size from cases is similar to that leading to the identification of *ZNF804A* (7). We must bear in mind that our power is increased by the use of common nsSNPs. This led to a strong reduction of the number of tests to be corrected for and increases the probability of analysis of functional variants instead of marker SNPs in partial LD with them. Other limits are the lack of a standardized assessment of the psychiatric phenotype and the use of unscreened control subjects for the discovery sample. An additional caveat is that several common nsSNPs were not included in the Affymetrix GeneChip Human 20k cSNP Kit due to technological limitation or because they were not identified at the time of design. Finally, using a less stringent threshold for inclusion of nsSNPs in the replication phase would increase the probability of finding additional associations. In spite of these limitations, we have detected one nsSNP in *SLC39A8* associated with schizophrenia in European populations. Further work must be done to investigate the involved mechanism, with metal ions homeostasis in brain being a plausible candidate.

This work has been funded by Grants FIS/FEDER 08/1522 from ISCIII (Instituto de Salud Carlos III) and INCITE08PXIB9101149PR from Xunta de Galicia to JC and by the REGENPSI (Red de Genética de Enfermedades Neurológicas y Psiquiátricas) network (Xunta de Galicia). Genotyping was performed at the Santiago de Compostela node of Centro Nacional de Genotipado.

We thank the health care professionals of Complejo Hospitalario de Santiago and the staff of the Centro de Transfusión de Galicia who assisted in the recruitment of study participants.

The authors reported no biomedical financial interests or potential conflicts of interest.

Supplementary material cited in this article is available online.

- Sullivan PF, Kendler KS, Neale MC (2003): Schizophrenia as a complex trait: Evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry* 60:1187–1192.
- 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* 373:234–239.
- 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* 460:748–752.
- Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Pe'er I, *et al.* (2009): Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* 460:753–757.
- Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D, *et al.* (2009): Common variants conferring risk of schizophrenia. *Nature* 460:744–747.
- Ikeda M, Aleksic B, Kinoshita Y, Okochi T, Kawashima K, Kushima I, *et al.* (2011): Genome-wide association study of schizophrenia in a Japanese population. *Biol Psychiatry* 69:472–478.
- O'Donovan MC, Craddock N, Norton N, Williams H, Peirce T, Moskvina V, *et al.* (2008): Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet* 40:1053–1055.
- Riley B, Thiselton D, Maher BS, Bigdeli T, Wormley B, McMichael GO, *et al.* (2010): Replication of association between schizophrenia and *ZNF804A* in the Irish Case-Control Study of Schizophrenia sample. *Mol Psychiatry* 15:29–37.
- Steinberg S, Mors O, Borglum AD, Gustafsson O, Werge T, Mortensen PB, *et al.* (2011): Expanding the range of *ZNF804A* variants conferring risk of psychosis. *Mol Psychiatry* 16:59–66.
- International Schizophrenia Consortium (2008): Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 455:237–241.
- Stefansson H, Rujescu D, Cichon S, Pietilainen OP, Ingason A, Steinberg S, *et al.* (2008): Large recurrent microdeletions associated with schizophrenia. *Nature* 455:232–236.
- Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, *et al.* (2008): Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 320:539–543.
- 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* 40:880–885.
- 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* 106:9362–9367.
- Ke X, Taylor MS, Cardon LR (2008): Singleton SNPs in the human genome and implications for genome-wide association studies. *Eur J Hum Genet* 16:506–515.
- Evans DM, Barrett JC, Cardon LR (2008): To what extent do scans of non-synonymous SNPs complement denser genome-wide association studies? *Eur J Hum Genet* 16:718–723.
- Roig B, Virgos C, Franco N, Martorell L, Valero J, Costas J, *et al.* (2007): The discoidin domain receptor 1 as a novel susceptibility gene for schizophrenia. *Mol Psychiatry* 12:833–841.
- McGuffin P, Farmer A, Harvey I (1991): A polydiagnostic application of operational criteria in studies of psychotic illness. Development and reliability of the OPCRIT system. *Arch Gen Psychiatry* 48:764–770.
- Hardenbol P, Yu F, Belmont J, Mackenzie J, Bruckner C, Brundage T, *et al.* (2005): Highly multiplexed molecular inversion probe genotyping: Over 10,000 targeted SNPs genotyped in a single tube assay. *Genome Res* 15:269–275.
- Pritchard JK, Stephens M, Donnelly P (2000): Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
- Devlin B, Roeder K (1999): Genomic control for association studies. *Biometrics* 55:997–1004.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, *et al.* (2007): PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575.
- Storey JD, Tibshirani R (2003): Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 100:9440–9445.
- Skol AD, Scott LJ, Abecasis GR, Boehnke M (2006): Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 38:209–213.
- Higgins JP, Thompson SG (2002): Quantifying heterogeneity in a meta-analysis. *Stat Med* 21:1539–1558.
- Abramson JH (2004): WINPEPI (PEPI-for-Windows): Computer programs for epidemiologists. *Epidemiol Perspect Innov* 1:6.
- Allen NC, Bagade S, McQueen MB, Ioannidis JP, Kavvoura FK, Khoury MJ, *et al.* (2008): Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: The SzGene database. *Nat Genet* 40:827–834.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, *et al.* (2000): Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25:25–29.
- Girijashanker K, He L, Soleimani M, Reed JM, Li H, Liu Z, *et al.* (2008): *SLC39A14* gene encodes ZIP14, a metal/bicarbonate symporter: Similarities to the ZIP8 transporter. *Mol Pharmacol* 73:1413–1423.
- Taylor KM, Nicholson RI (2003): The LZT proteins; the LIV-1 subfamily of zinc transporters. *Biochim Biophys Acta* 1611:16–30.
- Waterworth DM, Ricketts SL, Song K, Chen L, Zhao JH, Ripatti S, *et al.* (2010): Genetic variants influencing circulating lipid levels and risk of coronary artery disease. *Arterioscler Thromb Vasc Biol* 30:2264–2276.
- Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, *et al.* (2010): Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 42:937–948.
- Begum NA, Kobayashi M, Moriwaki Y, Matsumoto M, Toyoshima K, Seya T (2002): Mycobacterium bovis BCG cell wall and lipopolysaccharide

- induce a novel gene, BIGM103, encoding a 7-TM protein: Identification of a new protein family having Zn-transporter and Zn-metalloprotease signatures. *Genomics* 80:630–645.
34. He L, Girijashanker K, Dalton TP, Reed J, Li H, Soleimani M, Nebert DW (2006): ZIP8, member of the solute-carrier-39 (SLC39) metal-transporter family: Characterization of transporter properties. *Mol Pharmacol* 70: 171–180.
 35. Takeda A, Sotogaku N, Oku N (2003): Influence of manganese on the release of neurotransmitters in rat striatum. *Brain Res* 965:279–282.
 36. Bitanirwe BK, Cunningham MG (2009): Zinc: The brain's dark horse. *Synapse* 63:1029–1049.
 37. Nakashima AS, Dyck RH (2009): Zinc and cortical plasticity. *Brain Res Rev* 59:347–373.
 38. Sassine MP, Mergler D, Bowler R, Hudnell HK (2002): Manganese accentuates adverse mental health effects associated with alcohol use disorders. *Biol Psychiatry* 51:909–921.
 39. Aschner M, Guilarte TR, Schneider JS, Zheng W (2007): Manganese: Recent advances in understanding its transport and neurotoxicity. *Toxicol Appl Pharmacol* 221:131–147.
 40. Dalton TP, He L, Wang B, Miller ML, Jin L, Stringer KF, *et al.* (2005): Identification of mouse SLC39A8 as the transporter responsible for cadmium-induced toxicity in the testis. *Proc Natl Acad Sci U S A* 102:3401–3406.
 41. Au C, Benedetto A, Aschner M (2008): Manganese transport in eukaryotes: The role of DMT1. *Neurotoxicology* 29:569–576.
 42. Koolen DA, Pfundt R, de Leeuw N, Hehir-Kwa JY, Nillesen WM, Neefs I, *et al.* (2009): Genomic microarrays in mental retardation: A practical workflow for diagnostic applications. *Hum Mutat* 30:283–292.
 43. Baum AE, Hamshere M, Green E, Cichon S, Rietschel M, Noethen MM, *et al.* (2008): Meta-analysis of two genome-wide association studies of bipolar disorder reveals important points of agreement. *Mol Psychiatry* 13:466–467.
 44. Ollila HM, Soronen P, Silander K, Palo OM, Kiesepa T, Kaunisto MA, *et al.* (2009): Findings from bipolar disorder genome-wide association studies replicate in a Finnish bipolar family-cohort. *Mol Psychiatry* 14: 351–353.
 45. Muglia P, Tozzi F, Galwey NW, Francks C, Upmanyu R, Kong XQ, *et al.* (2010): Genome-wide association study of recurrent major depressive disorder in two European case-control cohorts. *Mol Psychiatry* 15:589–601.
 46. Dufner-Beattie J, Huang ZL, Geiser J, Xu W, Andrews GK (2005): Generation and characterization of mice lacking the zinc uptake transporter ZIP3. *Mol Cell Biol* 25:5607–5615.

Estudio 6. Análisis de SNPs no sinónimos de baja frecuencia en genes candidatos.

Effect of low frequency missense SNPs on schizophrenia risk

Noa Carrera^{a,b}, Manuel Arrojo^c, Ramón Ramos-Ríos^c, Eduardo Paz^c, Mario Páramo^c,
Santiago Agra^c, Julio Brenlla^c, Silvia Martínez^c, Ángel Carracedo^{a,b,d}, Javier Costas^a

^aFundación Pública Galega de Medicina Xenómica-SERGAS, Hospital Clínico
Universitario, Santiago de Compostela, Spain

^bDpto. Anatomía Patolóxica e Ciencias Forenses, Universidade de Santiago de
Compostela, Spain

^cServicio de Psiquiatría-SERGAS, Complexo Hospitalario Universitario, Santiago de
Compostela, Spain

^dCIBERER, Centro Nacional de Genotipado, Universidade de Santiago de Compostela,
Spain

Corresponding author complete contact information:

Javier Costas

Fundación Pública Galega de Medicina Xenómica-SERGAS

Hospital Clínico Universitario de Santiago

Edif. Consultas Planta -2

E-15706 Santiago de Compostela, Spain

Phone: +34-981951490

Fax: +34-981951473

E-mail: javier.costas@usc.es

Abstract

Low frequency variants may contribute to susceptibility to complex diseases. We genotyped 3554 missense SNPs at frequency $< 5\%$ in 476 schizophrenic patients and 447 controls, followed by a joint analysis at single genes and candidate gene-sets, incorporating predictions of functional effect and frequency to increase statistical power. Our main results are associations found at *MUTYH* and *AH11*, a gene previously associated with schizophrenia, although neither of the results resists Bonferroni correction. We did not find evidence of the polygenic mutation-selection balance model suggesting that most low-frequency missense SNPs in current SNP panels are of minor effect in schizophrenia risk.

Keywords: psychosis; rare variants; nsSNPs; gene-set analysis; VT test; mutation-selection balance

1. Introduction

Genome-wide association studies (GWAs) in schizophrenia have identified several common risk variants, although they explain a very small proportion of the heritability (Purcell et al., 2009; Stefansson et al., 2009; Steinberg et al., 2011). This may be due in part to the existence of rare or low frequency polymorphisms of higher risk (Manolio et al., 2009), in agreement with the polygenic mutation-selection balance model. This model may be the most appropriate to understand genetic risk to schizophrenia, due to the large number of genes that may be involved in schizophrenia risk and the reproductive disadvantage of schizophrenic patients (Keller and Miller, 2006; Purcell et al., 2009; Uher, 2009). The polygenic mutation-selection balance model is well supported from population genetics data (Barreiro et al., 2008; Gorlov et al., 2008; Kryukov et al., 2007). For instance, Kryukov et al (2007) showed that a substantial proportion of missense SNPs confers deleterious effects but these effects are weak enough to avoid removal by natural selection. Thus, these SNPs are segregating in populations at low frequencies, probably contributing to common diseases.

The analysis of rare or low frequency variants represents a statistical challenge due to the reduced power of single variant analysis. To increase power, joint analysis of the different variants at individual genes or gene-sets has been suggested, based on collapsing rare variants to classify individuals as carriers versus non-carriers or, alternatively, scoring individuals according to their load of rare variants (Bansal et al., 2010; Manolio et al., 2009). Adding information about the putative functionality of the SNPs or about their frequency to weight the scores may increase the chance to test causal SNPs (Madsen and Browning, 2009; Price et al., 2010).

With the arrival of next generation sequencing and resources such as the 1000 genomes project (The 1000 Genomes Project Consortium, 2010), data on rare and private variants will be available. In the meantime, dbSNP presents multiple SNPs at low frequency that may be useful to test the role of less common variants, usually excluded in GWAs analysis, in schizophrenia susceptibility.

In our previous work, we presented data of common (minor allele frequency (MAF) > 5%) missense SNPs in 476 schizophrenic patients and 447 controls using the Affymetrix 20k cSNPs array, specifically designed to genotyped the missense SNPs presented in dbSNP v124 (Carrera et al., 2011). Here, we analyzed the low frequency missense SNPs of this dataset and 129 additional low frequency missense SNPs of predicted functionality in candidate schizophrenia genes.

2. Materials and Methods

Four hundred seventy six patients meeting DSMIV diagnosis of schizophrenia and 447 controls from Galicia (NW Spain) that passed quality control procedures in our previous analysis (Carrera et al., 2011) were included in the present study.

Genotyping was done using the Affymetrix 20k cSNPs panel, as described previously (Carrera et al., 2011). In addition, 129 rare missense SNPs mainly from HapMap CEU data, chosen based on their putative impact on functionality according to PolyPhen (Ramensky et al., 2002) were genotyped by Sequenom (Supplementary Table 1). Based on 18 SNPs common to both platforms, the genotyping concordance rate was estimated as 0.994.

Four exclusion criteria for SNPs were applied as genotyping quality control: i) genotyping call rate lower than 95%; ii) significant difference in call rates between cases and controls ($P < 0.001$); iii) departure from Hardy-Weinberg equilibrium in controls ($P < 0.001$); and iv) MAF > 5% in cases and controls.

Single gene analysis was done with the help of CCRaVAT (Lawrence et al., 2010). This software collapses all the missense SNPs at MAF > 5% at each gene. Tests of excess of carriers of rare variants in cases versus controls were done by one-tailed Fisher's exact tests. Gene-set analysis was done with VTtest (Price et al., 2010). This program generates individual scores based on the number of missense SNPs per sample, alternatively weighted by putative functionality, as predicted by PolyPhen 2.1 (Adzhubei et al., 2010). The final score of the test is the difference in the sum of individual scores in cases minus the sum of individual scores in controls. A variable frequency threshold is used to maximize the value of the test. Significance of the results was assessed by 10000 permutations. All the analyses were restricted to autosomes.

Five different sets of genes were considered based on functional relevance (Supplementary Table 2): 1) genes that interact with *DISC1* at the protein level, collected from STRING, Entrez Gene and bibliographic searches; 2) all the genes

present in SchizophreniaGene, a database of genetic association studies in schizophrenia (Allen et al., 2008); 3) genes detected in the postsynaptic density (PSD) from human neocortex (Bayes et al., 2010); considered of relevance in schizophrenia from the glutamatergic dysfunction hypothesis; 4) genes from the Kyoto Encyclopaedia of Genes and Genomes (KEGG) Cell adhesion molecule (CAM) pathway (pathway ID: hsa04514), recently associated to schizophrenia susceptibility (O'Dushlaine et al., 2010); 5) genes with a role in synaptic transmission and/or neurodevelopment according to Gene Ontology (Ashburner et al., 2000). These gene-sets partially overlap (Supplementary Figure 1).

3. Results and Discussion

3.1. Analysis gene by gene: CCRaVAT

After QC procedure, the final set comprised 3554 low frequency missense SNPs at 2608 autosomal genes.

The most significant association was found at *MUTYH* (Supplementary Table 3). There was just one low-frequency missense SNP at this gene, rs3219484. Fifty seven of 472 cases (12.1%) versus 27 of 446 controls (6.1%) were carriers of the low-frequency variant (one-tailed Fisher's exact test $P = 0.001$; OR, 2.13; 95% CI, 1.30-3.58). The most significant association of a gene with several low-frequency missense SNPs was at *MCF2L2*. There were three low-frequency missense SNPs at this gene, rs9826325, rs3732602, and rs2293203. Thirty four of 463 cases (7.3%) versus 17 of 438 controls (3.9%) were carriers of at least one rare variant (one-tailed Fisher's exact test $P = 0.017$; OR, 1.96; 95% CI, 1.05-3.80). One of the genes with stronger evidence of association to schizophrenia was present among the top results, *AHII*, rank order 26 (Ingason et al., 2010). The unique low-frequency missense SNP genotyped at this gene, rs6940875, was absent from controls but present in 5 cases (one-tailed Fisher's exact test $P = 0.036$). Nineteen of the 49 SNPs at significant genes are predicted to be functional by PolyPhen 2.0, including the missense SNPs at *MUTYH* and *AHII* (Supplementary Table 3).

Nevertheless, none of the results were significant after Bonferroni's correction for multiple tests. In order to interpret these results, we have to take into account the power of our study. As shown in Figure 1, we have 80% power to detect association if low-frequency variants are at frequencies of 0.005, 0.01, 0.025, or 0.05 in controls and 0.06, 0.075, 0.105, or 0.15 in cases, respectively. Therefore, we cannot discard risk variants of lower effect.

3.2. Gene set analysis: VT test

In order to increase power, we performed further analysis at specific gene sets based on their putative functional relationship to schizophrenia. The use of missense SNPs has the advantage of unambiguous assignment of SNP to gene, as well as relatively accurate prediction of functionality. This allows specific analysis such as collapsing SNPs in

pathways or weighting SNPs based on putative functionality. In spite of this strategy, our main results discarded an accumulation of low frequency missense SNPs in cases in any of the gene sets (Table 1).

This result do not necessarily imply the irrelevance of these gene sets or the absence of low frequency missense SNPs of susceptibility to schizophrenia, but the underrepresentation of these variants in dbSNP and current genotyping arrays. Rare variants were underrepresented in dbSNP, although there are new data from the 1000 genome project since dbSNP v.132. In addition, there may be a problem of ascertainment bias. Most variants in dbSNP have been detected in general population, thus putatively lacking many variants at low frequency present mainly in patients. Furthermore, most samples contributing to dbSNP data are Northern Europeans, while our study is based on Southwestern Europeans. Under the polygenic mutation-selection balance model, risk variants are expected to be very young, increasing the probability to be population-specific (Uher, 2009). The application of next generation sequencing to case-control samples in combination with collapsing techniques such as those applied here, will allow to test the role of rarer variants in schizophrenia risk. Other reasons for our negative results may be the putative role of regulatory regions, gene x gene or gene x environment interactions (Costas et al., 2011; van Os et al., 2008). Fail to consider these interactions may explain the limited success in the identification of the genetic basis of schizophrenia.

In conclusion, in spite of the large number of missense SNPs tested and the application of statistical analyses specifically design to increase power in study of rare variants, our genome-wide survey of low-frequency missense SNPs did not detect evidence of the polygenic mutation-selection balance model. This fact may be explained by ascertainment bias in dbSNP or by the role of regulatory low frequency SNPs, rarer SNPs, or gene x gene and gene x environment interactions.

4. References

- Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., Sunyaev, S.R., 2010. A method and server for predicting damaging missense mutations. *Nat Methods* 7(4), 248-249.
- Allen, N.C., Bagade, S., McQueen, M.B., Ioannidis, J.P., Kavvoura, F.K., Khoury, M.J., Tanzi, R.E., Bertram, L., 2008. Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat Genet* 40, 827-834.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M.,

- Sherlock, G., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25(1), 25-29.
- Bansal, V., Libiger, O., Torkamani, A., Schork, N.J., 2010. Statistical analysis strategies for association studies involving rare variants. *Nat Rev Genet* 11(11), 773-785.
- Barreiro, L.B., 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.
- Bayes, A., van de Lagemaat, L.N., Collins, M.O., Croning, M.D., Whittle, I.R., Choudhary, J.S., Grant, S.G., 2010. Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nat Neurosci* 14(1), 19-21.
- Carrera, N., Arrojo, M., Sanjuán, J., Ramos-Ríos, R., Paz, E., Suárez-Rama, J.J., Páramo, M., Agra, S., Brenlla, J., Martínez, S., Rivero, O., Collier, D.A., Palotie, A., Cichon, S., Nöthen, M.M., Rietschel, M., Rujescu, D., Stefansson H, Steinberg S, Sigurdsson E, St Clair, D, Tosato S, Werge, T., Stefansson, K., González, J.C., Valero, J., Gutiérrez-Zotes, A., Labad, A., Martorell, L., Vilella, E., Carracedo, A., Costas, J., 2011. Association study of non-synonymous single nucleotide polymorphisms in schizophrenia. *Biol Psychiatry*. *in press*
- Costas, J., Sanjuan, J., Ramos-Rios, R., Paz, E., Agra, S., Tolosa, A., Paramo, M., Brenlla, J., Arrojo, M., 2011. Interaction between COMT haplotypes and cannabis in schizophrenia: a case-only study in two samples from Spain. *Schizophr Res* 127(1-3), 22-27.
- Faul, F., Erdfelder, E., Lang, A.G., Buchner, A., 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 39(2), 175-191.
- Gorlov, I.P., Gorlova, O.Y., Sunyaev, S.R., Spitz, M.R., Amos, C.I., 2008. Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. *Am J Hum Genet* 82(1), 100-112.
- Ingason, A., Giegling, I., Cichon, S., Hansen, T., Rasmussen, H.B., Nielsen, J., Jurgens, G., Muglia, P., Hartmann, A.M., Strengman, E., Vasilescu, C., Muhleisen, T.W., Djurovic, S., Melle, I., Lerer, B., Moller, H.J., Francks, C., Pietilainen, O.P., Lonnqvist, J., Suvisaari, J., Tuulio-Henriksson, A., Walshe, M., Vassos, E., Di Forti, M., Murray, R., Bonetto, C., Tosato, S., Cantor, R.M., Rietschel, M., Craddock, N., Owen, M.J., Peltonen, L., Andreassen, O.A., Nothen, M.M., St Clair, D., Ophoff, R.A., O'Donovan, M.C., Collier, D.A., Werge, T., Rujescu, D., 2010. A large replication study and meta-analysis in European samples provides further support for association of AH11 markers with schizophrenia. *Hum Mol Genet* 19(7), 1379-1386.
- Keller, M.C., Miller, G., 2006. Resolving the paradox of common, harmful, heritable mental disorders: which evolutionary genetic models work best? *Behav Brain Sci* 29(4), 385-404; discussion 405-352.

- Kryukov, G.V., Pennacchio, L.A., Sunyaev, S.R., 2007. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am J Hum Genet* 80(4), 727-739.
- Lawrence, R., Day-Williams, A.G., Elliott, K.S., Morris, A.P., Zeggini, E., 2010. CCRaVAT and QuTie-enabling analysis of rare variants in large-scale case control and quantitative trait association studies. *BMC Bioinformatics* 11, 527.
- Madsen, B.E., Browning, S.R., 2009. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet* 5(2), e1000384.
- Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorf, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F., McCarroll, S.A., Visscher, P.M., 2009. Finding the missing heritability of complex diseases. *Nature* 461(7265), 747-753.
- O'Dushlaine, C., Kenny, E., Heron, E., Donohoe, G., Gill, M., Morris, D., Corvin, A., 2010. Molecular pathways involved in neuronal cell adhesion and membrane scaffolding contribute to schizophrenia and bipolar disorder susceptibility. *Mol Psychiatry* 16(3), 286-292.
- Price, A.L., Kryukov, G.V., de Bakker, P.I., Purcell, S.M., Staples, J., Wei, L.J., Sunyaev, S.R., 2010. Pooled association tests for rare variants in exon-resequencing studies. *Am J Hum Genet* 86(6), 832-838.
- Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., Sklar, P., 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460(7256), 748-752.
- Ramensky, V., Bork, P., Sunyaev, S., 2002. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* 30(17), 3894-3900.
- Stefansson, H., Ophoff, R.A., Steinberg, S., Andreassen, O.A., Cichon, S., Rujescu, D., Werge, T., Pietilainen, O.P., Mors, O., Mortensen, P.B., Sigurdsson, E., Gustafsson, O., Nyegaard, M., Tuulio-Henriksson, A., Ingason, A., Hansen, T., Suvisaari, J., Lonnqvist, J., Paunio, T., Borglum, A.D., Hartmann, A., Fink-Jensen, A., Nordentoft, M., Hougaard, D., Norgaard-Pedersen, B., Bottcher, Y., Olesen, J., Breuer, R., Moller, H.J., Giegling, I., Rasmussen, H.B., Timm, S., Mattheisen, M., Bitter, I., Rethelyi, J.M., Magnusdottir, B.B., Sigmundsson, T., Olason, P., Masson, G., Gulcher, J.R., Haraldsson, M., Fossdal, R., Thorgeirsson, T.E., Thorsteinsdottir, U., Ruggeri, M., Tosato, S., Franke, B., Strengman, E., Kiemenev, L.A., Melle, I., Djurovic, S., Abramova, L., Kaleda, V., Sanjuan, J., de Frutos, R., Bramon, E., Vassos, E., Fraser, G., Ettinger, U., Picchioni, M., Walker, N., Touloupoulou, T., Need, A.C., Ge, D., Yoon, J.L., Shianna, K.V., Freimer, N.B., Cantor, R.M., Murray, R., Kong, A., Golimbet, V., Carracedo, A., Arango, C., Costas, J., Jonsson, E.G., Terenius, L., Agartz, I., Petursson, H., Nothen, M.M., Rietschel, M., Matthews, P.M., Muglia, P., Peltonen, L., St Clair, D.,

Goldstein, D.B., Stefansson, K., Collier, D.A., 2009. Common variants conferring risk of schizophrenia. *Nature* 460, 744-747.

Steinberg, S., de Jong, S., Andreassen, O.A., Werge, T., Borglum, A.D., Mors, O., Mortensen, P.B., Gustafsson, O., Costas, J., Pietilainen, O.P., Demontis, D., Papiol, S., Huttenlocher, J., Mattheisen, M., Breuer, R., Vassos, E., Giegling, I., Fraser, G., Walker, N., Tuulio-Henriksson, A., Suvisaari, J., Lonnqvist, J., Paunio, T., Agartz, I., Melle, I., Djurovic, S., Strengman, E., Jurgens, G., Glenthøj, B., Terenius, L., Hougaard, D.M., Orntoft, T., Wiuf, C., Didriksen, M., Hollegaard, M.V., Nordentoft, M., van Winkel, R., Kenis, G., Abramova, L., Kaleda, V., Arrojo, M., Sanjuan, J., Arango, C., Sperling, S., Rossner, M., Ribolsi, M., Magni, V., Siracusano, A., Christiansen, C., Kiemeneý, L.A., Veldink, J., van den Berg, L., Ingason, A., Muglia, P., Murray, R., Nothen, M.M., Sigurdsson, E., Petursson, H., Thorsteinsdottir, U., Kong, A., Rubino, I.A., De Hert, M., Rethelyi, J.M., Bitter, I., Jonsson, E.G., Golimbet, V., Carracedo, A., Ehrenreich, H., Craddock, N., Owen, M.J., O'Donovan, M.C., Ruggeri, M., Tosato, S., Peltonen, L., Ophoff, R.A., Collier, D.A., St Clair, D., Rietschel, M., Cichon, S., Stefansson, H., Rujescu, D., Stefansson, K., 2011. Common variants at VRK2 and TCF4 conferring risk of schizophrenia. *Hum Mol Genet*.

The 1000 Genomes Project Consortium., 2010. A map of human genome variation from population-scale sequencing. *Nature* 467(7319), 1061-1073.

Uher, R., 2009. The role of genetic variation in the causation of mental illness: an evolution-informed framework. *Mol Psychiatry* 14(12), 1072-1082.

van Os, J., Rutten, B.P., Poulton, R., 2008. Gene-environment interactions in schizophrenia: review of epidemiological findings and future directions. *Schizophr Bull* 34(6), 1066-1082.

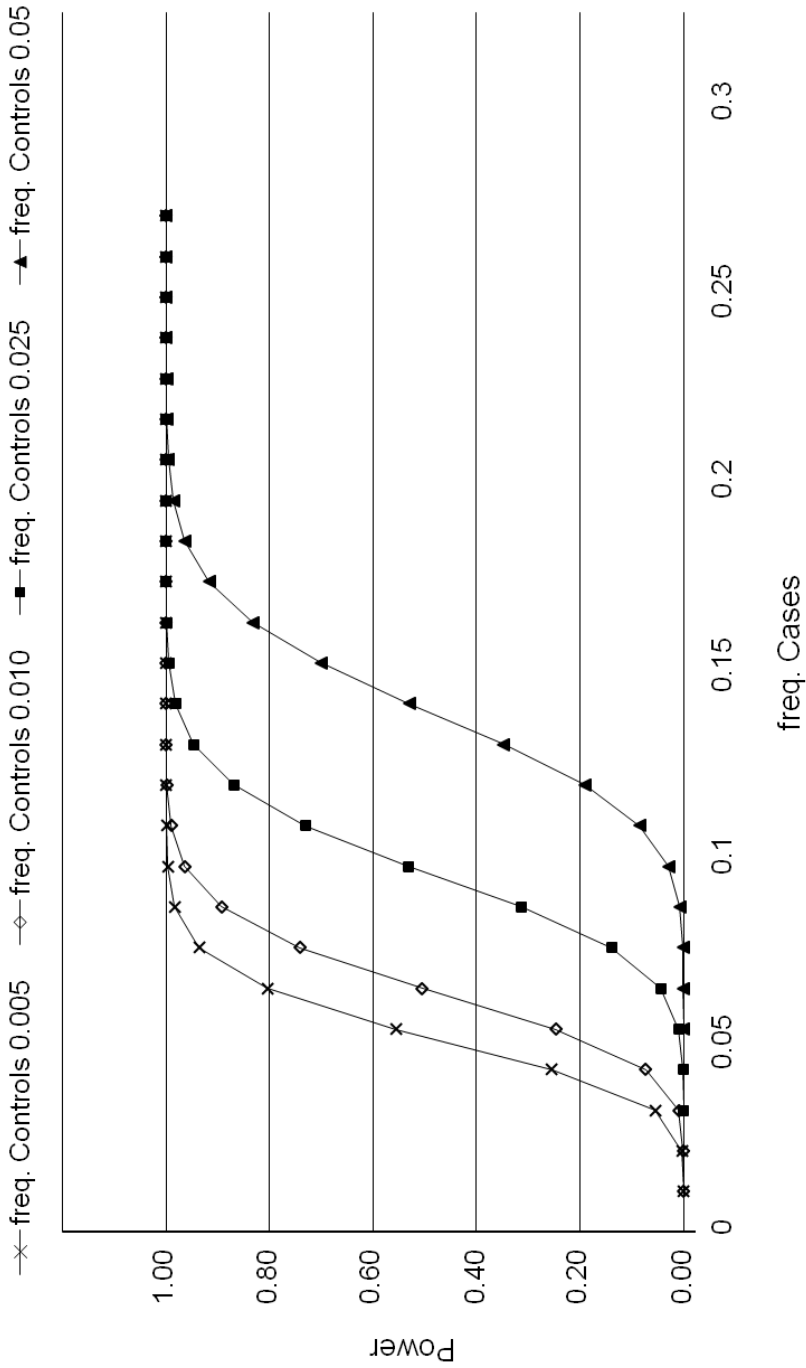


Figure 1. Power calculations at single gene analysis. The calculations were done assuming Bonferromi's correction for 3554 tests, using G*Power 3.0.10 (Faul et al., 2007).

Table 1. Gene set analysis of low frequency variants.

Gene set	# genes ^a	# low freq missense SNPs	VT p value ^b	PolyPhen VT p value ^c
DISC1 interactome	18	32	0.87	0.95
SchizophreniaGene	159	236	0.85	0.74
CAM	23	35	0.44	0.49
PSD	67	86	0.58	0.53
GO	63	115	1	1

^aAutosomal genes with low frequency missense SNPs.

^bOne-sided p value of the VT test

^cOne-sided p value of the VT test with incorporation of PolyPhen-2 scores. For those missense SNPs without PolyPhen prediction, a “neutral” score of 0.5 was used, as suggested (Price et al., 2010).

Supplementary Table 1. List of missense SNPs genotyped by Sequenom.

Supplementary Table 2. Gene sets analyzed in the present study.

Supplementary Table 3. Significant results at single gene analysis.

Supplementary Figure 1. Venn's diagram representing the overlap between the different gene sets.

Discusión general

La identificación de factores de predisposición a padecer enfermedades complejas es uno de los grandes retos de la genética médica actual. En el caso de la esquizofrenia, y en general de las enfermedades psiquiátricas, esta búsqueda se ha visto dificultada por la ausencia de un sustrato biológico conocido que permita guiar claramente esta búsqueda.

En un intento por aportar alguna variante de riesgo con una funcionalidad clara, el objetivo central de esta tesis ha sido la identificación de variantes de predisposición a padecer la enfermedad mediante el análisis de variantes con alta probabilidad *a priori* de ser funcionales. Atendiendo al modo en que hemos abordado esta búsqueda de susceptibilidad genética, el conjunto de los seis trabajos que son la parte sustantiva de esta memoria puede ser dividido en dos partes.

En la primera, en la que se incluyen los cuatro primeros trabajos, buscamos variantes que hubiesen sido objeto de selección natural reciente, y por tanto funcionalmente diferentes, en genes candidatos.

En la segunda parte, en la que se incluyen los dos últimos trabajos, exploramos sistemáticamente el papel de SNPs no sinónimos, polimorfismos con elevada probabilidad de ser funcionales, en la etiología del trastorno.

Como fue comentado en la introducción, la esquizofrenia está asociada a elevada heredabilidad y a una desventaja reproductiva, por lo que se puede decir que está sometida a selección negativa. Siendo esto así, y de acuerdo a la hipótesis neutralista de evolución molecular, parece poco probable que puedan existir variantes comunes de susceptibilidad con un efecto moderado o fuerte sobre el fenotipo. Sin embargo se han propuesto escenarios evolutivos en los que esta situación podría darse, como el que propone el modelo de susceptibilidad ancestral o el de pleiotropismo antagónico. En ambos modelos la selección natural reciente juega un papel clave, ya que serían variantes sometidas a selección reciente que no han llegado a fijarse en la población las que intervendrían en la diferente susceptibilidad a padecer enfermedades complejas asociadas a una fertilidad reducida. Para tratar de averiguar si en alguna medida la esquizofrenia puede ser explicada con estos modelos, decidimos seguir una estrategia evolutiva centrada en la búsqueda de variantes sometidas a selección natural reciente en genes candidatos. Este tipo de variantes comunes, sometidas a selección natural reciente, son el sustrato perfecto para la hipótesis enfermedad común / variante común, ya que son variantes funcional y fenotípicamente diferentes.

En el primero de los trabajos (estudio 1) confirmamos la existencia de un haplotipo común de protección frente a la enfermedad, identificado previamente por nuestro grupo (Dominguez et al., 2007), en un total de 794 casos y 1078 controles pertenecientes a tres poblaciones independientes (resultado de la prueba Mantel-Haenszel de cuatro muestras independientes: $p = 0,00227$; OR = 0,79, IC 95% = 0,68 - 0,92). A pesar de haber secuenciado los exones y regiones conservadas del gen en 10 individuos homocigotos para el haplotipo protector no hemos podido identificar ninguna variante funcional asociada a este haplotipo. Detectamos, además, mediante la aplicación de la prueba LRH (Sabeti et al., 2002), señales de selección natural reciente sobre *DRD3* en poblaciones no africanas. Es probable que la diana de este evento de selección sea el SNP no sinónimo Ser9Gly (rs6280). El haplotipo protector está en desequilibrio de ligamiento con este SNP ($D'=1$), por lo que es posible que el haplotipo protector haya alcanzado tanta frecuencia por arrastre, debido a la acción de la selección natural reciente sobre rs6280.

Un trabajo posterior intentó reproducir en población japonesa la asociación que nosotros identificamos con este haplotipo de protección (Nunokawa et al., 2009). Para ello genotiparon los mismos SNPs marcadores en una primera muestra de 595 pacientes y 598 controles y en una segunda muestra independiente de 2126 pacientes y 2228 controles. En la primera muestra la frecuencia del haplotipo protector era menor en casos que en controles (27% vs. 29%), aunque la diferencia no era significativa ($p = 0,261$; OR = 0,91, IC 95% = 0,76 - 1,08), y en la segunda población no se observó diferencia (28% vs. 28%; $p = 0,793$; OR = 1,00, IC 95% = 0,91 - 1,10). En el análisis combinado de todas las muestras, incluyendo las europeas, tampoco se observó

diferencia de frecuencia (OR = 0,93, IC 95% = 0,87 – 1,00). Por tanto, este estudio no confirmó que en la población japonesa analizada el haplotipo fuese de protección. Sin embargo, tal y como indican los autores en la discusión del artículo, este hecho no descarta que en población europea sí sea de protección, pues ésta puede depender de interacciones epistáticas o de interacciones con factores ambientales ausentes en población asiática.

Desde los inicios de la investigación en esquizofrenia *DRD3* parecía un gen candidato funcional prometedor, ya que es una importante diana para muchos antipsicóticos (Sokoloff et al., 1990) y se localiza en la región 3q13.3, región para la que se ha descrito alguna señal de ligamiento a esquizofrenia (Kaneko et al., 2007). Sin embargo, a pesar de haberse estudiado en multitud de poblaciones y haberse hecho meta-análisis, hasta la fecha no se ha logrado identificar ninguna variante funcional asociada a la enfermedad de manera consistente. Lo que sí se ha descrito son alteraciones de los niveles de expresión de este gen en pacientes esquizofrénicos. Por ejemplo, recientemente Zhang y colaboradores (Zhang et al., 2011a) analizaron este gen en población China Han y observaron menor expresión del gen en pacientes libres de tratamiento que en controles. Otros estudios han identificado diferencias en la expresión del gen, aunque no todos en la misma dirección (Gurevich et al., 1997; Ilani et al., 2001; Meador-Woodruff et al., 1997; Schmauss et al., 1993). Estas observaciones son consistentes con la alteración de la ruta metabólica de la dopamina en la enfermedad, como postula la hipótesis dopaminérgica (Pani, 2002). Esta alteración puede estar causada por modificaciones en el propio gen o en otros genes que participen en la ruta. Estudios funcionales de expresión en los que se evalúe el efecto del haplotipo de protección que hemos identificado podrían aportar algún resultado interesante.

Con respecto a la señal de selección en el entorno del polimorfismo rs6280 que identificamos en población europea, el trabajo de Zhang y colaboradores (Zhang et al., 2011a) antes citado investiga si esta señal existe también en población asiática. Para ello, y al igual que en nuestro trabajo, aplicaron la prueba LRH sobre datos de muestra asiática de HapMap. Este análisis no identificó evidencias claras de barridos selectivos incompletos. Sin embargo, al comparar la frecuencia del alelo derivado del polimorfismo entre poblaciones africana, asiática y europea (tabla II del artículo (Zhang et al., 2011a) , vieron que la frecuencia del alelo en población europea es significativamente mayor que en población africana ($P < 0,001$), que ocurre lo mismo entre la población asiática y africana ($P < 0,001$), pero no entre población asiática y europea ($P = 0,188$). Esto sugiere que la ausencia de señal según la prueba LRH puede deberse a que la selección sea más antigua en población asiática que en europea, de modo que procesos de recombinación han acortado el tamaño del haplotipo lo suficiente como para que no destaque al compararlo con otros haplotipos de frecuencia similar, o que este proceso de acortamiento del haplotipo haya sido más rápido en población asiática. Por otra parte, hay que tener en cuenta que la ausencia de señal significativa de selección no es una prueba decisiva de la ausencia de eventos selectivos, ya que, como comentaremos más adelante, las pruebas de este tipo

presentan una serie de limitaciones y no están exentas de falsos negativos (Kelley et al., 2006; Sabeti et al., 2006; Voight et al., 2006). Entre otros motivos, al desconocer qué porcentaje del genoma está sometido a selección natural, nuestra capacidad para determinar un umbral preciso que acote variantes con mayor probabilidad de haber sufrido selección es limitada.

En el segundo trabajo (estudio 2) decidimos probar si el modelo de susceptibilidad ancestral (Di Rienzo, 2006; Di Rienzo y Hudson, 2005) podría explicar parte de la susceptibilidad a padecer esquizofrenia. El modelo propone que alelos que en condiciones ancestrales resultaban ventajosos o neutros, en la actualidad podrían ser deletéreos debido a una mala adaptación a condiciones nuevas. Para ello el gen *MAOB*, gen para el que se han detectado fuertes señales de selección (The International HapMap Consortium, 2005), es decir, que alberga variantes funcionalmente diferentes, y que por su función podría estar implicado en la etiología de la enfermedad, al participar en la desaminación oxidativa de aminas biogénicas, incluyendo la dopamina, y aminas xenobióticas, presentes en diferentes alimentos (Strolin Benedetti et al., 2007).

Los resultados de este trabajo confirman la existencia de un haplotipo en el gen *MAOB* –marcado por el alelo derivado del polimorfismo rs5905512– sometido a selección natural reciente, y sugieren que los haplotipos ancestrales de esta región –marcados por el alelo ancestral de rs5905512– incrementan el riesgo (en hombres) a sufrir la enfermedad ($p = 0,024$; OR=1,41, IC 95% 1,01–1,90), tal y como predice el modelo de susceptibilidad ancestral.

Este mismo polimorfismo ha sido estudiado por otro grupo de investigación en una población independiente catalana (Mas et al., 2009). En este trabajo la frecuencia del alelo ancestral en hombres era claramente superior en casos que en controles, a pesar de que la diferencia no llegaba a ser significativa. De todos modos, tal y como los autores apuntan en su artículo, la muestra empleada no alcanza una potencia suficiente para identificar asociación con una variante de riesgo en torno a 1,41, que es el riesgo atribuido según nuestra muestra al alelo ancestral de rs5905512. Es por ello que nos pareció de interés hacer un análisis conjunto para confirmar o descartar la asociación que identificamos en una muestra más grande y aportar así mayor robustez al estudio.

Como se muestra en la tabla 3, al hacer el meta-análisis el valor p de la prueba de asociación obtenido en hombres es menor que en nuestro estudio ($p = 0,014$ frente a $p = 0,024$; OR=1,37, IC 95% 1,07 – 1,76), por lo que hay razones para pensar que puede tratarse, efectivamente, de una asociación verdadera (resultados no publicados).

Tabla 3. Resultados del meta-análisis para el polimorfismo rs5905512

	Casos (%)		Controles (%)		valor p^1	OR (IC 95%) ¹
	G	A	G	A		
Hombres						
Santiago	99 (60,7)	64 (39,3)	87 (50,3)	86 (49,7)		
Valencia	92 (51,1)	88 (48,9)	79 (44,4)	99 (55,6)		
Barcelona	89 (51,7)	83 (48,3)	62 (45,6)	74 (54,4)	0,014	1,37 (1,07 - 1,76)
Mujeres						
Santiago	106 (57,0)	80 (43,0)	134 (50,8)	130 (49,2)		
Valencia	84 (45,2)	102 (54,8)	114 (50,0)	114 (50,0)		
Barcelona	169 (54,2)	143 (45,8)	231 (51,6)	217 (48,4)	0,504	1,07 (0,88 - 1,30)

¹ Prueba Mantel-Haenszel, considerando 3 poblaciones

En el tercer trabajo presentado (estudio 3) intentamos probar el modelo de pleiotropismo antagónico en esquizofrenia. Este modelo propone que variantes genéticas que incrementan el riesgo a sufrir una enfermedad asociada a fertilidad reducida podrían alcanzar elevada frecuencia en la población si éstas implican también la aparición de un fenotipo con éxito reproductivo.

Decidimos, por tanto, estudiar genes con variantes que se hubiesen relacionado previamente con la enfermedad y con rasgos susceptibles de ser objeto de selección natural reciente. Tras hacer una revisión bibliográfica escogimos los siguientes genes: *DAOA*, para el que se ha descrito la existencia de un haplotipo a ~80 kb en dirección 3'-5' de la región codificante asociado a una mejor fluidez semántica y a esquizofrenia (Opgen-Rhein et al., 2008); *PP1R1B*, para el que se ha sugerido la existencia de un haplotipo común asociado a una mayor eficacia en el procesamiento estriatal y otros rasgos cognitivos en personas sanas, además de susceptibilidad a padecer esquizofrenia (Meyer-Lindenberg et al., 2007); y *APOLI*, en el que se han encontrado señales de selección natural reciente (Costas et al., 2005; Frazer et al., 2007), que está implicado en procesos de protección frente a patógenos (Perez-Morga et al., 2005; Vanhamme et al., 2003) y con cierta evidencia, aunque no bien establecida, de que podría ser un gen de susceptibilidad a padecer esquizofrenia (McGhee et al., 2005; Mimmack et al., 2002; Takahashi et al., 2008).

Con respecto a la búsqueda de señales de selección mediante la aplicación de la prueba LRH, en el caso de *DAOA* y *PP1R1B* no encontramos indicios de desviación de la neutralidad en la secuencia de estos genes o en su entorno. En *APOLI* identificamos evidencia clara de selección en población africana pero menos evidente

en población europea. De todos modos, a pesar de que la situación de pleiotropismo antagónico no se mantiene indefinidamente en el tiempo, resolviéndose con la fijación en la población de uno de los dos alelos, es posible que determinados haplotipos hayan permanecido en la población el tiempo suficiente como para que procesos de recombinación hayan difuminando la señal de barrido selectivo en poblaciones europeas.

Para el estudio de asociación, la muestra disponible consistió en 301 pacientes esquizofrénicos y 604 controles del área de Santiago de Compostela. Las variantes que analizamos fueron, en el caso de *DAOA* y *PPP1R1B*, variantes marcadoras de los haplotipos previamente asociados a la enfermedad y a rasgos cognitivos; en cuanto a *APOLI*, incluimos el SNP marcador del haplotipo seleccionado en población africana, así como tagSNPs para analizar toda la variabilidad común del gen.

En nuestra muestra, los resultados de este estudio parecen descartar la relación de *APOLI* con la enfermedad.

En el caso de *DAOA* y *PPP1R1B* no conseguimos confirmar en nuestra muestra las asociaciones descritas previamente. De todos modos, no podemos descartar que existan otras variantes en estos genes de susceptibilidad en nuestra muestra, ya que nuestro estudio se centró en el análisis de los haplotipos asociados previamente a diferentes rasgos cognitivos y a esquizofrenia, al ser los que encajaban con el modelo de pleiotropismo antagónico, y no analizamos toda su variabilidad común. Por ejemplo, las principales asociaciones positivas descritas con *DAOA* son con dos polimorfismos en dirección 5'-3' del gen, rs778293 y rs3916971, (Corvin et al., 2007; Chumakov et al., 2002; Korostishevsky et al., 2004; Ma et al., 2009; Ohi et al., 2009; Shin et al., 2007; Wang et al., 2004), polimorfismos que no incluimos en nuestro estudio ya que el haplotipo que nosotros encontramos de interés analizar está localizado a 80kb, dirección 3'-5'.

Hay que considerar, además, que los cálculos de potencia se hicieron asumiendo un riesgo para las variantes comunes de al menos 1,5 y, actualmente, según los estudios de genoma completo, parece poco probable que existan variantes comunes de riesgo mayor de 1,3 en esquizofrenia. Por tanto, es posible que nuestra muestra no alcanzase la potencia necesaria para identificar asociación con las variantes analizadas.

Resulta interesante destacar que los resultados de un estudio reciente (Genovese et al., 2010) sugieren que variantes del gen *APOLI* protagonizan, en población africana, una situación de pleiotropismo antagónico como la que nosotros pretendíamos identificar: la incidencia en América de determinadas enfermedades del renales (glomeruloesclerosis focal y segmentaria e hipertensión atribuida a enfermedad del riñón) es sustancialmente mayor en población americana de ascendencia africana (AA) que en población americana con ancestralidad europea. Tratando de encontrar una explicación a esta diferencia, un estudio reciente descubrió que en individuos AA, estas enfermedades están asociadas a dos variantes independientes del gen *APOLI*, factor sérico con capacidad para lisar el *Trypanosoma* (OR = 7,3 y 10,5), y vieron que de todas las variantes de *APOLI*, sólo las que incrementan el riesgo a padecer estas

enfermedades en individuos AA tienen capacidad para lisar el *Trypanosoma brucei rhodesiense*, presente únicamente en el continente africano. Curiosamente estas variantes de riesgo son comunes en África pero no están presentes en Europa, según datos de HapMap, y además se localizan en haplotipos en los que se han identificado señales de selección natural reciente en África, pero no en Europa. Así pues, estas observaciones parecen indicar que la elevada frecuencia en poblaciones africanas y en AA de estas variantes podría ser el resultado de la acción de la selección natural reciente sobre las variantes que lisar el *Trypanosoma*, que en población africana suponen una ventaja, aunque tienen el efecto pleiotrópico de incrementar el riesgo de padecer enfermedades del riñón (también presentes en estas poblaciones). En individuos AA estas variantes no suponen una ventaja, ya que no existe *Trypanosoma* en Norte América, y sólo resultan perjudiciales para las enfermedades renales, por lo que sería una situación de susceptibilidad ancestral.

En el último trabajo de esta línea de investigación (estudio 4) analizamos de forma más general el papel de la selección natural reciente en el espectro alélico de la enfermedad, considerando inicialmente una lista extensa de genes candidatos. En líneas generales, tras hacer una revisión bibliográfica (Arion et al., 2007; Camargo et al., 2007; Gratacos et al., 2009; Jungerius et al., 2008; Moises et al., 2002) y consultar diferentes bases de datos -Schizophrenia Gene,(Allen et al., 2008) Gene Ontology (Ashburner et al., 2000), Panther (Thomas et al., 2003) y Synapse DB (Zhang et al., 2007b)-, seleccionamos un total de 1749 genes. La selección se hizo atendiendo a los siguientes criterios: genes presentes en la base de datos Schizophrenia Gene y genes implicados en funciones potencialmente relevantes para la enfermedad, como la transmisión o plasticidad sináptica, desarrollo del sistema nervioso central, expresión diferencial en esquizofrenia o interacción con *DISC1*. En la primera fase del estudio buscamos barridos selectivos incompletos en la región de estos genes en población europea mediante la aplicación de la prueba LRH (Sabeti et al., 2002) a los datos genotípicos de la población CEU de HapMap. Un total de 163 genes de esta lista mostraban posibles señales de selección (valor p de la prueba de selección $< 0,05$ al menos a dos de las cuatro distancias medidas; tabla suplementaria 1 incluida en el disco compacto adjunto). A continuación, asumiendo que la lista inicial de genes está muy enriquecida en genes involucrados en la etiología de la enfermedad, y teniendo presente el tamaño de nuestra muestra y la necesidad de corregir los resultados del posterior estudio de asociación por el número de pruebas realizadas, decidimos centrar la atención en los genes con las señales más extremas de selección y evitar así un número demasiado grande de pruebas, que restaría potencia al estudio. Así pues, establecimos una serie de filtros para destacar las señales más evidentes de selección. De los 163 genes candidatos, 12 superaron estos filtros.

En la segunda fase del estudio analizamos la relación de estos haplotipos, que probablemente albergan variantes funcionalmente diferentes, con el fenotipo

esquizofrénico. Para ello seleccionamos SNPs marcadores de cada uno de ellos, los genotipamos en una muestra de 391 pacientes y 659 controles e hicimos un estudio de asociación caso-control. Los resultados del estudio sugieren que el haplotipo de *PTPN13*, que incluye varios SNPs no sinónimos (figura 1 del estudio 4), confiere protección frente a la enfermedad (valor p de la prueba de asociación de tendencia = 0,002; tras corrección por Bonferroni = 0,024). La proteína codificada por este gen pertenece a la familia de las tirosina-fosfatasa. A partir de las proteínas con las que interacciona se deduce que participa, entre otros procesos, en el desarrollo y funcionamiento del sistema nervioso central, procesos relevantes en la etiología de la esquizofrenia. La proteína tiene estructura modular, típica de proteínas centrales de ensamblaje de complejos proteicos (*scaffolding protein*), con varios dominios funcionales consecutivos (figura 2 del estudio 4). Dado que el haplotipo identificado abarca varios de estos dominios, podría ser que la combinación concreta de variantes presentes en el haplotipo seleccionado modifique de alguna manera la formación de complejos proteicos, y que esta modificación resulte ventajosa y reduzca el riesgo de padecer la enfermedad. Hasta la fecha no se han hecho muchas investigaciones en torno a este gen en relación a esquizofrenia u otras enfermedades psiquiátricas. Sólo existe cierta evidencia de que podría estar sobrerrepresentado en la corteza prefrontal de individuos esquizofrénicos (Arion et al., 2007), por lo que, tal y como indicamos en la discusión del artículo, este resultado está pendiente de un segundo estudio que confirme o descarte la relación de *PTPN13* con la enfermedad.

Para los otros 11 haplotipos analizados en el estudio caso-control no detectamos asociación en nuestra muestra tras corrección por pruebas múltiples.

En los escenarios evolutivos que proponen los modelos considerados en este trabajo, susceptibilidad ancestral y pleiotropismo antagónico, los mismos loci sobre los que la selección ha actuado son los que intervienen en la diferente susceptibilidad a padecer la enfermedad, como vimos que ocurría en *MAOB* y *PTPN13*. Sin embargo, otro escenario posible en el que la frecuencia de alelos de susceptibilidad se vería también modificada por eventos de selección, aunque de modo indirecto, sería aquel en el que la diana de la selección natural fuese una variante en desequilibrio de ligamiento con algún alelo de susceptibilidad, como sugerimos en el estudio 1 que podría ocurrirle al haplotipo de protección de *DRD3*. Dado que nuestro estudio se centró en el análisis de los haplotipos sometidos a selección positiva y no analizamos toda la variabilidad de cada uno, es posible que esto ocurra en alguno de los 11 genes.

Por ejemplo, para *IL3*, se han descrito asociaciones positivas entre la enfermedad y otros polimorfismos del gen en población europea y asiática (Chen et al., 2007; Luo et al., 2008). Según datos de HapMap, los polimorfismos asociados no están en desequilibrio de ligamiento con el haplotipo seleccionado; sin embargo, dado que un mismo gen de susceptibilidad puede presentar distintos alelos de predisposición, es posible que existan otros que sí estén en desequilibrio. En el caso de *SNCG*, estudios previos sí analizaron el mismo polimorfismo que nosotros (marcador único del haplotipo sometido a selección natural reciente) en muestras de pacientes esquizofrénicos y con trastorno bipolar, pero al igual que nosotros, no encontraron

indicios de asociación con la enfermedad (Fallin et al., 2005). Otro de los 12 genes, *GRIK2*, se estudió previamente, no los polimorfismos incluidos en nuestro estudio, y tampoco se encontró asociación (Bah et al., 2004; Shibata et al., 2002).

Como se ha comentado, además de estos 12 genes, otros 151 presentaban señales positivas de selección según la prueba LRH (tabla suplementaria 1), y aunque no satisfacían los criterios estrictos que establecimos, son lo bastante fuertes como para considerar a estos genes buenos candidatos de haber sido objeto de selección natural reciente. Es muy probable, pues, que hayan participado en diferentes procesos de adaptación y que alberguen variantes funcionales con algún efecto fenotípico, por lo que su estudio puede resultar interesante.

Entre estos genes se encuentran *PDE4B*, *GABRB2* (en el que ya se había detectado selección en un trabajo anterior (Lo et al., 2007)), *NRG1*, *PLXNA2* o *GRIN2B*, para los que numerosos estudios han descrito asociación con la enfermedad, y en los que según los meta-análisis realizados en Schizophrenia Gene, existen polimorfismos asociados de manera significativa a la enfermedad. También están el gen *TXNRD2*, localizado en la región de la microdelección 22q11.21 y con evidencia de estar asociado a esquizofrenia según el mega meta-análisis del *Psychiatric Genome-Wide Association Study Consortium* (tabla S10 en (Ripke et al., 2011), o *CACNA1C* y *ODZ4*, para los que este mismo estudio y el trabajo paralelo en trastorno bipolar han identificado asociaciones muy consistentes con estas enfermedades psiquiátricas (*CACNA1C* a trastorno bipolar y esquizofrenia, *ODZ4* a trastorno bipolar). Es interesante destacar que en el trabajo de Barreiro y colaboradores (Barreiro et al., 2008) también detectan señales de selección en *ODZ4*. En este trabajo analizaron el genoma completo humano para localizar señales de selección natural. Para ello emplearon el estadístico *Fst*, que cuantifica la diferenciación entre dos poblaciones y permite identificar procesos de selección natural distintivos, específicos de poblaciones, que reflejarían adaptación a diferentes condiciones ambientales. Según este estudio, 2 SNPs localizados en *ODZ4* presentan valores extremos de *Fst*, lo que incrementa la veracidad del evento selectivo que nosotros identificamos.

Estas observaciones sugieren que podría ser interesante profundizar en la relación entre estos 151 genes y las enfermedades psiquiátricas desde una perspectiva evolutiva.

En la bibliografía existen numerosos artículos que describen señales de selección en genes relacionados con la esquizofrenia. Por ejemplo, en la región MHC (asociada a esquizofrenia), en la que procesos de selección balanceadora, relacionados probablemente con la resistencia a diferentes patógenos, mantienen en la población múltiples haplotipos comunes desde hace millones de años (Hughes y Nei, 1988; Traherne, 2008). Otro ejemplo es el gen *MDGA1*, en el que el análisis ya citado de Barreiro y colaboradores (Barreiro et al., 2008) encuentra señales de selección; este gen estaba incluido en nuestra lista inicial de 1749 genes, y se ha asociado a esquizofrenia y trastorno bipolar en muestras europeas y asiáticas de más de 700 casos y 1.000 controles (Kahler et al., 2008; Li et al., 2011), con tres SNPs asociados de

manera significativa a esquizofrenia según los meta-análisis realizados en Schizophrenia Gene.

Además de esto, nos parece interesante destacar aquí que en nuestro estudio de SNPs no sinónimos, los resultados más destacados implican a dos genes para los que también existe evidencia de haber sido objeto de selección natural reciente: *SLC39A8* y *AHII*.

Con respecto a *SLC39A8*, en un estudio reciente en el que analizaron la estructura poblacional de Islandia (Price et al., 2009) identificaron ocho polimorfismos cuya frecuencia en población islandesa era significativamente diferente de la frecuencia en población escocesa ($p < 10^{-8}$) y sugieren que esta diferencia, al menos para alguno de ellos, podría ser debida a la acción de la selección natural regional. Además de estos ocho polimorfismos, los autores sugieren un polimorfismo adicional, el SNP rs13107325 de *SLC39A8*, asociado a esquizofrenia en nuestro estudio de SNPs no sinónimos comunes, como candidato de haber sufrido selección natural, a pesar de que la diferencia de frecuencia no llega a ser significativa (valor $p = 2 \times 10^{-7}$; $p = 0,06$ tras corrección por 284.191 pruebas). Según datos actualizados de HapMap, el rs13107325 es un SNP cuyo alelo derivado está presente en población europea a una frecuencia en torno al 10% y está presente, aunque es casi inexistente, en poblaciones africanas y asiáticas. En nuestro trabajo de SNPs no sinónimos comunes analizamos esta variante en una serie de poblaciones europeas. La frecuencia de este alelo se sitúa entre 4,4 y 9,9% en todas las muestras salvo en países nórdicos, en los que es considerablemente inferior, por debajo de 1,6%. Estos datos muestran que la frecuencia del polimorfismo es claramente superior en población europea que en africana o asiática y, dentro de Europa, bastante inferior en población islandesa o finlandesa.

Esta diferencia de frecuencia podría ser, tal y como sugieren Price y colaboradores, el resultado de la acción de la selección natural reciente a nivel regional. Ésta habría incrementado la frecuencia del alelo menor/derivado en determinadas poblaciones europeas. Otra explicación podría ser la siguiente. Dado que en poblaciones de tamaño efectivo pequeño, como son las poblaciones islandesa o finlandesa, la selección natural no es demasiado eficaz y la evolución es sobre todo por deriva genética, aunque el evento selectivo actuase en todas las poblaciones europeas, aquellas se verían afectadas en menor medida por él.

Sea como sea, si es cierto que la selección es la responsable de la diferencia de frecuencia observada, sería un claro ejemplo de antagonismo pleiotrópico ya que, según el resultado de nuestro estudio, el alelo derivado del SNP no sinónimo incrementa el riesgo a padecer esquizofrenia en población europea.

En cuanto a *AHII*, gen necesario para el desarrollo cerebelar y cortical (Ferland et al., 2004) y asociado firmemente a esquizofrenia (se han descrito asociaciones con varios SNPs y haplotipos; menor valor $p = 10^{-10}$; Ingason et al., 2010), hay múltiples evidencias que indican que ha sido objeto de selección natural reciente (Barreiro et al., 2008; Ferland et al., 2004; Torri et al., 2010; Yu et al., 2009). Ferland y colaboradores

fueron los primeros en detectar selección sobre este gen (Ferland et al., 2004) al comprobar que la tasa de sustituciones no sinónimas por sitio no sinónimo frente a sinónimas por sitio sinónimo (K_A/K_S) para *AHII* era de 1,26 en humanos, y significativamente superior a la de otros primates analizados, lo que indica una evolución acelerada específica del linaje humano. En el estudio de Barreiro y colaboradores, 12 SNPs no sinónimos de este gen presentan valores de F_{st} extremos (Barreiro et al., 2008). Los trabajos de Yu y colaboradores y Torri y colaboradores aportan aún más evidencia de selección (Torri et al., 2010; Yu et al., 2009).

De modo que todas estas observaciones parecen confirmar la hipótesis con la que hemos enfocado nuestros cuatro primeros trabajos: eventos de selección natural reciente han participado en la evolución de variantes de susceptibilidad a padecer esquizofrenia.

La estrategia empleada en ellos para localizar regiones sometidas a selección natural reciente ha sido la prueba LRH (Sabeti et al., 2002), que como se ha comentado, permite identificar barridos selectivos incompletos (Sabeti et al., 2006). Existen otros métodos diseñados para la detección de estos barridos, por ejemplo el Integrated Haplotype Store test (*iHS*), en el que se mide el valor de desequilibrio de ligamiento de SNPs individuales, en lugar de haplotipos (Voight et al., 2006). Dado que el efecto de un SNP puede verse incrementado en determinado contexto genético (Camps et al., 2007), para nosotros la prueba LRH tiene la ventaja de que, al analizar haplotipos, permite identificar posibles combinaciones de variantes que puedan resultar o resultaron ventajosas (selección basada en haplotipos), no sólo SNPs individuales. De hecho, resultados experimentales sugieren que la existencia de mayor desequilibrio de ligamiento y conservación de bloques haplotípicos en regiones génicas frente a intergénicas es debida a la acción de la selección purificadora, que eliminaría combinaciones de SNPs deletéreas (Guryev et al., 2006).

Debemos tener presente que las pruebas que pretenden identificar regiones sometidas a selección natural reciente no están exentas de errores, por lo que una proporción de las señales identificadas pueden corresponder a falsos positivos y muchas de las regiones descartadas pueden, en realidad, haber sufrido selección natural y ser falsos negativos (Kelley et al., 2006; Sabeti et al., 2006; Voight et al., 2006). Por ejemplo, la acción de la selección sobre el locus “Duffy” está bien establecida (Hamblin et al., 2002), pero la mayor parte de los test de selección no identifican ninguna señal significativa en la región (Sabeti et al., 2006). Por otra parte, señales que antes de la aparición de este tipo de pruebas empíricas (es decir, pruebas basadas en datos genómicos y no en simulaciones de la historia evolutiva de poblaciones humanas) se consideraron indicadoras de selección, ahora se ve que no destacan entre la diversidad del genoma, como es el caso de *CCR5-Δ32* (Sabeti et al., 2005). De todos modos, dado que desconocemos qué porcentaje de nuestro genoma está o ha estado sometido a selección natural, nuestra capacidad para establecer un umbral adecuado con el que determinar la probabilidad de que una de estas señales refleje un evento selectivo real es limitada. Por tanto, la observación de una de estas

señales por sí misma no es suficiente para confirmar un evento selectivo y es necesario aportar alguna prueba adicional que respalde la funcionalidad de la variante. La asociación de una región con señal de selección a un fenotipo concreto sería una prueba importante de funcionalidad de la variante (Sabeti et al., 2006).

Hay más limitaciones que pueden haber afectado a los resultados de estos cuatro trabajos. Dado que se desconoce el sustrato biológico de la enfermedad, la capacidad para seleccionar buenos genes candidatos funcionales es limitada. Además, dada la evidencia del carácter altamente poligénico del trastorno, probablemente genes con contribución causal importante para su desarrollo no hayan sido incluidos en ninguno de estos cuatro trabajos.

Asimismo, el tamaño de la muestra empleada puede no haber sido suficiente para identificar asociaciones débiles.

A pesar de estas limitaciones, el empleo de esta estrategia ha permitido establecer evidencia de asociación entre la enfermedad y dos haplotipos sometidos a selección natural reciente en los genes *MAOB* y *PTPN13*, y un haplotipo protector en *DRD3* que probablemente se haya visto afectado por la acción de la selección sobre una variante en desequilibrio de ligamiento. Esto, unido a las ya comentadas aportaciones que relacionan genes asociados a la enfermedad con eventos selectivos, como son los casos de *SLC39A8*, *AH11*, *ODZ4*, o la región MHC, valida nuestra estrategia de trabajo.

Por tanto, creemos que el empleo de esta estrategia evolutiva de manera sistemática –esto es, la consideración de cualquier gen en el estudio, la combinación de diferentes pruebas de selección, la reducción del umbral de las pruebas de selección para considerar regiones en el posterior estudio de asociación, el análisis de variantes en el entorno de las señales de asociación identificadas y el empleo de muestras de tamaño suficiente para poder analizar un número elevado de loci– podría dar lugar a la identificación de nuevas variantes funcionales comunes de susceptibilidad en genes importantes para la etiología de la enfermedad.

Siguiendo con nuestro objetivo central de analizar variantes con elevada probabilidad de ser funcionales, en la segunda etapa de esta tesis decidimos explorar sistemáticamente el papel de SNPs no sinónimos en la etiología de la enfermedad. Este tipo de polimorfismos están localizados en exones e implican un cambio de aminoácido en las proteínas, por lo que presentan elevada probabilidad *a priori* de ser funcionales y de estar implicados en la aparición de enfermedades (Hindorff et al., 2009). Restringir las variantes analizadas a esta clase de polimorfismos podría enriquecerlo en variantes funcionales.

Para ello empleamos el panel de Affymetrix 20K (Affymetrix GeneChip Human 20k cSNP kit), que incluye alrededor de 14.000 SNPs no sinónimos distribuidos a lo largo del genoma. Entre ellos se incluyen casi todos los SNPs no

sinónimos validados en la versión 124 de dbSNP que se pudieron genotipar con éxito con esta tecnología.

En una primera fase de análisis centramos nuestra atención en variantes comunes (frecuencia mayor del 5% en nuestra muestra) mediante una prueba de asociación χ^2 y un análisis de agregación basado en perfiles de riesgo para la eventual identificación del componente poligénico en nuestra muestra. Posteriormente analizamos las variantes menos frecuentes que resultaron de la genotipación del panel 20K, así como de SNPs no sinónimos adicionales en genes candidatos, mediante métodos diseñados específicamente para el análisis de variantes raras (frecuencia menor del 5% en casos o controles en nuestra muestra).

Análisis de SNPs no sinónimos comunes

En los últimos años la búsqueda de variantes comunes de susceptibilidad se ha visto revolucionada con la aparición de los paneles de genotipado de genoma completo (Hindorff et al., 2009). Estos paneles permiten analizar simultáneamente millones de polimorfismos distribuidos a lo largo de todo el genoma de manera que se obtiene información de la mayor parte de la variabilidad genética común. Esto hace que el estudio no dependa de una hipótesis previa acerca de la etiología de la enfermedad, más allá de la heredabilidad de la misma; son los propios resultados del estudio los generadores de nuevas hipótesis. Como se ha descrito en la introducción de esta tesis, la aplicación de esta estrategia al estudio genético de la esquizofrenia ha permitido identificar variantes comunes de manera consistente y ha revelado, además, la existencia del componente poligénico.

A pesar de este importante avance, dado que la estrategia se basa en el mapeo por desequilibrio de ligamiento y a que muchas de las señales de asociación se localizan en regiones no codificantes de funcionalidad todavía desconocida, en ocasiones resulta complicado establecer de manera clara las variantes causales responsables de todas las señales de asociación (Cirulli y Goldstein, 2010).

El análisis de los resultados de los estudios del genoma completo en enfermedades complejas revela que los SNPs no sinónimos son la clase funcional con mayor probabilidad de estar asociada con la enfermedad (OR = 3,9; 95% IC, 2,2 – 7,0; $p = 3,5 \times 10^{-7}$) (Hindorff et al., 2009).

Tomando esto en consideración y dada la cantidad de variantes comunes que podrían estar implicadas en la etiología de la enfermedad y que cada estudio tiene potencia para identificar un número limitado de variantes de riesgo, restringir el estudio a SNPs no sinónimos podría ayudar a focalizar la atención en variantes funcionales en genes concretos, cuya interpretación funcional y biológica podría ser más inmediata, además de ser económicamente más viable. Es por ello que nos pareció

interesante emplear un panel de genotipado que incluyese exclusivamente SNPs no sinónimos, en concreto el panel de Affymetrix 20K. Dado que el número de SNPs incluidos es muy inferior al de otros paneles empleados para hacer estudios del genoma completo, su uso nos permitiría analizar variantes codificantes distribuidas a lo largo de todo el genoma y reducir de manera significativa el número de polimorfismos analizados con menos probabilidad *a priori* de participar en la etiología de la enfermedad, aumentando así la potencia del estudio.

La estrategia seguida consistió en el análisis de SNPs no sinónimos de manera individual en una muestra inicial gallega seguido de un análisis de los polimorfismos con resultado más interesante en muestras adicionales independientes. Además, realizamos un análisis de todos los SNPs no sinónimos como grupo (multilocus) para hacer “perfiles de riesgo” e investigar el componente poligénico en nuestra muestra.

Análisis individual

El análisis individual de SNPs no sinónimos comunes dio lugar a la identificación de una variante en el transportador de iones metálicos *SLC39A8* asociada de manera consistente a un incremento en el riesgo a padecer esquizofrenia en población europea. En la primera etapa del análisis estadístico, evaluamos 5.100 SNPs no sinónimos en una muestra de 476 pacientes esquizofrénicos y 447 controles gallegos (SNPs y muestras que superaron los filtros de calidad) mediante una prueba de asociación de tendencia (Cochran–Armitage). No encontramos evidencias de estratificación (λ sin corregir = 1,0348; λ corregido por los cuatro componentes principales de ancestralidad = 1,0343; figura 1 del artículo). A continuación seleccionamos los polimorfismos con mayor evidencia de estar asociados a la enfermedad –18 polimorfismos– y los analizamos en 4.069 casos y 15.128 controles adicionales procedentes de una serie de muestras independientes de origen europeo. Muchas de estas muestras se emplearon en estudios previos de genoma completo en esquizofrenia, incluyendo los publicados en el 2009 ((ISC, 2009; Shi et al., 2009; Stefansson et al., 2009) y el último mega meta-análisis llevado a cabo por el *Psychiatric GWAs Consortium* (Ripke et al., 2011). El meta-análisis de todas las muestras (4.545 casos y 15.575 controles) dio lugar a la identificación de un polimorfismo fuertemente asociado a la enfermedad, el SNP rs13107325 en el gen *SLC39A8* ($p = 2,7 \times 10^{-6}$). El índice de heterogeneidad calculado para este polimorfismo fue de $I^2=16,7\%$, lo que indica baja heterogeneidad entre las muestras de las distintas poblaciones analizadas (Higgins y Thompson, 2002). La señal de asociación para este SNP en la población inicial gallega era de $p = 3,2 \times 10^{-4}$, por lo que se incrementó en dos órdenes de magnitud con el meta-análisis. El valor de la asociación final tras corrección por Bonferroni es de $p = 0,0048$, corrección conservativa ya que, como indicamos en el estudio 5, muchos de los polimorfismos analizados están en desequilibrio de ligamiento. A pesar de que para los estudios convencionales de genoma completo – en los que se interrogan cientos de miles de variantes – se haya establecido un valor de significación menor de 5×10^{-8} para

declarar significativa una asociación, en nuestro estudio este umbral sería excesivo y carente de sentido, ya que únicamente analizamos 5100 variantes.

Como hemos comentado en esta discusión, existen indicios de que el SNP rs13107325 ha sufrido selección natural en población europea (Price et al., 2009). Esta observación y la asociación encontrada en nuestro estudio, sugieren que se trata de una variante funcional y podría ser interesante hacer estudios funcionales con ella.

Tal y como indicamos en la discusión del estudio 5, otros genes de la misma familia –SLC39– se asociaron con otros trastornos psiquiátricos, lo que podría significar que exista alguna relación entre las vías metabólicas en las que participa esta familia de genes y estos trastornos. Estos genes son transportadores de iones metálicos que actúan, entre otros lugares, en la placenta y la barrera hemato–encefálica, por lo que es posible que la alteración de la homeostasia de los metales en el cerebro, durante el desarrollo embrionario y/o durante la vida de un individuo incrementa el riesgo a desarrollar trastornos psiquiátricos, abriendo una nueva línea de investigación.

Componente poligénico

Tal y como se describe en la introducción de esta tesis, una parte del riesgo a padecer la enfermedad viene dado por la acumulación en un mismo individuo de múltiples de variantes comunes de riesgo muy bajo, lo que se denomina el componente poligénico (ISC, 2009; Ripke et al., 2011). Es decir, que polimorfismos que individualmente no predisponen de manera clara a padecer la enfermedad, sí contribuyen de manera colectiva a incrementar el riesgo de una persona a padecerla.

El empleo de paneles de genotipado con cientos de SNPs permite identificarlos (ISC, 2009; Ripke et al., 2011): en una primera muestra de casos y controles (“training set”) se hace una prueba de asociación y se ordenan los polimorfismos de menor a mayor valor p ; se establece entonces un umbral de significación (uno o varios, por ejemplo $p < 0,5$, $<0,1$, etc.), y todas las variantes que hayan obtenido un valor p inferior al umbral fijado se emplean para evaluar el “riesgo genético” en individuos de una segunda muestra independiente (“testing set”). El riesgo genético –o perfil de riesgo– de cada individuo se mide cuantificando el número de alelos de riesgo (ponderados o no) presentes en cada uno; finalmente se evalúa si los individuos del grupo de casos acumulan mayor riesgo que los controles, es decir, si son portadores de un número significativamente mayor de alelos de riesgo. El riesgo que aporta cada variante puede ponderarse, por ejemplo, por su OR estimado en el “training set”. Una forma de evaluar la eficacia de estos “perfiles de riesgo” en cuanto a su capacidad para identificar casos o controles es cuantificar la proporción de varianza explicada; esto se puede hacer mediante un análisis de regresión logística del estatus caso–control y los perfiles de riesgo estimados.

Empleando este tipo de estrategia basada en perfiles de riesgo, el Consorcio Internacional de Esquizofrenia (ISC) aportó en el año 2009 la primera evidencia experimental de la existencia de este sustrato poligénico al observar que SNPs con un

efecto muy bajo (diferentes umbrales: $p < 0,1; 0,2; 0,3; 0,4; 0,5$) identificados en una primera muestra de pacientes esquizofrénicos estaban significativamente sobrerrepresentados en otras 3 muestras independientes de pacientes esquizofrénicos y en dos muestras independientes de individuos con trastorno bipolar, no encontrándose, sin embargo, sobrerrepresentación de estos alelos en enfermedades no psiquiátricas (diabetes tipo 1, hipertensión, enfermedad de Crohn, artritis reumatoide y enfermedad coronaria arterial) (ISC, 2009). Este resultado es, además, una prueba adicional de la vulnerabilidad genética compartida entre diferentes entidades diagnósticas psiquiátricas. El porcentaje máximo de varianza explicada que se alcanzó en este estudio fue de 3,2% (muestra: MGS-EA; umbral significación: $p < 0,5$). Este porcentaje se supone que es mayor (es decir, que las variantes de riesgo explicarían un porcentaje mayor de la varianza), ya que la mayor parte de los SNPs no son las variantes causales, sino variantes en desequilibrio de ligamiento.

Inicialmente algunos autores se mostraron escépticos y plantearon la posibilidad de que esta observación fuese en realidad un artefacto producto de la existencia de estratificación críptica en la muestra (McClellan y King, 2010). Para tratar de aclarar la situación, Ruderfer y colaboradores (Ruderfer et al., 2011) emplearon una estrategia basada en familias –estrategia que no se ve influida por la presencia de estratificación– para determinar si los alelos de riesgo identificados en la muestra inicial del trabajo del ISC estaban sobre–transmitidos de padres a hijos esquizofrénicos en una muestra independiente. Así pues emplearon la muestra inicial del ISC como “training set” (2.794 casos y 2.976 controles) y 694 tríos independientes de Bulgaria como “testing set”. Los resultados de este análisis mostraron que la tasa de transmisión de los alelos de riesgo era mayor que la tasa de no transmisión (valor mínimo de la prueba de asociación de transmisión: $p = 1,03 \times 10^{-12}$; umbral de significación: $p < 0,5$; varianza explicada: $> 4,5\%$), con lo que se demostró que el componente poligénico no es un artefacto debido a estratificación poblacional.

El *Psychiatric Genome-Wide Association Study Consortium* reevaluó la existencia de este componente añadiendo muestras nuevas a las ya analizadas en el trabajo del ISC y confirmaron la observación inicial (Ripke et al., 2011). La varianza explicada alcanzó en este caso el 5,8% (umbral de significación: $p < 0,3, 0,4$ y $0,5$; (Ripke et al., 2011).

Por otra parte, en un trabajo más modesto en cuanto al tamaño de muestra, Ikeda y colaboradores (Ikeda et al., 2011) confirmaron la presencia de este componente en población japonesa, y evidenciaron que muchas de las variantes que constituyen este componente son las mismas en diferentes continentes al ver que, empleando una muestra de población europea como “training set” y una muestra de población japonesa como “testing set” (y viceversa), el grupo de pacientes presentaba perfiles de riesgo significativamente mayores que el de controles.

Para investigar la presencia de este componente poligénico en nuestra muestra de pacientes gallegos en base al panel de genotipado que empleamos, decidimos aplicar una estrategia basada en validaciones cruzadas. Para ello dividimos la muestra

en diez partes y empleamos secuencialmente 9/10 como “training set” y 1/10 como “testing set” (figura 2).

Este análisis se hizo considerando el grupo de SNPs no sinónimos autosómicos incluidos en el panel, así como un subgrupo del mismo restringido a los polimorfismos localizados en genes candidatos funcionales, a saber, todos los genes presentes en la base de datos Schizophrenia Gene y genes de la base de datos Gene Ontology bajo las categorías “transmisión sináptica” y / o “neurodesarrollo”.

La estrategia seguida permitió confirmar la presencia del componente poligénico en nuestra muestra: para umbrales de significación de $p < 0,05$ y $p < 0,1$, el valor de la prueba para la acumulación de variantes fue de $p = 0,0068$ y $p = 0,0002$, respectivamente. En cuanto a la proporción de varianza explicada, el máximo valor alcanzado fue de 2,1% (umbral de significación: $p < 0,1$).

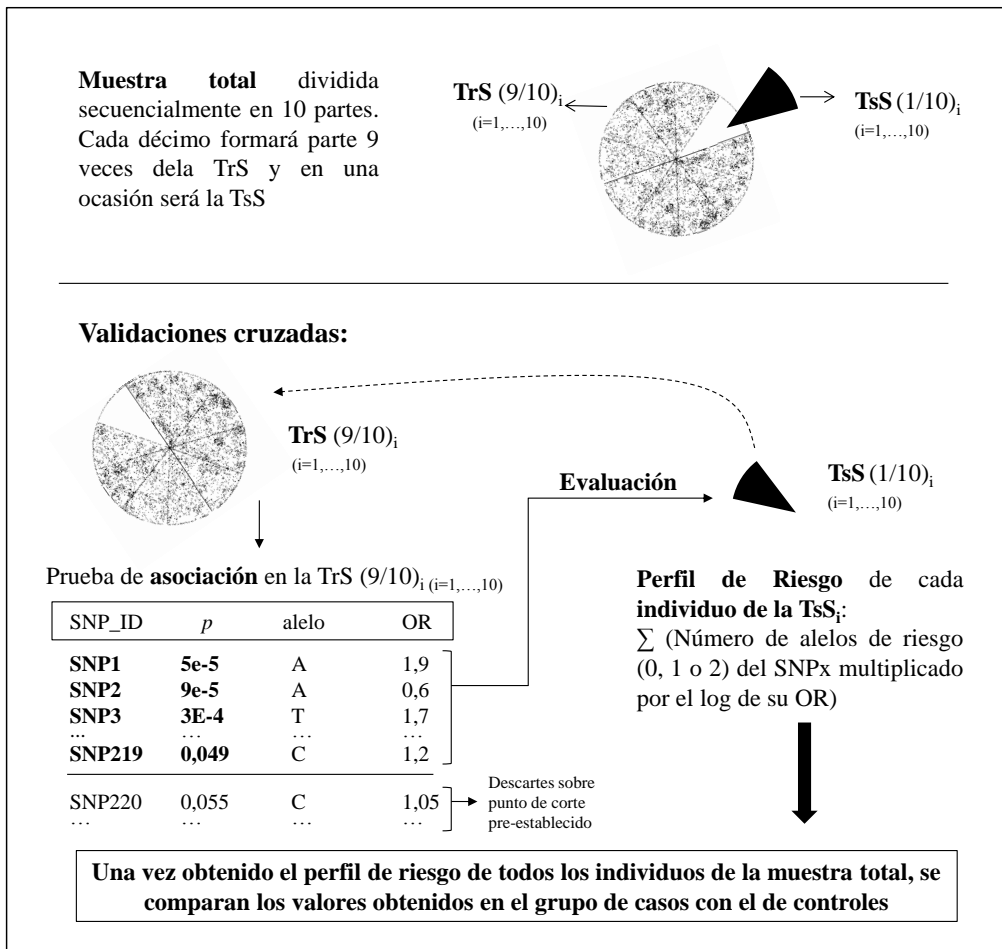


Figura 2. Estrategia basada en validaciones cruzadas para la identificación del componente poligénico en nuestra muestra. TrS = Training Set; TsS = Testing Set

La restricción del análisis a polimorfismos en genes candidatos funcionales no mejoró el resultado, lo que sugiere que un número importante de genes de susceptibilidad son insospechados y por el momento desconocidos.

En resumen, nuestro estudio de SNPs no sinónimos comunes a lo largo del genoma ha permitido confirmar la presencia del componente poligénico en las muestras de pacientes esquizofrénicos gallegos (aun empleando un número mucho menor de SNPs que el empleado habitualmente en estudios de genoma completo, 5.100 frente a varios cientos de miles) y la identificación de un SNP fuertemente asociado a la enfermedad en población europea en un nuevo gen de susceptibilidad, *SLC39A8*, lanzando una nueva hipótesis sobre posibles mecanismos que podrían relacionarse con la aparición de la enfermedad, en concreto, la alteración de la homeostasia de iones metálicos durante el desarrollo embrionario o durante el funcionamiento del cerebro.

En esquizofrenia, los estudios de genoma completo en muestras de tamaño suficiente y la combinación de diferentes muestras para hacer meta-análisis potentes han permitido identificar más de 10 loci de susceptibilidad con significación genómica, muchos de ellos de vulnerabilidad también a padecer trastorno bipolar (tabla 1, página 27). Sin embargo, teniendo en cuenta el sustrato altamente poligénico de la enfermedad y que cada estudio tiene potencia para identificar un número limitado de variantes de susceptibilidad, es probable que existan muchas más y que una proporción de las variantes ya analizadas sean realmente variantes implicadas en la etiología de la enfermedad, aunque no hayan superado el estricto umbral de asociación establecido. Cálculos teóricos estiman que el incremento en cuatro veces del tamaño actual de las muestras empleadas permitiría identificar entre 30 y 60 nuevos loci de susceptibilidad a padecer esquizofrenia y trastorno bipolar (Sullivan, 2011). Por este motivo, para seguir avanzando en la identificación de nuevas variantes comunes de susceptibilidad, en este momento es importante la recolección de nuevos pacientes sobre los que hacer estudios de genoma completo (nuestro trabajo es una prueba de esto). Por otra parte, los estudios de genoma completo son una herramienta eficaz para la detección de CNVs, por lo que permiten identificar este tipo de variantes raras en casos concretos.

Además, dado que los genes identificados en estos estudios participan claramente en la etiología del trastorno, es posible que existan otras variantes en su secuencia, independientes de las ya identificadas, cuyo efecto incrementa también el riesgo a padecerlo. Estas nuevas variantes podrían ser comunes o raras, en cuyo caso el efecto podría ser mayor, por lo que el estudio de estas regiones mediante resecuenciación en muestras de tamaño suficiente podría dar lugar a su identificación (Rivas et al., 2011).

A medida que se vaya incrementando el número de loci de susceptibilidad asociados de manera consistente a la enfermedad y se vayan localizando las variantes

causales responsables de las asociaciones, paralelo a un avance en la anotación de regiones funcionales (Lindblad-Toh et al., 2011), se podrá avanzar hacia un mejor conocimiento de la enfermedad, con la localización de vías metabólicas o procesos biológicos relacionados con su aparición. Para todo esto es fundamental la colaboración internacional entre todos los grupos de investigación y la continuidad de este tipo de estrategia, como se ha hecho hasta el momento, unida a un riguroso control de la calidad de los resultados (estratificación, porcentaje de genotipado en muestras y SNPs, etc.)

Análisis de SNPs no sinónimos poco frecuentes

A pesar de que los esfuerzos por modelar el riesgo de las enfermedades complejas se han centrado principalmente en el análisis de variantes comunes, en base a la hipótesis enfermedad común / variante común, cada vez hay más evidencias de que variantes menos frecuentes (< 5%) de efecto mayor podrían ser importantes en la etiología de estas enfermedades (Gorlov et al., 2010; Keller y Miller, 2006; Pritchard, 2001).

Además de mutaciones raras que individualmente incrementen el riesgo a padecer estas enfermedades, recientemente se ha sugerido que la acumulación de mutaciones de baja frecuencia moderadamente deletéreas en genomas individuales podría tener un efecto global dañino (Kryukov et al., 2007), en base a un modelo poligénico de mutación-selección. Datos experimentales parecen dar sustento a esta propuesta: se ha descrito que el grupo de SNPs de frecuencia menor al 6% está enriquecido en variantes funcionales (Wong et al., 2003). Li y colaboradores (Li et al., 2010) secuenciaron la regiones exónicas de 200 individuos procedentes de Dinamarca y vieron un exceso de mutaciones no sinónimas sobre sinónimas de 1,8 veces en el rango de variantes a baja frecuencia (entre 2 y 5%) y que en general las sustituciones en regiones muy conservadas tienen una menor frecuencia que mutaciones en regiones no conservadas, lo que indica la acción de la selección purificadora y sugiere que una proporción importante de variantes no sinónimas sean muy probablemente deletéreas. Este trabajo indica que la mayor parte de la variación que afecta a enfermedades con bajo éxito reproductivo corresponde a variantes de baja frecuencia. Por otra parte, se ha estimado que la mayor parte de las mutaciones que surgen en regiones codificantes y que dan lugar a un cambio no sinónimo son sólo moderadamente deletéreas (Kryukov et al., 2007), es decir, que no son lo suficientemente dañinas como para ser eliminadas completamente por selección purificadora y puedan llegar a alcanzar cierta frecuencia, llegando a ser variantes de baja frecuencia en la población y contribuir a la aparición de enfermedades complejas (Gorlov et al., 2008).

En el caso de la esquizofrenia, esta hipótesis es plausible: el sustrato altamente poligénico es compatible con un modelo poligénico de selección-mutación; por otra parte, es una enfermedad que aparece temprano en la edad reproductiva y está asociada a una tasa de reproducción reducida y a un incremento en la mortalidad (Laursen y Munk-Olsen, 2010), por lo que parece lógico que, en general, variantes con un efecto algo mayor sobre el fenotipo se vean afectadas por la presión negativa que existe sobre el fenotipo esquizofrénico y estén limitadas a una frecuencia baja en la población, o que sean incluso eliminadas si el efecto es muy grande (Cirulli y Goldstein, 2010; Gorlov et al., 2010). Hasta el momento esta hipótesis de acumulación de variantes raras de susceptibilidad no se ha evaluado de manera sistemática, ya que los estudios de genoma completo en general suelen excluir todas las variantes de frecuencia baja (<5%) y los estudios de secuenciación hasta el momento se han restringido al análisis de uno o unos pocos genes en cada estudio. Por estas razones nos pareció que podría ser interesante analizar los SNPs no sinónimos de baja frecuencia que resultaron de la genotipación del panel de Affymetrix en nuestra muestra. Incluimos, además, 129 SNPs no sinónimos de baja frecuencia, localizados en genes candidatos funcionales y con elevada probabilidad de ser funcionales según PolyPhen (tabla suplementaria 1 del estudio 6).

La estrategia de análisis que seguimos fue la evaluación del riesgo asociado a la acumulación de alelos raros de SNPs no sinónimos en genes individuales y en diferentes grupos de genes –genes candidatos funcionales en general y genes implicados en una misma vía funcional o interactoma concreto (tabla suplementaria 2 y figura suplementaria 1 del estudio 6).

Para el análisis de genes individuales empleamos el programa CCRaVAT (Lawrence et al., 2010), que permite identificar genes concretos con una proporción significativamente mayor de alelos raros en el grupo de casos o de controles (frecuencia del alelo menor < 5%). La prueba estadística empleada para el exceso de variantes raras en casos fue una prueba de Fisher de una cola.

Para el análisis de grupos de genes empleamos el programa VTtest (Price et al., 2010). Este programa realiza varios análisis para la evaluación de la acumulación de alelos raros en casos frente a controles, atendiendo tanto a umbrales de frecuencia como al peso atribuido a cada alelo raro. En particular, empleamos la posibilidad que proporciona el programa de utilizar un umbral de frecuencia variable, que pretende determinar un umbral desconocido por debajo del cual la probabilidad de que las variantes sean de susceptibilidad se incrementa significativamente en comparación con las variantes que superan ese umbral. Asimismo, ponderamos alternativamente el efecto de cada alelo por funcionalidad, en base a las predicciones de PolyPhen (Adzhubei et al., 2010). Para cada persona analizada el programa genera una puntuación basada en el número de alelos raros que alberga. El resultado final de la prueba es la diferencia entre la suma de las puntuaciones de cada persona del grupo de casos y la suma de puntuaciones de los individuos del grupo de controles. El nivel de significación de la prueba se calculó mediante 10.000 permutaciones.

En relación al análisis de genes individuales, encontramos una acumulación de variantes de baja frecuencia significativamente mayor en casos que en controles en ninguno de los genes analizados tras corrección por pruebas múltiples. A pesar de esto, nos parece particularmente interesante el resultado obtenido para uno de ellos, el gen *AH11*. De las variantes analizadas en este trabajo, sólo una se localiza en este gen, el SNP rs6940875; el alelo menor de esta variante está presente en cinco de los pacientes analizados y en ningún control; como dijimos, el resultado de la prueba de asociación no resiste la corrección por pruebas múltiples (prueba de Fisher de una cola sin corregir: $p = 0,036$); sin embargo, dado que nuestra muestra no tenía potencia suficiente para analizar variantes individuales muy raras, el análisis no permite ni aceptar ni rechazar la asociación, y existen una serie de evidencias que sugieren que la variante podría ser relevante para la enfermedad. Como se ha comentado anteriormente, este gen está asociado de manera consistente a esquizofrenia (Ingason et al., 2010; Rivero et al., 2010) y a autismo (Alvarez Retuerto et al., 2008). Se han descrito además mutaciones en el gen que dan lugar al síndrome de Joubert (Valente et al., 2006), un trastorno raro, de herencia recesiva, caracterizado por malformaciones congénitas en el cerebro, en el que el 25% de los pacientes muestran rasgos del trastorno autista (Ozonoff et al., 1999). Por otra parte, predicciones bioinformáticas para esta variante indican que es muy probable que el cambio sea funcional (“probablemente deletéreo” según PolyPhen <http://genetics.bwh.harvard.edu/pph2/>; (Adzhubei et al., 2010), y “deletéreo” según SIFT -<http://sift.jcvi.org/> (Kumar et al., 2009));. El alelo menor de este SNP es muy raro en europeos y no está presente en ninguna de las poblaciones de HapMap europeas o asiáticas, pero en poblaciones africanas es más común, llegando a alcanzar una frecuencia cercana al 10%. Curiosamente, al consultar la ancestralidad de los cinco pacientes portadores de la variante rara -ancestralidad inferida con el programa STRUCTURE (v2.3.1, (Pritchard et al., 2000); <http://pritch.bsd.uchicago.edu/structure.html>) a partir del análisis de 130 SNPs marcadores de ancestralidad presentes en el panel de Affymetrix 20k y genotipados en las poblaciones de HapMap- observamos que uno de ellos tiene una probabilidad de tener ancestralidad africana del 1,5%, probabilidad ligeramente superior a la del resto de las muestras analizadas; para las otras cuatro muestras la probabilidad de ancestralidad africana se sitúa entre el 0,2 y 0,7%, similar al del resto de las muestras y similar a su probabilidad de tener ancestralidad asiática. Como en el caso comentado del SNP de *SLC39A8*, es posible que la diferencia de frecuencia del alelo menor de esta variante sea el resultado de la acción de la selección natural, en cuyo caso su funcionalidad sería clara.

Así pues, teniendo en cuenta la implicación confirmada de este gen en la etiología de diferentes enfermedades mentales y la elevada probabilidad de que el SNP analizado, rs6940875, sea una variante funcional, es posible que este SNP tenga alguna repercusión en el riesgo a padecer esquizofrenia; su estudio en muestras más grandes, o en población africana, permitirá confirmar o descartar su implicación.

Con respecto al análisis de SNPs no sinónimos en grupos de genes, no encontramos acumulación significativa de variantes en casos para ninguno de los

grupos analizados. Tal y como comentamos en la discusión del estudio 6, creemos que el estudio pudo verse afectado por algún tipo de sesgo. Por ejemplo, sesgo relacionado con las variantes incluidas en el panel de genotipado: estas variantes proceden de la versión 124 de la base de datos dbSNP, del año 2005; en aquel momento, la mayor parte de las variantes catalogadas fueron identificadas en personas de población general, por lo que es muy posible que el conjunto de variantes incluidas en el panel 20K esté enriquecido en variantes neutras, y exista una baja representación de variantes de riesgo de baja frecuencia, presentes sobre todo en pacientes.

Por otra parte, no podemos descartar que el resultado se deba a un fallo de la propia estrategia, ya que a pesar de que el estudio de la acumulación de variantes raras se sustenta en evidencias experimentales (Li et al., 2010) y tiene mayor potencia que el análisis individual de este tipo de variantes, también presenta una serie de limitaciones (Basu y Pan, 2011; Li y Leal, 2008; Morris y Zeggini, 2009). Por ejemplo, si la mayoría de las variantes raras analizadas no tienen relación con la enfermedad, al combinarlas con variantes causales se introduce “ruido” y se reduce la potencia del estudio. En este sentido, a pesar de haber restringido el análisis a variantes localizadas en genes candidatos funcionales, es posible que no hayamos conseguido un enriquecimiento en variantes causales.

Además, podría ser que entre las posibles variantes relacionadas con la diferente susceptibilidad a padecer esquizofrenia, algunas de ellas predispongan y otras que protejan (una variante ventajosa o de protección frente a un fenotipo concreto podría mantenerse a baja frecuencia en la población si además ejerce un efecto pleiotrópico antagónico fuerte sobre otro rasgo fenotípico, de modo que la selección purificadora impida que prospere en la población). Así pues, si existiesen variantes raras de protección en nuestra muestra, al analizarlas en conjunto con variantes de riesgo las señales de asociación podrían disminuir o neutralizarse.

A pesar de que nuestro estudio no aporta nuevas evidencias, cada vez van apareciendo más pruebas que apoyan la implicación de variantes poco frecuentes en la etiología de la esquizofrenia, como la traslocación de *DISC1*, deleciones en 22q11 –el mayor factor de riesgo genético descrito hasta la fecha–, y otras CNVs raras y largas (> 100 Kb) que se presentan con mayor frecuencia en casos frente a controles (Gejman et al., 2011). Aunque el riesgo asociado a estas variantes es mayor que el de las variantes comunes, éstas explican un porcentaje muy pequeño de la heredabilidad total estimada debido a su baja frecuencia (< 0,2% las CNVs Visscher et al., 2011). Además, el número de variantes descritas es bajo por el momento, debido probablemente a un número insuficiente de muestras analizadas, así como a una estrategia poco eficaz para su identificación. Por ejemplo, hasta hace poco la búsqueda de variantes de riesgo poco frecuentes que sólo se pueden localizar mediante secuenciación, como deleciones o duplicaciones de unos pocos nucleótidos o mutaciones puntuales, se restringía a la secuenciación de genes candidatos en un número pequeño de individuos; teniendo en cuenta el sustrato poligénico de la enfermedad, la probabilidad de que las personas analizadas en cada estudio tengan mutaciones precisamente en el gen o genes analizados es mínima, por lo que la

estrategia resulta poco eficaz. Esto hace que por el momento no se sepa hasta qué punto contribuyen este tipo de variantes a la etiología de la enfermedad.

Esta situación está cambiando gracias a la aparición y a los avances de la resecuenciación a gran escala o ultra-secuenciación (Cirulli y Goldstein, 2010).

Hasta el momento se han publicado dos trabajos de ultra-secuenciación de exoma completo en esquizofrenia. Ambos estudios se diseñaron en base a la misma hipótesis, que mutaciones exónicas *de novo* juegan un papel destacado en la etiología de la enfermedad. Para contrastar esta hipótesis, Girard y colaboradores (Girard et al., 2011) analizaron el exoma de 14 tríos (formados por un paciente esquizofrénico esporádico y sus padres sanos); identificaron 15 mutaciones *de novo* en 8 casos y sugirieron que los casos esporádicos de esquizofrenia tienen una mayor probabilidad de albergar alguna mutación *de novo* codificante y deletérea que los controles. En el segundo trabajo, Xu y colaboradores analizan una muestra mayor, 53 tríos de casos y 22 tríos no afectados. Detectaron 40 mutaciones *de novo* en 27 pacientes y 7 en 7 controles. La diferencia no es significativa; sin embargo, vieron que, en los casos analizados, la tasa de mutaciones *de novo* no sinónimas frente a sinónimas (NS:S) era de 16:1, superior a lo esperado por azar (NS:S de 2,85:1) y que esta misma tasa para variantes heredadas nuevas (no descritas) era significativamente menor (NS:S de 1,61:1; $p = 0.0002$). Los resultados de este mismo análisis en el grupo de controles no se desviaron de lo esperado por azar, con una tasa NS:S para variantes *de novo* de 1,33:1, similar a la tasa NS:S para variantes heredadas (1,69:1; $p = 0.81$). Así pues, los autores sostienen que las variantes identificadas tienen alta probabilidad de relacionarse con la esquizofrenia por lo que los resultados obtenidos confirman su hipótesis inicial y concluyen que más de la mitad de los casos esporádicos de esquizofrenia podrían ser el resultado de la aparición de mutaciones puntuales *de novo* de efecto mayor (Xu et al., 2011).

Sin embargo, tal y como muchos investigadores han comentado, hay que ser cautos con los resultados obtenidos, ya que el tamaño de las muestras empleadas es por el momento muy pequeño, y si algo se ha demostrado en el estudio de enfermedades complejas es que es necesario un tamaño muestral elevado para poder obtener conclusiones de carácter general (comentarios de Daniel Weinberger, Patrick Sullivan y Jin Szatkiewicz sobre el trabajo: Girard SL et al. Increased exonic *de novo* mutation rate in individuals with schizophrenia. *Nat Genet.* 2011 Sep, disponibles en: <http://www.schizophreniaforum.org/pap/annotation.asp?powID=151310#{9491DFBF-FE0C-4A8C-BC15-492BBE55F74E}>, y comentario realizado por Patrick Sullivan y Jin Szatkiewicz sobre el trabajo: Xu B et al. Exome sequencing supports a *de novo* mutational paradigm for schizophrenia. *Nat Genet.* 2011, disponible en: <http://www.schizophreniaforum.org/pap/annotation.asp?powID=152096#>).

Por otra parte, los autores de ambos trabajos asumen una relación de causalidad entre las variantes identificadas y la enfermedad y, dado que están en heterocigosis, asumen además un modelo dominante de herencia para todas ellas. Aunque esto pueda ser cierto, el hecho de encontrar una variante *de novo* en un paciente, aunque se trate

de una mutación no sinónima con alta probabilidad de ser funcional, no indica necesariamente que se relacione con la patología, por lo que sin otro tipo de pruebas funcionales adicionales no se puede confirmar que esto sea así.

Por estos motivos, todavía es pronto para poder establecer la contribución de este tipo de variantes a la arquitectura alélica de la enfermedad. Estos dos trabajos han identificado una serie de variantes exónicas *de novo* en pacientes esporádicos esquizofrénicos, y es posible que muchas de ellas sean relevantes para la enfermedad. A medida que se vayan publicando más trabajos de secuenciación y se vayan acumulando datos de miles de pacientes y controles se podrán confirmar o descartar estas observaciones y se podrá determinar la contribución relativa de variantes de baja frecuencia a la etiología del trastorno de manera más firme.

En cuanto a las perspectivas futuras de la investigación genética en esquizofrenia, dado el carácter poligénico del trastorno, parece que las estrategias con mayor probabilidad de éxito serán aquellas que analicen sistemáticamente variantes a lo largo del genoma de cada individuo, con el objetivo de definir su perfil genético de riesgo. Será interesante también profundizar en el estudio de loci ya identificados mediante investigaciones complementarias, como análisis de expresión, cultivos celulares, estudios de neuroimagen funcional y estructural, modelos animales o estudios de interacción gen-gen, gen-ambiente y epigenéticos. En concreto, entender el modo en que los diferentes factores genéticos y ambientales interaccionan entre sí para dar lugar al fenotipo clínico constituye uno de los mayores retos (Akil et al., 2010).

Conclusiones

1. La estrategia evolutiva diseñada para la búsqueda de variantes comunes de susceptibilidad ha permitido la identificación de 3 genes asociados a esquizofrenia; en concreto, hemos identificado 3 haplotipos sometidos a selección natural reciente en los genes *DRD3*, *MAOB* y *PTPN13*, y hemos comprobado mediante estudios de asociación que los haplotipos ancestrales incrementan el riesgo a padecer esquizofrenia. Estos resultados encajan en las predicciones del modelo de susceptibilidad ancestral propuesto para explicar la existencia de variantes comunes de susceptibilidad a padecer enfermedades complejas.

2. Con respecto a los genes *DAOA*, *PPP1R1B* o *APOLI*, no encontramos señales de selección natural reciente en población europea. Tampoco encontramos evidencia de asociación de las variantes analizadas con esquizofrenia, por lo que concluimos que las variantes analizadas no encajan en el modelo de pleiotropismo antagónico.

3. La búsqueda de señales de selección natural reciente en 1749 genes candidatos muestra que 163 genes contienen señales bastante evidentes de selección, por lo que es muy probable que alberguen variantes funcional y fenotípicamente diferentes; entre ellos se encuentran genes previamente asociados a esquizofrenia y a otros trastornos psiquiátricos. El estudio de estas variantes sometidas a selección natural podría dar lugar a la identificación de variantes funcionales de predisposición a padecer esquizofrenia.

4. A pesar de que las asociaciones identificadas necesitan ser reproducidas en muestras de mayor tamaño, los resultados obtenidos y también la revisión bibliográfica realizada validan nuestra estrategia evolutiva de búsqueda de variantes comunes de susceptibilidad y permiten concluir que la selección natural reciente ha modulado el espectro alélico de genes de susceptibilidad a padecer esquizofrenia.

5. El análisis de SNPs no sinónimos distribuidos a lo largo del genoma en una muestra de tamaño suficiente permite identificar variantes comunes de susceptibilidad de manera consistente y generar nuevas hipótesis acerca de la etiología del trastorno. Así, nuestro estudio de SNPs no sinónimos comunes a lo largo del genoma ha dado lugar a la identificación de una variante en el transportador de iones metálicos *SLC39A8* asociada de manera consistente a un incremento en el riesgo a padecer esquizofrenia en población europea; otros genes de la misma familia se asociaron con otros trastornos psiquiátricos, por lo que es posible que la alteración de la homeostasia de los metales durante el desarrollo embrionario y / o en el cerebro durante la vida de un individuo incremente el riesgo a desarrollar este tipo de trastornos, abriendo una nueva línea de investigación.

6. El análisis multilocus confirmó la existencia del componente poligénico en nuestra muestra de casos, es decir, que una parte del riesgo a padecer esquizofrenia está determinado por el efecto global de variantes individualmente poco significativas.

7. Dada la participación confirmada del gen *AHI1* en la etiología de la esquizofrenia y otras enfermedades psiquiátricas y la elevada probabilidad de que el SNP no sinónimo rs6940875, localizado en este gen y analizado en nuestro trabajo, sea una variante funcional y haya sido objeto de selección natural reciente, el resultado de nuestro estudio sugiere que esta variante podría tener alguna repercusión en el riesgo a padecer esquizofrenia.

Bibliografía

- Addington, J., Cadenhead, K.S., Cannon, T.D., Cornblatt, B., McGlashan, T.H., Perkins, D.O., Seidman, L.J., Tsuang, M., Walker, E.F., Woods, S.W., Heinssen, R., 2007. North American Prodrome Longitudinal Study: a collaborative multisite approach to prodromal schizophrenia research. *Schizophr Bull* 33, 665-672.
- Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., Sunyaev, S.R., 2010. A method and server for predicting damaging missense mutations. *Nat Methods* 7, 248-249.
- Akil, H., Brenner, S., Kandel, E., Kendler, K.S., King, M.C., Scolnick, E., Watson, J.D., Zoghbi, H.Y., 2010. Medicine. The future of psychiatric research: genomes and neural circuits. *Science* 327, 1580-1581.
- Alkelai, A., Lupoli, S., Greenbaum, L., Giegling, I., Kohn, Y., Sarner-Kanyas, K., Ben-Asher, E., Lancet, D., Rujescu, D., Macciardi, F., Lerer, B., 2011a. Identification of new schizophrenia susceptibility loci in an ethnically homogeneous, family-based, Arab-Israeli sample. *FASEB J* 25, 4011-4023.
- Alkelai, A., Lupoli, S., Greenbaum, L., Kohn, Y., Kanyas-Sarner, K., Ben-Asher, E., Lancet, D., Macciardi, F., Lerer, B., 2011b. DOCK4 and CEACAM21 as novel schizophrenia candidate genes in the Jewish population. *Int J Neuropsychopharmacol*, 1-11.
- Altshuler, D.M., Gibbs, R.A., Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu, F., Bonnen, P.E., de Bakker, P.I., Deloukas, P., Gabriel, S.B., Gwilliam, R., Hunt, S., Inouye, M., Jia, X., Palotie, A., Parkin, M., Whittaker, P., Chang, K., Hawes, A., Lewis, L.R., Ren, Y., Wheeler, D., Muzny, D.M., Barnes, C., Darvishi, K., Hurler, M., Korn, J.M., Kristiansson, K., Lee, C., McCarroll, S.A., Nemes, J., Keinan, A., Montgomery, S.B., Pollack, S., Price, A.L., Soranzo, N., Gonzaga-Jauregui, C., Anttila, V., Brodeur, W., Daly, M.J., Leslie, S., McVean, G., Moutsianas, L., Nguyen, H., Zhang, Q., Ghorri, M.J., McGinnis, R., McLaren, W., Takeuchi, F., Grossman, S.R., Shlyakhter, I., Hostetter, E.B., Sabeti, P.C., Adebamowo, C.A., Foster, M.W., Gordon, D.R., Licinio, J., Manca, M.C., Marshall, P.A., Matsuda, I., Ngare, D., Wang, V.O., Reddy, D., Rotimi, C.N., Royal, C.D., Sharp, R.R., Zeng, C., Brooks, L.D., McEwen, J.E., 2010. Integrating common and rare genetic variation in diverse human populations. *Nature* 467, 52-58.
- Alvarez Retuerto, A.I., Cantor, R.M., Gleeson, J.G., Ustaszewska, A., Schackwitz, W.S., Pennacchio, L.A., Geschwind, D.H., 2008. Association of common variants in the Joubert syndrome gene (AHI1) with autism. *Hum Mol Genet* 17, 3887-3896.
- Allen, N.C., Bagade, S., McQueen, M.B., Ioannidis, J.P., Kavvoura, F.K., Khoury, M.J., Tanzi, R.E., Bertram, L., 2008. Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat Genet* 40, 827-834.
- American Psychiatric Association 2002. Diagnostic and Statistical Manual of Mental Disorders.
- Andreasen, N.C., 2007. DSM and the death of phenomenology in america: an example of unintended consequences. *Schizophr Bull* 33, 108-112.

- Arion, D., Unger, T., Lewis, D.A., Levitt, P., Mirmics, K., 2007. Molecular evidence for increased expression of genes related to immune and chaperone function in the prefrontal cortex in schizophrenia. *Biol Psychiatry* 62, 711-721.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25, 25-29.
- Athanasou, L., Mattingsdal, M., Kahler, A.K., Brown, A., Gustafsson, O., Agartz, I., Giegling, I., Muglia, P., Cichon, S., Rietschel, M., Pietilainen, O.P., Peltonen, L., Bramon, E., Collier, D., Clair, D.S., Sigurdsson, E., Petursson, H., Rujescu, D., Melle, I., Steen, V.M., Djurovic, S., Andreassen, O.A., 2010. Gene variants associated with schizophrenia in a Norwegian genome-wide study are replicated in a large European cohort. *J Psychiatr Res* 44, 748-753.
- Badner, J.A., Gershon, E.S., 2002. Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. *Mol Psychiatry* 7, 405-411.
- Bah, J., Quach, H., Ebstein, R.P., Segman, R.H., Melke, J., Jamain, S., Rietschel, M., Modai, I., Kanas, K., Karni, O., Lerer, B., Gourion, D., Krebs, M.O., Etain, B., Schurhoff, F., Szoke, A., Leboyer, M., Bourgeron, T., 2004. Maternal transmission disequilibrium of the glutamate receptor GRIK2 in schizophrenia. *Neuroreport* 15, 1987-1991.
- Barreiro, L.B., Laval, G., Quach, H., Patin, E., Quintana-Murci, L., 2008. Natural selection has driven population differentiation in modern humans. *Nat Genet* 40, 340-345.
- Bartley, M., 1994. Unemployment and ill health: understanding the relationship. *J Epidemiol Community Health* 48, 333-337.
- Basu, S., Pan, W., 2011. Comparison of statistical tests for disease association with rare variants. *Genet Epidemiol* 35, 606-619.
- Batel, P., 2000. Addiction and schizophrenia. *Eur Psychiatry* 15, 115-122.
- Bigos, K.L., Mattay, V.S., Callicott, J.H., Straub, R.E., Vakkalanka, R., Kolachana, B., Hyde, T.M., Lipska, B.K., Kleinman, J.E., Weinberger, D.R., 2010. Genetic variation in CACNA1C affects brain circuitries related to mental illness. *Arch Gen Psychiatry* 67, 939-945.
- Bilder, R.M., Howe, A., Novak, N., Sabb, F.W., Parker, D.S., 2011. The genetics of cognitive impairment in schizophrenia: a phenomic perspective. *Trends Cogn Sci* 15, 428-435.
- Blackwood, D.H., Fordyce, A., Walker, M.T., St Clair, D.M., Porteous, D.J., Muir, W.J., 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, 428-433.
- Blake, D.J., Forrest, M., Chapman, R.M., Tinsley, C.L., O'Donovan, M.C., Owen, M.J., 2010. TCF4, schizophrenia, and Pitt-Hopkins Syndrome. *Schizophr Bull* 36, 443-447.

- Boos, H.B., Aleman, A., Cahn, W., Hulshoff Pol, H., Kahn, R.S., 2007. Brain volumes in relatives of patients with schizophrenia: a meta-analysis. *Arch Gen Psychiatry* 64, 297-304.
- Botto, L.D., May, K., Fernhoff, P.M., Correa, A., Coleman, K., Rasmussen, S.A., Merritt, R.K., O'Leary, L.A., Wong, L.Y., Elixson, E.M., Mahle, W.T., Campbell, R.M., 2003. A population-based study of the 22q11.2 deletion: phenotype, incidence, and contribution to major birth defects in the population. *Pediatrics* 112, 101-107.
- Boulanger, L.M., 2009. Immune proteins in brain development and synaptic plasticity. *Neuron* 64, 93-109.
- Bramon, E., McDonald, C., Croft, R.J., Landau, S., Filbey, F., Gruzelier, J.H., Sham, P.C., Frangou, S., Murray, R.M., 2005. Is the P300 wave an endophenotype for schizophrenia? A meta-analysis and a family study. *Neuroimage* 27, 960-968.
- Brockschmidt, A., Todt, U., Ryu, S., Hoischen, A., Landwehr, C., Birnbaum, S., Frenck, W., Radlwimmer, B., Lichter, P., Engels, H., Driever, W., Kubisch, C., Weber, R.G., 2007. Severe mental retardation with breathing abnormalities (Pitt-Hopkins syndrome) is caused by haploinsufficiency of the neuronal bHLH transcription factor TCF4. *Hum Mol Genet* 16, 1488-1494.
- Brody, A.L., Mandelkern, M.A., Jarvik, M.E., Lee, G.S., Smith, E.C., Huang, J.C., Bota, R.G., Bartzokis, G., London, E.D., 2004. Differences between smokers and nonsmokers in regional gray matter volumes and densities. *Biol Psychiatry* 55, 77-84.
- Brown, A.S., 2011. The environment and susceptibility to schizophrenia. *Prog Neurobiol* 93, 23-58.
- Brune, M., 2004. Schizophrenia-an evolutionary enigma? *Neurosci Biobehav Rev* 28, 41-53.
- Bundy, H., Stahl, D., MacCabe, J.H., 2011. A systematic review and meta-analysis of the fertility of patients with schizophrenia and their unaffected relatives. *Acta Psychiatr Scand* 123, 98-106.
- Burns, J., 2006. The social brain hypothesis of schizophrenia. *World Psychiatry* 5, 77-81.
- Camargo, L.M., Collura, V., Rain, J.C., Mizuguchi, K., Hermjakob, H., Kerrien, S., Bonnert, T.P., Whiting, P.J., Brandon, N.J., 2007. Disrupted in Schizophrenia 1 Interactome: evidence for the close connectivity of risk genes and a potential synaptic basis for schizophrenia. *Mol Psychiatry* 12, 74-86.
- Camps, M., Herman, A., Loh, E., Loeb, L.A., 2007. Genetic constraints on protein evolution. *Crit Rev Biochem Mol Biol* 42, 313-326.
- Cantor-Graae, E., Selten, J.P., 2005. Schizophrenia and migration: a meta-analysis and review. *Am J Psychiatry* 162, 12-24.
- Capon, F., Allen, M.H., Ameen, M., Burden, A.D., Tillman, D., Barker, J.N., Trembath, R.C., 2004. A synonymous SNP of the corneodesmosin gene leads to increased mRNA stability and demonstrates association with psoriasis across diverse ethnic groups. *Hum Mol Genet* 13, 2361-2368.

- Cardno, A.G., Marshall, E.J., Coid, B., Macdonald, A.M., Ribchester, T.R., Davies, N.J., Venturi, P., Jones, L.A., Lewis, S.W., Sham, P.C., Gottesman, II, Farmer, A.E., McGuffin, P., Reveley, A.M., Murray, R.M., 1999. Heritability estimates for psychotic disorders: the Maudsley twin psychosis series. *Arch Gen Psychiatry* 56, 162-168.
- Cardon, L.R., Bell, J.I., 2001. Association study designs for complex diseases. *Nat Rev Genet* 2, 91-99.
- Carless, M.A., Glahn, D.C., Johnson, M.P., Curran, J.E., Bozaoglu, K., Dyer, T.D., Winkler, A.M., Cole, S.A., Almasy, L., MacCluer, J.W., Duggirala, R., Moses, E.K., Goring, H.H., Blangero, J., 2011. Impact of DISC1 variation on neuroanatomical and neurocognitive phenotypes. *Mol Psychiatry* 16, 1096-1104, 1063.
- Carpenter, W.T., 2011. The facts of schizophrenia: a personal commentary. *Schizophr Res* 128, 3-4.
- Carter, M., Watts, C.A., 1971. Possible biological advantages among schizophrenics' relatives. *Br J Psychiatry* 118, 453-460.
- Cirulli, E.T., Goldstein, D.B., 2010. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* 11, 415-425.
- Collier, D.A., Li, T., 2003. The genetics of schizophrenia: glutamate not dopamine? *Eur J Pharmacol* 480, 177-184.
- Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P., Fitzgerald, T., Hu, M., Ihm, C.H., Kristiansson, K., Macarthur, D.G., Macdonald, J.R., Onyiah, I., Pang, A.W., Robson, S., Stirrups, K., Valsesia, A., Walter, K., Wei, J., Tyler-Smith, C., Carter, N.P., Lee, C., Scherer, S.W., Hurles, M.E., 2009. Origins and functional impact of copy number variation in the human genome. *Nature* 464, 704-712.
- Cooper, B., 2005. Immigration and schizophrenia: the social causation hypothesis revisited. *Br J Psychiatry* 186, 361-363.
- Corvin, A., McGhee, K.A., Murphy, K., Donohoe, G., Nangle, J.M., Schwaiger, S., Kenny, N., Clarke, S., Meagher, D., Quinn, J., Scully, P., Baldwin, P., Browne, D., Walsh, C., Waddington, J.L., Morris, D.W., Gill, M., 2007. Evidence for association and epistasis at the DAOA/G30 and D-amino acid oxidase loci in an Irish schizophrenia sample. *Am J Med Genet B Neuropsychiatr Genet* 144B, 949-953.
- Costas, J., Salas, A., Phillips, C., Carracedo, A., 2005. Human genome-wide screen of haplotype-like blocks of reduced diversity. *Gene* 349, 219-225.
- Coyle, J.T., 2006. Glutamate and schizophrenia: beyond the dopamine hypothesis. *Cell Mol Neurobiol* 26, 365-384.
- Crespi, B., Summers, K., Dorus, S., 2007. Adaptive evolution of genes underlying schizophrenia. *Proc Biol Sci* 274, 2801-2810.
- Crow, T.J., 1997. Is schizophrenia the price that Homo sapiens pays for language? *Schizophr Res* 28, 127-141.

- Crow, T.J., 2008. The 'big bang' theory of the origin of psychosis and the faculty of language. *Schizophr Res* 102, 31-52.
- Curtis, D., Vine, A.E., McQuillin, A., Bass, N.J., Pereira, A., Kandaswamy, R., Lawrence, J., Anjorin, A., Choudhury, K., Datta, S.R., Puri, V., Krasucki, R., Pimm, J., Thirumalai, S., Queded, D., Gurling, H.M., 2011. Case-case genome-wide association analysis shows markers differentially associated with schizophrenia and bipolar disorder and implicates calcium channel genes. *Psychiatr Genet* 21, 1-4.
- Chanock, S.J., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D.J., Thomas, G., Hirschhorn, J.N., Abecasis, G., Altshuler, D., Bailey-Wilson, J.E., Brooks, L.D., Cardon, L.R., Daly, M., Donnelly, P., Fraumeni, J.F., Jr., Freimer, N.B., Gerhard, D.S., Gunter, C., Guttmacher, A.E., Guyer, M.S., Harris, E.L., Hoh, J., Hoover, R., Kong, C.A., Merikangas, K.R., Morton, C.C., Palmer, L.J., Phimister, E.G., Rice, J.P., Roberts, J., Rotimi, C., Tucker, M.A., Vogan, K.J., Wacholder, S., Wijsman, E.M., Winn, D.M., Collins, F.S., 2007. Replicating genotype-phenotype associations. *Nature* 447, 655-660.
- Chen, J., Lee, G., Fanous, A.H., Zhao, Z., Jia, P., O'Neill, A., Walsh, D., Kendler, K.S., Chen, X., 2011. Two non-synonymous markers in PTPN21, identified by genome-wide association study data-mining and replication, are associated with schizophrenia. *Schizophr Res* 131, 43-51.
- Chen, X., Wang, X., Hossain, S., O'Neill, F.A., Walsh, D., van den Oord, E., Fanous, A., Kendler, K.S., 2007. Interleukin 3 and schizophrenia: the impact of sex and family history. *Mol Psychiatry* 12, 273-282.
- Chumakov, I., Blumenfeld, M., Guerassimenko, O., Cavarec, L., Palicio, M., Abderrahim, H., Bougueleret, L., Barry, C., Tanaka, H., La Rosa, P., Puech, A., Tahri, N., Cohen-Akenine, A., Delabrosse, S., Lissarrague, S., Picard, F.P., Maurice, K., Essioux, L., Millasseau, P., Grel, P., Debailleul, V., Simon, A.M., Caterina, D., Dufaure, I., Malekzadeh, K., Belova, M., Luan, J.J., Bouillot, M., Sambucy, J.L., Primas, G., Saumier, M., Boubkiri, N., Martin-Saumier, S., Nasroune, M., Peixoto, H., Delaye, A., Pinchot, V., Bastucci, M., Guillou, S., Chevillon, M., Sainz-Fuertes, R., Meguenni, S., Aurich-Costa, J., Cherif, D., Gimalac, A., Van Duijn, C., Gauvreau, D., Ouellette, G., Fortier, I., Raelson, J., Sherbatich, T., Riazanskaia, N., Rogaev, E., Raeymaekers, P., Aerssens, J., Konings, F., Luyten, W., Macchiardi, F., Sham, P.C., Straub, R.E., Weinberger, D.R., Cohen, N., Cohen, D., 2002. Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia. *Proc Natl Acad Sci U S A* 99, 13675-13680.
- Dajas-Bailador, F., Jones, E.V., Whitmarsh, A.J., 2008. The JIP1 scaffold protein regulates axonal development in cortical neurons. *Curr Biol* 18, 221-226.
- Devlin, B., Risch, N., 1995. A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* 29, 311-322.
- Di Giorgio, A., Blasi, G., Sambataro, F., Rampino, A., Papazacharias, A., Gambi, F., Romano, R., Caforio, G., Rizzo, M., Latorre, V., Papolizio, T., Kolachana, B., Callicott, J.H., Nardini, M., Weinberger, D.R., Bertolino, A., 2008. Association of the SerCys DISC1 polymorphism with human hippocampal formation gray matter and function during memory encoding. *Eur J Neurosci* 28, 2129-2136.

- Di Rienzo, A., 2006. Population genetics models of common diseases. *Curr Opin Genet Dev* 16, 630-636.
- Di Rienzo, A., Hudson, R.R., 2005. An evolutionary framework for common diseases: the ancestral-susceptibility model. *Trends Genet* 21, 596-601.
- Dominguez, E., Loza, M.I., Padin, F., Gesteira, A., Paz, E., Paramo, M., Brenlla, J., Pumar, E., Iglesias, F., Cibeira, A., Castro, M., Caruncho, H., Carracedo, A., Costas, J., 2007. Extensive linkage disequilibrium mapping at HTR2A and DRD3 for schizophrenia susceptibility genes in the Galician population. *Schizophr Res* 90, 123-129.
- Dong, Z., Zhou, L., Del Villar, K., Ghanevati, M., Tashjian, V., Miller, C.A., 2005. JIP1 regulates neuronal apoptosis in response to stress. *Brain Res Mol Brain Res* 134, 282-293.
- Driscoll, D.A., Salvin, J., Sellinger, B., Budarf, M.L., McDonald-McGinn, D.M., Zackai, E.H., Emanuel, B.S., 1993. Prevalence of 22q11 microdeletions in DiGeorge and velocardiofacial syndromes: implications for genetic counselling and prenatal diagnosis. *J Med Genet* 30, 813-817.
- Dudbridge, F., Gusnanto, A., 2008. Estimation of significance thresholds for genomewide association scans. *Genet Epidemiol* 32, 227-234.
- Dwyer, S., Williams, H., Jones, I., Jones, L., Walters, J., Craddock, N., Owen, M.J., O'Donovan, M.C., 2011. Investigation of rare non-synonymous variants at ABCA13 in schizophrenia and bipolar disorder. *Mol Psychiatry* 16, 790-791.
- Edelmann, L., Pandita, R.K., Morrow, B.E., 1999. Low-copy repeats mediate the common 3-Mb deletion in patients with velo-cardio-facial syndrome. *Am J Hum Genet* 64, 1076-1086.
- Ekelund, J., Hennah, W., Hiekkalinna, T., Parker, A., Meyer, J., Lonqvist, J., Peltonen, L., 2004. Replication of 1q42 linkage in Finnish schizophrenia pedigrees. *Mol Psychiatry* 9, 1037-1041.
- Erk, S., Meyer-Lindenberg, A., Schnell, K., Opitz von Boberfeld, C., Esslinger, C., Kirsch, P., Grimm, O., Arnold, C., Haddad, L., Witt, S.H., Cichon, S., Nothen, M.M., Rietschel, M., Walter, H., 2010. Brain function in carriers of a genome-wide supported bipolar disorder variant. *Arch Gen Psychiatry* 67, 803-811.
- Esslinger, C., Walter, H., Kirsch, P., Erk, S., Schnell, K., Arnold, C., Haddad, L., Mier, D., Opitz von Boberfeld, C., Raab, K., Witt, S.H., Rietschel, M., Cichon, S., Meyer-Lindenberg, A., 2009. Neural mechanisms of a genome-wide supported psychosis variant. *Science* 324, 605.
- Falconer DS, Mackay TFC. 1996. *Introduction to Quantitative Genetics*, 4th edn. Longman: Harlow, Essex, UK.
- Fallin, M.D., Lasseter, V.K., Avramopoulos, D., Nicodemus, K.K., Wolyniec, P.S., McGrath, J.A., Steel, G., Nestadt, G., Liang, K.Y., Haganir, R.L., Valle, D., Pulver, A.E., 2005. Bipolar I disorder and schizophrenia: a 440-single-nucleotide polymorphism screen of 64 candidate genes among Ashkenazi Jewish case-parent trios. *Am J Hum Genet* 77, 918-936.
- Fearon, P., Kirkbride, J.B., Morgan, C., Dazzan, P., Morgan, K., Lloyd, T., Hutchinson, G., Tarrant, J., Fung, W.L., Holloway, J., Mallett, R., Harrison,

- G., Leff, J., Jones, P.B., Murray, R.M., 2006. Incidence of schizophrenia and other psychoses in ethnic minority groups: results from the MRC AESOP Study. *Psychol Med* 36, 1541-1550.
- Ferland, R.J., Eyaid, W., Collura, R.V., Tully, L.D., Hill, R.S., Al-Nouri, D., Al-Rumayyan, A., Topcu, M., Gascon, G., Bodell, A., Shugart, Y.Y., Ruvolo, M., Walsh, C.A., 2004. Abnormal cerebellar development and axonal decussation due to mutations in *AHI1* in Joubert syndrome. *Nat Genet* 36, 1008-1013.
- Ferreira, M.A., O'Donovan, M.C., Meng, Y.A., Jones, I.R., Ruderfer, D.M., Jones, L., Fan, J., Kirov, G., Perlis, R.H., Green, E.K., Smoller, J.W., Grozeva, D., Stone, J., Nikolov, I., Chambert, K., Hamshere, M.L., Nimgaonkar, V.L., Moskvina, V., Thase, M.E., Caesar, S., Sachs, G.S., Franklin, J., Gordon-Smith, K., Ardlie, K.G., Gabriel, S.B., Fraser, C., Blumenstiel, B., Defelice, M., Breen, G., Gill, M., Morris, D.W., Elkin, A., Muir, W.J., McGhee, K.A., Williamson, R., MacIntyre, D.J., MacLean, A.W., St, C.D., Robinson, M., Van Beck, M., Pereira, A.C., Kandaswamy, R., McQuillin, A., Collier, D.A., Bass, N.J., Young, A.H., Lawrence, J., Ferrier, I.N., Anjorin, A., Farmer, A., Curtis, D., Scolnick, E.M., McGuffin, P., Daly, M.J., Corvin, A.P., Holmans, P.A., Blackwood, D.H., Gurling, H.M., Owen, M.J., Purcell, S.M., Sklar, P., Craddock, N., 2008. Collaborative genome-wide association analysis supports a role for *ANKK1* and *CACNA1C* in bipolar disorder. *Nat Genet* 40, 1056-1058.
- Feuk, L., Carson, A.R., Scherer, S.W., 2006. Structural variation in the human genome. *Nat Rev Genet* 7, 85-97.
- Flora, A., Garcia, J.J., Thaller, C., Zoghbi, H.Y., 2007. The E-protein Tcf4 interacts with *Math1* to regulate differentiation of a specific subset of neuronal progenitors. *Proc Natl Acad Sci U S A* 104, 15382-15387.
- Fourgeaud, L., Boulanger, L.M., 2010. Role of immune molecules in the establishment and plasticity of glutamatergic synapses. *Eur J Neurosci* 32, 207-217.
- Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., Pasternak, S., Wheeler, D.A., Willis, T.D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhou, J., Gabriel, S.B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R.C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Sun, W., Wang, H., Wang, Y., Xiong, X., Xu, L., Wayne, M.M., Tsui, S.K., Xue, H., Wong, J.T., Galver, L.M., Fan, J.B., Gunderson, K., Murray, S.S., Oliphant, A.R., Chee, M.S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J.F., Phillips, M.S., Roumy, S., Sallee, C., Verner, A., Hudson, T.J., Kwok, P.Y., Cai, D., Koboldt, D.C., Miller, R.D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L.C., Mak, W., Song, Y.Q., Tam, P.K., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C.P., Delgado, M., Dermitzakis, E.T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B.E., Whittaker, P., Bentley, D.R., Daly, M.J., de Bakker, P.I., Barrett, J., Chretien, Y.R., Maller, J., McCarroll, S., Patterson, N., Pe'er, I.,

- Price, A., Purcell, S., Richter, D.J., Sabeti, P., Saxena, R., Schaffner, S.F., Sham, P.C., Varilly, P., Stein, L.D., Krishnan, L., Smith, A.V., Tello-Ruiz, M.K., Thorisson, G.A., Chakravarti, A., Chen, P.E., Cutler, D.J., Kashuk, C.S., Lin, S., Abecasis, G.R., Guan, W., Li, Y., Munro, H.M., Qin, Z.S., Thomas, D.J., McVean, G., Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L.R., Clarke, G., Evans, D.M., Morris, A.P., Weir, B.S., Mullikin, J.C., Sherry, S.T., Feolo, M., Skol, A., Zhang, H., Matsuda, I., Fukushima, Y., Macer, D.R., Suda, E., Rotimi, C.N., Adebamowo, C.A., Ajayi, I., Aniagwu, T., Marshall, P.A., Nkwodimmah, C., Royal, C.D., Leppert, M.F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I.F., Knoppers, B.M., Foster, M.W., Clayton, E.W., Watkin, J., Muzny, D., Nazareth, L., Sodergren, E., Weinstock, G.M., Yakub, I., Birren, B.W., Wilson, R.K., Fulton, L.L., Rogers, J., Burton, J., Carter, N.P., Clee, C.M., Griffiths, M., Jones, M.C., McLay, K., Plumb, R.W., Ross, M.T., Sims, S.K., Willey, D.L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J.C., L'Archeveque, P., Bellemare, G., Saeki, K., An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A.L., Brooks, L.D., McEwen, J.E., Guyer, M.S., Wang, V.O., Peterson, J.L., Shi, M., Spiegel, J., Sung, L.M., Zacharia, L.F., Collins, F.S., Kennedy, K., Jamieson, R., Stewart, J., 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449, 851-861.
- Fribourg, M., Moreno, J.L., Holloway, T., Provasi, D., Baki, L., Mahajan, R., Park, G., Adney, S.K., Hatcher, C., Eltit, J.M., Ruta, J.D., Albizu, L., Li, Z., Umali, A., Shim, J., Fabiato, A., Mackerell, A.D., Jr., Brezina, V., Sealfon, S.C., Filizola, M., Gonzalez-Maeso, J., Logothetis, D.E., 2011. Decoding the Signaling of a GPCR Heteromeric Complex Reveals a Unifying Mechanism of Action of Antipsychotic Drugs. *Cell* 147, 1011-1023.
- Friedman, J.I., Vrijenhoek, T., Markx, S., Janssen, I.M., van der Vliet, W.A., Faas, B.H., Knoers, N.V., Cahn, W., Kahn, R.S., Edelman, L., Davis, K.L., Silverman, J.M., Brunner, H.G., van Kessel, A.G., Wijmenga, C., Ophoff, R.A., Veltman, J.A., 2008. CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy. *Mol Psychiatry* 13, 261-266.
- Fullerton, S.M., Clark, A.G., Weiss, K.M., Nickerson, D.A., Taylor, S.L., Stengard, J.H., Salomaa, V., Vartiainen, E., Perola, M., Boerwinkle, E., Sing, C.F., 2000. Apolipoprotein E variation at the sequence haplotype level: implications for the origin and maintenance of a major human polymorphism. *Am J Hum Genet* 67, 881-900.
- Gejman, P.V., Sanders, A.R., Kendler, K.S., 2011. Genetics of Schizophrenia: New Findings and Challenges. *Annu Rev Genomics Hum Genet* 12, 121-144.
- Genovese, G., Friedman, D.J., Ross, M.D., Lecordier, L., Uzureau, P., Freedman, B.I., Bowden, D.W., Langefeld, C.D., Oleksyk, T.K., Uscinski Knob, A.L., Bernhardt, A.J., Hicks, P.J., Nelson, G.W., Vanhollebeke, B., Winkler, C.A., Kopp, J.B., Pays, E., Pollak, M.R., 2010. Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science* 329, 841-845.
- Geyer, M.A., Vollenweider, F.X., 2008. Serotonin research: contributions to understanding psychoses. *Trends Pharmacol Sci* 29, 445-453.

- Girard, S.L., Gauthier, J., Noreau, A., Xiong, L., Zhou, S., Jouan, L., Dionne-Laporte, A., Spiegelman, D., Henrion, E., Diallo, O., Thibodeau, P., Bachand, I., Bao, J.Y., Tong, A.H., Lin, C.H., Millet, B., Jaafari, N., Joobar, R., Dion, P.A., Lok, S., Krebs, M.O., Rouleau, G.A., 2011. Increased exonic de novo mutation rate in individuals with schizophrenia. *Nat Genet* 43, 860-863.
- Girirajan, S., Eichler, E.E., 2010. Phenotypic variability and genetic susceptibility to genomic disorders. *Hum Mol Genet* 19, R176-187.
- Girirajan, S., Rosenfeld, J.A., Cooper, G.M., Antonacci, F., Siswara, P., Itsara, A., Vives, L., Walsh, T., McCarthy, S.E., Baker, C., Mefford, H.C., Kidd, J.M., Browning, S.R., Browning, B.L., Dickel, D.E., Levy, D.L., Ballif, B.C., Platky, K., Farber, D.M., Gowans, G.C., Wetherbee, J.J., Asamoah, A., Weaver, D.D., Mark, P.R., Dickerson, J., Garg, B.P., Ellingwood, S.A., Smith, R., Banks, V.C., Smith, W., McDonald, M.T., Hoo, J.J., French, B.N., Hudson, C., Johnson, J.P., Ozmore, J.R., Moeschler, J.B., Surti, U., Escobar, L.F., El-Khechen, D., Gorski, J.L., Kussmann, J., Salbert, B., Lacassie, Y., Biser, A., McDonald-McGinn, D.M., Zackai, E.H., Deardorff, M.A., Shaikh, T.H., Haan, E., Friend, K.L., Fichera, M., Romano, C., Gecz, J., DeLisi, L.E., Sebat, J., King, M.C., Shaffer, L.G., Eichler, E.E., 2010. A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat Genet* 42, 203-209.
- Glahn, D.C., Therman, S., Manninen, M., Huttunen, M., Kaprio, J., Lonnqvist, J., Cannon, T.D., 2003. Spatial working memory as an endophenotype for schizophrenia. *Biol Psychiatry* 53, 624-626.
- Gleibermann, L. 1973. Blood pressure and dietary salt in human populations. *Ecol. Food Nutr.* 2, 143-156
- Glessner, J.T., Reilly, M.P., Kim, C.E., Takahashi, N., Albano, A., Hou, C., Bradfield, J.P., Zhang, H., Sleiman, P.M., Flory, J.H., Imielinski, M., Frackelton, E.C., Chiavacci, R., Thomas, K.A., Garris, M., Otieno, F.G., Davidson, M., Weiser, M., Reichenberg, A., Davis, K.L., Friedman, J.I., Cappola, T.P., Margulies, K.B., Rader, D.J., Grant, S.F., Buxbaum, J.D., Gur, R.E., Hakonarson, H., 2010. Strong synaptic transmission impact by copy number variations in schizophrenia. *Proc Natl Acad Sci U S A* 107, 10584-10589.
- Goldman, A.L., Pezawas, L., Mattay, V.S., Fischl, B., Verchinski, B.A., Zolnick, B., Weinberger, D.R., Meyer-Lindenberg, A., 2008. Heritability of brain morphology related to schizophrenia: a large-scale automated magnetic resonance imaging segmentation study. *Biol Psychiatry* 63, 475-483.
- Gomez-Ospina, N., Tsuruta, F., Barreto-Chang, O., Hu, L., Dolmetsch, R., 2006. The C terminus of the L-type voltage-gated calcium channel Ca(V)1.2 encodes a transcription factor. *Cell* 127, 591-606.
- Gonzalez-Maeso, J., Ang, R.L., Yuen, T., Chan, P., Weisstaub, N.V., Lopez-Gimenez, J.F., Zhou, M., Okawa, Y., Callado, L.F., Milligan, G., Gingrich, J.A., Filizola, M., Meana, J.J., Sealfon, S.C., 2008. Identification of a serotonin/glutamate receptor complex implicated in psychosis. *Nature* 452, 93-97.
- Gorlov, I.P., Gorlova, O.Y., Frazier, M.L., Spitz, M.R., Amos, C.I., 2010. Evolutionary evidence of the effect of rare variants on disease etiology. *Clin Genet* 79, 199-206.

- Gorlov, I.P., Gorlova, O.Y., Sunyaev, S.R., Spitz, M.R., Amos, C.I., 2008. Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. *Am J Hum Genet* 82, 100-112.
- Gottesman, II, Gould, T.D., 2003. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry* 160, 636-645.
- Gottesman, II, Shields, J., 1967. A polygenic theory of schizophrenia. *Proc Natl Acad Sci U S A* 58, 199-205.
- Gratacos, M., Costas, J., de Cid, R., Bayes, M., Gonzalez, J.R., Baca-Garcia, E., de Diego, Y., Fernandez-Aranda, F., Fernandez-Piqueras, J., Guitart, M., Martin-Santos, R., Martorell, L., Menchon, J.M., Roca, M., Saiz-Ruiz, J., Sanjuan, J., Torrens, M., Urretavizcaya, M., Valero, J., Vilella, E., Estivill, X., Carracedo, A., 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, 808-816.
- Green, E.K., Grozeva, D., Jones, I., Jones, L., Kirov, G., Caesar, S., Gordon-Smith, K., Fraser, C., Forty, L., Russell, E., Hamshere, M.L., Moskvina, V., Nikolov, I., Farmer, A., McGuffin, P., Holmans, P.A., Owen, M.J., O'Donovan, M.C., Craddock, N., 2010. The bipolar disorder risk allele at CACNA1C also confers risk of recurrent major depression and of schizophrenia. *Mol Psychiatry* 15, 1016-1022.
- Greenwood, T.A., Braff, D.L., Light, G.A., Cadenhead, K.S., Calkins, M.E., Dobie, D.J., Freedman, R., Green, M.F., Gur, R.E., Gur, R.C., Mintz, J., Nuechterlein, K.H., Olincy, A., Radant, A.D., Seidman, L.J., Siever, L.J., Silverman, J.M., Stone, W.S., Swerdlow, N.R., Tsuang, D.W., Tsuang, M.T., Turetsky, B.I., Schork, N.J., 2007. Initial heritability analyses of endophenotypic measures for schizophrenia: the consortium on the genetics of schizophrenia. *Arch Gen Psychiatry* 64, 1242-1250.
- Gulacsi, A.A., Anderson, S.A., 2008. Beta-catenin-mediated Wnt signaling regulates neurogenesis in the ventral telencephalon. *Nat Neurosci* 11, 1383-1391.
- Gurevich, E.V., Bordelon, Y., Shapiro, R.M., Arnold, S.E., Gur, R.E., Joyce, J.N., 1997. Mesolimbic dopamine D3 receptors and use of antipsychotics in patients with schizophrenia. A postmortem study. *Arch Gen Psychiatry* 54, 225-232.
- Guryev, V., Smits, B.M., van de Belt, J., Verheul, M., Hubner, N., Cuppen, E., 2006. Haplotype block structure is conserved across mammals. *PLoS Genet* 2, e121.
- Haahr, U., Friis, S., Larsen, T.K., Melle, I., Johannessen, J.O., Opjordsmoen, S., Simonsen, E., Rund, B.R., Vaglum, P., McGlashan, T., 2008. First-episode psychosis: diagnostic stability over one and two years. *Psychopathology* 41, 322-329.
- Hafner, H., 2003. Gender differences in schizophrenia. *Psychoneuroendocrinology* 28 Suppl 2, 17-54.
- Hafner, H., Riecher-Rossler, A., Maurer, K., Fatkenheuer, B., Loffler, W., 1992. First onset and early symptomatology of schizophrenia. A chapter of epidemiological and neurobiological research into age and sex differences. *Eur Arch Psychiatry Clin Neurosci* 242, 109-118.

- Hamblin, M.T., Thompson, E.E., Di Rienzo, A., 2002. Complex signatures of natural selection at the Duffy blood group locus. *Am J Hum Genet* 70, 369-383.
- Hamshere, M.L., Bennett, P., Williams, N., Segurado, R., Cardno, A., Norton, N., Lambert, D., Williams, H., Kirov, G., Corvin, A., Holmans, P., Jones, L., Jones, I., Gill, M., O'Donovan, M.C., Owen, M.J., Craddock, N., 2005. Genomewide linkage scan in schizoaffective disorder: significant evidence for linkage at 1q42 close to DISC1, and suggestive evidence at 22q11 and 19p13. *Arch Gen Psychiatry* 62, 1081-1088.
- Harrison, P.J., Weinberger, D.R., 2005. Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. *Mol Psychiatry* 10, 40-68; image 45.
- Hattori, T., Shimizu, S., Koyama, Y., Yamada, K., Kuwahara, R., Kumamoto, N., Matsuzaki, S., Ito, A., Katayama, T., Tohyama, M., 2010. DISC1 regulates cell-cell adhesion, cell-matrix adhesion and neurite outgrowth. *Mol Psychiatry* 15, 778, 798-809.
- Hawks, J., Wang, E.T., Cochran, G.M., Harpending, H.C., Moyzis, R.K., 2007. Recent acceleration of human adaptive evolution. *Proc Natl Acad Sci U S A* 104, 20753-20758.
- Higgins, J.P., Thompson, S.G., 2002. Quantifying heterogeneity in a meta-analysis. *Stat Med* 21, 1539-1558.
- Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., Manolio, T.A., 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* 106, 9362-9367.
- Hirschhorn, J.N., Daly, M.J., 2005. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 6, 95-108.
- Horrobin, D.F., 1998. Schizophrenia: the illness that made us human. *Med Hypotheses* 50, 269-288.
- Huang, F.L., Huang, K.P., Wu, J., Boucheron, C., 2006. Environmental enrichment enhances neurogranin expression and hippocampal learning and memory but fails to rescue the impairments of neurogranin null mutant mice. *J Neurosci* 26, 6230-6237.
- Hughes, A.L., Nei, M., 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335, 167-170.
- Huxley, J., Mayr, E., Osmond, H., Hoffer, A., 1964. Schizophrenia as a Genetic Morphism. *Nature* 204, 220-221.
- Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., Lee, C., 2004. Detection of large-scale variation in the human genome. *Nat Genet* 36, 949-951.
- Ikeda, M., Aleksic, B., Kinoshita, Y., Okochi, T., Kawashima, K., Kushima, I., Ito, Y., Nakamura, Y., Kishi, T., Okumura, T., Fukuo, Y., Williams, H.J., Hamshere, M.L., Ivanov, D., Inada, T., Suzuki, M., Hashimoto, R., Ujike, H., Takeda, M., Craddock, N., Kaibuchi, K., Owen, M.J., Ozaki, N., O'Donovan, M.C., Iwata,

- N., 2011. Genome-wide association study of schizophrenia in a Japanese population. *Biol Psychiatry* 69, 472-478.
- Ikeda, M., Aleksic, B., Kirov, G., Kinoshita, Y., Yamanouchi, Y., Kitajima, T., Kawashima, K., Okochi, T., Kishi, T., Zaharieva, I., Owen, M.J., O'Donovan, M.C., Ozaki, N., Iwata, N., 2010. Copy number variation in schizophrenia in the Japanese population. *Biol Psychiatry* 67, 283-286.
- Ilani, T., Ben-Shachar, D., Strous, R.D., Mazor, M., Sheinkman, A., Kotler, M., Fuchs, S., 2001. A peripheral marker for schizophrenia: Increased levels of D3 dopamine receptor mRNA in blood lymphocytes. *Proc Natl Acad Sci U S A* 98, 625-628.
- Ingason, A., Giegling, I., Cichon, S., Hansen, T., Rasmussen, H.B., Nielsen, J., Jurgens, G., Muglia, P., Hartmann, A.M., Strengman, E., Vasilescu, C., Muhleisen, T.W., Djurovic, S., Melle, I., Lerer, B., Moller, H.J., Francks, C., Pietilainen, O.P., Lonnqvist, J., Suvisaari, J., Tuulio-Henriksson, A., Walshe, M., Vassos, E., Di Forti, M., Murray, R., Bonetto, C., Tosato, S., Cantor, R.M., Rietschel, M., Craddock, N., Owen, M.J., Peltonen, L., Andreassen, O.A., Nothen, M.M., St Clair, D., Ophoff, R.A., O'Donovan, M.C., Collier, D.A., Werge, T., Rujescu, D., 2010. A large replication study and meta-analysis in European samples provides further support for association of AHI1 markers with schizophrenia. *Hum Mol Genet* 19, 1379-1386.
- Ingraham, L.J., Kety, S.S., 2000. Adoption studies of schizophrenia. *Am J Med Genet* 97, 18-22.
- Insel, T.R., 2010. Rethinking schizophrenia. *Nature* 468, 187-193.
- Ioannidis, J.P., 2008. Why most discovered true associations are inflated. *Epidemiology* 19, 640-648.
- Iritani, S., 2007. Neuropathology of schizophrenia: a mini review. *Neuropathology* 27, 604-608.
- ISC, 2008. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 455, 237-241.
- ISC, 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460, 748-752.
- Ishizuka, K., Kamiya, A., Oh, E.C., Kanki, H., Seshadri, S., Robinson, J.F., Murdoch, H., Dunlop, A.J., Kubo, K., Furukori, K., Huang, B., Zeledon, M., Hayashi-Takagi, A., Okano, H., Nakajima, K., Houslay, M.D., Katsanis, N., Sawa, A., 2011. DISC1-dependent switch from progenitor proliferation to migration in the developing cortex. *Nature* 473, 92-96.
- Itsara, A., Wu, H., Smith, J.D., Nickerson, D.A., Romieu, I., London, S.J., Eichler, E.E., 2010. De novo rates and selection of large copy number variation. *Genome Res* 20, 1469-1481.
- Jarvik, L.F., Deckard, B.S., 1977. The Odyssean personality. A survival advantage for carriers of genes predisposing to schizophrenia? *Neuropsychobiology* 3, 179-191.
- Jungerius, B.J., Hoogendoorn, M.L., Bakker, S.C., Van't Slot, R., Bardoel, A.F., Ophoff, R.A., Wijmenga, C., Kahn, R.S., Sinke, R.J., 2008. An association

- screen of myelin-related genes implicates the chromosome 22q11 PIK4CA gene in schizophrenia. *Mol Psychiatry* 13, 1060-1068.
- Kahler, A.K., Djurovic, S., Kulle, B., Jonsson, E.G., Agartz, I., Hall, H., Opjordsmoen, S., Jakobsen, K.D., Hansen, T., Melle, I., Werge, T., Steen, V.M., Andreassen, O.A., 2008. Association analysis of schizophrenia on 18 genes involved in neuronal migration: MDGA1 as a new susceptibility gene. *Am J Med Genet B Neuropsychiatr Genet* 147B, 1089-1100.
- Kalscheuer, V.M., Feenstra, I., Van Ravenswaaij-Arts, C.M., Smeets, D.F., Menzel, C., Ullmann, R., Musante, L., Ropers, H.H., 2008. Disruption of the TCF4 gene in a girl with mental retardation but without the classical Pitt-Hopkins syndrome. *Am J Med Genet A* 146A, 2053-2059.
- Kaminsky, Z.A., Tang, T., Wang, S.C., Ptak, C., Oh, G.H., Wong, A.H., Feldcamp, L.A., Virtanen, C., Halfvarson, J., Tysk, C., McRae, A.F., Visscher, P.M., Montgomery, G.W., Gottesman, II, Martin, N.G., Petronis, A., 2009. DNA methylation profiles in monozygotic and dizygotic twins. *Nat Genet* 41, 240-245.
- Kamiya, A., Kubo, K., Tomoda, T., Takaki, M., Youn, R., Ozeki, Y., Sawamura, N., Park, U., Kudo, C., Okawa, M., Ross, C.A., Hatten, M.E., Nakajima, K., Sawa, A., 2005. A schizophrenia-associated mutation of DISC1 perturbs cerebral cortex development. *Nat Cell Biol* 7, 1167-1178.
- Kaneko, N., Muratake, T., Kuwabara, H., Kurosaki, T., Takei, M., Ohtsuki, T., Arinami, T., Tsuji, S., Someya, T., 2007. Autosomal linkage analysis of a Japanese single multiplex schizophrenia pedigree reveals two candidate loci on chromosomes 4q and 3q. *Am J Med Genet B Neuropsychiatr Genet* 144B, 735-742.
- Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A.M., Pletikos, M., Meyer, K.A., Sedmak, G., Guennel, T., Shin, Y., Johnson, M.B., Krsnik, Z., Mayer, S., Fertuzinhos, S., Umlauf, S., Lisgo, S.N., Vortmeyer, A., Weinberger, D.R., Mane, S., Hyde, T.M., Huttner, A., Reimers, M., Kleinman, J.E., Sestan, N., 2011. Spatio-temporal transcriptome of the human brain. *Nature* 478, 483-489.
- Karayorgou, M., Morris, M.A., Morrow, B., Shprintzen, R.J., Goldberg, R., Borrow, J., Gos, A., Nestadt, G., Wolyniec, P.S., Lasseter, V.K., et al., 1995. Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. *Proc Natl Acad Sci U S A* 92, 7612-7616.
- Karayorgou, M., Simon, T.J., Gogos, J.A., 2010. 22q11.2 microdeletions: linking DNA structural variation to brain dysfunction and schizophrenia. *Nat Rev Neurosci* 11, 402-416.
- Karksson, J.L., 1970. Genetic association of giftedness and creativity with schizophrenia. *Hereditas* 66, 177-182.
- Keller, M.C., Miller, G., 2006. Resolving the paradox of common, harmful, heritable mental disorders: which evolutionary genetic models work best? *Behav Brain Sci* 29, 385-404; discussion 405-352.

- Kelley, J.L., Madeoy, J., Calhoun, J.C., Swanson, W., Akey, J.M., 2006. Genomic signatures of positive selection in humans and the limits of outlier approaches. *Genome Res* 16, 980-989.
- Kempton, M.J., Ruberto, G., Vassos, E., Tatarelli, R., Girardi, P., Collier, D., Frangou, S., 2009. Effects of the CACNA1C risk allele for bipolar disorder on cerebral gray matter volume in healthy individuals. *Am J Psychiatry* 166, 1413-1414.
- Keshavan, M.S., Tandon, R., Boutros, N.N., Nasrallah, H.A., 2008. Schizophrenia, "just the facts": what we know in 2008 Part 3: neurobiology. *Schizophr Res* 106, 89-107.
- Khaitovich, P., Lockstone, H.E., Wayland, M.T., Tsang, T.M., Jayatilaka, S.D., Guo, A.J., Zhou, J., Somel, M., Harris, L.W., Holmes, E., Paabo, S., Bahn, S., 2008. Metabolic changes in schizophrenia and human brain evolution. *Genome Biol* 9, R124.
- Kimura, M., 1979. The neutral theory of molecular evolution. *Sci Am* 241, 98-100, 102, 108 passim.
- Kirov, G., Grozeva, D., Norton, N., Ivanov, D., Mantripragada, K.K., Holmans, P., Craddock, N., Owen, M.J., O'Donovan, M.C., 2009a. Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Hum Mol Genet* 18, 1497-1503.
- Kirov, G., Gumus, D., Chen, W., Norton, N., Georgieva, L., Sari, M., O'Donovan, M.C., Erdogan, F., Owen, M.J., Ropers, H.H., Ullmann, R., 2008. Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. *Hum Mol Genet* 17, 458-465.
- Kirov, G., O'Donovan, M.C., Owen, M.J., 2005. Finding schizophrenia genes. *J Clin Invest* 115, 1440-1448.
- Kirov, G., Zaharieva, I., Georgieva, L., Moskvina, V., Nikolov, I., Cichon, S., Hillmer, A., Toncheva, D., Owen, M.J., O'Donovan, M.C., 2009b. A genome-wide association study in 574 schizophrenia trios using DNA pooling. *Mol Psychiatry* 14, 796-803.
- Kleinjan, D.A., van Heyningen, V., 2005. Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am J Hum Genet* 76, 8-32.
- Kobrynski, L.J., Sullivan, K.E., 2007. Velocardiofacial syndrome, DiGeorge syndrome: the chromosome 22q11.2 deletion syndromes. *Lancet* 370, 1443-1452.
- Kordeli, E., Lambert, S., Bennett, V., 1995. AnkyrinG. A new ankyrin gene with neural-specific isoforms localized at the axonal initial segment and node of Ranvier. *J Biol Chem* 270, 2352-2359.
- Korostishevsky, M., Kaganovich, M., Cholostoy, A., Ashkenazi, M., Ratner, Y., Dahary, D., Bernstein, J., Bening-Abu-Shach, U., Ben-Asher, E., Lancet, D., Ritsner, M., Navon, R., 2004. Is the G72/G30 locus associated with schizophrenia? single nucleotide polymorphisms, haplotypes, and gene expression analysis. *Biol Psychiatry* 56, 169-176.
- Kraft, P., Hunter, D.J., 2009. Genetic risk prediction--are we there yet? *N Engl J Med* 360, 1701-1703.

- Kringlen, E., Cramer, G., 1989. Offspring of monozygotic twins discordant for schizophrenia. *Arch Gen Psychiatry* 46, 873-877.
- Krueger, R.F., 1999. The structure of common mental disorders. *Arch Gen Psychiatry* 56, 921-926.
- Krug, A., Krach, S., Jansen, A., Nieratschker, V., Witt, S.H., Shah, N.J., Nothen, M.M., Rietschel, M., Kircher, T., 2011. The Effect of Neurogranin on Neural Correlates of Episodic Memory Encoding and Retrieval. *Schizophr Bull* doi: 10.1093/schbul/sbr076.
- Krug, A., Nieratschker, V., Markov, V., Krach, S., Jansen, A., Zerres, K., Eggermann, T., Stocker, T., Shah, N.J., Treutlein, J., Muhleisen, T.W., Kircher, T., 2010. Effect of CACNA1C rs1006737 on neural correlates of verbal fluency in healthy individuals. *Neuroimage* 49, 1831-1836.
- Kryukov, G.V., Pennacchio, L.A., Sunyaev, S.R., 2007. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am J Hum Genet* 80, 727-739.
- Kumar, P., Henikoff, S., Ng, P.C., 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 4, 1073-1081.
- Lahti, A.C., Weiler, M.A., Tamara Michaelidis, B.A., Parwani, A., Tamminga, C.A., 2001. Effects of ketamine in normal and schizophrenic volunteers. *Neuropsychopharmacology* 25, 455-467.
- Lander, E.S., 1996. The new genomics: global views of biology. *Science* 274, 536-539.
- Laursen, T.M., Munk-Olsen, T., 2010. Reproductive patterns in psychotic patients. *Schizophr Res* 121, 234-240.
- Lawrence, R., Day-Williams, A.G., Elliott, K.S., Morris, A.P., Zeggini, E., 2010. CCRaVAT and QuTie-enabling analysis of rare variants in large-scale case control and quantitative trait association studies. *BMC Bioinformatics* 11, 527.
- Lawrie, S.M., Hall, J., McIntosh, A.M., Owens, D.G., Johnstone, E.C., 2010. The 'continuum of psychosis': scientifically unproven and clinically impractical. *Br J Psychiatry* 197, 423-425.
- Lencz, T., Morgan, T.V., Athanasiou, M., Dain, B., Reed, C.R., Kane, J.M., Kucherlapati, R., Malhotra, A.K., 2007. Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. *Mol Psychiatry* 12, 572-580.
- Lennertz, L., Quednow, B.B., Benninghoff, J., Wagner, M., Maier, W., Mossner, R., 2011. Impact of TCF4 on the genetics of schizophrenia. *Eur Arch Psychiatry Clin Neurosci* 261 Suppl 2, S161-165.
- Leung, A., Chue, P., 2000. Sex differences in schizophrenia, a review of the literature. *Acta Psychiatr Scand Suppl* 401, 3-38.
- Levinson, D.F., Duan, J., Oh, S., Wang, K., Sanders, A.R., Shi, J., Zhang, N., Mowry, B.J., Olincy, A., Amin, F., Cloninger, C.R., Silverman, J.M., Buccola, N.G., Byerley, W.F., Black, D.W., Kendler, K.S., Freedman, R., Dudbridge, F., Pe'er, I., Hakonarson, H., Bergen, S.E., Fanous, A.H., Holmans, P.A., Gejman,

- P.V., 2011. Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. *Am J Psychiatry* 168, 302-316.
- Lewis, C.M., Levinson, D.F., Wise, L.H., DeLisi, L.E., Straub, R.E., Hovatta, I., Williams, N.M., Schwab, S.G., Pulver, A.E., Faraone, S.V., Brzustowicz, L.M., Kaufmann, C.A., Garver, D.L., Gurling, H.M., Lindholm, E., Coon, H., Moises, H.W., Byerley, W., Shaw, S.H., Mesen, A., Sherrington, R., O'Neill, F.A., Walsh, D., Kendler, K.S., Ekelund, J., Paunio, T., Lonnqvist, J., Peltonen, L., O'Donovan, M.C., Owen, M.J., Wildenauer, D.B., Maier, W., Nestadt, G., Blouin, J.L., Antonarakis, S.E., Mowry, B.J., Silverman, J.M., Crowe, R.R., Cloninger, C.R., Tsuang, M.T., Malaspina, D., Harkavy-Friedman, J.M., Svrakic, D.M., Bassett, A.S., Holcomb, J., Kalsi, G., McQuillin, A., Brynjolfson, J., Sigmundsson, T., Petursson, H., Jazin, E., Zoega, T., Helgason, T., 2003. Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. *Am J Hum Genet* 73, 34-48.
- Li, B., Leal, S.M., 2008. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet* 83, 311-321.
- Li, J., Liu, J., Feng, G., Li, T., Zhao, Q., Li, Y., Hu, Z., Zheng, L., Zeng, Z., He, L., Wang, T., Shi, Y., 2011. The MDGA1 gene confers risk to schizophrenia and bipolar disorder. *Schizophr Res* 125, 194-200.
- Li, L.Y., Liu, M.Y., Shih, H.M., Tsai, C.H., Chen, J.Y., 2006. Human cellular protein VRK2 interacts specifically with Epstein-Barr virus BHRF1, a homologue of Bcl-2, and enhances cell survival. *J Gen Virol* 87, 2869-2878.
- Li, Y., Vinckenbosch, N., Tian, G., Huerta-Sanchez, E., Jiang, T., Jiang, H., Albrechtsen, A., Andersen, G., Cao, H., Korneliussen, T., Grarup, N., Guo, Y., Hellman, I., Jin, X., Li, Q., Liu, J., Liu, X., Sparso, T., Tang, M., Wu, H., Wu, R., Yu, C., Zheng, H., Astrup, A., Bolund, L., Holmkvist, J., Jorgensen, T., Kristiansen, K., Schmitz, O., Schwartz, T.W., Zhang, X., Li, R., Yang, H., Wang, J., Hansen, T., Pedersen, O., Nielsen, R., 2010. Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nat Genet* 42, 969-972.
- Lichtenstein, P., Yip, B.H., Bjork, C., Pawitan, Y., Cannon, T.D., Sullivan, P.F., Hultman, C.M., 2009. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* 373, 234-239.
- Lieberman, J.A., 2006. Neurobiology and the natural history of schizophrenia. *J Clin Psychiatry* 67, e14.
- Lieberman, J.A., Perkins, D., Belger, A., Chakos, M., Jarskog, F., Boteva, K., Gilmore, J., 2001. The early stages of schizophrenia: speculations on pathogenesis, pathophysiology, and therapeutic approaches. *Biol Psychiatry* 50, 884-897.
- Lieberman, J.A., Stroup, T.S., McEvoy, J.P., Swartz, M.S., Rosenheck, R.A., Perkins, D.O., Keefe, R.S., Davis, S.M., Davis, C.E., Lebowitz, B.D., Severe, J., Hsiao, J.K., 2005. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N Engl J Med* 353, 1209-1223.

- Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M.F., Parker, B.J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E., Ward, L.D., Lowe, C.B., Holloway, A.K., Clamp, M., Gnerre, S., Alföldi, J., Beal, K., Chang, J., Clawson, H., Cuff, J., Di Palma, F., Fitzgerald, S., Flicek, P., Guttman, M., Hubisz, M.J., Jaffe, D.B., Jungreis, I., Kent, W.J., Kostka, D., Lara, M., Martins, A.L., Massingham, T., Moltke, I., Raney, B.J., Rasmussen, M.D., Robinson, J., Stark, A., Vilella, A.J., Wen, J., Xie, X., Zody, M.C., Baldwin, J., Bloom, T., Chin, C.W., Heiman, D., Nicol, R., Nusbaum, C., Young, S., Wilkinson, J., Worley, K.C., Kovar, C.L., Muzny, D.M., Gibbs, R.A., Cree, A., Dihn, H.H., Fowler, G., Jhangiani, S., Joshi, V., Lee, S., Lewis, L.R., Nazareth, L.V., Okwuonu, G., Santibanez, J., Warren, W.C., Mardis, E.R., Weinstock, G.M., Wilson, R.K., Delehaunty, K., Dooling, D., Fronik, C., Fulton, L., Fulton, B., Graves, T., Minx, P., Sodergren, E., Birney, E., Margulies, E.H., Herrero, J., Green, E.D., Haussler, D., Siepel, A., Goldman, N., Pollard, K.S., Pedersen, J.S., Lander, E.S., Kellis, M., 2011. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478, 476-482.
- Lo, W.S., Xu, Z., Yu, Z., Pun, F.W., Ng, S.K., Chen, J., Tong, K.L., Zhao, C., Xu, X., Tsang, S.Y., Harano, M., Stober, G., Nimgaonkar, V.L., Xue, H., 2007. Positive selection within the Schizophrenia-associated GABA(A) receptor beta(2) gene. *PLoS One* 2, e462.
- Luo, X.J., Diao, H.B., Wang, J.K., Zhang, H., Zhao, Z.M., Su, B., 2008. Association of haplotypes spanning PDZ-GEF2, LOC728637 and ACSL6 with schizophrenia in Han Chinese. *J Med Genet* 45, 818-826.
- Lupski, J.R., 2009. Genomic disorders ten years on. *Genome Med* 1, 42.
- Ma, J., Sun, J., Zhang, H., Zhang, R., Kang, W.H., Gao, C.G., Liu, H.S., Ma, X.H., Min, Z.X., Zhao, W.X., Ning, Q.L., Wang, S.H., Zhang, Y.C., Guo, T.W., Lu, S.M., 2009. Evidence for transmission disequilibrium at the DAOA gene locus in a schizophrenia family sample. *Neurosci Lett* 462, 105-108.
- Ma, X., Deng, W., Liu, X., Li, M., Chen, Z., He, Z., Wang, Y., Wang, Q., Hu, X., Collier, D.A., Li, T., 2011. A genome-wide association study for quantitative traits in schizophrenia in China. *Genes Brain Behav* 10, 734-739.
- Macgregor, S., Visscher, P.M., Knott, S.A., Thomson, P., Porteous, D.J., Millar, J.K., Devon, R.S., Blackwood, D., Muir, W.J., 2004. A genome scan and follow-up study identify a bipolar disorder susceptibility locus on chromosome 1q42. *Mol Psychiatry* 9, 1083-1090.
- MacIntyre, D.J., Blackwood, D.H., Porteous, D.J., Pickard, B.S., Muir, W.J., 2003. Chromosomal abnormalities and mental illness. *Mol Psychiatry* 8, 275-287.
- Magri, C., Sacchetti, E., Traversa, M., Valsecchi, P., Gardella, R., Bonvicini, C., Minelli, A., Gennarelli, M., Barlati, S., 2010. New copy number variations in schizophrenia. *PLoS One* 5, e13422.
- Manning, G., Whyte, D.B., Martinez, R., Hunter, T., Sudarsanam, S., 2002. The protein kinase complement of the human genome. *Science* 298, 1912-1934.
- Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin,

- M., Valle, D., Whittemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F., McCarroll, S.A., Visscher, P.M., 2009. Finding the missing heritability of complex diseases. *Nature* 461, 747-753.
- Mas, S., Bernardo, M., Parellada, E., Garcia-Rizo, C., Gasso, P., Alvarez, S., Lafuente, A., 2009. ARVCF single marker and haplotypic association with schizophrenia. *Prog Neuropsychopharmacol Biol Psychiatry* 33, 1064-1069.
- Maynard, T.M., Haskell, G.T., Peters, A.Z., Sikich, L., Lieberman, J.A., LaMantia, A.S., 2003. A comprehensive analysis of 22q11 gene expression in the developing and adult brain. *Proc Natl Acad Sci U S A* 100, 14433-14438.
- McCarroll, S.A., 2010. Copy number variation and human genome maps. *Nat Genet* 42, 365-366.
- McClellan, J., King, M.C., 2010. Genomic analysis of mental illness: a changing landscape. *JAMA* 303, 2523-2524.
- McClellan, J.M., Susser, E., King, M.C., 2007. Schizophrenia: a common disease caused by multiple rare alleles. *Br J Psychiatry* 190, 194-199.
- McGhee, K.A., Morris, D.W., Schwaiger, S., Nangle, J.M., Donohoe, G., Clarke, S., Meagher, D., Quinn, J., Scully, P., Waddington, J.L., Gill, M., Corvin, A., 2005. Investigation of the apolipoprotein-L (APOL) gene family and schizophrenia using a novel DNA pooling strategy for public database SNPs. *Schizophr Res* 76, 231-238.
- McGrath, J., Saha, S., Chant, D., Welham, J., 2008. Schizophrenia: a concise overview of incidence, prevalence, and mortality. *Epidemiol Rev* 30, 67-76.
- Meador-Woodruff, J.H., Haroutunian, V., Powchik, P., Davidson, M., Davis, K.L., Watson, S.J., 1997. Dopamine receptor transcript expression in striatum and prefrontal and occipital cortex. Focal abnormalities in orbitofrontal cortex in schizophrenia. *Arch Gen Psychiatry* 54, 1089-1095.
- Messias, E.L., Chen, C.Y., Eaton, W.W., 2007. Epidemiology of schizophrenia: review of findings and myths. *Psychiatr Clin North Am* 30, 323-338.
- Meyer-Lindenberg, A., 2010. From maps to mechanisms through neuroimaging of schizophrenia. *Nature* 468, 194-202.
- Meyer-Lindenberg, A., Straub, R.E., Lipska, B.K., Verchinski, B.A., Goldberg, T., Callicott, J.H., Egan, M.F., Huffaker, S.S., Mattay, V.S., Kolachana, B., Kleinman, J.E., Weinberger, D.R., 2007. Genetic evidence implicating DARPP-32 in human frontostriatal structure, function, and cognition. *J Clin Invest* 117, 672-682.
- Meyer-Lindenberg, A.S., Olsen, R.K., Kohn, P.D., Brown, T., Egan, M.F., Weinberger, D.R., Berman, K.F., 2005. Regionally specific disturbance of dorsolateral prefrontal-hippocampal functional connectivity in schizophrenia. *Arch Gen Psychiatry* 62, 379-386.
- Millar, J.K., Pickard, B.S., Mackie, S., James, R., Christie, S., Buchanan, S.R., Malloy, M.P., Chubb, J.E., Huston, E., Baillie, G.S., Thomson, P.A., Hill, E.V., Brandon, N.J., Rain, J.C., Camargo, L.M., Whiting, P.J., Houslay, M.D., Blackwood, D.H., Muir, W.J., Porteous, D.J., 2005. DISC1 and PDE4B are

- interacting genetic factors in schizophrenia that regulate cAMP signaling. *Science* 310, 1187-1191.
- Millar, J.K., Wilson-Annan, J.C., Anderson, S., Christie, S., Taylor, M.S., Semple, C.A., Devon, R.S., St Clair, D.M., Muir, W.J., Blackwood, D.H., Porteous, D.J., 2000. Disruption of two novel genes by a translocation co-segregating with schizophrenia. *Hum Mol Genet* 9, 1415-1423.
- Mimmack, M.L., Ryan, M., Baba, H., Navarro-Ruiz, J., Iritani, S., Faull, R.L., McKenna, P.J., Jones, P.B., Arai, H., Starkey, M., Emson, P.C., Bahn, S., 2002. Gene expression analysis in schizophrenia: reproducible up-regulation of several members of the apolipoprotein L family located in a high-susceptibility locus for schizophrenia on chromosome 22. *Proc Natl Acad Sci U S A* 99, 4680-4685.
- Mitchell, K.J., Porteous, D.J., 2011. Rethinking the genetic architecture of schizophrenia. *Psychol Med* 41, 19-32.
- 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, 685-694.
- Moises, H.W., Zoega, T., Gottesman, II, 2002. The glial growth factors deficiency and synaptic destabilization hypothesis of schizophrenia. *BMC Psychiatry* 2, 8.
- Morris, A.P., Zeggini, E., 2009. An evaluation of statistical approaches to rare variant analysis in genetic association studies. *Genet Epidemiol* 34, 188-193.
- Muir, W.J., Pickard, B.S., Blackwood, D.H., 2006. Chromosomal abnormalities and psychosis. *Br J Psychiatry* 188, 501-503.
- Mulle, J.G., Dodd, A.F., McGrath, J.A., Wolyniec, P.S., Mitchell, A.A., Shetty, A.C., Sobreira, N.L., Valle, D., Rudd, M.K., Satten, G., Cutler, D.J., Pulver, A.E., Warren, S.T., 2010. Microdeletions of 3q29 confer high risk for schizophrenia. *Am J Hum Genet* 87, 229-236.
- Murphy, K.C., Jones, L.A., Owen, M.J., 1999. High rates of schizophrenia in adults with velo-cardio-facial syndrome. *Arch Gen Psychiatry* 56, 940-945.
- Murray, R.M., Lewis, S.W., 1987. Is schizophrenia a neurodevelopmental disorder? *Br Med J (Clin Res Ed)* 295, 681-682.
- Mutsuddi, M., Morris, D.W., Waggoner, S.G., Daly, M.J., Scolnick, E.M., Sklar, P., 2006. Analysis of high-resolution HapMap of DTNBP1 (Dysbindin) suggests no consistency between reported common variant associations and schizophrenia. *Am J Hum Genet* 79, 903-909.
- Nakajima, T., Wooding, S., Sakagami, T., Emi, M., Tokunaga, K., Tamiya, G., Ishigami, T., Umemura, S., Munkhbat, B., Jin, F., Guan-Jun, J., Hayasaka, I., Ishida, T., Saitou, N., Pavelka, K., Lalouel, J.M., Jorde, L.B., Inoue, I., 2004. Natural selection and population history in the human angiotensinogen gene (AGT): 736 complete AGT sequences in chromosomes from around the world. *Am J Hum Genet* 74, 898-916.
- Need, A.C., Ge, D., Weale, M.E., Maia, J., Feng, S., Heinzen, E.L., Shianna, K.V., Yoon, W., Kasperaviciute, D., Gennarelli, M., Strittmatter, W.J., Bonvicini, C.,

- Rossi, G., Jayathilake, K., Cola, P.A., McEvoy, J.P., Keefe, R.S., Fisher, E.M., St Jean, P.L., Giegling, I., Hartmann, A.M., Moller, H.J., Ruppert, A., Fraser, G., Crombie, C., Middleton, L.T., St Clair, D., Roses, A.D., Muglia, P., Francks, C., Rujescu, D., Meltzer, H.Y., Goldstein, D.B., 2009. A genome-wide investigation of SNPs and CNVs in schizophrenia. *PLoS Genet* 5, e1000373.
- Neel, J.V., 1962. Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"? *Am J Hum Genet* 14, 353-362.
- Nettle, D., Clegg, H., 2006. Schizotypy, creativity and mating success in humans. *Proc Biol Sci* 273, 611-615.
- Ng, M.Y., Levinson, D.F., Faraone, S.V., Suarez, B.K., DeLisi, L.E., Arinami, T., Riley, B., Paunio, T., Pulver, A.E., Irmansyah, Holmans, P.A., Escamilla, M., Wildenauer, D.B., Williams, N.M., Laurent, C., Mowry, B.J., Brzustowicz, L.M., Maziade, M., Sklar, P., Garver, D.L., Abecasis, G.R., Lerer, B., Fallin, M.D., Gurling, H.M., Gejman, P.V., Lindholm, E., Moises, H.W., Byerley, W., Wijsman, E.M., Forabosco, P., Tsuang, M.T., Hwu, H.G., Okazaki, Y., Kendler, K.S., Wormley, B., Fanous, A., Walsh, D., O'Neill, F.A., Peltonen, L., Nestadt, G., Lasseter, V.K., Liang, K.Y., Papadimitriou, G.M., Dikeos, D.G., Schwab, S.G., Owen, M.J., O'Donovan, M.C., Norton, N., Hare, E., Raventos, H., Nicolini, H., Albus, M., Maier, W., Nimgaonkar, V.L., Terenius, L., Mallet, J., Jay, M., Godard, S., Nertney, D., Alexander, M., Crowe, R.R., Silverman, J.M., Bassett, A.S., Roy, M.A., Merette, C., Pato, C.N., Pato, M.T., Roos, J.L., Kohn, Y., Amann-Zalcenstein, D., Kalsi, G., McQuillin, A., Curtis, D., Brynjolfson, J., Sigmundsson, T., Petursson, H., Sanders, A.R., Duan, J., Jazin, E., Myles-Worsley, M., Karayiorgou, M., Lewis, C.M., 2009. Meta-analysis of 32 genome-wide linkage studies of schizophrenia. *Mol Psychiatry* 14, 774-785.
- Nichols, R.J., Wiebe, M.S., Traktman, P., 2006. The vaccinia-related kinases phosphorylate the N' terminus of BAF, regulating its interaction with DNA and its retention in the nucleus. *Mol Biol Cell* 17, 2451-2464.
- Nunokawa, A., Watanabe, Y., Kaneko, N., Sugai, T., Yazaki, S., Arinami, T., Ujike, H., Inada, T., Iwata, N., Kunugi, H., Sasaki, T., Itokawa, M., Ozaki, N., Hashimoto, R., Someya, T., 2009. The dopamine D3 receptor (DRD3) gene and risk of schizophrenia: case-control studies and an updated meta-analysis. *Schizophr Res* 116, 61-67.
- Nyegaard, M., Demontis, D., Foldager, L., Hedemand, A., Flint, T.J., Sorensen, K.M., Andersen, P.S., Nordentoft, M., Werge, T., Pedersen, C.B., Hougaard, D.M., Mortensen, P.B., Mors, O., Borglum, A.D., 2010. CACNA1C (rs1006737) is associated with schizophrenia. *Mol Psychiatry* 15, 119-121.
- O'Donovan, M.C., Craddock, N., Norton, N., Williams, H., Peirce, T., Moskvina, V., Nikolov, I., Hamshere, M., Carroll, L., Georgieva, L., Dwyer, S., Holmans, P., Marchini, J.L., Spencer, C.C., Howie, B., Leung, H.T., Hartmann, A.M., Moller, H.J., Morris, D.W., Shi, Y., Feng, G., Hoffmann, P., Propping, P., Vasilescu, C., Maier, W., Rietschel, M., Zammit, S., Schumacher, J., Quinn, E.M., Schulze, T.G., Williams, N.M., Giegling, I., Iwata, N., Ikeda, M., Darvasi, A., Shifman, S., He, L., Duan, J., Sanders, A.R., Levinson, D.F., Gejman, P.V., Cichon, S., Nothen, M.M., Gill, M., Corvin, A., Rujescu, D.,

- Kirov, G., Owen, M.J., Buccola, N.G., Mowry, B.J., Freedman, R., Amin, F., Black, D.W., Silverman, J.M., Byerley, W.F., Cloninger, C.R., 2008. Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet* 40, 1053-1055.
- Ohi, K., Hashimoto, R., Yasuda, Y., Yoshida, T., Takahashi, H., Iike, N., Fukumoto, M., Takamura, H., Iwase, M., Kamino, K., Ishii, R., Kazui, H., Sekiyama, R., Kitamura, Y., Azechi, M., Ikezawa, K., Kurimoto, R., Kamagata, E., Tanimukai, H., Tagami, S., Morihara, T., Ogasawara, M., Okochi, M., Tokunaga, H., Numata, S., Ikeda, M., Ohnuma, T., Ueno, S., Fukunaga, T., Tanaka, T., Kudo, T., Arai, H., Ohmori, T., Iwata, N., Ozaki, N., Takeda, M., 2009. Association study of the G72 gene with schizophrenia in a Japanese population: a multicenter study. *Schizophr Res* 109, 80-85.
- Opgen-Rhein, C., Lencz, T., Burdick, K.E., Neuhaus, A.H., DeRosse, P., Goldberg, T.E., Malhotra, A.K., 2008. Genetic variation in the DAOA gene complex: impact on susceptibility for schizophrenia and on cognitive performance. *Schizophr Res* 103, 169-177.
- Owen, M.J., Craddock, N., O'Donovan, M.C., 2005. Schizophrenia: genes at last? *Trends Genet* 21, 518-525.
- Owen, M.J., O'Donovan, M.C., Thapar, A., Craddock, N., 2011. Neurodevelopmental hypothesis of schizophrenia. *Br J Psychiatry* 198, 173-175.
- Owens, S.F., Picchioni, M.M., Rijdsdijk, F.V., Stahl, D., Vassos, E., Rodger, A.K., Collier, D.A., Murray, R.M., Toulopoulou, T., 2011. Genetic overlap between episodic memory deficits and schizophrenia: results from the Maudsley Twin Study. *Psychol Med* 41, 521-532.
- Ozeki, Y., Tomoda, T., Kleiderlein, J., Kamiya, A., Bord, L., Fujii, K., Okawa, M., Yamada, N., Hatten, M.E., Snyder, S.H., Ross, C.A., Sawa, A., 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, 289-294.
- Ozonoff, S., Williams, B.J., Gale, S., Miller, J.N., 1999. Autism and autistic behavior in Joubert syndrome. *J Child Neurol* 14, 636-641.
- Palmer, B.A., Pankratz, V.S., Bostwick, J.M., 2005. The lifetime risk of suicide in schizophrenia: a reexamination. *Arch Gen Psychiatry* 62, 247-253.
- Pani, L., 2002. Clinical implications of dopamine research in schizophrenia. *Curr Med Res Opin* 18 Suppl 3, s3-7.
- Perez-Morga, D., Vanhollebeke, B., Paturiaux-Hanocq, F., Nolan, D.P., Lins, L., Homble, F., Vanhamme, L., Tebabi, P., Pays, A., Poelvoorde, P., Jacquet, A., Brasseur, R., Pays, E., 2005. Apolipoprotein L-I promotes trypanosome lysis by forming pores in lysosomal membranes. *Science* 309, 469-472.
- Petronis, A., 2004. The origin of schizophrenia: genetic thesis, epigenetic antithesis, and resolving synthesis. *Biol Psychiatry* 55, 965-970.
- Piton, A., Gauthier, J., Hamdan, F.F., Lafreniere, R.G., Yang, Y., Henrion, E., Laurent, S., Noreau, A., Thibodeau, P., Karemera, L., Spiegelman, D., Kuku, F., Duguay, J., Destroismaisons, L., Jolivet, P., Cote, M., Lachapelle, K., Diallo, O., Raymond, A., Marineau, C., Champagne, N., Xiong, L., Gaspar, C.,

- Riviere, J.B., Tarabeux, J., Cossette, P., Krebs, M.O., Rapoport, J.L., Addington, A., Delisi, L.E., Mottron, L., Joober, R., Fombonne, E., Drapeau, P., Rouleau, G.A., 2011. Systematic resequencing of X-chromosome synaptic genes in autism spectrum disorder and schizophrenia. *Mol Psychiatry* 16, 867-880.
- Poliak, S., Peles, E., 2003. The local differentiation of myelinated axons at nodes of Ranvier. *Nat Rev Neurosci* 4, 968-980.
- Price, A.L., Helgason, A., Palsson, S., Stefansson, H., St Clair, D., Andreassen, O.A., Reich, D., Kong, A., Stefansson, K., 2009. The impact of divergence time on the nature of population structure: an example from Iceland. *PLoS Genet* 5, e1000505.
- Price, A.L., Kryukov, G.V., de Bakker, P.I., Purcell, S.M., Staples, J., Wei, L.J., Sunyaev, S.R., 2010. Pooled association tests for rare variants in exon-resequencing studies. *Am J Hum Genet* 86, 832-838.
- Pritchard, J.K., 2001. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* 69, 124-137.
- Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics* 155, 945-959.
- Raznahan, A., Lee, Y., Long, R., Greenstein, D., Clasen, L., Addington, A., Rapoport, J.L., Giedd, J.N., 2010. Common functional polymorphisms of DISC1 and cortical maturation in typically developing children and adolescents. *Mol Psychiatry* 16, 917-926.
- Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., Gonzalez, J.R., Gratacos, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Armengol, L., Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W., Hurles, M.E., 2006. Global variation in copy number in the human genome. *Nature* 444, 444-454.
- Rees, E., Moskvina, V., Owen, M.J., O'Donovan, M.C., Kirov, G., 2011. De novo rates and selection of schizophrenia-associated copy number variants. *Biol Psychiatry* 70, 1109-1114.
- Reich, D.E., Lander, E.S., 2001. On the allelic spectrum of human disease. *Trends Genet* 17, 502-510.
- Riecher-Rossler, A., Hafner, H., 2000. Gender aspects in schizophrenia: bridging the border between social and biological psychiatry. *Acta Psychiatr Scand Suppl*, 58-62.
- Rietschel, M., Mattheisen, M., Degenhardt, F., Kahn, R.S., Linszen, D.H., Os, J.V., Wiersma, D., Bruggeman, R., Cahn, W., de Haan, L., Krabbendam, L., Myin-Germeys, I., Muhleisen, T.W., Kirsch, P., Esslinger, C., Herms, S., Demontis, D., Steffens, M., Strohmaier, J., Haenisch, B., Breuer, R., Czerski, P.M., Giegling, I., Strengman, E., Schmael, C., Mors, O., Mortensen, P.B., Hougaard, D.M., Orntoft, T., Kapelski, P., Priebe, L., Basmanav, F.B.,

- Forstner, A.J., Hoffmann, P., Meier, S., Nikitopoulos, J., Moebus, S., Alexander, M., Mossner, R., Wichmann, H.E., Schreiber, S., Rivandeneira, F., Hofman, A., Uitterlinden, A.G., Wienker, T.F., Schumacher, J., Hauser, J., Maier, W., Cantor, R.M., Erk, S., Schulze, T.G., Stefansson, H., Steinberg, S., Gustafsson, O., Sigurdsson, E., Petursson, H., Kong, A., Stefansson, K., Pietilainen, O.P., Tuulio-Henriksson, A., Paunio, T., Lonnqvist, J., Suvisaari, J., Peltonen, L., Ruggeri, M., Tosato, S., Walshe, M., Murray, R., Collier, D.A., Clair, D.S., Hansen, T., Ingason, A., Jakobsen, K.D., Duong, L., Werge, T., Melle, I., Andreassen, O.A., Djurovic, S., Bitter, I., Rethelyi, J.M., Abramova, L., Kaleda, V., Golimbet, V., Jonsson, E.G., Terenius, L., Agartz, I., Winkel, R.V., Kenis, G., Hert, M.D., Veldink, J., Wiuf, C., Didriksen, M., Craddock, N., Owen, M.J., O'Donovan, M.C., Borglum, A.D., Rujescu, D., Walter, H., Meyer-Lindenberg, A., Nothen, M.M., Ophoff, R.A., Cichon, S., 2011. Association between genetic variation in a region on chromosome 11 and schizophrenia in large samples from Europe. *Mol Psychiatry* doi: 10.1038/mp.2011.80.
- Riley, B., Kendler, K.S., 2006. Molecular genetic studies of schizophrenia. *Eur J Hum Genet* 14, 669-680.
- Riley, B., Thiselton, D., Maher, B.S., Bigdeli, T., Wormley, B., McMichael, G.O., Fanous, A.H., Vladimirov, V., O'Neill, F.A., Walsh, D., Kendler, K.S., 2010. Replication of association between schizophrenia and ZNF804A in the Irish Case-Control Study of Schizophrenia sample. *Mol Psychiatry* 15, 29-37.
- Ripke, S., Sanders, A.R., Kendler, K.S., Levinson, D.F., Sklar, P., Holmans, P.A., Lin, D.Y., Duan, J., Ophoff, R.A., Andreassen, O.A., Scolnick, E., Cichon, S., St Clair, D., Corvin, A., Gurling, H., Werge, T., Rujescu, D., Blackwood, D.H., Pato, C.N., Malhotra, A.K., Purcell, S., Dudbridge, F., Neale, B.M., Rossin, L., Visscher, P.M., Posthuma, D., Ruderfer, D.M., Fanous, A., Stefansson, H., Steinberg, S., Mowry, B.J., Golimbet, V., De Hert, M., Jonsson, E.G., Bitter, I., Pietilainen, O.P., Collier, D.A., Tosato, S., Agartz, I., Albus, M., Alexander, M., Amdur, R.L., Amin, F., Bass, N., Bergen, S.E., Black, D.W., Borglum, A.D., Brown, M.A., Bruggeman, R., Buccola, N.G., Byerley, W.F., Cahn, W., Cantor, R.M., Carr, V.J., Catts, S.V., Choudhury, K., Cloninger, C.R., Cormican, P., Craddock, N., Danoy, P.A., Datta, S., de Haan, L., Demontis, D., Dikeos, D., Djurovic, S., Donnelly, P., Donohoe, G., Duong, L., Dwyer, S., Fink-Jensen, A., Freedman, R., Freimer, N.B., Friedl, M., Georgieva, L., Giegling, I., Gill, M., Glenthøj, B., Godard, S., Hamshere, M., Hansen, M., Hansen, T., Hartmann, A.M., Henskens, F.A., Hougaard, D.M., Hultman, C.M., Ingason, A., Jablensky, A.V., Jakobsen, K.D., Jay, M., Jurgens, G., Kahn, R.S., Keller, M.C., Kenis, G., Kenny, E., Kim, Y., Kirov, G.K., Konnerth, H., Konte, B., Krabbendam, L., Krasucki, R., Lasseter, V.K., Laurent, C., Lawrence, J., Lencz, T., Lerer, F.B., Liang, K.Y., Lichtenstein, P., Lieberman, J.A., Linszen, D.H., Lonnqvist, J., Loughland, C.M., Maclean, A.W., Maher, B.S., Maier, W., Mallet, J., Malloy, P., Mattheisen, M., Mattingsdal, M., McGhee, K.A., McGrath, J.J., McIntosh, A., McLean, D.E., McQuillin, A., Melle, I., Michie, P.T., Milanova, V., Morris, D.W., Mors, O., Mortensen, P.B., Moskvina, V., Muglia, P., Myin-Germeys, I., Nertney, D.A., Nestadt, G., Nielsen, J., Nikolov, I., Nordentoft, M., Norton, N., Nothen, M.M., O'Dushlaine, C.T., Olincy, A., Olsen, L., O'Neill, F.A., Orntoft, T.F.,

- Owen, M.J., Pantelis, C., Papadimitriou, G., Pato, M.T., Peltonen, L., Petursson, H., Pickard, B., Pimm, J., Pulver, A.E., Puri, V., Quedsted, D., Quinn, E.M., Rasmussen, H.B., Rethelyi, J.M., Ribble, R., Rietschel, M., Riley, B.P., Ruggeri, M., Schall, U., Schulze, T.G., Schwab, S.G., Scott, R.J., Shi, J., Sigurdsson, E., Silverman, J.M., Spencer, C.C., Stefansson, K., Strange, A., Strengman, E., Stroup, T.S., Suvisaari, J., Terenius, L., Thirumalai, S., Thygesen, J.H., Timm, S., Toncheva, D., van den Oord, E., van Os, J., van Winkel, R., Veldink, J., Walsh, D., Wang, A.G., Wiersma, D., Wildenauer, D.B., Williams, H.J., Williams, N.M., Wormley, B., Zammit, S., Sullivan, P.F., O'Donovan, M.C., Daly, M.J., Gejman, P.V., 2011. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet* 43, 969-976.
- Risch, N., 1990a. Genetic linkage and complex diseases, with special reference to psychiatric disorders. *Genet Epidemiol* 7, 3-16; discussion 17-45.
- Risch, N., 1990b. Linkage strategies for genetically complex traits. I. Multilocus models. *Am J Hum Genet* 46, 222-228.
- Risch, N., Merikangas, K., 1996. The future of genetic studies of complex human diseases. *Science* 273, 1516-1517.
- Risch, N.J., 2000. Searching for genetic determinants in the new millennium. *Nature* 405, 847-856.
- Rivas, M.A., Beaudoin, M., Gardet, A., Stevens, C., Sharma, Y., Zhang, C.K., Boucher, G., Ripke, S., Ellinghaus, D., Burt, N., Fennell, T., Kirby, A., Latiano, A., Goyette, P., Green, T., Halfvarson, J., Haritunians, T., Korn, J.M., Kuruvilla, F., Lagace, C., Neale, B., Lo, K.S., Schumm, P., Torkvist, L., Dubinsky, M.C., Brant, S.R., Silverberg, M.S., Duerr, R.H., Altshuler, D., Gabriel, S., Lettre, G., Franke, A., D'Amato, M., McGovern, D.P., Cho, J.H., Rioux, J.D., Xavier, R.J., Daly, M.J., 2011. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat Genet* 43, 1066-1073.
- Rivero, O., Reif, A., Sanjuan, J., Molto, M.D., Kittel-Schneider, S., Najera, C., Topner, T., Lesch, K.P., 2010. Impact of the AHI1 gene on the vulnerability to schizophrenia: a case-control association study. *PLoS One* 5, e12254.
- Robin, N.H., Shprintzen, R.J., 2005. Defining the clinical spectrum of deletion 22q11.2. *J Pediatr* 147, 90-96.
- Roth, T.L., Lubin, F.D., Sodhi, M., Kleinman, J.E., 2009. Epigenetic mechanisms in schizophrenia. *Biochim Biophys Acta* 1790, 869-877.
- Ruderfer, D.M., Kirov, G., Chambert, K., Moran, J.L., Owen, M.J., O'Donovan, M.C., Sklar, P., Purcell, S.M., 2011. A family-based study of common polygenic variation and risk of schizophrenia. *Mol Psychiatry* 16, 887-888.
- Ryan, A.K., Goodship, J.A., Wilson, D.I., Philip, N., Levy, A., Seidel, H., Schuffenhauer, S., Oechsler, H., Belohradsky, B., Prieur, M., Aurias, A., Raymond, F.L., Clayton-Smith, J., Hatchwell, E., McKeown, C., Beemer, F.A., Dallapiccola, B., Novelli, G., Hurst, J.A., Ignatius, J., Green, A.J., Winter, R.M., Brueton, L., Brondum-Nielsen, K., Scambler, P.J., et al., 1997. Spectrum of clinical features associated with interstitial chromosome 22q11 deletions: a European collaborative study. *J Med Genet* 34, 798-804.

- Sabb, F.W., Bearden, C.E., Glahn, D.C., Parker, D.S., Freimer, N., Bilder, R.M., 2008. A collaborative knowledge base for cognitive phenomics. *Mol Psychiatry* 13, 350-360.
- Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., Ackerman, H.C., Campbell, S.J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., Lander, E.S., 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419, 832-837.
- Sabeti, P.C., Schaffner, S.F., Fry, B., Lohmueller, J., Varily, P., Shamovsky, O., Palma, A., Mikkelsen, T.S., Altshuler, D., Lander, E.S., 2006. Positive natural selection in the human lineage. *Science* 312, 1614-1620.
- Sabeti, P.C., Walsh, E., Schaffner, S.F., Varily, P., Fry, B., Hutcheson, H.B., Cullen, M., Mikkelsen, T.S., Roy, J., Patterson, N., Cooper, R., Reich, D., Altshuler, D., O'Brien, S., Lander, E.S., 2005. The case for selection at CCR5-Delta32. *PLoS Biol* 3, e378.
- 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, 1123-1131.
- Saha, S., Chant, D., Welham, J., McGrath, J., 2005. A systematic review of the prevalence of schizophrenia. *PLoS Med* 2, e141.
- Saha, S., Chant, D.C., Welham, J.L., McGrath, J.J., 2006. The incidence and prevalence of schizophrenia varies with latitude. *Acta Psychiatr Scand* 114, 36-39.
- Salem, J.E., Kring, A.M., 1998. The role of gender differences in the reduction of etiologic heterogeneity in schizophrenia. *Clin Psychol Rev* 18, 795-819.
- Salvatore, P., Baldessarini, R.J., Tohen, M., Khalsa, H.M., Sanchez-Toledo, J.P., Zarate, C.A., Jr., Vieta, E., Maggini, C., 2009. McLean-Harvard International First-Episode Project: two-year stability of DSM-IV diagnoses in 500 first-episode psychotic disorder patients. *J Clin Psychiatry* 70, 458-466.
- Scott, L.J., Muglia, P., Kong, X.Q., Guan, W., Flickinger, M., Upmanyu, R., Tozzi, F., Li, J.Z., Burmeister, M., Absher, D., Thompson, R.C., Francks, C., Meng, F., Antoniadis, A., Southwick, A.M., Schatzberg, A.F., Bunney, W.E., Barchas, J.D., Jones, E.G., Day, R., Matthews, K., McGuffin, P., Strauss, J.S., Kennedy, J.L., Middleton, L., Roses, A.D., Watson, S.J., Vincent, J.B., Myers, R.M., Farmer, A.E., Akil, H., Burns, D.K., Boehnke, M., 2009. Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry. *Proc Natl Acad Sci U S A* 106, 7501-7506.
- Schmauss, C., Haroutunian, V., Davis, K.L., Davidson, M., 1993. Selective loss of dopamine D3-type receptor mRNA expression in parietal and motor cortices of patients with chronic schizophrenia. *Proc Natl Acad Sci U S A* 90, 8942-8946.
- Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M., Navin, N., Lucito, R., Healy, J., Hicks, J., Ye, K., Reiner, A., Gilliam, T.C., Trask, B., Patterson, N., Zetterberg, A., Wigler, M., 2004. Large-scale copy number polymorphism in the human genome. *Science* 305, 525-528.

- Seeman, P., Lee, T., 1975. Antipsychotic drugs: direct correlation between clinical potency and presynaptic action on dopamine neurons. *Science* 188, 1217-1219.
- Selemon, L.D., Goldman-Rakic, P.S., 1999. The reduced neuropil hypothesis: a circuit based model of schizophrenia. *Biol Psychiatry* 45, 17-25.
- Shaikh, T.H., Kurahashi, H., Saitta, S.C., O'Hare, A.M., Hu, P., Roe, B.A., Driscoll, D.A., McDonald-McGinn, D.M., Zackai, E.H., Budarf, M.L., Emanuel, B.S., 2000. Chromosome 22-specific low copy repeats and the 22q11.2 deletion syndrome: genomic organization and deletion endpoint analysis. *Hum Mol Genet* 9, 489-501.
- Shi, J., Levinson, D.F., Duan, J., Sanders, A.R., Zheng, Y., Pe'er, I., Dudbridge, F., Holmans, P.A., Whittemore, A.S., Mowry, B.J., Olincy, A., Amin, F., Cloninger, C.R., Silverman, J.M., Buccola, N.G., Byerley, W.F., Black, D.W., Crowe, R.R., Oksenberg, J.R., Mirel, D.B., Kendler, K.S., Freedman, R., Gejman, P.V., 2009. Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* 460, 753-757.
- Shi, Y.Y., He, G., Zhang, Z., Tang, W., Zhang, J., Jr., Zhao, Q., Zhang, J., Sr., Li, X.W., Xi, Z.R., Fang, C., Zhao, X.Z., Feng, G.Y., He, L., 2008. A study of rare structural variants in schizophrenia patients and normal controls from Chinese Han population. *Mol Psychiatry* 13, 911-913.
- Shibata, H., Shibata, A., Ninomiya, H., Tashiro, N., Fukumaki, Y., 2002. Association study of polymorphisms in the GluR6 kainate receptor gene (GRIK2) with schizophrenia. *Psychiatry Res* 113, 59-67.
- Shifman, S., Johannesson, M., Bronstein, M., Chen, S.X., Collier, D.A., Craddock, N.J., Kendler, K.S., Li, T., O'Donovan, M., O'Neill, F.A., Owen, M.J., Walsh, D., Weinberger, D.R., Sun, C., Flint, J., Darvasi, A., 2008. Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women. *PLoS Genet* 4, e28.
- Shin, H.D., Park, B.L., Kim, E.M., Lee, S.O., Cheong, H.S., Lee, C.H., Kim, S.G., Sohn, J.W., Park, C.S., Kim, J.W., Kim, B.H., Kim, I.Y., Choi, I.G., Woo, S.I., 2007. Association analysis of G72/G30 polymorphisms with schizophrenia in the Korean population. *Schizophr Res* 96, 119-124.
- Siever, L.J., Davis, K.L., 2004. The pathophysiology of schizophrenia disorders: perspectives from the spectrum. *Am J Psychiatry* 161, 398-413.
- Silber, J., Lim, D.A., Petritsch, C., Persson, A.I., Maunakea, A.K., Yu, M., Vandenberg, S.R., Ginzinger, D.G., James, C.D., Costello, J.F., Bergers, G., Weiss, W.A., Alvarez-Buylla, A., Hodgson, J.G., 2008. miR-124 and miR-137 inhibit proliferation of glioblastoma multiforme cells and induce differentiation of brain tumor stem cells. *BMC Med* 6, 14.
- Sklar, P., Ripke, S., Scott, L.J., Andreassen, O.A., Cichon, S., Craddock, N., Edenberg, H.J., Nurnberger, J.I., Jr., Rietschel, M., Blackwood, D., Corvin, A., Flickinger, M., Guan, W., Mattingsdal, M., McQuillin, A., Kwan, P., Wienker, T.F., Daly, M., Dudbridge, F., Holmans, P.A., Lin, D., Burmeister, M., Greenwood, T.A., Hamshere, M.L., Muglia, P., Smith, E.N., Zandi, P.P., Nievergelt, C.M., McKinney, R., Shilling, P.D., Schork, N.J., Bloss, C.S., Foroud, T., Koller, D.L., Gershon, E.S., Liu, C., Badner, J.A., Scheftner, W.A.,

- Lawson, W.B., Nwulia, E.A., Hipolito, M., Coryell, W., Rice, J., Byerley, W., McMahon, F.J., Schulze, T.G., Berrettini, W., Lohoff, F.W., Potash, J.B., Mahon, P.B., McInnis, M.G., Zollner, S., Zhang, P., Craig, D.W., Szelinger, S., Barrett, T.B., Breuer, R., Meier, S., Strohmaier, J., Witt, S.H., Tozzi, F., Farmer, A., McGuffin, P., Strauss, J., Xu, W., Kennedy, J.L., Vincent, J.B., Matthews, K., Day, R., Ferreira, M.A., O'Dushlaine, C., Perlis, R., Raychaudhuri, S., Ruderfer, D., Hyoun, P.L., Smoller, J.W., Li, J., Absher, D., Thompson, R.C., Meng, F.G., Schatzberg, A.F., Bunney, W.E., Barchas, J.D., Jones, E.G., Watson, S.J., Myers, R.M., Akil, H., Boehnke, M., Chambert, K., Moran, J., Scolnick, E., Djurovic, S., Melle, I., Morken, G., Gill, M., Morris, D., Quinn, E., Muhleisen, T.W., Degenhardt, F.A., Mattheisen, M., Schumacher, J., Maier, W., Steffens, M., Propping, P., Nothen, M.M., Anjorin, A., Bass, N., Gurling, H., Kandaswamy, R., Lawrence, J., McGhee, K., McIntosh, A., McLean, A.W., Muir, W.J., Pickard, B.S., Breen, G., St Clair, D., Caesar, S., Gordon-Smith, K., Jones, L., Fraser, C., Green, E.K., Grozeva, D., Jones, I.R., Kirov, G., Moskvina, V., Nikolov, I., O'Donovan, M.C., Owen, M.J., Collier, D.A., Elkin, A., Williamson, R., Young, A.H., Ferrier, I.N., Stefansson, K., Stefansson, H., Thorgeirsson, T., Steinberg, S., Gustafsson, O., Bergen, S.E., Nimgaonkar, V., Hultman, C., Landen, M., Lichtenstein, P., Sullivan, P., Schalling, M., Osby, U., Backlund, L., Frisen, L., Langstrom, N., Jamain, S., Leboyer, M., Etain, B., Bellivier, F., Petursson, H., Sigur Sson, E., Muller-Mysok, B., Lucae, S., Schwarz, M., Schofield, P.R., Martin, N., Montgomery, G.W., Lathrop, M., Oskarsson, H., Bauer, M., Wright, A., Mitchell, P.B., Hautzinger, M., Reif, A., Kelsoe, J.R., Purcell, S.M., 2011. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet* 43, 977-983.
- Smrt, R.D., Szulwach, K.E., Pfeiffer, R.L., Li, X., Guo, W., Pathania, M., Teng, Z.Q., Luo, Y., Peng, J., Bordey, A., Jin, P., Zhao, X., 2010. MicroRNA miR-137 regulates neuronal maturation by targeting ubiquitin ligase mind bomb-1. *Stem Cells* 28, 1060-1070.
- Sokoloff, P., Giros, B., Martres, M.P., Bouthenet, M.L., Schwartz, J.C., 1990. Molecular cloning and characterization of a novel dopamine receptor (D3) as a target for neuroleptics. *Nature* 347, 146-151.
- Song, W., Li, W., Feng, J., Heston, L.L., Scaringe, W.A., Sommer, S.S., 2008. Identification of high risk DISC1 structural variants with a 2% attributable risk for schizophrenia. *Biochem Biophys Res Commun* 367, 700-706.
- St Clair, D., Blackwood, D., Muir, W., Carothers, A., Walker, M., Spowart, G., Gosden, C., Evans, H.J., 1990. Association within a family of a balanced autosomal translocation with major mental illness. *Lancet* 336, 13-16.
- Stefansson, H., Ophoff, R.A., Steinberg, S., Andreassen, O.A., Cichon, S., Rujescu, D., Werge, T., Pietilainen, O.P., Mors, O., Mortensen, P.B., Sigurdsson, E., Gustafsson, O., Nyegaard, M., Tuulio-Henriksson, A., Ingason, A., Hansen, T., Suvisaari, J., Lonnqvist, J., Paunio, T., Borglum, A.D., Hartmann, A., Fink-Jensen, A., Nordentoft, M., Hougaard, D., Norgaard-Pedersen, B., Bottcher, Y., Olesen, J., Breuer, R., Moller, H.J., Giegling, I., Rasmussen, H.B., Timm, S., Mattheisen, M., Bitter, I., Rethelyi, J.M., Magnusdottir, B.B., Sigmundsson, T., Olason, P., Masson, G., Gulcher, J.R., Haraldsson, M., Fossdal, R., Thorgeirsson, T.E., Thorsteinsdottir, U., Ruggeri, M., Tosato, S., Franke, B.,

- Strengman, E., Kiemeny, L.A., Melle, I., Djurovic, S., Abramova, L., Kaleda, V., Sanjuan, J., de Frutos, R., Bramon, E., Vassos, E., Fraser, G., Ettinger, U., Picchioni, M., Walker, N., Touloupoulou, T., Need, A.C., Ge, D., Yoon, J.L., Shianna, K.V., Freimer, N.B., Cantor, R.M., Murray, R., Kong, A., Golimbet, V., Carracedo, A., Arango, C., Costas, J., Jonsson, E.G., Terenius, L., Agartz, I., Petursson, H., Nothen, M.M., Rietschel, M., Matthews, P.M., Muglia, P., Peltonen, L., St Clair, D., Goldstein, D.B., Stefansson, K., Collier, D.A., 2009. Common variants conferring risk of schizophrenia. *Nature* 460, 744-747.
- Stefansson, H., Rujescu, D., Cichon, S., Pietilainen, O.P., Ingason, A., Steinberg, S., Fossdal, R., Sigurdsson, E., Sigmundsson, T., Buizer-Voskamp, J.E., Hansen, T., Jakobsen, K.D., Muglia, P., Francks, C., Matthews, P.M., Gylfason, A., Halldorsson, B.V., Gudbjartsson, D., Thorgeirsson, T.E., Sigurdsson, A., Jonasdottir, A., Bjornsson, A., Mattiasdottir, S., Blondal, T., Haraldsson, M., Magnusdottir, B.B., Giegling, I., Moller, H.J., Hartmann, A., Shianna, K.V., Ge, D., Need, A.C., Crombie, C., Fraser, G., Walker, N., Lonnqvist, J., Suvisaari, J., Tuulio-Henriksson, A., Paunio, T., Touloupoulou, T., Bramon, E., Di Forti, M., Murray, R., Ruggeri, M., Vassos, E., Tosato, S., Walshe, M., Li, T., Vasilescu, C., Muhleisen, T.W., Wang, A.G., Ullum, H., Djurovic, S., Melle, I., Olesen, J., Kiemeny, L.A., Franke, B., Sabatti, C., Freimer, N.B., Gulcher, J.R., Thorsteinsdottir, U., Kong, A., Andreassen, O.A., Ophoff, R.A., Georgi, A., Rietschel, M., Werge, T., Petursson, H., Goldstein, D.B., Nothen, M.M., Peltonen, L., Collier, D.A., St Clair, D., Stefansson, K., 2008. Large recurrent microdeletions associated with schizophrenia. *Nature* 455, 232-236.
- Steinberg, S., de Jong, S., Andreassen, O.A., Werge, T., Borglum, A.D., Mors, O., Mortensen, P.B., Gustafsson, O., Costas, J., Pietilainen, O.P., Demontis, D., Papiol, S., Huttenlocher, J., Mattheisen, M., Breuer, R., Vassos, E., Giegling, I., Fraser, G., Walker, N., Tuulio-Henriksson, A., Suvisaari, J., Lonnqvist, J., Paunio, T., Agartz, I., Melle, I., Djurovic, S., Strengman, E., Jurgens, G., Glenthøj, B., Terenius, L., Hougaard, D.M., Orntoft, T., Wiuf, C., Didriksen, M., Hollegaard, M.V., Nordentoft, M., van Winkel, R., Kenis, G., Abramova, L., Kaleda, V., Arrojo, M., Sanjuan, J., Arango, C., Sperling, S., Rossner, M., Ribolsi, M., Magni, V., Siracusano, A., Christiansen, C., Kiemeny, L.A., Veldink, J., van den Berg, L., Ingason, A., Muglia, P., Murray, R., Nothen, M.M., Sigurdsson, E., Petursson, H., Thorsteinsdottir, U., Kong, A., Rubino, I.A., De Hert, M., Rethelyi, J.M., Bitter, I., Jonsson, E.G., Golimbet, V., Carracedo, A., Ehrenreich, H., Craddock, N., Owen, M.J., O'Donovan, M.C., Ruggeri, M., Tosato, S., Peltonen, L., Ophoff, R.A., Collier, D.A., St Clair, D., Rietschel, M., Cichon, S., Stefansson, H., Rujescu, D., Stefansson, K., 2011a. Common variants at VRK2 and TCF4 conferring risk of schizophrenia. *Hum Mol Genet* 20, 4076-4081.
- Steinberg, S., Mors, O., Borglum, A.D., Gustafsson, O., Werge, T., Mortensen, P.B., Andreassen, O.A., Sigurdsson, E., Thorgeirsson, T.E., Bottcher, Y., Olason, P., Ophoff, R.A., Cichon, S., Gudjonsdottir, I.H., Pietilainen, O.P., Nyegaard, M., Tuulio-Henriksson, A., Ingason, A., Hansen, T., Athanasiu, L., Suvisaari, J., Lonnqvist, J., Paunio, T., Hartmann, A., Jurgens, G., Nordentoft, M., Hougaard, D., Norgaard-Pedersen, B., Breuer, R., Moller, H.J., Giegling, I., Glenthøj, B., Rasmussen, H.B., Mattheisen, M., Bitter, I., Rethelyi, J.M., Sigmundsson, T., Fossdal, R., Thorsteinsdottir, U., Ruggeri, M., Tosato, S.,

- Strengman, E., Kiemeny, L.A., Melle, I., Djurovic, S., Abramova, L., Kaleda, V., Walshe, M., Bramon, E., Vassos, E., Li, T., Fraser, G., Walker, N., Touloupoulou, T., Yoon, J., Freimer, N.B., Cantor, R.M., Murray, R., Kong, A., Golimbet, V., Jonsson, E.G., Terenius, L., Agartz, I., Petursson, H., Nothen, M.M., Rietschel, M., Peltonen, L., Rujescu, D., Collier, D.A., Stefansson, H., St Clair, D., Stefansson, K., 2011b. Expanding the range of ZNF804A variants conferring risk of psychosis. *Mol Psychiatry* 16, 59-66.
- Strolin Benedetti, M., Tipton, K.F., Whomsley, R., 2007. Amine oxidases and monoxygenases in the in vivo metabolism of xenobiotic amines in humans: has the involvement of amine oxidases been neglected? *Fundam Clin Pharmacol* 21, 467-480.
- Sullivan, P., 2011. Don't give up on GWAS. *Mol Psychiatry* 17, 2-3.
- Sullivan, P.F., 2005. The genetics of schizophrenia. *PLoS Med* 2, e212.
- Sullivan, P.F., Kendler, K.S., Neale, M.C., 2003. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry* 60, 1187-1192.
- Sullivan, P.F., Lin, D., Tzeng, J.Y., van den Oord, E., Perkins, D., Stroup, T.S., Wagner, M., Lee, S., Wright, F.A., Zou, F., Liu, W., Downing, A.M., Lieberman, J., Close, S.L., 2008. Genomewide association for schizophrenia in the CATIE study: results of stage 1. *Mol Psychiatry* 13, 570-584.
- Szulwach, K.E., Li, X., Smrt, R.D., Li, Y., Luo, Y., Lin, L., Santistevan, N.J., Li, W., Zhao, X., Jin, P., 2010. Cross talk between microRNA and epigenetic regulation in adult neurogenesis. *J Cell Biol* 189, 127-141.
- Takahashi, S., Cui, Y.H., Han, Y.H., Fagerness, J.A., Galloway, B., Shen, Y.C., Kojima, T., Uchiyama, M., Faraone, S.V., Tsuang, M.T., 2008. Association of SNPs and haplotypes in APOL1, 2 and 4 with schizophrenia. *Schizophr Res* 104, 153-164.
- Taki, Y., Kinomura, S., Sato, K., Goto, R., Inoue, K., Okada, K., Ono, S., Kawashima, R., Fukuda, H., 2006. Both global gray matter volume and regional gray matter volume negatively correlate with lifetime alcohol intake in non-alcohol-dependent Japanese men: a volumetric analysis and a voxel-based morphometry. *Alcohol Clin Exp Res* 30, 1045-1050.
- Tam, G.W., van de Lagemaat, L.N., Redon, R., Strathdee, K.E., Croning, M.D., Malloy, M.P., Muir, W.J., Pickard, B.S., Deary, I.J., Blackwood, D.H., Carter, N.P., Grant, S.G., 2010. Confirmed rare copy number variants implicate novel genes in schizophrenia. *Biochem Soc Trans* 38, 445-451.
- Tandon, R., Nasrallah, H.A., Keshavan, M.S., 2010. Schizophrenia, "just the facts" 5. Treatment and prevention. Past, present, and future. *Schizophr Res* 122, 1-23.
- The International HapMap Consortium, 2005. A haplotype map of the human genome. *Nature* 437, 1299-1320.
- Thomas, P.D., Campbell, M.J., 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, 2129-2141.

- Tishkoff, S.A., Verrelli, B.C., 2003. Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annu Rev Genomics Hum Genet* 4, 293-340.
- Torrey, E.F., Buka, S., Cannon, T.D., Goldstein, J.M., Seidman, L.J., Liu, T., Hadley, T., Rosso, I.M., Bearden, C., Yolken, R.H., 2009. Paternal age as a risk factor for schizophrenia: how important is it? *Schizophr Res* 114, 1-5.
- Torri, F., Akelai, A., Lupoli, S., Sironi, M., Amann-Zalcenstein, D., Fumagalli, M., Dal Fiume, C., Ben-Asher, E., Kanyas, K., Cagliani, R., Cozzi, P., Trombetti, G., Strik Lievers, L., Salvi, E., Orro, A., Beckmann, J.S., Lancet, D., Kohn, Y., Milanese, L., Ebstein, R.B., Lerer, B., Macciardi, F., 2010. Fine mapping of AHI1 as a schizophrenia susceptibility gene: from association to evolutionary evidence. *FASEB J* 24, 3066-3082.
- Toulopoulou, T., Goldberg, T.E., Mesa, I.R., Picchioni, M., Rijdsdijk, F., Stahl, D., Cherny, S.S., Sham, P., Faraone, S.V., Tsuang, M., Weinberger, D.R., Seidman, L.J., Murray, R.M., 2010. Impaired intellect and memory: a missing link between genetic risk and schizophrenia? *Arch Gen Psychiatry* 67, 905-913.
- Toulopoulou, T., Picchioni, M., Rijdsdijk, F., Hua-Hall, M., Ettinger, U., Sham, P., Murray, R., 2007. Substantial genetic overlap between neurocognition and schizophrenia: genetic modeling in twin samples. *Arch Gen Psychiatry* 64, 1348-1355.
- Traherne, J.A., 2008. Human MHC architecture and evolution: implications for disease association studies. *Int J Immunogenet* 35, 179-192.
- Uher, R., 2009. The role of genetic variation in the causation of mental illness: an evolution-informed framework. *Mol Psychiatry* 14, 1072-1082.
- Valente, E.M., Brancati, F., Silhavy, J.L., Castori, M., Marsh, S.E., Barrano, G., Bertini, E., Boltshauser, E., Zaki, M.S., Abdel-Aleem, A., Abdel-Salam, G.M., Bellacchio, E., Battini, R., Cruse, R.P., Dobyns, W.B., Krishnamoorthy, K.S., Lagier-Tourenne, C., Magee, A., Pascual-Castroviejo, I., Salpietro, C.D., Sarco, D., Dallapiccola, B., Gleeson, J.G., 2006. AHI1 gene mutations cause specific forms of Joubert syndrome-related disorders. *Ann Neurol* 59, 527-534.
- Vallender, E.J., Lahn, B.T., 2004. Positive selection on the human genome. *Hum Mol Genet* 13 Spec No 2, R245-254.
- van Os, J., Kapur, S., 2009. Schizophrenia. *Lancet* 374, 635-645.
- van Os, J., Kenis, G., Rutten, B.P., 2010. The environment and schizophrenia. *Nature* 468, 203-212.
- Vanhamme, L., Paturiaux-Hanocq, F., Poelvoorde, P., Nolan, D.P., Lins, L., Van Den Abbeele, J., Pays, A., Tebabi, P., Van Xong, H., Jacquet, A., Moguilevsky, N., Dieu, M., Kane, J.P., De Baetselier, P., Brasseur, R., Pays, E., 2003. Apolipoprotein L-I is the trypanosome lytic factor of human serum. *Nature* 422, 83-87.
- Visscher, P.M., Goddard, M.E., Derks, E.M., Wray, N.R., 2011. Evidence-based psychiatric genetics, AKA the false dichotomy between common and rare variant hypotheses. *Mol Psychiatry* doi: 10.1038/mp.2011.65.

- Voight, B.F., Kudravalli, S., Wen, X., Pritchard, J.K., 2006. A map of recent positive selection in the human genome. *PLoS Biol* 4, e72.
- Vrijenhoek, T., Buizer-Voskamp, J.E., van der Stelt, I., Strengman, E., Sabatti, C., Geurts van Kessel, A., Brunner, H.G., Ophoff, R.A., Veltman, J.A., 2008. Recurrent CNVs disrupt three candidate genes in schizophrenia patients. *Am J Hum Genet* 83, 504-510.
- Walker, E., Kestler, L., Bollini, A., Hochman, K.M., 2004. Schizophrenia: etiology and course. *Annu Rev Psychol* 55, 401-430.
- Walsh, T., McClellan, J.M., McCarthy, S.E., Addington, A.M., Pierce, S.B., Cooper, G.M., Nord, A.S., Kusenda, M., Malhotra, D., Bhandari, A., Stray, S.M., Rippey, C.F., Roccanova, P., Makarov, V., Lakshmi, B., Findling, R.L., Sikich, L., Stromberg, T., Merriman, B., Gogtay, N., Butler, P., Eckstrand, K., Noory, L., Gochman, P., Long, R., Chen, Z., Davis, S., Baker, C., Eichler, E.E., Meltzer, P.S., Nelson, S.F., Singleton, A.B., Lee, M.K., Rapoport, J.L., King, M.C., Sebat, J., 2008. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 320, 539-543.
- Walters, J.T., Corvin, A., Owen, M.J., Williams, H., Dragovic, M., Quinn, E.M., Judge, R., Smith, D.J., Norton, N., Giegling, I., Hartmann, A.M., Moller, H.J., Muglia, P., Moskvina, V., Dwyer, S., O'Donoghue, T., Morar, B., Cooper, M., Chandler, D., Jablensky, A., Gill, M., Kaladjieva, L., Morris, D.W., O'Donovan, M.C., Rujescu, D., Donohoe, G., 2010. Psychosis susceptibility gene ZNF804A and cognitive performance in schizophrenia. *Arch Gen Psychiatry* 67, 692-700.
- Wang, H., Feng, R., Phillip Wang, L., Li, F., Cao, X., Tsien, J.Z., 2008. CaMKII activation state underlies synaptic labile phase of LTP and short-term memory formation. *Curr Biol* 18, 1546-1554.
- Wang, X., He, G., Gu, N., Yang, J., Tang, J., Chen, Q., Liu, X., Shen, Y., Qian, X., Lin, W., Duan, Y., Feng, G., He, L., 2004. Association of G72/G30 with schizophrenia in the Chinese population. *Biochem Biophys Res Commun* 319, 1281-1286.
- Wang, Z., Wei, J., Zhang, X., Guo, Y., Xu, Q., Liu, S., Shi, J., Yu, Y., Ju, G., Li, Y., Shen, Y., 2006. A review and re-evaluation of an association between the NOTCH4 locus and schizophrenia. *Am J Med Genet B Neuropsychiatr Genet* 141B, 902-906.
- Ward, M.A., Carlsson, C.M., Trivedi, M.A., Sager, M.A., Johnson, S.C., 2005. The effect of body mass index on global brain volume in middle-aged adults: a cross sectional study. *BMC Neurol* 5, 23.
- Weinberger, D.R., 1987. Implications of normal brain development for the pathogenesis of schizophrenia. *Arch Gen Psychiatry* 44, 660-669.
- Weinberger, D.R., Berman, K.F., Zec, R.F., 1986. Physiologic dysfunction of dorsolateral prefrontal cortex in schizophrenia. I. Regional cerebral blood flow evidence. *Arch Gen Psychiatry* 43, 114-124.
- Weinberger, D.R., McClure, R.K., 2002. Neurotoxicity, neuroplasticity, and magnetic resonance imaging morphometry: what is happening in the schizophrenic brain? *Arch Gen Psychiatry* 59, 553-558.

- Weiss, K.M., Clark, A.G., 2002. Linkage disequilibrium and the mapping of complex human traits. *Trends Genet* 18, 19-24.
- Williams, H.J., Norton, N., Dwyer, S., Moskvina, V., Nikolov, I., Carroll, L., Georgieva, L., Williams, N.M., Morris, D.W., Quinn, E.M., Giegling, I., Ikeda, M., Wood, J., Lencz, T., Hultman, C., Lichtenstein, P., Thiselton, D., Maher, B.S., Malhotra, A.K., Riley, B., Kendler, K.S., Gill, M., Sullivan, P., Sklar, P., Purcell, S., Nimgaonkar, V.L., Kirov, G., Holmans, P., Corvin, A., Rujescu, D., Craddock, N., Owen, M.J., O'Donovan, M.C., 2011. Fine mapping of ZNF804A and genome-wide significant evidence for its involvement in schizophrenia and bipolar disorder. *Mol Psychiatry* 16, 429-441.
- Williams, H.J., Owen, M.J., O'Donovan, M.C., 2009. Schizophrenia genetics: new insights from new approaches. *Br Med Bull* 91, 61-74.
- Wong, G.K., Yang, Z., Passey, D.A., Kibukawa, M., Paddock, M., Liu, C.R., Bolund, L., Yu, J., 2003. A population threshold for functional polymorphisms. *Genome Res* 13, 1873-1879.
- Worden, F.G., Childs, B., Matthyse, S., Gershon, E.S., 1976. Frontiers of psychiatric genetics. *Neurosci Res Program Bull* 14, 8-86.
- World Health Organization. 1992 ICD-10, The International Statistical Classification of Diseases and Related Health Problems, 10th revision. Geneva.
- Wray, N.R., Visscher, P.M., 2010. Narrowing the boundaries of the genetic architecture of schizophrenia. *Schizophr Bull* 36, 14-23.
- Wright, A., Charlesworth, B., Rudan, I., Carothers, A., Campbell, H., 2003. A polygenic basis for late-onset disease. *Trends Genet* 19, 97-106.
- Wu, S.J., Chiang, F.T., Chen, W.J., Liu, P.H., Hsu, K.L., Hwang, J.J., Lai, L.P., Lin, J.L., Tseng, C.D., Tseng, Y.Z., 2004. Three single-nucleotide polymorphisms of the angiotensinogen gene and susceptibility to hypertension: single locus genotype vs. haplotype analysis. *Physiol Genomics* 17, 79-86.
- Xu, B., Roos, J.L., Dexheimer, P., Boone, B., Plummer, B., Levy, S., Gogos, J.A., Karayiorgou, M., 2011. Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nat Genet* 43, 864-868.
- Xu, B., Roos, J.L., Levy, S., van Rensburg, E.J., Gogos, J.A., Karayiorgou, M., 2008. Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet* 40, 880-885.
- Yamada, K., Iwayama, Y., Hattori, E., Iwamoto, K., Toyota, T., Ohnishi, T., Ohba, H., Maekawa, M., Kato, T., Yoshikawa, T., 2011. Genome-wide association study of schizophrenia in Japanese population. *PLoS One* 6, e20468.
- Yang, J., Visscher, P.M., Wray, N.R., 2010. Sporadic cases are the norm for complex disease. *Eur J Hum Genet* 18, 1039-1043.
- Yao, W.D., Spealman, R.D., Zhang, J., 2008. Dopaminergic signaling in dendritic spines. *Biochem Pharmacol* 75, 2055-2069.

- Yu, F., Keinan, A., Chen, H., Ferland, R.J., Hill, R.S., Mignault, A.A., Walsh, C.A., Reich, D., 2009. Detecting natural selection by empirical comparison to random regions of the genome. *Hum Mol Genet* 18, 4853-4867.
- Zhabotinsky, A.M., Camp, R.N., Epstein, I.R., Lisman, J.E., 2006. Role of the neurogranin concentrated in spines in the induction of long-term potentiation. *J Neurosci* 26, 7337-7347.
- Zhang, F., Fan, H., Xu, Y., Zhang, K., Huang, X., Zhu, Y., Sui, M., Sun, G., Feng, K., Xu, B., Zhang, X., Su, Z., Peng, C., Liu, P., 2011a. Converging evidence implicates the dopamine D3 receptor gene in vulnerability to schizophrenia. *Am J Med Genet B Neuropsychiatr Genet* 156B, 613-619.
- Zhang, F., Gu, W., Hurles, M.E., Lupski, J.R., 2009. Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet* 10, 451-481.
- Zhang, J., Vinuela, A., Neely, M.H., Hallett, P.J., Grant, S.G., Miller, G.M., Isacson, O., Caron, M.G., Yao, W.D., 2007a. Inhibition of the dopamine D1 receptor signaling by PSD-95. *J Biol Chem* 282, 15778-15789.
- Zhang, R., Lu, S.M., Qiu, C., Liu, X.G., Gao, C.G., Guo, T.W., Valenzuela, R.K., Deng, H.W., Ma, J., 2011b. Population-based and family-based association studies of ZNF804A locus and schizophrenia. *Mol Psychiatry* 16, 360-361.
- Zhang, W., Zhang, Y., Zheng, H., Zhang, C., Xiong, W., Olyarchuk, J.G., Walker, M., Xu, W., Zhao, M., Zhao, S., Zhou, Z., Wei, L., 2007b. SynDB: a Synapse protein DataBase based on synapse ontology. *Nucleic Acids Res* 35, D737-741.
- Zhang, X.S., Hill, W.G., 2003. Multivariate stabilizing selection and pleiotropy in the maintenance of quantitative genetic variation. *Evolution* 57, 1761-1775.
- Zhong, L., Cherry, T., Bies, C.E., Florence, M.A., Gerges, N.Z., 2009. Neurogranin enhances synaptic strength through its interaction with calmodulin. *EMBO J* 28, 3027-3039.
- Zondervan, K.T., Cardon, L.R., 2004. The complex interplay among factors that influence allelic association. *Nat Rev Genet* 5, 89-100.
- Zweier, C., de Jong, E.K., Zweier, M., Orrico, A., Ousager, L.B., Collins, A.L., Bijlsma, E.K., Oortveld, M.A., Ekici, A.B., Reis, A., Schenck, A., Rauch, A., 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, 655-666.