



## ADAMTSL3 as a candidate gene for schizophrenia: Gene sequencing and ultra-high density association analysis by imputation

David J. Dow<sup>a</sup>, Julie Huxley-Jones<sup>b</sup>, Jamie M. Hall<sup>a</sup>, Clyde Francks<sup>b</sup>, Peter R. Maycox<sup>a</sup>, James N.C. Kew<sup>a</sup>, Israel S. Gloger<sup>a</sup>, Nalini A.L. Mehta<sup>a</sup>, Fiona M. Kelly<sup>a</sup>, Pierandrea Muglia<sup>b</sup>, Gerome Breen<sup>c</sup>, Sarah Jugurnauth<sup>c</sup>, Inti Pederoso<sup>c</sup>, David St.Clair<sup>d</sup>, Dan Rujescu<sup>e</sup>, Michael R. Barnes<sup>b,c,\*</sup>

<sup>a</sup> Molecular Discovery Research, GlaxoSmithKline Pharmaceuticals, New Frontiers Science Park (North), Third Avenue, Harlow, CM19 5AW, UK

<sup>b</sup> Computational Biology and Genetics, Quantitative Sciences, Drug Discovery, GlaxoSmithKline Pharmaceuticals, Gunnels Wood Road, Stevenage, SG1 2NY, UK

<sup>c</sup> Division of Psychological Medicine and Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College, London, UK

<sup>d</sup> Department of Mental Health, University of Aberdeen, Aberdeen, United Kingdom

<sup>e</sup> Division of Molecular and Clinical Neurobiology, Department of Psychiatry, Ludwig-Maximilians-University, Munich, Germany

### ARTICLE INFO

#### Article history:

Received 27 July 2010

Received in revised form 29 November 2010

Accepted 11 December 2010

Available online 15 January 2011

#### Keywords:

Schizophrenia

ADAMTSL3

Extracellular matrix

Imputation

Association

Sequencing

### ABSTRACT

We previously reported an association with a putative functional variant in the *ADAMTSL3* gene, just below genome-wide significance in a genome-wide association study of schizophrenia. As variants impacting the function of *ADAMTSL3* (a disintegrin-like and metalloprotease domain with thrombospondin type I motifs-like-3) could illuminate a novel disease mechanism and a potentially specific target, we have used complementary approaches to further evaluate the association. We imputed genotypes and performed high density association analysis using data from the HapMap and 1000 genomes projects. To review all variants that could potentially cause the association, and to identify additional possible pathogenic rare variants, we sequenced *ADAMTSL3* in 92 schizophrenics. A total of 71 *ADAMTSL3* variants were identified by sequencing, many were also seen in the 1000 genomes data, but 26 were novel. None of the variants identified by re-sequencing was in strong linkage disequilibrium (LD) with the associated markers. Imputation analysis refined association between *ADAMTSL3* and schizophrenia, and highlighted additional common variants with similar levels of association. We evaluated the functional consequences of all variants identified by sequencing, or showing direct or imputed association. The strongest evidence for function remained with the originally associated variant, rs950169, suggesting that this variant may be causal of the association. Rare variants were also identified with possible functional impact. Our study confirms *ADAMTSL3* as a candidate for further investigation in schizophrenia, using the variants identified here. The utility of imputation analysis is demonstrated, and we recommend wider use of this method to re-evaluate the existing canon of suggestive schizophrenia associations.

© 2011 Elsevier B.V. All rights reserved.

### 1. Background

Despite evidence for high heritability, the identification of genes conferring schizophrenia risk has consistently proved problematic. Early linkage studies have largely failed to replicate in independent

\* Corresponding author. Computational Biology, GlaxoSmithKline Pharmaceuticals, Gunnels Wood Road, Stevenage, SG1 2NY, UK. Tel.: +44 1279 622652; fax: +44 1279 622200.

E-mail addresses: [David.J.Dow@gsk.com](mailto:David.J.Dow@gsk.com) (D.J. Dow), [Julie.X.Huxley-jones@gsk.com](mailto:Julie.X.Huxley-jones@gsk.com) (J. Huxley-Jones), [James.M.Hall@gsk.com](mailto:James.M.Hall@gsk.com) (J.M. Hall), [clyde@well.ox.ac.uk](mailto:clyde@well.ox.ac.uk) (C. Francks), [Peter.R.Maycox@gmx.co.uk](mailto:Peter.R.Maycox@gmx.co.uk) (P.R. Maycox), [Israel.s.Gloger@gsk.com](mailto:Israel.s.Gloger@gsk.com) (I.S. Gloger), [Nalini.A.L.Mehta@gsk.com](mailto:Nalini.A.L.Mehta@gsk.com) (N.A.L. Mehta), [Fiona.M.Kelly@gsk.com](mailto:Fiona.M.Kelly@gsk.com) (F.M. Kelly), [pam@neurosearch.dk](mailto:pam@neurosearch.dk) (P. Muglia), [gerome.breen@kcl.ac.uk](mailto:gerome.breen@kcl.ac.uk) (G. Breen), [sarah.jugurnauth@iop.kcl.ac.uk](mailto:sarah.jugurnauth@iop.kcl.ac.uk) (S. Jugurnauth), [inti.pedroso\\_rovira@kcl.ac.uk](mailto:inti.pedroso_rovira@kcl.ac.uk) (I. Pederoso), [D.Stclair@abdn.ac.uk](mailto:D.Stclair@abdn.ac.uk) (D. St.Clair), [Dan.Rujescu@med.uni-muenchen.de](mailto:Dan.Rujescu@med.uni-muenchen.de) (D. Rujescu), [Dr.Michael.R.Barnes@gmail.com](mailto:Dr.Michael.R.Barnes@gmail.com) (M.R. Barnes).

cohorts (Brzustowicz, 2007), while a great deal of effort has also been expended on candidate gene association studies, which have also been largely non replicable. The focus is now shifting to well-powered genome-wide association studies (GWAS) (Purcell et al., 2009; Shi et al., 2009; Stefansson et al., 2009). We previously carried out a GWAS for schizophrenia susceptibility in a cohort of 871 cases and 863 controls (Need et al., 2009). The strongest association was localised to the *ADAMTSL3* (a disintegrin-like and metalloprotease domain with thrombospondin type I motifs-like) gene. Three single nucleotide polymorphisms (SNPs) in complete linkage disequilibrium, rs2135551, rs950169 and rs1911155, were highly significantly associated. The lowest combined p-value across the original and first replication study was  $1.3 \times 10^{-7}$  (rs2135551, odds ratio: 0.68, case/control minor allele freq: 0.23/0.30). However after applying a Bonferroni correction, the genome-wide threshold for 5% significance was  $1.6 \times 10^{-7}$ , confirming that the association was suggestive, but

fell short of this threshold (McCarthy et al., 2008). The case for *ADAMTSL3* in schizophrenia was subsequently strengthened by in vivo and in vitro evidence that the rs950169 minor allele mediated alternative splicing, resulting in a truncated PLAC (protease and lacunin) domain in the carboxy terminus of the *ADAMTSL3* protein. This suggested a direct causal relationship between rs950169 (combined  $p = 3.14 \times 10^{-7}$ , OR: 0.68) and the observed schizophrenia association, a degree of evidence that is rarely obtained in genetic association studies. Despite the promising functional and biological rationale, the most strongly associated SNP (rs2135551) failed to replicate in three independent populations (although the putative causal SNP was only tested by LD proxy). The replication populations included 394 cases and 524 controls from Italy ( $p = 0.311$ ), 589 schizophrenia cases and 11,491 controls from Iceland ( $p = 0.12$ ) and 179 cases and 267 controls of European-American ancestry ( $p = 0.19$ ) (Need et al., 2009). There are several possible explanations for the failure to replicate the *ADAMTSL3* association in these modest replication cohorts, however before considering these further; we consider it pertinent to explore the biological rationale that might support a role for *ADAMTSL3* in schizophrenia.

*ADAMTSL3* is a member of the ADAMTS superfamily of proteins, encompassing the 19 human ADAMTS metalloproteases and the seven non-proteolytic ADAMTS-like proteins. The superfamily show diverse functions in physiological processes such as angiogenesis, haemostasis, and particularly in the extracellular matrix (ECM) (Huxley-Jones et al., 2005). The ADAMTS proteases contain a metalloprotease domain, and an ancillary domain that determines specificity of substrate-binding, as the protease domain alone is unable to process native substrates (Cal et al., 2002). Members of the ADAMTS-like subfamily lack a metalloprotease domain and hence, lack proteolytic activity, but they otherwise have a similar structural organization including the ADAMTS ancillary domain, which has an essential substrate binding function in the extracellular space (Hall et al., 2003). Biologically related themes of extracellular matrix function and inhibition of angiogenesis emerge for *ADAMTSL3*, the later probably being mediated by the former (Porter et al., 2005). Recent genome-wide analyses have demonstrated involvement of *ADAMTSL3* in homeostatic regulation of colorectal epithelium (Koo et al., 2007), while somatic mutations and reduced *ADAMTSL3* expression have been reported in colorectal cancer (Sjoblom et al., 2006). The gene has also been shown to be strongly up regulated in optic nerve head astrocytes in patients with primary open angle glaucoma (Rudzinski et al., 2008). *ADAMTS20* the closest ADAMTS homologue to *ADAMTSL3* (which might be expected to show similar substrate binding) has also been shown to mediate beta-catenin signalling, a key component of WNT signalling (Zhang et al., 2008), which has been widely linked to schizophrenia pathology (De Ferrari and Moon, 2006). *ADAMTSL3* itself has been shown to be an important ECM component, for example in bone ECM *ADAMTSL3* variants (rs10906982) have been associated with adult human height (Weedon et al., 2008; Soranzo et al., 2009).

The neural ECM has a specific molecular composition, with chondroitin sulfate proteoglycans (CSPGs) organised as perineuronal nets (PNNs), enveloping neuronal soma and dendrites (Pantazopoulos et al., 2010). PNNs play a crucial role in the regulation of neuronal functions in adults and have been shown to be abnormally organised in schizophrenic subjects, with overexpression of CSPGs in PNN supporting glial cells in the amygdala, leading to abnormal cytoarchitecture and neuronal migration (Pantazopoulos et al., 2010). CSPGs are one of the primary substrates of ADAMTS superfamily members, for example *ADAMTS4* enhances neurite outgrowth directly at the cell surface by processing of CSPGs in the ECM (Hamel et al., 2008). Both *ADAMTS1* and *ADAMTS4* have been shown to be important in synaptogenesis and remodelling of the dentate gyrus and amygdala after excitotoxic lesion in rats, by cleavage of PNN CSPGs (Mayer et al., 2005; Yuan et al., 2002). *ADAMTSL3* is widely expressed in the brain

from the foetus to the adult (Supplementary Fig. 1a and b). Although the gene has not been directly investigated in synaptogenesis or synaptic remodelling, considering the known biology of other superfamily members in the neural ECM, we propose that *ADAMTSL3* could plausibly be involved in the PNN abnormalities seen in schizophrenia (Pantazopoulos et al., 2010).

Considering the possible role of *ADAMTSL3* in the development and remodelling of the neuronal architecture of the brain, the results of a recent evolutionary study of Neanderthal and modern human genomes may be significant (Green et al., 2010). *ADAMTSL3* alleles were shown to rank in the top 5% of genes showing evidence of a selective sweep in modern humans. This suggests that the gene may fit within the evolutionary hypotheses that have proposed the development of schizophrenia and other forms of psychosis as a by-product of social brain evolution in *Homo sapiens* (Burns, 2006).

In this study we are seeking to address a generic problem in schizophrenia genetics, namely failure to replicate (Mitchell and Porteous, 2010). A true association can fail to replicate for multiple reasons including, insufficient sample size or variability in phenotype definitions across independent samples. Allele frequencies of variants may also differ across studies due to sampling error or population differences. In one study replication was shown to fail due to interaction between distinct functional polymorphisms and differences in allele frequency were also shown to lead to a reversal of allelic effects where a protective allele became a risk factor in replication studies (Greene et al., 2009). This re-enforces the point that variants that fail to replicate should be checked for interactions with other polymorphisms, particularly when samples are collected from groups with distinct ethnic backgrounds or different geographic regions. Mindful of these recommendations, we have further evaluated the association between *ADAMTSL3* and schizophrenia. Our original association study gave only limited coverage of variants in the *ADAMTSL3* gene, leaving open the possibility that additional, un-genotyped variants might show stronger association and have a higher chance of replication in independent cohorts. We have also screened *ADAMTSL3* for rare highly penetrant mutations or structural variants, which might act independently to the observed association, as described in other genes originally identified by schizophrenia association analysis (Sebat et al., 2009). To address these goals, we sequenced *ADAMTSL3* in 92 schizophrenia cases from the original GWAS (Need et al., 2009). We identified all *ADAMTSL3* promoter and exonic sequence variation in our subjects, performed LD analysis and also predictive functional analysis to identify potential causal variants. Finally with the public availability of ultra-high density variant information from the 1000 genomes project ([www.1000genomes.org](http://www.1000genomes.org)), we used a genotype imputation method to repeat association analysis using all available variation information.

## 2. Methods

### 2.1. Samples and DNA extraction

Whole blood samples were randomly selected from 92 schizophrenia patients from both the Munich and Aberdeen subjects described in our original GWAS study (Need et al., 2009). Genomic DNA was extracted by SDS/Proteinase K treatment and salt/isopropanol precipitation using standard methods. All cases and controls gave informed consent. The study was approved by both local and multiregional academic ethical committees.

### 2.2. DNA sequencing and analysis

M13 tailed PCR primers were designed against a gene sequence constructed from the following NCBI entries: NM\_207517, BC128390, AK092324 and H26489 and NCBI\_36 genomic sequence, using the Primer 3 software package. Coverage included 30 exons,

approximately 100 base pairs of intronic sequence flanking each exon and 2 kb of sequence upstream of the transcription initiation site. A total of 41 fragments of approximately 550 bases were designed. PCR was performed using HotStarTaq Master Mix PCR Kit (Qiagen) using a PTC-225TM MJC Research programmable thermocycler (Bio-Rad) (Initiation step: 95 °C for 15 min, followed by 35 cycles: 95 °C for 20 s, 55–68 °C for 30 s, and 72 °C for 45 s). PCR products were purified by standard methods using an Ampure SPRI system (Agencourt). The purified PCR products were checked qualitatively and quantitatively on an E-Gel (96 well 2% agarose GP, Invitrogen) using a 1 kb Plus DNA Ladder (Invitrogen).

M13 universal primers were used for sequencing. The purified PCR products were cycle sequenced in a 10 µl reaction using BigDye Terminator Enzyme 3.1 and BigDye Terminator 5× Sequencing buffer (Applied Biosystems) (Initiation step: 94 °C for 5 min, followed by 30 cycles: 94 °C for 15 s, 50 °C for 30 s, and 60 °C for 4 m). The cycle sequenced products were purified using the CleanSeq SPRI purification method and loaded using 'direct-inject' magnets according to the manufacturers recommended protocols (Agencourt Biosciences). The purified sequencing reactions were analysed on a 3730xl DNA Analyzer (Applied Biosystems, Inc., Foster City, CA) according to standard protocols. All sequence data were analysed by the Seqscape software package (Applied Biosystems, Inc., Foster City, CA) to identify and genotype any sequence variation present. All polymorphisms were deposited in the NCBI dbSNP database.

### 2.3. Multipoint imputation and association analysis

We performed imputation of typed and untyped markers with the software IMPUTE version 2.0 using default options (Marchini et al., 2007) and used as reference panel Caucasian genotypes from the 1000 Genomes Project ([www.1000genomes.org](http://www.1000genomes.org)) and HapMap phase 3 genotypes. Software and genotype files used for imputation can be found at [https://mathgen.stats.ox.ac.uk/impute/impute\\_v2.html](https://mathgen.stats.ox.ac.uk/impute/impute_v2.html). We excluded 1798 out of 2617 imputed SNPs with a Hardy–Weinberg p-value  $<1 \times 10^{-6}$  in cases and controls, MAF  $<1\%$  in total sample or posterior probability  $<0.90$ . Genetic association with imputed genotypes was carried out with SNPTTEST under an additive model

with the proper flag to account for genotype uncertainty (Marchini et al., 2007).

### 2.4. Characterisation of ADAMTSL3 gene and variants

ADAMTSL3 protein domains were characterised based upon the literature (Hall et al., 2003; Huxley-Jones et al., 2005) and InterProScan (Zdobnov and Apweiler, 2001). Signal peptide predictions were performed using SignalP (Bendtsen et al., 2004). Transcription factor binding site prediction was performed using Promo (Farre et al., 2003). The SNP substitution predictions were based upon a published matrix (Betts and Russell, 2003) and PolyPhen (Ramensky et al., 2002).

## 3. Results

### 3.1. Analysis of ADAMTSL3 splicing and gene structure

A detailed review of the genomic context of ADAMTSL3 and available transcript data was completed using the UCSC genome browser (Karolchik et al., 2007). These identified four transcript variants (Fig. 1), the longest isoform, which we refer to as ADAMTSL3a, are identical to the NCBI peptide accession NP\_997400 and contains the canonical ADAMTSL domain structure, including thrombospondin type I repeats, ADAM spacer regions, IGCAM and Protease and Lacunin (PLAC) domains. Two additional isoforms exhibited changes at the C-terminus resulting from alternate splicing. These sequences were annotated as ADAMTSL3b and ADAMTSL3c, where the latter was found to be identical to the GenBank protein AAI28391. Due to this alternative splicing, the PLAC domain is severely truncated in ADAMTSL3b and absent in ADAMTSL3c (Fig. 1). The PLAC domain is a six-cysteine region of around 40 residues, commonly present with thrombospondin type I repeats. Domain structure analysis of these C-terminal sequences did not predict other known domains where the PLAC domain was missing. Although the function of the PLAC domain is currently not understood, there is good evidence to suggest that it is essential for ADAMTSL3 function. This assertion is supported by a strikingly analogous example in ADAMTSL4, where heterozygous nonsense mutations truncate the ADAMTSL4 carboxy terminal PLAC

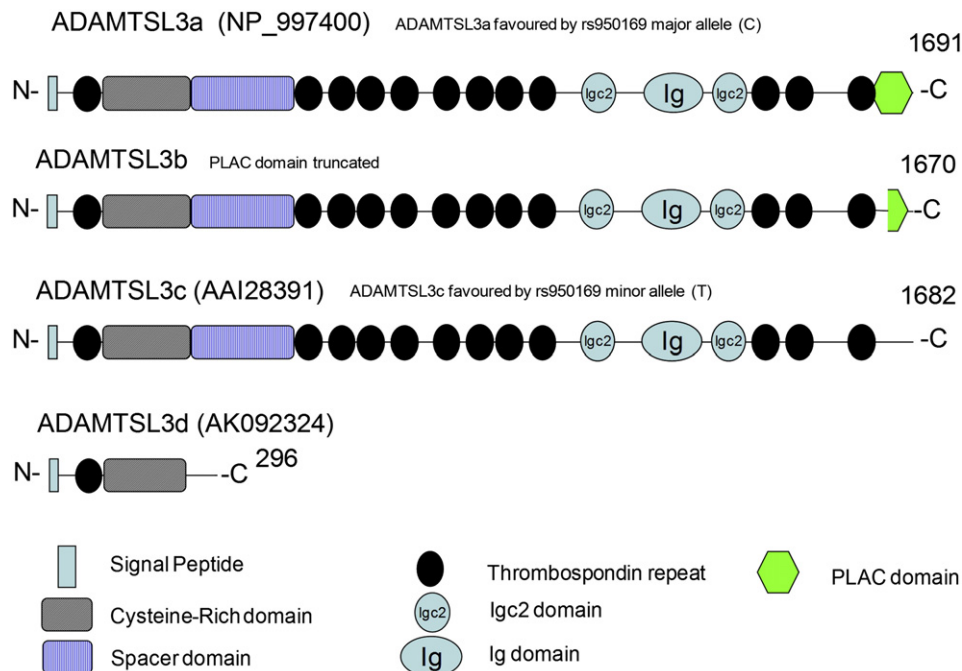


Fig. 1. Schematic view of the protein structure and domain annotation of ADAMTSL3 splice variants.

domain causing Ectopia Lentis, a congenital lens disorder (Aragon-Martin et al., 2010). The fourth transcript, known as *ADAMTSL3d*, produced a truncated isoform, lacking the ADAM-spacer, Thrombospondin type I repeat region, Immunoglobulin-like and PLAC domain, and contains a unique 29 amino acid C-terminal sequence encoded by a unique final exon. Evidence of this truncated transcript was supported by identical matches to two cDNAs (AK092324 and AX747482). The C-terminal sequence of *ADAMTSL3d* appears to be unique to this transcript with no detectable homology to any other sequence or domain. As the protein encoded by the *ADAMTSL3d* transcript lacks most of the protein folding domains, it is likely to have a significantly different protein structure and probably differs in function from other *ADAMTSL3* isoforms.

### 3.2. Sequencing data

All 30 exons, intron/exon boundaries and the promoter region of the *ADAMTSL3* gene were sequenced. We were unable to sequence two areas of the predicted promoter region due to >80% GC ratios, including a 492 bp stretch from positions 82112388 to 82112879 and a 254 bp section from 82113416 to 82113670 (positions based on NCBI\_36). We identified a total of 71 variants in the *ADAMTSL3* gene (Supplementary Table 1), six were located in the promoter region and thirteen encoded non-synonymous changes. A total of 26 polymorphisms were novel based on a search of dbSNP (build 131) and data from the 1000 genomes project. We analysed LD between all the polymorphisms and none showed meaningful LD ( $r^2 > 0.5$ ) with the GWAS associated SNPs (rs950169 and rs2135551) (Supplementary Fig. 2).

### 3.3. Imputation analysis

We performed imputation analysis on previously typed and untyped markers using reference Caucasian genotypes from the 1000 genomes project and the phase 3 HapMap (distributed as part of IMPUTE, see Section 2.3). Imputation results confirmed the association of the typed markers (with concordance ranging from 0.93 to 1) and highlighted additional common variants in high LD with similar levels of association (Fig. 2; Supplementary Table 2). The magnitude of association was not higher in newly tested variants, suggesting that the variants identified in our previous study (Need et al., 2009) may fully explain the association. Notably, the magnitude of association at rs950169 (the putative causal variant) increased to almost match the p-value at rs2135551, the most highly associated variant in the original GWAS. Had the association at rs950169 been supported by imputed data alone, we would clearly need to genotype the marker in the laboratory, however rs950169 has already been successfully genotyped in our cohort and its imputation provided genotypes of high concordance (0.982), confirming that its association statistics from imputed data are in agreement with those obtained with the original genotypes.

### 3.4. Predictive functional analysis of variation in the *ADAMTSL3* gene

All polymorphisms that were identified in this study were assessed for potential functional impact at the level of protein function, gene regulation and splicing (Supplementary Tables 1 and 3). It has been previously demonstrated that rs950169, which encodes a Thr1660Ile amino acid change, also mediates alternative splicing in the 3' region of *ADAMTSL3*, favouring the *ADAMTSL3b* isoform, resulting in a truncated PLAC domain (Need et al., 2009). After examining all imputed variants showing association ( $p < 0.001$ ) from the 1000 genomes project, we did not find any additional variants with compelling evidence for functional impact on *ADAMTSL3* or any other gene (Supplementary Table 3). We noted an un-bridged 50 kb gap in the genomic assembly downstream of *ADAMTSL3* (Fig. 2,

chr15:84984474–85034473). Based on LD and imputation analysis there is no evidence to suggest that the association extends into this region. Overall, we conclude that rs950169 is the most likely causative variant for the GWAS association.

A secondary objective of this sequencing study was to identify rare penetrant variants in *ADAMTSL3* which might independently cause schizophrenia. We found several variants that may impact splicing, mRNA structural stability and protein function. These are summarised in Supplementary Table 1 and discussed in Section 4.2.

## 4. Discussion

### 4.1. Is *ADAMTSL3* associated with schizophrenia?

This study supports the suggestive association between *ADAMTSL3* and schizophrenia and provides a detailed characterisation of variation across the gene. Our sequencing study found no additional variants in strong LD with the associated variants and our imputation analysis also showed that the highest levels of association remain in rs950169 and rs2135551. We also saw association extending to other SNPs at slightly reduced levels, although none showed evidence of functional impact (Supplementary Table 3). We consider it unlikely (although not impossible) that additional polymorphisms to rs950169 are contributing to the original GWAS association signal. This is important information for further studies of *ADAMTSL3* in schizophrenia, as the putative causal variant should be the focus of replication studies. However subsequent analysis should ideally capture all the imputation-associated variants that we identified (Supplementary Table 3), to account for other possible interactions between variants. Ultimately, conclusive replication or refutation of this association is likely to require a substantial increase in statistical power that would only be achieved with tens of thousands of carefully phenotyped subjects.

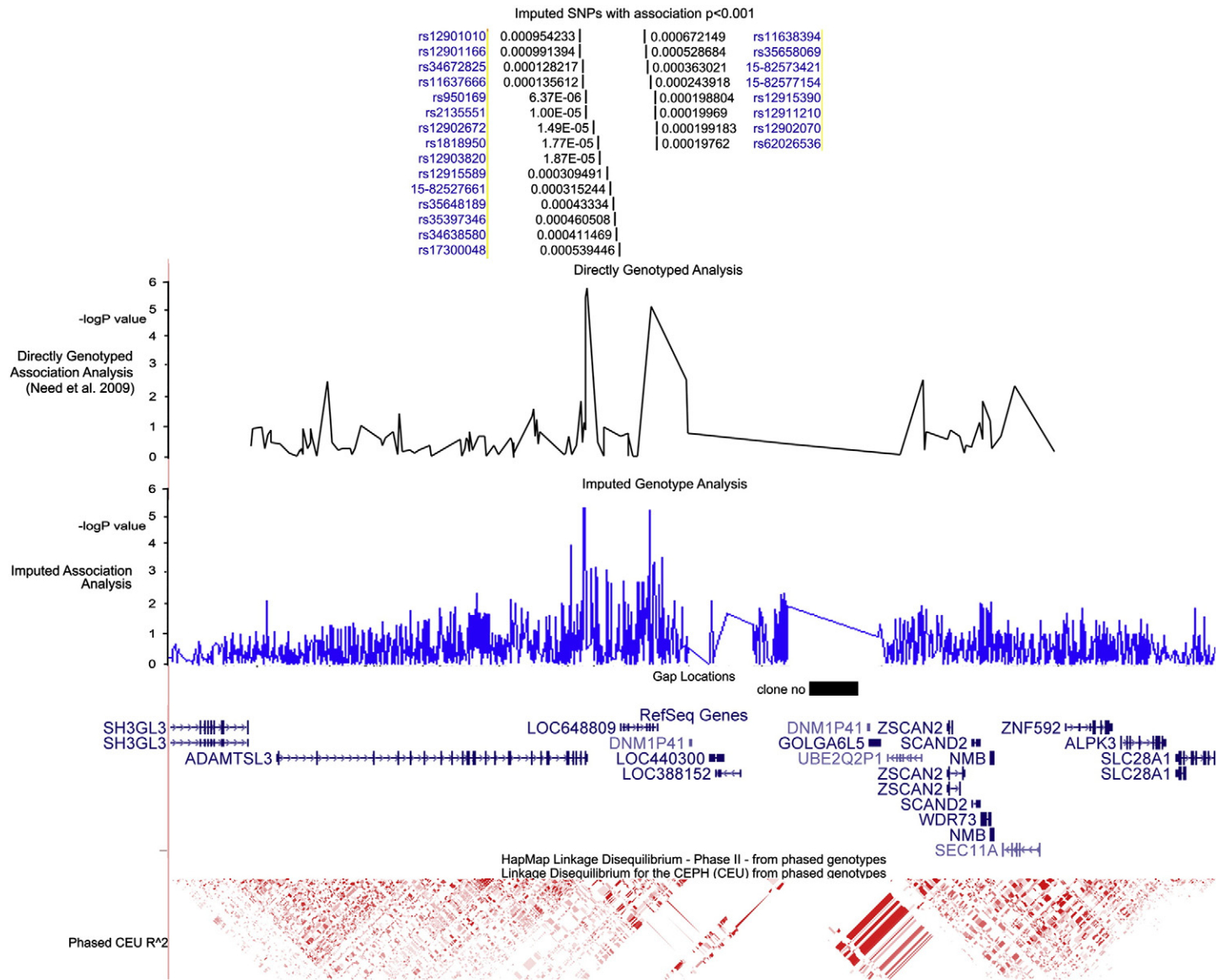
If truly associated with the disease, it may be possible to therapeutically target *ADAMTSL3* substrate binding in the brain ECM, or protease activity of related ADAMTS proteins acting on the same substrates in the same pathway. ADAMTS family members demonstrate a narrow substrate specificity, making them potentially safe pharmaceutical targets (Tortorella et al., 2009). This could open a new paradigm for schizophrenia therapy and address some of the considerable unmet medical need for schizophrenia therapeutics with a more limited side-effect profile.

### 4.2. Do rare pathogenic variants exist in *ADAMTSL3*?

If *ADAMTSL3* is a true schizophrenia susceptibility gene, it is possible that rare pathogenic variants might be present in our cases. In our sequencing study, we detected several novel rare variants that have potential for functional impact, some of which were seen as homozygotes (Supplementary Table 1). Two of the novel non-synonymous variants (ss252452302 and ss252452308) that we identified are non-conservative in nature, although none are homozygous. In a 17 bp section of the promoter region, four SNPs, two of which are seen as rare homozygotes (ss252452250, ss252452251, ss252452252, and rs2730079), are located in a transcriptionally active open chromatin region in the UCSC genome browser (Boyle et al., 2008; Karolchik et al., 2007). Multiple transcription factor binding sites map to this region and are predicted to be altered by these variants (data not shown). Potentially these SNPs could impact gene expression, although this would need to be verified by in vitro expression assays.

### 4.3. Using genotype imputation to revisit suggestive associations

By combining high density genotype data from two Caucasian reference populations ([www.1000genomes.org](http://www.1000genomes.org) and [www.hapmap.org](http://www.hapmap.org)).



**Fig. 2.** Overview of the *ADAMTSL3* locus in the UCSC genome browser, including association analysis based on lab-determined and imputed genotypes. SNPs showing  $p < 0.001$  by imputation are plotted at the top of the figure. Association results ( $-\log p$ -values) from the original study by [Need et al. \(2009\)](#) are shown, followed by association results ( $-\log p$ -values) from the imputation analysis. An un-bridged gap in the human genome sequence is highlighted above NCBI RefSeq transcripts across the region. Finally LD  $r^2$  from Caucasian HapMap data is plotted at the bottom of the figure.

org) with our own disease genotype data, we have been able to impute genotypes in our schizophrenia subjects for thousands of additional variants across the *ADAMTSL3* locus (2617 SNPs averaging 1 SNP/495 bp), far beyond those genotyped directly in the original GWAS (112 SNPs averaging 1 SNP/10 kb) (Need et al., 2009). Although genotype imputation clearly has great value, the method also has some limitations, the leading one being the reference panel used. A general reference population is unlikely to fully match the LD structure of the study population and will not match the disease status; this may have an important bearing on accuracy and sensitivity. Our use of the IMPUTE v2 combined panel of 1000 genomes and HapMap phase 3 data increases coverage of variants, providing additional information on haplotype structure but with moderate sample size. In particular, the 1000 genomes panel offers an opportunity to test for association with rare variants, but such rarity carries with it reduced imputation accuracy. In the current study, accuracy is quantified by means of concordance data given in Supplementary Table 2. Finally it should be emphasised that results based upon imputation alone, should be followed up by genotyping before strong inferences are drawn. Further, we would recommend genotyping of the most strongly associated SNP identified by imputation only (rs12905952) to evaluate concordance and further test the robustness of our conclusion that rs950169 is the most likely causative variant.

Putting aside the caveats of imputation, the method has enabled us to computationally localise the disease associated region in the *ADAMTSL3* gene with higher precision and no additional genotyping cost. Theoretically, data generated from the 1000 genomes project should provide information on almost all variants with a frequency of >1% in the Caucasian populations studied. Our own *ADAMTSL3* sequencing data broadly confirms this. The majority of novel variants we identified showed frequencies of <0.05%, only 6/26 novel variants showed a minor allele frequency >1%. Our results add to those of others (Li et al., 2009) to strongly argue for a more widespread and methodical revisiting of previous genetic associations with schizophrenia, using genome re-sequencing data from reference populations. Further sequencing or genotyping of disease samples, could be greatly reduced by the application of this approach, saving precious sample resources. In some cases, this might directly identify causal variants, and in others it would at least pinpoint small regions (perhaps one or two exons) for focused mutation discovery (again saving sample resources). This would enable researchers to better focus their experimentation on the variants that are most likely to play a functionally causative role in disease.

#### 4.4. Should the existing canon of schizophrenia candidates be revisited?

An extensive back-catalogue of un-replicated genetic associations has been generated for almost every disease showing complex genetic inheritance, schizophrenia being no exception (So et al., 2009). The majority of these associations are likely to be false, due to lack of power and issues of multiple testing. It is tempting to discount these associations in favour of newly generated, genome-wide significant GWAS associations. However a certain number of previously associated candidate genes may indeed represent true associations which could give valuable insight into disease pathology. More importantly, some of the groundwork towards building an understanding the molecular basis of an association may already have been completed as candidate genes have been characterised (e.g. transgenic models, site-directed mutagenesis, reporter gene assays, etc.). This argues strongly for re-evaluation of previous candidate gene associations with the benefit of new data. Imputation offers a computational method for revisiting and refining associations that minimise the use of precious disease sample resources. We recommend wider use of this in silico method to re-evaluate the existing canon of suggestive schizophrenia associations, building on the

tremendous volume of potentially disease relevant data that has been generated to date.

Supplementary materials related to this article can be found online at doi:10.1016/j.schres.2010.12.009.

#### Role of funding source

This study was funded primarily from the GlaxoSmithKline R&D budget with additional funding (GB, SJ, IP) from the NIHR Biomedical Research Centre for Mental Health at the South London and Maudsley NHS Foundation Trust and Institute of Psychiatry, Kings College London. GlaxoSmithKline funded employees designed the study; collected, analysed and interpreted the data; wrote the manuscript; and made the decision to submit the paper for publication.

#### Contributors

DJD—planned the study, analysed sequence data and wrote the manuscript. JHJ—performed computational biology analyses of the sequence data and wrote the manuscript. JMH—conducted PCR, sequencing and SNP detection. CF—planned the study and edited the manuscript. PRM—planned the study and edited the manuscript. JNCK—planned the study and reviewed the manuscript. ISG—planned the study and reviewed the manuscript. NALM—analysed sequence data and wrote a section of the manuscript. FMK—analysed gene expression data and edited the manuscript. PM—led the GSK/Academic collaboration and edited the manuscript. GB—supervised imputation analysis and wrote a section of the manuscript. IP—performed genotype imputation and wrote a section of the manuscript. SJ performed genotyping and wrote a section of the manuscript. DS—provided samples, advised on experimental planning and reviewed the manuscript. DR—provided samples, advised on experimental planning and reviewed the manuscript. MRB—planned the study, wrote and revised the manuscript. All authors contributed to and have approved the final manuscript.

#### Conflict of interest

David Dow, Julie Huxley-Jones, James Hall, James Kew, Israel Gloger, Nalini Mehta, Fiona Kelly and Michael Barnes are full time employees and shareholders of GlaxoSmithKline. Clyde Francks, Peter Maycox and Pierandrea Muglia are former employees of GlaxoSmithKline.

#### Acknowledgements

IP, SJ and GB were funded by the NIHR Biomedical Research Centre for Mental Health at the South London and Maudsley NHS Foundation Trust and Institute of Psychiatry, Kings College London.

#### References

- Aragon-Martin, J.A., Ahnood, D., Charteris, D.G., Saggat, A., Nischal, K.K., Comeglio, P., Chandra, A., Child, A.H., Arno, G., 2010. Role of *ADAMTSL4* mutations in *FBN1* mutation-negative ectopia lentis patients. *Hum. Mutat.* 31, E1622–E1631.
- Bendtsen, J.D., Neilsen, H., von Heijne, G., Brunak, S., 2004. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 340, 783–795.
- Betts, M.J., Russell, R.B., 2003. Amino acid properties and consequences of substitutions. In: Barnes, M.R., Gray, I.C. (Eds.), *Bioinformatics for geneticists*. Wiley.
- Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., Crawford, G.E., 2008. High-resolution mapping and characterization of open chromatin across the genome. *Cell* 132, 311–322.
- Brzustowicz, L.M., 2007. Size matters: the unexpected challenge of detecting linkage in large cohorts. *Am. J. Psychiatry* 164, 192–194.
- Burns, J.K., 2006. Psychosis: a costly by-product of social brain evolution in *Homo sapiens*. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 30, 797–814.
- Cal, S., Obaya, A.J., Llamazares, M., Garabaya, C., Quesada, V., Lopez-Otin, C., 2002. Cloning, expression analysis, and structural characterization of seven novel human *ADAMTS*s, a family of metalloproteinases with disintegrin and thrombospondin-1 domains. *Gene* 283, 49–62.
- De Ferrari, G.V., Moon, R.T., 2006. The ups and downs of Wnt signaling in prevalent neurological disorders. *Oncogene* 25, 7545–7553.
- Farre, D., Roset, R., Huerta, M., Adsuara, J.E., Rosello, L., Alba, M.M., Messeguer, X., 2003. Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. *Nucleic Acids Res.* 31, 3651–3653.
- Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H., Hansen, N.F., Durand, E.Y., Malaspina, A.S., Jensen, J.D., Marques-Bonet, T., Alkan, C., Prufer, K., Meyer, M., Burbano, H.A., Good, J.M., Schultz, R., Aximu-Petri, A., Butthof, A., Hober, B., Hoffner, B., Siegemund, M., Weihmann, A., Nusbaum, C., Lander, E.S., Russ, C., Novod, N., Affourtit, J., Egholm, M., Verna, C., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Doronichev, V.B., Golovanova, L.V., Lalueza-Fox, C., de la Rasilla, M., Fortea, J., Rosas, A., Schmitz, R.W., Johnson, P.L., Eichler, E.E., Falush, D., Birney, E., Mullikin, J.C., Slatkin, M., Nielsen, R., Kelso, J., Lachmann, M., Reich, D., Paabo, S., 2010. A draft sequence of the Neandertal genome. *Science* 328, 710–722.
- Greene, C.S., Penrod, N.M., Williams, S.M., Moore, J.H., 2009. Failure to replicate a genetic association may provide important clues about genetic architecture. *PLoS ONE* 4, e5639.

- Hall, N.G., Klenotoic, P., Anand-Apte, B., Apte, S.S., 2003. ADAMTSL-3/punctin-2, a novel glycoprotein in extracellular matrix related to the ADAMTS family of metalloproteinases. *Matrix Biol.* 22, 501–510.
- Hamel, M.G., Ajmo, J.M., Leonardo, C.C., Zuo, F., Sandy, J.D., Gottschall, P.E., 2008. Multimodal signaling by the ADAMTSs (a disintegrin and metalloproteinase with thrombospondin motifs) promotes neurite extension. *Exp. Neurol.* 210, 428–440.
- Huxley-Jones, J., Apte, S.S., Robertson, D.L., Boot-Handford, R.P., 2005. The characterisation of six ADAMTS proteases in the basal chordate *Ciona intestinalis* provides new insights into the vertebrate ADAMTS family. *Int. J. Biochem. Cell Biol.* 37, 1838–1845.
- Karolchik, D., Hinrichs, A.S., Kent, W.J., 2007. The UCSC genome browser. *Curr. Protoc. Bioinform.* 1 1.4.
- Koo, B.H., Hurskainen, T., Mielke, K., Aung, P.P., Casey, G., Autio-Harmainen, H., Apte, S.S., 2007. ADAMTSL3/punctin-2, a gene frequently mutated in colorectal tumors, is widely expressed in normal and malignant epithelial cells, vascular endothelial cells and other cell types, and its mRNA is reduced in colon cancer. *Int. J. Cancer* 121, 1710–1716.
- Li, Y., Willer, C., Sanna, S., Abecasis, G., 2009. Genotype imputation. *Annu. Rev. Genomics Hum. Genet.* 10, 387–406.
- Marchini, J., Howie, B., Myers, S., McVean, G., Donnelly, P., 2007. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* 39, 906–913.
- Mayer, J., Hamel, M.G., Gottschall, P.E., 2005. Evidence for proteolytic cleavage of brevican by the ADAMTSs in the dentate gyrus after excitotoxic lesion of the mouse entorhinal cortex. *BMC Neurosci.* 6, 52.
- McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P., Hirschhorn, J.N., 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* 9, 356–369.
- Mitchell, K.J., Porteous, D.J., 2010. Rethinking the genetic architecture of schizophrenia. *Psychol. Med.* 1–14.
- 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, J.H., 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.
- Pantazopoulos, H., Woo, T.U., Lim, M.P., Lange, N., Berretta, S., 2010. Extracellular matrix-glia abnormalities in the amygdala and entorhinal cortex of subjects diagnosed with schizophrenia. *Arch. Gen. Psychiatry* 67, 155–166.
- Porter, S., Kevorkian, L., Edwards, D.R., 2005. The ADAMTS metalloproteinases. *Biochem. J.* 386, 15–27.
- 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, 748–752.
- Ramensky, V., Bork, P., Sunayev, S., 2002. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res.* 30, 3894–3900.
- Rudzinski, M.N., Chen, L., Hernandez, M.R., 2008. Antiangiogenic characteristics of astrocytes from optic nerve heads with primary open-angle glaucoma. *Arch. Ophthalmol.* 126, 679–685.
- Sebat, J., Levy, D.L., McCarthy, S.E., 2009. Rare structural variants in schizophrenia: one disorder, multiple mutations; one mutation, multiple disorders. *Trends Genet.* 25, 528–535.
- 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.
- Sjoberg, T., Jones, J., Wood, L.D., Parsons, D.W., Lin, J., Barber, T., Mandelker, D., Leary, R.J., Ptak, J., Silliman, N., Szabo, S., Buckhaults, P., Farrell, C., Meeh, P., Markowitz, S.D., Willis, J., Dawson, D., Willson, J.K., Gazdar, A.F., Hartigan, J., Wu, L., Liu, C., Parmigiani, G., Park, B.H., Bachman, K.E., Papadopoulos, N., Vogelstein, B., Kinzler, K.W., Velculescu, V.E., 2006. The consensus coding sequences of human breast and colorectal cancers. *Science* 317, 268–274.
- So, H.C., Chen, E.Y., Sham, P.C., 2009. Genetics of schizophrenia spectrum disorders: looking back and peering ahead. *Ann. Acad. Med. Singapore* 38, 436–439.
- Soranzo, N., Rivadeneira, F., Chinapen-Horsley, U., Malkina, I., Richards, J.B., Hammond, N., Stolk, L., Nica, A., Inouye, M., Hofman, A., Stephens, J., Wheeler, E., Arp, P., Gwilliam, R., Jhamai, P.M., Potter, S., Chaney, A., Ghori, M.J., Ravindrarajah, R., Ermakov, S., Estrada, K., Pols, H.A., Williams, F.M., McArdle, W.L., van Meurs, J.B., Loos, R.J., Dermitzakis, E.T., Ahmadi, K.R., Hart, D.J., Ouweland, W.H., Wareham, N.J., Barroso, I., Sandhu, M.S., Strachan, D.P., Livshits, G., Spector, T.D., Uitterlinden, A.G., Deloukas, P., 2009. Meta-analysis of genome-wide scans for human adult stature identifies novel loci and associations with measures of skeletal frame size. *PLoS Genet.* 5, e1000445.
- 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 F.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, C.D., Goldstein, D.B., Stefansson, K., Collier, D.A., 2009. Common variants conferring risk of schizophrenia. *Nature* 460, 744–747.
- Tortorella, M.D., Malfait, F., Barve, R.A., Shieh, H.S., Malfait, A.M., 2009. A review of the ADAMTS family, pharmaceutical targets of the future. *Curr. Pharm. Des.* 15, 2359–2374.
- Weedon, M.N., Lango, J., Lindgren, C.M., Wallace, C., Evans, D.M., Mangino, M., Freathy, R.M., Perry, J.R., Stevens, S., Hall, A.S., Samani, N.J., Shields, B., Prokopenko, I., Farrall, M., Dominiczak, A., Diabetes Genetics Initiative, Wellcome Trust Case Control Consortium, Johnson, T., Bergmann, S., Beckmann, J.S., Vollenweider, P., Waterworth, D.M., Mooser, V., Palmer, C.N., Morris, A.D., Ouweland, W.H., Cambridge GEM Consortium, Zhao, J.H., Li, S., Loos, R.J., Barroso, I., Deloukas, P., Sandhu, M.S., Wheeler, E., Soranzo, N., Inouye, M., Wareham, N.J., Caulfield, M., Munroe, P.B., Hattersley, A.T., McCarthy, M.I., Frayling, T.M., 2008. Genome-wide association analysis identifies 20 loci that influence adult height. *Nat. Genet.* 40, 575–583.
- Yuan, W., Matthews, R.T., Sandy, J.D., Gottschall, P.E., 2002. Association between protease-specific proteolytic cleavage of brevican and synaptic loss in the dentate gyrus of kainate-treated rats. *Neuroscience* 114, 1091–1101.
- Zdobnov, E.M., Apweiler, R., 2001. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17, 847–848.
- Zhang, Y., Andl, T., Yang, S.H., Teta, M., Liu, F., Seykora, J.T., Tobias, J.W., Piccolo, S., Schmidt-Ullrich, R., Nagy, A., Taketo, M.M., Dlugosz, A.A., Millar, S.E., 2008. Activation of beta-catenin signaling programs embryonic epidermis to hair follicle fate. *Development* 135, 2161–2172.