



Published in final edited form as:

Nat Genet. 2014 August ; 46(8): 881–885. doi:10.1038/ng.3039.

Most genetic risk for autism resides with common variation

Trent Gaugler¹, Lambertus Klei², Stephan J. Sanders^{3,4}, Corneliu A. Bodea¹, Arthur P. Goldberg^{5,6,7}, Ann B. Lee¹, Milind Mahajan⁸, Dina Manaa⁸, Yudi Pawitan⁹, Jennifer Reichert^{5,6}, Stephan Ripke¹⁰, Sven Sandin⁹, Pamela Sklar^{6,7,8,11,12}, Oscar Svantesson⁹, Abraham Reichenberg^{5,6,13}, Christina M. Hultman⁹, Bernie Devlin², Kathryn Roeder^{1,14}, and Joseph D. Buxbaum^{5,6,8,11,15,16}

¹Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

²Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA

³Department of Psychiatry, University of California San Francisco, San Francisco, California, USA

⁴Department of Genetics, Yale University School of Medicine, New Haven, Connecticut, USA

⁵Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, New York, USA

⁶Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York, USA

⁷Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, New York, USA

⁸Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA

⁹Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, SE-171 77 Stockholm, Sweden

¹⁰Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA

¹¹Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA

¹²Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai, New York, New York, USA

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence to: Kathryn Roeder; Joseph D. Buxbaum.

Author Contributions

AR, CMH, BD, KR, and JDB conceived the project and designed its components. AR, CMH, BD, KR, and JDB identified funding for the study. SS, OS, and CHM were responsible for ascertaining case samples and PS and CHM for ascertaining control samples. JR, DM, and MM were responsible for genotyping the case samples. APG organized and managed the data files and TG and APG carried out quality control of the SNP data. SJS carried out simulations and additional analyses to assess contribution of rare variation to variance in liability, while SR carried out imputation to IKG. TG, CAB, and ABL carried out statistical analyses under the guidance of BD, LK and KR, while YP, SS, OS, AR, and CHM carried out epidemiological analyses. BD, KR, and JDB took the lead in writing the manuscript and all authors reviewed and approved the manuscript.

Competing Financial Interests

The authors have no competing financial interests.

¹³Department of Preventive Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA

¹⁴Ray and Stephanie Lane Center for Computational Biology, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

¹⁵Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, New York, USA

¹⁶The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA

Abstract

A key component of genetic architecture is the allelic spectrum influencing trait variability. For autism spectrum disorder (henceforth autism) the nature of its allelic spectrum is uncertain. Individual risk genes have been identified from rare variation, especially *de novo* mutations^{1–8}. From this evidence one might conclude that rare variation dominates its allelic spectrum, yet recent studies show that common variation, individually of small effect, has substantial impact en masse^{9,10}. At issue is how much of an impact relative to rare variation. Using a unique epidemiological sample from Sweden, novel methods that distinguish total narrow-sense heritability from that due to common variation, and by synthesizing results from other studies, we reach several conclusions about autism's genetic architecture: its narrow-sense heritability is $\approx 54\%$ and most traces to common variation; rare *de novo* mutations contribute substantially to individuals' liability; still their contribution to variance in liability, 2.6%, is modest compared to heritable variation.

Autism is a neurodevelopmental disorder typified by striking deficits in social communication, and genetically by a mix of *de novo* and inherited variation contributing to liability. Rare variants clearly play a role, with the contribution of *de novo* variation being the most obvious and easy to characterize, but inherited variation also has a role in liability^{11,12}. The contribution from inherited common variation is less substantiated. A handful of genome-wide association studies (GWAS) have been conducted, significant findings have been few, and those are specific to a single GWAS^{13–15}. The results mirror the early GWAS for schizophrenia, which in retrospect were underpowered, as witnessed by replicable associations involving common variants found from studies of tens of thousands of subjects¹⁰. In another parallel with early GWAS for schizophrenia, one of the first rays of hope for understanding how common variants impact liability was through the use of genetic scores, which were built from a large number of common variants and were shown to predict liability reliably¹⁶. For autism, scores also predict risk¹³. Some of the SNPs conferring risk to schizophrenia appear to confer risk to autism¹⁷, a result that compliments results for copy number variants (CNV)^{18,19}.

A natural complement to genetic scores from common variants is the estimation of narrow-sense heritability from the same variants. Two recent studies estimate heritability attributable to common variation to be substantial^{9,10}, yet one estimates heritability at roughly 50% (Fig. 1a), the other at 17%. As described in the reports^{9,10}, there are several

technical reasons for these differences, one being quite different study designs and another being ascertainment. As we shall show here, 50% appears more realistic.

In any case, neither of these values approaches early twin studies, which place autism heritability close to 1 (Supplementary Table 1 for data and discussion). Still results from early studies could be compatible with estimates from common variants (Fig. 2). The key issue is that early twin studies of autism assume that the genetic covariance of monozygotic twins is determined solely by additive effects, and that non-additive and *de novo* effects on monozygotic similarity can be ignored. These are dubious assumptions, creating ample room for the discrepancy between study designs. In contrast, a recent, large study of twins places heritability at 38%²⁰ (Figs. 1a, 2) under the same assumptions.

To resolve this conundrum we are evaluating a population sample by a variety of genetic analyses to estimate the relative contribution of rare, common, inherited and *de novo* variation to overall liability. We have ascertained subjects with strict autism (“autistic disorder”) from a Swedish epidemiological sample (“Population-based Autism Genetics and Environment Study” or PAGES).

Concurrently a comprehensive study of autism in Sweden has been ongoing and it recently reported the largest study of familial risk to date²¹. This “Swedish family study”, a population-based cohort of all Swedish children born from 1982–2007 and a registry of all diagnoses prior to 2010, includes more than 1.6 million families with at least two children, yielding 5,799,875 cousin pairs, 2,642,064 full sibling pairs, 432,281 maternal half-sibling pairs, 445,531 paternal half-sibling pairs and 37,570 twins. Of the 14,516 cases of broad autism, 5,689 (39%) have a strict diagnosis. This massive homogeneous sample permits precise estimation of relative recurrence risk of autism, given the diagnosis in relatives from monozygotic twins to first cousins, and after modeling covariates such as sex, birth year, parental psychiatric history and parental age at birth. By analyzing these recurrence risks for additive and non-additive genetic effects and shared and non-shared environmental effects, the best model consists of only additive genetic and non-shared environmental effects and yields quite precise estimates of the narrow-sense heritability of autism ($h^2=54\%$, $SE=5\%$).

The Swedish family study provides a sound foundation from which to address other questions about the genetic architecture of autism using PAGES. There are no major differences in the population and samples underlying both studies. To estimate heritability for PAGES, controls were sampled from the Swedish population and both cases and controls were genotyped on a common genotyping platform. After genotyping and quality control we analyzed data from 531,906 SNPs characterized on 3046 subjects, 466 with autism and 2580 subjects¹⁰ not known to be affected.

We used GCTA²² to estimate heritability due to common variants, i.e., SNP-based heritability. To ensure that all cases and controls were essentially unrelated (no pairs with kinship greater than 5th degree relatives), 151 individuals were excluded. The resulting estimate of total variance in liability explained by measured SNPs was 49.4% ($SE=9.6$) (Fig. 1a). This heritability estimate compared remarkably well with findings based on independent data from population samples and similar methods⁹. The common variation

imparting this heritability was distributed roughly uniformly across chromosomes (Fig. 1b), an expectation of polygenic inheritance that is reflected in the significant correlation between heritability per chromosome and its size ($r=0.49$, $p\text{-value}=0.018$). Prevalence of strict autism, required to calculate heritability, was set to 0.3% (Fig. 1c; Supplementary Fig. 1) for these heritability calculations. The estimate is a lower bound for total narrow-sense heritability because it includes contributions from causal variants not tagged by the measured SNPs. Although synthetic association²³ – a pileup of rare risk variants in linkage disequilibrium with a common variant – could account for a small fraction of this heritability, it cannot be large, as described below and previously^{24,25}.

To obtain an estimate of heritability due to both common and rare variation, we next included more closely related individuals. In a traditional analysis of heritability, e.g., the Swedish family study, the relationship matrix is given. Instead we estimated these relationships from the SNPs genotyped for PAGES. Because estimates of relationships from SNP genotypes tend to be noisy, we used treelet covariance smoothing²⁶ to improve estimates of pairwise relationships, especially for more distantly related individuals, and thereby refine estimates of heritability²⁶. When we included relatives, albeit mostly distant (Supplementary Fig. 2), the estimated total heritability was 52.4% (SE=9.5). While estimates were somewhat sensitive to prevalence, the differences between SNP-based heritability versus those based on estimated relatedness were insensitive to prevalence (Fig. 1d): at any prevalence, the difference was approximately 3%. To evaluate how successful this approach could be at partitioning sources of heritability, we performed a simple simulation experiment that demonstrates that we can successfully partition heritability into the portion explained by common and rare variants (Methods).

Previous work has shown that autism heritability could fluctuate substantially when the bulk of the sample was comprised of simplex families, which lowers heritability, versus multiplex families, which raises heritability⁹. These observations are consistent with liability being a classical quantitative genetic trait. Because PAGES is population based, it has no obvious simplex/multiplex ascertainment bias. Still there could be sources of more subtle bias. Based on the conjecture that subjects with intellectual disability and autism have a greater fraction of liability determined by *de novo* variants, one obvious bias would be misclassification of individuals who are comorbid for intellectual disability and autism as one or the other. To evaluate this issue we first determined the diagnostic classification of Swedes, according to governmental records. In this population 43.6% of subjects with autism also have intellectual disability, a rate comparable to other populations (note that strict autism has a higher rate of comorbid intellectual disability than broad autism). Next, to determine if IQ has a substantial impact on heritability, we contrast two estimates from data from the Autism Genome Project: using the full sample, heritability equaled 51.1% (SE=4.8, N=2097); whereas, for subjects with IQ > 80, it was slightly but not significantly larger (59.3%, SE=7.8, N=871). Subjects in the sample meet broad criteria for an autism diagnosis; for the subset of individuals given a strict diagnosis, heritability was 52.3% (SE=6.2, N=1242).

Next we asked how much of the variance in liability to autism could be explained by *de novo* mutations, applying the standard liability model to reported rates of *de novo* CNVs and

loss-of-function (LoF) mutations in autistic subjects and their siblings from the Simons Simplex Collection. In this sample, structured to enrich for *de novo* CNV and LoF mutations, their contribution to the variance in liability is 2.6% (Supplementary Note; Supplementary Tables 2–3). Yet *de novo* events can have a large impact on liability and 14% of subjects carry such mutations: roughly 80% of subjects that are carriers of a *de novo* CNV would not be affected if they were not carriers; likewise, for carriers of *de novo* LoF mutations, 57% would not be affected (Supplementary Note).

The estimate of heritability could indirectly include dominant or non-additive effects but should not include the impact of recessive inheritance. A recent study estimated the contribution from rare, recessive variation to be about 3%²⁷, a contribution similar to that from additive effects of rare variants. Rare hemizygous LoF mutations accounted for another 2% of liability.

We conclude that inherited rare variation explains a smaller fraction of total heritability compared to common variation (Fig. 2). While uncertainty is inherent in all of these estimates (Fig. 1), results converge on total heritability in the range of 50–60% with common variants explaining the bulk of it. Our analyses illustrate an approach to identify the contribution of rare and common variation to the heritability of any phenotype²⁸. Estimating the total contribution of genetic variation to variation in liability, which includes non-additive effects and *de novo* variation, is more challenging. If the only non-additive effects of genes were due solely to recessive inheritance, this would add rough 5% to the total, but that estimate could be low based on both theoretical and empirical grounds^{29,30}. And, while 14% of affected subjects carry *de novo* CNV and LoF mutations, the contribution of these mutations to the variance in liability is only 2.6%. Summing these estimates suggests that genetic variation accounts for roughly 60% of the variation in risk for autism in Sweden, implying that the majority of risk is due to genetic variation.

By contrast, a recent twin study finds that shared twin environment accounts for the majority of the variation in risk, 55%, based on a population sample of Californians from the USA. These different populations could have different genetic architectures or there could be an unknown sampling bias. Alternatively the California study fits many parameters to a relatively small data set – concordance rates on 54 monozygotic pairs and 138 dizygotic pairs – from which the study selects the best model based on statistical criteria. For small samples, however, the correct model, the one truly generating the data, can be quite different in structure from the selected model and yet the two can have only small differences in likelihood. It is possible that the distinct conclusion of the California study, versus others of its design, is due to a modest stochastic difference that altered model selection. In this regard a cautionary note for all such studies, including ours, is worthwhile: while we assume here a simple model structure, ours is but one of many possible models that could underlie trait covariance (e.g.³¹); the assumed model can alter inference, sometimes substantially; and many of these models can fit the data almost equally well. Nonetheless, that all Swedish studies, regardless of design, converge on similar estimates of heritability lends strong support for our conclusion that the bulk of risk for autism arises from genetic variation.

Data used in the preparation of this article reside in the NIH-supported National Database for Autism Research (NDAR), in [dataset identifier].

Online Methods

Ascertainment of subjects

We developed an epidemiological sample of autism or, more precisely, autistic disorder taking advantage of the detailed birth and medical registries and universal access to health care. Our sample frame was the medical birth register including all births in Sweden, where there is mandatory screening of all children at age 4 for neurodevelopmental disorders. The medical registries included all individuals diagnosed with autistic disorder at any time. Cases with autistic disorder (ICD-9 codes: 299A or ICD-10 codes: F84.0–F84.1), henceforth autism, were identified from the Swedish National Patient Register (NPR). Controls free from schizophrenia or bipolar disease were recruited from the general Swedish population matched by county, gender and birth year. Prevalence was 30 cases per 10,000 for autism and approximately 100 cases per 10,000 for the more inclusive broad autism diagnosis (Supplementary Figure 1). Inclusion criteria were diagnosis of autism in the NPR; born in Sweden; both parents born in a Nordic country; between age 10–65 years; and signed consent by a parent or a legal guardian (or by the subject, when possible and appropriate). Exclusion criteria were individuals with a diagnosis of autism, but who also had a genetic disorder known to be associated with autistic features (e.g., Fragile X, Down and Klinefelter syndromes); or medical or psychiatric history that could mitigate a confident diagnosis of autism. In this way 536 autism subjects were recruited from 12 counties in Sweden.

Genetic characterization

Samples were genotyped on the Illumina HumanOmniExpressExome BeadChip. Here we analyzed only the OmniExpress content of > 715,000 single nucleotide polymorphisms (SNPs) across the genome. Duplicate samples and samples with genotype completion rates < 98% were removed, resulting in a final sample of 3046 individuals, of which 466 were autism cases and 2580 were controls. We controlled for more subtle population structure using seven significant dimensions of ancestry as covariates in all subsequent analyses (n=3044, omitting one of each set of twins).

Heritability

To estimate heritability the Swedish family study relied on an extended sibling design, which included full siblings, half siblings, cousins and twins. The design facilitated estimation of additive and non-additive genetic sources of variance, as well as shared and non-shared environmental sources of variance.

For all genetic analyses of heritability, SNPs with minor allele frequency $MAF > 0.05$ were evaluated using the program GCTA²² to produce an estimated genetic relationship matrix (GRM). As described further in the Supplementary Note, we then modeled the case-control status via the mixed linear model $y = X\beta + g + e$, where y is the vector of case-control status, β is the vector of coefficients for the fixed effects (7 ancestry dimensions) with associated design matrix X , g , is the vector of random additive genetic effects associated with SNPs,

and e is a vector of random errors, which were assumed to be independent. To obtain estimates of the heritability, variance of the phenotypes was expressed as $Var(y) = A\sigma_g^2 + I\sigma_e^2$ (where A is the genetic relationship matrix (GRM) and I is an identity matrix, while σ_g^2 and σ_e^2 partition the total phenotypic variation into pieces attributable to additive genetic effects and random error, respectively) and heritability calculated as $h^2 = \frac{\sigma_g^2}{(\sigma_g^2 + \sigma_e^2)}$ on the observed scale, which is transformed to the liability scale as a function of the population prevalence (K).

To estimate heritability due to common SNPs we used GCTA, applying to a GRM calculated based on a sample of essentially unrelated individuals ($A < .025$). To estimate *total* narrow-sense heritability we included all sampled individuals, computed the GRM, smoothed this matrix using Treelet Covariance smoothing²⁶ (TCS), and then computed heritability from the GCTA package. See Supplementary Note for implementation of TCS. We used simulations to assess accuracy of this procedure of estimating heritability (see Supplementary Note for complete details). We started with phased genomes (haplotypes) of individuals from the HapMap 3 database, selecting two populations of European ancestry (CEU and TSI), and utilizing the available haplotypes we generated a large sample of haplotypes, representative of those that might be sampled from unrelated founders of a population. After generating haplotype pairs, chromosomes were randomly assigned to founders in each of 100 families and the founder chromosomes dropped through a five-generation pedigree. One hundred sets of independent pedigrees, including 20 individuals sampled per pedigree, were combined to generate the full genotype sample of size 2000. For the given set of genotypes, 50 independent vectors of phenotypes were simulated. For each simulation a random set of causal variants were chosen: 1000 rare (MAF < .01) and 1000 common variants. These two classes of SNPs generated 25% and 50% of the heritability (h^2), respectively, for a total of $h^2 = 75\%$. Using GCTA to estimate h^2 solely from common variant genotypes – after removing relatives – the mean $h^2 = 50.7\%$ (SE = 3.5). That this estimate is close to the simulated value for common variants suggests that the impact of synthetic association is minimal. Next, applying GCTA to genotypes from common variants and the full sample, including relatives, produces mean $h^2 = 72.4\%$ (SE = 1.2) with TCS and mean $h^2 = 70.6\%$ (SE = 1.2) without TCS. Both capture most of the h^2 due to rare variation.

Impact of clinical features on estimates of heritability, exemplified by diagnosis and intellectual function

Consistent with quantitative genetics theory it has already been shown that families who are multiplex for autism carry a larger load of liability alleles relative to simplex families (defined as families with only one affected subject within the set of first and second degree relatives). Clinical phenotypes could also affect heritability/genetic load, although how much impact they might have is an open question. To evaluate this question we evaluated two phenotypes thought to have major impact on the genetics of autism, namely diagnosis *per se* and higher versus lower functioning, as measured by IQ. First, by linking registry data from Sweden, an estimate of the fraction of subjects with autism and intellectual disability (IQ < 70) was obtained to determine its comparability to other population samples.

To assess the impact of diagnosis we use Autism Genome Project (AGP) data and follow the AGP by analyzing strict autism, as defined by meeting criteria for autism on the ADI-R and ADOS, versus broad autism, which includes autism disorder and subjects who meet looser criteria for a spectrum diagnosis (see Supplementary Note). For IQ we target subjects with IQ ≥ 80 , beyond the bound for intellectual disability. After quality control, there were 2097 AGP cases¹³ and 1663 HABC controls⁹ genotyped for 828,352 markers. After analysis using GCTA we observed heritabilities of $51.1 \pm 4.8\%$, $52.3 \pm 6.2\%$, and $59.3 \pm 7.8\%$ for broad autism, strict autism, and autism with IQ ≥ 80 , respectively.

Meta analysis of heritability

A meta estimate of h^2 due to common variants can be derived by taking a weighted average of two estimates of this quantity obtained from two independent samples: the PAGES study ($h^2 = 49.4\%$, $SE = 9.5$) and, the other, 1242 strict autism subjects from AGP data ($h^2 = 52.3\%$, $SE = 6.2$). We did not use the estimate based on the SSC (provided in Figure 1) because the SSC ascertainment of only simplex families induces a negative bias on the estimate. Meta-analysis produced $h^2 = 51.4\%$ ($SE = 5.2$) and corresponding 95% confidence interval (41.0–61.8). Contrasting this with the total h^2 obtained from the Swedish family study ($h^2 = 54\%$, $SE = 5$) produced an estimate of h^2 due to rare variants $h_r^2 = .2.6\%$ ($SE = 7.2$, $CI = (0, 17\%)$).

Estimating the contribution of *de novo* mutations and heritable variation to liability and variation in liability to autism

For motivation and computational methods, see Supplementary Note. To estimate the variance in liability explained by *de novo* variation, results from the SSC were analyzed, contrasting the rate of *de novo* copy number variants (CNVs), loss of function (LoF) mutations and missense mutations. All three have been shown to be significantly in excess in autism probands, relative to their unaffected siblings, although not all studies found *de novo* missense variation to be in excess^{1–5,7,34}. For inference, we assumed the excess proportion of cases carrying *de novo* mutations, relative to control siblings, conferred liability.

De novo CNVs—As described further in the Supplementary Note, 75 *de novo* CNVs were found in 858 probands and 19 *de novo* CNVs in 863 sibling controls³⁴ (relative risk=4.25). Assuming an ‘exposure rate’ of 0.022 (=19/863), the classical liability model determined that *de novo* CNVs accounted for 1.46% of the variability on the liability scale.

De novo LoF mutations—72 of 599 autism probands had a *de novo* LoF mutation compared with 32 of the 599 sibling controls^{1,4,35} (relative risk =2.42). For exposure rate of 0.053, *de novo* LoF mutations accounted for 1.11% of the variance in liability.

De novo missense mutations—253 out of 599 probands had at least one *de novo* missense mutation compared with 238 of 599 sibling controls (relative risk =1.11). For exposure rate of 0.397, *de novo* missense mutations accounted for negligible variance in liability (0.04%).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

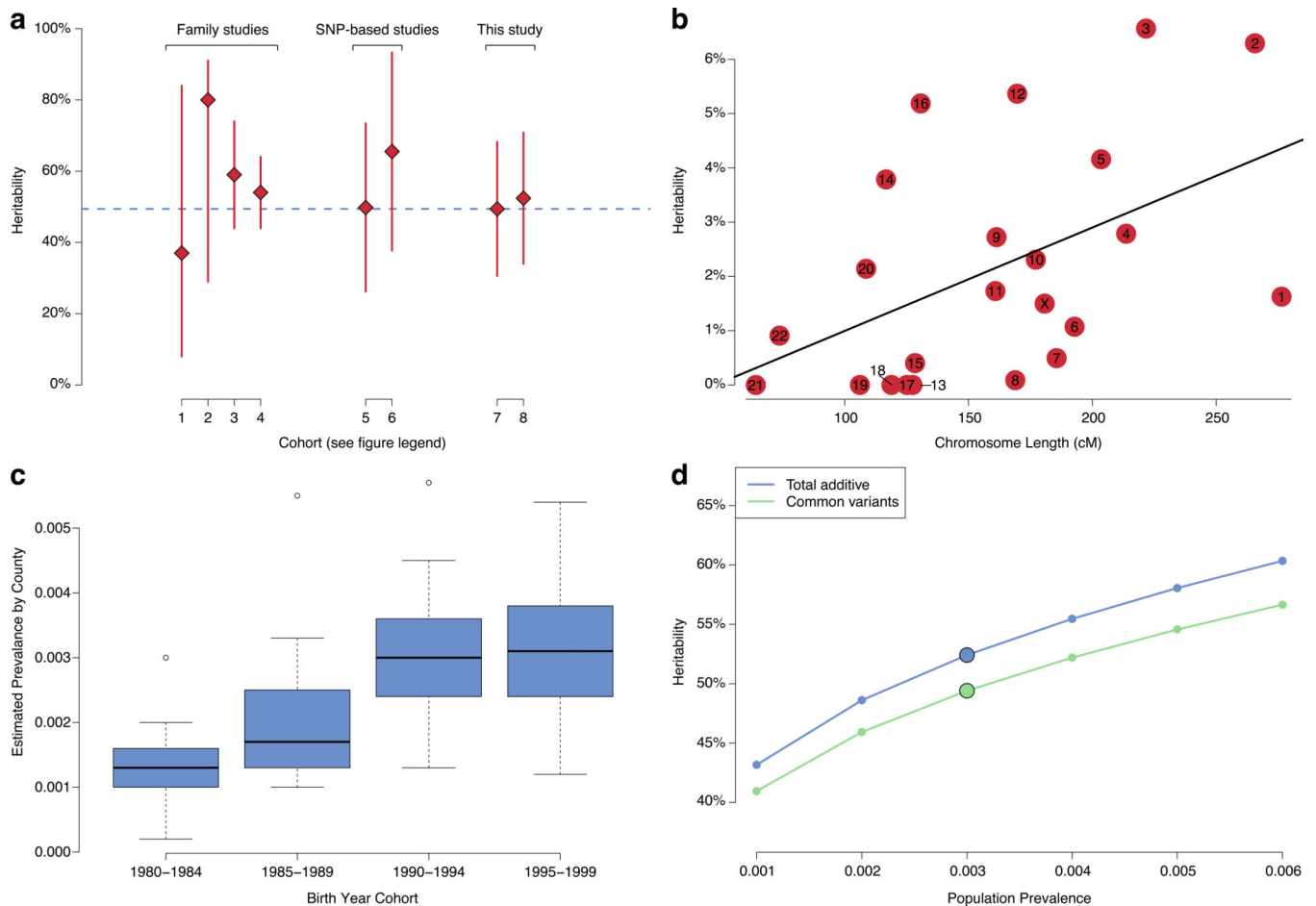
Acknowledgements

This study was supported by National Institute of Mental Health (NIMH) Grants MH057881 and MH097849, and also in part through the computational resources and staff expertise provided by the Scientific Computing Facility at the Icahn School of Medicine at Mount Sinai. We thank the Mount Sinai Genomics Core Facility for carrying out Illumina bead array genotyping. We thank Drs. David Cutler, Mark Daly and Shaun Purcell for comments on the manuscript and Drs. Daly and Patrick Sullivan for facilitating access to control samples, collected and genotyped by Drs. Daly, Hultman, Sklar, and Sullivan, and supported by NIMH grants MH095034 and MH077139. We also thank the nurses, Ann-Kristin Sundberg and Ann-Britt Holmgren, for their hard work in collecting the samples. This manuscript reflects the views of the authors and does not reflect the opinions or views of the NIH.

References

1. Iossifov I, et al. De novo gene disruptions in children on the autistic spectrum. *Neuron*. 2012; 74:285–299. [PubMed: 22542183]
2. Neale BM, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*. 2012; 485:242–245. [PubMed: 22495311]
3. O’Roak BJ, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature*. 2012; 485:246–250. [PubMed: 22495309]
4. Sanders SJ, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*. 2012; 485:237–241. [PubMed: 22495306]
5. Sebat J, et al. Strong association of de novo copy number mutations with autism. *Science*. 2007; 316:445–449. [PubMed: 17363630]
6. Glessner JT, et al. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature*. 2009; 459:569–573. [PubMed: 19404257]
7. Sanders SJ, et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron*. 2011; 70:863–885. [PubMed: 21658581]
8. Pinto D, et al. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am J Hum Genet*. 2014; 94:677–694. [PubMed: 24768552]
9. Klei L, et al. Common genetic variants, acting additively, are a major source of risk for autism. *Mol Autism*. 2012; 3:9. [PubMed: 23067556]
10. Lee SH, et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet*. 2013; 45:984–994. [PubMed: 23933821]
11. State MW, Levitt P. The conundrums of understanding genetic risks for autism spectrum disorders. *Nat Neurosci*. 2011; 14:1499–1506. [PubMed: 22037497]
12. Devlin B, Scherer SW. Genetic architecture in autism spectrum disorder. *Curr Opin Genet Dev*. 2012; 22:229–237. [PubMed: 22463983]
13. Anney R, et al. Individual common variants exert weak effects on the risk for autism spectrum disorders. *Hum Mol Genet*. 2012; 21:4781–4792. [PubMed: 22843504]
14. Wang K, et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature*. 2009; 459:528–533. [PubMed: 19404256]
15. Weiss LA, et al. A genome-wide linkage and association scan reveals novel loci for autism. *Nature*. 2009; 461:802–808. [PubMed: 19812673]
16. International Schizophrenia, C. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009; 460:748–752. [PubMed: 19571811]
17. Cross-Disorder Group of the Psychiatric Genomics, C. *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet*. 2013; 45:984–994. [PubMed: 23933821]

18. Stefansson H, et al. CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature*. 2014; 505:361–366. [PubMed: 24352232]
19. Talkowski ME, et al. Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. *Cell*. 2012; 149:525–537. [PubMed: 22521361]
20. Hallmayer J, et al. Genetic heritability and shared environmental factors among twin pairs with autism. *Arch Gen Psychiatry*. 2011; 68:1095–1102. [PubMed: 21727249]
21. Sandin S, et al. The familial risk of autism. *J Am Med Assoc*. 2014 in press.
22. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011; 88:76–82. [PubMed: 21167468]
23. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. *PLoS Biol*. 2010; 8:e1000294. [PubMed: 20126254]
24. Orozco G, Barrett JC, Zeggini E. Synthetic associations in the context of genome-wide association scan signals. *Hum Mol Genet*. 2010; 19:R137–R144. [PubMed: 20805105]
25. Wray NR, Purcell SM, Visscher PM. Synthetic associations created by rare variants do not explain most GWAS results. *PLoS Biol*. 2011; 9:e1000579. [PubMed: 21267061]
26. Crossett A, Lee AB, Klei L, Devlin B, Roeder K. Refining genetically inferred relationships using treelet covariance smoothing. *Annals of Applied Statistics*. 2013; 7:669–690. [PubMed: 24587841]
27. Lim ET, et al. Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. *Neuron*. 2013; 77:235–242. [PubMed: 23352160]
28. Agarwala V, Flannick J, Sunyaev S, Altshuler D. Evaluating empirical bounds on complex disease genetic architecture. *Nat Genet*. 2013; 45:1418–1427. [PubMed: 24141362]
29. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A*. 2012; 109:1193–1198. [PubMed: 22223662]
30. Lynch, M.; Walsh, B. *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer Associates; 1998.
31. Devlin B, Daniels M, Roeder K. The heritability of IQ. *Nature*. 1997; 388:468–471. [PubMed: 9242404]
32. Lichtenstein P, Carlstrom E, Rastam M, Gillberg C, Anckarsater H. The genetics of autism spectrum disorders and related neuropsychiatric disorders in childhood. *Am J Psychiatry*. 2010; 167:1357–1363. [PubMed: 20686188]
33. Lundstrom S, et al. Autism spectrum disorders and autistic like traits: similar etiology in the extreme end and the normal variation. *Arch Gen Psychiatry*. 2012; 69:46–52. [PubMed: 22213788]
34. Levy D, et al. Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron*. 2011; 70:886–897. [PubMed: 21658582]
35. Willsey AJ, et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell*. 2013; 155:997–1007. [PubMed: 24267886]

**Figure 1.**

Results for PAGES (Population-based Autism Genetics and Environment Study), the Swedish study of the heritability of autism. **(a)** Heritability estimate (95% confidence interval) compared across study designs and analytical methods. Horizontal reference is the PAGES estimate of heritability from SNP genotypes. Twin studies: 1 California twins for strict autism (95% confidence interval: 8–84%), the largest twin study to date using diagnosis only²⁰; 2, Swedish twins 9–12 years old (95% confidence interval: 29–91%)³²; 3, Swedish twins 9–12 years old characterized for a quantitative measure of autism (most extreme cutoff; 95% confidence interval: 44–74%)³³. SNP-based estimates of heritability: 4, Swedish family study (95% confidence interval: 44–64%)²¹; 5, simplex cases versus population controls (95% confidence interval: 26–73%)⁹; 6, multiplex autism cases versus population controls (95% confidence interval: 38–93%)⁹. SNP-based estimates from the PAGES study, assuming prevalence $K=0.3\%$; 7, heritability due to common variants using autism cases versus population controls (95% confidence interval: 31–69%); 8, total narrow-sense heritability due to both common and rare variation using smoothed estimates of relatedness (95% confidence interval: 35–71%). **(b)** Heritability per chromosome versus length in cM. **(c)** Prevalence by county for all 21 counties in Sweden tallied by birth year cohort. Each boxplot has a lower tail that extends from the minimum county-level prevalence to the 25th percentile; a central box that begins at the 25th percentile and ends at

the 75th percentile, with a line demarcating the median prevalence; and an upper tail that extends from the 75th percentile to either (1) the maximum county-level prevalence (in the absence of any outliers) or (2) to a value of the 75th percentile + 1.5 times the vertical distance covered by the box – in this case, any outliers that exceed this end of the tail are noted by circular points on the plot. **(d)** PAGES heritability versus population prevalence of autism for two estimators of heritability: case-control contrast using SNP genotypes (green); total heritability from smoothed relationships amongst subjects, based on SNP genotypes (blue). Beyond the analysis of the PAGES study we applied meta-analysis of selected h^2 estimates (Methods) to obtain $h^2 = 51.4\%$ (SE = 5.2), which corresponds to a 95% confidence interval of (41.0, 61.8). Contrasting this with the comprehensive estimate of h^2 obtained from the Swedish family study ($h^2 = 54\%$, SE = 5) produces an estimate of h^2 due to rare variants: $h^2 = 2.6\%$ (SE = 7.2, 95% confidence interval: 0–17%). Hence we conclude that common variants explain the bulk of the heritability for autism, at least 41% of the variability, and rare variants explain at most 17%, based on the upper and lower bounds of the respective 95% confidence intervals.

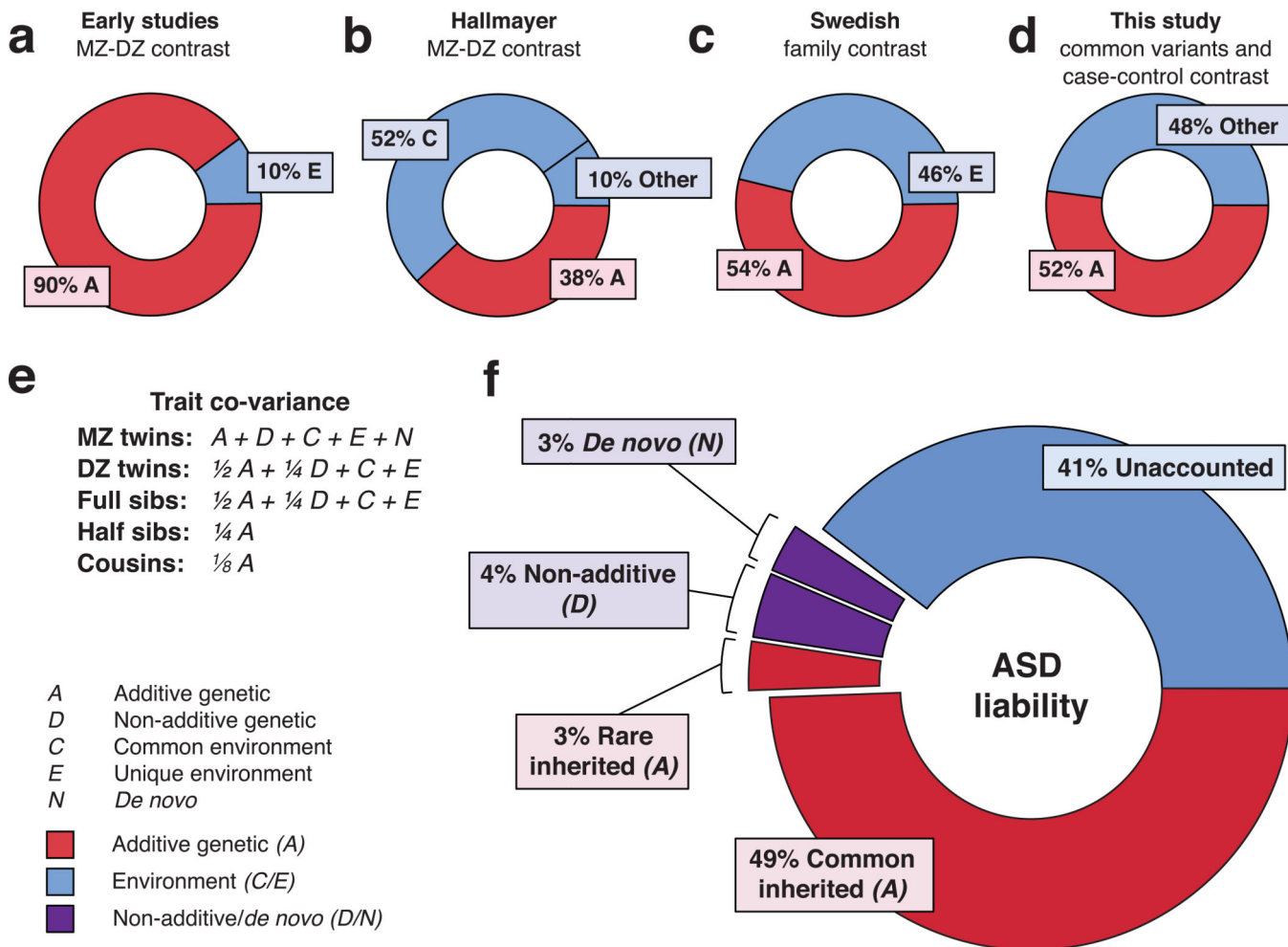


Figure 2. Results regarding the genetic architecture of autism spectrum disorder. The variance of autism liability is determined by genetic and environmental factors. The genetic factors include: ‘A’ additive effects, ‘D’ non-additive effects (dominant, recessive, epistatic), and ‘N’ *de novo* mutations. The environmental factors are split between ‘C’ common or shared environment and ‘E’ stochastic or unique environment. (a) Early autism twin studies estimate ‘A’ from the contrast of monozygotic (MZ) and dizygotic (DZ) correlations while assuming that ‘D’ and ‘N’ are zero. These are common assumptions for ‘ACE’ heritability models, but are unlikely to be appropriate for autism. (b) Applying the ACE model to the largest autism twin study to date yields a lower estimate of additive heritability. (c) Heritability results using a more extensive set of family relationships and based on much of the population of Sweden. (d) Results from the PAGES study (see Fig. 1). (e) Contribution of the various factors to the variance of autism liability according to family relationship. *De novo* variation should not be shared in dizygotic twins, and when it appears to be, it is almost surely inherited variation from a parent with gonadal mosaicism because the chance of the same mutation appearing *de novo* in the dizygotic twins is negligible. Most twin studies assume ‘C’ is the same for monozygotic and dizygotic twins, although that approximation has been debated. Of note, the excess covariance of monozygotic twins

relative to dizygotic twins is $1/2 A + 3/4 D + N$ as opposed to the $1/2 A$ assumed in the ACE model. (f) Synthesis of results for the genetic architecture of autism.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript