bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Assessment of functional convergence across study designs in autism

Authors:

Sara Ballouz,[1]
Jesse Gillis*,[1]

Affiliations:

1 The Stanley Institute for Cognitive Genomics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 11724, USA

* Corresponding author: Dr J Gillis, The Stanley Institute for Cognitive Genomics, Cold Spring Harbor Laboratory, Cold Spring Harbor 11724, NY, USA

JG: jgillis@cshl.edu

SB: sballouz@cshl.edu

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Background

Disagreements over genetic signatures associated with disease have been particularly prominent in the field of psychiatric genetics, creating a sharp divide between disease burdens attributed to common and rare variation, with study designs independently targeting each. Meta-analysis within each of these study designs is routine, whether using raw data or summary statistics, but no method for combining the functional conclusions across study designs exist.

Method

In this work, we develop a general solution which integrates the disparate genetic contributions constrained by their observed effect sizes to determine functional convergence in the underlying architecture of complex diseases, which we illustrate on autism spectrum disorder (ASD) data. Our approach looks not only for similarities in the functional conclusions drawn from each study type individually but also those which are consistent with the known effect sizes across these studies. We name this the "functional effect size trend" and it can be understood as a generalization of a classic meta-analytic method, the funnel plot test.

Results

We detected remarkably significant trends in aggregate (p~ 1.92e-31) with 20 individually significant properties (FDR<0.01), many in areas researchers have targeted based on different reasoning, such as the fragile X mental retardation protein (*FMRP*) interactor enrichment (FDR~0.006). We are also able to detect novel technical effects and we see that network enrichment from protein-protein interaction data is heavily confounded with study design, arising readily in control data.

Conclusions

As a novel means for meta-analysis across study designs, this method reveals a convergent functional signal in ASD from all sources of genetic variation. Meta-analytic approaches explicitly accounting for

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

different study designs can be adapted to other diseases to discover novel functional associations and increase statistical power.

Keywords

autism spectrum disorder; rare variation; common variation; loss-of-function; recurrence; effect sizes; functional enrichment; gene candidate score; meta-analysis; funnel-test

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Background

Autism spectrum disorder (ASD [MIM 209850]) is a neurodevelopmental disease commonly characterized by behavioral traits such as poor social and communication skills[1]. In more severe cases, ASD is comorbid with mild to severe intellectual disability, facial and cranial dysmorphology and gastrointestinal disorders. Perhaps because of grouping these multiple and sometimes distinct phenotypes into one disorder, and the complexity of behavior as a trait, understanding the genetic architecture of this cognitive disease is non-trivial [2]. The genetic component of ASD is estimated to be 50-60%[3], however there are still a substantial number of cases where the underlying genetic factors of the disease are unknown. Researchers have used chromosomal analysis arrays[4-7], genotyping arrays[8-11], whole-exome sequencing (WES)[12-17], and whole genome sequencing (WGS)[18, 19], to identify risk loci and alleles. What has been demonstrated through these analyses is that all humans carry a substantial mutational load, some of which is most likely deleterious[20]. There are varying levels of genetic risk, affected in part by epigenetic and environmental factors, that all contribute to the variation in penetrance of disease gene candidates [21]. Typically, overlapping functional properties of candidate disease genes are exploited to sort out the positive and negative results. Candidate genes are prioritized based on enrichment analyses in pathways related to the phenotype (e.g., neuronal activity regulation) or some other disease feature shared by the genes (e.g., expression in the brain). If these methods return no significant results, more complex methods are performed to extract common features from the disease gene set [22], such as co-regulatory module detection from co-expression networks[23] or binding from protein-protein interaction (PPI) networks[24]. Without further rigorous molecular validation, it is difficult to establish whether these signals are disease specific, or artifacts of the study design or ascertainment method of the patients and populations.

Ideally, the combination of these study designs would explain both the contribution of rare and common variation to autism, i.e., the genetic architecture of the disease. In particular, one would like to combine

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the results across these to understand the functional mechanisms of the disease and the genetic risks involved. Yet doing so is challenging, mainly because the different study designs are carefully calibrated to enrich for meaningful signals and thus reflect different effects which cannot be naively compared – common variants are limited to regions of the genome with known variation (a SNP is known) but only reach significance with large numbers, while rare or ultra-rare variants are conditioned on not being in this list of common variants. Trio and quad studies are used mainly in WES and WGS study designs, while large case and control cohorts are required for signals in genome-wide association studies (GWAS). For each study design, we are asking distinct questions that relate to the population prevalence, disease mechanism, burden and risk. These different study designs create a structured dependency among the resulting candidate gene sets: not only in the relative contribution of the genes and variants they implicate, but also in their false discovery rates, technical errors, and ascertainment biases. Rather than treating discrepancies in study designs as a basis for argument within the field (e.g., see [25, 26]), they should be appreciated as a necessary consequence of study design, which must be exploited to uncover a greater portion of the genetic architecture. Like blind men describing an elephant, the different results still reflect a unified whole, if only they could be combined. We seek to unify the individual functional signals meta-analytically, in a way which will identify individual biases and take advantage of the same biological features to which study designs are tailored.

What we can exploit and also agree upon are disease burdens or effect sizes for different candidate disease genes, as measured by their variant class or mutation type, and disease risk. For instance, disease candidates from GWAS have low relative risks (and therefore low effect sizes) as they are inherited common variation in the population. On the other hand, *de novo* mutations are a form of genetic variation which evolutionary forces have not had time to act upon[27], and are of high risk (and high effect sizes) i.e., a mutation generally implies disease. Studies also suffer from type I errors (false positives), and this too should be reflected in an aggregate disease signal of the candidate genes, as

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

quantified by their common functional properties. A set of genes with *de novo* mutations will show a strong aggregate disease signal, while we might expect a weak signal from the gene candidates from GWAS [28]. Measuring their 'functional' effect size, as determined by a gene set enrichment test or network analysis, we can thus exploit our knowledge of gene candidates' effect sizes and false positive rates. For a real disease property, we expect the correlation between gene set effect size and functional effect size to be strong, and for a weak or artifactual property, we expect no significant correlation. We call the presence of this tendency the 'functional effect size trend', and it can be rigorously assessed by repurposing fundamental meta-analytic techniques.

A simple and common method to detect bias in meta-analytic studies, the funnel graph [29], is based in a similar comparison of effect sizes to a measure of precision. We illustrate this in **Fig 1**. The underlying idea is to visualize the relationships between effect size estimates and sample size in the individual studies of a meta-analysis. This is done by plotting the odds ratio against the standard error or other related measure of variance (e.g., precision, standard error ratios). An inverted symmetric funnel would indicate that the data has no study selection biases. Small studies with small or no effect, as they are typically underpowered, should sit close to zero on both axes. Larger studies that are effective (high effect sizes) have low variance. However, a plot with asymmetry indicates likely systematic or publication biases and this is measured by either a rank correlation test [30] or a regression analysis [31]. The elegance of the test is in its simplicity: a correlation will indicate that there is asymmetry and bias, while symmetry implies heterogeneous data and publication results. There are two points of generalization from the traditional funnel test to meta-analysis across study designs [16]. First, we are not looking to a fixed effect size (no correlation) but to a known trend in effect sizes (positive correlation). Second, we consider many functional tests and then determine the subset which are significantly concordant with the effect sizes (after multiple hypothesis test correction). Our analysis is a generalized complement to the funnel test: rather than looking for correlations with bad properties

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

when we know there should none, we look for correlations with good properties when we know there should be one. This lack of heterogeneity in the funnel test is similar to their being hidden classes within the data – i.e., study designs – which we can detect [32].In both cases, though, expectations about effect size trends are being exploited, it is just that the expectations are different due to the differing types of variation alternate study designs target as causal for autism.

In order to integrate the varying data and extract the convergent disease functional signal, we have run a meta-analytic study on ASD candidates across numerous genetic studies and tested a wide range of gene properties. We first construct several disease gene candidate collections, with genes of varying levels of risk, as determined by their odds ratios and relative risks. On every gene collection, we run a number of analyses, calculating the functional effect size using standard enrichment methods, and more complex network analysis enrichments. By exploiting trends in targeted genetic variation and their known effect sizes, we demonstrate it is possible to discriminate biologically convergent signals from technical artifacts at a very fine resolution. The disease properties with strong trend signals are largely consistent with the known literature on ASD (e.g., FMRP interactor enrichment) but we also see a few otherwise interesting properties are unlikely to be disease specific, particularly protein-protein interaction networks and some co-expression networks, which extract artifactual signals from the study design. Our focus here is on autism due to our interest in the disorder, and also its phenotypic and genetic heterogeneity, but the method is generally applicable to determine functional replication across studies in a formal way that is absent in the field.

Methods

Study overview

An overview of our study design and method is shown in **Fig 2**. Briefly, we start by characterizing the ASD gene sets collected for this analysis (**Fig 2**, part 1). Each study's results (rare and common variants)

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

were collapsed individually into a set of genes, with an estimated average effect size for that candidate set (**Fig 2**, part 2). To combine the implicated variants and obtain a functional signal from the disparate studies, we calculate whether there is a functional effect which correlates with the estimated effect size of that variant class. More specifically, the set of genes with high effect sizes have strong relative functional effect sizes as measured by a functional enrichment of some disease property across them, and those with low effect sizes, have weaker functional signals (e.g., statistical overlaps with known functions). We apply this test to numerous functional properties on candidate gene sets from a variety of study designs (**Fig 2**, part 3). We then use the resulting functional properties that are consistent with this trend to construct a unifying gene score for autism specific genes (**Fig 2**, part 4). Throughout our work we refer to the "effect size" as the disease burden or risk of a gene candidate (or the average of such values within a gene set), and the "functional effect size" as the significance of a functional test for a disease gene set after controlling for the set size.

Data collection

Disease gene candidate sets

We first collected candidate disease gene sets from available autism studies. We selected the largest study of whole-exome sequencing (WES) of families from the Simons Simplex Collection (SSC)[33]. We defined different sets of genes from over 2000 gene candidates, splitting into recurrent (>2 probands having the mutation) and non-recurrent mutations, according to mutation type (loss-of-function, missense and silent mutations). We selected copy number variant (CNV) data also from the individuals in the SSC [34], and parsed it into similar sets. We then used the CNVs as parsed by Gilman et al.[35], which prioritized genes with their NETBAG algorithm. For GWAS gene sets, we searched for the term "Autism" in the GWAS catalog from NCBI (date February 2015, **Additional file 1: Table S1**). A total of 9 GWAS were available, and we used their SNP-gene mapping strategies to define our candidate sets: first

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

selected genes with associated SNPs, and then expanding that set to include the nearest gene for intergenic regions, and finally a gene set with all adjacent genes. We also filtered on significance of the SNPs (i.e., using the *p*=5e-8 threshold), and then without significance filters (selecting all reported SNPs). Each GWAS was of relatively small size (up to ~7000 individuals), and varied in strategy (family trio or case-control), and had few candidate genes, all of which were of similar effect sizes. As there were overlaps between the cohorts in the individual studies, the total number of individuals was estimated to be roughly 10,000, but over ~40K if summed naively. For the final ASD candidate list, we included OMIM[36] disease candidates (search on "Autism", date February 2015), all genes associated with "autism" reported in Phenocarta [37] (date January 2014), genes from the SFARI database[38] (date February 2015) and genes associated with "autism" from the HuGE navigator resource[39] (date Feburary 2015). We also filtered the SFARI database on "high-confidence" evidence as described in Shohat et al,[40] and included a curated ASD list from Betancur et al,[41]. All data sets are made available in the supplement and online. For additional analyses, we collected whole-exome sequencing data sets across other similar diseases (**Additional file 1: Table S5**).

Co-expression networks

The majority of recent studies used co-expression networks from BrainSpan to illustrate network convergence among disease genes of ASD (e.g., see [42-44]). In a similar fashion, we generated brain specific networks from the BrainSpan RNA-seq data (578 samples) and from an aggregate of 28 brain tissue and cell specific microarray experiments (3,362 samples). For more general networks, we used our aggregate RNA-seq and microarray co-expression networks as previously described in Ballouz et al, [45]. In brief, these are the aggregates of 50 networks (1,970 samples) and 43 networks (5,134) samples respectively, across various tissues, cell types and conditions.

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Functional gene sets

We considered common functional gene sets and neurological specific sets, as used in numerous studies, as gene sets to test for ASD candidate enrichment. These included the post synaptic density (HPSD) gene set[46], synapse sets [47], the synaptosome[48], chromatin remodelling set[49], fragile X mental retardation protein (FMRP) set[50], and gene essentiality[51]. For more standard sets, we also took the Gene Ontology [52] (GO) terms (April 2015) and KEGG pathways [53]. For each GO term, we only used evidence codes that were not inferred electronically, propagated annotations through the ontology (parent node terms inherited the genes of their leaf node terms). To minimize redundancy from GO, we restricted our enrichment analyses to GO terms groups with sizes between 20 to 1000 genes. While these GO terms and KEGG groups are used in the enrichment analyses (with the full multiple hypothesis test correction penalty), only the subset that intersected with any of the disease and control gene sets have effect size trends calculated leaving us with 69 GO groups and 23 KEGG pathways. This is primarily a visualization issue; the FDR values calculated used throughout remain the same.

Disease gene score sets

We used disease gene scoring methods that rank genes according to likely having damaging effects if they are mutated. This included the Residual Variation Intolerance Score (RVIS)[54], haploinsufficiency (HI) scores[55], and mutational rates[42].

Expression data

To obtain expression and differential expression information, we used three common and large sample size brain-specific transcriptomic sets. These included the Human Brain Transcriptome (GSE25159)[56], BrainSpan[57] and the Human Prefrontal Cortex transcriptome (GSE30272)[58]. We divided the samples into fetal (post-conception week – PCW) and post-birth stages, and performed a straightforward differential expression (DE) fold change analysis (averaging across these stages)[59].

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Protein-protein interaction data

We used all the physical protein-protein interactions from BIOGRID (version 3.2.121)[60] and created a binary protein-protein interaction network, where each protein was a node and each protein-protein interaction is an edge. Because of the sparseness of the network, we extended the network by modelling indirect connections[61], taking the inverse minimum path length between two proteins as the weighted edge, with a maximum distance of 6 jumps roughly as described in Gillis et al,[62].

Candidate gene set effect sizes and their ranks

From the total list of candidate gene sets we collated, we chose 14 representative sets for our effect size analyses (**Table 1**). For each candidate disease gene set, we ranked the set according to the overall or average "effect size" of the genes within it. To calculate this effect size for the GWAS results, we took the average odds ratios from the individual studies of each the SNP, which ranged between 1.1-1.2. For the *de novo* mutation candidates, we took the ratio of observed counts of mutations to silent mutations within the study for that class of mutations, and then the ratio of those odds between siblings to probands (as calculated in Sanders et al,[13]). For the control sets (siblings and the silent mutations), we took the effect size to be null. We then ranked the sets based on these overall effect sizes. After these calculations, we end up with three general classes: null effects (as controls), weak effects (missense and common variants) and strong effects (rarer loss-of-function and copy number variants).

Calculating functional effect sizes

Our functional tests, described below, return *p*-values which are dependent on the size of the gene set being considered. The statistical tests differ depending on the mode of analysis (e.g., enrichment or network), but by 'functional effect size' we simply mean significance (*p*-value) after correcting for the set size, typically by downsampling. For the downsampling, we took a subset of genes, recalculated the *p*-value and then took geometric means of the adjusted *p*-values. Throughout, where we write 'functional effect size' it is possible to read '*p*-value after correcting for set size'.

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Network connectivity

We measure the clustering of sets of genes within networks through the use of a network modularity calculation. We compare the degree of connections a gene has to all the genes in the network (global node degree), and to those of interest within the sub-network they form (local node degree). The null expectation is that genes will be connected equally well to genes within the sub-network as to those outside. Genes with large positive residuals have more weighted internal connections than external connections, implying a well inter-connected module. We test the significance of this distribution of residuals to a null set (random similarly sized set of genes, Mann-Whitney-Wilcoxon test, wilcox.test in R) to determine our test statistic.

Gene set enrichment testing

As a way to determine the level of enrichment of the candidate gene sets within other functional sets, we used a hypergeometric test with multiple test correction (phyper in R). The downsampled *p*-value was used as the functional effect size measure.

Disease gene property testing

For the disease gene scoring properties, we tested the significance of the scores of the candidate genes using the Mann-Whitney-Wilcoxon test (wilcox.test in R). The functional effect size was the *p*-value of this test.

Functional size effect trends

For each gene property tested, we then measured the "trend" by calculating the slope and intercept of the functional effect sizes of our gene sets, whereby the gene sets are ordered according to their effect size ranks. A positive slope is one where the function tested is correlated with our ordering, and an intercept close to zero indicating no effect or one with enrichment in the controls. We computed this using both ranked and unranked data, as well as correlations. We generally report results in terms of

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

slopes to capture the degree of variation, but the significant subsets identified are generally robust to choice of measurement metric, as the three metrics are well correlated themselves (r=0.7 with the raw slopes and r=0.93 with the ranked slopes).

Aggregate gene score

We then selected the top ten disease properties that had high and predictive positive slopes (**Additional file 2: Table S2**). We then took all known protein coding genes from GENCODE [63] (version 19, approximately 19,000 genes) that had at least one annotation across all our properties. For each disease gene property, we gave each gene a ranking within that property. For gene scores (i.e., mutational rates or the RVIS), this was simply the rank of their score. For the gene sets, we ranked the genes and gave a tied score to the NAs (i.e., GO terms, synaptic sets). For the networks, we used the node degree of the genes as their rank, but as their "real" score is dependent on a gene set, this was only a proxy measure.

Results

Little overlap of autism candidates across gene sets of different effect sizes

Genes with loss-of-function *de novo* mutations are not unusually likely to be found in the GWA studies, with 4 candidate genes overlapping those two sets (**Fig 3,** hypergeometric test p=0.62). Interestingly, the more recurrent genes in the loss-of-function *de novo* set, the more unlikely they are to be found in other gene sets (**Additional file 1: Fig S1**). For gene sets with the lower average effect sizes (e.g., the genes with missense mutations), their overlap with other gene sets is greater, in particular with the control sets (**Fig 3,** hypergeometric tests p~4.4e-3 to 2.4e-6). The *de novo* variants are conditioned on being rare (low frequency) and novel by not appearing in the parents. The SNPs used in GWAS are generally conditioned on being common by having minor allele frequencies greater than 0.05 [64]. Even if this filtering is done on the variant level, and not on the gene level, it still creates selection trends within our observations of variants and thus genes. This is possibly a version of Berkson's effect[65] – where

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

selecting for an outcome generates negative correlations between potential causes for it. An additional cause is largely technical; since we've conditioned on frequency, genes with higher mutability are depleted in our rare lists, and enriched in our common lists. Thus the lack of overlap is at least potentially not largely reflective of underlying genetics or biology, but likely due to the selection bias in obtaining them. There is also poor overlap within the rarer variation itself, for instance of genes within CNVs and those with loss-of-function SNVs (3 genes, p=0.37); there is generally a discrepancy between study designs focused on (different) sources of rare variation, and not just rare versus common. It should be noted that whether biological or technical, the lack of overlap does nothing to discredit either common or rare variation as a contributor to the disease – but highlights the need for a framework to combine and analyze the results of these studies that is aware of these biases and can distinguish biology from technical effects.

Modularity of *de novo* autism candidates exhibits a functional effect size trend

While enrichment analysis is comparatively straightforward, we demonstrate an example in **Fig 4A** using the genes with *de novo* loss-of-function mutations from Iossifov et al,[33] (341 genes) and their overlap with essential genes (see Methods). In **Fig 4A**, we represent this enrichment test as a Venn diagram of the overlap of the candidate disease gene set with the essential gene set, and calculate the significance of the overlap with a hypergeometric test (n=82, p~7.11e-11). We continue this analysis on the other candidate disease gene sets from recent ASD studies, varying across study designs and technologies (WES, GWAS and arrays). Splitting each gene set by mutational class, recurrence and gender, we perform the same hypergeometric tests. To make comparable assessments between studies and gene sets, we calculate the functional effect size by downsampling – selecting a subset of genes within that set and averaging the results over a 1000 permutations (schematic in **Fig 4B**). Taking a representative set of studies (**Table 1**), we use the degree of disease effect to rank these sets, noting that recurrence leads to a higher effect size even for variation and study designs of the same class by reducing the number of

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

false positives. Placing the controls sets on the far left, and the highest disease rank set (recurrent *de novo* loss-of-function genes) on the far right, and plotting their functional effect values, we observe an upward trend (**Fig 4C**, black line, slope=0.9, Spearman's ρ=0.9, Fisher's transformation *p*<1e-16). The slope of this trend line represents the "effect size trend", with higher slopes indicating higher functional effects.

A less common (likely due to complexity) yet important functional test is network connectivity. Genes that are co-regulated or form parts of a functional unit, protein complex or pathway, are preferentially co-expressed, and this information is captured in co-expression networks. We next demonstrate how network-style effect sizes can be similarly calculated through a modularity analysis. In **Fig 4D**, we plot the global node degrees (x-axis) against their connectivity to the remainder of genes in the set (y-axis). In the null (grey line), the genes would be connected to other autism genes in proportion to the incidence of those genes within the genome. Deviations from this null across all genes generate excess modularity within this set (studentized residuals shown in inset **Fig 4D**) and determine the statistical results reported for the set overall (Wilcoxon test). A large number of genes are highly interconnected in this set, as shown by the number of points above the line (Wilcoxon test on the studentized residuals, *p*~1.13e-42). It is important to note that this network analysis is calculated against the empirical null for each gene individually (x-axis) and so is unaffected by any gene-specific bias (such as length). Only higher-order topological properties across gene-gene relationships for a given gene can produce a signal. Even assortativity, the tendency for genes of high node degree to preferentially interact, is quite low within this data (r=0.064, see **Additional file 1: Table S3**).

As in the previous steps, we repeat the network connectivity tests across all gene sets (**Fig 4E**), also downsampling to calculate the functional effect size. Once again, gene sets with higher proportion of

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

burden genes correlate with functional effect size tests (**Fig 4E**, black line, slope=0.59, Spearman's ρ=0.59, Fisher's transformation *p*<0.03).

Autism gene sets exhibit a highly significant effect size trend across all study designs

Extending our analysis to other disease gene property tests (**Additional file 2: Table S2**), we calculated their effect size slopes (**Fig 5A**). We observe that most slopes are positive but close to zero (i.e., weak or low), with a few highly sloped lines (tail in distribution Student's T-test, *p*<1e-9). The highest sloped property was network connectivity in the BrainSpan co-expression network; however, all disease gene sets had a significant functional effect size with Brainspan, indicating that in addition to the real signal, there is a background signal affecting even control data. In particular, the signal from the silent recurrent mutations in the probands (functional effect size *p*=7.5e-7) shows that control data subject to only one study design may select genes in a highly non-random pattern. Most top scoring disease properties are consistent with the literature on autism candidates such as average RVIS and haploinsufficiency scores, along with CDS length and enrichment for FMRP interactors. RVIS scores are highly enriched in the loss-of-function recurrent set and the CNVs, but not significant in any of the other sets (**Fig 5B**); as with any meta-analysis significance in any one set is not necessary for aggregate significance. Genes with high haploinsufficiency scores – those that cannot maintain normal function with a single copy - are overrepresented in the loss-of-function recurrent genes, and there is also a significant effect in the GWAS results. Many interaction networks and traditional functional categories appear to be poor candidates to determine convergence in disease genes, as they cluster control gene sets and sets of low effects as well as those of disease genes. For instance, the extended PPI network has a high effect in the sibling controls sets (e.g., silent functional effect size *p*~1e-10, **Fig 5C**).GO terms and KEGG pathways typically do not survive correcting for multiple testing, although there is a general deviation from the null (as described in the following section) and the extremal GO functions are concordant with the known literature (e.g., GO:0051276 chromosome organization, hypergeometric test

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

p~3.96e-4 for the *de novo* recurrent set). So although effect size trends are concentrated in more clearly disease related -properties such as RVIS, traditional functional categories from, e.g., GO remain of modest use.

Functional properties cluster strongly in relation to autism

We calculated the null distributions for the variation across effect sizes in two ways: by randomly sampling genes to derive new meaningless gene sets or by permuting the estimated effect size for each real set and rerunning our analysis. The two nulls almost perfectly overlap (**Fig 6A**), indicating that significance must arise specifically from concordance across studies that is directional. Indeed, the slopes are much more likely to be positive than either randomly sampled sets, or sets where effect sizes were randomly estimated (**Fig 6A**, Wilcoxon test, *p*~1.92e-31 and *p*~1.56e-32). The long tail in raw slopes (**Fig 5A**) is visible as a much broader shift toward significance in the ranked values (**Fig 6B**), indicating the significant extremal values arise largely through effects distributed across multiple disease sets. See "robustness and relative contributions" section for more detail on statistical assessment.

Each property can be defined by its vector of effect sizes across gene sets (e.g., **Fig 5**) and so we can cluster the properties by their Euclidean distance in this space. The properties split into approximately 6 major clusters (**Fig 6B**). Clusters 1 and 2 have random signals across the sets and consist of mostly the GO, KEGG and neural gene enrichment tested properties. These are functionally enriched in some disease gene sets (e.g., the CNV and missense gene sets) but are not significantly enriched for any of the genes in the *de novo* recurrent gene set and are thus not showing a substantially positive functional effect size trend. Cluster 3 similarly contains GO groups which have some enrichment signals from the *de novo* and CNV set. These GO groups are protein demethylation and related activities (e.g., GO:0008214, GO:0006482). Cluster 4 contains an array of properties with no consistent trend across the data sets. The more interesting clusters are 5 and 6 which have the highest slopes (as depicted by the

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

dark purple scale), and where there is a stronger significant signal from the *de novo* set (white/yellow in heatmap). Cluster 5, specifically, has the most consistent trends and contains the expression analyses (overexpression and fold change) and the gene essentiality scores. Cluster 6 has the co-expression networks clustered, and the mutational probabilities.

Robustness and relative contributions

In order to determine whether the functional effect size trend rose preferentially from a subset of studies, we conducted a series of robustness analyses (**Additional file 1: Fig S2**). Ideally, the significant functional effect size trend we see is due only to effect size estimates across studies which are themselves robust. Nor do we want the trends to be strongly affected by ordering of the gene sets with similar effect sizes. Even though the average effect sizes for the GWAS sets were the same, the number of false positives within these sets varies, and this was incorporated into the ranking scheme. It is also arguable that the silent mutations in the probands may have some regulatory effect, or are false negatives. Our previous analyses ranked our gene sets with these assumptions in mind and incorporated an ordering for control sets and the GWAS sets. With this ranking of the gene sets, 20 of the functional effect size trends are significant at an FDR of 0.01 (**Additional file 1: Fig S2B,C**). If we instead ignore the ordering provided by false negative rates and regulatory impact and repeat our analysis (treating the same effect size sets as having tied ranks), there is a modest loss of significance in the distribution of slopes (7 functional tests at an FDR<0.01, **Additional file 1: Fig S2A,C**), while the overall deviation from the null remains very significant (Wilcoxon test p~4.76e-33). Generally, using ranking as we have is slightly conservative in impact, as expected, and using the raw effect sizes raises the number of significant slopes (FDR<0.01) to 23. As a more stringent test, we removed whole classes of variants from the analyses (e.g., all the controls or all the common/weaker gene sets) and calculated the trends once again (**Additional file 1: Fig S2D-F**). This is a negative control experiment in the sense that if the functional effect size trend arises meta-analytically, it should be largely robust to changing things we are

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

not certain about (e.g., as above, whether effect sizes are 1.1 or 1.09) and not robust to changing things we are certain about (common variants play some role in autism). Removing either controls, genes sets with the highest effects or the common variants from the trend analyses removes almost all the number of significant slopes (FDR 0.01, **Additional file 1: Fig S2C**), although some deviation from the null remained. When rare variants are excluded, the distribution of slopes is most similar to the null (Wilcoxon p~5.95e-6), while the total significance of the test is closest to the full version when common variation is excluded (Wilcoxon p~1.54e-26 compared to p~1.91e-31). Since our common data is likely the weakest due to the tremendous focus of autism data collection toward rare variation in the SSC, this makes sense, but common variation still contributes substantial joint signal. These tests confirm that the approximate order of gene sets by effect sizes correctly drives the results and that we are robust to minor variation in the exact effect sizes listed, but do rely on the joint use of the extremely divergent study results (rare and common) within the meta-analysis to attain significant results.

Promiscuous or absent enrichment have both historically been problematic within disease gene data; both diminish the specificity of functional results. When too many functions are returned from an analysis, we need to cherry pick and with too few, we have no “leads” and are left in the dark. We suggest that the strong aggregate effect we see and small number of significant functions is likely near the most useful and biologically plausible type of specificity for downstream analysis, as suggested by the fact that ad hoc filtering (i.e., top ten lists) usually are at about this level when not constrained by significance. Our set of functional tests and results are shown in **Additional file 2: Table S2** and the full data set is available online.

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Re-prioritizing ASD candidates emphasizes comorbidity among neurodevelopmental disorders

Our final analyses were to see if the most significant disease properties score ASD candidates in a notably distinct way. We should expect an important reordering particularly in the case of non-recurrent *de novo* derived candidate genes, where the aggregate effect size is a combination of very high effect true positives and unknown effect (likely low) false positives. Using the previous analyses, we construct a meta-gene score by first selecting a set of top properties from clusters 5 and 6 (**Fig 6A**) that held the upward trend (both high slope and high correlation) and removing redundant properties (**Additional file 2: Table S2**). In order to keep analyses straightforward and avoid overfitting, we did not weight among these functions and considered only binary functional characterizations (excluding few properties). We then ranked the genes by their aggregate rank across the number of properties they possessed for the set of functions which we had identified as having very high functional effect size trends.

To assess this score, we took 1,002 ASD candidate genes (**Additional file 2: Table S4**) from publically available databases (OMIM[36], Phenocarta[37], HuGE[39], SFARI[38] and the Betancur et al,[41] list) and ranked them according to this meta-gene score. We see a low but significant correlation with the aggregate score (Pearson's r= 0.19, Fisher's transformation p= 2.787e-09) and the incidence of these genes in the databases. Most of the top ranked candidates are comorbid with other neurological disorders, indicating that their disease candidacy is strongly associated with a cognitive disorder. The top candidate gene here is the ubiquitin protein ligase *HUWE1* [MIM 300697], which regulates neural differentiation and proliferation, and is known disease gene for mental retardation syndromic X-linked Turner type (MRXST [MIM 300706])[66] along with being implicated in schizophrenia[67] and autism[68]. *NSD1* (Nuclear Receptor Binding SET Domain Protein 1 [MIM 606681]) is implicated in Sotos syndrome [MIM 117550] and Weaver syndrome [MIM 277590], both comorbid with ASD[69], and the histone acetyltransferase p300, *EP300* [MIM 602700],[70] is involved in Rubinstein–Taybi syndrome

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(RSTS [MIM 613684]). Other top candidates included the neurexins (*NRXN2* [MIM 600566]and *NRXN3* [MIM 600567][71]) that are neuronal cell adhesion molecules and receptors which have important roles in synapse development and function. Candidates at the bottom of the list have a different disease profile. Some are uncharacterized proteins with no known functions (i.e., *C14orf64* and *C20orf144*), or have non-neuronal phenotypes. For instance, cytoskeletal protein *ODF3L2* is involved in sperm tail motility and *RHOXF1* [MIM 300446] is a transcription factor involved in reproductive processes (**Additional file 2: Table S4**).

Reassessing the *de novo* loss-of-function recurrent genes with this score, we aimed to distinguish the high effect contributing genes which are thus the best candidates for eventual recurrence as study sizes grow. The top ranked gene here is anykrin 2 (*ANK2* [MIM 106410]),a multifunctional scaffolding protein, which plays a major role in brain development, including paranodal junction assembly in the CNS[72]. Genes *KATNAL2* [MIM 614697] (katanin p60 subunit A-like 2) and *TCF7L2* [MIM 602228] (transcription factor 7-like 2) have high negatively associated property scores, and share very few properties with the other genes. *KATNAL2* has no known function, while its mouse ortholog is involved in sperm formation and male fertility[73]. *TCF7L2,* on the other hand, is a type II diabetes risk gene[74], and is implicated in numerous cancers[75]– implying it may be "good" disease risk gene overall, but not specific to ASD.

Discussion

Advances in sequencing technologies have vastly expanded our potential candidate list for autism linked genes. Different study designs target different genetic mechanisms, and so we believe will find quite distinct signals at the gene-level. This has engendered substantial disagreement among researchers as to the relative contribution and value of different study designs and their associated classes of gene variation (for instance *DISC1* [MIM 605210] in schizophrenia [25, 26]). Even within some study designs, the level of contribution of the type of variation is uncertain i.e., *de novo* or inherited truncating variants

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[76], or whether the variation is structural (i.e., CNVs, indels) or protein truncating point mutations, or both [77]. However, in most cases, common functional signals are assessed using virtually identical methods (i.e., enrichment of gene groups or network connections). Combining the candidate gene signals and looking for convergence at the functional level would be a desirable assessment, yet simply doing so by placing all candidate genes in the same 'bin' as the simplest analysis might suggest, would be clearly problematic. First, controls vary and artifactual trends may differ, so that case-control comparisons will not hold across study designs. Second, the more distributed the genic effect the study design targets, the higher the sample count necessary to obtain significance, and the more genes (weakly) implicated there will be, meaning rare variation more tightly linked to autism (but with less explanatory power in aggregate) will be dominated by diffuse trends which only link any single gene in a weak sense (the gene is more likely involved in numerous other processes). To overcome these, we have proposed conducting meta-analysis at the 'output' stage for any given study, after functional convergence is determined, and then additionally incorporating effect sizes across study designs as an additional constraint.

In specific, we have shown that the estimated effect size of a disease gene candidate set, as determined primarily by their mutational class, is a good guide to determining which disease properties are biological and which may be driven by artifact. We show that most ASD disease genes display a neurodevelopmental signature, with top candidates genes involved in neuronal migration and synapse formation. Epigenetic and chromatin remodelers rank lower, and may not be a specific feature of the disease, or play lesser roles. One particular point of interest is whether or not this scoring system is specific to ASD, or has largely picked out a signature of neurodevelopmental disease. While data availability is generally higher in autism, we checked for enrichment and functional effect size trends of other available GWAS and WES studies (see supplement for more details). The functional effect size

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

trends remain consistent; although the ASD studies functional effect size trends are generally greater than other cognitive disorders (see **Additional file 1: Fig S3**).

Our meta-analysis solely exploits variation across studies and therefore is partially independent of conventional analyses. Thus, when we consider PPI network clustering an artifactual property, it is not because we don't see significant clustering of gene sets in PPI data; in fact, we do. Rather, there is no functional effect trend consistent with effect size, a fact made clearest by the control sets also clustering. Another interesting example arises in one of the few properties we identified as functionally relevant that might seem artifactual: gene size. In fact, this has been hotly debated recently [40, 78], and may reflect real biology on top of technical effects [79] . In our analysis, CDS exhibits a likely bias (e.g., CNVs showing enrichment) but its significant variation in effect size is wholly independent of this (in fact, CNVs are the strongest negative outlier in the trend) and the majority of the significance arises from sets less plausibly length biased (particularly recurrent *de novo* mutations; while WES is obviously length-biased, recurrence is stringent enough to likely render this irrelevant). As a final point of interest, we note that for most of the co-expression network analyses, the control data shows significant enrichment, and therefore interpreting co-expression must be done with more aware analyses, as in this work. More broadly, our analysis can be seen as providing validation between study designs for their functional results, and doing so in ways that would be hard for them to accomplish independently. Individual studies have tended to rely on ever more stringent criteria (e.g., replication of family-wise error rates) to minimize bias, but this is obviously limited where the bias reflects the study design which is, itself, replicated. Meta-analysis is the classic approach to integrate potentially weak evidence to a stronger conclusion, but this is only possible once effect size expectations are factored into analysis. While improvements in both GWAS and rare variant data will undoubtedly improve their individual signals, cross-study analysis is a particularly stringent means of assessing functional outputs for consistency.

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

While the method we propose fills a new niche within analysis of genetic data, researchers informally look for cross-study convergence all the time. In this sense, our work is most similar to testing for recurrence of enrichment of functional properties that are associated with a disease. Like our analysis, this, too, would involve generalization to the gene-level (from variants) to look for overlaps. Such recurrence calculations would need to, at the least, incorporate some knowledge of the quite different properties of GWAS data and rare variant studies. We suggest researchers currently do this in an ad hoc way; for example, family-wise error rate control is common in GWAS data while false discovery rate control is common in rare variant analysis. This makes perfect sense, from our perspective, since it implicitly weights the degree to which it is worthwhile reporting a gene by the study design's likely effect sizes (and then corrected p-values within the study). This will tend to make recurrence of enrichment across study designs more similar to the formal test we propose. Before our work, attaining such field-wide means of consensus has been entirely ad hoc. Nonetheless, it is encouraging many of the properties we arrive at are considered of pre-existing interest for autism; making such assessments concrete offers considerable opportunity to winnow through hard-to-parse genetic signals.

As with any meta-analysis, the hope is that by incorporating multiple data, the aggregate signal may be stronger. While our work suggests an approach to do this and shows strong statistical trends, we anticipate that divisions in the field of genetics may play some role in the interpretation of this work. GWAS researchers may question the power of our GWAS analysis and assume that if it just had a large enough 'N', it alone would be the dominant player in understanding psychiatric disease. Similarly, they may suspect rare variants of reflecting 'anomalous' versions of the disorder and thus think they are less likely to be specifically linked to autism. Rare variant researchers may question the precision of the data underlying our rare variant analysis and assume that if we just had large enough 'N' to remove false positives, it would be the dominant player in understanding psychiatric disease. Similarly, they may suspect that GWAS data is affected by confounds such as population stratification. We respond to both

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

groups: we tested whether there were functional signal commons to both analyses which follows the known effect sizes. Such signals existed and made sense. Thus, both data support one another rather than being in conflict. Of course, it is possible that our approach misses real signals which are common or rare variant specific, although our robustness analysis suggests this is not a major factor. We note that our assessment of robustness of the analysis is, in some sense, more stringent even than replication cohorts or approaches such as cross-validation [80-82] since we have both positive and negative controls when holding out data. Because our test is so simple, our observation of statistically significant deviations from the null is immediately interpretable; this is critical given the extreme scale of the analysis which, in essence, permutes every functional test through every previously collected data set. This methodological simplicity implies that the approach will generalize well.

Indeed, incorporating effect size as a meta-analytic constraint offers a diverse range of novel applications. As more data from different studies is collected for autism and other complex diseases, we can see this approach as a basis for determining properties and functional signals from the disease candidates in an integrative way. That integration may be across study designs and classes of variation, as we have done, or may involve phenotype or other properties. So, for example, one could determine functional effect size trends that grow or shrink depending on how patients were classified, or even broken down in a sex-specific manner for interpreting protective effects. More broadly, as data and the means for obtaining it grows, techniques to statistically assess its structured dependencies will grow more useful and important. Our robustness analysis speaks to this in that while we are robust to modest losses of data, it is clear that more data will only improve the signals of the individual classes. More finely-tuned effect size estimates and better separations of the gene sets and variant classifications will also help refine the distinction between biological and artifactual signals, ideally allowing us conduct yet more focused study designs in a productive feedback loop.

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Conclusions

In summary, we have developed a novel means for combining studies of quite different types at the functional level. As evidence for autism and other disorders continues to develop and continues to be heterogeneous with respect to ascertainment biases and study designs, we suspect approaches based on the one we describe will be of increasing importance. Most importantly, because we largely exploit variation across studies to determine functional signals, the overlap between our results and those identified in individual studies (e.g., co-expression, RVIS) is partially independent; the functional effect size trends we observe provide evidence as to how far the field has collectively come in the interpretation of autism genetics and offers a means which, we hope, will aid the field in continuing this progress.

Declarations

Abbreviations

ASD: autism spectrum disorder
FMRP: fragile X mental retardation protein
PPIN: protein-protein interaction network
GWAS: genome-wide association studies
WES: whole exome sequencing
SSC: Simons Simplex Collection
LoF: Loss-of-function mutations

Availability of data and materials

The datasets supporting the conclusions of this article are included within the article and in the additional files.

Additional files

Additional file 1: Fig S1 Overlap of genes across gene sets.

Additional file 1: Fig S2 Trend line robustness analysis.

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Additional file 1: Fig S3 Effect sizes of other disease gene groups.

Additional file 1: Table S1 GWA studies used in the common variant gene sets

Additional file 2: Table S2 Gene functional properties and tests

Additional file 1: Table S3 Connectivity and assortativity of networks and their results in the trend tests

Additional file 2: Table S4 ASD candidates and scores

Additional file 1: Table S5 Whole-exome studies used to test aggregate score

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by a grant from T. and V. Stanley.

Authors' contributions

SB wrote the manuscript, and conducted the experiments. JG wrote the manuscript and designed the experiments. All authors read and approved the final manuscript.

Acknowledgments

We would like to thank members of the CSHL Wigler lab for access to their data. We thank Paul Pavlidis for helpful comments on a draft of the manuscript.

References

1. Miles JH: **Autism spectrum disorders--a genetics review**. *Genet Med* 2011, **13**(4):278-294.
2. Devlin B, Scherer SW: **Genetic architecture in autism spectrum disorder**. *Curr Opin Genet Dev* 2012, **22**(3):229-237.
3. Gibson G: **Rare and common variants: twenty arguments**. *Nat Rev Genet* 2012, **13**(2):135-145.

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J *et al*: **Strong Association of De Novo Copy Number Mutations with Autism**. *Science* 2007, **316**(5823):445-449.
5. Christian SL, Brune CW, Sudi J, Kumar RA, Liu S, KaraMohamed S, Badner JA, Matsui S, Conroy J, McQuaid D *et al*: **Novel Submicroscopic chromosomal abnormalities detected in Autism Spectrum Disorder**. *Biol Psychiatry* 2008, **63**(12):1111-1117.
6. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, Shago M, Moessner R, Pinto D, Ren Y *et al*: **Structural Variation of Chromosomes in Autism Spectrum Disorder**. *Am J Hum Genet* 2008, **82**(2):477-488.
7. Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S, Zhang H, Estes A, Brune CW, Bradfield JP *et al*: **Autism genome-wide copy number variation reveals ubiquitin and neuronal genes**. *Nature* 2009, **459**(7246):569-573.
8. Ma DQ, Salyakina D, Jaworski JM, Konidari I, Whitehead PL, Andersen AN, Hoffman JD, Slifer SH, Hedges DJ, Cukier HN *et al*: **A genome-wide association study of autism reveals a common novel risk locus at 5p14.1**. *Annals of human genetics* 2009, **73**(Pt 3):263-273.
9. Xia K, Guo H, Hu Z, Xun G, Zuo L, Peng Y, Wang K, He Y, Xiong Z, Sun L *et al*: **Common genetic variants on 1p13.2 associate with risk of autism**. *Mol Psychiatry* 2014, **19**(11):1212-1219.
10. Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS, Salyakina D, Imielinski M, Bradfield JP, Sleiman PMA *et al*: **Common genetic variants on 5p14.1 associate with autism spectrum disorders**. *Nature* 2009, **459**(7246):528-533.
11. Weiss LA, Arking DE, The Gene Discovery Project of Johns Hopkins the Autism C: **A genome-wide linkage and association scan reveals novel loci for autism**. *Nature* 2009, **461**(7265):802-808.
12. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, Karakoc E, MacKenzie AP, Ng SB, Baker C: **Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations**. *Nat Genet* 2011, **43**(6):585-589.
13. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, Ercan-Sencicek AG, DiLullo NM, Parikshak NN, Stein JL: **De novo mutations revealed by whole-exome sequencing are strongly associated with autism**. *Nature* 2012, **485**(7397):237-241.
14. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, Yamrom B, Lee Y-h, Narzisi G, Leotta A *et al*: **De Novo Gene Disruptions in Children on the Autistic Spectrum**. *Neuron* 2012, **74**(2):285-299.
15. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, Lin CF, Stevens C, Wang LS, Makarov V *et al*: **Patterns and rates of exonic de novo mutations in autism spectrum disorders**. *Nature* 2012, **485**(7397):242-245.
16. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN: **Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease**. *Nat Genet* 2003, **33**(2):177-182.
17. Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS, Salyakina D, Imielinski M, Bradfield JP, Sleiman PM *et al*: **Common genetic variants on 5p14.1 associate with autism spectrum disorders**. *Nature* 2009, **459**(7246):528-533.
18. Yuen RKC, Thiruvahindrapuram B, Merico D, Walker S, Tammimies K, Hoang N, Chrysler C, Nalpathamkalam T, Pellecchia G, Liu Y *et al*: **Whole-genome sequencing of quartet families with autism spectrum disorder**. *Nat Med* 2015, **21**(2):185-191.
19. Jiang Y-h, Yuen Ryan KC, Jin X, Wang M, Chen N, Wu X, Ju J, Mei J, Shi Y, He M *et al*: **Detection of Clinically Relevant Genetic Variants in Autism Spectrum Disorder by Whole-Genome Sequencing**. *Am J Hum Genet* 2013, **93**(2):249-263.
20. Henn BM, Botigue LR, Bustamante CD, Clark AG, Gravel S: **Estimating the mutation load in human genomes**. *Nat Rev Genet* 2015, **16**(6):333-343.

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

21. Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H: **Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease**. *Hum Genet* 2013, **132**(10):1077-1130.
22. Bauer-Mehren A, Bundschus M, Rautschka M, Mayer MA, Sanz F, Furlong LI: **Gene-Disease Network Analysis Reveals Functional Modules in Mendelian, Complex and Environmental Diseases**. *PLoS ONE* 2011, **6**(6):e20284.
23. Ben-David E, Shifman S: **Networks of Neuronal Genes Affected by Common and Rare Variants in Autism Spectrum Disorders**. *PLoS Genet* 2012, **8**(3):e1002556.
24. Sakai Y, Shaw CA, Dawson BC, Dugas DV, Al-Mohtaseb Z, Hill DE, Zoghbi HY: **Protein Interactome Reveals Converging Molecular Pathways Among Autism Disorders**. *Science Translational Medicine* 2011, **3**(86):86ra49.
25. Porteous DJ, Thomson PA, Millar JK, Evans KL, Hennah W, Soares DC, McCarthy S, McCombie WR, Clapcote SJ, Korth C *et al*: **DISC1 as a genetic risk factor for schizophrenia and related major mental illness: response to Sullivan**. *Mol Psychiatry* 2014, **19**(2):141-143.
26. Sullivan PF: **Questions about DISC1 as a genetic risk factor for schizophrenia**. *Mol Psychiatry* 2013, **18**(10):1050-1052.
27. Sullivan PF, Daly MJ, O'Donovan M: **Genetic architectures of psychiatric disorders: the emerging picture and its implications**. *Nat Rev Genet* 2012, **13**(8):537-551.
28. Schizophrenia Working Group of the Psychiatric Genomics C: **Biological insights from 108 schizophrenia-associated genetic loci**. *Nature* 2014, **511**(7510):421-427.
29. Light RJ, Pillemer DB: **Summing up: The science of reviewing research**: Harvard University Press; 1984.
30. Begg CB, Mazumdar M: **Operating characteristics of a rank correlation test for publication bias**. *Biometrics* 1994, **50**(4):1088-1101.
31. Egger M, Davey Smith G, Schneider M, Minder C: **Bias in meta-analysis detected by a simple, graphical test**. *BMJ : British Medical Journal* 1997, **315**(7109):629-634.
32. Lau J, Ioannidis JPA, Terrin N, Schmid CH, Olkin I: **The case of the misleading funnel plot**. *BMJ : British Medical Journal* 2006, **333**(7568):597-600.
33. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, Stessman HA, Witherspoon K, Vives L, Patterson KE *et al*: **The contribution of de novo coding mutations to autism spectrum disorder**. *Nature* 2014, **515**(7526):216-221.
34. Levy D, Ronemus M, Yamrom B, Lee YH, Leotta A, Kendall J, Marks S, Lakshmi B, Pai D, Ye K *et al*: **Rare de novo and transmitted copy-number variation in autistic spectrum disorders**. *Neuron* 2011, **70**(5):886-897.
35. Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, Vitkup D: **Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses**. *Neuron* 2011, **70**(5):898-907.
36. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A: **OMIM.org: Online Mendelian Inheritance in Man (OMIM(®)), an online catalog of human genes and genetic disorders**. *Nucleic Acids Res* 2015, **43**(Database issue):D789-D798.
37. Portales-Casamar E, Ch'ng C, Lui F, St-Georges N, Zoubarev A, Lai A, Lee M, Kwok C, Kwok W, Tseng L *et al*: **Neurocarta: aggregating and sharing disease-gene relations for the neurosciences**. *BMC Genomics* 2013, **14**(1):129.
38. Basu SN, Kollu R, Banerjee-Basu S: **AutDB: a gene reference resource for autism research**. *Nucleic Acids Res* 2009, **37**(Database issue):D832-836.
39. Yu W, Gwinn M, Clyne M, Yesupriya A, Khoury MJ: **A navigator for human genome epidemiology**. *Nat Genet* 2008, **40**(2):124-125.
40. Shohat S, Shifman S: **Bias towards large genes in autism**. *Nature* 2014, **512**(7512):E1-E2.

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

41. Betancur C: **Etiological heterogeneity in autism spectrum disorders: More than 100 genetic and genomic disorders and still counting**. *Brain Res* 2011, **1380**(0):42-77.
42. Willsey AJ, Sanders Stephan J, Li M, Dong S, Tebbenkamp Andrew T, Muhle Rebecca A, Reilly Steven K, Lin L, Fertuzinhos S, Miller Jeremy A *et al*: **Coexpression Networks Implicate Human Midfetal Deep Cortical Projection Neurons in the Pathogenesis of Autism**. *Cell* 2013, **155**(5):997-1007.
43. Sugathan A, Biagioli M, Golzio C, Erdin S, Blumenthal I, Manavalan P, Ragavendran A, Brand H, Lucente D, Miles J *et al*: **CHD8 regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors**. *Proceedings of the National Academy of Sciences* 2014, **111**(42):E4468-E4477.
44. Hormozdiari F, Penn O, Borenstein E, Eichler EE: **The discovery of integrated gene networks for autism and related disorders**. *Genome Res* 2015, **25**(1):142-154.
45. Ballouz S, Verleyen W, Gillis J: **Guidance for RNA-seq co-expression network construction and analysis: safety in numbers**. *Bioinformatics* 2015, **31**(13):2123-2130.
46. Ronan JL, Wu W, Crabtree GR: **From neural development to cognition: unexpected roles for chromatin**. *Nature reviews Genetics* 2013, **14**(5):347-359.
47. Bayés À, Collins MO, Croning MDR, van de Lagemaat LN, Choudhary JS, Grant SGN: **Comparative Study of Human and Mouse Postsynaptic Proteomes Finds High Compositional Conservation and Abundance Differences for Key Synaptic Proteins**. *PLoS ONE* 2012, **7**(10):e46683.
48. Collins MO, Husi H, Yu L, Brandon JM, Anderson CNG, Blackstock WP, Choudhary JS, Grant SGN: **Molecular characterization and comparison of the components and multiprotein complexes in the postsynaptic proteome**. *J Neurochem* 2006, **97 Suppl 1**:16-23.
49. Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB: **Genic Intolerance to Functional Variation and the Interpretation of Personal Genomes**. *PLoS Genet* 2013, **9**(8):e1003709.
50. Georgi B, Voight BF, Bućan M: **From Mouse to Human: Evolutionary Genomics Analysis of Human Orthologs of Essential Genes**. *PLoS Genet* 2013, **9**(5):e1003484.
51. Huang N, Lee I, Marcotte EM, Hurles ME: **Characterising and Predicting Haploinsufficiency in the Human Genome**. *PLoS Genet* 2010, **6**(10):e1001154.
52. Verleyen W, Ballouz S, Gillis J: **Measuring the wisdom of the crowds in network-based gene function inference**. *Bioinformatics* 2015, **31**(5):745-752.
53. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes**. *Nucleic Acids Res* 2000, **28**(1):27-30.
54. Darnell JC, Van Driesche SJ, Zhang C, Hung KYS, Mele A, Fraser CE, Stone EF, Chen C, Fak JJ, Chi SW *et al*: **FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism**. *Cell* 2011, **146**(2):247-261.
55. Lips ES, Cornelisse LN, Toonen RF, Min JL, Hultman CM, Holmans PA, O'Donovan MC, Purcell SM, Smit AB, Verhage M *et al*: **Functional gene group analysis identifies synaptic gene groups as risk factor for schizophrenia**. *Mol Psychiatry* 2012, **17**(10):996-1006.
56. Levy D, Ronemus M, Yamrom B, Lee Y-h, Leotta A, Kendall J, Marks S, Lakshmi B, Pai D, Ye K *et al*: **Rare De Novo and Transmitted Copy-Number Variation in Autistic Spectrum Disorders**. *Neuron* 2011, **70**(5):886-897.
57. Cho S-C, Yoo HJ, Park M, Cho IH, Kim B-N, Kim J-W, Shin M-S, Park T-W, Son J-W, Chung U-S *et al*: **Genome-Wide Association Scan of Korean Autism Spectrum Disorders with Language Delay: A Preliminary Study**. *Psychiatry Investigation* 2011, **8**(1):61-66.
58. Colantuoni C, Lipska BK, Ye T, Hyde TM, Tao R, Leek JT, Colantuoni EA, Elkahloun AG, Herman MM, Weinberger DR *et al*: **Temporal dynamics and genetic control of transcription in the human prefrontal cortex**. *Nature* 2011, **478**(7370):519-523.

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

59. Chang J, Gilman SR, Chiang AH, Sanders SJ, Vitkup D: **Genotype to phenotype relationships in autism spectrum disorders**. *Nat Neurosci* 2015, **18**(2):191-198.
60. Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A, Tyers M: **BioGRID: a general repository for interaction datasets**. *Nucleic Acids Research* 2006, **34**(suppl 1):D535-D539.
61. Chua HN, Sung W-K, Wong L: **Exploiting indirect neighbours and topological weight to predict protein function from protein–protein interactions**. *Bioinformatics* 2006, **22**(13):1623-1630.
62. Gillis J, Pavlidis P: **The role of indirect connections in gene networks in predicting function**. *Bioinformatics* 2011, **27**(13):1860-1866.
63. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S *et al*: **GENCODE: The reference human genome annotation for The ENCODE Project**. *Genome Res* 2012, **22**(9):1760-1774.
64. Bush WS, Moore JH: **Chapter 11: Genome-Wide Association Studies**. *PLoS Comput Biol* 2012, **8**(12):e1002822.
65. Westreich D: **Berkson's bias, selection bias, and missing data**. *Epidemiology (Cambridge, Mass)* 2012, **23**(1):159-164.
66. Froyen G, Corbett M, Vandewalle J, Jarvela I, Lawrence O, Meldrum C, Bauters M, Govaerts K, Vandeleur L, Van Esch H *et al*: **Submicroscopic Duplications of the Hydroxysteroid Dehydrogenase HSD17B10 and the E3 Ubiquitin Ligase HUWE1 Are Associated with Mental Retardation**. *Am J Hum Genet* 2008, **82**(2):432-443.
67. McCarthy SE, Gillis J, Kramer M, Lihm J, Yoon S, Berstein Y, Mistry M, Pavlidis P, Solomon R, Ghiban E *et al*: **De novo Mutations in Schizophrenia Implicate Chromatin Remodeling and Support a Genetic Overlap with Autism and Intellectual Disability**. *Mol Psychiatry* 2014, **19**(6):652-658.
68. Nava C, Lamari F, Heron D, Mignot C, Rastetter A, Keren B, Cohen D, Faudet A, Bouteiller D, Gilleron M *et al*: **Analysis of the chromosome X exome in patients with autism spectrum disorders identified novel candidate genes, including TMLHE**. *Transl Psychiatry* 2012, **2**:e179.
69. Buxbaum JD, Cai G, Nygren G, Chaste P, Delorme R, Goldsmith J, Råstam M, Silverman JM, Hollander E, Gillberg C *et al*: **Mutation analysis of the NSD1 gene in patients with autism spectrum disorders and macrocephaly**. *BMC Med Genet* 2007, **8**:68-68.
70. Vaags AK, Lionel AC, Sato D, Goodenberger M, Stein QP, Curran S, Ogilvie C, Ahn JW, Drmic I, Senman L *et al*: **Rare deletions at the neurexin 3 locus in autism spectrum disorder**. *Am J Hum Genet* 2012, **90**(1):133-141.
71. Gauthier J, Siddiqui T, Huashan P, Yokomaku D, Hamdan F, Champagne N, Lapointe M, Spiegelman D, Noreau A, Lafrenière R *et al*: **Truncating mutations in NRXN2 and NRXN1 in autism spectrum disorders and schizophrenia**. *Hum Genet* 2011, **130**(4):563-573.
72. Chang K-J, Zollinger DR, Susuki K, Sherman DL, Makara MA, Brophy PJ, Cooper EC, Bennett V, Mohler PJ, Rasband MN: **Glial Ankyrins Facilitate Paranodal Axoglial Junction Assembly**. *Nat Neurosci* 2014, **17**(12):1673-1681.
73. O'Donnell L, Rhodes D, Smith SJ, Merriner DJ, Clark BJ, Borg C, Whittle B, O'Connor AE, Smith LB, McNally FJ *et al*: **An Essential Role for Katanin p80 and Microtubule Severing in Male Gamete Production**. *PLoS Genet* 2012, **8**(5):e1002698.
74. Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, Helgason A, Stefansson H, Emilsson V, Helgadottir A: **Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes**. *Nat Genet* 2006, **38**(3):320-323.
75. Tang W, Dodge M, Gundapaneni D, Michnoff C, Roth M, Lum L: **A genome-wide RNAi screen for Wnt/beta-catenin pathway components identifies unexpected roles for TCF transcription factors in cancer**. *Proc Natl Acad Sci U S A* 2008, **105**(28):9697-9702.

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

76. Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K, Raja A, Coe BP, Stessman HA, He Z-X: **Excess of rare, inherited truncating mutations in autism**. *Nat Genet* 2015, **47**(6):582-588.
77. Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L, Thiruvahindrapuram B, Xu X, Ziman R, Wang Z *et al*: **Convergence of Genes and Cellular Pathways Dysregulated in Autism Spectrum Disorders**. *The American Journal of Human Genetics* 2014, **94**(5):677-694.
78. Zylka MJ, Philpot BD, King IF: **Zylka et al. reply**. *Nature* 2014, **512**(7512):E2-E2.
79. Raychaudhuri S, Korn JM, McCarroll SA, Altshuler D, Sklar P, Purcell S, Daly MJ, The International Schizophrenia C: **Accurately Assessing the Risk of Schizophrenia Conferred by Rare Copy-Number Variation Affecting Genes with Brain Function**. *PLoS Genet* 2010, **6**(9):e1001097.
80. Molinaro AM, Simon R, Pfeiffer RM: **Prediction error estimation: a comparison of resampling methods**. *Bioinformatics* 2005, **21**(15):3301-3307.
81. Bernau C, Riester M, Boulesteix A-L, Parmigiani G, Huttenhower C, Waldron L, Trippa L: **Cross-study validation for the assessment of prediction algorithms**. *Bioinformatics* 2014, **30**(12):i105-i112.
82. Verleyen W, Ballouz S, Gillis J: **Positive and negative forms of replicability in gene network analysis**. *Bioinformatics* 2015.

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figures

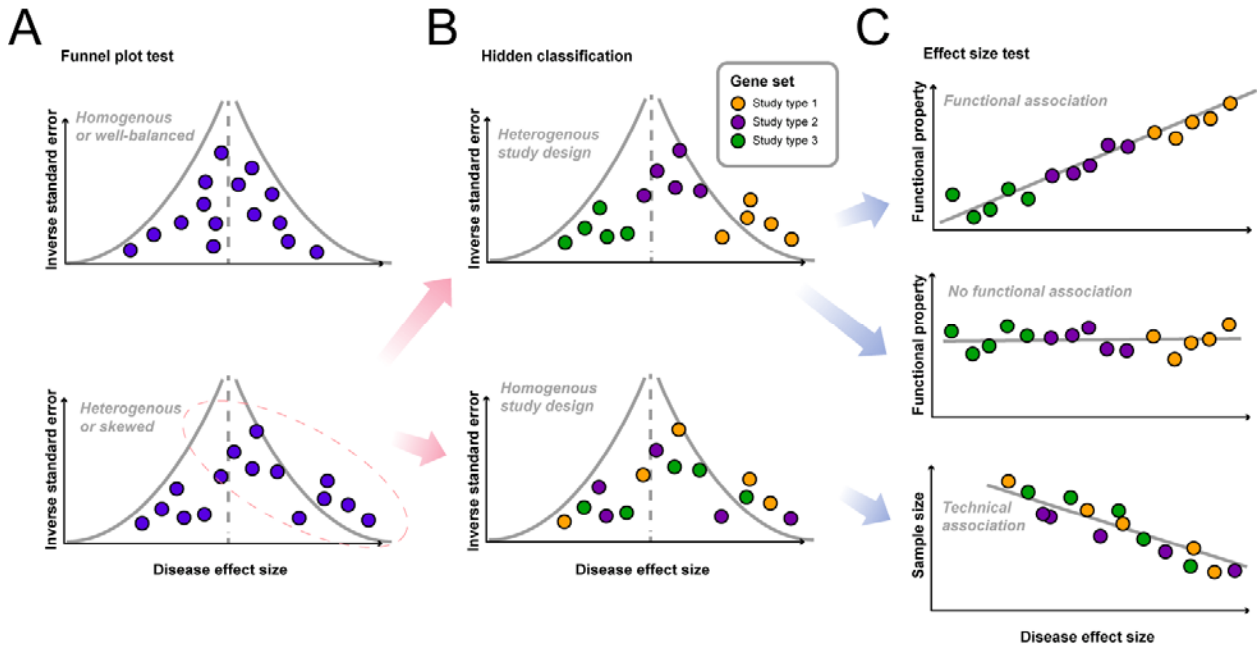

Fig 1 Hidden classes in meta-analysis assessments using funnel plots.
(A) Funnel plot tests are used in determining homogeneity in the studies of a meta-analysis. Typically, effect of the study (x-axis) is plotted against a form of error or variance (e.g., inverse standard error, y-axis). The studies should fit within this "funnel" for a homogenous set of studies (top panel). However, in cases where this is an obvious heterogeneity (skew in bottom panel, dashed ellipse), there are likely hidden biases. (B) Hidden classes within the data e.g., study types, could be the cause for such skew. In this case (top panel), the heterogeneous study forces skew. In the bottom panel, this is not the case, and there could be some other hidden class. (C) Given the lack of heterogeneity, we can exploit this "failure" (i.e., class association) through an effect size test for functional associations (top panel), or lack thereof (middle panel). For homogenous data sets, we can observe any correlations to potential confounds, which in this case is shown for sample size (bottom panel).

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

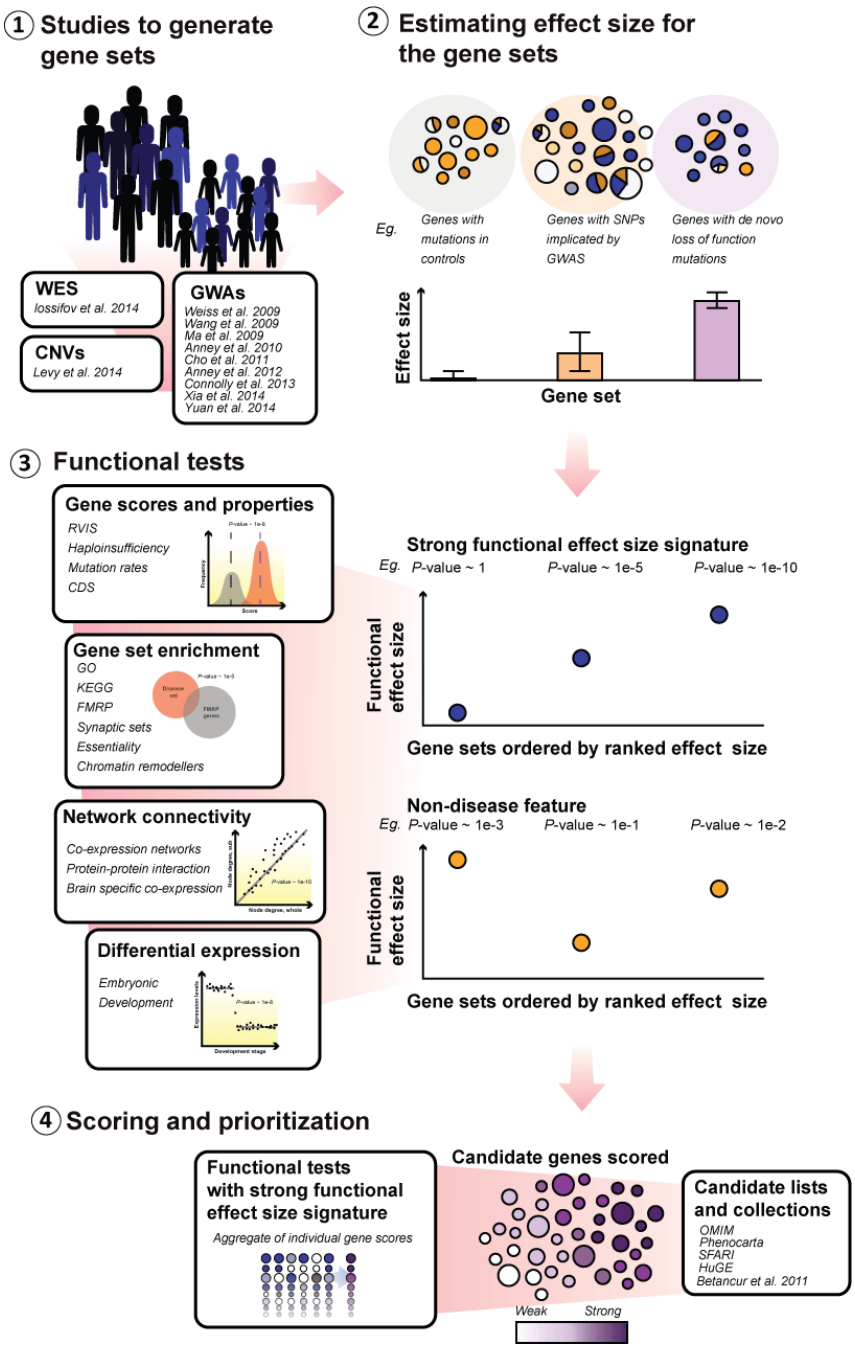

Fig 2 Schematic of functional effect size trend calculation.
(1) Starting with disease gene set collections, (2) each is then rank each by the average effect size of the genes within that set. (3) We then run functional tests on these genes sets and calculate a functional effect size for each. Then, using the ranking of the disease gene sets, we measure the functional effect size trend – the slope of the trend line of the functional effect sizes versus the rank. With these positive slopes (strong functional signatures), we determine which functions are likely disease-specific. (4) This allows us to generate a score for each gene, which we can use to identify and validate candidate genes.

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

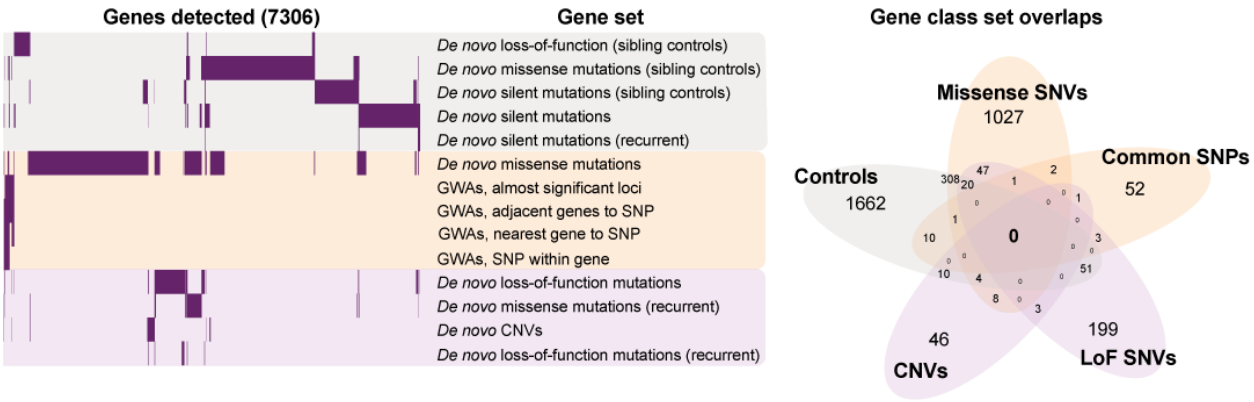

Fig 3 Overlap of genes across the different gene sets.
There is little overlap in the disease gene sets as shown in the heatmap and as enumerated in the Venn diagram beside it. There are some significant overlaps. Control gene sets overlap significantly with missense genes (333 genes hypergeometric test p=2.54e-6), common SNPs gene sets (11 genes, p=4.5e-3), and the loss-of-function (LoF) SNV gene sets (71 genes, p=3.2e-3) but not the CNV gene sets (14 genes, p=0.03). Missense and common SNPs overlap significantly (4 genes, p=2.4e-4). However, loss-of-function SNVs do not overlap significantly with either common (4 genes, p=0.62), missense (68 genes, p=0.75), or CNVs (3 genes, p=0.37).

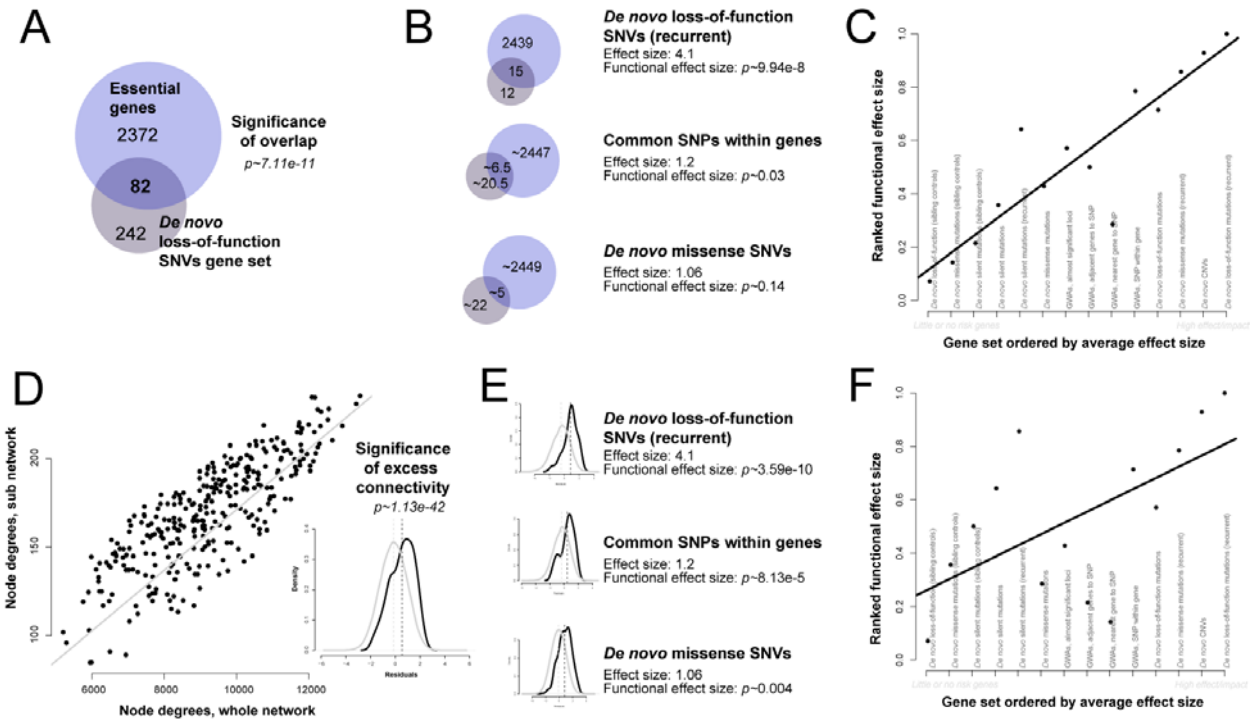

Fig 4 Functional properties of disease gene sets are tested using gene set enrichment (top panels) and co-expression network connectivity (bottom panels).
(A) Gene set enrichment is calculated with a hypergeometric test. A large number (34%) of the genes in the *de novo* loss-of-function set overlap with essential genes (hypergeometric test p~7.11e-11). (B) This is repeated across all disease gene sets, (a subset shown here). Sample size is controlled through downsampling. Gene sets with the

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

higher effect sizes also have the higher functional effect sizes. (C) We can now demonstrate how to calculate the functional effect size trend for the "essential genes" test. The disease gene sets are ordered by an estimate of the average effect size of genes within the set (from low to high on the x-axis) and the functional effect size is of that disease gene set is plotted (y-axis). A trend between the effect size of the candidate genes and their essentiality can be clearly observed (FDR<0.01). The network connectivity functional test (D) consists of calculating the ratio of disease genes' total connectivity (node degrees calculated from the whole network; sum of their connections) to their internal connectivity (node degrees of their subnetwork; sum of their connections to one another). The line (in grey) reflects the expected values if there is no preferential connectivity within the set. We see that a large number (72%) of the genes lie above the identity line. The Wilcoxon P-value of the mean residuals is shown in the inset (p~1.13e-42) (E) Once again, controlling for sample size through downsampling, the functional effect size of each gene set is calculated (subset shown). (F) A weak trend (FDR>0.01) between the effect size of the candidate genes and their degree of co-expression is visible. Empirical nulls are calculated by permuting disease gene sets and FDRs through a Benjamini-Hochberg correction against the resultant 158 functional effect size trends.

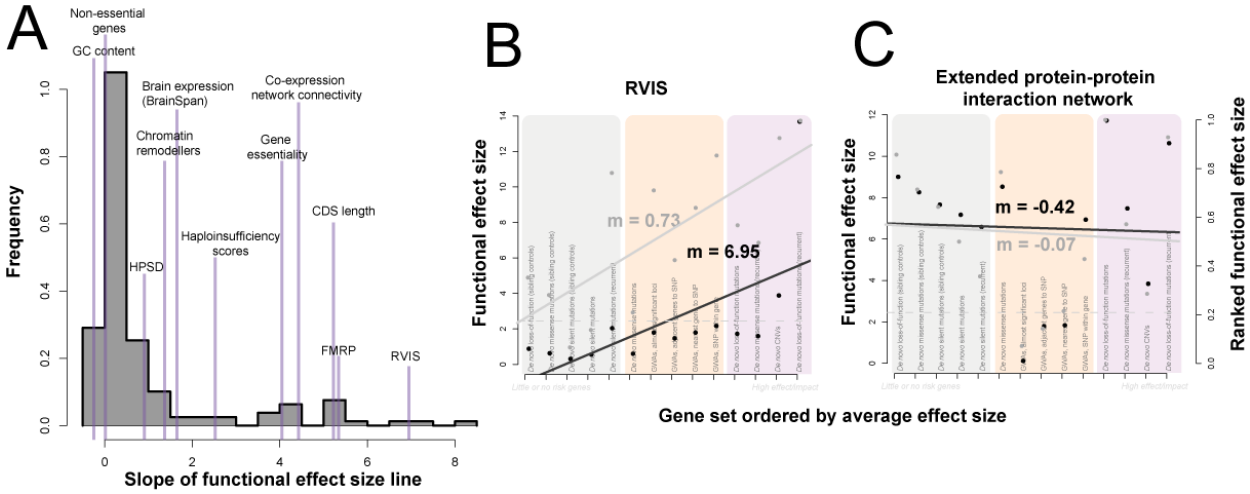

Fig 5 Functional convergence of the disease gene sets.
(A) Distribution of the effect size slopes of functional properties commonly tested for by disease candidate methods. The (B) RVIS enrichment test has a very high slope across the same gene sets (black lines)(C) extended PPIN has mostly artifactual signal as demonstrated by the flat slope, highly elevated from the null but in no effect-correlated way and suggestive of consistent biases. The grey lines are the ranked functional effect size trend slopes (to show the trends are robust to non-parametric assessment, see Fig 6A).

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

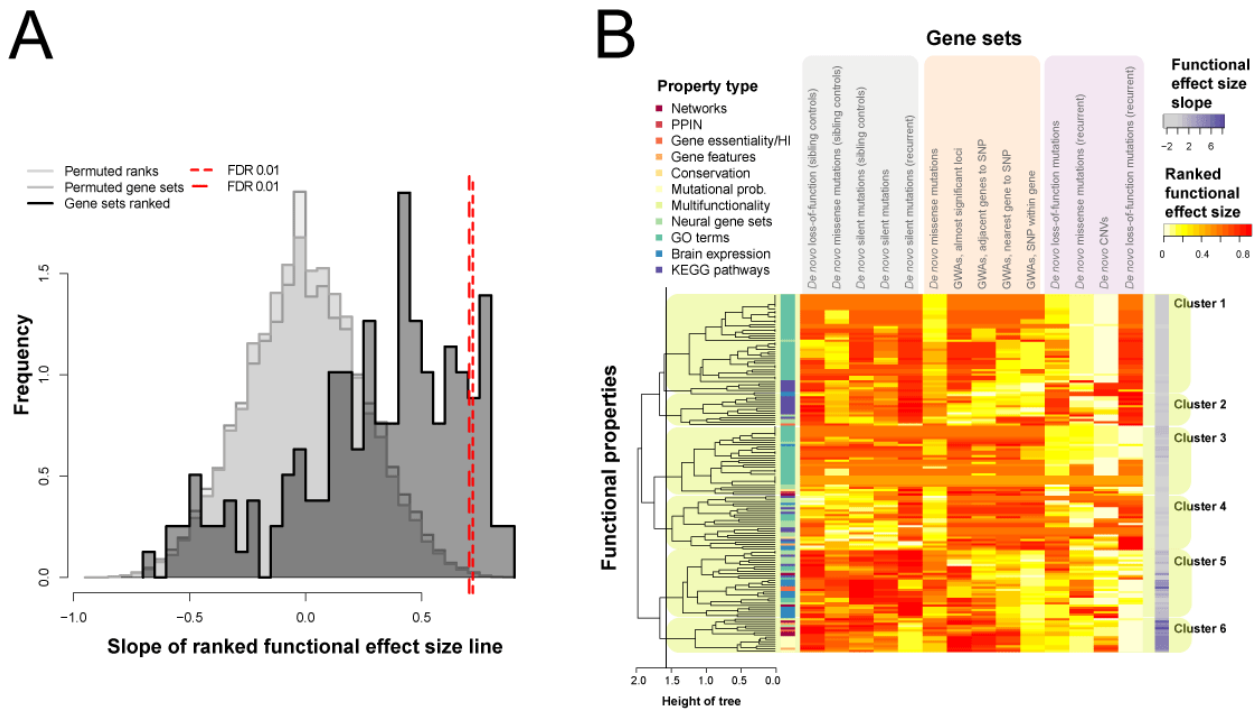

Fig 6 Clustering disease property tests by their functional effect size trends.
(A) The slopes of the ranked effect size trends are significantly different from null distributions for both effect size permutations and random gene set sampling (Wilcoxon test p~1.92e-31 and p~1.56e-32 respectively).We've drawn the two red lines to indicate the false discovery rates (FDRs) of 0.01, where 20 and 21 functions are significant. (B) A heatmap of all the ranked scores of the test gene sets (columns) for 158 properties tested (rows). The properties clustered into 6 groups when we cut the dendrogram at a height of ~1.6. Their effect size slopes (unranked) show that most highly sloped effect size lines cluster (in clusters 5 and 6). White/yellow is a high rank, red is low. A high slope is shaded purple, low/negative slopes are grey. The property type is color coded as described in the figure key.

bioRxiv preprint doi: https://doi.org/10.1101/043422; this version posted March 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1

2 # Tables and captions

3 **Table 1 Disease gene sets used in the study**

Gene set	Set size (genes)	Odds Ratios/ effect sizes	"Rank"
WES results.			
Varying effect sizes depending on mutation and recurrence.			
Odds ratio is calculated as in Sanders et al,[13] *(ratio of observed counts of mutations to silent mutations, and then ratio of those odds between siblings to probands).*			
***De novo* loss-of-function, recurrent**	27	4.1	1
***De novo* CNVs**	72	3.95	2
***De novo* missense, recurrent**	153	1.6	3
***De novo* loss-of-function**	341	1.5	4
***De novo* missense**	1339	1.06	9
GWAS. Multiple hit models, low effect size per gene. Odds ratios range between 1.1 and 1.2.			
GWAS, SNP within genes	50	1.2	5
GWAS, nearest gene to SNP	71	1.2	6
GWAS, adjacent genes to SNP	89	1.2	7
GWAS, almost significant loci	71	1.1	8
Control sets, no effect			
***De novo* silent, recurrent**	25	NA	10
***De novo* silent**	590	NA	11
Control groups, no disease			
***De novo* loss-of-function (sibling controls)**	174	NA	12
***De novo* missense (sibling controls)**	1066	NA	13
***De novo* silent (sibling controls)**	468	NA	14
Databases and ASD collections			
SFARI	376		
SFARI-high confidence set	104		
Phenocarta	679		
OMIM	79		
HUGE – meta	451		
Betancur et al.	103		

4

5