Cell Systems

A convergent molecular network underlying autism and congenital heart disease

Highlights

- A convergent molecular network is shared between autism and congenital heart disease
- 101 genes likely carry a risk for both disorders and validate in an independent cohort
- Ion channels, a key convergent pathway, are linked to early brain and heart development
- When disrupted, shared risk genes cause brain and heart abnormalities in *Xenopus*

Authors

Sara Brin Rosenthal, Helen Rankin Willsey, Yuxiao Xu, ..., Arthur Jeremy Willsey, Kathleen M. Fisch, Trey Ideker

Correspondence

jeremy.willsey@ucsf.edu (A.J.W.), kfisch@ucsd.edu (K.M.F.), tideker@ucsd.edu (T.I.)

In brief

Rosenthal et al., use network genetics to identify a convergent molecular network underlying autism and congenital heart disease, two co-morbid disorders. This network is impacted by damaging genetic variants from both disorders in multiple independent cohorts of patients, pinpointing 101 genes with shared genetic risk. The network implicates a large family of ion channels, including the sodium transporter *SCN2A*, in which the authors show disruptions cause defects in Xenopus heart and brain development.



Cell Systems



Article

A convergent molecular network underlying autism and congenital heart disease

Sara Brin Rosenthal,^{1,7} Helen Rankin Willsey,^{2,7} Yuxiao Xu,² Yuan Mei,³ Jeanselle Dea,² Sheng Wang,⁶ Charlotte Curtis,³ Emily Sempou,⁴ Mustafa K. Khokha,⁴ Neil C. Chi,⁵ Arthur Jeremy Willsey,^{2,6,*} Kathleen M. Fisch,^{1,*} and Trey Ideker^{3,8,*} ¹Center for Computational Biology & Bioinformatics, Department of Medicine, University of California, San Diego, La Jolla, CA 92093, USA ²Department of Psychiatry and Behavioral Sciences, Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA 94158, USA

³Division of Genetics, Department of Medicine, University of California, San Diego, La Jolla, CA 92093, USA

⁴Pediatric Genomics Discovery Program, Department of Pediatrics and Genetics, Yale University School of Medicine, New Haven, CT 06510, USA

⁵Division of Cardiology, Department of Medicine, University of California, San Diego, La Jolla, CA 92093, USA

⁶Quantitative Biosciences Institute (QBI), University of California, San Francisco, San Francisco, CA 94158, USA

⁷These authors contributed equally

⁸Lead contact

*Correspondence: jeremy.willsey@ucsf.edu (A.J.W.), kfisch@ucsd.edu (K.M.F.), tideker@ucsd.edu (T.I.) https://doi.org/10.1016/j.cels.2021.07.009

SUMMARY

Patients with neurodevelopmental disorders, including autism, have an elevated incidence of congenital heart disease, but the extent to which these conditions share molecular mechanisms remains unknown. Here, we use network genetics to identify a convergent molecular network underlying autism and congenital heart disease. This network is impacted by damaging genetic variants from both disorders in multiple independent cohorts of patients, pinpointing 101 genes with shared genetic risk. Network analysis also implicates risk genes for each disorder separately, including 27 previously unidentified genes for autism and 46 for congenital heart disease. For 7 genes with shared risk, we create engineered disruptions in *Xenopus tropicalis*, confirming both heart and brain developmental abnormalities. The network includes a family of ion channels, such as the sodium transporter *SCN2A*, linking these functions to early heart and brain development. This study provides a road map for identifying risk genes and pathways involved in co-morbid conditions.

INTRODUCTION

Autism spectrum disorder (ASD), a heterogeneous neurodevelopmental condition affecting behavior and communication in approximately 1 of every 59 children (CDC, 2018), commonly arises in combination with other disorders including congenital heart disease (CHD) (Zaidi and Brueckner, 2017). Likewise, children with CHD are at increased risk for neurodevelopmental disorders (NDDs), with a likelihood that strongly correlates with CHD severity and the presence of de novo mutations, especially those impacting chromatin modifiers (Zaidi and Brueckner, 2017; Jin et al., 2017). Genetics and adverse events during postnatal heart surgery have been proposed as explanations for the observed ASD-CHD co-morbidity (Homsy et al., 2015; Willsey et al., 2018; Gaynor et al., 2015; Morton et al., 2017; Jin et al., 2017), but the precise molecular mechanisms remain unclear (Zaidi and Brueckner, 2017; Homsy et al., 2015). As surgical techniques for newborns with CHD improve, with concomitant increases in survival, research is shifting to improving quality of life (Andonian et al., 2018; Zaidi and Brueckner, 2017) including timely monitoring and potential intervention for ASD and other

NDDs (Homsy et al., 2015; Jin et al., 2017; Zaidi and Brueckner, 2017). Critical to this goal is to better understand the shared genetic and molecular components of the cardiac and neurological phenotypes (Morton et al., 2017).

Both rare and common variants contribute to ASD and CHD genetic risk (Glessner et al., 2014; Fakhro et al., 2011; Sanders et al., 2012, 2015; Zaidi and Brueckner, 2017; Jin et al., 2017; Grove et al., 2019; Sestan and State, 2018; Satterstrom et al., 2020; Zaidi et al., 2013; O'Roak et al., 2012; Sebat et al., 2007; De Rubeis et al., 2014). While studies have identified some of the same genes harboring ASD and CHD rare variants (Homsy et al., 2015; Gelb and Chung, 2014; de la Torre-Ubieta et al., 2016; lossifov et al., 2014; Jin et al., 2017; Willsey et al., 2018), the majority of risk genes for these two disorders have not overlapped. For example, only 5 genes overlap between the 65 ASD risk genes and the 66 CHD risk genes identified in previous studies (Sanders et al., 2015; Jin et al., 2017). Therefore, it has been proposed that much larger patient cohorts, as well as genome-wide analyses of common variants, will be required to truly understand the risk genes, effect sizes, and, most critically, core biological mechanisms underlying the



Cell Systems Article

two disorders (O'Roak et al., 2012; Zaidi and Brueckner, 2017). Even with very large sample sizes, variants with particular effect sizes, population frequencies, or recessive inheritance have been difficult to uncover by current modes of genetic analysis (State, 2010), which prioritize heterozygous or non-inherited variants.

In any quest to identify genes and mechanisms underlying complex conditions, a necessary consideration is that the various genes and loci are hardly independent but act synergistically or antagonistically within gene networks (O'Roak et al., 2012; Willsey et al., 2013; Parikshak et al., 2013; Menche et al., 2015; Liu et al., 2014; Lage et al., 2012). While genetic variants at any single locus may be relatively rare, they can be more readily recognized by their convergence on commonly altered multi-gene systems, including protein complexes, signaling pathways, and higher-order developmental and metabolic processes. Integration of genomic variants with molecular networks, thus, provides a complementary genetic analysis strategy, with the ability to identify constellations of risk genes and variants across a large range of effect sizes, frequencies, and inheritance patterns (Barabási et al., 2011; Navlakha and Kingsford, 2010; Califano et al., 2012; Flint and Ideker, 2019; Civelek and Lusis, 2014).

Here, we integrate human molecular network knowledge with large-scale patient sequencing data to identify a common ASD-CHD network underlying both disorders. Further analysis of this network reveals convergent pathways in brain and heart development, including chromatin modification and MAPK/Notch signaling, as well as an unexpected role for ion transport. We functionally validate 7 shared ASD-CHD genes, including an ion transport gene, sodium voltage-gated channel alpha subunit 2 (SCN2A), confirming that ion transport plays an earlier role in development than expected.

RESULTS

A human molecular network integrating *de novo* variants

Our network analysis was based on the parsimonious composite network (PCNet), a resource of 2.7 million physical and functional associations among human genes (Huang et al., 2018). PCNet is formulated from a consensus of 21 molecular interaction databases and integrates multiple lines of evidence across tissues, including protein-protein interactions, co-expression, literature curation, and other measures (Figure 1A).

We first confirmed that genes with putatively damaging *de novo* variants (dDNV) identified in ASD, or separately in CHD, were significantly interconnected in the PCNet network (Jin et al., 2017; Homsy et al., 2015) (Figures S1A and S1B, ASD $p = 4 \times 10^{-83}$, CHD $p = 3 \times 10^{-29}$, degree-matched permutation test, STAR Methods). These observations were consistent with previous findings that genes with dDNVs in patients with ASD are more connected in protein-protein interaction networks than expected by chance (Chang et al., 2015; O'Roak et al., 2012; Liu et al., 2014) and that CHD risk factors functionally converge in protein networks (Lage et al., 2012). In contrast, genes with dDNVs from unaffected siblings of ASD probands were not significantly interconnected in the network (p = 0.20, Figure S1C).

Network propagation identifies risk genes for ASD or CHD

As a quantitative means of scoring network proximity among a set of disorder risk genes, we turned to the framework of network propagation, a general mathematical tool that has been repeatedly and successfully applied in biological analyses (Cowen et al., 2017) (Figure 1A). Here, we used network propagation to spread the signal from well-established risk genes to other genes in the network neighborhood, with the goal of implicating a broader network of genes for which the propagated genetic signal was above a set threshold (we used $z_D \ge 2$, where z_D is the network proximity score for disorder D, STAR Methods). The expanded network could include direct network neighbors of established genes or those at greater distances, depending on the convergent signal. Risk genes, defined previously by unbiased whole-exome sequencing studies based on a diagnosis of ASD or CHD, were used to seed the analysis: 65 high confidence ASD risk genes (Sanders et al., 2015) and 66 genes with recurrent dDNVs in CHD probands (Jin et al., 2017) (Figure 1B). Individual disorder subnetworks were then created, separately for ASD (1,583 genes; Table S1) and CHD (1,081 genes; Table S2), using network propagation seeded from the ASD or CHD risk genes.

Since the majority of dDNVs occur in genes not yet confidently associated with disease ("non-seed" genes, Figure 1C), we sought to determine whether non-seed genes were identified within our ASD or CHD subnetworks were enriched for dDNVs from probands. Indeed, dDNVs from individuals affected with ASD or with CHD were enriched within the disorder-specific subnetworks (Figures S1D–S1F; ASD $p = 1 \times 10^{-9}$, CHD $p = 2 \times 10^{-7}$, hypergeometric tests), but not dDNVs from unaffected siblings (ASD p = 0.10; CHD p = 0.27 Figure S1G).

Given this enrichment, we used the ASD and CHD disorderspecific networks for risk gene discovery, revealing 27 network-implicated autism risk genes and 46 network-implicated congenital heart disease risk genes (exome-wide Bonferroni-corrected p value ≤ 0.1 ; Table S1 and S2). We found that these putative risk genes appear inherently different from established risk genes, in that they have significantly lower pLI (probability of loss-of-function intolerance; ASD p = 0.02; CHD $p = 1 \times 10^{-4}$), Miz_Z (missense Z; ASD p = 0.01; CHD p = 0.001), and S_{Het} (selection coefficient; ASD p = 0.003; CHD p = 0.01; Wilcoxon rank-sum tests) (Lek et al., 2016).

Analysis of gene sets specific to brain region and development period (Xu et al., 2014) revealed that the broader set of networkimplicated ASD risk genes (excluding seed genes) was significantly enriched for genes with early and mid-fetal expression signatures in the amygdala, cerebellum, cortex, striatum, and thalamus (Table S3; STAR Methods), consistent with previous findings (Satterstrom et al., 2020; Willsey et al., 2013; Parikshak et al., 2013; Shohat et al., 2017). Young adulthood and adolescence cortex were also significantly enriched, although to a much lesser extent. Furthermore, the broader set of ASD risk genes implicated by the network included 19 genes that had been newly identified as ASD risk genes in the latest sequencing study (Satterstrom et al., 2020) ($p = 3 \times 10^{-8}$, hypergeometric test), validating the predictive value of our network approach and suggesting that additional network-implicated risk gene candidates are likely to be identified in future studies. Similarly,



Figure 1. Overview of genetic data and workflow

(A) Study workflow. We determined the network intersection of two sets of disorder risk genes, autism spectrum disorder (ASD, left, green) and congenital heart disease (CHD, right, pink); red denotes network proximity to both. Gene variants in the shared network are over-represented in patients from an independent replication cohort with co-morbid neurodevelopmental disorders (NDDs) and heart conditions, and disruptions to these genes give rise to brain and heart defects in a *Xenopus tropicalis* model of development. Analysis of ASD-CHD network architecture reveals a hierarchical organization of functional modules. SSC, Simons simplex collection; PCGC/PHN, Pediatric Cardiac Genomics Consortium/Pediatric Heart network.

(B) Numbers of patients in the ASD and CHD cohorts with neurodevelopmental and heart conditions, with numbers of *de novo* coding variants and established disorder risk genes. The ASD cohort has not been systematically examined for heart phenotypes.

(C) Bar chart displaying the fraction of patients in ASD and CHD cohorts with putatively damaging *de novo* variants (dDNVs) in ASD or CHD risk genes, dDNVs outside of these "risk" genes, and with no dDNVs yet identified.

the broader set of network-implicated CHD risk genes (Table S2, excluding seed genes) were significantly enriched for genes that are associated with CHD in mouse models (Izarzugaza et al., 2020) ($p = 2 \times 10^{-18}$, hypergeometric test).

Network intersection identifies ASD-CHD shared risk genes

To identify genes with network proximity to both ASD and CHD, we computed a combined ASD-CHD network proximity score, $z_{ASD-CHD}$, by taking the product of the separate network propagation scores for each disorder ($z_{ASD} \times z_{CHD}$) (Figures 2A, 2B, and S2A; STAR Methods). The number of genes at the intersection of the ASD and CHD networks was much larger than ex-

pected by chance, resulting in a distinct ASD-CHD network of 844 genes (empirical $p = 4 \times 10^{-141}$, Figures 2B and 2C; Table S4). Within this list, we further prioritized 398 high confidence genes with both $z_{ASD} > 1.5$ and $z_{CHD} > 1.5$ and a joint corrected p value < 0.1, reflecting genes that are highly connected to both disorders (Table S4).

To confirm that the ASD-CHD network represents genes carrying risk for both disorders, we assessed whether dDNVs present in CHD probands with NDDs (Homsy et al., 2015) (Figure 1B) were enriched among the ASD-CHD network genes (excluding seed genes). Indeed, we observed significant over-representation of these so-called "dual-condition" dDNVs within the ASD-CHD subnetwork (hypergeometric $p = 3 \times 10^{-4}$), but not

CellPress

Cell Systems Article





Figure 2. ASD and CHD de novo risk variants converge on common network regions

(A) Scatterplot of human proteins based on network proximity to established ASD risk genes or established CHD risk genes (y versus x). Proteins at the network intersection are shown in orange, with high confidence (HC; adjusted p < 0.1 and $z_{ASD} > 1.5$ and $z_{CHD} > 1.5$) proteins shown in red.

(B) Visualization of ASD-CHD network intersection, with the largest connected component shown. Node size is mapped to the number of variants observed for that gene, node color is the minimum of the brain or heart percentile expression from GTEx (The GTEx Consortium, 2015). Triangles indicate genes harboring at least one dual-condition variant. Edges are drawn for gene pairs with a network neighborhood similarity > 0.95 (percentile cosine similarity), with a purple-yellow color gradient indicating percentile above this threshold. Network layout was determined by a spring-embedding algorithm with edge bundling in cytoscape (Shannon et al., 2003). The spring layout was subsequently modified to enable visualization of the ASD and CHD seed nodes, where the seed nodes were manually translated to the left (ASD), right (CHD), and down (ASD and CHD) from the original layout. Selected genes are labeled. High confidence network intersection genes are indicated with a white border.

(C) Number of genes found at the ASD-CHD network intersection compared with the number expected by chance. ***empirical $p = 4 \times 10^{-141}$.

(D) Enrichment for dDNVs associated with dual NDD and CHD phenotypes is shown for ASD-specific network genes ($z_{ASD} \ge 2$, $z_{CHD} < 2$, $z_{ASD-CHD} < 3$), CHD-specific network genes ($z_{CHD} \ge 2$, $z_{ASD} < 2$, $z_{ASD} < 2$, $z_{ASD-CHD} < 3$), or genes at the ASD-CHD network intersection, and high confidence network intersection *** p < 0.001, **p < 0.01, *p < 0.05. Vertical bars indicate one standard deviation.

Cell Systems Article

CellPress



Figure 3. Validation of the ASD-CHD network in independent patient cohorts

(A) Distribution of minimum percentile mRNA expression of ASD-CHD network genes in GTEx brain or heart tissues (orange), compared with all expressed genes (gray).

(B) Distribution of minimum percentile mRNA expression of ASD-CHD network genes in mouse developmental brain or heart tissues (orange), compared with all expressed genes (gray).

(C) Heatmap showing the degree of network intersection among ASD rare variants, CHD rare variants, ASD common variants, and common variants from an unrelated control disease, atherosclerosis. Color bar shows enrichment in units of log 2 (observed/expected) number of genes.

* denotes larger-than-expected intersection at p < 0.05.

(D) Receiver operator characteristic (ROC) showing recovery of damaging variants associated with dual brain and heart phenotypes in the independent DECIPHER cohort. Genes are ranked by network proximity to ASD and CHD established risk genes ($z_{ASD-CHD}$). Results for all ranked genes in maroon (AUC = 0.71, p = 5 × 10⁻²²); results excluding established risk genes in salmon (AUC = 0.66, p = 5 × 10⁻¹⁴, empirical p values). Black circles highlight the sets of genes at the $z_{ASD-CHD} \ge 3$ cutoffs used to define the ASD-CHD network intersection in previous figures. Examined variants are damaging single nucleotide variants and indels.

(E) The number of dual-condition variants in DECIPHER is listed for all ASD-CHD network genes with four or more such variants. High confidence ASD-CHD network genes not previously identified in one or both disorders (Table S4) labeled in red.

within the ASD-specific (p = 0.75) or CHD-specific subnetworks (p = 0.21, STAR Methods; Figure 2D). As further evidence in support of this network, ASD-CHD network genes had higher brain and heart expression than would be expected by chance in both adult human tissue (The GTEx Consortium, 2015) (p = 3 × 10^{-70} , rank-sum test, Figure 3A) and in developing mouse tissue (Homsy et al., 2015) (p = 3 × 10^{-37} , rank-sum test; Figure 3B; embryonic day [E] 14.5 in heart; E9.5 in the brain).

We also examined genes with network proximity to one disorder only (STAR Methods). We found that genes proximal only to CHD risk genes were significantly enriched for dDNVs in CHD patients without symptoms of NDDs ($p = 1 \times 10^{-4}$, Figures S2B and S2C; Table S2), and they were not enriched for ASD dDNVs (p = 0.62). Reciprocally, genes proximal only to ASD risk genes were significantly enriched for dDNVs found in ASD patients (p = 0.04, Figures S2D and S2E; Table S1), and they were not enriched for CHD dDNVs (p = 0.10).

Network links between ASD rare and common variation

We next assessed the intersection between the top 65 ASD associated genes, based on *de novo* and rare inherited variants (Sanders et al., 2015), and genes implicated by proximity to common single nucleotide polymorphisms (SNPs) identified in the most recent ASD genome-wide association study (GWAS) (Grove et al., 2019). The size of the network intersection between these genes identified by rare or common variants was

CelPress

significantly and substantially larger than expected by chance ($p = 7 \times 10^{-34}$ with > 2-fold effect, Figure 3C), suggesting that these two types of variation impact a common molecular network. This finding extends a previous report that identified both rare and common variants in modules of genes co-expressed in the human brain (Ben-David and Shifman, 2012). While a large GWAS for CHD does not currently exist, in the future it will be interesting to see whether the same finding holds between CHD rare and common variation. Regardless, we saw a weaker, but significant, network intersection between the ASD genes identified by common variation and the CHD risk genes ($p = 2 \times 10^{-24}$), suggesting that common variant risk may also overlap across ASD and CHD.

To examine the specificity of our findings, we selected atherosclerosis as a negative control (Piñero et al., 2017), which has a similar number of risk genes as ASD and CHD and significant localization in PCNet (Figures S3A–S3D) but is not co-morbid with either ASD or CHD. We observed no significant network overlap of ASD or CHD variants with common variants for this condition (Figure 3C; STAR Methods), suggesting the observed overlaps do, in fact, reflect shared biology between common and rare variation.

Shared ASD-CHD risk genes replicate in an independent cohort

To validate the 844 genes of the ASD-CHD gene network, we asked whether they had damaging variants in an independent cohort of 2,628 individuals having both "Abnormality of the nervous system" and "Abnormality of the cardiovascular system" extracted from 29,069 individuals in the DECIPHER database (Firth et al., 2009) (STAR Methods; Tables S5 and S6). For this validation cohort, we indeed observed an over-representation of likely damaging variants within the 844 genes (Figures 3D and S4, empirical $p = 5 \times 10^{-22}$). This effect was maintained when excluding the established ASD and CHD risk genes (Figures 3D and S4, empirical $\rho = 5 \times 10^{-14}$. STAR Methods). Some of these genes contained variants observed in large numbers of DECIPHER probands. For example, KANSL1, ANKRD11, CREBBP, and KAT6A each harbored damaging variants from ten or more dual-phenotype individuals (Figure 3E). Altogether, we identified 101 genes in the ASD-CHD high confidence network with at least one damaging variant observed in a dual-phenotype DECIPHER or Pediatric Cardiac Genomics Consortium (PCGC)/Pediatric Heart Network (PHN) proband, which were not seed genes for both disorders. Of these, 98 genes had not been previously identified in one or both disorders (Table S4).

Prioritized ASD-CHD genes reveal atypical brain and heart development in *Xenopus tropicalis*

Candidate ASD-CHD risk genes with the highest number of dualcondition variants in the DECIPHER database were prioritized for validation (Table S5). Excluding genes with prior evidence in both ASD and CHD, we proceeded with validation for 3 genes, which were seeds for CHD but not ASD (*KMT2A*, *PTPN11*, and *KMT2D*) and 3 genes that were not seeds for either disorder (*KANSL1*, *KAT6A*, and *MAPT*) (Table S7; with the additional gene SCN2A, described below, a total of 7 genes were validated). Validation testing leveraged the *Xenopus tropicalis*, a powerful *in vivo*

Cell Systems Article

vertebrate model for uncovering fundamental mechanisms of human development and pathobiology (Kaltenbrun et al., 2011; Hwang et al., 2019; Sater and Moody, 2017; Blum and Ott, 2018), including ASD and CHD (Willsey et al., 2018b, 2018a, 2021; Garfinkel and Khokha, 2017; Exner and Willsey, 2021). We tested whether disruption of each gene was able to phenocopy known abnormalities in the brain and/or heart development for ASD and CHD, respectively (Willsey et al., 2021; Garfinkel and Khokha, 2017; Duncan and Khokha, 2016). We conducted CRISPR-Cas9 mutagenesis in the F0 generation via injection of Cas9 protein and a single guide RNA (sgRNA) targeting the gene of interest at the two-cell stage (Figure 4A). Animals were phenotyped at tadpole stages by fluorescence microscopy, with researchers blinded to the gene name (Figure 4A). Hearts were assessed for gross abnormalities including situs inversus and size differences, as these phenotypes have previously been observed for CHD risk genes in Xenopus (Garfinkel and Khokha, 2017; Duncan and Khokha, 2016). Similarly, brains were phenotyped for increased variation in forebrain size, as has been observed in Xenopus for multiple large-effect ASD risk genes (Willsey et al., 2018, 2020, 2021).

Targeting a negative control pigmentation gene did not cause any brain or heart anatomical phenotypes (*slc45a2*, Figures 4B and 4C), whereas positive control genes are known to have roles in both ASD and CHD (*nsd1* and *ankrd11*) (Satterstrom et al., 2020; Jin et al., 2017; Homsy et al., 2015; Zaidi and Brueckner, 2017) resulted in severe morphological changes to both the heart and brain (Figures 4B, 4C, and S5). Notably, for all 7 genes undergoing validation, we observed known brain and heart phenotypes previously associated with ASD or CHD in *Xenopus* at a greater frequency than in controls, albeit with different degrees of penetrance and severity by gene (Figures 4B, 4C, and S5A– S5C; Table S8). The most common phenotypes among the genes were an increase in telencephalon size variance (Figure S5A), changes in the heart size (Figure S5B), and reversed heart *situs* (Figure S5C).

Hierarchical organization of the ASD-CHD network implicates shared molecular mechanisms

To chart the functional organization of the ASD-CHD network, we applied the technique of network community detection (Kramer et al., 2014; Yu et al., 2019) to identify modular systems of genes enriched for high densities of molecular interactions (STAR Methods). We identified 120 gene systems organized hierarchically, with small specific systems contained within those that are progressively larger and more general as one moves upward in the hierarchy (Figure 5A). Systems corresponding to known biological functions were labeled by alignment to multiple gene knowledge bases (STAR Methods), revealing major functional branches related to chromatin and histone modification, ion channels and transport, Notch signaling, and MAP kinase signaling (Figure 5A). Approximately, a quarter of systems corresponded to well-characterized biological functions (27%), while many others (55%) expanded a known function by adding genes or by factoring that function into multiple distinct subsystems (Figure 5B; Table S9). The remaining 22 systems (18%) did not clearly enrich for a known function, representing putative functional gene assemblages. A convergence of replicated genes (genes with at least one

Cell Systems Article

CellPress



Figure 4. Functional validation in Xenopus tropicalis

(A) CRISPR mutagenesis strategy. CRISPR reagents and a tracer dye (red) are injected at the two-cell stage, either into both cells (bilateral mutants, heart phenotyping) or into one cell (unilateral mutants, brain phenotyping). Animals are grown to tadpole stages and phenotyped for the brain (top) or heart (bottom) anatomy. The brain is normally bilaterally symmetric. Note leftward heart looping direction. Telencephalon (tel), outflow tract (OFT), ventricle (V), right atrium (RA), and left atrium (LA).

(B) Quantification of brain and heart phenotype penetrance by gene mutant (percent observed in mutagenized animals). Negative control pigmentation gene *slc45a2* shows few heart or brain phenotypes, while positive control genes *ankrd11* and *nsd1* show both phenotypes. All predicted ASD-CHD network genes show some evidence of heart and brain phenotypes, albeit at differing penetrances. Sample sizes are presented per gene, with brain phenotyping sample size first, followed by heart phenotyping sample size.

(C) Brain telencephalon (top, β-tubulin staining in green) and heart (bottom, phalloidin actin staining in pink) images for three gene disruptions representative of the twelve total gene disruptions conducted and quantified in (B). Gene symbols are indicated at the lower right of each image. Shown is a negative control pigmentation gene *slc45a2* (far left), positive control gene *nsd1* (middle), and ASD-CHD network-implicated genes (right). For brain images, note the change in brain size on the injected half (red dotted outline) compared with control (white dotted outline). For heart categories, note changes in heart ventricle size (white dotted outline) and/or outflow tract looping direction (arrow) and size compared with control. Size differences in the total heart and outflow tract size can be readily seen in *nsd1*, *ptpn11*, and *kansl1* heart images. Scale bars are 100 μm. Sample sizes and quantification of telencephalon size variance and heart phenotypes are in Figure S5.

dual-condition dDNV in DECIPHER) was observed in select chromatin modification subsystems (e.g., System 886, 9/17 genes replicated, Figure S6A), supporting results from previous studies (Homsy et al., 2015). Convergence of replicated genes was also observed in brain morphogenesis subsystems (e.g., System 920, 19/37 genes replicated, Figure S6B) and in system 901 for which 7/22 genes replicated in DECIPHER (Figure S6C; Tables S1 and S10). While it is not yet obvious how the various genes in system 901 interrelate within the broader supersystem function of chromatin modification (*CHD8*, *SIN3B*, *ARID1A*, and *RNF44*, among others), it is clear that they have many more interconnections than would be expected by chance, suggesting

CellPress

Cell Systems Article



Figure 5. Hierarchical organization of the ASD-CHD network

(A) Representation of the network data as a hierarchy of densely interacting gene systems. Nodes represent distinct systems, with the size showing the number of genes (min 6, max 844) and color showing the fraction of these genes in which dual-condition variants were identified. Large nodes with bold outlines contain at least one *Xenopus* validated gene. Systems mentioned elsewhere in the text are indicated with a gray or blue background. The gene systems are labeled with a unique identification (represented as a number next to each system). Parent-child containment relationships were defined by the CliXO algorithm, with genes in child systems contained completely within parent systems. Systems mentioned elsewhere in the text are indicated with a gray or blue background.
(B) Pie chart demonstrating the fractions of identified systems that are well-characterized, new/extended subprocesses, new superprocesses, or new processes.
(C) Genes and interactions from four ion-channel modules are highlighted in (A). Triangles indicate genes in which dual-condition variants have been identified. High confidence network intersection genes are indicated with a white border. Node color mapped to the minimum brain or heart percentile expression for that gene (blue gradient). Edges with network cosine similarity >0.95 are shown, with color indicating the magnitude of similarity above this threshold.

a common functional role within the scope of chromatin modification (Figures S6D and S6E).

SCN2A and other ion channels underlie ASD-CHD co-morbidity

A prominent collection of systems in the ASD-CHD systems map was related to ion channels and transport (Figure 5C). While individual genes in these systems had only 1–2 replicated dual-condition variants each when considering the network structure, the number of dual-condition validated genes within ion transport systems was much higher than expected by chance ($p = 1 \times 10^{-6}$, hypergeometric test). This discovery highlights the utility of a systems-level approach in which convergence is identified among a group of functionally interconnected genes, in the absence of recurrent gene-level variants in dual-phenotype patients. Although ion channels have known roles in neuronal function and heart rhythm (Ackerman, 1998; Colbert and Pan, 2002), their role in heart and brain morphogenesis is less clear. While some DECIPHER-replicated ion-channel genes had been previously associated with ASD, none had previously been linked to CHD. These include *GRIN2A, GRIN2B, KCNQ2, SCN2A*, and *SLC6A1* (Figure 5C).

Of particular note was *SCN2A*, which had been previously reported to be expressed exclusively in the adult prefrontal cortex where it promotes the firing of mature excitatory neurons (Sanders et al., 2018; Spratt et al., 2019). Given the strong network proximity to both CHD and ASD and its replication in DECIPHER, we added *SCN2A* to the list of genes to be phenotyped in *Xenopus* following CRISPR-Cas9 mutagenesis. We observed marked defects during brain and heart development, including frank *situs inversus* of heart laterality, as well as

scn2a

Brain

Eye

Heart

Intestine

Kidney

Skin

TPM

9.0

0.0

Cell Systems Article







Figure 6. SCN2A is required for heart and brain development

CellPress

(A) Unilateral mutagenesis of scn2a leads to an increase in telencephalon size on the mutated (right) side. B-tubulin staining.

(B) High magnification of boxed region in (A). Note increase in the size of the olfactory bulb (ob) and rest of telencephalon (tel).

(C) Control tadpole with proper organ situs (gallbladder, GB, on the image left and gut on image right) and proper heart looping (toward image left).

(D) High magnification of boxed region in (C) showing heart. Note outflow tract loops toward the image left.

(E) scn2a bilateral mutant shows organ situs inversion (GB and gut in opposite positions from control) and heart looping defects.

(F) High magnification of boxed region in (E) showing heart. Note outflow tract defects (elongated, looping toward image right) and abnormal ventricle morphology.

(G) Adult human expression of SCN2A by tissue, using data from GTEx. Note strong expression in the brain but not the heart.

(H) Adult Xenopus laevis expression of scn2a by tissue. Note strong expression in the brain but not the heart.

(I) Embryonic expression of scn2a in Xenopus tropicalis. Numbers on x axis represent the developmental stage. Note expression during mesoderm specification (pink, late blastula/early gastrula stages) as well as during left-right patterning and heart morphogenesis.

(J) Embryonic human expression of SCN2A. Note expression during the blastocyst stage when mesoderm specification occurs (pink), similar to Xenopus.

(Yandım and Karakülah, 2019: Owens et al., 2016). Moreover, we found that SCN2A was expressed in the developing mouse heart and brain at similar levels (28th percentile in heart, E14.5; 23rd percentile in the brain, E9.5) (Homsy et al., 2015). Therefore, in addition to the function of SCN2A in mature differentiated cells (Sanders et al., 2018; Spratt et al., 2019), the data indicate a second conserved role for SCN2A in vertebrate heart and brain development and possibly other organ systems.

DISCUSSION

Risk variants often converge not on single genes but larger-scale biological

heart looping defects (Figures 4B, 6A-6F, and S5). While SCN2A is indeed highly expressed in adult brains but not hearts of humans (Figure 6G) (The GTEx Consortium, 2015) or Xenopus (Figure 6H) (Session et al., 2016), it is also expressed at early embryonic stages during germ layer specification, leftright patterning, and/or heart morphogenesis and neural differentiation in Xenopus and human (Figures 6I and 6J) structures and functions (Barabási et al., 2011; Navlakha and Kingsford, 2010; Califano et al., 2012), with the consequence that greater power to detect associations can be obtained by examining levels above the gene. For this purpose, large knowledge networks have been and continue to be assembled, including data from ongoing systematic gene and protein interaction screens and curation of previous literature (Huttlin et al., 2015;



Szklarczyk et al., 2015; Huang et al., 2018; Luck et al., 2020; Li et al., 2017). Here, we have demonstrated that by leveraging such networks, it is possible to identify risk genes and pinpoint shared molecular mechanisms between co-morbid but otherwise physiologically distinct disorders. As the diagnosis of CHD is typically made *in utero*, the ability to predict the subset of CHD patients at high risk for ASD moves us closer to prioritizing newborns with CHD for early screening, behavioral and therapeutic interventions, and/or prospective observational studies (Willsey et al., 2018).

The network-implicated ASD and CHD genes appear to represent a different class of risk genes from those identified in current large-scale patient sequencing studies, which typically focus on rare, heterozygous variants for gene discovery and therefore tend to identify haploinsufficient genes intolerant to mutation (e.g., high pLl, Mis_Z, and S_{Het}). In contrast, the network-implicated risk genes tend to have a lower pLl and lower heterozygous selection coefficients (S_{het}), consistent with recessive inheritance patterns (i.e., requiring homozygous mutations for a clinical phenotype). These results suggest that the putative risk genes may be enriched for haplosufficient genes with recessive inheritance patterns. On the other hand, we acknowledge that low loss-of-function intolerance alone is only suggestive of a recessive inheritance pattern, and it is also possible that a number of these network-implicated genes are false positives.

Further, our results show that rare and common variants in ASD converge on the same molecular network, implying shared function. Notably, the overlap between rare and common variants discovered here occurs not at the level of individual genes but across sets of genes interconnected by the molecular network. This result is consistent with the lack of significant enrichment of GWAS signals within ASD risk genes identified by whole-exome sequencing (Satterstrom et al., 2020), though a nominally significant overlap was observed in the study by Grove et al. (2019). An area of future research will be to see if variants in genes causing more severe symptoms impact separate network neighborhoods and pathways from variants in genes causing milder symptoms.

Our study uncovered a strong link between ion channels and ASD and CHD. The genetic signal from de novo variants was thinly spread over many different ion-channel genes, however, suggestive of why this signal has previously been underappreciated. This result again demonstrates the power of multi-scale network analysis, as each individual ion channel is only rarely impacted but, when considering a larger family of genes within a molecular network, the association becomes clear. We used the Xenopus model to test genetic disruptions in one of these ion-channel genes, SCN2A, resulting in major heart and brain structural defects. Based on these results, a crucial future area of investigation could be to study the molecular mechanisms by which ion transport factors jointly influence heart and brain development. SCN2A perturbation led to laterality defects, suggesting an impact on global left-right patterning. Such patterning would be consistent with evidence that ion transport regulates the left-right axis (Levin et al., 2002), although the mechanisms remain controversial (Walentek et al., 2012; Delling et al., 2016). While sodium channels like SCN2A would typically be ascribed a postnatal role in action-potential neuronal physiology or cardiac rhythm (Spratt et al., 2019), our results suggest earlier

Cell Systems Article

expression and function during embryonic development. Such a developmental role has implications for pathobiology and the timing of gene therapy approaches. For example, gene therapies delivered postnatally may not be able to reverse defects introduced during embryogenesis. Beyond ion channels, the *Xenopus* model adds to a growing body of evidence implicating chromatin remodelers in both brain and heart development (Satterstrom et al., 2020; Jin et al., 2017), including *KMT2A, KAT6A, KMT2D*, and *KANSL1*. The findings presented here also facilitate future prioritization of risk genes for further study in model organisms, with the expectation that such genes may vary in their effect sizes, with some genes causing only mild conditions, and others causing major abnormalities, due in part to haploinsufficiency.

Some of the genes validated in *Xenopus* have previous links to cardiac abnormalities and/or NDD in general; the disruption of *PTPN11* is known to cause Noonan syndrome for example (Tartaglia et al., 2002). However, here we associate these genes specifically with ASD and CHD and provide direct experimental evidence that these dual-phenotype genes phenocopy reliable and published phenotypes for ASD and CHD risk genes in *Xenopus* (Willsey et al., 2021; Garfinkel and Khokha, 2017; Duncan and Khokha, 2016). While there has been debate in the field regarding the extent to which ASD genes differ from NDD genes, and while some genes certainly carry a risk for both NDD and ASD, recent work supports the idea that there are discrete sets of ASD-predominant and NDD-predominant genes (Satterstrom et al., 2020).

We expect that the predictive value of network approaches will continue to improve as molecular interaction networks become more complete. Currently, large-scale protein-protein interaction mapping has been completed in a yeast-two-hybrid setting (Luck et al., 2020) and is progressing for a first human cell-line model, e.g., BioPlex (Huttlin et al., 2015) or CellMap (Go et al., 2019) in HEK293 cells. Other ongoing efforts are focused on mapping the specific molecular and genetic interactions underlying ASD and other NDDs (Willsey et al., 2018). As these and other projects continue to fill gaps in network knowledge, one might anticipate that the ability to associate genotype with phenotype will also improve markedly (Willsey et al., 2018). Such associations will also be improved by other factors, such as more complete patient phenotyping and the inclusion of additional classes of genetic variants (Young, 2019).

Finally, while the focus of this work has been on ASD and CHD, one can envisage the application of a similar methodology to any group of genetic disorders for which there is suspected comorbidity. The list of documented co-morbid conditions is already quite extensive, not only among developmental disorders like ASD and CHD but for common disorders of adulthood. Indeed, diagnoses such as cancer (Bluethmann et al., 2016) and type II diabetes (Schram et al., 2014) typically present not in isolation but as embeddings within a complex web of conditions, many of which have a suspected but poorly characterized shared genetic component. Understanding the molecular connections between these conditions will not only shed light on the etiology of these diseases but may also help predict offtarget effects of therapeutics and repurpose therapeutics across disorders. In this light, perhaps it is not so surprising that a web of phenotypes would be tied intimately to a web of molecular





interactions. Insofar, as that precept continues to hold, our experience here argues that (network) knowledge is power.

STAR***METHODS**

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Animal models
- METHOD DETAILS
 - Data acquisition
 - Xenopus mutagenesis and genotyping
 - Fluorescence staining, microscopy, and phenotyping
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Measuring network localization of a gene set
 - Network propagation
 - Joint network propagation to identify network intersection between two disorders
 - Network visualization
 - Enriched brain regions
 - ASD common variant analysis
 - Control disorder network overlap
 - O Replication of genetic variants in DECIPHER
 - Gene discovery
 - Construction of the ASD-CHD systems hierarchy
 - O Classification of systems by prior literature support
 - Selection of candidate genes for functional validation in *Xenopus*

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j. cels.2021.07.009.

ACKNOWLEDGMENTS

We thank Dr Matthew State (UCSF) for generous infrastructural support and advice. We thank Xenbase (RRID:SCR_003280) for essential daily reference and the National *Xenopus* Resource (RRID:SCR_013731) for wild-type *X. tropicalis* and technical advice. Funding: this work is a component of the NIMH Convergent Neuroscience Initiative and the Psychiatric Cell Map Initiative (pcmi.ucsf.edu) and was supported by NIH grant 1U01MH115747-01A1 to A.J.W. and T.I. We also gratefully acknowledge support for this study from the National Institute for General Medical Sciences (P41 GM103504); the National Human Genome Research Institute (R01 HG009979); the Clinical and Translational Science Awards Program (CTSA, UL1TR001442); the National Heart Lung and Blood Institute (R01HL149746 to M.K. and E.S.); and the Overlook International Foundation.

AUTHORS CONTRIBUTIONS

S.B.R. performed most of the computational analyses, interpreted the results, and co-wrote the manuscript. H.R.W. designed the experiments, interpreted the results, and co-wrote the manuscript. Y.X. and J.D. performed the experiments, and Y.X. participated in results interpretation. S.W. participated in computational analyses. C.C. participated in figure preparation. M.K., E.S., M.W.S., N.C., S.W., and Y.M. participated in results interpretation and manu-

script review. A.J.W., K.M.F., and T.I. jointly supervised the analysis and cowrote the manuscript.

DECLARATION OF INTERESTS

T.I. is co-founder of Data4Cure, is on the Scientific Advisory Board, and has an equity interest. T.I. is on the Scientific Advisory Board of Ideaya BioSciences, has an equity interest, and receives sponsored research funding. The terms of these arrangements have been reviewed and approved by the University of California San Diego in accordance with its conflict of interest policies.

Received: October 23, 2020 Revised: May 10, 2021 Accepted: July 28, 2021 Published: August 18, 2021

REFERENCES

Abrahams, B.S., Arking, D.E., Campbell, D.B., Mefford, H.C., Morrow, E.M., Weiss, L.A., Menashe, I., Wadkins, T., Banerjee-Basu, S., and Packer, A. (2013). SFARI gene 2.0: a community-driven KnowledgeBase for the autism spectrum disorders (ASDs). Mol. Autism *4*, 36. https://doi.org/10.1186/2040-2392-4-36.

Ackerman, M.J. (1998). The long QT syndrome: ion channel diseases of the heart. Mayo Clin. Proc. Mayo Clin. 73, 250–269.

Andonian, C., Beckmann, J., Biber, S., Ewert, P., Freilinger, S., Kaemmerer, H., Oberhoffer, R., Pieper, L., and Neidenbach, R.C. (2018). Current research status on the psychological situation of adults with congenital heart disease. Cardiovasc. Diagn. Ther. 8, 799–804.

Barabási, A.L., Gulbahce, N., and Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. Nat. Rev. Genet. *12*, 56–68.

Ben-David, E., and Shifman, S. (2012). Networks of neuronal genes affected by common and rare variants in autism spectrum disorders. PLoS Genet. *8*, e1002556.

Bluethmann, S.M., Mariotto, A.B., and Rowland, J.H. (2016). Anticipating the 'silver tsunami': prevalence trajectories and comorbidity burden among older cancer survivors in the United States. Cancer Epidemiol. Biomarkers Prev. *25*, 1029–1036.

Blum, M., and Ott, T. (2018). Xenopus: an undervalued model organism to study and model human genetic disease. Cells Tissues Organs 205, 303–313.

Brinkman, E.K., Chen, T., Amendola, M., and van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Res. *42*, e168.

Califano, A., Butte, A.J., Friend, S., Ideker, T., and Schadt, E. (2012). Leveraging models of cell regulation and GWAS data in integrative network-based association studies. Nat. Genet. *44*, 841–847.

CDC (2018). Data & statistics on autism spectrum disorder. (ASD) | NCBDDD | CDC, Centers for Disease Control and Prevention. https://www.cdc.gov/ ncbddd/autism/data.html.

Chang, J., Gilman, S.R., Chiang, A.H., Sanders, S.J., and Vitkup, D. (2015). Genotype to phenotype relationships in autism spectrum disorders. Nat. Neurosci. *18*, 191–198.

Civelek, M., and Lusis, A.J. (2014). Systems genetics approaches to understand complex traits. Nat. Rev. Genet. *15*, 34–48.

Colbert, C.M., and Pan, E. (2002). Ion channel properties underlying axonal action potential initiation in pyramidal neurons. Nat. Neurosci. *5*, 533–538.

Cowen, L., Ideker, T., Raphael, B.J., and Sharan, R. (2017). Network propagation: A universal amplifier of genetic associations. Nat. Rev. Genet. *18*, 551–562.

de la Torre-Ubieta, L., Won, H., Stein, J.L., and Geschwind, D.H. (2016). Advancing the understanding of autism disease mechanisms through genetics. Nat. Med. *22*, 345–361.

De Rubeis, S., He, X., Goldberg, A.P., Poultney, C.S., Samocha, K., Cicek, A.E., Kou, Y., Liu, L., Fromer, M., Walker, S., et al. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. Nature *515*, 209–215.



Cell Systems Article

DeLay, B.D., Corkins, M.E., Hanania, H.L., Salanga, M., Deng, J.M., Sudou, N., Taira, M., Horb, M.E., and Miller, R.K. (2018). Tissue-specific gene inactivation in Xenopus laevis: knockout of lhx1 in the kidney with CRISPR/Cas9. Genetics 208, 673–686.

Delling, M., Indzhykulian, A.A., Liu, X., Li, Y., Xie, T., Corey, D.P., and Clapham, D.E. (2016). Primary cilia are not calcium-responsive mechanosensors. Nature *531*, 656–660.

Doan, R.N., Lim, E.T., De Rubeis, S., Betancur, C., Cutler, D.J., Chiocchetti, A.G., Overman, L.M., Soucy, A., Goetze, S., Autism Sequencing Consortium, et al. (2019). Recessive gene disruptions in autism spectrum disorder. Nat. Genet. *51*, 1092–1098.

Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K., and Liu, X. (2015). Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum. Mol. Genet. *24*, 2125–2137.

Duncan, A.R., and Khokha, M.K. (2016). Xenopus as a model organism for birth defects-congenital heart disease and heterotaxy. Semin. Cell Dev. Biol. *51*, 73–79.

Exner, C.R.T., and Willsey, H.R. (2021). Xenopus leads the way: frogs as a pioneering model to understand the human brain. Genesis *59*, e23405.

Fakhro, K.A., Choi, M., Ware, S.M., Belmont, J.W., Towbin, J.A., Lifton, R.P., Khokha, M.K., and Brueckner, M. (2011). Rare copy number variations in congenital heart disease patients identify unique genes in left-right patterning. Proc. Natl. Acad. Sci. USA *108*, 2915–2920.

Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R.M., and Carter, N.P. (2009). DECIPHER: database of chromosomal imbalance and phenotype in humans using Ensembl resources. Am. J. Hum. Genet. *84*, 524–533.

Fischbach, G.D., and Lord, C. (2010). The Simons simplex collection: a resource for identification of autism genetic risk factors. Neuron 68, 192–195.

Flint, J., and Ideker, T. (2019). The great hairball gambit. PLoS Genet. 15, e1008519.

Garfinkel, A.M., and Khokha, M.K. (2017). An interspecies heart-to-heart: using Xenopus to uncover the genetic basis of congenital heart disease. Curr. Pathobiol. Rep. 5, 187–196.

Gaynor, J.W., Stopp, C., Wypij, D., Andropoulos, D.B., Atallah, J., Atz, A.M., Beca, J., Donofrio, M.T., Duncan, K., Ghanayem, N.S., et al. (2015). Neurodevelopmental outcomes after cardiac surgery in infancy. Pediatrics *135*, 816–825.

Gelb, B.D., and Chung, W.K. (2014). Complex genetics and the etiology of human congenital heart disease. Cold Spring Harb. Perspect. Med. 4, a013953.

Glessner, J.T., Bick, A.G., Ito, K., Homsy, J., Rodriguez-Murillo, L., Fromer, M., Mazaika, E., Vardarajan, B., Italia, M., Leipzig, J., et al. (2014). Increased frequency of de novo copy number variants in congenital heart disease by integrative analysis of single nucleotide polymorphism array and exome sequence data. Circ. Res. *115*, 884–896.

Go, C.D., Knight, J.D.R., Rajasekharan, A., Rathod, B., Hesketh, G.G., Abe, K.T., Youn, J.-Y., et al. (2019). A proximity biotinylation map of a human cell. bioRxiv. https://doi.org/10.1101/796391.

Grove, J., Ripke, S., Als, T.D., Mattheisen, M., Walters, R.K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O.A., Anney, R., et al. (2019). Identification of common genetic risk variants for autism spectrum disorder. Nat. Genet. *51*, 431–444.

Guney, E., Menche, J., Vidal, M., and Barábasi, A.-L. (2016). Network-based in silico drug efficacy screening. Nat. Commun. 7, 10331.

Harris, M.A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., et al. (2004). The gene ontology (GO) database and informatics resource. Nucleic Acids Res. *32*, D258–D261.

Homsy, J., Zaidi, S., Shen, Y., Ware, J.S., Samocha, K.E., Karczewski, K.J., DePalma, S.R., McKean, D., Wakimoto, H., Gorham, J., et al. (2015). De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. Science *350*, 1262–1266.

Huang, J.K., Carlin, D.E., Yu, M.K., Zhang, W., Kreisberg, J.F., Tamayo, P., and Ideker, T. (2018). Systematic evaluation of molecular networks for discovery of disease genes. Cell Syst. *6*, 484–495.e5.

Huttlin, E.L., Ting, L., Bruckner, R.J., Gebreab, F., Gygi, M.P., Szpyt, J., Tam, S., Zarraga, G., Colby, G., Baltier, K., et al. (2015). The BioPlex network: a systematic exploration of the human interactome. Cell *162*, 425–440. https://doi. org/10.1016/j.cell.2015.06.043.

Hwang, W.Y., Marquez, J., and Khokha, M.K. (2019). Xenopus: driving the discovery of novel genes in patient disease and their underlying pathological mechanisms relevant for organogenesis. Front. Physiol. *10*, 953.

lossifov, I., O'Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E., et al. (2014). The contribution of de novo coding mutations to autism spectrum disorder. Nature *515*, 216–221.

Izarzugaza, J.M.G., Ellesøe, S.G., Doganli, C., Ehlers, N.S., Dalgaard, M.D., Audain, E., Dombrowsky, G., Banasik, K., Sifrim, A., Wilsdon, A., et al. (2020). Systems genetics analysis identifies calcium-signaling defects as Novel cause of congenital heart disease. Genome Med. *12*, 76.

Jin, S.C., Homsy, J., Zaidi, S., Lu, Qiongshi, Morton, S., DePalma, S.R., Zeng, X., Qi, H., Chang, W., Sierant, M.C., et al. (2017). Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. Nat. Genet. *49*, 1593–1601.

Kaltenbrun, E., Tandon, P., Amin, N.M., Waldron, L., Showell, C., and Conlon, F.L. (2011). Xenopus: an emerging model for studying congenital heart disease. Birth Defects Res. A Clin. Mol. Teratol. *91*, 495–510.

Karimi, K., Fortriede, J.D., Lotay, V.S., Burns, K.A., Wang, D.Z., Fisher, M.E., Pells, T.J., James-Zorn, C., Wang, Y., Ponferrada, V.G., et al. (2018). Xenbase: a genomic, epigenomic and transcriptomic model organism database. Nucleic Acids Res. *46*, D861–D868.

Kramer, M., Dutkowski, J., Yu, M., Bafna, V., and Ideker, T. (2014). Inferring gene ontologies from pairwise similarity data. Bioinformatics *30*, i34–i42.

Lage, K., Greenway, S.C., Rosenfeld, J.A., Wakimoto, H., Gorham, J.M., Segrè, A.V., Roberts, A.E., Smoot, L.B., Pu, W.T., Pereira, A.C., et al. (2012). Genetic and environmental risk factors in congenital heart disease functionally converge in protein networks driving heart development. Proc. Natl. Acad. Sci. USA *109*, 14035–14040.

Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z., and Bergmann, S. (2016). Fast and rigorous computation of gene and pathway scores from SNP-based summary statistics. PLoS Comput. Biol. *12*, e1004714.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285–291.

Levin, M., Thorlin, T., Robinson, K.R., Nogi, T., and Mercola, M. (2002). Asymmetries in H+/K+-ATPase and cell membrane potentials comprise a very early step in left-right patterning. Cell *111*, 77–89.

Li, T., Wernersson, R., Hansen, R.B., Horn, H., Mercer, J., Slodkowicz, G., Workman, C.T., Rigina, O., Rapacki, K., Stærfeldt, H.H., et al. (2017). A scored human protein-protein interaction network to catalyze genomic interpretation. Nat. Methods *14*, 61–64.

Liu, L., Lei, J., Sanders, S.J., Willsey, A.J., Kou, Y., Cicek, A.E., Klei, L., Lu, C., He, X., Li, M., et al. (2014). DAWN: a framework to identify autism genes and subnetworks using gene expression and genetics. Mol. Autism *5*, 22.

Luck, K., Kim, D.-K., Lambourne, L., Spirohn, K., Begg, B.E., Bian, W., Brignall, R., Cafarelli, T., Campos-Laborie, F.J., Charloteaux, B., et al. (2020). A reference map of the human binary protein interactome. Nature *580*, 402–408.

Menche, J., Sharma, A., Kitsak, M., Ghiassian, S.D., Vidal, M., Loscalzo, J., and Barabási, A.-L. (2015). Disease networks. Uncovering disease-disease relationships through the incomplete interactome. Science *347*, 1257601.

Moreno-Mateos, M.A., Vejnar, C.E., Beaudoin, J.-D., Fernandez, J.P., Mis, E.K., Khokha, M.K., and Giraldez, A.J. (2015). CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. Nat. Methods *12*, 982–988.





Morton, P.D., Ishibashi, N., and Jonas, R.A. (2017). Neurodevelopmental abnormalities and congenital heart disease: insights Into altered brain maturation. Circ. Res. *120*, 960–977.

Navlakha, S., and Kingsford, C. (2010). The power of protein interaction networks for associating genes with diseases. Bioinformatics *26*, 1057–1063. https://doi.org/10.1093/bioinformatics/btq076.

Nieuwkoop, P.D., and Faber, J. (1994). In Normal table of *Xenopus laevis* (Daudin): A Systematical and Chronological Survey of the Development from the Fertilized Egg Till the End of Metamorphosis, P.D. Nieuwkoop and J. Faber, eds. (Garland Publishing).

O'Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P., Levy, R., Ko, A., Lee, C., Smith, J.D., et al. (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature *485*, 246–250.

Owens, N.D.L., Blitz, I.L., Lane, M.A., Patrushev, I., Overton, J.D., Gilchrist, M.J., Cho, K.W.Y., and Khokha, M.K. (2016). Measuring absolute RNA copy numbers at high temporal resolution reveals transcriptome kinetics in development. Cell Rep. *14*, 632–647.

Parikshak, N.N., Luo, R., Zhang, A., Won, H., Lowe, J.K., Chandran, V., Horvath, S., and Geschwind, D.H. (2013). Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell *155*, 1008–1021.

Paull, E.O., Carlin, D.E., Niepel, M., Sorger, P.K., Haussler, D., and Stuart, J.M. (2013). Discovering causal pathways linking genomic events to transcriptional states using tied diffusion Through interacting events (TieDIE). Bioinformatics 29, 2757–2764.

Piñero, J., Bravo, À., Queralt-Rosinach, N., Gutiérrez-Sacristán, A., Deu-Pons, J., Centeno, E., García-García, J., Sanz, F., and Furlong, L.I. (2017). DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Res. *45*, D833–D839.

Pratt, D., Chen, J., Welker, D., Rivas, R., Pillich, R., Rynkov, V., Ono, K., Miello, C., Hicks, L., Szalma, S., et al. (2015). NDEx, the network data exchange. Cell Syst. *1*, 302–305.

Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H., and Vilo, J. (2016). g:Profiler-a web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res. 44, W83–W89.

Resnik, P. (1995). Using information content to evaluate semantic similarity in a taxonomy. arXiv http://arxiv.org/abs/cmp-lg/9511007.

Sanders, S.J., Campbell, A.J., Cottrell, J.R., Moller, R.S., Wagner, F.F., Auldridge, A.L., Bernier, R.A., Catterall, W.A., Chung, W.K., Empfield, J.R., et al. (2018). Progress in understanding and treating SCN2A-mediated disorders. Trends Neurosci. *41*, 442–456.

Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L., et al. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature *485*, 237–241.

Sanders, S.J., He, X., Willsey, A.J., Ercan-Sencicek, A.G., Samocha, K.E., Cicek, A.E., Murtha, M.T., Bal, V.H., Bishop, S.L., Dong, S., et al. (2015). Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. Neuron *87*, 1215–1233.

Sater, A.K., and Moody, S.A. (2017). Using Xenopus to understand human disease and developmental disorders. Genesis 55, e22997. https://doi.org/10. 1002/dvg.22997.

Satterstrom, F.K., Kosmicki, J.A., Wang, Jiebiao, Breen, M.S., De Rubeis, S., An, J.-Y., Peng, M., Collins, R., Grove, J., Klei, L., et al. (2020). Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. Cell *180*, 568–584.e23.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nature Methods 9, 671–675.

Schram, M.T., Sep, S.J.S., van der Kallen, C.J., Dagnelie, P.C., Koster, A., Schaper, N., Henry, R.M.A., and Stehouwer, CoenD.A. (2014). The Maastricht study: an extensive phenotyping study on determinants of type 2 diabetes, its complications and its comorbidities. Eur. J. Epidemiol. *29*, 439–451.

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J., et al. (2007). Strong association of de novo copy number mutations with autism. Science *316*, 445–449.

Session, A.M., Uno, Y., Kwon, T., Chapman, J.A., Toyoda, A., Takahashi, S., Fukui, A., Hikosaka, A., Suzuki, A., Kondo, M., et al. (2016). Genome evolution in the allotetraploid frog Xenopus laevis. Nature *538*, 336–343.

Sestan, N., and State, M.W. (2018). Lost in translation: traversing the complex path from genomics to therapeutics in autism spectrum disorder. Neuron *100*, 406–423.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Shohat, S., Ben-David, E., and Shifman, S. (2017). Varying intolerance of gene pathways to mutational classes explain genetic convergence across neuro-psychiatric disorders. Cell Rep. *18*, 2217–2227.

Spratt, P.W.E., Ben-Shalom, R., Keeshen, C.M., Burke, K.J., Jr., Clarkson, R.L., Sanders, S.J., and Bender, K.J. (2019). The autism-associated gene Scn2a contributes to dendritic excitability and synaptic function in the prefrontal cortex. Neuron *103*, 673–685.e5.

State, M.W. (2010). The genetics of child psychiatric disorders: focus on autism and Tourette syndrome. Neuron 68, 254–269.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Res. *43*, D447–D452.

Tamassia, R. (2013). Handbook of Graph Drawing and Visualization (CRC Press).

Tartaglia, M., Kalidas, K., Shaw, A., Song, X., Musat, D.L., van der Burgt, I., Brunner, H.G., Bertola, D.R., Crosby, A., Ion, A., et al. (2002). PTPN11 mutations in Noonan syndrome: molecular spectrum, genotype-phenotype correlation, and phenotypic heterogeneity. Am. J. Hum. Genet. *70*, 1555–1563.

The GTEx Consortium (2015). Human genomics. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science *348*, 648–660.

Vanunu, O., Magger, O., Ruppin, E., Shlomi, T., and Sharan, R. (2010). Associating genes and protein complexes with disease via network propagation. PLoS Comput. Biol. *6*, e1000641.

Walentek, P., Beyer, T., Thumberger, T., Schweickert, A., and Blum, M. (2012). ATP4a is required for Wnt-dependent Foxj1 expression and leftward flow in Xenopus left-right development. Cell Rep. *1*, 516–527.

Willsey, A.J., Morris, M.T., Wang, S., Willsey, H.R., Sun, Nawei, Teerikorpi, N., Baum, T.B., Cagney, G., Bender, K.J., Desai, T.A., et al. (2018a). The psychiatric cell map initiative: a convergent systems biological approach to illuminating key molecular pathways in neuropsychiatric disorders. Cell *174*, 505–520.

Willsey, A.J., Sanders, S.J., Li, M., Dong, S., Tebbenkamp, A.T., Muhle, R.A., Reilly, S.K., Lin, L., Fertuzinhos, S., Miller, J.A., et al. (2013). Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell *155*, 997–1007.

Willsey, H.R., Exner, C.R.T., Xu, Y., Everitt, A., Sun, N., Wang, B., Dea, J., Schmunk, G., Zaltsman, Y., Teerikorpi, N., et al. (2021). Parallel in vivo analysis of large-effect autism genes implicates cortical neurogenesis and estrogen in risk and resilience. Neuron *109*, 788–804.e8.

Willsey, H.R., Walentek, P., Exner, C.R.T., Xu, X., Lane, A.B., Harland, R.M., Heald, R., and Santama, N. (2018b). Katanin-like protein Katnal2 is required for ciliogenesis and brain development in Xenopus embryos. Dev. Biol. *442*, 276–287.

Willsey, H.R., Xu, Y., Everitt, A., Dea, J., Exner, C.R.T., Willsey, A.J., State, M.W., and Harland, R.M. (2020). Neurodevelopmental disorder risk gene DYRK1A is required for Ciliogenesis and brain size in Xenopus embryos. Development. https://doi.org/10.1242/dev.189290.

CellPress

Cell Systems Article

Xu, X., Wells, A.B., O'Brien, D.R., Nehorai, A., and Dougherty, J.D. (2014). Cell type-specific expression analysis to identify putative cellular mechanisms for neurogenetic disorders. J. Neurosci. *34*, 1420–1431.

Yandım, C., and Karakülah, G. (2019). Expression dynamics of repetitive DNA in early human embryonic development. BMC Genomics *20*, 439.

Young, A.I. (2019). Solving the missing heritability problem. PLoS Genet. 15, e1008222.

Yu, M.K., Ma, J., Ono, K., Zheng, F., Fong, S.H., Gary, A., Chen, J., Demchak, B., Pratt, D., and Ideker, T. (2019). DDOT: a Swiss army knife for investigating data-driven biological ontologies. Cell Syst. *8*, 267–273.e3.

Zaidi, S., and Brueckner, M. (2017). Genetics and genomics of congenital heart disease. Circ. Res. *120*, 923–940.

Zaidi, S., Choi, M., Wakimoto, H., Ma, L., Jiang, J., Overton, J.D., Romano-Adesman, A., Bjornson, R.D., Breitbart, R.E., Brown, K.K., et al. (2013). De novo mutations in histone-modifying genes in congenital heart disease. Nature *498*, 220–223.





STAR * METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-β-Tubulin	DSHB	E7
Experimental models: Organisms/Strains		
Xenopus tropicalis	National Xenopus Resource	Superman strain; RRID: SCR_013731
Xenopus tropicalis	Khokha Lab (Yale)	Superman strain
Xenopus tropicalis	Nasco, Fort Atkinson, WI	LM00822
Oligonucleotides		
sgRNAs; See Table S8	This paper	N/A
Genotyping primers; See Table S8	This paper	N/A
Deposited data		
ASD-CHD network intersection	This paper	Ndexbio.org, UUID: cedacca2-8f2c-11ea- aaef-0ac135e8bacf
ASD-CHD systems map	This paper	Ndexbio.org, UUID: 5109757e-a5d6-11ea- aaef-0ac135e8bacf
PCNet interactome	Huang et al. 2018	https://public.ndexbio.org/#/network/ f93f402c-86d4-11e7-a10d-0ac135e8bacf
CHD de novo damaging variants	Jin et al. 2017	Table S9
ASD de novo damaging variants	Homsy et al. 2015	Table S8
DECIPHER shared condition variants	https://decipher.sanger.ac.uk/	Version 11.1
Software and algorithms		
ImageJ	Schneider et al., 2012	https://imagej.nih.gov/ij/
Python 2.7		https://www.python.org/download/ releases/2.7/
Clixo 1.0	Kramer et al., 2014	https://github.com/fanzheng10/CliXO-1.0
Networkx 1.11		https://networkx.github.io/
Combined network propagation	This paper	https://doi.org/10.5281/zenodo.5048355 https://github.com/ucsd-ccbb/ ASD_combined_network_analysis
Gprofiler	Reimand et al., 2016	https://biit.cs.ut.ee/gprofiler/
Cytoscape 3.8.0	Shannon et al., 2003	https://cytoscape.org
CSEA	Xu et al., 2014	http://genetics.wustl.edu/jdlab/csea- tool-2/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to the Lead Contact, Trey Ideker (tideker@ucsd.edu)

Materials availability

This study did not generate new materials.

Data and code availability

• This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table.

CellPress



- All original code has been deposited at https://github.com/ucsd-ccbb/ASD_combined_network_analysis, and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animal models

Xenopus tropicalis adult breeding animals originated in the Khokha lab (Yale University, wildtype *Superman* strain), in the National *Xenopus* Resource (RRID:SCR_013731, wildtype *Superman* strain) or from Nasco (Fort Atkinson, WI, wildtype). Animals were maintained in a recirculating system and used in accordance with approved UCSF IACUC protocols. Embryo stages 1-46 (Nieuwkoop and Faber, 1994) were used, as indicated in the manuscript. Males and females were used and clutch mates were always used as controls.

METHOD DETAILS

Data acquisition

The PCNet molecular interaction network was downloaded from NDex on November 13, 2018 (https://public.ndexbio.org/#/network/ f93f402c-86d4-11e7-a10d-0ac135e8bacf). Putatively damaging de novo variants were downloaded from the supplemental materials of previous studies (Homsy et al., 2015; Jin et al., 2017); the 750 ASD de novo damaging variants came from the Simons Simplex Collection (SSC), the 827 CHD de novo damaging variants came from the Pediatric Cardiac Genomics Consortium (PCGC) and the Pediatric Heart Network (PHN), and the 201 control de novo damaging variants came from SSC unaffected siblings. Putatively damaging de novo variants for CHD were defined in Jin et al. (2017) (Jin et al., 2017) as minor allele frequency $\leq 0.04\%$ across the cohort, and an annotation of nonsense, canonical splice site disruption, frameshift insertion-deletion, start loss, stop loss, or probably damaging (aka D-mis or Missense 3 (Dong et al., 2015)). Putatively damaging de novo variants for ASD probands and sibling controls were defined as in lossifov et al. (2014) (lossifov et al., 2014)-namely allele frequency $\leq 0.3\%$ in parents (to remove likely false positives), and an annotation of nonsense, canonical splice site disruption, frameshift insertion-deletion, start loss, stop loss, or D-mis. ASD and CHD established risk genes used as the seeds for our analyses were identified in previous studies (65 ASD risk genes (Sanders et al., 2015), 66 CHD risk genes (Jin et al., 2017)). ASD established risk genes were defined based on a false discovery rate for ASD association < 0.1, as per Sanders et al. (2015) (Sanders et al., 2015), and CHD established risk genes were defined as those with recurrent dDNVs in unrelated probands with CHD (Jin et al., 2017). For independent replication of variants, we accessed DECIPHER (DatabasE of genomiC variation and Phenotype in Humans using Ensembl Resources, https://decipher.sanger.ac.uk/, version 11.1), an interactive web database. ASD common variant summary statistics (Grove et al., 2019) were acquired from https://ipsych.dk/en/research/downloads/ on September 25, 2019. For analysis of adult human gene expression, data were downloaded from the Genotype-Tissue Expression (GTEx) Portal v8 (The GTEx Consortium, 2015). Human developmental data were downloaded from the supplemental materials of a previous study (Yandum and Karakülah, 2019). Xenopus developmental data were accessed at Xenbase (Karimi et al., 2018) (http://www.xenbase.org/, RRID:SCR_003280) as generated in a prior study (Session et al., 2016).

Xenopus mutagenesis and genotyping

Single guide RNAs (sgRNAs) were designed by CRISPRscan (Moreno-Mateos et al., 2015) and synthesized *in vitro* (Engen) and purified (Zymo). While off-target effects from using only one sgRNA per gene are possible, sgRNAs were designed according to the most stringent criteria within the algorithm CRISPRscan (Moreno-Mateos et al., 2015), with zero off-target sites predicted. 800 ng of sgRNA was injected with 1.5 ng of Cas9-NLS protein (MacroLabs, UC Berkeley) with a fluorescent dextran into one cell (brain phenotyping) or two cells (heart phenotyping) at the two-cell stage. A subset of the bilaterally injected animals were genotyped to ensure efficient mutagenesis by the sgRNA. Briefly, genomic DNA was extracted and PCR amplified using the sequence surrounding the PAM site. Sanger sequencing with sequence deconvolution by TIDE (Tracking of INDEIs) was used to determine mutation efficiency (Brinkman et al., 2014; DeLay et al., 2018). The full list of sgRNA sequences, genotyping primers and mutation efficiencies has been made available (Table S8).

Fluorescence staining, microscopy, and phenotyping

For brain imaging, animals were stained according to a previous method (Willsey et al., 2018) with a β -Tubulin primary antibody (DSHB, clone E7, 1:100) which labels all mature neurons and a fluorescence-conjugated secondary antibody (Life Technologies). For heart imaging, animals were processed similarly, with the omission of the bleaching step, and incubated with a fluorescent phalloidin (Invitrogen, 1:50) for 2 hours at room temperature instead of primary or secondary antibodies. Phalloidin labels actin, allowing for visualization of heart muscle tissue, rich in actin. Animals were imaged on a Zeiss AxioZoom V16 stereomicroscope with 1X objective, apotome, and a Zeiss AxioCam 512. Images were scored blinded to gene name and predicted phenotype. Phenotype severity was scored as "none", "mild," or "severe" per animal for every gene for both phenotypes. Telencephalon size was measured as a 2D area (um²) in Fiji and converted to a size ratio by dividing the injected side area by the uninjected side. Variance in telencephalon size





for each sample was determined in R. A Shapiro-Wilks test verified that no condition violated assumptions of normality, and qq plots for standard distribution were reviewed and did not appear skewed. A 95% confidence interval was derived from the F test function in R, as was variance ratio (var.test). Differences in telencephalon size variance compared to control CRISPR *slc45a2* were determined by a two-sided F-test, with p < 0.05 considered significant. Qualitative differences in heart *situs* and size were noted for each image and summed per condition (Figures S5B and S5C). sgRNA mutational efficiencies, sample sizes, and full phenotype quantification for every gene for both phenotypes is presented in Table S8 and Figure S5.

QUANTIFICATION AND STATISTICAL ANALYSIS

Measuring network localization of a gene set

We downloaded the Parsimonious Composite Network (PCNet, UUID: f93f402c-86d4-11e7-a10d-0ac135e8bacf), a resource of 2.7 million gene-gene associations (Huang et al., 2018). PCNet is formulated from integration of the STRING molecular interaction database (Szklarczyk et al., 2015) with a consensus of 21 additional molecular interaction databases. The interactions in PCNet cover a variety of interaction evidence types, including physical protein-protein, genetic (including synthetic lethality and epistasis), co-expression, and co-citation (where an interaction between genes is scored if they are mentioned frequently in the same papers, compared to how often they receive separate mentions in the literature) (Figures S6 and S7). Interactions were included in PCNet if they were supported by entries from two or more databases, the threshold identified by the authors to result in optimal performance in risk gene recovery. The majority of PCNet interactions were in fact supported by two or more types of evidence, with co-citation, co-expression, and protein-protein interactions making up the largest contributors. PCNet has been shown to enrich for genetic variants for a wide variety of diseases, to a higher degree than each database taken separately. We measured the localization of a disorder risk gene set in a gene network by calculating the number of PCNet interactions connecting pairs of genes in the set. A gene set which is localized in the network will have more interactions than a randomly selected gene set. We compared this localization measure, computed for the true disorder risk gene sets, to 1000 random gene sets of identical size and network degree distribution (Guney et al., 2016) (Figures S1 and S3).

Network propagation

Network propagation refers to a family of techniques (Cowen et al., 2017) which simulate how heat diffuses through a network by traversing its edges, starting from an initially hot set of 'seed' nodes. Here, we applied network propagation to the PCNet gene network (see above) with seed nodes defined as the risk genes for ASD (N=65) and CHD (N=66) (Figure 1A). At each iterative step of the algorithm, a unit of heat is added to each of the seeds and is spread to neighbors. A constant fraction of heat is then removed from all genes so that heat is conserved in the system. After a number of iterations, the heat on the genes converges to a stable value. This final heat vector is a proxy for the closeness of each gene in the network to the initial seed set. For example, if a gene interacts directly with two or more seed genes, it will tend to have a high final heat value, whereas if it is far from a seed gene, it will tend to have a low final heat value. This process is described according to the equation (Vanunu et al., 2010) Equation 1:

$$F^{t} = \alpha W' F^{t-1} + (1 - \alpha) Y$$
 (Equation 1)

where F^t is the heat vector at time t, Y is an indicator vector labeling the seed nodes, W' is the column-normalized adjacency matrix representation of the network under study, and $\alpha \in (0, 1)$ is the fraction of total heat retained at every time step. Following network propagation, gene scores are indexed against a null score distribution for each gene. This null distribution is constructed by randomly selecting a set of genes with size equal to the number of seed genes, and which match the degree of the seed gene set. We implemented a previously proposed approach (Guney et al., 2016) in which genes are binned such that each bin has at least 10 genes of similar degree. Network propagation is carried out 5,000 times with different random seed gene sets, resulting in a null distribution of random scores for each gene. A gene-level z-score is then calculated representing the number of standard deviations of the true score for that gene above (positive z) or below (negative z) the mean of the null distribution. This score defines the 'network proximity', \vec{z}_D for disorders D = ASD or CHD, to the seed genes Equation 2:

$$z_{n,D} = \left(log\left(F_{n,D}\right) - \left\langle log\left(\overrightarrow{F}_{n,rand}\right) \right\rangle \right) / \sigma\left(log\left(\overrightarrow{F}_{n,rand}\right) \right)$$
(Equation 2)

Where $F_{n,D}$ is the propagation score of gene *n*, seeded by established disorder seed genes; $\overline{F}_{n,rand}$ is the vector of propagation scores of gene *n* seeded by randomly selected degree-matched genes, () denotes an average of gene *n*'s propagation score over *N*randomly sampled sets, and σ denotes the standard deviation of the random distribution. Note the proximity vectors are log transformed so as to be approximately normally distributed. For example, a gene with *z* = 2 is two standard deviations more proximal to the established disorder risk genes than the average gene, and we may calculate an empirical *p* = 0.023 corresponding to the likelihood of observing this value by chance. Genes with Bonferroni corrected p-value < 0.1 defined the high confidence single disorder network risk genes.



Cell Systems Article

Joint network propagation to identify network intersection between two disorders

To prioritize genes related to two disorders, we implemented a joint network propagation procedure taking inspiration from prior work (Paull et al., 2013). Network propagation was carried out from ASD risk genes, and then CHD risk genes, separately. We then took the product of the two resulting proximity vectors in Equation 3,

$$Z_{n,ASD-CHD} = Z_{n,ASD} * Z_{n,CHD}$$
 (Equation 3)

where $z_{n,ASD-CHD}$ represents the vector of combined network proximity across all genes. Note the subscript n was dropped in the main text, for conciseness. We then selected genes for which $z_{n,ASD-CHD}$ was greater than a threshold ($z_{n,ASD-CHD} \ge 3$, Figure 2A) and $z_{n,ASD}$ and $z_{n,CHD}$ were non-negative. This operation selected 844 genes, which were interrelated by quantitative gene-gene similarities based on the cosine similarity of each gene pair in the PCNet as per a prior procedure (Yu et al., 2019), forming the "ASD-CHD gene network" reported in the main text (with edges thresholded at cosine similarity ≥ 0.95 , Figure 2B). This network is available for download from the Network Data Exchange (NDEx) (Pratt et al., 2015) as a community resource:https://ndexbio.org/viewer/ networks/cedacca2-8f2c-11ea-aaef-0ac135e8bacf

This gene set was further refined to a set of 398 high confidence genes, with both $z_{ASD} > 1.5$ and $z_{CHD} > 1.5$ and a joint corrected p value < 0.1, reflecting genes that are highly connected to both disorders. The z-score product identifies genes proximal to both ASD and CHD. We observed that the gene set overlap between ASD and CHD remained significant even when excluding the 24 CHD seed genes previously labeled as "syndromic", i.e. genes in which variants have been associated with phenotypes beyond CHD ($p = 5 \times 10^{-68}$; Table S2, identified as such in OMIM). A reciprocal analysis (i.e. non-syndromic ASD only) was not performed since most of the samples sequenced for the ASD gene discovery work were ascertained, in part, specifically based on not having a (known) syndromic phenotype (Fischbach and Lord, 2010). The combined network proximity score improves upon simpler methods to calculate the network overlap, such as computing the direct overlap among the neighbors of ASD and CHD seed genes. Since PCNet is a relatively dense interactome, with 2.7 million edges, there are a large number (7,974) of common neighbor genes between ASD and CHD seed genes. While these common neighbor genes are also significantly enriched for shared-condition DNV genes, the usefulness of such a large gene set, containing 40% of all genes in the interactome, is limited due to its low specificity. The combined network proximity scores allow us to prioritize a more specific set of neighbor genes highly relevant to both seed gene sets.

ASD and CHD specific gene networks were defined by genes which were proximal to one disorder, but not proximal to the other, and which were not in the ASD-CHD gene network. The ASD-specific gene network was defined by $z_{n,ASD} \ge 2$, $z_{n,CHD} < 2$, and $z_{n,ASD-CHD} < 3$, while the CHD-specific gene network was defined by $z_{n,ASD} \ge 2$, $z_{n,CHD} < 2$, and S2).

Network visualization

A spring-embedded layout algorithm (Prefuse force-directed layout) from Cytoscape was used to determine the positions of genes in Figures 2B, 5C, S6A–S6E, and S7A–S7C. This layout algorithm simulates the equilibrium positions of network objects when modeling nodes as balls of fixed mass and edges as springs of fixed spring constant (Tamassia, 2013). In Figures S6B–S6E and S7C, we display interaction edge type, where available. To identify edge type we query those databases in PCNet where interaction type is available (mainly STRING) since PCnet does not include interaction type information. Interaction types where available are shown on each edge in the PCnet subgraphs.

Enriched brain regions

Genes in the ASD-proximal subnetwork ($z_{n,ASD} \ge 2$), excluding seed genes, were tested for enrichment for specific brain regions and developmental time points using the CSEA tool (Xu et al., 2014); http://genetics.wustl.edu/jdlab/csea-tool-2/. Results are included for sensitivity threshold 0.05 (Table S3).

ASD common variant analysis

Summary statistics for common Single Nucleotide Polymorphisms (SNPs) were downloaded from a recent ASD genome wide association study (GWAS) (Grove et al., 2019) and were mapped to genes using PASCAL (Lamparter et al., 2016), a positional mapping tool which accounts for linkage disequilibrium. Genes were called as significant if they had a PASCAL determined $p < 2.510^{-4}$, resulting in 57 ASD common variant genes. These 57 genes were then subjected to the same network propagation algorithm used for the other ASD and CHD gene sets (see above).

Control disorder network overlap

A disease not expected to have significant genetic overlap with ASD or CHD – atherosclerosis – was selected from the DisGeNet database (Piñero et al., 2017). Using the atherosclerosis gene set from DisGeNet, along with the ASD common variant gene set (see above), we calculated the network z-scores for all genes (see above) as well as the combined proximity score vector $z_{n,D1-D2}$ for every disorder pair (6 pairs in total). Based on this vector we also computed the size of the network overlap at a threshold of $z_{n,D1-D2} \ge 3$ and compared this size to a null distribution of sizes. This null distribution was obtained by permuting *z*-score labels over genes for each disorder 1000 times, in each case taking the product of these shuffled *z*-score vectors. Significance was assessed by comparing the observed size of the network intersection to the null distribution, by computing an empirical p-value from the *z*-score.

Cell Systems Article



Replication of genetic variants in DECIPHER

Genes in the ASD-CHD network were cross referenced to the DECIPHER database using its web interface (https://decipher.sanger. ac.uk/), filtering by patients presenting with the dual phenotypes of 'Abnormality of the nervous system' and 'Abnormality of cardiovascular system'. After searching for the two traits of interest (abnormality of the nervous system and abnormality of the cardiovascular system), we clicked the 'browser' tab in the results to view the summary of all returned variants by chromosome. SNVs and indels are shown as histograms along the top of each chromosome, while copy number gains and losses are shown below. We manually recorded the number of SNVs and indels reported for each gene, the large majority of which are classified as predicted loss of function, or protein changing, by DECIPHER. Our analysis was restricted to variants which were likely LOF or protein changing (Table S6). We separately manually queried the DECIPHER database for small CNVs (<1KB), for each gene in the ASD-CHD network individually, to find CNVs that are likely gene specific. The number of small CNVs, along with SNVs/indels and dual-condition DNVs from (Jin et al., 2017), were included in the criteria for selection of genes to assay in Xenopus (Table S5). These small CNVs were not included in the ROC analysis since we were not able to access the full set of these small CNVs associated with the dual conditions of abnormal nervous system and abnormal cardiovascular system. We calculated Receiver Operating Characteristic (ROC) curves for recovering these genes when ranking human genes by the combined ASD-CHD network proximity score z_{ASD-CHD} (Figure 3D). Based on the ROC, we reported the Area Under ROC Curve (AUC) along with an empirical p-value on this area by permuting the gene labels 1000 times. Results did not greatly depend on whether we used genes with at least 2 variants (Figure 3D) or all genes with at least one variant (Figure S4).

Gene discovery

Disorder-specific subnetwork genes with an exome-wide Bonferroni-corrected p value < 0.1 (p values derived from *z* scores, 18,500 genes were used as the correction factor) were considered to be a risk gene (Tables S1 and S2). Network-implicated ASD risk genes were classified as not previously identified if they had a Bonferroni adjusted p < 0.1, and they had not been previously identified in exome studies at an FDR ≤ 0.1 (Satterstrom et al., 2020; Sanders et al., 2015; Doan et al., 2019), and were not present in the SFARI gene database (Abrahams et al., 2013). Network-implicated CHD risk genes were classified as not previously identified if they had a Bonferroni adjusted p < 0.1, had not been previously identified in exome studies (Jin et al., 2017) and were not known CHD risk genes in humans or mice (Jin et al., 2017). See Tables S1 and S2 for details on classification. Genes not previously identified were compared to established seed genes for differences in pLI, Mis_Z, and S_{het} scores (Lek et al., 2016) and tested for significance by Wilcoxon rank-sum test. In considering risk genes generated from the ASD-CHD network intersection, risk genes not previously identified were determined as above (with Bonferroni correction of the p value generated from the joint $z_{ASD-CHD}$ score, corrected by 18,500 genes x 2 disorders) but with the additional criteria of z_{ASD} and z_{CHD} both ≥ 1.5 and at least 1 dual-condition DECIPHER or PCGC/PHN variant observed (Table S4).

Construction of the ASD-CHD systems hierarchy

The ASD-CHD systems hierarchy was constructed by applying the CliXO community detection algorithm (Kramer et al., 2014) to the ASD-CHD network (defined above in the section 'joint network propagation to identify network intersection between two disorders'). Prior to applying the community detection algorithm, we transformed the PCNet edges in the ASD-CHD network into normalized cosine neighborhood similarity scores, following a prior procedure (Yu et al., 2019). These cosine similarity scores represent the extent to which two genes in a pair interact with similar sets of genes and, as such, they help visually reveal the underlying clustering structure present in a network (Figure 2B). The ASD-CHD network with original PCNet interactions is included for reference (Figure S7A). CliXO parameters α and β were swept over a range of values $\alpha \in (0.01, 0.02, 0.03, 0.04, 0.05), \beta \in (0.4, 0.45, 0.5)$ and the values ($\alpha = 0.01, \beta = 0.45$) were selected as those which gave the best localization of replicated variants in the hierarchy, as measured by the Resnik semantic similarity (Resnik, 1995) averaged over all pairs of replicated DECIPHER variants. Resnik semantic similarity is an information-based measure of two items in a taxonomy, based on the distance to the nearest common ancestors. The resulting hierarchy, consisting of 120 gene communities or "systems", was aligned to the Biological Process branch of the Gene Ontology (Harris et al., 2004) following a prior procedure (Yu et al., 2019), resulting in 29 significantly aligned systems. The remaining un-annotated systems were manually annotated with GO, KEGG and REACTOME pathways with gProfiler (Reimand et al., 2016), a python tool which uses a hypergeometric test to assess whether genes in the un-annotated systems had more shared genes with a GO, KEGG or REACTOME pathway than expected by chance. We filtered to pathways with less than 500 terms to improve specificity of annotation. Systems that remained un-annotated could not be confidently associated with a GO, KEGG or REACTOME entity and thus were considered new processes. Links represent parent → child containment relationships, with genes in child systems contained completely within parent systems. Parent systems may contain additional genes outside of those covered by child systems. These parent-child relationships were defined by the CliXO (Kramer et al., 2014) algorithm for detecting multiscale community structure in networks. This ASD-CHD systems hierarchy has been deposited in NDEx (Pratt et al., 2015) as a community resource: https:// ndexbio.org/viewer/networks/5109757e-a5d6-11ea-aaef-0ac135e8bacf

Classification of systems by prior literature support

Systems were classified by prior literature support (Figure 5B; Table S9). A system was classified as either 1) well characterized, if it aligned to GO following the procedure in (Yu et al., 2019) or had a precision and recall of \geq 40% with a known gene set from GO, KEGG or REACTOME in over-representation analysis; 2) a putative <u>new/extended subprocess</u>, if \geq 90% of the genes in the system

CellPress

Cell Systems Article

were contained by a known gene set but with precision \leq 40% (the system is largely contained in the known pathway but more focused); 3) a putative <u>new superprocess</u> if two or more known gene sets with distinct genes (none in common) were significantly over-represented in the system (the system unites known pathways in a new way); or 4) a putative <u>new process</u> if no known gene set was found to be significantly over-represented among system genes. Known gene sets from GO, KEGG and REACTOME were defined as significantly over-represented in a system if they had gene set sizes of \leq 500 (to filter less general biological functions); if they had at least 3 genes contained by the ASD-CHD system; and if they had an enrichment *p*-value < 0.05 (hypergeometric test, after Bonferroni correction for multiple testing).

Selection of candidate genes for functional validation in Xenopus

ASD-CHD molecular network prioritized genes were selected for validation that had $z_{n,ASD-CHD} \ge 3$, and that ranked in the top 10 genes with the highest number of dual-condition variants in the DECIPHER database. Variants were included which were classified as predicted loss of function, protein changing, or if they were a small CNV < 1KB. The z-score $z_{n,ASD-CHD}$ was used to break ties. Upon literature review we identified 4 of these 10 genes that already had strong evidence for association with both ASD and CHD (*CREBBP, NSD1, ANKRD11, and CDK13*) (Satterstrom et al., 2020; Jin et al., 2017; Abrahams et al., 2013). We decided to treat 2 of these genes (*NSD1* and *ANKRD11*) as positive controls. We also selected one gene (*SCN2A*), which had only 1 dual-phenotype DECIPHER variant, to represent the ion channel system, since genes in this system had many fewer variants. This resulted in a total of 7 genes not previously identified in ASD and/or CHD which were explored for validation in *Xenopus* experiments. See Table S7 for a summary of these genes.