UC San Diego
UC San Diego Previously Published Works

Title

A convergent molecular network underlying autism and congenital heart disease

Permalink

https://escholarship.org/uc/item/5r2683dd

Journal

Cell Systems, 12(11)

ISSN

2405-4712

Authors

Rosenthal, Sara Brin
Willsey, Helen Rankin
Xu, Yuxiao
et al.

Publication Date

2021-11-01

DOI

10.1016/j.cels.2021.07.009

Peer reviewed

A convergent molecular network underlying autism and congenital heart disease

Sara Brin Rosenthal[†,1], **Helen Rankin Willsey**[†,2], **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], **Trey Ideker**[*,3,7]

[1]Center for Computational Biology & Bioinformatics, Department of Medicine, University of California, San Diego, La Jolla, California, 92093, USA.

[2]Department of Psychiatry and Behavioral Sciences, Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA, 94158, USA.

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

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

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

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

[7]Lead contact

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

---

[*]Correspondence: jeremy.willsey@ucsf.edu, kfisch@ucsd.edu, tideker@ucsd.edu.

[†]These authors contributed equally

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 roadmap for identifying risk genes and pathways involved in co-morbid conditions.

eTOC blurb

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.

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, 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; A. J. Willsey et al. 2018; Gaynor et al. 2015; Morton, Ishibashi, and Jonas 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 neurodevelopmental disorders (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, Ishibashi, and Jonas 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; 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; Sanders et al. 2015; 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; Iossifov et al. 2014; Jin et al. 2017; A. J. 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 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; A. J. 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 multigene 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, Gulbahce, and Loscalzo 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, *SCN2A* (*Sodium Voltage-Gated Channel Alpha Subunit 2*), 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 (Fig. 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) (Fig. S1A–B, ASD $p = 4 \times 10^{-83}$, CHD $p = 3 \times 10^{-29}$, degree-matched permutation test, 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 ($p = 0.20$, Fig. 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) (Fig. 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 2, where z_D is the network proximity score for disorder D, 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) (Fig. 1B). Individual disorder subnetworks were then created, separately for ASD (1,583 genes; Table S1) and for 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, Fig. 1C), we sought to determine whether non-seed genes 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 (Fig. S1D–F; ASD $p = 1\times10^{-9}$, CHD $p = 2\times10^{-7}$, hypergeometric tests), but not dDNVs from unaffected siblings (ASD $p = 0.10$; CHD $p = 0.27$ Fig. S1G).

Given this enrichment, we used the ASD and CHD disorder-specific 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; Tables S1,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 $p = 1\times10^{-4}$), 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 network-implicated ASD risk genes (excluding seed genes) was significantly enriched for genes with early and mid-fetal expression signatures in amygdala, cerebellum, cortex, striatum, and thalamus (Table S3, Methods), consistent with previous findings (Satterstrom et al. 2020; A. J. Willsey et al. 2013; Parikshak et al. 2013; Shohat, Ben-David, and Shifman 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) ($p = 3\times10^{-8}$, 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, 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, $z_{ASD\text{-}CHD}$, by taking the product of the separate network propagation scores for each disorder ($z_{ASD} \times z_{CHD}$) (Figs. 2A,B, S2A, Methods). The number of genes at the intersection of the ASD and CHD networks was much larger than expected by chance, resulting in a distinct ASD-CHD network of 844 genes (empirical $p = 4 \times 10^{-141}$, Figs. 2B,C, 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 neurodevelopmental disorders (NDD) (Homsy et al. 2015) (Fig. 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 within the ASD-specific ($p = 0.75$) or CHD-specific subnetworks ($p = 0.21$, Methods, Fig. 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) ($p = 3 \times 10^{-70}$, rank-sum test, Fig. 3A) and in developing mouse tissue (Homsy et al. 2015) ($p = 3 \times 10^{-37}$, rank-sum test; Fig. 3B; embryonic day 14.5 in heart; embryonic day 9.5 in brain).

We also examined genes with network proximity to one disorder only (Methods). We found that genes proximal only to CHD risk genes were significantly enriched for dDNVs in CHD patients without symptoms of neurodevelopmental disorders ($p = 1 \times 10^{-4}$, Fig. S2B–C, 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 ($p = 0.04$, Fig. S2D–E, 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 significantly and substantially larger than expected by chance ($p = 7 \times 10^{-34}$ with > 2-fold effect, Fig. 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 if 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 (Fig. S3A–D), 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 (Fig. 3C, 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) (Methods, Tables S5, S6). For this validation cohort, we indeed observed an over-representation of likely damaging variants within the 844 genes (Fig. 3D, S4, empirical $p = 5 \times 10^{-22}$). This effect was maintained when excluding the established ASD and CHD risk genes (Fig. 3D, S4, empirical $p = 5 \times 10^{-14}$, 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 (Fig. 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 PCGC/PHN proband, and 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 dual-condition 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, KMT2D*) and 3 genes which were not seeds for either disorder (*KANSL1, KAT6A, 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* vertebrate model for uncovering fundamental mechanisms of human development and pathobiology (Kaltenbrun et al. 2011; Hwang, Marquez, and Khokha 2019; Sater and Moody 2017; Blum and Ott 2018), including ASD and CHD (H. R. Willsey et al. 2018; A. J. Willsey et al. 2018; Garfinkel and Khokha 2017; Exner and Willsey 2021; H. R. Willsey et al. 2021). We tested whether disruption of each gene was able to phenocopy known abnormalities in brain and/or heart development for ASD and CHD, respectively (H. R. 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 sgRNA targeting the gene of interest at the two-cell stage (Fig. 4A).

Animals were phenotyped at tadpole stages by fluorescence microscopy, with researchers blinded to gene name (Fig. 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 (H. R. Willsey et al. 2018, 2020, 2021).

Targeting a negative control pigmentation gene did not cause any brain or heart anatomical phenotypes (*slc45a2*, Fig. 4B,C), whereas positive control genes known to have roles in both ASD and CHD (*nsd1, 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 (Fig. 4B,C, Fig. S5). Notably, for all 7 genes undergoing validation we observed known brain and heart phenotypes previously associated with ASD or CHD in *Xenopus* at greater frequency than in controls, albeit with different degrees of penetrance and severity by gene (Fig. 4B,C, Fig. S5A–C, Table S8). The most common phenotypes among the genes were an increase in telencephalon size variance (Fig. S5A), changes in the heart size (Fig. S5B), and reversed heart *situs* (Fig. 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 (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 upwards in the hierarchy (Fig. 5A). Systems corresponding to known biological functions were labeled by alignment to multiple gene knowledge bases (Methods), revealing major functional branches related to chromatin and histone modification, ion channels and transport, Notch signaling, and MAP kinase signaling (Fig. 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 (Fig. 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 dual-condition dDNV in DECIPHER) was observed in select chromatin modification subsystems (e.g. System 886, 9/17 genes replicated, Fig. 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, Fig. S6B) and in system 901 for which 7/22 genes replicated in DECIPHER (Fig. S6C, Tables S1,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 a common functional role within the scope of chromatin modification (Fig. S6D–E).

SCN2A and other ion channels underlie ASD-CHD comorbidity.

A prominent collection of systems in the ASD-CHD systems map was related to ion channels and transport (Fig. 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), a 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* (Fig. 5C).

Of particular note was *SCN2A* (*Sodium Voltage-Gated Channel Alpha Subunit 2*), which had been previously reported to be expressed exclusively in the adult prefrontal cortex where it promotes 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 heart looping defects (Figs. 4B, 6A–F, Fig. S5). While *SCN2A* is indeed highly expressed in adult brains but not hearts of humans (Fig. 6G) (The GTEx Consortium 2015) or *Xenopus* (Fig. 6H) (Session et al. 2016), it is also expressed at early embryonic stages during germ layer specification, left-right patterning, and/or heart morphogenesis and neural differentiation in *Xenopus* and human (Fig. 6I–J) (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, embryonic day 14.5; 23rd percentile in brain, embryonic day 9.5) (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 on larger-scale biological structures and functions (Barabási, Gulbahce, and Loscalzo 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 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 (A. J. 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 pLI, Mis_Z, and S_{Het}). In contrast, the network-implicated risk genes tend to have lower pLI 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 (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 spread thinly over many different ion channel genes, however, suggesting why this signal has previously been underappreciated. This result again demonstrates the power of a 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 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 neurodevelopmental disorders (NDD) in general; 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* (H. R. 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 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 (A. J. 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 (A. J. Willsey et al. 2018). Such associations will also be improved by other factors, such as more complete patient phenotyping and 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 application of a similar methodology to any group of genetic disorders for which there is suspected co-morbidity. 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, Mariotto, and Rowland 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 it may help predict off-target 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

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.

- 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 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 Iossifov _et al._ (2014) (Iossifov et al. 2014)--namely allele frequency 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 (Yandım 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 INDELs) 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 (H. R. 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 (um^2) 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 (Fig. S5B–C). sgRNA mutational efficiencies, sample sizes, and full phenotype quantification for every gene for both phenotypes is presented in Table S8 and Fig. 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) (Figs. S6, 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) (Figs. S1,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) (Fig. 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):

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

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:

$$z_{n,D} = (log(F_{n,D}) - \langle log(\vec{F}_{n,rand}) \rangle) / \sigma(log(\vec{F}_{n,rand}) \tag{2}$$

Where $F_{n,D}$ is the propagation score of gene n, seeded by established disorder seed genes; $\vec{F}_{n,rand}$ is the vector of propagation scores of gene n seeded by randomly selected degree-matched genes, $\langle \rangle$ 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.

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,

$$z_{n,ASD-CHD} = z_{n,ASD} * z_{n,CHD} \tag{3}$$

Where $z_{n,D-CHD}$ 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 $z_{n,D-CHD}$ was greater than a threshold ($z_{n,ASD-CHD}$ 3, Fig. 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, Fig. 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 ($p = 5\times10^{-68}$; 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 $z_{n,ASD}$ 2, $z_{n,CHD} < 2$, and $z_{n,ASD–CHD} < 3$, while the CHD-specific gene network was defined by $z_{n,CHD}$ 2, $z_{n,ASD} < 2$, and $z_{n,ASD–CHD} < 3$ (Tables S1, 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–E, S7A–C. 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–E, and Figure 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 ($z_{n,ASD}$ 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 $p < 2.5\times10^{-4}$, 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 $z_{n,D1-D2}$ 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.

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 $z_{ASD-CHD}$ (Fig. 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 (Fig. 3D) or all genes with at least one variant (Fig. 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,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 (Fig. 2B). The ASD-CHD network with original PCNet interactions is included for reference (Fig. 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 (Fig. 5B, Table S9). A system was classified as either 1) <u>well characterized</u>, if it aligned to GO following the procedure in (Yu et al. 2019) or had a

precision and recall of 40% with a known gene set from GO, KEGG or REACTOME in over-representation analysis; 2) a putative <u>new/extended subprocess</u>, if 90% of the genes in the system were contained by a known gene set but with precision 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 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 $z_{n,SD-CHD}$ 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 $z_{n,D-CHD}$ 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

References

Abrahams Brett S., Arking Dan E., Campbell Daniel B., Mefford Heather C., Morrow Eric M., Weiss Lauren A., Menashe Idan, Wadkins Tim, Banerjee-Basu Sharmila, and Packer Alan. (2013).

"SFARI Gene 2.0: A Community-Driven Knowledgebase for the Autism Spectrum Disorders (ASDs)." Molecular Autism. 10.1186/2040-2392-4-36.

Ackerman MJ (1998). "The Long QT Syndrome: Ion Channel Diseases of the Heart." Mayo Clinic Proceedings. Mayo Clinic 73 (3): 250–69.

Andonian Caroline, Beckmann Jürgen, Biber Sabina, Ewert Peter, Freilinger Sebastian, Kaemmerer Harald, Oberhoffer Renate, Pieper Lars, and Neidenbach Rhoia Clara. (2018). "Current Research Status on the Psychological Situation of Adults with Congenital Heart Disease." Cardiovascular Diagnosis and Therapy 8 (6): 799–804. [PubMed: 30740327]

Barabási Albert-László, Gulbahce Natali, and Loscalzo Joseph. (2011). "Network Medicine: A Network-Based Approach to Human Disease." Nature Reviews. Genetics 12 (1): 56–68.

Ben-David Eyal, and Shifman Sagiv. (2012). "Networks of Neuronal Genes Affected by Common and Rare Variants in Autism Spectrum Disorders." PLoS Genetics 8 (3): e1002556. [PubMed: 22412387]

Bluethmann Shirley M., Mariotto Angela B., and Rowland Julia H.. (2016). "Anticipating the 'Silver Tsunami': Prevalence Trajectories and Comorbidity Burden among Older Cancer Survivors in the United States." Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology 25 (7): 1029–36.

Blum Martin, and Ott Tim. (2018). "Xenopus: An Undervalued Model Organism to Study and Model Human Genetic Disease." Cells, Tissues, Organs 205 (5–6): 303–13. [PubMed: 30092565]

Brinkman Eva K., Chen Tao, Amendola Mario, and van Steensel Bas. (2014). "Easy Quantitative Assessment of Genome Editing by Sequence Trace Decomposition." Nucleic Acids Research 42 (22): e168. [PubMed: 25300484]

Bycroft Clare, Freeman Colin, Petkova Desislava, Band Gavin, Elliott Lloyd T., Sharp Kevin, Motyer Allan, et al. (2018). "The UK Biobank Resource with Deep Phenotyping and Genomic Data." Nature 562 (7726): 203–9. [PubMed: 30305743]

Califano Andrea, Butte Atul J., Friend Stephen, Ideker Trey, and Schadt Eric. (2012). "Leveraging Models of Cell Regulation and GWAS Data in Integrative Network-Based Association Studies." Nature Genetics 44 (8): 841–47. [PubMed: 22836096]

CDC. (2018). "Data and Statistics | Autism Spectrum Disorder (ASD) | NCBDDD | CDC." Centers for Disease Control and Prevention. 4 26, 2018. https://www.cdc.gov/ncbddd/autism/data.html.

Chang Jonathan, Gilman Sarah R., Chiang Andrew H., Sanders Stephan J., and Vitkup Dennis. (2015). "Genotype to Phenotype Relationships in Autism Spectrum Disorders." Nature Neuroscience 18 (2): 191–98. [PubMed: 25531569]

Civelek Mete, and Lusis Aldons J.. (2014). "Systems Genetics Approaches to Understand Complex Traits." Nature Reviews. Genetics 15 (1): 34–48.

Colbert Costa M., and Pan Enhui. (2002). "Ion Channel Properties Underlying Axonal Action Potential Initiation in Pyramidal Neurons." Nature Neuroscience 5 (6): 533–38. [PubMed: 11992119]

Cowen Lenore, Ideker Trey, Raphael Benjamin J., and Sharan Roded. (2017). "Network Propagation: A Universal Amplifier of Genetic Associations." Nature Reviews. Genetics 18 (9): 551–62.

DeLay Bridget D., Corkins Mark E., Hanania Hannah L., Salanga Matthew, Deng Jian Min, Sudou Norihiro, Taira Masanori, Horb Marko E., and Miller Rachel K.. (2018). "Tissue-Specific Gene Inactivation in Xenopus Laevis: Knockout of lhx1 in the Kidney with CRISPR/Cas9." Genetics 208 (2): 673–86. [PubMed: 29187504]

Delling M, Indzhykulian AA, Liu X, Li Y, Xie T, Corey DP, and Clapham DE. (2016). "Primary Cilia Are Not Calcium-Responsive Mechanosensors." Nature 531 (7596): 656–60. [PubMed: 27007841]

De Rubeis Silvia, He Xin, Goldberg Arthur P., Poultney Christopher S., Samocha Kaitlin, Cicek A. Erucment, Kou Yan, et al. (2014). "Synaptic, Transcriptional and Chromatin Genes Disrupted in Autism." Nature 515 (7526): 209–15. [PubMed: 25363760]

Doan Ryan N., Lim Elaine T., Rubeis Silvia De, Betancur Catalina, Cutler David J., Chiocchetti Andreas G., Overman Lynne M., et al. (2019). "Recessive Gene Disruptions in Autism Spectrum Disorder." Nature Genetics 51 (7): 1092–98. [PubMed: 31209396]

Dong Chengliang, Wei Peng, Jian Xueqiu, Gibbs Richard, Boerwinkle Eric, Wang Kai, and Liu Xiaoming. (2015). "Comparison and Integration of Deleteriousness Prediction Methods for Nonsynonymous SNVs in Whole Exome Sequencing Studies." Human Molecular Genetics 24 (8): 2125–37. [PubMed: 25552646]

Duncan Anna R., and Khokha Mustafa K.. (2016). "Xenopus as a Model Organism for Birth Defects-Congenital Heart Disease and Heterotaxy." Seminars in Cell & Developmental Biology 51 (March): 73–79. [PubMed: 26910255]

Exner Cameron R. T., and Willsey Helen Rankin. (2021). "Xenopus Leads the Way: Frogs as a Pioneering Model to Understand the Human Brain." Genesis 59 (1–2): e23405. [PubMed: 33369095]

Fakhro Khalid A., Choi Murim, Ware Stephanie M., Belmont John W., Towbin Jeffrey A., Lifton Richard P., Khokha Mustafa K., and Brueckner Martina. (2011). "Rare Copy Number Variations in Congenital Heart Disease Patients Identify Unique Genes in Left-Right Patterning." Proceedings of the National Academy of Sciences of the United States of America 108 (7): 2915–20. [PubMed: 21282601]

Firth Helen V., Richards Shola M., Bevan A. Paul, Clayton Stephen, Corpas Manuel, Rajan Diana, Van Vooren Steven, Moreau Yves, Pettett Roger M., and Carter Nigel P.. (2009). "DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources." American Journal of Human Genetics 84 (4): 524–33. [PubMed: 19344873]

Fischbach Gerald D., and Lord Catherine. (2010). "The Simons Simplex Collection: A Resource for Identification of Autism Genetic Risk Factors." Neuron 68 (2): 192–95. [PubMed: 20955926]

Flint Jonathan, and Ideker Trey. (2019). "The Great Hairball Gambit." PLoS Genetics 15 (11): e1008519. [PubMed: 31770365]

Garfinkel Alexandra Maccoll, and Khokha Mustafa K.. (2017). "An Interspecies Heart-to-Heart: Using Xenopus to Uncover the Genetic Basis of Congenital Heart Disease." Current Pathobiology Reports 5 (2): 187–96. [PubMed: 29082114]

Gaynor J. William, Stopp Christian, Wypij David, Andropoulos Dean B., Atallah Joseph, Atz Andrew M., Beca John, et al. (2015). "Neurodevelopmental Outcomes after Cardiac Surgery in Infancy." Pediatrics 135 (5): 816–25. [PubMed: 25917996]

Gelb Bruce D., and Chung Wendy K.. (2014). "Complex Genetics and the Etiology of Human Congenital Heart Disease." Cold Spring Harbor Perspectives in Medicine 4 (7): a013953. [PubMed: 24985128]

Glessner Joseph T., Bick Alexander G., Ito Kaoru, Homsy Jason, Rodriguez-Murillo Laura, Fromer Menachem, Mazaika Erica, 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." Circulation Research 115 (10): 884–96. [PubMed: 25205790]

Go Christopher D., Knight James D. R., Rajasekharan Archita, Rathod Bhavisha, Hesketh Geoffrey G., Abe Kento T., Youn Ji-Young, et al. (2019). "A Proximity Biotinylation Map of a Human Cell." bioRxiv. 10.1101/796391.

Grove Jakob, Ripke Stephan, Als Thomas D., Mattheisen Manuel, Walters Raymond K., Won Hyejung, Pallesen Jonatan, et al. (2019). "Identification of Common Genetic Risk Variants for Autism Spectrum Disorder." Nature Genetics 51 (3): 431–44. [PubMed: 30804558]

Guney Emre, Menche Jörg, Vidal Marc, and Barábasi Albert-László. (2016). "Network-Based in Silico Drug Efficacy Screening." Nature Communications 7 (February): 10331.

Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, et al. (2004). "The Gene Ontology (GO) Database and Informatics Resource." Nucleic Acids Research 32 (Database issue): D258–61. [PubMed: 14681407]

Homsy Jason, Zaidi Samir, Shen Yufeng, Ware James S., Samocha Kaitlin E., Karczewski Konrad J., DePalma Steven R, et al. (2015). "De Novo Mutations in Congenital Heart Disease with Neurodevelopmental and Other Congenital Anomalies." Science 350 (6265): 1262–66. [PubMed: 26785492]

Huang Justin K., Carlin Daniel E., Yu Michael Ku, Zhang Wei, Kreisberg Jason F., Tamayo Pablo, and Ideker Trey. (2018). "Systematic Evaluation of Molecular Networks for Discovery of Disease Genes." Cell Systems 6 (4): 484–95.e5. [PubMed: 29605183]

Huttlin Edward L., Ting Lily, Bruckner Raphael J., Gebreab Fana, Gygi Melanie P., Szpyt John, Tam Stanley, et al. (2015). "The BioPlex Network: A Systematic Exploration of the Human Interactome." Cell. 10.1016/j.cell.2015.06.043.

Hwang Woong Y., Marquez Jonathan, and Khokha Mustafa K.. (2019). "Xenopus: Driving the Discovery of Novel Genes in Patient Disease and Their Underlying Pathological Mechanisms Relevant for Organogenesis." Frontiers in Physiology 10 (July): 953. [PubMed: 31417417]

Iossifov Ivan, O'Roak Brian J., Sanders Stephan J., Ronemus Michael, Krumm Niklas, Levy Dan, Stessman Holly A., et al. (2014). "The Contribution of de Novo Coding Mutations to Autism Spectrum Disorder." Nature 515 (7526): 216–21. [PubMed: 25363768]

Izarzugaza Jose M. G., Ellesøe Sabrina G., Doganli Canan, Ehlers Natasja Spring, Dalgaard Marlene D., Audain Enrique, Dombrowsky Gregor, et al. (2020). "Systems Genetics Analysis Identifies Calcium-Signaling Defects as Novel Cause of Congenital Heart Disease." Genome Medicine 12 (1): 76. [PubMed: 32859249]

Jin Sheng Chih, Homsy Jason, Zaidi Samir, Lu Qiongshi, Morton Sarah, DePalma Steven R., Zeng Xue, et al. (2017). "Contribution of Rare Inherited and de Novo Variants in 2,871 Congenital Heart Disease Probands." Nature Genetics 49 (11): 1593–1601. [PubMed: 28991257]

Kaltenbrun Erin, Tandon Panna, Amin Nirav M., Waldron Lauren, Showell Chris, and Conlon Frank L.. (2011). "Xenopus: An Emerging Model for Studying Congenital Heart Disease." Birth Defects Research. Part A, Clinical and Molecular Teratology 91 (6): 495–510. [PubMed: 21538812]

Karimi Kamran, Fortriede Joshua D., Lotay Vaneet S., Burns Kevin A., Wang Dong Zhou, Fisher Malcom E., Pells Troy J., et al. (2018). "Xenbase: A Genomic, Epigenomic and Transcriptomic Model Organism Database." Nucleic Acids Research 46 (D1): D861–68. [PubMed: 29059324]

Kramer Michael, Dutkowski Janusz, Yu Michael, Bafna Vineet, and Ideker Trey. (2014). "Inferring Gene Ontologies from Pairwise Similarity Data." Bioinformatics 30 (12): i34–42. [PubMed: 24932003]

Lage Kasper, Greenway Steven C., Rosenfeld Jill A., Wakimoto Hiroko, Gorham Joshua M., Segrè Ayellet V., Roberts Amy E., et al. (2012). "Genetic and Environmental Risk Factors in Congenital Heart Disease Functionally Converge in Protein Networks Driving Heart Development." Proceedings of the National Academy of Sciences of the United States of America 109 (35): 14035–40. [PubMed: 22904188]

Lamparter David, Marbach Daniel, Rueedi Rico, Kutalik Zoltán, and Bergmann Sven. (2016). "Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics." PLoS Computational Biology 12 (1): e1004714. [PubMed: 26808494]

Lek Monkol, Karczewski Konrad J., Minikel Eric V., Samocha Kaitlin E., Banks Eric, Fennell Timothy, O'Donnell-Luria Anne H., et al. (2016). "Analysis of Protein-Coding Genetic Variation in 60,706 Humans." Nature 536 (7616): 285–91. [PubMed: 27535533]

Levin Michael, Thorlin Thorleif, Robinson Kenneth R., Nogi Taisaku, and Mercola Mark. (2002). "Asymmetries in H+/K+-ATPase and Cell Membrane Potentials Comprise a Very Early Step in Left-Right Patterning." Cell 111 (1): 77–89. [PubMed: 12372302]

Li Taibo, Wernersson Rasmus, Hansen Rasmus B., Horn Heiko, Mercer Johnathan, Slodkowicz Greg, Workman Christopher T., et al. (2017). "A Scored Human Protein-Protein Interaction Network to Catalyze Genomic Interpretation." Nature Methods 14 (1): 61–64. [PubMed: 27892958]

Liu Li, Lei Jing, Sanders Stephan J., Willsey Arthur Jeremy, Kou Yan, Cicek Abdullah Ercument, Klei Lambertus, et al. (2014). "DAWN: A Framework to Identify Autism Genes and Subnetworks Using Gene Expression and Genetics." Molecular Autism 5 (1): 22. [PubMed: 24602502]

Luck Katja, Kim Dae-Kyum, Lambourne Luke, Spirohn Kerstin, Begg Bridget E., Bian Wenting, Brignall Ruth, et al. (2020). "A Reference Map of the Human Binary Protein Interactome." Nature 580 (7803): 402–8. [PubMed: 32296183]

Menche Jörg, Sharma Amitabh, Kitsak Maksim, Ghiassian Susan Dina, Vidal Marc, Loscalzo Joseph, and Barabási Albert-László. (2015). "Disease Networks. Uncovering Disease-Disease Relationships through the Incomplete Interactome." Science 347 (6224): 1257601. [PubMed: 25700523]

Moreno-Mateos Miguel A., Vejnar Charles E., Beaudoin Jean-Denis, Fernandez Juan P., Mis Emily K., Khokha Mustafa K., and Giraldez Antonio J.. (2015). "CRISPRscan: Designing Highly Efficient

sgRNAs for CRISPR-Cas9 Targeting in Vivo." Nature Methods 12 (10): 982–88. [PubMed: 26322839]

Morton Paul D., Ishibashi Nobuyuki, and Jonas Richard A.. (2017). "Neurodevelopmental Abnormalities and Congenital Heart Disease: Insights Into Altered Brain Maturation." Circulation Research 120 (6): 960–77. [PubMed: 28302742]

Navlakha Saket, and Kingsford Carl. (2010). "The Power of Protein Interaction Networks for Associating Genes with Diseases." Bioinformatics. 10.1093/bioinformatics/btq076.

Nieuwkoop PD, and Faber J. (1994). Normal Table of Xenopus Laevis (Daudin): A Systematical and Chronological Survey of the Development from the Fertilized Egg till the End of Metamorphosis. Edited by Nieuwkoop PD and Faber J. Garland Publishing Inc.

O'Roak Brian J., Vives Laura, Girirajan Santhosh, Karakoc Emre, Krumm Niklas, Coe Bradley P., Levy Roie, et al. (2012). "Sporadic Autism Exomes Reveal a Highly Interconnected Protein Network of de Novo Mutations." Nature 485 (7397): 246–50. [PubMed: 22495309]

Owens Nick D. L., Blitz Ira L., Lane Maura A., Patrushev Ilya, Overton John D., Gilchrist Michael J., Cho Ken W. Y., and Khokha Mustafa K.. (2016). "Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development." Cell Reports 14 (3): 632–47. [PubMed: 26774488]

Parikshak Neelroop N., Luo Rui, Zhang Alice, Won Hyejung, Lowe Jennifer K., Chandran Vijayendran, Horvath Steve, and Geschwind Daniel H.. (2013). "Integrative Functional Genomic Analyses Implicate Specific Molecular Pathways and Circuits in Autism." Cell 155 (5): 1008–21. [PubMed: 24267887]

Paull Evan O., Carlin Daniel E., Niepel Mario, Sorger Peter K., Haussler David, and Stuart Joshua M.. (2013). "Discovering Causal Pathways Linking Genomic Events to Transcriptional States Using Tied Diffusion Through Interacting Events (TieDIE)." Bioinformatics 29 (21): 2757–64. [PubMed: 23986566]

Piñero Janet, Bravo Àlex, Queralt-Rosinach Núria, Gutiérrez-Sacristán Alba, Deu-Pons Jordi, Centeno Emilio, García-García Javier, Sanz Ferran, and Furlong Laura I.. (2017). "DisGeNET: A Comprehensive Platform Integrating Information on Human Disease-Associated Genes and Variants." Nucleic Acids Research 45 (D1): D833–39. [PubMed: 27924018]

Pratt Dexter, Chen Jing, Welker David, Rivas Ricardo, Pillich Rudolf, Rynkov Vladimir, Ono Keiichiro, et al. (2015). "NDEx, the Network Data Exchange." Cell Systems 1 (4): 302–5. [PubMed: 26594663]

Reimand Jüri, Arak Tambet, Adler Priit, Kolberg Liis, Reisberg Sulev, Peterson Hedi, and Vilo Jaak. (2016). "g:Profiler—a Web Server for Functional Interpretation of Gene Lists (2016 Update)." Nucleic Acids Research 44 (W1): W83–89. [PubMed: 27098042]

Resnik Philip. (1995). "Using Information Content to Evaluate Semantic Similarity in a Taxonomy." arXiv [cmp-Lg]. arXiv. http://arxiv.org/abs/cmp-lg/9511007.

Sanders Stephan J., Campbell Arthur J., Cottrell Jeffrey R., Moller Rikke S., Wagner Florence F., Auldridge Angie L., Bernier Raphael A., et al. (2018). "Progress in Understanding and Treating SCN2A-Mediated Disorders." Trends in Neurosciences 41 (7): 442–56. [PubMed: 29691040]

Sanders Stephan J., He Xin, A. Jeremy Willsey, Ercan-Sencicek A. Gulhan, Samocha Kaitlin E., Cicek A. Ercument, Murtha Michael T., et al. (2015). "Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci." Neuron 87 (6): 1215–33. [PubMed: 26402605]

Sanders Stephan J., Murtha Michael T., Gupta Abha R., Murdoch John D., Raubeson Melanie J., Willsey A. Jeremy, Ercan-Sencicek A. Gulhan, et al. (2012). "De Novo Mutations Revealed by Whole-Exome Sequencing Are Strongly Associated with Autism." Nature 485 (7397): 237–41. [PubMed: 22495306]

Sater Amy K., and Moody Sally A.. (2017). "Using Xenopus to Understand Human Disease and Developmental Disorders." Genesis 55 (1–2). 10.1002/dvg.22997.

Satterstrom F. Kyle, Kosmicki Jack A., Wang Jiebiao, Breen Michael S., Rubeis Silvia De, An Joon-Yong, Peng Minshi, et al. (2020). "Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism." Cell 180 (3): 568–84.e23. [PubMed: 31981491]

Schram Miranda T., Sep Simone J. S., van der Kallen Carla J., Dagnelie Pieter C., Koster Annemarie, Schaper Nicolaas, Henry Ronald M. A., and Stehouwer Coen D. A.. (2014). "The Maastricht Study: An Extensive Phenotyping Study on Determinants of Type 2 Diabetes, Its Complications and Its Comorbidities." European Journal of Epidemiology 29 (6): 439–51. [PubMed: 24756374]

Sebat Jonathan, Lakshmi B, Malhotra Dheeraj, Troge Jennifer, Lese-Martin Christa, Walsh Tom, Yamrom Boris, et al. (2007). "Strong Association of de Novo Copy Number Mutations with Autism." Science 316 (5823): 445–49. [PubMed: 17363630]

Session Adam M., Uno Yoshinobu, Kwon Taejoon, Chapman Jarrod A., Toyoda Atsushi, Takahashi Shuji, Fukui Akimasa, et al. (2016). "Genome Evolution in the Allotetraploid Frog Xenopus Laevis." Nature 538 (7625): 336–43. [PubMed: 27762356]

Sestan Nenad, and State Matthew W.. (2018). "Lost in Translation: Traversing the Complex Path from Genomics to Therapeutics in Autism Spectrum Disorder." Neuron 100 (2): 406–23. [PubMed: 30359605]

Shannon Paul, Markiel Andrew, Ozier Owen, Baliga Nitin S., Wang Jonathan T., Ramage Daniel, Amin Nada, Schwikowski Benno, and Ideker Trey. (2003). "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks." Genome Research 13 (11): 2498–2504. [PubMed: 14597658]

Shohat Shahar, Ben-David Eyal, and Shifman Sagiv. (2017). "Varying Intolerance of Gene Pathways to Mutational Classes Explain Genetic Convergence across Neuropsychiatric Disorders." Cell Reports 18 (9): 2217–27. [PubMed: 28249166]

Spratt Perry W. E., Ben-Shalom Roy, Keeshen Caroline M., Burke Kenneth J. Jr, Clarkson Rebecca L., Sanders Stephan J., and Bender Kevin J.. (2019). "The Autism-Associated Gene Scn2a Contributes to Dendritic Excitability and Synaptic Function in the Prefrontal Cortex." Neuron 103 (4): 673–85.e5. [PubMed: 31230762]

State Matthew W. (2010). "The Genetics of Child Psychiatric Disorders: Focus on Autism and Tourette Syndrome." Neuron 68 (2): 254–69. [PubMed: 20955933]

Szklarczyk Damian, Franceschini Andrea, Wyder Stefan, Forslund Kristoffer, Heller Davide, Huerta-Cepas Jaime, Simonovic Milan, et al. (2015). "STRING v10: Protein–protein Interaction Networks, Integrated over the Tree of Life." Nucleic Acids Research 43 (D1): D447–52. [PubMed: 25352553]

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

Tartaglia Marco, Kalidas Kamini, Shaw Adam, Song Xiaoling, Musat Dan L., van der Burgt Ineke, Brunner Han G., et al. (2002). "PTPN11 Mutations in Noonan Syndrome: Molecular Spectrum, Genotype-Phenotype Correlation, and Phenotypic Heterogeneity." American Journal of Human Genetics 70 (6): 1555–63. [PubMed: 11992261]

The GTEx Consortium. (2015). "The Genotype-Tissue Expression (GTEx) Pilot Analysis: Multitissue Gene Regulation in Humans." Science 348 (6235): 648–60. [PubMed: 25954001]

Torre-Ubieta Luis de la, Won Hyejung, Stein Jason L., and Geschwind Daniel H.. (2016). "Advancing the Understanding of Autism Disease Mechanisms through Genetics." Nature Medicine 22 (4): 345–61.

Vanunu Oron, Magger Oded, Ruppin Eytan, Shlomi Tomer, and Sharan Roded. (2010). "Associating Genes and Protein Complexes with Disease via Network Propagation." PLoS Computational Biology 6 (1): e1000641. [PubMed: 20090828]

Walentek Peter, Beyer Tina, Thumberger Thomas, Schweickert Axel, and Blum Martin. (2012). "ATP4a Is Required for Wnt-Dependent Foxj1 Expression and Leftward Flow in Xenopus Left-Right Development." Cell Reports 1 (5): 516–27. [PubMed: 22832275]

Willsey A. Jeremy, Morris Montana T., Wang Sheng, Willsey Helen R., Sun Nawei, Teerikorpi Nia, Baum Tierney B., et al. (2018). "The Psychiatric Cell Map Initiative: A Convergent Systems Biological Approach to Illuminating Key Molecular Pathways in Neuropsychiatric Disorders." Cell 174 (3): 505–20. [PubMed: 30053424]

Willsey A. Jeremy, Sanders Stephan J., Li Mingfeng, Dong Shan, Tebbenkamp Andrew T., Muhle Rebecca A., Reilly Steven K., et al. (2013). "Coexpression Networks Implicate Human Midfetal Deep Cortical Projection Neurons in the Pathogenesis of Autism." Cell 155 (5): 997–1007. [PubMed: 24267886]

Willsey Helen Rankin, Exner Cameron R. T., Xu Yuxiao, Everitt Amanda, Sun Nawei, Wang Belinda, Dea Jeanselle, et al. (2021). "Parallel in Vivo Analysis of Large-Effect Autism Genes Implicates Cortical Neurogenesis and Estrogen in Risk and Resilience." Neuron 109 (5): 788–804.e8. [PubMed: 33497602]

Willsey Helen Rankin, Walentek Peter, Exner Cameron R. T., Xu Yuxiao, Lane Andrew B., Harland Richard M., Heald Rebecca, and Santama Niovi. (2018). "Katanin-like Protein Katnal2 Is Required for Ciliogenesis and Brain Development in Xenopus Embryos." Developmental Biology 442 (2): 276–87. [PubMed: 30096282]

Willsey Helen Rankin, Xu Yuxiao, Everitt Amanda, Dea Jeanselle, Exner Cameron R. T., Willsey A. Jeremy, State Matthew W., and Harland Richard M.. (2020). "Neurodevelopmental Disorder Risk Gene DYRK1A Is Required for Ciliogenesis and Brain Size in Xenopus Embryos." Development, 5. 10.1242/dev.189290.

Xu Xiaoxiao, Wells Alan B., O'Brien David R., Nehorai Arye, and Dougherty Joseph D.. (2014). "Cell Type-Specific Expression Analysis to Identify Putative Cellular Mechanisms for Neurogenetic Disorders." The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 34 (4): 1420–31. [PubMed: 24453331]

Yandım Cihangir, and Gökhan Karakülah. (2019). "Expression Dynamics of Repetitive DNA in Early Human Embryonic Development." BMC Genomics 20 (1): 439. [PubMed: 31151386]

Young Alexander I. (2019). "Solving the Missing Heritability Problem." PLoS Genetics 15 (6): e1008222. [PubMed: 31233496]

Yu Michael Ku, Ma Jianzhu, Ono Keiichiro, Zheng Fan, Fong Samson H., Gary Aaron, Chen Jing, Demchak Barry, Pratt Dexter, and Ideker Trey. (2019). "DDOT: A Swiss Army Knife for Investigating Data-Driven Biological Ontologies." Cell Systems 8 (3): 267–73.e3. [PubMed: 30878356]

Zaidi Samir, and Brueckner Martina. (2017). "Genetics and Genomics of Congenital Heart Disease." Circulation Research 120 (6): 923–40. [PubMed: 28302740]

Zaidi Samir, Choi Murim, Wakimoto Hiroko, Ma Lijiang, Jiang Jianming, Overton John D., Angela Romano-Adesman, et al. (2013). "De Novo Mutations in Histone-Modifying Genes in Congenital Heart Disease." Nature 498 (7453): 220–23. [PubMed: 23665959]

Highlights:

- A convergent molecular network shared between autism and congenital heart disease

- 101 genes likely carry 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*
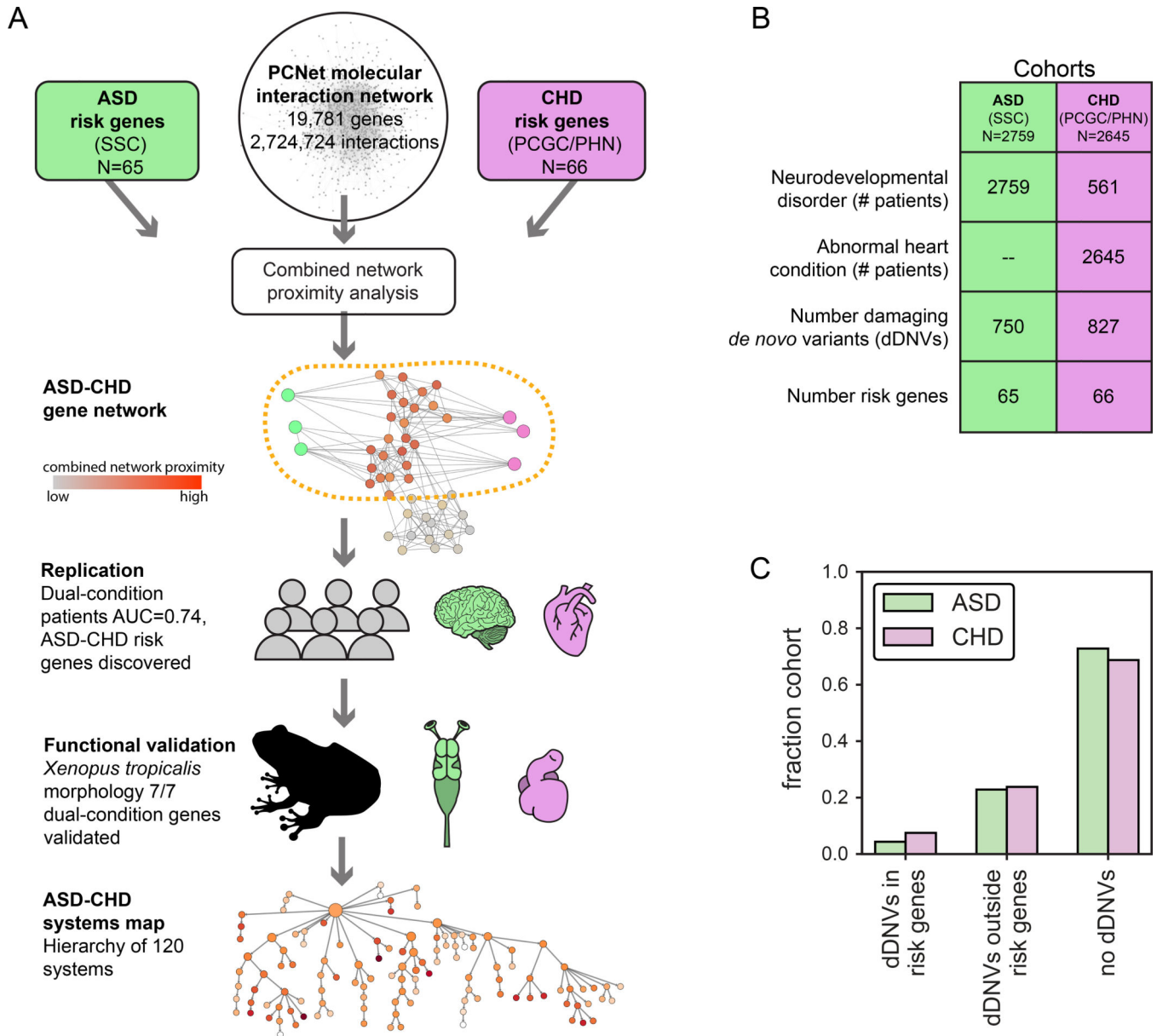
Fig. 1. Overview of genetic data and workflow.

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 (NDD) 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.
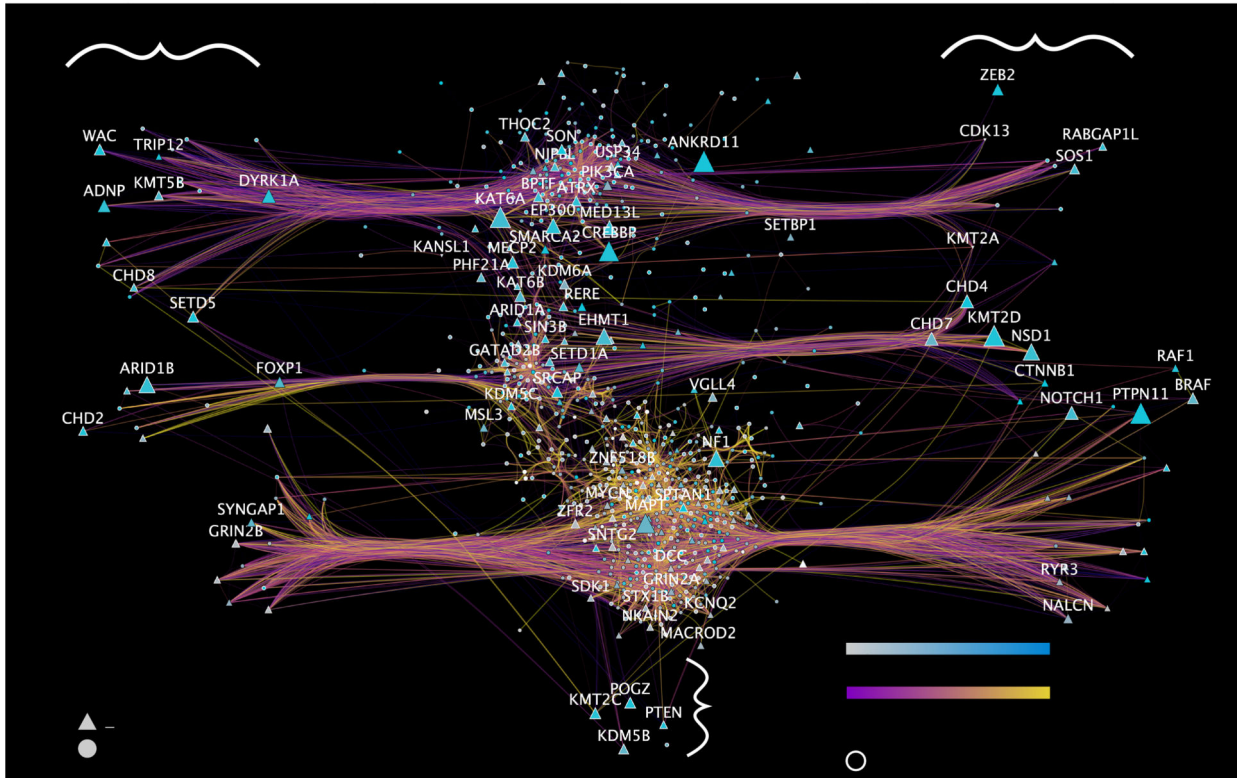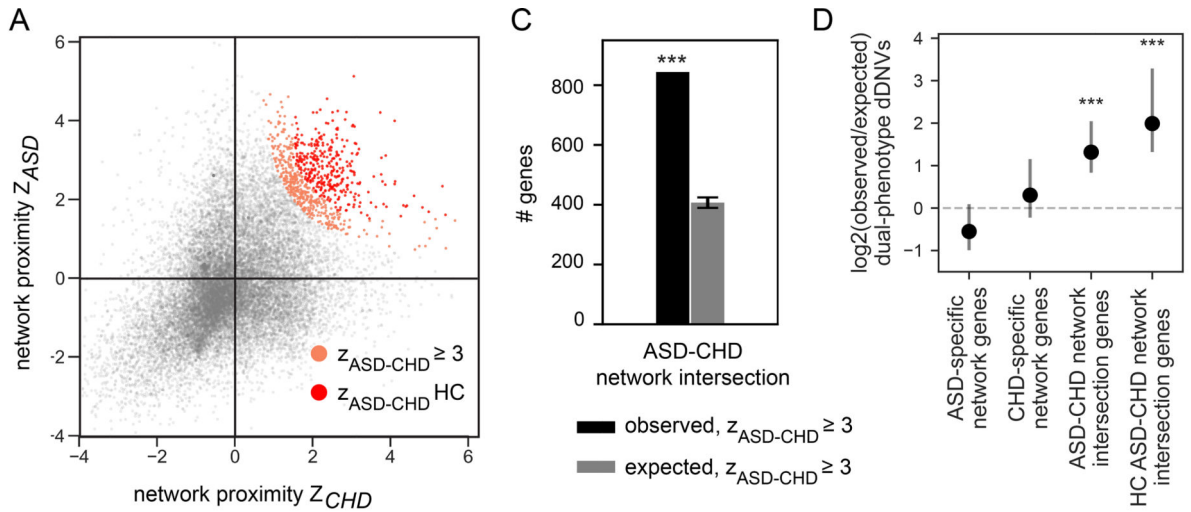
Fig. 2. ASD and CHD *de novo* risk variants converge on common network regions.
A) Scatterplot of human proteins based on network proximity to established ASD risk
genes or established CHD risk genes (y versus x). Proteins at the network intersection are
shown in orange, with high confidence (HC; adj. 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 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 to 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} 2, $z_{CHD} < 2$, $z_{ASD\text{-}CHD} < 3$); CHD-specific network genes (z_{CHD} 2, $z_{ASD} < 2$, $z_{ASD\text{-}CHD} < 3$); or genes at the ASD-CHD network intersection, and high confidence network intersection *** indicates $p < 0.001$, ** indicates $p < 0.01$, * indicates $p < 0.05$. Vertical bars indicate one standard deviation.
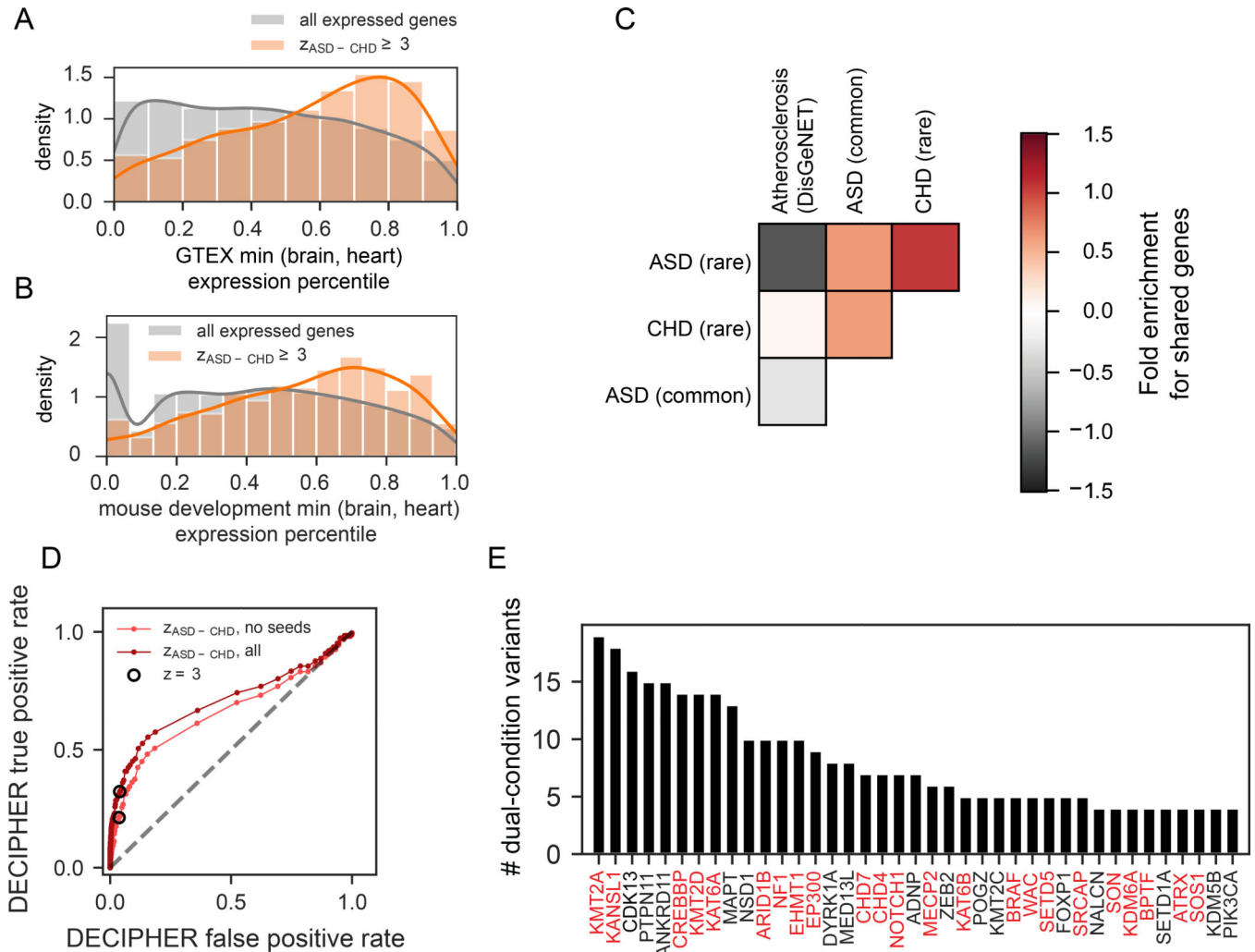
Fig. 3. Validation of the ASD-CHD network in independent patient cohorts.
A) Distribution of minimum percentile mRNA expression of ASD-CHD network genes in GTEx brain or heart tissues (orange), compared to all expressed genes (gray). **B)** Distribution of minimum percentile mRNA expression of ASD-CHD network genes in mouse developmental brain or heart tissues (orange), compared to 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. Colorbar 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 ($z_{ASD\text{-}CHD}$). Results for all ranked genes in maroon (AUC = 0.71, $p = 5 \times 10^{-22}$); results excluding established risk genes in salmon (AUC = 0.66, $p = 5 \times 10^{-14}$, empirical p-values). Black circles highlight the sets of genes at the $z_{ASD\text{-}CHD}$ 3 cutoff 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.
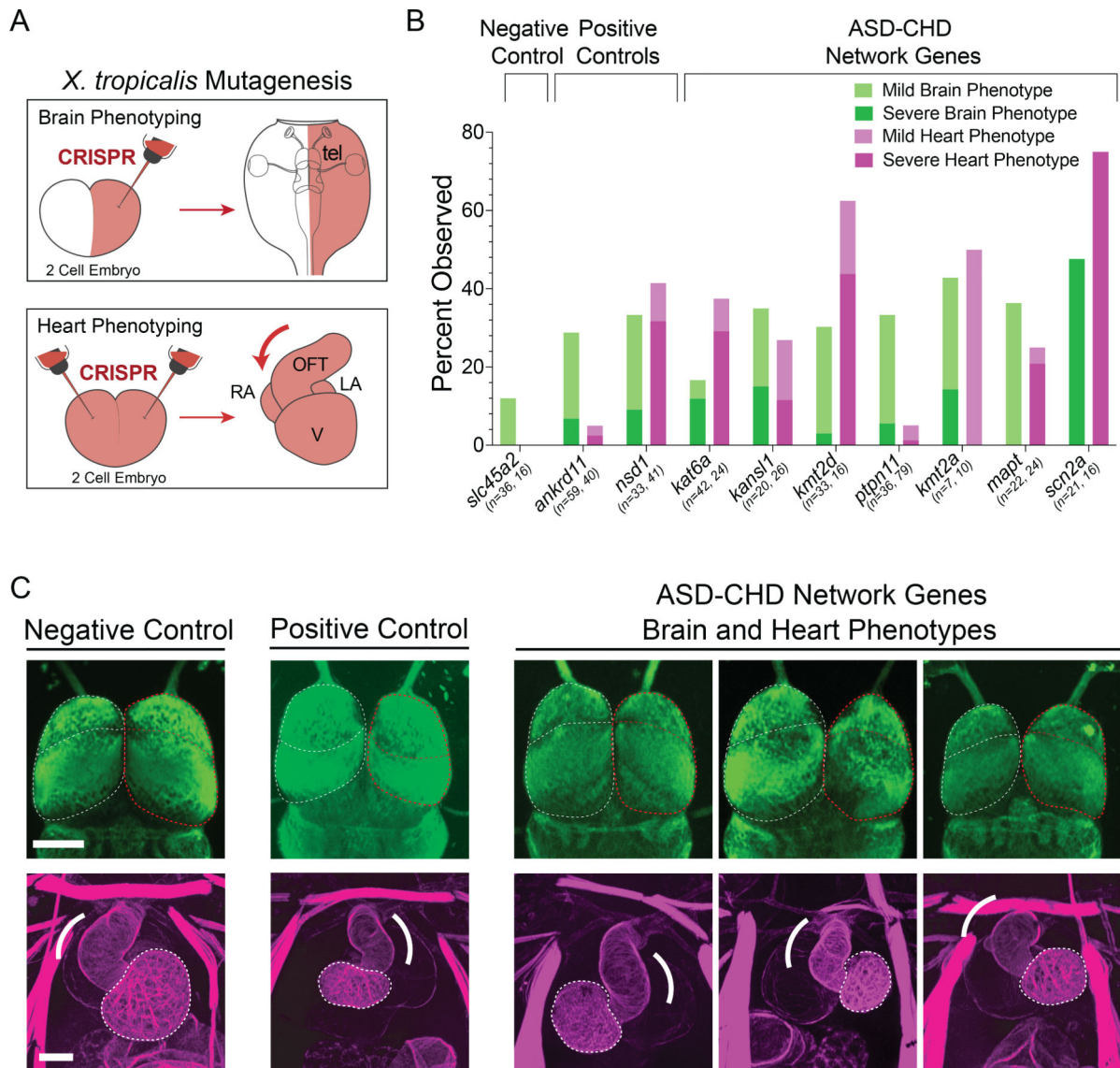
Fig. 4. Functional validation in *Xenopus tropicalis*.
A) CRISPR mutagenesis strategy. CRISPR reagents and a tracer dye (red) are injected
at the two-cell stage, either into both cells (bilateral mutants, heart phenotyping) or into
one cell (unilateral mutants, brain phenotyping). Animals are grown to tadpole stages
and phenotyped for brain (top) or heart (bottom) anatomy. Brain is normally bilaterally
symmetric. Note leftward heart looping direction. Telencephalon (tel), Outflow tract (OFT),
Ventricle (V), Right Atrium (RA), 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 change in brain size on the injected half (red dotted outline) compared to 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 to control. Size differences in the total heart and outflow tract size can be readily seen in *nsd1*, *ptpn11* and *kansl1* heart images. Sample sizes and quantification of telencephalon size variance and heart phenotypes are in Figure S5.
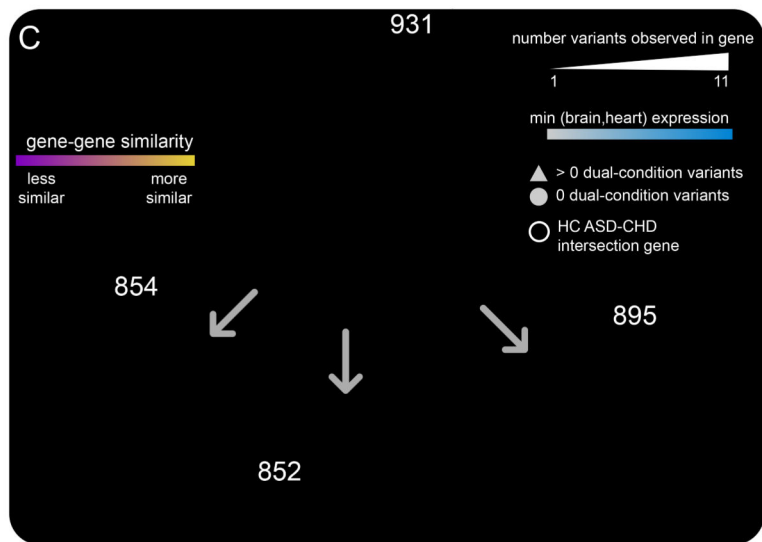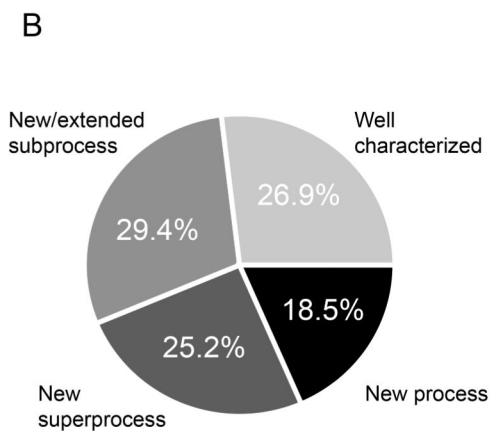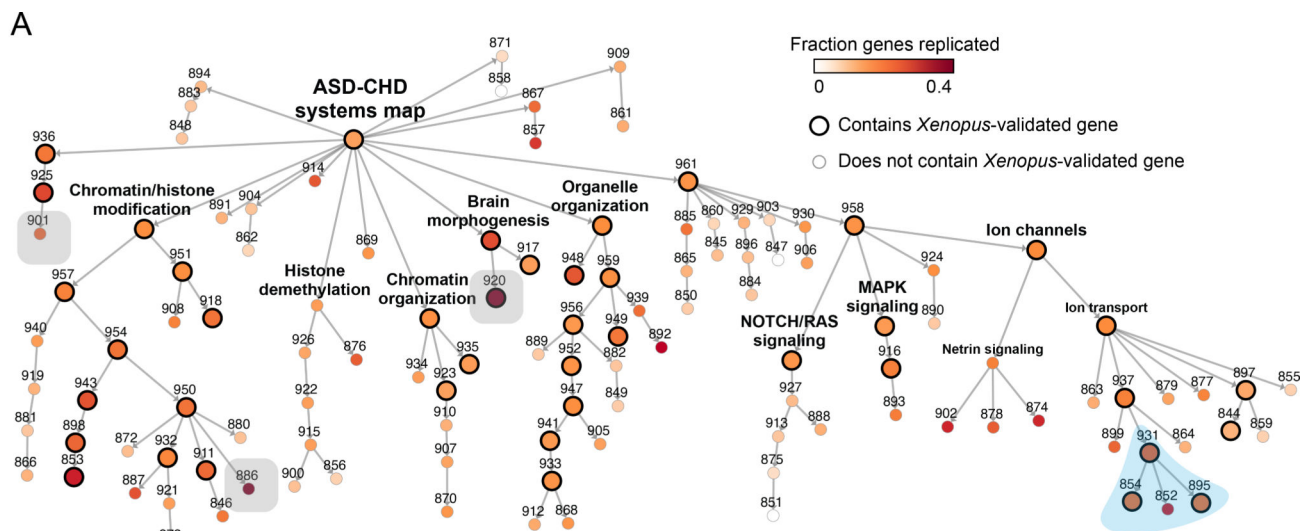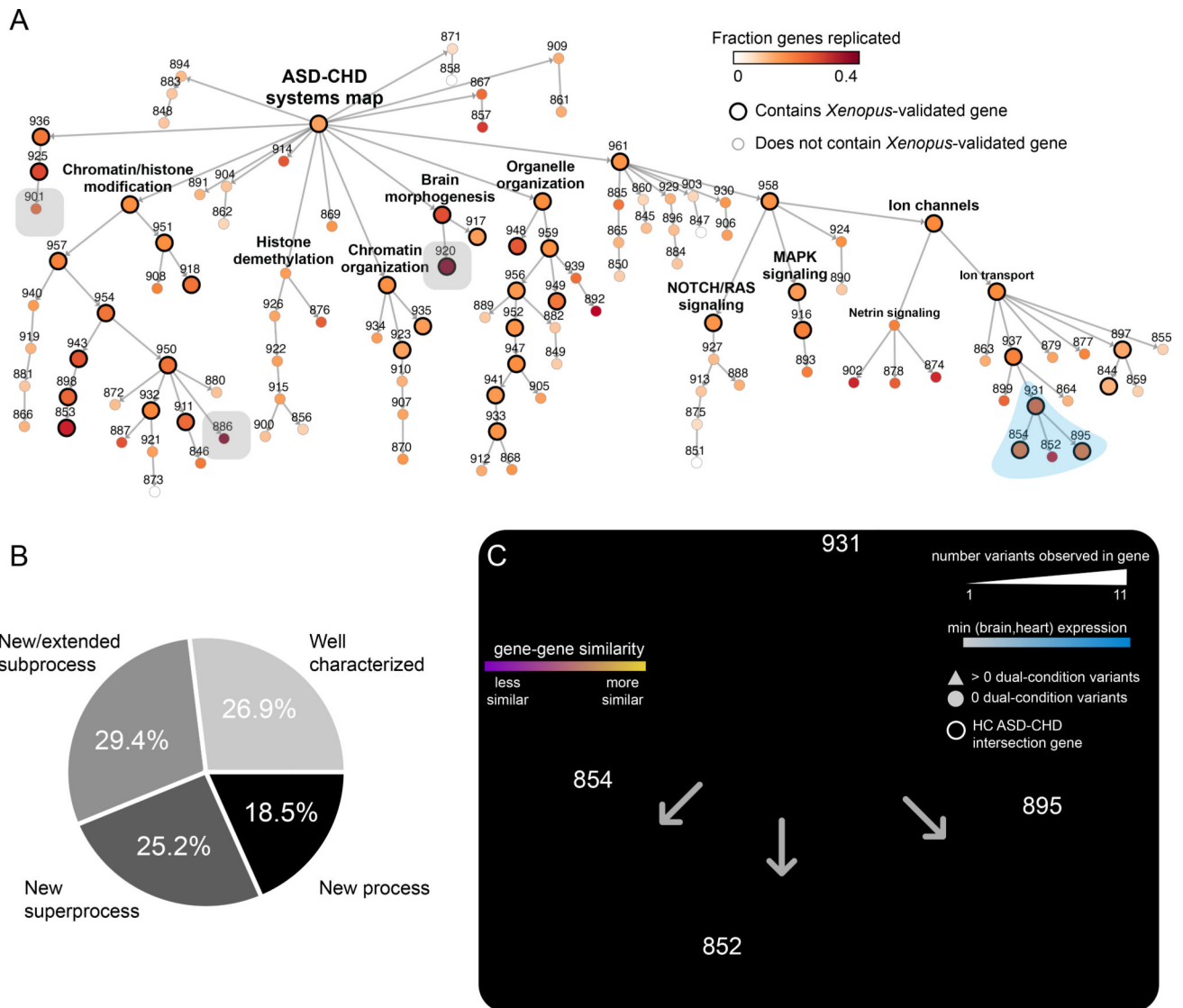
Fig. 5. Hierarchical organization of the ASD-CHD network.
A) Representation of the network data as a hierarchy of densely interacting gene systems. Nodes represent distinct systems, with size showing the number of genes (min 6, max 844) and color showing the fraction of these genes in which dual-condition 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 ID (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 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.
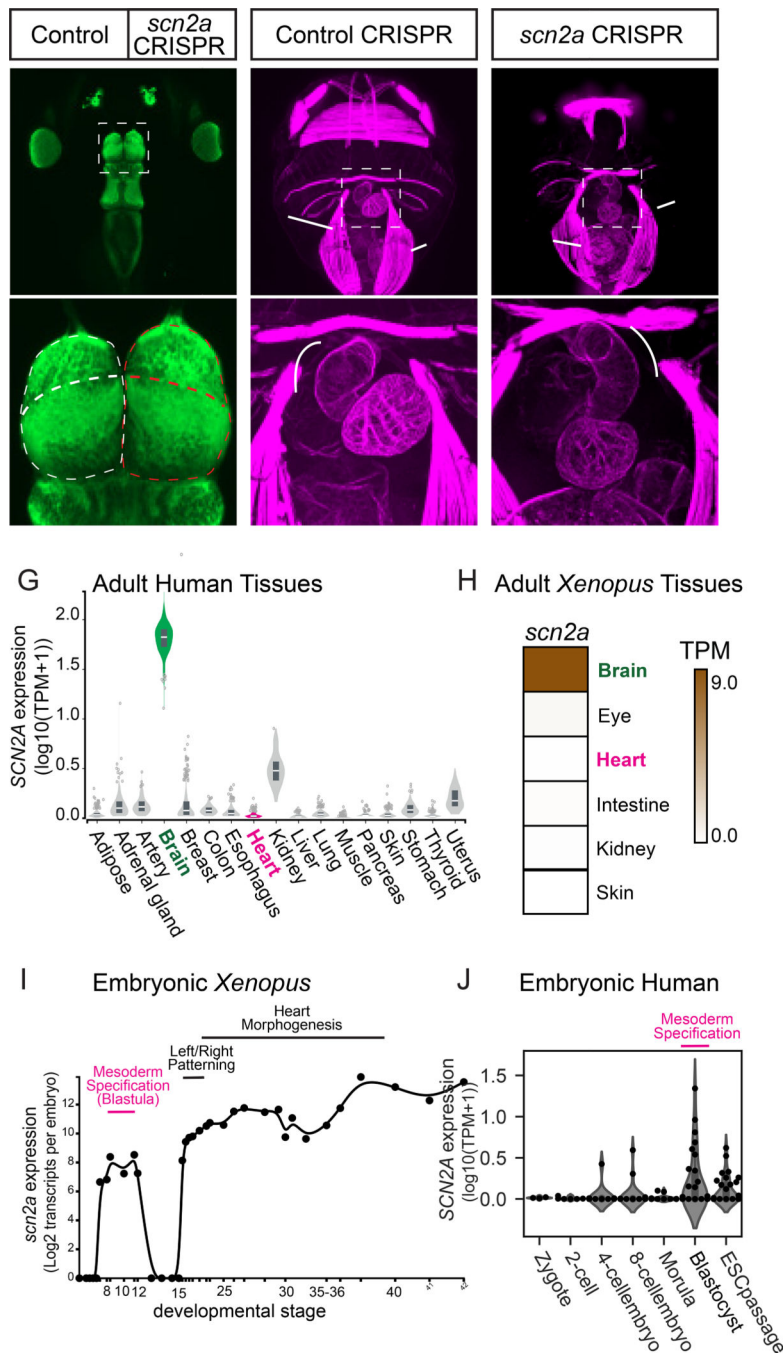
Fig. 6. *SCN2A* is required for heart and brain development.
A) Unilateral mutagenesis of *scn2a* leads to increase in telencephalon size on the mutated (right) side. β-Tubulin staining. **B)** High magnification of boxed region in A. Note increase in size of olfactory bulb (ob) and rest of telencephalon (tel). **C)** Control tadpole with proper organ situs (gallbladder, GB, on 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 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 brain but not heart. **H)** Adult *Xenopus laevis* expression of *scn2a* by tissue. Note strong expression in brain but not heart. **I)** Embryonic expression of *scn2a* in *Xenopus tropicalis*. Numbers on x-axis represent 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 blastocyst stage when mesoderm specification occurs (pink), similar to *Xenopus*.

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 (RRID: SCR_013731)	*Superman* strain
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 S08
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	DOI 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/