A network landscape of autism spectrum and other developmental disorders reveals shared molecular mechanisms

Sara Brin Rosenthal, Trey Ideker, Kathleen Fisch

Center for Computational Biology and Bioinformatics, University of California San Diego

Introduction

- Autism spectrum disorder (ASD) has a high comorbidity with other developmental and neuronal disorders, including congenital heart disorder (CHD) and epilepsy (EPI).
- Molecular and genetic basis of these disorders and the shared underlying mechanisms causing observed co-morbidity not fully understood.
- Integrative analysis to demonstrate that ASD, CHD, and EPI are disorders of networks, not genes, and build a deeper understanding of the molecular mechanisms involved by examining the network at the interface of one pair- ASD and CHD.

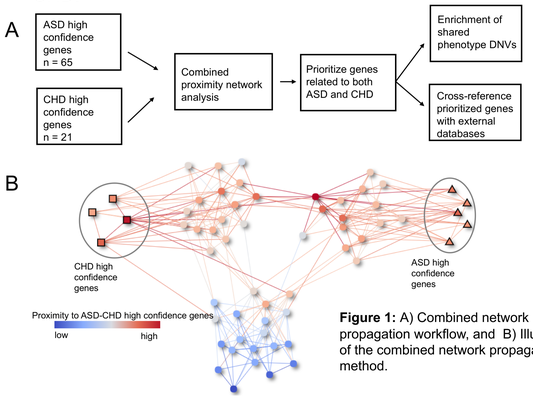

Figure 1: A) Combined network propagation workflow, and B) Illustration of the combined network propagation method.

Summary of phenotypes, DNVs, and cohorts

- The comorbidity between pairs of disorders is apparent by examining the overlap in phenotypes among the datasets
- Only a small fraction of probands in all three cohorts have **de novo variants (DNVs)** in any **high confidence disease-genes** (literature defined).

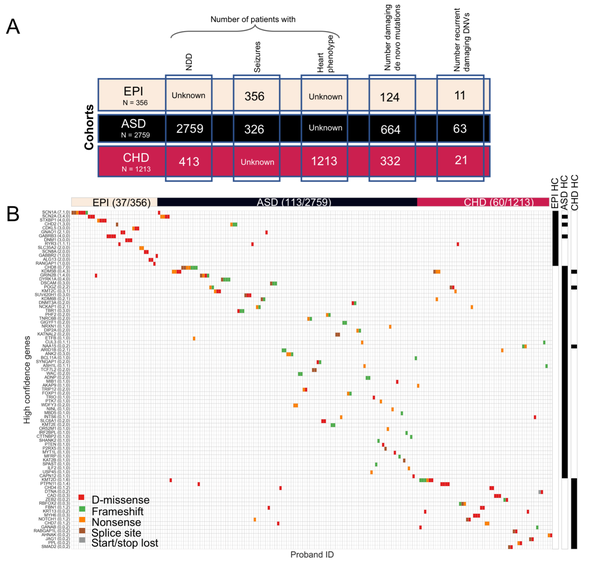

Figure 2: Summary of phenotypes, DNVs, and cohorts. A) Overlap of patients in CHD, ASD, and EPI cohorts with phenotypes of interest. B) High confidence genes found in epilepsy, autism, and congenital heart disorder cohorts.

ASD, EPI, and CHD de novo variants localized in network space

- **Network localization** of DNVs calculated. Highly localized gene sets indicate a strong signal, pointing to a likely shared function. Network localization defined by measuring the difference between the observed and expected largest connected components (Fig. 3A-C), as in (1).
- All three sets of de novo variants are significantly localized in network space, when compared to degree-matched, randomly selected gene sets.

- **Network propagation (2) used to prioritize likely disease-related DNVs**.
- Larger than expected number of DNVs were revealed in the network proximity around each of the high confidence gene sets (ASD p = 6.6E-9, CHD p = 8.7E-4, EPI p = 0.04, hypergeometric tests).
- Importantly, high confidence genes excluded from DNV recovery set, as they were used to seed the propagation, and are guaranteed to have high proximity scores.
- Finding an enrichment of DNVs in the network neighborhood proximal to genes we know to be involved in the disease serves two purposes.
  1. Validation that the method of network propagation and the choice of interactome are effective at extracting signal from noise in the DNV genes.
  2. The re-prioritized DNV list serves to identify novel disease gene candidates (Fig. 3E).

Network propagation methods prioritize genes at the interface of two disorders

- The high observed comorbidity between pairs of disorders suggests that an approach integrating information from two disorders is necessary to uncover more about their shared molecular mechanisms.
- To prioritize genes potentially involved in two disorders, we implemented a double network propagation procedure, where we use similar network propagation mechanisms as in (3,4). We perform two network propagation simulations separately from the ASD high confidence genes and the CHD high confidence genes, and find those genes which have higher than expected scores from both simulations, and thus are likely candidates for involvement in both disorders.
- Genes which are likely to be related to both diseases are found in the upper right corner of the Figure 3. These genes comprise the **ASD-CHD interactome**
- **Validation:**
  - Calculate the enrichment of DNVs from patients with symptoms of both disorders in the combined-prioritized genes (black triangles in Fig. 4A)
  - **Observe a larger than expected number of shared phenotype DNVs** in the upper right corner of the combined proximity graph.
  - Furthermore, **we recover more shared phenotype DNVs using the combined proximity ranking than either ranking by ASD-proximity, or CHD-proximity alone**, suggesting that we are successfully prioritizing those genes which contribute to aspects of both disorders (Fig. 4B).

Genes in ASD-CHD interactome linked to shared phenotype and developmental disorders in general

- **Further validation** of the genes in the ASD-CHD interactome by integrating with the DECIPHER database (5), an interactive web-based database which furthers the interpretation of genomic variants.
- We cross-referenced the 300 highest scoring genes (excluding seed genes) in the ASD-CHD interactome with the DECIPHER database, filtering by patients with phenotypes of both 'Abnormality of nervous system', and 'Abnormality of cardiovascular system. This filtered the patient set to 669 patients with 742 variants.
- A large fraction (288 out of the 669) of patients with the combined phenotype can be accounted for by the top 300 genes (including high confidence seed genes), compared to 152 for low z-score genes. When we include the top 300 genes without high confidence seed genes, 206 patients are accounted for (Figure 6A).
- What's more, the gap between the curves in Figure 6A with seeds and without seeds, is evidence that the non-seed genes are capturing a different set of patients than the seed set, suggesting that **our network prioritization method is contributing information which is distinct from what would be found with just the seed genes.**
- We further connected the ASD-CHD interactome genes with variants in other disorders in the denovo-db dataset (6), highlighting **the complex genetic relationships between neurological and developmental disorders** (Fig. 5C).

Multiple lines of evidence point to ASD-CHD disease-gene candidates

- By combining three pieces of validation for the prioritized ASD-CHD genes, consisting of non-recurrent shared-phenotype DNVs, shared-phenoytpe DECIPHER patients, and patients with developmental disorders from the DDD cohort, we find there are 39 non-seed genes for which there are multiple distinct lines of evidence (Fig. 6D).
- Among these, many were found in all three validation datasets, including **KAT6A, KAT6B**, and **EP300**, all of which have been linked to symptoms of neurological and developmental disorders, either separately or together (7—9).
- These 39 represent novel paired-disease-gene candidates, and as such may be good candidates for followup studies aimed at prediction of the ASD-CHD subtype, diagnosis of patients at risk for developing multiple disorders, and targets for novel therapies.

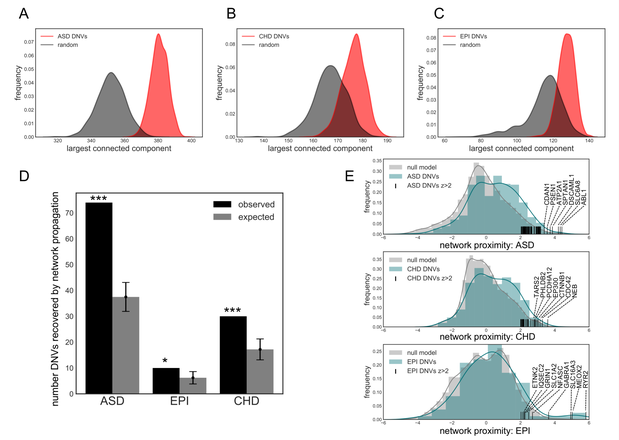

Figure 3: Localization and enrichment of DNVs. A-C) Localization of ASD DNVs (A), CHD DNVs (B), and EPI DNVs (C) in the interactome. D) DNVs are significantly enriched in genes identified with network propagation from individual high confidence gene sets. E) Distributions of network proximity scores for each DNV set, compared to the randomized null model. Top ranking DNVs are labeled.

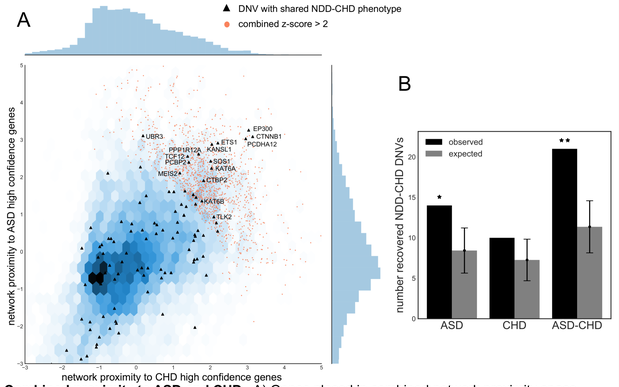

Figure 4: Combined proximity to ASD and CHD. A) Genes placed in combined network proximity space. Genes with combined proximity z-scores greater than 2.0 are shown in orange. Shared phenotype DNVs are indicated with black squares. The highest scoring 15 recovered DNVs are labeled. B) Shared phenotype DNVs are enriched in combined network propagation scores, when compared to randomly selected genes, and when compared to ASD and CHD alone. * p < 0.05, ** p<0.01

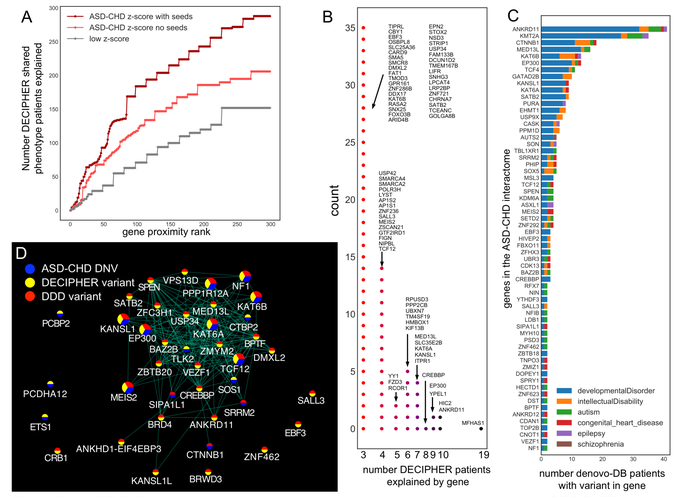

Figure 5: A) ASD-CHD interactome successfully explains a large fraction of patients in the DECIPHER database with abnormal nervous and cardiovascular symptoms. B) Many of the genes identified in A) explain multiple patients, including MFHAS1 which explains 19 different shared phenotype patients. C) The genes captured in the ASD-CHD interactome cross-referenced with all phenotypes in denovo-db. D) Summary of genes in the ASD-CHD interactome for which at least two lines of validation evidence exist.

Conclusions

- By integrating known, high confidence disease-related genes with tissue-specific networks, we successfully prioritized new, likely pathogenic, de novo variants, and proposed a new pathway of highly interconnected genes which represent the intersection of ASD and CHD.
- Many genes in this new pathway were validated in an independent, external database, with many of these genes having multiple lines of supporting evidence.
- Furthermore, many of those prioritized genes without multiple lines of evidence, have literature support linking them to a specific developmental disorder, but not the two at the focus of this analysis. This further emphasizes the interrelatedness of many developmental disorders. These genes represent good candidates for further experimental follow up and possible therapeutic targets.
- While the focus here has been on ASD and CHD, touching briefly on other developmental and neurological disorders, this method may be applied to any pair of diseases for which there exists sets of high confidence genes.
- Wide application of this type of paired network analysis to all developmental and neurological disorders may aid in filling in the genetic landscape of disease, and will allow for more precise diagnoses and treatments.

Literature cited

1. Menche, Jörg, et al. "Uncovering disease-disease relationships through the incomplete interactome." *Science* 347.6224 (2015): 1257601.
2. Cowen, Lenore, et al. "Network Propagation: A Universal Amplifier of Genetic Associations." *Nature Reviews Genetics* 18.9 (2017): 551–62.
3. Vanunu, Oron, et al. "Associating genes and protein complexes with disease via network propagation." *PLoS Comput Biol* 6.1 (2010): e1000641.
4. Paull, Evan O. et al. "Discovering Causal Pathways Linking Genomic Events to Transcriptional States Using Tied Diffusion Through Interacting Events (TieDIE)." *Bioinformatics* 29.21 (2013): 2757–64.
5. Firth, Helen et al. "The Deciphering Developmental Disorders (DDD) Study." *Developmental Medicine and Child Neurology* 53.8 (2011): 702–3.
6. Turner, Tychele et al. "Denovo-Db: A Compendium of Human de Novo Variants." *Nucleic Acids Research* 45.D1 (2017): D804–11.
7. Millan, Francisca et al. "Whole Exome Sequencing Reveals de Novo Pathogenic Variants in KAT6A as a Cause of a Neurodevelopmental Disorder." *American Journal of Medical Genetics*. Part A 170.7 (2016): 1791–98.
8. Babu, Aswini, et al. "Chemical and Genetic Rescue of an ep300 Knockdown Model for Rubinstein Taybi Syndrome in Zebrafish." *Biochimica et Biophysica Acta* 1864.4 (2018): 1203–15.
9. Marangi, Giuseppe et al. 2018. "A Novel Truncating Variant within Exon 7 of KAT6B Associated with Features of Both Say-Barber-Bieseker-Young-Simpson Syndrome and Genitopatellar Syndrome: Further Evidence of a Continuum in the Clinical Spectrum of KAT6B-Related Disorders." *American Journal of Medical Genetics*. Part A 176.2 (2018): 455–59.

Acknowledgments

This work was funded by the UC San Diego Health Sciences department, the National Resource for Network Biology grant P41GM103504, the San Diego Center for Systems Biology grant P50GM085764, the Clinical and Translational Research Institute grant UL1TR001442.