

# An oligogenic inheritance test detects risk genes and their interactions in congenital heart defects and developmental comorbidities

Maureen Pittman<sup>1,2</sup> Kihyun Lee<sup>1,3</sup>, Deepak Srivastava<sup>1,3,\*</sup>, Katherine S. Pollard<sup>1,2,4,\*</sup>

Affiliations:

1 Gladstone Institutes, San Francisco, CA 94158

2 Department of Epidemiology & Biostatistics, University of California, San Francisco, CA 94158

3 Department of Pediatrics, University of California, San Francisco, CA 94158

4 Chan Zuckerberg Biohub, San Francisco, CA 94158

\* Corresponding author

Running title: GCOD detects developmental risk gene interactions

Abstract: Exome sequencing of thousands of families has revealed many individual risk genes for congenital heart defects (CHD), yet most cases cannot be explained by a single causal mutation. Further, those who carry *de novo* and inherited mutations in known risk genes often demonstrate variable phenotypes even within the same family, indicating the presence of genetic modifiers. To explore oligogenic causes of CHD without assessing billions of variant combinations, we developed an efficient, simulation-based method to detect gene sets that carry damaging variants in probands at a higher rate than expected given parental genotypes. We implemented this approach in software called Gene Combinations in Oligogenic Disease (GCOD) and applied it to a cohort of 3382 trios with exome sequencing. This analysis detected 353 high-confidence risk genes in 202 pairs that appear together in multiple probands but rarely or never appear in combination in their unaffected parents. Stratifying analyses by specific CHD diagnosis and considering gene combinations of higher orders yielded an additional 244 gene sets. The oligogenic genes we discovered cluster in pathways specific to heart development and suggest new molecular disease mechanisms, such as arylsulfatase activity and *de novo* nucleotide biosynthesis. Finally, by combining CHD families with an autism spectrum disorder cohort, we were able to detect 925 oligogenic sets transmitted in renal disease, a known co-morbidity of both conditions. As genome sequencing is applied to more families and other disorders, GCOD will enable detection of increasingly large, novel gene combinations, shedding light on combinatorial causes of genetic diseases.

## Introduction

Many disease phenotypes are complex and genetically heterogeneous such that damaging variants at different genetic loci can lead to similar disease manifestations (Veenstra-Vanderweele, Christian, and Cook 2004; Akhirome et al. 2017). This genetic architecture

complicates our ability to identify causal variants and genes, despite increasingly large cohorts with exome or whole genome sequencing. For instance, congenital heart defects (CHD) have heritability estimates of 70-90%, but just 8% of probands have a damaging variant in a known CHD gene that is not shared by an unaffected family member (Jin et al. 2017). This indicates that although most defects can be explained by patient genetics, our current knowledge is insufficient to pinpoint the full genetic source and mechanism of most individual cases. Some risk genes remain cryptic due to incomplete penetrance, which reduces our statistical power to detect them; at the same time, previously-identified risk genes and variants may themselves exhibit incomplete penetrance such that carriers differ in the presence and severity of disease symptoms (Parker and Landstrom 2021). A potential explanation for these phenomena is oligogenic inheritance, or multifactorial inheritance in which a few variants are required in combination to cause a particular phenotype (Kousi and Katsanis 2015). For example, an unaffected father may carry a variant that is tolerated in most genetic backgrounds, and is therefore unlikely to be considered a candidate for causing disease; however, if that variant is transmitted to a child along with a *de novo* or maternally-derived damaging allele in another gene, this could destabilize key pathways and cause disease where the single variant would not.

Previous work has validated one instance of oligogenic inheritance in CHD (Gifford et al. 2019), as well as speculated a role for oligogenic inheritance in other developmental disorders like autism spectrum disorder (ASD) (Schaaf et al. 2011; Wenger et al. 2016). While these studies yielded mechanistic insights into specific gene and variant combinations, they relied on known risk genes and existing functional information to propose plausible hypotheses about specific gene combinations. Alternatively, one could enumerate and prioritize all damaging variant combinations in an automated and statistically rigorous manner. But this strategy is complicated by combinatorial explosion: given that each human genome typically contains 40,000 to 200,000 variants at <0.5% minor allele frequency (MAF) (1000 Genomes Project Consortium et al. 2015), an analysis of all possible rare variant pairs yields between 8 billion to nearly 20 trillion

combinations per individual. Nevertheless, there are existing methods that scan the genome for possible oligogenic disease combinations. ORVAL, for example, accepts a list of variants as input, and uses gene annotations to assess the potential for oligogenic combinations (Renaux et al. 2019). However, this tool only offers qualitative hypotheses for a single individual, which limits applicability across probands. In contrast, the Digenic method reduces the combinatorial search space by aggregating rare variants at the gene level and assesses the statistical significance of each gene pair based on its prevalence in a disease cohort (Kerner et al. 2020), but this method is limited to oligogenic combinations of only two genes (“digenic”). Another recent advance in the field is RareComb (Pounraja and Girirajan 2022), an algorithm that tests for greater frequency of gene combinations in cases compared to controls. However, neither of these computational methods incorporate parental sequencing data, which are especially useful in reducing false positives in simplex families, where an affected proband is born to unaffected parents. It logically follows that the highest-effect causal variants are likely transmitted separately from each parent or include at least one *de novo* mutation; thus, incorporating parental transmission data is potentially a powerful tool to reduce false positives, increase confidence in the phenotypic relevance of variants, and correct for population structure.

To address this gap, we developed the tool Gene Combinations in Oligogenic Disease (GCOD), an algorithmic framework that uses simulation experiments to identify disease gene sets that carry rare damaging variants significantly more often than expected given parental genotypes. GCOD automatically assesses every digenic pair observed in multiple probands, as well as sets of three, four, or more genes. Using information from unaffected parents limits the number of gene sets tested, which increases statistical power and accelerates analysis time. Applying GCOD to exome sequencing from families with CHD, we identified potential novel risk genes and pathways in disease. Jointly analyzing these families with an ASD cohort, we demonstrate the power of GCOD for detecting gene sets associated with comorbidities. These

results greatly expand candidate causal mechanisms for both disorders and establish a broad oligogenic component for developmental diseases.

## Results

### Computational identification of pathogenic gene set candidates

We developed an algorithm to predict pathogenic gene sets using exome or whole genome sequencing data and implemented it in software we called GCOD. The method involves four steps (**Figure 1**): select variants of interest based on constraint or other criteria, enumerate co-occurring genes, count their frequency across probands, and evaluate the statistical significance of these counts given the parental genotypes. An optional fifth step repeats the process for healthy siblings or pseudo-siblings (Yu and Deng 2011) as a control. GCOD can be viewed as an extension of the Transmission Disequilibrium Test (TDT) in which the unit being tested is an observed co-occurrence of variants in a group of genes, rather than observed transmissions at a single locus (Spielman, McGinnis, and Ewens 1993) or within a single gene (He et al. 2017). GCOD also differs from the TDT in its use of simulations to assess significance rather than a chi-squared statistic, increasing our power to detect rare gene combinations with very few expected observations. Details and options for each step of GCOD follow.

First, users provide a list of damaging variants meeting some criteria of interest from each individual in the dataset (**Figure 1A**). GCOD then enumerates all combinations of genes that contain co-occurring damaging variants in each proband (**Figure 1B**), retaining only those combinations observed in multiple families (referred to as “candidate sets”). By changing the definition of damaging, different scenarios can be explored, from small gene sets with very rare co-occurring variants of high impact to larger gene sets that include small-effect, possibly-

damaging variants that could modify the effects of other primary disease-causing variant(s) or otherwise contribute to a polygenic signal. We designed GCOD to automatically consider three tiers of variant severity (**Supplemental Table 1**), defined by residue-level characteristics such as minor allele frequency (MAF) and CADD score (Rentzsch et al. 2019), gene-level cutoffs for gnomAD observed-expected z-score (Karczewski et al. 2020), as well as gene  $s_{het}$  score (Cassa et al. 2017) and minimum expression level in relevant cell types (J. Cao et al. 2020). The variant severity tiers are nested such that more severe variants are included in each of the decreasingly strict tiers.

By default, a candidate gene set is only considered in a particular proband if the qualifying variants were derived from at least two sources (that is, “oligogenic transmission” where a healthy parent did not also carry that same combination of variants), though users can remove this requirement. The count of observed oligogenic transmissions comprise the test statistic. To reduce the number of tested hypotheses and the multiple testing correction burden where possible, we implemented an algorithm that begins with digenic pairs and merges these up to the “highest-order” set, defined as the largest unique candidate set observed in two or more probands. Finally, lower-order combinations are added back if an additional proband harbors this candidate set and not of any larger sets. For example, two probands carrying damaging variants in seven genes share 21 unique gene pairs, 35 unique gene trios and quartets, 21 unique gene pentads, and so on (**Supplemental Figure 1, Methods: Highest-order gene set enumeration**). If no other probands carry any of the sets, GCOD tests only the seven-gene set and none of the 119 lower-order combinations. But if a third proband were to carry one of the gene pairs but none of the higher-order candidate sets, the pair (with a count of three) is included along with the seven-gene set (with a count of two). This algorithm produces a co-occurrence count for each highest-order candidate set, recursively checking for additional transmissions at each level descending to pairs.

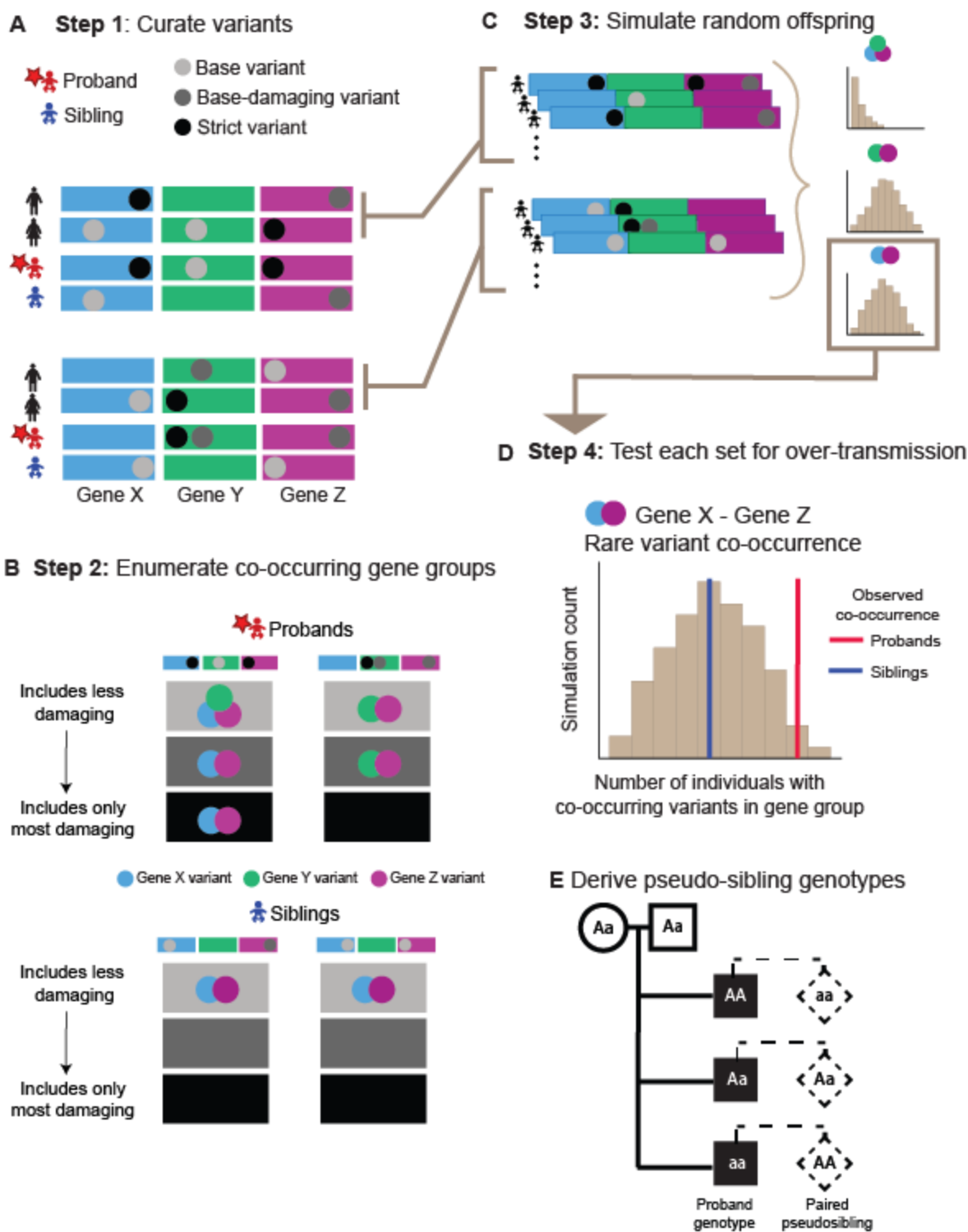
In the third step, GCOD performs simulations based on parental genotypes to assess the statistical significance of the candidate set counts (**Figure 1C**). The goal of the simulations is to estimate the distribution of counts expected by chance if there were no transmission disequilibrium. In each iteration, one hypothetical offspring is simulated for each parental pair as follows. Each damaging parental variant has a 50% chance of being passed to the simulated offspring. We treat variants independently and do not model linkage disequilibrium, because GCOD counts genes with at least one damaging variant rather than the number of variants per gene, and given that linkage disequilibrium decays by 70 kilobases (kb) in most regions of the human genome, genes are only very rarely in the same LD block (Bosch et al. 2009). Additionally, variants must be derived from at least two sources such that a candidate set will never be entirely inherited as a single LD block (i.e., a variant from the other parent or mutated *de novo* must have been observed). After simulating inherited variants, GCOD adds *de novo* variants using previous estimates of per-gene *de novo* mutation probabilities (Samocha et al. 2014). Henceforth, we use the term “oligogenic transmission” to describe an event in which a variant combination occurs in the offspring that was not seen in either parent, i.e. comprises at least one variant from each parent and/or a *de novo* mutation.

For each simulation, we tally the number of times a given gene set was transmitted with oligogenic inheritance in the simulated cohort. Repeating over many iterations generates a null distribution of oligogenic transmissions against which we compare the observed number in probands in the fourth step. We calculate the probability of no over-transmission for each candidate gene set as the proportion of iterations with a simulated count equal to or greater than the number of probands (**Figure 1D**). GCOD returns the full list of candidate sets along with the p-values associated with their transmission to probands (unadjusted, Benjamini-Hochberg corrected, and Bonferroni corrected).

In an optional fifth step, GCOD repeats this entire procedure for sibling controls. If healthy siblings have been sequenced, these may be used. Otherwise, GCOD generates a genotype for

a pseudo-sibling of each proband, a hypothetical sibling who has inherited the parental alleles that were not transmitted to the affected individual (Yu and Deng 2011) (**Figure 1E**). Siblings, unlike pseudo-siblings, carry some variants inherited by the proband. They also can be assessed for the phenotype, whereas pseudo-siblings cannot. Siblings without a diagnosis and pseudo-siblings are typically presumed to be unaffected, though this is not always true; siblings must be deeply phenotyped past the typical age of onset to rule out disease, and pseudo-siblings have unknown phenotypes. Hence, it can be helpful to use both when sibling data is available and to interpret results with this caveat in mind. Collectively, the sibling controls serve several purposes. Significant candidate sets from probands can be filtered to exclude any that also appear in (pseudo-)siblings. A comparison of the number of significant sets for probands versus (pseudo-)siblings can also be used to assess overall evidence that a phenotype is likely oligogenic, where probands would be expected to carry more over-transmitted gene sets than (pseudo-)siblings. Finally, the types of genes in significant sets can be compared between probands and (pseudo-)siblings, for example, using gene set enrichment, pathway analysis, or literature review. These analyses help to refine and prioritize significant sets from step three. We refer to the resulting statistically significant gene sets as “oligogenic sets”.





### **Figure 1. The GCOD approach to identify sets of genes that interact in oligogenic disease.**

**A:** Variants in each family meeting user-defined criteria of “damaging” are curated. GCOD by default provides three tiers of variant severity based on constraint annotations: Base, Base-damaging, and Strict. The variants considered in the most stringent tier of variant severity (Strict) are included in the analyses of the more lenient tiers. **B:** Next, GCOD enumerates the candidate sets of genes for which variants occur in multiple probands, by default only including combinations not transmitted from a single parent. This can be repeated for unaffected siblings and/or pseudo-siblings (see **E**). **C:** A user-defined number of simulations are performed for each pair of parents; variants have a 50% chance of being passed to the simulated offspring, and each gene has an associated per-chromosome probability of *de novo* mutation. This yields a distribution of simulated carriers for each candidate set. **D:** For a given candidate set, the simulated distribution is compared to the observed number of oligogenic transmissions to probands to compute a p-value. This is repeated for siblings and/or pseudo-siblings. **E:** Method for pseudo-sibling generation: for each input variant locus, the pseudo-sibling inherits the alleles that were not transmitted to the observed proband.

## Probands with a congenital heart defect carry more oligogenic gene sets than familial controls

To discover oligogenic disease genes for CHD, we applied GCOD to whole exome sequencing data from 3382 families in the Pediatric Cardiac Genomics Consortium (PCGC) (Pediatric Cardiac Genomics Consortium et al. 2013; Hoang et al. 2018; Morton et al. 2021). In each family, the PCGC generated exome variants for two asymptomatic parents and an affected child. We used parental data to compute pseudo-sibling genotypes for each trio (Yu and Deng 2011) (**Figure 1E**). We ran GCOD at three tiers of variant severity (**Supplemental Table 1**): the

Strict tier which includes only very rare mutations with a high probability of impairing protein function, plus two more lenient tiers of variant severity (Base-damaging and Base). Only oligogenic transmissions of a variant set were considered (see above). To reduce false positives caused by a single driver gene with spurious co-transmitted partners, we additionally filtered oligogenic sets to those with a logistic regression interaction coefficient greater than 1.0 (**Methods: Logistic regression**). In short, this step aims to restrict our analysis to sets in which a higher proportion of individuals with damaging mutations in the full oligogenic set have CHD compared to individuals with damaging mutations in each gene separately.

The GCOD approach revealed 202 oligogenic gene pairs comprising 353 CHD risk genes at the Strict and Base-damaging tiers (**Supplemental Table 2**). At each variant tier, we identified more oligogenic gene pairs in CHD probands compared to their pseudo-siblings (**Figure 2A**). These differences were all statistically significant (Binomial  $p \leq 1.4 \times 10^{-23}$ ). At the Strict tier, for example, we found 179 gene pairs significantly over-transmitted to probands, compared to just 38 gene pairs over-transmitted to their pseudo-siblings (Bonferroni-adjusted  $p < 0.05$ ) (**Table 1**). A similar nearly 5-fold enrichment was seen at the Base-damaging tier, and an almost 12-fold enrichment was detected at the Base tier. Oligogenic pairs are present in a substantial number of probands, with 9.2% of PCGC probands carrying at least one pair even at the Strict tier. Together these results provide evidence that CHD is broadly oligogenic.

As expected, for both probands and pseudo-siblings the number of oligogenic pairs increased with more lenient definitions of damaging variants (**Table 1**), with the Base tier yielding 22515 oligogenic pairs in probands (versus 2783 in pseudo-siblings). This many significant gene pairs means that 96.6% of probands carry at least one oligogenic pair at the Base tier. These gene pairs likely comprise a more diffuse polygenic risk compared to the 179 pairs at the Strict tier, which may play a more central role in the disease of the 312 probands who carry them. We

therefore focus the subsequent analyses on the oligogenic sets from the Strict and Base-damaging tiers, which are expected to have larger pathogenic effects per occurrence.

Tier	Disease Status	N Oligogenic Pairs	N unknown genes in Oligogenic Pairs (N Total Genes)	N (% of cohort) that carry at least one pair	Number of Single Genes	Number of CHD known Single Genes
Strict	Proband	179	295 (317)	312 (9.2%)	133	8
	Pseudo-sibling	38	70 (70)	79 (2.3%)	0	0
Base-damaging	Proband	189	312 (334)	328 (9.7%)	136	12
	Pseudo-sibling	38	70 (71)	76 (2.2%)	0	0
Base	Proband	22515	7627 (7782)	3279 (96.9%)	363	10
	Pseudo-sibling	1893	2498 (2541)	1693 (50.1%)	51	0

**Table 1.** GCOD counts of significantly over-transmitted oligogenic pairs and genes, categorized by membership to the list of curated human/mouse CHD genes in Jin et al. 2017 (Jin et al. 2017).

We next computed the highest-order gene sets shared among PCGC probands using Strict and Base-damaging variants (**Methods: Highest-order gene sets**). Highest-order sets ranged from groups of two genes to groups of twelve genes for both probands and pseudo-

siblings. As with gene pairs, we find a significantly greater count of over-transmitted sets in probands compared to pseudo-siblings (both binomial  $p < 1 \times 10^{-150}$ , **Figure 2B**). The highest-order oligogenic sets at the Base-damaging and Strict level yield an additional 1662 potential contributing risk genes.

We next investigated whether the oligogenic pairs identified in PCGC probands include any novel CHD genes that have been missed using smaller datasets and single-gene tests. First, we annotated which oligogenic sets contained one or more of the 253 curated human or mouse CHD genes reported by Jin et al. (Jin et al. 2017). While 18% of Base-damaging and Strict proband oligogenic pairs contain at least one of these CHD genes, the vast majority of risk genes captured by GCOD were not previously known (312 of 334 genes, **Table 1**). Next, we directly assessed the additional genes discovered by testing pairs versus single genes within the GCOD simulation framework (**Methods: Single-gene transmission simulation test**). This analysis is distinct from using known CHD genes, because prior studies used a variety of statistical tests, cohorts, and sample sizes. GCOD detects about a third as many genes in single-gene mode compared to testing oligogenic pairs at the Strict and Base-damaging tiers, and this difference is even more pronounced at the Base tier (**Table 1**). In all tiers, fewer known CHD genes are detected in single-gene mode compared to when oligogenic sets are run. These findings underscore the additional signal GCOD is able to capture by considering oligogenic transmissions.

## Oligogenic combinations are found together in canonical cardiac gene sets significantly more often in CHD probands

We have so far shown that proband cohorts have a greater number of oligogenic gene sets, and that these sets are enriched for known CHD risk genes. To test whether the CHD oligogenic pairs identified by GCOD comprise phenotype-relevant genes, we performed a Gene

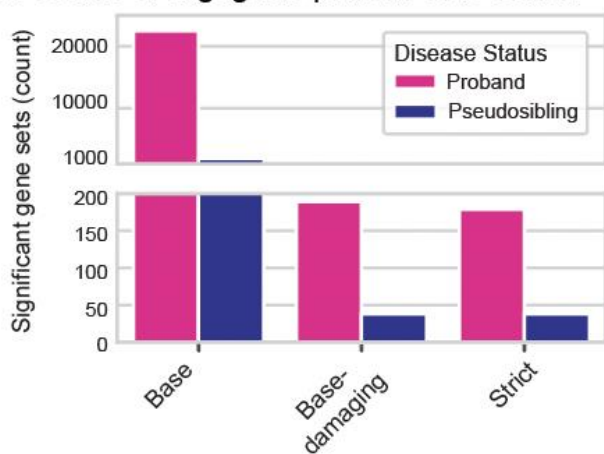
Ontology (GO) analysis. Strict tier oligogenic genes were enriched for many GO terms related to heart development, such as “atrioventricular canal development” and “cardiac atrium morphogenesis” (**Figure 2C**). In contrast, no GO terms were enriched in pseudo-sibling genes at the Strict or Base-damaging tiers. We therefore conclude that genes comprising proband oligogenic pairs are relevant to CHD, and genes over-transmitted to pseudo-siblings are largely false positives unrelated to any particular phenotype, confirming their utility as controls. Interestingly, pseudo-sibling sets at the more permissive Base tier are enriched for GO:0046872 metal ion binding and GO:0005524 ATP binding, potentially indicating that related polygenic risk for CHD is present in healthy members of PCGC families (**Supplemental Figure 2**).

The aforementioned analysis combines oligogenic sets to find common functions among our putative risk genes. We further hypothesized that genes *within* the same disease-associated oligogenic set would share functional annotations and also co-occur in pathways and protein complexes. To test this, we calculated the odds of an oligogenic combination being found in a proband and co-occurring in a Molecular Signatures Database (MSigDB) ontology gene set, and found a significant association for the combined GO collections (odds ratio = 1.6, Fisher exact  $p = 2.2e-33$ ) and for the Human Phenotype Ontology collections related to the heart (odds ratio = 3.8, Fisher exact  $p = 3.5e-12$ ). A similar enrichment was observed for the PCNet composite database of gene-gene interaction networks (S.-Y. Cao et al. 2021) (odds ratio = 1.8, Fisher exact  $p = 4.8e-05$ ). These results indicate that the genes comprising CHD oligogenic pairs in the PCGC cohort are functionally associated, perhaps indicating that the interaction of their gene products is necessary for shared functions in the context of disease pathogenesis.

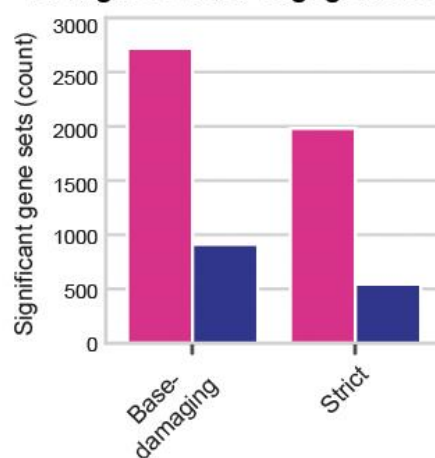
In summary, we have enumerated gene sets for which multiple CHD probands inherited an oligogenic set of variants and tested whether each set was transmitted together more often than expected. We showed that the genes in the resulting oligogenic sets are enriched for heart development functions and tend to occur together in canonical gene networks. By leveraging rare

variation from separate sources (two parents or inherited plus *de novo*), GCOD captures additional known and putative disease genes that other rare variant aggregation methods do not.

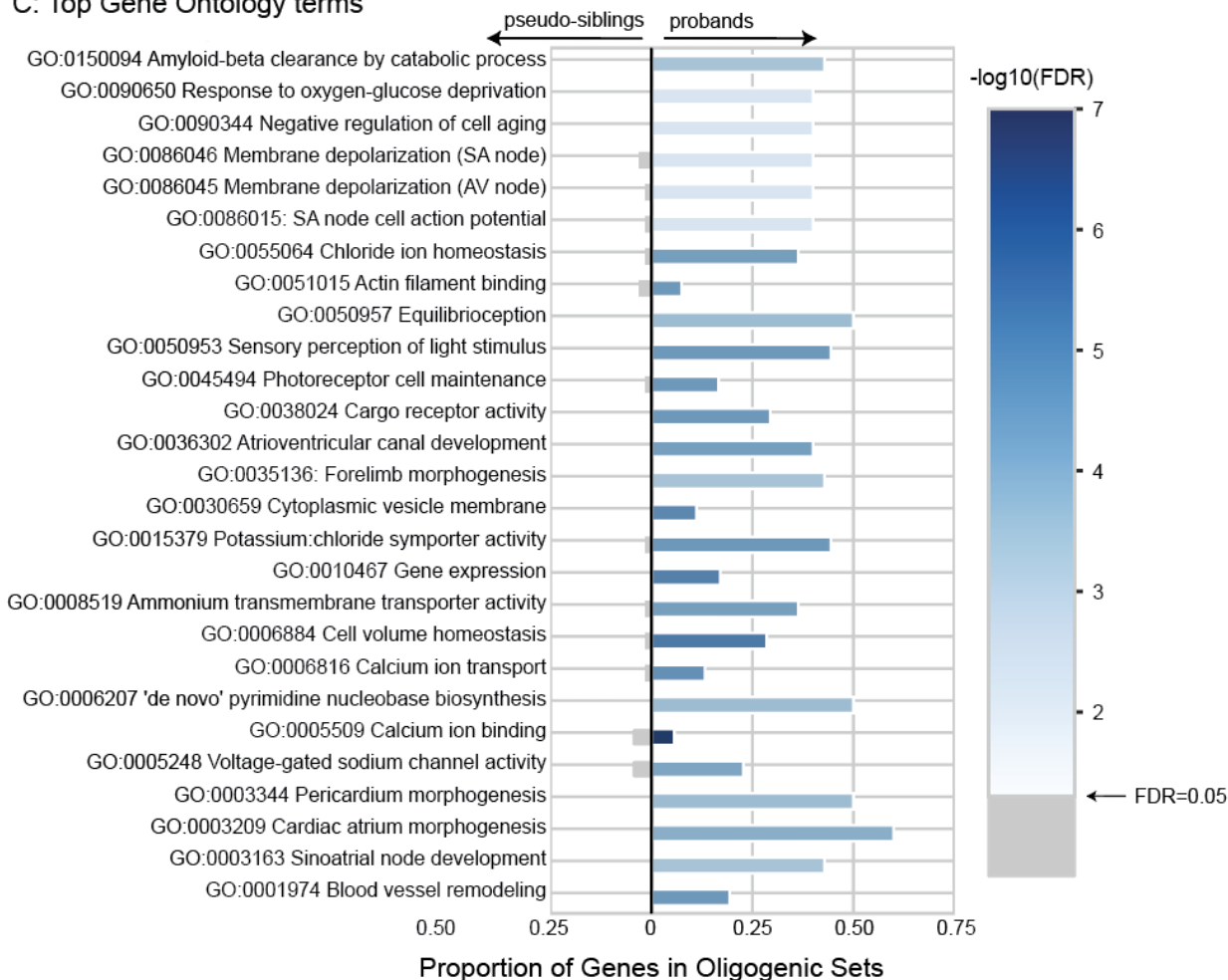
A: Counts of oligogenic pairs in CHD cohort



B: Highest-order oligogenic sets



### C: Top Gene Ontology terms



**Figure 2. Greater number and phenotypic relevance of co-occurring gene sets in congenital heart defect probands.** **A:** Counts of gene pairs that were transmitted to offspring more often than expected by chance in probands (magenta) and pseudo-siblings (blue). **B:** Counts of highest-order oligogenic sets for Strict and Base-damaging tiers in probands (magenta) and pseudo-siblings (blue). **C:** Top 15 GO terms with highest fold enrichment and 15 GO terms with lowest FDR for genes in Strict oligogenic pairs (total 27 terms: “Potassium:chloride symporter activity”, “Sensory perception of light stimulus”, and “Atrioventricular canal development” are among the top categories for both fold enrichment and FDR significance.)

## Novel risk genes and interactions in CHD



While up to this point we have analyzed oligogenic transmission frequency across the entirety of the exomes of the 3382 trios in the PCGC, these probands have been diagnosed with myriad distinct diagnoses. To explore the possibility that similar phenotypes are caused by more similar genetic etiology, we selected five diagnoses with relatively consistent phenotypic criteria (**Table 2**): Ebstein’s anomaly, pulmonary atresia with intact ventricular septum (PA-IVS), truncus arteriosus, tetralogy of Fallot, and hypoplastic left heart syndrome (HLHS). HLHS with mitral atresia and aortic atresia (HLHS-MA-AA) or mitral stenosis and aortic atresia (HLHS-MS-AA) were also analyzed separately. We hypothesized that these smaller analyses might have improved power to detect oligogenic sets that are specific to one or a few diagnoses but do not replicate in the entire cohort.

<b>CHD diagnosis</b>	<b>Proband Count (989)</b>	<b>Description</b>
Hypoplastic Left Heart Syndrome (HLHS) - all subtypes	427	The left ventricle is underdeveloped and too small, often with mitral and aortic valve atresia (absence of opening) or stenosis (narrowing).
HLHS - mitral atresia and aortic atresia (HLHS-MA-AA)	104	A severe form of HLHS accompanied by atresia of both mitral and aortic valves.
HLHS - mitral stenosis and aortic	41	A rare version of HLHS characterized by atresia of the aortic valve, while the mitral valve opens but is narrowed.

atresia (HLHS-MS-AA)		
Truncus Arteriosus	240	The aorta and pulmonary artery fail to separate completely during development and the corresponding conotruncal ventricular septum does not form (ventricular septal defect).
Tetralogy of Fallot	202	Tetralogy of Fallot comprises four co-occurring defects: ventricular septal defect, pulmonary valve stenosis, an aorta displaced over the ventricular septum, and increased muscle mass (hypertrophy) of the right ventricle.
Pulmonary Atresia, intact ventricular septum (PA-IVS)	66	Atresia of the pulmonary valve, with an intact ventricular septum (no ventricular septal defect).
Ebstein's anomaly	54	A displacement of the tricuspid valve into the right ventricle that often leads to valve regurgitation and right ventricular dysfunction.

**Table 2. Description of CHD sub-diagnoses and respective sample size of patients analyzed in this study.**

Counts of oligogenic sets including digenic pairs and highest-order sets in phenotypically similar sub-groups corroborated our finding from the full PCGC cohort that probands harbor significantly more oligogenic sets compared to pseudo-siblings (**Figure 3A**). Many of these sets were found in the broader PCGC as a whole, but we find 160, 181, and 7960 additional diagnosis-specific oligogenic sets at the Strict, Base-damaging, and Base tiers respectively (**Supplemental Table 3**). Many of these were found in the largest diagnostic categories, like HLHS (all subtypes)

and tetralogy of Fallot. However, even at the Strict level we detect the oligogenic pair *CAPN9-MGA*, a rare combination which was transmitted to two of just 66 individuals with PA-IVS. Almost eight percent of patients diagnosed with HLHS-MA-AA carry oligogenically-transmitted damaging mutations in one of the following oligogenic sets: *DNA2-FOXK2*, *MEGF6-SRRM2*, *SACS-TNC*, or *CRB1-GALNS*.

We cross-referenced canonical protein complexes sourced from the CORUM database (Giurgiu et al. 2019) with our oligogenic sets from all of PCGC (**Supplemental Table 2**) and from separate diagnoses (**Supplemental Table 3**), finding 34 proband oligogenic sets in which two or more genes physically interact. For example, the products of the *CREBBP* and *EP300* genes participate in the p300-CBP-p270-SWI/SNF complex, which regulates histone acetyltransferase activity during development (Chan and La Thangue 2001). Variants in these genes were transmitted oligogenically to two individuals in combination with the *COL6A3* gene (*COL6A3-CREBBP-EP300*, 2 probands,  $p = 0.016$ ). These three genes were also over-transmitted independently with other genes, which led to the discovery of a network of interconnected oligogenic sets that collectively harbor multiple hits to three complexes: p300-CBP-p270-SWI/SNF, AML1-HIPK2-p300, and pRb2/p130-multimolecular complex (**Figure 3**).

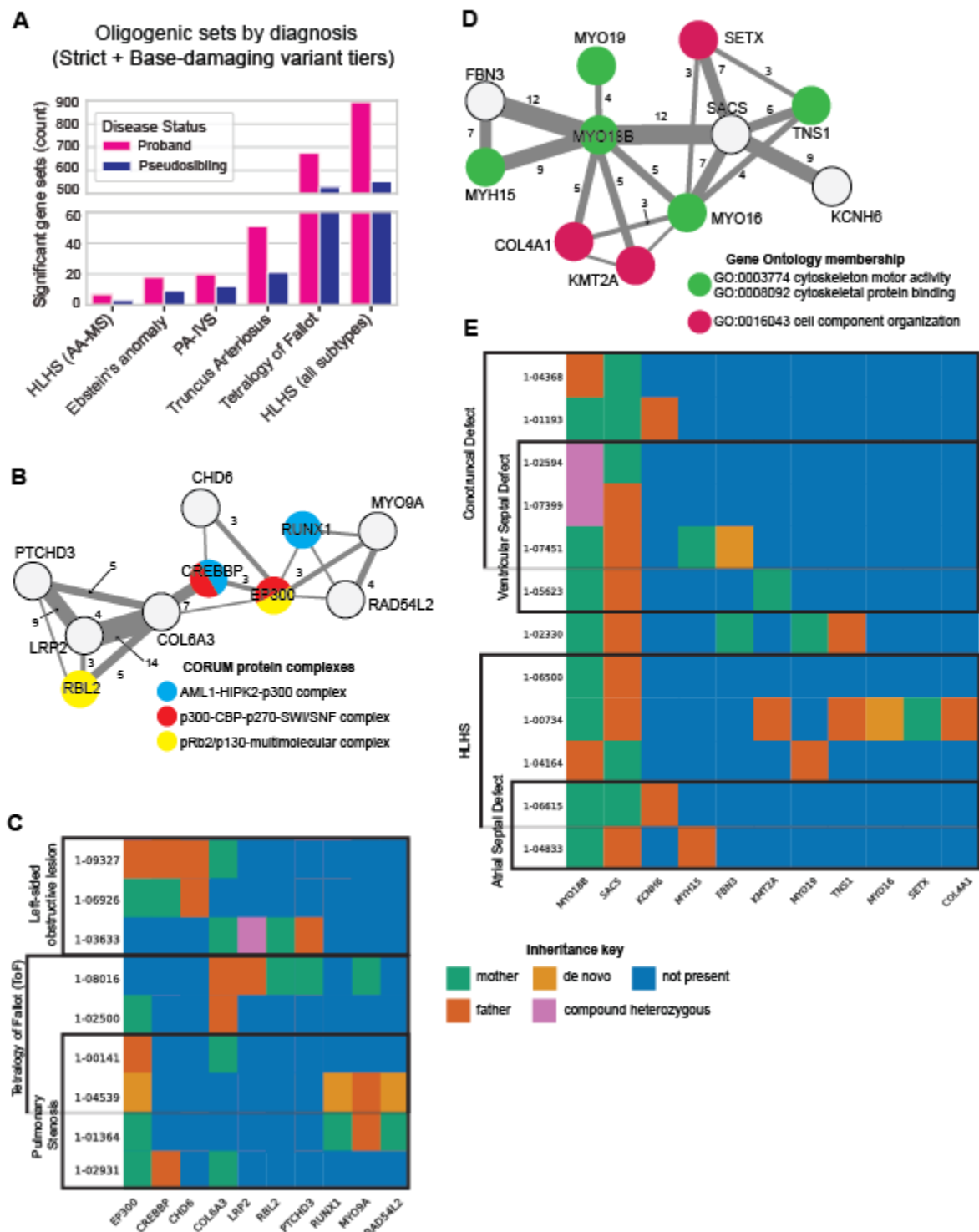
*EP300* co-occurs with *RUNX1* in the AML1-HIPK2-p300 complex, and although the digenic pair is observed in one unaffected parent, the highest-order oligogenic set including *MYO9A* from an alternate source is only observed in CHD probands, both diagnosed with pulmonary stenosis (**Figure 3C**, left; *EP300-MYO9A-RAD54L2-RUNX1*, 2 probands,  $p < 0.0001$ ). *RUNX1* phosphorylates *EP300* and *CREBBP* to activate acetyltransferase activity (Aikawa et al. 2006). The co-occurrence of multiple combinations of transcriptional regulators and chromatin modifiers suggests that they interact in gene regulatory networks that are necessary for normal heart development. Genes not annotated with these functions, such as *COL6A3* and *MYO9A*, could be target genes or have upstream or downstream regulatory functions. For example, the fibrillar collagen gene *COL6A3* has a role in TGF $\beta$  signaling as previously characterized in

cancers (Huang et al. 2018; Dankel et al. 2020). Given the involvement of TGF $\beta$  signaling in epithelial–mesenchymal transformations in cardiovascular development (Azhar et al. 2003) and its coactivation by the CBP/p300 complex (Janknecht, Wells, and Hunter 1998), co-occurring mutations in *COL6A3*, *CREBBP*, and *EP300* are likely pathogenic due to perturbation of this pathway.

Notably, twelve probands carried predicted-damaging variants in both the *MYO18B* gene and the *SACS* gene, which encode for the myosin-18B and saccin proteins respectively (**Figure 3B**, bottom; **Supplemental Table 3**). Ten of these observations were oligogenic transmissions in which the variants did not co-occur in an unaffected parent. The other two probands inherited *MYO18B-SACS* from their respective mothers, but additionally inherited mutations in *KCNH6* from the fathers (**Figure 3C**, right). We conclude that there is an underappreciated role for the *SACS* gene in heart development, likely related to the organization and activity of the cytoskeleton. This is consistent with previous research indicating that saccin regulates filament assembly, though this was determined in neurofilaments (Gentil et al. 2019). The *SACS* protein also acts as a chaperone in the binding of LRP1B (Marschang et al. 2004).

*LRP1B* was found to be significantly over-transmitted with *SACS* at the Base tier of variant severity (4 probands,  $p < 0.0001$ ). Furthermore, oligogenic transmission of *LRP1B* variants with both *MYH6* and *SRCAP* variants at the Strict tier was observed in three families. *SRCAP* is the Snf2-related CREBBP activator protein, and is known to be involved in histone modification and oxidative metabolism during prenatal heart development (Xu et al. 2021), and *MYH6* is a major component in the thick filaments of the sarcomere (Mahdavi, Periasamy, and Nadal-Ginard 1982; Warkman et al. 2012). Oligogenic transmission of Base-damaging/Strict variants in the *MYH6-SRCAP* pair itself occurred in nine CHD probands ( $p < 0.0001$ ). Other potential genetic modifiers of this interaction include the oxidase *DUOX1* (transmitted with *MYH6-SRCAP* in 4 of 9 probands) and the membrane-repair gene *DYSF* (2 of 9 probands). This result suggests a genetic interaction

of cell membrane proteins like LRP1B and its chaperone SACS with genes involved in cell motility and oxidative metabolism.



**Figure 3. Oligogenic sets in congenital heart defects. A:** Counts of oligogenic sets at Strict and Base-damaging tiers of variant severity by CHD diagnosis, with probands in magenta and pseudo-siblings in blue. **B:** Network visualization of genes in oligogenic sets containing one or more of the genes coding for members of AML1-HIPK2-p300 complex, p300-CBP-p270-SWI/SNF complex, or pRb2/p130-multimolecular complex. Node colors represent functional annotations and edge width indicates the number of probands with co-occurring mutations in each gene pair. Note that the edges can include observations that were inherited from a single parent (non-oligogenic transmission) if the gene pair is part of a higher-order oligogenic set. Non-annotated edges represent a co-occurrence count of two probands. **C:** Variant inheritance matrix for genes in B. Columns indicate genes, and rows indicate probands that harbor damaging mutations in each gene. Cell color represents a proband's inheritance of variant(s) in the specified gene column: blue if the proband does not carry a damaging mutation in that gene; green, orange, and pink for mother, father, and both parents respectively; and gold for *de novo* mutations. Multiple probands sharing a diagnostic category are indicated by black boxes. **D:** Network visualization of interconnected oligogenic sets, centered on the *MYO18B-SACS* digenic pair. Genes annotated with ontology terms related to cytoskeleton activity and cellular component organization are indicated in green and pink respectively. **E:** Variant inheritance matrix for genes in D.

## Novel risk genes and interactions in developmental comorbidities

CHD probands have an elevated risk of diagnosis with neurodevelopmental delay and ASD later in life, due to some combination of shared genetic etiology and downstream injury induced by the heart defect (Marino et al. 2012; McCarthy et al. 2015; Wernovsky and Licht 2016). ASD and CHD share 44 canonical risk genes (**Supplemental Table 4**), including genes implicated in chromatin remodeling like *CHD7* (Kasah, Oddy, and Basson 2018; Ji et al. 2020) and cell signaling like *PTPN11* (S. Wang et al. 2009). Furthermore, both of these developmental

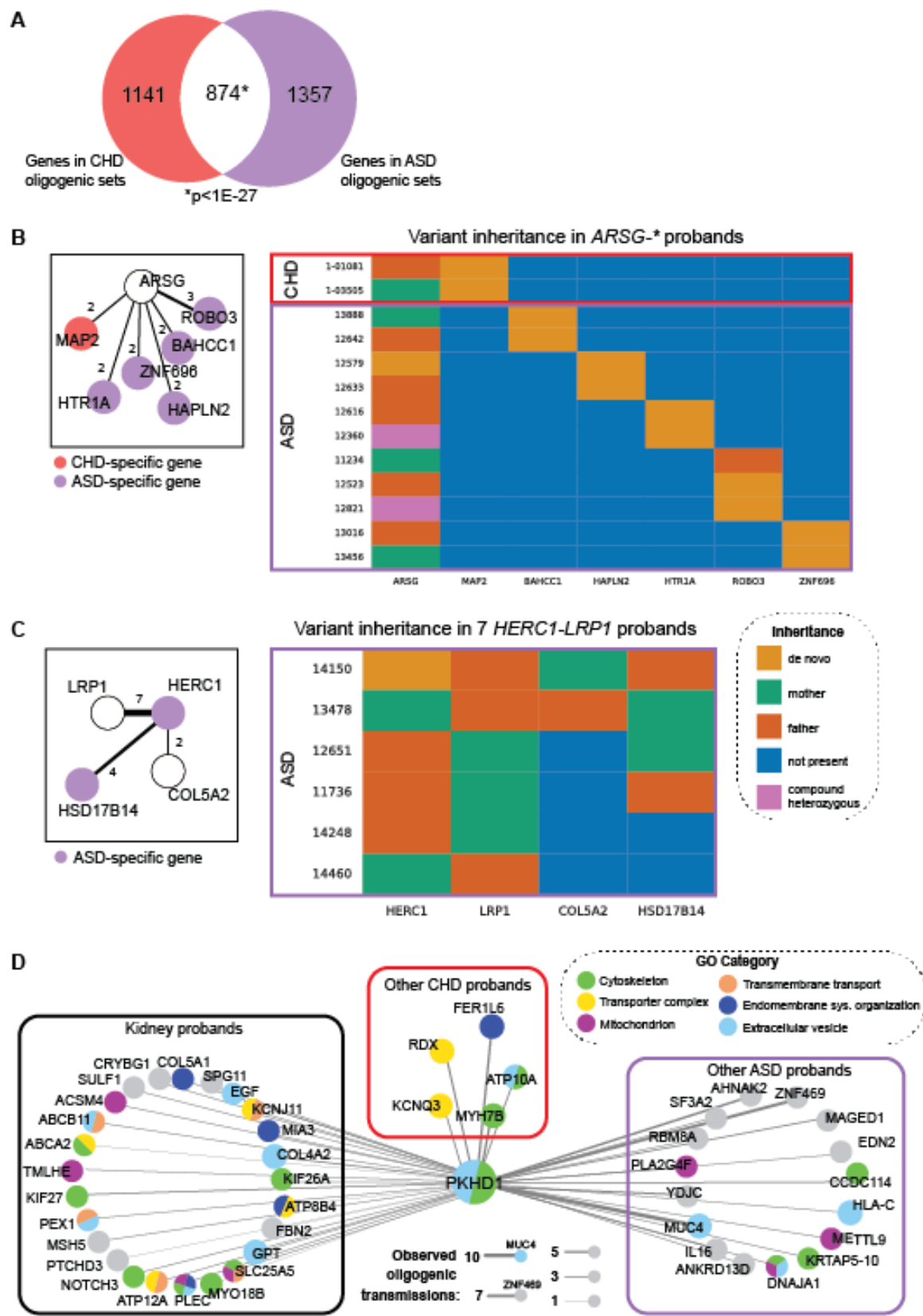
phenotypes are associated with secondary renal (San Agustin et al. 2016; Gabriel, Pazour, and Lo 2018; Clothier and Absoud 2021) and gastrointestinal (Feng et al. 2021; Niesler and Rappold 2021) comorbidities. We therefore sought to characterize the extent to which ASD and CHD probands share oligogenic gene sets and functions. We used GCOD to analyze the sequenced exomes of 1793 families from the Simons Foundation Autism Research Initiative (SFARI) (Fischbach and Lord 2010) and found 1358 Strict and 2866 Base-damaging tier oligogenic pairs. Since an unaffected sibling is sequenced in each family, we were able to further filter this set to those pairs not transmitted to any unaffected sibling (1023 Strict and 2060 Base-damaging pairs, **Supplemental Table 5**). Of the 1310 genes found in these ASD oligogenic pairs, 138 have been previously characterized in ASD, making the other 1172 novel risk gene candidates.

Interestingly, we found no oligogenic pairs or sets over-transmitted to both ASD and CHD probands at the Strict and Base-damaging tiers. However, the overlap of genes contained within these sets is greater than expected by chance (**Figure 4A**, Hypergeometric test  $p < 1 \times 10^{-27}$ ). GO terms significantly overrepresented by the overlapping genes include calcium ion binding, cell motility, and extracellular matrix activity (**Supplemental Table 6**). One novel gene shared between CHD and ASD probands is *ARSG*, which encodes for the arylsulfatase G protein. While its known functions are limited to arylsulfatase and heparin hydrolysis, homozygous variants play a causal role in Usher syndrome, which involves congenital deafness and vision loss (Khateb et al. 2018; Abad-Morales et al. 2020). In the CHD cohort, a *de novo* mutation in the microtubule assembly gene *MAP2* was co-transmitted with a rare inherited *ARSG* variant in two probands, more often than expected by chance (**Figure 4B** left,  $p < 0.0001$ ). Fifty unaffected parents carried a rare predicted-damaging variant in one of either gene, suggesting that a damaged copy of one is not sufficient to cause a defect. The probands with both *MAP2* and *ARSG* variants were born with valvular defects (aortic stenosis or Ebstein's anomaly, respectively), and have no known learning disability or autism spectrum diagnoses. In the SFARI autism cohort, *ARSG* was significantly over-transmitted to probands with five genes, namely *ROBO3*, *BAHCC1*, *ZNF696*,

*HAPLN2*, and *HTR1A* ( $p < 0.0001$  for all, **Figure 4B** left). *ROBO3*, *ZNF696*, and *HTR1A* have been previously implicated in autism (Anitha et al. 2008; Krumm et al. 2015; Sener et al. 2016), while *BAHCC1* is known to be necessary in neuronal commitment (Hezroni et al. 2020), and *HAPLN2* has been associated with neurological disorders (Q. Wang et al. 2019). We suspect that *ARSG* participates in embryonic development of heart and brain (as well as ear and eye), possibly through a shared function like calcium ion binding, cell motility, or extracellular matrix activity; additional modifier genes in each case may determine whether this manifests in heart or brain phenotypes. Notably, genes in oligogenic sets with *ARSG* were found to be mutated *de novo* in at least one proband with the combination (**Figure 4B** right), demonstrating that GCOD is able to recover combinations with *de novo* variants as well as the more common transmission from alternating parents.

Another gene found in both CHD and ASD oligogenic sets was *LRP1*, a large, commonly-mutated gene that encodes for the low-density lipoprotein (LDL) receptor-related protein, involved in Wnt signaling. While *LRP1* is known to be associated with ASD and CHD (Lillis et al. 2008; Torrico et al. 2019; Lin et al. 2020), we found 392 unaffected SFARI parents and siblings with one or more rare damaging *LRP1* variants, and therefore predict that *LRP1* is more likely to be an oligogenic modifier than a monogenic cause in any given case. Notably, 7 ASD probands had co-occurring variants in the digenic pair *HERC1-LRP1*, a gene combination that was never found in an unaffected sibling and observed only once in an unaffected mother (whose autistic child carries an additional *HERC1 de novo* mutation, **Figure 4C**). Many unaffected family members carry the same damaging *HERC1* mutation as the familial proband ( $n=51$  unaffected carriers), but the disease faithfully segregated seven times with the addition of a damaging *LRP1* variant. We therefore conclude that *LRP1* is likely to increase risk of developing ASD phenotypes, along with additional genetic modifiers determining final phenotypic outcome of *HERC1* variants.





**Figure 4. Signal of oligogenic inheritance in comorbidities of CHD.** **4A:** Venn diagram of overlap of genes in ASD oligogenic sets (purple) and CHD oligogenic sets (red). Of 2498 genes over-transmitted in combination with another gene in either CHD or ASD, 874 are common between the two diagnoses, an overlap that is significantly higher than expected from a random selection of the genome (hypergeometric  $p < 1E-27$ ). **4B, left:** Oligogenic pairs containing the *ARSG* gene in the CHD cohort (red) and the ASD cohort (purple), with edge widths representing the number of transmissions to probands for each pair. **4B, right:** Grid showing the provenance of variants contributing to an *ARSG*-\* oligogenic transmission for each proband. Cells are colored by the variant source (gold for a *de novo* mutation, green for maternal transmission, orange for paternal transmission, and pink for compound heterozygous transmission, i.e. one variant from each parent in the same gene). **4C, left:** A subset of the genes significantly over-transmitted with *HERC1*, with edge widths representing the number of transmissions to probands for each pair. Note: these counts represent the total number of oligogenic transmissions in the ASD cohort. **4C, right:** Grid showing the provenance of *HERC1-LRP1* digenic pair, as well as potential modifier genes *COL5A2* and *HSD17B14*. **4D:** Network visualization of genes significantly over-transmitted with *PKHD1*. Nodes are colored by GO membership to six biological categories and edges represent the number of probands with damaging variants in the *PKHD1*-\* pair. Genes are sorted by the cohort in which they were found to be significantly over-transmitted with *PKHD1*: from left to right, either the joint analysis of CHD and ASD probands with a congenital kidney diagnosis, solely the CHD cohort, or solely the ASD cohort.

We next examined genes associated with renal comorbidities in CHD and ASD. Of the 1468 CHD probands that were explicitly examined for renal abnormalities, 185 were determined to have a kidney phenotype. In ASD, 47 of 1748 probands have a recorded diagnosis of a kidney or urinary phenotype. Based on the hypothesis that renal abnormalities in CHD and ASD have the same underlying causal genes (albeit with additional phenotype-specific modifiers), we ran a

joint GCOD analysis with all 232 probands diagnosed with a kidney disease in either cohort (“kidney probands”). Candidate gene sets in kidney probands were further filtered by those with a logistic interaction coefficient greater than 1 to ensure that the combination of genes in these sets was more predictive of disease than any single constituent gene (**Methods: Logistic regression**). We found 925 oligogenic gene sets, consisting of 506 unique genes, which were significantly over-transmitted to renal probands and increased the odds of kidney disease when combined (**Supplemental Table 7**). Genes in these sets are enriched for functions in membrane transport, especially of calcium ions, and filament/motor activity (**Supplemental Table 8**). We recovered three CHD genes that were previously described to confer risk to kidney development, namely *LRP2*, *TMEM67*, *SLIT2* (Gabriel, Pazour, and Lo 2018), as well as 12 genes annotated with the process GO:0001822 (kidney development), namely *CAT*, *ADIPOQ*, *NOTCH3*, *WFS1*, *SULF1*, *COL4A3*, *FBN1*, *LRP4*, *DCHS2*, *COL4A4*, *PKHD1*, and *SMAD3*.

Previous studies have suggested ciliopathies as an underlying cause of renal-cardiac comorbidities (San Agustin et al. 2016; Gabriel, Pazour, and Lo 2018), and separately in autism and neuropsychiatric diagnoses (Louvi and Grove 2011; Marley and von Zastrow 2012; Guo et al. 2015). For example, bi-allelic *PKHD1* lesions are a known cause of polycystic kidney disease and heart defects, presumably related to its function in primary cilia in epithelial cells (Ward et al. 2002; Zhang et al. 2004; Chinali et al. 2019). However, we found that probands with rare damaging monoallelic *PKHD1* variants were less likely to have been diagnosed with a kidney phenotype than probands without a rare damaging *PKHD1* mutation (odds ratio = 0.55, Fisher exact  $p = 0.02$ ). This signal is driven mainly by a high proportion of CHD probands with a damaging *PKHD1* mutation but without a kidney diagnosis (54 of 59 *PKHD1*-carrying CHD probands). Combining the PCGC cohort with the SFARI autism cohort, GCOD found 55 genes significantly over-transmitted with *PKHD1* specifically to patients with renal complications, 22 of which were significantly over-transmitted as a digenic pair with *PKHD1* alone (**Figure 4D**, left). Once again we find that there are no overlapping oligogenic sets across each of the three

analyses (kidney cohort, CHD-only cohort, and ASD-only cohort). We conclude that *PKHD1* does indeed contribute to both cardiac and renal pathologies, but penetrance in the kidney appears to be modified by variants in other genes which largely differ between probands with comorbid kidney defects versus CHD-only or ASD-only. This is the first analysis to our knowledge that combines genetic data from two disparate phenotypes to increase detection power of a common comorbidity.

## Discussion

Developmental phenotypes affecting the heart and brain have profound impacts on the lives of patients and their families, and understanding the underlying cause in each case can reveal additional disease risks to monitor as the patient ages to adulthood. Evidence of oligogenic transmission of developmental disease has increased in recent years (Priest et al. 2016; Alsemari et al. 2018; Gifford et al. 2019; Mkaouar et al. 2021), necessitating a fast and accurate method to determine gene combinations of interest from patient data. We provide a computational framework called GCOD to test for the significance of oligogenic gene set transmission and report 202, 1023, and 922 significantly over-transmitted oligogenic pairs in CHD, ASD, and kidney disease respectively. These findings collectively contribute 2966 novel potential risk genes and modifiers from 7889 unique predicted-damaging higher-order sets across all three datasets. This software is unique compared to other recent developments in the field of oligogenic combination detection (Kerner et al. 2020; Pounraja and Girirajan 2022), because it uses family genotypes to limit statistical inference to gene sets with oligogenic transmission, thereby reducing computation and increasing our power to detect true phenotype associations.

Given that ten percent of probands carry an oligogenic set of predicted-damaging variants, our method potentially explains the genetic underpinnings of a substantial minority of cases with previously cryptic causal variants (**Table 1**). Experimental validation is required to determine how

many of these candidate variant sets are indeed sufficient to cause heart defects, or whether some level of additional background genetic risk is necessary. This is especially key as 97% of probands also carry combinations of less damaging variants that collectively confer some risk of CHD. A polygenic risk score approach (Wendt et al. 2020; Isgut et al. 2021) could be used to prioritize oligogenic sets that lead to disease even in otherwise relatively low-risk genomes.

We have established the concept of the “highest-order” gene set, which allows users to examine gene groups of three, four, and larger sets. This increases power by minimizing the number of individual tests. Such an analysis could be computationally prohibitive, but the GCOD algorithm efficiently leverages family sequencing data to filter out combinations seen in healthy parents. Parental sequencing also controls for the effects of population stratification, as gene set significance is conditional only on variant transmission from parents to offspring, and does not rely on relative frequencies in a population of potentially different ancestry.

Our method identified several protein complexes and functions associated with heart development, and further expanded these canonical sets to include potentially-interacting genes and risk modifiers. For example, oligogenic combinations including *SACS* hint at key protein interactions between the *SACS* filamental organizing protein with various myosin and membrane-related gene products. Exploring the function and interactions of such gene products could further elucidate cytoskeletal activity in the developing heart. Furthermore, genes that appear in CHD oligogenic sets are enriched for appropriate functions like atrioventricular canal development, as well as more novel associations like *de novo* pyrimidine synthesis and perception of light stimulus. As was recently explored by (Morton et al. 2021), individuals born with heart defects are at risk for additional complications later in life, including cancers. We similarly hypothesize a role for known bladder cancer gene *COL6A3* in CHD, and our results further indicate that eye and ear function in patients may be affected based on functional analysis. We recover several retinitis pigmentosa risk genes, suggesting future studies to determine if the eye health of CHD patients should be monitored as they age to adulthood.

The overlap of genes and functions related to cell motility and ECM signaling in both CHD and ASD further corroborates previous findings that the increased incidence of autism in CHD patients is partially due to pleiotropic effects of shared risk genes and is not solely attributable to impaired cardiac function, such as hypoxia affecting brain development (Bean Jaworski et al. 2017; Sigmon et al. 2019). Furthermore we have identified modifiers in CHD- and ASD-associated renal phenotypes, though we note that this analysis likely suffers from incomplete diagnostic information. For example, some probands without renal indications at birth or at the time of the study may develop symptoms later in life due to genetic risk, reducing our power. This makes the oligogenic sets we identified a conservative list of risk genes for congenital kidney abnormalities in CHD and ASD.

Our simulation method to assess significance of oligogenic sets relies on having sequencing data from enough families to observe multiple occurrences of rare gene combinations. However, we expect that clinicians working with one or just a few family pedigrees could make use of the first two steps of GCOD alone. Variant and gene combinations segregating with the disease but not observed in healthy family members can be enumerated to develop hypotheses for disease etiology where no clear Mendelian inheritance pattern exists.

GCOD as implemented here is not an exhaustive search for oligogenic sets in these families, as it does not take into account common variants (which are expected to confer less risk, but undoubtedly play a role in some cases). It is also important to note that kinship controls, while offering advantages in a conservative analysis, are likely not entirely unaffected by the risk variants they carry and may exhibit subclinical phenotypes. This is an especially important caveat in the SFARI autism cohort, as sets were filtered to those never observed in an unaffected sibling; however, the protein products of sibling gene sets might indeed interact and collaboratively cause disease contingent on other factors like sex or differences in environmental exposure between siblings. Finally, the method is more sensitive to *de novo* variation (given that an inherited event has a 50% probability of occurring, while most *de novo* events are exceedingly rare); results are

therefore biased towards discovering oligogenic sets where a non-synonymous *de novo* variant was observed.

Overall, we have created a framework to identify combinations of genes transmitted in CHD, ASD, and their renal comorbidities. We have provided researchers with these lists of prioritized, high-confidence gene and variant combinations for high-throughput screens or more precise mechanistic experiments, in order to move the field towards a more complete understanding of gene interactions during development, and how genetic variants combine in disease. GCOD is freely available for applications to other diseases and species.

## Methods

### Variant calling and filtering

Whole Exome Sequencing data from the 3382 CHD trios was processed as described in (Jin et al. 2017) and (Morton et al. 2021). Protein-coding mutations were filtered based on a Mapping Quality score  $> 59$  and Genotype Quality  $> 90$ . *De novo* variants were called using the TrioDeNovo program (Wei et al. 2015) and accepted if the minor allele frequency (MAF) and read-depth criteria described in (Homsy et al. 2015) are met. Namely, the in-cohort MAF of the variant must be below  $4 \times 10^{-4}$ , with a minimum of 5 alternative reads and 10 total reads in the proband, and a minimum of 10 reference reads in the parents (with a maximum alternate allele ratio of 3.5%). Variants in the SFARI autism cohort were called as described in (Krumm et al. 2015). In addition to the criteria above, we filtered this set for inherited variants with a Genotype Quality Mean  $> 85$  and plausible inheritance pattern (e.g. if a parent is 1/1 and a child is 0/0 without a *de novo* variant at this locus, the variant is disqualified). *De novo* variants for SFARI probands and unaffected siblings were provided by (Iossifov et al. 2014). All variants were called and are reported in GRCh37 coordinates.

For both datasets, we annotated variants using ANNOVAR version date 2021-06-08 (K. Wang, Li, and Hakonarson 2010). We restricted our analysis to variants annotated as <5% MAF in either the gnomAD (Karczewski et al. 2020) or ExAC non-psychiatric (Lek et al. 2016) databases, and below an in-cohort MAF of 10%. GCODE users can change these cutoffs where desired. Variants predicted to be damaging by at least one of MetaSVM, FATHMM, and SIFT models were considered in the Strict and Base-damaging tiers of variant severity (Liu et al. 2016; Rogers et al. 2017). Filtering criteria of variant severity is specified in Supplemental Table 1. We used gnomAD observed/expected (Karczewski et al. 2020),  $S_{het}$  (Zhu, Zhang, and Sha 2018), CADD scores (Rentzsch et al. 2019), and a minimum expression of TPM > 0.5 in any cardiac or brain cell type (for CHD and ASD genes, respectively) included in the DESCARTES developmental gene expression database (J. Cao et al. 2020). Cell types considered are listed below.

Cardiac cell types: 'Heart-Cardiomyocytes', 'Heart-CLC\_IL5RA positive cells', 'Heart-ELF3\_AGBL2 positive cells', 'Heart-Endocardial cells', 'Heart-Epicardial fat cells', 'Heart-Erythroblasts', 'Heart-Lymphatic endothelial cells', 'Heart-Lymphoid cells', 'Heart-Megakaryocytes', 'Heart-Myeloid cells', 'Heart-SATB2\_LRRC7 positive cells', 'Heart-Schwann cells', 'Heart-Smooth muscle cells', 'Heart-Stromal cells', 'Heart-Vascular endothelial cells', 'Heart-Visceral neurons.'

Brain cell types: 'Cerebellum-Astrocytes', 'Cerebellum-Granule neurons', 'Cerebellum-Inhibitory interneurons', 'Cerebellum-Microglia', 'Cerebellum-Oligodendrocytes', 'Cerebellum-Purkinje neurons', 'Cerebellum-SLC24A4\_PEX5L positive cells', 'Cerebellum-Unipolar brush cells', 'Cerebellum-Vascular endothelial cells', 'Cerebrum-Astrocytes', 'Cerebrum-Excitatory neurons', 'Cerebrum-Inhibitory neurons', 'Cerebrum-Limbic system neurons', 'Cerebrum-Megakaryocytes', 'Cerebrum-Microglia', 'Cerebrum-Oligodendrocytes', 'Cerebrum-SKOR2\_NPSR1 positive cells', 'Cerebrum-Vascular endothelial cells.'



## Pseudo-sibling genotype generation

Pseudo-sibling genotypes were generated from the two parental alleles not transmitted to the proband as described in (Yu and Deng 2011). In the rare case in which a proband *de novo* mutation occurred at a locus where a parent carried a qualifying variant, the pseudo-sibling inherited the parental rare allele if the proband did not. *De novo* variants in pseudo-siblings were randomly assigned to genes based on previously-derived protein-coding non-synonymous ('prot') mutability (Samocha et al. 2014). Since all such variants are automatically included at the Strict variant tier, GCOD does not predict specific amino acid substitutions to further qualify the *de novo* variant's CADD,  $S_{het}$ , or predicted-damaging status.

## Digenic pairs enumeration

In order to efficiently enumerate gene pairs in which multiple probands harbor co-occurring mutations at each tier of variant severity, we create a gene-by-family matrix where cell values indicate the presence of a variant meeting the relevant criteria. We use the following key:

0 = proband does not have a rare variant in this gene;

1 = proband harbors at least one variant at the Base tier of variant severity, and no variants in this gene at higher tiers;

2 = proband harbors at least one variant at the Base-damaging tier of variant severity, and no Strict-level variants in this gene;

3 = proband harbors at least one variant at the Strict tier of variant severity.

This can be repeated to create a separate matrix for (pseudo-)siblings. During generation of each gene-by-family matrix, we create three additional "provenance" matrices to indicate the inheritance of the variants meeting criteria at each of the three tiers. The provenance matrices indicate whether mother, father, both, or neither parents contributed a variant at or above this level of severity. These matrices are indexed during group enumeration to determine whether a

given combination of genes includes an oligogenically-inherited variant set (that is, the set of offspring variants in this gene combination are not also collectively carried by one of the unaffected parents).

Gene columns in the gene-by-family matrix were dropped if fewer than two probands (or siblings) carried a mutation at or above the relevant tier of severity. For remaining genes, all possible combinations of digenic pairs were enumerated and checked against the matrix for multiple co-occurrences at each severity tier. Note that each severity tier is inclusive of more-damaging variants, such that a variant meeting Strict criteria is included in Base and Base-damaging analyses. If the provenance matrix indicated an oligogenic transmission of a given set of variants in a digenic pair, and this was true of at least two probands (or siblings), the gene pair was considered a “candidate set.”

### **Highest-order gene set enumeration**

We defined the concept of a “highest-order” gene set, which comprises the largest unique candidate set observed in two or more probands (Supplemental Figure 1). To efficiently enumerate these sets, we begin with the previously-calculated digenic pairs and the list of offspring harboring them. A highest-order candidate set necessarily contains at least one qualifying digenic pair; therefore we reduce the computational search space by only examining pairs of offspring already known to carry a common digenic candidate. For every pair of probands sharing a digenic candidate, all additional damaging variant-harboring genes in common between the two probands comprise a highest-order set. In other words, the gene pair is expanded by adding all other shared genes. The final step is to check whether any lower-order sets for a given pair of probands is found in one or more additional probands. For example, Proband1 and Proband2 may have a highest-order set of *GeneA-GeneB-GeneC-GeneD*, while Proband1 and

Proband3 have a highest-order set of *GeneA-GeneB-GeneD-GeneE*. The initial algorithmic pass only recorded the following as highest order:

$$\begin{aligned} & \{ ( \textit{GeneA-GeneB-GeneC-GeneD} ) : [ \textit{Proband1, Proband2} ], \\ & ( \textit{GeneA-GeneB-GeneD-GeneE} ) : [ \textit{Proband1, Proband3} ], \\ & ( \textit{GeneA-GeneB-GeneD} ) : [ \textit{Proband 2, Proband3} ], \end{aligned}$$

where the true carriers of the latter set should be:

$$( \textit{GeneA-GeneB-GeneD} ) : [ \textit{Proband1, Proband2, Proband3} ].$$

GCOD therefore sorts the candidate sets by the number of participating genes, beginning with sets of two and extending to the largest highest-order set. GCOD sequentially checks whether a given highest-order candidate is a subset of another candidate, and if so, confirms oligogenic transmission of the variants in the additional probands only captured in the higher-order candidate set. Thus, since the *GeneA-GeneB-GeneD* combination is a subset of both *GeneA-GeneB-GeneC-GeneD* and *GeneA-GeneB-GeneD-GeneE* in the example above, we append Proband1 to the *GeneA-GeneB-GeneD* entry as long as the transmission of those variants in Proband1 meet criteria for oligogenic transmission.

## Logistic regression

Before simulation analysis, candidate sets are filtered based on their interaction coefficient in a logistic regression model. Disease status is predicted by  $n+1$  variables, where  $n$  is the number of genes in an oligogenic candidate set. For all parents and probands (and sequenced siblings where applicable), the presence/absence of a qualifying variant in each individual gene, as well as whether all genes in the combination harbor a variant in that individual, are denoted by 0/1. We use the binomial `glm()` function in R (version 3.6.1, (R Core Team 2020)) with 50 maximum iterations to test whether the coefficient of the gene interaction is greater than 1. This indicates

that the combination itself is broadly associated with disease in this dataset, and it removes spurious combinations that are driven by a single gene within the set.

### **Oligogenic transmission simulation test**

For each family, parental genotypes at loci participating in candidate sets are aggregated and used to simulate 10,000 random offspring. One of each allele from father and mother is randomly selected to comprise the simulated offspring genotype at each locus, and additional *de novo* mutations are simulated based on previously-derived mutability constants (Samocha et al. 2014). A gene-by-family matrix and parental provenance matrices (as described above in “Digenic pairs enumeration”) is created for each of the 10,000 simulated cohorts. For each oligogenic candidate set, we index the constituent genes and record which simulated offspring carry that combination, contingent on oligogenic transmission. For each candidate set, this generates a distribution of co-occurrence counts based on parental genotypes under the null hypothesis of no disease association (i.e., random transmission). The raw p-value is determined by the proportion of iterations with a simulated count as large or larger than the proband test statistic. GCOD returns the test statistic (number of probands observed with the oligogenic combination) and raw, Benjamini-Hochberg FDR-corrected, and Bonferroni-corrected p-values for each candidate set. In this study, we use a significance threshold of Bonferroni-corrected p-value less than 0.05.

### **Single-gene transmission simulation test**

From the 10,000 simulated gene-by-family matrices described in the “Oligogenic transmission simulation test” above, we additionally create marginal null distributions of the number of probands harboring variants in each gene and use these to compute p-values for individual genes. Significant p-values indicate the binary presence of any qualifying variant in a gene is significantly higher in probands than randomly simulated offspring. Note that this is distinct from a traditional

Transmission Disequilibrium Test in that the count of binary presence/absence across probands is the test statistic here, instead of all variant transmissions within the given gene.

## Gene Ontology analyses

We performed Gene Ontology (GO) analyses using the GOATOOLS package in Python3 (version 1.1.6, (Klopfenstein et al. 2018)), using the GOEnrichmentStudyNS() function with default settings (alpha = 0.05, correction method = fdr\_bh for Benjamini-Hochberg FDR-adjusted p-values). When reporting top enriched terms, we focused on GO categories with more than five gene members because small categories are prone to false positives.

## Oligogenic set network discovery and depiction

We selected specific oligogenic sets in CHD probands based on criteria detailed below. To visualize gene network nodes represent genes and are colored according to functional annotations. Edges indicate that genes carry damaging variants in the same proband, with edge width representing the number of probands with co-occurring mutations in that gene pair. Note that the edge counts can include probands in which the gene pair was inherited from one parent (non-oligogenic transmission), under the condition that a higher-order oligogenic set includes that transmission (i.e., at least one variant in another gene in the higher-order set was inherited from the other parent or mutated *de novo*). Edges are not drawn between nodes unless the two genes appear together in at least one significant oligogenic set. Oligogenic sets are reported in the accompanying text, as well as **Supplemental Table 2** for the full cohort and **Supplemental Table 4** for groups specific to a particular diagnosis. For brevity, only select significant sets of interest are included in the provenance matrices in these supplemental tables.

For Figure 3B, we discovered two oligogenic sets in which genes are known to physically interact in a canonical protein complex. We selected all other significant oligogenic sets containing at least one gene in these complexes, and visualized them using genes for nodes and co-occurrences as edges. For Figure 3D, we sought an oligogenic set with several counts of oligogenic transmissions but rarely any variant combinations seen in unaffected parents, discovering the *MYO18B-SACS* combination transmitted oligogenically 10 out of 12 times. We additionally incorporated genes appearing in significant oligogenic sets in any of these 12 probands. Figure 4A shows all significant sets containing *ARSG* in the CHD and ASD cohorts (6 pairs), and Figure 4B shows all significant sets containing *HERC1* in the ASD cohort (3 pairs).

## Acknowledgements

We thank the participants and research clinicians of the Pediatric Cardiac Genomics Consortium (PCGC) and Simons Foundation Autism Research Initiative. We thank Dr. Sarah Morton of the Harvard Medical School Division of Newborn Medicine for providing data and processing pipelines from the PCGC. CHD genomic data were generated by the PCGC, under the auspices of the National Heart, Lung, and Blood Institute's Bench to Bassinet Program <https://benthtobassinet.com>. This study was funded by NIH/NHLBI (grants P01 HL098707, P01 HL146366, and UM1 HL098179), Additional Ventures, and the L.K. Whittier Foundation. M.P and K.L were supported by NIH/NHLBI awards F31HL156439 and F32HL158263, respectively.

## Data and software access

PCGC variants are available under dbGaP Study Accession: phs000571.v6.p2 for qualifying researchers. Access to SFARI ASD variants can be requested at <https://base.sfari.org/>. GCOD software, including Jupyter notebooks with detailed analyses and use cases, is provided at <https://github.com/mepittman/gcod>.

## Competing interest statement

The authors have no competing interests to report.

## References

- 1000 Genomes Project Consortium, Adam Auton, Lisa D. Brooks, Richard M. Durbin, Erik P. Garrison, Hyun Min Kang, Jan O. Korb, et al. 2015. "A Global Reference for Human Genetic Variation." *Nature* 526 (7571): 68–74.
- Abad-Morales, Víctor, Rafael Navarro, Anniken Burés-Jelstrup, and Esther Pomares. 2020. "Identification of a Novel Homozygous ARSG Mutation as the Second Cause of Usher Syndrome Type 4." *American Journal of Ophthalmology Case Reports* 19 (September): 100736.
- Aikawa, Yukiko, Lan Anh Nguyen, Kyoichi Isono, Nobuyuki Takakura, Yusuke Tagata, M. Lienhard Schmitz, Haruhiko Koseki, and Issay Kitabayashi. 2006. "Roles of HIPK1 and HIPK2 in AML1- and p300-Dependent Transcription, Hematopoiesis and Blood Vessel Formation." *The EMBO Journal* 25 (17): 3955–65.
- Akhirome, Ehirole, Nephi A. Walton, Julie M. Noguee, and Patrick Y. Jay. 2017. "The Complex Genetic Basis of Congenital Heart Defects." *Circulation Journal: Official Journal of the Japanese Circulation Society* 81 (5): 629–34.
- Alsemari, Abdulaziz, Mohammed Alsuhaibani, Rawabi Alhathloul, and Bayan Mamdouh Ali. 2018. "Potential Oligogenic Disease of Mental Retardation, Short Stature, Spastic Paraparesis, and Osteopetrosis." *The Application of Clinical Genetics* 11 (November): 129–34.
- Anitha, A., Kazuhiko Nakamura, Kazuo Yamada, Shiro Suda, Ismail Thanseem, Masatsugu Tsujii, Yoshimi Iwayama, et al. 2008. "Genetic Analyses of Roundabout (ROBO) Axon Guidance Receptors in Autism." *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics* 147B (7): 1019–27.
- Azhar, Mohamad, Jo El J. Schultz, Ingrid Grupp, Gerald W. Dorn 2nd, Pierre Meneton, Daniel G. M.

- Molin, Adriana C. Gittenberger-de Groot, and Thomas Doetschman. 2003. "Transforming Growth Factor Beta in Cardiovascular Development and Function." *Cytokine & Growth Factor Reviews* 14 (5): 391–407.
- Bean Jaworski, Jessica L., Thomas Flynn, Nancy Burnham, Jesse L. Chittams, Therese Sammarco, Marsha Gerdes, Judy C. Bernbaum, et al. 2017. "Rates of Autism and Potential Risk Factors in Children with Congenital Heart Defects." *Congenital Heart Disease* 12 (4): 421–29.
- Bosch, Elena, Hafid Laayouni, Carlos Morcillo-Suarez, Ferran Casals, Andrés Moreno-Estrada, Anna Ferrer-Admetlla, Michelle Gardner, et al. 2009. "Decay of Linkage Disequilibrium within Genes across HGDP-CEPH Human Samples: Most Population Isolates Do Not Show Increased LD." *BMC Genomics* 10 (July): 338.
- Cao, Junyue, Diana R. O'Day, Hannah A. Pliner, Paul D. Kingsley, Mei Deng, Riza M. Daza, Michael A. Zager, et al. 2020. "A Human Cell Atlas of Fetal Gene Expression." *Science* 370 (6518). <https://doi.org/10.1126/science.aba7721>.
- Cao, Si-Yuan, Hui-Liang Shen, Lun Luo, Shu-Jie Chen, and Chunguang Li. 2021. "PCNet: A Structure Similarity Enhancement Method for Multispectral and Multimodal Image Registration." *arXiv [cs.CV]*. arXiv. <http://arxiv.org/abs/2106.05124>.
- Cassa, Christopher A., Donat Weyhorn, Daniel J. Balick, Daniel M. Jordan, David Nusinow, Kaitlin E. Samocha, Anne O'Donnell-Luria, et al. 2017. "Estimating the Selective Effects of Heterozygous Protein-Truncating Variants from Human Exome Data." *Nature Genetics* 49 (5): 806–10.
- Chan, H. M., and N. B. La Thangue. 2001. "p300/CBP Proteins: HATs for Transcriptional Bridges and Scaffolds." *Journal of Cell Science* 114 (Pt 13): 2363–73.
- Chinali, Marcello, Laura Lucchetti, Agnese Ricotta, Claudia Esposito, Carolina D'Anna, Gabriele Rinelli, Francesco Emma, and Laura Massella. 2019. "Cardiac Abnormalities in Children with Autosomal Recessive Polycystic Kidney Disease." *Cardiorenal Medicine* 9 (3): 180–89.
- Clothier, Joanna, and Michael Absoud. 2021. "Autism Spectrum Disorder and Kidney Disease." *Pediatric Nephrology* 36 (10): 2987–95.
- Dankel, Simon N., Elise Grytten, Jan-Inge Bjune, Hans Jørgen Nielsen, Arne Dietrich, Matthias Blüher, Jørn V. Sagen, and Gunnar Mellgren. 2020. "COL6A3 Expression in Adipose Tissue Cells Is



- Associated with Levels of the Homeobox Transcription Factor PRRX1." *Scientific Reports* 10 (1): 20164.
- Feng, Dan, Jason T. Christensen, Anji T. Yetman, Merry L. Lindsey, Amar B. Singh, and Jeffrey D. Salomon. 2021. "The Microbiome's Relationship with Congenital Heart Disease: More than a Gut Feeling." *Journal of Congenital Cardiology* 5 (1): 1–11.
- Fischbach, Gerald D., and Catherine Lord. 2010. "The Simons Simplex Collection: A Resource for Identification of Autism Genetic Risk Factors." *Neuron* 68 (2): 192–95.
- Gabriel, George C., Gregory J. Pazour, and Cecilia W. Lo. 2018. "Congenital Heart Defects and Ciliopathies Associated With Renal Phenotypes." *Frontiers in Pediatrics* 6 (June): 175.
- Gentil, Benoit J., Gia-Thanh Lai, Marie Menade, Roxanne Larivière, Sandra Minotti, Kalle Gehring, J-Paul Chapple, Bernard Brais, and Heather D. Durham. 2019. "Sacsin, Mutated in the Ataxia ARSACS, Regulates Intermediate Filament Assembly and Dynamics." *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology* 33 (2): 2982–94.
- Gifford, Casey A., Sanjeev S. Ranade, Ryan Samarakoon, Hazel T. Salunga, T. Yvanka de Soysa, Yu Huang, Ping Zhou, et al. 2019. "Oligogenic Inheritance of a Human Heart Disease Involving a Genetic Modifier." *Science* 364 (6443): 865–70.
- Giurgiu, Madalina, Julian Reinhard, Barbara Brauner, Irmtraud Dunger-Kaltenbach, Gisela Fobo, Goar Frishman, Corinna Montrone, and Andreas Ruepp. 2019. "CORUM: The Comprehensive Resource of Mammalian Protein Complexes-2019." *Nucleic Acids Research* 47 (D1): D559–63.
- Guo, Jiami, Holden Higginbotham, Jingjun Li, Jackie Nichols, Josua Hirt, Vladimir Ghukasyan, and E. S. Anton. 2015. "Developmental Disruptions Underlying Brain Abnormalities in Ciliopathies." *Nature Communications* 6 (July): 7857.
- He, Zongxiao, Di Zhang, Alan E. Renton, Biao Li, Linhai Zhao, Gao T. Wang, Alison M. Goate, Richard Mayeux, and Suzanne M. Leal. 2017. "The Rare-Variant Generalized Disequilibrium Test for Association Analysis of Nuclear and Extended Pedigrees with Application to Alzheimer Disease WGS Data." *American Journal of Human Genetics* 100 (2): 193–204.
- Hezroni, Hadas, Rotem Ben-Tov Perry, Noa Gil, Neta Degani, and Igor Ulitsky. 2020. "Regulation of Neuronal Commitment in Mouse Embryonic Stem Cells by the *Reno1/Bahcc1* Locus." *EMBO*

*Reports* 21 (11): e51264.

Hoang, Thanh T., Elizabeth Goldmuntz, Amy E. Roberts, Wendy K. Chung, Jennie K. Kline, John E.

Deanfield, Alessandro Giardini, et al. 2018. "The Congenital Heart Disease Genetic Network Study: Cohort Description." *PloS One* 13 (1): e0191319.

Homsy, Jason, Samir Zaidi, Yufeng Shen, James S. Ware, Kaitlin E. Samocha, Konrad J. Karczewski,

Steven R. DePalma, et al. 2015. "De Novo Mutations in Congenital Heart Disease with Neurodevelopmental and Other Congenital Anomalies." *Science* 350 (6265): 1262–66.

Huang, Yan, Gang Li, Kai Wang, Zhongyi Mu, Qingpeng Xie, Hongchen Qu, Hang Lv, and Bin Hu. 2018.

"Collagen Type VI Alpha 3 Chain Promotes Epithelial-Mesenchymal Transition in Bladder Cancer Cells via Transforming Growth Factor  $\beta$  (TGF- $\beta$ )/Smad Pathway." *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research* 24 (August): 5346–54.

Iossifov, Ivan, Brian J. O’Roak, Stephan J. Sanders, Michael Ronemus, Niklas Krumm, Dan Levy, Holly

A. Stessman, et al. 2014. "The Contribution of de Novo Coding Mutations to Autism Spectrum Disorder." *Nature* 515 (7526): 216–21.

Isgut, Monica, Jimeng Sun, Arshed A. Quyyumi, and Greg Gibson. 2021. "Highly Elevated Polygenic Risk

Scores Are Better Predictors of Myocardial Infarction Risk Early in Life than Later." *Genome Medicine* 13 (1): 13.

Janknecht, R., N. J. Wells, and T. Hunter. 1998. "TGF-Beta-Stimulated Cooperation of Smad Proteins

with the Coactivators CBP/p300." *Genes & Development* 12 (14): 2114–19.

Jin, Sheng Chih, Jason Homsy, Samir Zaidi, Qiongshi Lu, Sarah Morton, Steven R. DePalma, Xue Zeng,

et al. 2017. "Contribution of Rare Inherited and de Novo Variants in 2,871 Congenital Heart Disease Probands." *Nature Genetics* 49 (11): 1593–1601.

Ji, Weizhen, Dina Ferdman, Joshua Copel, Dustin Scheinost, Veronika Shabanova, Martina Brueckner,

Mustafa K. Khokha, and Laura R. Ment. 2020. "De Novo Damaging Variants Associated with Congenital Heart Diseases Contribute to the Connectome." *Scientific Reports* 10 (1): 7046.

Karczewski, Konrad J., Laurent C. Francioli, Grace Tiao, Beryl B. Cummings, Jessica Alföldi, Qingbo

Wang, Ryan L. Collins, et al. 2020. "The Mutational Constraint Spectrum Quantified from Variation in 141,456 Humans." *Nature* 581 (7809): 434–43.

- Kasah, Sahrnunizam, Christopher Oddy, and M. Albert Basson. 2018. "Autism-Linked CHD Gene Expression Patterns during Development Predict Multi-Organ Disease Phenotypes." *Journal of Anatomy* 233 (6): 755–69.
- Kerner, Gaspard, Matthieu Bouaziz, Aurélie Cobat, Benedetta Bigio, Andrew T. Timberlake, Jacinta Bustamante, Richard P. Lifton, Jean-Laurent Casanova, and Laurent Abel. 2020. "A Genome-Wide Case-Only Test for the Detection of Digenic Inheritance in Human Exomes." *Proceedings of the National Academy of Sciences of the United States of America* 117 (32): 19367–75.
- Khateb, Samer, Björn Kowalewski, Nicola Bedoni, Markus Damme, Netta Pollack, Ann Saada, Alexey Obolensky, et al. 2018. "A Homozygous Founder Missense Variant in Arylsulfatase G Abolishes Its Enzymatic Activity Causing Atypical Usher Syndrome in Humans." *Genetics in Medicine: Official Journal of the American College of Medical Genetics* 20 (9): 1004–12.
- Klopfenstein, D. V., Liangsheng Zhang, Brent S. Pedersen, Fidel Ramírez, Alex Warwick Vesztrocy, Aurélien Naldi, Christopher J. Mungall, et al. 2018. "GOATOOLS: A Python Library for Gene Ontology Analyses." *Scientific Reports* 8 (1): 10872.
- Kousi, Maria, and Nicholas Katsanis. 2015. "Genetic Modifiers and Oligogenic Inheritance." *Cold Spring Harbor Perspectives in Medicine* 5 (6). <https://doi.org/10.1101/cshperspect.a017145>.
- Krumm, Niklas, Tychele N. Turner, Carl Baker, Laura Vives, Kiana Mohajeri, Kali Witherspoon, Archana Raja, et al. 2015. "Excess of Rare, Inherited Truncating Mutations in Autism." *Nature Genetics* 47 (6): 582–88.
- Lek, Monkol, Konrad J. Karczewski, Eric V. Minikel, Kaitlin E. Samocha, Eric Banks, Timothy Fennell, Anne H. O'Donnell-Luria, et al. 2016. "Analysis of Protein-Coding Genetic Variation in 60,706 Humans." *Nature* 536 (7616): 285–91.
- Lillis, Anna P., Lauren B. Van Duyn, Joanne E. Murphy-Ullrich, and Dudley K. Strickland. 2008. "LDL Receptor-Related Protein 1: Unique Tissue-Specific Functions Revealed by Selective Gene Knockout Studies." *Physiological Reviews* 88 (3): 887–918.
- Lin, Jiuann-Huey I., Timothy N. Feinstein, Anupma Jha, Jacob T. McCleary, Juan Xu, Angelo B. Arrigo, Grace Rong, et al. 2020. "Mutation of LRP1 in Cardiac Neural Crest Cells Causes Congenital Heart Defects by Perturbing Outflow Lengthening." *Communications Biology* 3 (1): 1–12.

- Liu, Xiaoming, Chunlei Wu, Chang Li, and Eric Boerwinkle. 2016. "dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs." *Human Mutation* 37 (3): 235–41.
- Louvi, Angeliki, and Elizabeth A. Grove. 2011. "Cilia in the CNS: The Quiet Organelle Claims Center Stage." *Neuron* 69 (6): 1046–60.
- Mahdavi, V., M. Periasamy, and B. Nadal-Ginard. 1982. "Molecular Characterization of Two Myosin Heavy Chain Genes Expressed in the Adult Heart." *Nature* 297 (5868): 659–64.
- Marino, Bradley S., Paul H. Lipkin, Jane W. Newburger, Georgina Peacock, Marsha Gerdes, J. William Gaynor, Kathleen A. Mussatto, et al. 2012. "Neurodevelopmental Outcomes in Children with Congenital Heart Disease: Evaluation and Management: A Scientific Statement from the American Heart Association." *Circulation* 126 (9): 1143–72.
- Marley, Aaron, and Mark von Zastrow. 2012. "A Simple Cell-Based Assay Reveals That Diverse Neuropsychiatric Risk Genes Converge on Primary Cilia." *PloS One* 7 (10): e46647.
- Marschang, Peter, Jochen Brich, Edwin J. Weeber, J. David Sweatt, John M. Shelton, James A. Richardson, Robert E. Hammer, and Joachim Herz. 2004. "Normal Development and Fertility of Knockout Mice Lacking the Tumor Suppressor Gene LRP1b Suggest Functional Compensation by LRP1." *Molecular and Cellular Biology* 24 (9): 3782–93.
- McCarthy, Ann L., Madeline E. Winters, David R. Busch, Ernesto Gonzalez-Giraldo, Tiffany S. Ko, Jennifer M. Lynch, Peter J. Schwab, et al. 2015. "Scoring System for Periventricular Leukomalacia in Infants with Congenital Heart Disease." *Pediatric Research* 78 (3): 304–9.
- Mkaouar, Rahma, Lamia Cherif Ben Abdallah, Chokri Naouali, Saida Lahbib, Zinet Turki, Sahar Elouej, Yosra Bouyacoub, et al. 2021. "Oligogenic Inheritance Underlying Incomplete Penetrance of PROKR2 Mutations in Hypogonadotropic Hypogonadism." *Frontiers in Genetics* 12 (September): 665174.
- Morton, Sarah U., Akiko Shimamura, Peter E. Newburger, Alexander R. Opatowsky, Daniel Quiat, Alexandre C. Pereira, Sheng Chih Jin, et al. 2021. "Association of Damaging Variants in Genes With Increased Cancer Risk Among Patients With Congenital Heart Disease." *JAMA Cardiology* 6 (4): 457–62.

- Niesler, Beate, and Gudrun A. Rappold. 2021. "Emerging Evidence for Gene Mutations Driving Both Brain and Gut Dysfunction in Autism Spectrum Disorder." *Molecular Psychiatry* 26 (5): 1442–44.
- Parker, Lauren E., and Andrew P. Landstrom. 2021. "Genetic Etiology of Left-Sided Obstructive Heart Lesions: A Story in Development." *Journal of the American Heart Association* 10 (2): e019006.
- Pediatric Cardiac Genomics Consortium, Bruce Gelb, Martina Brueckner, Wendy Chung, Elizabeth Goldmuntz, Jonathan Kaltman, Juan Pablo Kaski, et al. 2013. "The Congenital Heart Disease Genetic Network Study: Rationale, Design, and Early Results." *Circulation Research* 112 (4): 698–706.
- Pounraja, Vijay Kumar, and Santhosh Girirajan. 2022. "A General Framework for Identifying Oligogenic Combinations of Rare Variants in Complex Disorders." *Genome Research*, March. <https://doi.org/10.1101/gr.276348.121>.
- Priest, James R., Kazutoyo Osoegawa, Nebil Mohammed, Vivek Nanda, Ramendra Kundu, Kathleen Schultz, Edward J. Lammer, et al. 2016. "De Novo and Rare Variants at Multiple Loci Support the Oligogenic Origins of Atrioventricular Septal Heart Defects." *PLoS Genetics* 12 (4): e1005963.
- R Core Team. 2020. "R: A Language and Environment for Statistical Computing." Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Renaux, Alexandre, Sofia Papadimitriou, Nassim Versbraegen, Charlotte Nachtegaele, Simon Boutry, Ann Nowé, Guillaume Smits, and Tom Lenaerts. 2019. "ORVAL: A Novel Platform for the Prediction and Exploration of Disease-Causing Oligogenic Variant Combinations." *Nucleic Acids Research* 47 (W1): W93–98.
- Rentzsch, Philipp, Daniela Witten, Gregory M. Cooper, Jay Shendure, and Martin Kircher. 2019. "CADD: Predicting the Deleteriousness of Variants throughout the Human Genome." *Nucleic Acids Research* 47 (D1): D886–94.
- Rogers, Mark F., Hashem A. Shihab, Matthew Mort, David N. Cooper, Tom R. Gaunt, and Colin Campbell. 2017. "FATHMM-XF: Accurate Prediction of Pathogenic Point Mutations via Extended Features." *Bioinformatics* 34 (3): 511–13.
- Samocha, Kaitlin E., Elise B. Robinson, Stephan J. Sanders, Christine Stevens, Aniko Sabo, Lauren M. McGrath, Jack A. Kosmicki, et al. 2014. "A Framework for the Interpretation of de Novo Mutation in

- Human Disease.” *Nature Genetics* 46 (9): 944–50.
- San Agustin, Jovenal T., Nikolai Klena, Kristi Granath, Ashok Panigrahy, Eileen Stewart, William Devine, Lara Strittmatter, et al. 2016. “Genetic Link between Renal Birth Defects and Congenital Heart Disease.” *Nature Communications* 7 (March): 11103.
- Schaaf, Christian P., Aniko Sabo, Yasunari Sakai, Jacy Crosby, Donna Muzny, Alicia Hawes, Lora Lewis, et al. 2011. “Oligogenic Heterozygosity in Individuals with High-Functioning Autism Spectrum Disorders.” *Human Molecular Genetics* 20 (17): 3366–75.
- Sener, Elif Funda, Merve Cıkılı Uytun, Keziban Korkmaz Bayramov, Gokmen Zararsiz, Didem Behice Oztop, Halit Canatan, and Yusuf Ozkul. 2016. “The Roles of CC2D1A and HTR1A Gene Expressions in Autism Spectrum Disorders.” *Metabolic Brain Disease* 31 (3): 613–19.
- Sigmon, Eric R., Michael Kelleman, Apryl Susi, Cade M. Nylund, and Matthew E. Oster. 2019. “Congenital Heart Disease and Autism: A Case-Control Study.” *Pediatrics* 144 (5).  
<https://doi.org/10.1542/peds.2018-4114>.
- Spielman, R. S., R. E. McGinnis, and W. J. Ewens. 1993. “Transmission Test for Linkage Disequilibrium: The Insulin Gene Region and Insulin-Dependent Diabetes Mellitus (IDDM).” *American Journal of Human Genetics* 52 (3): 506–16.
- Torrico, Bàrbara, Alex D. Shaw, Roberto Mosca, Norma Vivó-Luque, Amaia Hervás, Noèlia Fernández-Castillo, Patrick Aloy, et al. 2019. “Truncating Variant Burden in High-Functioning Autism and Pleiotropic Effects of LRP1 across Psychiatric Phenotypes.” *Journal of Psychiatry & Neuroscience: JPN* 44 (5): 350–59.
- Veenstra-Vanderweele, Jeremy, Susan L. Christian, and Edwin H. Cook Jr. 2004. “Autism as a Paradigmatic Complex Genetic Disorder.” *Annual Review of Genomics and Human Genetics* 5: 379–405.
- Wang, Kai, Mingyao Li, and Hakon Hakonarson. 2010. “ANNOVAR: Functional Annotation of Genetic Variants from High-Throughput Sequencing Data.” *Nucleic Acids Research* 38 (16): e164.
- Wang, Qinqin, Chunmei Wang, Bingyuan Ji, Jiawei Zhou, Chunqing Yang, and Jing Chen. 2019. “Hapln2 in Neurological Diseases and Its Potential as Therapeutic Target.” *Frontiers in Aging Neuroscience* 11 (March): 60.

- Wang, Siying, Wen-Mei Yu, Wanming Zhang, Keith R. McCrae, Benjamin G. Neel, and Cheng-Kui Qu. 2009. "Noonan Syndrome/leukemia-Associated Gain-of-Function Mutations in SHP-2 Phosphatase (PTPN11) Enhance Cell Migration and Angiogenesis." *The Journal of Biological Chemistry* 284 (2): 913–20.
- Ward, Christopher J., Marie C. Hogan, Sandro Rossetti, Denise Walker, Tam Sneddon, Xiaofang Wang, Vicky Kubly, et al. 2002. "The Gene Mutated in Autosomal Recessive Polycystic Kidney Disease Encodes a Large, Receptor-like Protein." *Nature Genetics* 30 (3): 259–69.
- Warkman, Andrew S., Samantha A. Whitman, Melanie K. Miller, Robert J. Garriock, Catherine M. Schwach, Carol C. Gregorio, and Paul A. Krieg. 2012. "Developmental Expression and Cardiac Transcriptional Regulation of Myh7b, a Third Myosin Heavy Chain in the Vertebrate Heart." *Cytoskeleton* 69 (5): 324–35.
- Wei, Qiang, Xiaowei Zhan, Xue Zhong, Yongzhuang Liu, Yujun Han, Wei Chen, and Bingshan Li. 2015. "A Bayesian Framework for de Novo Mutation Calling in Parents-Offspring Trios." *Bioinformatics* 31 (9): 1375–81.
- Wendt, Frank R., Carolina Muniz Carvalho, Gita A. Pathak, Joel Gelernter, and Renato Polimanti. 2020. "Polygenic Risk for Autism Spectrum Disorder Associates with Anger Recognition in a Neurodevelopment-Focused Phenome-Wide Scan of Unaffected Youths from a Population-Based Cohort." *PLoS Genetics* 16 (9): e1009036.
- Wenger, Tara L., Charly Kao, Donna M. McDonald-McGinn, Elaine H. Zackai, Alice Bailey, Robert T. Schultz, Bernice E. Morrow, Beverly S. Emanuel, and Hakon Hakonarson. 2016. "The Role of mGluR Copy Number Variation in Genetic and Environmental Forms of Syndromic Autism Spectrum Disorder." *Scientific Reports* 6 (January): 19372.
- Wernovsky, Gil, and Daniel J. Licht. 2016. "Neurodevelopmental Outcomes in Children With Congenital Heart Disease-What Can We Impact?" *Pediatric Critical Care Medicine: A Journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies* 17 (8 Suppl 1): S232–42.
- Xu, Mingjie, Jie Yao, Yingchao Shi, Huijuan Yi, Wukui Zhao, Xinhua Lin, and Zhongzhou Yang. 2021. "The SRCAP Chromatin Remodeling Complex Promotes Oxidative Metabolism during Prenatal Heart

Development.” *Development* 148 (8). <https://doi.org/10.1242/dev.199026>.

Yu, Zhaoxia, and Li Deng. 2011. “Pseudosibship Methods in the Case-Parents Design.” *Statistics in Medicine* 30 (27): 3236–51.

Zhang, Ming-Zhi, Weiyi Mai, Cunxi Li, Sae-Youll Cho, Chuanming Hao, Gilbert Moeckel, Runxiang Zhao, et al. 2004. “PKHD1 Protein Encoded by the Gene for Autosomal Recessive Polycystic Kidney Disease Associates with Basal Bodies and Primary Cilia in Renal Epithelial Cells.” *Proceedings of the National Academy of Sciences of the United States of America* 101 (8): 2311–16.

Zhu, Huanhuan, Shuanglin Zhang, and Qiuying Sha. 2018. “A Novel Method to Test Associations between a Weighted Combination of Phenotypes and Genetic Variants.” *PloS One* 13 (1): e0190788.