

# 1 **Genome Sequencing is Critical for Forecasting Outcomes** 2 **following Congenital Cardiac Surgery**

3  
4  
5 W. Scott Watkins<sup>1</sup>, E. Javier Hernandez<sup>2</sup>, Thomas A. Miller<sup>3</sup>, Nathan R. Blue<sup>4</sup>, Raquel  
6 Zimmerman<sup>2</sup>, Eric R. Griffiths<sup>5</sup>, Erwin Frise<sup>6</sup>, Daniel Bernstein<sup>7</sup>, Marko T. Boskovski<sup>8</sup>,  
7 Martina Brueckner<sup>9,10</sup>, Wendy K. Chung<sup>11</sup>, J. William Gaynor<sup>12</sup>, Bruce D. Gelb<sup>13</sup>, Elizabeth  
8 Goldmuntz<sup>14</sup>, Peter J. Gruber<sup>15</sup>, Jane W. Newburger<sup>16</sup>, Amy E. Roberts<sup>16</sup>, Sarah U. Morton<sup>17</sup>,  
9 John E. Mayer Jr<sup>18</sup> Christine E. Seidman<sup>19,20</sup>, Jonathan G. Seidman<sup>19</sup>, Yufeng Shen<sup>21</sup>,  
10 Michael Wagner<sup>22</sup>, H. Joseph Yost<sup>23</sup>, Mark Yandell<sup>1\*</sup>, and Martin Tristani-Firouzi<sup>24,\*</sup>

11  
12 <sup>1</sup>Department of Human Genetics, University of Utah, Salt Lake City, UT, 84112, USA.

13 <sup>2</sup>Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, 84108,  
14 USA.

15 <sup>3</sup>Department of Pediatrics, Maine Medical Center, Portland, ME, USA.

16 <sup>4</sup>Department of Obstetrics and Gynecology, University of Utah, Salt Lake City, UT, 84112,  
17 USA;

18 <sup>5</sup>Pediatric Cardiothoracic Surgery, University of Utah, Salt Lake City, Utah, USA

19 <sup>6</sup>Fabric Genomics Inc., Oakland, CA, 94612, USA

20 <sup>7</sup>Department of Pediatrics, Stanford University School of Medicine, Palo Alto, CA 94305,  
21 USA.

22 <sup>8</sup>Department of Surgery, University of California, San Francisco, CA 94143, USA

23 <sup>9</sup>Department of Pediatrics, Yale University School of Medicine, New Haven, CT, 06416,  
24 USA.

25 <sup>10</sup>Department of Genetics, Yale University School of Medicine, New Haven, CT, 06416,  
26 USA.

27 <sup>11</sup>Departments of Pediatrics, Boston Children's Hospital, Harvard Medical School, Boston,  
28 MA 02115, USA.

29 <sup>12</sup>Division of Cardiothoracic Surgery, Children's Hospital of Philadelphia, Philadelphia, PA,  
30 USA.

31 <sup>13</sup>Mindich Child Health and Development Institute, Departments of Pediatrics and Genetics  
32 & Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029,  
33 USA.

34 <sup>14</sup>Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania,  
35 Philadelphia, Pennsylvania, USA.

36 <sup>15</sup>Department of Surgery, Yale University, New Haven, CT, USA.

37 <sup>16</sup>Department of Cardiology, Boston Children's Hospital, and Department of Pediatrics,  
38 Harvard Medical School, Boston, MA 02115, USA.

39 <sup>17</sup>Division of Newborn Medicine, Department of Medicine, Boston Children's Hospital,  
40 Boston, MA, USA

41 <sup>18</sup>Department of Cardiovascular Surgery, Boston Children's Hospital, Boston, MA 02115,  
42 USA

43 <sup>19</sup>Departments of Genetics and Medicine, Harvard Medical School, Boston 02115, USA.

44 <sup>20</sup>Howard Hughes Medical Institute, Harvard Medical School, Boston 02115, USA.

45 <sup>21</sup>Departments of Systems Biology and Biomedical Informatics, Columbia University, New  
46 York, NY 10032, USA.

47 <sup>22</sup>Division of Biomedical Informatics, Department of Pediatrics, Cincinnati Children's  
48 Hospital Medical Center, Cincinnati, OH, USA.

49 <sup>23</sup>Molecular Medicine Program, University of Utah, Salt Lake City, UT 84112, USA.

50 <sup>24</sup>Nora Eccles Harrison Cardiovascular Research and Training Institute, and Division of  
51 Pediatric Cardiology, Salt Lake City, Utah, 84108, USA.

52  
53 *email: scott.watkins@genetics.utah.edu, edgarh@genetics.utah.edu,*  
54 *Thomas.A.Miller@mainehealth.org, Nathan.Blue@hsc.utah.edu,*  
55 *RAQUEL.REISINGER@hsc.utah.edu, Eric.Griffiths@hsc.utah.edu,*  
56 *efrise@fabricgenomics.com, danb@stanford.edu, marko.boskovski@ucsf.edu,*  
57 *martina.brueckner@yale.edu, wendy.chung@childrens.harvard.edu,*  
58 *GAYNOR@chop.edu, bruce.gelb@mssm.edu, peter.gruber@yale.edu,*  
59 *GOLDMUNTZ@email.chop.edu, jane.newburger@cardio.chboston.org,*  
60 *Amy.Roberts@cardio.chboston.org, sarah.morton@childrens.harvard.edu,*  
61 *John.Mayer@cardio.chboston.org, cseidman@genetics.med.harvard.edu,*  
62 *seidman@genetics.med.harvard.edu, ys2411@columbia.edu, Michael.Wagner@cchmc.org,*  
63 *jyost@genetics.utah.edu, myandell@genetics.utah.edu, and Martin.Tristani@utah.edu*

64  
65 \*Corresponding authors

## 66 **Abstract**

67 While genome sequencing has transformed medicine by elucidating the genetic underpinnings  
68 of both rare and common complex disorders, its utility to predict clinical outcomes remains  
69 understudied. Here, we used artificial intelligence (AI) technologies to explore the predictive  
70 value of genome sequencing in forecasting clinical outcomes following surgery for congenital  
71 heart defects (CHD). We report results for a cohort of 2,253 CHD patients from the Pediatric  
72 Cardiac Genomics Consortium with a broad range of complex heart defects, pre- and post-  
73 operative clinical variables and exome sequencing. Damaging genotypes in chromatin-  
74 modifying and cilia-related genes were associated with an elevated risk of adverse post-  
75 operative outcomes, including mortality, cardiac arrest and prolonged mechanical ventilation.  
76 The impact of damaging genotypes was further amplified in the context of specific CHD  
77 phenotypes, surgical complexity and extra-cardiac anomalies. The absence of a damaging  
78 genotype in chromatin-modifying and cilia-related genes was also informative, reducing the risk  
79 for adverse postoperative outcomes. Thus, genome sequencing enriches the ability to forecast  
80 outcomes following congenital cardiac surgery.

81

## 82 **Introduction**

83 Congenital heart defects (CHD) represent a complex class of often life-threatening disorders that  
84 affect more than 40,000 newborns in the U.S. annually. The prevalence of CHD is approximately  
85 1 per 100 live births, with an incidence that varies according to the specific CHD lesion<sup>1-3</sup>. The  
86 genetic architecture of CHD has been the focus of several large-scale sequencing efforts<sup>4-9</sup>,  
87 demonstrating that the genetic landscape of syndromic and sporadic CHD differ, with sporadic  
88 forms characterized by considerable locus and allelic heterogeneity<sup>7</sup>. More recently, work by the  
89 National Heart, Lung and Blood Institute (NHLBI)-funded Pediatric Cardiac Genomics  
90 Consortium (PCGC) has shown that dominantly and recessively inherited forms of CHD have  
91 distinct genetic and phenotypic landscapes, whereby dominant forms of CHD are significantly  
92 enriched for damaging variants in chromatin-modifying genes, while recessive forms are enriched  
93 for damaging variants in cilia-related biallelic genotypes and heterotaxy phenotypes<sup>4,5,8,9</sup>.

94       Recent work has also demonstrated the value of genetic testing for outcomes prediction  
95 for specific types of CHD and within specific clinical contexts<sup>10-14</sup>. Broader investigations,  
96 however, have faced difficulties in assaying genetic contributions across multiple CHD  
97 phenotypes and clinical contexts, in part due to the widely varying severity of CHD lesions and  
98 the complex medical and surgical interventions necessary for survival. Here, we demonstrate that  
99 condensing heterogenous CHD phenotypes into five major clinically relevant phenotypic  
100 categories using anatomic descriptors<sup>15</sup> renders these data amenable for outcomes analyses. We  
101 also show that the high allelic and locus heterogeneity characteristic of CHD can be overcome  
102 using an artificial intelligence (AI) genome interpretation tool<sup>16</sup>, followed by categorization of  
103 damaging genotypes into molecular pathways or gene categories. This two-pronged approach of  
104 phenotypic and genotypic classification, when combined with probabilistic graphical models,

105 enables clinically relevant and highly personalized risk estimates in patients undergoing congenital  
106 cardiac surgery.

## 107 **Methods**

108 **Human subjects.** All patients were diagnosed, phenotyped, and recruited by PCGC centers and  
109 participating regional hospitals into the PCGC Congenital Heart Disease Network Study (CHD  
110 GENES: ClinicalTrials.gov identifier NCT01196182; [<https://clinicaltrials.gov/>]). Informed  
111 consent was obtained from all participants or the participants' guardians. Approval for human  
112 subjects research was obtained by the institutional review boards of participating centers,  
113 including Boston's Children's Hospital, Brigham and Women's Hospital, Great Ormond Street  
114 Hospital, Children's Hospital of Los Angeles, Children's Hospital of Philadelphia, Columbia University  
115 Medical Center, Icahn School of Medicine at Mount Sinai, Rochester School of Medicine and  
116 Dentistry, Steven and Alexandra Cohen Children's Medical Center of New York, Lucile Packard  
117 Children's Hospital Stanford, University of California-San Francisco, University of Utah, and Yale  
118 School of Medicine. Automated CHD phenotype classification was performed on 14,765 PCGC  
119 participants. A subset of these participants who had both exome sequencing and perioperative  
120 data (2,253) was used for network analyses.

121  
122 **Clinical phenotypes.** Cardiac diagnoses were obtained from review of echocardiogram, cardiac  
123 MRI, catheterization, and operative reports at the time of enrollment into the PCGC<sup>4,5</sup>. Detailed  
124 cardiac diagnoses for each patient were coded using the Fyler system<sup>15</sup>. Extra-cardiac anomalies  
125 (ECAs) were identified at the time of PCGC enrollment<sup>4,5</sup> (Supplemental Table 1). Any  
126 structural anomaly that was not acquired was classified as an extra cardiac anomaly (ECA).

127

128 **Post-operative variables.** For patients undergoing open heart surgery, surgical and  
129 hospitalization data were obtained from participating centers using the local data collected for  
130 submission to the Society of Thoracic Surgeons Congenital Heart Surgery Database (STS-  
131 CHSD)<sup>17</sup>. A total of 59 surgical complication variables were extracted for analysis. The size of  
132 the final data set was constrained to 2,253 patients, such that all patients had WES and surgical  
133 variables had no more than 10% missing data. Most patients had multiple cardiac surgeries. A  
134 patient was scored as having an adverse event or surgical complication (e.g., prolonged  
135 mechanical ventilation) if that event occurred for any surgery at any age (Supplemental Table  
136 2).

137  
138 **Surgical complexity.** Surgical complexity is a well-known driver of mortality and morbidity.  
139 In response, the STS-European Association for Cardio-Thoracic Surgery (STAT) has created  
140 risk assessment categories in which procedures are grouped based on similar mortality rates<sup>18</sup>.  
141 STAT categories range from 1 to 5, with STAT1 representing the procedures with the lowest  
142 mortality rates and STAT5 representing the procedures with the highest mortality rates.

143  
144 **CHD classification.** The PCGC has classified cardiac diagnoses for over 14,000 CHD probands  
145 using the Fyler coding system, which describes the congenitally malformed heart using a  
146 vocabulary of more than 3,000 possible phenotypic descriptors<sup>15</sup>. While this system allows for  
147 highly granular descriptions of heart defects, we hypothesized that condensing these terms into a few  
148 clinically relevant phenotypic categories might render them more tractable for outcomes analyses.  
149 Thus, we sought to automate cardiac phenotype classification across the entire PCGC cohort,  
150 assigning each patient to a single category. To do so, we used five major cardiac categories derived  
151 from a previous PCGC study<sup>5</sup>: left ventricular outflow tract obstructions (LVO), laterality and

152 heterotaxy defects (HTX), atrioventricular canal defects (AVC), conotruncal defects (CTD), and  
153 other defects (OTH), which includes simple atrial septal defects and more complex heart defects  
154 not assigned to the other four categories<sup>5,15</sup>. Each participant was assigned uniquely to one of the  
155 five phenotypic categories.

156 A gradient-boosted decision tree model was built to automatically classify PCGC  
157 probands (14,765) into one of these five CHD categories (see Supplemental Methods,  
158 Supplemental Figures 1, 2). Model learning and classification was performed using an ensemble-  
159 based method using the XGBoost library (v1.5)<sup>19</sup>. We created an analysis toolkit to streamline  
160 XGBoost training, grid-based parameter optimization, and performance evaluation. The truth set  
161 for training the classifier included 3,000 CHD patients, 2,752 PCGC patients previously assigned  
162 into the five CHD categories<sup>5</sup> and 248 randomly selected PCGC patients that were manually  
163 reviewed and assigned to a CHD phenotype category. A gradient-boosted probabilistic patient  
164 classifier was built with XGBoost<sup>19</sup> using the 3,000 patients and 698 Fyler features  
165 (Supplemental Table 3). Model training was performed with five-fold cross-validation and  
166 replacement subsampling (Supplemental Table 4). Model accuracy was assessed by comparing  
167 the patient's known phenotype category label derived from the literature<sup>5</sup> to its predicted  
168 phenotype category label (Supplemental Figure 3, Supplemental Table 5). Single-class  
169 prediction accuracy for the training data was higher for HTX (98.9%), AVC (97.8%), and LVO  
170 (97.7%) than for CTD (95.3%) and OTH (91.9%), where a low level of ambiguity occurred.  
171 Overall classification accuracy was 97.7% with a specificity of 99.3%, and sensitivity of  
172 97.7% (Supplemental Tables 6). We then applied the trained classifier to 14,765 PCGC CHD  
173 patients with Fyler descriptors (Supplemental Tables 7, 8). Final classification of the 3,000  
174 training patients was identical to the original training predictions. The five phenotype categories  
175 remained generally proportional between the training data and the full data set, with a maximum

176 observed difference of 4.2% for CTD patients.

177

178 **AI-based scoring of predicted damaging genetic variants.** AI-based identification of  
179 candidate disease-causing genotypes was performed using Fabric GEM<sup>16</sup> (Fabric Genomics,  
180 Oakland, CA). GEM incorporates Human Phenotype Ontology (HPO) terms, sex, genotype  
181 frequency (gnomAD), evolutionary conservation, Online Mendelian Inheritance in Man  
182 (OMIM), GnomAD, and ClinVar information in a probabilistic AI framework to identify the  
183 most likely genetic variant or genotype that explains the patient's disease phenotype. Because  
184 WES are difficult substrates for CNV calling, we restricted our analyses to SNVs and short  
185 indels. HPO terms utilized in the GEM analyses were based on each patient's Fyler phenotypes,  
186 which were mapped to HPO terms using the Clinithink software package (Clinithink, London).

187 GEM's gene scores are  $\log_{10}$  transformed Bayes factors<sup>20</sup> that summarize the relative  
188 support for the hypothesis that the prioritized genotype damages the gene in which it resides and  
189 explains the patient's phenotype versus the hypothesis that the variant neither damages the gene  
190 nor explains the patient's phenotype. We used a stringent GEM score of  $\geq 1.0$  to represent a  
191 likely pathogenic genotype. A recent genomic analysis of critically ill newborns<sup>16</sup> showed that  
192 a GEM score of  $\geq 1.0$  identified 90% of all true positive damaging variants, with a median of  
193 two candidate variants per patient<sup>16</sup>. Gene penetrance for GEM calculations was set to 0.95 to  
194 enforce strict consideration of known dominant and recessive disorders. For downstream  
195 analyses, damaging genetic variants were classified as *de novo*, dominant, or recessive/biallelic  
196 variants based on their inheritance pattern in trios. Dominant and *de novo* damaging variants  
197 were required to have a frequency of  $< 1/10000$  in gnomAD databases (v2.1, v3.1) and most  
198 variants were not observed. Overall, we identified damaging *de novo* or recessive genotypes in  
199 10.56% of the study cohort (Supplemental Tables 9, 10), in line with previous studies that utilized



200 different methods of defining pathogenicity<sup>4,5,8,9</sup>. Damaging genetic variants were assigned to  
201 several functional gene pathways. Gene lists for gene pathways were obtained using the  
202 reactome pathway browser. Gene lists are shown in Supplemental Table 11 and have been  
203 previously described<sup>5,8,9</sup> There is overlap between gene lists, with some genes represented in  
204 more than one gene pathway/category (Supplemental Figure 4).

205  
206 **Probabilistic graphical models.** Probabilistic graphical models (PGMs) provide a robust  
207 explainable AI methodology capable of discovering and quantifying additive and synergistic  
208 effects amongst broad classes of variables. For the work presented here, we used a form of PGMs  
209 known as a Bayesian networks<sup>21</sup>. Bayesian networks are fully transparent, and their graphical  
210 representation offers an intuitive and visual mechanism for understanding the relationships  
211 between variables and the impacts of multiple variables on outcomes of interest<sup>22-29</sup>. Moreover,  
212 Bayesian networks offer practical advantages over regression approaches, by capturing the entire  
213 joint probability distribution of the data, encompassing all interrelationships among the variables  
214 incorporated in the non-linear model. Thus, a single network can be used to explore any  
215 combination of variables as a target outcome in one query and then as a risk factor for a different  
216 target outcome, all within the same model. For more on these points see <sup>21,30,31</sup>.

217  
218 **Feature selection.** Single variables, such as damaging genetic variants in chromatin-modifying  
219 genes, were tested for conditional dependency with phenotype variables using exact Bayesian  
220 Networks. Each conditional probability [e.g., the probability of LVO given a damaging *de novo*  
221 chromatin variant:  $P(\text{LVO} \mid \text{chromatin } dGV)$ ] was estimated as the median conditional  
222 probability from 1000 independent networks. Conditional probability estimates were divided by  
223 the baseline probability for each respective phenotype to obtain absolute risk ratios. Associations

224 with absolute risk ratios  $\geq 1.0$  were selected for further analyses. Surgery-related variables  
225 associated with gene categories and CHD phenotypes were identified in a similar way. Each  
226 surgical feature was tested individually as a conditional variable with genetic variables and each  
227 of the five specific CHD phenotypes. All conditional variables, individually or as composite  
228 variables (e.g., mortality and ECA), were required to have at least six events.

229  
230 **Network construction.** Bayesian networks were created for each CHD phenotype category.  
231 Each network included genetic and surgical variables identified in the feature selection stage.  
232 All-cause mortality was also included in each network. All input conditional variables were coded  
233 as presence/absence. A small number of missing surgical values ( $< 10\%$  for any single variable)  
234 was imputed using a K-nearest neighbors approach ( $k = 10$ ). The structure of each network was  
235 learned with the Silander-Myllymaki exact algorithm with Bayesian information criterion (BIC)  
236 scoring<sup>32</sup> or by a greedy hill-climbing method for the large networks with more than 15 nodes.  
237 Posterior probabilities were network propagated using exact inference. Network structure  
238 learning and belief propagation were performed with the bnstruct and gRain R packages<sup>33,34</sup>.  
239 To improve convenience and functionality, we created the BayesNetExplorer.jl package which  
240 implements network structure learning and belief propagation methods and provides tools for  
241 feature selection, risk estimation, graphics, and other network analysis tasks.

242  
243 **Risk Calculations.** We used two risk ratios (RR) to summarize our risk estimates, absolute and  
244 relative RR. For example, the absolute RR of a LVO phenotype given a damaging genotype in  
245 a chromatin-modifying gene is as follows:  $Absolute\ RR_{LVO|dGVchromatin} =$   
246  $\frac{P(LVO = true | dGV\ chromatin=true)}{P(LVO = true)}$ . The relative RR estimates the relative change in mortality risk  
247 for LVO patients with damaging mutations in chromatin-modifying genes, compared to similar

248 patients without a damaging chromatin genotype:  $Relative RR_{mortality + LVO | dGV chromatin} =$   
249  $\frac{P(mortality = true, LVO = true | dGV chromatin = true)}{P(mortality = true, LVO = true | GV chromatin = false)}$ . Final risk ratios and their confidence intervals  
250 are reported as the median and 95% confidence interval from an empirical distribution of risk ratio  
251 estimates. The empirical distributions are created by randomly resampling the data set with  
252 replacement and recreating 1000 independent networks and risk estimates. A Laplace correction  
253 ( $k = 1$ ) or network smoothing value was used to prevent zero-state probability estimates during  
254 bootstrapping.

255

## 256 **Results**

257 **Refining the genetic architecture of CHD.** For genetic and outcomes analysis, the study  
258 population consisted of 2,253 PCGC probands (1992 trios, 12 duos, 245 singletons) with both  
259 exome sequencing and surgical outcomes data, classified into five CHD phenotype categories.  
260 The AI-based genome analysis tool GEM<sup>16</sup> identified predicted damaging *de novo* genotypes in  
261 238 participants (10.6% of the cohort). A total of 131 damaging *de novo* / dominant and 198  
262 damaging recessive/biallelic genetic variants were discovered (Supplemental Tables 9, 10).  
263 There were 17 genes with damaging *de novo* variants in two or more patients. The most  
264 commonly recurrent *de novo* variants were in known CHD-related genes such as *KMT2D* (11),  
265 *CHD7* (6), *RAF1* (3), *JAG1* (3), and *TAB2* (3). Biallelic damaging genotypes were observed in  
266 multiple patients for several genes including *DYNC2H1* (3), *DNAH5* (3), *LAMA2* (3), *GDF1* (2),  
267 and *IFT140* (2).

268 We discovered that CHD phenotype categories were enriched for damaging  
269 genetic variants in specific gene pathways/categories (Table 1). For example, the LVO class  
270 was enriched 1.61-fold (CI 1.41-1.81) for damaging *de novo* genotypes in chromatin-modifying

271 genes, with this signal driven primarily by patients with hypoplastic left heart syndrome  
272 (HLHS), a subset of LVO where these genotypes were enriched 1.92-fold (CI 1.32-2.52). While  
273 previous studies implicated damaging chromatin-modifying gene variants in CHD cohorts at  
274 large<sup>4,5,8,9</sup>, our analyses here help to define the specific CHD subtypes most influenced by  
275 damaging variants in chromatin-modification genes. The LVO phenotype class was also  
276 enriched for *de novo* genotypes in *WNT* genes (2.13-fold, CI 1.86-2.40), signal transduction  
277 genes (1.51-fold, CI 1.01-2.01), and a curated list of genes known to cause CHD (1.22-fold, CI  
278 1.11-1.33; see Supplemental Tables 9 and 10). Notably, damaging genotypes in these pathways  
279 were not enriched in HLHS patients, further underscoring the complex genetic landscape  
280 underlying CHD.

281 The HTX phenotype class was enriched for damaging recessive/biallelic variants in cilia-  
282 related genes (2.63-fold, CI 2.06-3.20) and showed proportionally higher enrichment in the  
283 subset of motile cilia genes modulated by *FOXJ1* (6.89-fold, 3.30-10.36), findings consistent  
284 with previous reports<sup>5,8,35</sup>. The OTHER phenotype class was enriched for damaging *de novo*  
285 variants in chromatin-modifying genes (1.85-fold, CI 1.44-2.26) and the curated CHD genes  
286 (1.59-fold, CI 1.38-1.80) lists. We were underpowered to detect enrichment in damaging  
287 genotypes in the AVC and CTD phenotype classes.

288

### 289 **Damaging genotypes impact surgical outcomes.**

290 To further explore the relationships between genetic and clinical variables, we utilized Bayesian  
291 networks, a powerful statistical framework that can model complex dependencies, including  
292 non-linear relationships and indirect associations, in a probabilistic manner. In Bayesian  
293 networks, variables are depicted as nodes in a graph and conditional dependencies between  
294 variables are represented by the edges connecting those nodes. Once a network is constructed,

295 the impact of any combination variables on any selected outcome can be quantified, while  
296 controlling for the effects of other variables incorporated into the network. The networks  
297 describing the conditional dependencies between damaging genotypes, CHD phenotypes and  
298 post-operative variables are shown in Figure 1a, b.

299 Damaging genotypes in chromatin-modifying and cilia-related genes (defined by a  
300 GEM<sup>16</sup> score  $\geq 1.0$ ) increased the probability of severe adverse clinical outcomes following  
301 congenital cardiac surgery, including mortality, cardiac arrest, and prolonged mechanical  
302 ventilation ( $> 7$  days post-surgery). For example, damaging *de novo* chromatin genetic variants  
303 increased the probability (relative risk) of mortality 1.81-fold (CI 1.50-3.21), cardiac arrest 1.74-  
304 fold (CI 1.40-2.94) and prolonged ventilation 1.65-fold (CI 1.41-2.27). Likewise, damaging  
305 recessive/biallelic cilia genotypes increased the probability of mortality 1.40-fold (CI 1.09-  
306 2.10), cardiac arrest 1.50-fold (CI 1.13-2.34) and prolonged ventilation 1.43-fold (CI 1.09-1.96).  
307 Reciprocally, the absence of a damaging genotype was protective for these adverse post-  
308 operative outcomes. Thus, for a proband without a damaging *de novo* chromatin genotype, the  
309 relative risk ratio for mortality was 0.55 (CI 0.31-0.69), 0.58 (CI 0.34-0.72) for cardiac arrest  
310 and 0.61 (CI 0.44-0.72) for prolonged ventilation. For a proband without at damaging  
311 recessive/biallelic cilia genotype, the relative risk ratio for mortality was 0.72 (CI 0.48-0.91),  
312 0.63 (CI 0.43-0.88) for cardiac arrest and 0.70 (CI 0.51-0.92) for prolonged ventilation.

313

### 314 **Damaging genotypes impact surgical outcomes in the context of surgical mortality risk** 315 **category.**

316 We discovered that damaging chromatin and cilia genotypes were associated with an  
317 increased risk of mortality for probands undergoing the highest risk surgical procedures (Figure  
318 1c). Thus, probands who died after a STAT4 or STAT5 surgical procedure were 1.80-fold (CI

319 1.47-4.13) more likely to harbor a damaging chromatin variant. Similarly, those who died after  
320 a STAT4 surgery were 1.73-fold (CI 1.24-2.47) more likely to harbor a damaging  
321 recessive/biallelic cilia genotype (Figure 1). Damaging chromatin and cilia genotypes were  
322 overrepresented in probands experiencing cardiac arrest or prolonged mechanical ventilation  
323 following the most complex surgical procedures (Figure 1c).

324

### 325 **Damaging genotypes impact surgical outcomes in the context of CHD phenotypes**

326 More broadly, considering mortality in the context of CHD phenotypes, LVO patients  
327 who died were 2.26-fold (CI 1.22-2.94) more likely to harbor a damaging *de novo* chromatin  
328 genotype, while HTX patients who died were 2.80-fold (CI 2.45-3.15) more likely to harbor a  
329 damaging recessive/biallelic cilia genotype (Figure 1c). Similarly, damaging chromatin or cilia  
330 genotypes were overrepresented in probands with LVO, HLHS and HTX who experienced  
331 cardiac arrest or prolonged post-operative ventilation (Figure 1c). Specifically, HTX patients  
332 who arrested post-operatively were 3.34-fold (CI 1.44-5.49) more likely to harbor a damaging  
333 recessive/biallelic cilia genotype, compared to similar patients without a damaging cilia  
334 genotype. Collectively, these findings demonstrate that genome sequencing data are critical for  
335 predicting severe post-operative events in the context of specific CHD phenotypes and the  
336 highest risk congenital heart surgeries.

337

### 338 **Damaging genotypes impact surgical outcomes in the context of extracardiac phenotypes.**

339 Given the recognized impact of ECAs on outcomes following congenital cardiac  
340 surgery<sup>10,14,36</sup>, we also explored the relationship between ECAs and adverse post-operative  
341 outcomes in the context of genotypes and CHD phenotypes. ECAs increased the probability

342 (relative risk) of mortality 2.85-fold (CI 1.47-2.86) and prolonged ventilation 1.72-fold (CI 1.63-  
343 1.72) following congenital cardiac surgery. Consistent with previous findings<sup>10</sup>, damaging *de*  
344 *novo* genetic variants, including *de novo* variants in chromatin-modifying genes, were enriched in  
345 probands with ECAs 1.47-fold (CI 1.44-1.48) and 2.09-fold (CI 2.08-2.11), respectively. By  
346 contrast, damaging recessive/biallelic cilia genotypes were not enriched in probands with ECAs  
347 (0.84-fold; CI 0.85-1.13).

348 We also examined reciprocal effects, i.e., impact of predicted damaging genotypes on  
349 ECAs and adverse outcomes. For example, probands with damaging *de novo* chromatin genotypes  
350 identified by GEM were 2.49-fold (CI 2.17-4.99) more likely to have an ECA and die, compared  
351 to probands without a damaging chromatin genotype, and 2.44-fold (CI 1.97-3.41) more likely to  
352 have an ECA and prolonged ventilation (Figure 2). Moreover, a damaging recessive/biallelic cilia  
353 genotype identified by GEM increased the probability of mortality in probands with an ECA 1.48-  
354 fold (CI 1.02-2.85), compared to similar probands without a damaging cilia genotype, and  
355 increased the probability of prolonged ventilation in the presence of an ECA 1.52-fold (CI 1.06-  
356 2.51). Additionally, a damaging cilia genotype increased the probability of prolonged ventilation  
357 in HTX patients with an ECA 4.01-fold (CI 1.67-10.6), compared to similar patients without a  
358 damaging cilia genotype (Figure 2). Taken together, these findings demonstrate that damaging  
359 genotypes in chromatin and cilia genes predict severe post-operative events in the setting of  
360 ECAs.

361 The number of probands experiencing adverse outcomes and harboring damaging gene  
362 pathway variants in the AVC, CTD, and OTHER categories was too low to warrant generation of  
363 Bayesian networks for outcomes prediction in these CHD phenotypes. Consequently, larger  
364 cohorts are necessary to adequately predict the impact of genetics on outcomes for these  
365 CHD phenotypes. However, damaging genotypes in a number of gene pathways/categories,

366 such as FOXP1-controlled genes, high murine heart expression genes (HHE), WNT signaling  
367 genes, NOTCH signaling genes, and genes in a curated CHD gene list were predictive of  
368 mortality for the most complex surgical categories (Figure 3). Damaging genotypes in signal  
369 transduction and TGF- $\beta$  pathways were not predictive of mortality. Taken together, these  
370 findings highlight the value of genomic data for predicting adverse outcomes following  
371 congenital cardiac surgery, especially in the context of CHD phenotypes, ECAs and surgical  
372 complexity.

373

## 374 **Discussion**

375

376 Assessing the impact of genetics on patient outcomes in CHD is complicated by the intrinsic  
377 severity of the cardiac lesion, the complex medical and surgical interventions necessary for  
378 survival, and the high degree of phenotypic, locus and allelic heterogeneity. The NHBLI-funded  
379 PCGC is one of the world's largest collections of genetic, phenotypic, and clinical variables for  
380 CHD and thus provides an excellent resource for exploring the utility of genomics data for  
381 outcomes prediction. In this study, we implemented an explainable AI-based analysis  
382 framework to automatically classify CHD patients into phenotype categories and identify  
383 damaging genetic variants and genotypes. This approach allowed us to explore how damaging  
384 genotypes impact outcomes following congenital cardiac surgery, in the context of specific CHD  
385 phenotypes, ECAs, and surgical complexity, providing precise risk estimates for specific clinical  
386 contexts.

387 *De novo* variants associated with CHD have been shown to be enriched in genes related  
388 to chromatin regulation<sup>4,5,8,9</sup>. Our results identify LVO lesions and confirm HLHS as a principal  
389 driver of the chromatin signal in this cohort. HLHS is one of the most severe forms of CHD and



390 associated with substantial morbidity and mortality. Our results show that the subset of HLHS  
391 patients with damaging genetic variants in chromatin genes has even greater risk (up to 2.57-  
392 fold) for severe post-operative outcomes in the context of the most complex surgical procedures.

393 Our findings also reinforce previous studies showing that damaging recessive/biallelic  
394 genotypes in cilia-related genes are overrepresented in the heterotaxy/laterality phenotype  
395 category<sup>5,8</sup>. Our results here demonstrate the additional utility of genetic findings for outcomes  
396 predictions. Damaging recessive/biallelic cilia genotypes increase the risk of severe adverse  
397 post-operative outcomes in the context of surgical complexity, HTX phenotype and the presence  
398 of an ECA. For example, damaging recessive/biallelic cilia genotype substantially increase  
399 (4.01-fold) the risk of prolonged ventilation for HTX patients with an ECA. These findings are  
400 consistent with an emerging body of literature implicating cilia dysfunction, HTX, and  
401 respiratory complications following congenital cardiac surgery<sup>37,38</sup>.

402 Established and emerging literature has highlighted the impact of genetics on mortality  
403 and other adverse outcomes following congenital cardiac surgery, mostly focusing on the impact  
404 of copy number variants.<sup>10-14</sup> Damaging *de novo* genic variants were associated with worse  
405 transplant-free survival and longer times to final extubation in a previously reported subset of  
406 the PCGC cohort (n = 1268)<sup>10</sup>. Here, we expand upon these findings in the largest study to date  
407 relating genotypes to CHD surgical outcomes. Our analyses reveal that damaging genotypes in  
408 specific gene pathways/categories impact post-operative outcomes across CHD phenotypic  
409 categories in specific and quantifiable ways.

410 Our AI approach allowed us to unravel the conditional dependencies among diverse  
411 clinical and genetic variables and to discover their impacts, either in isolation or in combination,  
412 on post-operative outcomes. These findings define a critical role for genome sequencing in  
413 outcomes prediction for congenital cardiac surgeries, especially in the context of higher risk

414 surgical procedures, specific CHD phenotypes and ECAs. Importantly, the absence of damaging  
415 genotypes was protective for adverse outcomes following congenital cardiac surgery. Thus,  
416 genomic information is informative whether or not a proband has an identified damaging  
417 genotype.

418 Nevertheless, there are limitations inherent to this study. For example, the PCGC  
419 population is not an inception cohort and thus is likely depleted for genetic lesions that  
420 predispose to early death, meaning our morbidity estimates are likely lower bounds. Although  
421 the PCGC cohort reflects a broad spectrum of CHD, recruitment of severe CHD forms was  
422 favored, leaving us under-powered to investigate the impact of genome sequencing for less  
423 severe CHD phenotypes. Additionally, while large clinical registries, such as the STS database,  
424 are invaluable resources for outcomes research, these databases, despite the inclusion of auditing  
425 features, may suffer from data quality issues, variability in the abstraction of data, batch effects,  
426 and missing data<sup>39-41</sup> that might impact the interpretation of the results presented here. Finally,  
427 we do not yet have access to an independent validation cohort with genomic data and similar  
428 clinical variables.

429 Looking to the future, a more complete description of the genetic and outcomes  
430 landscape of CHD could be enabled through clinical genome sequencing of CHD patients at  
431 even greater scales, together with initiatives by major consortia to collect and distribute genomic  
432 and clinical data more broadly. Given the rapid decline in costs, the increasing availability and  
433 quick turn-a-round time, genome sequencing is now poised to become the standard of care for  
434 all critically ill newborns.<sup>42,43</sup> Our findings make it clear that genome sequencing of all  
435 newborns with complex CHD will empower personalized risk-stratification for outcomes  
436 following congenital cardiac surgery.

437

## 438 **Acknowledgments**

439  
440 We thank the patients and their families for participating in the Pediatric Cardiac Genomics  
441 Consortium and Bench to Bassinet research programs. This research would not be possible  
442 without the clinical professionals involved in patient recruitment, and we thank those at the  
443 following institutions including the Columbia Medical School: D. Awad, C. Breton, K. Celia, C.  
444 Duarte, D. Etwaru, N. Fishman, E. Griffin, M. Kaspakoval, J. Kline, R. Korsin, A. Lanz, E.  
445 Marquez, D. Queen, A. Rodriguez, J. Rose, J.K. Sond, D. Warburton, A. Wilpers, and R. Yee; 2)  
446 the Children’s Hospital of Los Angeles: J. Ellashek and N. Tran; 3) the Children’s Hospital of  
447 Philadelphia: S. Edman, J. Garbarini, J. Tusi, and S. Woyciechowski; 4) the Harvard Medical  
448 School: J. Geva and M. Borensztein; 5) the Icahn School of Medicine at Mount Sinai: A. Julian,  
449 M. Mac Neal, Y. Mendez, T. Mendiz-Ramdeen, and C. Mintz; 6) the University College London:  
450 B. McDonough, K. Flack, L. Panesar, and N. Taylor; 7) the University of Rochester School of  
451 Medicine and Dentistry: E. Taillie; and 8) the Yale School of Medicine: N. Cross. We thank  
452 Nick Felicelli, Prakash Velayutham, and the staff at the Cincinnati Children’s Hospital Medical  
453 Center for computational support and for providing access to the HeartsMart database. We thank  
454 Carson Holt, Shawn Rynearson, Barry Moore at the Utah Center for Genomic Discovery and the  
455 staff at the Utah Center for High Performance Computing for high-throughput processing of  
456 patient sequence data.

457  
458 The Pediatric Cardiac Genomics Consortium (PCGC) program is funded by the National Heart,  
459 Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human  
460 Services through grants UM1HL128711, UM1HL098162, UM1HL098147, UM1HL098123,  
461 UM1HL128761, U01-HL098153, U01-HL098163, and U01HL131003. This manuscript was  
462 prepared in collaboration with investigators of the PCGC and has been reviewed and/or  
463 approved by the PCGC. PCGC investigators are listed at  
464 [\[https://benchtobassinet.com/?page\\_id=119\]](https://benchtobassinet.com/?page_id=119)

## 465 **Data availability**

466 Genetic and phenotypic data used in the paper are provided in the supplementary tables. Exome  
467 sequencing data have been deposited in the database of Genotypes and Phenotypes (dbGaP)  
468 under accession numbers [phs000571.v1.p1](#), [phs000571.v2.p1](#) and [phs000571.v3.p2](#).

469

## 470 **Online resources**

471 <https://heartsmart.pcgcid.org/>

472 <https://github.com/dmlc/XGBoost.jl>

473 <https://github.com/ScottWatkins/BayesNetExplorer>

474 <https://julialang.org/>

475 <https://www.r-project.org/>

476 <https://reactome.org>

477

## 478 **Author contributions**

479 W.S.W., E.J.H., M.Y., and M.T.F. conceived and planned the project; M.T.F. and W.S.W.  
480 analyzed the data and created the manuscript; E.J.H., R.Z., advised the network analysis and  
481 presentation; T.A.M. provided clinical interpretation and analysis support; E.F. contributed to  
482 the GEM implementation; M.T.F. and M.Y. oversaw all aspects of the network and statistical  
483 analyses; M.W. provided network access to data; W.S.W., E.J.H., T.A.M., R.Z., D.B., M.B.,  
484 W.K.C., J.W.G., B.D.G., E.G., P.G., J.W.N., S.U.M., A.E.R., C.E.S., J.G.S., Y.S., H.J.Y.,  
485 N.R.B., E.R.G., M.T.B, J.E.M., M.Y., and M.T.F. edited the final manuscript.

486

## 487 **Competing interests**

488 The authors declare the following competing interests: M.Y. -- GEM commercialization through  
489 Fabric Genomics, Inc; E.F. is an employee of Fabric Genomics.  
490

## 491 References

- 492
- 493 1. Hoffman, J.I. & Kaplan, S. The incidence of congenital heart disease. *J Am Coll Cardiol*
- 494 **39**, 1890-1900 (2002).
- 495 2. Liu, Y., *et al.* Global birth prevalence of congenital heart defects 1970-2017: updated
- 496 systematic review and meta-analysis of 260 studies. *Int J Epidemiol* **48**, 455-463 (2019).
- 497 3. Reller, M.D., Strickland, M.J., Riehle-Colarusso, T., Mahle, W.T. & Correa, A. Prevalence
- 498 of congenital heart defects in metropolitan Atlanta, 1998-2005. *J Pediatr* **153**, 807-813
- 499 (2008).
- 500 4. Homsy, J., *et al.* *De novo* mutations in congenital heart disease with neurodevelopmental
- 501 and other congenital anomalies. *Science* **350**, 1262-1266 (2015).
- 502 5. Jin, S.C., *et al.* Contribution of rare inherited and *de novo* variants in 2,871 congenital heart
- 503 disease probands. *Nat Genet* **49**, 1593-1601 (2017).
- 504 6. Richter, F., *et al.* Genomic analyses implicate noncoding *de novo* variants in congenital
- 505 heart disease. *Nat Genet* **52**, 769-777 (2020).
- 506 7. Sifrim, A., *et al.* Distinct genetic architectures for syndromic and nonsyndromic congenital
- 507 heart defects identified by exome sequencing. *Nat Genet* **48**, 1060-1065 (2016).
- 508 8. Watkins, W.S., *et al.* *De novo* and recessive forms of congenital heart disease have distinct
- 509 genetic and phenotypic landscapes. *Nat Commun* **10**, 4722 (2019).
- 510 9. Zaidi, S., *et al.* *De novo* mutations in histone-modifying genes in congenital heart disease.
- 511 *Nature* **498**, 220-223 (2013).
- 512 10. Boskovski, M.T., *et al.* *De Novo* Damaging Variants, Clinical Phenotypes, and Post-
- 513 Operative Outcomes in Congenital Heart Disease. *Circ Genom Precis Med* **13**, e002836
- 514 (2020).
- 515 11. Carey, A.S., *et al.* Effect of copy number variants on outcomes for infants with single
- 516 ventricle heart defects. *Circ Cardiovasc Genet* **6**, 444-451 (2013).
- 517 12. Geddes, G.C., Przybylowski, L.F., 3rd & Ware, S.M. Variants of significance: medical
- 518 genetics and surgical outcomes in congenital heart disease. *Curr Opin Pediatr* **32**, 730-738
- 519 (2020).
- 520 13. Kim, D.S., *et al.* Burden of potentially pathologic copy number variants is higher in
- 521 children with isolated congenital heart disease and significantly impairs covariate-adjusted
- 522 transplant-free survival. *J Thorac Cardiovasc Surg* **151**, 1147-1151 e1144 (2016).
- 523 14. Landis, B.J., *et al.* Learning to Crawl: Determining the Role of Genetic Abnormalities on
- 524 Postoperative Outcomes in Congenital Heart Disease. *J Am Heart Assoc* **11**, e026369
- 525 (2022).
- 526 15. Newburger, J.W., Fyler, J. & Colan, S.D. Donald Charles Fyler: 1924-2011. *Cardiol Young*
- 527 **21**, 485-489 (2011).
- 528 16. De La Vega, F.M., *et al.* Artificial intelligence enables comprehensive genome
- 529 interpretation and nomination of candidate diagnoses for rare genetic diseases. *Genome*
- 530 *Med* **13**, 153 (2021).
- 531 17. Nathan, M., *et al.* Completeness and Accuracy of Local Clinical Registry Data for Children
- 532 Undergoing Heart Surgery. *Ann Thorac Surg* **103**, 629-636 (2017).
- 533 18. Jacobs, M.L., *et al.* Updating an Empirically Based Tool for Analyzing Congenital Heart
- 534 Surgery Mortality. *World J Pediatr Congenit Heart Surg* **12**, 246-281 (2021).
- 535 19. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. in *Proceedings of*
- 536 *the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data*
- 537 *Mining* (Association for Computing Machinery, San Francisco, California, USA, 2016).

- 538 20. Goodman, S.N. Toward evidence-based medical statistics. 2: The Bayes factor. *Ann Intern*  
539 *Med* **130**, 1005-1013 (1999).
- 540 21. Pearl, J. *Causality : models, reasoning, and inference*, (Cambridge University Press,  
541 Cambridge, U.K. ; New York, 2000).
- 542 22. Amann, J., *et al.* Explainability for artificial intelligence in healthcare: a multidisciplinary  
543 perspective. *BMC Med Inform Decis Mak* **20**, 310 (2020).
- 544 23. Franz L SY & B., P. A deep learning pipeline for patient diagnosis prediction using  
545 electronic health records. *arXiv preprint* (2020).
- 546 24. Heckerman, D., Geiger, D. & Chickering, D.M. Learning Bayesian Networks: The  
547 Combination of Knowledge and Statistical Data. *arXiv preprint* (2013).
- 548 25. London, A.J. Artificial Intelligence and Black-Box Medical Decisions: Accuracy versus  
549 Explainability. *Hastings Cent Rep* **49**, 15-21 (2019).
- 550 26. Miotto, R., Li, L., Kidd, B.A. & Dudley, J.T. Deep Patient: An Unsupervised  
551 Representation to Predict the Future of Patients from the Electronic Health Records. *Sci*  
552 *Rep* **6**, 26094 (2016).
- 553 27. Payrovnaziri, S.N., *et al.* Explainable artificial intelligence models using real-world  
554 electronic health record data: a systematic scoping review. *J Am Med Inform Assoc* **27**,  
555 1173-1185 (2020).
- 556 28. Rajkomar, A., *et al.* Scalable and accurate deep learning with electronic health records.  
557 *NPJ Digit Med* **1**, 18 (2018).
- 558 29. Wang, H., *et al.* Predicting Hospital Readmission via Cost-Sensitive Deep Learning.  
559 *IEEE/ACM Trans Comput Biol Bioinform* **15**, 1968-1978 (2018).
- 560 30. Miller, T.A., *et al.* Genetic, demographic and clinical variables act synergistically to impact  
561 neurodevelopmental outcomes in children with single ventricle heart disease. *medRxiv*,  
562 2022.2010.2001.22280594 (2023).
- 563 31. Wesolowski, S., *et al.* An explainable artificial intelligence approach for predicting  
564 cardiovascular outcomes using electronic health records. *PLOS Digit Health* **1**(2022).
- 565 32. Silander, T. & Myllymäki, P. A Simple Approach for Finding the Globally Optimal  
566 Bayesian Network Structure. *arXiv* **1206.6875** (2012).
- 567 33. Franzin, A., Sambo, F. & Di Camillo, B. bnstruct: an R package for Bayesian Network  
568 structure learning in the presence of missing data. *Bioinformatics* **33**, 1250-1252 (2017).
- 569 34. Hojsgaard, S. Graphical Independence Networks with the gRain Package for R. *J Stat*  
570 *Softw* **46**, 1-26 (2012).
- 571 35. Burnicka-Turek, O., *et al.* Cilia gene mutations cause atrioventricular septal defects by  
572 multiple mechanisms. *Human molecular genetics* **25**, 3011-3028 (2016).
- 573 36. Vodiskar, J., *et al.* Impact of Extracardiac Anomalies on Mortality and Morbidity in Staged  
574 Single Ventricle Palliation. *Ann Thorac Surg* **115**, 1197-1204 (2023).
- 575 37. Garrod, A.S., *et al.* Airway ciliary dysfunction and sinopulmonary symptoms in patients  
576 with congenital heart disease. *Ann Am Thorac Soc* **11**, 1426-1432 (2014).
- 577 38. Nakhleh, N., *et al.* High prevalence of respiratory ciliary dysfunction in congenital heart  
578 disease patients with heterotaxy. *Circulation* **125**, 2232-2242 (2012).
- 579 39. Kim, K.M., *et al.* The Society of Thoracic Surgeons Adult Cardiac Surgery Database: 2022  
580 Update on Outcomes and Research. *Ann Thorac Surg* **115**, 566-574 (2023).
- 581 40. Jacobs, J.P., *et al.* Successful linking of the Society of Thoracic Surgeons Database to  
582 Social Security data to examine the accuracy of Society of Thoracic Surgeons mortality  
583 data. *J Thorac Cardiovasc Surg* **145**, 976-983 (2013).
- 584 41. Brown, M.L., Lench, J.R. & Schaff, H.V. Variability in data: the Society of Thoracic  
585 Surgeons National Adult Cardiac Surgery Database. *J Thorac Cardiovasc Surg* **140**, 267-

- 586 273 (2010).  
587 42. Dimmock, D., *et al.* Project Baby Bear: Rapid precision care incorporating rWGS in 5  
588 California children's hospitals demonstrates improved clinical outcomes and reduced costs  
589 of care. *American journal of human genetics* **108**, 1231-1238 (2021).  
590 43. Group, N.I.S., *et al.* Effect of Whole-Genome Sequencing on the Clinical Management of  
591 Acutely Ill Infants With Suspected Genetic Disease: A Randomized Clinical Trial. *JAMA*  
592 *Pediatr* **175**, 1218-1226 (2021).  
593  
594  
595



596  
597

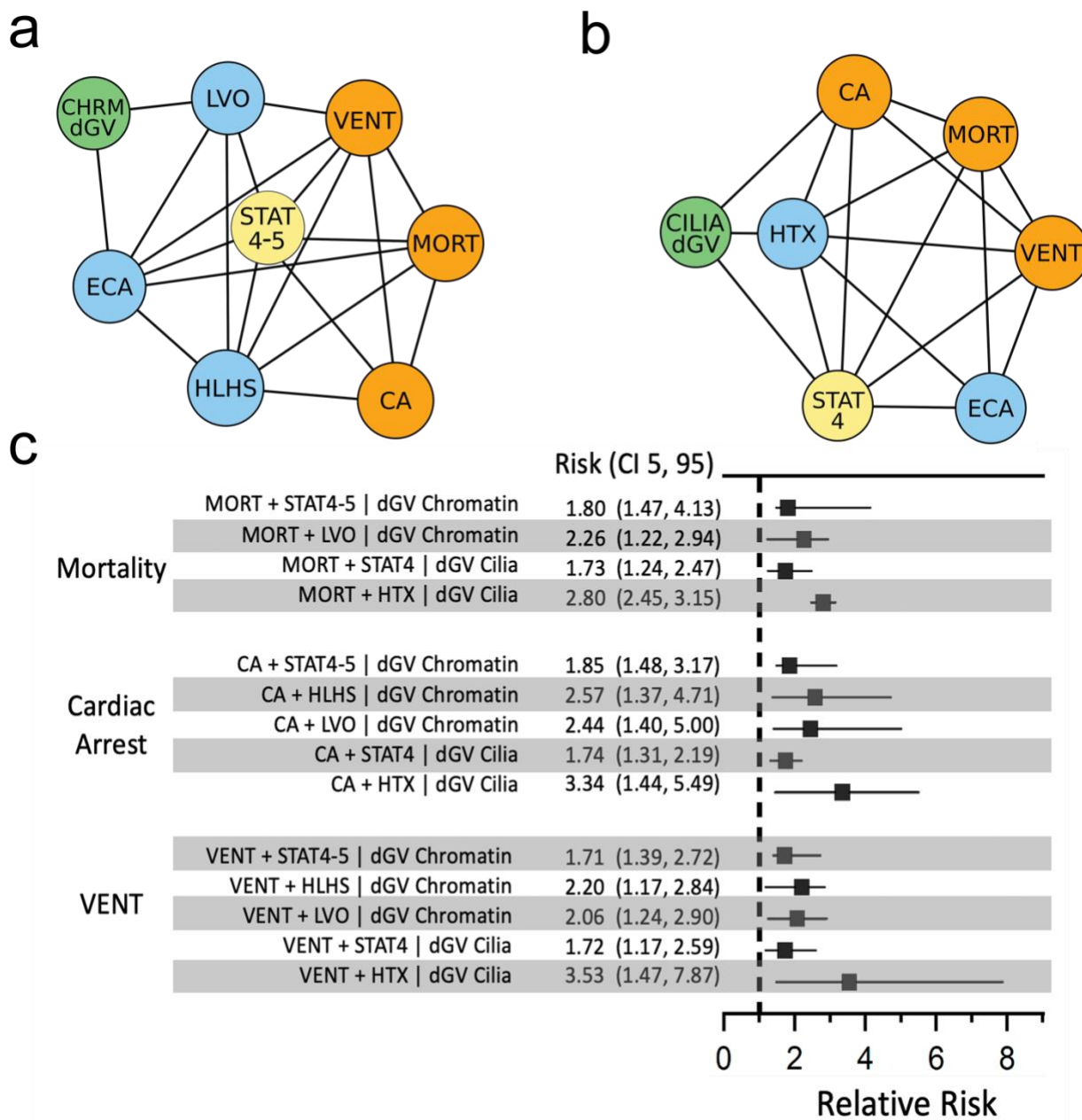
**Table 1. Absolute risk ratios for CHD phenotypes by gene pathway**

Gene pathway	n	Cardiac phenotype				
		AVC	CTD	HTX	LVO	OTH
<b>Chromatin genes</b> ( <i>de novo</i> )	28	-	0.52 (0.32, 0.72)	-	<b>1.61</b> (1.41, 1.81)	<b>1.85</b> (1.44, 2.26)
<b>Cilia genes</b> ( <i>recessive</i> )	35	3.02 (0.00, 6.97)	0.55 (0.41, 0.69)	<b>2.63</b> (2.06, 3.20)	1.09 (0.90, 1.28)	0.67 (0.18, 1.16)
<b>HHE genes</b> ( <i>de novo</i> )	9	-	0.82 (0.00, 1.74)	1.58 (0.00, 9.05)	0.84 (0.00, 4.88)	1.93 (0.00, 4.29)
<b>Wnt genes</b> ( <i>de novo</i> )	18	-	0.54 (0.16, 0.92)	-	<b>2.13</b> (1.86, 2.40)	0.99 (0.00, 2.25)
<b>FoxJ1 genes</b> ( <i>recessive</i> )	6	-	0.56 (0.00, 3.54)	<b>6.89</b> (3.30, 10.36)	0.78 (0.00, 4.82)	0.81 (0.00, 3.54)
<b>Notch1 genes</b> ( <i>de novo</i> )	8	-	1.49 (0.96, 2.02)	-	1.36 (0.00, 2.86)	-
<b>Signal trans genes</b> ( <i>de novo</i> )	14	3.56 (0.00, 20.50)	0.69 (0.24, 1.14)	-	<b>1.51</b> (1.01, 2.01)	1.26 (0.00, 2.84)
<b>TGF-β genes</b> ( <i>de novo</i> )	13	-	1.11 (0.74, 1.48)	-	1.09 (0.32, 1.86)	1.37 (0.00, 3.05)
<b>CHD genes</b> ( <i>de novo</i> )	51	-	0.86 (0.78, 0.94)	0.27 (0.00, 1.43)	<b>1.22</b> (1.11, 1.33)	<b>1.59</b> (1.38, 1.80)

598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612

Each ratio is reported as the mean and 95% confidence interval from 1000 bootstrap replicates fitted to a t-distribution. Absolute risk ratios with 95% CIs >1.00 are bolded. n indicates the number of patients with damaging genetic variants/genotypes (GEM score ≥ 1) found in that gene pathway. Dashes indicate no patients with damaging genotypes. Phenotype categories are atrioventricular canal defects (AVC, n=64), conotruncal defects (CTD, n=934), heterotaxy/laterality defects (HTX, n=219), left ventricular outflow tract obstructions (LVO, n=647), and all other defects (OTH, n=389). HHE, high heart expression genes in the developing mouse heart; CHD genes, a curated list of genes reported to cause CHD (see Supplemental Table 11).

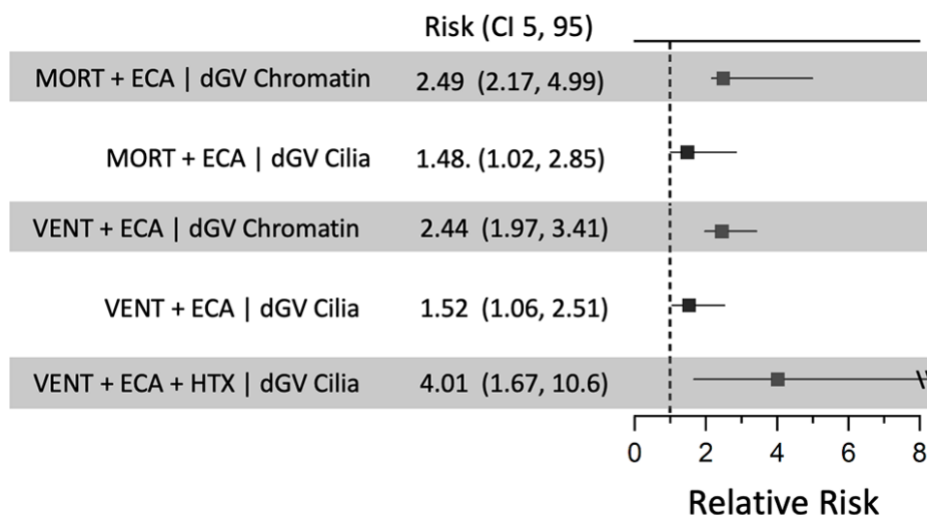
613 **Figure 1.** Damaging chromatin and cilia genotypes predict adverse post-operative outcomes in  
 614 the context of CHD phenotypes and surgical complexity.



615  
 616  
 617  
 618  
 619  
 620  
 621  
 622  
 623

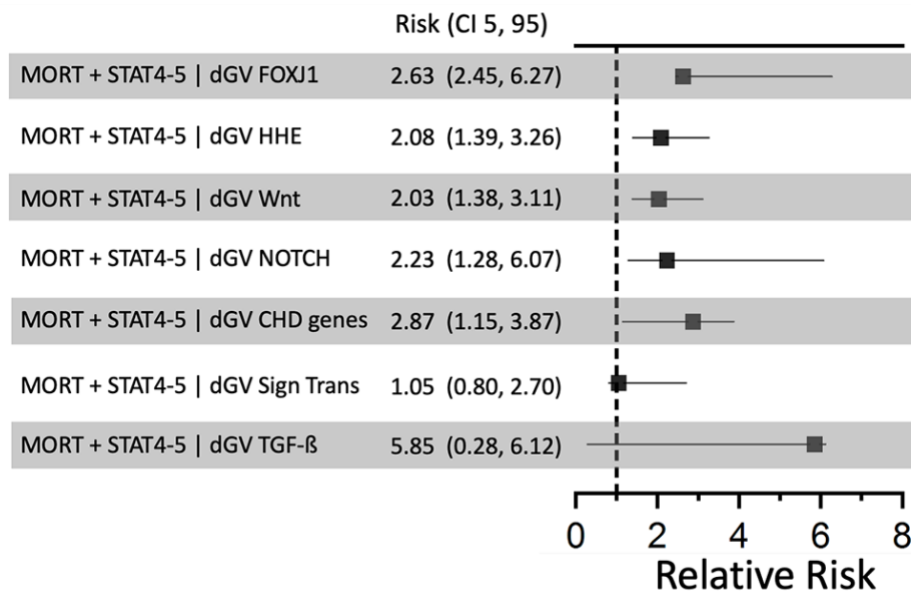
624  
625  
626  
627

**Figure 2.** Damaging chromatin and cilia genotypes predict adverse post-operative outcomes in the context of extracardiac anomalies.



628  
629  
630  
631  
632  
633  
634

**Figure 3.** Damaging genotypes in various gene categories/pathways are predictive of mortality for the most complex surgical procedures.



635  
636

637 **Figure Legends**

638

639 **Figure 1. Damaging chromatin and cilia genotypes predict adverse post-operative outcomes**

640 **in the context of CHD phenotypes and surgical complexity. a,** An exact Bayesian network

641 depicting the relationship among damaging *de novo* genetic variants in chromatin-modifying genes

642 (green), phenotypes: LVO, HLHS, and ECAs (blue), surgical STAT4 or STAT5 category (red),

643 and surgical outcomes (orange). **b,** An exact Bayesian network depicting the relationship among

644 damaging recessive genetic variants in cilia genes (green), phenotypes: laterality defects (HTX)

645 and extra cardiac anomalies (ECAs) (blue), surgical STAT4 category (yellow), and surgical

646 outcomes (orange). Directed acyclic graphs were moralized and displayed as non-directional

647 networks. **c,** Relative risk ratios for adverse post-operative outcomes and CHD phenotypes or

648 surgical complexity, comparing probands with and without damaging genotypes. Empirical ninety-

649 five percent confidence intervals (CI 5, 95) are based on 1000 resampled network-based

650 probabilities (see Methods). Abbreviations: Chromatin dGV: *de novo* damaging genotypes in

651 chromatin-modifying genes, LVO: left ventricular outflow tract obstruction, HLHS: hypoplastic

652 left heart syndrome Cilia dGV: biallelic damaging genotypes in cilia-related genes, ECA: extra

653 cardiac anomaly, HTX: heterotaxy/laterality defects, MORT: mortality, STAT4: surgical STAT4

654 category, STAT4-5: surgical STAT 4 or STAT5 category, VENT: post-operative ventilation time >7

655 days.

656

657 **Figure 2. Damaging chromatin and cilia genotypes predict adverse post-operative outcomes**

658 **in the context of extracardiac anomalies.** Relative risk ratios for adverse post-operative outcomes

659 and extracardiac anomalies (ECAs), comparing probands with and without damaging genotypes in

660 chromatin-modifying or cilia-related genes. The \\ symbol represents CI95 that exceeds the x-axis  
661 range.

662

663 **Figure 3. Damaging genotypes in various gene categories/pathways are predictive of**  
664 **mortality for the most complex surgical procedures.** Relative risk ratios for adverse post-  
665 operative outcomes and surgical complexity, comparing probands with and without damaging  
666 genotypes in various gene pathways or categories. Gene lists are described in Supplemental Table  
667 11 and have been previously published.<sup>5,8,9</sup> There is overlap between gene lists, with some genes  
668 represented in more than one gene pathway/category (Supplemental Figure 4).