# SUPPLEMENTAL MATERIALS

**Table of Contents**

**Supplemental Methods**

*Patient characteristics*

Pregnancy history was obtained by parental interview at time of enrollment and review of medical records. Patients who were adopted or conceived with donor gametes were excluded. The echocardiogram, catheterization, and operative reports were reviewed to determine cardiac phenotypes. Extracardiac structural anomalies were obtained from the medical records as previously defined[4]. and review of medical records

*Whole genome sequencing and variant identification*

Whole genome sequencing was completed on 1812 PCGC trios (patient and parents). DNVs (coding and noncoding) arising in CHD probands were identified by comparing CHD probands' WGS to their parents' WGS as previously described[20]. In brief, genomic DNAs from venous blood or saliva were prepared for sequencing using a PCR-free library preparation or SK2-IES library preparation. All samples were sequenced on an Illumina Hi-Seq X Ten with 150-bp paired reads to a median depth >30x per individual. Reads were aligned to GRCh37 or GRCh38 with the Burrows-Wheeler Aligner (BWA-MEM)[51]. GATK Best Practices recommendations were implemented for base quality score recalibration (QSR), indel realignment, and duplicate removal[52]. Standard hard filtering parameters were used for SNV and indel discovery across all, followed by N+1 joint genotyping and variant QSR[53,54].

*Identification and confirmation of de novo variants*

DNV identification was jointly performed for both cases and controls. First, candidate DNVs were identified separately by three computational pipelines from PCGC members at three centers- Columbia, Harvard and Mount Sinai. Seven hundred forty-eight of the CHD patients were previously described[20]. DNVs in the remaining 1064 CHD patients were identified after minor modifications of the previously reported pipeline. At Columbia, parameters for DNV identification were proband genotype quality (GQ) ≥70, parent GQ ≥30, parent read depth ≥10, Fisher's exact test strand bias >25, variant quality by depth > 2 for SNVs and > 1 for indels, ReadPosRankSum <-3 (indel), proband alternative allele depth >5,

proband allele balance ≥0.2, parent alternative allele ratio ≤3.5%, variant allele count (AC) <3, position AC <4, and allele frequency < 1E-04 in the Genome Aggregation Database (gnomAD) and 1000 Genomes databases[55,56]. Harvard required GATK PASS, proband GQ >= 60, parent GQ >= 30, proband alternative allele depth >6, parent depth >= 10, AC < 3 across all trios, allele frequency < 1E-04 in the gnomAD and 1000 Genomes databases[55,56] Mount Sinai required variants to be GATK PASS AC <3, GQ > 30, and proband allele balance 0.2-0.8, proband alternative allele depth >4, parent alternative allele ratio <= 0.1. MUC genes, HLA genes, and non-standard chromosomes were excluded. Variants within regions UCSC Genome Browser Tracks[57] identified by RepeatMasker (mappability >1 in 300 bp, low complexity, segmental duplications) or DAC and Duke exclusion lists were also excluded. After the initial candidate DNV identification, all candidate DNVs identified by any of these three pipelines were next collaboratively re-evaluated by both FreeBayes[58] and DeepVariant[59]. Variants from any pipeline that were FreeBayes PASS and DeepVariant score >30 were considered true. After the initial candidate DNV identification, all candidate next collaboratively both  from any pipeline

*Genomic Risk Score*

Common biallelic single nucleotide polymorphisms were identified by filtering for PASS filter, minimum depth 10, minimum genotype quality 20, and minor allele frequency > 0.05. Positions were then filtered to remove genotypes with >10% missingness or Hardy-Weinberg equilibrium threshold of 1e-6. Polygenic scores for Type 2 Diabetes[27] (PGS000014) and hypertension[28] (PGS000706) were downloaded from the PGS Catalog[60] (pgscatalog.org) and lifted from hg19 to hg38. Plink 1.9 was then used to determine genomic risk scores (GRS) for mothers. After filtering, GRS were calculated for 1730 patients. Comparisons between patient or maternal GRS and DNV counts were made using a Poisson linear model as described below.

*Pathogenic CHD Variants*

One hundred thirty-eight human dominant CHD genes were identified from the ClinVar and Online Mendelian Inheritance in Man databases[15,16] (Supplemental Table 4). Rare coding variants in exome and genome sequencing were identified as previously described[1,17]. A proband was considered to have a

pathogenic CHD variant if a *de novo* loss-of function (stopgain, frameshift or canonical splice site) variant, or rare inherited loss-of-function variant, was identified in a dominant CHD gene.

*Statistical Analysis*

Prevalence of gestational diabetes was compared with published US birth cohorts[18,19]. Pre-pregnancy maternal body mass index (BMI) was calculated as $kg/m^2$. Odds ratios were calculated using a Fischer's exact test. P-values were calculated either with two-sided binomial unpaired t-tests, Fisher's exact tests, or Poisson generalized linear model as indicated. The R package *stats* was used for the Poisson generalized linear model to determine the relationship of patient DNV count with maternal diabetes status, maternal BMI status of <25 or >30, maternal age at patient birth, father age at patient birth, or the interaction between these factors as indicated. 'Significant' p-values were identified using a Bonferroni correction for the number of independent tests per comparison, summing all tests in related Tables and Supplemental Tables as specified in each caption.

# Supplemental Table I. Cohort Summary

| Cohort | Probands, Number | Mother Age in Years, Mean (Range) | Father Age in Years, Mean (Age) | Female, Number (%) | Binomial p-value for Number Female, compared | With ECA, Number (Percent) | Maternal BMI >30, Number (Percent) | Maternal Diabetes, Number (Percent) | Hypoplastic Left Heart Syndrome, Number (percent) | Tetralogy of Fallot, Number (percent) | Left-sided Conotruncal Defect, Number (percent) | Heterotaxy, Number (percent) | Atrial Septal Defect, Number (percent) | Other, Number (percent) | T-test p-value for CHD Type Counts |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **All PCGC** | 12842 | 30.1 (13.0-55.3) | 32.6 (13.6-68.9) | 5823/12842 (45%) | - | 4043/12817 (32%) | 3453/10994 (31%) | 1212/12842 (9%) | 687 (5%) | 1964 (15%) | 5484 (43%) | 344 (3%) | 659 (5%) | 3704 (29%) | - |
| **With Diabetes during Pregnancy** | 1212 | 32.5 (16.0-48.4) | 34.9 (18.0-68.9) | 534/1212 (44%) | 0.42* | 495/1212 (41%) | 533/1078 (49%) | 1212/1212 (100%) | 53 (4%) | 203 (17%) | 512 (42%) | 44 (4%) | 48 (4%) | 352 (29%) | 0.07 |
| **Without Diabetes, during Pregnancy or Historical** | 11471 | 29.8 (13.0-55.3) | 32.4 (13.6-66.7) | 5218/11471 (45%) | 0.87* | 3518/11447 (31%) | 2874/9775 (29%) | 0/11471 (0%) | 634 (6%) | 1730 (15%) | 4911 (43%) | 291 (3%) | 600 (5%) | 3305 (29%) | 0.84 |
| **With BMI >30** | 1605 | 30.7 (14.0-48.4) | 33.2 (16.4-60.7) | 706/1605 (44%) | 0.45* | 594/1601 (37%) | 1605/1605 (100%) | 399/1605 (25%) | 101 (6%) | 250 (16%) | 674 (42%) | 47 (3%) | 50 (3%) | 483 (30%) | 0.08 |
| **With BMI <25** | 7066 | 29.8 (13.0-55.3) | 32.3 (13.6-66.6) | 3222/7066 (46%) | 0.8* | 2100/7048 (30%) | 0/7066 (0%) | 390/7066 (6%) | 368 (5%) | 1071 (15%) | 3061 (43%) | 177 (2%) | 387 (5%) | 2002 (28%) | 0.35 |
| **PCGC WGS** | 1812 | 31.1 (15.4-47.3) | 33.5 (16.5-60.3) | 722/1811 (40%) | - | 536/1812 (30%) | 378/1693 (21%) | 188/1812 (10%) | 128 (7%) | 369 (20%) | 784 (43%) | 54 (3%) | 107 (6%) | 370 (20%) | - |
| **With Diabetes** | 188 | 33.1 (20.3-42.7) | 35.9 (20.6-60.3) | 72/188 (38%) | 0.81** | 69/188 (37%) | 89/170 (52%) | 188/188 (100%) | 6 (4%) | 40 (22%) | 86 (48%) | 4 (2%) | 10 (5%) | 33 (18%) | 0.06 |
| **Without Diabetes** | 1605 | 30.9 (15.4-47.3) | 33.2 (16.5-59.8) | 643/1604 (40%) | 0.94** | 462/1605 (29%) | 278/1506 (18%) | 0/1605 (0%) | 121 (8%) | 321 (20%) | 689 (43%) | 45 (3%) | 97 (6%) | 332 (21%) | 0.82 |
| **With BMI >30** | 259 | 31.2 (18.4-45.2) | 33.8 (18.2-60.3) | 95/259 (37%) | 0.49** | 94/259 (36%) | 259/259 (100%) | 71/259 (27%) | 19 (7%) | 52 (20%) | 113 (44%) | 10 (4%) | 10 (4%) | 53 (20%) | 0.07 |
| **With BMI <25** | 1075 | 30.9 (15.4-47.3) | 33.2 (16.5-58.0) | 426/1074 (40%) | 0.95** | 312/1075 (29%) | 0/1075 (0%) | 56/1075 (5%) | 76 (7%) | 217 (20%) | 456 (42%) | 28 (3%) | 76 (7%) | 222 (21%) | 0.38 |

Abbreviations: BMI, body mass index; CHD, congenital heart disease; ECA, extracardiac anomaly; PCGC, Pediatric Cardiac Genomics Consortium

* compared to All PCGC; ** compared to PCGC WGS; p-value threshold for significance 0.00625 based on 8 comparisons

Supplemental Table II. Extracardiac Anomalies by Maternal Age

Mothers 20-30 years old

| Exposure Status | Extracardiac Anomalies Present, Number | Extracardiac Anomalies Absent, Number | Fisher OR (95% CI, p-value) |
|---|---|---|---|
| **GDM Exposed** | 88 | 158 | **1.39 (1.24-1.56, 1.84E-08)** |
| **GDM Non-Exposed** | 1444 | 3417 | - |
| **Maternal BMI >30** | 231 | 431 | **1.34 (1.12-1.61, 1.38E-03)** |
| **Maternal BMI <25** | 859 | 2154 | - |

Mothers 30-40 years old

| Exposure Status | Extracardiac Anomalies Present, Number | Extracardiac Anomalies Absent, Number | Fisher OR (95% CI, p-value) |
|---|---|---|---|
| **GDM Exposed** | 202 | 301 | **1.52 (1.25-1.84, 1.47E-05)** |
| **GDM Non-Exposed** | 1581 | 3586 | - |
| **Maternal BMI >30** | 294 | 490 | **1.40 (1.18-1.65, 6.82E-05)** |
| **Maternal BMI <25** | 989 | 2302 | - |

Abbreviations: BMI, body mass index; CI, confidence interval; GDM, gestational diabetes mellitus; OR, odds ratio

Bonferroni p-value threshold: 7.14E-03 (7 total comparisons in Main and Supplemental Tables)

Supplemental Table III. Pairwise GLM with Interaction Term

| GLM Exclusing Patients with ASD or "other" diagnoses | Diabetes Exposure, P-Value (Parameter Esimate, Standard Error) | Father Age, P-Value (Parameter Esimate, Standard Error) | Mother Age, P-Value (Parameter Esimate, Standard Error) |
|---|---|---|---|
| **Mother Age** | x | x | **<2E-16 (0.03, 5.9E-4)** |
| **Father Age** | x | **<2E-16 (0.02, 4.7E-04)** | x |
| **Mother Age + Father Age** | x | **<2E-16 (0.02, 7.3E-04)** | **<2E-16 (0.01, 8.9E-04)** |
| **Diabetes** | **3.93E-09 (0.06, 0.01)** | x | x |
| **Mother Age + Diabetes** | 0.95 (6.8E-04, 0.01) | x | **<2E-16 (0.03, 5.9E-04)** |
| **Father Age + Diabetes** | 0.65 (-4.6E-03, 0.01) | **<2E-16 (0.02, 4.7E-04)** | x |
| **Mother Age + Father Age + Diabetes** | 0.38 (-0.01, 0.01) | **<2E-16 (0.02, 7.2E-04)** | **<2E-16 (0.01, 8.9E-04)** |

Abbreviations: ASD, atrial septal defect; GLM, generalized lienar model

GLM comparison Bonferroni p-value threshold 0.0038 (13 total comparisons including Main and Supplemental Tables)

Supplemental Table IV. Parental Age Interaction By Generalized Linear Model

| GLM Co-Variates | Diabetes Exposure, P-Value (Parameter Esimate, Standard Error) | Father Age, P-Value (Parameter Esimate, Standard Error) | Mother Age, P-Value (Parameter Esimate, Standard Error) | Interaction Term, P-Value (Parameter Esimate, Standard Error) |
|---|---|---|---|---|
| **Mother + Father Age + Interaction** | x | **<2E-16 (0.03, 2.2E-03)** | **6.43E-15 (0.02, 2.2E-03)** | **1.58E-05 (-2.8E-04, 6.4E-05)** |
| **Diabetes + Father Age + Interaction** | 0.85 (0.01, 0.05) | **<2E-16 (0.02, 4.3E-04)** | x | 0.80 (-3.3E-04, 1.3E-03) |
| **Diabetes + Mother Age + Interaction** | 0.63 (0.03, 0.06) | x | **<2E-16 (0.03, 5.3E-04)** | 0.68 (-7.2E-04, 1.7E-03) |

Abbreviations: GLM, generalized linear model

GLM comparison Bonferroni p-value threshold 0.0038 (13 total comparisons including Main and Supplemental Tables)

# Supplemental Table V. CHD Genes

| geneName | |
|---|---|
| ABCC9 | MED13L |
| ACTB | MEIS2 |
| ACVR2B | MYH6 |
| ADAMTS10 | NF1 |
| ADNP | NFATC1 |
| ANKRD11 | NIPBL |
| ARHGAP31 | NKX2-5 |
| ARID1A | NODAL |
| ARID1B | NOTCH1 |
| ASXL1 | NOTCH2 |
| BBS1 | NR2F2 |
| BCOR | NRAS |
| BRAF | NSD1 |
| CACNA1C | NSDHL |
| CBL | OFD1 |
| CDK13 | PBX1 |
| CDKN1C | PITX2 |
| CFC1 | PKD1 |
| CHD4 | PKD2 |
| CHD7 | PTPN11 |
| CITED2 | RAD21 |
| COL1A1 | RAF1 |
| COL2A1 | RAI1 |
| COL3A1 | RBFOX2 |
| COL5A1 | RIT1 |
| COL5A2 | RPL11 |
| COX7B | RPL35A |
| CREBBP | RPL5 |
| CRELD1 | RPS10 |
| DGCR2 | RPS17 |
| DLL4 | RPS19 |
| DYNC2H1 | RPS24 |
| ECE1 | RPS26 |
| EFTUD2 | RPS7 |
| EHMT1 | SALL1 |
| ELN | SEMA3E |
| EP300 | SETBP1 |
| ETS1 | SF3B4 |
| FBN1 | SHH |
| FBN2 | SHOC2 |
| FGF8 | SKI |
| FGFR1 | SMAD2 |
| FLT4 | SMAD3 |
| FMR1 | SMAD4 |
| FOXC1 | SMAD6 |
| FOXC2 | SMARCA4 |
| FOXF1 | SMARCB1 |
| GATA4 | SMARCE1 |
| GATA5 | SMC3 |
| GATA6 | SMS |
| GDF1 | SON |
| GJA1 | SOS1 |
| GLI3 | SOX2 |
| HAND1 | SOX9 |
| HAND2 | TAB2 |
| HCCS | TBX1 |
| HRAS | TBX20 |
| JAG1 | TBX3 |
| KANSL1 | TBX5 |
| KAT6A | TCOF1 |
| KAT6B | TFAP2B |
| KDM6A | TGFBR1 |
| KMT2A | TGFBR2 |
| KMT2D | TLL1 |
| KRAS | TSC1 |
| LBR | TSC2 |
| LEFTY2 | TWIST1 |
| MAP2K1 | ZEB2 |
| MAP2K2 | ZFPM2 |

Supplemental Table VI. Association of Exposures with Presence of Loss-of-Function Variant in Dominant Human CHD Gene

| Exposure Status | Patients with LoF CHD Gene Variant | Patients without LoF CHD Gene Variant | Fisher OR (95% CI, p-value) |
|---|---|---|---|
| **Diabetes Exposure** | 17 | 338 | 1.0 (0.56-1.68, 1.0) |
| **No Diabetes Exposure** | 157 | 3123 | |

Abbreviations: CHD, congenital heart disease; CI, confidence interval; LoF, loss of function; OR, odds ratio