**Supplementary Table 1. Rare *de novo* protein-altering variants identified from exome sequencing data of children with biliary atresia and their parents enrolled in the National Birth Defects Prevention Study, 1997-2011**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Child** | **Gene** | **Variant** | **Variant type** | **REVEL** | **CADD** | **Clinvar variation ID** | **AAF†** | **Case AAF** | **Control AAF** | **p-value‡** |
| 1 | *MICALL1* | NM\_033386.4:c.2257C>T(p.Gln753Ter) | stop gain | . | 40.0 | . | 0 | 0.009 | 0.0 | 1.6x10-7 |
| 2 | *C10orf118* | NM\_018017.4:c.2616A>T(p.Glu872Asp) | missense | 0.1 | 24.4 | . | 0 | 0.009 | 0.0 | 2.7x10-7 |
| 2 | *EDC4* | NM\_014329.5:c.872T>C(p.Ile291Thr) | missense | 0.3 | 24.1 | . | 1.2x10-5 | 0.009 | 0.0 | 2.7x10-7 |
| 3 | *COL2A1* | NM\_001844.5:c.976C>T(p.Arg326Cys) | missense | 1.0 | 32.0 | . | 0 | 0.009 | 0.0 | 1.6x10-7 |
| 3 | *OR6Y1* | NM\_001005189.1:c.724G>C(p.Ala242Pro) | missense | 0.3 | 26.6 | . | 0 | 0.009 | 0.0 | 1.6x10-7 |
| 4 | *BZRAP1* | NM\_004758.4:c.4907A>G(p.Tyr1636Cys) | missense | 0.8 | 24.7 | . | 0 | 0.009 | 0.0 | 1.6x10-7 |
| 5 | *FAT1* | NM\_005245.4:c.6452\_6453del(p.Leu2151ArgfsTer26) | frameshift deletion | . | . | . | 0 | 0.009 | 0.0 | 2.7x10-7 |
| 6 | *CATSPER1* | NM\_053054.4:c.1916G>A(p.Ser639Asn) | missense | 0.3 | 15.4 | . | 0 | 0.009 | 0.0 | 2.7x10-7 |
| 7 | *SGK223* | NM\_001080826.3:c.2941C>T(p.His981Tyr) | missense | 0.1 | 25.2 | . | 0 | 0.009 | 0.0 | 2.7x10-7 |
| 8 | *COL4A3* | NM\_000091.5:c.248T>C(p.Leu83Pro) | missense | 0.4 | 23.5 | . | 0 | 0.009 | 0.0 | 2.7x10-7 |
| 8 | *NEFL* | NM\_006158.5:c.418G>C(p.Glu140Gln) | missense | . | 21.3 | . | 0 | 0.009 | 0.0 | 2.7x10-7 |
| 9 | *CASC5* | NM\_170589.5:c.4292T>C(p.Val1431Ala) | missense | 0.0 | 8.9 | . | 0 | 0.009 | 0.0 | 2.7x10-7 |
| 9 | *NOTCH2* | NM\_024408.4:c.5194C>T(p.Gln1732Ter) | stop gain | . | 40.0 | . | 0 | 0.009 | 0.0 | 2.7x10-7 |
| 9 | *THAP10* | NM\_020147.4:c.346C>T(p.Gln116Ter) | stop gain | . | 34.0 | . | 0 | 0.009 | 0.0 | 2.7x10-7 |
| 10 | *SPATS1* | NM\_145026.4:c.807A>G(p.Ile269Met) | missense | 0.2 | 23.6 | . | 0 | 0.009 | 0.0 | 2.7x10-7 |
| 11 | *CRIM1* | NM\_016441.3:c.2542C>T(p.Leu848Phe) | missense | 0.4 | 27.1 | . | 0 | 0.009 | 0.0 | 2.7x10-7 |
| 12 | *GRM1* | NM\_001278064.2:c.2902T>C(p.Phe968Leu) | missense | 0.2 | 19.5 | . | 0 | 0.009 | 0.0 | 2.7x10-7 |
| 13 | *ALS2CR11* | NM\_001168221.2:c.5326A>G(p.Asn1776Asp) | missense | 0.1 | 17.7 | . | 0 | 0.009 | 0.0 | 1.6x10-7 |
| 13 | *RLN2* | NM\_134441.3:c.551T>A(p.Phe184Tyr) | missense | 0.1 | 0.0 | . | 4.3x10-6 | 0.009 | 0.0 | 1.6x10-7 |
| 14 | *CASP10* | NM\_032977.4:c.1510G>A(p.Ala504Thr) | missense | 0.1 | 22.7 | . | 0 | 0.009 | 0.0 | 1.6x10-7 |
| 15 | *STXBP4* | NM\_178509.6:c.1414C>T(p.Arg472Cys) | missense | 0.1 | 24.3 | . | 2.4x10-5 | 0.009 | 0.0 | 2.7x10-7 |
| 16 | *CAPN2* | NM\_001748.5:c.691G>A(p.Ala231Thr) | missense | 0.9 | 27.6 | . | 0 | 0.009 | 0.0 | 2.7x10-7 |
| 16 | *MMRN2* | NM\_024756.3:c.1553C>A(p.Thr518Asn) | missense | 0.1 | 13.3 | . | 0 | 0.009 | 0.0 | 2.7x10-7 |
| 17 | *COMMD8* | NM\_017845.5:c.376C>T(p.Leu126Phe) | missense | 0.3 | 32.0 | . | 0 | 0.009 | 0.0 | 1.6x10-7 |
| 18 | *ARHGEF4* | NM\_015320.4:c.1778A>G(p.Gln593Arg) | missense | 0.5 | 24.7 | . | 0 | 0.009 | 0.0 | 2.7x10-7 |
| 18 | *CDC20* | NM\_001255.3:c.761A>T(p.Asp254Val) | missense | 0.7 | 32.0 | . | 0 | 0.009 | 0.0 | 2.7x10-7 |
| 18 | *MARVELD3* | NM\_052858.6:c.35C>T(p.Ala12Val) | missense | 0.0 | 18.7 | . | 0 | 0.009 | 0.0 | 2.7x10-7 |
| 18 | *SCN1A* | NM\_006920.6:c.4055T>A(p.Ile1352Asn) | missense | 1.0 | 28.3 | 189856 | 0 | 0.009 | 0.0 | 2.7x10-7 |
| 19 | *PHIP* | NM\_017934.7:c.623A>G(p.Lys208Arg) | missense | 0.4 | 26.7 | . | 0 | 0.009 | 0.0 | 1.6x10-7 |
| 20 | *CMTM4* | NM\_178818.3:c.472T>C(p.Phe158Leu) | missense | 0.3 | 27.2 | . | 4.0x10-6 | 0.009 | 0.0 | 2.7x10-7 |
| 20 | *E2F7* | NM\_203394.3:c.266T>C(p.Met89Thr) | missense | 0.6 | 23.5 | . | 0 | 0.009 | 0.0 | 2.7x10-7 |
| 20 | *ZNF845* | NM\_138374.3:c.1681C>T(p.Arg561Cys) | missense | 0.1 | 13.9 | . | 8.0x10-6 | 0.009 | 0.0 | 2.7x10-7 |
| 21 | *GCN1L1* | NM\_006836.2:c.2051C>G(p.Ser684Cys) | missense | 0.2 | 24.4 | . | 0 | 0.009 | 0.0 | 2.7x10-7 |
| 21 | *LRSAM1* | NM\_001005374.3:c.1871G>A(p.Arg624Gln) | missense | 0.1 | 19.1 | . | 8.0x10-6 | 0.009 | 0.0 | 2.7x10-7 |
| 21 | *SVIL* | NM\_021738.3:c.1597C>G(p.His533Asp) | missense | 0.1 | 12.0 | . | 0 | 0.009 | 0.0 | 2.7x10-7 |
| 22 | *CRIPAK* | ENST00000324803.4:c.1196A>T(p.Lys399Ile) | missense | 0.1 | 14.1 | . | 0 | 0.009 | 0.0 | 1.6x10-7 |
| 22 | *FBLN1* | NM\_006486.3:c.1697+1G>T | splicing | . | 34.0 | . | 0 | 0.009 | 0.0 | 1.6x10-7 |
| 23 | *P2RY2* | NM\_002564.4:c.124G>T(p.Val42Leu) | missense | 0.0 | 13.6 | . | 0 | 0.009 | 0.0 | 1.6x10-7 |
| 24 | *ATOH8* | NM\_032827.7:c.740G>A(p.Ser247Asn) | missense | 0.7 | 25.7 | . | 0 | 0.009 | 0.0 | 1.7x10-7 |
| 25 | *RBBP8* | NM\_002894.3:c.2329G>A(p.Val777Met) | missense | 0.1 | 24.9 | . | 0 | 0.009 | 0.0 | 1.6x10-7 |
| 26 | *CNRIP1* | NM\_015463.3:c.430G>A(p.Glu144Lys) | missense | 0.6 | 32.0 | . | 0 | 0.009 | 0.0 | 2.7x10-7 |
| 27 | *WSB1* | NM\_015626.10:c.1091G>A(p.Ser364Asn) | missense | 0.1 | 22.6 | . | 0 | 0.009 | 0.0 | 1.6x10-7 |

†Average alternate allele frequency based on gnomAD v2.1.1 database

‡P-values based on case-control single variant Score test adjusting for biological sex and the first five principal components

Abbreviations: AAF, alternate allele frequency; REVEL, rare exome variant ensemble learner score; CADD, phred-scaled combined annotation dependent depletion score

**Supplementary Table 2. Rare homozygous protein-altering variants identified from exome sequencing data of children with biliary atresia and their parents enrolled in the National Birth Defects Prevention Study, 1997-2011**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Child** | **Gene** | **Variant** | **Variant type** | **REVEL** | **CADD** | **Clinvar variation ID** | **AAF†** | **Case AAF** | **Control AAF** | **p-value‡** |
| 11 | *CEL* | NM\_001807.6:c.2026G>T(p.Ala676Ser) | missense | 0.0 | 0.0 | 1328315 | 0.0001 | 0.009 | 0.0003 | 4.0x10-4 |
| 28 | *CEL* | NM\_001807.6:c.2026G>T(p.Ala676Ser) | missense | 0.0 | 0.0 | 1328315 | 0.0001 | 0.009 | 0.0003 | 4.0x10-4 |

†Average alternate allele frequency based on gnomAD v2.1.1 database

‡P-values based on case-control single variant Score test adjusting for biological sex and the first five principal components

Abbreviations: AAF, alternate allele frequency; REVEL, rare exome variant ensemble learner score; CADD, phred-scaled combined annotation dependent depletion score

**Supplementary Table 3. Rare compound heterozygous protein-altering variants identified from exome sequencing data of children with biliary atresia and their parents enrolled in the National Birth Defects Prevention Study, 1997-2011**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Child** | **Gene** | **Variant** | **Variant type** | **REVEL** | **CADD** | **Clinvar variation ID** | **AAF†** | **Case AAF** | **Control AAF** | **p-value‡** |
| 10 | *NEB* | NM\_004543.5:c.11370G>A(p.Met3790Ile) | Missense | 0.07 | 22.30 | . | 0 | 0.009 | 0.0000 | 2.7x10-7 |
| NM\_004543.5:c.9318C>A(p.Ser3106Arg) | Missense | 0.45 | 17.11 | . | 2.41x10-5 | 0.009 | 0.0000 | 7.9x10-1 |
| 18 | *NEB* | NM\_004543.5:c.15287A>G(p.Gln5096Arg) | Missense | 0.10 | 17.73 | . | 0 | 0.009 | 0.0000 | 2.7x10-7 |
| NM\_004543.5:c.8063A>G(p.His2688Arg) | Missense | 0.07 | 19.83 | 705083 | 4.00x10-4 | 0.009 | 0.0007 | 5.7x10-3 |
| 23 | *PDK1L1* | NM\_138295.5:c.7552G>A(p.Ala2518Thr) | Missense | 0.14 | 9.64 | . | 2.51x10-5 | 0.009 | 0.0000 | 1.6x10-7 |
| NM\_138295.5:c.8485G>C(p.Glu2829Gln) | Missense | 0.08 | 14.41 | . | 0 | 0.009 | 0.0000 | 1.6x10-7 |
| 29 | *PDK1L1* | NM\_138295.5:c.731C>T(p.Pro244Leu) | Missense | 0.05 | 7.61 | 787669 | 4.20x10-3 | 0.027 | 0.0060 | 7.9x10-3 |
| NM\_138295.5:c.6473+2\_6473+3del | frameshift deletion | . | . | 235796 | 4.00x10-4 | 0.009 | 0.0003 | 4.6x10-4 |
| 9 | *RNF123* | NM\_022064.5:c.986G>A(p.Arg329His) | Missense | 0.14 | 22.80 | . | 3.19x10-5 | 0.009 | 0.0000 | 2.7x10-7 |
| NM\_022064.5:c.3485G>A(p.Arg1162His) | Missense | 0.07 | 20.30 | . | 2.39x10-5 | 0.009 | 0.0000 | 2.7x10-7 |
| 22 | *RNF123* | NM\_001256071.3:c.1815A>T(p.Lys605Asn) | Missense | 0.01 | 0.03 | 710511 | 3.00x10-4 | 0.018 | 0.0003 | 2.7x10-9 |
| NM\_001256071.3:c.5707G>A(p.Glu1903Lys) | Missense | 0.34 | 24.90 | 1438428 | 3.60x10-5 | 0.009 | 0.0000 | 1.6x10-7 |

†Average alternate allele frequency based on gnomAD v2.1.1 database

‡P-values based on single variant case-control Score test adjusting for biological sex and the first five principal components

Abbreviations: AAF, alternate allele frequency; REVEL, rare exome variant ensemble learner score; CADD, phred-scaled combined annotation dependent depletion score

**Supplementary Table 4. Rare non-synonymous variants in *IFRD2* gene identified from the sequence kernel-based association analysis from exome sequencing data of children with biliary atresia enrolled in the National Birth Defects Prevention Study, 1997-2011**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variant** | **Variant type** | **AAF†** | **Case AAF** | **Control AAF** | **p-value‡** |
| NM\_006764.5:c.1016C>T(p.Ser339Phe) | missense | 0 | 0.009 | 0 | 3.8x10-8 |
| NM\_006764.5:c.427G>A(p.Gly143Ser) | missense | 3.62x10-5 | 0.009 | 0 | 1.3x10-6 |
| NM\_006764.5:c.791G>A(p.Arg264Gln) | missense | 4.37x10-4 | 0.009 | 0.0003 | 9.9x10-5 |
| NM\_006764.5:c.109G>T(p.Ala37Ser) | missense | 1.27x10-3 | 0.009 | 0.002 | 0.06 |
| NM\_006764.5:c.934A>G(p.Thr312Ala) | missense | 1.89x10-3 | 0.02 | 0.002 | 0.1 |
| NM\_006764.5:c.1313G>A(p.Arg438Gln) | missense | 4.90x10-3 | 0 | 0.008 | 0.3 |
| NM\_006764.5:c.237G>C(p.Glu79Asp) | missense | 1.24x10-2 | 0.03 | 0.02 | 0.3 |
| NM\_006764.5:c.962G>A(p.Arg321His) | missense | 3.62x10-3 | 0.009 | 0.004 | 0.4 |
| NM\_006764.5:c.181G>T(p.Gly61Trp) | missense | 0 | 0 | 0.0003 | 0.5 |
| NM\_006764.5:c.647A>G(p.Tyr216Cys) | missense | 4.09x10-3 | 0 | 0.003 | 0.6 |
| NM\_006764.5:c.1000C>T(p.Arg334Cys) | missense | 1.28x10-3 | 0 | 0.0003 | 0.7 |
| NM\_006764.5:c.1033T>C(p.Cys345Arg) | missense | 2.56x10-4 | 0 | 0.001 | 0.8 |
| NM\_006764.5:c.697G>A(p.Gly233Ser) | missense | 1.02x10-5 | 0 | 0.0007 | 0.8 |
| NM\_006764.5:c.304G>A(p.Ala102Thr) | missense | 2.60x10-4 | 0 | 0.0007 | 0.8 |
| NM\_006764.5:c.1072T>C(p.Tyr358His) | missense | 6.87x10-5 | 0 | 0.0003 | 0.8 |
| NM\_006764.5:c.671G>T(p.Ser224Ile) | missense | 3.50x10-4 | 0 | 0.0003 | 0.9 |
| NM\_006764.5:c.347G>A(p.Arg116His) | missense | 3.15x10-4 | 0 | 0.0003 | 0.9 |
| NM\_006764.5:c.1091G>A(p.Arg364Gln) | missense | 2.02x10-5 | 0 | 0.0003 | 0.9 |
| NM\_006764.5:c.842G>A(p.Gly281Asp) | missense | 5.77x10-4 | 0 | 0.0003 | 0.9 |
| NM\_006764.5:c.1097G>A(p.Arg366Gln) | missense | 4.04x10-5 | 0 | 0.0003 | 0.9 |
| NM\_006764.5:c.343C>T(p.Arg115Cys) | missense | 2.42x10-5 | 0 | 0.0003 | 0.9 |

†Average alternate allele frequency based on gnomAD v2.1.1 database

‡P-values based on case-control single variant Score test adjusting for biological sex and the first five principal components

Abbreviations: AAF, alternate allele frequency

**Supplementary Table 5. Common variants identified from a genome-wide association analysis in exome sequencing data of children with biliary atresia enrolled in the National Birth Defects Prevention Study, 1997-2011**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Variant** | **Variant type** | **Gene** | **AAF†** | **Case AAF** | **Control AAF** | **p-value (adjusted‡)** |
| NM\_001845.6:c.3189A>T(p.Arg1063=) | synonymous | *COL4A1* | 0.354 | 0.56 | 0.36 | 1.29x10-5 (0.21) |
| NM\_001845.6:c.3183G>A(p.Gly1061=) | synonymous | *COL4A1* | 0.354 | 0.56 | 0.36 | 1.30x10-5 (0.21) |
| NM\_007374.3:c.421C>A(p.His141Asn) | missense | *SIX6* | 0.531 | 0.34 | 0.59 | 1.49x10-5 (0.24) |

†Average alternate allele frequency based on gnomAD v2.1.1 database

‡P-values adjusted for multiple testing using Bonferroni correction

Abbreviations: AAF, alternate allele frequency

**Supplementary Figure 1. Quantile-quantile plot based on the sequence kernel-based association test (SKAT) analysis among rare non-synonymous variants with significant gene highlighted in red, National Birth Defects Prevention Study, 1997-2011. Horizontal threshold line indicates the Bonferroni corrected p-value**.



**Supplementary Figure 2. Quantile-quantile plot based on the sequence kernel-based association test (SKAT) analysis among rare synonymous variants, National Birth Defects Prevention Study, 1997-2011. Horizontal threshold line indicates the Bonferroni corrected p-value**.



**Supplementary Figure 3. Quantile-quantile plot based on the genome-wide association analysis among common variants, National Birth Defects Prevention Study, 1997-2011. Horizontal threshold line indicates the Bonferroni corrected p-value**.

