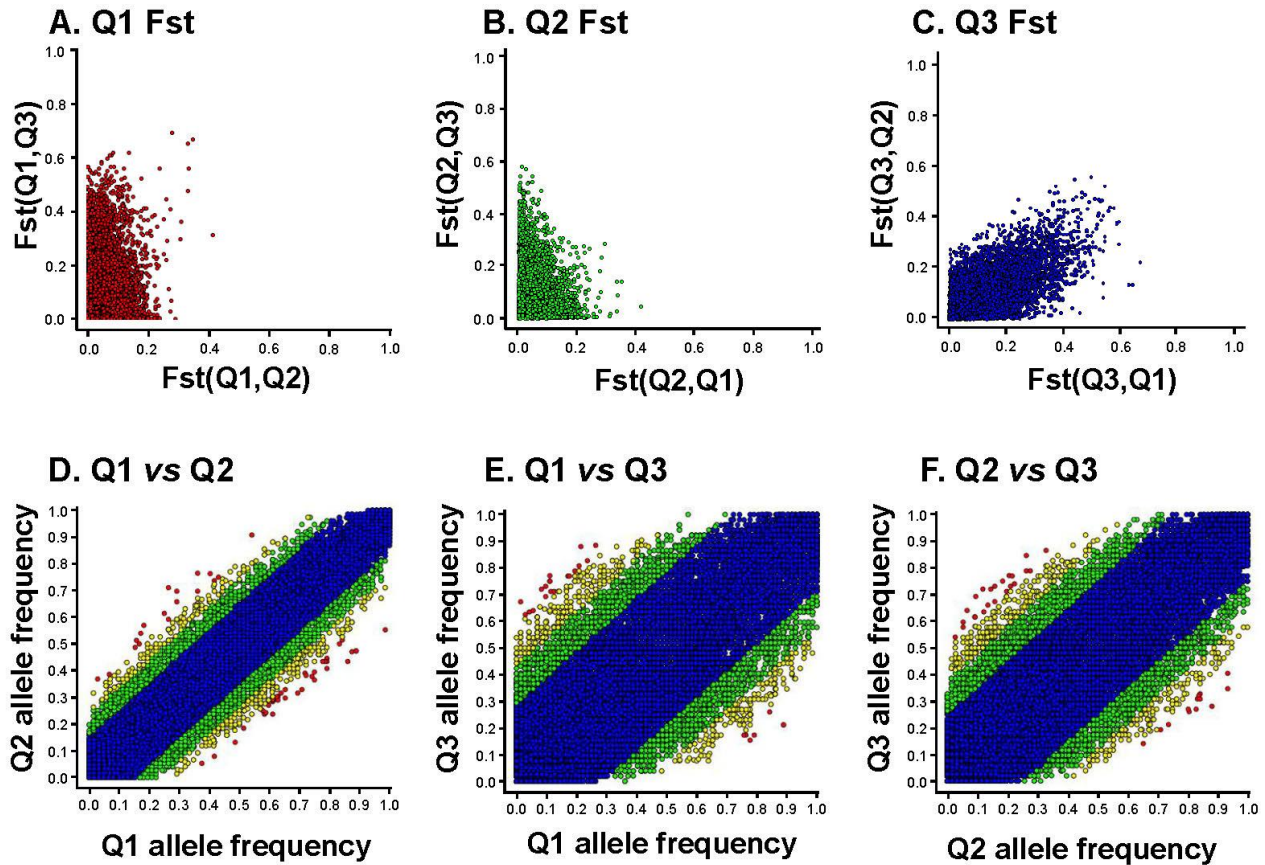
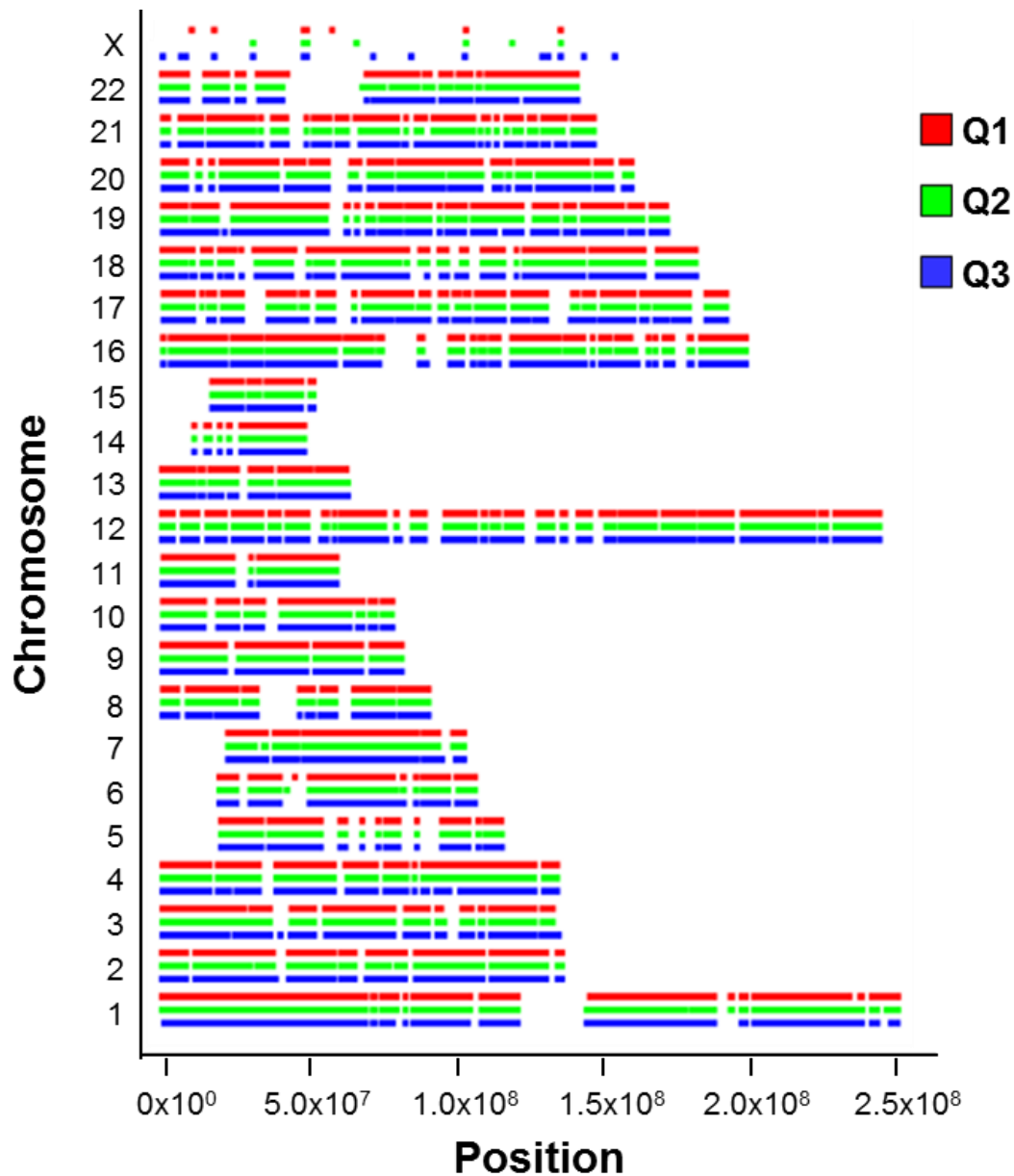


Supp. Figure S1. Site frequency spectrums for the total Qatari population. Shown in log₁₀ scale is the number of variants in the Qatari population on the Y-axis, and the minor allele frequency in Qatari on the X-axis.



Supp. Figure S2. Comparison of the allele frequencies among the Qatari Q1, Q2 and Q3 genetic groups. To identify variants that differentiate Q1, Q2 and Q3 Qatari subpopulations, the allele frequency was compared for 95,840 coding variants using two methods. **A-C.** F_{st} ; and **D-F.** allele frequency difference. **A-C.** F_{st} was calculated for comparison of Q1 vs Q2, Q1 vs Q3, and Q2 vs Q3. **A.** is the F_{st} of Q1 compared to Q2 (x axis) and Q3 (y axis); **B.** F_{st} of Q2 compared to Q1 (x axis) and Q3 (y axis); and **C.** F_{st} of Q3 compared to Q1 (x axis) and Q2 (y axis). **D-F.** The standard deviation in absolute value of allele frequency calculated for each pair of populations. **D.** Q1 vs Q2; **E.** Q1 vs Q3; and **F.** Q2 vs Q3. The variants were classified into 4 bins based on standard deviations; variants with allele frequency difference <4 (blue), 4-6 (green), 6-8 (yellow) and >8 (red) standard deviations.



Supp. Figure S3. The Qatari exome. In order to provide a major-allele reference exome for future studies of each Qatari subpopulation, variants were identified where the major allele in a Qatari subpopulation is the non-reference allele. Genome-wide distribution of 11,043 variants where the major allele in Q1 is the non-reference allele (red), 11,033 variants where the major allele in Q2 is the non-reference allele (green), and 11,183 variants where the major allele in Q3 is the nonreference allele (blue). The x axis is chromosome position and the y axis is the chromosome.

Supp. Table S1. SNP Mendelian disease variants directly genotyped in the panel of genetic tests conducted in the Laboratory Medicine and Pathology, Hamad Medical Corporation, Doha, Qatar¹

Disorder	Test
Fragile X	<i>FRAXA</i> (PCR & Southern blot analysis)
Cystic Fibrosis	Full <i>CFTR</i> sequencing of coding + exon-intron boundaries, CF29 platform, p.Ile1234Val, MLPA analysis
Spinal Muscular Atrophy	SMN1 and SMN2 dosage analysis + sequencing exons 7 & 8
Deafness	Connexin 26 (gap junction protein)
Wilson Disease	Full <i>ATP7B</i> exon sequencing + MLPA analysis
Achondroplasia	1138 G>A / 1138 G>C of <i>FGFR3</i> gene
Hypochondroplasia	1620C>A or 1620C>G of <i>FGFR3</i> gene
Craniosynostosis	Mutation analysis of fibroblast growth factor receptor, <i>FGFR</i> 1,2,3
Homocystinuria	Cystathionine β -synthase (<i>CBS</i>) deficiency due to p.Arg336Cys and p.Asp234Asn. Also <i>CBS</i> sequencing of coding + exon-intron boundaries
Hereditary Multiple Exostoses	Full sequencing of the <i>EXT1</i> , <i>EXT2</i> coding + exon-intron boundaries
Prader-Willi / Angelman Syndromes	Methylation analysis, <i>UBE3A</i> sequencing
Arterial tortuosity Syndrome	<i>SLC2A10</i> , p.Ser81Arg (c.243C>G)
Ehlers-Danlos Syndrome	<i>B4GALT7</i> , p.Arg270Cys (c.808C>T)
Hemoglobinopathies	Variants of hemoglobin <i>Hbs</i> (Sickle Cell Anemia) β -thalassemia: full hemoglobin, beta sequencing (UTR's, exon & intron) α -thalassemia: 3.7kb deletion and alpha T-Saudi
Thrombophilia	Factor V Leiden Prothrombin 20210 G>A <i>MTHFR</i> 677 C>T and 1298 A>C
Blood group genotyping	RhD genotyping (including heterozygosity testing)
Late infantile neuronal ceroid lipofuscinosis (LINCL)	Full <i>CLN</i> 5,6,8 sequencing of coding + exon-intron boundaries.
Male Infertility	Y chromosome microdeletions
Long QT syndrome	Full <i>KCNQ1</i> sequencing of coding + exon-intron boundaries
SRY/TDF analysis	SRY / X / Y genotyping
Patients infected with HCV	<i>IL28B</i> genotyping for pegIFN / RBV treatment decision making
Philadelphia chromosome	Chronic Myelocytic Leukemia (major CML) & acute lymphocytic leukemia (minor ALL) BCR/ABL test - available in mid 2012
RNA / DNA banking	From blood
DNA banking	From blood / buccal cells / tissue / Guthrie card / amniotic fluid / CVS / whole genome amplification
Tests upon request	For most disorders with known mutation(s). For inquiries please contact the laboratory.

¹ List of genetic tests provided by the Molecular Genetics Laboratory of Hamad Medical Corporation in Doha Qatar. Shown is the disorder and for each disorder a list of genes and specific mutations tested. For premarital genetic screening, tests include Cystic Fibrosis (p.Ile1234Val), Homocystinuria (p.Arg336Cys) and Spinal Muscular Atrophy with follow ups for beta- and alpha-thalassemia. For prenatal genetic diagnosis, tests include beta-thalassemia; spinal muscular atrophy; trisomies 13, 18, 21; X/Y genotyping; RhD genotyping; cystic fibrosis and Ehlers-Danlos Syndrome.

Supp. Table S2 is available as a separate Excel file under the Supporting Information for this article.

Supp. Table S3. Review of Literature for Variants in Table 3

GJB2 p.Trp24Ter – Gap junction protein, beta 2 (MIM# 121011), also known as connexin-26, functions in cell to cell channels with adherent cells.(Bruzzone et al., 1996) Mutations in GJB2 are associated with hereditary deafness (MIM# 220290). The GJB2 (c.286G>A) p.Trp24Ter variant (rs104894396) functions as an autosomal recessive previously identified in Pakistani and Indian families and in Spanish Gypsies.(Kelsell et al., 1997; Maheshwari et al., 2003; Alvarez et al., 2005) The GJB2 p.Trp24Ter variant was observed in the Q3 subpopulation. Autosomal recessive deafness was previously observed in other Arab populations.(Tadmouri et al., 2006l) [category - Auditory].

F5 p.Arg534Gln – Coagulation factor V (MIM# 612309), functions as a cofactor in the thrombinase complex; activated factor X requires calcium and activated factor V to convert prothrombin to thrombin.(Kane and Davie, 1986) A deficiency of F5 is associated with predisposition to hemorrhage, while some variants predispose to thrombosis. The F5 p.Arg506Gln variant (referred to in dbSNP as c.1601G>A, p.Arg534Gln, rs6025), observed only in the Q2 population, is the classic factor V Leiden variant that predisposes to thrombophilia (MIM# 188055). This variant is referred to as p.Arg534Gln or p.Arg506Gln, the numeric position discrepancy depends on inclusion of the 28 amino acid leader peptide.(Jenny et al., 1987) [category - Hematologic].

HBB p.Glu6Val, p.Glu26Lys, and p.Glu121Gln – Hemoglobin beta locus (MIM# 141900) encodes one of two polypeptides that combine to form adult hemoglobin (HbA). Three autosomal recessive disease variants were observed in the Qatari at the HBB locus, including p.Glu6Val (referred to in dbSNP as c.70A>T, p.Glu7Val, rs77121243) associated with sickle cell anemia (HbS; MIM# 603903), p.Glu26Lys (referred to in dbSNP as c.129G>A, p.Glu27Lys, rs33950507) associated with hemoglobin E beta-plus-thalassemia (MIM# 613985), and p.Glu121Gln (referred to in dbSNP as c.414C>G, p.Glu122Gln, rs33946267) associated with hemoglobin D Punjab. The single-position discrepancy for the three variants is due to the standard of counting the first methionine, a standard adopted after the variants were discovered. Hence the literature refers to Glu6Val but the database refers to Glu7Val; these are the same mutations.(Bender and Hobbs, 2012) The p.Glu6Val variant is the classic sickle cell anemia variant, the biochemical manifestation is sickle-shaped red blood cells, and the clinical manifestation is related to poor circulation and includes bone pain, fatigue, and

jaundice. Sickle cell anemia has been previously observed in the Q2 and Q3 subpopulations, and was previously observed in many other Arab population.(Tadmouri et al., 2006k) The Glu26Lys variant causes beta-plus-thalassemia, characterized biochemically an imbalance in globin chain production and abnormal erythropoiesis and characterized clinically by paleness, abnormal development and premature death. This variant was observed only in the Q3 subpopulation, and was not observed in 1000 Genomes. Hemoglobin D Punjab is the most common abnormal hemoglobin in Asia, consistent with it being observed in the Q2 subpopulation. The clinical manifestation is of atypical hemolytic anemia. [category - Hematologic].

MPL p.Lys39Asn - Myeloproliferative leukemia virus oncogene (MIM# 159530) encodes the receptor for thrombopoietin. Heterozygosity is associated with thrombocytosis (MIM# 601977) and homozygotes have severe thrombocytosis. This variant (c.162G>T, p.Lys39Asn, rs17292650) was found only in the Q3 population, consistent with the knowledge that it common in the African-American population. [category - Hematologic].

MEFV p.Pro369Ser, p.Met694Val – The Mediterranean fever protein (“pyrin”) (MIM# 249100) is expressed in leukocytes and functions in inflammation and host defense.(Centola et al., 2000) Mutations in MEFV cause Mediterranean fever, sometimes in association with systemic amyloidosis, as is the case with the common autosomal recessive (c.2120A>G) p.Met694Val variant (rs61752717), with renal failure as a common complication. The MEFV p.Met694Val variant was observed only in the Q2 subpopulation, and the (c.1145C>T) p.Pro369Ser variant (rs11466023) was observed only in the Q1 population. Familial Mediterranean fever has been previously observed in many other Arab populations.(Tadmouri et al., 2006j) [category - Hematologic].

EVC p.Arg443Gln – Elis-van Creveld (MIM# 604831) variants are the cause of Ellis-van Creveld syndrome (MIM# 225500), an autosomal recessive disorder associated with skeletal dysplasia characterized by short limbs, short ribs, postaxial polydactyly and dysplastic nails and teeth.(Ruiz-Perez et al., 2000) The EVC protein acts together with EVC2 as a positive regulator of the IHH (Indian Hedgehog) signaling pathway of the cartilage growth plate.(Ruiz-Perez et al., 2007) The (c.1512G>A) p.Arg443Gln variant (rs35953626), found in the Q2 and Q3 subpopulations, functions as an autosomal dominant.(Ruiz-Perez et al., 2000) Ellis-van Creveld Syndrome has been previously observed in other Arab populations.(Tadmouri et al., 2006i)-[category - Bone].

TGIF p.Glu107Leu – Transforming growth factor-beta-induced factor (MIM# 602630) is a homeobox protein with DNA-binding and transcription factor activity.(Bertolino et al., 1995) The (c.636A>T) p.Glu107Leu variant (rs28939693) is an autosomal recessive associated with holoprosencephaly (MIM# 142946). The TGIF1 p.Glu107Leu variant was specific for the Q1 subpopulation. Holoprosencephaly was previously observed in other Arab populations.(Tadmouri et al., 2006h) [category - Bone].

WNT10A p.Phe228Ile – Wingless-type mouse mammary tumor virus integration site family, member 10A (MIM# 606268), a member of the WNT gene family, is a secreted signaling molecule that binds to G protein coupled receptors with functions in development.(Rawadi and Roman-Roman, 2005) The (c.1145T>A) p.Phe228Ile variant (rs121908120) is an autosomal dominant associated with odontoonychodermal dysplasia,(Kantaputra and Sripathomsawat, 2011) with clinical manifestation including anomalous teeth, skin, hair and nails. The WNT10A p.Phe228Ile variant was found only in the Q1 subpopulation. [category - Bone].

ABCA4 p.Val931Met – ATP-binding cassette, subfamily A, member 4 (MIM# 601691) is a member of the ATP-binding cassette superfamily transmembrane protein expressed exclusively in retinal photoreceptors involved in clearance of all-trans-retinal aldehyde, a byproduct of the retinoid cycle of vision. The (c.2895G>A) p.Val931Met variant (rs58331765), associated with Stargardt macular dystrophy (MIM# 248200), was observed only in the Q3 subpopulation. Stargardt disease has been previously observed in other Arab populations.(Tadmouri et al., 2006g) [category - Eye].

CERKL p.Arg257Ter – ceramide kinase-like (MIM# 608381) converts the sphingolipid metabolite ceramide into ceramide 1-phosphate, a key product related to cell apoptosis and survival. The (c.870C>T) p.Arg257Ter variant (rs121909398) is associated with autosomal recessive retinitis pigmentosa (MIM# 608380). The p.Arg257Ter variant prematurely truncates the protein within the predicted catalytic domain. This variant was found only in the Q2 subpopulation. [category - Eye].

CYP1B1 p.Arg368His – Cytochrome p450, subfamily I, polypeptide I (MIM# 601771) functions to metabolize steroid hormones and is expressed during eye development.(Vasiliou and Gonzalez, 2008) The (c.1506G>A) p.Arg368His variant (rs79204362) is linked to primary congenital glaucoma

(MIM# 231300) in Saudi Arabians.(Bejjani et al., 2000) It is also observed in a digenic form of juvenile-onset glaucoma, in combination with mutations in MYOC.(Vincent et al., 2002) Primary congenital glaucoma was previously observed in other Arab populations.(Tadmouri et al., 2006f) [category - Eye].

RP1 p.Thr373Ile – Oxygen-regulated photoreceptor protein 1 (MIM# 603937) is a microtubule-binding protein expressed in retinal photoreceptors.(Liu et al., 2002) Variants of RP1 are associated with retinitis pigmentosa (MIM# 180100). Most RP1 variants function mostly as autosomal dominants, but the (c.1266C>T) p.Thr373Ile variant (rs77775126) has been described as an autosomal recessive disorder in 2 consanguineous Pakistani families. Consistent with this, the RP1 p.Thr373Ile variant was specific for the Q2 subpopulation in the 100 Qatari exomes, and also observed at 0.015 prevalence in the validation cohort of Q1. [category - Eye].

CAV3 p.Thr78Met – Caveolin 3 (MIM# 601253) is a caveolin family member that functions as a component of caveolae plasma membranes.(Minetti et al., 1998) The (c.310C>A) p.Thr78Met variant (rs72546668) functions as an autosomal dominant that causes long QT syndrome 9 (MIM# 611818).(Vatta et al., 2006) The CAV3 p.Thr78Met variant was found in the Q1 and Q2 subpopulations. [category - Cardiovascular].

NKX2-5 p.Arg25Cys – Homeobox protein NKX -2.5 (MIM# 600584) plays a role in cardiac development. The (c.302C>T) p.Arg25Cys variant (rs28936670) is inherited as an autosomal recessive associated variably with tetralogy of Fallot, interrupted aortic arch, truncus arteriosus, hypoplastic left heart and valvular defects (MIM# 187500). The NKX-2.5 p.Arg25Cyl variant was found in the Q1 and Q3 subpopulations. Tetralogy of Fallot was previously observed in other Arab populations.(Tadmouri et al., 2006e) [category - Cardiovascular].

SLC2A10 p.Ser81Arg – Solute carrier family 2 (facilitated glucose transporter), member 10 (MIM# 606145) is a member of the glucose transporter family.(McVie-Wylie et al., 2001) The (c.340C>G) p.Ser81Arg variant (rs121908120) is an autosomal recessive associated with the arterial tortuosity syndrome (MIM# 208050), previously described in 10 Qatari, 8 in the same Bedouin tribes (e.g., likely Q1)(Faiyaz-Ul-Haque et al., 2008). Consistent with this, the SLC2A10 p.Ser81Arg variant was specific for the Q1 subpopulation. Arterial tortuosity was previously reported in Qatar and other Arab

populations.(Tadmouri et al., 2006d) [category - Cardiovascular].

NCF2 p.Arg395Trp – Neutrophil cytosolic factor 2 (MIM# 608515), also referred to as “p67-phox” for “phagocyte oxidase”, is a component of the NADPH oxidase complex. The (c.1458C>T) p.Arg395Trp variant (rs13306575) has been associated with autosomal recessive chronic granulomatous disease (MIM# 233710). This was found only in the Q3 subpopulation. Autosomal recessive chronic granulomatous disease has been previously observed in other Arab populations.(Tadmouri et al., 2006c) [category - Inflammatory].

NLRP12 p.Arg284Ter – NLR family, pyrin domain containing 12 (MIM# 609648) implicated in the activation of proinflammatory caspases as part of the inflammasome complex.(Tschopp et al., 2003) NLRP12 has a N-terminal pyrin domain.(Pinheiro et al., 2011) The (c.1070C>T) p.Arg284Ter variant (rs104895564) is an autosomal recessive associated with cold autoinflammatory syndrome-2 (MIM# 611762), with episodic fever, arthralgias, myalgias, and urticaria triggered by cold exposure.(Jeru et al., 2008) The NLRP12 p.Arg284Ter variant was found only in the Q3 population. [category -Inflammatory]

NLRP3 p.Val198Met – NLR family, pyrin domain-containing 3 (MIM# 606416) encodes a pyrin-like protein expressed predominantly in blood leukocytes. The (c.1344G>A) p.Val198Met variant (rs121908147) is linked to familial cold autoinflammatory syndrome (MIM# 120100). [category – Inflammatory].

MASP2 p.Asp105Gly – Mannan-binding lectin serine protease 2 (MIM# 605102) plays a role in innate immunity in an antibody-independent pathway by the MASP2 protein binding to carbohydrates, resulting in complement activation. Homozygote inheritance of the (c.380A>G) p.Asp105Gly variant (rs72550870) leads to MASP2 deficiency (MIM# 613791), and is associated with frequent infections, chronic inflammatory disease and pulmonary fibrosis. [category - Immunity].

SLC7A9 p.Ala182Thr – Solute carrier family 7 (cationic amino acid transporter, y+ system), member 9 (MIM# 604144) is a cationic amino acid transporter that functions in the reabsorption of cysteine in the kidney tubules.(Mattoo and Goldfarb, 2008) The (c.661G>A) p.Ala182Thr variant

(rs79389353) is an autosomal recessive associated with mild cystinuria (MIM# 220100) and kidney stones, first described in 4 Spanish patients and 1 Italian patient.(Feliubadalo et al., 1999) The SLC7A9 p.Ala182Thr variant was specific for the Q2 subpopulation. Cystinuria was previously observed in other Arab populations.(Tadmouri et al., 2006b) [category - Kidney].

CIRH1A p.Arg565Trp –Cirhin (MIM# 607456) binds to the zinc-finger protein cirip; together they bind to NF-KB promoter regulatory sequences.(Yu et al., 2009) Homozygosity of the (c.1789C>T) p.Arg565Trp variant (rs119465999) weakens cirhin-cirip binding and is associated with North American Indian childhood cirrhosis (MIM# 604901), a severe autosomal recessive intrahepatic cholestasis(Chagnon et al., 2002) observed in the Ojibway-Cree tribe of Quebec.(Chagnon et al., 2002) [category –Liver].

CTNS p.Val42Ile – Cystinosin (MIM# 606272) is a lysosomal membrane protein strongly expressed in pancreas, kidney and skeletal muscle.(Taranta et al., 2012) The (c.594G>A) p.Val42Ile variant (rs35086888) is associated with a mild form of cystinosis (MIM# 219800). In its most severe form, cystinosis is characterized by proximal renal tubulopathy and end stage renal disease in infants due to intra-lysosomal accumulation of cysteine.(Schnaper et al., 1992) The p.Val42Ile variant is localized in the N-terminus of the protein and partially alters CTNS transport capacity. Given the mild severity of the phenotype, this disease is classified as recessive. Nephropathic cystinosis has been previously observed in other Arab populations.(Tadmouri et al., 2006m) [category - Metabolic].

KLF11 p.Thr220Met – Kruppel-like factor 11 (MIM# 603301) is a transcription factor enriched in the pancreas.(Niu et al., 2007) The (c.821C>T) p.Thr220Met variant (rs34336420) functions as an autosomal dominant associated with maturity-onset diabetes of the young, type 7 (MODY7; MIM# 610508). The variant was found in the Q2 and Q3 populations. Both of the Qatari heterozygous for this variant were diagnosed with type-2 diabetes. [category - Metabolic]

LPL p.Asp9Asn – Lipoprotein lipase (MIM# 609708) is the p.Asp9Asn variant (referred to in dbSNP as c.476G>A, p.Asp36Asn, rs1801177) linked to familial combined hyperlipidemia (MIM# 144250) in an autosomal dominant inheritance mode. In a study of the Dutch population all carriers of the p.Asp9Asn variant had elevated levels of cholesterol, apoB, and triglycerides. [category - Metabolic].

FECH p.Met267Ile and p.Gly55Cys – Ferrochelatase (heme synthetase) (MIM# 612386) is an enzyme that catalyzes the terminal step in the biosynthesis of heme, catalyzing the insertion of iron into protoporphyrin IX converting into heme.(Dailey, 2002) Mutations of FECH, such as the (c.918G>A) p.Met267Ile (rs118204037) and the (c.280G>T) p.Gly55Cys (rs3848519) variants, are associated as autosomal recessives with erythropoietic protoporphyria (MIM# 177000), a disease characterized by elevated protoporphyrin levels in red blood cells and photosensitivity to visible light. The FECH p.Met267Ile variant was observed only in the Q1 subpopulation. [category - Metabolic].

TYR p.Pro406Leu – Tyrosinase (MIM# 606933) catalyzes the oxidation of tyrosine forming melanin.(Hearing and Ekel, 1976) Like other TYR mutations, the (c.1299C>T) p.Pro406Leu (rs104894313) variant of TYR causes type I oculocutaneous albinism (MIM# 606952). The p.Pro406Leu variant has been previously reported in the Amish population.(Tripathi et al., 1992) In the Qataris, it was observed only in the Q2 population. Oculocutaneous albinism (MIM# 203100) has been previously observed in Qatar and in other Arab populations.(Tadmouri et al., 2006n) [category - Metabolic].

UPB1 p.Ala85Glu – Beta-ureidopropionase (MIM# 606673) catalyzes the last step in pyrimidine degradation. The (c.375C>A) p.Ala85Glu variant (rs34035085) is associated in an autosomal recessive mode with beta-ureidopropionase deficiency (MIM# 613161), a malfunction of the pyrimidine degradation pathway that affects cleavage of N-carbamyl-beta-alanine and results in convulsive crises and physical collapse. [category - Metabolic].

DYSF p.Ile1298Val – Dysferlin (MIM# 603009) is a member of the family of genes coding for transmembrane proteins related to calcium-dependent membrane fusion events; dysferlin plays a role in muscle fiber repair. The (c.4268A>G) p.Ile1298Val variant (rs121908954) is associated with Miyoshi myopathy (MIM# 254130), limb girdle type 2B muscular dystrophy. The p.Ile1298Val variant was observed as a compound heterozygote with p.Arg2042Cys, and is classified as an autosomal recessive.(Liu et al., 1998) Myoshi myopathy has been previously observed in other Arab populations.(Tadmouri et al., 2006a) [category - Muscular].

POMGT1 p.Asp556Asn – Protein O-mannose beta-1,2-N-acetylglucosaminyltransferase (MIM# 606822) participates in O-mannosyl glycan synthesis. The (c.1666G>A) p.Asp556Asn variant (rs74374973) is an autosomal recessive associated with limb-girdle type C3 muscular dystrophy-dystroglycanopathy (MIM# 613157). [category - Muscular].

AGTR2 p.Arg324Gln – Angiotensin II receptor, type 2 (MIM# 300034) is one of 2 major receptors for the blood pressure regulation hormone angiotensin. ARG2 is primarily expressed in brain and adrenal medulla, and functions to mediate programmed cell death.(Yamada et al., 1996) The (c.1178G>A) p.Arg324Gln variant (rs35474657) is associated with mental retardation in an X-linked inheritance mode (MIM# 300852). The variant and phenotype was observed only in males,(Vervoort et al., 2002) and in Qatari was observed in a Q3 female in the heterozygous state. [category – Neurologic].

RNASEH2C p.Arg69Trp – subunit C of the ribonuclease H2 complex (MIM# 610330) which cleaves ribonucleotides from RNA:DNA duplexes. The (c.385C>T) p.Arg69Trp variant (rs78635798) is an autosomal recessive associated with Aicardi-Goutieres syndrome 3 (MIM# 610329), an encephalopathy that manifests as calcification of basal ganglia, lymphocytosis of the cerebrospinal fluid, and abnormal white matter.(Crow et al., 2006) This variant has been previously described in Pakistani families, consistent with Qatari specificity within the Q2 population. [category - Neurologic].

BMP15 p.Ala180Thr – Bone morphogenetic protein 15 (MIM# 300247) is a growth factor expressed in oocytes during follicular development.(Otsuka et al., 2000) The (c.587G>A) p.Ala180Thr variant (rs104894767) is an X-linked dominant associated with premature ovarian failure (MIM# 300510). The disease allele was previously observed in the heterozygous state in Indian women, consistent with presence in the Q2 population. [category - Reproductive].

POF1B p.Arg329Gln – Actin binding protein (MIM# 300603) The (c.1132G>A) p.Arg329Gln variant (rs75398746) is an X-linked recessive associated with premature ovarian failure (MIM# 300604). The variant was previously observed in the homozygous state in Lebanese women, consistent with presence in the Q1 population. [category - Reproductive]

Supp. Table S4. OMIM + HGMD Variants Excluded¹

Gene	Amino acid change²	DbSNP rsID	Disease allele frequency in Qatar³	Reason for exclusion⁴
ABCA1	p.Arg219Lys	rs2230806	0.475	Greater than 5% disease allele frequency in Qatari
ABCA4	p.Gly1961Glu	rs1800553	0.025	Association / risk / susceptibility variant
ABCA4	p.Arg943Gln	rs1801581	0.010	Association / risk / susceptibility variant
ABCB1	3435C-T	rs1045642	*	Disease allele unclear
ABCC6	p.Arg1268Gln	rs2238472	0.165	Greater than 5% disease allele frequency in Qatari
ACADS	p.Gly185Ser	rs1799958	0.258	Greater than 5% disease allele frequency in Qatari
ACADVL	p.Pro65Leu	rs28934585	0.025	Compound with other variants in cases
ADA	p.Lys80Arg	rs11555566	0.130	Greater than 5% disease allele frequency in Qatari
ADRB2	p.Gln27Glu	rs1042714	0.261	Greater than 5% disease allele frequency in Qatari
ADRB2	p.Thr164Ile	rs1800888	0.010	Molecular basis not known for phenotype
AGT	p.Met235Thr	rs699	0.612	Greater than 5% disease allele frequency in Qatari
AGTR2	*	rs12917810	*	SNPEFF and HGMD disagree
ALAD	p.Lys59Asn	rs1800435	0.085	Greater than 5% disease allele frequency in Qatari
AMPD1	p.Gln12Ter	rs17602729	0.026	Compound with other variants in cases
ANK2	p.Leu1622Ile	rs35530544	0.010	Observed in controls
APC	p.Glu1317Gln	rs1801166	0.020	Molecular basis not known for phenotype
APOC2	p.Lys55Gln	rs5126	0.010	Polymorphism with no known functional impact
APOC2	p.Glu38Lys	rs5122	0.005	Molecular basis not known for phenotype
APOC3	p.Arg19Ter	rs76353203	0.005	Association / risk / susceptibility variant
ASPA	p.Tyr231Ter	rs12948217	*	Disease allele unclear
ATM	p.Met1040Val	rs3092857	0.005	Somatic mutation in cancer
ATP6V0A4	p.Met580Thr	rs3807153	0.202	Greater than 5% disease allele frequency in Qatari
ATP8B1	p.Asp70Asn	rs34719006	0.027	Observed in controls
BBS1	p.Glu234Lys	rs35520756	0.020	Compound with other variants in cases
BBS2	p.Asn70Ser	rs4784677	0.005	Snpeff and OMIM disagree
BCHE	p.Ala539Thr	rs1803274	0.197	Greater than 5% disease allele frequency in Qatari
BCHE	p.Asp70Gly	rs1799807	0.005	Molecular basis not known for phenotype
BCHE	p.Gly390Val	rs28933390	0.005	Molecular basis not known for phenotype

Gene	Amino acid change²	DbSNP rsID	Disease allele frequency in Qatar³	Reason for exclusion⁴
BDNF	p.Val66Met	rs6265	0.125	Greater than 5% disease allele frequency in Qatari
BDNF	p.Thr2Ile	rs8192466	0.010	Observed in controls
BMP4	p.Arg287His	rs121912768	0.005	Modifier not causative
BRCA2	p.Asn372His	rs144848	0.361	Greater than 5% disease allele frequency in Qatari
BTD	p.Ala171Thr	rs13078881	*	Disease allele unclear
BUB1B	p.Thr40Met	rs56079734	0.005	Somatic mutation in cancer
C1GALT1C1	p.Asp131Glu	rs17261572	0.101	Greater than 5% disease allele frequency in Qatari
C2	p.Glu318Asp	rs9332739	0.040	Association / risk / susceptibility variant
C3	p.Leu314Pro	rs1047286	0.890	Greater than 5% disease allele frequency in Qatari
C3	p.Arg102Gly	rs2230199	0.145	Greater than 5% disease allele frequency in Qatari
C8B	p.Arg374Ter	rs41286844	0.005	Compound with other variants in cases
CA2	p.Asn252Asp	rs2228063	0.017	Insufficient evidence to determine causality
CASP8	p.Asp302His	rs1045485	0.065	Greater than 5% disease allele frequency in Qatari
CASR	p.Ala986Ser	rs1801725	0.265	Greater than 5% disease allele frequency in Qatari
CAV3	p.Asn33Lys	rs1008642	*	Disease allele unclear
CAV3	p.Gly56Ser	rs72546667	0.010	Conflicting reports in literature
CETP	p.Ile405Val	rs5882	0.460	Greater than 5% disease allele frequency in Qatari
CFH	*	rs2274700	*	Disease allele unclear
CFH	p.Ile62Val	rs800292	0.714	Greater than 5% disease allele frequency in Qatari
CFH	p.Tyr402His	rs1061170	0.305	Greater than 5% disease allele frequency in Qatari
CFTR	p.Met470Val	rs213950	0.416	Greater than 5% disease allele frequency in Qatari
CLEC7A	p.Tyr238Ter	rs16910526	0.040	Association / risk / susceptibility variant
CLN6	p.Glu72Ter	rs104894483	*	Disease allele unclear
CPN1	p.Gly178Asp	rs61751507	0.060	Greater than 5% disease allele frequency in Qatari
CPOX	p.Arg447Cys	rs28931603	0.005	Insufficient evidence to determine causality
CTH	*	rs1021737	*	Disease allele unclear
CTH	p.Thr67Ile	rs28941785	0.005	Insufficient evidence to determine causality
CYB5R3	p.Thr117Ser	rs1800457	0.065	Greater than 5% disease allele frequency in Qatari
CYBA	p.His72Tyr	rs4673	0.429	Greater than 5% disease allele frequency in Qatari

Gene	Amino acid change ²	DbSNP rsID	Disease allele frequency in Qatar ³	Reason for exclusion ⁴
CYP11B2	p.Arg181Trp	rs61757294	*	Disease allele unclear
CYP2A6	p.Leu160His	rs1801272	0.041	Association / risk / susceptibility variant
CYP2C19	p.Trp212Ter	rs4986893	0.010	Drug metabolism variant not linked to disease
CYP2C9	p.Arg144Cys	rs1799853	0.122	Greater than 5% disease allele frequency in Qatari
CYP2C9	p.Ile359Leu	rs1057910	0.015	Drug metabolism variant not linked to disease
DBT	p.Gly323Ser	rs12021720	0.160	Greater than 5% disease allele frequency in Qatari
DMD	p.His2921Arg	rs1800279	0.036	Insufficient evidence to determine causality
DMD	p.Asn2912Asp	rs1800278	0.025	Compound with other variants in cases
DMD	p.Glu2910Val	rs41305353	0.025	Compound with other variants in cases
DPYD	p.Cys29Arg	rs1801265	0.276	Greater than 5% disease allele frequency in Qatari
EDAR	p.Val370Ala	rs3827760	0.015	Association / risk / susceptibility variant
EIF2B2	p.Arg183Ter	rs104894427	0.005	Compound with other variants in cases
ENPP1	p.Lys121Gln	rs1044498	0.290	Greater than 5% disease allele frequency in Qatari
ENPP1	p.Arg774Cys	rs28933977	0.025	Reclassified by OMIM as variant of unknown significance
EPHX1	p.Tyr113His	rs1051740	0.278	Greater than 5% disease allele frequency in Qatari
EPHX1	p.His139Arg	rs2234922	0.155	Greater than 5% disease allele frequency in Qatari
ERBB2	p.Val655Ile	rs1136201	0.867	Greater than 5% disease allele frequency in Qatari
F12	*	rs1801020	*	Disease allele unclear
F13A1	p.Val34Leu	rs5985	0.151	Greater than 5% disease allele frequency in Qatari
F13B	p.His95Arg	rs6003	0.190	Greater than 5% disease allele frequency in Qatari
FAH	p.Arg341Trp	rs11555096	0.011	Molecular basis not known for phenotype
FBN2	p.Val964Ile	rs154001	0.720	Greater than 5% disease allele frequency in Qatari
FGB	p.Arg448Lys	rs4220	0.145	Greater than 5% disease allele frequency in Qatari
FKTN	p.Gly125Ser	rs34006675	0.035	Reclassified by OMIM as variant of unknown significance
FMO3	p.Val257Met	rs1736557	0.117	Greater than 5% disease allele frequency in Qatari
FMO3	p.Glu308Gly	rs2266780	0.050	Greater than 5% disease allele frequency in Qatari
FSHR	p.Thr307Ala	rs6165	0.546	Greater than 5% disease allele frequency in Qatari
FSHR	p.Asn680Ser	rs6166	0.444	Greater than 5% disease allele frequency in Qatari
FUCA1	p.Gln281Arg	rs13551	0.192	Greater than 5% disease allele frequency in Qatari

Gene	Amino acid change²	DbSNP rsID	Disease allele frequency in Qatar³	Reason for exclusion⁴
FUT6	p.Glu247Lys	rs17855739	0.144	Greater than 5% disease allele frequency in Qatari
G6PD	p.Ser188Phe	rs5030868	0.025	Compound with other variants in cases
G6PD	p.Val68Met	rs1050828	0.021	Modifier not causative
GALT	p.Leu218Leu	rs2070075	*	Disease allele unclear
GALT	p.Asn314Asp	rs2070074	0.075	Greater than 5% disease allele frequency in Qatari
GCNT2	p.Ala169Thr	rs56106312	0.005	Gene not expressed in disease tissue
GDNF	p.Arg93Trp	rs36119840	0.005	Molecular basis not known for phenotype
GHR	p.Leu526Ile	rs6180	0.541	Greater than 5% disease allele frequency in Qatari
GIF	p.Gln5Arg	rs35211634	0.091	Greater than 5% disease allele frequency in Qatari
GJB2	p.Met34Thr	rs35887622	0.005	Observed in controls
GLA	p.Asp313Tyr	rs28935490	0.005	Molecular basis not known for phenotype
GNPTG	*	rs193302860	0.005	Snpeff and HGMD disagree
H6PD	p.Arg453Gln	rs6688832	0.280	Greater than 5% disease allele frequency in Qatari
HBB	p.His143Gln	rs36020563	*	Disease allele unclear
HBB	p.His2Gln	rs713040	*	Disease allele unclear
HEXB	p.Lys121Arg	rs11556045	0.122	Greater than 5% disease allele frequency in Qatari
HFE	p.His63Asp	rs1799945	0.108	Greater than 5% disease allele frequency in Qatari
HFE	p.Val53Met	rs28934889	0.005	Molecular basis not known for phenotype
HPD	p.Ala33Thr	rs1154510	0.095	Greater than 5% disease allele frequency in Qatari
HR	p.Thr1022Ala	rs7014851	0.085	Greater than 5% disease allele frequency in Qatari
HSD17B4	p.Arg106Pro	rs25640	*	Disease allele unclear
HTRA2	p.Gly399Ser	rs72470545	0.005	Observed in controls
IL10RB	p.Glu47Lys	rs2834167	0.645	Greater than 5% disease allele frequency in Qatari
IL7R	p.Ile138Val	rs1494555	0.374	Greater than 5% disease allele frequency in Qatari
INSR	p.Val985Met	rs1799816	0.025	Observed in controls
ITGB3	p.Leu33Pro	rs5918	0.167	Greater than 5% disease allele frequency in Qatari
KCNE2	p.Gln9Glu	rs16991652	0.010	Association / risk / susceptibility variant
KCNE2	p.Ile57Thr	rs74315448	0.010	Genotype of cases not specified
KCNJ11	p.Glu23Lys	rs5219	0.210	Greater than 5% disease allele frequency in Qatari

Gene	Amino acid change²	DbSNP rsID	Disease allele frequency in Qatar³	Reason for exclusion⁴
KDR	p.Cys482Arg	rs34231037	0.030	Observed in controls
KLKB1	p.Asn124Ser	rs3733402	0.540	Greater than 5% disease allele frequency in Qatari
LEPR	p.Lys656Asn	rs8179183	0.208	Greater than 5% disease allele frequency in Qatari
LEPR	p.Lys109Arg	rs1137100	0.105	Greater than 5% disease allele frequency in Qatari
LIG4	p.Thr9Ile	rs1805388	0.075	Greater than 5% disease allele frequency in Qatari
LIG4	p.Ala3Val	rs1805389	0.020	Association / risk / susceptibility variant
LPL	p.Asp250Asn	rs118204068	0.005	Compound with other variants in cases
LRP5	p.Val667Met	rs4988321	0.082	Greater than 5% disease allele frequency in Qatari
MECP2	p.Gly428Ser	rs61753971	0.005	Insufficient evidence to determine causality
MEFV	p.Pro369Ser	rs11466023	0.010	Reclassified by OMIM as variant of unknown significance
MEFV	p.Arg408Gln	rs11466024	0.010	Reclassified by OMIM as variant of unknown significance
MLH3	p.Val741Phe	rs28756990	0.020	Observed in controls
MSR1	p.Asp174Tyr	rs72552387	0.010	Observed in controls
MTHFR	*	rs1801131	*	Disease allele unclear
MTHFR	*	rs1801133	*	Disease allele unclear
MTP	p.Ile128Thr	rs3816873	0.392	Greater than 5% disease allele frequency in Qatari
MTRR	p.Ile22Met	rs1801394	0.392	Greater than 5% disease allele frequency in Qatari
MYO1A	p.Gly662Glu	rs33962952	0.035	Insufficient evidence to determine causality
NDUFV2	p.Ala29Val	rs906807	0.156	Greater than 5% disease allele frequency in Qatari
NOD2	p.Gly908Arg	rs2066845	0.035	Association / risk / susceptibility variant
NPHS2	p.Arg229Gln	rs61747728	0.005	Association / risk / susceptibility variant
NR3C1	p.Asn363Ser	rs56149945	0.005	Insufficient evidence to determine causality
NTRK1	p.Tyr604His	rs6336	0.967	Greater than 5% disease allele frequency in Qatari
OAT	p.Asn378Asn	rs11461	*	Disease allele unclear
OCA2	p.Arg305Trp	rs1800401	0.165	Greater than 5% disease allele frequency in Qatari
OPTN	p.Met98Lys	rs11258194	0.055	Greater than 5% disease allele frequency in Qatari
OTC	p.Lys46Arg	rs1800321	0.217	Greater than 5% disease allele frequency in Qatari
PARK2	p.Ala82Glu	rs55774500	0.005	Compound with other variants in cases
PARK7	p.Asp149Ala	rs74315352	0.005	Insufficient evidence to determine causality

Gene	Amino acid change²	DbSNP rsID	Disease allele frequency in Qatar³	Reason for exclusion⁴
PCSK1	p.Asn221Asp	rs6232	0.050	Greater than 5% disease allele frequency in Qatari
PHYH	p.Pro29Ser	rs28938169	0.061	Greater than 5% disease allele frequency in Qatari
PHKB	p.Met185Ile	rs56257827	0.005	Insufficient evidence to determine causality
PI	*	rs6647	*	Disease allele unclear
PI	p.Glu376Asp	rs1303	0.258	Greater than 5% disease allele frequency in Qatari
PI	p.Arg101His	rs709932	0.185	Greater than 5% disease allele frequency in Qatari
PLA2G7	p.Ala379Val	rs1051931	0.175	Greater than 5% disease allele frequency in Qatari
PLA2G7	p.Ile198Thr	rs1805018	0.121	Greater than 5% disease allele frequency in Qatari
PLA2G7	p.Val279Phe	rs76863441	0.005	Association / risk / susceptibility variant
PLAU	p.Pro141Leu	rs2227564	0.070	Greater than 5% disease allele frequency in Qatari
PNP	p.Ser51Gly	rs1049564	*	Disease allele unclear
PPARG	p.His449His	rs3856806	*	Disease allele unclear
PPARG2	p.Pro12Ala	rs1801282	0.045	Association / risk / susceptibility variant
PPP1R3	p.Asp905Tyr	rs1799999	0.088	Greater than 5% disease allele frequency in Qatari
PRNP	p.Glu219Lys	rs1800014	0.015	Association / risk / susceptibility variant
PRNP	p.Asn171Ser	rs16990018	0.010	Association / risk / susceptibility variant
PRODH	p.Thr466Met	rs2870984	0.011	Association / risk / susceptibility variant
PTH	p.Arg83Ter	rs6256	*	Disease allele unclear
PTPRJ	p.Gln276Pro	rs1566734	0.202	Greater than 5% disease allele frequency in Qatari
RET	p.Ala45Ala	rs1800858	*	Disease allele unclear
RHAG	p.Val270Ile	rs16879498	0.040	Compound with other variants in cases
RHCE	p.Pro226Ala	rs609320	0.806	Greater than 5% disease allele frequency in Qatari
RNASEL	p.Arg462Gln	rs486907	0.235	Greater than 5% disease allele frequency in Qatari
RP1	p.Asn985Tyr	rs2293869	0.400	Greater than 5% disease allele frequency in Qatari
RPGRIP1L	p.Ala229Thr	rs61747071	0.070	Greater than 5% disease allele frequency in Qatari
RPL5	p.Gly140Ser	rs121434406	0.005	Observed in controls
RYR1	p.Arg2676Trp	rs35180584	*	Disease allele unclear
SBDS	*	rs113993993	*	Disease allele unclear
SCNN1A	p.Trp493Arg	rs5742912	0.010	Observed in controls

Gene	Amino acid change ²	DbSNP rsID	Disease allele frequency in Qatar ³	Reason for exclusion ⁴
SDHB	p.Ser163Pro	rs33927012	0.005	Genotype of cases not specified
SDHD	p.Gly12Ser	rs34677591	0.010	Molecular basis not known for phenotype
SEMA4A	p.Arg713Gln	rs41265017	0.065	Greater than 5% disease allele frequency in Qatari
SERPINE1	p.Ala15Thr	rs6092	0.060	Greater than 5% disease allele frequency in Qatari
SLC45A2	p.Phe374Leu	rs16891982	0.909	Greater than 5% disease allele frequency in Qatari
SLC45A2	p.Glu272Lys	rs26722	0.210	Greater than 5% disease allele frequency in Qatari
SLC4A1	p.Lys56Glu	rs5036	0.075	Greater than 5% disease allele frequency in Qatari
SLC6A20	p.Thr199Met	rs17279437	0.010	Compound with other variants in cases
SPINK1	p.Asn34Ser	rs17107315	0.005	Molecular basis not known for phenotype
SPTA1	p.Asp791Glu	rs7418956	0.010	Observed in heterozygous state in cases
STOX1	p.Tyr153His	rs1341667	0.560	Greater than 5% disease allele frequency in Qatari
TAP1	p.Ile333Val	rs1057141	0.275	Greater than 5% disease allele frequency in Qatari
TAP1	p.Asp637Gly	rs1135216	0.255	Greater than 5% disease allele frequency in Qatari
TBG	p.Leu283Phe	rs1804495	0.078	Greater than 5% disease allele frequency in Qatari
TBG	p.Asp171Asn	rs1050086	0.010	Insufficient evidence to determine causality
TCN2	p.Pro259Arg	rs1801198	0.550	Greater than 5% disease allele frequency in Qatari
TF	p.Pro570Ser	rs1049296	0.207	Greater than 5% disease allele frequency in Qatari
TF	p.Gly277Ser	rs1799899	0.080	Greater than 5% disease allele frequency in Qatari
TG	p.Met1027Val	rs853326	0.586	Greater than 5% disease allele frequency in Qatari
TG	p.Ser734Ala	rs180223	0.571	Greater than 5% disease allele frequency in Qatari
TINF2	p.Ser245Tyr	rs142777869	0.024	Observed in controls
TLR4	p.Thr399Ile	rs4986791	0.040	Association / risk / susceptibility variant
TLR4	p.Asp299Gly	rs4986790	0.036	Association / risk / susceptibility variant
TNFRSF1A	p.Arg92Pro	rs4149584	*	Disease allele unclear
TNNI3	p.Pro82Ser	rs77615401	0.011	Modifier not causative
TOR1A	p.Asp216His	rs1801968	0.130	Greater than 5% disease allele frequency in Qatari
TPMT	p.Tyr240Cys	rs1142345	0.010	Insufficient evidence to determine causality
TPMT	p.Arg215His	rs56161402	0.005	Molecular basis not known for phenotype
TTR	p.Gly6Ser	rs1800458	0.010	Polymorphism with no known functional impact

Gene	Amino acid change ²	DbSNP rsID	Disease allele frequency in Qatar ³	Reason for exclusion ⁴
TTR	p.Val122Ile	rs76992529	0.005	Association / risk / susceptibility variant
TYR	p.Arg402Gln	rs1126809	0.130	Greater than 5% disease allele frequency in Qatari
TYR	p.Ser192Tyr	rs1042602	0.072	Greater than 5% disease allele frequency in Qatari
TYR	p.Arg217Trp	rs63159160	0.005	Compound with other variants in cases
UGT1A1	p.Gly71Arg	rs4148323	0.005	Association / risk / susceptibility variant
WFS1	*	rs10010131	*	Disease allele unclear
WISP3	p.Cys78Arg	rs17073260	0.985	Greater than 5% disease allele frequency in Qatari
WRD36	p.Ala449Thr	rs35703638	0.010	Observed in controls
WRD36	p.Asp658Gly	rs34595252	0.005	Observed in controls
ZEB1	p.Asn78Thr	rs80194531	*	Disease allele unclear
ZFP57	p.Cys241Ter	rs61730328	*	Disease allele unclear
ZFYVE27	p.Gly191Val	rs35077384	0.020	Conflicting reports in literature

¹ The exomes of 100 Qataris were sequenced, and variants present in dbSNP were annotated using OMIM (<http://omim.org>). Out of 251 variants that were linked to a record in the OMIM or HGMD database, the list was pruned down to 37 variants based on a combination of computational and manual filtering. Shown are the 214 excluded variants, including information on the gene, disease allele frequency in 100 Qatari exomes, and reason for exclusion.

² Amino acid substitution is shown for coding coding variants, with wild-type allele listed first, followed by polypeptide position and disease allele. (*) indicates variant is non-coding, the amino acid substitution is not specified in the OMIM record, or multiple databases disagree on the coding function.

³ Allele frequency is shown for the disease allele when the disease allele could be determined.

⁴ Variants were excluded based on one or more of the listed reasons: greater than 5% disease allele frequency in qatari (n=95), disease allele unclear (n=27), association / risk / susceptibility variant (n=20), compound with other variants in cases (n=13), observed in controls (n=14), molecular basis not known for phenotype (n=12), insufficient evidence to determine causality (n=11), reclassified by OMIM as variant of unknown significance (n=4), conflicting reports in literature (n=2), drug metabolism variant not linked to disease (n=2), genotype of cases not specified (n=2), modifier not causative (n=3), polymorphism with no known functional impact (n=2), somatic mutation in cancer (n=2), gene not expressed in disease tissue (n=1), observed in heterozygous state in cases (n=1), SNPEFF and OMIM (or HGMD) disagree (n=3).

Supp. References

Alvarez A, del C, I, Villamar M, Aguirre LA, Gonzalez-Neira A, Lopez-Nevot A, Moreno-Pelayo MA, Moreno F. 2005. High prevalence of the W24X mutation in the gene encoding connexin-26 (GJB2) in Spanish Romani (gypsies) with autosomal recessive non-syndromic hearing loss. *Am J Med Genet A* 137A:255-258.

Bejjani BA, Stockton DW, Lewis RA, Tomey KF, Dueker DK, Jabak M, Astle WF, Lupski JR. 2000. Multiple CYP1B1 mutations and incomplete penetrance in an inbred population segregating primary congenital glaucoma suggest frequent de novo events and a dominant modifier locus. *Hum Mol Genet* 9:367-374.

Bender MA, Hobbs W. 2012. *Sickle Cell Disease*.

Bertolino E, Reimund B, Wildt-Perinic D, Clerc RG. 1995. A novel homeobox protein which recognizes a TGT core and functionally interferes with a retinoid-responsive motif. *J Biol Chem* 270:31178-31188.

Bruzzone R, White TW, Paul DL. 1996. Connections with connexins: the molecular basis of direct intercellular signaling. *Eur J Biochem* 238:1-27.

Centola M, Wood G, Frucht DM, Galon J, Aringer M, Farrell C, Kingma DW, Horwitz ME, Mansfield E, Holland SM, O'Shea JJ, Rosenberg HF, et al. 2000. The gene for familial Mediterranean fever, MEFV, is expressed in early leukocyte development and is regulated in response to inflammatory mediators. *Blood* 95:3223-3231.

Chagnon P, Michaud J, Mitchell G, Mercier J, Marion JF, Drouin E, Rasquin-Weber A, Hudson TJ, Richter A. 2002. A missense mutation (R565W) in cirhin (FLJ14728) in North American Indian childhood cirrhosis. *Am J Hum Genet* 71:1443-1449.

Crow YJ, Leitch A, Hayward BE, Garner A, Parmar R, Griffith E, Ali M, Semple C, Aicardi J, Babul-Hirji R, Baumann C, Baxter P, et al. 2006. Mutations in genes encoding ribonuclease H2 subunits cause Aicardi-Goutieres syndrome and mimic congenital viral brain infection. *Nat Genet* 38:910-916.

Dailey HA. 2002. Terminal steps of haem biosynthesis. *Biochem Soc Trans* 30:590-595.

Faiyaz-Ul-Haque M, Zaidi SH, Wahab AA, Eltohami A, Al-Mureikhi MS, Al-Thani G, Peltekova VD, Tsui LC, Teebi AS. 2008. Identification of a p.Ser81Arg encoding mutation in SLC2A10 gene of arterial tortuosity syndrome patients from 10 Qatari families. *Clin Genet* 74:189-193.

Feliubadalo L, Font M, Purroy J, Rousaud F, Estivill X, Nunes V, Golomb E, Centola M, Aksentijevich I, Kreiss Y, Goldman B, Pras M, et al. 1999. Non-type I cystinuria caused by mutations in SLC7A9, encoding a subunit (bo,+AT) of rBAT. *Nat Genet* 23:52-57.

Hearing VJ, Ekel TM. 1976. Mammalian tyrosinase. A comparison of tyrosine hydroxylation and melanin formation. *Biochem J* 157:549-557.

Jenny RJ, Pittman DD, Toole JJ, Kriz RW, Aldape RA, Hewick RM, Kaufman RJ, Mann KG. 1987. Complete cDNA and derived amino acid sequence of human factor V. *Proc Natl Acad Sci U S A* 84:4846-4850.

Jeru I, Duquesnoy P, Fernandes-Alnemri T, Cochet E, Yu JW, Lackmy-Port-Lis M, Grimprel E, Landman-Parker J, Hentgen V, Marlin S, McElreavey K, Sarkisian T, et al. 2008. Mutations in NALP12 cause hereditary periodic fever syndromes. *Proc Natl Acad Sci U S A* 105:1614-1619.

Kane WH, Davie EW. 1986. Cloning of a cDNA coding for human factor V, a blood coagulation factor homologous to factor VIII and ceruloplasmin. *Proc Natl Acad Sci U S A* 83:6800-6804.

Kantaputra P, Sripathomsawat W. 2011. WNT10A and isolated hypodontia. *Am J Med Genet A* 155A:1119-1122.

Kelsell DP, Dunlop J, Stevens HP, Lench NJ, Liang JN, Parry G, Mueller RF, Leigh IM. 1997. Connexin 26 mutations in hereditary non-syndromic sensorineural deafness. *Nature* 387:80-83.

Liu J, Aoki M, Illa I, Wu C, Fardeau M, Angelini C, Serrano C, Urtizberea JA, Hentati F, Hamida MB, Bohlega S, Culper EJ, et al. 1998. Dysferlin, a novel skeletal muscle gene, is mutated in Miyoshi myopathy and limb girdle muscular dystrophy. *Nat Genet* 20:31-36.

Liu Q, Zhou J, Daiger SP, Farber DB, Heckenlively JR, Smith JE, Sullivan LS, Zuo J, Milam AH, Pierce EA. 2002. Identification and subcellular localization of the RPI protein in human and mouse photoreceptors. *Invest Ophthalmol Vis Sci* 43:22-32.

Maheshwari M, Vijaya R, Ghosh M, Shastri S, Kabra M, Menon PS. 2003. Screening of families with autosomal recessive non-syndromic hearing impairment (ARNSHI) for mutations in GJB2 gene: Indian scenario. *Am J Med Genet A* 120A:180-184.

Mattoo A, Goldfarb DS. 2008. Cystinuria. *Semin Nephrol* 28:181-191.

McVie-Wylie AJ, Lamson DR, Chen YT. 2001. Molecular cloning of a novel member of the GLUT family of transporters, SLC2a10 (GLUT10), localized on chromosome 20q13.1: a candidate gene for NIDDM susceptibility. *Genomics* 72:113-117.

Minetti C, Sotgia F, Bruno C, Scartezzini P, Broda P, Bado M, Masetti E, Mazzocco M, Egeo A, Donati MA, Volonte D, Galbiati F, et al. 1998. Mutations in the caveolin-3 gene cause autosomal dominant limb-girdle muscular dystrophy. *Nat Genet* 18:365-368.

Niu X, Perakakis N, Laubner K, Limbert C, Stahl T, Brendel MD, Bretzel RG, Seufert J, Path G. 2007. Human Kruppel-like factor 11 inhibits human proinsulin promoter activity in pancreatic beta cells. *Diabetologia* 50:1433-1441.

Otsuka F, Yao Z, Lee T, Yamamoto S, Erickson GF, Shimasaki S. 2000. Bone morphogenetic protein-15. Identification of target cells and biological functions. *J Biol Chem* 275:39523-39528.

Pinheiro AS, Eibl C, Ekman-Vural Z, Schwarzenbacher R, Peti W. 2011. The NLRP12 pyrin domain: structure, dynamics, and functional insights. *J Mol Biol* 413:790-803.

Rawadi G, Roman-Roman S. 2005. Wnt signalling pathway: a new target for the treatment of osteoporosis. *Expert Opin Ther Targets* 9:1063-1077.

Ruiz-Perez VL, Blair HJ, Rodriguez-Andres ME, Blanco MJ, Wilson A, Liu YN, Miles C, Peters H, Goodship JA. 2007. Evc is a positive mediator of Ihh-regulated bone growth that localises at the base of chondrocyte cilia. *Development* 134:2903-2912.

Ruiz-Perez VL, Ide SE, Strom TM, Lorenz B, Wilson D, Woods K, King L, Francomano C, Freisinger P, Spranger S, Marino B, Dallapiccola B, et al. 2000. Mutations in a new gene in Ellis-van Creveld syndrome and Weyers acrorenal dysostosis. *Nat Genet* 24:283-286.

Schnaper HW, Cattel J, Merrill S, Marcusson E, Kissane JM, Shackelford GD, So SK, Nelson RD, Cole BR, Smith ML, . 1992. Early occurrence of end-stage renal disease in a patient with infantile nephropathic cystinosis. *J Pediatr* 120:575-578.

Tadmouri GO, Al Ali MT, Al-Haj AS, Al KN. 2006e. CTGA: the database for genetic disorders in Arab populations. *Nucleic Acids Res* 34:D602-D606.

Taranta A, Petrini S, Citti A, Boldrini R, Corallini S, Bellomo F, Levchenko E, Emma F. 2012. Distribution of cystinosis-LKG in human tissues. *Histochem Cell Biol* 138:351-363.

Tripathi RK, Strunk KM, Giebel LB, Weleber RG, Spritz RA. 1992. Tyrosinase gene mutations in type I (tyrosinase-deficient) oculocutaneous albinism define two clusters of missense substitutions. *Am J Med Genet* 43:865-871.

Tschopp J, Martinon F, Burns K. 2003. NALPs: a novel protein family involved in inflammation. *Nat Rev Mol Cell Biol* 4:95-104.

Vasiliou V, Gonzalez FJ. 2008. Role of CYP1B1 in glaucoma. *Annu Rev Pharmacol Toxicol* 48:333-358.

Vatta M, Ackerman MJ, Ye B, Makielski JC, Ughanze EE, Taylor EW, Tester DJ, Balijepalli RC, Foell JD, Li Z, Kamp TJ, Towbin JA. 2006. Mutant caveolin-3 induces persistent late sodium current and is associated with long-QT syndrome. *Circulation* 114:2104-2112.

Vervoort VS, Beachem MA, Edwards PS, Ladd S, Miller KE, de M, X, Clarkson K, DuPont B, Schwartz CE, Stevenson RE, Boyd E, Srivastava AK. 2002. AGTR2 mutations in X-linked mental retardation. *Science* 296:2401-2403.

Vincent AL, Billingsley G, Buys Y, Levin AV, Priston M, Trope G, Williams-Lyn D, Heon E. 2002. Digenic inheritance of early-onset glaucoma: CYP1B1, a potential modifier gene. *Am J Hum Genet* 70:448-460.

Yamada T, Horiuchi M, Dzau VJ. 1996. Angiotensin II type 2 receptor mediates programmed cell death. *Proc Natl Acad Sci U S A* 93:156-160.

Yu B, Mitchell GA, Richter A. 2009. Cirhin up-regulates a canonical NF-kappaB element through strong interaction with Cirip/HIVEP1. *Exp Cell Res* 315:3086-3098.