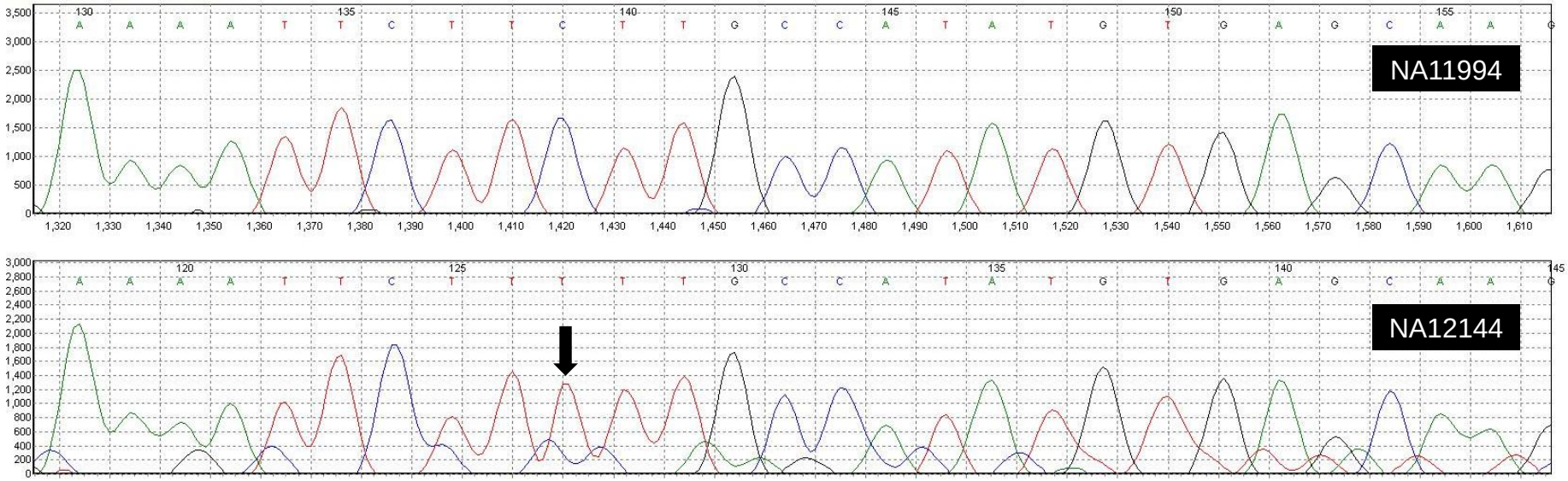
The American Journal of Human Genetics, *91*

Supplemental Data

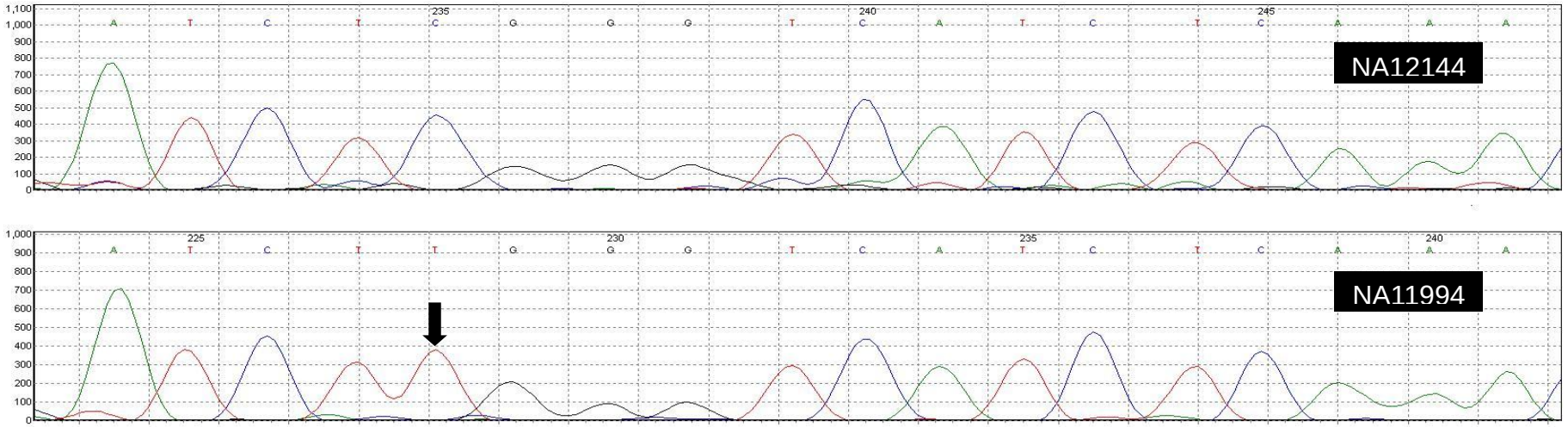
Deleterious- and Disease-Allele Prevalence in Healthy Individuals: Insights from Current Predictions, Mutation Databases, and Population-Scale Resequencing

Yali Xue, Yuan Chen, Qasim Ayub, Ni Huang, Edward V. Ball, Matthew Mort, Andrew D. Phillips, Katy Shaw, Peter D. Stenson, David N. Cooper, Chris Tyler-Smith, and the 1000 Genomes Project Consortium

A



B



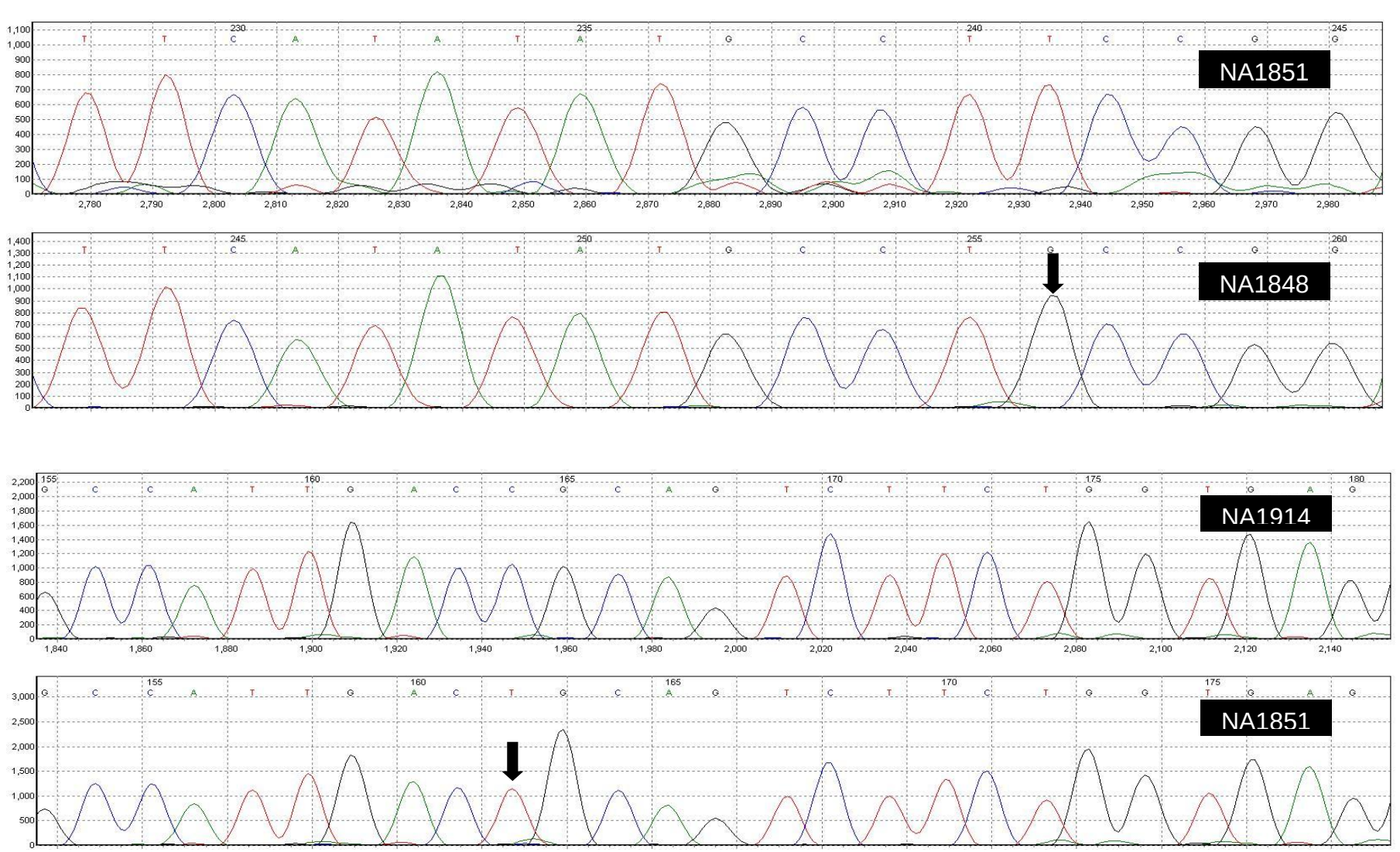
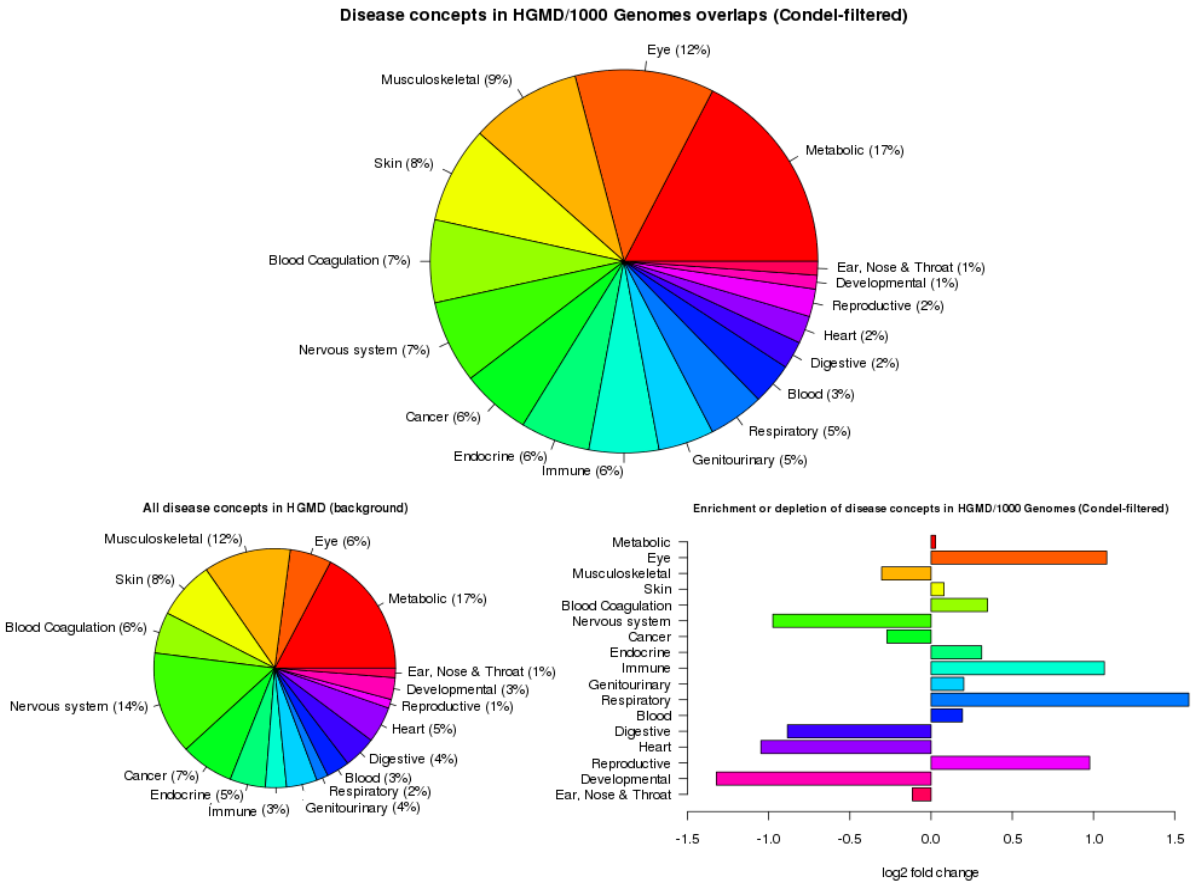


Figure S1. Capillary Sequencing Validation Results for Homozygous Damaging DMs

Disease mutations (arrows) for (A) CM00653, (B) CM022976, (C) CM025932 and (D) CM034771 in the samples indicated.



FigureS2. Disease Concepts in Overlaps of HGMD and 1000 Genomes (Condel Filtered)

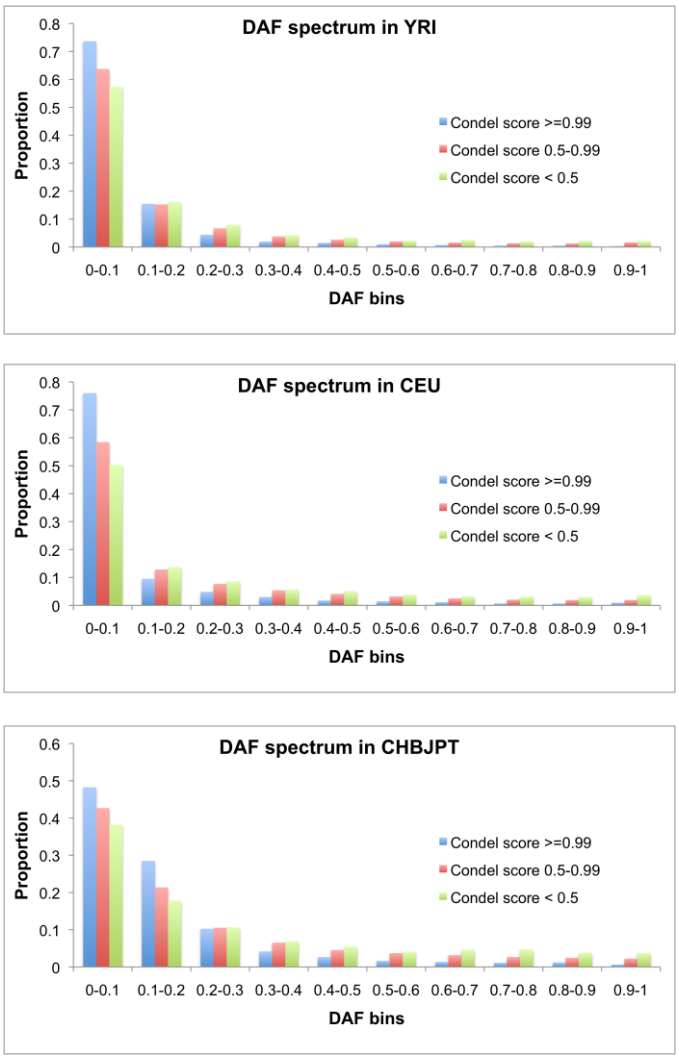


Figure S3. Derived Allele Frequency Spectra of Different Condel Classes

Table S5. Correlation of DAF with Condel Score

	Spearman				Kendall		
	rho	S	P value		tau	Z	P value
YRI	-0.148	2.60E+12	< 2.2e-16		-0.101	-22.8784	< 2.2e-16
CEU	-0.198	1.04E+12	< 2.2e-16		-0.136	-26.1619	< 2.2e-16
CHBJPT	-0.151	5.53E+11	< 2.2e-16		-0.103	-18.1524	< 2.2e-16

Supplemental Data on HGMD Variants

CM000653	Ataxia telangiectasia Leu-Phe	*ATM* 1420	c.4258C>T

HGMD - DM, no question mark in phenotype.

Original reference - Found in compound heterozygosity (described as L1420P), second mutation not identified. No screening of controls. Secondary reference reports association with breast cancer risk.

Disease - #208900 – autosomal recessive. Mental retardation, growth retardation, abnormal eye movements, skin lesions, and immune deficiency (ataxia becomes evident at the end of the first year of life, telangiectasia become evident between the second and eighth year of life).

Other data - Conserved in dog, chicken, mouse and African clawed frog (Valine in zebrafish). Predicted to be damaging by SIFT, predicted to be medium risk by MutPred.

Conclusion - Likely to be pathogenic. Change phenotype to "Breast cancer, increased risk".

CM001809	Usher syndrome 2a? Gly-Arg	*USH2A* 713	c.2137G>C

HGMD - DM, question mark in phenotype.

Original reference - Found in compound heterozygosity with frameshift, not identified in 100 control alleles.

Extra references - Identified in 2/200 controls (Nájera (2002) Hum Mutat 20: 76).
Loss of usherin/collagen IV binding in functional study (Bhattacharya (2004) J Cell Sci 117: 233).
Characterised as severe recessive disease-causing mutation (Bell (2011) Sci Transl Med 3: 65ra4). Identified in homozygosity in patient (Vozzi (2011) Mol Vis 17: 1662).
"Probable damaging effect" and "Damaging" according to PolyPhen-2 and SIFT (Song (2011) Invest Ophthalmol Vis Sci 52: 9053).

Disease - #276901 – autosomal recessive. Hearing loss and progressive retinitis pigmentosa.

Other data - No conservation data.
Predicted to be damaging by SIFT, predicted to be very high risk by MutPred.

Conclusion - Likely to be pathogenic (delete question mark in phenotype from HGMD).

CM005424	Cerebrotendinous xanthomatosis	*CYP27A1* Pro-Leu	c.1151C>T 384

HGMD - DM, no question mark in phenotype.

Original reference - Identified homozygously in one family, and heterozygously in three families. Not found in 100 control alleles. Conserved in rabbit, rat and mouse.

Extra references - Predicted to be deleterious by both PolyPhen and SIFT (Wang (2009) Drug Metab Dispos 37: 977).
Characterised as severe recessive disease-causing mutation (Bell (2011) Sci Transl Med 3: 65ra4).

Disease - #213700 – autosomal recessive. Juvenile cataracts, angina, respiratory insufficiency, osteoporosis, tuberous xanthomas, dementia, cerebellar ataxia, mental retardation, spasticity and peripheral neuropathy.

Other data - Conserved in pig, chicken, mouse and African clawed frog. Predicted to be damaging by SIFT, predicted to be very high risk by MutPred.

Conclusion - Likely to be pathogenic (checked as part of 500 Exomes overlap in December 2010).

CM010436 Diastrophic dysplasia *SLC26A2* c.1474C>T Arg-Trp 492

HGMD - DM, no question mark in phenotype.

Original reference - Identified in "multiple patients", not found in >100 controls.

Extra references - Characterised as severe recessive disease-causing mutation (Bell (2011) Sci Transl Med 3: 65ra4).

Disease - #222600 – autosomal recessive. Short-limb dwarfism identifiable at birth, hearing loss, cleft palate and skeletal abnormalities.

Other data - Conserved in dog, chicken, mouse and zebrafish.
Predicted to be damaging by SIFT, predicted to be high risk by MutPred.

Conclusion - Likely to be pathogenic.

CM014821 Colorectal cancer, non-polyposis ? *MLH1* c.2152C>T His-Tyr 718

HGMD - DM, question mark in phenotype.

Original reference - Identified in three cases and one control.

Extra references - Compromised function in yeast model (Weber (2001) AJHG 69: A416).
β-gal activity not significantly different from wild-type in functional assays (Kondo (2003) Cancer Res 63: 3302).
Mismatch repair activity similar to wild-type (Takahashi (2007) Cancer Res 67: 4595).
Binding to PMS2 and EXO1 comparable to wild-type (Ou (2007) Hum Mutat 28: 1047).
Identified in 1/932 cases and 1/1000 controls, described as benign (Barnetson (2008) Hum Mutat 29: 367).
No significant difference in mismatch repair compared to wild-type (Martinez (2010) Proc Natl Acad Sci U S A 107: 5070).

	Found in four controls and with MSH2 mutation in one case, described as likely nonpathogenic (Moussa (2011) Int J Colorectal Dis 26: 455).		
Disease -	#609310 – autosomal dominant. Early-age onset colorectal cancer and extra-colonic malignancies.		
Other data -	Conserved in dog, chicken, mouse and zebrafish. Predicted to be damaging by SIFT, no MutPred data.		
Conclusion -	Disease status unclear.		
CM015009	Photoreceptor protein deficiency?	*RP1* Arg-Term	c.5797C>T 1933
HGMD - Original reference -	DM, question mark in phenotype. Found only in control.		
Conclusion -	Disease status unclear (re-tag as FTV)		
CM020047	Enteropeptidase deficiency Ser-Term	*TMPRSS15* 712	c.2135C>G
HGMD -	DM, no question mark in phenotype.		
Original reference -	Compound heterozygosity with second nonsense mutation in two siblings. Predicted to result in abolishment of enzymatic function.		
Disease -	#226200 – autosomal recessive. Failure to thrive, diarrhoea, hypoproteinaemic oedema.		
Conclusion -	Likely to be pathogenic.		
CM022677	Loose anagen hair syndrome?	*KRT75* Glu-Lys	c.1009G>A 337
HGMD -	DM, question mark in phenotype.		
Original reference -	Identified in six affected and three unaffected individuals from three families. Not found in 50 controls.		
Disease -	#600628 - autosomal dominant with variable expressivity. **Loose** actively growing, relatively sparse, short **hair (childhood onset).**		
Other data -	Conserved in cow, chicken, mouse and chimp. Predicted to be damaging by SIFT, no MutPred data.		
Conclusion -	Disease status unclear.		

CM022976	Nephronophthisis 4	*NPHP4* Arg-Trp	c.2542C>T 848

HGMD - DM, no question mark in phenotype.
Original reference - Compound heterozygosity with nonsense mutation. Absent from 92–96 unaffected controls.
Extra references - Characterised as severe recessive disease-causing mutation (Bell (2011) Sci Transl Med 3: 65ra4).

Disease - #606966 – autosomal recessive. Growth retardation, anaemia, n**ephronophthisis, e**nd stage renal disease (age 6-35 years), renal tubular cell atrophy with corticomedullary cysts and renal interstitial fibrosis.

Other data - Conserved in dog, cow, mouse and chimp.
Predicted to be damaging by SIFT, predicted to be medium risk by MutPred.

Conclusion - Likely to be pathogenic.

CM025213	Dysfibrinogenaemia ? Pro-Leu	*FGB* 265	c.794C>T

HGMD - DM, question mark in phenotype.
Original reference - Identified in three patients, clot turbidity and lysis ~70% of wild-type in functional assay.
Extra references - Found in 5% of controls (Brennan (2000) Blood 95: 1709).

Disease - Defective fibrin clot formation, may lead to an increased tendency to bleeding or thrombosis.

Other data - Conserved in dog, chicken, mouse and chimp.
Predicted to be damaging by SIFT, predicted to be medium risk by MutPred.

Conclusion - Disease status unclear. Re-tag as DP.

CM025932	Cushing syndrome	*MC2R* Phe-Cys	c.833T>G 278

HGMD - DM, no question mark in phenotype.
Original reference - Associated with elevated basal cAMP accumulation compared to wild-type, impaired desensitization and internalization.

Disease - #202200 – autosomal recessive. Tall stature, failure to thrive, hyperpigmentation, seizures, coma, recurrent hypoglycaemic episodes, frequent and severe infections.

Other data - Conserved in dog, chicken, mouse and chimp (Leucine in cow).
No SIFT data, high risk in MutPred.

Conclusion - Likely to be pathogenic.

CM030964	Parkinson disease	*SNCAIP*	c.1861C>T
		Arg-Cys	621

HGMD - DM, no question mark in phenotype.
Original reference - Found in two patients and not in 702 control chromosomes. Decreased aggregation-forming capacity compared to wild-type.
Extra references - Identified in 4/600 patients and 10/824 controls (Myhre (2008) BMC Med Genet 9: 19).

Disease - Autosomal dominant and recessive forms. Parkinsonism, tremor, personality changes, depression and dementia.

Other data - Conserved in dog, chicken, mouse and chimp.
No SIFT data, medium risk in MutPred.

Conclusion - Disease status unclear (add question mark to phenotype in HGMD)

CM031328	Liver glycogenosis ?	*PHKB*	c.2309A>G
	Tyr-Cys	770	

HGMD - DM, question mark in phenotype.
Original reference - Found in one patient, not in 52 control chromosomes.
Extra references - Found in 2/40 controls

Disease - #261750 – autosomal recessive. Short stature (post natal onset), hepatomegaly, diarrhoea, hypotonia and mild muscle weakness.

Other data - Conserved in dog, chicken, mouse and chimp.
Predicted to be tolerated by SIFT, no MutPred data.

Conclusion - Disease status unclear.

CM032598	Cardiomyopathy, hypertrophic	*MYBPC3*	c.1519G>A
		Gly-Arg	507

HGMD - DM, no question mark in phenotype.
Original reference - Found in one patient, no family members available for study (*i.e.* co-segregation with HCM could not be demonstrated).

Disease - #115197 – autosomal dominant. Hypertrophic cardiomyopathy.

Other data - Conserved in dog, chicken, mouse and chimp.
Predicted to be damaging by SIFT, high risk in MutPred.

Conclusion - Likely to be pathogenic.

CM033370	Dihydropyrimidine dehydrogenase deficiency ?	*DPYD*	c.1601G>A	Ser-Asn 534

HGMD - DM, question mark in phenotype.
Original reference - Found in two patients (with other missense changes) and in heterozygosity in 3/157 controls.
Extra references - Statistically significant deviation from the median DPD activity of the population (Seck (2005) Clin Cancer Res 11: 5886).
Characterised as likely polymorphism (Bell (2011) Sci Transl Med 3: 65ra4).

Disease - #274270 – autosomal recessive. Failure to thrive, growth retardation, microcephaly, seizures, motor retardation, mental retardation, hyperactivity, autism (onset usually in infancy).

Other data - Conserved in dog, cow, mouse and chimp.
Predicted to be damaging by SIFT, no MutPred data.

Conclusion - Disease status unclear.

CM034771	Low phospholipid associated cholelithiasis	*ABCB4*	c.2363G>A	Arg-Gln 788

HGMD - DM, no question mark in phenotype.
Original reference - Found in heterozygosity (patient showed at least two symptoms from onset < 40 years of age, recurrence after cholecystectomy, presence of intrahepatic spots), not identified in 28 controls.

Disease - #602347 – autosomal recessive. Intrahepatic cholestasis, jaundice, hepatomegaly, cirrhosis, end-stage liver disease before adulthood, splenomegaly and diarrhoea.

Other data - Conserved in dog, cow, mouse and chimp.
Predicted to be damaging by SIFT, high risk in MutPred.

Conclusion - Likely to be pathogenic.

CM040093	Multiple epiphyseal dysplasia ?	*MATN3*	c.754G>A	Glu-Lys 252

HGMD - DM, question mark in phenotype.
Original reference - Identified in affected and unaffected members of four families and in 10/418 controls.

Disease - #607078 – autosomal dominant. Early onset osteoarthritis, multiple epiphyseal dysplasia, delayed carpal and tarsal ossification.

Other data - Conserved in dog, chicken, mouse and chimp.
Predicted to be damaging by SIFT, no MutPred data.

Conclusion - Disease status unclear (delete from HGMD).

CM041235 Adenosine monophosphate deaminase deficiency *AMPD1* c.959A>T Lys-Ile 320

HGMD - DM, no question mark in phenotype.

Original reference - Identified in 6.4% of patients and ~3% of controls. ~50% catalytic activity of wild-type in functional assay.

Disease - +102770 – autosomal dominant. Exercise-induced myopathy, limp infant, benign congenital hypotonia.

Other data - Conserved in dog, chicken, mouse and chimp.
Predicted to be damaging by SIFT, no MutPred data.

Conclusion - Disease status unclear (re-tag as DFP in HGMD).

CM041734 Muscular dystrophy, limb girdle *CAPN3* c.479C>G p.Ala160Gly

HGMD - DM, no question mark in phenotype, no contradictory extra references.

Original reference - Phenotype should be HyperCKemia. Asymptomatic.

Disease - #253600. Autosomal recessive. Elevated creatine kinase.

Other data - SIFT predicts damaging. MutPred predicts high risk.
Conserved in mouse, dog, cow, chicken.

Recommendation – Likely to be pathogenic.
Change phenotype to HyperCKemia. Retain as DM.

CM044022 Colorectal cancer and Nijmegen breakage syndrome *NBN* c.643C>T p.Arg215Trp

HGMD - DM, no question mark in phenotype, no contradictory extra references.

Original reference - Colorectal cancer. Found in a higher frequency in various cancers than in controls.

Extra references - Nibrin-Trp215, the product of the R215W allele was found to be much less abundant than wildtype nibrin-Arg215 as shown by Western blot analysis of lymphoblastoid cell extracts. This difference may indicate a lower expression level of the Trp allele, a shorter half life of the mRNA or, more likely, reduced stability of the nibrin-Trp215 protein. This might in turn reflect an inability of the mutant protein to associate correctly with its cellular partners, Mre11 and Rad50, and subsequent degradation of non-bound nibrin monomer. In cells with only a truncated nibrin as an alternative, the R215W mutation might

therefore lead to a reduction in active trimeric complex below a critical level (Seemanova (2006) J Med Genet 43: 218.
Results indicate that heterozygous c.657-661del, p.I171V and p.R215W mutations in the NBN gene do not significantly impair main nibrin functions. Cells containing a homozygous c.657-661del displayed a significantly higher number of chromatid breaks, residual c-H2AX foci and intra-S-phase checkpoint defects following irradiation, which resulted in increased radiosensitivity, cells with heterozygous c.657-661del, p.I171V and p.R215W mutations behaved similarly to control cells (Dzikiewicz-Krawczyk (2011), Mutagenesis (PMID:22131123).

Disease - #251260. Autosomal recessive. The Nijmegen breakage syndrome is a chromosomal instability syndromes characterised by microcephaly, growth retardation, immunodeficiency, and predisposition to cancer.

Other data - SIFT predicts damaging. MutPred predicts medium risk.
Conserved in mouse, dog, cow. His in chicken.

Conclusion - Likely to be pathogenic

CM050736 Usher syndrome 1f *PCDH15* c.4024C>A p.Gln1342Lys

HGMD - DM, no question mark in phenotype, no contradictory extra references.

Original reference - Compound heterozygote with c.5601_5603delAAC. Mutation located in pre-transmembrane region. Not found in 100 ethnically matched normal controls.

Disease - #602083. Autosomal recessive. Characterised by sensorineural hearing loss and retinitis pigmentosa.

Other data - SIFT - no prediction. MutPred predicts very low risk.
Conserved in mouse, dog, cow, chicken.

Conclusion - Likely to be pathogenic.

CM050770 Glaucoma, primary open angle *WDR36* c.1586G>A p.Arg529Gln

HGMD - DM, no question mark in phenotype, no contradictory extra references.

Original reference - Absent in at least 200 normal controls. Conserved in chimp, dog, rat and mouse.

Extra references - Functional study supports pathogenicity. Mutation of yeast UTP21 with homologous glaucoma-associated variant Arg529Gln of WDR36 underlies cell growth defects resulting from altered protein function (Footz (2009) Hum Mol Genet 18:1276.

Disease - Autosomal dominant. Adult onset leading cause of blindness.

Other data - SIFT predicts damaging. MutPred predicts medium risk.

	Conserved in mouse, dog, cow, chicken.			
Conclusion -	Likely to be pathogenic.			
CM060475	Sucrase isomaltase deficiency?	*SI*	c.5234T>G	p.Phe1745Cys
HGMD -	? in phenotype			
Original reference -	Described as potentially causative.			
	Extra references -Found in heterozygous patient with C1229Y. F1745C was misfolded and could not exit the endoplasmic reticulum (Alfalah (2009) Gastroenterology 136: 883.			
Disease -	#222900. Autosomal recesssive. Disorder leading to maldigestion of disaccharides.			
Other data -	SIFT predicts damaging. No MutPred prediction. Conserved in dog, cow, chicken.			
Conclusion -	Likely to be pathogenic. Remove ? From phenotype			
CM064935	Ichthyosis, harlequin	*ABCA12*	c.3889C>T	p.Arg1297Term
HGMD -	DM, no question mark in phenotype, no contradictory extra references.			
Original reference -	Compound heterozygous with c.2025delG			
Disease -	#242500. Autosomal recessive. Severe form of ichthyosis often resulting in death in first few weeks.			
Other data -	SIFT not applicable. Mutpred not applicable.			
Conclusion -	Likely to be pathogenic			
CM065101	Carnitine palmitoyltransferase 2 deficiency?	*CPT2*	c.1055T>G	p.Phe352Cys
HGMD -	DM. ? in phenotype			
Original reference - with	Described as potentially disease causing. Sole sequence variant in three cases			
	myopathic symptoms and clearly reduced residual skeletal muscle CPT II activity.			
Extra references -	Described as polymorphism in Japanese (found in 21 of 100 normal alleles. Not found in Caucasians. Did not alter CPT II activity in transfected cells (Wataya (1998) Hum Mutat 11: 377. Found as compound heterozygote with p.Val368Ile in patient with influenza-associated encephalopathy. Showed reduced			

activities and thermal instability compared with wildtype (Yao (2008) Hum Mutat 29: 718).

Disease - Autosomal recessive. Carnitine palmitoyltransferase II is one of the most common defects of oxidative lipid metabolism in humans. It has three distinct clinical phenotypes: a common adult onset myopathy (MIM255110) characterised by episodes of muscle pain, cramps, elevated serum creatine kinase levels and myoglobinuria triggered by prolonged exercise or fasting; a severe late infantile form with hypoketotic hypoglycemia and multiple organ system involvement including liver failure, cardiomyopathy, and peripheral myopathy; and a rare lethal neonatal form with dysmorphic features, renal cysts, and additional symptoms of the late-infantile form .

Other data - SIFT predicts damaging. No MutPred prediction.
Conserved in mouse, dog, cow. Tyr in chicken.

Conclusion - Disease status unclear

CM065405 Kallmann syndrome? *PROKR2* c.802C>T p.Arg268Cys

HGMD - DM. ? in phenotype.

Original reference - Heterozygote. Sporadic case.

Extra references - Decreased signalling activities and impaired G-protein coupling of receptor. Results argue against a dominant negative effect of the mutation in vivo, and therefore supports the hypothesis of a digenic or oligogenic inheritance of Kallmann syndrome in patients heterozygous for PROKR2 mutations (Monnier (2009) Hum Mol Genet 18: 75)

Disease - #244200. Autosomal recessive. Kallmann syndrome combines hypogonadotropic hypogonadism and anosmia or hyposmia .

Other data - SIFT predicts damaging. No MutPred prediction.
Conserved in mouse, dog, cow, chicken.

Conclusion - Disease status unclear

CM066043 Impaired efavirenz metabolism *CYP2B6* c.983T>C p.Ile328Thr

HGMD - Retagged as FP on 05/08/2010..

Conclusion - Functional polymorphism.

CM066133 Colorectal cancer, non-polyposis *MLH1* c.1742C>T p.Pro581Leu

HGMD - DM, no question mark in phenotype, no contradictory extra references.

Original reference - No additional/functional evidence

Extra references - The carboxyl-terminal replacement P581L in hMLH1 resulted in complete loss of activity in both yeast two-hybrid and coimmunoprecipitation tests and considered as pathogenic (Fan (2007) Clin Cancer Res 13: 7515

Disease - #609310. Autosomal dominant.

Other data - SIFT predicts damaging. MutPred predicts high risk.
Conserved in mouse, dog, cow, chicken.

Conclusion - Likely to be pathogenic

CM067435 Spermatogenic failure *NLRP14* c.322A>T p.Lys108Term

HGMD - DM, no question mark in phenotype, no contradictory extra references.

Original reference - Found in heterozygous state in one patient. Not present in 158 normospermic controls. Mutation leads to one inactive allele and likely to a reduction in expression. This truncated protein lacks the functional NACHT and LRR domains.
NALP14

Disease - Subfertility, that is the inability to conceive after 1 year of unprotected intercourse, affects 10–15% of all couples. In ~50% of cases, this subfertility is due to spermatogenic failure

Other data - SIFT not applicable. MutPred not applicable.

Conclusion - Likely to be pathogenic

CM071966 Potential protein deficiency *MOK* c.280C>T p.Arg94Term

HGMD - Tagged as FTV

Conclusion - Truncating variant

CM074454 Muscle-eye-brain syndrome/Fukuyama CMD *POMT2* c.1238G>C p.Arg413Pro

HGMD - DM, no question mark in phenotype, no contradictory extra references.

Original reference - Compound heterozygous with Tyr666Cys

Disease - #613150. Autosomal recessive. Congenital muscular dystrophy-dystroglycanopathy with brain abnormality less severe than that seen with WWS. MRI findings include pachygyria with preferential fronto-parietal involvement, polymicrogyria, cerebellar hypoplasia, cerebellar dysplasia and frequent flattening of the pons and brainstem. Eye abnormalities are often seen and include congenital glaucoma, progressive myopia, retinal atrophy and

juvenile cataracts. Individuals may, rarely, acquire the ability to walk although this is delayed.

Other data - SIFT predicts damaging. MutPred predicts very high risk.
Conserved in mouse, dog, cow, chicken.

Conclusion - Likely to be pathogenic

CM074661 Stargardt disease *ABCA4* c.1928T>G p.Val643Gly

HGMD - DM, no question mark in phenotype, no contradictory extra references.

Original reference - Found in one allele.

Disease - #248200. Autosomal recessive. Stargardt disease is one of the most frequent causes of macular degeneration in childhood. It has onset between 7 and 12 years, a rapidly progressive course, and a poor final visual outcome..

Other data - SIFTpredicts damaging. MutPred predicts high risk.
Conserved in mouse, dog, cow, chicken.

Conclusion - Likely to be pathogenic

CM074671 Joubert syndrome ? *AHI1* c.2488C>T p.Arg830Trp

HGMD - DM, ? in phenotype.

Original reference - R830W was enriched in patients who harbored NPHP1 mutations with neurologic symptoms (5/26) compared with those without neurologic symptoms (3/152) or with healthy control subjects (4/276 alleles). The R830W variant affects an evolutionarily conserved amino acid located in the WD40 repeat domain, a domain that is known to allow specific protein complexes to assemble.

Extra references - Characterised as likely polymorphism (Bell (2011) Sci Transl Med 3: 65ra4).

Disease - #608629. Autosomal recessive. Joubert syndrome is described in patients with cerebellar ataxia, mental retardation, hypotonia, and neonatal respiratory dysregulation.

Other data - SIFT predicts damaging. No MutPred prediction.
Conserved in mouse, dog, cow, chicken.

Conclusion - Disease status unclear

CM083699 Renal cell carcinoma *FLCN* c.715C>T p.Arg239Cys

HGMD - DM, no question mark in phenotype, no contradictory extra references.

Original reference - The p.Arg239Cys missense substitution was identified in a female patient who developed multicentric cRCC age 55 years after a contralateral cRCC 9 years previously. This novel missense substitution occurred at a highly conserved residue and was not detected in control samples.

Extra references - Disrupts protein stability (Nahorski (2011) Hum Mutat 32: 921

Disease - Autosomal dominant. Renal cell carcinoma accounts for ~2% of all cancers in the Western World. RCC is histologically heterogeneous, with ~80% classified as clear cell (conventional; cRCC) and among the nonclear cell types, papillary (chromophilic) and hromophobic tumors are the most frequent (1). Only 2% to 3% of all cases of RCC are familial.

Other data - SIFT predicts damaging. MutPred predicts high risk.
Conserved in mouse, dog, cow, chicken.

Conclusion - Likely to be pathogenic

CM085330 Achromatopsia *CNGA3* c.1405G>A p.Ala469Thr

HGMD - DM, no question mark in phenotype, no contradictory extra references.

Original reference - Found in patients from a family heterozygous with known pathogenic mutation R277C. Not found in 100 healthy controls. Supporting functional evidence of pathogenicity.

Disease - #216500. Autosomal recessive. Total colorblindness, also referred to as rod monochromacy or complete achromatopsia, is a rare congenital autosomal recessive disorder characterised by photophobia, reduced visual acuity, nystagmus, and the complete inability to discriminate between colors.

Other data - SIFT predicts damaging. MutPred predicts low risk.
Conserved in mouse, dog, cow, chicken.

Conclusion - Likely to be pathogenic.

CM830003 Antitrypsin alpha 1 deficiency SERPINA1 1096G>A GAG-AAG Glu-Lys 342

HGMD - DM, no question mark, no “association with”.

Original ref - Disease causing. Antitrypsin alpha 1 deficiency predisposes to degenerative lung disease. 'Z' allele.

Extra refs - Suggest pathogenicity.

Alignment - Conserved in sheep, dog, rat, cow, pig, frog and zebrafish.

Disease - #613490 Antitrypsin alpha 1 deficiency - most common manifestation is emphysema in 3rd/4th decade. Also liver disease which is less common.

Autosomal recessive.

Other data - SIFT - Damaging. MutPred - Very high risk.

Conclusion - Probably pathogenic.

CM860003 Apolipoprotein E deficiency APOE 526C>T CGC-TGC Arg-Cys 176

HGMD - Retagged as DFP in 2010. rs7412.

Conclusion - Functional polymorphism.

CM870016 Phenylketonuria PAH 1222C>T CGG-TGG Arg-Trp 408

HGMD - DM, no question mark, no "association with".

Original ref - Disease causing with functional evidence. PKU is autosomal recessive.

Extra refs - Functional evidence to suggest pathogenicity.

Alignment - Conserved in dog, rat, cow, zebrafish and frog.

Disease - #261690. Autosomal recessive.

SIFT - Damaging.

MutPred - Very high risk.

Conclusion - Probably pathogenic.

CM880008 Apolipoprotein E deficiency APOE 487C>T CGT-TGT Arg-Cys 163

HGMD - DM, no question mark, no "association with".

Original ref - No functional evidence to prove pathogenicity.

Extra refs - Suggest probable pathogenicity.

Alignment - Conserved in dog, mouse, rat, and pig.

Disease - +107741 Apolipoprotein E deficiency.

Other data - SIFT - Damaging. MutPred - Very high risk.

Conclusion - Probably pathogenic.

CM890060 Heparin cofactor 2 deficiency SERPIND1 623G>A CGC-CAC Arg-His 208

HGMD - DM, no question mark, no “association with”.

Original ref - Found heterozygously in 'healthy' patient. Functional studies show that mutation causes heparin cofactor 2 deficiency.

Disease - #612356. Heparin cofactor 2 deficiency is caused by heterozygous SERPIND1 mutation. HCF2 deficiency is a risk factor for thrombophilia. Autosomal dominant.

Other data - Conserved in panda, cow, zebrafish, mouse, rabbit and rat. SIFT - Damaging. MutPred - High risk.

Conclusion - Probably pathogenic.

CM890097 Antitrypsin alpha 1 deficiency SERPINA1 863A>T GAA-GTA Glu-Val 288

HGMD - DM, no question mark, no “association with”.

Original ref - Suggests pathogenicity with some functional evidence. 'S' allele.

Disease - #613490 Antitrypsin alpha 1 deficiency - most common manifestation is emphysema in 3rd/4th decade. Also liver disease which is less common. Autosomal recessive.

Other data - Conserved in dog, cow, rat and pig. SIFT - Damaging. MutPred - High risk.

Conclusion - Probably pathogenic.

CM900007 Adenosine deaminase deficiency ADA 643G>A GCC-ACC Ala-Thr 215

HGMD - DM, no question mark, no “association with”.

Original ref - Suggests pathogenicity.
Extra refs - Functional studies suggest pathogenicity.

Disease - #102700. Severe combined immunodeficiency due to adenosine deaminase deficiency. Autosomal recessive.

Other data - Conserved in dog, cow, rat, mouse and frog. SIFT - Damaging. MutPred - Very high risk.

Conclusion - Probably pathogenic.

CM910070 Cystic Fibrosis CFTR 1000C>T CGG-TGG Arg-Trp 334

HGMD - DM, no question mark, no “association with”.

Original ref - Not found in controls. No functional evidence to suggest pathogenicity.

Extra refs -	Described as severe recessive disease-causing mutation in 1 ref. Mild mutation in another ref.
Disease -	#219700 Cystic fibrosis. Autosomal recessive.
Other data -	Conserved in dog, cow, rat, mouse, sheep and pig. SIFT - Damaging. MutPred - Very high risk.
Conclusion -	Probably pathogenic.

CM910395	Von Willebrand, Normandy variant VWF 2561G>A CGG-CAG Arg-Gln 854
HGMD -	DM, no question mark, no “association with”.
Original ref -	Suggests pathogenicity. Found heterozygously in parents and homozygous / compound heterozygous in patients.
Extra refs -	Suggest pathogenicity.
Disease -	#613554 Von Willebrand disease, type 2N. Autosomal recessive.
Other data -	Conserved in dog, cow, rat, mouse and cat. SIFT - Damaging. MutPred - Medium risk.
Conclusion -	Probably pathogenic.

CM920026	Adenosine monophosphate deaminase deficiency AMPD1 242C>T CCG-CTG Pro-Leu 81
HGMD -	DM, no question mark, no “association with”.
Original ref -	Found in cis with Q45X in all patients. No effect on catalytic activity in functional study.
Extra refs -	Described as likely polymorphism.
Alignment -	Conserved in dog, cow, rat, mouse and pig.
Disease -	+102770 Myoadenylate deaminase deficiency. Autosomal recessive.
Other data -	N/A
Conclusion -	Possibly not pathogenic. Change tag to DM? or remove.

CM930629	McArdle disease PYGM 148C>T CGA-TGA Arg-Term 50
HGMD -	DM, no question mark, no “association with”.

Original ref -	Nonsense mutation – probably pathogenic
Extra refs -	None.
Disease -	#232600 **McArdle disease** - relatively benign disorder, except for possible renal failure as a complication of myoglobinuria. Symptoms include exercise intolerance & muscle cramps. Autosomal recessive.
Other data -	N/A.
Conclusion -	Probably pathogenic.

CM930731	Von Willebrand disease 2a VWF 4517C>T TCG-TTG Ser-Leu 1506
HGMD -	DM, no question mark, no “association with”.
Original ref -	Found heterozygously with a pseudogene polymorphism. No functional data.
Extra refs -	Functional study – increased proteolysis.
Disease -	#613554. Usually autosomal dominant though can be recessive.
Other data - risk.	Conserved in dog, cow, rat, mouse, cat and pig. SIFT - N/A. MutPred - Very high
Conclusion -	Probably pathogenic.

CM931158	Cystic Fibrosis CFTR 4056G>C CAG-CAC Gln-His 1352
HGMD -	DM, no question mark, no “association with”.
Original ref -	Not found in controls. Study looking at chronic pancreatitis (not cystic fibrosis). No functional data.
Extra refs -	Additional phenotypes but no functional data. 1 paper describes mutation as non-CF causing and found the mutation in chronic pancreatitis patients and controls.
Disease -	#219700 Cystic fibrosis. Autosomal recessive. #167800 Chronic pancreatitis.
Other data -	Conserved in dog, cow, rat, mouse, sheep and pig. SIFT - Damaging. MutPred - Very high risk.
Conclusion -	Possibly pathogenic. Change phenotype to chronic pancreatitis and change tag to DM?

CM940961 Histatin 2 deficiency HTN3 141 T>A TAT-TAA Tyr-Term 47

HGMD - FP. rs17147990.

Conclusion - Functional polymorphism.

CM950247 Cystic Fibrosis CFTR 2002C>T CGT-TGT Arg-Cys 668

HGMD - DM, phenotype with question mark.

Original ref - Study looking at congenital bilateral absence of the vas deferens (not cystic fibrosis). No functional evidence. Found compound heterozygously in 4 patients.

Extra refs - No functional data. Described as severe recessive disease-causing mutation in 1 ref.

Disease - #219700 Cystic fibrosis. Autosomal recessive.
#277180 CBAVD. X-linked recessive or autosomal dominant male-limited inheritance.

Other data - Conserved in dog, cow, cat, mouse, sheep and pig.

Conclusion - Possibly pathogenic. Change phenotype to congenital bilateral absence of the vas deferens and keep the question mark.

CM950959 Haemolytic anaemia PKLR 1468C>T CGG-TGG Arg-Trp 490

HGMD - DM, no question mark, no “association with”.

Original ref - Suggests pathogenicity.

Disease - #266200 Pyruvate kinase deficiency (causing haemolytic anaemia). Autosomal dominant?

Other data - Conserved in dog, cow, mouse, rat and zebrafish. SIFT - Damaging. MutPred-Medium risk.

Conclusion - Possibly pathogenic. Possibly change to DM?

CM951041 Protein S deficiency PROS1 119G>T CGT-CTT Arg-Leu 40

HGMD - DM, no question mark, no “association with”.

Original ref - Described as probably deleterious. Found in 2 patients.

Disease - #612336 Thrombophilia due to protein S deficiency.

Other data - Conserved in dog, cow, mouse, rat and pig. SIFT - Damaging. MUTPRED - Medium risk.

Conclusion - Possibly pathogenic. Possibly change to DM?

CM951045 Protein S deficiency *PROS1* c.233C>T ACG-ATG Thr-Met 78

HGMD - DM, no question mark, no "association with", no contradictory extra references.

Original reference - Table 2. Probably deleterious mutation. Patient exhibited first DVT age 25. Family study not performed.

Disease - #612336 and #614514. Dominant and recessive types. Main symptom is recurrent thrombophilia. Adult onset. Probable environmental components.

Other data - SIFT - Damaging MUTPRED - HIGH RISK.

Conclusion - Likely deleterious.

CM970003 Stargardt disease *ABCA4* c.2588G>C GGA-GCA Gly-Ala 863

HGMD - DM, no question mark, no "association with". Extra reference is functional study in support of pathogenicity - the G863A mutation was found to have significant attenuation of the rates of nucleotide hydrolysis and binding affinities in E. coli expression system.

Original reference - Found in 3 families. Absent in 100 controls. No other information.

Disease - #248200. Stargardt disease. Autosomal recessive. Onset within first 2 decades.

Other data - SIFT – Damaging. MUTPRED - HIGH RISK.

Conclusion - Likely deleterious.

CM970206 Complement C9 deficiency *C9* c.162C>A TGC-TGA Cys-Term 54

HGMD - DM, no question mark, no "association with".

Original reference- Segregates with phenotype in family studied.

Disease - #613825. Deficiency state. No overt disease susceptibility.

Other data - N/A.

Conclusion - Very likely deleterious

CM970401 Pulmonary alveolar proteinosis *CSF2RB* c.1807C>A CCT-ACT Pro-Thr 603

HGMD - DM, no question mark, no “association with”, no contradictory extra references.

Original reference -Sequences of 20 healthy random controls showed no polymorphism at this codon.

Disease - #614370. Pulmonary alveolar proteinosis is a rare autosomal recessive cause of respiratory failure. It is a heterogenous disorder of acquired or genetic etiology and is characterised by occurrence of severe respiratory distress and rapid progression to death.

Other data - SIFT – Damaging. MUTPRED - MEDIUM RISK.

Conclusion - Likely deleterious.

CM983332 Myeloperoxidase deficiency *MPO* c.518A>G TAC-TGC Tyr-Cys 173

HGMD - DM, no question mark, no “association with”, no contradictory extra references.

Original reference - Found during routine screening of otherwise healthy 27 year old female. Functional study supports pathogenicity.

Disease - #254600. Autosomal recessive. Majority of individuals with phenotype are healthy.

Other data - SIFT – Damaging. MUTPRED - VERY HIGH RISK.

Conclusion - Likely deleterious.

CM985505 Retinitis pigmentosa *SAG* c.250C>T CGC-TGC Arg-Cys 84

HGMD - DM, no question mark, no “association with”. Extra reference is functional study in support of pathogenicity (as Arg80Cys).

Original reference - Found in single patient. No co-segregation. Author concludes this is unlikely to be pathogenic.

Disease - #613758. Autosomal dominant in patient, but OMIM has as recessive type for this gene.

Other data - SIFT – Damaging. MUTPRED - HIGH RISK.

Conclusion - Possibly deleterious. Add question mark.

CM985506 Retinitis pigmentosa *SAG* c.1091C>T CCT-CTT Pro-Leu 364

HGMD - DM, no question mark, no “association with”. Extra reference is functional study, but it is not conclusive.

Original reference - Found in heterozygously in single patient diagnosed with recessive retinitis pigmentosa. Author concludes this is unlikely to be pathogenic.

Disease - #613758. Autosomal recessive but only found heterozygously in patient.

Other data - SIFT – Damaging. MUTPRED - HIGH RISK.

Conclusion - Possibly non-pathogenic. Add question mark.

CM990302 Muscular dystrophy, limb girdle? *CAPN3* c.551C>T ACG-ATG Thr-Met 184

HGMD - DM, with question mark, no "association with". Extra reference reports that variant is found in-cis with Ala236Thr.

Original reference - Found in one homozygous patient.

Disease - #253600. Autosomal recessive. Onset ranges from early childhood to adulthood (usually before age 15). Wheelchair use at 20-30 years. Gradual progression

Other data - SIFT – Damaging. MUTPRED - HIGH RISK.

Conclusion - Possibly deleterious (when found in-cis).

CM990718 Haemochromatosis, association with *HFE* c.193A>T AGT-TGT Ser-Cys 65

HGMD - DP, with "association with". Not a DM.

Original reference - S65C substitution was significantly enriched in probands. Mild form of phenotype.

Disease - #235200. Autosomal recessive. Incomplete penetrance?

Other data - SIFT – Damaging. MUTPRED - VERY HIGH RISK.

Conclusion - Likely reduced penetrance pathogenic allele.

CM990837 Mediterranean fever, familial *MEFV* c.1105C>T CCC-TCC Pro-Ser 369

HGMD - DM, no question mark, no "association with". Extra reference indicates mutation has also been found in fibromyalgia syndrome. Additional functional study is negative, but reports that P369S/R408Q substitutions are associated with a highly variable phenotype, and found in both symptomatic and asymptomatic family members.

Original reference - Found in ~8% of patients and 20% of carriers (Askenazi Jewish population). Also found at low frequency in control population.

Disease - #134610 and #249100. Autosomal dominant and recessive forms. Homozygous healthy individuals sometimes found indicating incomplete penetrance.

Other data - SIFT – Damaging. MUTPRED - LOW RISK.

Conclusion - Likely deleterious.

CM991062 Colorectal cancer, non-polyposis *PMS1* c.501G>A GGA-AGA Gly-Arg 501

HGMD - DM, no question mark, no “association with”, no additional references.

Original reference - Found in cis with Met394Thr in same individual. Not found in 50 healthy controls. Pathogenicity unconfirmed.

Disease - Autosomal dominant. Variable onset.

Other data - SIFT – Damaging. MUTPRED - LOW RISK.

Conclusion - Possibly deleterious (when found in-cis).

CD993068 Haemophagocytic lymphohistiocytosis, familial *PRF1* c.50delT

HGMD - DM, no question mark, no “association with”, no additional references.

Original reference - Found as homozygote in single patient. No other information.

Disease - #603553. Autosomal recessive. Infantile onset. Rapidly fatal if left untreated.

Other data - N/A

Conclusion - Very likely deleterious.

CS020995 Berardinelli-Seip lipodystrophy*AGPAT2* c.589-2A>G

HGMD - DM, no question mark, no “association with”.

Original reference - Identified as common ancestral mutation Reported splicing effect is Gln196fs*228.

Disease - #608594. Autosomal recessive. Early onset.

Other data - N/A

Conclusion - Very likely deleterious.

CS030995 Deafness, non-syndromic *SLC26A5* c.-53-2A>G

HGMD - DM, no question mark, no “association with”, no extra references.

Original reference - Found in 2 deaf probands. Predicted to affect splicing. In addition 7 out of 220 white deaf probands were found to be heterozygous for the same mutation with variable deafness symptoms. Found in 1/150 white controls.

Disease - #613865. Autosomal recessive. Onset from birth. Possible dominant traits with variable onset.

Other data - N/A

Conclusion - Very likely deleterious.

CS910444 Tay-Sachs disease *HEXA* c.1073+1G>A

HGMD - DM, no question mark, no "association with". Additional reference is Bell *et al* paper confirming genotype.

Original reference - Found as heterozygote in one patient. Other allele is unidentified.

Disease - #272800. Autosomal recessive. Infantile onset. Usually fatal by age 5 years.

Other data - N/A

Conclusion - Very likely deleterious.

CS961669 Proline-rich protein PRB4 deficiency *PRB4* c.762+1G>C

HGMD - DM, no question mark, no "association with", no extra references.

Original reference - Polymorphic null allele.

Disease - No overt disease.

Other data - N/A

Conclusion - Very likely deleterious. Re-tag to reflect polymorphic nature.

CS982392 Obesity, severe, with diabetes *UCP3* c.824+1G>A

HGMD - DM, no question mark, no "association with", no extra references.

Original reference - Found as compound heterozygote in morbidly obese and diabetic patient. Noted that any putative protein resulting from the mutation would be identical to that encoded by the short transcript of UCP3. 10% minor allele frequency in African populations studied.

Disease - #601665. Autosomal recessive. Variable onset and penetrance.

Other data - N/A

Conclusion -	Very likely deleterious.		
CS992248	Lp(a) deficiency	*LPA*	c.4289+1G>A
HGMD -	DM, no question mark, no “association with”, no extra references.		
Original reference -	Reported as common null mutation with approximately 6% frequency.		
Disease -	+152200. Autosomal dominant. Risk factor in heart disease.		
Other data -	N/A		
Conclusion -	Very likely deleterious.		
CS951368	Complement C6 deficiency, subtotal	*C6*	c.2381+2T>C
HGMD -	DM, no question mark, no “association with”, no extra references.		
Original reference -	Found in 3 individuals.		
Disease - infection.	#612446. Deficiency state. No overt disease. Possible increased susceptibility to		
Other data -	N/A		
Conclusion -	Very likely deleterious.		