Cross-comparison of the genome sequences from human, chimpanzee, Neanderthal and a Denisovan hominin identifies novel potentially compensated mutations

Guojie Zhang,[1] Zhang Pei,[1] Edward V. Ball,[2] Matthew Mort,[2] Hildegard Kehrer-Sawatzki[3] and David N. Cooper[2]*

[1]Bioinformatics Department, Beijing Genomics Institute at Shenzhen, Shenzhen 518083, China
[2]Institute of Medical Genetics, School of Medicine, Cardiff University, Heath Park, Cardiff CF14 4XN, UK
[3]Institute of Human Genetics, University of Ulm, Albert-Einstein-Allee 11, 89081 Ulm, Germany
**Correspondence to:* Tel: +86 0755 25273794; Fax: +86 0755 25273114; E-mail: zhanggj@genomics.org.cn

Date received (in revised form): 22nd March 2011

Abstract

The recent publication of the draft genome sequences of the Neanderthal and a ~50,000-year-old archaic hominin from Denisova Cave in southern Siberia has ushered in a new age in molecular archaeology. We previously cross-compared the human, chimpanzee and Neanderthal genome sequences with respect to a set of disease-causing/disease-associated missense and regulatory mutations (Human Gene Mutation Database) and succeeded in identifying genetic variants which, although apparently pathogenic in humans, may represent a 'compensated' wild-type state in at least one of the other two species. Here, in an attempt to identify further 'potentially compensated mutations' (PCMs) of interest, we have compared our dataset of disease-causing/disease-associated mutations with their corresponding nucleotide positions in the Denisovan hominin, Neanderthal and chimpanzee genomes. Of the 15 human putatively disease-causing mutations that were found to be compensated in chimpanzee, Denisovan or Neanderthal, only a solitary *F5* variant (Val1736Met) was specific to the Denisovan. In humans, this missense mutation is associated with activated protein C resistance and an increased risk of thromboembolism and recurrent miscarriage. It is unclear at this juncture whether this variant was indeed a PCM in the Denisovan or whether it could instead have been associated with disease in this ancient hominin.

Keywords: *Human, chimpanzee, Neanderthal, Denisovan hominin, genome sequence, potentially compensated mutations, disease*

Introduction

The recent publication of the draft sequence of the Neanderthal genome[1] ushered in a new age in molecular archaeology.[2,3] This achievement was followed closely by the publication of the draft genome sequence (1.9-fold coverage) of a ~50,000-year old archaic hominin from Denisova Cave in southern Siberia.[4] This hominin (a 'Denisovan') is thought to have been a member of a sister group of hominins to the Neanderthals with whom they lived sympatrically during the Upper Pleistocene.[4–7] Denisovans appear to be more closely related to Neanderthals than humans, having diverged from Neanderthals about 640,000 years ago and from extant Africans about 804,000 years ago.[4]

Access to DNA sequence data from ancient hominins not only promises to revolutionise our knowledge of hominin relationships, but is also potentially informative in the context of exploring the molecular basis of human genetic disease.[8,9] We have previously cross-compared the human, chimpanzee and Neanderthal genome sequences with a set of disease-causing/disease-associated missense and regulatory mutations in order to identify genetic variants which, although apparently pathogenic in humans, may represent a 'compensated' wild-type state in at least one of the other two species ('potentially compensated mutations' [PCMs]).[10] PCMs correspond to variants that may have been deleterious for a certain period of evolutionary time but which persisted long enough in a given population or species to have become positively selected upon the introduction of a 'compensatory' nucleotide change.[8,11–14] Such compensatory changes are thought to be localised in the same gene as the PCM.[15] Not only do PCMs represent excellent candidates for recent population-specific selection (with different alleles having exhibited differential functional importance in different environments), but they may also furnish us with new insights into the genetic basis of susceptibility to common diseases.[8,14] Here, in an attempt to identify further PCMs of interest, we have compared a dataset of human mutations of putative pathological significance with their corresponding nucleotide positions in the Neanderthal, Denisovan and chimpanzee genomes.

Methods

Human Gene Mutation Database (HGMD) dataset

A total of 46,060 disease-causing (DMs) or disease-associated mutations had been obtained from the HGMD[16] (http://www.hgmd.org) as of 13th May 2010. These data comprised 44,348 missense mutations from within the coding regions of 2,628 genes, and 1,712 single base-pair substitutions from within the regulatory regions (5′ and 3′ untranslated/flanking regions) of 807 genes. Some 42,595 of the mutations were disease-causing (41,960 missense and 635 regulatory), whereas 3,465 represented disease-associated or functional polymorphisms (2,388 missense and 1,077 regulatory) (Table 1). The latter were further ascribed to three distinct subcategories: (1) DPs, comprising variants reported to be in statistically significant ($p < 0.05$) association with a particular human disease state but lacking experimental evidence of functionality — for example, from expression studies; (2) disease-associated polymorphisms with experimental evidence of functionality (DFPs) such as, for example, altered in *vitro* gene expression or protein function; (3) FPs that have been shown in *vitro* or in *vivo* to affect the structure, function or expression of the gene or gene product but for which no statistically significant disease association has yet been reported (see http://www.hgmd.cf.ac.uk/docs/poly.html for further information).

Identification of PCMs

A total of 8,280,851 nucleotide positions at which the Denisovan genome differs from either the human (NCBI36/hg18) or chimpanzee genome were downloaded from the website of the Max Planck Institute for Evolutionary Anthropology (http://bioinf.eva.mpg.de/download/DenisovaGenome/Denisova_Neandertal_catalog.tgz).[1,4] The human and the Denisovan hominin were found to exhibit the same nucleotide at 7,283,268 positions (87.95 per cent), so that the human–chimpanzee mismatches must have arisen before the divergence of modern

Table 1. Missense and regulatory mutations from the HGMD used in this study, categorised by mutation type and putative role in disease aetiology

Mutation/ polymorphism type	Type and putative role in disease aetiology				
	DM	DP	DFP	FP	Total
Coding sequence	41,960	942	295	1,151	44,348
Regulatory	635	340	391	346	1,712
Total	42,595	1,282	686	1,497	46,060

DM, disease-causing mutation; DP, disease-associated polymorphism lacking functional evidence; DFP, disease-associated polymorphism with functional evidence; FP, polymorphism with functional evidence but lacking a reported disease association as yet.

humans and Denisovans (termed a 'derived' or 'D' state in the Denisovan). A total of 941,947 positions (11.38 per cent) displayed the same nucleotide in both Denisovan and chimpanzee, suggesting that the respective substitutions were human specific ('ancestral' or 'A' state in the Denisovan). The remaining 55,636 positions, which display different nucleotides in modern humans, Denisovans and chimpanzees, were termed 'undefined' ('N' state). Of the 8,280,851 Denisovan nucleotide positions investigated here, there were 5,205,736 positions at which the Neanderthal was found to differ from at least one of modern human, chimpanzee and Denisovan. From these 5,205,736 sites, we identified 197 sites for which the apparent wild-type nucleotide in

Table 2. HGMD-derived mutations identified as PCMs in the Denisovan, Neanderthal and/or chimpanzee genomes

	Mutation type and basis of disease aetiology					
Mutation/ regulatory type	**PCM state**	**DM**	**DP**	**FP**	**DFP**	**Total**
Coding sequence	Human	5/5	38/43	11/11	17/18	71/77
	Neanderthal	0/0	1/1	0/0	0/0	1/1
	Denisovan	1/1	0/0	0/0	0/0	1/1
	Ancient	0/0	1/1	2/4	0/0	3/5
	Chimpanzee	4/4	7/8	2/4	0/0	13/16
	Denisovan and chimpanzee	3/3	4/5	0/0	0/0	7/8
	Neanderthal and chimpanzee	2/2	4/6	1/1	1/1	8/10
	Others	1/1	0/1	0/0	0/0	1/2
	Total	16/16	55/65	16/20	18/19	105/120
Regulatory	Human	0	23	10	13	46
	Neanderthal	0	0	2	1	3
	Denisovan	0	0	0	3	3
	Ancient	0	2	0	1	3
	Chimpanzee	0	5	5	4	14
	Denisovan and chimpanzee	0	4	1	1	6
	Neanderthal and chimpanzee	0	1	1	0	2
	Others	0	0	0	1	1
	Total	0	35	19	24	78

'Human': The Denisovan nucleotide, Neanderthal nucleotide and chimpanzee nucleotide were identical to a human DM/disease-associated mutation; 'Neanderthal': The Neanderthal nucleotide was identical to the human DM/disease-associated mutation, whereas both the chimpanzee nucleotide and the Denisovan nucleotide were identical to the human wild-type nucleotide; 'Denisovan': The Denisovan nucleotide was identical to the human DM/disease-associated mutation, whereas both the chimpanzee nucleotide and the Neanderthal nucleotide were identical to the human wild-type nucleotide; 'Ancient': Both the Denisovan nucleotide and the Neanderthal nucleotide were identical to the human DM/disease-associated P mutation, whereas the chimpanzee nucleotide was identical to the human wild-type nucleotide. 'Chimpanzee': The chimpanzee nucleotide was identical to the human DM/disease-associated mutation, whereas both the Neanderthal nucleotide and the Denisovan nucleotide were identical to the modern human wild-type nucleotide. 'Denisovan and chimpanzee': Both the Denisovan nucleotide and the chimpanzee nucleotide were identical to the human DM/disease-associated mutation, whereas the Neanderthal nucleotide was identical to the human wild-type nucleotide; 'Neanderthal and chimpanzee': Both the Neanderthal nucleotide and the chimpanzee nucleotide were identical to the human DM/disease-associated mutation, whereas the Denisovan nucleotide was identical to the human wild-type nucleotide. Under coding sequence, 'a/b' means that there were a total number of 'b' mutations, of which 'a' were non-synonymous mutations (there were some synonymous mutations within the coding sequence; eg CM068190, CM077900).

PCM, potentially compensated mutations; DM, disease-causing mutation; DP, disease-associated polymorphism with functional evidence; FP, polymorphism with functional evidence but lacking a reported disease association as yet; DFP, disease-associated polymorphism with functional evidence.

Denisovan, Neanderthal or chimpanzee was logged in the HGMD as disease causing or disease associated in modern humans (Table 2). From the remaining 3,075,115 sites, we identified 117 sites for which the apparent wild-type nucleotide in the Denisovan or chimpanzee was logged in the HGMD as disease causing or disease associated in either the Denisovan or chimpanzee (Table 3).

Gene ontology (GO) enrichment analysis

A GO enrichment analysis of PCM-containing genes against a background of 2,688 human disease-causing genes was performed using the DAVID bioinformatics tool.[17] The statistical significance of a particular GO term was calculated using Fisher's exact test, which was then adjusted to allow for multiple testing by means of the Benjamini–Hochberg correction.[18]

Table 3. HGMD-derived mutations identified as PCMs in the Denisovan genome and/or chimpanzee genome

	Mutation type and basis of disease aetiology					
Mutation/ regulatory type	**PCM state**	**DM**	**DP**	**FP**	**DFP**	**Total**
Coding sequence	Ancestral	5/5	24/29	9/9	5/7	43/50
	Derived	4/4	7/7	4/5	4/4	19/20
	Denisovan	0/0	4/6	2/2	0/0	6/8
	Others	0/0	1/1	0/0	0/0	1/1
	Total	9/9	36/43	15/16	9/11	69/79
Regulatory	Ancestral	2	6	9	12	29
	Derived	1	2	1	2	6
	Denisovan	0	1	0	2	3
	Total	3	9	10	16	38

Ancestral: Both the Denisovan nucleotide and the chimpanzee nucleotide were identical to the human DM/disease-associated mutation; Derived: The chimpanzee nucleotide was identical to the human DM/disease-associated mutation, whereas the Denisovan nucleotide was identical to the human wild-type nucleotide; Denisovan: The Denisovan nucleotide was identical to the human DM/disease-associated mutation, whereas the chimpanzee nucleotide was identical to the human wild-type nucleotide. Under coding sequence, 'a/b' means there were a total number of 'b' mutations, of which 'a' were non- synonymous mutations.
PCM, potentially compensated mutations; DM, disease-causing mutation; DP, disease-associated polymorphism with functional evidence; FP, polymorphism with functional evidence but lacking a reported disease association as yet; DFP, disease-associated polymorphism with functional evidence.

Calculation of Wright's fixation index (F_{ST}) values

The F_{ST} measures the proportion of genetic diversity in a subdivided population that is attributable to allele frequency differences between subpopulations. Pairwise F_{ST} values have also been used as a measure of genetic distance between populations. In this context, the allele frequencies of polymorphic ancestral PCMs in selected populations were obtained from HapMap (http://hapmap.ncbi.nlm.nih.gov/) and pairwise F_{ST} values were estimated for each polymorphism using the small sample estimate proposed by Weir and Hill.[19] The significance of individual F_{ST} values was then assessed by reference to the empirical distribution of F_{ST} among all single nucleotide polymorphisms (SNPs) in HapMap.

Results and discussion

Identification of PCMs in the Denisovan, Neanderthal and/or chimpanzee genomes

A total of 44,348 missense mutations from 2,628 genes and 1,712 putative regulatory mutations from 807 genes, which have been recorded in the HGMD as being either causative of (or associated with) a human inherited disease state, were cross-compared with the corresponding nucleotide positions in the Neanderthal, Denisovan and chimpanzee genomes.

When the 197 PCMs covered by both the Denisovan and the Neanderthal sequences were considered, these included 129 of 143 PCMs identified in the Neanderthal genome (10/12 DMs, 65/73 DPs, 25/26 FPs, 29/32 DFPs), and 123 (62 per cent) PCMs for which the Denisovan, Neanderthal and chimpanzee wild-type nucleotides were identical to the human disease-causing/disease-associated mutant allele. Of the 117 PCMs covered only by the Denisovan sequence, there were 79 (67.5 per cent) for which both the Denisovan nucleotide and the chimpanzee nucleotide were identical to a human DM/disease-associated mutation. This may be indicative of either a bottleneck effect or selection during the evolution of the modern human lineage. Of the 197 PCMs, there was one mutation

Table 4. Human DMs identified as PCMs

Category	HGMD Acc. No	Chr	Chrom. location	Strand	Disease	Gene	Mutation	HGVS (cDNA) nomenclature	HGVS (protein) nomenclature	Type
Covered by both the Neanderthal and the Denisovan sequence[a]	CM993347	Chr1	67633930	+	Atopy	*IL12RB2*	A > G:GAA	NM_001559.2: c.2159A > G	NP_001550.1: p.H720R	Chimpanzee
	CM042258	Chr1	94337039	–	Stargardt disease	*ABCA4*	T > G:GGT	NM_000350.2: c.667A > C	NP_000341.2: p.K223Q	Denisovan and chimpanzee
	CM070090	Chr1	167765599	–	Thrombosis?	*F5*	C > T:CTC	NM_000130.4: c.5290G > A	NP_000121.2: p.V1764M	Denisovan
	CM099258	Chr15	40468491	+	Muscular dystrophy?	*CAPN3*	G > A:AAA	NM_000070.2: c.706G > A	NP_000061.1: p.A236T	Human
	CM085365*	Chr15	43185730	–	Hypothyroidism	*DUOX2*	T > C:CCC	NM_014080.4: c.2033A > G	NP_054799.4: p.H678R	Human
	CM984025*	Chr19	18047618	–	Mycobacterial infection	*IL12RB1*	T > C:CCT	NM_005535.1: c.641A > G	NP_005526.1: p.Q214R	Denisovan and chimpanzee
	CM044918	Chr19	41022117	–	Congenital nephrotic syndrome, Finnish type	*NPHS1*	C > G:GGG	NM_004646.1: c.2971G > C	NP_004637.1: p.V991L	Human
	CM064230	Chr19	43656115	+	Malignant hyperthermia	*RYR1*	A > G:GAA	NM_000540.2: c.4024A > G	NP_000531.2: p.S1342G	Chimpanzee
	CM961339*	Chr22	30836050	+	Glucose/galactose malabsorption	*SLC5A1*	C > G:GGC	NM_000343.1: c.1845C > G	NP_000334.1: p.H615Q	Denisovan and chimpanzee
	CM980573	Chr5	149341414	+	Achondrogenesis 1B	*SLC26A2*	A > T:TAT	NM_000112.3: c.2065A > T	NP_000103.2: p.T689S	Neanderthal and chimpanzee

Continued

Table 4. Continued

Category	HGMD Acc. No	Chr	Chrom. location	Strand	Disease	Gene	Mutation	HGVS (cDNA) nomenclature	HGVS (protein) nomenclature	Type
	CM043093	Chr6	25958824	–	Glycogen storage disease 1c?	*SLC17A3*	C > T:TCC	NM_006632.3: c.601G > A	NP_006623.2: p.G201R	Chimpanzee
	CM072814	Chr7	86894112	–	Intrahepatic cholestasis, familial progressive?	*ABCB4*	T > C:CCC	NM_000443.3: c.1954A > G	NP_000434.1: p.R652G	Human
	CM050323	Chr7	107129530	+	Pendred syndrome?	*SLC26A4*	T > G:GTG	NM_000441.1: c.1826T > G	NP_000432.1: p.V609G	Neanderthal and chimpanzee
	CM983990	Chr8	22032655	–	Alopecia universalis?	*HR*	T > C:CCC	NM_005144.3: c.3064A > G	NP_005135.2: p.T1022A	Human
	CM099178*	Chr8	118899878	–	Multiple osteochondromas	*EXT1*	C > T:TCC	NM_000127.2: c.1609G > A	NP_000118.2: p.V537I	Chimpanzee
	CM085353*	ChrX	149390017	+	Hypospadias	*MAMLD1*	T > C:CYC	NM_005491.2: c.1514T > C	NP_005482.2: p.V505A	Others
Covered only by the Denisovan sequence[b]	CM043273	Chr1	195670491	+	Retinitis pigmentosa	*CRB1*	G > A:AG	NM_201253.1: c.2875G > A	NP_957705.1: p.G959S	Chimpanzee
	CM067436	Chr11	7020956	+	Spermatogenic failure	*NLRP14*	G > A:AG	NM_176822.3: c.1123G > A	NP_789792.1: p.A375T	Chimpanzee
	CM043536	Chr11	47326617	–	Cardiomyopathy, hypertrophic?	*MYBPC3*	T > C:CT	NM_000256.3: c.706A > G	NP_000247.2: p.S236G	Chimpanzee
	CM082943	Chr11	118720796	–	Primary angle-closure glaucoma?	*MFRP*	C > T:TT	NM_031433.1: c.770G > A	NP_113621.1: p.R257H	Ancestral
	CM091988	Chr12	32913201	–	Arrhythmogenic right ventricular cardiomyopathy	*PKP2*	A > G:GG	NM_004572.3: c.1097T > C	NP_004563.2: p.L366P	Ancestral
	CM044579	Chr13	51413355	–	Wilson disease?	*ATP7B*	A > G:GG	NM_000053.2: c.3419T > C	NP_000044.2: p.V1140A	Ancestral

Continued

Table 4. Continued

Category	HGMD Acc. No	Chr	Chrom. location	Strand	Disease	Gene	Mutation	HGVS (cDNA) nomenclature	HGVS (protein) nomenclature	Type
	CM073339	Chr17	24310977	–	Febrile seizures?	*SEZ6*	T > C:CC	NM_178860.4: c.1636A > G	NP_849191.3: p.T546A	Ancestral
	CM101950	Chr2	98363138	+	Progressive cone dystrophy?	*CNGA3*	C > T:TC	NM_001298.2: c.284C > T	NP_001289.1: p.P95L	Chimpanzee
	CM961335	Chr22	30817700	+	Glucose/galactose malabsorption	*SLC5A1*	G > A:AA	NM_000343.1: c.1231G > A	NP_000334.1: p.A411T	Ancestral
	CR080762	Chr1	15645754	+	Pancreatitis, chronic?	*CTRC*	T > C:CC	rs75456156:T > C	NA	Ancestral
	CR080761	Chr1	15645757	+	Pancreatitis, chronic?	*CTRC*	A > G:GG	rs760937:A > G	NA	Ancestral
	CR962526	Chr8	41774321	–	Spherocytosis	*ANK1*	A > G:GA	rs77173848:A > G	NA	Chimpanzee

[a]Mutation type: modern human wild-type > modern human mutation: chimpanzee nucleotide, Denisovan nucleotide, Neanderthal nucleotide (both Neanderthal and Denisovan sequence covered). Y denotes pyrimidine.
[b]Modern human wild-type > modern human mutation: chimpanzee nucleotide, Denisovan nucleotide (only Denisovan sequence covered).
*Previously reported by Zhang *et al.*[10]

that was compensated only in the Neanderthal, one that was compensated only in the Denisovan, five that were compensated in both Neanderthal and Denisovan and 16 that were compensated only in the chimpanzee. There were also 18 mutations that differed between the Neanderthal and the Denisovan, which could imply that such mutations were identical-by-state (Tables 2 and 3).

Disease-causing PCMs

There were 16 human DMs that were found to be potentially compensated in the chimpanzee, Denisovan or Neanderthal (covered by both the Neanderthal and the Denisovan sequence) and 12 human DMs potentially compensated in the chimpanzee or Denisovan (covered only by the Denisovan sequence) (Table 4).

Of the human DMs that were potentially compensated in the chimpanzee, Denisovan or Neanderthal, only the putatively pathological *F5* variant was specific to the Denisovan. In humans, this missense mutation, Val1736Met, is associated with activated protein C resistance and an increased risk of thromboembolism and recurrent miscarriage.[20,21] It is unclear at this juncture whether this variant was indeed a PCM in the Denisovan or whether it could instead have been associated with disease in this archaic hominin.

Even though Denisovans appear to be more closely related to Neanderthals than humans, the Neanderthal and Denisovan were discrepant with respect to certain PCMs (eg the *SLC5A1* H615Q variant associated with glucose–galactose malabsorption). In this case, the Denisovan (and the chimpanzee) possessed the allele that was mutant in humans (G), whereas the Neanderthal possessed the allele (C) which was wild-type in humans. In this context, it may be pertinent to mention that *SLC5A1* is located on chromosome 22q12.3 within a region of putative gene flow from Neanderthal to Eurasian.[1]

Some of the PCMs listed in Table 4 may well have been misclassified by the original authors as disease-causing in human (especially those variants which have been allocated a '?' by the HGMD; see

Table 4) when they were actually neutral polymorphisms; however, this is much less likely in the case of the 16 human disease-causing mutations that are covered by both the Neanderthal and Denisovan sequences. These mutant alleles would have had to have been maintained in both Neanderthal and Denisovan populations for ~640,000 years, when these two hominins last shared a common ancestor, and this would have been unlikely if such variants had been neutral polymorphisms.

Statistically enriched GO terms were identified for genes containing human DMs identified as PCMs (Table 4) against a background of known disease-causing genes (from the HGMD) and are shown in Table S1. Five significantly enriched GO terms were found; all relate to the plasma membrane.

With respect to the DPs/FPs, 100 DPs, 39 FPs and 43 DFPs were covered by both the Neanderthal and Denisovan sequences (Table S2), while 52 DPs, 26 FPs and 27 DFPs were covered by the Denisovan but not the Neanderthal sequence (Table S3); these DPs/FPs may be relevant to human genetic disease.

Human variants with significantly different population frequencies at sites of PCMs

The F_{ST} was used to quantify the allele frequency differences for the different polymorphic PCMs between extant African, Asian and European populations. Alleles that have been the target of localised positive selection tend to exhibit unusually high F_{ST} values.[22,23] We therefore compared the F_{ST} values of the ancestral polymorphic PCMs with the empirical F_{ST} distribution derived from all HapMap SNPs (International HapMap Consortium, 2007),[24] to assess the significance of individual F_{ST} values. We identified six PCMs with significantly elevated F_{ST} values (Table 5).

Although four of these PCMs had already been identified in our previous comparative analysis of the human, chimpanzee and Neanderthal genomes,[10] two novel PCMs were identified in the putative cation exchanger *SLC24A5* (DP) gene and in the alcohol dehydrogenase *ADH1B* (FP) gene. These genes have in common the GO terms

Table 5. PCMs (disease-causing and disease-related) with significantly different genotype frequencies in different HapMap populations

Gene	rs	HGMD Acc	WT	PCM	Asian		European		African		Pair-wise F_{ST} (p value)		
					f_{WT}	n	f_{WT}	n	f_{WT}	n	Asian-African	European-African	Asian-European
SLC24A5	rs1426654	CM054862	A	G	0.01	178	1.00	116	0.03	120	−0.001 (0.8490)	0.974 (0.0054)	0.987 (0.0010)
*TP53BP1**	rs2602141	CM067476	T	G	0.52	176	0.69	120	0.00	120	0.470 (0.2830)	0.689 (0.0489)	0.054 (0.5701)
*CAPN3**	rs1801449	CM099258	G	A	0.91	178	0.94	120	0.23	120	0.653 (0.2234)	0.143 (0.3877)	0.680 (0.0026)
*TP53BP1**	rs560191	CM067475	G	C	0.52	178	0.69	120	0.00	120	0.475 (0.2981)	0.689 (0.0489)	0.051 (0.5536)
ADH1B	rs1229984	CM890003	T	C	0.75	178	0.00	120	0.00	118	0.715 (0.1576)	NA (NA)	0.717 (0.0197)
*ENPP1**	rs1044498	CM993455	A	C	0.94	180	0.87	118	0.00	120	0.927 (0.0314)	0.873 (0.0110)	0.020 (0.6004)

*Previously reported by Zhang *et al.*[10]
rs: reference number, dbSNP, WT: wild type, f_{WT}: frequency of the wild-type allele, NA: Not applicable.

GO:0046872, GO:0043169 and GO:0043167, terms which relate to metal ion binding, cation binding and ion binding, respectively. The *SLC24A5* variant appears to be associated with increased skin pigmentation and predominates in African/East Asian populations.[25,26]

In conclusion, using the newly reported genome sequence from a Denisovan hominin, we have identified a number of PCMs in the chimpanzee, Neanderthal and Denisovan. Those human PCMs that were ancestral (ie both the Denisovan nucleotide and the chimpanzee nucleotide were identical to the human DM/disease-associated mutation) could potentially be indicative of either the human lineage-specific loss of compensatory nucleotide changes within the respective genes carrying the PCM, or adaptive differences between modern humans and Denisovans.

References

1. Green, R.E., Krause, J., Briggs, A.W., Maricic, T. *et al.* (2010), 'A draft sequence of the Neanderthal genome', *Science* Vol. 328, pp. 710–722.
2. Noonan, J.P. (2010), 'Neanderthal genomics and the evolution of modern humans', *Genome Res.* Vol. 20, pp. 547–553.
3. Gibbons, A. (2010), 'Paleogenetics. Close encounters of the prehistoric kind', *Science* Vol. 328, pp. 680–684.
4. Reich, D., Green, R.E., Kircher, M., Krause, J. *et al.* (2010), 'Genetic history of an archaic hominin group from Denisova Cave in Siberia', *Nature* Vol. 468, pp. 1053–1060.
5. Reich, D., Green, R.E., Kircher, M., Krause, J. *et al.* (2010), 'The complete mitochondrial DNA genome of an unknown hominin from southern Siberia', *Nature* Vol. 468, pp. 1053–1060.
6. Krause, J., Fu, Q., Good, J.M., Viola, B. *et al.* (2010), 'The complete mitochondrial DNA genome of an unknown hominin from southern Siberia', *Nature* Vol. 464, pp. 894–897.
7. Martinón-Torres, M., Dennell, R. and Bermúdez de Castro, J.M. (2011), 'The Denisova hominin need not be an out of Africa story', *J. Hum. Evol.* Vol. 60, pp. 251–255.
8. Di Rienzo, A. and Hudson, R.R. (2005), 'An evolutionary framework for common diseases: The ancestral-susceptibility model', *Trends Genet.* Vol. 21, pp. 596–601.
9. Crespi, B.J. (2010), 'The origins and evolution of genetic disease risk in modern humans', *Ann. N. Y. Acad. Sci.* Vol. 1206, pp. 80–109.
10. Zhang, G., Pei, Z., Krawczak, M., Ball, E.V. *et al.* (2010), 'Triangulation of the human, chimpanzee, and Neanderthal genome sequences identifies potentially compensated mutations', *Hum. Mutat.* Vol. 31, pp. 1286–1293.
11. Gao, L. and Zhang, J. (2003), 'Why are some human disease-associated mutations fixed in mice?', *Trends Genet.* Vol. 19. pp. 678–681.
12. Azevedo, L., Suriano, G., van Asch, B., Harding, R.M. and Amorim, A. (2006), 'Epistatic interactions: How strong in disease and evolution?', *Trends Genet.* Vol. 22, pp. 581–585.
13. Ferrer-Costa, C., Orozco, M. and de la Cruz, X. (2007), 'Characterization of compensated mutations in terms of structural and physico-chemical properties'. *J. Mol. Biol.* Vol. 365, pp. 249–256.
14. Corona, E., Dudley, J.T. and Butte, A.J. (2010), 'Extreme evolutionary disparities seen in positive selection across seven complex diseases', *PLoS One* Vol. 5, p. e12236.
15. Baresić, A., Hopcroft, L.E., Rogers, H.H., Hurst, J.M. *et al.* (2010), 'Compensated pathogenic deviations: Analysis of structural effects', *J. Mol. Biol.* Vol. 396, pp. 19–30.
16. Stenson, P.D., Mort, M., Ball, E.V., Howells, K. *et al.* (2009), 'The Human Gene Mutation Database: 2008 update', *Genome Med.* Vol. 1, p. 13.
17. Huang da, W., Sherman, B.T. and Lempicki, R.A. (2009), 'Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources', *Nat. Protoc.* Vol. 4, pp. 44–57.
18. Benjamini, Y. and Hochberg, Y. (1995), 'Controlling the false discovery rate: A practical and powerful approach to multiple testing', *J. R. Stat. Soc. Series B* Vol. 57, pp. 289–300.
19. Weir, B.S. and Hill, W.G. (2002), 'Estimating F-statistics', *Annu. Rev. Genet.* Vol. 36, pp. 721–750.
20. Dawood, F., Mountford, R., Farquharson, R. and Quenby, S. (2007), 'Genetic polymorphisms on the factor V gene in women with recurrent miscarriage and acquired APCR'. *Hum. Reprod.* Vol. 22, pp. 2546–2553.
21. Chegeni, R., Kazemi, B., Hajifathali, A., Pourfathollah, A. *et al.* (2007), 'Factor V mutations in Iranian patients with activated protein C resistance and venous thrombosis', *Thromb. Res.* Vol. 119, pp. 189–193.
22. Holsinger, K.E. and Weir, B.S. (2009), 'Genetics in geographically structured populations: Defining, estimating and interpreting F_{ST}', *Nat. Rev. Genet.* Vol. 10, pp. 639–650.
23. Thornton, K.R. and Jensen, J.D. (2007), 'Controlling the false-positive rate in multilocus genome scans for selection', *Genetics* Vol. 175, pp. 737–750.
24. International HapMap Consortium, Frazer, K.A., Ballinger, D.G., Cox, D.R. *et al.* (2007), 'A second generation human haplotype map of over 3.1 million SNPs'. *Nature* Vol. 449, pp. 851–861.
25. Lamason, R.L., Mohideen, M.A., Mest, J.R., Wong, A.C. *et al.* (2005), '*SLC24A5*, a putative cation exchanger, affects pigmentation in zebrafish and humans', *Science* Vol. 310, pp. 1782–1786.
26. Stokowski, R.P., Pant, P.V., Dadd, T., Fereday, A. *et al.* (2007), 'A genomewide association study of skin pigmentation in a South Asian population', *Am. J. Hum. Genet.* Vol. 81, pp. 1119–1132.

Table S1. Significantly enriched GO terms (Benjamini-corrected p-value < 0.05) for human genes containing DMs identified as PCMs (listed in Table 4) against a background of known disease-causing genes. No significantly enriched GO terms were found to relate to biological processes or molecular function

GO Term	Category	Description	Fold enrichment	p-Value	Genes
GO:0031224	Cellular component	Intrinsic to membrane	2.12	4.29E-03	*SLC5A1, DUOX2, CNGA3, ABCA4, SLC26A2, MFRP, ABCB4, IL12RB2, SLC26A4, IL12RB1, CRB1, SLC17A3, PKP2, NPHS1, RYR1, EXT1, SEZ6, ATP7B*
GO:0016021	Cellular component	Integral to membrane	2.21	4.59E-03	*SLC5A1, DUOX2, CNGA3, ABCA4, SLC26A2, MFRP, ABCB4, IL12RB2, SLC26A4, IL12RB1, CRB1, SLC17A3, PKP2, NPHS1, RYR1, EXT1, SEZ6, ATP7B*
GO:0005886	Cellular component	Plasma membrane	2.17	5.49E-03	*SLC5A1, DUOX2, ABCA4, SLC26A2, ABCB4, IL12RB2, SLC26A4, IL12RB1, ANK1, CRB1, SLC17A3, F5, PKP2, NPHS1, RYR1, SEZ6, ATP7B*
GO:0031226	Cellular component	Intrinsic to plasma membrane	3.02	3.92E-02	*IL12RB2, IL12RB1, SLC17A3, SLC5A1, NPHS1, RYR1, ABCA4, SLC26A2, ATP7B, ABCB4*
GO:0005887	Cellular component	Integral to plasma membrane	3.11	3.93E-02	*IL12RB2, IL12RB1, SLC17A3, SLC5A1, NPHS1, RYR1, ABCA4, SLC26A2, ATP7B, ABCB4*

Table S2. PCMs covered by both the Denisovan sequence and the Neanderthal sequence

HGMD Acc	Chr	Location	Strand	Tag	Disease	Gene	Mutation	AA seq	Type
CM031993	Chr1	9246497	+	DFP	Cortisone reductase deficiency, partial	*H6PD*	G > A:AAA	Arg-Gln	Human
CM040788	Chr1	11828655	−	DP	Stroke, increased risk, association with	*NPPA*	A > G:GGG	Term-Arg	Human
CM100611	Chr1	12005513	+	DFP	Breast cancer, reduced risk, association with	*MIIP*	A > G:GGG	Lys-Glu	Human
CM980072	Chr1	21767322	+	DFP	Hypophosphatasia, association with	*ALPL*	T > C:CCC	Tyr-His	Human
CM056598	Chr1	31865112	+	DP	Polydipsia–hyponatraemia, association with	*HCRTR1*	A > G:GAA	Ile-Val	Chimpanzee
CM994122	Chr1	35033356	+	DFP	Atherosclerosis, association with	*GJA4*	C > T:TTT	Pro-Ser	Human
CM065514	Chr1	55410663	−	DP	Parkinson's disease, risk, association with	*USP24*	G > A:AAA	Thr-Ile	Human
CM073141	Chr1	67457975	+	DP	Psoriasis, increased risk, association with	*IL23R*	T > C:CCC	Leu-Pro	Human
CM993347	Chr1	67633930	+	DM	Atopy	*IL12RB2*	A > G:GAA	His-Arg	Chimpanzee
CM067986	Chr1	86873963	+	DP	Chloride channel deficiency, association with	*CLCA3P*	C > G:GGG	Tyr-Term	Human
CM042258	Chr1	94337039	−	DM	Stargardt disease	*ABCA4*	T > G:GGT	Lys-Gln	Denisova and chimpanzee
CM067656	Chr1	156491643	+	DP	Guillain–Barré syndrome, reduced risk, association with?	*CD1A*	C > G:GGC	Cys-Trp	Denisova and chimpanzee
CM070090	Chr1	167765599	−	DM	Thrombosis?	*F5*	C > T:CTC	Val-Met	Denisovan
CM099896	Chr1	173615346	−	DP	Schizophrenia, association with	*TNR*	C > T:TTT	Arg-Lys	Human
CM023569	Chr1	199313698	−	DP	Hypokalaemic periodic paralysis, association with?	*CACNA1S*	G > A:GRA	Gly-Gly	Unsure
CM920010	Chr1	228912417	−	DP	Hypertension, association with	*AGT*	A > G:GGG	Met-Thr	Human

Continued

Table S2. Continued

HGMD Acc	Chr	Location	Strand	Tag	Disease	Gene	Mutation	AA seq	Type
CM065155	Chr1	240108924	+	DP	Colorectal cancer, increased risk, association with	*EXO1*	G > A:AAA	Glu-Lys	Human
CM033447	Chr10	42926693	+	DP	Hirschsprung disease, association with	*RET*	A > G:GGG	Ala-Ala	Human
CM068190	Chr10	54198272	−	FP	Increased serum mannose-binding lectin (MBL) level, association with?	*MBL2*	C > G:CGG	Leu-Leu	Ancient
CM033482	Chr10	64085190	+	DP	Uric acid nephrolithiasis, association with	*znf365d*	G > A:GGA	Ala-Thr	Neanderthal
CM067461	Chr10	81691702	−	DP	Lung cancer, susceptibility to, association with	*SFTPD*	T > C:CCC	Thr-Ala	Human
CM035804	Chr11	524242	−	DP	Bladder cancer, association with?	*HRAS*	A > G:GAG	His-His	Neanderthal and chimpanzee
CM025891	Chr11	74585230	+	FP	Decreased enzyme activity, association with	*SLCO2B1*	C > T:TTT	Ser-Phe	Human
CM080415	Chr11	113308238	+	FP	Altered receptor function, association with	*HTR3B*	A > C:CCC	Tyr-Ser	Human
CM950862	Chr12	5473868	+	DP	Schizophrenia, severe, increased risk, association with	*NTF3*	G > A:AGG	Gly-Glu	Chimpanzee
CM093840	Chr12	6023795	−	DP	von Willebrand disease, quantitative type, association with	*VWF*	T > C:CCC	Thr-Ala	Human
CM994637	Chr12	6327323	−	DFP	Hypertension, reduced risk, association with	*SCNN1A*	T > C:CCC	Thr-Ala	Human
CM003671	Chr12	14884706	−	FP	Dombrock blood group variation	*ART4*	T > C:TCC	Asn-Asp	Ancient
CM077900	Chr12	70659129	+	FP	Increased mRNA expression, association with?	*TPH2*	G > A:GAA	Pro-Pro	Ancient

Continued

Table S2. Continued

HGMD Acc	Chr	Location	Strand	Tag	Disease	Gene	Mutation	AA seq	Type
CM085048	Chr12	78539038	−	DP	Schizophrenia in females, association with	*PAWR*	A > C:CCC	Ile-Met	Human
CM033453	Chr12	107542027	−	DFP	Coronary heart disease, decreased risk, in African Americans, association with	*SELPLG*	C > T:TTT	Met-Ile	Human
CM022034	Chr13	32526193	+	DP	Age-related phenotypes, association with	*KL*	G > C:CGG	Cys-Ser	Chimpanzee
CM033777	Chr14	24170122	−	DP	Apoptosis, unable to induce, association with	*GZMB*	A > G:GGG	Tyr-His	Human
CM070246	Chr14	60993992	+	DFP	Cerebral infarction, association with	*PRKCH*	G > A:AAA	Val-Ile	Human
CM067476	Chr15	41511938	−	DP	Lung cancer, susceptibility to, association with	*TP53BP1*	T > G:GGG	Lys-Gln	Human
CM067475	Chr15	41555066	−	DP	Lung cancer, susceptibility to, association with	*TP53BP1*	G > C:CCC	Asp-Glu	Human
CM085365	Chr15	43185730	−	DM	Hypothyroidism	*DUOX2*	T > C:CCC	His-Arg	Human
CM054862	Chr15	46213776	+	DP	Increased skin pigmentation, association with	*SLC24A5*	A > G:GGG	Thr-Ala	Human
CM057869	Chr15	76704628	−	FP	Altered function, association with	*CHRNB4*	T > C:CTT	Met-Val	Chimpanzee
CM031698	Chr15	97295748	+	DP	Increased longevity, association with?	*IGF1R*	G > A:AGG	Glu-Glu	Chimpanzee
CM057585	Chr16	1442858	−	DP	Lower femoral neck bone mineral density in women, association with	*CLCN7*	C > T:TCT	Val-Met	Neanderthal and chimpanzee
CM983400	Chr16	27263704	+	DFP	Asthma, atopic, association with	*IL4R*	A > G:GGG	Ile-Val	Human
CM067985	Chr16	87788983	+	DP	Cadherin deficiency, association with	*CDH15*	C > A:AAA	Tyr-Term	Human

Continued

Table S2. Continued

HGMD Acc	Chr	Location	Strand	Tag	Disease	Gene	Mutation	AA seq	Type
CM057933	Chr17	4585312	−	DP	Atherosclerotic stenosis, increased, association with	*CXCL16*	G > A:AAG	Ala-Val	Denisova and chimpanzee
CM077855	Chr17	7532893	+	DP	Breast cancer, oestrogen receptor (ER) negative, association with?	*WRAP53*	C > G:GGG	Arg-Gly	Human
CM087381	Chr17	7987497	−	FP	Increased sex hormone-binding globulin levels, association with	*PER1*	C > G:GGG	Ala-Pro	Human
CM067489	Chr17	16468520	−	DP	Lung cancer, susceptibility to, association with	*ZNF624*	C > A:AAA	Lys-Asn	Human
CM030773	Chr17	19753133	−	DP	Cardiac disease, susceptibility to, association with	*AKAP10*	T > C:CCC	Ile-Val	Human
CM067336	Chr17	19802050	−	DP	Lung cancer, susceptibility to, association with	*AKAP10*	C > T:TTT	Arg-His	Human
CM096315	Chr17	38498462	−	DFP	Cervical cancer, decreased risk, association with	*BRCA1*	G > A:AAA	Pro-Leu	Human
CM093418	Chr17	39581073	+	DP	Hip bone mineral density, association with?	*C17orf53*	A > C:CCC	Thr-Pro	Human
CM032397	Chr17	41432502	+	DP	Progressive supranuclear palsy, association with	*STH*	A > G:GAA	Gln-Arg	Chimpanzee
CM064363	Chr17	45788957	+	DP	Organ involvement in pseudoxantnoma elasticum (PXE), association with	*XYLT2*	T > C:CCC	Tyr-Tyr	Human
CM092499	Chr17	76468818	+	FP	Altered splicing, association with?	*KIAA1303*	A > G:GAA	Gln-Gln	Chimpanzee
CM080431	Chr19	11091881	+	FP	Increased plasma low-density lipoprotein cholesterol, association with	*LDLR*	T > C:CTT	Val-Val	Chimpanzee

Continued

Table S2. Continued

HGMD Acc	Chr	Location	Strand	Tag	Disease	Gene	Mutation	AA seq	Type
CM984025	Chr19	18047618	−	DM	Mycobacterial infection	*IL12RB1*	T > C:CCT	Gln-Arg	Denisova and chimpanzee
CM044918	Chr19	41022117	−	DM	Congenital nephrotic syndrome, Finnish type	*NPHS1*	C > G:GGG	Val-Leu	Human
CM073386	Chr19	50087554	+	DP	Alzheimer's disease, late-onset, association with?	*TOMM40*	T > C:CCC	Phe-Phe	Human
CM004814	Chr19	50546759	−	DFP	Basal cell carcinoma, reduced risk, association with	*ERCC2*	T > G:GGG	Lys-Gln	Human
CM096319	Chr2	11276571	−	DP	Chronic kidney disease in individuals with low triglycerides, association with	*ROCK2*	G > T:TGT	Thr-Asn	Neanderthal and chimpanzee
CM052876	Chr2	49043425	−	DP	Menstrual cycle length, association with	*FSHR*	C > T:TTT	Ser-Asn	Human
CM073086	Chr2	85634047	−	DP	Higher body mass index, association with	*GGCX*	C > T:TCC	Arg-Gln	Chimpanzee
CM087379	Chr2	100957736	+	FP	Higher testosterone levels, association with	*NPAS2*	A > G:GGG	Thr-Ala	Human
CM004559	Chr2	227369287	−	DP	Diabetes, type 2, association with	*IRS1*	T > C:CCT	Ala-Ala	Denisova and chimpanzee
CM085146	Chr2	227839413	+	DP	Chronic obstructive pulmonary disease, association with	*COL4A3*	A > G:GAA	His-Arg	Chimpanzee
CM014824	Chr20	4653718	+	DP	Creutzfeldt–Jakob disease, association with	*PRND*	C > T:TCT	Thr-Met	Neanderthal and chimpanzee
CM064121	Chr20	44075813	+	DP	Leukaemia, risk, association with	*MMP9*	G > C:CCC	Arg-Pro	Human
CM035699	Chr21	14403236	−	FP	Plasma high-density lipoprotein (HDL) cholesterol, association with	*LIPI*	G > T:TTT	Asp-Glu	Human

Continued

Table S2. Continued

HGMD Acc	Chr	Location	Strand	Tag	Disease	Gene	Mutation	AA seq	Type
CM057711	Chr21	33536125	+	DP	Multiple sclerosis, susceptibility to, association with	*IFNAR2*	T > G:GGG	Phe-Val	Human
CM025479	Chr21	44534334	+	DP	Alopecia universalis, association with	*AIRE*	C > G:GGG	Ser-Arg	Human
CM057927	Chr22	21957369	+	DP	Bipolar disorder, association with?	*BCR*	A > G:GGG	Asn-Ser	Human
CM065332	Chr22	24489289	+	DP	Colorectal cancer, increased risk, association with	*MYO18B*	G > A:AAA	Gly-Glu	Human
CM961339	Chr22	30836050	+	DM	Glucose/galactose malabsorption	*SLC5A1*	C > G:GGC	His-Gln	Denisova and chimpanzee
CM096696	Chr22	35792882	−	DP	Iron status and erythrocyte volume, association with	*TMPRSS6*	A > G:GGG	Val-Ala	Human
CM092918	Chr22	37827350	+	FP	Increased antiretroviral activity, association with	*APOBEC3H*	G > C:CCC	Gly-Arg	Human
CM910052	Chr22	49410905	−	DP	Phenotype modifier, association with?	*ARSA*	G > C:CCC	Thr-Ser	Human
CM023348	Chr3	336508	+	DP	Schizophrenia, association with	*CHL1*	C > T:TTT	Leu-Phe	Human
CM096382	Chr3	46476217	−	DFP	Periodontitis, aggressive, association with	*LTF*	T > C:CCC	Lys-Arg	Human
CM066581	Chr3	126109714	−	DP	Ulcerative colitis, association with	*MUC13*	T > G:GGG	Arg-Ser	Human
CM941277	Chr3	172214994	−	DP	Diabetes, type 2, association with	*SLC2A2*	G > A:AAG	Thr-Ile	Denisova and chimpanzee
CM065290	Chr3	187925712	+	DP	Nephropathy, reduced risk, association with	*KNG1*	T > C:CCC	Met-Thr	Human
CM025429	Chr4	2960297	+	FP	Increased enzymatic activity, association with	*GRK4*	G > T:TTT	Arg-Leu	Human
CM094340	Chr4	38476105	−	DFP	Leprosy, association with	*TLR1*	T > C:CCC	Asn-Ser	Human

Continued

Table S2. Continued

HGMD Acc	Chr	Location	Strand	Tag	Disease	Gene	Mutation	AA seq	Type
CM890003	Chr4	100458342	−	FP	Alcohol dehydrogenase beta variant	*ADH1B*	T > C:CCC	His-Arg	Human
CM092574	Chr4	123756413	−	DFP	Asthma, atopic, association with	*IL21*	G > A:AAA	Cys-Cys	Human
CM031390	Chr4	141708518	−	DP	Waist-to-hip ratio, association with	*UCP1*	C > T:TTT	Ala-Thr	Human
CM004732	Chr5	1464412	−	DP	Parkinson's disease, protection against, association with?	*SLC6A3*	T > C:CTC	Ser-Ser	Neanderthal and chimpanzee
CM094298	Chr5	96165006	−	DFP	Cervical carcinoma survival, association with	*ERAP1*	C > G:GGG	Arg-Pro	Human
CM0910115	Chr5	131424377	+	DP	Graves disease, association with	*IL3*	C > T:TTT	Pro-Ser	Human
CM043093	Chr6	25958824	−	DM	Glycogen storage disease 1c?	*SLC17A3*	C > T:TCC	Gly-Arg	Chimpanzee
CM074911	Chr6	39433056	−	DP	Coronary heart disease, association with	*KIF6*	A > G:GGG	Trp-Arg	Human
CM020385	Chr6	74550153	+	FP	Gov platelet antigen variation	*CD109*	A > C:CCC	Tyr-Ser	Human
CM993455	Chr6	132214061	+	DFP	Insulin resistance, association with	*ENPP1*	A > C:CCC	Lys-Gln	Human
CM060415	Chr6	150156438	+	FP	Reduced stability, association with	*PCMT1*	A > G:AGG	Ile-Val	Ancient
CM072043	Chr6	160462998	+	FP	Reduced metformin uptake, association with	*SLC22A1*	C > T:TCC	Ser-Phe	Chimpanzee
CM005460	Chr7	17345635	+	FP	Higher induced cytochrome P-450 (CYP) 1A1 activity, association with	*AHR*	G > A:AAA	Arg-Lys	Human
CM055287	Chr7	45899194	+	DP	Renal function in diabetes, association with	*IGFBP1*	A > G:GGA	Ile-Met	Denisova and chimpanzee
CM072814	Chr7	86894112	−	DM	Intrahepatic cholestasis, familial progressive?	*ABCB4*	T > C:CCC	Arg-Gly	Human

Continued

Table S2. Continued

HGMD Acc	Chr	Location	Strand	Tag	Disease	Gene	Mutation	AA seq	Type
CM064968	Chr7	91468556	+	DP	Colorectal cancer, increased risk, association with	*AKAP9*	G > T:TTT	Met-Ile	Human
CM930596	Chr7	94775382	−	DFP	Longevity, association with	*PON1*	T > C:CCC	Gln-Arg	Human
CM050323	Chr7	107129530	+	DM	Pendred syndrome?	*SLC26A4*	T > G:GTG	Val-Gly	Neanderthal and chimpanzee
CM060083	Chr7	122422409	−	DP	Alcohol dependence, risk, association with	*TAS2R16*	A > C:CAA	Asn-Lys	Chimpanzee
CM031370	Chr7	141319073	−	DP	Phenylthiocarbamide taste sensitivity, association with	*TAS2R38*	T > C:CCC	Ile-Val	Human
CM031368	Chr7	141319814	−	DP	Phenylthiocarbamide taste sensitivity, association with	*TAS2R38*	C > G:GGG	Ala-Pro	Human
CM081694	Chr8	6466450	+	DP	Cranial volume, association with	*MCPH1*	C > T:TTT	Ala-Val	Human
CM024569	Chr8	18124476	+	FP	Increased enzymatic activity, association with	*NAT1*	T > G:GTG	Ser-Ala	Neanderthal and chimpanzee
CM983990	Chr8	22032655	−	DM	Alopecia universalis?	*HR*	T > C:CCC	Thr-Ala	Human
CM057431	Chr8	27518398	−	DP	Preeclampsia & essential hypertension, association with?	*CLU*	A > G:GGG	His-His	Human
CM950017	Chr8	37942955	−	DFP	Hyperinsulinaemia, association with	*ADRB3*	A > G:GGG	Trp-Arg	Human
CM099178	Chr8	118899878	−	DM	Multiple osteochondromas	*EXT1*	C > T:TCC	Val-Ile	Chimpanzee
CM081761	Chr8	143758933	+	DFP	Gastric cancer, diffuse-type, association with	*PSCA*	C > T:TTT	Thr-Met	Human
CM094855	Chr9	14712477	−	DP	Low bone mineral density, association with	*CER1*	G > C:CGC	Ala-Gly	Neanderthal and chimpanzee
CM940804	Chr9	34639442	+	DFP	Galactosaemia, Duarte variant	*GALT*	A > G:GAG	Asn-Asp	Neanderthal and chimpanzee

Continued

Table S2. Continued

HGMD Acc	Chr	Location	Strand	Tag	Disease	Gene	Mutation	AA seq	Type
CM071685	Chr9	89511843	+	DP	Inactivation of extracellular signal-regulated kinase (ERK)-induced apoptosis, association with	*DAPK1*	A > G:AGG	Asn-Ser	Ancient
CM990005	Chr9	106626574	−	FP	Higher plasma HDL cholesterol, association with	*ABCA1*	T > C:CCC	Ile-Met	Human
CM0910114	ChrX	77414973	−	DP	Asthma, association with	*CYSLTR1*	G > A:AAA	Phe-Phe	Human
CM085353	ChrX	149390017	+	DM	Hypospadias	*MAMLD1*	T > C:CYC	Val-Ala	Unsure
CR043164	Chr1	43575707	+	DP	Platelet count, association with?	*MPL*	C > A:AAA		Human
CR060579	Chr1	111020443	−	DP	Low insulin sensitivity, association with	*KCNA3*	T > C:TCC		Ancient
CR057791	Chr1	111571946	+	FP	Increased promoter activity, association with	*CHI3L2*	G > T:GGT		Neanderthal
CR031479	Chr1	170894121	+	DFP	Systemic lupus erythematosus (SLE), association with	*FASLG*	C > T:TTT		Human
CR025943	Chr1	228917021	−	DP	Increased angiotensinogen levels, association with?	*AGT*	G > A:AGG		Chimpanzee
CR102882	Chr10	64279946	−	DFP	SLE, association with	*EGR2*	C > T:TCC		Chimpanzee
CR102883	Chr10	64280724	−	DFP	SLE, association with	*EGR2*	T > C:CTT		Chimpanzee
CR072313	Chr10	94452862	+	DP	Diabetes, type 2, association with?	*HHEX*	C > T:TCC		Chimpanzee
CR942079	Chr10	104587142	−	DP	Polycystic ovaries, association with	*CYP17A1*	A > G:GGG		Human
CR012509	Chr11	34416293	+	DP	Hypertension, susceptibility to, association with	*CAT*	G > A:AGA		Neanderthal and chimpanzee
CR072303	Chr11	44212190	+	DP	Diabetes, type 2, reduced risk, association with?	*EXT2*	C > T:CTT		Ancient

Continued

Table S2. Continued

HGMD Acc	Chr	Location	Strand	Tag	Disease	Gene	Mutation	AA seq	Type
CR035965	Chr11	45863406	+	DFP	Alzheimer's disease, association with	*MAPK8IP1*	A > G:GGG		Human
CR094845	Chr11	74539529	+	FP	Increased mRNA expression, association with	*SLCO2B1*	G > A:AAA		Human
CR045957	Chr11	102101690	−	DFP	Preterm premature rupture of membranes, association with?	*MMP8*	G > A:GGA		Neanderthal
CR025510	Chr11	102331749	−	FP	Increased transcriptional activity, association with	*MMP13*	C > T:TCC		Chimpanzee
CR031478	Chr12	10203556	−	DP	Alzheimer disease, reduced risk, association with	*OLR1*	G > A:AAG		Denisova and chimpanzee
CR082031	Chr12	55796928	−	DP	Schistosomiasis infection, association with	*STAT6*	C > T:TTT		Human
CR087739	Chr13	42046024	+	DFP	Bone mineral density in osteoporosis, association with?	*TNFSF11*	C > T:CTC		Denisovan
CR080758	Chr13	45577313	−	FP	Increased promoter activity, association with	*CPB2*	T > C:CTT		Chimpanzee
CR994765	Chr13	112807756	+	DFP	Reduced plasma F7 levels, association with	*F7*	G > T:CTT		Unsure
CR066661	Chr15	49336891	−	DP	Alzheimer's disease in apolipoprotein E4 (APOE4) carriers, increased risk, association with	*CYP19A1*	G > A:AAA		Human
CR002154	Chr15	56511231	+	DP	Dyslipidaemia and insulin resistance, association with	*LIPC*	G > A:AGG		Chimpanzee
CR993820	Chr15	72828970	+	DFP	Increased activity in smokers, association with	*CYP1A2*	C > A:AAA		Human

Continued

Table S2. Continued

HGMD Acc	Chr	Location	Strand	Tag	Disease	Gene	Mutation	AA seq	Type
CR102187	Chr16	13921167	+	DFP	Bladder cancer, increased risk, association with	*ERCC4*	A > C:CAA		Chimpanzee
CR066332	Chr16	54244319	+	DFP	Attention-deficit hyperactivity disorder, association with	*SLC6A2*	A > T:ATA		Denisovan
CR000229	Chr16	55552737	+	DFP	Higher HDL cholesterol level, association with	*CETP*	C > A:AAA		Human
CR084012	Chr17	25549137	−	FP	Increased expression, association with	*SLC6A4*	A > C:CCC		Human
CR035881	Chr17	29706729	+	FP	Increased monocyte chemoattractant protein-4 (MCP-4) plasma levels, association with	*CCL13*	C > T:CCT		Neanderthal
CR003707	Chr17	31231893	−	DFP	Atopic dermatitis, association with	*CCL5*	C > T:TTT		Human
CR078280	Chr17	35323475	−	DP	Asthma, increased risk, association with?	*GSDMB*	C > T:TTC		Denisova and chimpanzee
CR090198	Chr17	38531642	−	FP	Promoter activity, association with	*BRCA1*	T > C:CCC		Human
CR052976	Chr17	43163827	+	DP	Asthma, aspirin-induced, association with	*TBX21*	T > C:CCC		Human
CR084013	Chr17	43178034	+	DP	Genital herpes simplex virus-2 (HSV-2) infection, association with?	*TBX21*	G > A:AAA		Human
CR051707	Chr19	7718733	−	DFP	Dengue disease, protection against, association with	*CD209*	A > G:GGG		Human
CR095376	Chr19	40464739	+	DP	Increased liver iron concentration	*HAMP*	A > G:GGA		Denisova and chimpanzee
CR050427	Chr19	46188969	+	FP	CYP2B6 expression, association with	*CYP2B6*	T > C:CCC		Human

Continued

Table S2. Continued

HGMD Acc	Chr	Location	Strand	Tag	Disease	Gene	Mutation	AA seq	Type
CR051274	Chr19	54149750	+	DFP	Disease progression, chronic lymphocytic leukaemia, association with	*BAX*	G > A:GAG		Denisovan
CR010588	Chr19	60077416	+	DP	Immunoglobulin A nephropathy, association with	*FCAR*	T > C:CCC		Human
CR051277	Chr2	69468799	−	DP	Obesity, association with	*GFPT1*	C > T:TTT		Human
CR025220	Chr2	234330398	+	DFP	Hyperbilirubinaemia, association with	*UGT1A1*	T > G:GGG		Human
CR075263	Chr20	17370063	+	DP	Diabetes, type 2, association with	*PCSK2*	T > C:CCC		Human
CR077665	Chr20	44066518	+	FP	Increased expression, association with?	*MMP9*	C > T:TTT		Human
CR078166	Chr21	33619134	+	FP	Increased expression, association with	*IFNAR1*	T > C:CCT		Denisova and chimpanzee
CR054260	Chr21	38590628	+	FP	Promoter activity, association with	*KCNJ15*	T > G:GTT		Chimpanzee
CR096274	Chr21	42492734	+	DFP	Coronary artery disease, severity, association with	*ABCG1*	T > G:GGT		Denisova and chimpanzee
CR032439	Chr3	12328198	+	DFP	Increased height/lipid metabolism, association with	*PPARG*	C > G:GGG		Human
CR066664	Chr3	129680794	−	DP	Coronary artery disease, association with	*GATA2*	G > A:AGG		Chimpanzee
CR014438	Chr3	185572960	−	DP	Myocardial infarction, association with	*THPO*	C > T:TTT		Human
CR004797	Chr4	26101320	−	DP	Higher percentage body fat, association with	*CCKAR*	C > A:ACC		Chimpanzee
CR045948	Chr4	69995928	+	FP	Promoter activity, association with	*UGT2B7*	G > A:AAA		Human
CR025435	Chr4	111053559	+	DFP	Malignant melanoma, association with	*EGF*	A > G:GGG		Human

Continued

Table S2. Continued

HGMD Acc	Chr	Location	Strand	Tag	Disease	Gene	Mutation	AA seq	Type
CR057903	Chr4	155703465	+	DFP	Cerebral infarction, association with	*FGB*	C > T:TTT		Human
CR071281	Chr4	156348632	+	DP	Obesity, association with	*NPY2R*	C > T:TTT		Human
CR071289	Chr5	1499389	−	DP	Attention-deficit hyperactivity disorder, association with	*SLC6A3*	A > G:GGG		Human
CR086597	Chr5	110434641	+	FP	Increased promoter activity, association with	*TSLP*	C > T:TCC		Chimpanzee
CR035513	Chr5	131436741	+	DP	Reduced severity in atopic dermatitis, association with	*CSF2*	A > C:CCC		Human
CR015845	Chr5	132020708	+	DP	Asthma, association with	*IL13*	C > T:TTC		Denisova and chimpanzee
CR082018	Chr6	78227843	−	DFP	Aggressive behaviour, association with	*HTR1B*	C > T:TCC		Chimpanzee
CR073540	Chr6	131935252	+	DP	Myocardial infarction, association with	*ARG1*	G > T:TTT		Human
CR052970	Chr6	132254387	+	DP	Obesity, association with	*ENPP1*	A > G:GGG		Human
CR075243	Chr6	132314950	−	DFP	Systemic sclerosis, association with	*CTGF*	C > G:CGG		Ancient
CR075274	Chr6	133077018	−	DP	HDL cholesterol concentration, association with	*VNN1*	A > C:CCC		Human
CR077383	Chr6	154401054	+	FP	Increased promoter activity, association with	*OPRM1*	A > G:GGG		Human
CR066667	Chr7	30969948	+	DP	Breast cancer, decreased risk, association with	*GHRHR*	C > T:TTT		Human
CR092300	Chr7	111902894	+	DFP	Severity in cystic fibrosis, association with	*IFRD1*	C > T:TTT		Human
CR068449	Chr7	128381961	+	DP	SLE, association with?	*IRF5*	C > T:TTT		Human

Continued

Table S2. Continued

HGMD Acc	Chr	Location	Strand	Tag	Disease	Gene	Mutation	AA seq	Type
CR022507	Chr7	136351848	+	DP	Major depression in women, association with	*CHRM2*	T > A:AAA		Human
CR971950	Chr8	19840951	+	FP	Lower plasma triglyceride level, association with	*LPL*	T > G:GGG		Human
CR023703	Chr8	120034205	−	DP	Decreased bone mineral density, association with?	*TNFRSF11B*	C > T:TTT		Human
CR084001	Chr9	70877744	+	DP	Myocardial infarction, association with	*FXN*	C > T:TTT		Human
CR102176	Chr9	100952292	+	DFP	Breast cancer, association with	*TGFBR1*	A > G:GGG		Human
CR020827	Chr9	106730271	−	DP	Increased risk of coronary artery disease, association with	*ABCA1*	G > A:AAA		Human
CR045560	Chr9	106730659	−	FP	Reduced plasma HDL cholesterol, association with	*ABCA1*	C > G:GGG		Human
CR091269	Chr9	116608587	−	DFP	Crohn's disease, susceptibility to, association with	*TNFSF15*	A > G:GGG		Human
CR034594	Chr9	124172343	+	FP	Inhibition of prostaglandin H2 formation, association with?	*PTGS1*	A > G:GAG		Neanderthal and chimpanzee
CR054255	Chr9	127043845	−	DP	Bipolar disorder, association with?	*HSPA5*	T > C:CCC		Human
CR077381	ChrX	113724838	+	FP	Reduced promoter activity, association with	*HTR2C*	G > C:CGG		Chimpanzee
CR063398	ChrX	135554616	+	FP	Increased soluble CD40 ligand (CD40L) levels, association with	*CD40LG*	A > G:GGG		Human

Table S3. PCMs covered by the Denisovan sequence but not the Neanderthal sequence

Acc	Chr	Location	Strand	Tag	Disease	Gene	Mutation	AA seq	Type
CM062419	chr1	19483828	−	DP	Leukaemia, risk, association with	*AKR7A3*	C > T:CT	Asp-Asn	Denisovan
CM098300	chr1	24074507	−	DFP	Eating disorders, association with	*CNR2*	T > C:CC	Gln-Arg	Ancestral
CM066774	chr1	110267989	+	DP	Periodontitis, association with?	*CSF1*	T > C:CC	Leu-Pro	Ancestral
CM094244	chr1	111656412	+	FP	Increased enzyme activity, association with?	*CHIA*	A > G:GA	Asn-Asp	Derived
CM094243	chr1	111656461	+	DFP	Asthma, protection against, association with?	*CHIA*	G > T:TG	Arg-Met	Derived
CM084968	chr1	150552554	−	DP	Psoriasis, increased risk, association with	*FLG*	G > A:AG	Pro-Ser	Derived
CM067657	chr1	156591049	+	DP	Guillain–Barré syndrome, reduced risk, association with	*CD1E*	A > G:GG	Gln-Arg	Ancestral
CM033904	chr1	169444714	+	FP	Flavin-containing monooxygenase 2 (FMO2) gene variant	*FMO2*	T > C:CC	Term-Gln	Ancestral
CM043273	chr1	195670491	+	DM	Retinitis pigmentosa	*CRB1*	G > A:AG	Gly-Ser	Derived
CM024366	chr1	224093029	+	DFP	Preeclampsia, association with	*EPHX1*	A > G:GA	His-Arg	Derived
CM994344	chr10	115795046	+	FP	Gain of function, association with	*ADRB1*	G > C:CC	Gly-Arg	Ancestral
CM067436	chr11	7020956	+	DM	Spermatogenic failure	*NLRP14*	G > A:AG	Ala-Thr	Derived
CM043536	chr11	47326617	−	DM	Cardiomyopathy, hypertrophic?	*MYBPC3*	T > C:CT	Ser-Gly	Derived
CM035848	chr11	57739196	+	FP	Olfactory receptor deficiency?	*OR1S1*	G > A:GA	Arg-His	Denisovan
CM087504	chr11	102218830	−	DP	Blood pressure, association with	*MMP3*	T > C:CC	Lys-Glu	Ancestral
CM041241	chr11	112776038	+	FP	Reduced dopamine D2 receptor (DRD2) receptor density, association with?	*ANKK1*	G > A:AA	Glu-Lys	Ancestral

Continued

Table S3. Continued

Acc	Chr	Location	Strand	Tag	Disease	Gene	Mutation	AA seq	Type
CM082943	chr11	118720796	−	DM	Primary angle-closure glaucoma?	*MFRP*	C > T:TT	Arg-His	Ancestral
CM075018	chr11	130255852	−	DP	Coronary heart disease, association with	*SNX19*	A > C:CC	Leu-Arg	Ancestral
CM091988	chr12	32913201	−	DM	Arrhythmogenic right ventricular cardiomyopathy	*PKP2*	A > G:GG	Leu-Pro	Ancestral
CM087618	chr12	56152088	+	DFP	Inflammatory bowel disease, association with	*GLI1*	G > C:CC	Glu-Gln	Ancestral
CM098354	chr12	120099486	+	FP	Altered function, association with	*P2RX7*	G > A:AA	Ala-Thr	Ancestral
CM065186	chr13	38162690	+	DP	Colorectal cancer, increased risk, association with	*FREM2*	T > C:CC	Phe-Ser	Ancestral
CM063919	chr13	45546095	−	FP	Higher thrombin-activatable fibrinolysis inhibitor (TAFI) antigen levels, association with	*CPB2*	C > T:TT	Ala-Thr	Ancestral
CM044579	chr13	51413355	−	DM	Wilson disease?	*ATP7B*	A > G:GG	Val-Ala	Ancestral
CM063843	chr14	19994994	+	DFP	Amyotrophic lateral sclerosis, association with	*APEX1*	T > G:GG	Asp-Glu	Ancestral
CM073244	chr14	20010446	+	DP	Faster cognitive decline in Alzheimer's disease, association with	*NP*	G > A:AG	Gly-Ser	Derived
CM068495	chr15	49316404	−	DP	Increased cortical bone mass density, association with	*CYP19A1*	T > C:CC	Val-Val	Ancestral
CM045806	chr15	83248435	+	FP	Reduced affinity for gemcitabine, association with	*SLC28A1*	G > A:AG	Val-Ile	Derived
CM102885	chr16	10908349	+	DP	Multiple sclerosis, increased risk, association with	*CIITA*	G > C:CC	Gly-Ala	Ancestral

Continued

Table S3. Continued

Acc	Chr	Location	Strand	Tag	Disease	Gene	Mutation	AA seq	Type
CM093131	chr16	55950234	+	DP	*Helicobacter pylori*-related gastric carcinoma, association with	*CCL22*	A > C:CC	Asp-Ala	Ancestral
CM067679	chr17	7858004	+	DP	Lung cancer, susceptibility to, association with	*GUCY2D*	T > A:AA	Leu-His	Ancestral
CM073339	chr17	24310977	−	DM	Febrile seizures?	*SEZ6*	T > C:CC	Thr-Ala	Ancestral
CM057951	chr17	37960432	+	DP	Endometriosis, association with	*HSD17B1*	A > G:AG	Ser-Gly	Denisovan
CM994214	chr17	39808591	−	DP	Reduced post-stroke mortality, association with	*ITGA2B*	A > C:GC	Ile-Ser	Unsure
CM091892	chr17	42363569	+	DP	Hypertension, association with	*GOSR2*	G > A:AG	Arg-Lys	Derived
CM091876	chr17	73642170	+	DP	Epidermodysplasia verruciformis, susceptibility in HIV, association with	*TMC8*	A > T:TA	Asn-Ile	Derived
CM000831	chr19	3546794	−	DP	Bronchial asthma, association with	*TBXA2R*	A > G:GG	Tyr-Tyr	Ancestral
CM030470	chr19	18041451	−	DP	Tuberculosis, susceptibility to, association with	*IL12RB1*	A > G:GG	Met-Thr	Ancestral
CM044082	chr19	18407678	−	DP	Spina bifida, reduced risk, association with	*isyna1*	T > C:CC	Leu-Leu	Ancestral
CM057586	chr19	40534926	+	DP	Increased beta-cell function, association with	*FFAR1*	G > A:AA	Arg-His	Ancestral
CM057545	chr19	50560149	−	DP	Lung adenocarcinoma, increased risk, association with	*ERCC2*	G > T:GT	Arg-Arg	Denisovan
CM044227	chr19	60088712	+	DP	Aggressive periodontitis, reduced risk, assoc with	*FCAR*	A > G:GG	Arg-Arg	Ancestral
CM003809	chr2	38155681	−	DP	Breast or lung cancer, association with	*CYP1B1*	C > A:AA	Ala-Ser	Ancestral

Continued

Table S3. Continued

Acc	Chr	Location	Strand	Tag	Disease	Gene	Mutation	AA seq	Type
CM101950	chr2	98363138	+	DM	Progressive cone dystrophy?	*CNGA3*	C > T:TC	Pro-Leu	Derived
CM092797	chr2	169550992	−	FP	Alternate splicing, association with	*ABCB11*	T > C:CT	Gly-Gly	Derived
CM066575	chr2	218738088	−	DP	AIDS progression, protection, association with	*IL8RA*	A > C:CC	Met-Arg	Ancestral
CM057769	chr2	234266408	+	FP	Altered enzyme activity, association with	*UGT1A6*	T > G:GG	Ser-Ala	Ancestral
CM910018	chr2	241466189	+	DP	Hyperoxaluria, association with	*AGXT*	A > G:GG	Ile-Met	Ancestral
CM053304	chr20	54257212	+	DP	Obesity, association with	*MC3R*	C > A:AA	Thr-Lys	Ancestral
CM970391	chr22	18331207	+	DFP	Schizoaffective disorder, association with	*COMT*	C > G:GG	Leu-Leu	Ancestral
CM961335	chr22	30817700	+	DM	Glucose/galactose malabsorption	*SLC5A1*	G > A:AA	Ala-Thr	Ancestral
CM930187	chr22	40853887	−	DP	Parkinson's disease, association with	*CYP2D6*	G > A:GA	Arg-Cys	Denisovan
CM099899	chr22	41888870	+	FP	Increased pregnenolone levels, association with	*TSPO*	A > G:GG	Thr-Ala	Ancestral
CM025430	chr4	2975841	+	FP	Activity, association with	*GRK4*	C > T:TT	Ala-Val	Ancestral
CM013959	chr4	23424760	−	DP	Diabetes, type 2, association with	*PPARGC1A*	C > T:TC	Gly-Ser	Derived
CM033593	chr4	100479812	−	DP	Alcoholism, increased risk, association with?	*ADH1C*	T > C:CC	Ile-Val	Ancestral
CM064956	chr4	109893565	−	DP	Colorectal cancer, increased risk, association with	*AGXT2L1*	A > G:GG	Ser-Pro	Ancestral
CM030066	chr4	149576925	−	FP	Reduced expression, association with	*NR3C2*	T > C:TC	Ile-Val	Denisovan
CM080365	chr4	155711209	+	DP	Increased clot stiffness, association with	*FGB*	G > A:AA	Arg-Lys	Ancestral

Continued

Table S3. Continued

Acc	Chr	Location	Strand	Tag	Disease	Gene	Mutation	AA seq	Type
CM057405	chr4	156355126	+	DP	Severe obesity, in men, association with	*NPY2R*	C > T:TT	Ile-Ile	Ancestral
CM067358	chr5	22114341	−	DP	Lung cancer, susceptibility to, association with	*CDH12*	C > T:TT	Val-Met	Ancestral
CM094788	chr5	121441107	−	DFP	Breast cancer, increased risk, in African American women, association with	*LOX*	C > T:TT	Arg-Gln	Ancestral
CM013815	chr5	147461148	+	DP	Atopy, maternally inherited, association with	*SPINK5*	G > A:GA	Glu-Lys	Denisovan
CM083577	chr6	24611569	+	DFP	Impaired cognitive function, association with	*ALDH5A1*	C > T:TT	His-Tyr	Ancestral
CM086146	chr6	25921129	−	DP	Uric acid concentration, association with	*SLC17A1*	G > A:AA	Thr-Ile	Ancestral
CM052232	chr6	80683094	−	DP	Age-related maculopathy, association with	*ELOVL4*	T > C:CT	Met-Val	Derived
CM073245	chr7	34784638	+	DP	Panic disorder, in males, association with	*NPSR1*	A > T:TT	Asn-Ile	Ancestral
CM084696	chr7	87017537	−	DFP	Parkinson's disease, association with	*ABCB1*	A > G:GG	Gly-Gly	Ancestral
CM091200	chr7	129737976	+	DP	Prostate cancer, aggressive early-onset, association with	*CPA4*	G > T:TT	Gly-Cys	Ancestral
CM952203	chr7	142350235	−	FP	Kell blood group variation	*KEL*	A > G:GA	Leu-Pro	Derived
CM073993	chr7	150188598	+	FP	Reduced activity, association with	*ABP1*	C > G:GG	His-Asp	Ancestral
CM973386	chr8	18124281	+	FP	Increased activity, association with	*NAT1*	G > A:AG	Val-Ile	Derived
CM099895	chr8	24412708	+	DP	Schizophrenia, association with	*ADAM7*	A > C:CC	Asn-His	Ancestral

Continued

Table S3. Continued

Acc	Chr	Location	Strand	Tag	Disease	Gene	Mutation	AA seq	Type
CM064954	chr8	26683945	−	DP	Hypertension, association with?	*ADRA1A*	A > G:GG	Cys-Arg	Ancestral
CM033767	chr8	27414422	+	DFP	Coronary heart disease, in Caucasians, association with	*EPHX2*	A > G:GA	Lys-Arg	Derived
CM034886	chr8	91059655	−	DP	Lung cancer, association with?	*NBN*	C > G:GG	Glu-Gln	Ancestral
CM045665	chr8	120033233	−	DP	Osteoporotic fractures, association with	*TNFRSF11B*	G > C:CG	Asn-Lys	Derived
CM093465	chr9	2181309	+	DFP	Schizophrenia, association with	*SMARCA2*	C > G:GC	Asp-Glu	Derived
CM073190	chrX	43475980	+	DP	Bipolar disorder, association with?	*MAOA*	T > G:TG	Arg-Arg	Denisovan
CR072321	chr1	11841858	−	DFP	Diabetes, type, 2, reduced risk, association with	*NPPB*	A > G:GG		Ancestral
CR080762	chr1	15645754	+	DM	Pancreatitis, chronic?	*CTRC*	T > C:CC		Ancestral
CR080761	chr1	15645757	+	DM	Pancreatitis, chronic?	*CTRC*	A > G:GG		Ancestral
CR016187	chr1	87101113	−	FP	Increased selenocysteine insertion sequence (SECIS) efficiency, association with	*sep15*	C > T:TT		Ancestral
CR092707	chr1	201194130	−	DFP	Lower insulin resistance, association with	*ADIPOR1*	C > T:TT		Ancestral
CR034628	chr10	26545502	+	DP	Obesity, association with?	*GAD2*	G > A:GA		Denisovan
CR061340	chr11	35397552	−	DFP	Progressing stroke, increased risk, association with	*SLC1A2*	T > G:GG		Ancestral
CR068212	chr11	59612604	+	DFP	Asthma, aspirin-intolerant	*MS4A2*	T > C:CC		Ancestral
CR063407	chr14	50069895	−	DP	Diabetes, type 2, reduced risk, association with	*MAP4K5*	G > A:AA		Ancestral
CR077666	chr15	71712835	−	DFP	Schizophrenia, reduced risk, association with?	*NPTN*	C > A:CA		Denisovan

Continued

Table S3. Continued

Acc	Chr	Location	Strand	Tag	Disease	Gene	Mutation	AA seq	Type
CR084880	chr17	35697157	+	DFP	Hepatocellular carcinoma, reduced risk, association with	*CDC6*	A > G:GG		Ancestral
CR087465	chr17	39785770	+	DFP	Frontotemporal dementia, association with	*GRN*	C > T:TT		Ancestral
CR035036	chr18	647685	+	FP	Transcriptional activity, association with	*TYMS*	G > C:CC		Ancestral
CR032436	chr18	45342041	+	DP	High-density lipoprotein (HDL) cholesterol levels, association with?	*LIPG*	A > C:CA		Derived
CR087182	chr19	44589133	+	DFP	Rheumatoid arthritis, shorter duration, association with	*ZFP36*	A > G:GG		Ancestral
CR035033	chr19	46188301	+	FP	Cytochrome P-450 (CYP) 2B6 expression, association with?	*CYP2B6*	T > C:CC		Ancestral
CR068525	chr2	69467665	−	DFP	Diabetes, type 2, association with	*GFPT1*	A > G:GG		Ancestral
CR077669	chr2	85748849	−	FP	Increased promoter activity, association with	*SFTPB*	T > G:GG		Ancestral
CR093507	chr2	168743982	−	DFP	Hypertension, association with	*STK39*	A > G:GG		Ancestral
CR093026	chr2	169465787	+	DFP	Increased insulin secretion, association with	*G6PC2*	G > A:AA		Ancestral
CR073559	chr2	224174588	−	DFP	Hypertension, association with	*SCG2*	C > T:TT		Ancestral
CR053505	chr20	4653756	+	DP	Creutzfeldt–Jakob disease, association with?	*PRND*	T > C:CC		Ancestral
CR015272	chr22	40858326	−	FP	Intermediate metaboliser, association with?	*CYP2D6*	C > G:GG		Ancestral
CR055620	chr4	75938792	−	FP	Promoter activity, association with	*BTC*	C > G:GG		Ancestral

Continued

Table S3. Continued

Acc	Chr	Location	Strand	Tag	Disease	Gene	Mutation	AA seq	Type
CR093469	chr6	2945302	+	DFP	Breast cancer, decreased risk, association with	*NQO2*	A > C:CA		Derived
CR035882	chr6	78230101	−	DFP	Suicidal ideation, in major depression, association with	*HTR1B*	A > C:CA		Derived
CR025333	chr6	137582213	−	DFP	Malaria, susceptibility, association with	*IFNGR1*	A > G:GG		Ancestral
CR093919	chr6	153121754	+	DP	Pulmonary arterial hypertension, idiopathic, association with?	*VIP*	T > C:CC		Ancestral
CR016149	chr7	22732746	+	FP	Altered transcriptional activity, association with	*IL6*	A > G:GG		Ancestral
CR053504	chr7	91995822	−	FP	Gene expression, association with	*PEX1*	A > G:GA		Derived
CR041138	chr7	99192235	−	DP	Prostate cancer, low aggressiveness, association with	*CYP3A4*	G > A:AG		Derived
CR072316	chr7	128376663	+	FP	Shorter transcript, association with	*IRF5*	G > A:AA		Ancestral
CR962526	chr8	41774321	−	DM	Spherocytosis	*ANK1*	A > G:GA		Derived
CR098013	chr9	22109195	+	DFP	Coronary artery disease, association with	*CDKN2BAS*	C > T:CT		Denisovan
CR044772	chr9	99499399	−	DP	Lung adenocarcinoma, risk, association with	*XPA*	T > C:CC		Ancestral
CR020828	chr9	106730356	−	DP	Reduced risk of coronary artery disease, association with	*ABCA1*	G > C:CC		Ancestral
CR052068	chr9	136911887	+	FP	Promoter activity, association with	*FCN2*	A > G:GG		Ancestral
CR042847	chr9	138995962	+	DP	HDL cholesterol, association with?	*PTGDS*	A > C:CC		Ancestral