Exploring the potential relevance of human-specific genes to complex disease

David N. Cooper[1*] *and Hildegard Kehrer-Sawatzki*[2]

[1]Institute of Medical Genetics, School of Medicine, Cardiff University, Cardiff, UK
[2]Institute of Human Genetics, University of Ulm, Ulm, Germany
*Correspondence to: Tel: +44 2920 744062; Fax: +44 2920 746551; E-mail: cooperDN@cardiff.ac.uk

Date received (in revised form): 21st November 2010

Abstract

Although human disease genes generally tend to be evolutionarily more ancient than non-disease genes, complex disease genes appear to be represented more frequently than Mendelian disease genes among genes of more recent evolutionary origin. It is therefore proposed that the analysis of human-specific genes might provide new insights into the genetics of complex disease. Cross-comparison with the Human Gene Mutation Database (http://www.hgmd.org) revealed a number of examples of disease-causing and disease-associated mutations in putatively human-specific genes. A sizeable proportion of these were missense polymorphisms associated with complex disease. Since both human-specific genes and genes associated with complex disease have often experienced particularly rapid rates of evolutionary change, either due to weaker purifying selection or positive selection, it is proposed that a significant number of human-specific genes may play a role in complex disease.

Keywords: *human-specific genes, evolution, disease genes, complex disease, mutations, polymorphisms*

Human 'disease genes' have been known for some time to differ significantly from 'non-disease genes' in terms of their higher degree of evolutionary conservation.[1–3] Further, with respect to their evolutionary age, human disease genes appear not to be simply a random subset of all genes in the genome but are instead biased toward being of ancient (early metazoan) origin.[4] Concomitantly, a pronounced paucity of human lineage-specific genes is also evident among disease genes.[4] These initial findings were subsequently confirmed and elaborated upon by Cai *et al.*,[5] who determined the approximate age of evolutionary emergence of all human genes and then proceeded to compare disease genes with non-disease genes with respect to whether they were 'young', 'middle-aged' or 'old-aged'. For the purposes of their study, the origin of a given gene was determined by retracing its orthologues back to the species most distantly related to human. Genes that originated during the period since the adaptive radiation of the Laurasiatheria were described as 'young', the term 'middle-aged' was employed to describe those genes whose origin went back to the bony fish, and genes that emerged at some stage between yeast and *Ciona* (a tunicate) were ascribed the term 'old-aged'. Using these fairly crude descriptors of gene age, Cai *et al.*[5] confirmed that there was a tendency for Mendelian disease genes (ie those genes underlying single gene disorders) to be of more ancient evolutionary origin than non-disease genes. With Mendelian disease genes, the 'old-aged' genes were in the majority, closely followed by the 'middle-aged' genes. By contrast, most genes involved in the aetiology of complex disease were found to reside in the 'middle-aged' category. Although both Mendelian and complex disease genes were found to be under-represented in the 'young' category, the frequency of complex disease genes in this category was found to be more than twice that exhibited by the Mendelian disease genes.[5] Given this finding, we speculated that closer examination of the most recently acquired

(ie human-specific) genes might well provide new insights into the genetics of complex disease.

Considerable efforts have been made to identify those genes that have been inactivated in the human lineage but which are still present in other higher primate species, including chimpanzee.[6–9] However, somewhat less attention has been paid so far either to human genes of relatively recent origin that are specific to the human lineage[10,11] or to those genes that have been retained in the human genome despite having been lost in other primate species.[12]

The first reported attempts to identify putative human-specific gene duplications were those of Fortna *et al.*[13] and Cheng *et al.*[14] The dataset of Cheng *et al.* comprised 88 complete gene duplications that were considered to have occurred since the divergence of human and chimpanzee. Of these, it was found that 13 have been reported in association with human inherited disease (see Human Gene Mutation Database (HGMD)[15]). These are the genes encoding Fc fragment of IgG, high affinity Ia, receptor (*FCGR1A*), DEAD box 1 protein (*DDX11*), cholinergic nicotinic acetylcholine receptor, alpha 7 subunit, exons 5–10 and family with sequence similarity 7A, exons A–E fusion (*CHRFAM7A*), aryl sulphotransferase (*SULT1A1*), *CC* chemokine ligand 4-like 1 (*CCL4L1*), killer cell immunoglobulin-like receptor, three domains, long cytoplasmic tail, 1 (*KIR3DL1*), mammalian STE20-like kinase 1 (*MST1*), dopamine receptor D5 (*DRD5*), succinate dehydrogenase complex (*SDHA*), survivor motor neurone 2 (*SMN2*), general transcription factor 2-I repeat domain-containing protein 2 (*GTF2IRD2*), neutrophil cytosolic factor 1 (*NCF1*), and aquaporin 7 (*AQP7*). Despite Cheng *et al.*[14] presenting expression data to support their claim that their 88 gene duplications involved functional duplicated gene copies, however, a question mark has remained over whether or not some of these genes may actually represent pseudogenes.

Probably, the most reliable dataset of human-specific gene duplications so far produced is that of Itan *et al.*[16] These workers identified 138 human-specific complete gene duplications that appear to have occurred since the divergence of human and chimpanzee. It was found that four of these human-specific genes are listed in the HGMD as having been reported in association with human inherited disease (Table 1). Indeed, these genes have been shown to harbour a number of different disease-causing mutations (DMs; including missense mutations and copy number variations) or disease-associated polymorphisms that may confer increased risk of a given disease state. In two of the four cases, the reported disease association was between a disease-associated polymorphism and a complex disease phenotype (ie susceptibility to infectious disease).

Intuitively, it might be supposed that those human genes that are present in more than one copy, as a consequence of a human lineage-specific increase in copy number, are better protected against the consequences of mutation in their functional parent genes by virtue of their newly acquired genetic redundancy.[27,28] Using fairly stringent selection criteria, at least 27 human genes have so far been identified as having experienced a human lineage-specific increase in copy number.[13,29–32] Contrary to expectation, nine of these genes (*AQP7*), cadherin 12, type 2 (*CDH12*), *CHRFAM7A*, *DRD5*, *FCGR1A*, *GTF2IRD2*, neuronal apoptosis inhibitory protein (*NAIP*), *NCF1*, and occludin (*OCLN*) are listed in the HGMD, although only five of them (*FCGR1A*, *GTF2IRD2*, *NAIP*, *NCF1*, *OCLN*) have been reported to harbour mutations that actually cause inherited disease (ie DMs; Table 2). It would be of considerable interest to ascertain (retrospectively) whether the particular patients in whom these mutations were described do indeed harbour extra functional copies of the relevant genes or whether perhaps these mutations have come to clinical attention precisely because the disease genes are effectively single copy in these particular individuals.[27,28] An alternative explanation could, however, involve the loss of genetic redundancy through the post-duplication functional divergence of the gene copies leading to diversification through sub- or neo-functionalisation.[42–45] It is premature to speculate as to which of these postulates might provide an explanation for the

Table 1. Human-specific genes that have been associated with inherited disease. The four genes listed were taken from a total of 138 human-specific genes identified by Itan *et al.*[16]

Gene symbol	Gene name	Chrom. loc.	Disease-associated mutation (DM)	HGMD accession No.	HGMD entry tag	dbSNP No.	Associated disease state	Reference No.
CCL3L1	Chemokine (C-C motif) ligand 3-like 1	17q12	CNV (duplication of entire gene)	CN052767	DP	—	HIV/AIDS susceptibility, association with	17
							Kawasaki disease	18
							Rheumatoid arthritis, association with	19
							Chronic hepatitis C infection, susceptibility to	20
FOXD4	Forkhead box D4	9p24.3	Missense (Trp148Arg)	CM073074	DM	—	Dilated cardiomyopathy, obsessive-compulsive disorder and suicidality	21
MRC1	Mannose receptor, C type 1	10p12.33	Missense (Gly396Ser)	CM099897	DP	rs1926736	Leprosy, protection against/association with	22
SMN2	Survival of motor neurone 2, centromeric	5q13.2	CNV (duplication of entire gene)	CN082433	DM	—	Spinal muscular atrophy 3	23
						—	Spinal muscular atrophy, modifier of	24
			Missense (Gly287Arg)	CM095436	DM			25
			Deletion (exons 7 and 8)	CG015657	DM	—	Spinal muscular atrophy	26

Abbreviations: Chrom. loc., chromosomal localisation; CNV, copy number variant; dbSNP, the Single Nucleotide Polymorphism database; DP, disease-associated polymorphism in statistically significant association with a particular disease state but lacking experimental evidence of functionality.

observation that five of the nine genes listed in Table 2 can harbour clinically important mutations.

Another approach to this whole issue is to identify a set of human genes whose orthologues have been lost from the chimpanzee genome and then to ascertain how many of these genes are known to be involved in human inherited disease. The initial analysis of the chimpanzee genome (Chimpanzee Sequencing and Analysis Consortium, 2005) identified a total of 55 genes that are present in human but have been lost (or irrevocably disrupted) in chimpanzee. It should be noted that these genes may not necessarily be *bona fide* human-specific genes, since although they are absent from the chimpanzee genome, they may still be present in the genomes of other higher primates. This notwithstanding, cross-checking with the HGMD revealed that eight of these genes are known to be associated with either human inherited disease or disease susceptibility (Table 3). In seven of these eight cases, the reported disease association was between a disease-associated polymorphism and a complex disease phenotype (susceptibility to infectious disease or autoimmune disease). Although the numbers involved are clearly small, it would appear that there is at least a tendency for human-specific genes to harbour a greater-than-expected number of examples of polymorphisms associated with complex disease by comparison with the number of mutations causing Mendelian disease (bearing in mind that disease-associated polymorphisms constitute only a small minority of the lesions logged in the HGMD; see legend to Table 3).

Another intriguing finding emerges if Tables 1–3 are considered together: *NAIP*, *OCLN* and *SMN2* are all located at 5q13.2; *GTF2IRD2* and *NCF1* are both located at 7q11.23, whereas butyrophilin-like 2 (*BTNL2*), HLA complex P5 (*HCP5*) and MHC class I polypeptide-related sequence A (*MICA*) are all located at 6p21.3. The most likely explanation for the clustering of the human-specific genes at 5q13.2 and 7q11.23 is that both of these chromosomal regions became duplicated specifically during the human lineage. On the other hand, the clustering of

Table 2. Human genes identified as having experienced a human lineage-specific increase in copy number and which harbour mutations causing, or associated with, inherited disease. The nine genes listed were taken from a total of 27 human genes identified as having experienced a human lineage-specific increase in copy number.[13,30–32]

Gene symbol	Gene name	Chrom. loc.	Disease-associated mutation (DM)	HGMD accession No.	HGMD entry tag	dbSNP No.	Associated disease state	Reference No.
AQP7	Aquaporin 7	9p13.3	Missense (Gly264Val)	CM023765	DP	rs62542743	No exercise-induced glycerol increase, association with	33
CDH12	Cadherin 12, type 2	5p14.3	Missense (Val68Met)	CM067358	DP	rs4371716	Lung cancer susceptibility, association with	34
CHRFAM7A	*CHRNA7* (cholinergic nicotinic acetylcholine receptor, alpha 7 subunit, exons 5–10) and *FAM7A* (family with sequence similarity 7A, exons A-E) fusion	15q13.2	Micro-deletion (coding region)	CD025514	DFP	—	P50 sensory gating deficit, association with	35
DRD5	Dopamine receptor D5	4p16.1	Nonsense (Cys335Term)	CM995180	FP	—	Dopamine receptor D5 deficiency	36
FCGR1A	Fc fragment of IgG, high affinity Ia, receptor	1q21.2	Nonsense (Arg92Term)	CM950456	DM	—	IgG receptor I, phagocytic, deficiency	37
GTF2IRD2	General transcription factor 2-I repeat domain-containing protein 2	7q11.23	Complete gene deletion	CG044469	DM	—	Williams–Beuren syndrome features	38
NAIP	NLR family, apoptosis inhibitory protein	5q13.2	Complete gene deletion Partial gene deletion	CG952277 CG952278	DM DM	—	Spinal muscular atrophy	39
NCF1	Neutrophil cytosolic factor 1	7q11.23	Nonsense (Gln91Term)	CM065336	DM	—	Chronic granulomatous disease	40
OCLN	Occludin	5q13.2	Missense (Phe219Ser)	CM105655	DM	—	Band-like calcification, simplified gyration and polymicrogyria	41

Abbreviations: Chrom. loc., chromosomal localisation; dbSNP, the Single Nucleotide Polymorphism database; DP, disease-associated polymorphism in statistically significant association with a particular disease state but lacking experimental evidence of functionality; FP, functional polymorphism.

Table 3. Human genes associated with an inherited disease that have been lost from the chimpanzee genome. The eight genes listed were taken from a total of 55 genes that are present in human but which have been lost or been irrevocably disrupted in chimpanzee.[46]

Gene symbol	Gene name	Chrom. loc.	Disease-associated mutation (DM)	HGMD accession No.	HGMD entry tag	dbSNP No.	Associated disease state	Reference No.
APOL1	Apolipoprotein L1	22q12.3	Missense (Ser342Gly) Missense (Ile384Met) Microdeletion (6 bp)	CM105041 CM105043 CD105044	DFP DFP DFP	rs73885319 rs60910145 rs71785313	Resistance to trypanosomes	47
BTNL2	Butyrophilin-like 2	6p21.32	Splice site mutation (IVS5 DS -1 G > A)	CS051245	DFP	rs2076530	Sarcoidosis, association with	48
CD24	CD24 molecule	6q21	Missense (Ala57Val)	CM035761	DFP	rs52812045	Multiple sclerosis, modifier of susceptibility to Lupus erythematosus, susceptibility to Chronic hepatitis B infection, risk of	49 50 51
HCP5	HLA complex P5	6p21.3	Missense (Val112Gly)	CM074273	DP	rs2395029	Reduced HIV viral load, association with	52
MICA	MHC class I polypeptide-related sequence A	6p21.33	Missense (Lys196Glu) Missense (Val152Met)	CM0910116 CM058256	DP DP	rs1051794 rs1051792	Rheumatoid arthritis, association with Ankylosing spondylitis, early onset, association with	53 54
NBEA	Neurobeachin	13q13.3	Balanced translocation t(5;13)(q12.1;q13.2)	CP035454	DM	—	Autism, idiopathic	55

Continued

Table 3. Continued

Gene symbol	Gene name	Chrom. loc.	Disease-associated mutation (DM)	HGMD accession No.	HGMD entry tag	dbSNP No.	Associated disease state	Reference No.
RNASE3	Ribonuclease, RNase A family, 3	14q11.2	Missense (Thr124Arg)	CM025442	DFP	rs2073342	Expression of allergic symptoms	56
							Schistosoma mansoni infection, risk of	57
SFTPA1	Surfactant protein A1	10q22.3	Missense (Leu50Val)	CM033024	DP	rs1136450	Idiopathic pulmonary fibrosis, association with	58
			Missense (Arg219Trp)	CM033026	DP	rs4253527	Idiopathic pulmonary fibrosis, association with	58
							Tubersclerosis, susceptibility to	59

Abbreviations: Chrom. loc., chromosomal localisation; dbSNP, the Single Nucleotide Polymorphism database; DFP, disease-associated polymorphism with supporting evidence of functionality; DP: disease-associated polymorphism in statistically significant association with a particular disease state but lacking experimental evidence of functionality. Of 105,018 unique mutations currently listed in the HGMD (October 2010), 99,716 are DMs, 2,065 are DPs, 2,238 are functional polymorphisms (FPs) and 999 are DFPs. Of the 3,851 different human genes listed in the HGMD, 2,629 contain DMs, while 1,222 contain only DFPs or FPs (in the absence of DMs).

the *BTNL2*, *HCP5* and *MICA* genes at 6p21.3 is likely to be due to the loss of this chromosomal region in chimpanzee. The clustering of human-specific genes at 5q13.2 has been noted previously.[13] The present data, however, suggest that at least two additional genomic regions (7q11.23 and 6p21.3) contain clusters of human-lineage specific genes.

Despite the above cited examples, there still appear to be relatively few 'young' (human-specific) genes among human disease genes. It may therefore be inferred that only a few young genes perform functions of sufficient importance to ensure that they have immediately come to clinical attention when mutated. This view is certainly supported by the observation that younger human, primate and mammalian genes tend to evolve more rapidly and are subject to weaker purifying selection than their more ancient counterparts.[45,60–62] If it is assumed that the underlying mutation rate does not differ markedly between these categories of gene, there are essentially two potential explanations for very young genes experiencing a particularly rapid rate of evolutionary change: weaker purifying selection or an increase in positive selection. The available evidence suggests that both explanations are likely to pertain.[62–65] The majority of newly duplicated genes experience a period of relaxed purifying selection, while only a relatively small proportion exhibits a signature of positive selection consequent to the acquisition of new biological functions.[66] Presumably, in between the initial post-duplicational redundancy and eventual neofunctionalisation through genetic divergence, there is a 'half-way house' state in which a gene is relatively free to explore the acquisition of new functions while still being constrained to some extent by selection against the loss of those functions it has already acquired.[67]

Whereas mutant alleles responsible for single gene disorders are usually under negative selection, alleles associated with complex disease appear to have been either under much less stringent purifying selection or may even have been subject to positive selection.[3,68,69] Genes that have experienced positive selection during the human lineage are likely to be characterised by human-specific functional adaptations. Since genes that have been subject to positive selection during human evolution have frequently also been implicated in disease,[70] we may surmise that the underlying pathological mutations may sometimes have interfered with these newly acquired functions. Consistent with this interpretation, Lappalainen *et al.*[71] have reported a statistically significant correlation between those regions of the human genome that have experienced recent positive selection in northern European populations and those regions that have been implicated in complex disease. This observation is also compatible with the tendency observed by the present authors for human-specific genes to harbour a disproportionate number of polymorphic variants associated with complex disease.

In terms of their expression profiles, there appears to be a tendency for rapidly evolving younger genes to be tissue-specific, whereas the more slowly evolving older genes are more broadly expressed.[72–76] It remains to be seen whether complex disease genes differ from Mendelian disease genes in terms of their expression characteristics.

More generally, it has been proposed that 'taxonomically restricted genes' may play a role in the generation of morphological diversity, thereby enabling organisms to adapt to changing environmental conditions.[77] If this is also true for the human lineage-specific genes discussed here, it follows that the acquisition of mutations in these genes may well reduce an individual's capacity to deal with a rapidly changing environment.[78] This chimes well with the (essentially unrelated) idea that some forms of common disease susceptibility may be a consequence of ancient human adaptations to a long-term stable environment ('thrifty alleles'). These ancestral alleles may now increase the risk of common disease in a changed environment consequent to the recent shift to a modern lifestyle.[79] The present authors therefore suggest that it would be a worthwhile exercise to construct a complete lexicon of human-specific genes, since these loci may well provide a happy hunting

ground for those seeking to identify genes that play a key role in complex disease.

References

1. Huang, H., Winter, E.E., Wang, H., Weinstock, K.G. *et al.* (2004), 'Evolutionary conservation and selection of human disease gene orthologs in the rat and mouse genomes', *Genome Biol.* Vol. 5, p. R47.
2. Tu, Z., Wang, L., Xu, M., Zhou, X. *et al.* (2006), 'Further understanding human disease genes by comparing with housekeeping genes and other genes', *BMC Genomics* Vol. 7, p. 31.
3. Blekhman, R., Man, O., Herrmann, L., Boyko, A.R. *et al.* (2008), 'Natural selection on genes that underlie human disease susceptibility', *Curr. Biol.* Vol. 18, pp. 883–889.
4. Domazet-Lošo, T. and Tautz, D. (2008), 'An ancient evolutionary origin of genes associated with human genetic diseases', *Mol. Biol. Evol.* Vol. 25, pp. 2699–2707.
5. Cai, J.J., Borenstein, E., Chen, R. and Petrov, D.A. (2009), 'Similarly strong purifying selection acts on human disease genes of all evolutionary ages', *Genome Biol. Evol.* Vol. 1, pp. 131–144.
6. Kehrer-Sawatzki, H. and Cooper, D.N. (2007), 'Understanding the recent evolution of the human genome: insights from human-chimpanzee genome comparisons', *Hum. Mutat.* Vol. 28, pp. 99–130.
7. Yu, G. (2009), 'An exon-based comparative variant analysis pipeline to study the scale and role of frameshift and nonsense mutation in the human-chimpanzee divergence', *Comp. Funct. Genomics*, Article ID 406421.
8. Schrider, D.R., Costello, J.C. and Hahn, M.W. (2009), 'All human-specific gene losses are present in the genome as pseudogenes', *J. Comput. Biol.* Vol. 16, pp. 1419–1427.
9. Kim, H.L., Igawa, T., Kawashima, A., Satta, Y. *et al.* (2009), 'Divergence, demography and gene loss along the human lineage', *Philos. Trans. R. Soc. B* Vol. 365, pp. 2451–2457.
10. Nahon, J.L. (2003), 'Birth of "human-specific" genes during primate evolution', *Genetica* Vol. 118, pp. 193–208.
11. Stahl, P.D. and Wainszelbaum, M.J. (2009), 'Human-specific genes may offer a unique window into human cell signalling', *Sci. Signal.* Vol. 2, p. e59.
12. Knowles, D.G. and McLysaght, A. (2009), 'Recent *de novo* origin of human protein-coding genes', *Genome Res.* Vol. 19, pp. 1752–1759.
13. Fortna, A., Kim, Y., MacLaren, E., Marshall, K. *et al.* (2004), 'Lineage-specific gene duplication and loss in human and great ape evolution', *PLoS Biol.* Vol. 2, p. E207.
14. Cheng, Z., Ventura, M., She, X., Khaitovich, P. *et al.* (2005), 'A genome-wide comparison of recent chimpanzee and human segmental duplications', *Nature* Vol. 437, pp. 88–93.
15. 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.
16. Itan, Y., Bryson, K. and Thomas, M.G. (2010), 'Detecting gene duplications in the human lineage', *Ann. Hum. Genet.* Vol. 74, pp. 555–565.
17. Gonzalez, E., Kulkarni, H., Bolivar, H., Mangano, A. *et al.* (2005), 'The influence of *CCL3L1* gene-containing segmental duplications on HIV-1/AIDS susceptibility', *Science* Vol. 307, pp. 1434–1440.
18. Mamtani, M., Matsubara, T., Shimizu, C., Furukawa, S. *et al.* (2010), 'Association of *CCR2-CCR5* haplotypes and *CCL3L1* copy number with Kawasaki disease, coronary artery lesions, and IVIG responses in Japanese children', *PLoS One* Vol. 5, p. e11458.
19. McKinney, C., Merriman, M.E., Chapman, P.T., Gow, P.J. *et al.* (2008), 'Evidence for an influence of chemokine ligand 3-like 1 (*CCL3L1*) gene copy number on susceptibility to rheumatoid arthritis', *Ann. Rheum. Dis.* Vol. 67, pp. 409–413.
20. Grünhage, F., Nattermann, J., Gressner, O.A., Wasmuth, H.E. *et al.* (2010), 'Lower numbers of the chemokine *CCL3L1* gene in patients with chronic hepatitis C', *J. Hepatol.* Vol. 52, pp. 153–159.
21. Minoretti, P., Arra, M., Emanuele, U., Olivieri, V. *et al.* (2007), 'A W148R mutation in the human *FOXD4* gene segregating with dilated cardiomyopathy, obsessive-compulsive disorder, and suicidality', *Int. J. Mol. Med.* Vol. 19, pp. 369–372.
22. Alter, A., de Léséleuc, L., Van Thuc, N., Thai, V.H. *et al.* (2010), 'Genetic and functional analysis of common *MRC1* exon 7 polymorphisms in leprosy susceptibility', *Hum. Genet.* Vol. 127, pp. 337–348.
23. Jedrzejowska, M., Ryniewicz, B., Kabzińska, D., Drac, H. *et al.* (2008), 'A patient with both Charcot-Marie-Tooth disease (CMT 1A) and mild spinal muscular atrophy (SMA 3)', *Neuromuscul. Disord.* Vol 18, pp. 339–341.
24. Prior, T.W., Krainer, A.R., Hua, Y., Swoboda, K.J. *et al.* (2009), 'A positive modifier of spinal muscular atrophy in the *SMN2* gene', *Am. J. Hum. Genet.* Vol. 85, pp. 408–413.
25. Vezain, M., Saugier-Veber, P., Goina, E., Touraine, R. *et al.* (2010), 'A rare *SMN2* variant in a previously unrecognized composite splicing regulatory element induces exon 7 inclusion and reduces the clinical severity of spinal muscular atrophy', *Hum. Mutat.* Vol. 31, pp. E1110–E1125.
26. Srivastava, S., Mukherjee, M., Panigrahi, I., Shanker Pandey, G. *et al.* (2001), '*SMN2*-deletion in childhood-onset spinal muscular atrophy', *Am. J. Med. Genet.* Vol. 101, pp. 198–202.
27. Hsiao, T.-L. and Vitkup, D. (2008), 'Role of duplicate genes in robustness against deleterious human mutations', *PLoS Genet.* Vol. 4, p. e1000014.
28. Ng, P.C., Levy, S., Huang, J., Stockwell, T.B. *et al.* (2008), 'Genetic variation in an individual human exome', *PLoS Genet.* Vol. 4, p. e1000160.
29. Popesco, M.C., Maclaren, E.J., Hopkins, J., Dumas, L. *et al.* (2006), 'Human lineage-specific amplification, selection, and neuronal expression of DUF1220 domains', *Science* Vol. 313, pp. 1304–1307.
30. Dumas, L., Kim, Y.H., Karimpour-Fard, A., Cox, M. *et al.* (2007), 'Gene copy number variation spanning 60 million years of human and primate evolution', *Genome Res.* Vol. 17, pp. 1266–1277.
31. Armengol, G., Knuutila, S., Lozano, J.J., Madrigal, I. *et al.* (2010), 'Identification of human specific gene duplications relative to other primates by array CGH and quantitative PCR', *Genomics* Vol. 95, pp. 203–209.
32. Sudmant, P.H., Kitzman, J.O., Antonacci, F., Alkan, C. *et al.* (2010), 'Diversity of human copy number variation and multicopy genes', *Science* Vol. 330, pp. 641–646.
33. Kondo, H., Shimomura, I., Kishida, K., Kuriyama, H. *et al.* (2002), 'Human aquaporin adipose (AQPap) gene. Genomic structure, promoter analysis and functional mutation', *Eur. J. Biochem.* Vol. 269, pp. 1814–1826.
34. Rudd, M.F., Webb, E.L., Matakidou, A., Sellick, G.S. *et al.* (2006), 'Variants in the GH-IGF axis confer susceptibility to lung cancer', *Genome Res.* Vol. 16, pp. 693–701.
35. Raux, G., Bonnet-Brilhault, F., Louchart, S., Houy, E. *et al.* (2002), 'The -2 bp deletion in exon 6 of the "alpha 7-like" nicotinic receptor subunit gene is a risk factor for the P50 sensory gating deficit', *Mol. Psychiatry* Vol. 7, pp. 1006–1011.
36. Cravchik, A. and Gejman, P.V. (1999), 'Functional analysis of the human D5 dopamine receptor missense and nonsense variants: differences in dopamine binding affinities', *Pharmacogenetics* Vol. 9, pp. 199–206.
37. van de Winkel, J.G., de Wit, T.P., Ernst, L.K., Capel, P.J. *et al.* (1995), 'Molecular basis for a familial defect in phagocyte expression of IgG receptor I (CD64)', *J. Immunol.* Vol. 154, pp. 2896–2903.
38. Tipney, H.J., Hinsley, T.A., Brass, A., Metcalfe, K. *et al.* (2004), 'Isolation and characterisation of *GTF2IRD2*, a novel fusion gene and member of the TFII-I family of transcription factors, deleted in Williams-Beuren syndrome', *Eur. J. Hum. Genet.* Vol. 12, pp. 551–560.
39. Roy, N., Mahadevan, M.S., McLean, M., Shutler, G. *et al.* (1995), 'The gene for neuronal apoptosis inhibitory protein is partially deleted in individuals with spinal muscular atrophy', *Cell* Vol. 80, pp. 167–178.
40. Roos, D., de Boer, M., Köker, M.Y., Dekker, J. *et al.* (2006), 'Chronic granulomatous disease caused by mutations other than the common GT deletion in *NCF1*, the gene encoding the pp47phox component of the phagocyte NADPH oxidase', *Hum. Mutat.* Vol. 27, pp. 1218–1229.

41. O'Driscoll, M.C., Daly, S.B., Urquhart, J.E., Black, G.C. *et al.* (2010), 'Recessive mutations in the gene encoding the tight junction protein occludin cause band-like calcification with simplified gyration and polymicrogyria', *Am. J. Hum. Genet.* Vol. 87, pp. 354–364.
42. Lynch, M. and Force, A. (2009), 'The probability of duplicate gene preservation by subfunctionalization', *Genetics* Vol. 154, pp. 459–473.
43. Shakhnovich, B.E. and Koonin, E.V. (2006), 'Origins and impact of constraints in evolution of gene families', *Genome Res.* Vol. 16, pp. 1529–1536.
44. Conrad, B. and Antonarakis, S.E. (2007), 'Gene duplication: A drive for phenotypic diversity and cause of human disease', *Annu. Rev. Genomics Hum. Genet.* Vol. 8, pp. 17–35.
45. Han, M.V., Demuth, J.P., McGrath, C.L., Casola, C. *et al.* (2009), 'Adaptive evolution of young gene duplicates in mammals', *Genome Res.* Vol. 19, pp. 859–867.
46. Chimpanzee Sequencing and Analysis Consortium (2005), 'Initial sequence of the chimpanzee genome and comparison with the human genome', *Nature* Vol. 437, pp. 69–87.
47. Genovese, G., Friedman, D.J., Ross, M.D., Lecordier, L. *et al.* (2010), 'Association of trypanolytic ApoL1 variants with kidney disease in African Americans', *Science* Vol. 329, pp. 841–845.
48. Valentonyte, R., Hampe, J., Huse, K., Rosenstiel, P. *et al.* (2005), 'Sarcoidosis is associated with a truncating splice site mutation in *BTNL2*', *Nat. Genet.* Vol. 37, pp. 357–364.
49. Zhou, Q., Rammohan, K., Lin, S., Robinson, N. *et al.* (2003), 'CD24 is a genetic modifier for risk and progression of multiple sclerosis', *Proc. Natl. Acad. Sci USA* Vol. 100, pp. 15041–15046.
50. Sánchez, E., Abelson, A.K., Sabio, J.M., González-Gay, M.A. *et al.* (2007), 'Association of a *CD24* gene polymorphism with susceptibility to systemic lupus erythematosus', *Arthritis Rheum.* Vol. 56, pp. 3080–3086.
51. Li, D., Zheng, L., Jin, L., Zhou, Y. *et al.* (2009), '*CD24* polymorphisms affect risk and progression of chronic hepatitis B virus infection', *Hepatology* Vol. 50, pp. 735–742.
52. Fellay, J., Shianna, K.V., Ge, D., Colombo, S. *et al.* (2007), 'A whole-genome association study of major determinants for host control of HIV-1', *Science* Vol. 317, pp. 944–947.
53. Kirsten, H., Petit-Teixeira, E., Scholz, M., Hasenclever, D. *et al.* (2009), 'Association of *MICA* with rheumatoid arthritis independent of known *HLA-DRB1* risk alleles in a family-based and a case control study', *Arthritis Res. Ther.* Vol. 11, p. R60.
54. Amroun, H., Djoudi, H., Busson, M., Allat, R. *et al.* (2005), 'Early-onset ankylosing spondylitis is associated with a functional *MICA* polymorphism'. *Hum. Immunol.* Vol. 66, pp. 1057–1061.
55. Castermans, D., Wilquet, V., Parthoens, E., Huysmans, C. *et al.* (2003), 'The neurobeachin gene is disrupted by a translocation in a patient with idiopathic autism', *J. Med. Genet.* Vol. 40, pp. 352–356.
56. Jönsson, U.B., Byström, J., Stålenheim, G. and Venge, P. (2002), 'Polymorphism of the eosinophil cationic protein-gene is related to the expression of allergic symptoms', *Clin. Exp. Allergy* Vol. 32, pp. 1092–1095.
57. Eriksson, J., Reimert, C.M., Kabatereine, N.B., Kazibwe, F. *et al.* (2007), 'The 434(G>C) polymorphism within the coding sequence of eosinophil cationic protein (ECP) correlates with the natural course of *Schistosoma mansoni* infection', *Int. J. Parasitol.* Vol. 37, pp. 1359–1366.
58. Selman, M., Lin, H.M., Montaño, M., Jenkins, A.L. *et al.* (2003), 'Surfactant protein A and B genetic variants predispose to idiopathic pulmonary fibrosis', *Hum. Genet.* Vol. 113, pp. 542–550.
59. Malik, S., Greenwood, C.M., Eguale, T., Kifle, A. *et al.* (2006), 'Variants of the *SFTPA1* and *SFTPA2* genes and susceptibility to tuberculosis in Ethiopia', *Hum. Genet.* Vol. 118, pp. 752–759.
60. Albà, M.A. and Castresana, J. (2004), 'Inverse relationship between evolutionary rate and age of mammalian genes', *Mol. Biol. Evol.* Vol. 22, pp. 598–606.
61. Wolf, Y.I., Novichkov, P.S., Karev, G.P., Koonin, E.V. *et al.* (2009), 'The universal distribution of evolutionary rates of genes and distinct characteristics of eukaryotic genes of different apparent ages', *Proc. Natl. Acad. Sci. USA* Vol. 106, pp. 7273–7280.
62. Cai, J.J. and Petrov, D.A. (2010), 'Relaxed purifying selection and possibly high rate of adaptation in primate lineage-specific genes', *Genome Biol. Evol.* Vol. 2, pp. 393–409.
63. De, S., Lopez-Bigas, N. and Teichmann, S.A. (2008), 'Patterns of evolutionary constraints on genes in humans', *BMC Evol. Biol.* Vol. 8, p. 275.
64. Nguyen, D.Q., Webber, C., Hehir-Kwa, J., Pfundt, R. *et al.* (2008), 'Reduced purifying selection prevails over positive selection in human copy number variant evolution', *Genome Res.* Vol. 18, pp. 1711–1723.
65. Kelley, J.L. and Swanson, W.J. (2008), 'Positive selection in the human genome: From genome scans to biological significance', *Annu. Rev. Genomics Hum. Genet.* Vol. 9, pp. 143–160.
66. Lynch, M. and Conery, J.S. (2000), 'The evolutionary fate and consequences of duplicate genes', *Science* Vol. 290, pp. 1151–1155.
67. Chung, W.Y., Albert, R., Albert, I., Nekrutenko, A. *et al.* (2006), 'Rapid and asymmetric divergence of duplicate genes in the human gene coexpression network', *BMC Bioinformatics* Vol. 7, p. 46.
68. Nielsen, R., Hubisz, M.J., Hellmann, I., Torgerson, D. *et al.* (2009), 'Darwinian and demographic forces affecting human protein coding genes', *Genome Res.* Vol. 19, pp. 838–849.
69. Beekman, M., Nederstigt, C., Suchiman, H.E.D., Kremer, D. *et al.* (2010), 'Genome-wide association study (GWAS)-identified disease risk alleles do not compromise human longevity', *Proc. Natl. Acad. Sci. USA* Vol. 107, pp. 18046–18049.
70. Vamathevan, J.J., Hasan, S., Emes, R.D., Amrine-Madsen, H. *et al.* (2008), 'The role of positive selection in determining the molecular cause of species differences in disease', *BMC Evol. Biol.* Vol. 8, p. 273.
71. Lappalainen, T., Salmela, E., Andersen, P.M., Dahlman-Wright, K. *et al.* (2010), 'Genomic landscape of positive natural selection in northern European populations', *Eur. J. Hum. Genet.* Vol. 18, pp. 471–478.
72. Winter, E.E., Goodstadt, L. and Ponting, C.P. (2004), 'Elevated rates of protein secretion, evolution, and disease among tissue-specific genes', *Genome Res.* Vol. 14, pp. 54–61.
73. Zhang, L. and Li, W-H. (2004), 'Mammalian housekeeping genes evolve more slowly than tissue-specific genes', *Mol. Biol. Evol.* Vol. 21, pp. 236–239.
74. Liao, B.Y. and Zhang, J. (2006), 'Low rates of expression profile divergence in highly expressed genes and tissue-specific genes during mammalian evolution', *Mol. Biol. Evol.* Vol. 23, pp. 1119–1128.
75. Hao, L., Ge, X., Wan, H., Hu, S. *et al.* (2010), 'Human functional genetic studies are biased against the medically most relevant primate-specific genes', *BMC Evol. Biol.* Vol. 10, p. 316.
76. Park, S.G. and Choi, S.S. (2010), 'Expression breadth and expression abundance behave differently in correlations with evolutionary rates', *BMC Evol. Biol.* Vol. 10, p. 241.
77. Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R. *et al.* (2009), 'More than just orphans: are taxonomically-restricted genes important in evolution?', *Trends Genet.* Vol. 25, pp. 404–413.
78. Blekhman, R., Oshlack, A., Chabot, A.E., Smyth, G.K. *et al.* (2008), 'Gene regulation in primates evolves under tissue-specific selection pressures', *PLoS Genet.* Vol. 4, p. e1000271.
79. 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.