Gene conversion: mechanisms, evolution and human disease

Jian-Min Chen[*‡], *David N. Cooper*[§], *Nadia Chuzhanova*[||], *Claude Férec*[*‡¶#] *and George P. Patrinos*[**]

Abstract | Gene conversion, one of the two mechanisms of homologous recombination, involves the unidirectional transfer of genetic material from a 'donor' sequence to a highly homologous 'acceptor'. Considerable progress has been made in understanding the molecular mechanisms that underlie gene conversion, its formative role in human genome evolution and its implications for human inherited disease. Here we assess current thinking about how gene conversion occurs, explore the key part it has played in fashioning extant human genes, and carry out a meta-analysis of gene-conversion events that are known to have caused human genetic disease.

Unequal crossover
A recombination event between non-allelic sequences on non-sister chromatids of a pair of homologous chromosomes.

Homologous recombination
The process by which segments of DNA are exchanged between two DNA duplexes that share high sequence similarity.

*INSERM, U613, 29220 Brest, France.
‡Etablissement Français du Sang — Bretagne, 46 rue Félix Le Dantec, 29220 Brest, France.
**Erasmus University Medical Center, Faculty of Medicine and Health Sciences, MGC-Department of Cell Biology and Genetics, PO BOX 2040, 3000 CA, Rotterdam, The Netherlands.
Correspondence to J.M.C. or G.P.P.
e-mails: Jian-Min.Chen@univ-brest.fr; g.patrinos@erasmusmc.nl
doi:10.1038/nrg2193
Published online 11 September 2007

Gene duplication and amplification have contributed to the marked expansion of vertebrate genomes, which has in turn led to the proliferation of regulatory and developmental processes and enzymatic reactions. Following gene duplication, the newly generated gene copies are prone to reciprocal unequal crossover or unidirectional gene-conversion events by virtue of the high degree of homology between them. As a consequence of the latter process, the 'acceptor' sequence is replaced, wholly or partly, by a sequence that is copied from the 'donor', whereas the sequence of the donor remains unaltered. A common phenomenon in fungi, gene conversion has also long been noted in mammalian cells, with the human haemoglobin genes *HBG1* and *HBG2* genes being the first characterized examples[1].

The past few years have witnessed significant progress towards an understanding of the mechanism of gene conversion and its impact on human genome evolution. This has been made possible by advances in several fields: the burgeoning characterization of homologous recombination-associated factors, sperm typing, and the availability of large-scale human population genetic variation data and complete genome sequences from multiple species. Gene conversion has also increasingly been found to underlie human inherited disease. Here we bring together the many recent developments in this rapidly evolving field. Although current models of gene conversion are largely derived from work carried out in model systems (particularly yeast), our discussion, whether in an evolutionary or a pathological context, focuses almost exclusively on human data. In this context, it is important to bear in mind an important caveat: direct evidence for gene conversion in humans is always lacking because it is impossible to analyse both products of a single recombination event in humans.

The mechanistic basis of gene conversion

In eukaryotes, gene conversion constitutes the main form of homologous recombination that is initiated by DNA double-strand breaks (DSBs). During meiosis, DSBs are created by a topoisomerase-like enzyme (SPO11), whereas during mitosis they can be induced by radiation, stalled replication forks or specialized endonucleases (for example, the site-specific HO endonuclease in the switching of yeast mating-type (*MAT*) genes; reviewed in REFS 2,3). Gene conversion mediates the transfer of genetic information from intact homologous sequences to the region that contains the DSB, and it can occur between sister chromatids, homologous chromosomes or homologous sequences on either the same chromatid or different chromosomes.

Current models of gene conversion. Our current understanding of how gene conversion occurs is summarized in FIG. 1. According to the seminal double-strand break repair (DSBR) model of Szostak and colleagues[4], the ends of the DSB are resected by 5′→3′ exonucleases, resulting in the formation of two 3′ ssDNA tails. These tails actively 'scan' the genome for homologous sequences; one of them invades the homologous DNA duplex to form a displacement (D)-loop, which is then extended by DNA synthesis, having been primed from this single-end invasion. The extended D-loop then

Double-strand break
Breaks in opposite DNA strands that lie within ~10–20 bp of each other.

Holliday junction
A point at which the strands of two dsDNA molecules exchange partners, an event that occurs as an intermediate in crossing over or gene conversion.

Mismatch repair
A natural enzymatic process that replaces a mispaired nucleotide within a DNA duplex to yield perfect Watson–Crick base pairing.

Orthologue
A homologous gene that is derived from a speciation event or by vertical descent.

pairs with the other 3′ ssDNA tail (second-end capture), while DNA synthesis at the newly captured strand followed by ligation of the nicks results in the formation of an intermediate with two Holliday junctions (HJs). Random cleavage of the two HJs by an HJ resolvase yields either a non-crossover (that is, gene conversion) or a crossover product (reviewed in REFS 2,3). During this process, most gene conversion is derived from the mismatch repair of the heteroduplex DNA that is formed between the donor and acceptor DNA sequences, rather than from double-strand gap repair, as was originally thought[4]. The mismatch correction probably occurs before the resolution of the double HJs. It is the broken strand that is usually corrected using the intact strand as a template (reviewed in REF. 5).

The DSBR model has provided a satisfactory explanation for several features of meiotic recombination, and HJs have been physically identified as intermediates in meiotic recombination (for example, REF. 6). However, because the DSBR model predicts that an equal number of crossover and non-crossover outcomes are generated, it fails to account for the extremely low occurrence of crossover events (<8%) in mitotic DSB-induced recombination in several model systems[7]. To account for this, the synthesis-dependent strand-annealing (SDSA) model was put forward (FIG. 1): after strand invasion and D-loop extension, the newly synthesized strand is displaced from the template and anneals to the other 3′ ssDNA tail; this is followed by DNA synthesis and ligation of nicks. The SDSA model has two characteristic features: it generally yields only non-crossover products, and all DNA synthesis occurs on the receiving strand[5,7]. In yeast, the SDSA pathway may be promoted by the Srs2 helicase, which is thought to facilitate displacement of the invaded strand and the newly synthesized DNA from the template by removing Rad51 (see BOX 1 for information about the proteins involved in gene conversion) during the strand-exchange process[8,9], a view that is supported by many studies[10–13]. In addition, the *Drosophila melanogaster* orthologue of the Bloom syndrome helicase (BLM), a member of the highly conserved RecQ family, promotes SDSA by unwinding a D-loop intermediate[14–16]; in support of this, BLM has also been shown to efficiently displace the invading strand of mobile D-loops[17].

Rad54 has a novel DNA branch-migration activity[18], and might therefore be able to promote branch migration of DNA junctions at the D-loop in either direction, away from or towards the DSB. The former would stabilize the D-loop, thereby facilitating the capture of the second 3′ ssDNA tail and subsequent formation of double HJs; the latter would destabilize the D-loop, thereby facilitating the displacement of the invaded strand and the newly synthesized DNA from the template[18].

Another mechanism, known as double-HJ dissolution[8,19], shares the two defining characteristic features of SDSA in terms of outcome. However, it generates its non-crossover product from the convergent migration of the two HJs towards each other, leading to the collapse of the double HJs (FIG. 1). This mechanism is promoted by the coordinated action of BLM and topoisomerase IIIα (by BLM in humans[19], Sgs1 in yeast[8,9,20] and the fly homologue of BLM in *D. melanogaster*[21,22]). More recently, the BLM-associated protein BLAP75 (also known as RMI1) has been identified as a third component of the double HJ 'dissolvasome'[23,24].

At least in yeast, the control of crossover and non-crossover recombination is differentially timed during meiosis[25,26]. It would appear that "…once the DSB-repair reaction has committed to a resolution pathway that involves double Holliday junctions, the intermediate is already destined to be resolved in a specific orientation that leads to crossover."[27] This, when considered together with the double-HJ dissolution and SDSA models, may pose a serious challenge to the DSBR model: the non-crossover events that are presumed to result from random resolution of double HJs (FIG. 1) might in fact come from either the HJ dissolution or SDSA pathways.

Our current thinking points to the existence of several mutually exclusive pathways leading to the occurrence of gene conversion (FIG. 1). It is apparent that these pathways are finely regulated, as evidenced by the observations that SDSA can be promoted by several different proteins (that is, Srs2, BLM and Rad54), the same protein can act in different pathways (that is, BLM in both SDSA and double-HJ dissolution) and the same protein (that is, Rad54) can promote bidirectional branch migration of HJs, thereby facilitating either SDSA or the formation of double HJs.

Characteristics of gene conversion. Efficient gene conversion generally requires homology between interacting sequences. Our meta-analysis of the pathogenic gene-conversion events (see below) revealed that, with one exception (88% between the folate receptor gene *FOLR1* and its pseudogene), the homology between the interacting sequences is always >92% and usually >95% (data not shown). In addition, the frequency of gene conversion is inversely proportional to the distance between the interacting sequences *in cis* (for example, REF. 28), and a ~15-fold higher frequency has been observed between linked as opposed to unlinked gene pairs in mice and rats[29]. Furthermore, the rate of gene conversion is directly proportional to the length of the uninterrupted-sequence tract in the putatively converted region: the minimal efficient processing segment (MEPS) for efficient meiotic homologous recombination in mouse cells is >200 bp (REFS 30,31), whereas in humans it is estimated to be in the range of 337–456 bp (REF. 32).

Author addresses

[§]Institute of Medical Genetics, Cardiff University, Heath Park, Cardiff, CF14 4XN, UK.
[‖]Department of Biological Sciences, University of Central Lancashire, Preston PR1 2HE, UK.
[¶]Faculté de Médecine de Brest et des Sciences de la Santé, Université de Bretagne Occidentale, 29238 Brest, France.
[#]Laboratoire de Génétique Moléculaire et d'Histocompatibilité, Centre Hospitalier Universitaire de Brest, Hôpital Morvan, 29220 Brest, France.

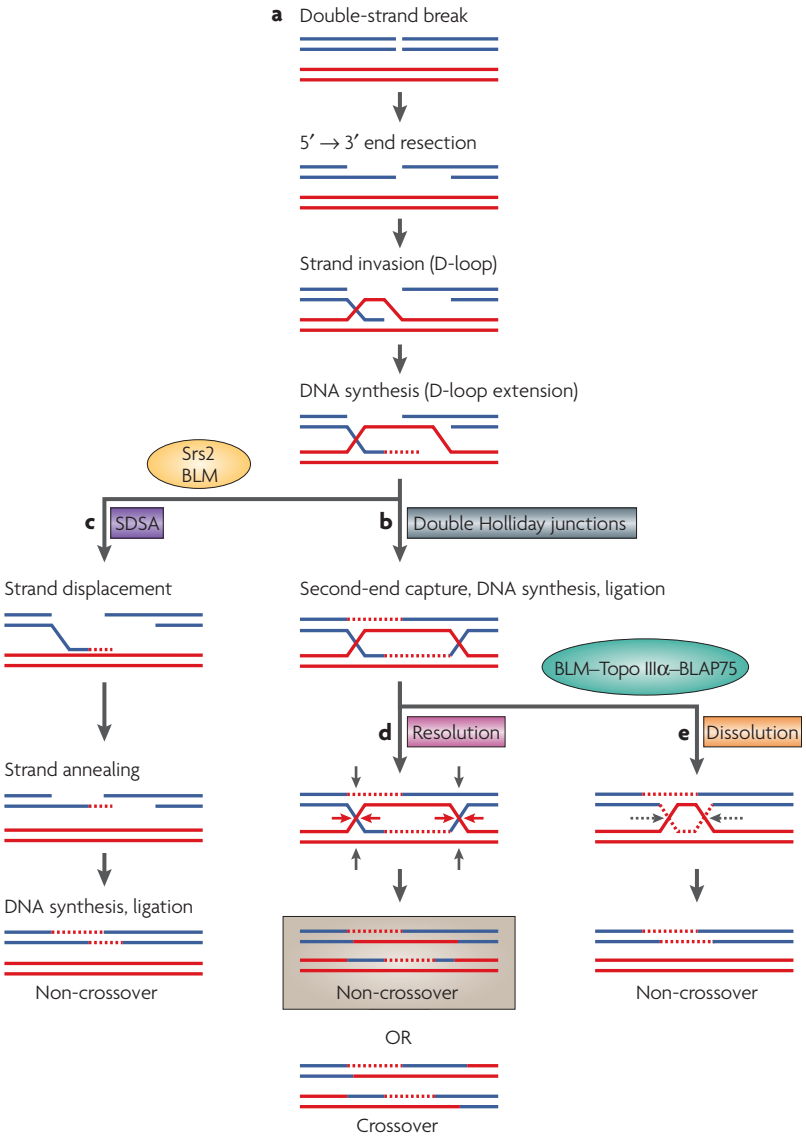


Figure 1 | **Mechanisms of gene conversion.** The double-strand break repair (DSBR; **a**–**b**–**d**), synthesis-dependent strand-annealing (SDSA; **a**–**c**; refer to REF. 5 for other possible SDSA pathways) and double Holliday junction (HJ) dissolution (**a**–**b**–**e**) models are illustrated. All models share a common initiating step: the 5′ ends of the double-strand break (DSB) are resected to form 3′ ssDNA tails; the tails actively 'scan' the genome for homologous sequences, and one tail invades the homologous DNA duplex forming a displacement (D)-loop, which is then extended by DNA synthesis. SDSA diverges from the other two pathways after D-loop extension: the invading strand and the newly synthesized DNA are displaced from the template (through the action of the Srs2 helicase in yeast, or the Bloom syndrome protein (BLM) in humans and flies) and anneal to the other 3′ end of the DSB, leading to the formation of only gene-conversion events. Otherwise, the other 3′ end of the DSB is captured, and DNA synthesis and ligation of nicks lead to the formation of double HJs. According to the dissolution model, BLM, topoisomerase IIIα (Topo IIIα) and the BLM-associated protein BLAP75 (also known as RMI1) act together to remove the double HJs via convergent branch migration (indicated by dotted arrows at both HJs) leading exclusively to gene conversion. In DSBR, the resolution of the double HJs by an HJ resolvase is predicted to generate an equal number of non-crossover (indicated by red arrows at both HJs) and crossover (indicated by black arrows at one HJ and red arrows at the other HJ) events. Identification of the distinct SDSA and double-HJ dissolution pathways that result in gene conversion challenges the concept of DSBR: the non-crossover events in DSBR (highlighted within the coloured box) might in fact come either from SDSA or double-HJ dissolution. See BOX 1 for more information on the proteins that are involved in gene conversion.

In yeast, mitotic gene-conversion tracts (often larger than 4 kb) are generally much longer than meiotic ones (typically 1–2 kb; for example, REF. 33). In mammals, gene-conversion tracts are usually short, of the order of 200 bp to 1 kb in length. For example, estimates range from 113–2,266 bp for the human globin genes[34], to 1–1,365 bp for two Yq-located human endogenous retroviral (HERV) sequences[35], to 54–132 bp in single-sperm analysis of the human leukocyte antigen *HLA-DPB1* locus[36] and 55–290 bp in various gene-conversion hotspots[37]. Our own review of pathological gene-conversion events has revealed that the lengths of the maximally converted tracts rarely exceed 1 kb (TABLE 1; see Supplementary Information S1 (box) for details of the meta-analysis). In practice, the length of the converted tract has been used to distinguish a gene-conversion event from a double-crossover event (BOX 2).

It has long been assumed that specific motifs surrounding DNA sequences involved in gene conversion can either promote or inhibit gene-conversion and branch-migration events (for example, REF. 34). Such motifs include alternating purine and pyrimidine or polypurine and polypyrimidine tracts, palindrome-like sequences, minisatellite sequences, *chi* (χ)-like sequences and sequences that can adopt tetraplex, Z-DNA, slipped and triplex structures, although many examples of gene-conversion events are not associated with any of the above sequence motifs[35]. Our analysis of gene-conversion tracts that have been reported in the context of human disease reveals that short (5 bp) and long (≥10 bp) polypurine and polypyrimidine tracts are over-represented at the 1% level of significance within the minimal and maximal converted tracts, in comparison with a simulated data set that matched the corresponding original minimal and maximal converted tracts in terms of their length and nucleotide composition (N.C., unpublished observations). Several sequences that can form tetraplex, triplex, Z-DNA and slipped structures were overrepresented ($p \leq 0.01$) in the regions flanking the minimal but not the maximal converted tracts (FIG. 2d). Bearing in mind the spatial coincidence of non-B DNA structures with deletion breakpoints[38], it is tempting to speculate that the DSBs that initiated gene conversion occurred either within or very close to the regions flanking the minimal converted tracts. Neither the χ nor the human minisatellite or χ-like elements were overrepresented, even at the 5% level of significance, within converted tracts (N.C., unpublished observations).

Gene-conversion events can be non-allelic (also known as interlocus) or interallelic (FIG. 2a–c). Non-allelic gene conversion often shows biased directionality. For example, whereas proximal-to-distal (relative to the centromere) gene conversion between two directly repeated HERV elements on the long arm of the human Y chromosome occurs at a rate of between 2.4×10^{-4} and 1.2×10^{-3} events per generation, the rate of distal-to-proximal gene conversion is some 20-fold lower[35]. In some cases (for example, in human globin genes[34]), the directionality of gene conversion correlates with the relative levels of expression of the participating genes,

Box 1 | **Proteins involved in gene conversion**

Gene conversion is initiated by the processing of DNA ends to yield molecules with 3′ single-stranded tails that are subsequently coated with replication protein A (RPA), an ssDNA binding protein. Rad51, which shares sequence similarity with the *Escherichia coli* RecA strand-exchange protein[105], has a key role in recombination. The Rad51 paralogues (XRCC2, XRCC3, Rad51B, Rad51C and Rad51D) are present in two distinct complexes in which Rad51C is the common component. The Rad51B–Rad51C–Rad51D–XRCC2 complex binds ssDNA, single-stranded gaps in dsDNA and nicks in duplex DNA[106]; the Rad51C–XRCC3 complex also has a DNA binding activity and associates with a Holliday junction resolvase[107].

Other key components of the gene-conversion machinery are Rad52 and the dsDNA-dependent ATPase Rad54. Both proteins are accessory partners of Rad51, but they operate at different stages; Rad52 functions early on to assist loading of Rad51 onto resected DNA ends, whereas Rad54 functions later when the intact double-stranded template has become engaged in the gene-conversion process[108].

The meiotic recombination protein 11 (Mre11)–Rad50–Nijmegen breakage syndrome protein 1 (Nbs1) complex, or MRN complex, is a central player in homologous recombination. The existence of Mre11 nuclease and the Rad50 ATPase homologues in different organisms (from yeast to humans) suggests that this complex is fundamental for genomic stability. An essential biological function of the complex in the context of double-strand break (DSB) repair is to tether DNA ends through a DNA-driven conformational change[109,110].

The breast cancer susceptibility proteins BRCA1 and BRCA2 also participate in gene conversion. BRCA1 interacts with Rad51 and the MRN complex[111], whereas BRCA2 interacts with both Rad51 (REF. 112) and BRCA1 (REF. 113). Biochemical evidence implies that BRCA2 is likely to act preferentially at the interface between dsDNA and ssDNA to displace RPA from the overhang and assist with loading of Rad51. Following DNA damage and initial DSB processing, BRCA2 is required for the formation of Rad51 foci[114]. Recent data suggest that the C-terminal region of BRCA2 specifically interacts with multimeric Rad51 and binds an intersubunit interface of Rad51 that is present in the Rad51 multimer and in the nucleoprotein filament[115]. This interaction is crucial for the protection of Rad51–DNA filaments from disassembly during recombination-mediated DNA repair[116].

Gene-conversion tract
In theory, this is the portion of the 'acceptor' sequence that is copied from the 'donor'. Because in practice the length of the tract cannot be precisely known, it must be expressed in terms of the lengths of the minimal and maximal converted tracts: the former refers to the entire region spanned by converted discriminant nucleotides and the latter refers to the region that is delimited by the two nearest unconverted discriminant nucleotides between the donor and acceptor sequences.

Double crossover
Two crossovers that occur in a chromosomal region between highly homologous genes, resulting in reciprocal sequence exchange between them.

Chi (χ) sequence
An 8-bp sequence (5′-GCTGGTGG-3′) that acts as a recombination hotspot in *Escherichia coli*.

with the gene that is expressed at a higher level (the 'master' gene) converting the gene that has the lower expression level (the 'slave' gene).

The master and slave gene rule should be interpreted with caution because it was originally formulated in an evolutionary context. Indeed, increased transcription of a gene seems to increase the likelihood of its use not only as a donor (for example, REF. 39), but also as an acceptor (for example, REF. 40). Moreover, multiple pseudogene-mediated conversion events that self-evidently do not follow this rule have been shown to cause human inherited disease (TABLE 1). Thus, although it seems clear that gene conversion can occur in the opposite direction (from slave to master), such events are much less likely to be observed in the human genome because they would probably generate defective alleles that would tend to become extinct.

Evolutionary consequences of gene conversion

Since the first reported characterization of gene-conversion events in the human globin genes[1], interlocus gene conversion has been implicated in the concerted evolution of many human gene families including the Rh blood group antigen genes *RHD* and *RHCE*[41], gonadotropin hormone β-subunit gene (*CGB*)[42], red (*OPN1LW*) and green (*OPN1MW*) opsin genes[43], olfactory receptor genes[44], α-interferon gene[45], γ-crystallin gene[46] and the chemokine receptor genes *CCR2* and *CCR5* (REF. 47). The key role of gene conversion in maintaining the high level of sequence homogeneity between the tandemly repeated ribosomal RNA (rRNA) genes in eukaryotes has been reviewed elsewhere[48]. Below we discuss gene conversion in the context of some well-studied examples. We then discuss how recent population genetic studies have provided new insights into how gene conversion is shaping fine-scale structural variation in the human genome, and how these findings promise to improve the design of disease association studies.

The role of interlocus gene conversion in concerted evolution. Interlocus gene conversion has an important role in the concerted evolution of multigene families and highly repeated DNA sequences (see also BOX 3). A hallmark of its action is that paralogous gene sequences become more closely related to each other than they are to their orthologous counterparts in a closely related species[49]. The dramatic impact of this process is perhaps best exemplified by the multi-copy testis-expressed gene families that lie within the eight large palindromes that comprise 25% of the euchromatic portion of the male-specific region (MSY) on the human Y chromosome[50]. All of the known genes within these palindromes have an identical or near-identical copy (with a typical sequence similarity of 99.97%) on the opposite arms of the palindrome. Such a high degree of similarity would normally be indicative of a recent duplication event; however, human–chimpanzee sequence comparison revealed that the MSY palindromes pre-dated the split of these two lineages (~5 million years ago)[50]. Furthermore, analysis of single-nucleotide differences between extant MSY palindromes provided evidence of recurrent arm-to-arm gene conversion in humans: as many as 600 bp (from the 5.4 Mb contained within MSY palindromes) in each newborn male have been estimated to undergo Y–Y gene conversion[50].

The arm-to-arm sequence divergence values that have been calculated for the human and chimpanzee palindromes are 0.028% and 0.021%, respectively. By contrast, an average sequence divergence of 1.44% was observed between the orthologous palindrome arms in the two species[50]. Interlocus gene conversion has homogenized the paralogous palindrome sequences within each species while diversifying the orthologous palindrome sequences between the two species. So, how precisely does gene conversion influence sequence similarity or divergence between paralogues and orthologues? On the basis of simulations involving two paralogous HERV15 proviral sequences that flank the Y-chromosomal azoospermia factor A (*AZFA*, also known as *USP9Y*) locus in three primate species (that is, humans, chimpanzees and gorillas)[51], it seems that the higher the conversion rate, the higher the degree of homogenization between paralogous sequences and the greater the extent of the divergence between orthologous sequences. Moreover, the higher the degree of sequence divergence before speciation, the greater the subsequent divergence between both the paralogous and the orthologous sequences. Finally, although the biased directionality of interlocus gene conversion does

Table 1 | **Interlocus gene-conversion events that cause human inherited disease* (part 1)**

Disease/phenotype	Donor gene‡	Acceptor gene	Chromosomal localization	Direction-ality§	Mutation	Converted tract length (bp)‖	Refs
Atypical haemolytic uraemic syndrome	*CFHR1*§§	*CFH*	1q32	3′>5′	c.[3572C>T;3590T>C]¶	19–331	131
Congenital adrenal hyperplasia	*CYP21A1P*	*CYP21A2*	6p21.3	5′>3′	[-209T>C;-198C>T;-189/-188insT]	21–155	132
					[-4C>T;92C>T;118T>C;138A>C]	142–523	133
					[1380T>A;1383T>A;1389T>A; IVS6+12_13AC>GT]	42–210	134
					[1688G>T;1767_1768insT]¶	80–202	133
Syndrome of corticosterone methyloxidase II deficiency	*CYP11B1*§§	*CYP11B2*	8q21–q22	3′>5′	Conversion of exons 3 and 4	446–626	135
Increased 18-hydroxycortisol production	*CYP11B1*§§	*CYP11B2*	8q21–q22	3′>5′	Conversion of two nucleotides separated by two bases in exon 8	4–56	136
Autosomal dominant cataract	*CRYBP1*	*CRYBB2*	22q11.2–q12.1	3′>5′	c.[475C>T;483C>T]	9–104	137
Neural tube defects	*FOLR1P*	*FOLR1*	11q13.3–q14.1	5′>3′	7497_7662 including 13 discriminant nucleotides¶	166–215	138
					[7539T>C;7541G>A]¶	3–65	138
Gaucher disease	*GBAP*	*GBA*	1q21	3′>5′	[del55bp;D409H;L444P;A456P; V460V]	604–974	139
					[D409H;L444P;A456P;V460V]	525–848	140;141
					[L444P;A456P]	36–175	139
					[L444P;A456P;V460V]	50–475	141;142
Short stature	*GH2*§§	*GH1*	17q22–q24	3′>5′	Conversion involving 12 discriminant nucleotides in the promoter region	40–218	143
Mild microcytosis	*HBB*§§	*HBD*	11p15.5	3′>5′	Conversion involving exons 1 and 2	≥212–≤348#	144
Hereditary persistence of fetal haemoglobin	*HBG2*§§	*HBG1*	11p15.5	5′>3′	Conversion involving 9 discriminant nucleotides in the promoter region¶	423–1554	145
Agammaglobulinaemia	*IGLL3*	*IGLL1*	22q11.23	Inverted**	c.[393T>C;420T>C;425C>T]	33–152	146
Chronic granulomatous disease	*NCF1B* or *NCF1C*	*NCF1*	7q11.23	5′ or inverted**	[C>T;ΔGT]	124–1474	147
					[C>T;ΔGT;ins20bp]	369–1529	147
					[ΔGT;ins20bp]	247–423	148
Blue cone monochromacy	*OPN1MW*§§	*OPN1LW*	Xq28	3′>5′	3′ limit of the maximal converted region cannot be defined owing to the lack of information on an intervening discriminant nucleotide	≥636–≤3676#	149
Autosomal dominant polycystic kidney disease	?‡‡	*PKD1*	16p13.3	?‡‡	[8446T>G;8490T>C;8493G>C; 8502T>C]	57–126	150
					[8446T>G;8490T>C;8493G>C; 8498C>G;8502T>C]	57–126	150
					[8639G>T;8651G>A;8658T>C; 8662C>T]	24–230	151
Chronic pancreatitis	*PRSS2*§§	*PRSS1*	7q35	3′>5′	Conversion involving 22 discriminant nucleotides	289–457	88

* Only interlocus events that were informative with respect to the converted tracts, and that comprised at least two neighbouring but non-consecutive markers, were collected. Some mutations were re-named or re-interpretated. In mutation nomenclature, 'c.' refers to coding nucleotides. The full selection criteria used in the collation of these mutations and the details of how they were annotated are provided in the Supplementary Information S1 (box) online. See the main text for pathogenic interallelic gene-conversion events and gene-conversion mutations in cancer. ‡Functional donor genes are indicated by §§. The solitary case of a 'partially functional' donor gene, *SMN2*, is indicated by ‖‖. Other donor genes are pseudogenes. §The directionality of sequence transfer from the donor gene to the acceptor gene (in the context of the sense strand) is specified whenever the donor and acceptor pairs represent tandem duplications. ‖Limits to the length of the converted tract, minimal to maximal (see also FIG. 2d). ¶*De novo* mutations. #Lengths of the minimal and maximal converted tracts cannot be unequivocally assigned owing to the lack of information on certain internal marker(s); see Supplementary Information S1 (box) for details. **Gene-conversion events between gene copies that are not tandem duplications. ‡‡Donors cannot be assigned to current human genome sequence assemblies, because of either copy-number variation or the presence of a 'gap' (there are at least six *PKD1* pseudogenes at 16p13).

Table 1 | **Interlocus gene-conversion events that cause human inherited disease* (part 2)**

Disease/ phenotype	Donor gene‡	Acceptor gene	Chromosomal localization	Direction-ality§	Mutation	Converted tract length (bp)‖	Refs
Shwachman–Bodian–Diamond syndrome	*SBDSP*	*SBDS*	7q11.22	Inverted**	c.[129-443A>G;129-433G>A]	11–104	152
					c.[141C>T;183_184TA>CT]	44–217	152
					c.[141C>T;183_184TA>CT;201A>G]	61–276	153
					c.[141C>T;183_184TA>CT;201A>G;258+2T>C]¶	120–398	152
					c.[183_184TA>CT;201A>G]	19–118	153
					c.[183_184TA>CT;201A>G;258+2T>C]	78–240	154
					c.[201A>G;258+2T>C]	60–197	152
Spinal muscular atrophy	*SMN2*‖‖	*SMN1*	5q13.2	5′>3′	Conversion of exon 7 and intron 7	264–776	155
von Willebrand disease	*VWFP*	*VWF*	22q11.22–q11.23 (*VWFP*)/ 12p13.3 (*VWF*)	Not applicable	[IVS27-45C>T;IVS27-36C>T;3686T>G; 3692A>C]	63–131	156
					c.[3686T>G;3692A>C]	7–95	157
					c.[3686T>G;3692A>C;3735G>A;3789G>A; 3797C>T]	112–195	90
					c.[3789G>A;3797C>T]	9–99	158;159
					c.[3789G>A;3797C>T;3835G>A]	47–195	157
					c.[3789G>A;3797C>T;3835G>A;3931C>T; 3951C>T]	163–291	90
					c.[3835G>A;3931C>T;3951C>T]	117–229	90
					c.[3835G>A;3931C>T;3951C>T;4027A>G; 4079T>C;4105T>A]	271–335	90
					c.[3931C>T;3951C>T]	21–191	90
					c.[3931C>T;3951C>T;4027A>G;4079T>C; 4105T>A]	175–297	160

* Only interlocus events that were informative with respect to the converted tracts, and that comprised at least two neighbouring but non-consecutive markers, were collected. Some mutations were re-named or re-interpretated. In mutation nomenclature, 'c.' refers to coding nucleotides. The full selection criteria used in the collation of these mutations and the details of how they were annotated are provided in the Supplementary Information S1 (box) online. See the main text for pathogenic interallelic gene-conversion events and gene-conversion mutations in cancer. ‡Functional donor genes are indicated by §§. The solitary case of a 'partially functional' donor gene, *SMN2*, is indicated by ‖‖. Other donor genes are pseudogenes. §The directionality of sequence transfer from the donor gene to the acceptor gene (in the context of the sense strand) is specified whenever the donor and acceptor pairs represent tandem duplications. ‖Limits to the length of the converted tract, minimal to maximal (see also FIG. 2d). ¶*De novo* mutations. **Gene-conversion events between gene copies that are not tandem duplications.

Z-DNA
One of several possible double-helical structures of DNA. Z-DNA is a left-handed double-helical structure in which the double helix winds to the left in a zig-zag pattern, rather than to the right, as occurs in the more common B-DNA form.

Concerted evolution
The process by which repetitive DNA sequences are homogenized such that the individual members of a given DNA repeat or multigene family in one species come to show a higher degree of sequence identity with each other than they do with members of the same DNA repeat or multigene family in another species.

not affect the degree to which paralogues are homogenized, it does appear to cause smaller perturbations of orthologous sequence divergence in the preferred donor than in the preferred acceptor gene[51].

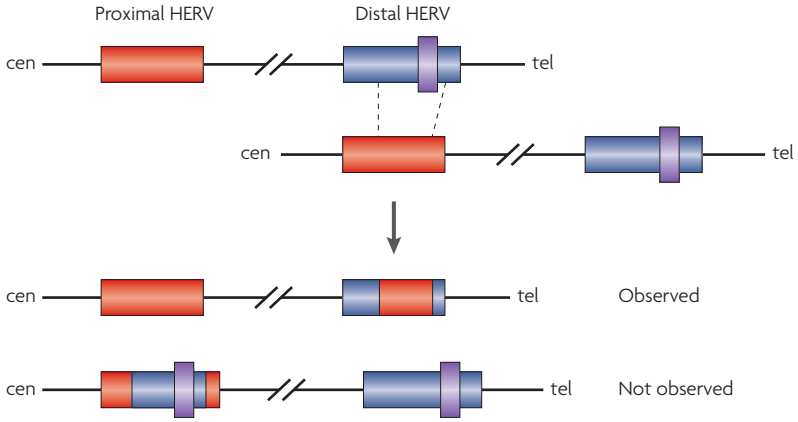
The above notwithstanding, the impact of interlocus gene conversion on the evolution of multigene families has certainly been affected by other factors, most notably selection. As an obvious consequence of this interaction, interlocus gene conversion has facilitated the spread of advantageous mutations. For example, most of the coding polymorphisms in human *OPN1LW* and *OPN1MW* genes, which are tandemly arranged on the X chromosome, lie within exon 3 (7 out of 11 polymorphisms in *OPN1LW* and 6 out of 8 polymorphisms in *OPN1MW*)[52]. That 6 out of 7 of the exon 3 polymorphisms are shared between the two genes is indicative of the magnitude of the imprint of interlocus gene conversion[52]. Human *OPN1LW* replacement SNPs serve to shift λ_{max} into the 'red–orange' portion of the visual spectrum[53]. This and other observations[43] suggest that polymorphisms in exon 3 of *OPN1LW* might enhance colour discrimination.

The spread of an interlocus gene-conversion-derived advantageous mutation through the population might be accelerated by interallelic gene conversion. This view is supported by a recent study that traced the evolution of *HLA-E* in eight primate and rodent species[54].

Interallelic gene conversion generates allelic diversity. Some regions of the human genome are remarkably polymorphic. For example, the human *ABO* blood group locus, apart from the heterogeneity that underlies its three major alleles (*A*, *B* and *O*), contains extensive heterogeneity within the various alleles that constitute the ABO subgroups. Haplotype analysis implies that some of these alleles might be generated by gene conversion between the parental alleles (reviewed in REF. 55).

The best known example of interallelic gene conversion is probably the HLA class II region containing *HLA-DR*, *HLA-DQ* and *HLA-DP*; more than 500 *HLA-DRB1* alleles have been reported in the human population (see the International Immunogenetics Project HLA database). Most of the polymorphisms lie in exon 2, which encodes part of the antigen-recognition

Box 2 | **Example of a double-crossover event**



In humans, gene conversion can never be formally and unambiguously distinguished from double crossover because only one of the products of recombination can be observed. In practice, double crossover is commonly assumed to have occurred when the observed sequence change is considered to be too long (for example, >3 kb) to have been caused by gene conversion. This is exemplified by the sequence exchange between the two ~9-kb human endogenous retroviral (HERV) repeats (shown in red and blue) on human chromosome Yq, in which a ~5-kb stretch, which includes a 1.5-kb insertion (shown as a purple bar) that is present in only the distal repeat copy, was presumed to result from a double crossover between misaligned chromatids (or chromosomes in the case of repeats that are located on autosomes or X chromosomes), rather than from a gene-conversion event[76]. cen, centromere; tel, telomere.

site of class II proteins. This high degree of polymorphism has been under positive selection in response to the ongoing immunological challenge from diverse pathogens. Direct evidence for interallelic gene conversion has been provided by sperm analysis of the *HLA-DPB1* locus[36]; furthermore, a recent large-scale population genetic study has suggested that novel alleles in the *HLA* locus are being continually generated by gene conversion[56].

Gene conversion and human genome variation. On the basis of the realization that interlocus gene conversion can generate allelic diversity, Hurles[57] challenged the then prevailing assumption that "…there is no reason to expect that polymorphic variation is increased within duplicated regions."[58] He further speculated that gene conversion could have contributed to the formation of the much greater than expected number (~100,000) of SNPs found within segmental duplications in the human genome[58]. Indeed, many duplicon SNPs have recently been shown to be true SNPs (that is, allelic variants) with only a few of them being paralogous sequence variants[59]. In the meantime, the role of gene conversion in the evolution of segmental duplications has been increasingly recognized, both at specific loci (for example, REFS 35,42,50,60) and at the genomic level[61]. In particular, analysis of 24 human duplicon families that together span >8 Mb of DNA (representing some 5% of all known human duplicons[62]) indicates that most of the segmental duplications contain a significant excess of sites that show signatures of concerted evolution, as compared with the number that would be expected as a result of nucleotide substitution alone[63].

The impact of gene conversion on the human genome has not been limited to duplicons. There is growing evidence that gene conversion has had an important role in shaping fine-scale patterns of linkage disequilibrium (LD) in the human genome[64,65]. For example, nearly all pairs of loci that are separated by 124 bp on average would be expected to be in complete LD because recombination does not usually occur over such a short interval. The fact that a significant fraction of locus pairs showed only partial LD suggests that gene conversion has increased the apparent rate of recombination between nearby loci[64].

The public database established by the International HapMap Project[66], which comprises >3 million SNPs obtained from four different human populations, was originally intended to be a resource of tag SNPs for genome-wide association mapping studies in any human population[67,68]. However, for technical reasons, duplicated regions in the human genome (which constitute at least 5% of the genome) are generally poorly covered by the HapMap. Given the widespread presence of gene conversion and the short length of the conversion tracts that are usually involved in gene conversion, LD between gene-conversion-generated SNPs and the tagging markers may often be weak, thereby reducing the efficacy of LD-based association studies[69]. Indeed, it is anticipated that the now widely used tagging approach will be replaced by more economical whole-genome sequencing in the near future[70].

High gene-conversion activity is a common feature of both allelic[37] and non-allelic recombination hotspots[71,72] (at least 25,000 recombination hotspots have been identified across the human genome[73]). A better understanding of the role of gene conversion in generating or eliminating recombination hotspots in the human genome[71,72,74] should improve our ability to predict the location of unstable genomic regions[75]. Finally, gene-conversion-mediated sequence homogenization of duplicons increases the likelihood of non-allelic homologous recombination between these sequences, potentially leading to genomic disorders[76,77].

The ratio of gene conversion to crossover. Several studies have attempted to estimate the average rate of gene conversion, *f*, expressed as its ratio to crossover (see REF. 78 for a recent review), using genome-wide population genetic variation data. In a study of 21,840 biallelic SNPs on 20 independent copies of human chromosome 21, *f* was estimated to be 1.6 assuming a conversion-tract length of 500 bp, and 9.4 assuming a mean conversion-tract length of 50 bp (REF. 79). In a broadly similar study, *f* was calculated to be 7.3 for a mean tract length of 500 bp (REF. 80). Another study estimated an *f* value of between 3 and 10, with a ratio of 6 providing the best fit[64]. However, the estimated *f* is only around 0.3 for Europeans and 1.0 for African Americans[65]. Setting aside the influence of the assumptions about the length of the converted tract, the wide variation in *f* between

Paralogue
One of a set of homologous genes in the same species that have evolved from a gene duplication, and that can be associated with a subsequent divergence of function.

Segmental duplication
A segment of DNA of larger than 1 kb that occurs in two or more copies per haploid genome, with the different copies sharing >90% sequence identity.

Linkage disequilibrium
A statistical association between particular alleles at two or more neighbouring loci on the same chromosome that results from a specific ancestral haplotype being common in the population under study.

Tag SNPs
SNPs that are correlated with, and can therefore serve as a proxy for, much of the known remaining common variation in a region.

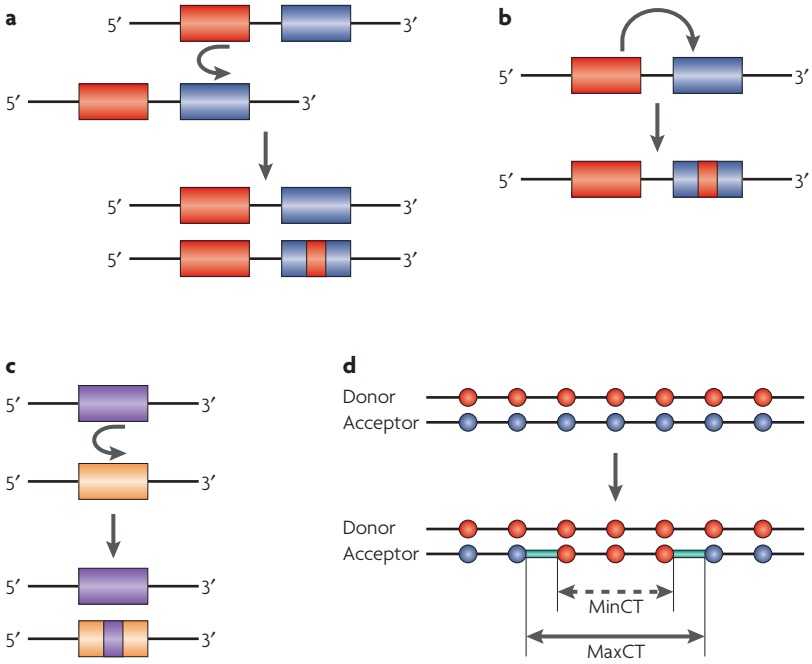


Figure 2 | **Types of gene conversion and demarcation of the converted tract.** **a** | Non-allelic (or interlocus) gene conversion *in trans*, shown as an event occurring between paralogous sequences (represented as red and blue boxes) that reside on sister chromatids or on homologous chromosomes. Gene-conversion events that occur between homologous sequences that reside on different chromosomes are not shown. **b** | Non-allelic gene-conversion events *in cis* (between non-allelic gene copies that reside on the same chromosome). Gene-conversion events, which are depicted in **a** and **b**, are virtually indistinguishable from each other. **c** | Interallelic gene-conversion events occurring between alleles (shown in purple and orange) that reside on homologous chromosomes. **d** | Maximal and minimal converted tracts of a given gene-conversion event. Although the length of the minimal converted tract (MinCT) is usually shorter than the true tract, the length of the maximal converted tract (MaxCT) is usually longer than the true tract. The initiating and terminating points of gene conversion can lie anywhere within the two regions that are marked in green.

the studies might reflect the different populations and genomic regions that are under investigation. Indeed, sperm-typing experiments show that *f* is between 4 and 15 in the *DNA3* hotspot[37], ~0.3 in the nidogen 1 (*NID1*) hotspot[81] and <0.1 in the β-globin gene (*HBB*) region[82] (reviewed in REF. 78).

Biased gene conversion. Gene conversion seems to favour some alleles over others, a process known as biased gene conversion (BGC). BGC arises as a consequence of the GC-biased repair of A:C and G:T mismatches that are formed in heteroduplex recombination intermediates during meiosis[83]. Increasing the probability of the fixation of G and C alleles leads to the GC-enrichment of the sequences involved. BGC has therefore exerted an important influence on GC content during the evolution of mammalian genomes. Indeed, members of multigene families that are closely linked and highly homologous to each other (and hence presumably undergo gene conversion) have a significantly higher GC content than gene singletons (which are presumed to undergo less gene conversion than members of gene families)[84]. Furthermore, chromosomal regions that exhibit a high rate of recombination become enriched in GC[85]. Thus, G and C allele frequencies tend to be higher at sites of SNPs that lie in the vicinity of recombination hotspots[86]. Nevertheless, using a context-dependent mutation model, Hernandez and colleagues[87] recently concluded that much of the evidence from human population genetic data that suggests a recent fixation bias favouring GC-content could have been compromised by the misidentification of the ancestral state of each SNP.

Gene conversion and disease

Gene-conversion events, predominantly of the interlocus variety (TABLE 1), have been implicated as the molecular cause of an increasing number of human inherited diseases. Although the imprint of gene conversion is not invariably unambiguous, some important features, such as a high degree of homology (in the range of 92–99%) between the sequences that are presumed to be involved and the substitution of at least two neighbouring but non-consecutive markers within a short sequence tract, can point to gene conversion rather than single-nucleotide substitution or double crossover as the underlying mutational mechanism.

Nearly 50% of the donor genes are functional or partially functional. Pathogenic gene conversion is often viewed as having resulted almost exclusively from the transfer of genetic information from non-functional pseudogenes to their closely related functional counterparts (see also BOX 4). However, our meta-analysis reveals that, of the 17 acceptor genes known to have been involved in pathogenic interlocus gene-conversion events, 7 had a functional donor counterpart; an additional gene, survival of motor neuron 1 (*SMN1*), had a partially functional counterpart (TABLE 1). Acceptor genes with donors that were functional usually carried only a single mutation, whereas those with donor sequences that were non-functional carried multiple mutations (TABLE 1). This might be at least partially explained by the fact that non-functional genes generally present a more extensive pathogenic template than functional genes, simply because non-functional genes tend to carry multiple types of gene-disrupting mutation (for example, nonsense and frameshift).

Functional consequences of interlocus gene-conversion events causing human inherited disease. Our systematic survey of the 44 interlocus gene-conversion events that cause human inherited disease (TABLE 1) reveals that all events in which pseudogenes have acted as donors resulted in the functional loss of the respective acceptor genes through the introduction of frameshifting, aberrant splicing, nonsense mutations, deleterious missense mutations and so on. Functional loss of the respective acceptor gene was also associated with nearly all of the gene-conversion events in which a functional gene acted as the donor sequence. The sole exception was *PRSS1*, which encodes cationic trypsinogen, the major isoform of trypsinogen; in this case, gene conversion led to the replacement of at least 289 nucleotides of *PRSS1* by those of *PRSS2*, which encodes anionic trypsinogen,

Box 3 | **Gene conversion involving *Alu* and LINE-1 sequences**

Amplified to >1.1 million copies and comprising >10% of the genome mass, the *Alu* repeat constitutes the most abundant short interspersed element family in the human genome. Despite its short length (<300 bp), *Alu*-mediated gene conversion is not rare: a genome-wide analysis revealed that some 10–20% of the sequence variation in the Ya5 subfamily (one of the youngest *Alu* subfamilies) was attributable to gene conversion[117]. A more recent study estimated that some 15,000–85,000 point mutations in the human genome can be attributed to sequence exchanges between neighbouring *Alu* elements, mainly by gene conversion[118].

This high gene-conversion rate is related to the dense distribution of *Alu* elements in the human genome (on average, one element every 3 kb), their high GC content (~63%) and the remarkable sequence similarity between *Alu* elements (70–100%). Sequence homogenization between *Alu* elements would be expected to potentiate *Alu*-mediated genomic rearrangements in both an evolutionary and a pathological context[119].

Long interspersed elements-1 (LINE-1 or L1) constitute ~17% of the human genome sequence. Given the greater length of L1 elements (up to 6–7 kb) and their prevalence in the human genome (>500,000 copies), L1-mediated gene conversion would not be expected to occur infrequently. L1 elements have been experimentally demonstrated to efficiently mediate gene conversion *in vitro* and in trangenic mouse lines[120,121]. However, only a limited number of gene-conversion events have been documented in humans[122,123]. Several reasons might account for this. First, although a full-length L1 element is typically of ~6 kb, more than 90% of the L1 copies in the human genome are heterogeneously truncated from their 5′ end, and many are either shorter than 2 kb and/or rearranged (for example, REF. 124). Second, the degree of sequence divergence between L1 elements is generally much higher than that between *Alu* sequences: L1 elements have been amplified in vertebrate genomes for around 170 million years (Myr), beginning before the mammalian radiation, whereas the amplification of *Alu* sequences has been specific to primate genomes over the last 65 Myr. By comparison, even those L1 elements that have been amplified over the past 70 Myr have diverged by 32% (REF. 125). Third, by comparison with *Alu* sequences, L1 elements are GC-poor (~43%) and are characterized by a higher average distance between them (one element every 6.3 kb)[119]. Fourth, L1 elements tend to occur in AT-rich, low-recombination and gene-poor regions of the human genome.

the second major isoform of trypsinogen. *In vitro* functional analysis showed that trypsin levels are increased (through enhanced autocatalytic activation) as a result of a specific missense substitution, N29I, which is brought about by the conversion[88].

The functional dissection of the only example of a partially functional gene acting as the donor sequence in gene conversion (that is, *SMN2*) has contributed significantly to our current understanding of the determinants of exon recognition. Mutations in *SMN1* cause autosomal recessive spinal muscular atrophy (SMA). *SMN2* is located centromeric to *SMN1*, with a sequence similarity of ~99%. *SMN2* is considered partially functional because its expression is impaired by a translationally silent SNP at position 6 of exon 7 (C in *SMN1* versus T in *SMN2*, the only difference between the coding sequences of the two genes), which causes skipping of exon 7 in most *SMN2* transcripts; only a minority of *SMN2* transcripts are correctly spliced and hence capable of encoding a protein identical to SMN1. By contrast, most of the wild-type *SMN1* transcripts are full-length. When a minigene was created in which C was artificially replaced by T at position 6 of exon 7 to mimic the SMA-causing gene-conversion mutation in *SMN1*, most of the mutant *SMN1* transcripts also lacked exon 7 (REF. 89). Therefore C→T at position 6 of exon 7 of the *SMN* genes is a key determinant of exon identity and alternative splicing.

Intrachromosomal versus interchromosomal exchange. In nearly all known cases of disease-causing interlocus gene conversion, the acceptor and donor genes are located on the same chromosome. In only one case do the acceptor and donor genes reside on different chromosomes (TABLE 1): the von Willebrand factor gene (*VWF*; 12p13.3) and its pseudogene (*VWFP*; 22q11.22–q11.23). The pseudogene spans exons 23–34 of its functional counterpart and is 97% homologous to it.

This example establishes the precedent for the occurrence, in a pathological context, of gene conversion between unlinked loci in the human genome. However, it is surprising that *VWF* conversions are as common as they are. Of the 17 acceptor genes involved in interlocus gene-conversion events that cause human inherited disease, *VWF* has the highest number (up to 10) of different mutations reported (TABLE 1). In one study, 8 out of 56 (14%) von Willebrand disease (VWD) patients were shown to carry mutations that were compatible with the occurrence of pseudogene-mediated gene conversion[90]. This unexpectedly high frequency can be at least partially accounted for by two considerations: first, the *VWF* gene lies at the tip of chromosome 12p, and chromosome ends are known to be hotspots of recombination and DSBs[91,92]; second, as a bleeding disorder, VWD readily comes to clinical attention and is one of the most extensively studied human inherited diseases.

Interlocus versus interallelic events. In stark contrast to the frequent detection of pathogenic interlocus gene conversion (TABLE 1), the occurrence of interallelic gene conversion causing human inherited disease is rare. One putative example involves the T364M missense mutation in the α-L-iduronidase gene (*IDUA*), which was found in the homozygous state in a patient with mucopolysaccharidosis type I (REF. 93); however, only the patient's father carried the mutation, whereas his mother did not. Analysis of additional polymorphic markers suggested that either a *de novo* interallelic gene conversion (with the paternal allele acting as the donor) or a *de novo* genomic deletion (of the maternal allele) had occurred, leading to real or apparent homozygosity for T364M[93].

Analogously, apparent homozygosity for a mutation in exon 10 of the follicle-stimulating hormone receptor gene (*FSHR*) in a patient with hypergonadotrophic hypogonadism could in principle have resulted from either a *de novo* interallelic gene conversion (with the paternal allele acting as the donor) or a *de novo* genomic deletion of the maternal allele[94]. In the first case, gene conversion must have occurred postzygotically at an early stage of embryonic development. In the second case, the genomic deletion would probably have occurred in the maternal germ cells. It should be possible to distinguish between these possibilities by quantitative PCR because, in the first case, but not the second case, the patient would be expected to show a degree of mosaicism.

Box 4 | **Pseudogene-mediated interlocus gene conversion during evolution**

Immunoglobulin (Ig) diversification in mammalian B cells is generated mainly by somatic hypermutation and class switch recombination. By contrast, gene conversion is the primary mechanism used by chicken B cells to generate Ig gene diversity (reviewed in REFS 126,127). Chickens have a limited number of functional V regions; and the 'reservoir' of potential genetic changes is contained within a series of non-functional V genes. Thus, although chickens have only a single functional light-chain V segment (V_λ), after successful recombination with a single J_λ segment early on in B-cell development, V_λ undergoes diversification by copying sequence from any of its 25 upstream pseudo-V_λ genes[126,127]. Gene conversion is also used, albeit rarely, for Ig diversification by the B cells of rodents (for example, REF. 128) and other mammals including humans (for example, REF. 129).

Introduction of pathogenic mutations into functional genes by pseudogene-mediated gene conversion is a well-known mutational mechanism in human genetic disease (TABLE 1). By analogy, could it be that pseudogenes have also served as templates from which multiple, potentially advantageous changes in their single-copy functional source genes have been derived during the course of human evolution? As long as pseudogene-templated changes in functional genes were not deleterious, they might eventually become fixed. In principle, therefore, pseudogenes could act as a reservoir of sequence variants that could then be transferred to the functional gene in new hitherto untested combinations for selection to act on.

The human sialic acid binding Ig-like lectin 11 (*SIGLEC11*) gene might provide one example of just such an evolutionarily significant gene conversion. Its 5′ upstream region and the exons that encode the sialic acid recognition domain (~2 kb) have been converted by the closely flanking *SIGLECP16* pseudogene[130]. This event came to light through a comparison of human *SIGLEC11* and its pseudogene with their homologues in the chimpanzee, bonobo, gorilla and orangutan. The observations that this gene conversion occurred in only the human lineage, that brain cortex acquired prominent *SIGLEC11* expression specifically in the human lineage and that human SIGLEC11 shows altered substrate binding as compared with its chimpanzee counterpart, are suggestive of an adaptive change that could have been important in the evolution of the genus *Homo*[130].

How frequently this mechanism of pseudogene-mediated gene conversion has contributed to functional and adaptive changes during human evolution remains unknown, but the *SIGLEC11* example suggests that it can introduce genetic changes that might then become positively selected. The contribution of pseudogene-mediated gene conversion to promoting gene evolution has so far scarcely been explored.

Somatic hypermutation
A process that occurs after immunoglobulin gene rearrangement, whereby the base sequences of part of the immunoglobulin variable regions are mutated more frequently than the rest of the genome. This sequence variation is subject to a selection process in the immune system that favours those cells that express immunoglobulins with the highest affinity for an antigen.

Class switch recombination
The somatic recombination process by which immunoglobulin isotypes are switched to IgG, IgA or IgE, without altering antigen specificity.

The study of a patient with campomelic dysplasia, bearing a homozygous nonsense mutation (Y440X) in SRY-box 9 (*SOX9*), led to the identification of an apparent *de novo* interallelic gene-conversion event[95]. Unlike the previous examples, this mutation was absent in both parents. Analysis of intragenic polymorphisms, quantitative real-time PCR of the Y440X-harbouring region and the study of the mosaicism that was evident in the patient's leukocytes led the authors to propose an elegant mechanism for how this unique case of spontaneous homozygosity could have been generated. First, a *de novo* Y440X mutation in *SOX9* occurred in the maternal germ line, yielding a zygote that was heterozygous for the mutant maternal *SOX9* allele and that bore a wild-type paternal *SOX9* allele. Following several cell divisions, a somatic interallelic gene conversion resulted in the replacement of at least 440 nucleotides of the wild-type paternal allele by the mutant maternal allele, resulting in somatic mosaicism[95].

Gene conversion in cancer. Only a few examples of gene-conversion events have been well documented in cancer. The first was noted in a Burkitt lymphoma cell line, in which a ~2kb DNA segment at the *D6S347* locus was replaced by a partially homologous sequence; this event was presumed to result from a somatic interlocus gene conversion involving an unidentified gene[96]. More recently, multiplex ligation-dependent probe amplification analysis showed that somatic mutL homologue 1 (*MLH1*) and mutS homologue 2 (*MSH2*) deletions are identical to their germline counterparts in up to 55% of hereditary nonpolyposis colorectal cancer (HNPCC) specimens[97]. Given that none of these tumours showed allele loss at markers that flanked the respective gene loci, gene conversion was invoked to account for some of the observed changes.

More recently, a germline gene-conversion event in the gene postmeiotic segregation increased 2 (*PMS2*), involving transfer of 3–23 bp from its centromeric pseudogene, *PMS2CL*, has been reported in a HNPCC patient and his affected sister[98]. *MLH1*, *MSH2* and *PMS2* are all DNA mismatch repair genes, defects of which lead to elevated levels of recombination[99].

Given that the detection of somatic mutations is technically much more demanding than that of their germline counterparts, it is likely that the occurrence of somatic gene-conversion events in cancer has been seriously underestimated.

De novo pathogenic gene-conversion events. The fairly high prevalence of *de novo* gene-conversion events (6 out of the 44 interlocus events listed in TABLE 1, and all the somatic events in both inherited disease and cancer) is indicative of the dynamic nature of gene conversion. TABLE 2 lists these known *de novo* pathogenic gene-conversion events.

Mutation spreading on different chromosomal backgrounds. Although recurrent mutation can provide one explanation for the occurrence of a specific mutation on different chromosomal backgrounds in different ethnic groups, interallelic gene conversion provides an equally plausible mechanism. Several human *HBB* mutations (reviewed in REF. 34) and a *CBS* mutation, which cause β-thalassaemia and autosomal recessive cystathionine β-synthase (CBS) deficiency (known as homocystinuria), respectively, are two well-known examples.

CBS deficiency is an inborn error of sulphur metabolism. The most common pathogenic *CBS* variant in Western Eurasians is c.833T>C (p.I278T) in exon 8. However, only a small proportion of human chromosomes carry the pathogenic c.833C mutation on its own; a much larger proportion contains a non-pathogenic combination of two mutations *in cis* c.[833C; 844_845ins68] (REF. 100). The non-pathogenic c.[833C; 844_845ins68] chromosomes are common in sub-Saharan Africa (up to 40% of control chromosomes), but less frequent in Europe and America, where the pathogenic c.[833C;–] chromosomes are more prevalent. Haplotype analysis of 69 pathogenic c.[833C;–] chromosomes of predominantly European origin, using 12 intragenic *CBS* polymorphic markers, revealed three unrelated haplotypes, suggesting that the three pathogenic and comparatively prevalent c.[833C;–] (c. stands for coding nucleotides) chromosomes probably originated by recurrent interallelic gene conversion in which

Table 2 | ***De novo* pathogenic gene-conversion events***

Disorder	Donor	Acceptor	Mutation	Type	Origin	Alternative mechanism	Refs
Inherited disease							
Atypical haemolytic uraemic syndrome	*CFHR1*	*CFH*	c.[3572C>T;3590T>C]	Interlocus	Germline	–	131
Congenital adrenal hyperplasia	*CYP21A1P*	*CYP21A2*	[1688G>T;1767_1768insT]	Interlocus	Germline	–	133
Neural tube defects	*FOLR1P*	*FOLR1*	7497_7662 including 13 discriminant nucleotides	Interlocus	Germline	–	138
Neural tube defects	*FOLR1P*	*FOLR1*	[7539T>C;7541G>A]	Interlocus	Germline	–	138
Hereditary persistence of fetal haemoglobin	*HBG2*	*HBG1*	Conversion involving 9 discriminant nucleotides in the promoter region	Interlocus	Germline	–	145
Shwachman–Diamond syndrome	*SBDSP*	*SBDS*	c.[141C>T;183_184TA>CT; 201A>G;258+2T>C]	Interlocus	Germline	–	152
Mucopolysaccharidosis type I (Hurler–Scheie syndrome)	*IDUA* (Paternal allele)	*IDUA* (Maternal allele)	Homozygosity of T364M plus five other markers	Interallelic	Somatic	*De novo* deletion in the maternally inherited allele	93
Hypergonadotrophic hypogonadism	*FSHR* (Paternal allele)	*FSHR* (Maternal allele)	Homozygosity of a C>G transversion within exon 10 plus a downstream single nucleotide polymorphism	Interallelic	Somatic	Maternally inherited allele harbours a *de novo* genomic deletion	94
Campomelic dysplasia	*SOX9* (Maternal allele)	*SOX9* (Paternal allele)	Homozygosity of Y440X	Interallelic	Somatic	–	95
Cancer							
Burkitt lymphoma	Unknown	*D6S347*	Replacement of ~2 kb DNA segment	Interlocus	Somatic	–	96
Hereditary nonpolyposis colorectal cancer	*MLH1* or *MSH2*	*MLH1* or *MSH2*	Somatic deletions	Interallelic	Somatic	–	97

*Arranged in the same order as in the text. In mutation nomenclature, 'c.' refers to coding nucleotides.

the common non-pathogenic c.[833C; 844_845ins68] chromosomes were used as a template[101].

Therapeutic gene conversion. Gene conversion of an allele bearing an inherited pathogenic mutation can occasionally lead to the reversion of the disease phenotype, a phenomenon that has been dubbed 'natural gene therapy'. This is best exemplified by the mitotic gene conversion that underlies the unusual phenotype of an individual with epidermolysis bullosa, who presented with patches of skin that seemed to be unaffected[102]. In theory, mitotic natural gene therapy can occur only in either dominantly inherited diseases or recessively inherited diseases that are caused by different mutations in the two alleles, resulting in somatic mosaicism[49]. In the first case, the mutant allele is reverted to normal by the wild-type allele whereas, in the latter case, one mutant allele is converted back to wild type by the other mutant allele. In short, mitotic natural gene therapy seems to invariably involve interallelic gene conversion. An extension to this concept is provided by the exceptional example of the *SMN1* and *SMN2* gene pair. Instances of conversion of *SMN2* by *SMN1* with respect to C→T at nucleotide 6 of exon 7, which would serve to convert the partially functional *SMN2* to a fully functional *SMN2*, have been detected in the general population[103]. A similar process could also occur in some SMA patients, provided that they carried at least one *SMN1* allele in which C at nucleotide 6 of exon 7 remained intact. The resulting increased production of SMN protein would modulate the clinical manifestation of SMA.

Natural gene therapy raises the exciting prospect of harnessing gene conversion for therapeutic purposes and has, in part, been responsible for fuelling the development of oligonucleotide-based strategies for gene therapy[49]. However, despite considerable progress over the past few years, there is much work to be done before this approach becomes a clinical reality[104].

Conclusions and future directions

A sizeable body of data has accumulated to show that, although gene conversion has been an important driving force in human genome evolution, it can also be the cause of devastating genetic conditions. There is now considerable evidence to support the view that gene conversion derives from distinct pathways, rather than from a random resolution of double HJs. Yet we are only beginning to unravel the basic mechanisms that underlie gene conversion and crossover events in eukaryotic cells. Many important questions remain to be answered. When and how is a recombination event 'pre-determined' to a given fate? How does one and the same protein

Multiplex ligation-dependent probe amplification analysis
A semi-quantitative PCR-based method that allows multiple targets to be amplified with only a single primer pair, and that is widely used for detecting copy-number variations.

participate in distinct pathways leading to gene conversion? What determines the selective use of different proteins to promote SDSA? Are the different pathways of gene conversion conserved across the eukaryotic kingdom? What is the nature of the sequences that can promote gene conversion? Which factors influence or determine the frequency, tract length and directionality of gene conversion? Answers to these questions should enhance our understanding of the significance of gene conversion and homologous recombination in ensuring the fidelity of DSB repair and the maintenance of genome integrity.

In the context of evolution, studies of the role of gene conversion in creating specific multigene families have been greatly facilitated by the availability of complete genome sequences from multiple species. The use of population genetic data to make inferences about the impact of gene conversion at the genome level is gaining popularity. These approaches, when combined with new or improved statistical methods, are expected to provide better estimates of the prevalence and spatial distribution of gene-conversion events along human chromosomes. Further integration of data from genome-wide studies, sperm typing and disease analysis should provide a clearer picture of the impact that gene conversion has had on the evolution of human and other mammalian genomes.

Undoubtedly, the spectrum of gene-conversion events causing human disease will continue to expand. In this regard, it should be noted that our meta-analysis deliberately excluded many point mutations that could well have been bona fide gene-conversion events, thereby inadvertently understating the actual contribution of this mutational mechanism to human genetic disease.

Considering the available disease data and the basic mechanism that leads to gene conversion, we speculate that somatic mosaicism resulting from interallelic gene conversion might, until now, have largely escaped our attention. We propose two factors that might account for this possibility. First, leukocyte genomic DNA is usually used for disease-associated mutation screening, and this tissue might be an inappropriate choice if somatic mosaicism is to be considered. Second, even when genomic DNA has been prepared from the tissue that is most relevant to the disease in question, somatic mosaicism is still likely to be overlooked if quantitative PCR is not carried out. Given that a restoration of only 5% of the function of an affected gene can significantly ameliorate the clinical symptoms of a disease (for example, cystic fibrosis), somatic mosaicism caused by interallelic gene conversion could turn out to be a potentially new and important disease modifier.

1. Slightom, J. L., Blechi, A. E. & Smithies, O. Human fetal Gγ- and Aγ-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell* **21**, 627–638 (1980).
2. Paques, F. & Haber, J. E. Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **63**, 349–404 (1999).
3. Krogh, B. O. & Symington, L. S. Recombination proteins in yeast. *Annu. Rev. Genet.* **38**, 233–271 (2004).
4. Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J. & Stahl, F. W. The double-strand-break repair model for recombination. *Cell* **33**, 25–35 (1983).
5. Haber, J. E., Ira, G., Malkova, A. & Sugawara, N. Repairing a double-strand chromosome break by homologous recombination: revisiting Robin Holliday's model. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **359**, 79–86 (2004).
This review describes how the seminal DSBR and SDSA models have evolved over time.
6. Hunter, N. & Kleckner, N. The single-end invasion: an asymmetric intermediate at the double-strand break to double-Holliday junction transition of meiotic recombination. *Cell* **106**, 59–70 (2001).
7. Ira, G., Satory, D. & Haber, J. E. Conservative inheritance of newly synthesized DNA in double-strand break-induced gene conversion. *Mol. Cell. Biol.* **26**, 9424–9429 (2006).
8. Ira, G., Malkova, A., Liberi, G., Foiani, M. & Haber, J. E. Srs2 and Sgs1–Top3 suppress crossovers during double-strand break repair in yeast. *Cell* **115**, 401–411 (2003).
This study suggests that, whereas Srs2 promotes the SDSA pathway, Sgs1 and the topoisomerase Top3 remove double HJs, both leading to the generation of only gene-conversion events.
9. Robert, T., Dervins, D., Fabre, F. & Gangloff, S. Mrc1 and Srs2 are major actors in the regulation of spontaneous crossover. *EMBO J.* **25**, 2837–2846 (2006).
10. Aylon, Y., Liefshitz, B., Bitan-Banin, G. & Kupiec, M. Molecular dissection of mitotic recombination in the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **23**, 1403–1417 (2003).
11. Krejci, L. *et al.* DNA helicase Srs2 disrupts the Rad51 presynaptic filament. *Nature* **423**, 305–309 (2003).
12. Veaute, X. *et al.* The Srs2 helicase prevents recombination by disrupting Rad51 nucleoprotein filaments. *Nature* **423**, 309–312 (2003).
13. Krejci, L. *et al.* Role of ATP hydrolysis in the antirecombinase function of *Saccharomyces cerevisiae* Srs2 protein. *J. Biol. Chem.* **279**, 23193–23199 (2004).
14. Adams, M. D., McVey, M. & Sekelsky, J. J. *Drosophila* BLM in double-strand break repair by synthesis-dependent strand annealing. *Science* **299**, 265–267 (2003).
15. McVey, M., Larocque, J. R., Adams, M. D. & Sekelsky, J. J. Formation of deletions during double-strand break repair in *Drosophila* DmBLM mutants occurs after strand invasion. *Proc. Natl Acad. Sci. USA* **101**, 15694–15699 (2004).
16. Weinert, B. T. & Rio, D. C. DNA strand displacement, strand annealing and strand swapping by the *Drosophila* Bloom's syndrome helicase. *Nucleic Acids Res.* **35**, 1367–1376 (2007).
17. Bachrati, C. Z., Borts, R. H. & Hickson, I. D. Mobile D-loops are a preferred substrate for the Bloom's syndrome helicase. *Nucleic Acids Res.* **34**, 2269–2279 (2006).
18. Bugreev, D. V., Mazina, O. M. & Mazin, A. V. Rad54 protein promotes branch migration of Holliday junctions. *Nature* **442**, 590–593 (2006).
This work identifies a novel function (that is, promotion of bidirectional DNA branch migration) of the Rad54 protein, and suggests that it could facilitate either the SDSA pathway or the formation of double HJs.
19. Wu, L. & Hickson, I. D. The Bloom's syndrome helicase suppresses crossing over during homologous recombination. *Nature* **426**, 870–874 (2003).
This work shows that BLM and human topoisomerase IIIα suppress crossing over through the mechanism of double-HJ dissolution.
20. Jessop, L., Rockmill, B., Roeder, G. S. & Lichten, M. Meiotic chromosome synapsis-promoting proteins antagonize the anti-crossover activity of sgs1. *PLoS Genet.* **2**, e155 (2006).
21. Plank, J. L., Wu. J. & Hsieh, T. S. Topoisomerase IIIα and Bloom's helicase can resolve a mobile double Holliday junction substrate through convergent branch migration. *Proc. Natl Acad. Sci. USA* **103**, 11118–11123 (2006).
22. Johnson-Schlitz, D. & Engels, W. R. Template disruptions and failure of double Holliday junction dissolution during double-strand break repair in *Drosophila* BLM mutants. *Proc. Natl Acad. Sci. USA* **103**, 16840–16845 (2006).
23. Wu, L. *et al.* BLAP75/RMI1 promotes the BLM-dependent dissolution of homologous recombination intermediates. *Proc. Natl Acad. Sci. USA* **103**, 4068–4073 (2006).
This work identifies BLAP75 as the third component of the double-HJ dissolvasome; this finding was confirmed concurrently in reference 24.
24. Raynard, S., Bussen, W. & Sung, P. A double Holliday junction dissolvasome comprising BLM, topoisomerase IIIα, and BLAP75. *J. Biol. Chem.* **281**, 13861–13864 (2006).
25. Allers, T. & Lichten, M. Differential timing and control of noncrossover and crossover recombination during meiosis. *Cell* **106**, 47–57 (2001).
26. Borner, G. V., Kleckner, N. & Hunter, N. Crossover/noncrossover differentiation, synaptonemal complex formation, and regulatory surveillance at the leptotene/zygotene transition of meiosis. *Cell* **117**, 29–45 (2004).
27. Liu, Y. & West, S. C. Happy Hollidays: 40th anniversary of the Holliday junction. *Nature Rev. Mol. Cell Biol.* **5**, 937–944 (2004).
An historical review of the HJ that remains the basis of our thinking about homologous recombination.
28. Schildkraut, E., Miller, C. A. & Nickoloff, J. A. Gene conversion and deletion frequencies during double-strand break repair in human cells are controlled by the distance between direct repeats. *Nucleic Acids Res.* **33**, 1574–1580 (2005).
29. Ezawa, K., Oota, S. & Saitou, N. Genome-wide search of gene conversions in duplicated genes of mouse and rat. *Mol. Biol. Evol.* **23**, 927–940 (2006).
30. Liskay, R. M., Letsou, A. & Stachelek, J. Homology requirement for efficient gene conversion between duplicated chromosomal sequences in mammalian cells. *Genetics* **115**, 161–167 (1987).
31. Waldman, A. S. & Liskay, R. M. Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology. *Mol. Cell. Biol.* **8**, 5350–5357 (1988).
32. Reiter, L. T. *et al.* Human meiotic recombination products revealed by sequencing a hotspot for homologous strand exchange in multiple *HNPP* deletion patients. *Am. J. Hum. Genet.* **62**, 1023–1033 (1998).
33. Judd, S. R. & Petes, T. D. Physical lengths of meiotic and mitotic gene conversion tracts in *Saccharomyces cerevisiae*. *Genetics* **118**, 401–410 (1988).

34. Papadakis, M. N. & Patrinos, G. P. Contribution of gene conversion in the evolution of the human β-like globin gene family. *Hum. Genet.* **104**, 117–125 (1999).
35. Bosch, E., Hurles, M. E., Navarro, A. & Jobling, M. A. Dynamics of a human interparalog gene conversion hotspot. *Genome Res.* **14**, 835–844 (2004).
36. Zangenberg, G., Huang, M. M., Arnheim, N. & Erlich, H. New *HLA-DPB1* alleles generated by interallelic gene conversion detected by analysis of sperm. *Nature Genet.* **10**, 407–414 (1995).
37. Jeffreys, A. J. & May, C. A. Intense and highly localized gene conversion activity in human meiotic crossover hot spots. *Nature Genet.* **36**, 151–156 (2004).
This work provides evidence for hotspots of human interallelic gene conversion with the potential to exert profound effects on haplotype diversity.
38. Bacolla, A. *et al.* Breakpoints of gross deletions coincide with non-B DNA conformations. *Proc. Natl Acad. Sci. USA* **101**, 14162–14167 (2004).
39. Schildkraut, E., Miller, C. A. & Nickoloff, J. A. Transcription of a donor enhances its use during double-strand break-induced gene conversion in human cells. *Mol. Cell. Biol.* **26**, 3098–3105 (2006).
40. Gonzalez-Barrera, S., Garcia-Rubio, M. & Aguilera, A. Transcription and double-strand breaks induce similar mitotic recombination events in *Saccharomyces cerevisiae*. *Genetics* **162**, 603–614 (2002).
41. Innan, H. A two-locus gene conversion model with selection and its application to the human *RHCE* and *RHD* genes. *Proc. Natl Acad. Sci. USA* **100**, 8793–8798 (2003).
The results presented here provide an indication of the strength of selection that is required to balance gene conversion in maintaining the observed pattern of allelic variation in this two-locus system.
42. Hallast, P., Nagirnaja, L., Margus, T. & Laan, M. Segmental duplications and gene conversion: human luteinizing hormone/chorionic gonadotropin β gene cluster. *Genome Res.* **15**, 1535–1546 (2005).
43. Verrelli, B. C. & Tishkoff, S. A. Signatures of selection and gene conversion associated with human color vision variation. *Am. J. Hum. Genet.* **75**, 363–375 (2004).
By means of population genetic and statistical analyses, these authors showed how a combination of natural selection and gene conversion has shaped sequence diversity in the *OPN1LW* gene.
44. Sharon, D. *et al.* Primate evolution of an olfactory receptor cluster: diversification by gene conversion and recent emergence of pseudogenes. *Genomics* **61**, 24–36 (1999).
45. Woelk, C. H., Frost, S. D., Richman, D. D., Higley, P. E. & Kosakovsky Pond, S. L. Evolution of the interferon α gene family in eutherian mammals. *Gene* **397**, 38–50 (2007).
46. Plotnikova, O. V. *et al.* Conversion and compensatory evolution of the γ-crystallin genes and identification of a cataractogenic mutation that reverses the sequence of the human *CRYGD* gene to an ancestral state. *Am. J. Hum. Genet.* **81**, 32–43 (2007).
47. Vazquez-Salat, N., Yuhki, N., Beck, T., O'Brien, S. J. & Murphy, W. J. Gene conversion between mammalian *CCR2* and *CCR5* chemokine receptor genes: a potential mechanism for receptor dimerization. *Genomics* **90**, 213–224 (2007).
48. Eickbush, T. H. & Eickbush, D. G. Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics* **175**, 477–485 (2007).
49. Hurles, M. E. in: *Encyclopedia of Life Sciences* (John Wiley & Sons, Chichester, 2003).
50. Rozen, S. *et al.* Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* **423**, 873–876 (2003).
Having estimated that an average of ~600 nucleotides in each newborn male have undergone Y–Y gene conversion, the authors highlighted an important role for gene conversion in the evolution of multi-copy testis-expressed gene families in the male-specific region of the human Y chromosome.
51. Hurles, M. E., Willey, D., Matthews, L. & Hussain, S. S. Origins of chromosomal rearrangement hotspots in the human genome: evidence from the *AZFa* deletion hotspots. *Genome Biol.* **5**, R55 (2004).
The authors carried out multiple simulations to explore how gene conversion homogenizes paralogous sequences at the same time that it diversifies orthologous sequences.
52. Winderickx, J., Battisti, L., Hibiya, Y., Motulsky, A. G. & Deeb, S. S. Haplotype diversity in the human red and green opsin genes: evidence for frequent sequence exchange in exon 3. *Hum. Mol. Genet.* **2**, 1413–1421 (1993).
53. Carroll, J., Neitz, J. & Neitz, M. Estimates of L:M cone ratio from ERG flicker photometry and genetics. *J. Vis.* **2**, 531–542 (2002).
54. Joly, E. & Rouillon, V. The orthology of *HLA-E* and *H2-Qa1* is hidden by their concerted evolution with other MHC class I molecules. *Biol. Direct* **1**, 2 (2006).
55. Yip, S. P. Sequence variation at the human *ABO* locus. *Ann. Hum. Genet.* **66**, 1–27 (2002).
56. von Salome, J., Gyllensten, U. & Bergstrom, T. F. Full-length sequence analysis of the *HLA-DRB1* locus suggests a recent origin of alleles. *Immunogenetics* **59**, 261–271 (2007).
57. Hurles, M. Are 100,000 'SNPs' useless? *Science* **298**, 1509 (2002).
58. Bailey, J. A. *et al.* Recent segmental duplications in the human genome. *Science* **297**, 1003–1007 (2002).
59. Fredman, D. *et al.* Complex SNP-related sequence variation in segmental genome duplications. *Nature Genet.* **36**, 861–866 (2004).
One of the first papers to report structural variation in the human genome, providing evidence for a role for gene conversion in promoting the variability of duplicon sequences.
60. Pavlicek, A., House, R., Gentles, A. J., Jurka, J. & Morrow, B. E. Traffic of genetic information between segmental duplications flanking the typical 22q11.2 deletion in velo-cardio-facial syndrome/DiGeorge syndrome. *Genome Res.* **15**, 1487–1495 (2005).
61. Cheng, Z. *et al.* A genome-wide comparison of recent chimpanzee and human segmental duplications. *Nature* **437**, 88–93 (2005).
62. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
63. Jackson, M. S. *et al.* Evidence for widespread reticulate evolution within human duplicons. *Am. J. Hum. Genet.* **77**, 824–840 (2005).
64. Ardlie, K. *et al.* Lower-than-expected linkage disequilibrium between tightly linked markers in humans suggests a role for gene conversion. *Am. J. Hum. Genet.* **69**, 582–589 (2001).
65. Ptak, S. E., Voelpel, K. & Przeworski, M. Insights into recombination from patterns of linkage disequilibrium in humans. *Genetics* **167**, 387–397 (2004).
66. International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
67. Conrad, D. F. *et al.* A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nature Genet.* **38**, 1251–1260 (2006).
68. de Bakker, P. I. *et al.* Transferability of tag SNPs in genetic association studies in multiple populations. *Nature Genet.* **38**, 1298–1303 (2006).
69. Wall, J. D. Close look at gene conversion hot spots. *Nature Genet.* **36**, 114–115 (2004).
70. Need, A. C. & Goldstein, D. B. Genome-wide tagging for everyone. *Nature Genet.* **38**, 1227–1228 (2006).
71. Lindsay, S. J., Khajavi, M., Lupski, J. R. & Hurles, M. E. A chromosomal rearrangement hotspot can be identified from population genetic variation and is coincident with a hotspot for allelic recombination. *Am. J. Hum. Genet.* **79**, 890–902 (2006).
Together with reference 72, this work reveals the imprint of gene conversion by resequencing mutation-prone disease-associated low copy repeat (LCR) regions (commented on in reference 75).
72. Raedt, T. D. *et al.* Conservation of hotspots for recombination in low-copy repeats associated with the *NF1* microdeletion. *Nature Genet.* **38**, 1419–1423 (2006).
73. Myers, S., Bottolo, L., Freeman, C., McVean, G. & Donnelly, P. A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**, 321–324 (2005).
74. Coop, G. & Myers, S. R. Live hot, die young: transmission distortion in recombination hotspots. *PLoS Genet.* **3**, e35 (2007).
75. Myers, S. R. & McCarroll, S. A. New insights into the biological basis of genomic disorders. *Nature Genet.* **38**, 1363–1364 (2006).
76. Blanco, P. *et al.* Divergent outcomes of intrachromosomal recombination on the human Y chromosome: male infertility and recurrent polymorphism. *J. Med. Genet.* **37**, 752–758 (2000).
77. Forbes, S. H., Dorschner, M. O., Le, R. & Stephens, K. Genomic context of paralogous recombination hotspots mediating recurrent *NF1* region microdeletion. *Genes Chrom. Cancer* **41**, 12–25 (2004).
78. Hellenthal, G. & Stephens, M. Insights into recombination from population genetic variation. *Curr. Opin. Genet. Dev.* **16**, 565–572 (2006).
79. Padhukasahasram, B., Marjoram, P. & Nordborg, M. Estimating the rate of gene conversion on human chromosome 21. *Am. J. Hum. Genet.* **75**, 386–397 (2004).
80. Frisse, L. *et al.* Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. *Am. J. Hum. Genet.* **69**, 831–843 (2001).
81. Jeffreys, A. J. & Neumann, R. Factors influencing recombination frequency and distribution in a human meiotic crossover hotspot. *Hum. Mol. Genet.* **14**, 2277–2287 (2005).
82. Holloway, K., Lawson, V. E. & Jeffreys, A. J. Allelic recombination and *de novo* deletions in sperm in the human β-globin gene region. *Hum. Mol. Genet.* **15**, 1099–1111 (2006).
83. Marais, G. Biased gene conversion: implications for genome and sex evolution. *Trends Genet.* **19**, 330–338 (2003).
84. Galtier, N. Gene conversion drives GC content evolution in mammalian histones. *Trends Genet.* **19**, 65–68 (2003).
85. Spencer, C. C. *et al.* The influence of recombination on human genetic diversity. *PLoS Genet.* **2**, e148 (2006).
86. Spencer, C. C. Human polymorphism around recombination hotspots. *Biochem. Soc. Trans.* **34**, 535–536 (2006).
87. Hernandez, R. D., Williamson, S. H., Zhu, L. & Bustamante, C. D. Context-dependent mutation rates may cause spurious signatures of a fixation bias favoring higher GC-content in humans. *Mol. Biol. Evol.* 26 Jul 2007 (doi:10.1093/molbev/msm149).
88. Teich, N. *et al.* Gene conversion between functional trypsinogen genes *PRSS1* and *PRSS2* associated with chronic pancreatitis in a six-year-old girl. *Hum. Mutat.* **25**, 343–347 (2005).
This paper reports a unique gene-conversion event that occurred between two functional genes, resulting in a gain of function.
89. Lorson, C. L., Hahnen, E., Androphy, E. J. & Wirth, B. A single nucleotide in the *SMN* gene regulates splicing and is responsible for spinal muscular atrophy. *Proc. Natl Acad. Sci. USA* **96**, 6307–6311 (1999).
90. Gupta, P. K. *et al.* Gene conversions are a common cause of von Willebrand disease. *Br. J. Haematol.* **130**, 752–758 (2005).
91. Linardopoulou, E. V. *et al.* Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. *Nature* **437**, 94–100 (2005).
92. Rudd, M. K. *et al.* Elevated rates of sister chromatid exchange at chromosome ends. *PLoS Genet.* **3**, e32 (2007).
93. Lee-Chen, G. J. & Wang, T. R. Mucopolysaccharidosis type I: identification of novel mutations that cause Hurler/Scheie syndrome in Chinese families. *J. Med. Genet.* **34**, 939–941 (1997).
94. Allen, L. A. *et al.* A novel loss of function mutation in exon 10 of the FSH receptor gene causing hypergonadotrophic hypogonadism: clinical and molecular characteristics. *Hum. Reprod.* **18**, 251–256 (2003).
95. Pop, R., Zaragoza, M. V., Gaudette, M., Dahrmann, U. & Scherer, G. A homozygous nonsense mutation in *SOX9* in the dominant disorder campomelic dysplasia: a case of mitotic gene conversion. *Hum. Genet.* **117**, 43–53 (2005).
This work reports the first fully characterized somatic interallelic gene-conversion event in the context of human inherited disease.
96. Hauptschein, R. S. *et al.* An apparent interlocus gene conversion-like event at a putative tumor suppressor gene locus on human chromosome 6q27 in a Burkitt's lymphoma cell line. *DNA Res.* **7**, 261–272 (2000).
97. Zhang, J. *et al.* Gene conversion is a frequent mechanism of inactivation of the wild-type allele in cancers from *MLH1*/*MSH2* deletion carriers. *Cancer Res.* **66**, 659–664 (2006).
98. Auclair, J. *et al.* Novel biallelic mutations in *MSH6* and *PMS2* genes: gene conversion as a likely cause of *PMS2* gene inactivation. *Hum. Mutat.* 7 Jun 2007 (doi:10.humu.20569).
99. Chambers, S. R., Hunter, N., Louis, E. J. & Borts, R. H. The mismatch repair system reduces meiotic homeologous recombination and stimulates recombination-dependent chromosome loss. *Mol. Cell. Biol.* **16**, 6110–6120 (1996).
100. Romano, M. *et al.* Regulation of 3′ splice site selection in the 844ins68 polymorphism of the cystathionine β-synthase gene. *J. Biol. Chem.* **277**, 43821–43829 (2002).

101. Vyletal, P. *et al.* Haplotype diversity of cystathionine β-synthase alleles bearing the most common homocystinuria mutation c.833T > C: a possible role for gene conversion. *Hum. Mutat.* **28**, 255–264 (2007).
102. Jonkman, M. F. *et al.* Revertant mosaicism in epidermolysis bullosa caused by mitotic gene conversion. *Cell* **88**, 543–551 (1997).
103. Ogino, S., Gao, S., Leonard, D. G., Paessler, M. & Wilson, R. B. Inverse correlation between *SMN1* and *SMN2* copy numbers: evidence for gene conversion from *SMN2* to *SMN1*. *Eur. J. Hum. Genet.* **11**, 275–277 (2003).
104. Fichou, Y. & Férec, C. The potential of oligonucleotides for therapeutic applications. *Trends Biotechnol.* **24**, 563–570 (2006).
105. Shinohara, A., Ogawa, H. & Ogawa, T. Rad51 protein involved in repair and recombination in *S. cerevisiae* is a RecA-like protein. *Cell* **69**, 457–470 (1992).
106. Masson, J. Y. *et al.* Identification and purification of two distinct complexes containing the five RAD51 paralogs. *Genes Dev.* **15**, 3296–3307 (2001).
107. Liu, Y., Masson, L. Y., Shah, R., O'Regan, P. & West, S. C. RAD51C is required for Holliday junction processing in mammalian cells. *Science* **303**, 243–246 (2004).
108. Lisby, M. & Rothstein, R. DNA repair: keeping it together. *Curr. Biol.* **14**, R994–R996 (2004).
109. De Jager, M. *et al.* Human Rad50/Mre11 is a flexible complex that can tether DNA ends. *Mol. Cell* **8**, 1129–1135 (2001).
110. Moreno-Herrero, F. *et al.* Mesoscale conformational changes in the DNA-repair complex Rad50/Mre11/Nbs1 upon binding DNA. *Nature* **437**, 440–443 (2005).
111. Scully, R. *et al.* Association of BRCA1 with Rad51 in mitotic and meiotic cells. *Cell* **88**, 265–275 (1997).
112. Sharan, S. K. *et al.* Embryonic lethality and radiation hypersensitivity mediated by Rad51 in mice lacking BRCA2. *Nature* **386**, 804–810 (1997).
113. Chen, J. *et al.* Stable interaction between the products of the *BRCA1* and *BRCA2* tumor suppressor genes in mitotic and meiotic cells. *Mol. Cell* **2**, 317–328 (1998).
114. Yuan, S. S. *et al.* BRCA2 is required for ionizing radiation-induced assembly of Rad51 complex *in vivo*. *Cancer Res.* **59**, 3547–3551 (1999).
115. Esashi, F., Galkin, V. E., Yu, X., Egelman, E. H. & West, S. C. Stabilization of RAD51 nucleoprotein filaments by the C-terminal region of BRCA2. *Nature Struct. Mol. Biol.* **14**, 468–474 (2007).
116. Davies, O. R. & Pellegrini, L. Interaction with the BRCA2 C terminus protects RAD51–DNA filaments from disassembly by BRC repeats. *Nature Struct. Mol. Biol.* **14**, 475–483 (2007).
117. Roy, A. M. *et al.* Potential gene conversion and source genes for recently integrated *Alu* elements. *Genome Res.* **10**, 1485–1495 (2000).
118. Zhi, D. Sequence correlation between neighboring *Alu* instances suggests post-retrotransposition sequence exchange due to *Alu* gene conversion. *Gene* **390**, 117–121 (2007).
119. Sen, S. K. *et al.* Human genomic deletions mediated by recombination between *Alu* elements. *Am. J. Hum. Genet.* **79**, 41–53 (2006).
120. Tremblay, A., Jasin, M. & Chartrand, P. A double-strand break in a chromosomal LINE element can be repaired by gene conversion with various endogenous LINE elements in mouse cells. *Mol. Cell. Biol.* **20**, 54–60 (2000).
121. Wurtele, H., Gusew. N., Lussier, R. & Chartrand, P. Characterization of *in vivo* recombination activities in the mouse embryo. *Mol. Genet. Genomics* **273**, 252–263 (2005).
122. Myers, J. S. *et al.* A comprehensive analysis of recently integrated human Ta L1 elements. *Am. J. Hum. Genet.* **71**, 312–326 (2002).
123. Vincent, B. J. *et al.* Following the LINEs: an analysis of primate genomic variation at human-specific LINE-1 insertion sites. *Mol. Biol. Evol.* **20**, 1338–1348 (2003).
124. Chen, J. M., Férec, C. & Cooper, D. N. Mechanism of *Alu* integration into the human genome. *Genome Med.* **1**, 9–17 (2007).
125. Khan, H., Smit, A. & Boissinot, S. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res.* **16**, 78–87 (2006).
126. Maizels, N. Immunoglobulin gene diversification. *Annu. Rev. Genet.* **39**, 23–46 (2005).
127. Tang, E. S. & Martin, A. Immunoglobulin gene conversion: Synthesizing antibody diversification and DNA repair. *DNA Repair* 27 Jun 2007 (doi:10.1016/j.dnarep.2007.05.002).
128. D'Avirro, N., Truong, D., Xu, B. & Selsing, E. Sequence transfers between variable regions in a mouse antibody transgene can occur by gene conversion. *J. Immunol.* **175**, 8133–8137 (2005).
129. Darlow, J. M. & Stott, D. I. Gene conversion in human rearranged immunoglobulin genes. *Immunogenetics* **58**, 511–522 (2006).
130. Hayakawa, T. *et al.* A human-specific gene in microglia. *Science* **309**, 1693 (2005).
This work describes a human-specific gene-conversion event that might have been significant in the evolution of the *Homo* genus.
131. Heinen, S. *et al. De novo* gene conversion in the *RCA* gene cluster (1q32) causes mutations in complement factor H associated with atypical hemolytic uremic syndrome. *Hum. Mutat.* **27**, 292–293 (2006).
132. Lee, H. H., Tsai, F. J., Lee, Y. J. & Yang, Y. C. Diversity of the CYP21A2 gene: a 6.2-kb TaqI fragment and a 3.2-kb TaqI fragment mistaken as CYP21A1P. *Mol. Genet. Metab.* **88**, 372–277 (2006).
133. Friaes, A. *et al. CYP21A2* mutations in Portuguese patients with congenital adrenal hyperplasia: identification of two novel mutations and characterization of four different partial gene conversions. *Mol. Genet. Metab.* **88**, 58–65 (2006).
134. Higashi, Y., Tanae, A., Inoue, H., Hiromasa, T. & Fujii-Kuriyama, Y. Aberrant splicing and missense mutations cause steroid 21-hydroxylase [P-450(C21)] deficiency in humans: possible gene conversion products. *Proc. Natl Acad. Sci. USA* **85**, 7486–7490 (1988).
135. Fardella, C. E. *et al.* Gene conversion in the *CYP11B2* gene encoding P450c11AS is associated with, but does not cause, the syndrome of corticosterone methyloxidase II deficiency. *J. Clin. Endocrinol. Metab.* **81**, 321–326 (1996).
136. Nicod, J., Dick, B., Frey, F. J. & Ferrari, P. Mutation analysis of *CYP11B1* and *CYP11B2* in patients with increased 18-hydroxycortisol production. *Mol. Cell. Endocrinol.* **214**, 167–174 (2004).
137. Sarhadi, V. *et al.* A unique form of autosomal dominant cataract explained by gene conversion between β-crystallin B2 and its pseudogene. *J. Med. Genet.* **38**, 392–396 (2001).
138. De Marco, P. *et al.* Folate pathway gene alterations in patients with neural tube defects. *Am. J. Med. Genet.* **95**, 216–223 (2000).
139. Hatton, C. E., Cooper, A., Whitehouse, C. & Wraith, J. E. Mutation analysis in 46 British and Irish patients with Gaucher's disease. *Arch. Dis. Child.* **77**, 17–22 (1997).
140. Latham, T., Grabowski, G. A., Theophilus, B. D. & Smith, F. I. Complex alleles of the acid β-glucosidase gene in Gaucher disease. *Am. J. Hum. Genet.* **47**, 79–86 (1990).
141. Eyal, N., Wilder, S. & Horowitz, M. Prevalent and rare mutations among Gaucher patients. *Gene* **96**, 277–283 (1990).
142. Hong, C. M., Ohashi, T., Yu, X. J., Weiler, S. & Barranger, J. A. Sequence of two alleles responsible for Gaucher disease. *DNA Cell Biol.* **9**, 233–241 (1990).
143. Millar, D. S. *et al.* Novel mutations of the growth hormone 1 (*GH1*) gene disclosed by modulation of the clinical selection criteria for individuals with short stature. *Hum. Mutat.* **21**, 424–440 (2003).
144. Adams, J. G. III, Marrison, W. T. & Steinberg, M. H. Hemoglobin Parchman: double crossover within a single human gene. *Science* **218**, 291–293 (1982).
145. Patrinos, G. P. *et al.* The Cretan type of non-deletional hereditary persistence of fetal hemoglobin [Aγ −158 C > T] results from two independent gene conversion events. *Hum. Genet.* **102**, 629–634 (1998).
146. Minegishi, Y. *et al.* Mutations in the human λ5/14.1 gene results in B cell deficiency and agammaglobulinemia. *J. Exp. Med.* **187**, 71–77 (1998).
147. Roesler, J. *et al.* Recombination events between the p47-phox gene and its highly homologous pseudogenes are the main cause of autosomal recessive chronic granulomatous disease. *Blood* **95**, 2150–2156 (2000).
148. Vazquez, N. *et al.* Mutational analysis of patients with p47-phox-deficient chronic granulomatous disease: the significance of recombination events between the p47-phox gene (*NCF1*) and its highly homologous pseudogenes. *Exp. Hematol.* **29**, 234–243 (2001).
149. Reyniers, E. *et al.* Gene conversion between red and defective green opsin gene in blue cone monochromacy. *Genomics* **29**, 323–328 (1995).
150. Watnick, T. J., Gandolph, M. A., Weber, H., Neumann, H. P. & Germino, G. G. Gene conversion is a likely cause of mutation in *PKD1*. *Hum. Mol. Genet.* **7**, 1239–1243 (1998).
151. Inoue, S. *et al.* Mutation analysis in *PKD1* of Japanese autosomal dominant polycystic kidney disease patients. *Hum. Mutat.* **19**, 622–628 (2002).
152. Nicolis, E., Bonizzato, A., Assael, B. M. & Cipolli, M. Identification of novel mutations in patients with Shwachman–Diamond syndrome. *Hum. Mutat.* **25**, 410 (2005).
153. Nakashima, E. *et al.* Novel *SBDS* mutations caused by gene conversion in Japanese patients with Shwachman–Diamond syndrome. *Hum. Genet.* **114**, 345–348 (2004).
154. Boocock, G. R. *et al.* Mutations in *SBDS* are associated with Shwachman–Diamond syndrome. *Nature Genet.* **33**, 97–101 (2003).
155. Bussaglia, E. *et al.* A frame-shift deletion in the survival motor neuron gene in Spanish spinal muscular atrophy patients. *Nature Genet.* **11**, 335–337 (1995).
156. Eikenboom, J. C., Castaman, G., Vos, H. L., Bertina, R. M. & Rodeghiero, F. Characterization of the genetic defects in recessive type 1 and type 3 von Willebrand disease patients of Italian origin. *Thromb. Haemost.* **79**, 709–717 (1998).
157. Eikenboom, J. C., Vink, T., Briët, E., Sixma, J. J. & Reitsma, P. H. Multiple substitutions in the von Willebrand factor gene that mimic the pseudogene sequence. *Proc. Natl Acad. Sci. USA* **91**, 2221–2224 (1994).
158. Zhang, Z. P., Blombäck, M., Nyman, D. & Anvret, M. Mutations of von Willebrand factor gene in families with von Willebrand disease in the Aland Islands. *Proc. Natl Acad. Sci. USA* **90**, 7937–7940 (1993).
159. Holmberg, L. *et al.* von Willebrand factor mutation enhancing interaction with platelets in patients with normal multimeric structure. *J. Clin. Invest.* **91**, 2169–2177 (1993).
160. Surdhar, G. K., Enayat, M. S., Lawson, S., Williams, M. D. & Hill, F. G. Homozygous gene conversion in von Willebrand factor gene as a cause of type 3 von Willebrand disease and predisposition to inhibitor development. *Blood* **98**, 248–250 (2001).

Acknowledgements
We regret that, owing to space limitations, much of the relevant work could not be cited. We are indebted to R. Kanaar for critically reading the manuscript and for his valuable and detailed comments. We also thank the three anonymous reviewers for their constructive criticism of the manuscript. This work was partially supported by the INSERM (Institut National de la Santé et de la Recherche Médicale), France, and the Erasmus University Medical Center, Rotterdam, The Netherlands.

Competing interests statement
The authors declare no competing financial interests.

DATABASES
Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
ABO | *AZFA* | CBS | *CCR2* | *CCR5* | *CGB* | *FOLR1* | *FSHR* | *HBB* | *HBG1* | *HBG2* | *HLA-DPB1* | *IDUA* | *MLH1* | *MSH2* | *NID1* | *OPN1LW* | *OPN1MW* | *PMS2* | *PMS2CL* | *PRSS1* | *PRSS2* | *RHCE* | *RHD* | *SIGLEC11* | *SMN1* | *SMN2* | *SOX9* | *VWF* | *VWFP*
OMIM: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
β-thalassaemia | homocystinuria | SMA | VWD
UniProtKB: http://ca.expasy.org/sprot
BLAP75 | BLM | Rad51 | Rad54 | SPO11 | Srs2

FURTHER INFORMATION
The Patrinos laboratory homepage: http://www2.eur.nl/fgg/ch1/cellbiology/patrinos
International HapMap Project: http://www.hapmap.org
International Immunogenetics Project HLA database: http://www.ebi.ac.uk/imgt/hla

SUPPLEMENTARY INFORMATION
See online article: S1 (box)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF.