The 1.4-Mb CMT1A Duplication/HNPP Deletion Genomic Region Reveals Unique Genome Architectural Features and Provides Insights into the Recent Evolution of New Genes

Ken Inoue,[1] Ken Dewar,[3] Nicholas Katsanis,[1] Lawrence T. Reiter,[1,4] Eric S. Lander,[3] Keri L. Devon,[3] Dudley W. Wyman,[3] James R. Lupski,[1,2,5] and Bruce Birren[3]

[1]Department of Molecular and Human Genetics, and [2]Department of Pediatrics, Baylor College of Medicine, Houston, Texas 77030, USA; [3]Whitehead Institute for Biomedical Research/MIT Center for Genome Research, Cambridge, Massachusetts 02141, USA

Duplication and deletion of the 1.4-Mb region in 17p12 that is delimited by two 24-kb low copy number repeats (CMT1A–REPs) represent frequent genomic rearrangements resulting in two common inherited peripheral neuropathies, Charcot-Marie-Tooth disease type 1A (CMT1A) and hereditary neuropathy with liability to pressure palsy (HNPP). CMT1A and HNPP exemplify a paradigm for genomic disorders wherein unique genome architectural features result in susceptibility to DNA rearrangements that cause disease. A gene within the 1.4-Mb region, *PMP22*, is responsible for these disorders through a gene-dosage effect in the heterozygous duplication or deletion. However, the genomic structure of the 1.4-Mb region, including other genes contained within the rearranged genomic segment, remains essentially uncharacterized. To delineate genomic structural features, investigate higher-order genomic architecture, and identify genes in this region, we constructed PAC and BAC contigs and determined the complete nucleotide sequence. This CMT1A/HNPP genomic segment contains 1,421,129 bp of DNA. A low copy number repeat (LCR) was identified, with one copy inside and two copies outside of the 1.4-Mb region. Comparison between physical and genetic maps revealed a striking difference in recombination rates between the sexes with a lower recombination frequency in males (0.67 cM/Mb) versus females (5.5 cM/Mb). Hypothetically, this low recombination frequency in males may enable a chromosomal misalignment at proximal and distal CMT1A–REPs and promote unequal crossing over, which occurs 10 times more frequently in male meiosis. In addition to three previously described genes, five new genes (*TEKT3, HS3ST3B1, NPD008/CGI–148, CDRT1,* and *CDRT15*) and 13 predicted genes were identified. Most of these predicted genes are expressed only in embryonic stages. Analyses of the genomic region adjacent to proximal CMT1A–REP indicated an evolutionary mechanism for the formation of proximal CMT1A–REP and the creation of novel genes by DNA rearrangement during primate speciation.

Submicroscopic duplications/deletions represent genomic rearrangements that can be responsible for inherited diseases. These are not visible by conventional karyotype assays and are thus likely to involve rearranged fragments smaller than 1–2 Mb. Disorders with these types of rearrangements may be caused by dosage effects of a single or multiple genes. Inherited diseases resulting from such genomic rearrangement may be categorized as genomic disorders in contrast to classic Mendelian diseases caused by point mutations in the causative genes (for review, see Lupski 1998b; Shaffer and Lupski 2000).

Charcot-Marie-Tooth disease type 1A (CMT1A) is one of the first and best-characterized examples of a submicroscopic genomic disorder. CMT1A is the most common inherited peripheral neuropathy and accounts for 70% of CMT type 1 inherited demyelinating neuropathy (for review, see Lupski and Garcia 2001). Molecular genetic approaches have identified a submicroscopic duplication of the 1.4-Mb genomic region in chromosome band 17p12 in the majority of the CMT1A cases (Lupski et al. 1991; Raeymaekers et al. 1991; Wise et al. 1993; Nelis et al. 1996; Roa et al. 1996). A submicroscopic deletion of the same region results in hereditary neuropathy with liability to pressure palsy (HNPP), a distinct form of inherited peripheral neuropathy with episodic and milder manifestations (Chance et al. 1993, 1994). The CMT1A duplication and HNPP deletion represent products of unequal crossing over and a reciprocal recombination between flanking 24-kb homologous sequences termed CMT1A–REPs (Lupski 1998a). Subsequently, a gene encoding PMP22, a major component of the peripheral

nervous system myelin, was mapped in the middle of this 1.4-Mb region (Matsunami et al. 1992; Patel et al. 1992; Timmerman et al. 1992; Valentijn et al. 1992). Several lines of evidence indicate that gain of one copy of *PMP22* is responsible for CMT1A, whereas loss of one copy of *PMP22* results in HNPP through a *PMP22* gene dosage effect as the mechanism for these disorders (Lupski et al. 1992).

Although duplication and deletion of *PMP22* is the event responsible for CMT1A and HNPP, respectively, as many as 30 to 50 other genes may be contained in this 1.4-Mb region on the basis of its genomic size (Murakami et al. 1997b). A question remains as to why only *PMP22* is dosage sensitive, whereas other genes in the region are apparently not. In addition, the clinical phenotypes of patients having the same 1.4-Mb duplication are quite variable. A formal possibility exists that minor dosage effect of genes other than *PMP22* in this 1.4-Mb region somehow contribute to the variability of phenotypic manifestations or a combination of phenotypes (e.g., CMT + connective tissue disorder). Furthermore, there are rare case reports of smaller duplications (Ionasescu et al. 1993; Palau et al. 1993; Valentijn et al. 1993) or deletion (Chapon et al. 1996), raising the question as to whether such rare recombination events are mediated by other repeat units in this region.

To characterize the genomic architecture of this region, we constructed PAC and BAC contigs and produced a finished sequence across this 1.4-Mb interval. We defined a 1,421,129-bp genomic interval as the CMT1A duplication/HNPP deletion region. Here we report the identification of low-copy number repeats (LCRs), the comparison of genetic and physical maps, the identification and characterization of genes, and a mechanism for the evolution of new mammalian genes by DNA rearrangements.

RESULTS

Sequencing the 1.4-Mb CMT1A Duplication/HNPP Deletion Region

A contig of overlapping bacterial clones was constructed on the basis of marker content by use of pre-existing and newly generated STSs. Restriction fragment fingerprinting (Marra et al. 1997) verified the order of clones within the contig and identified a set of minimally overlapping BAC and PAC-tiling path of clones for genomic characterization. Individual clones were subjected to shotgun sequencing, assembly, and finishing. A path of 12 overlapping clones contains the complete region bounded by the CMT1A–REPs, and this is part of a larger 15-clone path analyzed in this study (Fig. 1). Previously, we have predicted the size of
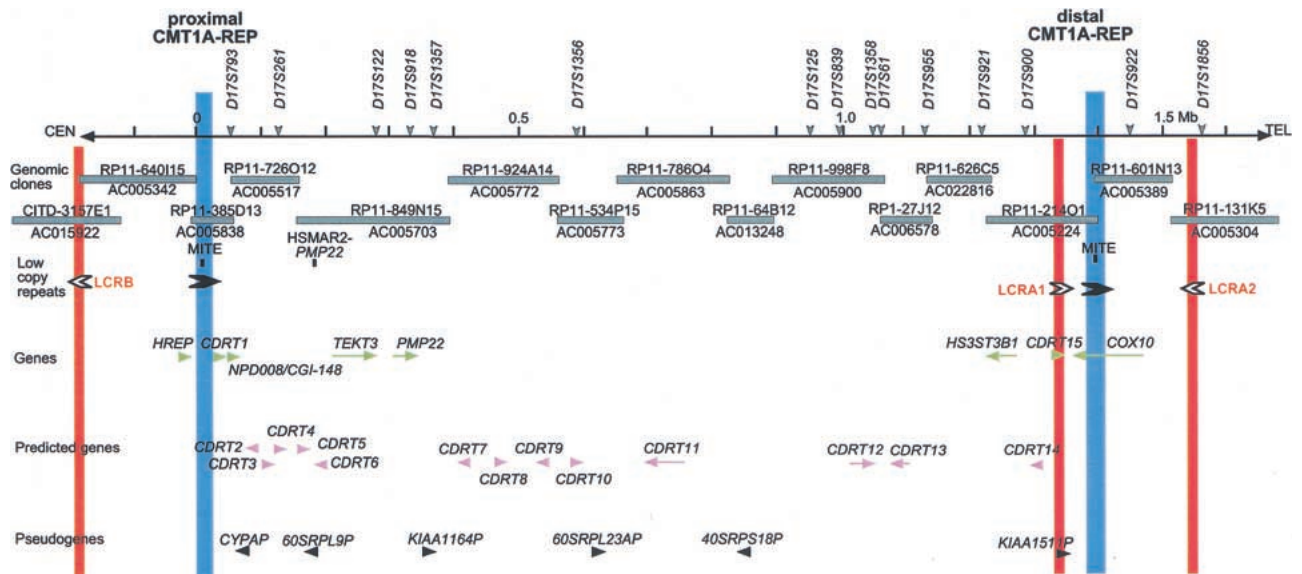

Figure 1 The genomic sequence map of the CMT1A duplication/HNPP deletion region in 17p12. The *top* solid horizontal line represents the genomic sequence of the CMT1A/HNPP region in the centromere to telomere orientation. Position 0 is assigned to the first base of the proximal CMT1A–REP and vertical markings are placed every 100 kb for reference. The STR polymorphic genetic markers are shown *above.* Shaded horizontal boxes *below* depict the large insert clones used to derive genomic sequences. Clones are identified by their individual names and GenBank accession numbers. Proximal and distal CMT1A–REPs are shown as vertical blue boxes, and newly identified low copy repeats (LCRA1, LCRA2, LCRB) are represented as vertical red bars. Arrowheads indicate the orientation/direction of each repeat unit. Underneath are shown known genes (green), predicted genes (purple), and pseudogenes (black), with arrows pointing in the direction of transcription. (MITE and HSMAR2-*PMP22*) *mariner* transposon-like elements; (*CYPAP*) cyclophilin A pseudogene; (*60SRPL9P*) 60S ribosomal protein L9 pseudogene; (*60SRPL23AP*) 60S ribosomal protein L23A pseudogene; (*40SRPS18P*) 40S ribosomal protein S18 pseudogene.

this genomic region to be 1.5 Mb on the basis of physical mapping data obtained by pulsed-field gel electrophoresis (PFGE) and Southern blotting analyses (Pentao et al. 1992). Our completed sequence indicates that the entire region from the first nucleotide of the proximal CMT1A–REP to the last nucleotide of the distal CMT1A–REP is 1,421,129 bp.

Repetitive Elements

`RepeatMasker` indicates that high copy number retrotransposable elements and simple tandem repeats (STRs) account for 43.37% of the entire CMT1A/HNPP region (Table 1). The repetitive elements consist of 9.97% *Alu* sequences and 13.43% LINE1 elements, which is comparable in distribution with that of chromosome 21, but in contrast to that of chromosome 22, which contains 16.8% of *Alu* sequences and 9.73% of LINE1 elements (Dunham et al. 1999; Hattori et al. 2000).

There is a *mariner* insect transposon-like element 140-kb centromeric to *PMP22*, termed HSMAR2–PMP22 (Fig 1). This *mariner* element is interrupted by an insertion of an *Alu* element, indicating that it is no longer active. However, we observed both 5′ and 3′ inverted terminal repeats (ITRs), suggesting that this *mariner* element has the potential to act as a *cis*-acting substrate to promote double-strand DNA breakage (Reiter et al. 1996, 1999).

We identified 53 STRs with repeating units >11.

Nine STRs (*D17S793*, *D17S261*, *D17S122*, *D17S1357*, *D17S1356*, *D17S839*, *D17S1358*, *D17S955*, and *D17S921*) were mapped previously to this region, two (*D17S918* and *D17S900*) were mapped to the region but not known to be within the CMT1A/HNPP interval, and forty-two represent newly identified potential polymorphic markers. The new STRs include 26 dinucleotide [21 $(CA)_n$, 2 $(GA)_n$, 1 $(TA)_n$, 1 $(TA)_n(CA)_n$, and 1 $(TG)_n(GA)_n$], 2 trinucleotide [2 $(CAA)_n$], 10 tetranucleotide [6 $(TTTA)_n$ and 4 $(TTTC)_n$], and 4 pentanucleotide [1 $(TTTTC)_n$, 1 $(CAATA)_n$, 1 $(CGATA)_n$, and 1 $(TTTTA)_n$] elements. Fifteen of these STRs have been shown to reveal significant polymorphic variation in different ethnic populations (Badano et al. 2001).

Low Copy Repeats: An 11-kb Element

In addition to the previously defined CMT1A–REPs (24,011 bp of 98.7% nucleotide identity, Reiter et al. 1997), other low copy repeats were identified (Fig.1). LCRA1 and LCRA2, located 32-kb centromeric and 140-kb telomeric to the distal CMT1A–REP in inverted orientaion, are highly similar 11-kb low copy number repeat segments. We also found a 4-kb truncated copy of this repeat, termed LCRB, ~180 kb centromeric to the proximal CMT1A–REP (Fig. 1). Therefore, one copy of this repeat is located within the 1.4-Mb region and the other two are located outside of this region. LCRA1 and LCRA2 are highly similar throughout the 11 kb (97% identity), whereas LCRB aligns only with a 4-kb interior portion (95% identity to LCRA1) (Fig. 2A). Further sequence comparisons revealed one small region (132 bp) that represents DNA rearrangements between these LCRs (Fig. 2A). LCRA1 contains three contiguous fragments (25, 89, and 18 bp) that involve small tandem repeat units (14- and 9-bp monomer). The corresponding region in LCRA2 contains a duplication of the 25-bp monomer as well as a deletion of the 18-bp fragment, probably resulting from polymerase slippage at the 14- and 9-bp repeat units flanking these 25- and 18-bp fragments, respectively, in LCRA1. Furthermore, the recombination breakpoint of the LCRB is located in this small region between the 14- and 9-bp repeat units, resulting in truncation of the 89-bp fragment and loss of the 25-bp fragment. No 18-bp deletion was found in

Table 1. The Interspersed Repeat Content of the CMT1A/HNPP Region

Repeat type	Total no. of elements	Coverage (bp)	Coverage (%)
SINEs	**703**	**168,239 bp**	**11.84%**
*Alu*s	511	141,709 bp	9.97%
MIRs	192	26,530 bp	1.87%
LINEs	**463**	**262,132 bp**	**18.45%**
LINE1	245	190,833 bp	13.43%
LINE2	190	59,880 bp	4.21%
Other LINEs	28	11,419 bp	0.81%
LTR elements	**246**	**115,140 bp**	**8.10%**
MaLRs	120	49,012 bp	3.45%
Retroviral	50	21,746 bp	1.53%
MER4 group	43	29,094 bp	2.05%
Other LTRs	33	15,288 bp	1.07%
DNA elements	**194**	**48,268 bp**	**3.40%**
MER1 type	135	28,178 bp	1.98%
MER2 type	26	11,345 bp	0.80%
Mariners	8	4,332 bp	0.30%
Other DNA elements	25	4,413 bp	0.32%
Unclassified	**2**	**319 bp**	**0.02%**
Total interspersed repeats	**1608**	**594,098 bp**	**41.80%**
Small RNA	5	668 bp	0.05%
Simple repeats	247	13,610 bp	0.96%
Low complexity	144	8,140 bp	0.57%
Total	2004	616,328 bp	43.37%
Total sequence length	1,421,129 bp		
GC %	41.57%		

the LCRB. This genomic evidence indicated that the LCRA1 is likely the progenitor and the other two LCRs are derivatives of LCRA1. A duplication event that results in LCRB may have been followed by another duplication that generated LCRA2.

Searches of the high throughput human genome sequence revealed the presence of multiple copies of this LCR in the genome. After elimination of the highly repetitive 4.4-kb flanking sequences from this 11-kb fragment, BLAST searches with the 6.6-kb region identified 29 BAC clones assigned to 9 different chromosomes; 1, 4, 8, 9, 11, 13, 16, 17, and 22 (data not shown). Electronic PCR analyses (Schuler 1998) of each BAC clone showed STSs from multiple chromosomes matching a single BAC sequence, whereas the 11-kb

LCRA1 only contains a chromosome 17-specific STS, suggesting the repeat structures involving these loci in the genome are complex. Further mapping and characterization are required to elucidate the nature of these repeat structures involving multiple loci in the genome.

BLAST searches of this 6.6-kb region against the human EST database revealed a number of clones homologous to this portion of the LCRA1 low copy repeat. There are two different genes or groups of genes; one homologous to the 3kb–4kb region from the centromeric side (named *CDRT15*, see details in the following section) and the other to the 4.5–6.3-kb region. Further database searching revealed that the latter is a processed pseudogene of *KIAA1511*, which encodes a protein of unknown function and maps to chromosome 1 (GenBank accession no. AB040944). Interestingly, ESTs belonging to the former group have various levels of homology, suggesting that these ESTs may be transcribed from multiple loci in the genome. Further sequence comparison of these EST clones to the genomic sequence database mapped them to at least nine different genomic loci.

Figure 2 Genomic structure of LCRA1, LCRA2, and LCRB. (*A*) The genomic architecture of three LCRs is shown. High copy number retrotransposable elements including *Alu*, L1, MaLR, and MER2 type DNA elements, which are conserved between these LCR, are boxed in gray. Each exon of *CDRT15* is shown as a solid black box. A small genomic region in the 5′ upstream of *CDRT15* exon I representing DNA rearrangements (enlarged) between the three LCRs. The LCRA1 has a 132-bp region that is further divided into 25 bp (blue), 89 bp (pink) and 18 bp (green) segments. There are short tandem repeat sequences, 14 bp (orange) and 9 bp (red), flanking these subdivided fragments. In comparison with the same region in the LCRA2, the 25-bp monomer is tandemly duplicated and the 18-bp fragment is deleted in LCRA2. Furthermore, the distal boundary of 4.4-kb LCRB is located in this short genomic region between 14- and 9-bp repeat (arrow). These 18 bp are present in LCRB. (*B*) Hypothetical inversion involving LCRA1 and LCRA2 results in flipping of distal CMT1A–REP. The proximal and distal CMT1A–REPs are depicted by thick black arrows with their relative orientation given by the directions of arrows. LCRA1 and LCRA2 are depicted as open arrows. The genome architecture as determined by sequencing this specific genomic library alone placed the CMT1A–REPs in a direct orientation and thus the region in between is susceptible to duplication/deletion. *Below* is shown the hypothetical orientation of CMT1A–REPs if an inversion occurs via homologous recombination using LCRA1 and LCRA2 as substrates. Note that this orientation of CMT1A–REPs will prevent the formation of duplication/deletion event (Lupski 1998b).

Comparison between the Physical and Genetic Maps

In previous efforts to identify the *CMT1A* gene by linkage analysis, the CMT1A region was estimated to be much larger than 1.4 Mb on the basis of the genetic distance between linked markers (Patel et al. 1990; Timmerman et al. 1990). However, subsequent physical mapping with PFGE and YAC-based STS content mapping revealed a physical size of 1.5 Mb (Pentao et al. 1992). One hypothesis to explain the observed discrepancy between genetic and physical distances has been that a potential recombination hotspot exists within the CMT1A genomic region in addition to the positional recombination hotspot located within CMT1A–REP (Reiter et al. 1996). To evaluate the actual recombination frequency, we
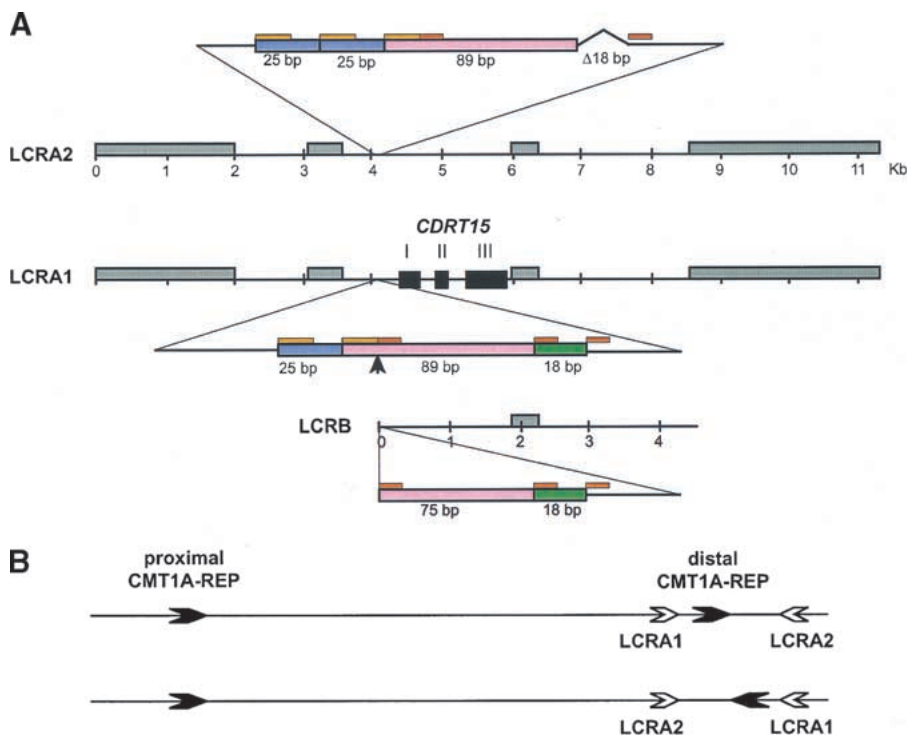
systematically compared the genetic map and genome sequence-based physical map of the CMT1A duplication/HNPP deletion region by integrating the Marshfield genetic mapping data into our physical map. Eight polymorphic microsatellite markers (*D17S900*, *D17S921*, *D17S955*, *D17S839*, *D17S918*, *D17S122*, *D17S261*, and *D17S793*) were found in both the Marshfield genetic map and genomic sequence from the 1.4-Mb region (Broman et al. 1998). Of these, two markers (*D17S900* and *D17S918*) were not mapped inside this region in the previous physical maps (Murakami and Lupski 1996; Boerkoel et al. 1999). Three markers identified previously in the CMT1A region (*D17S1356*, *D17S1357*, and *D17S1358*) were not included in the Marshfield study (Blair et al. 1995).

We generated a genetic/physical map correlation (Fig. 3A) and compared it with the flanking 1.5-Mb regions. Physical distances in the proximal regions include estimates based on BAC physical mapping data at 100-Kb resolution on the centromeric side (J.R. Lupski and B. Birren, unpubl.) and fully finished sequence on the telomeric side. These genetic/physical map comparisons indicate that the recombination frequency of an at least 4.5-Mb region including the CMT1A duplication/HNPP deletion region is low in males. In sharp contrast, this region recombines frequently in females. The cM/Mb ratio of the entire 4.5-Mb region is 5.5 for female, 0.67 for male, and 3.3 for the sex-averaged map. As a result of this contrast, this region has a high female/male recombination frequency ratio, which is steeply increasing toward the centromere (Fig. 3B). Neither CMT1A–REP regions nor the entire CMT1A/HNPP region have a higher recombination frequency than flanking regions. The 820-kb region between *D17S1843* and *D17S918*, which spans the proximal CMT1A–REP, revealed no recombination in the families examined in both male and female meiosis (Broman et al. 1998). There is also a low recombination region in both sexes telomeric to distal CMT1A–REP for >1 Mb.

Genes in the 1.4-Mb Region

Sequence analysis was performed by the use of NIX (nucleotide identification of unknown sequences), which incorporates a number of independent gene prediction

tools (Fig. 1; Table 2). Each gene was further characterized by additional database searches and expression analyses. We categorized the genes into three groups; (I) genes for which we have biological evidence including cDNA sequences, gene structures, similarity to other genes, or multiple spliced ESTs, matching gene predictions with complete gene structure; (II) predicted
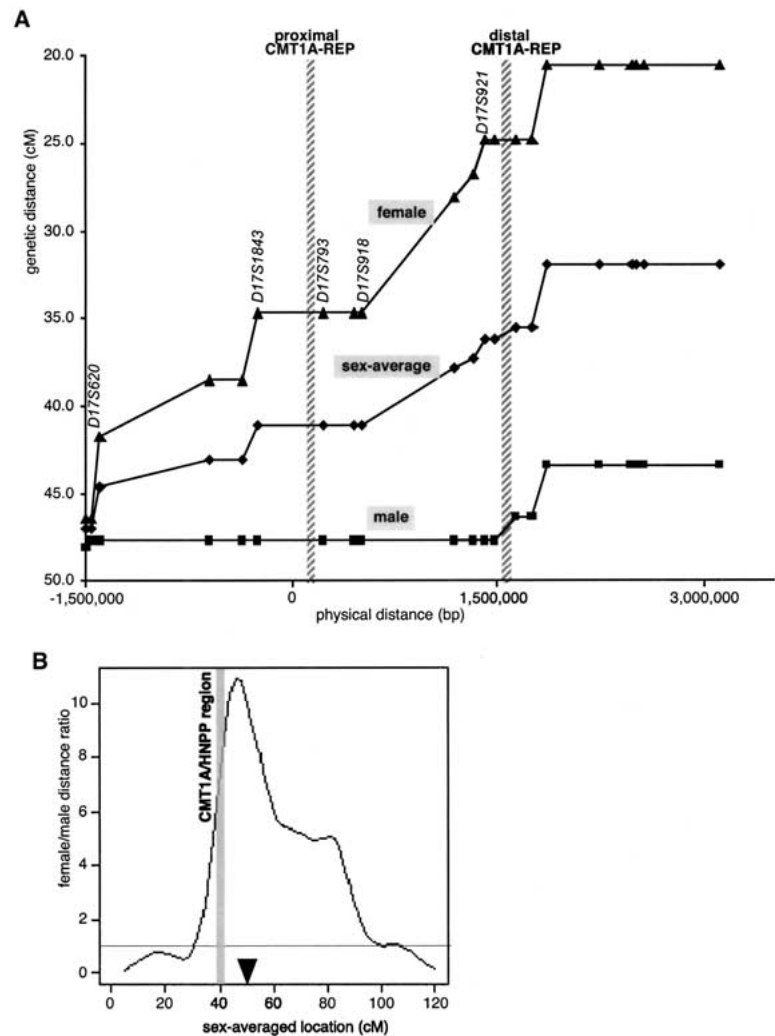

Figure 3 Sex-specific recombination frequencies in the CMT1A/HNPP genomic region. (*A*) The relationship between genetic and physical distance. The STR markers in the 17p12 CMT1A/HNPP region from the Marshfield genetic map were aligned to the nucleotide sequence-based physical map. The marker order is as follows: centromere-*D17S1794-D17S620-D17S2196-D17S1857-D17S953-D17S1843-D17S793-D17S122-D17S918-D17S839-D17S955-D17S921-D17S900-D17S922-D17S1856-D17S947-D17S936-D17S639-D17S799-D17S1808-D17S1803*-telomere (markers within the CMT1A/HNPP genomic region are underlined). Both CMT1A–REPs are shown as hatched bars. As the sequencing of the centromeric side of the map has not yet been completed, the physical distance was calculated on the basis of the BAC contig map (K. Inoue, K. Dewar, N. Katsanis, L.T. Reiter, E.S. Lander, K.L. Devon, D.W. Wyman, J.R. Lupski, and B. Birren, unpubl.). (*B*) Female/male distance ratio (*vertical* axis) is plotted along the sex-averaged genetic map (*horizontal* axis). The histogram was obtained from the Marshfield Center for Genetics (http://research.marshfieldclinic.org/genetics/Map_Markers/maps/IndexMapFrames.html). The CMT1A/HNPP genomic region is shown as a shaded vertical bar. (▼) The predicted position of the centromere.

Table 2. Summary of the Genes, Predicted Genes, and Pseudogenes in the CMT1A/HNPP Region

Group	Gene	UniGene/ESTs accession no.	Exons	Core promoter	Poly-A signal	Kozak consensus	mRNA	Predicted amino acids	Northern/RT-PCR profiles	Domain	Other features
Genes	PMP22	L03203	6	+	+	+	1716 bp	160			(Resemberg et al. 1997)
	HS3ST3B1	AF105377	2	+	+	+	2032 bp	290	N/ubiqitous		No domain
	NPD008/CGI-148	Hs.6776	7	+	+	+	1878 bp	205			
	TEKT3	AF334676	8	+	+	+	1707 bp	490	testis	tektin	
	CDRT1	AF337810	1	weak	+	−	800 bp	243	N/pancreas		No domain
	CDRT15	AF355097	3	weak	+	−	778 bp	188	adult(−) fetus(+)		No domain
Predicted genes	CDRT2	No EST	4	N/A	+	−	591 bp	196	adult(−) fetus(+)		No domain
	CDRT3	Hs.147742	4	N/A	+	N/A	N/A	N/A	adult(−) fetus(+)	N/A	
	CDRT4	Hs.164595	>1	N/A	+	N/A	>428 bp	>141	adult(−) fetus(+)		No domain
	CDRT5	Hs.199583	3	N/A	+	+	807 bp	210	adult(−) fetus(+)		No domain
	CDRT6	AL037161	N/A	N/A	N/A	N/A	>653 bp	N/A	adult(−) fetus(+)	N/A	
	CDRT7	Hs.147654	3	N/A	+	N/A	>1035 bp	179	N/A		No domain
	CDRT8	Hs.98684	N/A	N/A	+	N/A	N/A	N/A	N/pancreas	N/A	
	CDRT9	AI220152	N/A	N/A	+	N/A	N/A	N/A	N/low ubiqitous	N/A	
	CDRT10	N/A	4/5	N/A	+	−	720/755 bp	117/150	N/Sk.muscle	N/A	
	CDRT11	Hs.98605	>4	N/A	+	N/A	>878 bp	N/A	N/A		No domain
	CDRT12	N/A	7	N/A	+	+	2410 bp	257	N/low pancreas		No domain
	CDRT13	N/A	8	N/A	+	−	973 bp	154	N/A		No domain
	CDRT14	AI678032	>3	N/A	N/A	N/A	>560 bp	>81	adult(−) fetus(+)		No domain
Pseudogenes	CYPAP										Processed/rearranged
	60SRPL9P										Processed
	KIAA1164P										Processed/rearranged
	60SRPL23AP										Processed
	40SRPS18P										Processed
	KIAA1511P										Processed

Genes, predicted genes, and pseudogenes are listed accordingly. GenBank accession nos. for UniGene (if possible), or one most representative EST, are listed. Core promoter activities, poly-A signals, and functional domains were determined as described in Methods. Northern blotting assay was used primarily for the expresion studies, as noted with a capital N. In cases in which no signal was detected, RT-PCR assays were performed. (N/A) not available for determination.

genes with limited information such as multiple EST matches and/or predicted exonic structures, but complete gene structural information is not available, and; (III) pseudogenes. Overall, we identified 21 genes or predicted genes (Groups I and II) in this region.

Genes

Of the eight genes in this group, four are known: *HREP*, *PMP22*, *HS3ST3B1*, and *COX10*. Of these, *HS3ST3B1* is the only gene newly mapped to this region. *COX10* and *HREP* are located in the CMT1A–REP regions in which complete sequence data were available previously (Reiter et al. 1997; Kennerson et al. 1997, 1998; Murakami et al. 1997a). We thus describe the genomic structures of *PMP22* and *HS3ST3B1* in further detail. Four previously unknown genes were also identified, *NPD008/CGI-148*, *tektin3* (*TEKT3*), *CDRT1* (CMT1A duplicated region transcript 1), and *CDRT15*.

PMP22

PMP22, the gene responsible for CMT1A and HNPP, has four coding exons and two alternatively utilized exons I (Suter et al. 1994; Sabéran-Djoneidi et al. 2000). *PMP22* spans 35 kb and is transcribed toward the telomere (Fig. 4A). One trinucleotide repeat sequence was found in intron 3, which matched a previously known STR, *D17S918*. This STR contains 12 CAG repeats.

HS3ST3B1

The cDNA sequence for *HS3ST3B1* (heparan sulfate D-glucosaminyl) 3-O-sulfotransferase 3B1) was described previously, but the genomic structure was unknown (Shworak et al. 1999). *HS3ST3B1* encodes a 390 amino acid enzyme that catalyzes sulfation of heparan sulfate (Liu et al. 1999) and contains two coding exons that are separated by a 43-kb intron (Fig. 4B). *HS3ST3B1* is part of a gene family that includes *HS3ST1*, *HS3ST2*, *HS3ST3A1*, *HS3ST3B1*, and *HS3ST4* (Shworak et al. 1999). Both *HS3ST3A1* and *HS3ST3B1*, which are highly similar within their sulfotransferase domains (99.2%), map to 17p12. *HS3ST3A1* is located ~700-kb telomeric to *HS3ST3B1*, and these genes flank the distal CMT1A–REP (Fig. 4B). *HS3ST3A1* also contains two coding exons separated by a large 100-kb intron, but is transcribed in the opposite direction. Nucleotide sequence analysis of the 100-kb *HS3ST3A1* intron and the 43-kb *HS3ST3B1* intron revealed no homology, suggesting that these genomic regions are not conserved between the two genes, and thus, if these genes arose through duplication, the event is evolutionarily very ancient.

NPD008/CGI–148

This transcript has a 615-bp ORF, encoding a predicted 205 amino acid protein. The structure of this gene is shown in Figure 4C. The cDNA sequence reveals an almost complete match with two genes in the data-
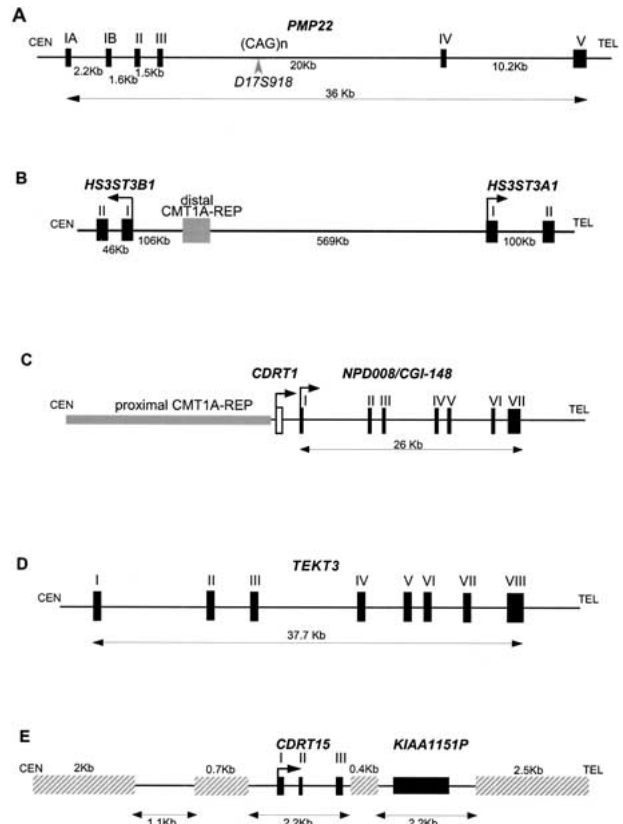

Figure 4 Genomic structure of the genes in the CMT1A/HNPP region. (*A*) *PMP22*. Arrowhead indicates marker *D17S918*, which contains polymorphic CAG repeats. Distances for each intron are shown at *bottom*, whereas individual exons are numbered at *top*. We hypothesized that in rare families with CMT accompanied by anticipation, it may be related to an expanded allele of this triplet repeat. To test this hypothesis, we obtained DNA samples from one such family for which the disease locus mapped to 17p11.2–17p12 by linkage analysis (Kovach et al. 1999). We examined the number of CAG repeat in members of this family, but failed to identify any expansion of the triplet repeats in affected individuals (data not shown). A point mutation in *PMP22* was subsequently found to segregate with the disease phenotype in this family (Kovach et al. 1999). (*B*) *HS3ST3B1* and *HS3ST3A1*. *HS3STB1* is located inside the CMT1A/HNPP genomic region, whereas *HS3STA1* is 569-kb telomeric to the distal CMT1A–REP. Each exon of the two genes is indicated as a box, and arrows show the direction of the transcription. (*C*) *CDRT1* and *NPD008/CGI-148*. The horizontal shaded rectangle indicates the proximal CMT1A–REP. An open box shows the single exon of *CDRT1* and solid boxes indicate seven exons of *NPD008/CGI-148* that span 26 kb. (*D*) *TEKT3* contains eight exons (solid boxes) including an untranslated first exon. The putative initiating methionine is in exon II. The exon/intron boundaries from exons III to VI, which were not determined by database analyses, were experimentally confirmed by RT–PCR and sequence analyses using single-stranded testis cDNA as the template DNA. (*E*) *CDRT15* in the 11-kb low-copy repeat unit, LCRA1. Hatched horizontal bars indicate repetitive elements. Pseudogene for *KIAA1151* is also shown.

base, *NPD008* (GenBank accession no. AF223467) and *CGI-148* (GenBank accession no. AF151906). *NPD008* was isolated from pituitary glands, whereas *CGI-148* was reconstructed by a comparative EST database

search between human and *Caenorhabditis elegans* (Lai et al. 2000). Ubiquitous expression was observed by Northern blotting and RT–PCR analyses (Fig. 5A,B), but embryonic tissues showed higher expression levels, except for the brain. Database searches identified putative orthologs in various species, including *Drosophila*

Figure 5 Expression studies of genes identified in the CMT1A/HNPP region. (*A*) Multiple-tissue Northern blot analyses. Tissues are indicated at *top* of each lane (*He*) Heart; (*Br*) brain; (*Pl*) placenta; (*Lu*) lung; (*Li*) liver; (*Mu*) skeletal muscle; (*Ki*) kidney; (*Pa*) pancreas. Marker sizes are 9.5, 7.5, 4.4, 2.4, and 1.35 kb. (*B*) Multiple-tissue RT–PCR analyses. *NP008/CGI-148* is expressed in a wide variety of tissues. A 2-kb major transcript and two minor transcripts are observed. *TEKT3* revealed no expression by Northern blotting (data not show), but high expression in testis by RT–PCR. Faint expression is also observed in both ovary and pancreas by RT–PCR. Fetal tissues reveal low level but ubiquitous expression. *CDRT1* shows a major 2-kb and a minor 1-kb transcripts in pancreas. Faint expression is also identified in the heart. *CDRT8* reveals two short transcript (0.7 and 1 kb) in the pancreas. A 3-kb transcript of *CDRT9* represents low-level ubiquitous expression. *CDRT10* is expressed in the skeletal muscle as a 1.3-kb transcript. *CDRT12* reveals a 2.8-kb transcript in the pancreas at very low expression levels. *CDRT2*, *CDRT3*, *CDRT4*, *CDRT5*, *CDRT6*, *CDRT14*, and *CDRT15* reveal no expression in adult tissues by Northern blotting (data not shown) and RT–PCR analyses, but show obvious expression in fetal tissues.

melanogaster, *C. elegans*, *Schizosaccharomyces pombe*, *Saccharomyces cervisiae*, and *Arabidopsis thaliana*, but no information is available with regard to function. A mouse ortholog was reconstructed from EST sequences and was found to encode a predicted protein of 205 amino acids with 91% identity to the human protein.

The *NPD008/CGI-148* gene product is likely a membrane-bound protein with three possible transmembrame domains. The orthologs in other species are likely to have a similar structure. Through GenBank searches, we also identified three processed pseudogenes on 2p13, 7q21– 7q22, and 16p13.

TEKT3

TEKT3 (*Tektin3*), located 50-kb centromeric to *PMP22*, spans 37.7 kb (Fig. 4D). Its eight exons encode a 490 amino acid protein with significant homology to the tektin protein families. The closest homology was to the sea urchin protein, tektin A1, suggesting that this gene is likely to encode a human ortholog for tektin A1, termed *TEKT3*. As observed in other members of the tektin family, TEKT3 also has a highly conserved tektin domain, RSNVELCRD (underlined residues were conserved in TEKT3) (Norrander et al. 1998; Iguchi et al. 1999). Although Northern blotting analysis failed to show expression in an 8-tissue panel (data not shown), extensive RT–PCR-based expression studies revealed that *TEKT3* is primarily expressed in adult testis with low-level widespread expression observed in embryonic tissues (Fig. 5B).

CDRT1

CDRT1 is located 1.3-kb telomeric to proximal CMT1A–REP (Fig. 4C). Multiple human and mouse EST alignments reveal a single exon gene encoding a 243 amino acid protein with unknown function. The upstream 1.3-kb region has weak but potential promoter sequence motifs estimated by the promoter prediction programs TSSW and NNPP. Northern blotting identified a major 2-kb and a minor 1-kb transcript in the pancreas and a faint 2-kb transcript in the heart (Fig. 5A). Further evolutionary analysis of this gene is described in a subsequent section.

CDRT15

CDRT15 is located within the LCRA1. The 778-bp cDNA sequence is divided into three exons, encoding an 188 amino acid protein of unknown function (Fig. 4E). As mentioned above, there are at least eight paralogous copies of this gene in the human genome. Submitted sequences include one full-length cDNA clone encoding an unknown protein (GenBank accession no.
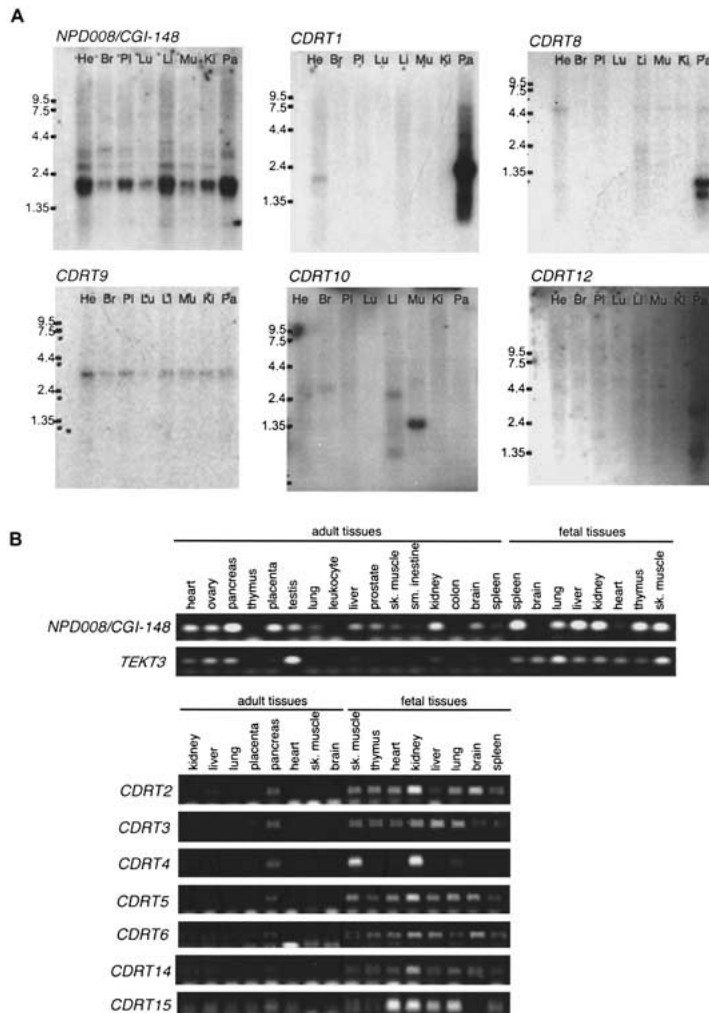
AF038169) and numerous partial sequences. We reconstructed complete coding cDNA sequences by aligning these ESTs with each other. At least three cDNA clones were found to contain ORFs with possible exon/intron structures. Interestingly, they have insertion/deletion mutations that result in frameshifts of the ORF, thus encoding totally different proteins; others have insertions/deletions that appear to result in early termination. It is not clear which gene copies are producing functional proteins and which are transcribed pseudogenes.

Predicted Genes

We identified 13 predicted genes (Fig. 1; Table 2). Each of these has incomplete information to determine full-length cDNA sequence. However, substantive evidence, including matching UniGene clusters, matching ESTs with intron structure, and significant scores by gene prediction programs, suggest these represent bona fide genes. Interestingly, Northern blotting analyses of these genes by use of an adult tissue panel revealed minimal expression, whereas RT–PCR analysis indicated substantial expression in embryonic tissues (Fig. 5). Results of the database and expression analyses for these 13 genes are summarized in Table 2.

Pseudogenes

Six pseudogenes were identified in the CMT1A/HNPP region (Fig. 1; Table 2). Each locus reveals evidence for absent introns and disrupted coding sequence by mutations, suggesting that they are processed pseudogenes. The pseudogene for *cyclophilin A* (*CYPAP*) revealed deletion of a region corresponding to the first 180 bp of cDNA sequence. The pseudogene for *KIAA1164* showed deletion for the first 2 kb of original 4 kb cDNA, inversion of a 1-kb region, and insertion of an L1 element.

Evolution of New Genes by DNA Rearrangement During Speciation: Origin of *HREP* and *CDRT1*

Database searches to identify mouse orthologs of human genes in this region provided evidence of an additional ancestral rearrangement with functional consequences. Searches with human *CDRT1* sequences identified mouse ESTs with coding sequences extending 5′ upstream from the initiation site for the human gene (Fig. 6A). Human sequence corresponding to this 5′ extension is not found in the genomic sequence from the *CDRT1* region. In fact, the mouse EST sequences that extend 298-bp 5′ from the start of the human *CDRT1* gene do not match any sequence in the human genome. However, additional sequences further 5′ in the mouse EST contig show similarity to the
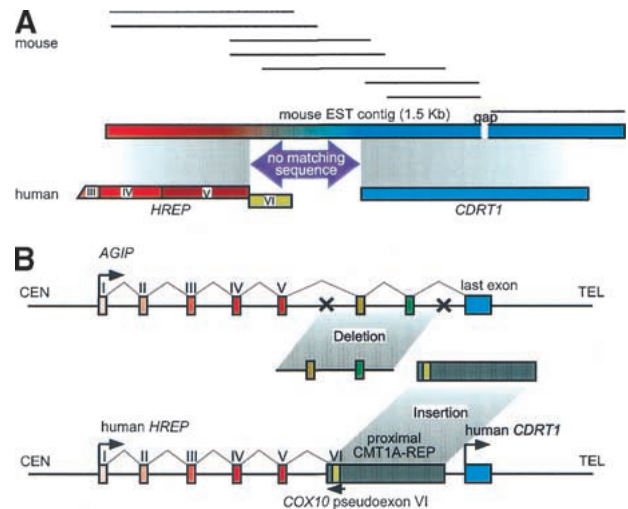

Figure 6 Gene evolution surrounding the proximal CMT1A–REP region. (*A*) A comparison of EST contigs between mouse and human. Eight mouse ESTs (shown as solid horizontal bars; GenBank accession nos. AI606691, AA089107, AA982328, AI181334, AI882388, AI429210, AI447154, and AI506955, from *top* to *bottom*) were reconstructed into a 1.5-kb contig with a 24-bp gap (horizontal rectangle with gradient colors) that aligns with two human genes, *HREP* (the numbers represent each exon of *HREP*; the exon VI does not align with the mouse EST contig) and *CDRT1*. Between the alignment of these two genes, there is a 269-bp region in the mouse clone that does not match any human sequence (purple arrow). The conceptual translation of this region does not identify a known protein functional motif. (*B*) A model for the evolution of new genes and the genomic structure surrounding the proximal CMT1A–REP region. *Top* figure represents the genomic structure of a hypothetical ancient gene *AGIP* (Ancestral Gene before the Integration of Proximal CMT1A–REP) modeled in mice. One or more exons originally contained in the *AGIP* are predicted to be lost by the integration of the proximal CMT1A–REP. *Bottom* figure shows human genomic structure in which *HREP* and *CDRT1* are separated by the inserted proximal CMT1A–REP (dark rectangle). The pseudoexon of *COX10* is utilized as the last exon of *HREP* from the opposite direction (green box).

human *HREP* gene. The human *HREP* gene is located centromeric to the proximal CMT1A–REP and, like the human *CDRT1* gene, is transcribed in the telomeric direction, ending within the proximal CMT1A–REP (Kennerson et al. 1997, 1998) (Fig. 1). In searching for a mouse ortholog for *HREP*, we identified a 759-bp continuous fragment of mouse *HREP* partial mRNA sequence. The first 269 bp of this sequence aligns with the human cDNA and corresponds to human exons IV and V. However, the remainder of the mouse mRNA does not align with human *HREP* exon VI, but instead the sequences at the 3′ end of this mouse *HREP* EST contig contain *CDRT1* sequences. Exon VI of human *HREP* is located inside the proximal CMT1A–REP and utilizes complementary sequence of *COX10* pseudoexon VI. Mice do not have the proximal CMT1A–REP; the proximal CMT1A–REP appeared during primate speciation between gorilla and chimpanzee (Kiyosawa and Chance 1996; Reiter et al. 1997; Boerkoel et al.

1999; Keller et al. 1999). These data suggest that in the mouse, sequences corresponding to human *HREP* and *CDRT1* are part of a single gene. The fact that 298 bp from within the mouse ortholog of *HREP* does not match genomic sequence on either side of the proximal CMTA1–REP suggests that the primate progenitor to human lost some genome sequence when the proximal CMT1A–REP integrated into this region (Fig. 6B).

DISCUSSION

Human 17p12 is a genomic region prone to DNA rearrangement (the CMT1A duplication and HNPP deletion) and has undergone relatively recent evolutionary changes during primate speciation (the 24-kb duplicated CMT1A–REPs). Although extensive studies have been performed to elucidate the molecular mechanism for the CMT1A duplication and HNPP deletion, an unequal crossing-over event via homologous recombination utilizing the flanking CMT1A–REPs as substrates, less information has been available for the 1.4-Mb CMT1A/HNPP genomic region between the CMT1A–REPs (Murakami and Lupski 1996; Murakami et al. 1997b; Boerkoel et al. 1999). The finished genomic sequence of this 1.4-Mb region has allowed the elucidation of the genes within the genomic interval and has provided information regarding the genomic architecture of the CMT1A/HNPP region. Our analyses uncovered new LCRs, revealed male-specific reduced recombination, identified novel genes, and shown a mechanism for the evolution of new genes through DNA rearrangement. Our findings suggest that the human genome is in a state of flux with DNA rearrangements apparently responsible for a significant amount of genomic evolution.

LCRs

Large genomic rearrangements mediated by LCR units are associated with a number of human genomic disorders (Lupski 1998b; Shaffer and Lupski 2000). In the CMT1A/HNPP region, in addition to the previously reported CMT1A–REP (Pentao et al. 1992; Reiter et al. 1996, 1997), we have identified three copies of a novel LCR, LCRA1, LCRA2, and LCRB. Interestingly, the genomic organization of LCRA1 and LCRA2 consists of inverted repeats flanking the 200-kb region containing the distal CMT1A–REP (Fig. 1). This genomic structure may allow flipping or inversion of the 200-kb genomic fragment in between, thus resulting in the CMT1A–REPs having an inverted orientation (Fig. 2B). Such a genomic arrangement may prevent the interchromosomal unequal crossing over that results in CMT1A duplication and HNPP deletion, making such individuals less susceptible to de novo duplication/deletion. This hypothesis is directly

testable by determining the CMT1A–REP orientation in the parent of origin for the de novo rearrangement.

A nucleotide sequence comparison between these LCRs revealed that the LCRA1 is likely a progenitor and the other two arose from subsequent duplication events. Two features indicate that the LCRB was probably generated first by local duplication followed by another duplication event to generate LCRA2 from LCRA1. First, the 18-bp deletion only exists in LCRA2 and the sequence homology between LCRA1/LCRB is lower than that between LCRA1/LCRA2. Secondly, a corresponding copy of *CDRT15* in LCRA2 has premature termination and thus is likely a pseudogene of *CDRT15*.

Multiple copies of LCRs are distributed throughout the human genome. Some BAC clones containing these LCRs map to the Smith-Magenis syndrome (SMS) region on 17p11.2. SMS–REP is a large (>200 kb) low copy region-specific repeat that acts as an homologous recombination substrate and is responsible for a large (~4 Mb) genomic deletion and duplication associated with human disorders (Chen et al. 1997; Potocki et al. 2000). Six copies of the LCRs were also mapped in 22q11.2, but not in the chromosome 22-specific LCRs (Dunham et al. 1999). Therefore, this LCR family manifests complex divergence throughout the human genome. Because copies of this LCR family are located close to the recombination breakpoints of SMS in 17p12, this LCR family may potentially be involved in the mechanism generating other genomic disorders.

Furthermore, these genome-wide repeat units also involve a gene family that reveals multiple transcripts from different loci. At least three copies of the transcript with no premature termination have been isolated. Further characterization of the sequences of these genomic loci as well as determination of the function of *CDRT15* and its paralogs will clarify the complicated structure of these LCRs.

Comparison of Genetic and Physical Maps of the CMT1A Duplication/HNPP Deletion Region

We hypothesized previously that the *mariner* transposon-like element MITE, which is located ~500 bp proximal to the preferential region for strand exchange or hotspot for unequal crossing over in the CMT1A–REPs, may promote double-strand DNA breaks and stimulate the homologous recombination (Reiter et al. 1996, 1998). Multiple studies from CMT1A duplication and HNPP deletion patients in different world populations confirm a positional hotspot for recombination within an ~500-bp region of the 24,011-bp homologous CMT1A–REPs (Kiyosawa et al. 1995; Lopes et al. 1996; Reiter et al. 1996; Timmerman et al. 1997; Yamamoto

et al. 1997; Chang et al. 1998). It has been suggested that CMT1A–REPs may also mediate high-frequency homologous recombination of this region at a genomic level.

To investigate this latter hypothesis, we examined the relationship between genetic and physical distances using 21 known STS markers that span this portion of the genome (Fig. 3A). Although we expected increased recombination frequency at some specific *cis*-acting sequence, such as CMT1A–REPs or HSMAR2–*PMP22*, there is no significant change in the recombination frequency throughout the region. Instead, we observed evidence for reduced recombination in the 820-kb region between *D17S1843* and *D17S918* that contains the proximal CMT1A–REP and two of three HSMAR2 elements. These data indicate that the HSMAR2 elements may not increase the frequency of the recombination in the germ line, or the resolution and sensitivity to detect their effect on recombination ratio may be below the lower limit of detection in this study.

Interestingly, in male meiosis, the genomic region with low recombination frequency extended beyond the CMT1A region in both the proximal and distal directions. As shown in chromosome 7, high female/male distance ratio in the genetic versus physical map is likely the result of reduced recombination in males, not of enhanced recombination in females (Broman et al. 1998). There was no recombination identified in the male meiotic map between *D17S921* and *D17S620* (~3 Mb), whereas in females this same physical distance revealed a 20-cM genetic distance. This reduced male recombination frequency may result in an extended region of two allelic chromosomes without crossing over or synapse formation in meiosis. Such an absence of synapse formation could in turn allow the chromosomes to slip on each other, thus enabling an unequal crossover involving the tandem repeat units, CMT1A–REPs. On the other hand, frequent interchromosomal equal crossovers may provide anchors to prevent chromosomal slipping and reduce the chance of unequal crossovers between the proximal and distal CMT1A–REPs. In support of this hypothesis, de novo CMT1A duplication events occur 10 times more frequently in males than females (Palau et al. 1993; Lopes et al. 1997). Therefore, we hypothesize that one of the mechanisms for the male sex preference in de novo CMT1A duplication may result from the male sex-specific low recombination frequency throughout the region. Interestingly, in the studies of human trisomies, significant reduction of genetic recombination was observed in the trisomy-generating meiosis, and it was suggested that absence of pairing and/or recombination contributes to nondisjunction (Lamb et al. 1996). In the context of the hypothesis that decreased recombination may increase the unequal crossover at the proximal and distal CMT1A–REPs, individuals with reduced meiotic recombination may have an increased propensity to generate unequal reciprocal recombination products.

Han et al. (2000) reported recently that the frequency of unequal crossover between the proximal and distal CMT1A–REPs is almost identical to that of the average equal crossover in the human genome by use of sperm DNA analysis. This hypothesis also indicates that the CMT1A–REPs do not contain a genomic recombination hotspot for the unequal crossover. In the same study, Han et al (2000) localized the recombination breakpoint in the same hotspot identified previously by the analysis of patient DNA. Together with the fact that the CMT1A–REPs do not contain a genomic hotspot for equal crossover according to the comparison of the genetic and physical maps in this study, the hotspot in the CMT1A–REP should be defined as a hotspot for the position preference, not for recombination frequency (Han et al. 2000).

Genes in the CMT1A Duplication/HNPP Deletion Region

In the 1.4-Mb CMT1A duplication/HNPP deletion region, we identified five genes and 13 predicted genes in addition to three previously mapped genes. The current estimated average number of human genes per Mb is between 9.6 and 12.9 (International Human Genome Sequence Consortium 2001). Previous studies suggested that chromosome 17 is gene-rich by a factor of 1.44 (Deloukas et al. 1998), which increases the estimated number of the genes on chromosome 17 to be between 13.8 and 18.6 per Mb. The combination of the eight confirmed and 13 predicted genes within this 1.4-Mb region yields a density of 15 genes/Mb, well within this estimate.

In addition to *PMP22*, we mapped one previously characterized and two uncharacterized genes to this region, *HS3ST3B1*, *NPD008/CGI-148*, and *TEKT3*. *HS3ST3B1* is one of the five isoforms of genes encoding heparan sulphate biosynthesizing enzymes, heparan sulphate sulphotransferases (HS3STs). Heparan sulphate binds to specific proteins such as antithrombin and several growth factors, and thereby regulates various biological processes including anticoagulation and angiogenesis (Rosenberg et al. 1997). HS3STs catalyze sulfation of monosaccharide sequences of heparan sulphate, which is believed to be critical for binding to the target proteins. *HS3ST3B1* has a closely related isoform, *HS3ST3A1*, which also has similar patterns of tissue expression and encodes a protein with similar enzymatic activity. Together with the nature of this type of catalytic enzyme, wherein changes in dosage usually do not affect the

system, existence of a paralog with similar enzymatic properties suggest that duplication or deletion of one allele of *HS3ST3B1* may not affect heparan sulphate biosynthesis.

Tektin includes a family of proteins and represents one of the components of motile and primary cilia associating with the major structural component of cilia, microtubules (Linck and Langevin 1982; Linck et al. 1985; Steffen and Linck 1988). Tektins have been best studied in sea urchins, a species in which three isoforms have been isolated; tektin A1, tektin B1, and tektin C1. Mammalian homologs for tektin B1 and tektin C1 have been isolated (GenBank accession no. NM_014466, NM_011902 and NM_011569) (Norrander et al. 1998; Iguchi et al. 1999). In the CMT1A/HNPP region, we identified *TEKT3* as the first homolog for *tektin A1* in mammals. Like other tektin homologs, it is preferentially expressed in testis. Tektin A1 and tektin B1 are thought to be assembled as heterodimers to comprise the tektin filament, and interact with tubulins to form the basis of the high degree of stability of doublet microtubules (Pirner and Linck 1994). In the mouse sperm, the tektin B1 homologous protein tekt2 is localized in flagella, strongly suggesting that tektins may play essential roles in formation of sperm and in sperm motility (Iguchi et al. 1999). Loss of *TEKT3* may reduce the motility of the sperm of HNPP patients because of their haploid nature.

Relevance to CMT1A/HNPP Genomic Disorders

Of the new LCRs found in the CMT1A/NHPP region, LCRA2 and LCRB are present in a tandem orientation and flank *PMP22*, suggesting that they have the potential to be substrates for unequal homologous recombination leading to duplication or deletion of *PMP22*. Four families with alternate size duplication or deletion were reported previously (Ionasescu et al. 1993; Palau et al. 1993; Valentijn et al. 1993; Chapon et al. 1996). Genetic studies with a few markers showed that the proximal break points of these cases are located close to or within the proximal CMT1A–REP, and the distal break points mapped between *PMP22* and *D17S125* (Ionasescu et al. 1993; Palau et al. 1993; Valentijn et al. 1993; Chapon et al. 1996). Therefore, at least in these cases, recombination between the LCRs found in this study are unlikely to be involved in the small duplication or deletion. Additional analyses for LCR in this region failed to identify any significant stretches of homologous sequence (>1 kb) that may serve as substrates for such alternative homologous recombination events.

Most of the genes identified in this study revealed extremely low expression in adult tissues but obvious expression in fetal tissues. It is surprising that these embryonic genes have no developmental effect on the

individuals with duplication or deletion of the 1.4-Mb region. The observation that to date *PMP22* is the only gene responsible for CMT1A/HNPP due to the mechanism of gene dosage accompanied by duplication or deletion of this region suggests that dosage sensitivity may be a unique property of *PMP22* but not of the other genes in the 1.4-Mb region. The sequence of most of these genes contains insufficient information to estimate their function. However, the cumulative data suggest that only 1 in 21 genes, at least in this portion of the human genome, is sensitive to dosage effects.

Evolution of New Genes, *HREP* and *CDRT1*, by DNA Rearrangement

Identification of the *COX10* gene spanning the distal CMT1A–REP and only one exon (pseudoexon VI) in the proximal CMT1A–REP indicates that the distal copy is the original and the proximal CMT1A–REP represents a duplicated copy (Murakami et al. 1997a; Reiter et al. 1997). Evolutionary studies reveal that this insertional event occurred between gorilla and chimpanzee (Kiyosawa and Chance 1996; Reiter et al. 1997; Boerkoel et al. 1999; Keller et al. 1999). Subsequently, another gene, *HREP*, was identified close to the proximal CMT1A–REP (Kennerson et al. 1997, 1998). *HREP* is transcribed toward the telomere from outside the proximal CMT1A–REP and terminates within the proximal CMT1A–REP. The last exon of *HREP* occurs at the same position, but on the complementary strand of *COX10* pseudoexon VI (Kennerson et al. 1997).

Interestingly, we found that a mouse gene homologous to human *HREP* does not share the region after exon V with human *HREP*, but instead matches *CDRT1*, which is adjacent to the proximal CMT1A–REP on the telomeric side. Therefore, *CDRT1* and *HREP* are likely to be parts of an Ancestral Gene before the Integration of Proximal CMT1A–REP (*AGIP*) (Fig. 6). The CMT1A–REP insertional event, which is estimated to have occurred during primate speciation between gorilla and chimpanzee, divided *AGIP* into two genes, *HREP* and *CDRT1*. These findings show an example of evolution of new genes by DNA rearrangement during mammalian genome evolution. The first half of *AGIP* became *HREP* utilizing a part of CMT1A–REP as a new terminating exon, whereas the last exon of *AGIP* became a single exon gene *CDRT1*. Interestingly, expression profiles of these two genes are different; *HREP* is expressed in heart and skeletal muscle, whereas the major expression of *CDRT1* is observed in pancreas. Furthermore, a region in *AGIP* between the *HREP* syntenic portion and *CDRT1* syntenic portion was likely to be lost during the CMT1A–REP integration, suggesting that this insertional genomic rearrangement was accompanied by loss of a genomic fragment. Further evo-

lutionary analysis of the genomic region surrounding proximal CMT1A–REP in chimpanzee and gorilla may elucidate the mechanism of integration of the CMT1A–REP.

In conclusion, we have evaluated the 1.4-Mb finished genomic sequence of the CMT1A/HNPP region. Data obtained from this genome-sequencing study enable new insights into human genome architecture and mammalian genome evolution, show evolution of new genes by genome rearrangements during primate speciation, and add to the plethora of information being created by the complete nucleotide sequencing of the human genome.

METHODS

Construction of Physical Maps of the 1.4-Mb CMT1A/HNPP Region

We implemented two independent approaches to construct the physical map of the CMT1A/HNPP genomic region. The first approach utilized STS content-based mapping performed at Baylor College of Medicine. We used the end sequences of the multiple cosmid clones from a previously constructed cosmid contig of this region (Murakami and Lupski 1996) to screen PAC (P1 artificial chromosome; RPCI-1 Rosewell Park Cancer Institute, Buffalo, NY) and BAC (bacterial artificial chromosome; CITB California Institute for Technology) libraries by PCR on DNA pools and/or by filter hybridization. Eight known genetic markers and the *PMP22* gene were also used as probes. Overlaps of each large insert genomic clone were evaluated by *Eco*RI fingerprinting by use of a FluorImager (Molecular Dynamics), as described elsewhere (Marra et al. 1997).

A parallel and alternative approach used YAC-based mapping conducted at the Whitehead Institute Center for Genome Research as a part of the effort to sequence the entire human chromosome 17. To create reliable physical maps despite significant amounts of low-copy repetitive sequence, we used a high density of unique markers. In addition to pre-existing markers, new markers were generated from shotgun sequences derived from pulsed-field gel-purified YACs. Overlapping YACs from the CEPH Mega-YAC library (Chumakov et al. 1995) that were not known to be chimeric based on STS content (Hudson et al. 1995) were selected from the CMT1A region (Pentao et al. 1992). Each YAC was fractionated and subcloned separately into M13. Single-sequencing reactions were performed on several hundred subclones from each YAC and the resulting sequences contained from 20%–60% yeast DNA, depending on the YAC. Thirty-eight base pair overgos were designed (Ross et al. 1999) and further tested by hybridization to eliminate probes that contained highly or moderately repetitive sequences that escaped detection during their design. BAC library (RPCI-11) screening was by hybridization with pools of up to 40 overgos derived from a single YAC, with an average density of 30 overgos per Mb of genomic region. Positive clones from the library screen were streaked on agar plates to obtain single colonies and one clone from each positive address was rearrayed into new 96-well plates. To generate marker content maps, replica filters made from the 96-well plates were hybridized individually with each of the overgos used in the library screen, as well as overgos derived from overlapping YACs, and overgos representing other markers mapped in the region. Markers that hybridized to greater than the expected number of clones were not included in the final map, nor were markers that were not linked by at least two clones. Clones that did not share at least two markers with an overlapping clone were not included in the map. The final density of markers in the BAC map of the region was ~1 marker every 10 kb. This high-density physical mapping generated an overlapping contig with 8- to 10-fold coverage. Combining these two physical maps, clones with a minimal tiling path were selected for sequencing (Fig. 1).

Shotgun Library Construction, DNA Sequencing, and Sequence Data Analyses

Subclone libraries were constructed for each human genome containing bacterial clone and shotgun sequencing, assembly and finishing was performed as described (International Human Genome Sequencing Consortium 2001). A single annotated gap remains in the sequence of RP11–726O12 (AC005517). PCR amplification of template DNA from the corresponding large-insert genomic clone followed by sequencing revealed that the gap contains 439 bp with an extremely high content of GA repeat. The repeat content is probably responsible for the difficulties encountered in cloning and sequencing this gap region. The sequence from each BAC/PAC clone was assembled into a larger sequence contig by use of Sequencher (Gene Codes). These data were analyzed by the NIX analysis program (Nucleotide Identification of unknown sequences, UK MRC Human Genome Mapping Project; http://www.hgmp.mrc.ac.uk), a Web-based package of gene analysis software (including GRAIL, Fex, Hexon, MZEF, Genemark, Genefinder, FGene, BLAST, Polyah, RepeatMasker and TRNAscan). Each region that contained a potential gene was individually analyzed by additional gene prediction and protein analysis programs, by use of the ExPASy proteomics server (Expert Protein Analysis System; http://www.expasy.ch). Putative core promoter and transcription-binding sites were analyzed by TESS (http://www.cbil.upenn.edu/tess/index.html), Human Core-Promoter Finder (http://sciclio.cshl.org/genefinder/CPROMOTER/human.htm), TSSG, and TSSW (BCM GeneFinder; http://dot.imgen.bcm.tmc.edu:9331/gene-finder/gf.html). RepeatMasker was independently run to identify interspersed repeat sequences. A genetic map of chromosome 17 with raw data from polymorphic genetic markers within this region was obtained from the Marshfield Web site (http://www.marshmed.org/genetics) to evaluate genetic/physical map correlations (Broman et al. 1998).

Northern Blotting and RT–PCR Analyses

Expression profiles and the size of each transcript was determined by multiple tissue Northern blotting (Clontech). Primers from the unique 3′ untranslated region of each isolated gene were designed by use of web-based software, Primer3 (http://www-genome.wi.mit.edu/genome_software/other/primer3.html). Corresponding BAC/PAC clones were used as template DNA for PCR to generate probes to minimize the chance of amplification of gene family members and pseudogenes mapping elsewhere in the genome. RT–PCR was performed for some of the predicted genes by use of first-strand cDNA from various adult and fetal tissues (Clontech).

ACKNOWLEDGMENTS

REFERENCES

Badano, J.L., Inoue, K., Katsanis, N., and Lupski, J.R. New polymorphic short tandem repeats for PCR-based Charcot-Marie-Tooth disease type 1A duplication diagnosis. *Clin. Chem.* (In press).

Blair, I.P., Kennerson, M.L., and Nicholson, G.A. 1995. Detection of Charcot-Marie-Tooth type 1A duplication by the polymerase chain reaction. *Clin. Chem.* **41:** 1105–1108.

Boerkoel, C.F., Inoue, K., Reiter, L.T., Warner, L.E., and Lupski, J.R. 1999. Molecular mechanisms for CMT1A duplication and HNPP deletion. *Ann. NY Acad. Sci.* **883:** 22–35.

Broman, K.W., Murray, J.C., Sheffield, V.C., White, R.L., and Weber, J.L. 1998. Comprehensive human genetic maps: Individual and sex-specific variation in recombination. *Am. J. Hum. Genet.* **63:** 861–869.

Chance, P.F., Alderson, M.K., Leppig, K.A., Lensch, M.W., Matsunami, N., Smith, B., Swanson, P.D., Odelberg, S.J., Disteche, C.M., and Bird, T.D. 1993. DNA deletion associated with hereditary neuropathy with liability to pressure palsies. *Cell* **72:** 143–151.

Chance, P.F., Abbas, N., Lensch, M.W., Pentao, L., Roa, B.B., Patel, P.I., and Lupski, J.R. 1994. Two autosomal dominant neuropathies result from reciprocal DNA duplication/deletion of a region on chromosome 17. *Hum. Mol. Genet.* **3:** 223–228.

Chang, J.-G., Jong, Y.-J., Wang, W.-P., Wang, J.-C., Hu, C.-J., Lo, M.-C., and Chang, C.-P. 1998. Rapid detection of a recombinant hotspot associated with Charcot-Marie-Tooth disease type 1A duplication by a PCR-based DNA test. *Clin. Chem.* **44:** 270–274.

Chapon, F., Diraison, P., Lechevalier, B., Chazot, G., Viader, F., Bonnebouche, C., Vandenberghe, A., Timmerman, V., and Van Broeckhoven, C. 1996. Hereditary neuropathy with liability to pressure palsies with a partial deletion of the region often duplicated in Charcot-Marie-Tooth disease, type 1A. *J. Neurol. Neurosurg. Psychiatry* **61:** 535–536.

Chen, K.-S., Manian, P., Koeuth, T., Potocki, L., Zhao, Q., Chinault, A.C., Lee, C.C., and Lupski, J.R. 1997. Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome. *Nat. Genet.* **17:** 154–163.

Chumakov, I.M., Rigault, P., Le Gall, I., Bellanné-Chantelot, C., Billault, A., Guillou, S., Soularue, P., Guasconi, G., Poullier, E., Gros, I., et al. 1995. A YAC contig map of the human genome. *Nature* **377:** 175–297.

Deloukas, P., Schuler, G.D., Gyapay, G., Beasley, E.M., Soderlund, C., Rodriguez-Tomé, P., Hui, L., Matise, T.C., McKusick, K.B., Beckmann, J.S., et al. 1998. A physical map of 30,000 human genes. *Science* **282:** 744–746.

Dunham, I., Hunt, A.R., Collins, J.E., Bruskiewich, R., Beare, D.M., Clamp, M., Smink, L.J., Ainscough, R., Almeida, J.P., Babbage, A., et al. 1999. The DNA sequence of human chromosome 22. *Nature* **402:** 489–495.

Han, L.-L., Keller, M.P., Navidi, W., Chance, P.F., and Arnheim, N. 2000. Unequal exchange at the Charcot-Marie-Tooth disease type 1A recombination hot-spot is not elevated above the genome average rate. *Hum. Mol. Genet.* **9:** 1881–1889.

Hattori, M., Fujiyama, A., Taylor, T.D., Watanabe, H., Yada, T., Park, H.-S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D.-K., et al. 2000. The DNA sequence of human chromosome 21. The chromosome 21 mapping and sequencing consortium. *Nature* **405:** 311–319.

Hudson, T.J., Stein, L.D., Gerety, S.S., Ma, J., Castle, A.B., Silva, J., Slonim, D.K., Baptista, R., Kruglyak, L., Xu, S.-H., et al. 1995. An STS-based map of the human genome. *Science* **270:** 1945–1954.

Iguchi, N., Tanaka, H., Fujii, T., Tamura, K., Kaneko, Y., Nojima, H., and Nishimune, Y. 1999. Molecular cloning of haploid germ cell-specific tektin cDNA and analysis of the protein in mouse testis. *FEBS Lett.* **456:** 315–321.

International Human Genome Sequence Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921.

Ionasescu, V.V., Ionasescu, R., Searby, C., and Barker, D.F. 1993. Charcot-Marie-Tooth neuropathy type 1A with both duplication and non-duplication. *Hum. Mol. Genet.* **2:** 405–410.

Keller, M.P., Seifried, B.A., and Chance, P.F. 1999. Molecular evolution of the CMT1A-REP region: A human- and chimpanzee-specific repeat. *Mol. Biol. Evol.* **16:** 1019–1026.

Kennerson, M.L., Nassif, N.T., Dawkins, J.L., DeKroon, R.M., Yang, J.G., and Nicholson, G.A. 1997. The Charcot-Marie-Tooth binary repeat contains a gene transcribed from the opposite strand of a partially duplicated region of the *COX10* gene. *Genomics* **46:** 61–69.

Kennerson, M.L., Nassif, N.T., and Nicholson, G.A. 1998. Genomic structure and physical mapping of *C17orf1*: A gene associated with the proximal element of the CMT1A-REP binary repeat. *Genomics* **53:** 110–112.

Kiyosawa, H. and Chance, P.F. 1996. Primate origin of the CMT1A-REP repeat and analysis of a putative transposon-associated recombinational hotspot. *Hum. Mol. Genet.* **5:** 745–753.

Kiyosawa, H., Lensch, M.W., and Chance, P.F. 1995. Analysis of the CMT1A-REP repeat: Mapping crossover breakpoints in CMT1A and HNPP. *Hum. Mol. Genet.* **4:** 2327–2334.

Kovach, M.J., Lin, J.-P., Boyadjiev, S., Campbell, K., Mazzeo, L., Herman, K., Rimer, L.A., Frank, W., Llewellyn, B., Wang Jabs, E., et al. 1999. A unique point mutation in the *PMP22* gene is associated with Charcot-Marie-Tooth disease and deafness. *Am. J. Hum. Genet.* **64:** 1580–1593.

Lai, C.-H., Chou, C.-Y., Ch'ang, L.-Y., Liu, C.-S., and Lin, W. 2000. Identification of novel human genes evolutionarily conserved in *Caenorhabditis elegans* by comparative proteomics. *Genome Res.* **10:** 703–713.

Lamb, N.E., Freeman, S.B., Savage-Austin, A., Pettay, D., Taft, L., Hersey, J., Gu, Y., Shen, J., Saker, D., May, K.M., et al. 1996. Susceptible chiasmate configurations of chromosome 21 predispose to non-disjunction in both maternal meiosis I and meiosis II. *Nat. Genet.* **14:** 400–405.

Linck, R.W. and Langevin, G.L. 1982. Structure and chemical composition of insoluble filamentous components of sperm flagellar microtubules. *J. Cell. Sci.* **58:** 1–22.

Linck, R.W., Amos, L.A., and Amos, W.B. 1985. Localization of tektin filaments in microtubules of sea urchin sperm flagella by immunoelectron microscopy. *J. Cell. Biol.* **100:** 126–135.

Liu, J., Shworak, N.W., Sinäy, P., Schwartz, J.J., Zhang, L., Fritze, L.M.S., and Rosenberg, R.D. 1999. Expression of heparan sulfate D-glucosaminyl 3-*O*-sulfotransferase isoforms reveals novel substrate specificities. *J. Biol. Chem.* **274:** 5185–5192.

Lopes, J., LeGuern, E., Gouider, R., Tardieu, S., Abbas, N., Birouk, N., Gugenheim, M., Bouche, P., Agid, Y., and Brice, A. 1996. Recombination hot spot in a 3.2-kb region of the

Charcot-Marie-Tooth type 1A repeat sequences: New tools for molecular diagnosis of hereditary neuropathy with liability to pressure palsies and of Charcot-Marie-Tooth type 1A. French CMT Collaborative Research Group. *Am. J. Hum. Genet.* **58:** 1223–1230.

Lopes, J., Vandenberghe, A., Tardieu, S., Ionasescu, V., Lévy, N., Wood, N., Tachi, N., Bouche, P., Latour, P., Brice, A., et al. 1997. Sex-dependent rearrangements resulting in CMT1A and HNPP. *Nat. Genet.* **17:** 136–137.

Lupski, J.R. 1998a. Charcot-Marie-Tooth disease: Lessons in genetic mechanisms. *Mol. Med.* **4:** 3–11.

———. 1998b. Genomic disorders: Structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet.* **14:** 417–422.

Lupski, J.R. and Garcia, C.A. 2001. Charcot-Marie-Tooth peripheral neuropathies and related disorders. In *The metabolic and molecular basis of inherited diseases* (eds. C.R. Scriver, A.L. Beaudet, W.S. Sly, and D. Valle), pp. 5759–5788. McGraw-Hill, New York, NY.

Lupski, J.R., de Oca-Luna, R.M., Slaugenhaupt, S., Pentao, L., Guzzetta, V., Trask, B.J., Saucedo-Cardenas, O., Barker, D.F., Killian, J.M., Garcia, C.A., et al. 1991. DNA duplication associated with Charcot-Marie-Tooth disease type 1A. *Cell* **66:** 219–232.

Lupski, J.R., Wise, C.A., Kuwano, A., Pentao, L., Parke, J.T., Glaze, D.G., Ledbetter, D.H., Greenberg, F., and Patel, P.I. 1992. Gene dosage is a mechanism for Charcot-Marie-Tooth disease type 1A. *Nat. Genet.* **1:** : 29–33.

Marra, M.A., Kucaba, T.A., Dietrich, N.L., Green, E.D., Brownstein, B., Wilson, R.K., McDonald, K.M., Hillier, L.W., McPherson, J.D., and Waterston, R.H. 1997. High throughput fingerprint analysis of large-insert clones. *Genome Res.* **7:** 1072–1084.

Matsunami, N., Smith, B., Ballard, L., Lensch, M.W., Robertson, M., Albertsen, H., Hanemann, C.O., Müller, H.W., Bird, T.D., White, R., et al. 1992. Peripheral myelin protein-22 gene maps in the duplication in chromosome 17p11.2 associated with Charcot-Marie-Tooth 1A. *Nat. Genet.* **1:** 176–179.

Murakami, T. and Lupski, J.R. 1996. A 1.5-Mb cosmid contig of the CMT1A duplication/HNPP deletion critical region in 17p11.2-p12. *Genomics* **34:** 128–133.

Murakami, T., Reiter, L.T., and Lupski, J.R. 1997a. Genomic structure and expression of the human heme A:farnesyltransferase (*COX10*) gene. *Genomics* **42:** 161–164.

Murakami, T., Sun, Z.S., Lee, C.C., and Lupski, J.R. 1997b. Isolation of novel genes from the CMT1A duplication/HNPP deletion critical region in 17p11.2-p12. *Genomics* **39:** 99–103.

Nelis, E., Van Broeckhoven, C., De Jonghe, P., Löfgren, A., Vandenberghe, A., Latour, P., Le Guern, E., Brice, A., Mostacciuolo, M.L., Schiavon, F., et al. 1996. Estimation of the mutation frequencies in Charcot-Marie-Tooth disease type 1 and hereditary neuropathy with liability to pressure palsies: A European collaborative study. *Eur. J. Hum. Genet.* **4:** 25–33.

Norrander, J., Larsson, M., Ståhl, S., Höög, C., and Linck, R. 1998. Expression of ciliary tektins in brain and sensory development. *J. Neurosci.* **18:** 8912–8918.

Palau, F., Löfgren, A., De Jonghe, P., Bort, S., Nelis, E., Sevilla, T., Martin, J.-J., Vilchez, J., Prieto, F., and Van Broeckhoven, C. 1993. Origin of the *de novo* duplication in Charcot-Marie-Tooth disease type 1A: Unequal nonsister chromatid exchange during spermatogenesis. *Hum. Mol. Genet.* **2:** 2031–2035.

Patel, P.I., Franco, B., Garcia, C., Slaugenhaupt, S.A., Nakamura, Y., Ledbetter, D.H., Chakravarti, A., and Lupski, J.R. 1990. Genetic mapping of autosomal dominant Charcot-Marie-Tooth disease in a large French-Acadian kindred: Identification of new linked markers on chromosome 17. *Am. J. Hum. Genet.* **46:** 801–809.

Patel, P.I., Roa, B.B., Welcher, A.A., Schoener-Scott, R., Trask, B.J., Pentao, L., Snipes, G.J., Garcia, C.A., Francke, U., Shooter, E.M., et al. 1992. The gene for the peripheral myelin protein PMP-22 is a candidate for Charcot-Marie-Tooth disease type 1A. *Nat. Genet.* **1:** 159–165.

Pentao, L., Wise, C.A., Chinault, A.C., Patel, P.I., and Lupski, J.R. 1992. Charcot-Marie-Tooth type 1A duplication appears to arise from recombination at repeat sequences flanking the 1.5 Mb monomer unit. *Nat. Genet.* **2:** 292–300.

Pirner, M.A. and Linck, R.W. 1994. Tektins are heterodimeric polymers in flagellar microtubules with axial periodicities matching the tubulin lattice. *J. Biol. Chem.* **269:** 31800–31806.

Potocki, L., Chen, K.-S., Park, S.-S., Osterholm, D.E., Withers, M.A., Kimonis, V., Summers, A.M., Meschino, W.S., Anyane-Yeboa, K., Kashork, C.D., et al. 2000. Molecular mechanism for duplication 17p11.2–-the homologous recombination reciprocal of the Smith-Magenis microdeletion. *Nat. Genet.* **24:** 84–87.

Raeymaekers, P., Timmerman, V., Nelis, E., De Jonghe, P., Hoogendijk, J.E., Baas, F., Barker, D.F., Martin, J.J., De Visser, M., Bolhuis, P.A., et al. 1991. Duplication in chromosome 17p11.2 in Charcot-Marie-Tooth neuropathy type 1a (CMT 1a). The HMSN Collaborative Research Group. *Neuromuscl. Disord.* **1:** 93–97.

Reiter, L.T., Murakami, T., Koeuth, T., Pentao, L., Muzny, D.M., Gibbs, R.A., and Lupski, J.R. 1996. A recombination hotspot responsible for two inherited peripheral neuropathies is located near a *mariner* transposon-like element. *Nat. Genet.* **12:** 288–297.

Reiter, L.T., Murakami, T., Koeuth, T., Gibbs, R.A., and Lupski, J.R. 1997. The human *COX10* gene is disrupted during homologous recombination between the 24 kb proximal and distal CMT1A-REPs. *Hum. Mol. Genet.* **6:** 1595–1603.

Reiter, L.T., Hastings, P.J., Nelis, E., De Jonghe, P., Van Broeckhoven, C., and Lupski, J.R. 1998. 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.

Reiter, L.T., Liehr, T., Rautenstrauss, B., Robertson, H.M., and Lupski, J.R. 1999. Localization of *mariner* DNA transposons in the human genome by PRINS. *Genome Res.* **9:** 839–843.

Roa, B.B., Greenberg, F., Gunaratne, P., Sauer, C.M., Lubinsky, M.S., Kozma, C., Meck, J.M., Magenis, R.E., Shaffer, L.G., and Lupski, J.R. 1996. Duplication of the *PMP22* gene in 17p partial trisomy patients with Charcot-Marie-Tooth type-1A neuropathy. *Hum. Genet.* **97:** 642–649.

Rosenberg, R.D., Shworak, N.W., Liu, J., Schwartz, J.J., and Zhang, L. 1997. Heparan sulfate proteoglycans of the cardiovascular system. Specific structures emerge but how is synthesis regulated? *J. Clin. Invest.* **99:** 2062–2070.

Ross, M.T., LaBrie, S., McPherson, J., and Stanton, V.P.J. 1999. Screening large-insert libraries by hybridization. In *Current protocols in human genetics* (eds. N.C. Dracopoli, J.L. Haines, B.R. Korf, D.T. Moir, C.C. Morton, C.E. Seidman, J.G. Seidman, and D.R. Smith), pp. 5.6.1–5.6.52. John Wiley and Sons, New York, NY.

Sabéran-Djoneidi, D., Sanguedolce, V., Assouline, Z., Lévy, N., Passage, E., and Fontés, M. 2000. Molecular dissection of the Schwann cell specific promoter of the *PMP22* gene. *Gene* **248:** 223–231.

Schuler, G.D. 1998. Electronic PCR: Bridging the gap between genome mapping and genome sequencing. *Trends Biotechnol.* **16:** 456–459.

Shaffer, L.G. and Lupski, J.R. 2000. Molecular mechanisms for constitutional chromosomal rearrangements in humans. *Annu. Rev. Genet.* **34:** 297–329.

Shworak, N.W., Liu, J., Petros, L.M., Zhang, L., Kobayashi, M., Copeland, N.G., Jenkins, N.A., and Rosenberg, R.D. 1999. Multiple isoforms of heparan sulfate D-glucosaminyl 3–*O*-sulfotransferase. Isolation, characterization, and expression of human cDNAs and identification of distinct genomic loci. *J. Biol. Chem.* **274:** 5170–5184.

Steffen, W. and Linck, R.W. 1988. Evidence for tektins in centrioles and axonemal microtubules. *Proc. Natl. Acad. Sci.* **85:** 2643–2647.

Suter, U., Snipes, G.J., Schoener-Scott, R., Welcher, A.A., Pareek, S., Lupski, J.R., Murphy, R.A., Shooter, E.M., and Patel, P.I. 1994. Regulation of tissue-specific expression of alternative peripheral myelin protein-22 (*PMP22*) gene transcripts by two promoters. *J. Biol. Chem.* **269:** 25795–25808.

Timmerman, V., Raeymaekers, P., De Jonghe, P., De Winter, G., Swerts, L., Jacobs, K., Gheuens, J., Martin, J.-J., Vandenberghe, A., and Van Broeckhoven, C. 1990. Assignment of the Charcot-Marie-Tooth neuropathy type 1 (CMT 1a) gene to 17p11.2-p12. *Am. J. Hum. Genet.* **47:** 680–685.

Timmerman, V., Nelis, E., Van Hul, W., Nieuwenhuijsen, B.W., Chen, K.L., Wang, S., Ben Othman, K., Cullen, B., Leach, R.J., Hanemann, C.O. et al., 1992. The peripheral myelin protein gene *PMP-22* is contained within the Charcot-Marie-Tooth disease type 1A duplication. *Nat. Genet.* **1:** 171–175.

Timmerman, V., Rautenstrauss, B., Reiter, L.T., Koeuth, T., Löfgren, A., Liehr, T., Nelis, E., Bathke, K.D., De Jonghe, P., Grehl, H., et al. 1997. Detection of the CMT1A/HNPP recombination hotspot in unrelated patients of European descent. *J. Med. Genet.* **34:** 43–49.

Valentijn, L.J., Bolhuis, P.A., Zorn, I., Hoogendijk, J.E., van den Bosch, N., Hensels, G.W., Stanton, V.P., Jr., Housman, D.E., Fischbeck, K.H., Ross, D.A., et al. 1992. The peripheral myelin gene *PMP-22/GAS-3* is duplicated in Charcot-Marie-Tooth disease type 1A. *Nat. Genet.* **1:** 166–170.

Valentijn, L.J., Baas, F., Zorn, I., Hensels, G.W., de Visser, M., and Bolhuis, P.A. 1993. Alternatively sized duplication in Charcot-Marie-Tooth disease type 1A. *Hum. Mol. Genet.* **2:** 2143–2146.

Wise, C.A., Garcia, C.A., Davis, S.N., Heju, Z., Pentao, L., Patel, P.I., and Lupski, J.R. 1993. Molecular analyses of unrelated Charcot-Marie-Tooth (CMT) disease patients suggest a high frequency of the CMT1A duplication. *Am. J. Hum. Genet.* **53:** 853–863.

Yamamoto, M., Yasuda, T., Hayasaka, K., Ohnishi, A., Yoshikawa, H., Yanagihara, T., Ikegami, T., Yamamoto, T., Ohashi, H., Nishimura, T., et al. 1997. Locations of crossover breakpoints within the CMT1A-REP repeat in Japanese patients with CMT1A and HNPP. *Hum. Genet.* **99:** 151–154.