Evolutionary dynamics of the primate LRRC37 gene family

Giuliana Giannuzzi,1 Priscillia Siswara,2 Maika Malig,2 Tomas Marques-Bonet,3 NISC Comparative Sequencing Program,4 James C. Mullikin,4 Mario Ventura,1,2 and Evan E. Eichler2,5,6

1 Dipartimento di Biologia, Università degli Studi di Bari “Aldo Moro,” Bari 70126, Italy; 2 Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington 98195, USA; 3 IBE, Institut de Biologia Evolutiva (UPF-CSIC), Universitat Pompeu Fabra, PRBB, 08003 Barcelona, Catalonia, Spain; 4 Genome Technology Branch and NIH Intramural Sequencing Center (NISC), National Human Genome Research Institute, Bethesda, Maryland 20892, USA; 5 Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA

Core duplicons in the human genome represent ancestral duplication modules shared by the majority of intrachromosomal duplication blocks within a given chromosome. These cores are associated with the emergence of novel gene families in the hominoid lineage, but their genomic organization and gene characterization among other primates are largely unknown. Here, we investigate the genomic organization and expression of the core duplcon on chromosome 17 that led to the expansion of LRRC37 during primate evolution. A comparison of the LRRC37 gene family organization in human, orangutan, macaque, marmoset, and lemur genomes shows the presence of both orthologous and species-specific gene copies in all primate lineages. Expression profiling in mouse, macaque, and human tissues reveals that the ancestral expression of LRRC37 was restricted to the testis. In the hominid lineage, the pattern of LRRC37 became increasingly ubiquitous, with significantly higher levels of expression in the cerebellum and thymus, and showed a remarkable diversity of alternative splice forms. Transfection studies in HeLa cells indicate that the human FLAG-tagged recombinant LRRC37 protein is secreted after cleavage of a transmembrane precursor and its overexpression can induce filipodia formation.

[Supplemental material is available for this article.]
Results

Human LRRC37 family organization

The LRRC37 family maps to 18 distinct loci in the human genome (NCBI36/hg18) on the q arm of chromosome 17. These copies derive from complete or partial duplication of the gene and contain both exons and introns. Further, a retrocopy is present on the p arm of chromosome 10 (Fig. 1A; Supplemental Table S1; Supplemental Note). We distinguished the copies in six different types (A–F) according to their location: B-type copies clustered at 17q11; F copy matched at 17q12; A-type copies, except A3, clustered at 17q21; C copy matched at 17q24; D copy matched on chr17 random; and E copy is the retrocopy on chromosome 10. LRRC37A, A2, A3, and B have complete gene structure, ReSeq coding RNA annotated, and intact open reading frames (ORFs), whereas the remaining copies have partial gene structure and/or disrupted reading frames. LRRC37A, A2, and A3 differ from LRRC37B in the length of exon 1 (2612 bp in A type, exon 1a; 1690 bp in B type, exon 1b) and exon 9 (1532 bp in A type; split in two exons of 141 and 293 bp in B type, named exons 9b and 9b). Terminal exons, downstream from exon 10, also differ: exons 11–14 in A type and exon 15 in B type (Fig. 1A,B).

Based on the ORF, the LRRC37A and LRRC37B putatively encoded proteins consist of 1700 and 947 residues and have a molecular weight of 189 and 106 kDa, respectively. They contain a predicted signal peptide and transmembrane helix encoded by exons 1 and 10, respectively. The LRRC37A protein has repetitive segments (RPT, RePeaT) encoded by exon 1 (Fig. 1B.2). The region from the end of exon 1 to exon 8 encodes six tandem extracellular LRR motifs, flanked by the LRR N- and C-terminal domains (LRR-NT and LRR-CT), according to the LRRscan program (Fig. 1B.2; Dolan et al. 2007). The motifs LRR1, LRR3, LRR4, and LRR5 are typical and match the consensus sequence LxLxLxLxNxL (where x can be any amino acid, and L positions can also be occupied by V, I, and F), whereas LRR2 and LRR6 are degenerate or atypical as they do not match perfectly the consensus sequence. Additionally, the last LRR before the LRR-CT domain contains only the first subdomain of eight residues. Notably, the LxLxLxLxNxL domain of each LRR motif spans a splicing junction (blue arrowheads in Fig. 1B.3). According to the protein structure and the extracellular nature of the LRR motifs, the LRRC37 family is thought to encode single-span transmembrane proteins with the N-terminal portion comprising the LRR region outside of the cell and a short C-terminal tail in the cytoplasm.

Evolutionary history of the LRRC37 family in primates

We analyzed the genomic organization of the LRRC37 family in various primate lineages, including orangutan (Pongo pygmaeus), rhesus macaque (Macaca mulatta), common marmoset (Callithrix jacchus), and ring-tailed lemur (Lemur catta) as representatives of great ape, Old World monkey, New World monkey, and prosimian lineages, respectively. Mammalian outgroup genomes included those of Mus musculus (NCBI37/mm9), Rattus norvegicus (Baylor 3.4/rn4), Canis familiaris (Broad CanFam3.1/canFam3), and Bos taurus (UMD_3.1/bosTaue6). Both the mouse and the rat genomes have two tandem copies of Lrc37 on chromosome 11 and chromosome 10, respectively. Mouse and rat Lrc37a1 and Lrc37a2 genomic regions show synteny to human chr17:40.9 Mb (LRRC37A4) and chr17:42.4 Mb (LRRC37A5) (hg18), respectively, based on the order of flanking genes. The cow genome has one copy of LRRC37 on chr19:45 Mb and the dog genome has one complete copy on chr9:10 Mb, both syntenic to human chr17:40.9 Mb (Supplemental Note). This comparison of mouse, rat, dog, and cow genomes suggests that the human LRRC37A4 (chr17:40.9 Mb) corresponds to the ancestral mammalian single-copy locus and that two tandem copies originated in the Euarchontoglires ancestor, likely corresponding to the human LRRC37A4 and LRRC37A5 (Fig. 2B).

Since recently duplicated genes are frequently collapsed or missing from the draft genome assemblies (She et al. 2004; Alkan et al. 2011b), we identified large-insert clones from each primate species by genomic library hybridization. We screened BAC libraries (five- to sixfold coverage) from orangutan (CHORI-253), rhesus macaque (CHORI-250), common marmoset (CHORI-259), and ring-tailed lemur (LB45-2) using a PCR-amplified probe (Anc409) corresponding to the exon 8/intron 8 canonical gene structure shared among most human duplicate copies. Due to the partial deletion of this locus, a second probe (Anc419, corresponding to sequence between introns 4 and 5) was used to screen the CH250 library (see below and Supplemental Note). We obtained a total of 46 (CH253), 36 (CH250), 28 (CH259), and 19 (LB-2) positive clones. Based on the genomic coverage of each library, this suggests the presence of seven, six, five, and three copies in the orangutan, macaque, marmoset, and lemur genomes, respectively. All clones were PCR amplified using conserved primers, and PCR products were sequenced. Based on the level of sequence divergence, we distinguished ten copies in the orangutan genome (PPY1–10) (Supplemental Table S2), seven copies in the macaque genome (MMU1–7) (Supplemental Table S3), seven copies in the marmoset genome (CJA1–7) (Supplemental Table S4), and twelve copies in the lemur genome (LCA1–12) (Supplemental Table S5; Supplemental Note). We identified the map location of each of these putative copies by BAC-end sequencing of clone inserts and mapping these sequences back to the human reference and respective primate genome sequence assemblies. We mapped BAC clones representing all identified copies on the respective primate metaphase chromosomes using FISH (fluorescent in situ hybridization) assays (Figs. 2A, 3C).

We determined that the orangutan has ten copies of LRRC37 on chromosome 17 (referred to here as PPY1–10) and a retrocopy on chromosome 10 (PPY11). FISH mapping identified three locations on 17q for orangutan copies from the centromere to the telomere: a first cluster with PPY2, 3, and 6, and 7; a second one with PPY1, 4, 5, 9, and 10; and a last location with PPY8 (Fig. 2A; Supplemental Fig. S2). For seven of the orangutan loci (PPY2, 4, 6, 7, 8, 9, and 10), the BAC-end sequence (BES) placements were all inconsistent with the ponAb2 reference, suggesting potential orangutan assembly errors. Using the higher quality human genome (hg18) as a reference, we found that several of the loci map discordantly when compared to human, suggestive of human-specific duplicative transpositions. PPY1 maps to the orthologous location with respect to human A-type (LRRC37A, A3, and A4) copies; PPY2 maps to the orthologous location with respect to human chr17:26.1–26.2 and 27.5–27.6 Mb, containing the B-type copies LRRC37B6, B3, B, and B5; and PPY6 and 7 have at least one end mapping to chr17:26–27 Mb, hg18. PPY3, 8, and 11 map to an orthologous location with respect to human LRRC37B4, C, and E copies, respectively, whereas PPY4, 5, 9, and 10 copies map to the orthologous location of human LRRC37A5.

The rhesus macaque genome has seven copies of LRRC37 on chromosome 16 (MMU1–7) and a retrocopy on chromosome 9 (MMU8)—syntenic to human chromosomes 17 and 10, respectively. MMU1, 2, and 7 map to a cluster (chr16:55 Mb, as annotated in
Figure 1. (Legend on next page)
Primate LRRC37 gene family evolution

Figure 1. (A) LRRC37 family organization in human. The structure of complete and partial LRRC37 genes according to the reference human genome (hg18) is shown. (Left) Chromosomal band location of each locus on human chromosome 17 ideogram (only LRRC37 maps to chromosome 10 [gray] and is a retrocopy; LRRC37D does not have a chromosomal assignment). Genes are indicated as expressed (red) or not expressed (blue) based on the analysis of EST (expressed sequence tag) data. (Right) Variation in LRRC37 structures. Exons are represented as blocks connected by horizontal lines with arrowheads depicting introns. Exon 1 of A, A2, A3, and A4 copies (1a, in red) is longer than exon 1 of the B copy (1b, in orange); the 8 and 82 copies have exon 9 sequence split into two, 9b’ and 9b”. Exon 1 with deletions and/or insertions is red-orange. Exon 9 is dark green; exon 9 with deletions and/or insertions is light green. LRRC37A copy has one additional exon located at 5’ but lacks exon 12, and exon 13 is not transcribed. LRRC37A copy lacks exons 6 and 7 but carries a tandem duplication of exons 9 and 10. LRRC37A has a deletion of a single nucleotide at the end of the exon 1 sequence, causing a frameshift of the reading frame and the formation of a stop codon at the end of exon 1. LRRC37B is a fusion gene: exons 7, 8, 9 match exons 9b’, 9b”, 10, and 15 of LRRC37, whereas exons 4, 5, and 6 derive from a duplication of exons 4, 5, and 6 of SMAD-specific EZ ubiquitin protein ligase 2 (SMURF2) (NM_022739) (in pink), which maps at 17q24.1. Exonated introns are gray striped. RefSeq genes annotated in these loci are reported, following the same display conventions as in the UCSC Genome Browser. (B) Human LRRC37A and LRRC37B gene and protein structures. (Top panel) LRRC37A and LRRC37B cDNA predict an ORF with the methionine start codon (exon 1) and a stop codon (exon 14 and exon 15, respectively). LRRC37 encodes a predicted protein of 1700 amino acids with a molecular weight of 188 kDa. LRRC37B encodes a predicted protein of 947 amino acids with a molecular weight of 106 kDa. (Middle panel) Predicted structure of human LRRC37A and LRRC37B proteins according to SMART and LRScan tools. Signal peptides (red), LR motifs (green), LRR-NT and LRR-CT domains (turquoise), transmembrane helices (pink), and other repetitive motifs not associated with a known domain (blue) are shown. Exon boundaries are indicated with vertical dashed lines. (Bottom panel) An alignment of LRR regions of human LRRC37A and LRRC37B. The six LRR motifs and the final LRR are shown with the conserved leucine and asparagine residues highlighted in red and green, respectively. Boundaries between LRR motifs are marked with vertical dashed lines. LRR-NT and LRR-CT domains are shown with the conserved cysteine residues highlighted in turquoise. Exon boundaries are indicated with blue arrowheads.
likelihood) using two noncoding regions corresponding to intron 1 and intron 8 (Fig. 4; Supplemental Note). We obtained sequences from reference genome assemblies (human and mouse hg18 and mm9, and CJA7 from WUGSC 3.2/calJac3), from PCR amplification and sequencing of BACs (orangutan, marmoset, and lemur), or from draft sequencing of BACs (macaque CH250, marmoset CH259, and lemur LB2) (Supplemental Note). The intron 1 phylogenetic tree included human, orangutan, macaque, marmoset, lemur, and mouse sequences, whereas the intron 8 phylogenetic tree included human, macaque, lemur, and mouse sequences. Because of the differences in gene structure, corresponding sequence was not obtained for all clones.

The topology of the phylogenetic trees confirmed the lineage-specific expansions and orthologous relationships previously inferred by BES and FISH mapping. For example, MMU3 and MMU4 are monophyletic and are part of a clade including HSA-F and CJA7 orthologs. In addition, the phylogeny confirms the orthologous relationship among HSA-C, PPY8, and MMU5-6 copies. HSA-C and HSA-D, as well as MMUS and MMU6, are derived from two recent duplication events in the human and macaque lineages, respectively. These two clades can be inferred by the tree topology since the copies are located in two separate loci that were unlikely affected by nonallelic gene conversion and as a result did not experience strong sequence homogenization.

In contrast, the increase in copy number within the great ape lineage appears to be driven by the expansion of A and B types. Interestingly, the process appears to have occurred independently in both human and orangutan lineages, maintaining a similar balance in copy number. In humans there are five A subtypes (A, A2, A3, A4, and A5), and in orangutan there are five A subtypes (PPY1, 4, 5, 9, and 10) with at least A4 and A5 existing in their common ancestor. Comparably, there are four human (B, B2, B3, and B7) and four orangutan (PPY2, 3, 6, and 7) B-type copies that belong to the same monophyletic group. Similar lineage-specific expansions are observed in each of the primate lineages. For example, copies CJA1–6 are specific to the marmoset lineage and expanded (or homogenized) after separation from the other primate lineages (bootstrap support = 100). Similarly, we have identified multiple copies of lemur LRRC37—all of

**Figure 2.** (A) Chromosomal organization of the LRRC37 family among primates. FISH results on primate metaphase chromosomes are shown for *Lemur catta* (LCA), *Callithrix jacchus* (CJA), *Macaca mulatta* (MMU), and *Pongo pygmaeus* (PPY). The following BACs were used as probes in each species: LCA chr17 (LB2-169E11 in red), CJA chr5 (CH259-152N22 in red and CH259-145C2 in green), MMU chr16 (CH250-219M3, CH250-221J22, and CH250-197J22 in red, green, and blue, respectively), and PPY chr17 (CH253-10O4, CH253-149E2, and CH253-167E20 in red, green, and blue, respectively). The corresponding LRRC37 family members are shown next to each chromosome. (B) Model of the LRRC37 family evolution. Different LRRC37 types are colored according to the human schematic. For other species, genes corresponding to syntenic locations or derived from lineage-specific duplications are colored corresponding to their human paralog. Since lemur copies are highly homogenized, we could not determine whether the ancestor of the human F copy emerged in the primate or the Haplorhini ancestor (dashed lines). Orthology relationships between human A type (A4 and A5) and marmoset 1–6 copies could not be determined (red-orange lines).
which are clustered and virtually identical at the sequence level (Figs. 3 and 4).

Evolutionary analysis of the ancestral locus

We analyzed the evolutionary conservation of the LRRC37 ancestral locus in mammals and aligned the corresponding orthologous coding sequences of cow LRRC37A, dog LRRC37A, mouse Lrrc37a1, rat Lrrc37a1, macaque MMU1, and human LRRC37A4, as well as the duplicated coding sequences of mouse Lrrc37a2, rat Lrrc37a2, and human LRRC37A. With the exception of human LRRC37A4, the exon-intron structure is conserved for exons 1–11 (Fig. 5). LRRC37 has two portions conserved differently in mammalian evolution: a variable in length, repetitive first exon and a more conserved portion corresponding to exons 2–8 containing the LRR region (Supplemental Note). We searched for evolutionary signals of natural selection on the LRRC37 ancestral protein coding sequence by using maximum likelihood models that estimate evolutionary rates for individual branches in the tree for coding sequence corresponding to exons 2–8 (Fig. 5). (Note: Due to the variable length and extensive amino acid replacement within exon 1, it was impossible to generate a high-quality sequence alignment for this portion.) Using a “one-ratio” model to test uniform $d_S/d_\Omega$ across all the branches of the phylogenetic tree, we noted that compared to a neutral model ($\omega = 1$) the protein evolution for the LRRC37 model (exons 2 to 8) across all lineages is under negative constraint ($\omega = 0.4175, P$-value < 0.001) (Fig. 5; Supplemental Table S6). Importantly, we find that the duplicated copies of mouse and rat (Lrc37a2) show greater selective constraint than that of the ancestral locus (Lrc37a1). We assessed branches for positive, neutral, and purifying selection and found that at the individual species level only the dog, cow, and macaque branches are significantly conserved ($P$-values 0.002, 0.04, and 0.048, respectively) (Supplemental Note).

Alternative splicing and copy-specific expression for human LRRC37

RT-PCR experiments showed a striking increase in the complexity in alternative splicing particularly among human tissues (Fig. 6). We subcloned and sequenced RT-PCR amplification products from
various sources (Fig. 7; Supplemental Note) revealing extensive alternative splicing of exons encoding the LRR region. In human, we detected the expression of LRRC37A, A2, A3, A4, and B copies, with no products derived from A5 and C copies, despite having both primer annealing sites. We characterized a total of 24 different splice variants in human and six in mouse and assigned them to duplicate copies wherever possible based on diagnostic paralogous sequence differences (Fig. 7): 19 were assigned to LRRC37A, A2, A3, and B paralogs with five distinct forms corresponding to LRRC37A4. It is notable that most of this splicing diversity was observed in tissues other than the testis. We cloned and sequenced only six splice forms from both human and mouse testis. The alternative splicing creates different transcripts encoding putative LRRC37 protein isoforms with a variable number of LRR motifs. This ranges from no motifs, when exon 1 is spliced directly with exon 9, to six LRR motifs, when no exon is removed. Notably, the exclusion of exon 8 in mouse and human LRRC37A, A2, A3, and B copies removes the LRR C-terminal capping domain and shifts the ORF, introducing a premature stop codon at the beginning of exon 9. The resulting transcripts may be subjected to nonsense-mediated decay (Wilusz et al. 2001) or encode shorter putative proteins that terminate after the LRR region and lack the transmembrane domain codified by exon 10. These proteins, if expressed, may be secreted as they retain the signal peptide at the N-terminus. Because of this potential functional difference, we categorize transcripts into two groups according to the presence or absence of exon 8: the long or a isoforms, with exon 8, and the short or b isoforms, without exon 8 and with the reading frame ending at the beginning of exon 9. The most abundant splice form in all samples tested was the complete one (1a), having all exons from 2 to 8. The mouse 12b variant has a different donor splice site in intron 7 that introduces a stop codon at the end of exon 7. The distribution of various splice products differs dramatically depending on RNA source material. In human testis, 88% of the sequenced transcripts show all exons (variant 1a), in contrast to cerebellum where this drops to 26%. In mouse testis, splice form 1a predominates with very little alternative splicing observed as suggested by the RT-PCR profile on agarose gel (Fig. 6B). Human LRRC37A4 transcripts are unique—the gene structure lacks exons 6 and 7 and has a different splice acceptor site within intron 8 creating a longer (147 bp) exon 9 (light blue box in Fig. 7). This additional sequence generates a frameshift in the reading frame of all LRRC37A4 transcripts, which, in turn, leads to a stop codon at the beginning of exon 9 and a putative truncated version of the protein. Sequencing suggests that expression levels of different human paralogs varied depending on cell line or tissue. We designed four distinct quantitative RT-PCR assays to evaluate the expression of LRRC37A/A2, A3, A4, and B copies (Supplemental

---

**Figure 4.** LRRC37 family phylogeny. Maximum likelihood phylogenetic trees were constructed using noncoding regions from intron 1 (left) and intron 8 (right). Trees are drawn to scale, with branch lengths measured in the number of substitutions per site and bootstrap values (1000 replicates) shown. Human and orangutan A- and B-type copies are shaded blue and yellow, respectively; the monophyletic clade of HSA-F paralogs is gray shaded; branches with HSA-C, HSA-D, PPY8, MMU5, and MMU6 orthologous copies are shaded green. Copies clearly mapping to syntenic locations are signed by a superscript.
Subcellular localization of LRRC37A

To determine the subcellular localization and provide some potential functional insights into the LRRC37 family, we generated recombinant constructs carrying N-terminal or C-terminal FLAG-tagged human LRRC37A full-length coding sequence and developed polyclonal antibodies able to recognize human LRRC37 proteins. HeLa cells transiently transfected with recombinant plasmids were analyzed through immunofluorescence and Western blot using anti-FLAG and anti-LRRC37 antibodies. Immuno-fluorescence analysis showed that LRRC37A protein localizes to cellular vesicles, including the Golgi apparatus, and to the plasma membrane (Fig. 8A). Western blot analysis of both cell lysates and conditioned media shows two principal bands with a molecular weight >260 kDa common to both antibodies (Fig. 8B). This suggests that LRRC37A (190 kDa) has a very low electrophoretic mobility, potentially conferred by a moiety encoded by exon 1 (Supplemental Note).

A time-course experiment (Supplemental Note) showed that C-terminal FLAG-tagged LRRC37A recombinant protein first localizes in the Golgi apparatus and then moves to the plasma membrane (Fig. 9). Our analysis indicates that the recombinant LRRC37A is routed to the endoplasmic reticulum through its signal peptide and transmembrane domain, then to the Golgi apparatus, and finally delivered to the plasma membrane by vesicles. Here, at the cell surface, it is likely cleaved within the extracellular juxta-membrane domain and released in the extracellular space (Supplemental Fig. S13; Supplemental Note). Interestingly, over-expression of LRRC37A in HeLa cells, which do not endogenously express this protein at detectable levels, causes a deformation of plasma membrane shape and the generation of filopodia-like protrusions (Fig. 9). At later stages of the expression there is an accumulation of circular anucleated elements stained by the anti-FLAG antibody, potentially as a consequence of protein over-expression (Fig. 9).

Discussion

Evolutionary dynamism of LRRC37 in primates

The organization of the LRRC37 family in various genomes reveals a remarkable dynamism with a trend toward recurrent duplication in primate lineages when compared to other mammalian genomes. Both FISH and phylogenetic analyses suggest independent expansions of LRRC37 in different primate lineages. The duplication pattern becomes increasingly interspersed among great apes and associated with intrachromosomal rearrangements on hominid chromosome 17 (Zody et al. 2006). In lemurs, for example, the duplication pattern is tandemly configured, reminiscent of nonprimate mammalian genomes (Tuzun et al. 2004; She et al. 2008; Liu et al. 2009). Independent duplications of both A- and B-type LRRC37 occurred specifically in the human and orangutan lineages becoming dispersed to more locations along the long arm of chromosome 17 (Figs. 1, 2) and the “cores” of more complex and larger blocks of segmental duplications composed of a greater number of duplons (Jiang et al. 2007). We note that human LRRC37A4 and A2 loci correspond to the inversion breakpoints of the MAPT (microtubule associated protein tau) 17q21.31 region (Zody et al. 2008), one of the most structurally complex and evolutionarily dynamic regions of the human genome (Cruts et al. 2005; Stefansson et al. 2005). This duplication and restructuring has been accompanied by local changes in gene structure. The human LRRC37F copy, for example, is partial and does not retain the exon 8/intron 8 region, in contrast to its macaque orthologs (MMU3 and MMU4), which maintain the canonical gene structure. Likewise, the human
LRRC37C copy retains all exons, but its macaque orthologs (MMU5 and MMU6) have lost a portion of exon 5. These data raise the possibility that different paralogs have assumed functional roles independently in different lineages as specific copies have become pseudogenized.

**Expanded LRRC37 transcriptional diversity in humans**

Changes in the duplication architecture of LRRC37 were accompanied by shifts in the transcriptional landscape and an overall increase in transcript diversity. In mouse, both Lrc37 copies are expressed; in humans, we find that at least 13 of the 18 copies produce transcripts. The tissue expression profile is testis-exclusive in mouse, whereas in macaque we observe a pattern that is predominantly testis-specific but with a low level of expression in other tissues. In human, the pattern of expression becomes more ubiquitous. Although the highest level of expression still occurs in the testis, in human, we observe high levels of expression in other tissues, such as the cerebellum and thymus. Notably, the higher expression level in the cerebellum appears to have emerged in the great ape ancestor and to have been conserved over the last 15 million years of great ape evolution (Supplemental Note). We propose that a testis-specific pattern of expression was the most likely ancestral state of mRNA expression and that during primate evolution a broader landscape was acquired through the acquisition of novel regulatory elements via segmental duplication (Bekpen et al. 2012). This is consistent with the observation that rat peptides corresponding to the first exon of Lrrc37a2 are processed, secreted into the lumen of the prostate, and are androgen regulated (Heyns et al. 1982; Hemschoote et al. 1988; De Clercq et al. 1992). We point out that the macaque tissues were derived from a pig-tailed macaque (Macaca nemestrina), whereas genomic organization data were from the rhesus macaque (Macaca mulatta). Although they are closely related species, potential differences in their genomic organization may exist.

Our sequencing results also suggest remarkable diversity in LRRC37 transcriptional splice products when compared to mouse. In human, much of the alternative splicing involves exons encoding the LRRs, whose sequences have been conserved during mammalian evolution. Since LRR motifs are involved in protein-ligand interaction, the translation of these splice variants would potentially generate an expanded repertoire of proteins with the potential to interact with different ligands or with the same ligand at different strengths and affinities. From our library of sequenced products, we observed some interesting patterns. We found, for example, coordinated splicing of exons 4 and 5, as well as of exons 6 and 7, in all tissues with the exception of the

**Figure 6.** Tissue expression patterns of the LRRC37 family. RT-PCR and RT-qPCR results of the amplification of cDNA prepared from a panel of human tissue total RNA (Clontech) and from macaque and mouse tissue total RNA. *UBE1* amplification was used as a control. (A) Human LRRC37A structure indicating the position of the primers used in RT-PCR experiments (red arrows). Forward primers were designed at the end of exon 1 and reverse primers at the beginning of exon 9 to amplify the LRR region. (B) Mouse Lrc37a is expressed only in testis. The length of the complete LRR region of Lrc37a2 is 767 bp. Minor products with shorter length are present. Degenerated primers have been designed for macaque and human to amplify all copies of the family. In both, the LRRC37 family is expressed in all the tissues tested, with the highest expression in testis. The expected size of the complete LRR region is 763 bp, but several products of smaller size are noted. The same forward primers and a reverse primer specific for the human LRRC37A4 copy have been tested in human tissue cDNA. There is no expression of the LRRC37A4 copy in liver and skeletal muscle. (C) Quantitative expression profiling of the LRRC37 family in macaque tissues and of LRRC37A, A2, A3, and A4 in human tissues. C value is shown and calculated by comparative C method (Livak and Schmittgen 2001) with UBE1 as the reference gene. In macaque, expression values are relative to heart. In human, the highest value of expression in the assay (testis) is set to 100. (Error bars) Standard error of the mean.
testis. Removal of exon 8 disrupts the ORF and generates a stop codon at the beginning of exon 9, potentially producing two different groups of LRRC37 proteins: long isoforms, with the complete ORF, carrying the final transmembrane domain versus short isoforms, without the transmembrane domain. Similarly, a novel splice-acceptor within intron 8 of LRRC37A4 (Eichler 2001) adds 147 bp to exon 9 resulting in the formation of a premature stop codon and a protein lacking the transmembrane domain. These data suggest a complex interplay between segmental duplication and transcriptional novelty.

Functional insights into the LRRC37 protein

LRRC37 proteins, in general, are predicted to carry both a signal peptide and a transmembrane domain. Our analysis of one member of this gene family, a full-length isoform of LRRC37A2, confirms subcellular localization to both the Golgi and plasma membrane. Despite the absence of a clear cleavage or secretory signal sequence, transfection experiments in HeLa cells suggest that this isoform of LRRC37A is cleaved and released extracellularly. Interestingly, functional data about the rat ortholog Lrrc37a2 support the evidence that mammalian LRRC37 proteins are secreted. Notwithstanding, rat Lrrc37a2 experienced a peculiar expansion of a 300-bp tandem unit in the exon 1, each unit codifying for a secreted polypeptide. It is noteworthy that rat prostate cells translate a 20.6-kbp mRNA to produce at the end numerous 38-residue polypeptides. In this context, the role of the evolutionarily highly conserved LRR region remains completely unresolved.

With the exception of the first exon, the mammalian LRRC37 structure (exons 2–8) has been generally maintained for more than 100 million years of evolution and shows significant evidence of purifying selection (Fig. 5) with both rat and human proteins showing experimental evidence of secretion (Heyns et al. 1982; Hemschoote et al. 1988). The rate, however, has not been constant. We note that these estimates are based on comparisons between lineages and are unable to disentangle alternating bouts of purifying and positive selection that might appear in aggregate neutral. The human copy corresponding to the ancestral locus (LRRC37A4) has undergone some of the most radical changes in its exon-intron structure, potentially encoding the most novel protein when compared to the mammalian archetype. We propose that these changes occurred as a consequence of even greater relaxation of selection after the emergence of additional expressed copies by segmental duplication. It is interesting that the same copy acquired a higher expression in the cerebellum.

Although there is evidence that the rat ventral prostate secretes LRRC37 polypeptides (Heyns et al. 1982; Hemschoote et al. 1988; De Clercq et al. 1992), we could not detect any band in a Western blot of different human secreted fluids (seminal plasma, tears, saliva, blood plasma, breast milk, and sweat) probed with the...
many protrusions that they ultimately lose their integrity and are observed. The significance of this observation is unclear. Although this cellular phenotype might be determined by the LRRC37A protein itself, it may also be an artifact of its overexpression in transfection assays. In this context, since LRR motifs in characterized proteins are used for intermolecular or intercellular interactions, their presence suggests the LRRC37 proteins may bind some ligands. Finally, our overall observations represent only the initial step toward understanding the function of LRRC37. Separate efforts are needed to elucidate the human tissue and cell type that produce LRRC37 proteins and to obtain insights into potential ancestral and novel functions that have emerged.

Conclusions
The LRRC37 family is a member of 14 “core” duplicons that have expanded in the primate lineage since the divergence of the human lineage from other mammalian species (Jiang et al. 2007; Marques-Bonet and Eichler 2009). These cores represent seeds of genomic instability upon which the complex interspersed segmental duplication architecture of great ape and human chromosomes has evolved (Jiang et al. 2007). We have hypothesized that the selective disadvantage of this architecture predisposing the human genome to recurrent rearrangement has been offset by the emergence of novel genes with different functions and expression profiles with respect to their ancestral genes (Marques-Bonet and Eichler 2009). Our detailed analysis of the evolutionary history of the LRRC37 family provides some support for this model. The increase in copy number of the LRRC37 duplicon appears to have evolved from a tandem organization (e.g., lemur) to one that is increasingly dispersed in the hominid lineage. Although different copies have expanded in various primate lineages, the LRRC37A and LRRC37B copies have specifically expanded in humans and great apes and are preferential sites of both recurrent inversion polymorphisms (Zody et al. 2008) as well as rearrangements associated with disease (Bengesser et al. 2010). Our detailed expression and transcript analyses indicate both greater diversity and a broader expression profile in humans with marked increases in the cerebellum and thymus when compared to a testis-only

Figure 8. Subcellular localization of N- and C-terminal FLAG-tagged LRRC37A. (A) HeLa cells transiently transfected with N- and C-terminal FLAG-tagged LRRC37A constructs and probed with anti-FLAG antibody. LRRC37A tagged at the N-terminus does not show strong localization, whereas LRRC37A tagged at the C-terminus shows a plasma membrane localization (magnification, 63X). (B) Western blot of lysates and conditioned media of HeLa cells transfected with FLAG-tagged LRRC37A. Lysates and conditioned media of HeLa cells transiently transfected with N-terminal (N) or C-terminal (C) FLAG-tagged LRRC37A, probed with anti-FLAG (lanes 1–3) and anti-LRRC37 (lanes 4–6) antibodies. Nontransfected HeLa cells represent the control. (C) Conditioned media were concentrated, and a comparable amount of lysates and media was loaded. Samples were probed with anti-LRRC37 antibody.

Figure 9. Time-course experiment of FLAG-tagged LRRC37A expression in HeLa cells. HeLa cells were transiently transfected with recombinant FLAG-tagged LRRC37A and probed with anti-FLAG antibody (magnification, 40×). The pictures depict the most frequent aspects shown by transfected cells at different intervals after transfection. Note the accumulation at the plasma membrane and the formation of filopodia-like protrusions. After 48 h, cells begin to lose their integrity, and enucleated cells are observed.
expression profile in the mouse. While we can only speculate on
the function of the gene, our data suggest that some members of
this gene family localize to the plasma membrane and are secreted.
One possibility may be that this protein family, similar to other
studied proteins corresponding to core duplicons (e.g., TBC1D3),
plays a role in regulating cell signaling, growth, and prolifer-
ation during development (Wainszelaum et al. 2008; Stahl and
Wainszelaum 2009). It is also intriguing that our cell transfection
studies suggest that LRRC37 may play a role in the formation of
filopodia protrusions similar to what has been reported for
other gene families that have expanded in the human lineage
(Linardopoulos et al. 2007; Guerrier et al. 2009). A critical step for-
ward will be accurately assessing genetic variation, including copy
number changes of these and other duplicate genes with respect to
human phenotypes (Sudmant et al. 2010; Alkan et al. 2011a).

Methods

Hybridization of high-density filters
Radioactive genomic hybridization of CHORI-253 (Sumatran orangutan), CHORI-250 (Indian rhesus macaque), CHORI-259
(common marmoset), and LBNL-2 (ring-tailed lemur) BAC libraries
was carried out according to the protocol available at CHORI
BACPAC resources (http://bapac.chori.org/highdensity.htm).
Anc409 and Anc419 probes were obtained by means of PCR am-
plification of human genomic DNA (Supplemental Table S7), BES
(BAC end sequences) were repeat-masked (Smit et al. 1996) and
mapped on human (hg18) and primate (ponAbe2, rheMac2, and
calJac3) genomes using the BLAT tool at the UCSC Genome
Browser (http://genome.ucsc.edu/cgi-bin/hgBlat).

Fluorescence in situ hybridization (FISH)
Metaphase spreads and interphase nuclei were prepared from a fi-
broblast cell line of Lemur catta and from lymphoblastoid cell lines
of Homo sapiens, Pongo pygmaeus, Macaca mulatta, and Callithrix
jacchus. FISH experiments were performed as previously described
(Lichter et al. 1990). Digital images were obtained using a Leica
epitfluorescence microscope equipped with a cooled CCD camera.
Fluorescence signals detected with Cy3, Fluorescein, and Cy5
filters and chromosomes and nuclei images detected with DAPI
folds were recorded separately as grayscale images. Pseudocolor-
ing and merging of images were performed using Adobe Photo-
shop software.

Phylogenetic and evolutionary analysis
A 2-kbp region from intron 1 and a 1.2-kbp region from intron 8
were PCR-amplified and sequenced from BAC clones representing
different loci in each species. We reconstituted human, chimpan-
zee, orangutan, and macaque complete gene models based on
RefSeq entries, genome assemblies, and BAC sequences. Multiple
sequence alignments were performed using Clustal W (Thompson
et al. 1994), and phylogenetic analyses were conducted in MEGA5
(Tamura et al. 2011). The evolutionary histories were inferred by
using the maximum likelihood method based on the Kimura two-
parameter model (Kimura 1980). A bootstrap test with 1000 rep-
licates was conducted to evaluate the statistical significance of each
node (Felsenstein 1985).

We aligned ancestral locus coding sequences using RevTrans
(Wernersson and Pedersen 2003). Tests of selection (ω = dS/dD) were
performed by maximum likelihood using PAML (Yang 1997) ap-
plying the Branch Model to calculate ω under different scenarios at
different times during evolution. The likelihood ratio test (LRT)
was used to assess the significance of different values of ω for dif-
ferent models.

Expression analyses
Mouse (wild-type mouse C57BL/6J) and macaque (Macaca nem-
estrina) tissue total RNA were extracted using the RNeasy Fibrous
Tissue Midi Kit (Qiagen), following manufacturer recommenda-
tions. HeLa cells and G248 human lymphoblastoid cell line total
RNA were extracted using the RNeasy Mini Kit (Qiagen), following
manufacturer recommendations. Human tissue total RNA panels
were from Clontech, Stratagene, and Capital Biosciences. 1 µg
of isolated total RNA as well as of commercial human tissue total RNA
(Clontech, Stratagene, and Capital Biosciences) was used for re-
verse transcription with Transcriptor High Fidelity cDNA Synthesis
Kit (Roche) using Oligo(dT)18 primers. PCR reactions were per-
formed with PCR Master (Roche), using standard PCR cycle con-
ditions (initial denaturation of 3 min at 94°C; 35 cycles 30 s at
94°C, 30 s at 55°C, 1 min at 72°C; final extension 10 min at 72°C). The
amplification with primers designed for UBE1 was used as a control.
Quantitative gene expression profiling studies were performed
using the LightCycler 480 SYBR Green System (Roche) with three replicates for each sample. The primers used are listed in
Supplemental Table S7. CT values were elaborated by the compar-
ative Ct method (Livak and Schmittgen 2001). All PCR and qPCR
products were sequence verified.

Sequencing
PCR products were sequenced using PCR primers or were cloned
and sequenced using vector primers (Supplemental Table S7; Supple-
mental Note). DNA from BAC clones was extracted using the
BAC Prep Protocol (Schein et al. 2004), and BAC ends were
sequenced (if not available in the NCBI Trace Archive) using
sp6_BAC and t7_BAC primers. Sequencing was performed on an
ABI PRISM 3100 Genetic Analyzer using the BigDye Terminator
v3.1 Chemistry (Applied Biosystems) according to manufacturer
instructions.

Generation of FLAG-fusion constructs
Human LRRC37A coding sequence was amplified from HeLa cells
cDNA using PhTurbo DNA polymerase (Stratagene) using touch-
down cycle conditions (initial denaturation 3 min at 94°C; 10
cycles 30 s at 94°C, 30 s at 65°C, 10 min at 72°C; 10 cycles 30 s at
94°C, 30 s at 60°C, 10 min at 72°C; 30 cycles 30 s at 94°C, 30 s at
55°C, 10 min at 72°C; final extension 10 min at 72°C) and sub-
cloned in pFLAG-CMV-6c and pFLAG-CMV-5.1 mammalian ex-
pression vectors (Sigma) to generate FLAG-tagged constructs at
amino and carboxyl terminus, respectively. The recombinant
plasmids were transformed into chemically competent E. coli cells
(Invitrogen). Recombinant plasmids were sequence verified using
primers listed in Supplemental Table S7.

Generation of anti-LRRC37 antibodies
Rabbit polyclonal antibody against human LRRC37 proteins was
produced by the Pocono Rabbit Farm and Laboratory Inc., using as
immunogen the peptide ETLEDIQSSSLQQEA (position 228–242 of
human LRRC37 protein).

Cell culture and transient transfection
HeLa cells were grown in DMEM High Glucose (Invitrogen) sup-
plemented with 10% fetal bovine serum, 100 µg/mL streptomycin,
and 100 μg/mL penicillin. FLAG-fusion constructs were extracted using the EndoFree Plasmid Maxi Kit (Qiagen), following manufacturer instructions. HeLa cells were transiently transfected with FuGENE HD Transfection Reagent (Roche), using a 6:2 ratio between reagent and DNA, according to manufacturer instructions.

Immunofluorescence
HeLa cells were seeded on coverslips and after 24 h were transfected. Transfected and nontransfected control cells were fixed with 4% paraformaldehyde in phosphate-buffered saline (PBS) for 10 min at room temperature (RT). Cells were washed with PBS and permeabilized with PBS containing 0.1% Saponin (Fluka) for 10 min at RT. Cells were then blocked with PBS, 0.1% Saponin, and 3% bovine serum albumin (BSA) for 1 h at RT and probed with rabbit polyclonal anti-FLAG antibody (Abcam) overnight at 4°C. Cells were washed extensively with PBS, 0.1% Saponin, and then incubated with Alexa Fluor 488 goat anti-rabbit IgG (H+L) secondary antibody (Molecular Probes) for 30 min at RT and washed. Coverslips were mounted on slides with ProLong Gold Antifade Reagent (Molecular Probes) and observed under a Leica DMRXA2 epifluorescence microscope equipped with a cooled CCD camera (Princeton Instruments).

Western blot
Transfected and nontransfected control cells were collected 24 h after transfection and lysed in RIPA buffer (Sigma) supplemented with Complete Protease Inhibitor Cocktail Tablets (Roche). The lysates were centrifuged for 1 h at 13,000 g for 1 h. Media were concentrated on Amicon Ultra-15 centrifugal filter devices (Millipore). Lysates and 100 μg/mL penicillin. FLAG-fusion constructs were extracted using the EndoFree Plasmid Maxi Kit (Qiagen), following manufacturer instructions. HeLa cells were transiently transfected with FuGENE HD Transfection Reagent (Roche), using a 6:2 ratio between reagent and DNA, according to manufacturer instructions.

Immunofluorescence
HeLa cells were seeded on coverslips and after 24 h were transfected. Transfected and nontransfected control cells were fixed with 4% paraformaldehyde in phosphate-buffered saline (PBS) for 10 min at room temperature (RT). Cells were washed with PBS and permeabilized with PBS containing 0.1% Saponin (Fluka) for 10 min at RT. Cells were then blocked with PBS, 0.1% Saponin, and 3% bovine serum albumin (BSA) for 1 h at RT and probed with rabbit polyclonal anti-FLAG antibody (Abcam) overnight at 4°C. Cells were washed extensively with PBS, 0.1% Saponin, and then incubated with Alexa Fluor 488 goat anti-rabbit IgG (H+L) secondary antibody (Molecular Probes) for 30 min at RT and washed. Coverslips were mounted on slides with ProLong Gold Antifade Reagent (Molecular Probes) and observed under a Leica DMRXA2 epifluorescence microscope equipped with a cooled CCD camera (Princeton Instruments).

Western blot
Transfected and nontransfected control cells were collected 24 h after transfection and lysed in RIPA buffer (Sigma) supplemented with Complete Protease Inhibitor Cocktail Tablets (Roche). The lysates were centrifuged for 1 h at 13,000 g for 1 h. Media were concentrated on Amicon Ultra-15 centrifugal filter devices (Millipore). Lysates and conditioned media of the same cells were collected and ultra-centrifuged at 200,000 × g for 1 h. Media were concentrated on Amicon Ultra-15 centrifugal filter devices (Millipore). Lysates and conditioned media were solubilized in SDS-PAGE sample buffer and separated by SDS-PAGE gel (4%–15%). Gels were electro-blotted to Hybond ECL nitrocellulose membranes (Amersham Biosciences), which were blocked and probed with rabbit polyclonal anti-FLAG antibody (Abcam) and anti-LRRC37 antibodies. Goat anti-rabbit HRP conjugated antibodies were from GE Healthcare. Protein expression levels were corrected for whole protein loading determined by staining membrane with Red Ponceau.

Data access
The sequence data from this study have been submitted to GenBank. This includes BAC clone sequencing of LRRC37 loci: orangutan AC206276.2, AC216100.2, AC210931.1, AC212980.2, AC212589.2, AC210533.4, and AC206550.4; macaque AC239129.2, AC239243.2, AC241896.2, AC240582.3, AC214495.5, AC241829.2, AC241249.2, AC242077.2, AC245843.1, and AC143065.2; marmoset AC239345.3, AC241896.2, AC240582.3, AC241249.2, AC242077.2, AC245843.1, and AC143065.2; lemur AC234058.2, AC232433.2, AC242077.2, AC245843.1, and AC143065.2; mouse AC239345.3, AC243473.2, and AC241431.2; lemur AC234058.2. In addition, sequenced PCR products used in the phylogenetic analysis are available from intron 1 (JQ582363-JQ582376) and intron 8 (JQ582377-JQ582394); sequenced RT-PCR products of the LRR region are available (JQ790561-JQ790788). BAC end sequences have been submitted to the Genome Survey Sequence (GSS) division (http://www.ncbi.nlm.nih.gov/nucgss) (Supplementary Table S8).

Acknowledgments
We thank NIH Intramural Sequencing Center (NISC) for BAC clone sequence data generation and T. Brown for manuscript editing. This work was supported by National Institutes of Health (NIH) grants HG002385 and GM058815 to E.E.E. and in part by the Intramural Research Program of the National Human Genome Research Institute, NIH. E.E.E. is an investigator of the Hughes Medical Institutes.

References
Alkan C, Coc BP, Eichler EE. 2011a. Genome structural variation discovery and genotyping. Nat Rev Genet 12: 363–376.
Alkan C, Sahildjian S, Eichler EE. 2011b. Limitations of next-generation genome sequence assembly. Nat Methods 8: 61–65.
Bailey JA, Eichler EE. 2006. Primate segmental duplications: Crucibles of evolution, diversity and disease. Nat Rev Genet 7: 552–564.
Bekpen C, Tastekin I, Siwara P, Akdis CA, Eichler EE. 2012. Primate segmental duplication creates novel promoters for the LRRC37 gene family within the 17q21.31 inversion polymorphism region. Genome Res 22: 1050–1058.
Bengesser K, Cooper DN, Steimann K, Kluwe L, Chuzhanova NA, Wimmer K, Tatarigba M, Tirschert S, Mautner VE, Kehrer-Sawatzki H. 2010. A novel third type of recurrent NF1 microdeletion mediated by nonallelic homologous recombination between LRRC37B-containing low-copy repeat sequences in 17q11.2. Hum Mutat 31: 742–751.
Bosch N, Caceres M, Cardone MF, Carreras A, Ballana E, Marmoll L, Estivill X. 2007. Characterization and evolution of the novel gene families originated by multiple duplication and rearrangement events. Hum Mol Genet 16: 2572–2582.
Bueno DE, Sunaga DY, Kobayashi GS, Aguena M, Raposo-Amaral CE, Masotti C, Cruz LA, Pearson PL, Passos-Bueno MR. 2011. Human stem cell cultures from cleft lip/palate patients show enrichment of transcripts involved in extracellular matrix modeling by comparison to controls. Stem Cell Rev 7: 446–457.
Chen Y, Aulia S, Li L, Tang BL. 2006. AMIGO and friends: An emerging family of brain-enriched, neuronal growth modulating, type I transmembrane proteins with leucine-rich repeats (LRR) and cell adhesion molecule motifs. Brain Res Rev 51: 265–274.
Ciccarelli FD, von Mering C, Suyama M, Hutterale R, Bork P. 2005. Complex genomic rearrangements lead to novel primate gene function. Genome Res 15: 343–351.
Cruts M, Rademakers R, Gijseleik I, van der Zee J, Dermaut B, de Poorter T, de Rijk P, Del-Faverio J, van Broeckhoven C. 2005. Genomic architecture of human 17q21 linked to frontotemporal dementia uncovers a highly homologous family of low-copy repeats in the r region. Hum Mol Genet 14: 1753–1762.
De Clercq N, Hemschoote K, Devos A, Peeters B, Heyns W, Rombaerts B. 1992. The 4.4-kilodalton proline-rich polypeptides of the rat ventral prostate are the proteolytic products of a 657-kilodalton protein displaying highly repetitive sequences and encoded in a single exon. J Biol Chem 267: 9884–9894.
Dokal J, Walseke K, Albiksh S, Hékamp K, O’Keeffe S, Okafujii T, Miller SF, Tear G, Mitchell JK. 2007. The extracellular leucine-rich repeat superfamily: A comparative survey and analysis of evolutionary relationships and expression patterns. BMC Genomics 8: 320. doi: 10.1186/1471-2164-8-320.
Eichler EE. 2001. Recent duplication, domain accretion and the dynamic mutation of the human genome. Trends Genet 17: 661–669.
Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39: 783–791.
Gentry LE, Twardzik DR, Lim GJ, Ranchalis JE, Lee DC. 1987. Expression and characterization of transforming growth factor-α precursor protein in transfected mammalian cells. Mol Cell Biol 7: 1585–1591.
Giannuzzi et al. 58 Genome Research www.genome.org...
Primate LRR37 gene family evolution

Jin H, Selje J, Whitehouse C, Morris JR, Solomon E, Roberts RG. 2004. Structural evolution of the BRCA1 genomic region in primates. *Genomics* **84**: 1071–1082.

Johnson ME, Viggiano L, Bailey JA, Abdul-Rauf M, Goodwin G, Rocchi M, Eichler EE. 2001. Positive selection of a gene family during the emergence of humans and African apes. *Nature* **413**: 514–519.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16**: 111–120.

Kobe B, Kajava AV. 2001. The leucine-rich repeat as a protein recognition motif. *Curr Opin Struct Biol* **11**: 725–732.

Lichter P, Tang CJ, Call K, Hermanson G, Evans GA, Houseman D, Ward DC. 1990. High-resolution mapping of human chromosome 11 by in situ hybridization with cosmide clones. *Science* **247**: 64–69.

Linardopoulou EV, Parghi SS, Friedman C, Osborn GE, Parkhurst SM, Trask BJ. 2007. Human subtelomeric WASH genes encode a new subclass of the WASP family. *PLoS Genet* **3**: e257. doi: 10.1371/journal.pgen.0030237.

Liu GE, Ventura M, Cellamare A, Chen L, Cheng Z, Zhu B, Li C, Song J, Eichler EE. 2009. Analysis of recent segmental duplications in the bovine genome. *BMC Genomics* **10**: 571. doi: 10.1186/1471-2164-10-571.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT method. *Methods* **25**: 402–408.

Locke DP, Hillier LW, Warren WC, Worley KC, Nazareth LV, Muzny DM, Yang SP, Wang Z, Chinwalla AT, Minx P, et al. 2011. Comparative and demographic analysis of orang-utan genomes. *Science* **330**: 641–646.

Marques-Bonet T, Kidd JM, Ventura M, Graves TA, Cheng Z, Hillier LW, Jiang L, Baker C, Malfavon-Borja R, Fulton LA, et al. 2009. A burst of segmental duplications in the genome of the African great ape ancestor. *Nature* **457**: 877–881.

Monteiro JC, Yuste L, Diaz-Rodriguez E, Esparr-Iogando A, Pandiela A. 2002. Mitogen-activated protein kinase-dependent and -independent routes control shedding of transmembrane growth factors through multiple secretases. *Biochem J* **363**: 211–221.

Nurnberger T, Brunner F, Kemmerling B, Piater L. 2004. Innate immunity in plants and animals: striking similarities and obvious differences. *Immunol Rev* **198**: 249–266.

Paulding CA, Ruvolo M, Haber DA. 2003. The Trez (USP6) oncogene is a hominoid-specific gene. *Proc Natl Acad Sci USA* **100**: 2507–2511.

Popescu MC, Maclaren EJ, Hopkins D, Dumas L, Cox M, Meltesen L, McGavran L, Wyckoff GJ, Sikela JM. 2006. Human lineage-specific amplification, selection, and neuronal expression of DUF1220 domains. *Science* **315**: 1304–1307.

Rettensnier CR, Roussel ME. 1988. Differential processing of colony-stimulating factor 1 precursors encoded by two human CDNAs. *Mol Cell Biol* **8**: 5026–5034.

Schein J, Kucaba T, Seihon M, Smaltdt L, Waterston R, Marra M. 2004. High-throughput BAC fingerprinting. *Methods Mol Biol* **255**: 143–156.

She X, Jiang Z, Clark RA, Liu G, Cheng Z, Tuzun E, Church DM, Sutton G, Halpern AL, Eichler EE. 2004. Shotgun sequence assembly and recent segmental duplications within the human genome. *Nature* **431**: 927–930.

She X, Liu G, Ventura M, Zhao S, Misceo D, Roberto R, Cardone MF, Rocchi M, Green ED, Archidiacono N, et al. 2006. A preliminary comparative analysis of primate segmental duplications shows elevated substitution rates and a great-ape expansion of intrachromosomal duplications. *Genome Res* **16**: 576–583.

Received February 13, 2012; accepted in revised form October 2, 2012.