Published in final edited form as:

Gene. 2007 September 1; 399(1): 33–45. doi:10.1016/j.gene.2007.04.028.

Conserved features of imprinted differentially methylated domains

Ariane Paoloni-Giacobino[#], Leonardo D'Aiuto, M. Cecilia Cirio, Bonnie Reinhart, and J. Richard Chaillet^{*}

University of Pittsburgh, Department of Molecular Genetics and Biochemistry, Pittsburgh, PA 15261, USA

Abstract

Genomic imprinting is a conserved epigenetic phenomenon in eutherian mammals, with regards both to the genes that are imprinted and the mechanism underlying the expression of just one of the parental alleles. Epigenetic modifications of alleles of imprinted genes are established during oogenesis and spermatogenesis, and these modifications are then inherited. Differentially methylated domains (DMDs) of imprinted genes are the genomic sites of these inherited epigenetic imprints. We previously showed that CpG-rich imperfect tandem direct repeats within three different mouse DMDs (*Snurf/Snrpn*, *Kcnq1* and *Igf2r*), each with a unique sequence, play a central role in maintaining the differential methylation. This finding implicates repeat-related DNA structure, not sequence, in the imprinting mechanism. To better define the important features of this signal, we compared sequences of these three DMD tandem repeats among mammalian species. All DMD repeats contain short indirect repeats, many of which are organized into larger unit repeats. Even though the larger repeat units undergo deletion and addition during evolution (most likely through unequal crossovers during meiosis), the size of DMD tandem repeated regions has remained remarkably stable during mammalian evolution. Moreover, all three DMD tandem repeats have a high-CpG content, an ordered arrangement of CpG dinucleotides, and similar predicted secondary structures. These observations suggest that a structural feature or features of these DMD tandem repeats is the conserved DMD imprinting signal.

Keywords

genomic imprinting; tandem repeats; primate; DNA methylation; CpG islands

1. Introduction

Imprinting and epigenetic reprogramming in mammalian germ cells, the zygote, and early embryos play crucial roles in regulating genome functions at critical stages of development. The majority of imprinted genes are distinguished by a clear difference in expression between parental alleles, such that one allele is transcriptionally active and the opposite allele is silent (Reik and Lewis 2005; Trasler 2006). This expression difference coincides with different

*Address for correspondence: Department of Molecular Genetics and Biochemistry, University of Pittsburgh, 200 Lothrop Street, Pittsburgh, PA 15261 USA, Tel: (001) 412 383 7974, fax: (001) 412 383 7984, Email address: chaillet@pitt.edu.

[#]Present address: Department of Genetic Medicine and Development, Geneva University Medical School, 1, Michel-Servet, 1211 Geneva 4, Switzerland

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

patterns of CpG methylation on the maternal and paternal alleles (Chaillet et al. 1991; Thorvaldsen et al. 1998; Reinhart and Chaillet 2005). These sex-specific patterns first appear during germ cell maturation in both parents. After fertilization, the maintenance of these gametic methylation patterns produces the difference in expression between the two parental alleles (Chaillet et al. 1991; Brandeis et al. 1993; Bartolomei and Tilghman 1997). In the germ cell lineage of the developing fetus, imprinted methylation patterns are erased, and this erasure allows for the subsequent placement of correct sex-specific methylation patterns during gametogenesis in sexually mature females and males. The net effect of periodic changes in DNA methylation patterns throughout gametogenesis and embryogenesis is the creation of an imprinting cycle that ensures a set of epigenetic imprints for every generation.

The parent-specific DNA methylation differences between alleles of imprinted genes are concentrated in small (2- to 5-kb) genomic regions called differentially methylated domains (DMDs) (Verona et al. 2003; Reinhart and Chaillet 2005). One type of DMD is established in the germ cells, such that the sperm sequence is methylated differently than the oocyte sequence (Stoger et al. 1993). This gamete-derived type of DMD is maintained in the embryo through inheritance of the gametic methylation patterns. A second type of DMD is established in the embryo, possibly as a consequence of a nearby gamete-derived DMD (Chaillet 1997; Lopes et al. 2003). Essential functions of inherited DMDs can be demonstrated by deleting them in the mouse. For example, the imprinted gene *Kcnq1* on mouse chromosome 7 is normally expressed only from the maternal allele. When the KvDMR1 DMD within an intron of *Kcnq1* is deleted from the paternally inherited chromosome, *Kcnq1* is also expressed from the paternal allele. That is, *Kcnq1* becomes biallelically expressed and not imprinted (Fitzpatrick et al. 2002). Similar losses of imprinting occur when DMDs are deleted from the mouse *Snurf/Snrpn*, *Igf2r*, *H19* and *Gtl2* genes (Thorvaldsen et al. 1998; Yang et al. 1998; Bressler et al. 2001; Wutz et al. 1997; Lin et al. 2003). In diploid somatic cells, inherited DMDs, such as those associated with the *Snurf/Snrpn*, *Kcnq1* and *Igf2r* genes, function to regulate the expression of many imprinted genes in a cluster, such that the removal of a DMD disrupts imprinting of the respective cluster of genes (Verona et al. 2003). Thus, at least some, if not all, inherited DMD sequences regulate imprinted gene expression in *cis* over long genomic distances.

The aforementioned two features of imprinted genes, the inheritance of sex-specific DNA methylation patterns and the regulation of imprinted gene expression over a long genomic distance, are fundamental molecular processes that define mammalian genomic imprinting. DMD sequences are involved in both of these processes. For example, the intronic DMD2 sequences of the mouse *Igf2r* gene become methylated in the oocyte, but remain unmethylated during spermatogenesis (Lucifero et al. 2002). Upon inheritance of these epigenetic states, DMD2 is perpetuated in all somatic cells. The paternally expressed non-coding *Air* antisense-transcript originates within DMD2. A targeted mutation of DMD2 that prematurely terminates the *Air* transcript resulted in a loss of imprinting of *Igf2r* and other, neighboring imprinted genes. However, the remaining sequences of this mutated DMD2, comprised of imperfect tandem direct repeats, retained their differential methylation (Sleutels et al. 2002). This suggests that the two fundamental molecular imprinting processes are physically separable, with the tandem repeats being associated with the establishment and maintenance of sex-specific methylation patterns. Two other studies support such a role for the tandem repeats. First, the differential methylation on the mouse KvDMR1 DMD roughly coincides with its tandem repeats (Mancini-Dinardo et al. 2006). Second, we showed that the tandem repeats of the *Snurf/Snrpn*, *Kcnq1* and *Igf2r* DMDs are all involved in establishing and maintaining parent-specific DNA methylation differences in a model transgene system (Reinhart et al. 2002; Reinhart et al. 2006). These observations are consistent with the finding that CpG islands of imprinted genes very often contain arrays of tandem repeats (Hutter et al. 2006).

Despite sharing a functional role in establishing and maintaining parental allele-specific methylation, each mouse DMD has its own unique sequence, including that of the tandem direct repeats. Moreover, roughly 75 million years of evolution from a common ancestor of mouse and human has resulted in marked sequence divergence of each DMD, akin to a degree expected of many non-coding gene regions (Boffelli et al. 2003). This suggests that structure, rather than primary sequence, of the tandem repeats of the DMDs may be important. To identify those features of the tandem repeats that are likely to be involved in this important aspect of imprinting, we searched for features of three different DMD tandem repeats (*Snurf/Snrpn*, *Kcnq1*, and *Igf2r*) that are conserved during mammalian evolution. Specifically, we compared orthologous tandem repeat sequences among a number of primate species, and performed subsequent methylation analysis to confirm that these genomic regions from non-human primates, like their human counterparts, are differentially methylated. This approach represents a detailed species-to-species comparison of functionally relevant regions of orthologous genes in small evolutionary intervals and provides insight into the roles of these conserved elements in genomic imprinting.

2. Material and Methods

2.1 Acquisition of orthologous DMD sequences in non-human primates

The phylogenetic primate DNA panel PRP00001 encompassing different primate species was obtained through the Coriell Cell Repositories. Genomic DNA samples from the following species were analyzed: *Pan troglodytes* (chimpanzee, #NG06939), *Pan paniscus* (bonobo, #NG05253), *Gorilla gorilla* (gorilla, #NG05251), *Pongo pygmaeus* (Sumatra orangutan, #NG12256), *Macaca mulata* (Rhesus monkey, #NG07109), *Cercopithecidae macaca* (pigtailed macaque, #NG08452), and *Saguinus labiatus* (red-chested mustached tamarin, #NG05308).

The human *SNURF/SNRPN* gene sequence used as a reference was available from Genbank (Accession no. [U41384](#)). PCR amplifications were performed in order to get two overlapping *SNURF/SNRPN* gene fragments. The amplification of the first fragment was conducted under the following reaction conditions: 100 ng of genomic DNA, 0.2 mM deoxynucleoside triphosphates (dNTPs), 1mM MgCl₂, 0.4 μM each primer, and 1.25 U of Taq polymerase (Gibco/BRL) and 1xPCR buffer (Gibco/BRL). Oligonucleotides SnrpnchF: 5'-TGG TAT CCT GTC CGC TCG CAT TG-3' and SnrpnchR: 5'-CAA TAC CCA CTG CCG CCG CCT TC-3' were used. The cycling conditions were a Touch Down 65–55 program as follows: 94°C for 5 minutes, 10 cycles of 94°C for 1 min, 65°C to 55°C decrease of 1°C per cycle, for 30 sec, 72°C for 1min; then 20 cycles of 94°C for 1 min, 55°C for 30 sec, 72 for 1min; and a final extension of 72°C for 10 minutes. The amplification of the second fragment was conducted under the following reaction conditions: 100 ng of genomic DNA, 0.2 mM dNTPs, 0.4 μM each primer, and 1.25 U Expand High Fidelity Taq polymerase (Roche) and the corresponding 1xPCR Buffer (Roche). Oligonucleotides Snrpn2: 5'-TTG TGT TGG TGG TCG CGG CGC G-3' and Snrpn3: 5'-GAT TCC ACG CAT TTC TGC AGC C-3' were used. The cycling conditions were a 65–55 TouchDown program as follows: 94°C for 5 minutes, 10 cycles of 94°C for 1 min, 65°C to 55°C decrease of 1°C per cycle, for 30 sec, 72°C for 1min; then 20 cycles of 94°C for 1 min, 55°C for 30 sec, 72 for 1 minute; and a final extension of 72°C for 10 minutes.

The human *KCNQ1* gene sequence used as a reference was available at Genbank (Accession no. [AJ006345](#)). Amplifications of *KCNQ1* gene fragments were conducted under the following reaction conditions: 100 ng of genomic DNA, 0.2 mM dNTPs, 0.4 μM each primer, and 1.25 U of Expand High Fidelity Taq polymerase (Roche) and corresponding 1xPCR Buffer (Roche) in a 50-μl reaction mixture. Oligonucleotides KcnqchA: 5'-CCG AAT TCT GAT CCC AAC TAG GA-3' and KcnqchB: 5'-CAG CGC GGC ACT CTC CCG GCG G-3' from the human *KCNQ1* sequence were used. The cycling conditions were a 65–55 TouchDown program as

follows: 94°C for 5 minutes, 10 cycles of 94°C for 1 min, 65°C to 55°C decrease of 1°C per cycle, for 30 sec, 72°C for 30sec; then 20 cycles of 94°C for 1 min, 55°C for 30 sec, 72 for 30 sec; and a final extension of 72°C for 10 minutes.

The human *IGF2R* gene sequence used as a reference was available at Genbank (Accession no. **X91880**). Amplifications for *IGF2R* gene fragments were conducted under the following reaction conditions for DNAs from chimpanzee and Sumatra orangutan: 100ng of genomic DNA, 0.2 mM dNTPs, 0.4 µM each primer, and 1.25 U of Expand High Fidelity Taq polymerase (Roche) and corresponding 1xPCR Buffer (Roche) in a 50-µl reaction mixture. Oligonucleotides Igf2rchA: 5'-TCC TGG GCC TCT TGC GCC TCC-3' and Igf2rchB: 5'-CCT TCT CCA TCT GCA ACA CCC A-3' were used. The cycling conditions were as follows: 94°C for 5 minutes, 30 cycles of 94°C for 1 min, 55°C for 30 sec, 72°C for 40 sec; and a final extension of 72°C for 10 min.

All PCR-amplified fragments were cloned into pCR2.1-TOPO vector (Invitrogen) through the TOPO-TA cloning system (Invitrogen) and purifications of color-selected colonies were performed with the Quiaprep Spin Miniprep (Quiagen). At least 2 positive clones with the insert of interest for each species for each gene were sequenced with universal oligonucleotides M13 Forward and M13 Reverse on the pCRII-TOPO vector. Sequences were determined using the Applied Biosystems Big Dye sequencing kit following the manufacturer's protocol on an ABI 3700 automated sequencer (Applied Biosystems).

2.2 Methylation analysis of DMD repeats

To address the methylation status of CpG dinucleotides within the tandem repeats we performed bisulfite analysis on the *SNURF/SNRPN* (from chimp and orangutan), *KCNQ1* (from chimp, orangutan and Rhesus monkey), and *IGF2R* (from chimp and orangutan) regions of interest. 1 µg of genomic DNA was resuspended in 100 µl of TE and treated by sodium bisulfite mutagenesis as generally described by Clark et al. (40), but using the EpiTect Bisulfite Kit (Qiagen). For the chimp and orangutan *SNURF/SNRPN* regions, PCR amplification was carried out with semi-nested oligonucleotides SnrpnF: 5'-TTT GTT TTG GAG AAT TAG ATT T-3, SnrpnR: 5'-AAT CAC TAT TAT ACA CCT ACC TA-3', and SnrpnN: 5'-TTG ATG GTA TTT TGT TTG TTT G-3'. For the chimp and orangutan *KCNQ1* regions, PCR amplification was carried out with the semi-nested oligonucleotides Kcnq1F: 5'-ATT AGG ATA TTT AAT TTA GA-3', Kcnq1R: 5'-ACA ATA CAA CCT AAC CTA AAC AA-3', and Kcnq1N: 5'-TCC CAT CTA CAC CTT ATA AA-3'. For the Rhesus monkey *KCNQ1* region, PCR amplification was then carried out with hemi-nested oligonucleotides Kcfa: 5'-TGA TTT AGA ATT ATT AAG TGG ATT TT-3', Kcra: 5'-CAA TAC AAC CTA ACC TAA ACA ATA C, and Kcrb: 5'-AAC AAT ACA ACC TAA CCT AAA CAA TAC-3'. For the chimp *IGF2R* region, PCR amplification was carried out with the semi-nested oligonucleotides Igf2rF: 5'-GTT TTA TAT TTT TTT ATT TTT GTG T-3', Igf2rChR: 5'-CCT AAA CAA CTC CCA CTA T-3', and Igf2rN: 5'-TTG TGT TTT TTT GTT TAT TTT TTG-3'. For the orangutan *IGF2R* region, PCR amplification was carried out with the semi-nested oligonucleotides Igf2rF: 5'-GTT TTA TAT TTT TTT ATT TTT GTG T-3', Igf2rOrR: 5'-CCT AAC AAC TCT CCC TAT CC-3', and Igf2rN: 5'-TTG TGT TTT TTT GTT TAT TTT TTG-3'. For all PCR amplifications on bisulfite-treated DNA the cycling conditions for the first PCR were as follows: 2 cycles of 94°C for 4 min, 52°C (55°C for Rhesus monkey only) for 2 min, and 72°C for 2 min, then 35 cycles of 94°C for 1 min, 52°C (55°C for Rhesus monkey only) for 2 min, and 72 for 2 min, followed by a final extension of 72°C for 10 min. For the second (semi-nested) PCR: 94°C for 5 min, then 35 cycles of 94°C for 1 min, 52°C (55°C for Rhesus monkey only) for 2 min, and 72 for 2 min; followed by a final extension of 72°C for 10 min. The resulting PCR products were cloned into the TOPO-TA cloning system (Invitrogen) and purifications of color-selected colonies were performed with the Quiaprep Spin Miniprep

(Quiagen). Clones with the insert of interest for each species for each gene were sequenced with universal oligonucleotide M13 Reverse on the pCRII-TOPO vector. Sequences were determined using the Applied Biosystems Big Dye sequencing kit following the manufacturer's protocol on an ABI 3700 automated sequencer (Applied Biosystems).

2.3 DNA sequence analysis

Sequence repeats analysis, homologies calculations and alignments were performed with the MacVector 7.2. Program (Accelrys). The parameters used for alignment were: Open Gap penalty: 10.0; Extend Gap Penalty: 5.0; Delay Divergent: 40%. CpG islands tools were obtained at <http://www.ebi.ac.uk/emboss/cpgplot/>. Comparison with human sequences was done at <http://www.ncbi.nlm.nih.gov/BLAST/>. All the obtained chimpanzee and Rhesus monkey sequences were compared to the publicly available chimpanzee and Rhesus monkey sequences at <http://genome.ucsc.edu>. Except for a large portion of the chimp *SNURF/SNRPN* DMD sequence that was missing from the public database, only minor differences were found between the database version and the sequence obtained in this work for the same DMD in the same species.

2.4 GenBank Accession numbers

The following accession numbers have been provisionally applied to the primate DMD sequences described.

[AY948946](#), [AY948947](#), [AY948948](#), [AY948949](#), [AY952471](#), [AY952472](#), [AY952473](#), [AY952475](#), [AY952477](#), [AY952478](#), [AY952480](#), [AY953509](#), [AY953510](#), [AY953511](#), [AY953512](#)

3. Results

3.1 Tandem repeats are conserved in homologous chromosomal regions of mammalian genomes

The position of DMDs within human and mouse imprinting genes is conserved, and the human counterparts, like the mouse *Snurf/Snrpn*, *Kcnq1* and *Igf2r* DMDs, also contain tandem repeat sequences (Verona et al. 2003; Reinhart and Chaillet 2005). Mouse and human DMD tandem repeats can be easily visualized by matrix analysis, in which the degree of homology between a strand of DNA and itself is visualized on a dotplot. When both mouse and human *Snurf/Snrpn*, *Kcnq1* and *Igf2r* repeat regions are plotted in this way, it can be seen that all DMDs contain prominent direct (but imperfect) repeats of different unit lengths (Figure 1 and Table 1). The overall lengths of these DMD repeated domains in mouse and human are remarkably similar, with a total size range of 425–740 nucleotides across both species and all three DMDs (Table 1). For some DMDs, such as the mouse *Kcnq1* DMD, populations of both short (unit lengths of approximately 20 nt) and long repeats (approximately 155-nt unit lengths) are evident on the dotplots (Figure 1). In contrast, only repeats within a narrow size range (40–75nt) are present in the human *IGF2R* DMD. The other DMDs also have unique arrangements and sizes of tandem repeats (Figure 1 and summarized in Table 1), consistent with an early analysis of some DMD-associated repeats (Neumann et al. 1995). Long indirect repeats were not seen on the dotplots.

The analysis of the three DMD-associated tandem repeats in both mouse and human using the Zuker mfold program (Mfold server at <http://www.bioinfo.rpi.edu/applications/mfold/>) showed that these sequences have the potential to fold into stable secondary structures (Figure 2). The significant similarities of these structures among the three DMDs and between mouse and human suggest that such structures may be related to a common function or functions of the DMD tandem repeats.

Repetitive sequences have been studied in many organisms, where they are present at a variety of different chromosomal locations, primarily centromeric and telomeric (McEachern et al. 2000; Henikoff et al. 2001; Sullivan et al. 2001). A recurring feature of many of these sequences is that, upon careful analysis, large unit repeat lengths are founded on shorter repeats (Fanning 1987; Lee et al. 1997; Jobse et al. 2005). To determine if there is a similar general organization to the DMD tandem repeats, we examined the human *IGF2R* tandem repeats in detail, and found that they are composed of two types of shorter repeats (subrepeats), primarily imperfect, organized in head-to-tail fashion (Figure 3A). It is evident from Figure 2 that many short inverted repeats can also form from the mixture of short and long imperfect direct repeats found within these regions.

Each subrepeat of the human *IGF2R* repeats (Figure 3A) contains stretches of cytosine-guanine basepairs. It has been shown that stretches of cytosine and guanine can adopt non-canonical structures *in vitro*, and these structures may also be present *in vivo* (Davey and Allan 2003; Nonin-Lecomte et al. 2005). In this regard, G-quartet structures are predicted structures in the repeated region of the human *IGF2R* DMD, as well as in the repeated regions of the human *SNURF/SNRPN* and *KCNQ1* DMDs (Figure 3B). Notably, two sequences within the human *SNURF/SNRPN* DMD are predicted to assume very stable G-quartet structures containing three guanine tetrads (G-scores > 30).

3.2 Comparison of DMD tandem repeat features among primate species

Orthologous mouse and human DMD tandem repeats are so divergent at the level of their primary sequence that little can be learned, through a comparison of orthologous sequences, about evolutionary processes that shape these important regions. Therefore, we instead compared the sequences of DMD orthologues among a group of more closely related species, namely primates (Figures 4 and 5). The panel of primate genomic DNAs used to analyze the phylogeny of tandem repeats encompasses more than 40 million years of evolution. The evolutionary separation between humans and non-human primates from their common ancestors corresponds to the following periods of time: apes 6 to 14 million years, old world monkeys 25 million years, and new world monkeys 40 million years (Goodman 1999). As a comparison, the separation between humans and mice from their last common ancestor is 75 million years (Botfelli et al. 2003). We focused on the tandem repeat sequences within three different DMDs, *SNURF/SNRPN*, *KCNQ1*, and *IGF2R*. Starting with known sequence of the human DMDs, we used PCR amplification to capture, from non-human primate species, fragments of the tandem-repeat regions of the three different DMDs.

The human *SNURF/SNRPN* gene sequence used for all comparisons is available through GenBank (Accession no. [U41384](#)). The PCR amplification was designed to produce the amplification of two overlapping fragments covering a total sequence length of 819 nt in human and the corresponding amplified sequence was 840 nt in chimpanzee, 820 nt in bonobo, 800 nt in gorilla, 806 nt in Sumatra orangutan, 782 nt in Rhesus monkey, 781 nt in the pigtailed macaque and 734 bp in the red-chested mustached tamarin, after exclusion of the portions corresponding to the oligonucleotides (data not shown). Focusing on just the two distinct regions of tandem repeats in the human DMD (defined by double-headed arrows in Figure 1), we were easily able to align the different primate sequences (Figure 4). Notably, in the distal repeated region of some of the analyzed non-human primate species, large gaps in the sequence alignments were evident (Figure 4). For example, a gap of 33 nucleotides was evident in the Rhesus monkey, pigtailed macaque and red-chested mustached tamarin sequences, and there was a 20-nt stretch of contiguous sequence unique to the chimpanzee. Because the sizes of the length changes roughly correspond to the average unit repeat length in this region, we interpret these differences as evolutionary additions and/or deletions of repeat copies. The homology among mammalian species for the combined regions is shown in Table 1. The CpG content

for the entire region of the *Snurf/Snrpn* tandem repeats in mouse and human is shown in Table 1 (3.4% for mouse and 7.7% for human). Alignments of the primate sequences show a high degree of conservation of CpG content and position (Table 1 and Figure 4). All the sequences analyzed fit the criteria for CpG islands (data not shown).

The human *KCNQ1* gene sequence used for all comparisons is available through GenBank (Accession no. [AJ006345](#)). Amplification of the *KCNQ1* tandem repeats produced fragments, after exclusion of the portions corresponding to the oligonucleotides, of nearly identical size: 502 bp from human DNA, 500 bp from chimpanzee DNA, 502 nt from bonobo DNA, 502 nt from gorilla DNA, 504 nt from Sumatra orangutan DNA, 504 nt from Rhesus monkey DNA, and 499 nt from red-chested mustached tamarin DNA. The pigtailed macaque DNA could not be amplified for this particular gene. The sequence alignment of the different species is shown in Figure 5A. As is the case for the *SNURF/SNRPN* regions, there is a high degree of conservation of CpG content and position within the tandem repeats (Table 1 and Figure 5A). Details of species' homologies and CpG content of the *KCNQ1* DMDs are shown in Table 1. Except for the red-chested mustached tamarin *KCNQ1* region, all the sequences analyzed fit the criteria for CpG islands (data not shown), and the change in degree of sequence homology over evolutionary time for *KCNQ1* is similar to that observed for *SNURF/SNRPN* (Table 1). However, in contrast to the *Snurf/Snrpn* tandem repeats, there were no observed additions or deletions of unit copies among the primate species examined, suggesting that the *KCNQ1* repeats within the region analyzed did not undergo changes in copy number during this evolutionary time period.

The human *IGF2R* sequence used for all comparisons is available through GenBank (Accession no. [X91880](#)). The PCR amplification with oligonucleotides Igf2rchA and Igf2rchB resulted in the amplification of a 704-nt fragment from human DNA, and the corresponding amplified fragment was 718 nt from chimpanzee and 853 nt from Sumatra orangutan, after exclusion of the portions corresponding to the oligonucleotides. We were unable to produce amplified products from other primate species (bonobo, gorilla, Rhesus monkey, pigtailed macaque and red-chested mustached tamarin). The alignment of the three different species' sequences is shown in Figure 5B. Details of species' homologies and CpG content of the *IGF2R* DMDs are shown in Table 1. The regions analyzed fit the criteria for CpG islands (data not shown), and the CpG content and position is well conserved. In a similar fashion to the analyzed *SNURF/SNRPN* tandem repeat homology across primate species, the *IGF2R* sequence comparison also indicates gaps in sequence alignments of 42 and 58 nucleotides. Again, because the size of these gaps is approximately equal to the size of the repeated unit copies evident in human DNA (Figure 1), we conclude that unit copies of the *IGF2R* tandem repeats have been added or deleted during primate evolution.

3.3 Methylation analysis of DMD repeats

To determine if the non-human primate tandem-repeat sequences are themselves differentially methylated tandem repeats, we analyzed the methylation of the *SNURF/SNRPN* and *IGF2R* orthologous tandem repeats in two different non-human primate species and the *KCNQ1* repeats in three different such species. Bisulfite mutagenesis analysis of *SNURF/SNRPN* and *IGF2R* tandem repeats was performed on chimpanzee and orangutan species, and the analysis of *KCNQ1* was performed on chimpanzee, orangutan and Rhesus monkey. The results of this analysis are summarized in Figure 6. Alleles of the *SNURF/SNRPN* tandem repeats from both chimpanzee and orangutan ranged from completely unmethylated to completely methylated, with approximately one-half of the alleles having the majority of their CpG dinucleotides methylated (11 out of 19 for chimpanzee and 11 out of 20 for orangutan). Similar results were obtained for the three analyzed *KCNQ1* tandem repeats (11 out of 17 for chimpanzee; 8 out of 16 for orangutan; and 7 out of 14 for Rhesus monkey), and for the chimpanzee *IGF2R* tandem

repeats (7 out of 15). For the orangutan *IGF2R* tandem repeats, the population of analyzed alleles was skewed to a larger percentage of unmethylated ones (only 5 out of 22 alleles were heavily methylated). However, a single-nucleotide polymorphism distinguished the five methylated alleles (cytosine base) from the other alleles (adenine base); this polymorphism is consistent with a maternally methylated orangutan DMD. For each of these tandem repeat regions, there was a general trend towards a division of the alleles into a heavily methylated population and a nearly unmethylated population. These findings are consistent with methylation on the maternal alleles and the absence of methylation on the paternal alleles in the chimpanzee and orangutan orthologues of the human *SNURF/SNRPN* and *IGF2R* tandem repeats, and in the chimpanzee, orangutan and Rhesus monkey orthologues of the human *KCNQ1* tandem repeats. We conclude from this that the orthologous non-human primate tandem repeats of the three human DMD sequences are functionally relevant, imprinted DMDs.

4. Discussion

4.1 Conservation of DMD tandem repeats

We analyzed the structures of three distinct DMDs (*Snurf/Snrpn*, *Kcnq1*, and *Igf2r*) among different species with the aim of identifying a conserved and, therefore, functional *cis*-acting imprinting signal. We chose to target our analysis to tandem repeats within these DMDs because of the demonstrated importance of these motifs for mouse and human imprinting (Neumann et al. 1995; Reinhart et al. 2002; Hutter et al. 2006; Reinhart et al. 2006). From our own detailed functional analysis in the mouse, these DMD tandem repeats are necessary for maintaining the differentially methylated state of the DMD following establishment of two different sex-specific methylation patterns in oocyte and sperm (Reinhart et al. 2002; Reinhart et al. 2006). Importantly, this maintenance function is shared by all three DMDs, although each is a unique sequence. This suggests that a structural feature of the tandem repeats, rather than primary sequence, is important for imprinting. For all three DMDs, their maternal-specific DNA methylation and genomic locations are conserved in the human and mouse (Smrzka et al. 1995; Engemann et al. 2000; Onyango et al. 2000; Paulsen et al. 2000). Based on these aspects of mouse and human DMDs, it is likely that their genomic position and function are also conserved in all non-human primate species.

In principle, features of the tandem repeats that change in the short evolutionary intervals represented by different primate species are not likely to be important, whereas features that do not change over the 75 million-year interval represented by evolution from a common ancestor to mouse and human are likely to be highly important (Cooper and Sidow 2003). That is, nonfunctional features are expected to accumulate mutations and exhibit divergence at a faster rate than functional features (Sidow 2002). We made use of this principle by first comparing the orthologous DNA sequences of the three different DMD tandem repeats. Interestingly, the rates of average sequence divergence of the three DMD tandem repeats (*SNURF/SNRPN*, *KCNQ1*, and *IGF2R*) were similar for human versus chimpanzee (1.9%), human versus gorilla (1.1%), and human versus orangutan (7.7%) to the rates in non-coding regions of the corresponding genomes. For example, based on the analysis of 53 autosomal intergenic nonrepetitive DNA segments, the average genome sequence divergence between human and chimpanzee, between human and gorilla, and between human and orangutan were shown to be 1.24%, 1.62% and 3.08%, respectively (Chen and Li 2001). The complete lack of sequence similarity between mouse and human DMDs of the same gene makes it difficult to estimate a degree of sequence similarity. However, the DMD sequence divergence between mouse and human certainly is much greater than that seen in coding sequences. We can assume from these data that the three tandem repeat regions of the DMDs are typical genomic elements in that they exhibit a rate of sequence divergence similar to that of non-coding sequences.

Despite the rates of evolutionary change of the three different DMD tandem repeats, perturbations of the repeated regions in the form of additions and/or deletions clearly occur in the relatively short intervals represented in primate evolution (Figures 4 and 5B). This suggests that the precise number of unit copies is not important, but rather that there is a minimum requirement for the number of repeats, which can be met with a subset of the complete tandem-repeat region (Reinhart et al. 2006). Additions and deletions would most likely occur by unequal crossover events between sister chromatids or homologous chromosomes during meiosis (Smith 1976; Charlesworth et al. 1994). One possibility is that additions and deletions occur to maintain repetitiveness and overall length of the repeated region. This argument is consistent with the maintenance of length of the different DMD tandem repeats in the 75 million years of evolution from a common ancestor of mouse and human (Table 1). These observations, taken together with the absence of sequence similarities for all three DMDs, are consistent with the notion that the tandem repetitiveness of the *Snurf/Snrpn*, *Kcnq1* and *Igf2r* DMDs is the most conserved feature of these elements throughout mammalian evolution.

4.2 Functionality of tandem repeats

Despite the absence of homology of DMD sequences over a long evolutionary time (i.e. rodent to primate), the intragenic locations and presence of tandem repeats are invariant features of the mouse and human *Snurf/Snrpn*, *Kcnq1*, and *Igf2r* DMDs. This suggests that the tandem repeats contained within the DMDs are of immense functional significance for the imprinting process. This notion is also supported by our studies analyzing the contribution of different DMD sequences to transgene imprinting (Reinhart et al. 2002; Reinhart et al. 2006).

What are the specific features of the DMD tandem repeats that provide their function in establishing and maintaining imprints in the mammalian genome? The likelihood that secondary DNA structures assumed by the DMD tandem repeats provide the signal for maintenance of inherited genomic imprints is high. This is because the short repetitive motifs embedded with larger unit copies (Figures 1 and 2), contain an asymmetric distribution of G and C residues, with one strand particularly G-rich. This organization is very reminiscent of telomeric and centromeric sequences in different species (Ferrer et al. 1995; Abad et al. 2000). These types of repeats are known to assume different types of secondary structures, such as G-quartets. Interestingly, Figures 2 and 3 show that the human *SNURF/SNRPN*, *KCNQ1*, and *IGF2R* DMDs may fold into similar secondary structures *in vivo*. The folding of two proximate guanine-rich tracts into hairpins may potentially lead to the pairing and to the formation of tetraplex structures (Vorlickova et al. 2005). We can speculate that, although the tandem repeats may show structural features of telomeres and centromeres, it is their specific euchromatic location away from telomeres and centromeres that determines their imprinting function.

The three analyzed DMD tandem repeats have been evolutionarily maintained with high CpG dinucleotide contents, consistent with a CpG island definition (Table 1). Because the maternal alleles of these DMD tandem repeats are highly methylated at all times except for early germ cell stages, and the paternal alleles are always unmethylated, they are unusual CpG islands. These observations are consistent with a specific function of the CpG dinucleotides within the tandem repeat sequences. There are two likely functions for these embedded CpG dinucleotides. First, the CpG dinucleotides are part of a repeat-dependent structure of the DMD tandem repeats that define them as epigenetic imprinting signals. Second, the CpG dinucleotides are within the DMD tandem repeats simply to provide a heritable epigenetic mark. In both cases, the density of embedded CpG dinucleotides is probably important. In the first case, position of the CpG dinucleotides within the DMD tandem repeats would be important, whereas position would probably not be important for the second model. The high CpG dinucleotide content of the tandem repeats of the *Snurf/Snrpn*, *Kcnq1* and *Igf2r* DMDs

cannot alone account for the imprinted phenotype of DMDs, as the vast majority of CpG islands are not imprinted. Other features of the DMDs must be associated with the imprinted phenotype of DMDs. In this regard, we have recently shown in mouse transgenes that tandem repeats of these three DMDs are likely to be *cis*-acting signals for DMD imprinting (Reinhart et al. 2002). Imprinting of these three DMD sequences involves a series of molecular steps, initiated by methylation of the maternal DMD allele during oogenesis, followed by the successful transmission of this epigenetic modification, and ultimately the persistent maintenance of a difference in the maternal and paternal methylation states. The *cis*-acting signals that direct this complex process must be successful at attracting methylation during oogenesis, and then at maintaining a strict difference in methylation between parental DMD alleles during all of post-fertilization development. The tandem repeats are not required for all of these stages in the cyclical imprinting process. Rather, as we have recently shown, the function of the tandem repeats is in maintaining allele-specific methylation differences after fertilization (Reinhart et al. 2006). Based on the analysis of mouse and primate tandem repeats described above, the tandem repetitiveness alone (or an organization conferred by the repetitiveness, such as the order of CpG dinucleotides) is the best candidate for providing this imprinting maintenance function.

Acknowledgments

This work was funded by the Litta Foundation and the Fondation Suisse des Bourses Médecine et Biologie (to APG), and by the NIH (to JRC).

Abbreviations

DMD	differentially methylated domain
dNTPs	deoxynucleoside triphosphates

References

- Abad JP, Villasante A. Searching for a common centromeric structural motif: Drosophila centromeric satellite DNAs show propensity to form telomeric-like unusual DNA structures. *Genetica* 2000;109:71–75. [PubMed: 11293798]
- Bartolomei MS, Tilghman SM. Genomic imprinting in mammals. *Annu Rev Genet* 1997;31:493–525. [PubMed: 9442905]
- Boffelli D, McAuliffe J, Ovcharenko D, Lewis KD, Ovcharenko I, Pachter L, Rubin EM. Phylogenetic shadowing of primate sequences to find functional regions of the human genome. *Science* 2003;299:1391–1394. [PubMed: 12610304]
- Brandeis M, Kafri T, Ariel M, Chaillet JR, McCarrey J, Razin A, Cedar H. The ontogeny of allele-specific methylation associated with imprinted genes in the mouse. *EMBO Journal* 1993;12:3669–3677. [PubMed: 7504628]
- Bressler J, Tsai TF, Wu MY, Tsai SF, Ramirez MA, Armstrong D, Beaudet AL. The SNRPN promoter is not required for genomic imprinting of the Prader-Willi/Angelman domain in mice. *Nat Genet* 2001;28:232–240. [PubMed: 11431693]
- Chaillet JR, Vogt TF, Beier DR, Leder P. Parental-specific methylation of an imprinted transgene is established during gametogenesis and progressively changes during embryogenesis. *Cell* 1991;66:77–83. [PubMed: 1649008]
- Chaillet, JR. Developmental regulation of genomic imprinting by DNA methylation. In: Reik, W.; Surani, MA., editors. *Frontiers in Molecular Biology: Genomic Imprinting in Mammals*. Oxford University Press; 1997. p. 21-37.

- Charlesworth B, Sniegowski P, Stephan W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 1994;371:215–220. [PubMed: 8078581]
- Chen FC, Li WH. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am J Hum Genet* 2001;68:444–456. [PubMed: 11170892]
- Clark SJ, Harrison J, Paul CL, Frommer M. High sensitivity mapping of methylated cytosines. *Nucleic Acids Res* 1994;22:2990–2997. [PubMed: 8065911]
- Cooper GM, Sidow A. Genomic regulatory regions: insights from comparative sequence analysis. *Curr Opin Genet Dev* 2003;13:604–610. [PubMed: 14638322]
- Davey C, Allan J. Nucleosome positioning signals and potential H-DNA within the DNA sequence of the imprinting control region of the mouse *Igf2r* gene. *Biochimica et Biophysica Acta* 2003;1630:103–116.
- Engemann S, Stroedicke M, Paulsen M, Franck O, Reinhardt R, Lane N, Reik W, Walter J. Sequence and functional comparison in the Beckwith-Weidemann region: Implications for a novel imprinting centre and extended imprinting. *Hum Mol Genet* 2000;9:2691–2706. [PubMed: 11063728]
- Fanning TG. Origin and evolution of a major feline satellite DNA. *Mol Biol* 1987;197:627–34.
- Ferrer N, Azorin F, Villasante A, Gutierrez C, Abad JP. Centromeric dodeca-satellite DNA sequences form fold-back structures. *J Mol Biol* 1995;245:8–21. [PubMed: 7823322]
- Fitzpatrick GV, Soloway PD, Higgins MJ. Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. *Nat Genet* 2002;32:426–431. [PubMed: 12410230]
- Goodman M. The genomic record of Humankind's evolutionary roots. *Am J Hum Genet* 1999;64:31–39. [PubMed: 9915940]
- Henikoff S, Ahmad K, Malik HS. The centromere paradox: Stable inheritance with rapidly evolving DNA. *Science* 2001;294:2478–248. [PubMed: 11770517]
- Hutter B, Helms V, Paulsen M. Tandem repeats in the CpG islands of imprinted genes. *Genomics* 2006;88:323–332. [PubMed: 16690248]
- Jobse C, Buntjer JB, Haagsma N, Breukelman HJ, Beintema JJ, Lenstral JA. Evolution and recombination of bovine DNA repeats. *J Mol Evol* 2005;41:277–283. [PubMed: 7563113]
- Lee C, Court DR, Cho C, Haslett JL, Lin CC. Higher-order organization of subrepeats and the evolution of cervid satellite I DNA. *J Mol Evol* 1997;44:327–335. [PubMed: 9060399]
- Lin SP, Youngson N, Takada S, Seitz H, Reik W, Paulsen M, Cavaille J, Ferguson-Smith AC. Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the Dlk1-Gtl2 imprinted cluster on mouse chromosome 12. *Nat Genet* 2003;35:97–102. [PubMed: 12937418]
- Lopes S, Lewis A, Hajkova P, Dean W, Oswald J, Forne T, Murrell A, Constancia M, Bartolomei M, Walter J, Reik W. Epigenetic modifications in an imprinting cluster are controlled by a hierarchy of DMRs suggesting long-range chromatin interactions. *Hum Mol Genet* 2003;12:295–305. [PubMed: 12554683]
- Lucifero D, Mertineit C, Clarke HJ, Bestor TH, Trasler JM. Methylation dynamics of imprinted genes in mouse germ cells. *Genomics* 2002;79:530–538. [PubMed: 11944985]
- Mancini-Dinardo D, Steele SJ, Levorse JM, Ingram RS, Tilghman SM. Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. *Genes Dev* 2006;20:1268–1282. [PubMed: 16702402]
- McEachern MJ, Krauskopf A, Blackburn EH. Telomeres and their control. *Annu Rev Genet* 2000;34:331–358. [PubMed: 11092831]
- Neumann B, Kubicka P, Barlow DP. Characteristics of imprinted genes. *Nat Genet* 1995;9:12–13. [PubMed: 7704015]
- Nonin-Lecomte S, Dardel F, Lestienne P. Self-organisation of an oligodeoxynucleotide containing the G- and C-rich stretches of the direct repeats of the human mitochondrial DNA. *Biochimie* 2005;87:725–735. [PubMed: 16054526]
- Onyango P, Miller W, Lehoczy J, Leung CT, Birren B, Wheelan S, Dewar K, Feinberg AP. Sequence and comparative analysis of the mouse 1-Megabase region orthologous to the human 11p15 imprinted domain. *Genome Res* 2000;10:1697–1710. [PubMed: 11076855]

- Paulsen M, El-Maarri O, Engemann S, Strodicke M, Franck O, Davies K, Reinhardt R, Reik W, Walter J. Sequence conservation and variability of imprinting in the Beckwith-Weidemann syndrome gene cluster in human and mouse. *Hum Mol Genet* 2000;9:1829–1841. [PubMed: 10915772]
- Reik W, Lewis A. Co-evolution of X-chromosome inactivation and imprinting in mammals. *Nat Rev Genet* 2005;6:403–410. [PubMed: 15818385]
- Reinhart B, Eljanne M, Chaillet JR. Shared role for differentially methylated domains of imprinted genes. *Mol Cell Biol* 2002;22:2089–2098. [PubMed: 11884597]
- Reinhart B, Paoloni-Giacobino A, Chaillet JR. Specific differentially methylated domain sequences direct the maintenance of methylation at imprinted genes. *Mol Cell Biol* 2006;26:8347–8356. [PubMed: 16954379]
- Reinhart B, Chaillet JR. Genomic imprinting: Cis-acting sequences and regional control. *Int Rev Cytol* 2005;243:173–213. [PubMed: 15797460]
- Sidow A. Sequence first. Ask questions later *Cell* 2002;111:13–16.
- Slutels F, Zwart R, Barlow DP. The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* 2002;415:810–813. [PubMed: 11845212]
- Smith GP. Evolution of repeated sequences by unequal crossover. *Science* 1976;191:528–535. [PubMed: 1251186]
- Smrzka OW, Faé I, Stöger R, Kurzbauer R, Fischer G, Henn T, Andreas Weith A, Barlow DP. Conservation of a maternal-specific methylation signal at the human *IGF2R* locus. *Hum Mol Genet* 1995;4:1945–1952. [PubMed: 8595419]
- Stoger R, Kubicka P, Liu CG, Kafri T, Razin A, Cedar H, Barlow DP. Maternal-specific methylation of the imprinted mouse *Igf2r* locus identifies the expressed locus as carrying the imprinting signal. *Cell* 1993;73:61–71. [PubMed: 8462104]
- Sullivan BA, Blower MD, Karpen GH. Determining centromere identity: Cyclical stories and forking paths. *Nat Rev Genet* 2001;2:584–596. [PubMed: 11483983]
- Thorvaldsen JL, Duran KL, Bartolomei MS. Deletion of the H19 differentially methylated domain results in loss of imprinted expression of *H19* and *Igf2*. *Genes Dev* 1998;12:3693–3702. [PubMed: 9851976]
- Trasler JM. Gamete imprinting: setting epigenetic patterns for the next generation. *Reprod Fertil Dev* 2006;18:63–69. [PubMed: 16478603]
- Verona RI, Mann MR, Bartolomei MS. Genomic imprinting: intricacies of epigenetic regulation in clusters. *Annu Rev Cell Dev Biol* 2003;19:237–259. [PubMed: 14570570]
- Vorlickova M, Chladkova J, Kejnovska I, Fialova M, Kypr J. Guanine tetraplex topology of human telomere DNA is governed by the number of (TTAGGG) repeats. *Nucleic Acids Res* 2005;33:5851–5860. [PubMed: 16221978]
- Wutz A, Smrzka OW, Schweifer N, Schellander K, Wagner EF, Barlow DP. Imprinted expression of the *Igf2r* gene depends on an intronic CpG island. *Nature* 1997;389:745–749. [PubMed: 9338788]
- Yang T, Adamson TE, Resnick JL, Leff S, Wevrick R, Francke U, Jenkins NA, Copeland NG, Brannan CI. A mouse model for Prader-Willi syndrome imprinting-centre mutations. *Nat Genet* 1998;19:25–31. [PubMed: 9590284]
- Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acid Res* 2003;31:3406–3415. [PubMed: 12824337]

Figure 1.

Dotplots of tandem repeats of the mouse and human *Snurf/Snrpn*, *Kcnq1* and *Igf2r* DMDs. Each species' plot is a comparison of DMD sequence (containing tandem repeats) to itself. Direct repeats are identified as lines offset from the diagonal, but parallel to the diagonal. Relevant parameters for the dotplot constructions are: Window size 30; Minimal percentage scores: A. 60%; B. 50%; C. 40%; D. 60%; E. 60%; and F. 60%. A dotplot similar to the one shown in panel E, but with window size 23 and minimal percentage score 70%, was previously published (Neumann et al. 1995). A dotplot similar to the one shown in panel F, but with window size 30 and minimal percentage score 65%, was previously published (Smrzka et al. 1995). Sequences used in alignments with sequences from other species, and in calculations

of CpG dinucleotide contents and the extents of homology are indicated by the double-headed arrows at the bottom of the panels. The thick line labeled *DME2* at the bottom of panel E refers to the minimal mouse *Igf2r* DMD sequence that is required for imprinting in a mouse transgene model (Reinhart et al. 2006).

Figure 2.

Secondary structures of human and mouse DMDs modeled by the Zuker Mfold program (Zuker 2003). The six panels correspond to the panels in Figure 1; for each structure, the complete sequence used to generate the dotplot was also used to generate the Mfold structure. Calculated free energies (kcal/mol) for the structures shown in panels A–F are as follows: A. –124.48; B. –90.11; C. –108.23; D. –81.51; E. –123.73; F. –117.41.

Figure 3. Small sequence motifs embedded within larger repeats of the human DMDs. A. The contiguous sequence of the human *IGF2R* DMD has been written to emphasize the small repeats and their organization within the DMD. The most conserved nucleotides are in blue, and CG dinucleotides are highlighted in yellow. The small repeats can be classified in two types (grouped into the red and green rectangles). Upper contiguous sequence corresponds to nucleotides 134–651, and lower contiguous sequence corresponds to nucleotides 666–754 of Figure 1, panel F. B. Motifs from the human DMDs sequences with predicted G-quartet structures. Sequence positions correspond to those of Figures 1 and 2. Structures predicted using the QGRS Mapper algorithm at <http://bioinformatics.ramapo.edu/QGRS/analyze.php>. The G-score is a relative measure of the stability (including the number of guanine tetrads) of the G-quartet, with larger values indicating greater stability. Guanine groups that form guanine tetrads are underlined. QGRS = Quadruplex-forming G-Rich Sequences. A maximum length of 30 nucleotides was considered for a QGRS.

A *KCNQ1* Alignments

Human	TACCTAATCTCAGAACACAACGGGGATTCCCACTCGATGTCCT-AATTGGGCCCTGTACTCAGAACCACTAAGTGGTGAATTC
Chimp	TACCTAATCTCAGAACACAACGGGGATTCCCACTCGATGTCCT-AACTGGGCCCTGTACTCAGAACCACTAAGTGGTGAATTC
Bonobo	TACCTAATCTCAGAACACAACGGGGATTCCCACTCGATGTCCT-AATTGGGCCCTGTACTCAGAACCACTAAGTGGTGAATTC
Gorilla	TACCTAATCTCAGAACACAACGGGGATTCCCACTCGATGTCCT-AATTGGGCCCTGTACTCAGAACCACTAAGTGGTGAATTC
OrangU	TACCTAATCTCAGAACCCAACGGGGATTCCCACTCGATGTCCTTAATTGGGCCCTGTACTCAGAACCACTAAGTGGTGAATTC
Rhesus	TACCTAATCTCAGAACACAACGGGGATTCCCACTCGATGTCCTAAATGGGCCCTGTACTCAGAACCACTAAGTGGTGAATTC
Tamarin	TACCTAATCTCAGAACACAACGGGGATTCCCACTCGATGTCCT-AATTGAGACTTGACTCAAACACTCATGTTGGTGGTTC *** * ***** *
Human	CGACTCGGTCGTCCAATTGGGGTTTGACTCAAAAACCAACATGGATTCCCAACTCCATCCCAATAGCGGC-GACTCGAGA
Chimp	CGACTCGGTCGTCCAATTGGGGTTTGACTCAAAAACCAACATGGATTCCCAACTCCATCCCAATAGCGGC-GACTCGAGA
Bonobo	CGACTCGGTCGTCCAATTGGGGTTTGACTCAAAAACCAACATGGATTCCCAACTCCATCCCAATAGCGGC-GACTCGAGA
Gorilla	CGACTCGGTCGTCCAATTGGGGTTTGACTCAAAAACCAACATGGATTCCCAACTCCATCCCAATAGCGGC-GACTCGAGA
OrangU	GACCCCCTGTCCTCAATTGGGGTTTGACTCAAAAACCAACATGGATTCCCAACTCCATCCCAATAGCGGC-GACTCGAGA
Rhesus	CGACTCCGATCGTCCAATTGGGGTTTGACTCAAAAACCAACATGGATTCCCAACTCCGTCGATAGGCGA-GACTCGAGA
Tamarin	TGACTCTGGTCGTCCAATTGAGCGCTTCACTCAAAAACATGACAGAGATTCTCAACTCCATCCCAATAGTGCCGACTCGAGA * * **** ** *
Human	ACCACAGATGCGGGATTCCAGACTCCAAAT
Chimp	ACCACAGATGCGGGATTCCAGACTCCAAAT
Bonobo	ACCACAGATGCGGGATTCCAGACTCCAAAT
Gorilla	ACCACAGATGCGGGATTCCAGACTCCAAAT
OrangU	ACCACAGATGCGGGATTCCAGACTCCAAAT
Rhesus	ACCACATGCGGGATTCCAGACTCCAAAT
Tamarin	ACCACATGCGGGATTCCAGACTCCAAAT ***** *****

B *IGF2R* Alignments

```

Human    CACCCGGCATCTCTGTGTGCCCATGCGCTACACCCCTCACGCTCCCTCTACCTGCATGACCCCGTGTGCTCTGTGTGCC
Chimp    CACCGGCATCTCTGTGTGCCCATGCGCTACACCCCT-TACCGCTCCCTCTGATCCGTCATGCCCGGGTGCTCTGTGTGCC
OrangU   CACCGGCGCTCTCTGTGTGCCCATGCGCTCTTCTCTGCGCTCCCTCTATACGTCGATGACCCCGTCTCGCCTCTGTGTGCC
*****
Human    CCACGCGCTCTCCCCCTCTCGCCGCT-----
Chimp    CCACGCGCTCTCCCCCTCTCGCCGCT-----
OrangU   CCACGCGCTCTCCCCCTCTCGCCGCTCCCTGTACCTTACATGCGCCCTGTGCTGCCGCCACCGCCGCTCTCCCCCGCGCC
*****
Human    -----CCTGTGTACCTGCTGATGCCCTGTGCCGCTGTGCCGCCCACGCGCTCTCCCCCTCTCGCCGCTGCC-----
Chimp    -----CCTGTTAACCTGCTGATGCCCTGTGCCGCTGTGCCGCCCACGCGCTCTCCCCCTCTCGCCGCTGCC-----
OrangU   TCCCTGTACCGCTGATGCCCGCGCGCTGTGCCGCCCACGCGCTCTCCCCCTCTCGCTCTCTGATCCCTGATGCGCC
*****
Human    -----CCTGTGCCGCTCCCTGTACCCCTGCATACCCCGTGAGCGCTGCTGCCGCCCATCGC-----
Chimp    -----CCTGTGCCGCTCTCGCGCCCACGCGCTCTCCCCCTCTGTACCGCTGATCGCCCGCGCGCTGCTGTGCGCCCATCGC
OrangU   GTGTGCCGCTGCTGCCGCCCACGCGCTCTCCCCCTCTGTACCGCTGATCGCCCGCGCGCTGCTGTGCGCCCATCGC
*****
Human    CTTCCCGCGCTCTCCAGCGCTCTCATCGCGCTCCAGCTGTGCGCATCTCTCCT
Chimp    CTTCCCGCGCTCTCGAGCGCCCATCGCTCTCGCTAGCTGTGCGCATCTCTCCT
OrangU   CTTCCCGCGCTCTCGAGCGCCCATCGCTCTCGCTAGCTGTGCGCATCTCTCCT
*****

```

Figure 5.

Sequence comparisons of (A) *Kcnq1* and (B) *Igf2r* DMD tandem repeats among primate species. The regions used for comparison are indicated by the double-headed arrows in panels D and F of Figure 1. CpG dinucleotides are highlighted in bold type face. The beginning of each sequence is closest to the 5' end of the human genes. The species are ordered by evolutionary distance. Asterisks below the alignments show the positions of nucleotide identities among the sequences of all species. Human and chimpanzee *KCNQ1* and chimpanzee *IGF2R* sequences can also be found in public genome sequence databases.

Figure 6.

Methylation status of individual DNA strands for the *SNURF/SNRPN* gene in chimpanzee and orangutan, for the *KCNQ1* gene in chimpanzee, orangutan and Rhesus monkey, and for the *IGF2R* gene in chimpanzee and orangutan, as determined by bisulfite mutagenesis analysis. Each line corresponds to an individual strand of analyzed DNA. Methylated CpGs are indicated as filled circles, and absence of circles indicates unmethylated CpGs. The vertical arrow at the bottom of the orangutan *IGF2R* alleles indicates the position of an A/C polymorphism. The 5 bottommost alleles contain a C at the position of the arrow, and the remaining alleles contain an A.

Table 1

Sequence and structural features of DMD tandem repeats

Size of repeated regions of DMDs ^a		
DMD	Mouse	Human
Snurf/Snrpn	640	455
Kcnq1	740	425
Igf2r	639	620
Unit repeat lengths ^b		
DMD	Mouse	Human
Snurf/Snrpn	60, 100, 115	25, 75
Kcnq1	20, 155	55, 80
Igf2r	30, 175	40, 60, 75
CpG content of repeated regions ^c		
DMD	Mouse	Human
Snurf/Snrpn	22/640 (3.4%)	23/300 (7.7%)
Kcnq1	50/740 (6.8%)	10/187 (5.3%)
Igf2r	42/630 (6.7%)	24/270 (8.9%)
Homology among primate species ^d		
DMD	Chimp	Gorilla
Snurf/Snrpn	4/300 (1.3%)	2/281 (0.7%)
Kcnq1	4/187 (2.1%)	3/187 (1.6%)
Igf2r	6/270 (2.2%)	33/270 (12.2%)
	Orangutan	Tamarin
		29/232 (12.5%)
	Rhesus	27/188 (14.4%)
	23/261 (8.8%)	
	19/187 (10.2%)	
	Bonobo	
	23/300 (7.7%)	
	11/187 (5.9%)	
	Gorilla	
	23/281 (8.2%)	
	11/187 (5.9%)	
	Orangu	
	22/284 (7.7%)	
	10/188 (5.3%)	
	38/370 (10.3%)	
	Rhesus	
	23/261 (8.8%)	
	7/187 (3.7%)	
	Macaque	
	23/261 (8.8%)	
	Tamarin	
	20/232 (8.6%)	
	5/188 (2.7%)	

^aEstimated from dotplots shown in Figure 1, and displayed as length in nucleotides.

^bLengths in nucleotides of the unit copies of DMD tandem repeats estimated from dotplots shown in Figure 1.

^cCalculated as percentage for regions marked by double-headed arrows in Figure 1.

^dCalculated for aligned sequences shown in Figures 3 and 4.