7 Simple Sequence Repeats in 5′ and 3′ Flanking Sequences of Cell Cycle Genes

Seema Trivedi

Abstract

Simple sequence repeats (SSRs) are hypermutable, and this instability leads to many disorders. Perhaps it is because of this reason SSRs are relatively rare in coding sequences. The present study was undertaken to explore SSRs in 5′ and 3′ flanking sequences (FS) of cell cycle genes (checkpoint; regulation; replication, repair, and recombination (RRR); and transition) in humans and eight mammalian orthologues. The present study shows more SSRs in FS of regulation genes compared to other gene groups. However, differences in repeat numbers between different groups of cell cycle genes are not significant. Trinucleotide repeats are generally more in 3′ FS of human cell cycle genes but not in other mammals (with some exceptions). On the other hand, in 5′ FS of cell cycle genes (except human genes), trinucleotide repeats are more in number compared to other repeat types in almost all mammals (with some exceptions). Repeat numbers do not differ significantly from other mammals except human and cow genes. Many repeats in FS of human genes are conserved, including rare repeats like CG/GC. CG motifs are conserved only in 5′ and 3′ FS of regulation genes but GC motifs are conserved in RRR genes. This paper presents characteristics of SSRs occurring in 5′ and 3′ FS of cell cycle genes, which may be potential mutational hotspots that could be used for further exploration of their potential roles in gene regulation or medical investigations.

Keywords

Cell cycle • Checkpoint • Microsatellites • Replication • Repair • Recombination • Simple sequence repeats (SSRs) • Transition • Flanking sequences

Abbreviations

Di	Dinucleotide
FS	Flanking sequence(s)

S. Trivedi (✉)
Department of Zoology, JN Vyas University, Jodhpur, Rajasthan 342011, India
e-mail: svtrived@hotmail.com

P.R. Sudhakaran (ed.), *Perspectives in Cancer Prevention – Translational Cancer Research*, DOI 10.1007/978-81-322-1533-2_7, © Springer India 2014

Penta	Penta-nucleotide
SSRs	Simple sequence repeats
Tetra	Tetranucleotide
Tri	Trinucleotide

7.1 Introduction

Arrays of DNA motifs of 1–8 base pairs repeated in tandem are known as simple sequence repeats (SSRs, or microsatellites) with density and length variations in different species or even chromosomes of the same species (Chambers and MacAvoy 2000; Toth et al. 2000; Trivedi 2006, 2010). Several factors affect SSR frequency and length in different organisms that include differences in nucleotide content (CG richness) of genome or repeat, repair machinery, mutational pressures, and distance of repeats from replication origin (Jacob and Eckert 2007; Eckert and Hile 2009; Choudhary and Trivedi 2010; Tian et al. 2011). Despite these differences, most organisms have higher SSR density in intergenic regions compared to coding sequences (Chambers and MacAvoy 2000; Toth et al. 2000) with exceptions in Archaea (Trivedi 2006). Moreover, in general trinucleotide and hexanucleotide repeats are more common in coding sequences compared to other repeat types. This may be due to risk of mutations in other repeat types that could lead to nonsense mutation and possible loss of gene expression (Metzgar et al. 2000).

SSRs mutations are mainly due to polymerase strand slippage (Levinson and Gutman 1987) and, to some extent, due to unequal recombination (Li et al. 2002), which may result in expansion or contraction of repeat length. SSR instability can increase by mutations affecting post-replication mismatch mutation repair (MMR) (Strand et al. 1993), or some other mechanism as some prokaryotes do not have MMR (Eisen and Hanawalt 1999).

SSRs may have functional roles (Kashi and King 2006; Lukusa and Fryns 2008; Bacolla and Wells 2009; Eckert and Hile 2009) and may repress transcription (Regelson et al. 2006) or affect efficiency of the cell cycle. For example, AT-rich repeats can affect replication initiation, nucleosome assembly, and DNA supercoiling (Bacolla and Wells 2009). DNA replication time during S-phase can be affected by repeats like (CA)*n* and (ACTG)*n* that are present in the regions flanking later replicating genes and (CATA)n repeats near earlier replicating genes. S-phase checkpoint proteins can influence SSR length mutations especially on SSRs present near the origin of replication or present on lagging strand (Dere et al. 2004).

Since SSRs may affect gene regulation, it might be expected that there would be fewer repeats in 5′ and 3′ flanking sequences (FS) of genes specially cell cycle genes as these regions may also have regulatory elements that may be important for gene expression and may affect cell cyle. Nonetheless, SSRs are present in many genes involved in DNA repair (Chang et al. 2001; Trivedi 2003, 2010). Repeats are also present in genes involved in FS of cell cycle checkpoint, like (CT)19/(CA)16 are present in G0S2 (member of G0S genes actively involved in G0/G1 switch). Mutations in some of these repeats are known to cause tumor or cancer in humans. For example, tumorigenesis is seen due to mutations in the mononucleotide repeat A(9) present in the CtIP gene that plays a role in DNA-damage-induced cell cycle checkpoint control at the G2/M transition and G1/S transition (Russell and Forsdyke 1991).

Though repeats in cell cycle genes have been reported, analysis of SSRs in 5′ and 3′ FS of all cell cycle genes has not been undertaken. The aim of present study is to analyze the distribution of repeats in 5′ and 3′ FS of human cell cycle genes (checkpoint; regulation; replication, repair, and recombination (RRR); and transition) and compare with orthologues in eight mammals. This may help in identification of candidate cell cycle genes that could be vulnerable to mutations in their FS and hence affect cell cycle regulation.

7.2 Material and Methods

7.2.1 Cell Cycle Gene Sequences

Gene Ontology (GO) annotation IDs for *Homo sapiens* cell cycle genes were obtained from AmiGO (Biological Process, all data source, evidence code all, search terms: cell cycle/cell division) (The Gene Ontology Consortium 2000). These GO IDs were used for obtaining *H. sapiens* gene IDs from the Ensembl Genome Browser version 55 (http://www.ensembl.org/biomart/martview/). Human orthologue IDs of chimpanzee (*Pan troglodytes*), orangutan (*Pongo pygmaeus*), macaque (*Macaca mulatta*), horse (*Equus caballus*), cow (*Bos taurus*), dog (*Canis familiaris*), mouse (*Mus musculus*), and rat (*Rattus norvegicus*) were also obtained from Ensembl by using human cell cycle genes' Ensembl gene IDs. Unspliced genes with 50 nucleotides (nts) upstream and downstream flanks from Ensembl were obtained by using these Ensembl gene IDs.

Note: Only GO IDs for genes that have a direct role in the cell cycle were used for further analysis. Further, IDs for genes not directly associated with cell cycle but associated with signaling pathways that affect cell cycle regulation were not included.

7.2.2 Grouping/Classification of Genes

Classification of human cell cycle-related genes was done on the basis of reported function/expression in the cell cycle as per GO "biological process" names (The Gene Ontology Consortium 2000), cell cycle base (Gauthier et al. 2008), reactome (Joshi-Tope et al. 2003; Matthews et al. 2007, 2009; Va Vastrik et al. 2007), WikiPathways (Pico et al. 2008; Kelder et al. 2009), KEGG pathways (http://www.genome.jp/kegg/pathway.html), EntrezGene (http://www.ncbi.nlm.nih.gov/gene/), and UniProtKB/Swiss-Prot (http://www.uniprot.org/uniprot/) into four groups: (1) replication, repair, and recombination (RRR); (2) checkpoint; (3) regulation (translation and transcription regulation excluded); and (4) transition.

7.2.3 Repeat Search and Representation

Repeat search program SPUTNIK (http://espressosoftware.com/sputnik/index.html), which looks for di-, tri-, tetra-, and penta-nucleotide repeats, was used for analysis of cell cycle gene sequences of all nine mammals. Heat map was generated to represent repeat motif frequencies by using web interface of "Matrix2png" (Pavlidis and Noble 2003).

7.2.4 SSR CG Richness and Length

For calculation of SSR lengths, total numbers of nucleotides for each SSR were counted. Further, this length was adjusted to the repeat-divisible value – i.e., if the length of a dinucleotide repeat was given as 13nt by SPUTNIK, this was adjusted to 12nt repeat length and for trinucleotide repeat, if length was given as 16nt, it was adjusted to 15nt.

7.2.5 Repeat Position

3′ and 5′ FS positions in genes were identified based on the positions obtained from Ensembl. Repeats that were present in 3′ and 5′ FS but also extended to either exon or intron of the gene were named as 3′ FS_H and 5′ FS_H, respectively. Repeat positions obtained from SPUTNIK were then compared with positions of 5′ and 3′ FS.

7.2.6 Conservation of Repeats

Ensembl version 55 was used for downloading human cell cycle genes and their aligned (PECAN) orthologues in eight mammals (if present). Each nucleotide of repeats present in 5′ and 3′ FS of human genes was compared with aligned orthologue sequences. Web interface of “Matrix2png” (Pavlidis and Noble 2003) was used to generate heat map to represent repeat motif conservation.

7.2.7 Statistical Analysis

One way ANOVA (analysis of variance) followed by Tukey’s HSD (honestly significant difference and Bonferroni correction) at 95 and 99 % confidence levels was done with the help of SPSS (version 16.0) to gauge significant differences between the nine mammals as well as differences between cell cycle genes within and between species.

Similarly, ANOVA followed by Tukey’s HSD (honestly significant difference and Bonferroni correction) at 95 and 99 % confidence levels was done to seek significance of differences between repeat types as well as different cell cycle gene groups in each mammal.

7.3 Results

7.3.1 Cell Cycle Genes 5′ and 3′ Flanking Sequences and Repeats

Repeat search shows more repeats in 5′ FS compared to 3′ FS in genes of all mammals except human genes. However, there is no consistency in distribution of repeat numbers in FS_H regions of all nine mammals (Fig. 7.1). Moreover, there are no significant differences between total repeats of nine mammals except between human and cow repeats ($P < 0.05$).

Distributions of SSRs types are different in different mammals (Fig. 7.2a and b). Trinucleotides are the most abundant repeat class in human 3′ FS but are not present in chimp, orangutan, horse, and cow genes. In 5′ FS, though tri- and penta-nucleotide repeats are present in all mammals, dinucleotide repeats are not present in dog genes, and tetranucleotide repeats are not present in human and horse genes. There are no significant differences between repeat types in nine mammals.

7.3.2 Cell Cycle Gene Groups and Repeats

Only regulation genes have repeats in all regions, i.e., 3′ and 5′ FS and FS_H of all mammals (Fig. 7.3). Distributions of repeats in different gene groups also show higher repeat numbers in FS of regulation genes in 5′ and 3′ FS compared to other gene groups. There are no significant differences in repeat numbers in gene groups of nine mammals except between regulation genes of human and cow ($P < 0.05$). Further, within each mammal, different gene groups do not show significant differences in repeat numbers.

Dinucleotide repeats are not present in 3′ FS of checkpoint genes in human, chimp, orangutan, and macaque and regulation genes of rat. Only human and orangutan RRR genes and transition genes of horse and mouse have dinucleotide repeats. 5′ FS has repeats only in macaque checkpoint genes and RRR genes of only orangutan. Human, horse, and dog regulation genes do not have dinucleotide repeats in 5′ FS. 3′ FS_H has dinucleotide repeats only in regulation genes of macaque, horse, and rat. 5′ FS_H has repeats only in macaque checkpoint genes, human and chimp regulation genes, and macaque and mouse transition genes (Fig. 7.4). Similarly, distributions of tri-, tetra-, and penta-nucleotide repeats are different in different gene groups of the nine mammals (Fig. 7.4). Trinucleotide repeats compared to other repeat types are more in regulation genes of all mammals (in particular non-primate). It is interesting to note that 5′ FS of checkpoint genes do not have tetranucleotide repeats in all mammals. Repeat types show no significant differences in each mammal.

7.3.3 Motifs in Cell Cycle Genes

Distribution of repeat motifs varies in different gene groups of all nine mammals (Fig. 7.5a and b). It is interesting to note that GC repeat motif in 3′ FS is present only in human RRR genes. CG dinucleotide repeats are present only in 5′ FS of orangutan regulation genes. Among trinucleotide repeats GC-rich motifs like CCG,

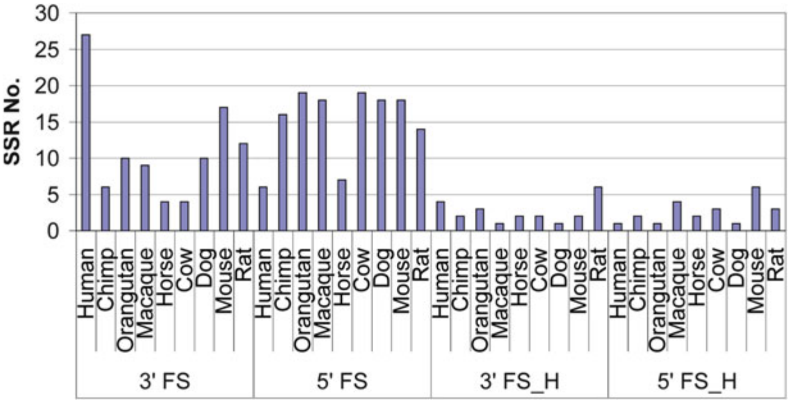

Fig. 7.1 Total repeats in flanking sequences (*FS*) of cell cycle genes in nine mammals. *FS_H* repeats that extend from either 3′ or 5′ flanking sequences to either exons or introns of gene sequences

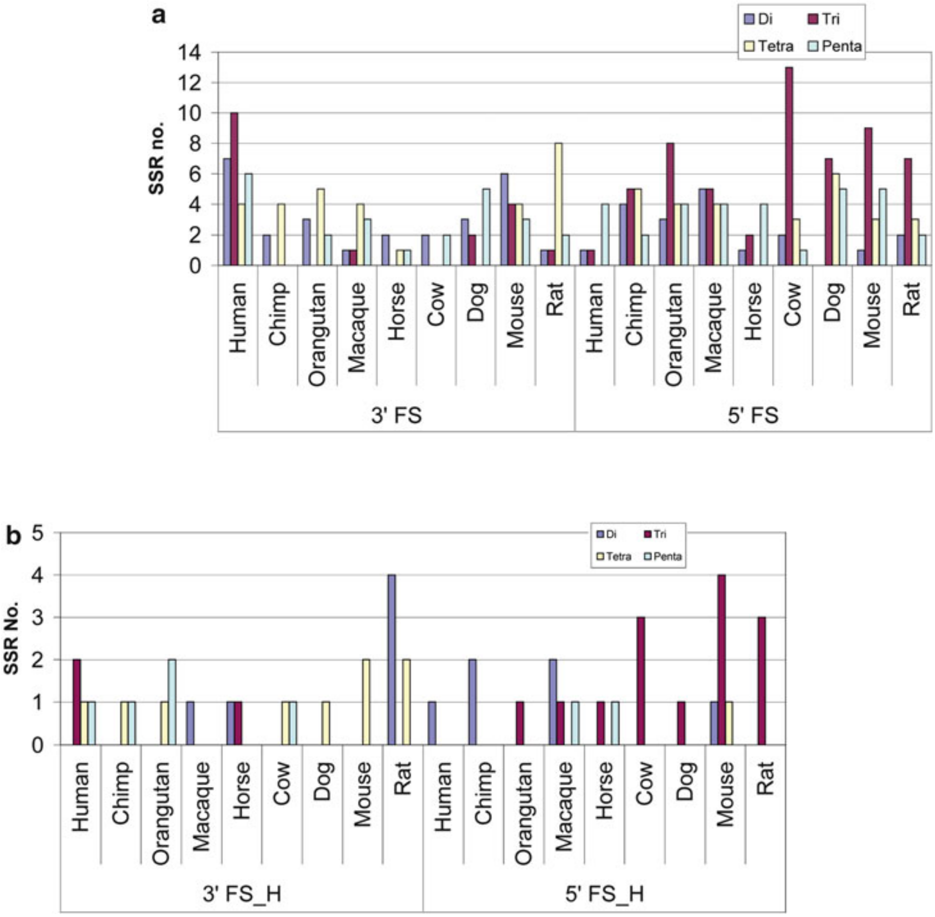

Fig. 7.2 (**a**) Repeat types in flanking sequences (*FS*) of cell cycle genes in nine mammals. *Di* – dinucleotide, *Penta* – penta-nucleotide, *Tri* – trinucleotide, *Tetra* – tetranucleotide. (**b**) Repeat types that extend from either 3′ or 5′ flanking sequences to either exons or introns of cell cycle gene sequences (*FS_H*) in nine mammals. *Di* – dinucleotide, *Penta* – penta-nucleotide, *Tri* – trinucleotide, *Tetra* – tetranucleotide

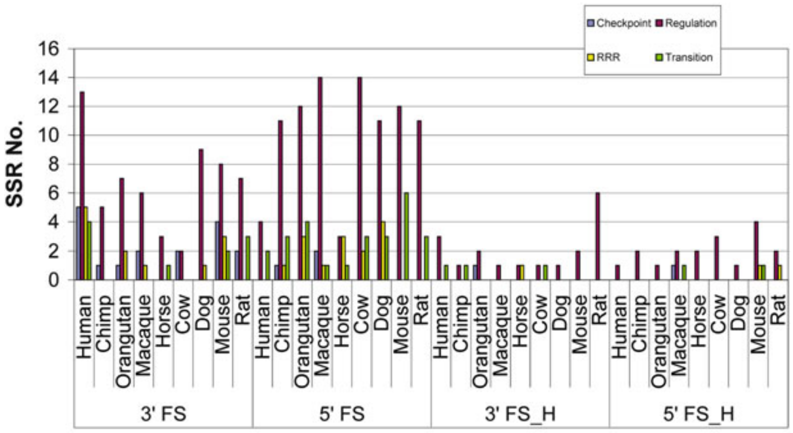

Fig. 7.3 Repeats in flanking sequences (*FS*) of different groups of cell cycle genes in nine mammals. *FS_H* repeats that extend from either 3′ or 5′ flanking sequences to either exons or introns of gene sequences, *RRR* replication, repair, and recombination

CCT, CGC, GCG, GGA, and GGC are more common in 5′ FS of regulation genes in nine mammals (with exceptions). Similarly, tetra- and penta-nucleotide repeat motifs have different distribution in 5′ and 3′ FS of different cell cycle genes in nine mammals (Fig. 7.5a and b).

7.3.4 Conservation of Repeats

In 5′ FS, only regulation and transitions genes show conservation of repeat motifs (CCTG and CCT, respectively). In 3′ FS GC motif is conserved in RRR genes of all mammals except horse. GA (except dog), TC (except macaque), and TG repeat motifs are conserved in regulation genes of all mammals. Dinucleotide repeats are not conserved in checkpoint and transition genes. Among trinucleotide repeat motifs, ATA, CCG, and TTC are conserved in checkpoint genes, CGC and TCC are conserved only in regulation genes, but GCG motifs are conserved in all gene groups except checkpoint genes. Tetra- and penta-nucleotide motifs show differences in conservation in different groups. Further, some motifs are not conserved in all mammals (Fig. 7.6). In 5′ FS_H only CG dinucleotide repeat is conserved (regulation genes) in all mammals except horse. In 3′ FS_H only trinucleotide motifs are conserved in regulation genes where CGC is conserved in all mammals except dog and CGG is conserved only in mouse and rat (data not shown).

7.4 Discussion

During DNA replication, strand slippage often leads to SSR flux which is corrected by replication repair and the checkpoint machinery (Chambers and MacAvoy 2000; Lahiri et al. 2004).

The present analysis shows repeats in human FS of cell cycle genes as well as in eight eutherian orthologues. Moreover, repeat frequency is higher in 5′ FS in genes of all mammals except human genes (3′ FS). Though many repeats may be neutral in nature, it is known that SSRs can form alternative non-B DNA structures and affect normal DNA replication or MMR system, leading to enhanced cancer susceptibility and neurodegenerative disorders (Lukusa and Fryns 2008; Bacolla and Wells 2009; Eckert and Hile 2009). Longer (AAAG)(*n*) repeats serve as binding site of many transcription factors in 5′-UTRs of the estrogen-related receptor gamma gene (ERR-γ) in breast cancer patients (Galindo et al. 2011).

It is known that repeat length alterations in CDKN2A and CCND1 result in dysplastic head and neck lesions (Tripathi Bhar et al. 2003). The present study finds penta-nucleotide repeat in 3′ FS of human and macaque CDKN2A gene besides repeats in other regions of the gene. Among human genes in the present study that contain repeats, some are listed in the Mendelian Inheritance in Man morbid (MIM) database (http://www.ncbi.nlm.nih.gov/omim/) (obtained

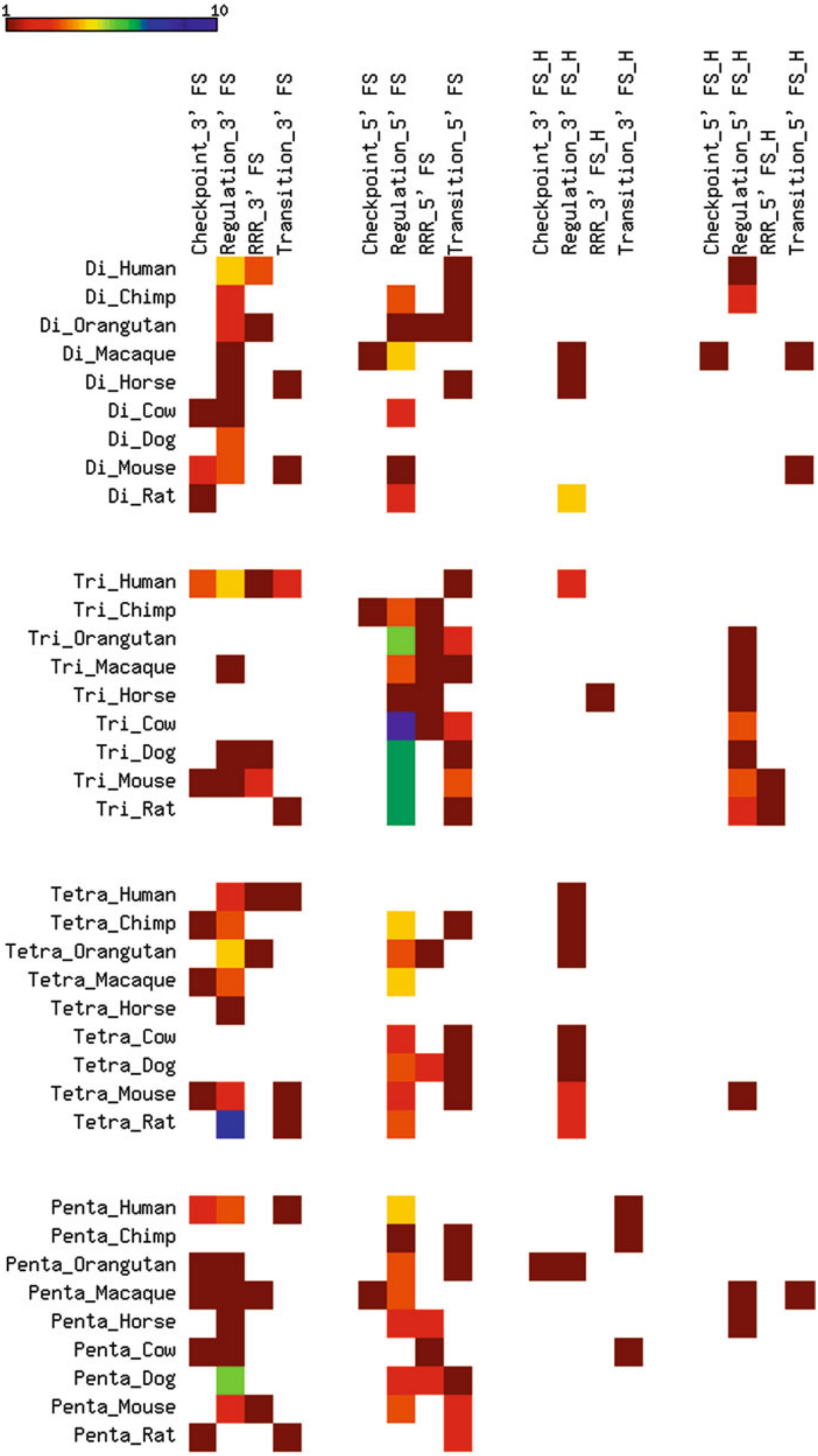

Fig. 7.4 Repeat types in flanking sequences of cell cycle gene groups

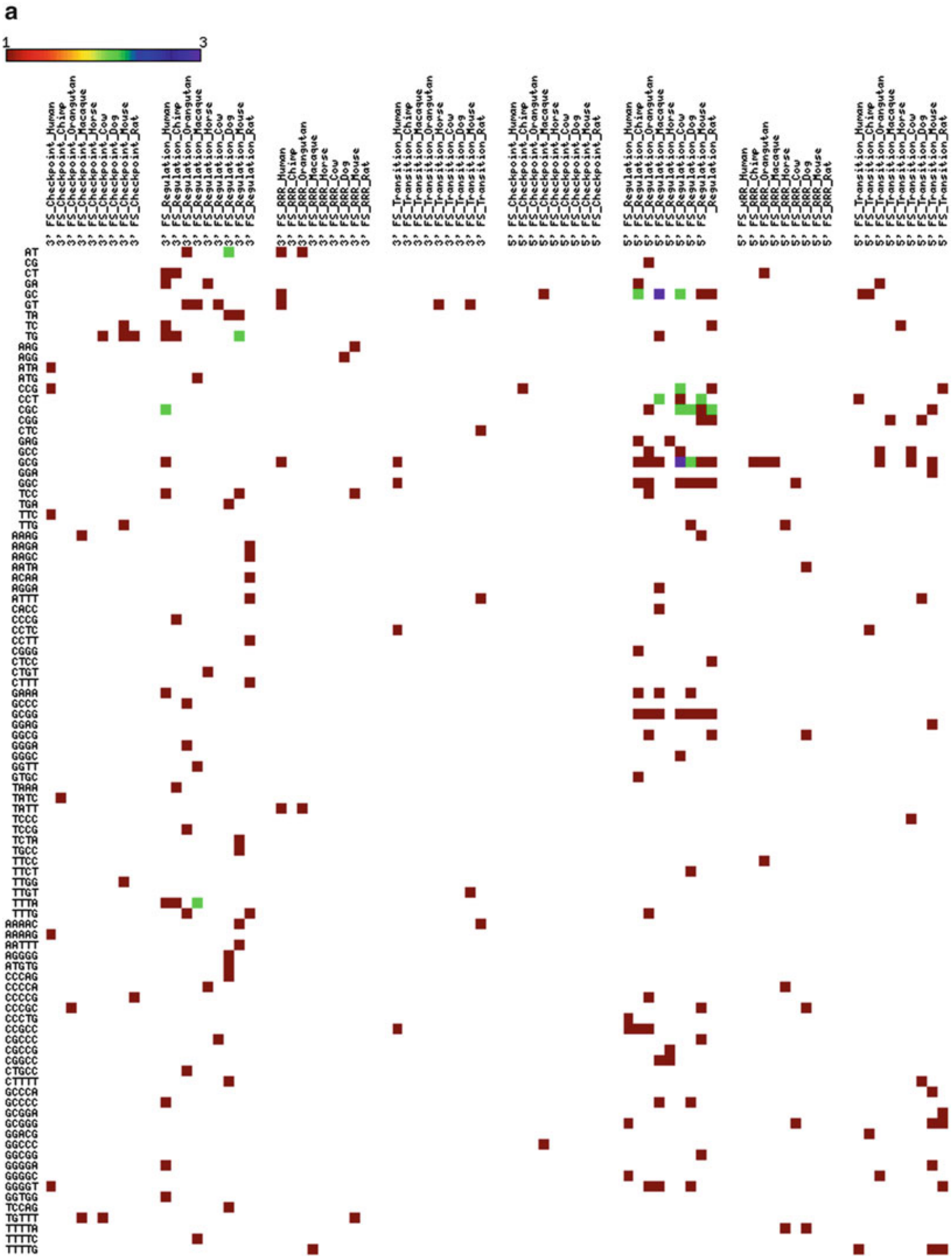

Fig. 7.5 (**a**) Repeat motifs in 3′ and 5′ flanking sequences of cell cycle genes in nine mammal. (**b**) Repeat motifs in flanking and overlapping sequences of cell cycle genes

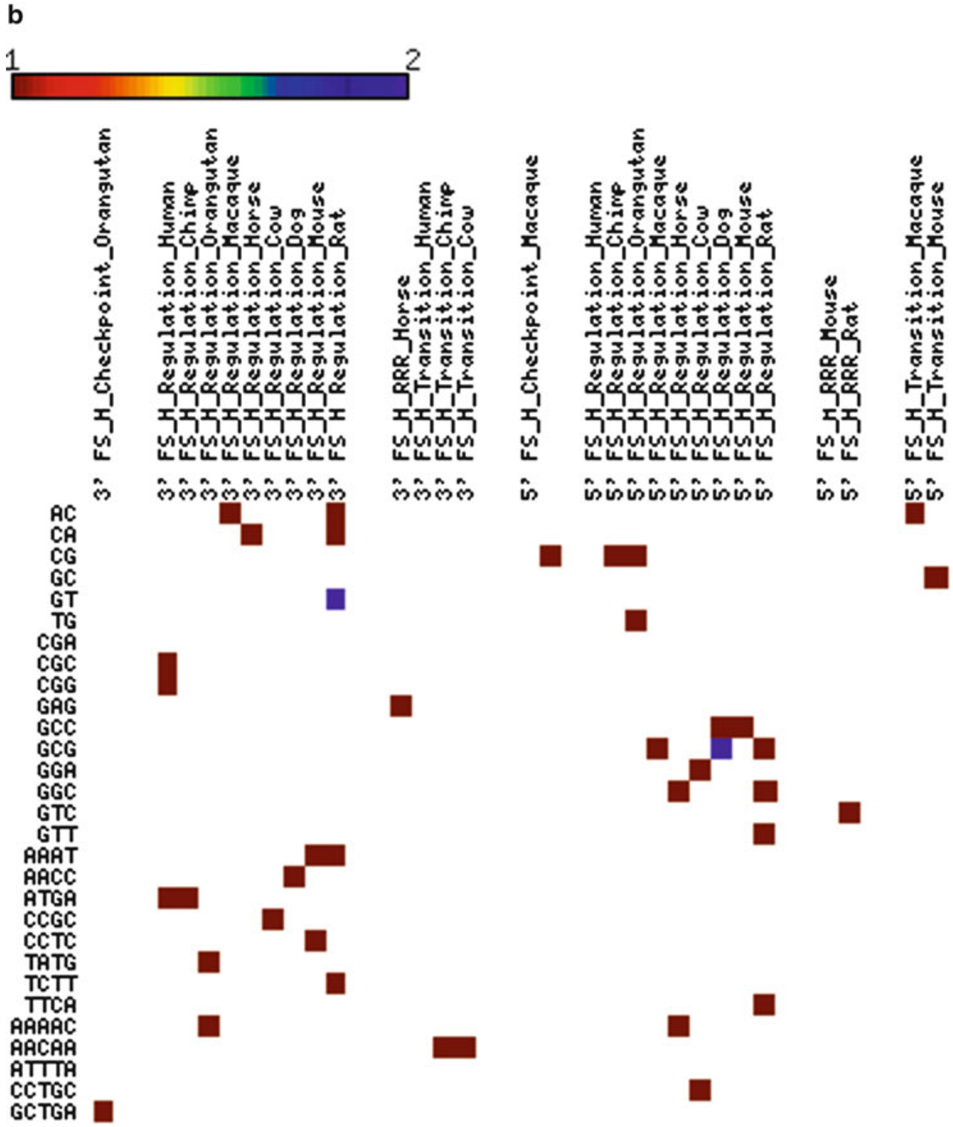

Fig. 7.5 (continued)

through Ensembl biomart version 55) (Table 7.1). Many disorders including different types of cancers are associated with these genes. These facts point to the possibility that SSRs in FS of cell cycle genes may contribute to disorders due to SSR mutations that may affect gene regulation.

Importance of SSRs in cell cycle 5′ and 3′ FS is further confirmed by establishment of association of mutations in SSRs present in 5′ and 3′-UTRs (that may be part of FS) with different types of cancers. Mutations in coding regions or methylation of promoters of MMR genes may lead to certain cancers like hematological malignancies though not always detected in all patients. In such cases, MMR deficiency may be due to mutation in the MLH1 3′-untranslated region (3′-UTR), though the exact mechanism is unknown (Mao et al. 2008). Poly (T)8 repeat deletion within the 3′-UTR of the CDK2-AP1 gene results in its decreased expression due to reduced mRNA stability and is associated with MSI human colorectal cancer (Shin et al.

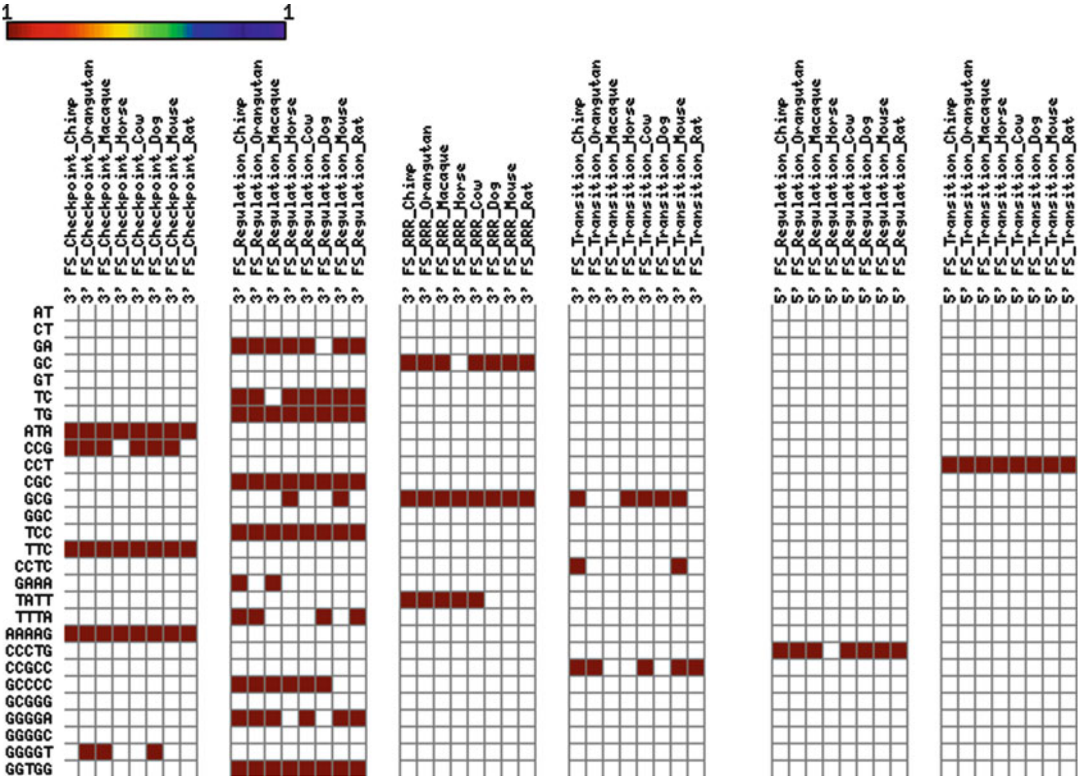

Fig. 7.6 Conserved motifs in human cell cycle genes flanking sequences. Blank cells indicate no conservation

Table 7.1 Repeat (in flanking sequences) containing human cell cycle gene names and MIM morbid description

Gene name	MIM morbid description
MCM6	Lactase persistence
TGFB1	Camurati-Engelmann disease
ACVR1	Fibrodysplasia ossificans progressiva
BMP4	Microphthalmia, syndromic 6
MTUS1	Hepatocellular carcinoma
MEN1	Hyperparathyroidism 1, multiple endocrine neoplasia, Type I
CDC73	Hyperparathyroidism 1 and 2, parathyroid carcinoma
AKT1	Breast cancer, breast-ovarian cancer, familial, colorectal cancer
CDKN2A	Li-Fraumeni syndrome 1, melanoma, cutaneous malignant, uveal, melanoma-astrocytoma syndrome, melanoma-pancreatic cancer syndrome
PTEN	Bannayan-Riley-Ruvalcaba syndrome, Cowden disease, endometrial cancer, glioma of brain familial, macrocephaly/autism syndrome, prostate cancer, Proteus syndrome, squamous cell carcinoma, head and neck, Vacterl association with hydrocephalus
EXT1	Chondrosarcoma, exostoses, multiple, Type I; trichorhinophalangeal syndrome, Type II
SEPT6	Amyotrophy, hereditary neuralgic

2007). Dinucleotide deletion in the 3′-UTR of CD24 also causes mRNA instability that may result in multiple sclerosis (MS) and systemic lupus erythematosus (SLE) (Wang et al. 2007). Microsatellite instability-high (MSI-H) tumors (in endometrial and colorectal carcinomas) show deletions of 3′-UTR polyA in epidermal growth factor receptor (EGFR) (Deqin et al. 2012), and its association is also seen in gastric cancer (Corso et al. 2011) and colon cancers (Yuan et al. 2009).

It is possible that despite risks of disorders, SSRs may exist in FS of cell cycle genes due to the possible involvement of repeats in processes like gene regulation and chromatin organization, or act as source of genetic variability that may be useful for rapid adaptation (Chang et al. 2001; Li et al. 2002; Fondon and Garner 2004; Kashi and King 2006). Repeats that can form loops and other secondary structures or cause unfolding of DNA can be advantageous for transcription (Li et al. 2002). Compared with CAG and CUG tracts, there is higher degree of stability of structure due to CGG repeats (Kiliszek et al. 2011). It is possible that some repeats found in the present study like CGG in CDKN2D and AKT1 are also associated with normal binding of transcription factors or other DNA binding proteins with or without formation of non-B DNA structures.

Single nucleotide polymorphism in dinucleotide repeat (TA)(*n*) in the promoter region of IL-28B affects transcriptional activity in length-dependent manner (Sugiyama et al. 2011). Therefore, it is possible that repeats in FS of cell cycle genes may also facilitate transcription during the division cycle especially during chromatin condensation. Z-DNA conformation is formed due to dinucleotide repeats present in the negative regulatory element (NRE, a transcriptional repressor) at the 5′-UTR of ADAM-12 gene (member of the multifunctional ADAM family of proteins linked to cancer, arthritis, and cardiac hypertrophy) and is essential for interaction with Z-DNA-binding protein that in turn repress transcription of ADAM-12 gene (Ray et al. 2011). In the present study, more dinucleotide repeats are present in 3′ FS compared to 5′ FS in human cell cycle gene but no motif type is particularly more in number. But the distribution is different in other mammals. Few genes in the present study contain CG/GC motifs or CG-rich repeats in FS. This could be because methylation and/or deamination of polymerase slippages is high in CG-rich motifs (Tian et al. 2011; Li and Chen 2011); thus, GC-rich SSRs are more unstable (Zahra et al. 2007; Tian et al. 2011), and, therefore, only essential CG-rich motifs may be retained in genes.

It is also possible that during G0 phase when activation of cell cycle genes (except recruitment for DNA damage) may not be necessary, SSRs in 5′ and 3′ FS may play a role in keeping these genes repressed until the cell cycle is reinitiated.

The present study indicates that the nine mammals are similar in terms of the repeat numbers in 5′ and 3′ FS of cell cycle genes, as there are no significant differences between them. Furthermore, many human repeats show conservation in FS of cell cycle genes all mammals studied here. The present study does not show presence of long repeats in primates except dinucleotide repeat GT (28nt) in 3′ FS of human RRR genes and pentanucleotide repeat GCCCC (35nt) in regulation genes. In 5′ FS also repeats are not long in primates except dinucleotide TG (44nt) in macaque regulation genes (data not shown). It is known that mutation rates are higher in long SSRs compared to shorter repeats. This is because the likelihood of substitutions that can interrupt repeat length is more in long repeats and can curtail their infinite growth (Kruglyak et al. 1998; Ellegren 2000; Xu et al. 2000; Dieringer and Schlotterer 2003). Further, short motifs are compared to motifs like tetra- and penta-nucleotide repeats are more unstable (Jacob and Eckert 2007; Eckert and Hile 2009). Some functional roles of tetranucleotide repeats are known (Li et al. 2004; Csink and Henikoff 1998); it is also known that repeat instability in tetranucleotide repeats may be a cause of cancer (Li et al. 2004). Possibly this could be the reason for conservation of very few tetranucleotide repeats in the present study.

7.5 Conclusion

The present study shows presence of repeats in 5′ and 3′ FS of cell cycle genes. SSRs in FS of regulation genes are present in all nine mammals studied but significance remains untested. There is no significant difference in total repeats between humans and the other eight mammals, except cow, and many repeats are conserved. The fact that some of these repeat-associated genes are in the MIM morbid database indicates that SSRs may act as “chink in the armor” for these critical genes. In future, it would be useful for investigators to recognize the potential role of instability of repeats and identify cycle-related disorders.

Acknowledgments I am grateful to Mr. Abhay Pendse for help with computational work. I am indebted to Ensembl workgroup especially Dr. Javier Herrero (Ensembl Compara Project Leader, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, UK) for their help in obtaining multiple aligned sequences. Further, I am obliged to all open-source journals, databases, and scientists who provided reprints and information related to this work.

References

Bacolla A, Wells RD (2009) Non-B DNA conformations as determinants of mutagenesis and human disease. Mol Carcinog 48(4):273–285

Chambers GK, MacAvoy ES (2000) Microsatellites: consensus and controversy. Comp Biochem Physiol B Biochem Mol Biol 126(4):455–476

Chang DK, Metzgar D, Wills C, Boland CR (2001) Microsatellites in the eukaryotic DNA mismatch repair genes as modulators of evolutionary mutation rate. Genome Res 11(7):1145–1146

Choudhary OP, Trivedi S (2010) Microsatellite or simple sequence repeat (SSR) instability depends on repeat characteristics during replication and repair. J Cell Molec Biol 8(2):21–34

Corso G, Velho S, Paredes J, Pedrazzani C, Martins D, Milanezi F, Pascale V, Vindigni C, Pinheiro H, Leite M, Marrelli D, Sousa S, Carneiro F, Oliveira C, Roviello F, Seruca R (2011) Oncogenic mutations in gastric cancer with microsatellite instability. Eur J Cancer 47(3):443–451

Csink AK, Henikoff S (1998) Something from nothing: the evolution and utility of satellite repeats. Trends Genet 14(5):200–204

Deqin M, Chen Z, Nero C, Patel KP, Daoud EM, Cheng H, Djordjevic B, Broaddus RR, Medeiros LJ, Rashid A, Luthra R (2012) Somatic Deletions of the PolyA Tract in the 3′ Untranslated Region of Epidermal Growth Factor Receptor Are Common in Microsatellite Instability-High Endometrial and Colorectal Carcinomas. Arch Pathol Lab Med 136(5):510–516

Dere R, Napierala M, Ranum LP, Wells RD (2004) Hairpin structure-forming propensity of the (CCTG.CAGG) tetranucleotide repeats contributes to the genetic instability associated with myotonic dystrophy type 2. J Biol Chem 279(40): 41715–41726

Dieringer D, Schlotterer C (2003) Two distinct modes of microsatellite mutation processes: evidence from the complete genomic sequences of nine species. Genome Res 13:2242–2251

Eckert KA, Hile SE (2009) Every microsatellite is different: Intrinsic DNA features dictate mutagenesis of common microsatellites present in the human genome. Mol Carcinog 48(4):379–388

Eisen JA, Hanawalt PC (1999) A phylogenomic study of DNA repair genes, proteins, and processes. Mut Res 435:171–213

Ellegren H (2000) Heterogeneous mutation processes in human microsatellite DNA sequences. Nat Genet 24:400–402

Fondon JW 3rd, Garner HR (2004) Molecular origins of rapid and continuous morphological evolution. Proc Natl Acad Sci USA 101(52):18058–18063

Galindo CL, McCormick JF, Bubb VJ, Abid Alkadem DH, Li LS, McIver LJ, George AC, Boothman DA, Quinn JP, Skinner MA, Garner HR (2011) A long AAAG repeat allele in the 5′-UTR of the ERR-γ gene is correlated with breast cancer predisposition and drives promoter activity in MCF-7 breast cancer cells. Breast Cancer Res Treat 130(1):41–48

Gauthier NP, Larsen ME, Wernersson R, de Lichtenberg U, Jensen LJ, Brunak S, Jensen TS (2008) Cyclebase.org-a comprehensive multi-organism online database of cell-cycle experiments. Nucleic Acids Res 36(Database issue):D854–9

Jacob KD, Eckert KA (2007) Escherichia coli DNA polymerase IV contributes to spontaneous mutagenesis at coding sequences but not microsatellite alleles. Mut Res 619(1–2):93–103

Joshi-Tope G, Vastrik I, Gopinath GR, Matthews L, Schmidt E, Gillespie M, D'Eustachio P, Jassal B, Lewis S, Wu G, Birney E, Stein L (2003) The genome knowledgebase: a resource for biologists and bioinformaticists. Cold Spring Harb Symp Quant Biol 68: 237–243

Kashi Y, King DG (2006) Simple sequence repeats as advantageous mutators in evolution. Trends Genet 22(5):253–259

Kelder T, Pico AR, Hanspers K, van Iersel MP, Evelo C, Conklin BR (2009) Mining biological pathways using WikiPathways web services. PLoS One 4(7)

Kiliszek A, Kierzek R, Krzyzosiak WJ, Rypniewski W (2011) Crystal structures of CGG RNA repeats with implications for fragile X-associated tremor ataxia syndrome. Nucleic Acids Res 39(16):7308–7315

Kruglyak S, Durrett RT, Schug MD, Aquadro CF (1998) Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. Proc Natl Acad Sci USA 95:10774–10778

Lahiri M, Gustafson TL, Majors ER, Freudenreich CH (2004) Expanded CAG repeats activate the DNA damage checkpoint pathway. Mol Cell 15(2):287–293

Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4:203–221

Li M, Chen SS (2011) The tendency to recreate ancestral CG dinucleotides in the human genome. BMC Evol Biol 11:3

Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol Ecol 11(12):2453–2465

Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function, and evolution. Mol Biol Evol 21(6):991–1007

Lukusa T, Fryns JP (2008) Human chromosome fragility. Biochim Biophys Acta 1779(1):3–16

Mao G, Pan X, Gu L (2008) Evidence that a mutation in the MLH1 3′-untranslated region confers a mutator phenotype and mismatch repair deficiency in patients with relapsed leukemia. J Biol Chem 283(6): 3211–3216

Matthews L, D'Eustachio P, Gillespie M, Croft D, de Bono B, Gopinath G, Jassal B, Lewis S, Schmidt E, Vastrik I, Wu G, Birney E, Stein L (2007) An introduction to the reactome knowledgebase of human biological pathways and processes. Bioinform Primer NCI/Nat Pathway Interact Datab. doi:10.1038/pid.2007.3

Matthews L, Gopinath G, Gillespie M, Caudy M, Croft D, de Bono B, Garapati P, Hemish J, Hermjakob H, Jassal B, Kanapin A, Lewis S, Mahajan S, May B, Schmidt E, Vastrik I, Wu G, Birney E, Stein L, D'Eustachio P (2009) Reactome knowledgebase of biological pathways and processes. Nucleic Acids Res 37(Database issue):D619–D622

Metzgar D, Bytof J, Wills C (2000) Selection against frameshift mutations limits microsatellite expansion in coding DNA. Genome Res 10(1):72–80

Pavlidis P, Noble WS (2003) Matrix2png: a utility for visualizing matrix data. Bioinformatics 19:295–296. doi:10.1093/bioinformatics/19.2.295

Pico AR, Kelder T, van Iersel MP, Hanspers K, Conklin BR, Evelo C (2008) WikiPathways: pathway editing for the people. PLoS Biol 6(7)

Ray BK, Dhar S, Shakya A, Ray A (2011) Z-DNA-forming silencer in the first exon regulates human ADAM-12 gene expression. Proc Natl Acad Sci USA 108(1):103–108

Regelson M, Eller CD, Horvath S, Marahrens Y (2006) A link between repetitive sequences and gene replication time. Cytogenet Genome Res 112(3–4):184–193

Russell L, Forsdyke DR (1991) A human putative lymphocyte G0/G1 switch gene containing a CpG-rich island encodes a small basic protein with the potential to be phosphorylated. DNA Cell Biol 10(8):581–591

Shin J, Yuan Z, Fordyce K, Sreeramoju P, Kent TS, Kim J, Wang V, Schneyer D, Weber TK (2007) A del T poly T (8) mutation in the 3′ untranslated region (UTR) of the CDK2-AP1 gene is functionally significant causing decreased mRNA stability resulting in decreased CDK2-AP1 expression in human microsatellite unstable (MSI) colorectal cancer (CRC). Surgery 142(2):222–227

Strand M, Prolla TA, Liskay RM, Petes TD (1993) Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature 365:274–276

Sugiyama M, Tanaka Y, Wakita T, Nakanishi M, Mizokami M (2011) Genetic variation of the IL-28B promoter affecting gene expression. PLoS One 6(10):e26620

The Gene Ontology Consortium (2000) Gene Ontology: tool for the unification of biology. Nature Genet 25:25–29. http://amigo.geneontology.org/cgi-bin/amigo/browse.cgi

Tian X, Strassmann JE, Queller DC (2011) Genome nucleotide composition shapes variation in simple sequence repeats. Mol Biol Evol 28(2):899–909

Toth G, Gaspari Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res 10:967–981

Tripathi Bhar A, Banerjee S, Chunder N, Roy A, Sengupta A, Roy B, Roychowdhury S, Panda CK (2003) Differential alterations of the genes in the CDKN2A-CCND1-CDK4-RB1 pathway are associated with the development of head and neck squamous cell carcinoma in Indian patients. J Cancer Res Clin Oncol 129(11):642–650

Trivedi S (2003) Do microsatellites have biased associations? Nucleus 46:61–76

Trivedi S (2006) Comparison of simple sequence repeats in 19 Archaea. Genet Mol Res 5(4):741–772

Trivedi S (2010) Do simple sequence repeats in replication, repair and recombination genes of mycoplasmas provide genetic variability? J Cell Mole Biol 7(2) & 7(2 & 8(1)):53–70

Va Vastrik I, D'Eustachio P, Schmidt E, Joshi-Tope G, Gopinath G, Croft D, de Bono B, Gillespie M, Jassal B, Lewis S, Matthews L, Wu G, Birney E, Stein L (2007) Reactome: a knowledge base of biologic pathways and processes. Genome Biol 8:R39

Wang L, Lin S, Rammohan KW, Liu Z, Liu JQ, Liu RH, Guinther N, Lima J, Zhou Q, Wang T, Zheng X, Birmingham DJ, Rovin BH, Hebert LA, Wu Y, Lynn DJ, Cooke G, Yu CY, Zheng P, Liu Y (2007) A dinucleotide deletion in CD24 confers protection against autoimmune diseases. PLoS Genet 3(4):e49

Xu X, Peng M, Fang Z (2000) The direction of microsatellite mutations is dependent upon allele length. Nat Genet 24:396–399

Yuan Z, Shin J, Wilson A, Goel S, Ling YH, Ahmed N, Dopeso H, Jhawer M, Nasser S, Montagna C, Fordyce K, Augenlicht LH, Aaltonen LA, Arango D, Weber TK, Mariadason JM (2009) An A13 repeat within the 3′-untranslated region of epidermal growth factor receptor (EGFR) is frequently mutated in microsatellite instability colon cancers and is associated with increased EGFR expression. Cancer Res 69(19): 7811–7818

Zahra R, Blackwood JK, Sales J, Leach DR (2007) Proofreading and secondary structure processing determine the orientation dependence of CAG × CTG trinucleotide repeat instability in Escherichia coli. Genetics 176(1): 27–41