PRIMARY RESEARCH

Open Access



Genome-scale portrait and evolutionary significance of human-specific core promoter tri- and tetranucleotide short tandem repeats

N. Nazaripanah¹, F. Adelirad¹, A. Delbari¹, R. Sahaf¹, T. Abbasi-Asl² and M. Ohadi^{1*}

Abstract

Background: While there is an ongoing trend to identify single nucleotide substitutions (SNSs) that are linked to inter/intra-species differences and disease phenotypes, short tandem repeats (STRs)/microsatellites may be of equal (if not more) importance in the above processes. Genes that contain STRs in their promoters have higher expression divergence compared to genes with fixed or no STRs in the gene promoters. In line with the above, recent reports indicate a role of repetitive sequences in the rise of young transcription start sites (TSSs) in human evolution.

Results: Following a comparative genomics study of all human protein-coding genes annotated in the GeneCards database, here we provide a genome-scale portrait of human-specific short- and medium-size (\geq 3-repeats) tri- and tetranucleotide STRs and STR motifs in the critical core promoter region between – 120 and + 1 to the TSS and evidence of skewing of this compartment in reference to the STRs that are not human-specific (Levene's test *p* < 0. 001). Twenty-five percent and 26% enrichment of human-specific transcripts was detected in the tri and tetra human-specific compartments (mid-*p* < 0.0002 and mid-*p* < 0.002, respectively).

Conclusion: Our findings provide the first evidence of genome-scale skewing of STRs at a specific region of the human genome and a link between a number of these STRs and TSS selection/transcript specificity. The STRs and genes listed here may have a role in the evolution and development of characteristics and phenotypes that are unique to the human species.

Keywords: Short tandem repeat, Core promoter, Human-specific, Trinucleotide, Tetranucleotide

Introduction

Speciation and evolution are, at least in part, due to the plasticity (expansion or contraction) of short tandem repeats (STRs)/microsatellites, which can function as "tuning knobs" in response to the environment or other genes [1-3]. In line with the above, certain STRs are directionally expanded in the human species or co-occur identically in related taxa such as primates [4-8]. Genes that contain STRs in their promoters have higher expression divergence compared to genes with fixed or no

* Correspondence: mi.ohadi@uswr.ac.ir; ohadi.mina@yahoo.com

¹Iranian Research Center on Aging, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran STRs in the gene promoters [9]. Recent reports indicate a role of repetitive sequences in the rise of young transcription start sites (TSSs) in human evolution [10-12].

Preliminary data on the sequencing of a number of "exceptionally long" STRs (\geq 6-repeats), which compose 1–2% of all human core promoter STRs [3], support critical evolutionary adaptive roles for a number of these STRs. Human specificity of the predominant allele of the *RIT2* core promoter STR in the human species, the presence of the shortest allele of this STR (5-repeat) in huntergatherer humans (BUSHMAN KB1: rs113265205), the lack of this allele in the agricultural modern humans (Genome Aggregation database: gnomad.broadinstitute. org), and its co-occurrence with schizophrenia provide the



© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

Full list of author information is available at the end of the article

first indication of STR allele selection in humans [13]. A link between the *CYTH4* core promoter STR (the longest tetranucleotide STR identified in a human gene core promoter) with the Old World monkeys and Apes and evidence of extreme "disease-only" genotypes at this STR with schizophrenia [14] provide the first link between a primate-specific STR and higher-order brain functions in human. The "exceptionally long" CA-repeat in the core promoter of *SCGB2B2* is another example of directional STR expansion in the Old World monkeys and Apes [5]. The *PAXBP1* gene is an extreme example in which expansion of a core promoter CT-repeat occurs in the Old World monkeys and reaches maximum length and complexity in human; OMIM: 617621 [4].

As "exceptionally long" STRs may be subject to natural selection, short- and medium-size alleles (\geq 3-repeats) might have had similar fate. This is indicated by the predominance of specific short- and medium-size pentaand hexanucleotide STRs and their cognate transcription factors (TFs) in the critical core promoter interval [15]. Indeed, shortening of a number of STRs and their identical co-occurrence is linked to the evolution of primates [8]. In line with the above findings, repeats associated with younger human TSSs tend to be shorter than those in older TSSs [10]. In the study reported here, we present genome-scale data on two categories of STRs, i. e., tri- and tetranucleotide STRs, and their implication in human evolution.

Materials and methods

The interval between – 120 and + 1 to the TSS of all human protein-coding genes annotated in the GeneCards database (version 3.0) (www.genecards.org) was screened for tri- and tetranucleotide STRs of \geq 3-repeats, based on the Ensembl database (versions 87-91) (asia.ensembl. org) and using the Microsatellite Repeats Finder at the following link: http://insilico.ehu.es/mini_tools/microsatellites/

The evolutionary status of the identified STRs was analyzed in 25 species (*N*), including primates (N = 5), non-primate mammals (N = 12), birds and reptiles (N = 5), amphibians (N = 1), and fish (N = 2), based on the Ensembl database.

Human specificity of transcripts was evaluated based on the multiple and pair-wise %identity scoring of the TSS-flanking 5' untranslated region (UTR), using the sequence alignment program Clustal Omega (https://www. ebi.ac.uk/Tools/msa/clustalo), and the overall composition of the transcript and encoded protein (i.e., length of the transcript, number of exons and amino acids). The threshold of sequence identity was set at 50%, which was based on the comparison of two randomly selected and unrelated sequences in the human genome. The p value for the skewing of the human-specific STR compartment was calculated using Levene's equality of variances test.

The *p* values for transcript enrichment were calculated using the two by two table analysis;

the human-specific tri- and tetranucleotide STR groups were compared against corresponding randomly selected STRs from the non-human-specific STRs. The comparison was set based on the sample size of the human-specific STRs (n) and the sample size of the non-human compartments (1.5n).

Results

Overall prevalence of tri- and tetranucleotide STR motifs across human protein-coding core promoter sequences

In total, 56 and 82 STR motifs were detected for the triand tetranucleotide repeats, respectively (Figs. 1 and 2). The most prevalent tri- and tetranucleotide STR motifs across the human protein-coding gene core promoters were GGC and GGGC, respectively (Figs. 1 and 2). In the category of non-GC STRs, GGA and TCCC were the most prevalent tri- and tetranucleotides, respectively.

Skewing of the human-specific core promoter tri- and tetranucleotide STRs

A significant skewing of the tri- and tetranucleotide STR distribution was found in the human-specific tri-(Fig. 1) and tetranucleotide (Fig. 2) compartments (Levene's p < 0.001). While the most prevalent tri- and tetranucleotide repeats in the non-human-specific category were the GGC- and GGGC-repeats, respectively, the most prevalent human-specific STRs were of the GCC and CTCC motifs, respectively. Disproportionate distribution of human-specific STRs was also detected in other STRs such as CCT, GAA, CTCC, GTTT, and GAAA.

The human-specific tri- and tetranucleotide STRs were of a wide range of motifs, e.g., the CCA motif in *ADCY6*, the TCCC motif in *ARHGEF35*, GCCC in *DRD2*, and GTTT in *MCTP2* (Tables 1 and 2).

In a number of instances, not only the STR, but also the genes containing those STRs, were human-specific (e.g., *ARHGEF35*, *AMY1C*, and *Clorf204*). Furthermore, a number of the tri- and tetranucleotide STRs were found to be unique to the human species at the specified interval of – 120 to + 1 TSS. For example, in the tetranucleotide compartment, CACC, GACA, CCGG, GATA, TCTG, GGCT, and TTTA STRs were detected in human only.

Enrichment of human-specific transcripts at the humanspecific STR compartment

Based on sequence comparison and the overall composition of the transcript and encoded protein, 25



and 26% of the transcripts in the tri and tetra human-specific compartments were found to be human-specific (mid-p < 0.0002 and mid-p < 0.002), respectively). The %identity score of multiple sequence alignment for the human-specific transcripts was 0 (exemplified in Fig. 3), and pair-wise analysis (exemplified in Fig. 4) resulted in %identity scores

ranging from 37 to 48%. In the trinucleotide category, 14 genes, *MPRIP*, *NPAS1*, *PAQR9*, *PRSS1*, *R3HDM2*, *TMEM99*, *ZSCAN30*, *C22orf24*, *ECSCR*, *AMY1C*, *DDX58*, *C1orf204*, *RGPD6*, and *LCE2B*, contained human-specific transcripts. In the tetranucleotide category, five genes, *DRD2*, *DUX4*, *TEAD4*, *ARL17B*, and *ARHGEF35*, contained human-specific transcripts.



 Table 1 Genome-scale human-specific core promoter trinucleotide STRs

ACC/6 ENTGU0003/0854 201 -48 (CA)3 AMP/IC ENTGU0003/08703703 701 -79 (ATT)3 AMP/ID ENTGU00031103.1/ 202 -11 (6CG)3 BMR2 (FAMSB) ENTGU00031539.4 201 -26 (6CG)3 CTC178 (FLRMSB) ENTGU0009771.12 701 -79 (ACG)3 -57 (CCC)5 CDH4 ENTGU0009372.12 701 -98 (GG)3 -57 (CCC)5 CDH4 ENTGU0009372.12 701 -98 (GG)3 -57 (CCC)5 CDH4 ENTGU0009372.12 701 -98 (GG)3 -57 (CCC)5 CDH4 ENTGU0009363.32 701 -98 (GG)3 -57 (CCC)3 CIMPAP ENTGU00093612.5 201 -11 (AG)3 -57 (CCC)3 CIMPAP ENTGU00093612.5 201 -13 (CCT)3 -58 (GG)3 CIMPAP ENTGU00093612.5 201 -33 (CCT)3 -57 (CCC)3 CIMPAP ENTGU00093612.5 201 -33 (CCT)3 -57 (CCC)3 CIMPAP ENTGU0009362.5 201 -35 (GC)3 CIMPAP <	Human gene symbol	Ensembl transcript ID	Variant no.	STR formula	
AMPC DNT00003/00/93 201 -79 (ATTB) AR81 DNT0000310517 202 -11 (SCG3 BMS2 (AMB3) DNT00003022.12 203 -35 (CT)3 CCC074 (C180/JS) DNT00003022.12 201 -39 (CT)3 CCMAND DNT00003022.12 201 -39 (CT)3 CMANND DNT00003022.12 201 -39 (CT)3 CMANND DNT0000302.1742.33 201 -38 (CA)3 CMANN2 DNT0000302.1742.33 201 -88 (CA)3 CMANN2 DNT0000302.1742.33 201 -88 (CA)3 CMANA DNT0000302.1742.33 201 -81 (CA)3 CMANA DNT0000302.1742.33 201 -81 (CA)3 CMANA DNT0000302.1742.35 201 -81 (CA)3 CMANA DNT0000305.075 202 -88 (CA)3 MNT01	ADCY6	ENST0000307885.4	201	- 48 (CCA)3	
A/BB1 DNTC000031051.7/ 202 -11 (GCG3 BNN2 (FAMS3) DNTC000044002. 233 -26 (GCG5 CCDC178 (C186r34) DNTC0000540332.7 201 -79 (AGC3 -57 (GGG5 CDDC178 (C186r34) DNTC000054033.7 201 -31 (CT)3 -57 (GGG5 CDM4 PNTC000054037.7 201 -88 (GCA3 -57 (GGG5 C1MAR2 PNTC00005402.7 201 -88 (GCA3 -57 (GGG5 C1MAR2 PNTC00005402.1 201 -88 (GCA3 -58 (GCA3 C1MAR2 DNTC00005462.41.1 202 -70 (AGA3 -51 (GGA3 C2MR1 DNTC00005462.41.1 202 -70 (AGA3 -51 (GGA3 CAMR1 DNTC00005462.41.1 202 -70 (AGA3 -51 (GGA3 CAMR1 DNTC00005462.41.1 202 -70 (GGA3 -51 (GGA3 CAMR1 DNTC000053.698.3 201 -67 (GCC3 -53 (GCA3 CAMR2 DNTC000053.698.3 201 -70 (GCA3 -55 (GCG3 CAMR2 DNTC000054.678.1 202 -97 (GC	AMY1C	ENST00000370079.3	201	— 79 (ATT)3	
BMR2 (FAMG8) ENST0000361339.4 201 -26 (GCG) BVS DNST000030102444082 203 -35 (CG) CCICLT/R (CBMF4) ENST00003014323.2 201 -31 (CG) CMIAN ENST00003014323.2 201 -39 (CG) CMIAN ENST0000311743.3 201 -88 (GA) CVFAH1 ENST0000311743.3 201 -88 (GA) CVFAH1 ENST000031683.8 201 -86 (CA) CVFAH1 ENST0000317473.3 201 -81 (GA) CVFAH1 202 -70 (AG) -70 (AG) CVFAH1 202 -70 (AG) -70 (AG) CVAR3 ENST000037968.5 201 -81 (CG) CVAR4 ENST00003797.5 202 -86 (GA) CVAR3 ENST0000397.5 203 -51 (GG) SR050 ENST0000397.5 202 -96 (GA) NR701/7078.5 203 -51 (GG) NR701/7078.5 201 -57 (CG) SR05 201 -97 (CG) NR701/707	APBB1	ENST00000311051.7	202	- 11 (GCC)3	
BVES ENST00000444082 203 -35 (CCT)3 CCDC178 (C180r/59) ENST0000030027.12 201 -79 (ACG)3 -57 (CCG)5 CMAIN ENST00000563341.1 202 -39 (CCT)3 - CMIMAP2 ENST0000051727.7 201 -88 (CG)3 - CMIMAP2 ENST0000031638.8 201 -88 (CG)3 - C147441 ENST0000036102.5 201 -11 (AG)3 - C14724 ENST0000036102.5 201 -11 (AG)3 - C14724 ENST0000036102.5 201 -33 (CCT)3 - C260124 ENST00000379668.5 201 -33 (CCT)3 - C260124 ENST000003697.5 201 -55 (GCG)3 - C5508 ENST000003797.5 202 -88 (CG)3 - INP448 ENST000003797.5 203 -81 (GG)3 - INP448 ENST0000037980.5 201 -58 (GA)3 - INP448 ENST0000037980.5 201 -96 (GA)3 - <td< td=""><td>BRINP2 (FAM5B)</td><td>ENST00000361539.4</td><td>201</td><td>– 26 (GGC)5</td><td></td></td<>	BRINP2 (FAM5B)	ENST00000361539.4	201	– 26 (GGC)5	
CCCC178 (C18ar/34) ENST000030227.12 201 -79 (AGQ3 -57 (CGQS CDH ENST000054323.22 201 -31 (CCT)3 -31 (CCT)3 CMPNI ENST0000563241.1 202 -99 (TGC)3 -57 (CGQS CVF4411 ENST000031658.8 201 -88 (CGA)3 -57 (CGC)3 CVF4411 ENST000036125 201 -88 (CGA)3 -57 (CGC)3 CVF4411 ENST00003625 201 -31 (CTA)3 -57 (CGC)3 CVF441 ENST0000379685 201 -31 (CTA)3 -57 (CGC)3 CVF441 ENST0000379685 203 -55 (GGC)3 -57 (CTA)3 GRN2D ENST0000379565 203 -61 (AGG)3 -57 (CTA)3 GRN2D ENST000005062285 203 -61 (AGG)3 -57 (CTA)3 KRA5 ENST000005062285 201 -59 (GAA)3 -57 (CTA)3 KRA5 ENST000003697865 201 -59 (GAA)3 -57 (CTA)3 KRA5 ENST000002698780.3 201 -59 (GAA)3 -57 (CTA)3 LCT2R ENST000002	BVES	ENST00000446408.2	203	- 35 (CCT)3	
CDH4 ENST00000543233.2 201 -31 (CCT)3 CMRN1 ENST0000056127.7 201 -39 (CT_3) CMMA2 ENST0000031072.3.3 201 -88 (CCA)3 CMMA11 ENST00000310938.8 201 -86 (CCT)3 C10r204 ENST00000310938.8 201 -86 (CCA)3 C20r204 ENST0000036126.2 201 -86 (CCA)3 C2ArA7 ENST000003626.2 201 -83 (CCA)3 C2ArA7 ENST000003626.2 201 -33 (CCT)3 C3A8 ENST000003626.2 201 -67 (CCA)3 C3650 ENST00000362.2 201 -67 (CCA)3 GIN2D ENST0000031193.6 202 -88 (CGG)3 GIN2D ENST0000031193.6 203 -67 (CCA)3 KM1211 ENST0000031193.6 202 -70 (CGA)9 KMA121 ENST0000031193.6 202 -89 (GA)3 LCF2B ENST0000031193.6 202 -70 (CGA)9 LCF2B ENST000003193.6 201 -108 (CCA)9 LCF2B ENST0000037885.5 201 -92 (CCA)3 LCF2B <td< td=""><td>CCDC178 (C18orf34)</td><td>ENST00000300227.12</td><td>201</td><td>– 79 (AGC)3</td><td>– 57 (CGC)5</td></td<>	CCDC178 (C18orf34)	ENST00000300227.12	201	– 79 (AGC)3	– 57 (CGC)5
CUPNN1 ENT00000363341.1 202 - 99 (CCT)3 CMTMAP2 ENT0000036727.2 201 - 98 (GG)3 CST4 ENT000003688 201 - 86 (CG)3 CMAD42 ENT0000036880.25 201 - 81 (CA)3 Clor204 ENT00003680.25 201 - 81 (CA)3 Clor204 ENT0000369865 201 - 81 (CA)3 Clor204 ENT000039865 201 - 61 (CC)3 Clor204 ENT000039865 201 - 67 (CC)3 ECSCR ENT000039875 202 - 67 (CC)3 ENT000039975 203 - 57 (CC)3 - RMP45 ENT000039975 202 - 80 (GA)3 KR7012 ENT0000695275 202 - 70 (CC)3 KR711 ENT0000695275 202 - 70 (CC)3 LC78 ENT00000695203 201 </td <td>CDH4</td> <td>ENST00000543233.2</td> <td>201</td> <td>- 31 (CCT)3</td> <td></td>	CDH4	ENST00000543233.2	201	- 31 (CCT)3	
CNTMAP2 ENST0000316727.7 201 - 98 (FGC)3 CST4 ENST00000316223.0 201 - 88 (GGA)3 CVP4A11 ENST000036802.5 201 - 61 (CA)3 C220f24 ENST000036802.5 201 - 70 (GGA)3 C220f24 ENST000036802.5 201 - 70 (GGA)3 C3AR1 ENST0000379683.5 201 - 67 (GCC)3 CSGR ENST0000379693.3 - 121 (CCA)3 - 121 (CCA)3 GNN2D ENST0000397563.5 203 - 55 (GGC)3 - 141 (FA)3 GNN2D ENST000039755.3 202 - 88 (GGC)3 - 141 (FA)3 GNN2D ENST0000397863.5 203 - 67 (GCC)3 - 141 (FA)3 KMA1211 ENST0000397803.5 202 - 88 (GGC)3 - 141 (FA)3 KMA1211 ENST0000397803.5 201 - 58 (GGA)3 - 162 (GCC)3 KMA1211 ENST0000397803.5 201 - 79 (CC1)3 - 108 (GCC)3 KMA1211 ENST0000397803.5 201 - 25 (GCC)7 - 108 (GCC)3 KMA1211 ENST0000	CIAPIN1	ENST00000563341.1	202	- 39 (CCT)3	
C374 ENST000032174233 201 -BS (GGA)3 C/PAA11 ENST00003306388 201 -BS (CT)5 ClarD204 ENST00003561025 201 -BS (CGA)3 C/AR1 ENST00003566241.1 202 -70 (AGA)3 DXSR ENST00003798685 201 -BS (CGA)3 DXSR ENST00003263633 201 -67 (GCG)3 GMR/D ENST00003263637 202 -88 (CGC)3 GMR/D ENST00000326275 202 -88 (CGC)3 GMR/D ENST00000309275 203 -81 (AGA)3 KAMA1211 ENST0000030935085 202 -88 (CGC)3 KAMA1211 ENST00003093085 202 -80 (GAA)3 LCTEU ENST00003095085 202 -97 (CT)3 -108 (GCO)3 KTN1 ENST00003506167 202 -97 (CT)3 -108 (GCO)3 LCCR (L100rh2) ENST00003506167 201 -35 (GCA)7 -108 (GCO)3 MRPIP ENST00003506167 201 -35 (GCA)7 -108 (GCO)3 MRPIP ENST00003506167 201 -35 (GCA)7 -108 (GCA)3 MRPIP<	CNTNAP2	ENST00000361727.7	201	- 98 (TGC)3	
CYP4A11 ENST00003106388 201 -86 (CCT)3 Chor204 ENST00003681025 201 -11 (AAG)3 C20r2A NM_013722.1 -70 (AGA)3 C3AR1 ENST00003796845 201 -33 (CCT)3 ECSCR ENST00003796835 201 -67 (CCC)3 ECSCR ENST000039755 203 -55 (GCG)3 MPPAB ENST000039755 202 -88 (CCG)3 KIRDD12 ENST000039757 204 -57 (CCT)3 KIRD12 ENST000039755 203 -81 (AAG)3 KIRD12 ENST0000398085 202 -89 (GCG)3 KIRD12 ENST00003998085 202 -90 (GCG)3 KIRD12 ENST00003998085 201 -79 (CCT)3 LCRB1 ENST00003667803 201 -79 (CCT)3 LCRB1 ENST00003667803 201 -91 (GCG)3 LCRB1 ENST0000375521.3 201 -82 (GC)3 LCRB1 ENST0000499656 201 -13 (GA)3 MAR9 ENST0000499621.3 203 -29 (GC)3 MPRP ENST0000499656 201 <t< td=""><td>CST4</td><td>ENST00000217423.3</td><td>201</td><td>– 83 (GGA)3</td><td></td></t<>	CST4	ENST00000217423.3	201	– 83 (GGA)3	
Clord204 ENST00003681025 201 -11 (AAG)3 C22nf24 MQ.153722.1 -98 (GCA)3 C3AR1 ENST00003796865 201 -31 (CT)3 EKSCR MM_01077693.1 -121 (CCA)3 - GRM2D ENST00003941756 203 -55 (GGA)3 - GRM4 ENST000039275 202 -88 (GGA)3 - KRTBD12 ENST000030476967 203 -57 (CT)3 - KRTBD12 ENST000047697 202 -89 (GAA)3 - KRTBD12 ENST0000407697 202 -89 (GAA)3 - KRTM1 ENST0000407697 202 -89 (GAA)3 - KRTM1 ENST0000407697 202 -97 (CT)3 - KRA41211 ENST0000407697 202 -97 (CT)3 - LC7BL1 ENST0000492826 201 -58 (GAA)3 - LC2B ENST0000492626 201 -138 (GAA)3 - LC2B ENST000049865 201 -138 (GAA)3 -	CYP4A11	ENST00000310638.8	201	- 86 (CCT)3	
C22x724 NM_015372.1 -98 (GCA)3 C3AR1 ENST0000546211.1 202 -70 (AGA)3 DX58 ENST0000379868.5 201 -33 (CTA)3 EGCR NM_01077693.1 -121 (CCA)3 GMN2D ENST0000253269.3 201 -67 (GCC)3 GSDM8 ENST000025327.5 202 -88 (GGC)3 INPP4B ENST000054228.5 203 -51 (AGC)3 KAA1211 ENST00005428.5 203 -58 (GAA)3 KAA1211 ENST000039308.5 202 -88 (GAA)3 KTN1 ENST000039508.5 202 -80 (GAA)3 KTN1 ENST0000366780.3 201 -55 (GCC)7 LC28 ENST000046616.2 201 -58 (GAA)3 LC28 ENST000046616.2 201 -97 (CT)3 -108 (GC)3 MARNTD3 (Gorl/2) ENST000046166.2 201 -135 (GAA)9 -108 (GC)3 MARNTD3 (Gorl/3) ENST0000439365.6 201 -108 (GC)3 -108 (GC)3 MARNTD3 (Gorl/3) ENST0000439365.6 201 -103 (GC)3 -108 (GC)3 MARP ENST0000439365.6 201 <td>C1orf204</td> <td>ENST00000368102.5</td> <td>201</td> <td>- 11 (AAG)3</td> <td></td>	C1orf204	ENST00000368102.5	201	- 11 (AAG)3	
C3AR1 202 -70 (AGA)3 DX68 ENST000037968.5 201 -33 (CCT)3 ECSCR NM_001077693.31 -67 (GCC)3 GRIN2D ENST0000391175.6 203 -55 (GC)3 SGDM8 ENST000039175.6 202 -88 (GC)3 INPP48 ENST00000407609.7 204 -57 (CCT)3 KRA1211 ENST0000391936.7 202 -80 (GA)3 KRA1211 ENST000039508.5 202 -70 (GCG)9 LC7BL1 ENST000039508.5 201 -55 (GA)3 LC7BL1 ENST000039508.5 201 -70 (CCG)9 LC7BL1 ENST000039508.5 201 -70 (GCG)9 LC7BL1 ENST0000374885.5 201 -79 (CCT)3 -108 (GC)3 MFRIP ENST0000374885.5 201 -25 (GC,7 -108 (GC)3 MFRIP ENST0000374885.5 201 <	C22orf24	NM_015372.2.1		- 98 (GCA)3	
DDX88ENST00000379868.5201-33 (CT)3ECSCRNM_001077693.3.1-121 (CCA)3GRN2DENST00000394175.6201-67 (GC)3GSDM8ENST00000394175.6202-88 (GCG)3KM7D12ENST000003927.5203-81 (AAG)3KMA1211ENST0000037609.7204-57 (CT)3KMA1211ENST000003927.5203-81 (AAG)3KMA1211ENST00000392.85202-89 (GA)3KMA1211ENST00000392.85202-99 (CG)3LATEL1ENST0000042692.86201-70 (GC)9LATEL1ENST0000046780.3201-97 (CT)3LCGR (C100r/12)ENST0000366186.2209-88 (GA)11MSANTD3 (C90rB2)ENST000039351.5201-15 (GA)9MRPIPENST00003904.81201-100 (GAT)3MRAT12 (C20rH2)ENST000039304.81201-100 (GAT)3MRAT12 (S0rH2)ENST00004905.1205-96 (GAT)3MRPD6ENST000049305.1205-96 (GAT)3MRPD6ENST000049305.1205-91 (GC)3MRPD6ENST000044873.21205-91 (GC)3SCN2BENST000029333.7201-123 (GCT)3SCN2BENST000039305.10201-123 (GCT)3SCN2BENST000029333.7201-92 (GC)3SCN2BENST000029333.7201-92 (GC)3SCN2BENST000029333.7201-92 (GC)3SCN2BENST000029333.7201-92 (GC)3SCN2BENST000029333.7 <t< td=""><td>C3AR1</td><td>ENST00000546241.1</td><td>202</td><td>– 70 (AGA)3</td><td></td></t<>	C3AR1	ENST00000546241.1	202	– 70 (AGA)3	
ECSCR NM_001077693.3.1 -121 (CCA)3 GRINZD ENST0000263269.3 201 -67 (GCC)3 GSDM8 ENST0000394175.6 203 -55 (GSC)3 INPF4B ENST000050927.5 202 -88 (GAC)3 K8TRD12 ENST00000407609.7 204 -57 (CCT)3 K4A1211 ENST0000311936.7 202 -89 (GAA)3 K7N1 ENST000035308.5 202 -70 (GCG)9 LCTBL1 ENST0000356780.3 201 -79 (CCT)3 LCZB ENST0000354885.5 201 -79 (CCT)3 LCR0 (C100r/12) ENST0000374885.5 201 -155 (GAC)7 MRPIP ENST0000374885.5 201 -155 (GAC)7 VASATT ENST0000374885.5 201 -155 (GAC)3 VARX1 ENST0000373521.3 201 -108 (GCC)3 VARX1 ENST0000493056.1 203 -29 (TCG)3 PAR9 ENST000049305.1 205 -90 (GCG)3 PAR9 ENST000049305.1 205 -90 (GCG)3 PRS1 ENST000049375.1 205 -90 (GCG)3 PRS1 ENST0000493	DDX58	ENST00000379868.5	201	- 33 (CCT)3	
GRIN2DENST00002632693201- 67 (GC)3GSDM8ENST0000394175.6203- 55 (GC)3INPP48ENST0000503927.5202- 88 (GC)3KBTBD12ENST00000504228.5203- 81 (AAG)3KAA1211ENST000030504228.5202- 70 (GC)9LACTEL1ENST000039505.5202- 70 (GC)9LACTEL1ENST0000368780.3201- 79 (CC)3LC2R (C10or12)ENST000036616.7202- 97 (CC)3LCOR (C10or12)ENST000036887.63201- 25 (GC)7MARJENST0000374885.5201- 25 (GC)7MASANTD3 (C9or/S0)ENST000043966.62201- 100 (GAT)3CMA11ENST0000374885.5201- 100 (GAT)3PABPC1L2BENST000043965.65201- 100 (GAT)3PABPC1L2BENST000049470.1203- 29 (TGC)3PARS1ENST000049490.1203- 29 (TGC)3PARS1ENST00004920.51.1205- 96 (GAT)3RGPD6ENST00004920.51.1205- 70 (GCG)5SRT12ENST00004920.51.7201- 123 (GCT)5SRT215ENST00004933.7201- 29 (TGC)3SGRA8ENST000004933.7201- 29 (GC)3SGRA9ENST0000045052.87201- 29 (GC)3SGRA9ENST000003502.87201- 29 (GC)3SGRA9ENST000003502.87201- 29 (GC)3SGRA9ENST000003502.87201- 29 (TG)3SGRA9ENST000003502.87201- 29 (GC)3<	ECSCR	NM_001077693.3.1		- 121 (CCA)3	
GSDMBENST0000394175.6203- 55 (GG)3INPP48ENST0000503927.5202- 88 (GG)3KBTBD12ENST000047609.7204- 57 (CT)3KIAA1211ENST000054228.5203- 81 (AAG)3KRASENST000039303.65202- 70 (GC)9KITN1ENST0000342692.66201- 58 (GA)3LCTBL1ENST000036878.03201- 79 (CT)3LCCB (C100/12)ENST000036878.03201- 79 (CT)3LCRA (C100/12)ENST00003697.03201- 108 (GC)3MPRIPENST0000374885.5201- 25 (GC)7NAS1ENST0000373521.3201- 100 (GAT)3PAGPENST0000373521.3201- 60 (GC)3PARPENST000044992.61203- 29 (TG)3PARPENST00004559.51201- 102 (GC)3RFI215ENST00004559.51201- 29 (TG)3RFI215ENST00004559.51203- 70 (GC)3SERPINB9ENST00004559.51203- 29 (TG)3SCN38ENST00004559.51201- 123 (GCT)5SERPINB9ENST00004559.51201- 29 (GC)3SERPINB9ENST00004559.51201- 29 (GC)3SERPINB9ENST0000350.62.7201- 29 (GC)3SERPINB9ENST0000350.56.7201- 29 (GC)3SERPINB9ENST0000350.56.7201- 29 (GC)3SERPINB9ENST0000350.56.7201- 29 (GC)3SERPINB9ENST0000350.56.7201- 29 (GC)3SERPINB9 <td>GRIN2D</td> <td>ENST00000263269.3</td> <td>201</td> <td>- 67 (GCC)3</td> <td></td>	GRIN2D	ENST00000263269.3	201	- 67 (GCC)3	
INPP48 ENST0000039327.5 202 - 88 (CGC)3 KBBD12 ENST0000407609.7 204 - 57 (CCT)3 KAA1211 ENST0000311936.7 202 - 89 (GAA)3 KRAS ENST000031936.7 202 - 70 (GCG)9 LACTBL1 ENST0000368780.3 201 - 58 (GAA)3 LCE28 ENST0000368780.3 201 - 79 (CCT)3 LCCR (C10orf12) ENST0000366780.3 201 - 79 (CCT)3 MPRIP ENST0000366780.3 201 - 25 (GCC)7 MSANTD3 (C9orf30) ENST0000374885.5 201 - 25 (GCC)7 NPAS1 ENST00003751.3 201 - 100 (GAT)3 PARPC1L28 ENST000049926.1 203 - 29 (GCG)3 PARPS1 ENST00004989.1 201 - 60 (GCC)3 PARPS1 ENST00004989.1 203 - 29 (GGG)3 RSS1 ENST00004989.1 203 - 29 (GGG)3 RSS1 ENST00004989.1 203 - 29 (GGG)3 RSS1 ENST000004989.1 201 - 123 (GCT)5	GSDMB	ENST00000394175.6	203	– 55 (GGC)3	
KBBD12 ENST000047609.7 204 -57 (CCT)3 KIAA1211 ENST0000504228.5 203 -81 (AAG)3 KRAS ENST0000311936.7 202 -80 (GAA)3 KTN1 ENST00003508.5 202 -70 (GCG)9 LACTBL1 ENST0000426928.6 201 -58 (GAA)3 LCEB ENST000036618.0 201 -79 (CCT)3 -108 (GCA)3 LCDR (C100rf12) ENST000036618.2 209 -88 (GCA)11 -108 (GCA)3 MFRIP ENST000037488.5 201 -25 (GCC)7 -108 (GCA)3 MSANTD3 (C90rJ30) ENST000037488.5 201 -155 (GAA)9 -100 (GAT)3 PARY ENST000037485.5 201 -100 (GAT)3 -100 (GAT)3 PARY ENST0000373521.3 201 -100 (GAT)3 -100 (GAT)3 PARPC1L28 ENST0000498470.1 203 -29 (TGC)3 -100 (GAT)3 RGPD6 ENST0000495595.1 205 -96 (GAT)3 -100 (GAT)3 -100 (GAT)3 SERPINB9 ENST0000215798.10 201 -123 (GCC)5 -100 (GAC)3 -100 (GAC)3 -100 (GAC)3 -100 (GAC)3 -100 (GAC)3	INPP4B	ENST00000503927.5	202	- 88 (CGC)3	
KMA1211 ENST000054228.5 203 -B1 (AAG)3 KRAS ENST0000311936.7 202 -B9 (GAA)3 KTN1 ENST0000395308.5 202 -70 (GCG)9 LACTBL1 ENST0000426928.6 201 -58 (GAA)3 LCE2B ENST0000368780.3 201 -79 (CCT)3 -108 (GCQ)3 LCOR (C10orf12) ENST0000466186.2 209 -88 (GCA)11 -108 (GCQ)3 MRIP ENST000046386.6 201 -125 (GCC)7 -108 (GCQ)3 MAX17D3 (C9orJ30) ENST000037488.5 201 -135 (GAA)9 -100 (GAT)3 ORAX1 ENST0000373521.3 201 -100 (GAT)3 -100 (GAT)3 PABPC1L2B ENST0000373521.3 201 -60 (GCC)3 -101 (GAT)3 PRSS1 ENST000049470.1 203 -29 (TGC)3 -101 (GAT)3 RGPD6 ENST0000495695.1 205 -70 (GCC)5 -101 (GAT)3 RGPD6 ENST0000215798.10 201 -123 (GCT)5 -101 (GAT)3 SCN3B ENST000029333.7 201 -29 (GGT)3 -101 (GAC)3 SERPINB9 ENST0000038628.7 201 -29	KBTBD12	ENST00000407609.7	204	- 57 (CCT)3	
KRAS ENST0000311936.7 202 -89 (GAA)3 KTN1 ENST0000395308.5 202 -70 (GCG)9 LACTBL1 ENST00000426928.6 201 -58 (GAA)3 LCE2B ENST00000368780.3 201 -79 (CCT)3 -108 (GCQ)3 LCC0R (C100r/12) ENST000036616.7 202 -97 (CCT)3 -108 (GCQ)3 MPRIP ENST000037488.5.5 201 -25 (GCQ)7 -108 (GCQ)3 MAX1 ENST000037488.5.5 201 -135 (GAA)9 -100 (GAT)3 ORAX1 ENST0000373521.3 201 -100 (GAT)3 -100 (GAT)3 PRS1 ENST0000498470.1 203 -29 (TGC)3 -100 (GAT)3 PRS1 ENST0000495695.1 205 -70 (GGC)5 -100 (GAT)3 RGPD6 ENST000048732.1 208 -27 (GCQ)3 -100 (GAT)3 SCN3B ENST000004873.21 208 -27 (GCQ)3 -100 (GAT)3 SCN3B ENST0000038698.4 201 -12 (GCA)3 -100 (GAC)3 SCRPINB9 ENST000038698.4 201 -29 (GCG)3 -100 (GCQ)3 SCRPINB9 ENST0000038698.4 201	KIAA1211	ENST00000504228.5	203	- 81 (AAG)3	
KN11 ENST0000395308.5 202 -70 (GCG)9 LACTBL1 ENST0000426928.6 201 -58 (GAA)3 LCE2B ENST0000368780.3 201 -79 (CT)3 -108 (GCC)3 LCDR (C10orf12) ENST0000366186.2 209 -88 (GCA)11 -100 (GCT)3 -108 (GCC)3 MPRIP ENST0000374885.5 201 -25 (GCC)7 -108 (GCA)3 VASATD3 (C9orf30) ENST0000373521.3 201 -100 (GAT)3 -100 (GAT)3 PABPC1L2B ENST0000373521.3 201 -60 (GCC)3 -100 (GAT)3 PAGP9 ENST0000498470.1 203 -29 (TGC)3 -100 (GAT)3 RGPD6 ENST0000498470.1 205 -70 (GGC)5 -100 (GAT)3 RF215 ENST00000498470.1 205 -70 (GGC)5 -100 (GAT)3 RAPD6 ENST0000049832.1 201 -123 (GCT)5 -100 (GAT)3 SCN38 ENST0000029933.7 201 -29 (GC)3 -100 (GAT)3 SCR19KB9 ENST00003608.68.4 201 -12 (GCA)3 -100 (GAT)3 SGLEC7 ENST00003608.69.4 201 -29 (GCG)3 -100 (GAT)3	KRAS	ENST00000311936.7	202	- 89 (GAA)3	
LACTBL1 ENST00004269286 201 -58 (GAA)3 LCE2B ENST0000368780.3 201 -79 (CCT)3 -108 (GCC)3 LCOR (C10orf12) ENST000046186.2 209 -88 (GA)11 -108 (GCC)3 MPRIP ENST0000374885.5 201 -25 (GCC)7 -108 (GCC)3 MSANTD3 (C9orf30) ENST0000374885.5 201 -135 (GAA)9 -100 (GAT)3 OR4X1 ENST0000373521.3 201 -60 (GCC)3 -100 (GAT)3 PABPC1L2B ENST0000498470.1 203 -29 (TGC)3 -100 (GAT)3 PAS51 ENST0000498470.1 203 -96 (GAT)3 -100 (GAT)3 RGPD6 ENST0000492062.1 205 -96 (GAT)3 -100 (GAT)3 RAPD16 ENST0000492062.1 205 -96 (GAT)3 -100 (GAT)3 RAPD6 ENST000048732.1 208 -27 (GCC)3 -100 (GAT)3 SCN38 ENST0000215798.10 201 -12 (GCA)3 -100 (GAT)3 SERPINB9 ENST0000330205.10 201 -80 (TTC)3 -100 (TGG)4 -90 (TGG)4 SMC11 (L21orf56) ENST0000330205.10 201 -37 (TGG)4 -9	KTN1	ENST00000395308.5	202	– 70 (GCG)9	
LC28 ENST0000368780.3 201 -79 (CT)3 -108 (GCC)3 LC0R (C100rf12) ENST0000356016.7 202 -97 (CT)3 -108 (GCC)3 MPRIP ENST00000374885.5 201 -25 (GCC)7 NPAS1 ENST00000374885.5 201 -100 (GAT)3 OR4X1 ENST0000373201.3 201 -60 (GCC)3 PABPC1L2B ENST0000498470.1 203 -29 (TGC)3 PARS1 ENST0000498470.1 205 -96 (GAT)3 RGPD6 ENST00001455695.1 201 -123 (GCT)5 RFL215 ENST0000215798.10 201 -123 (GCT)5 SK104002 ENST0000215798.10 201 -29 (GG)3 SCN3B ENST0000036698.4 201 -123 (GCT)5 SERPINB9 ENST0000035628.7 201 -29 (GG)3 SCR1A ENST0000035628.7 201 -29 (GG)3 SFATCHL (C21orf56) ENST000035052.10 202 -37 (TGG)4 -90 (TGG)4 SFATCHL (C21orf56) ENST0000330205.10 201 -12 (GCA)3 -90 (TGG)4 SFATCHL (C21orf56) ENST0000215948.8 201 -91 (GCC)3	LACTBL1	ENST00000426928.6	201	- 58 (GAA)3	
LCOR (C10orf12) ENST0000356016.7 202 -97 (CCT)3 -108 (GCC)3 MPRIP ENST0000466186.2 209 -88 (GCA)11 - MSANTD3 (C9orf30) ENST0000374885.5 201 -25 (GCC)7 - NPAS1 ENST000032048.1 201 -100 (GAT)3 - - OR4X1 ENST000032048.1 201 -60 (GCC)3 -	LCE2B	ENST00000368780.3	201	– 79 (CCT)3	
MRRIP ENST0000466186.2 209 -88 (GCA)11 MSANTD3 (C90rf30) ENST0000374885.5 201 -25 (GCQ)7 NPAS1 ENST0000439365.6 201 -135 (GAA)9 OR4X1 ENST0000320048.1 201 -60 (GCC)3 PABPC1L2B ENST0000373521.3 201 -60 (GCC)3 PAQR9 ENST0000498470.1 203 -29 (TGC)3 PRS51 ENST0000495695.1 205 -70 (GGC)5 RRPD6 ENST0000215798.10 201 -123 (GCT)5 RNF215 ENST0000248732.1 208 -27 (GCQ)3 SCN3B ENST0000248732.1 208 -27 (GCQ)3 SCN3B ENST0000299333.7 201 -123 (GCT)5 SCN3B ENST0000299333.7 201 -29 (GGT)3 SGLEC7 ENST000038058.4 201 -12 (GCA)3 SGLEC7 ENST0000330205.10 202 -37 (TGG)4 -90 (TGG)4 SVMF1 ENST0000219548.8 201 -91 (GCC)3 -90 (TGG)4 SUMF1 ENST0000219548.8 201 -102 (AGC)3 -90 (TGG)4 SUMF1 ENST000023058.4 <td>LCOR (C10orf12)</td> <td>ENST00000356016.7</td> <td>202</td> <td>- 97 (CCT)3</td> <td>- 108 (GCC)3</td>	LCOR (C10orf12)	ENST00000356016.7	202	- 97 (CCT)3	- 108 (GCC)3
MSANTD3 (C9orf30) ENST0000374885.5 201 -25 (GCC)7 NPAS1 ENST0000439365.6 201 -135 (GAA)9 OR4X1 ENST0000320048.1 201 -60 (GCC)3 PABPC1L2B ENST0000373521.3 201 -60 (GCC)3 PAQR9 ENST0000498470.1 203 -29 (TGC)3 PRS51 ENST000049206.1 205 -96 (GAT)3 RGPD6 ENST0000215798.10 201 -123 (GCT)5 RNF215 ENST0000215798.10 201 -123 (GCT)5 RSHDM2 ENST0000248732.1 208 -27 (GCC)3 SCN3B ENST000029933.7 201 -12 (GCA)3 SCN3B ENST0000386698.4 201 -12 (GCA)3 SGLEC7 ENST0000386698.4 201 -80 (TTC)3 SPATCIL (C21orf56) ENST0000330205.10 202 -37 (TGG)4 -90 (TGG)4 STUB1 ENST0000219548.8 201 -91 (GCC)3 -90 (TGG)4 STUB1 ENST0000219548.8 201 -102 (AGC)3 -90 (TGG)4 STUB1 ENST	MPRIP	ENST00000466186.2	209	- 88 (GCA)11	
NPAS1ENST0000439365.6201-135 (GAA)9OR4X1ENST000032048.1201-100 (GAT)3PABPC1L2BENST0000373521.3201-60 (GCC)3PAQR9ENST0000498470.1203-29 (TGC)3PRSS1ENST0000492062.1205-96 (GAT)3RGPD6ENST0000455695.1205-70 (GGC)5RNF215ENST0000215798.10201-123 (GCT)5RJHDM2ENST0000448732.1208-27 (GCC)3SCN3BENST00002933.7201-29 (GGT)3SERPINB9ENST000030698.4201-12 (GCA)3SGLEC7ENST0000305628.7201-80 (TTC)3STUB1ENST0000215948.8201-91 (GCQ)3SUMF1ENST000022902.9201-102 (AGC)3TX12ENST000028358.4201-102 (GCG)3	MSANTD3 (C9orf30)	ENST00000374885.5	201	- 25 (GCC)7	
OR4X1 ENST0000320048.1 201 -100 (GAT)3 PABPC1L2B ENST0000373521.3 201 -60 (GCC)3 PAQR9 ENST0000498470.1 203 -29 (TGC)3 PRSS1 ENST0000492062.1 205 -96 (GAT)3 RGPD6 ENST000045695.1 205 -70 (GGC)5 RNF215 ENST0000215798.10 201 -123 (GCT)5 R3HDM2 ENST0000248732.1 208 -27 (GCC)3 SCN3B ENST000029933.7 201 -29 (GGT)3 SERPINB9 ENST0000380698.4 201 -12 (GCA)3 SGLEC7 ENST00003305205.10 202 -37 (TGG)4 -90 (TGG)4 SPATC1L (C21orf56) ENST0000330205.10 202 -37 (TGG)4 -90 (TGG)4 STUB1 ENST000021594.8 201 -91 (GC)3 -90 (TGG)4 SUMF1 ENST000027902.9 201 -102 (AGC)3 -90 (TGG)4 TEX12 ENST0000280358.4 201 -102 (AGC)3 -90 (TGG)4	NPAS1	ENST00000439365.6	201	– 135 (GAA)9	
PABPC1L2BENST0000373521.3201-60 (GCC)3PAQR9ENST0000498470.1203-29 (TGC)3PRSS1ENST0000492062.1205-96 (GAT)3RGPD6ENST0000455695.1205-70 (GGC)5RNF215ENST0000215798.10201-123 (GCT)5R3HDM2ENST0000248732.1208-27 (GCC)3SCN3BENST0000380698.4201-29 (GGT)3SERPINB9ENST0000380698.4201-12 (GCA)3SIGLEC7ENST000035628.7201-80 (TTC)3STUB1ENST0000215748.8201-91 (GCC)3SUMF1ENST000027902.9201-102 (AGC)3TEX12ENST000028358.4201-105 (TGG)3	OR4X1	ENST00000320048.1	201	- 100 (GAT)3	
PAQR9 ENST0000498470.1 203 -29 (TGC)3 PRSS1 ENST0000492062.1 205 -96 (GAT)3 RGPD6 ENST0000455695.1 205 -70 (GC)5 RNF215 ENST0000215798.10 201 -123 (GCT)5 R3HDM2 ENST0000248732.1 208 -27 (GCC)3 SCN3B ENST0000299333.7 201 -29 (GGT)3 SERPINB9 ENST0000380698.4 201 -12 (GCA)3 SGLEC7 ENST0000380698.7 201 -80 (TTC)3 SPATC1L (C21orf56) ENST0000330205.10 202 -37 (TGG)4 -90 (TGG)4 STUB1 ENST0000219548.8 201 -91 (GCC)3 -90 (TGG)4 STUB1 ENST0000272902.9 201 -102 (AGC)3 -91 (TGC)4 TEX12 ENST0000280358.4 201 -105 (TGG)3 -90 (TGG)4	PABPC1L2B	ENST00000373521.3	201	- 60 (GCC)3	
PRSS1 ENST0000492062.1 205 -96 (GAT)3 RGPD6 ENST0000455695.1 205 -70 (GGC)5 RNF215 ENST0000215798.10 201 -123 (GCT)5 R3HDM2 ENST0000248732.1 208 -27 (GCC)3 SCN3B ENST0000299333.7 201 -29 (GGT)3 SERPINB9 ENST0000386698.4 201 -12 (GCA)3 SGLEC7 ENST0000330205.10 202 -37 (TGG)4 -90 (TGG)4 STUB1 ENST0000219548.8 201 -91 (GCC)3 -90 (TGG)4 STUB1 ENST0000272902.9 201 -102 (AGC)3 -91 (TGG)4 TEX12 ENST000028058.4 201 -102 (AGC)3 -90 (TGG)4	PAQR9	ENST00000498470.1	203	– 29 (TGC)3	
RGPD6 ENST0000455695.1 205 -70 (GGC)5 RNF215 ENST0000215798.10 201 -123 (GCT)5 R3HDM2 ENST0000448732.1 208 -27 (GCC)3 SCN3B ENST000029933.7 201 -29 (GGT)3 SERPINB9 ENST0000380698.4 201 -12 (GCA)3 SIGLEC7 ENST000035628.7 201 -80 (TTC)3 SPATC1L (C210rf56) ENST0000330205.10 202 -37 (TGG)4 -90 (TGG)4 STUB1 ENST0000219548.8 201 -91 (GCC)3 -90 (TGG)4 STUB1 ENST000022902.9 201 -102 (AGC)3 -122 (AGC)3 TEX12 ENST0000280358.4 201 -102 (AGC)3 -90 (TGG)4	PRSS1	ENST00000492062.1	205	- 96 (GAT)3	
RNF215 ENST0000215798.10 201 -123 (GCT)5 R3HDM2 ENST0000448732.1 208 -27 (GCC)3 SCN3B ENST0000299333.7 201 -29 (GGT)3 SERPINB9 ENST0000380698.4 201 -12 (GCA)3 SIGLEC7 ENST000035628.7 201 -80 (TTC)3 SPATC1L (C210r56) ENST0000330205.10 202 -37 (TGG)4 -90 (TGG)4 STUB1 ENST0000219548.8 201 -91 (GCC)3 -91 (GCC)3 TEX12 ENST0000280358.4 201 -102 (AGC)3 -91 (GCC)3	RGPD6	ENST00000455695.1	205	– 70 (GGC)5	
R3HDM2 ENST0000448732.1 208 -27 (GCC)3 SCN3B ENST0000299333.7 201 -29 (GGT)3 SERPINB9 ENST0000380698.4 201 -12 (GCA)3 SIGLEC7 ENST000035628.7 201 -80 (TTC)3 SPATC1L (C21orf56) ENST0000330205.10 202 -37 (TGG)4 -90 (TGG)4 STUB1 ENST0000219548.8 201 -91 (GCC)3 -90 (TGG)4 STUB1 ENST0000272902.9 201 -102 (AGC)3 -122 (AGC)3 TEX12 ENST000028358.4 201 -105 (TGG)3 -90 (TGG)4	RNF215	ENST00000215798.10	201	- 123 (GCT)5	
SCN3B ENST0000299333.7 201 - 29 (GGT)3 SERPINB9 ENST0000380698.4 201 - 12 (GCA)3 SIGLEC7 ENST0000305628.7 201 - 80 (TTC)3 SPATC1L (C210rf56) ENST0000330205.10 202 - 37 (TGG)4 -90 (TGG)4 STUB1 ENST0000219548.8 201 - 91 (GCC)3 - 91 (GCC)3 TEX12 ENST0000280358.4 201 - 102 (AGC)3 - 91 (GCC)3	R3HDM2	ENST00000448732.1	208	- 27 (GCC)3	
SERPINB9 ENST0000380698.4 201 -12 (GCA)3 SIGLEC7 ENST0000305628.7 201 -80 (TTC)3 SPATC1L (C21orf56) ENST0000330205.10 202 -37 (TGG)4 -90 (TGG)4 STUB1 ENST0000219548.8 201 -91 (GCC)3 - SUMF1 ENST0000272902.9 201 -102 (AGC)3 - TEX12 ENST000028358.4 201 -105 (TGG)3 -	SCN3B	ENST00000299333.7	201	– 29 (GGT)3	
SIGLEC7 ENST0000305628.7 201 -80 (TTC)3 SPATC1L (C21orf56) ENST0000330205.10 202 -37 (TGG)4 -90 (TGG)4 STUB1 ENST0000219548.8 201 -91 (GCC)3 -91 (GCC)3 SUMF1 ENST0000272902.9 201 -102 (AGC)3 -105 (TGG)3 TEX12 ENST0000280358.4 201 -105 (TGG)3	SERPINB9	ENST00000380698.4	201	- 12 (GCA)3	
SPATC1L (C21orf56) ENST0000330205.10 202 - 37 (TGG)4 -90 (TGG)4 STUB1 ENST0000219548.8 201 - 91 (GCC)3 - SUMF1 ENST0000272902.9 201 - 102 (AGC)3 - TEX12 ENST0000280358.4 201 - 105 (TGG)3	SIGLEC7	ENST00000305628.7	201	- 80 (TTC)3	
STUB1 ENST0000219548.8 201 -91 (GCC)3 SUMF1 ENST0000272902.9 201 -102 (AGC)3 TEX12 ENST0000280358.4 201 -105 (TGG)3	SPATC1L (C21orf56)	ENST00000330205.10	202	– 37 (TGG)4	90 (TGG)4
SUMF1 ENST0000272902.9 201 - 102 (AGC)3 TEX12 ENST0000280358.4 201 - 105 (TGG)3	STUB1	ENST00000219548.8	201	- 91 (GCC)3	
<i>TEX12</i> ENST0000280358.4 201 – 105 (TGG)3	SUMF1	ENST00000272902.9	201	- 102 (AGC)3	
	TEX12	ENST0000280358.4	201	– 105 (TGG)3	

Human gene symbol	Ensembl transcript ID	Variant no.	STR formula	
TMEM99	ENST00000301665.7	201	- 32 (CCG)3	- 47 (CCG)3
			– 59 (CCG)3	- 83 (CCG)3
			- 110 (CCG)3	– 125 (CCG)3
			- 48 (GCC)3	- 60 (GCC)3
			- 84 (GCC)3	- 126 (GCC)3
TNNC2	ENST00000372557.1	202	– 53 (GCC)3	
ТРТЕ	ENST00000427445.6	201	- 110 (GCG)3	
TRBJ2-7	ENST00000390419.1	201	– 120 (GGC)3	
TRGV5	NC_000007.14:TRGV5:u_t_1.1		- 48 (CTC)3	
TRIM39	ENST00000376656.8	201	– 58 (CCT)4	
UAP1	ENST00000367926.8	204	– 9 (CGT)3	
VNN2	ENST00000326499.10	201	- 31 (GAA)10	
WRN	NM_000553.4.1		- 67 (GCC)3 - 69 (CCG)4	- 92 (GCC)3
WRNIP1	ENST00000618555.4	205	- 67 (CCG)3	
ZDHHC21	XM_006716760.1.1		- 32 (AGG)3	
ZSCAN30	ENST0000639929.1	212	- 57 (GAA)3	

 Table 1 Genome-scale human-specific core promoter trinucleotide STRs (Continued)

The numbers before the brackets represent the start site of the STR in respect of the corresponding transcription start site. "Variant no" corresponds to the Ensembl isoform number

Table 2 Genome-scale human-specific core promoter tetranucleotide STRs

Human gene symbol	Ensembl transcript ID	Variant no.	STR formula
ARHGAP5	ENST00000345122.7	202	– 110 (GGGA)4
ARHGEF35	ENST00000378115.2	201	- 22 (TCCC)3
ARL17B	ENST00000622877.4	201	- 99 (CTCC)3
ATP7A	ENST00000343533.9	201	– 27 (GAGG)3
DRD2	ENST00000542616.1	207	- 54 (GCCC)3
DUX4	ENST00000565211.1	203	- 144 (GGCT)6
FAM83G	ENST00000388995.10	202	- 80 (TCTG)3
GTF2IRD2B	ENST0000614064.4	206	- 17 (GAAA)3
JCAD (KIAA1462)	ENST00000375377.1	201	- 13 (CCGG)3
MCTP2	ENST00000451018.7	203	– 102 (GTTT)3
METTL21C (C13orf39)	ENST00000267273.6	201	- 69 (CAGT)3
OR10G6	ENST00000307002.3	201	– 123 (GATA)13
PHYHD1	ENST00000308941.9	201	- 107 (TTTA)3
SAMD1	ENST00000269724.5	201	- 75 (CCGC)3
TEAD4	ENST00000540314.1	206	- 51 (CTCC)3
TRAJ49	ENST00000390488.1	201	- 124 (GCCT)7
TRDJ2	ENST00000390475.1	201	- 86 (CCAC)3
TRAV38-1	ENST00000390464.2	201	- 109 (CACC)3
TRAV7	ENST00000390429.3	201	– 111 (GACA)3

The numbers before the brackets represent the start site of the STR in respect of the corresponding transcription start site. "Variant no" corresponds to the Ensembl isoform number

Page	6	of	8
------	---	----	---

а		b	
NPAS1		DRD2	
Prairie Dog Opossum Zebrafish Human Gorilla Mouse Macaque Chimpanzee Horse		Mouse Annale Armadillo Zebrafish Chicken Human Dog Bushbaby Gibbon Rabbit	-ACG6CT6CCGGAGGGGCG6CCGTGCGT6GATGGGGCGGGAGCT6 GCCG6CTCTTG6CGCTTCATGCCACTAACTCCTCCACCT6CGTGAGCCTC
DDX58		TEAD4	
Gorilla Mouse Human Chimpanzee Prairie Horse Guinea Pig Rabbit		Mouse Human Dog Bushbaby Xenopus Flycatcher Chicken Armadillo Gibbon Zebrafish	
Fig. 3 Multi and tetranu	ple sequence alignment of the TSS-flanking 5'UTRs. Examples \mathfrak{c} cleotide (b) categories. Species inclusion was based on the info	of ClustAl O prmation av	mega sequence alignment are represented in the tri- (a) ailable in the Ensembl database

A number of the identified STRs were linked to noncanonical translation in the following genes, *TEAD4*, *ECSCR*, *MPRIP*, *PAQR9*, *PRSS1*, and *ZSCAN30*.

Discussion

There is an ever-growing literature on the biological and pathological implications of STRs at the interand intraspecies levels [16-27]. The STRs listed in the present study are genetic codes that are unique to humans and are likely to be responsible for the human-specific regulation of the relevant genes. The significant enrichment of human-specific transcripts at the human-specific STR compartment indicates a link to a mechanism for TSS selection and transcript specificity.

A number of the identified STRs such as GTTT have established repressor activity [6, 28, 29] and are differentially expanded in certain genes in the Old World

NPAS1 AATGGGAGTCATGGCAGCCATG-----GTGGTCAG Human Human Chimpanzee Prairie CACCTGCGACTCCTTCGCGCCTGAATTCTCGAGTCCCCGCCGTG--.. ** ** **. * * *** * - AAT GGG AGT CAT GGC AGC CAT GGT GGT CAG -------AATGGGAGTCATGGCAGCCATGGTGGTCAG Human Human Gorilla GTCTCCAAGTCTGCGAGGCCGAGGTGGGCGCCGAGAGCTGGGCGCCACAGACCGCGCGTC ..* ..****: *..***.:********. Opossum -----CTCCTGGGTTCCTG-----**.*** : **: ------ AA TGGGA-----GTC ATGGCAGCC ATGGTGGTC AG------AATGGGAGTCATGGCAGCCATGGTGGTCAG-------Human Human Macaque TCTCCAAGTCTGCGAGGCCGAGGTGGGCGCCGAGAGCCAGGCGCCACAGCCCGCGCGTCT .*:***** :**. .** .****: *..***.:***** :.. ---- - AA TGG GAG T CA TGG C AG C CA TGG TGG T CA G ------Human GGTCTCAAAGTCTGCGAGGCCGCCGTGGGCGCCGAGAGCCGGGCTCCCCAGCCCGCGCGC Mouse .*:.* . *...***.***.*** Human Dog CGGAGACCTGGGTTCCTGGCCTTGAGGGAGGGCAGCTCCGGACGCCAAAGCCTGGCTTTA *. **. **. **. ***. * **.** **. **. **** AATGGGAGTCATGGCAGCCATGGTGGTCAG Human Horse -----CTGCCCGCA---+.+++. Fig. 4 Pair-wise sequence comparison of the TSS-flanking 5'UTRs. %identity scoring was performed between human and other species. Asterisks represent sequence identity

monkeys and Apes [14]. Purine STRs such as GAAA repeats are also functional in gene expression regulation, and their link to certain diseases unique to humans were previously reported [30, 31]. While the CG-rich STRs (e. g., CCG, GGGC) are subject to DNA methylation and can repress gene expression activity [32], they can also form G4 quadruplex structures, which have significant functions in gene expression regulation [33]. Several other identified STRs can form G4 structures with high overlap fraction (e.g., AGGG/CCCT, GCCC/GGGC).

It is not possible to estimate the number of crucial events that have led to the emergence of the human species. However, only a few genetic changes are needed to spur the evolution of new species in general, exemplified by the highly restricted initial divergence in butterfly hybridization models [34]. Accelerated evolution of a number of the identified genes in the present study (e.g., *DRD2*) has a well-established role in the origin of *Homo sapiens* [35]. Remarkably, a human-specific 7-amino acid transcript of this gene is flanked by a human-specific GCCC-repeat. Human-specific transcripts are increasingly recognized of having a role in the pathogenesis of diseases unique to the human species, such as schizo-phrenia [25, 36].

In a number of instances, not only the STR and the transcript, but also the gene containing these STRs and transcripts, were unique to humans, e.g., *AMY1C*, which is indicated in the evolution of the human phenotype during the Pleistocene [37].

For a number of the identified genes, sparse literature is available on the relevant function and pathways (e.g., *ARH-GEF35*, *CXorf40A*, *C22orf24*, *TMEM99*, and *ARL17B*).

In a number of the identified genes, the STRs were linked to noncanonical (non-AUG) translation. Although the significance of this compartment is unknown for the most part, recent emerging data indicates likely biological functions [38].

The plasticity of STRs confers them unique ability to respond to adaptive evolutionary processes in a more efficient way than the quaternary codes provided by the SNSs. This potential aspect of STRs is vastly unknown at present, and it is expected that identification of STRs that have evolved differentially in humans vs. other species may pave the way for better understanding of the evolutionary implication of these highly mutable motifs.

This study warrants expansion to other vitally important gene regulatory sequences such as the distal promoter, 5'UTR, and 3'UTR. It is also necessary to sequence these STRs in characteristics and diseases that are unique to the human species. The recent reports of mass STR analysis using CRISPR/Cas9 [39] make it particularly more feasible to investigate STRs in the context of human evolution.

Conclusion

Our findings provide the first evidence of genomescale skewing of STRs at a specific region of the human genome, and support a link between STRs and TSS selection/transcript specificity. The genes and STRs listed here may have a role in the divergence of humans from other species through the development of characteristics and phenotypes that are unique to the human species.

Abbreviations

SNS: Single nucleotide substitution; STR: Short tandem repeat; TF: Transcription factor; TSS: Transcription start site; UTR: Untranslated region

Funding

This research was funded by the University of Social Welfare and Rehabilitation Sciences, Tehran, Iran.

Availability of data and materials

Please contact author for data requests.

Authors' contributions

NN, FA, and TAA carried out the bioinformatics studies. AD and RS were the advisors of the project and helped in coordination. MO conceived the study, designed the project, supervised the analysis and wrote the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Iranian Research Center on Aging, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran. ²Department of Biostatistics, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran.

Received: 19 December 2017 Accepted: 20 March 2018 Published online: 05 April 2018

References

- Press MO, Carlson KD, Queitsch C. The overdue promise of short tandem repeat variation for heritability. Trends Genet. 2014;30(11):504–12.
- King DG. Evolution of simple sequence repeats as mutable sites. Adv Exp Med Biol. 2012;769:10-25.
- Ohadi M, Mohammadparast S, Darvish H. Evolutionary trend of exceptionally long human core promoter short tandem repeats. Gene. 2012; 507(1):61–7.
- Mohammadparast S, Bayat H, Biglarian A, Ohadi M. Exceptional expansion and conservation of a CT-repeat complex in the core promoter of PAXBP1 in primates. Am J Primatol. 2014;76(8):747–56.
- Nikkhah M, Rezazadeh M, Khorshid HK, Biglarian A, Ohadi M. An exceptionally long CA-repeat in the core promoter of SCGB2B2 links with the evolution of apes and Old World monkeys. Gene. 2016;576(1):109–14.
- Rezazadeh M, Gharesouran J, Mirabzadeh A, Khorshid HK, Biglarian A, Ohadi M. A primate-specific functional GTTT-repeat in the core promoter of CYTH4 is linked to bipolar disorder in human. Prog Neuro-Psychopharmacol Biol Psychiatry. 2015;56:161–7.

- Namdar-Aligoodarzi P, Mohammadparast S, Zaker-Kandjani B, Kakroodi ST, Vesiehsari MJ, Ohadi M. Exceptionally long 5' UTR short tandem repeats specifically linked to primates. Gene. 2015;569(1):88–94.
- Ohadi M, Valipour E, Ghadimi-Haddadan S, Namdar-Aligoodarzi P, Bagheri A, Kowsari A, et al. Core promoter short tandem repeats as evolutionary switch codes for primate speciation. Am J Primatol. 2015;77(1):34–43.
- Sonay TB, Carvalho T, Robinson MD, Greminger MP, Krützen M, Comas D, et al. Tandem repeat variation in human and great ape populations and its impact on gene expression divergence. Genome Res. 2015;25(11):1591–9.
- Li C, Lenhard B, Luscombe NM. Integrated analysis sheds light on evolutionary trajectories of young transcription start sites in the human genome. bioRxiv. 2017. https://doi.org/10.1101/192757.
- Kramer M, Sponholz C, Slaba M, Wissuwa B, Claus RA, Menzel U, et al. Alternative 5'untranslated regions are involved in expression regulation of human heme oxygenase-1. PLoS One. 2013;8(10):e77224.
- Li Y, Seidel K, Marschall P, Klein M, Hope A, Schacherl J, et al. A polymorphic microsatellite repeat within the ECE-1c promoter is involved in transcriptional start site determination, human evolution, and Alzheimer's disease. J Neurosci. 2012;32(47):16807–20.
- Emamalizadeh B, Movafagh A, Darvish H, Kazeminasab S, Andarva M, Namdar-Aligoodarzi P, et al. The human RIT2 core promoter short tandem repeat predominant allele is species-specific in length: a selective advantage for human evolution? Mol Gen Genomics. 2017;292(3):611–7.
- Khademi E, Alehabib E, Shandiz EE, Ahmadifard A, Andarva M, Jamshidi J, et al. Support for "disease-only" genotypes and excess of homozygosity at the CYTH4 primate-specific GTTT-repeat in schizophrenia. Genet Test Mol Biomarkers. 2017;21(8):485–90.
- Bushehri A, Barez MM, Mansouri S, Biglarian A, Ohadi M. Genome-wide identification of human-and primate-specific core promoter short tandem repeats. Gene. 2016;587(1):83–90.
- Valipour E, Kowsari A, Bayat H, Banan M, Kazeminasab S, Mohammadparast S, et al. Polymorphic core promoter GA-repeats alter gene expression of the early embryonic developmental genes. Gene. 2013;531(2):175–9.
- Heidari A, Fam ZNS, Esmaeilzadeh-Gharehdaghi E, Banan M, Hosseinkhani S, Mohammadparast S, et al. Core promoter STRs: novel mechanism for interindividual variation in gene expression in humans. Gene. 2012;492(1):195–8.
- Bagshaw AT. Functional mechanisms of microsatellite DNA in eukaryotic genomes. Genome Biol Evol. 2017;9(9):2428–43.
- Bagshaw AT, Horwood LJ, Fergusson DM, Gemmell NJ, Kennedy MA. Microsatellite polymorphisms associated with human behavioural and psychological phenotypes including a gene-environment interaction. BMC Med Genet. 2017;18(1):12.
- Gymrek M, Willems T, Guilmatre A, Zeng H, Markus B, Georgiev S, et al. Abundant contribution of short tandem repeats to gene expression variation in humans. Nat Genet. 2016;48(1):22–9.
- Hammock EA, Young LJ. Microsatellite instability generates diversity in brain and sociobehavioral traits. Science. 2005;308(5728):1630–4.
- Carrat GR, Hu M, Nguyen-Tu M-S, Chabosseau P, Gaulton KJ, van de Bunt M, et al. Decreased STARD10 expression is associated with defective insulin secretion in humans and mice. Am J Hum Genet. 2017;100(2):238–56.
- 23. Abe H, Gemmell NJ. Evolutionary footprints of short tandem repeats in avian promoters. Sci Rep. 2016;6:19421.
- Fondon JW, Hammock EA, Hannan AJ, King DG. Simple sequence repeats: genetic modulators of brain function and behavior. Trends Neurosci. 2008; 31(7):328–34.
- Alizadeh F, Bozorgmehr A, Tavakkoly-Bazzaz J, Ohadi M. Skewing of the genetic architecture at the ZMYM3 human-specific 5' UTR short tandem repeat in schizophrenia. Mol Gen Genomics. 2018; https://doi.org/10.1007/ s00438-018-1415-8.
- Hannan AJ. Tandem repeats mediating genetic plasticity in health and disease. Nat Rev Genet. 2018; https://doi.org/10.1038/nrg.2017.115.
- Watts PC, Kallio ER, Koskela E, Lonn E, Mappes T, Mokkonen M. Stabilizing selection on microsatellite allele length at arginine vasopressin 1a receptor and oxytocin receptor loci. Proc Biol Sci. 2017;284(1869):20171896. https:// doi.org/10.1098/rspb.2017.1896.
- Andrioli LPM, Vasisht V, Theodosopoulou E, Oberstein A, Small S. Anterior repression of a Drosophila stripe enhancer requires three position-specific mechanisms. Development. 2002;129(21):4931–40.
- Andrioli LP, Oberstein AL, Corado MS, Yu D, Small S. Groucho-dependent repression by sloppy-paired 1 differentially positions anterior pair-rule stripes in the Drosophila embryo. Dev Biol. 2004;276(2):541–51.

- Darvish H, Heidari A, Hosseinkhani S, Movafagh A, Khaligh A, Jamshidi J, et al. Biased homozygous haplotypes across the human caveolin 1 upstream purine complex in Parkinson's disease. J Mol Neurosci. 2013;51(2):389–93.
- Heidari A, Hosseinkhani S, Talebi S, Meshkani R, Esmaeilzadeh-Gharedaghi E, Banan M, et al. Haplotypes across the human caveolin 1 gene upstream purine complex significantly alter gene expression: implication in neurodegenerative disorders. Gene. 2012;505(1):186–9.
- Quilez J, Guilmatre A, Garg P, Highnam G, Gymrek M, Erlich Y, et al. Polymorphic tandem repeats within gene promoters act as modifiers of gene expression and DNA methylation in humans. Nucleic Acids Res. 2016; 44(8):3750–62.
- Sawaya S, Bagshaw A, Buschiazzo E, Kumar P, Chowdhury S, Black MA, et al. Microsatellite tandem repeats are abundant in human promoters and are associated with regulatory elements. PLoS One. 2013;8(2):e54710.
- Kronforst MR, Hansen ME, Crawford NG, Gallant JR, Zhang W, Kulathinal RJ, et al. Hybridization reveals the evolving genomic architecture of speciation. Cell Rep. 2013;5(3):666–77.
- Dorus S, Vallender EJ, Evans PD, Anderson JR, Gilbert SL, Mahowald M, et al. Accelerated evolution of nervous system genes in the origin of Homo sapiens. Cell. 2004;119(7):1027–40.
- Li M, Jaffe AE, Straub RE, Tao R, Shin JH, Wang Y, et al. A human-specific AS3MT isoform and BORCS7 are molecular risk factors in the 10q24. 32 schizophrenia-associated locus. Nat Med. 2016;22(6):649–56.
- Hardy K, Brand-Miller J, Brown KD, Thomas MG, Copeland L. The importance of dietary carbohydrate in human evolution. Q Rev Biol. 2015;90(3):251–68.
- Na CH, Barbhuiya MA, Kim MS, Verbruggen S, Eacker SM, Pletnikova O, Troncoso JC, Halushka MK, Menschaert G, Overall CM, Pandey A. Discovery of noncanonical translation initiation sites through mass spectrometric analysis of protein N termini. Genome Res. 2018;28(1):25–36.
- Shin G, Grimes SM, Lee H, Lau BT, Xia LC, Ji HP. CRISPR–Cas9-targeted fragmentation and selective sequencing enable massively parallel microsatellite analysis. Nat Commun. 2017;8:14291.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at www.biomedcentral.com/submit

