Supplementary Material 1 Single amino acid repeats (SAARs) of selected GenBank entries are tabulated here.
The homopolymer location column indicates the residue number range of the SAAR.
General codon bias for the repeat amino acid in the human proteome and some relevant information are given in the last column.

Protein name and mRNA Accession No.; Human ONLY (but repeats are ubquitious)	*Amino acid (number in run)*	*Homoplymer location*	*Codon run pattern (runs of three or more are in Blue); number (highest to lowest);* ***(XXX indicates a nonpolymer codon)***	*Codon*	*Human genomewide codon bias (highest to lowest) for each amino acid, and comments*
Fibrosin-1 **NM_001105079.2**	Ala (19)	831-849	GCU1,GCC3,GCU1,GCC1,GCU3,GCC3,GCU1,GCC3,GCA2,GCC1; **GCC=11,GCU=6,GCA=2**		**Ala:** GCC = 0.40, GCU = 0.27, GCA = 0.23, GCG = 0.11
	Glu (21) Discontinous	107-135	GAG1,GAA1,GAG7,XXX6,GAA1,XXX,GAG3,CCU,GAG3,GAA,GAG4; **GAG=18,GAA=3**		**Glu:** GAG = 0.58, GAA = 0.42
	Ser (6)	54-59	UCC1,UCG1,UCC1,UCG3; **UCG=4,UCC=2**		**Ser:** AGC=0.24,UCC=0.22,UCU=0.19,UCA=0.15,AGU=0.15,UCG=0.05
	Ser (15) Discontinous	371-386	UCA2,UCC1,UCU1,UCG1,UCC4,UCA1,UCU1,GCC,UCC2,UCG1,UCC1; **UCC=8,UCA=3,UCU=2,UCG=2**		In these two Ser runs, AGC/U is disfavored
SKI/DACH domain-containing protein 1 SKIDA1; NM_207371.3	Ala (21)	224-244	GCA1,GCC8,GCU2,GCC10; **GCC=18.GCU=2,GCA=1**		
	Ala (16)	303-318	GCG8,GCA1,GCG4,GCC2,GCG1; **GCG=13,GCC=2,GCA=1**		Follows codon repeat, not codon bias, since GCG is lowest in bias.
	His (15)	337-351	CAC7,CAU1,CAC7; **CAC=14,CAU=1**		**His:** CAC = 0.58, CAU = 0.42
	Glu (E) (16) Discontinuous	411-428	GAG1,GAA1,GAG4,GGA,GAG5,GAA1,GAG4; **GAG=14,GAA=2**		(GAA1,GAG4) may be a micro-repeat
Transcription factor SPT20 homolog-like 1; SPT20HL1 **NM_001136234.1**	Ala (55) Discontinous (mostly Pro)	495-565	GCU4,AUU,GCU18,GCU1,CCU,GCU1,CCU,GCU1,CUA,GCU5,CCU,GCU1,CUA,GCU5,CCU, etc; **GCU=54,GCC=1** (Overall Ala codon use in the whole protein: GCU = 0.73, GCC = 0.19, GCA = 0.17, GCG = 0.07; but if the repeat part is deleted, GCU drops: GCU = 0.2, GCC = 0.18, GCA = 0.17, GCG = 0.07;		**Ala:** GCC = 0.40, GCU = 0.27, GCA = 0.23, GCG = 0.11 Again follows repeat, not human codon usage bias.
Ribosomal protein S6 kinase, 90kDa, polypeptide 3 variant: AB208933.1	Ala (22)	17-38	GCC3,GCU5,GCC1,GCU1,GCC1,GCU1,GCC8,GCU1,GCC1; **GCC=14,GCU=8**		Although the overall usage appears to match bias (GCC>GCU).
Runt-related transcription factor 2 (RUNX2), transcript variant X1 XM_011514960.2	Gln (Q) (23)	117-139	CAA1,CAG3,CAA1,CAG4,CAA1,CAG6,CAA1,CAG6 **CAG=19,CAA=4**		**Gln:** CAG = 0.73, CAA = 0.27
	Ala (17)	141-157	GCG4,GCU1,GCG5,GCU1,GCG3,GCA1,GCU1,GCA1; **GCG =12,GCU=3,GCA=2**		Follows repeat, not bias; GCG is highest here, but has lowest bias. GCC, with highest codon bias, is totally absent in the repeats.
Zinc finger protein 358 (ZNF358), NM_018083.4	Ala (17)	419-435	GCA1,GCC1,GCU2,GCA1,GCU1,GCC1,GCG1,GCC1,GCU1,GCA1,GCU1,GCA2,GCG1,GCC2; **GCC=5,GCU=5,GCA=5,GCG=2**		Follows codon bias!! No run of 3 or more!! **Ala:** GCC = 0.40, GCU = 0.27, GCA = 0.23, GCG = 0.11
Homeodomain transcription factor NBPHOX, NM_003924.3	Ala (20)	241-260	GCA3,GCG3,GCC1,GCG1,GCA1,GCG4,GCA1,GCG1,GCA1,GCG2,GCA1,GCU1; **GCG=11,GCA=7,GCC=1,GCU=1**		**Ala:** GCC = 0.40, GCU = 0.27, GCA = 0.23, GCG = 0.11 Repeat, not codon bias
Homeobox A13, BC075791.1	Ala (18)	116-133	GCU1,GCC2,GCU2,GCC1,GCG1,GCU1,GCC1,GCU1,GCA1,GCC6,GCG1; **GCC=10,GCU=5,GCG=2,GCA=1**		Appears codon bias, not repeat, but there is a GCC6 micro-repeat.
POU domain, class 3, transcription factor 3 (POU3F3) NM_006236.2	Gly (20) one Ala in the middle	30-50	GGC3,GGG1,GGU1,GGC6,GGG1,GGC2,GCA,GGG1,GGC1,GGG1,GGC3; **GGC=15,GGG=4,GGU=1**		**Gly:** GGC = 0.34, GGA = 0.25, GGG = 0.25, GGU = 0.16 Appears bias, not repeat.
	Ala (16)	189-194	GCG1,GCC2,GCU1,GCC1,GCA1,GCC1,GCA1,GCC7,GCG1; **GCC=11,GCG=2,GCA=2,GCU=1**		Appears codon bias, not repeat.
	His (8)	270-277	**CAC8**		His: CAU = 0.42, CAC = 0.58 By repeat, not codon bias.
Ribosomal protein L14 (RPL14); BC005134.2	Ala (17)	150-166	**GCU17**		Clearly repeat, not codon bias.
FOXL2 (mutant 103 isolate); DQ016609.1	Ala (25)	221-245	GCG1,GCA1,GCC1,GCA1,GCG1,GCU1,GCA2,GCU1,GCG1,GCU2,GCA1,GCC1,GCA1, GCG1,GCU1,GCA2,GCU1,GCG1,GCU1,GCA1,GCC1,GCG1; **GCA9,GCU7.GCG6.GCC3**		Like ZNF358; no run of 3 or more! However, not exactly bias either. **Ala:** GCC = 0.40, GCU = 0.27, GCA = 0.23, GCG = 0.11
PH domain and leucine rich repeat protein phosphatase 1 (PHLPP1); NM_194449.3	Ala (20)	26-45	GCG1,GCC2,GCU1,GCG1,GCA5,GCG2,GCC1,GCG2,GCU1,CUG,GCG3,GCC1; **GCG9,GCA5,GCC4,GCU2**		Not much of a repeat (except GCA5), but also not fully bias; e.g., GCG, the lowest in codon bias, is highest here.
Zinc finger SWIM-type containing 6 (ZSWIM6); **Arg1163Trp mutation in acromelic frontonasal dysostosis; NM_020928.1**	Ala (16)	169-184	GCC4,GCU1,GCC10,GCG1; **GCC=14,GCG=1,GCU=1**		**Ala:** GCC = 0.40, GCU = 0.27, GCA = 0.23, GCG = 0.11 Essentially codon bias, but medium and micro GCC repeats
Aristaless-related homeobox protein, ARX; NM_139058.2	Ala (16)	100-115	GCG1,GCA1,GCG10,GCA1,GCG1,GCC2; **GCG=12,GCC=2,GCA=2**		Repeat of GCG, not bias. Pattern: GCG repeat is often favored (not always) in Ala repeats, although GCG is of lowest bias in human codon usage.
Transcription factor SP9; NM_001145250.1	Ala (16)	455-470	GCG2,GCA1,GCG4,GCA1,GCG7,GCC1		**Ala:** GCC = 0.40, GCU = 0.27, GCA = 0.23, GCG = 0.11

			GCG=13,GCA=2,GCC=1	Repeat of GCG, not bias. Extensive GCG usage seems common in Ala repeats.
Zinc-finger DNA binding protein (MAZ); U33819.1	Ala (16), intrrupted by one Val	495-511	GCG2,GCA1,GCG1,GCA1,GCG2,GCA1,GCG1,GCA2,GCG1,GCA2,GUA,GCA1,GCC1; **GCG=7,GCA=8,GCC=1**	No codon run; again GCG is favored, GCC is least, reverse of bias.
NK2 homeobox 3 (NKX2-3); NM_145285.2	Ala (16)	271-286	GCC1,GCG1,GCC9,GCA2,GCG2,GCC1; **GCC=11,GCG=3,GCA=2**	Overall it is close to codon bias, but has a single run of GCC9.
GS homeobox 2; NM 133267.2	Ala (16)	147-162	GCC1,GCG3,GCA4,GCG4,GCC1,GCG2,GCC1; **GCG=9,GCA=4,GCC3**	Mixture of runs; GCG > GCC, unlike codon bias
bHLH protein 3 (BHLHB3) AY665466.1	Ala (16)	117-132	GCG1,GCA1,GCC1,GCG1,GCA1,GCC4,GCU1,GCC5,GCG1; **GCC=10,GCG=3,GCA=2,GCU=1**	**Ala:** GCC = 0.40, GCU = 0.27, GCA = 0.23, GCG = 0.11 Like codon bias, because GCC is the most favoered, which is rare.
Solute carrier family 12, member 2 (SLC12A2) NM_001046.2	Ala (15)	93-107	GCC1,GCU2,GCG7,GCA1,GCG3,GCU1; **GCG=10,GCU=3,GCC=1,GCA=1**	GCG is highest, as in most cases, with runs.
Serum amyloid A activating factor 2	Ala (16)	96-112, one Val inside	GCC1,GCG1,GCU2,GCG1,GCC1,GCU1,GCC2,GCU2,GCC2,GUC,GCU1,GCC1,GCG1; **GCC=7,GCU=6,GCG=3**	No codon run, close to codon usage bias, though GCA is missing.
AlkB homolog 5, RNA demethylase (ALKBH5); NM 017758.3	Ala (18), one Val inside	30-48	GCC3,GCU1,GCC1,GCA1,GCC2,GUA,GCC2,GCA1,GCC1,GCA1,GCC2,GCU1,GCC2; **GCC=13,GCU=2,GCA=3**	**Ala:** GCC = 0.40, GCU = 0.27, GCA = 0.23, GCG = 0.11 Matches codon usage bias: rare example.
Engrailed homeobox 1 (EN1)	Ala (19), one Val inside	199-218	GCU1,GCG5,GCC1,GCG2,GCA1.GUA,GCG5,GCC1,GCA2,GCC1; **GCG=12,GCC=3,GCA=3,GCU=1**	More repeat than codon usage bias, as GCG is high with two runs.
				Gln: CAG = 0.73, CAA = 0.27 (Ratio ~3:1)
Ataxin 2 (ATXN2); NM_002973.3	Gln (23)	166-188	CAG13,CAA1,CAG9; **CAG=22,CAA1**	Two runs of CAG, with a single CAA codon in between
Ataxin 8 (ATXN8); DQ641254.1	Gln (59)	2-60	**CAG59**	This protein is all Gln, and the initiator Met; and all CAG, no CAA. Clearly repeat.
Huntingtin A; L12392.1	Gln (21)	18-38	CAG21,CAA1,CAG1; **CAG=22,CAA=1**	Long run of CAG, with a single penultimate CAA
Huntingtin, partial; EU797047.1	Gln (46)	18-63	CAG44,CAA1,CAG1; **CAG=45,CAA=1**	Long run of CAG, with a single penultimate CAA
TATA-box binding protein (TBP); NM_003194.4	Gln (38)	58-95	CAG3,CAA3,CAG8,CAA1,CAG1,CAA1,CAG19,CAA1,CAG1 (then AVAAAAV, i.e. V inside A) **CAG=35,CAA=3**	CAG repeat overwhelms (12:1) codon usage bias (3:1); strong sign of repeat.
E1A binding protein p400 (EP400); NM_015409.4	Gln (29)	2056-2084	CAG27,CAA1,CAG1; **CAG=28,CAA=1**	The most common pattern: CAG dominates, with a penultimate CAA
Thyroid hormone receptor-associated protein complex component TRAP230; AF117755.1	Repeat 1: Q (26)	2086-2111	CAG5,CAA1,CAG2,CAA1,CAG1,CAA1,CAG5,CAA1,CAG1,CAA1,CAG7; **CAG=21,CAA=5**	CAG repeat (4:1), close to codon bias (3:1).
	Repeat 2: Q (33); one His inside	2125-2158	CAG3,CAA1,CAG1,CAA1,CAG7,CAA1,CAG1,CAA1,CAG4,CAA1,CAG1,CAA2,CAG1,CAA1,CAC,CAG2,CAA1,CAG2,CAA1,CAG1; **CAG=23,CAA=8**	CAG repeat resembles codon bias (3:1).
Cancer-amplified transcriptional coactivator ASC-2 (ASC2); AF177388.1	Q25	261-285	CAG4,CAA4,CAG8,CAA2,CAG1,CAA1,CAG2,CAA2,CAG1; **CAG=16,CAA=9**	CAG repeat resembles codon bias (3:1).
NCOA3 / AIB1; NM_181659.2	Q29	1248-1276	CAG6,CAA1,CAG9,CAA1,CAG1,CAA1,CAG1,CAA1,CAG1,CAA1,CAG2,CAA1,CAG2,CAA1; **CAG=22,CAA=7**	CAG repeat resembles codon bias (3:1).
Androgen receptor (AR); ADD26781.1	Q23	58-80	CAG22,CAA1	22 CAG in a row, then a single CAA (way above codon bias)
	G23	451-473	GGU3,GGG1,GGU2,GGC17; **GGC=17,GGU=5,GGG=1**	GGC bias is the highest and slightly > than GGA, but that does not explain 17 GGC in a row.
FoxP2; BC143867.1	Q40	152-191	**CAG4**,CAA1,**CAG4**,CAA2,CAG2,CAA2,**CAG3**,**CAA5**,CAG2,CAA2,**CAG5**,CAA1,**CAG5**,CAA1,CAG1; **CAG=26,CAA=14**	Codon bias
UDP-N-acetylglucosaminyltransferase subunit (ALG13): NM_001099922.2; Many isoforms (X1-X15) have similar repeats	P27	919-945	**CCA14,CCU13**	**Pro:** CCC = 0.31, CCU = 0.29, CCA = 0.28, CCG = 0.11
Piccolo presynaptic cytomatrix protein (PCLO); NM_033026.5	P22L1P3	2405-2430	CCU2,CCC1,CCU2,CCC1,CCU1,CCC1,CCU5,CCA3,CCC1,CCU2,CCC1,CCA2,CUU,CCU1,CCA2; **CCU=13,CCC=5,CCA=7**	
Zinc finger homeobox 4 (ZFHX4); NM_024721.4	P3T1P20	2038-2061	CCA2,CCC1,ACU,CCU1,CCC1,CCA4,CCU7,CCC2,CCA1,CCU2,CCA1,CCU1; **CCA=8,CCC=4,CCU=11**	
Formin like-2 (FMNL2); BC167159; Many isoforms (X1-X8) have similar repeats	P2M1P21	550-574	CCA1,CCU1,AUG,CCA1,CCG4,CCA1,CCC1,CCU2,CCA1,CCU3,CCC1,CCA1,CCG1,CCC1,CCU1,CCC1,CCU2; **CCA=5,CCU=9,CCG=5,CCC=4**	
Zinc finger protein 341 (ZNF341); NM_001282933.1	P10L1P6Q1P6	178-201	CCU1,CCA1,CCU3,CCA1,CCU2,CCA2,CUG,CCC1,CCA1,CCG1,CCA2,CCU1,CAG,CCU1,CCA2,CCU1,CCA1,CCC1; **CCU=9,CCA=10,CCC=2,CCG=1**	
Preproacrosin; Y00970.1	P2A2Q1P1R1P3S1P8A1S1P1L1P9	330-370	CCA1,CCG1,GCA,GCC,CAG,CCC1,CGA,CCC1,CCA1,CCU1,UCA,CCC1,CCG1,CCC1,CCA1,CCC1,CCA1,CCU1,CCA1,GCC,UCA,CCU1,UUA,CCC1,CCA1,CCC1,CCA1,CCC1,CCA1,CCC1,CCA1,CCU1; **CCA=9,CCC=9,CCG=2,CCU=4**	
				Arg: CGG=0.20, CGC=0.18, CGA=0.11,CGU=0.08; AGA=0.21,AGG=0.21
FMR1; M67468.1	R20	13-42	All **CGG**	Repeat, since CGC, the close 2nd highest bias is absent.
SR-repetitive matrix protein 3; NM_001291831.1	R11	385-395	AGG1,**CGG5**,CGU1,AGG1,CGG2,CGC1; **CGG=7,AGG=2,CGC=1,CGU=1**	Codon bias