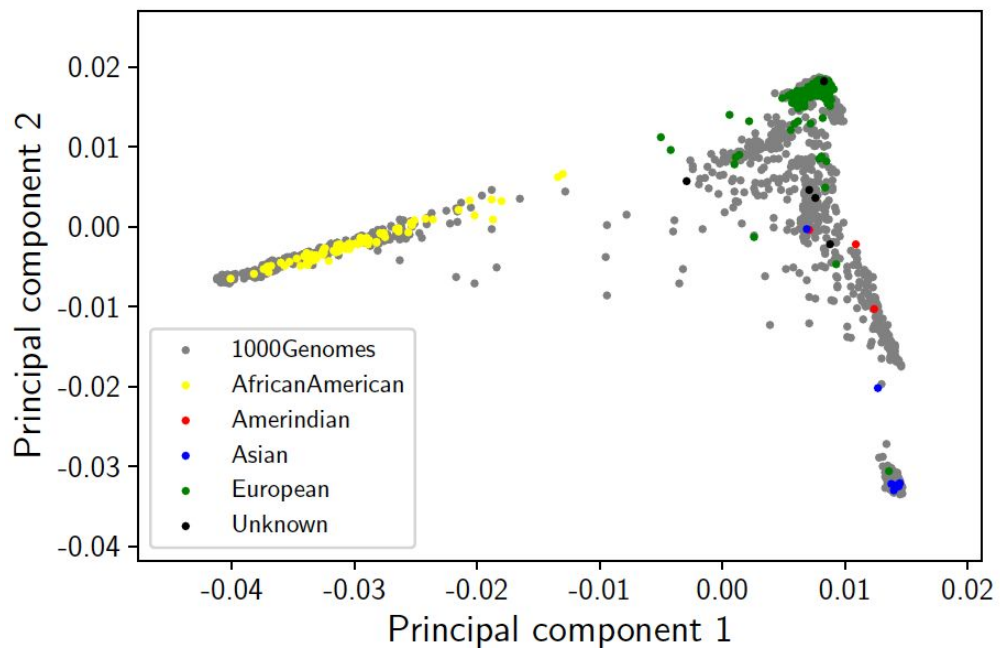


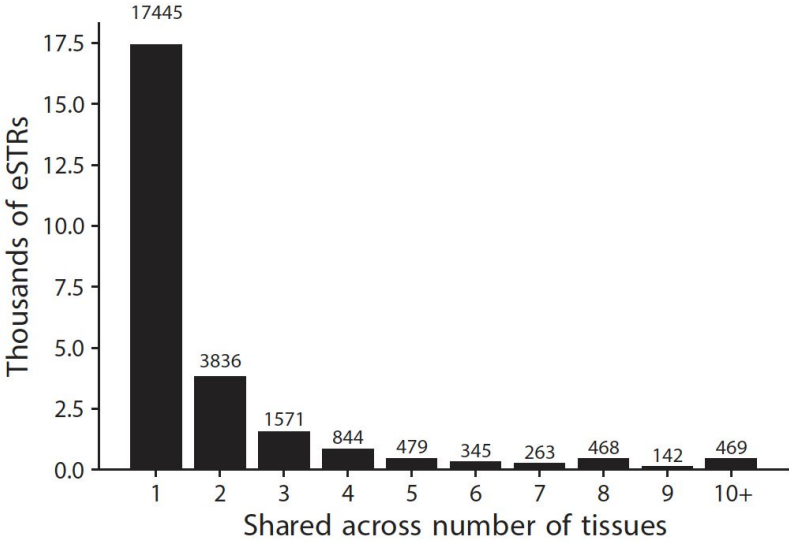
**Figure S1.** Related to Figure 1



**Analysis of GTEx population structure.** Principal component analysis was performed using SNP genotypes from the GTEx and 1000G cohorts. Samples from the 1000 Genomes project are shown in gray and GTEx samples are shown as colored dots based on ethnicity provided for each sample (yellow=African American; red=Amerindian; blue=Asian; green=European, black=Unknown).

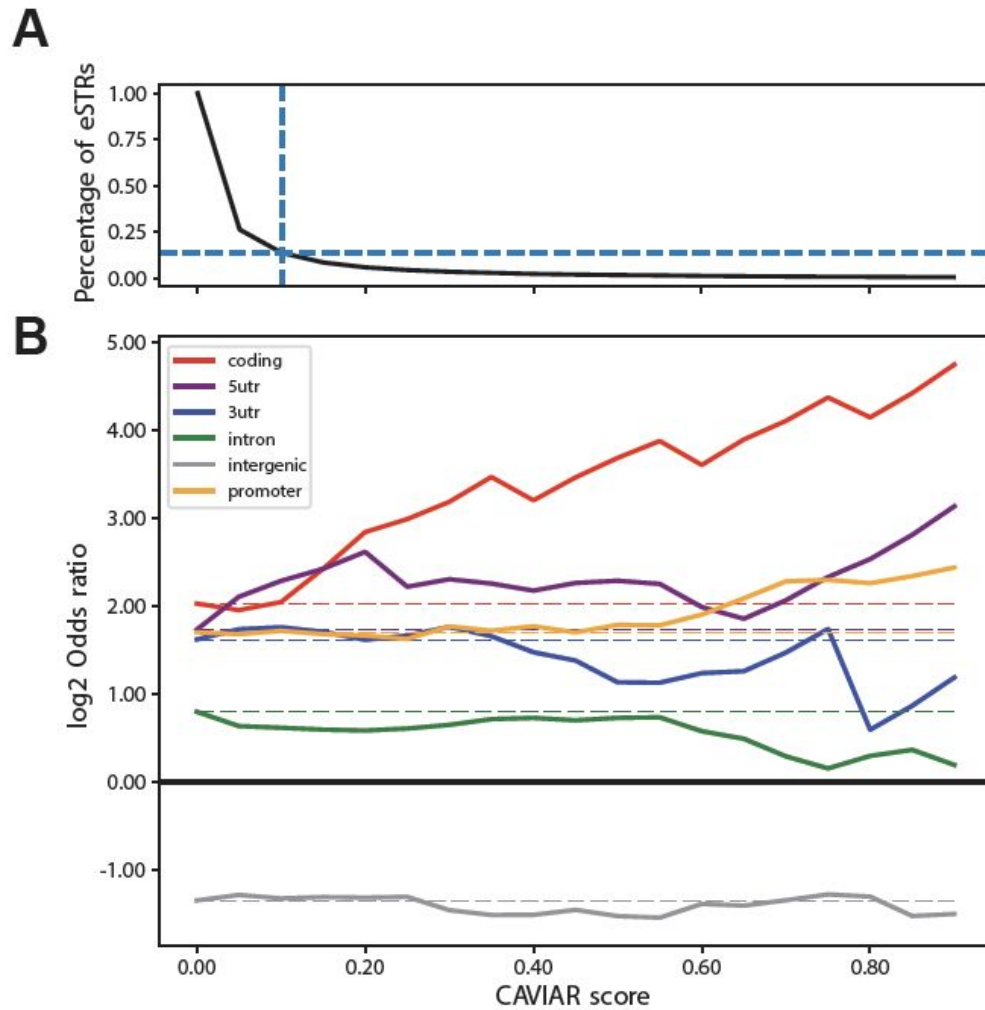


**Figure S3.** Related to Figure 1



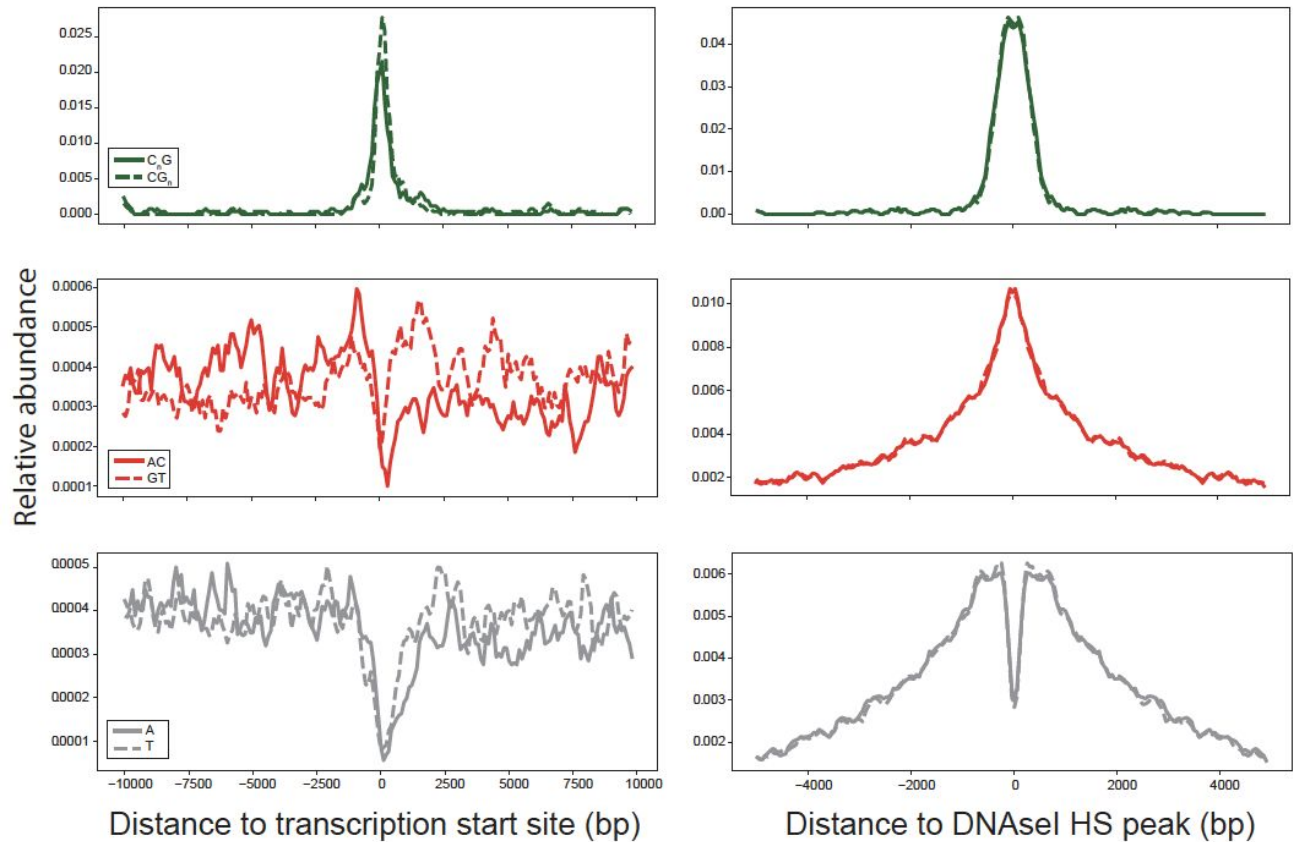
**Sharing of eSTRs across tissues.** The x-axis represents the number of tissues that share a given eSTR. The y-axis and annotated values for each bar represent the number of eSTRs shared across a given number of tissues.

Figure S4. Related to Figure 1



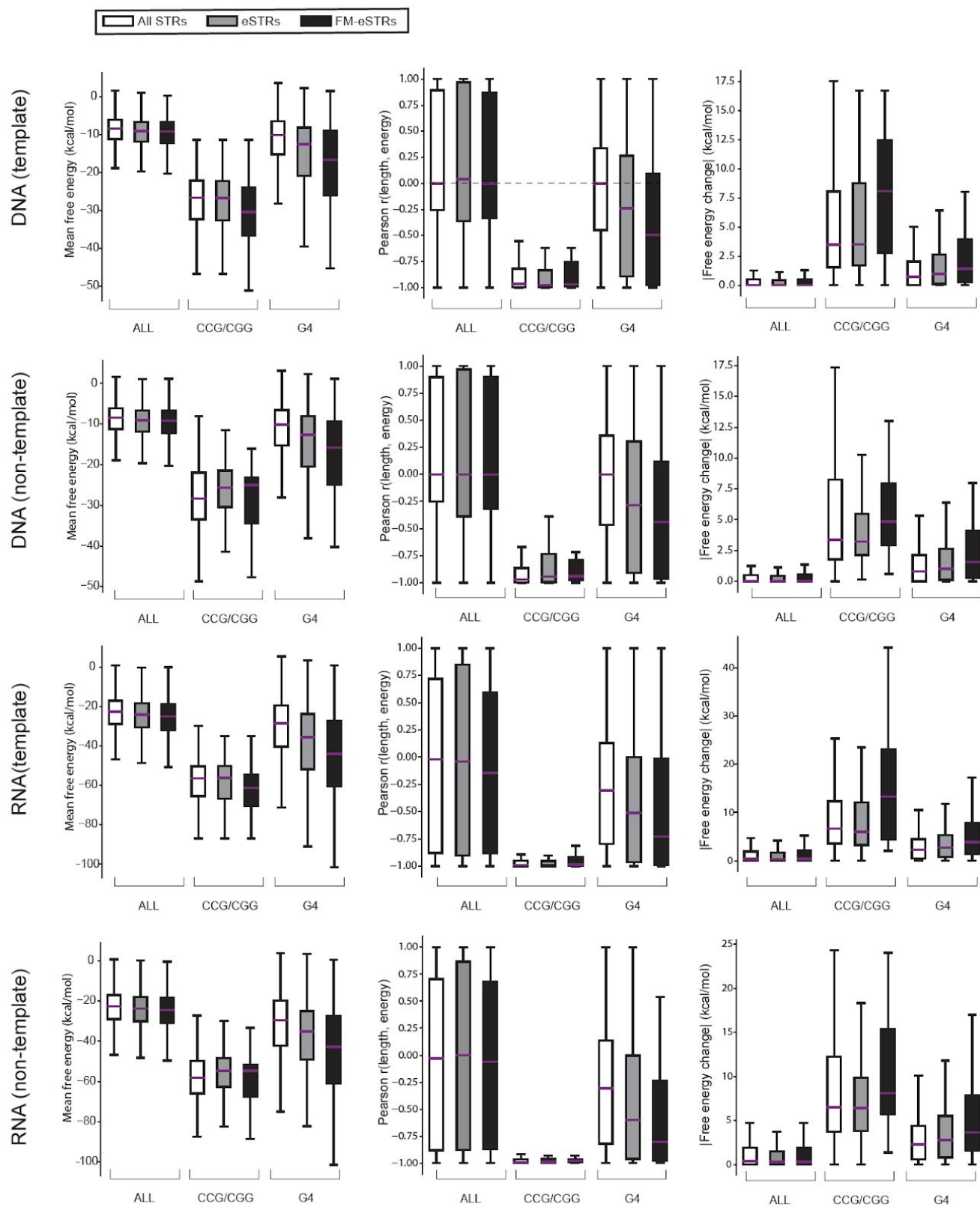
**FM-eSTRs are enriched in coding and regulatory regions. A. Percent of eSTRs that are annotated as FM-eSTRs as a function of CAVIAR threshold.** The x-axis represents CAVIAR scores and the y-axis gives the percentage of eSTRs with CAVIAR scores in at least one tissue above each threshold. The dashed horizontal line gives the percent of eSTRs with CAVIAR scores of at least 10%. **B. Enrichment of gene annotation categories as a function of CAVIAR threshold.** The y-axis represents the log<sub>2</sub> odds ratio for enrichment of each category in eSTRs passing each threshold (solid line). The dashed line represents the odds ratio when considering all eSTRs regardless of CAVIAR score. Red=coding, purple=5' UTR, blue=3' UTR, green=intron, gray=intergenic, orange=promoter (within 3kb upstream of each gene).

**Figure S5.** Related to Figure 2



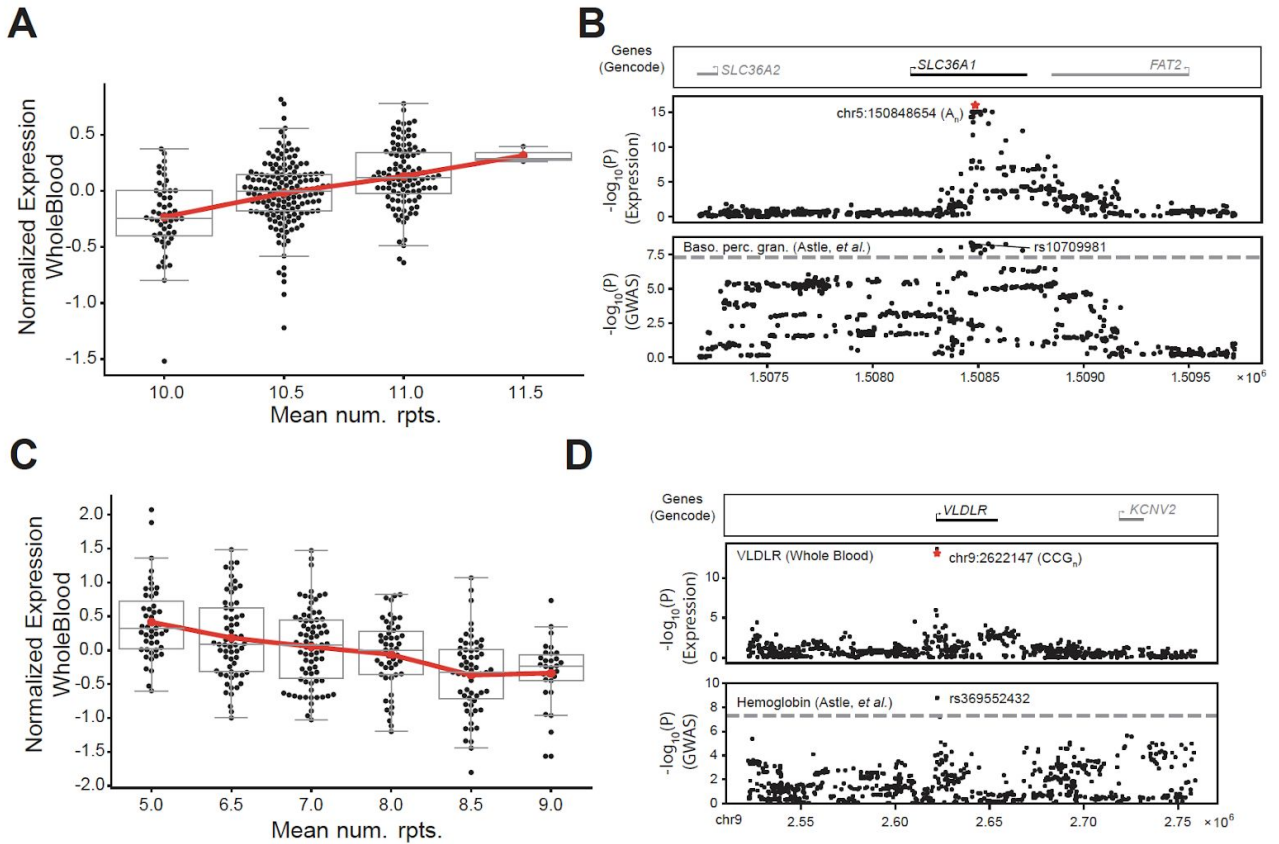
**Localization of all STRs around putative regulatory regions.** Left and right plots show localization around transcription start sites and DNaseI HS clusters, respectively. The y-axis denotes the relative number of STRs of each type in each bin. For promoters, the x-axis is divided into 100bp bins. For DNaseI HS sites, the x-axis is divided into 50bp bins. In each plot, values were smoothed by taking a sliding average of each four consecutive bins. Only STR-gene pairs passing all filters are considered. Each plot compares localization of the two possible sequences of a given repeat motif on the coding strand. *i.e.* top plots compare motifs of the form  $C_nG$  vs. their reverse complement on the opposite strand, middle plots compare AC vs. GT repeats, and bottom plots compare A vs. T repeats. The strand of each STR was determined based on the coding strand of each target gene.

**Figure S6.** Related to Figure 4



**GC-rich eSTRs modulate DNA or RNA secondary structure.** Plots are same as those shown in **Figure 4C-E** (top row), but are additionally shown for sequences on the non-template DNA strand (2nd row), template RNA strand (3rd row), and non-template RNA strand (4th row). **Left: Free energy of STRs and their context.** Boxplots show the distribution of free energy for each STR +/- 50bp of context sequence, computed as the average across all alleles at each STR. **Center: Pearson correlation between STR length and free energy.** Correlations were computed separately for each STR, and plots show the distribution of correlation coefficients across all STRs. **Right: Magnitude of free energy change across alleles.** For each STR the absolute value of the difference in energy between the shortest and longest allele was computed. Plots show the distribution of these values across all STRs. For **C-E**, purple lines give medians and boxes span from the 25th percentile (Q1) to the 75th percentile (Q3). Whiskers extend to  $Q1 - 1.5 * IQR$  (bottom) and  $Q3 + 1.5 * IQR$  (top), where IQR gives the interquartile range ( $Q3 - Q1$ ). White boxes show all STRs in each category, gray boxes show eSTRs, and black boxes show FM-eSTRs. Upper brackets denote significant differences for all STRs (white) across categories or for significant differences within each category between all STRs (white) and FM-eSTRs (black).

**Figure S7.** Related to Figure 5



**Example GWAS signals falling within FM-eSTRs. A.,C. eSTR associations for candidate FM-eSTRs underlying GWAS signals.** For each plot, the x-axis shows STR genotype as the mean number of repeats, and the y-axis shows normalized expression. Each point represents a single individual. Red lines show the mean expression for each x-axis value. Boxplots summarize the expression distribution for each genotype as described in **Figure 1D**. Gene diagrams are not drawn to scale. **A.** shows association between a homopolymer A/T STR (chr5:150848654) and *SLC36A1*. **B.** shows association between a CCG STR (chr9:2622147) and *VLDLR*. **B.,D. Summary statistics for gene expression and GWAS signals.** Top panels show genes in each region. The target gene for the eQTL associations is shown in black. Middle panels show the  $-\log_{10}$  p-values of association between each variant and the expression of the target gene. The FM-eSTR is denoted by a red star. Bottom panels show the  $-\log_{10}$  p-values of association from published GWAS for each trait. Dashed gray horizontal lines show the genome-wide significance threshold of  $5e-8$ . **B.** shows eQTL signals for *SLC36A1* (middle) and GWAS signals for basophil percentage of granulocytes (bottom). **D.** shows eQTL signals for *VLDLR* in whole blood (middle) and GWAS signals for hemoglobin (bottom).



**Table S1.** Related to Figure 1

<b>Tissue</b>	<b>Sample size</b>	<b># eSTRs (FDR&gt;10%)</b>	<b># FM-eSTRs (CAVIAR score&gt;0.1)</b>
Tibial Nerve	265	4,352	312
Thyroid	262	4,105	330
Sun-exposed Skin	297	3,827	298
Subcutaneous Adipose	270	3,587	292
Tibial Artery	276	3,454	269
Esophagus Mucosa	255	3,461	286
Skeletal Muscle	343	3,370	272
Transformed Fibroblasts	225	3,088	248
Lung	259	2,989	277
Esophagus Muscularis	214	2,824	256
Whole Blood	336	2,585	206
Aorta Artery	191	2,396	234
Non-exposed Skin	209	2,386	200
Visceral Adipose	193	1,840	175
Heart-Left Ventricle	199	1,731	212
Brain-Cerebellum	107	1,551	191
Brain-Caudate	108	675	103

**Summary of multi-tissue eSTR identification.**

**Table S2.** Related to Figure 1

Gene	STR (hg19)	Reference effect	GTE <sub>x</sub> eSTR evidence (p<0.01)
<i>CSTB</i> (Borel et al., 2012)	21:45196326 (CGGGGCGGGGCG <sub>n</sub> )	Expansion implicated in myoclonus epilepsy. Normal alleles show increasing effect in LCLs ( $\beta > 0$ ).	<b>Adipose-Visceral; p=2.5e-6; <math>\beta=0.34</math></b> <b>Transformedfibroblasts; p=1.5e-18; <math>\beta=0.55</math></b> <b>Skin-SunExposed; p=1.89e-14; <math>\beta=0.43</math></b> <b>Heart-LeftVentricle; p=1.6e-11; <math>\beta=0.46</math></b> <b>Esophagus-Muscularis; p=2.8e-9; <math>\beta=0.40</math></b> <b>Artery-Tibial; p= 2.3e-15; <math>\beta=0.46</math></b> <b>Artery-Aorta; p=2.0e-5; <math>\beta=0.31</math></b> <b>Adipose-Subcutaneous; p=1.9e-4; <math>\beta=0.23</math></b> <b>Skin-NotSunExposed; p=6.3e-7; <math>\beta=0.34</math></b> <b>Thyroid; p=4.1e-16; <math>\beta=0.48</math></b> <b>WholeBlood; p=2.1e-14; <math>\beta=0.41</math></b> <b>Nerve-Tibial; p=7.3e-14; <math>\beta=0.45</math></b> <b>Muscle-Skeletal; p=1.3e-21; <math>\beta=0.49</math></b>
<i>NOP56</i> (Matsuzono et al., 2017)	20:2633379 (GGCCTG <sub>n</sub> )	Expansion implicated in spinocerebellar ataxia 36. Large expansions show decreasing effect in SCA36 patient iPSC cell lines ( $\beta < 0$ ). No data on normal alleles.	Skin-SunExposed; p= 6.0e-05; $\beta=0.23$ Artery-Tibial; p=9.5e-5; $\beta=0/24$ Adipose-Subcutaneous; p=0.0011; $\beta=0.20$ Skin-NotSunExposed; p=0.0039; $\beta=0.21$ Thyroid; p=2.4e-8; $\beta=0.34$ Muscle-Skeletal; p=7.8e-12; $\beta=0.36$
<i>MMP9</i> (Shimajiri et al., 1999)	20:44637413 (AC <sub>n</sub> )	Increasing effect in esophageal carcinoma cell lines ( $\beta > 0$ ).	<b>Skin-SunExposed; p=0.0039; <math>\beta=0.18</math></b>
<i>EGFR</i> (Gebhardt et al., 1999)	7:55088254 (AC <sub>n</sub> )	<i>In vitro</i> decreasing effect; decreasing but non-linear effect across various cell lines ( $\beta < 0$ ).	<b>Heart-LeftVentricle; p=0.0073; <math>\beta=-0.21</math></b> Esophagus-Mucosa; p=0.0018; $\beta=0.21$ Muscle-Skeletal; p=0.0070; $\beta=0.16$
<i>TP53/3</i> (Contente et al., 2002)	2:24307211 (TGYCC <sub>n</sub> )	Increasing effect in presence of p53 in H1299 cells ( $\beta > 0$ ).	<b>Esophagus-Mucosa; p=0.0081; <math>\beta=0.18</math></b> <b>Thyroid; p=0.00060; <math>\beta=0.22</math></b> <b>Muscle-Skeletal; p=0.0068; <math>\beta=0.15</math></b>
<i>ALOX5</i> (Vikman et al., 2009)	10:45869549 (CGGGGG <sub>n</sub> )	Increasing effect in monocytes ( $\beta > 0$ ), decreasing effect in lymphocytes ( $\beta < 0$ ).	<b>Esophagus-Mucosa; p=4.35e-7; <math>\beta=0.32</math></b> <b>Brain-Cerebellum; p=0.0027; <math>\beta=0.29</math></b> <b>Skin-SunExposed; p=9.0e-4; <math>\beta=-0.19</math></b>
<i>IFNG</i> (Saha et al., 2005)	12:68552495 (AC <sub>n</sub> )	<i>In vitro</i> reporter assay showed increasing effect for CA12 vs. CA15 ( $\beta > 0$ ).	No GTE <sub>x</sub> eSTRs found
<i>UGT1A1</i> (Johnson et al., 2009)	2:234668880 (AT <sub>n</sub> )	Decreasing effect in human hepatoma ( $\beta < 0$ )	No GTE <sub>x</sub> eSTRs found

**STRs previously associated with expression that were analyzed in the GTEx cohort.** The right hand column reports STR by gene associations in the GTEx cohort with nominal  $p < 0.01$ . Nominal p-values and effect sizes ( $\beta$ ) are given for each eSTR. Bolded results in the right column indicate eSTRs that agree with the direction of effect reported previously. Note several previously reported eSTRs (e.g. a TG repeat in the promoter of *HMOX1* (Køllgaard et al., 2016) and a TC repeat in the promoter of *HMGA2* (Borrmann et al., 2003)) did not pass our quality filters and thus were not included.

**Table S3.** Related to Figure 2

Annotation	# STRs			eSTR enrichment		FM-eSTR enrichment	
	All	eSTRs	FM-eSTRs	P-val	OR	P-val	OR
Coding	217	75	16	4.0e-19	4.1	5.0e-6	4.1
5' UTR	657	197	56	1.9e-37	3.3	2.8e-20	4.9
3' UTR	1,660	468	100	3.1e-77	3.1	2.6e-23	3.4
Intron	74,501	10,998	1,759	1.4e-294	1.7	1.1e-33	1.5
Intergenic	91,325	6,458	1,016	p<1e-300	0.39	2.5e-142	0.40
Promoter (within 3kb upstream of TSS)	6,866	1,941	380	p<1e-300	3.2	7.9e-77	3.2

**Enrichment of eSTRs for genomic annotations.** For each annotation, a Fisher's exact test was performed to test whether eSTRs showed significantly more or less overlap compared to all STRs analyzed. FM-eSTRs represent all eSTRs with CAVIAR score of at least 10% as defined in the main text. OR=odds ratio and p-values are two-sided. Each STR was assigned to at most one category as described in the **Methods**.

**Table S4.** Related to Figure 2

Factor	# STRs			eSTR enrichment		FM-eSTR enrichment	
	All	eSTRs	FM-eSTRs	P-val	OR	P-val	OR
KDM5A	13	8	6	2.3E-05	12.3	7.1E-08	44.4
RDBP	9	5	3	1.7E-03	9.6	5.3E-04	25.9
THAP1	32	19	6	1.1E-10	11.3	2.8E-05	11.9
SAP30	170	72	27	2.0E-24	5.7	3.8E-17	9.8
SIX5	99	42	15	6.3E-15	5.7	7.0E-10	9.3
PHF8	462	198	68	2.4E-65	5.8	1.0E-38	9.1
ELK4	111	46	15	9.9E-16	5.5	3.6E-09	8.1
NRF1	127	47	17	8.0E-14	4.5	3.9E-10	8.0
KDM5B	321	137	42	1.8E-45	5.8	1.8E-22	7.9
SREBP1	38	17	5	2.8E-07	6.2	7.3E-04	7.8
CTCFL	62	16	8	1.9E-03	2.7	2.3E-05	7.7
NFYA	81	29	10	1.0E-08	4.3	3.3E-06	7.3
HMGN3	310	126	38	2.9E-39	5.3	1.7E-19	7.3
E2F6	416	159	50	5.0E-45	4.8	9.2E-25	7.2
CREB1	241	93	29	2.2E-27	4.9	5.1E-15	7.1
SP4	117	54	14	8.1E-21	6.6	5.7E-08	7.0
ZBTB7A	315	113	36	1.3E-29	4.3	1.5E-17	6.7
CEBPD	175	66	20	2.8E-19	4.7	2.1E-10	6.7
GTF2B	44	17	5	3.3E-06	4.9	1.4E-03	6.6
CCNT2	401	143	45	1.1E-36	4.3	3.1E-21	6.6
HDAC1	244	92	27	3.1E-26	4.7	3.3E-13	6.5
ELK1	141	50	15	9.1E-14	4.2	9.5E-08	6.2
TAF7	196	79	20	6.0E-25	5.2	1.5E-09	5.9

IRF1	430	147	43	2.7E-35	4.0	1.9E-18	5.8
SIN3A	585	185	58	1.1E-38	3.6	3.6E-24	5.8
E2F4	295	96	29	7.2E-22	3.7	9.3E-13	5.7
ZKSCAN1	72	16	7	8.5E-03	2.2	4.4E-04	5.6
SRF	159	41	15	5.6E-07	2.7	4.6E-07	5.4
TAF1	937	298	85	5.7E-62	3.6	6.4E-32	5.3
GABPA	586	180	53	7.4E-36	3.4	2.0E-20	5.2
E2F1	604	190	54	2.4E-39	3.6	1.5E-20	5.1
MAZ	992	314	85	1.1E-64	3.6	3.7E-30	4.9
RBBP5	710	216	61	5.6E-42	3.4	4.0E-22	4.9
NFYB	200	57	17	7.7E-11	3.1	3.6E-07	4.8
UBTF	273	88	23	7.0E-20	3.7	4.0E-09	4.8
ETS1	250	88	21	8.9E-23	4.2	2.0E-08	4.8
ZEB1	60	17	5	3.1E-04	3.0	5.7E-03	4.7
EGR1	658	186	54	1.3E-31	3.1	7.0E-19	4.7
SMARCB1	257	89	21	1.9E-22	4.1	3.2E-08	4.6
SP2	37	12	3	6.1E-04	3.7	3.3E-02	4.6
CHD1	512	144	41	9.9E-25	3.0	2.4E-14	4.5
CHD2	624	177	49	1.6E-30	3.1	1.6E-16	4.5
BRCA1	116	38	9	1.1E-09	3.8	3.9E-04	4.4
FOXP2	572	141	44	2.0E-18	2.5	1.1E-14	4.3
PML	642	175	49	7.3E-28	2.9	5.0E-16	4.3
HDAC2	472	132	36	1.4E-22	3.0	3.6E-12	4.3
TCF3	342	93	26	2.2E-15	2.9	3.6E-09	4.3
MXI1	786	218	59	2.0E-35	3.0	1.2E-18	4.3
ELF1	665	200	50	3.4E-38	3.3	4.6E-16	4.2
POU2F2	492	125	37	9.8E-18	2.6	2.8E-12	4.2

SIN3AK20	900	263	67	3.8E-47	3.2	8.9E-21	4.2
MBD4	165	46	12	9.3E-09	3.0	8.4E-05	4.1
TBP	834	218	59	1.6E-31	2.7	1.9E-17	4.0
PAX5	594	161	41	1.8E-25	2.9	2.9E-12	3.9
MEF2C	159	32	11	1.6E-03	1.9	2.5E-04	3.8
STAT1	392	107	27	1.1E-17	2.9	1.5E-08	3.8
STAT2	58	17	4	2.0E-04	3.2	2.4E-02	3.8
SMARCC2	60	15	4	3.4E-03	2.6	2.7E-02	3.7
GTF2F1	301	79	19	1.4E-12	2.7	6.7E-06	3.5
RFX5	430	97	27	9.4E-11	2.2	9.6E-08	3.5
BCLAF1	225	53	14	3.2E-07	2.4	1.2E-04	3.4
RELA	897	216	55	4.0E-26	2.5	8.4E-14	3.4
TFAP2A	390	83	24	2.8E-08	2.1	6.9E-07	3.4
SP1	758	191	46	9.3E-26	2.6	1.2E-11	3.4
YY1	1,187	278	71	6.5E-31	2.4	8.0E-17	3.3
REST	952	209	57	5.0E-20	2.2	8.0E-14	3.3
TBL1XR1	449	110	27	1.7E-14	2.5	2.3E-07	3.3
ZBTB33	234	64	14	3.7E-11	2.9	1.8E-04	3.3
MYBL2	438	109	26	7.2E-15	2.6	4.8E-07	3.3
POLR2A	4,645	1,118	262	1.1E-130	2.5	4.1E-54	3.3
FOSL1	170	35	10	6.4E-04	2.0	1.6E-03	3.2
BHLHE40	757	170	44	1.3E-17	2.2	1.3E-10	3.2
MTA3	353	83	20	2.7E-10	2.4	1.9E-05	3.1
MYC	1,984	460	110	4.0E-49	2.4	1.1E-22	3.1
RXRA	249	63	14	1.5E-09	2.6	3.4E-04	3.1
WRNIP1	254	63	14	4.5E-09	2.5	4.2E-04	3.0
NR2C2	145	30	8	1.5E-03	2.0	6.8E-03	3.0

ZNF143	620	139	34	1.3E-14	2.2	6.2E-08	3.0
MAX	1,718	429	92	6.9E-55	2.6	3.0E-18	3.0
USF2	279	64	15	8.0E-08	2.3	3.4E-04	2.9
NR3C1	467	94	25	8.1E-08	1.9	4.9E-06	2.9
STAT5A	474	96	25	4.4E-08	2.0	6.4E-06	2.9
USF1	665	153	35	5.2E-17	2.3	1.1E-07	2.9
TFAP2C	504	112	26	8.5E-12	2.2	6.2E-06	2.8
MEF2A	340	64	17	7.9E-05	1.8	3.3E-04	2.7
ARID3A	542	116	27	5.0E-11	2.1	7.8E-06	2.7
SMARCA4	120	24	6	6.1E-03	1.9	2.7E-02	2.7
RCOR1	805	177	40	3.1E-17	2.2	6.5E-08	2.7
NFIC	1,161	222	57	3.3E-14	1.8	1.9E-10	2.7
HNF4A	266	60	13	3.4E-07	2.2	1.9E-03	2.7
BACH1	226	54	11	1.7E-07	2.4	4.3E-03	2.6
CBX3	392	100	19	1.6E-14	2.6	2.3E-04	2.6
CTBP2	207	44	10	6.7E-05	2.1	6.6E-03	2.6
JUNB	228	51	11	3.2E-06	2.2	4.6E-03	2.6
EZH2	858	153	41	4.9E-08	1.7	2.0E-07	2.6
TCF12	813	167	38	1.2E-13	2.0	7.8E-07	2.6
RUNX3	1,193	238	55	2.6E-17	1.9	3.6E-09	2.5
SMARCC1	240	51	11	1.6E-05	2.1	6.7E-03	2.5
GTF3C2	111	29	5	1.7E-05	2.7	6.1E-02	2.4
ATF3	313	82	14	8.7E-13	2.7	3.0E-03	2.4
EBF1	682	129	30	1.7E-08	1.8	2.9E-05	2.4
FOXM1	548	123	24	4.6E-13	2.2	2.0E-04	2.4
TRIM28	228	56	10	3.2E-08	2.5	1.3E-02	2.4
BATF	390	64	17	4.0E-03	1.5	2.1E-03	2.4



GATA1	668	165	29	1.6E-21	2.5	4.9E-05	2.4
ATF1	161	43	7	9.1E-08	2.8	3.5E-02	2.4
FOSL2	719	145	31	2.1E-11	2.0	4.5E-05	2.3
ZNF263	814	167	35	1.2E-13	2.0	1.3E-05	2.3
FOXA2	584	108	25	6.7E-07	1.8	2.0E-04	2.3
PBX3	213	52	9	1.3E-07	2.5	2.1E-02	2.3
JUND	1,707	312	71	1.6E-16	1.7	1.9E-09	2.3
TCF7L2	1,256	224	52	3.8E-11	1.7	3.6E-07	2.3
HNF4G	225	47	9	5.5E-05	2.0	4.3E-02	2.2
TEAD4	987	187	39	9.5E-12	1.8	3.0E-05	2.1
ZZZ3	76	10	3	5.9E-01	1.2	1.8E-01	2.1
BCL3	948	182	37	5.1E-12	1.8	6.0E-05	2.1
SPI1	795	167	30	1.6E-14	2.1	5.6E-04	2.0
FOXA1	1,309	216	49	6.2E-08	1.5	1.5E-05	2.0
NR2F2	244	52	9	1.2E-05	2.1	5.4E-02	2.0
ATF2	697	147	25	4.5E-13	2.1	3.0E-03	1.9
SMC3	737	144	26	2.8E-10	1.9	2.7E-03	1.9
SETDB1	709	118	25	4.4E-05	1.5	3.4E-03	1.9
ESR1	284	43	10	6.2E-02	1.4	7.5E-02	1.9
EP300	2,558	413	88	1.4E-12	1.5	1.7E-07	1.9
NFATC1	551	109	18	2.0E-08	1.9	2.7E-02	1.7
IKZF1	252	46	8	1.5E-03	1.7	1.6E-01	1.7
CTCF	2,746	522	86	7.3E-31	1.8	1.1E-05	1.7
IRF4	317	49	10	3.4E-02	1.4	9.9E-02	1.7
GATA2	1,628	240	49	6.6E-05	1.3	1.9E-03	1.6
ZNF217	299	37	9	6.5E-01	1.1	1.9E-01	1.6
BCL11A	271	38	8	1.8E-01	1.3	1.8E-01	1.6

TAL1	309	64	9	2.8E-06	2.0	2.0E-01	1.6
SUZ12	421	73	12	4.0E-04	1.6	1.5E-01	1.5
JUN	923	157	26	6.1E-07	1.6	5.1E-02	1.5
CEBPB	2,329	405	64	2.9E-17	1.6	4.5E-03	1.5
KAP1	1,024	172	28	4.8E-07	1.6	6.4E-02	1.5
RPC155	117	23	3	8.7E-03	1.9	4.9E-01	1.4
GATA3	1,134	139	29	4.3E-01	1.1	1.0E-01	1.4
STAT3	1,069	159	27	7.5E-04	1.3	1.4E-01	1.3
RAD21	1,247	205	30	1.7E-07	1.5	1.8E-01	1.3
MAFF	646	94	13	1.9E-02	1.3	7.7E-01	1.1
FOS	2,045	278	39	3.4E-03	1.2	9.4E-01	1.0
MAFK	1,176	156	21	6.0E-02	1.2	9.1E-01	0.9

**Enrichment of eSTRs in binding sites for transcription factors and chromatin regulators annotated by ENCODE.** For each factor, a Fisher's exact test was performed to test whether eSTRs showed significantly more or less overlap compared to all STRs analyzed. FM-STRs represent all eSTRs with CAVIAR score of at least 10% as defined in the main text. OR=odds ratio and p-values are two-sided. Transcription factor annotations are described in **Methods**.

**Table S5.** Related to Figure 2

<b>Gene</b>	<b>Chrom</b>	<b>STR position (hg19)</b>	<b>Motif</b>	<b>Top CAVIAR score</b>	<b>Top tissue</b>
<i>DHRS1</i>	chr14	24769851	CCT	1.00	Thyroid ( $\beta=-0.42$ )
<i>AGTPBP1</i>	chr9	88356816	CCG	1.00	Cells-Transfor medfibroblasts ( $\beta=0.36$ )
<i>RPL14</i>	chr3	40503522	CTG	0.99	Nerve-Tibial ( $\beta=-0.32$ )
<i>SMPD1</i>	chr11	6411932	CGCTGG	0.79	Thyroid ( $\beta=0.27$ )
<i>POMC</i>	chr2	25384461	AGC	0.59	Muscle-Skelet al ( $\beta=0.35$ )
<i>HSPBP1</i>	chr19	55790888	CGG	0.58	Cells-Transfor medfibroblasts ( $\beta=-0.37$ )
<i>ROBO3</i>	chr11	124750441	AGCCGG	0.39	Thyroid ( $\beta=0.33$ )
<i>MAGEF1</i>	chr3	184429135	AGG	0.38	Thyroid ( $\beta=-0.33$ )
<i>PRR12A</i>	chr19	50093219	ACCCCC	0.36	Skin-SunExpo sed ( $\beta=0.51$ )
<i>EPOR</i>	chr19	11558342	CCT	0.29	Heart-LeftVent ricle ( $\beta=0.28$ )
<i>TOM1L2</i>	chr17	17697095	CTG	0.22	Lung ( $\beta=0.28$ )
<i>IST1</i>	chr16	71956508	ATGCCC	0.20	Artery-Aorta ( $\beta=0.27$ )
<i>MRPS26</i>	chr20	3026347	CCCCG	0.17	Nerve-Tibial ( $\beta=0.32$ )
<i>ZNF219</i>	chr14	21560753	AGCCTC	0.14	Adipose-Subcu taneous ( $\beta=0.22$ )

<i>DLX5</i>	chr7	96635365	CTG	0.12	Muscle-Skeletal ( $\beta=-0.21$ )
<i>MCM3AP</i>	chr21	47721987	ACC	0.11	Artery-Aorta ( $\beta=0.46$ )

**FM-eSTRs overlapping protein-coding regions.** The CAVIAR score and eSTR effect size ( $\beta$ ) are given for the tissue with highest CAVIAR score. FM-eSTRs are ordered based on the CAVIAR score.

**Table S6.** Related to Figure 2

Canonical motif	# STRs			eSTR enrichment		FM-eSTR enrichment	
	All	eSTRs	FM-eSTRs	P-val	OR	P-val	OR
CCCGG	13	7	3	2.4E-04	9.0	1.7E-03	15.5
CCCCG	41	18	9	1.8E-07	6.0	6.4E-08	14.6
CCCCCG	24	11	4	2.7E-05	6.5	1.0E-03	10.3
AGGCGG	18	8	3	4.6E-04	6.2	4.5E-03	10.3
AGCCCC	35	11	5	1.4E-03	3.5	5.0E-04	8.6
CCCCGG	23	11	3	1.6E-05	7.1	9.1E-03	7.8
CCG	210	66	22	1.0E-14	3.5	1.4E-10	6.1
AAAAG	213	39	11	3.4E-03	1.7	2.8E-03	2.8
AGC	293	55	10	3.1E-04	1.8	7.9E-02	1.8
AATC	271	48	9	2.2E-03	1.7	1.1E-01	1.8
C	1,163	165	38	4.8E-03	1.3	1.6E-03	1.8
AATT	268	50	8	5.3E-04	1.8	1.8E-01	1.6
ACC	250	32	7	4.9E-01	1.1	2.5E-01	1.5
AAAG	1,854	261	49	6.5E-04	1.3	2.6E-02	1.4
AAAT	8,806	1,306	218	2.5E-22	1.4	8.5E-05	1.3
AAAAC	3,289	485	77	1.2E-08	1.3	7.0E-02	1.2
AGAGGG	220	46	5	7.2E-05	2.0	6.2E-01	1.2
AAAC	7,779	1,158	173	2.1E-20	1.4	3.4E-02	1.2
AAAAG	491	67	11	1.4E-01	1.2	5.1E-01	1.2
ACAG	187	23	4	7.3E-01	1.1	7.8E-01	1.1
AGG	481	83	10	1.8E-04	1.6	7.4E-01	1.1
A	64,118	7,922	1,276	1.0E-17	1.1	3.4E-02	1.1
AAAAT	1,738	267	33	1.0E-06	1.4	1.0E+00	1.0
AAC	3,347	501	61	8.3E-10	1.4	8.0E-01	1.0

AC	48,732	4,389	841	4.2E-95	0.7	9.6E-04	0.9
AAGG	1,227	103	20	4.5E-04	0.7	6.0E-01	0.9
AATG	1,274	170	20	4.2E-02	1.2	4.7E-01	0.8
ATCC	930	79	14	3.3E-03	0.7	4.7E-01	0.8
AAAAAC	1,018	156	15	2.1E-04	1.4	4.2E-01	0.8
AAT	2,915	362	43	1.1E-01	1.1	1.0E-01	0.8
AGGG	348	39	5	9.3E-01	1.0	6.9E-01	0.8
AAAAAT	498	71	7	5.7E-02	1.3	5.1E-01	0.7
ACAT	656	91	9	5.7E-02	1.2	3.9E-01	0.7
AT	7,950	703	103	5.2E-15	0.7	2.5E-05	0.7
ATC	725	69	9	1.0E-01	0.8	2.2E-01	0.6
AGAT	3,461	187	42	1.4E-35	0.4	2.0E-03	0.6
AG	6,552	565	80	1.5E-14	0.7	1.7E-05	0.6
AAG	395	47	4	8.1E-01	1.0	2.6E-01	0.5

**Enrichment of STR motifs in eSTRs.** For each motif, a Fisher's exact test was performed to test whether eSTRs were significantly more or less likely to have each motif sequence compared to all STRs analyzed. FM-eSTRs represent all eSTRs with CAVIAR score of at least 10% as defined in the main text. OR=odds ratio and p-values are two-sided. Motifs give the canonicalized repeat sequence (**Methods**).

**Table S7.** Related to Figure 4

Motif	Number of FM-eSTRs
AACCCC/GGGGTT	1/1
AACCCT/AGGGTT	1/1
ACCCCC/GGGGGT	1/0
AGAGGG/CCCTCT	2/3
AGCCCC/CTGGGG	2/4
AGCCCG/CGGGCT	1/1
AGGG/CCCT	2/3
ATGCCC/ATGGGC	1/0
C/G	20/20
CCCCCG/CGGGGG	1/3
CCCCG/CGGGG	4/6
AGGGG/CCCCT	0/1
CCCG/CGGG	2/0
CCCGG/CCGGG	2/1
AGGGC/CCCTG	0/1
AGGGCC/CCCTGG	0/1
AAGGG/CCCTT	0/1
AGGCC/CCTGGG	0/1
AGCCCG/CGGGCT	0/1
AGCCC/CTGGG	0/1
ACCCC/GGGGT	0/1

**STR motifs with the ability to form G4 quadruplexes.** STR motifs matching the canonical G4 motif ( $G_3N_{1-7}G_3N_{1-7}G_3N_{1-7}G_3$ ). For each motif, the first sequence gives the canonical form of the motif and the second gives the reverse complement of the canonical form.

**Table S8.** Related to Figure 5

STR locus (hg19)	Trait	Motif	Lead SNP	SNP-STR LD ( $r^2$ )	Gene	Best CAVIAR score
1:23692212	Glomerular filtration rate	AATT	rs4525087	0.13	<i>ZNF436</i>	0.39 (Lung)
1:150314982	Eye morphology	AC	rs201983168	0.13	<i>APH1A</i>	0.13 (Cells-TransformedFibroblasts)
4:187139784	Blood protein levels	AC	rs12331051	0.09	<i>AC110771.1</i>	0.19 (Skin-NotSunExposed)
5:150848654	Basophil percentage of granulocytes	A	rs10709981	0.71	<i>SLC36A1</i>	0.43 (Skin-SunExposed)
6:42921475	Blood protein levels	AT	rs57736976	0.76	<i>GNMT</i> <i>PEX6</i>	0.13 (Adipose-Subcutaneous) 0.19 (Adipose-Subcutaneous)
8:144839147	Schizophrenia	A	rs11784536	0.71	<i>NRBP2</i>	0.15 (Thyroid)
9:2622147	Hematocrit, Hemoglobin concentration	CCG	rs369552432	0.15	<i>VLDLR</i>	0.20 (WholeBlood)
11:5297399	Red cell distribution width	AC	rs112176573	0.05	<i>HBG1</i>	0.12 (Adipose-Visceral)
11:11907700	Platelet count	CCG	rs7108857	0.16	<i>NLRX1</i>	0.19 (Cells-Transformedfibroblasts)
12:50480263	Achilles tendinopathy	ACC	rs57224706	0.72	<i>GPD1</i>	0.16 (Artery-Tibial)
16:28871190	Red blood cell count	C	rs12448902	0.78	<i>NFATC2IP</i> <i>TUFM</i>	0.13 (Artery-Aorta) 0.15 (Adipose-Subcutaneous)
19:7765393	Blood protein levels	AAAT	rs113893384	0.99	<i>FCER2</i>	0.13 (Lung)
19:10338986	Immature fraction of reticulocytes	AAAAC	rs140658666	1.00	<i>MRPL4</i>	0.13 (Skin-NotSunExposed)
19:45413225	Blood protein levels Free cholesterol levels in large LDL LDL cholesterol Serum total	C	rs1065853	0.10	<i>APOC1</i>	0.11 (Esophagus-Mucosa)



	cholesterol Total cholesterol in LDL					
19:52130149	Blood protein levels	AAAAC	rs147859411	1.0	<i>SIGLEC14</i> <i>SIGLEC5</i>	0.48 (Skin-SunExposed) 0.18 (WholeBlood)

**GWAS signals from the NHGRI/EBI GWAS catalog falling within FM-eSTRs.** LD was computed as the squared Pearson correlation between STR dosage and SNP genotype (0, 1, or 2 reference alleles) for each SNP. The CAVIAR score for each FM-eSTR is shown for highest-scoring tissue.

**Table S9.** Related to Figures 5 and 6

Trait	Gene	STR (hg19)	Motif	Lead SNP	SNP-STR LD ( $r^2$ )	Top CAVIAR score	Coloc
Height	<i>RFT1</i>	3:53128363	AC	rs2336725	0.85	0.70 (Artery-Aorta)	99.0%
Height	<i>FADS1</i>	11:61620629	AAC	rs174547	0.56	0.32 (Brain-Cerebellum)	97.7%
Height	<i>LUZP1</i>	1:23553614	AAC	rs1738475	0.22	0.48 (Adipose-Subcutaneous)	93.1%
Height	<i>PTPRCAP</i>	11:67238583	AAT	rs12795957	0.04	0.11 (Artery-Aorta)	90.5%
Height	<i>UBE2Z</i>	17:46965665	AGAT	rs318095	0.37	0.13 (Muscle-Skeletal)	79.0%
Height	<i>SLCO1C1</i>	12:20856340	AAC	rs10770705	0.79	0.16 (Skin-NotSun Exposed)	76.7%
Height	<i>ADAMTS3</i>	4:73471388	AG	rs16848425	0.0058	0.16 (Cells-TransformedFibroblasts)	58.1%
Schizophrenia	<i>PRR12</i>	19:50110619	AC	rs56873913	0.14	0.99 (Thyroid)	98.2%
Schizophrenia	<i>MED19</i>	11:57523883	AC	rs9420	0.63	0.47 (Adipose-Subcutaneous)	90.1%
Schizophrenia	<i>CEBPZ</i>	2:37526655	AC	rs2372993	0.74	0.28 (Lung)	86.7%
Schizophrenia	<i>FAM134A</i>	2:220082366	A	rs6707588	0.01	0.12 (Esophagus-Mucosa)	79.8%
Schizophrenia	<i>TM6SF2</i>	19:19424949	AAAT	rs2905424	0.82	0.11 (Adipose-Visceral)	77.3%
Schizophrenia	<i>TMEM81</i>	1:205076171	AC	rs16937	0.43	0.13 (Nerve-Tibial)	63.3%
Schizophrenia	<i>CCDC126</i>	7:23615272	A	rs227932	0.80	0.12 (Thyroid)	60.2%
Schizophrenia	<i>SPIRE2</i>	16:89829656	A	rs12449000	0.69	0.69 (Skin-NotSun Exposed)	58.7%
Schizophrenia	<i>VARS2</i>	6:30884231	AC	rs111782145	0.28	0.12 (Brain-Cerebellum)	54.1%

Schizophrenia	<i>UBAP2L</i>	1:154251259	A	rs7521837	0.95	0.11 (Esophagus- Muscularis)	54.1%
---------------	---------------	-------------	---	-----------	------	------------------------------------	-------

**FM-eSTRs co-located with height** (Yengo et al., 2018) **and schizophrenia** (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014) **GWAS signals**. LD was computed as the squared Pearson correlation between STR dosage and SNP genotype (0, 1, or 2 reference alleles) for each SNP. The CAVIAR score for each FM-eSTR is shown for highest-scoring tissue. Only FM-eSTRs that had the highest CAVIAR score for each gene are shown. Coloc column shows the posterior probability of a model where both gene expression and the GWAS trait are driven by the same causal variant.

**Table S10.** Related to Figure 6

Sequence ID	Description	Sequence
RFT1_0rpt	Synthetic construct for <i>RFT1</i> locus with 0 AC repeats.	AGCTCGCTAGCCTCGAGGATACCAGCCTGGACAACATAG TGAGACCATGTCTCTACCAAAAAGAAAAAGAAGGCAAG CCACTGCATCCCCCAAATGCCCAGGTATAGGTCCTTGGC TGATTCAGGCCAAGGTGCTGTGTGCTGTGGAGATACGCT TTATCAAAGTGGGTGGTAAAAATACTACACGCTCTCACAC ACAATACTATACTCCAACAGGAGAGATGTGCTTTCATCCG GAGTTCAGAGAGGGAGTTTATTTGCTTTTCAGTCCAGA GAGAGTCTCTTTAATGAAGTACGAGAGCCTGTACCTCA CAGGGGAAGTTTTTTCACTACTAATCAGTGATCAGAAAG GTCAACTTCAATCAAGATCTGGCCTCGGCG
RFT1_5rpt	Synthetic construct for <i>RFT1</i> locus with 5 AC repeats.	AGCTCGCTAGCCTCGAGGATACCAGCCTGGACAACATAG TGAGACCATGTCTCTACCAAAAAGAAAAAGAAGGCAAG CCACTGCATCCCCCAAATGCCCAGGTATAGGTCCTTGGC TGATTCAGGCCAAGGTGCTGTGTGCTGTGGAGATACGCT TTATCAAAGTGGGTGGTAAAAATACTACACGCACACACA CACTCTCACACACAATACTATACTCCAACAGGAGAGATG TGCTTTCATCCGGAGTTCCAGAGAGGGAGTTTATTTGCTT TTCAGTCCAGAGAGAGTCTCTTTAATGAAGTACGAGAG CCTGTACCTCACAGGGGAAGTTTTTTCACTACTAATCAGT GATCAGAAAGGTCAACTTCAATCAAGATCTGGCCTCGGC G
RFT1_10rpt	Synthetic construct for <i>RFT1</i> locus with 10 AC repeats.	AGCTCGCTAGCCTCGAGGATACCAGCCTGGACAACATAG TGAGACCATGTCTCTACCAAAAAGAAAAAGAAGGCAAG CCACTGCATCCCCCAAATGCCCAGGTATAGGTCCTTGGC TGATTCAGGCCAAGGTGCTGTGTGCTGTGGAGATACGCT TTATCAAAGTGGGTGGTAAAAATACTACACGCACACACA CACACACACACACTCTCACACACAATACTATACTCCAAC AGGAGAGATGTGCTTTCATCCGGAGTTCCAGAGAGGGGA GTTTATTTGCTTTTCAGTCCAGAGAGAGTCTCTTTAATG AAGTACGAGAGCCTGTACCTCACAGGGGAAGTTTTTTC CTACTAATCAGTGATCAGAAAGGTCAACTTCAATCAAGAT CTGGCCTCGGCG
RFT1eSTR_F	Primer for amplifying <i>RFT1</i> locus from genomic DNA (forward)	AGCTCGCTAGCCTCGAGGATACCAGCCTGGACAACATA GTG
RFT1eSTR_R	Primer for amplifying <i>RFT1</i> locus from genomic DNA (reverse)	CGCCGAGGCCAGATCTTGATTGAAGTTGACCTTTCTGA TCACTG
CRE	Positive control	AGCTCGCTAGCCTCGAGGATGCACCAGACAGTGACGTC AGCTGCCAGATCCCATGGCCGTCACTACTGTGACGTCTTT

		CAGACACCCCATTGACGTCAATGGGAGAACATCAAGATC TGGCCTCGGCG
--	--	--------------------------------------------------------

**Sequences used in reporter assay analyses.** Blue sequences show homology arms used for cloning into pGL4.27 firefly luciferase plasmid. Red shows the STR region that was modified. Black shows additional genomic context.