

Supplementary Table 12: Detailed documentation of annotation categories used for enrichment analyses of eQTNs.

Annotation	Abbreviation	Source/URL	Assembly	Citation	Comment
Exons of short non-coding RNAs					
Short RNAs	miRNA precursors, snoRNAs(C/D box and H/ACA box snoRNA)	UCSC table browser (track: sno/miRNA)	GRCh37/hg19	[1, 2, 3]	-
long non-coding RNAs					
Long non-coding RNAs(lncRNAs)	Gencode v13	ftp://ftp.sanger.ac.uk/pub/gencode/release_13/gencode.v13.long_noncoding_RNAs.gtf.gz	GRCh37/hg19	Gencode [4]	The original set of long non-coding RNAs as annotated in Gencode was reduced to a set of <i>bona fide</i> non-coding RNAs without any evidence for conserved ORFs (RNAcode [5]) nor exhibiting any sequence similarity to known amino acid sequences (RefSeq Mar 7 2012); categorized into exons and introns.
Large intergenic non-coding RNAs (lincRNAs)	Cabili et al.	http://www.broadinstitute.org/genome_bio/human_lincrnas/sites/default/files/lincRNA_catalog/lincRNAs_transcripts_stringentSet.bed	GRCh37/hg19	[6]	Stringent set; categorized into exons and introns.
LncRNadb	lncRNadb	http://lncrnadb.com	-	[7]	Coordinates in GRCh37/hg19 have been derived by BLAT [8] with paramters <code>-trimHardA -minIdentity=95</code> .
RNA transcripts found to be contained in chromatin					
Transcript located in intergenic space	CAR	-	NCBI36/hg18	[9]	Mapped to GRCh37/hg19 using liftOver [3].
Transcript located in introns of protein-coding genes	CAR	-	NCBI36/hg18	[9]	Mapped to GRCh37/hg19 using liftOver [3].
miRNAs target sites					
conserved miRNA target sites	TsmiRNA	UCSC table browser (track:TS site)	NCBI37/hg19	[10]	-
Bona fide non-coding RNAs regulated in cell cycle, TP53 pathway or STAT3 pathway					

Annotation	Abbreviation	Source/URL	Assembly	Citation	Comment
Cell cycle	CC	-	hg19	<i>Hackermüller et al.</i> ¹	Categorized according to transcript location, transcripts in intergenic space and transcript located in introns of protein-coding genes. Reduced to a set of <i>bona fide</i> non-coding RNAs without any evidence for conserved ORFs (RNAcode [5]) nor exhibiting any sequence similarity to known amino acid sequences (RefSeq Mar 7 2012).
Tumor protein 53 gene	TP53	-	hg19	<i>Hackermüller et al.</i> ¹	Categorized according to transcript location, transcripts in intergenic space and transcript located in introns of protein-coding genes. Reduced to a set of <i>bona fide</i> non-coding RNAs without any evidence for conserved ORFs (RNAcode [5]) nor exhibiting any sequence similarity to known amino acid sequences (RefSeq Mar 7 2012).
Signal transducer and activator of transcription 3	STAT3	-	hg19	<i>Hackermüller et al.</i> ¹	Categorized according to transcript location, transcripts in intergenic space and transcript located in introns of protein-coding genes. Reduced to a set of <i>bona fide</i> non-coding RNAs without any evidence for conserved ORFs (RNAcode [5]) nor exhibiting any sequence similarity to known amino acid sequences (RefSeq Mar 7 2012).
Bona fide non-coding genomic regions predicted to contain conserved secondary structure motifs					
RNAz	RNAz	-	GRCh37/hg19	<i>Smith et al.</i> ²	Categorized according to intronic and intergenic space.
SISSIz	SISSIz	-	GRCh37/hg19	<i>Smith et al.</i> ²	Categorized according to intronic and intergenic space.
Protein-coding gene annotation					
Coding exons	CDS	Gencode v13	GRCh37/hg19	Gencode [4]	-

¹Hackermüller J et al. "Cell cycle, oncogenic and tumor suppressor pathways regulate numerous long and macro RNAs." submitted.

²Smith MA, Gesell T, Stadler PF, Mattick JS. "Widespread Purifying Selection on RNA Structure in Mammals". submitted

Annotation	Abbreviation	Source/URL	Assembly	Citation	Comment
Introns	intron	Gencode v13	GRCh37/hg19	Gencode [4]	Defined as intronic nucleotides which do not overlap any exon of a protein-coding transcript annotation.
Intergenic	intergenic	Gencode v13	GRCh37/hg19	Gencode [4]	Defined as the complement of all known protein-coding transcript annotation.
UTRs	UTRs	Gencode v13	GRCh37/hg19	Gencode [4]	-
Pseudogenes					
Pseudogenes	Gencode v13		GRCh37/hg19	-	-
Novel transcripts with putative coding function					
Transcripts of uncertain coding potential	TUCP	http://www.broadinstitute.org/genome_bio/human_lincrnas/sites/default/files/TUCP_transcripts_catalog/TUCP_transcripts.gtf	GRCh37/hg19	[6]	Categorized according to intronic and intergenic space.
RNAcode	RNAcode	-	GRCh37/hg19	[11, 12, 13, 14]	Predicted ORF in intergenic space.
Regulatory sites					
CpG islands	CpG islands	UCSC table browser (track: CpG islands)	NCBI37/hg19	[10]	-
Most conserved sequences	MCS	UCSC table browser (track: MCS)	NCBI37/hg19	[10]	-
Open source for Regulatory Annotation	OregAnno	UCSC table browser (track: OregAnno)	NCBI37/hg19	[10]	-
Promoter regions			GRCh37/hg19		Defined as 2kb and 5kb upstream of 5'UTR.
H3K4 trimethylation	H3K4me3	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg18/encodeDCC/wgEncodeBroadChIPSeq/	NCBI36/hg18	[15, 3]	Chromatin-mark associated with promoter sites [16, 17]. Mapped to GRCh37/hg19 using <code>liftOver</code> [3].
DNaseI-hypersensitive sites	DNaseI	UCSC table browser (track: Dnase Clusters)	GRCh37/hg19	[15, 3]	Associated with transcription factor binding sites [17, 18].
Transcription factor binding sites (TFBs)	TFBs (Encode)	UCSC table browser (track: Txn Factor ChIP)	GRCh37/hg19	[15, 3]	Binding sites identified by ChIP-seq [15].
Transcription factor binding sites (TFBs)	TFBs (Transfac)	UCSC table browser (track: TFBS Conserved)	GRCh37/hg19	[19, 3]	Binding sites conserved in human, mouse and rat from Transfac Matrix Database (v7.0) [19].

Annotation	Abbreviation	Source/URL	Assembly	Citation	Comment
Pol-II binding sites	Pol-II	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg18/encodeDCC/wgEncodeBroadChIPseq/	NCBI36/hg18	[15, 3]	Pol-II binding sites derived by ChIP-seq [15]. Mapped to GRCh37/hg19 using liftOver [3].
H3K36 trimethylation	H3K36me3	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg18/encodeDCC/wgEncodeBroadChIPseq/	NCBI36/hg18	[15, 3]	Chromatin-mark associated with active regions of Pol-II transcripts [20]. Mapped to GRCh37/hg19 using liftOver [3].
H3K27 trimethylation	H3K27me3	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg18/encodeDCC/wgEncodeBroadChIPseq/	NCBI36/hg18	[15, 3]	Chromatin-mark associated with repressed regions of Pol-II transcripts [21]. Mapped to GRCh37/hg19 using liftOver [3].
H3K4 monomethylation	H3K4me1	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg18/encodeDCC/wgEncodeBroadChIPseq/	NCBI36/hg18	[15, 3]	Chromatin-mark associated with enhancer regions [22, 17]. Mapped to GRCh37/hg19 using liftOver [3].
H3K27 acetylation	H3K27ac	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg18/encodeDCC/wgEncodeBroadChIPseq/	NCBI36/hg18	[15, 3]	Chromatin-mark associated with enhancer and promoter sites [15, 23, 24]. Mapped to GRCh37/hg19 using liftOver [3].

References

- [1] Lestrade L, Weber MJ (2006) snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res* 34: D158–D162.
- [2] Griffiths-Jones S (2004) The microRNA Registry. *Nucleic Acids Res* 32: D109–D111.
- [3] Kuhn RM, Haussler D, Kent WJ (2012) The UCSC genome browser and associated tools. *Brief Bioinformatics* .
- [4] Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, et al. (2012) GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* 22: 1760–1774.
- [5] Washietl S, Findeiss S, Miller SA, Kalkhof S, von Bergen M, et al. (2011) RNAcode: robust discrimination of coding and noncoding regions in comparative sequence data. *RNA* 17: 578–594.
- [6] Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, et al. (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* 25: 1915–1927.
- [7] Amaral P, Clark M, Gascoigne D, Dinger M, Mattick J (2011) Incrnadb: a reference database for long noncoding rnas. *Nucleic Acids Res* 39: D146–D151.
- [8] Kent WJ (2002) BLAT—the BLAST-like alignment tool. *Genome Res* 12: 656–664.
- [9] Mondal T, Rasmussen M, Pandey GK, Isaksson A, Kanduri C (2010) Characterization of the rna content of chromatin. *Genome Res* 20: 899–907.
- [10] Meyer LR, Zweig AS, Hinrichs AS, Karolchik D, Kuhn RM, et al. (2012) The UCSC genome browser database: extensions and updates 2013. *Nucleic Acids Res* .
- [11] Gruber AR, Neubck R, Hofacker IL, Washietl S (2007) The RNAz web server: prediction of thermodynamically stable and evolutionarily conserved RNA structures. *Nucleic Acids Res* 35: W335–W338.
- [12] Gruber AR, Bernhart SH, Hofacker IL, Washietl S (2008) Strategies for measuring evolutionary conservation of RNA secondary structures. *BMC Bioinformatics* 9: 122.
- [13] Gruber A, Findeiss S, Washietl S, Hofacker I, Stadler P (2010) RNAz 2.0: Improved noncoding RNA detection. In: PSB in press.
- [14] Gruber AR, Findei S, Washietl S, Hofacker IL, Stadler PF (2010) RNAZ 2.0: IMPROVED NON-CODING RNA DETECTION. *Pac Symp Biocomput* 15: 69–79.
- [15] Birney E, Stamatoyannopoulos J, Dutta A, Guigo R, Gingeras T, et al. (2007) Identification and analysis of functional elements in 1by the encode pilot project. *Nature* 447: 799–816.
- [16] Roh TY, Cuddapah S, Cui K, Zhao K (2006) The genomic landscape of histone modifications in human T cells. *Proc Natl Acad Sci U S A* 103: 15782–15787.
- [17] Bulger M, Groudine M (2010) Enhancers: the abundance and function of regulatory sequences beyond promoters. *Dev Biol* 339: 250–257.
- [18] Xi H, Shulha HP, Lin JM, Vales TR, Fu Y, et al. (2007) Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS Genet* 3: e136.

- [19] Wingender E, Chen X, Hehl R, Karas H, Liebich I, et al. (2000) TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res* 28: 316–319.
- [20] Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, et al. (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448: 553–560.
- [21] Barski A, Cuddapah S, Cui K, Roh TY, Schonnes DE, et al. (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129: 823–837.
- [22] Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, et al. (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 39: 311–318.
- [23] Terrenoire E, McRonald F, Halsall JA, Page P, Illingworth RS, et al. (2010) Immunostaining of modified histones defines high-level features of the human metaphase epigenome. *Genome Biol* 11: R110.
- [24] Shin JH, Li RW, Gao Y, Baldwin R 6th, Li Cj (2012) Genome-wide ChIP-seq mapping and analysis reveal butyrate-induced acetylation of H3K9 and H3K27 correlated with transcription activity in bovine cells. *Funct Integr Genomics* 12: 119–130.