

Preferential Binding of a stable G3BP Ribonucleoprotein Complex to Intron-retaining Transcripts in Mouse Brain and Modulation of their Expression in the Cerebellum

Sophie Martin, Nicolas Bellora, Juan González-Vallinas, Manuel Irimia, Karim Chebli, Marion DeToledo, Monika Raabe, Eduardo Eyra, Henning Urlaub, Ben J. Blencowe and Jamal Tazi

Supplemental Figures Legends

Figure S1.

(A) G3BP1 IP was performed with a different G3BP1 antibody, which also revealed additional proteins immunoprecipitated under CLIP conditions, as seen in a silver staining of the SDS-PAGE gel. (B) G3BP1 and G3BP2 colocalize in primary hippocampal neurons (i), as well as in stress granules formed under arsenite treatment (ii). DNA is counterstained with Hoechst staining. *Scale bar represents 5 μ m.* (C) Different highly stringent washes did not permit to eliminate the presence of the G3BP partners and the separation of three ribonucleoproteic complexes under high RNase treatment (a). The numbers 1 to 4 are indicative of the four complexes detailed in Figure 2A and B. In (b), different RNase concentrations in normal CLIP washes conditions show G3BP-complex gradually shifting to high molecular weight as RNase treatment decreases.

Figure S2.

(A) Long motif over-represented in G3BP-complex clusters, part of a SINE-Alu-B1 transposable element. (B) Around 8 % of the clusters possess this sequence with an

occurrence >0, with a peak in the center of the clusters. When looking for motifs over-represented in G3BP-complex clusters compared to background which are not part of repeated sequences, several sequences are found, less enriched compared to the logo in Figure S2A, and less centered. (C) (1) Consensus from two 7mers: CACTCTG + GACTCTG. This motif is still part of Alu repeat elements, and may be missed by the RepeatMasker tool. (2) Consensus from two 7-mers: CCCTCCC + CCCACCC. (3) Consensus binding motif identified for G3BP2 by RNAcompete. (4) Center of consensus binding motif identified for G3BP1 by SELEX.

Figure S3. HITS-CLIP of G3BP2 in G3BP1 KO mouse brain.

G3BP CLIP in G3BP1 KO brain, leading to G3BP2a immunoprecipitation, revealed RBP: RNAs complexes on the autoradiogram, like in WT (A) Three complexes are also revealed with high (+++) RNase treatment, with the complex around 72kDa corresponding to G3BP2a:RNAs as revealed in Figure 1. (B) Repartition of the clusters identified after sequencing of RNAs extracted from the low RNase treated-G3BP2a complex ((a) in 2.a). Again, a majority of clusters were mapped into intronic regions. Overall, the distribution of clusters into the coding and non-coding transcripts reflects the distribution of G3BP clusters in WT.

Figure S4.

(A) Distribution of G3BP HITS-CLIP clusters within introns along the RNA, from 5' to 3', for G3BP1 and G3BP2, or G3BP1 only. The position is 5' biased (although not at

the most 5' decil, presumably because the first intron of a transcript is often very long). (B) Expression levels of genes with G3BPs clusters compared to expression levels of the genes in neural or non-neural tissues determined from RNA-Seq experiments (“Neural”: 6 RNAseq samples: 2 of whole brain, one of cerebellum, one retina, and isolated neurons (cerebellar granular neurons, neurons from Dorsal root ganglia), and non-neural tissues: “Rest”: 14 samples: 2 of Liver, 2 kidney, 2 Heart, Muscle, Myoblast 168h of differentiation, T cells, Testis, Lung fibroblasts, 3T3 cell line and 2x embryonic fibroblasts). 5 groups of expression were made from these RNA-Seq data, based on average corrected cRPKMs (RPKMs with some extra corrections): (i) (nearly) no expression (av. cRPKM<2), (ii) low expression (2-10), (iii) moderate expression (10-25), (iv) high expression (25-50), (v) very high expression (>50).

Figure S5.

Consensus sequences analysis in introns. A CTG-containing 8-mer motif was identified in at least 20 % of the intronic clusters (A), and a C-rich motif in at least 10 % of the clusters (B). Top graphs represent the relative abundance of the motif in the clusters according to the genomic position in the cluster, and the logo (obtained with MEME) is represented at the bottom. The clusters were also scanned for enriched heptamer motifs (C and D) as well as for the consensus of SFPQ, partner in the complex (E). The latter was not significantly enriched. On the contrary, by scanning the clusters for other RBP binding sequences, we found an enrichment of PCBP consensus binding motif (F). *Red line: whole HITS-CLIP clusters dataset filtered out*

of repeats (RepeatMasker); Black line: intronic clusters only; hashed red line: enrichment of the reverse complement sequence of the motif in the whole dataset; hashed black line: enrichment of the reverse complement sequence of the motif in the intronic clusters dataset.

Figure S6.

Images from UCSC genome browser showing the structure of the gene and the distribution of the clusters for representative genes of G3BP-complex HITS-CLIP. The sequences are represented either from 5' to 3' extremities or from 3' to 5' of the pre-mature transcripts, depending on the sense of transcription (indicated by the arrows from 5' to 3'). The numbers indicate exons numbers as attributed in NCBI database. The horizontal lines represent the introns. Several alternative transcripts are encoded by the genes. (A) Five transcripts that were studied more extensively by PCRs targeting different regions of the transcripts. (B) Structure of the *Grm5* and *Grm1* genes and distribution of clusters.

Figure S7. The G3BP-complex regulates immature transcripts in the cerebellum.

(A) G3BP HITS-CLIP performed in cerebellum. Overall, the distribution of clusters in protein-coding (left), as well as the different non-coding transcripts (right) reflects the distribution in the whole brain. (B) Expression of the different proteins of the G3BP-complex in the cerebrum and the cerebellum from WT and G3BP1 KO mouse. Total

homogenates were prepared from cerebrum and cerebellum of each genotype, and equal amounts of each protein lysates (confirmed by loading control actin) were analyzed by SDS-polyacrylamide gel electrophoresis, followed by immunoblotting. The Western blot is representative of three independent experiments with different mice of each genotype, n = 3.

Supplemental Methods

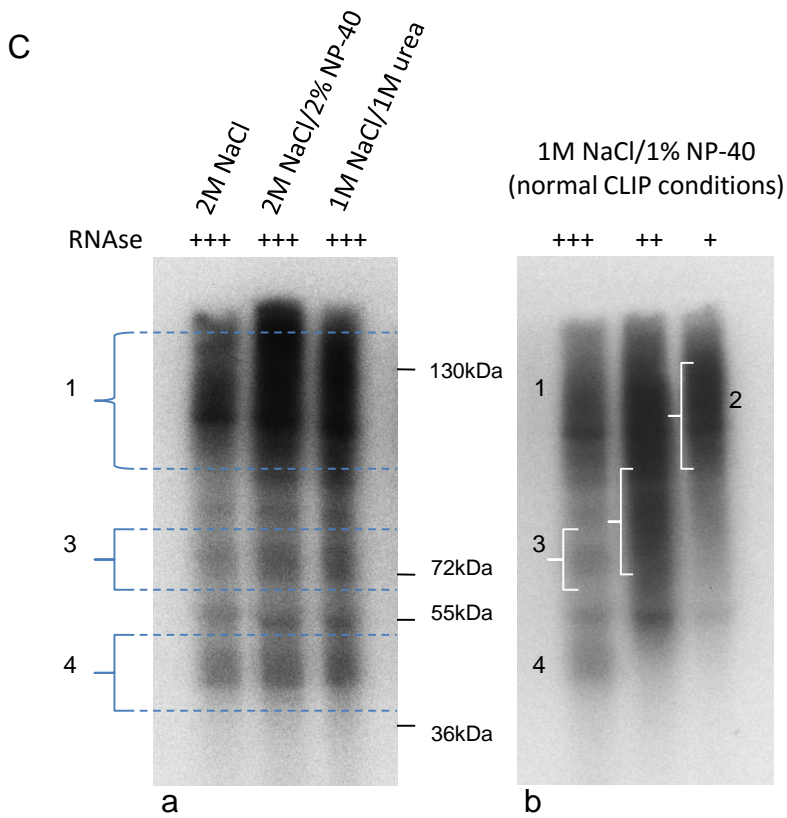
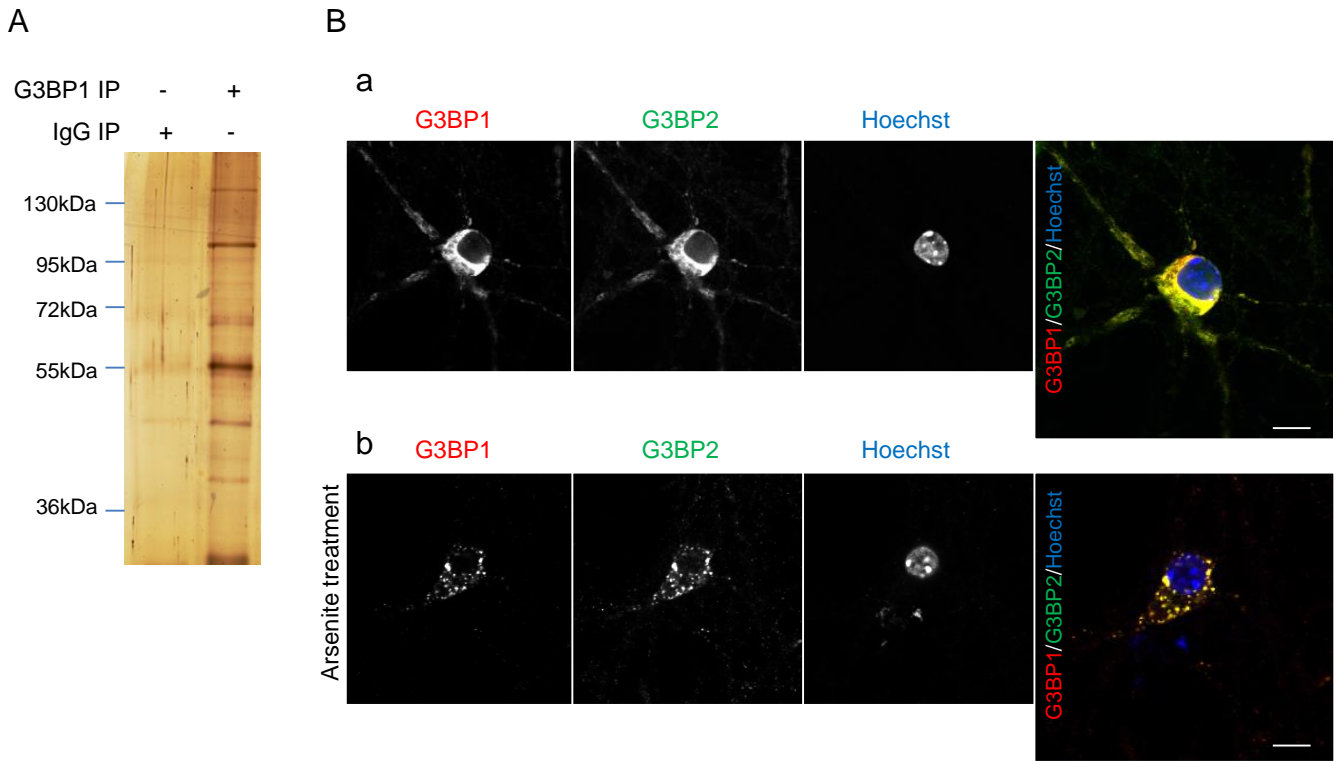
Evaluation of G3BP-complex binding motifs

Motif search was carried out in the reproduced transcripts from WT sample HITS-CLIP, using k-mer and MEME (Multiple Em for Motif Elicitation) approaches. A k-mer enrichment analysis was performed by calculating the ratio of the density of k-mers in G3BP-complex HITS-CLIP clusters versus the density in a background set of clusters (Agirre *et al.* 2015). The analysis was performed with the whole set of G3BP-complex clusters, or the set of intronic clusters only. To define the control background sets, we selected random regions, but of the same size as the clusters and obtained from the same gene regions, adjacent to the clusters (to avoid sequence composition biases from different parts of the genes). Concerning the introns, these random clusters were generated on the same introns. Furthermore, we checked that the G3BP-complex clusters were randomly distributed along the introns, which avoids any sequence bias due to the position, like polypyrimidine tracks. Ranked by p-values, the top k-mers were retrieved and they were sorted relative to their abundance in G3BP-complex clusters. The clusters were anchored and aligned on the position of k-mers in order to obtain sequence logos with MEME. Furthermore, we eliminated the clusters that were overlapping repeated and transposable elements. Those were obtained from the RepeatMasker-recovered annotations in NCBI UCSC browser. Around 32 % of the clusters were thus removed (mostly SINE and LINE sequences), however the RepeatMasker tool may miss some degraded ALU sequences that we were still able to detect. Finally, the clusters were scanned to detect the presence of the motifs which were present in a minimum of 10 or 20 % of the sequences, and plot

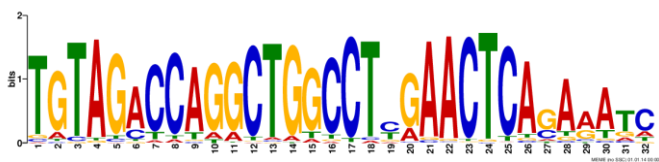
the graphs of the relative abundance of the motif in the G3BP-complex clusters datasets. We also scanned the clusters for the presence of previously identified consensus sequences for G3BP1 (SELEX motif (Tourrière *et al.* 2001)), G3BP2 (RNAcompete (Ray *et al.* 2013)), and SFPQ (RNAcompete).

Supplemental References

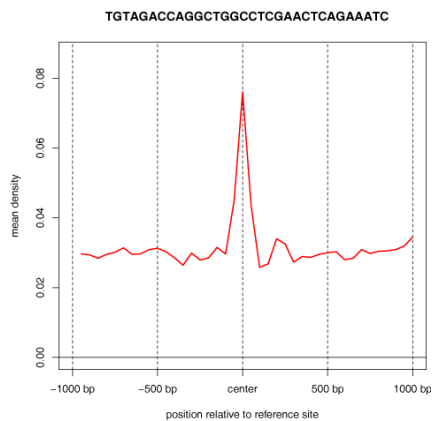
- Agirre E., Bellora N., Alló M., Pagès A., Bertucci P., Kornblihtt A. R., Eyrales E. (2015) A chromatin code for alternative splicing involving a putative association between CTCF and HP1 α proteins. *BMC Biol.* **13**, 31.
- Ray D., Kazan H., Cook K. B., Weirauch M. T., Najafabadi H. S., Li X., Gueroussov S., et al. (2013) A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**, 172–177.
- Tourrière H., Gallouzi I. E., Chebli K., Capony J. P., Mouaikel J., van der Geer P., Tazi J. (2001) RasGAP-associated endoribonuclease G3Bp: selective RNA degradation and phosphorylation-dependent localization. *Mol. Cell. Biol.* **21**, 7747–7760.



A

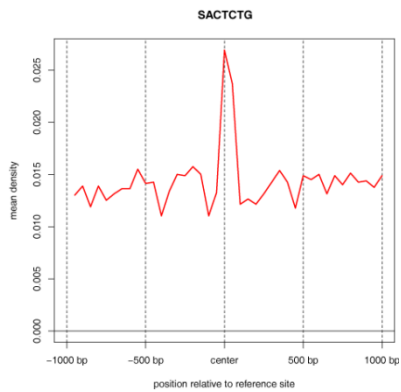


B

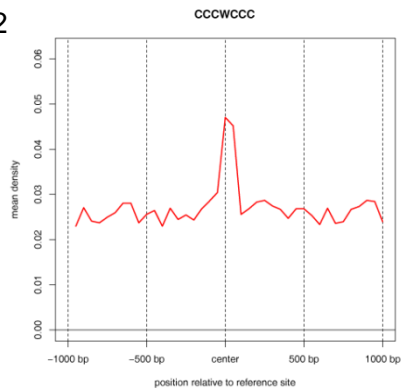


C

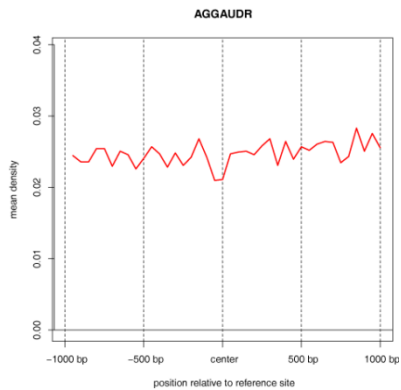
1



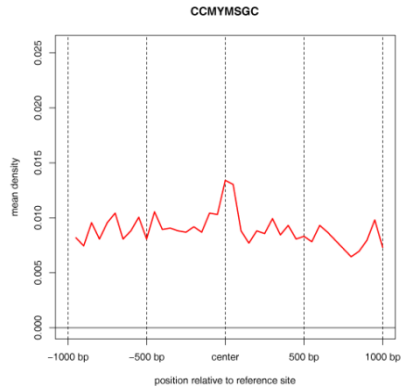
2



3



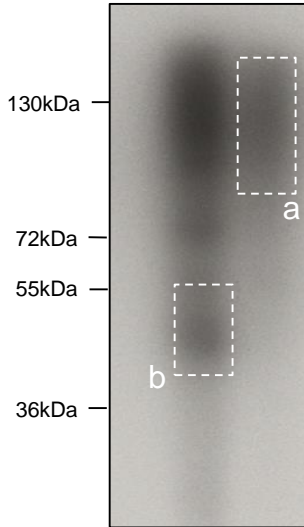
4



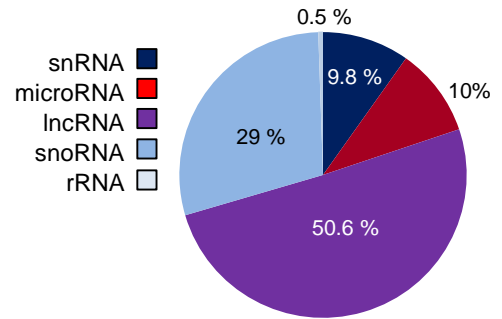
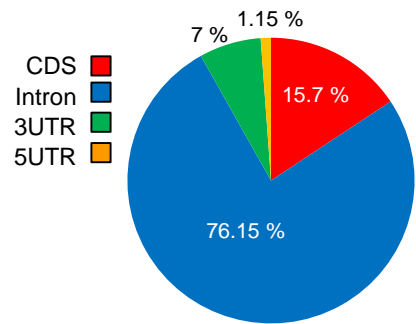
Martin_Suppl_Fig3

A

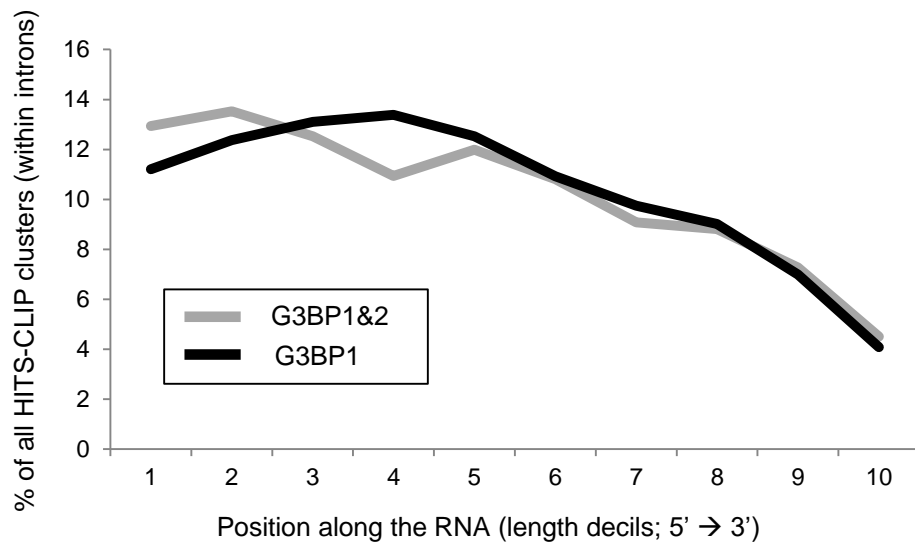
antiG3BP	+	+	+
UV	-	+	+
RNase	+	+++	+



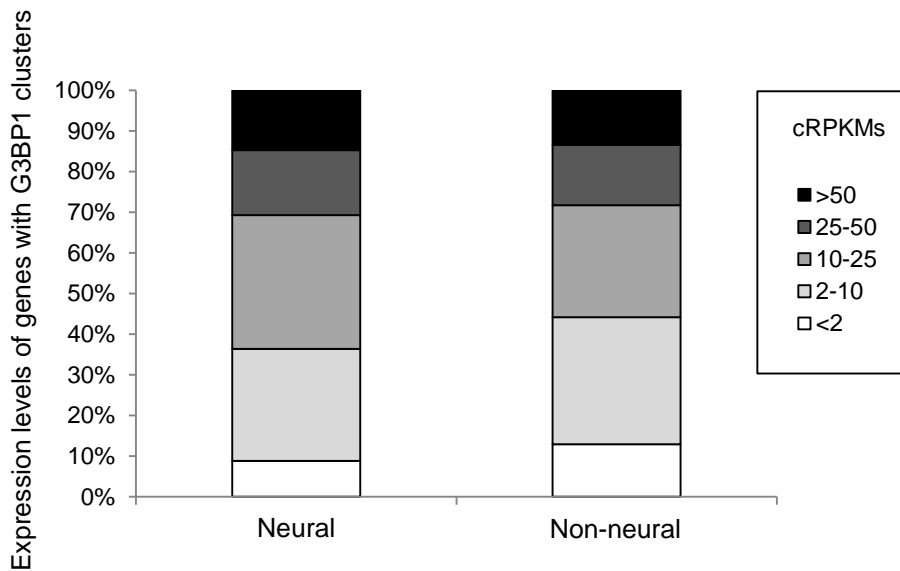
B

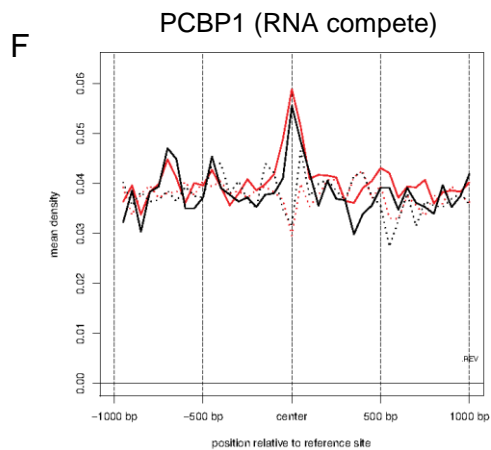
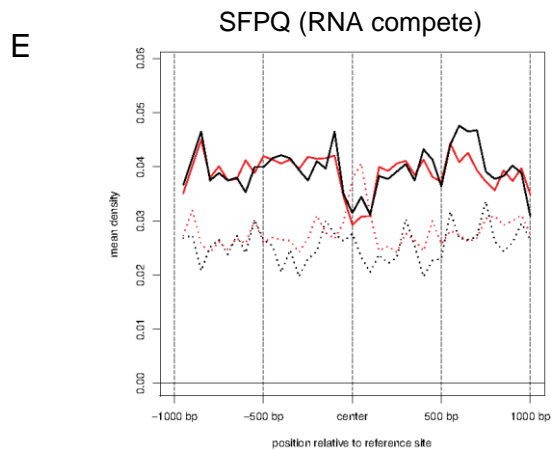
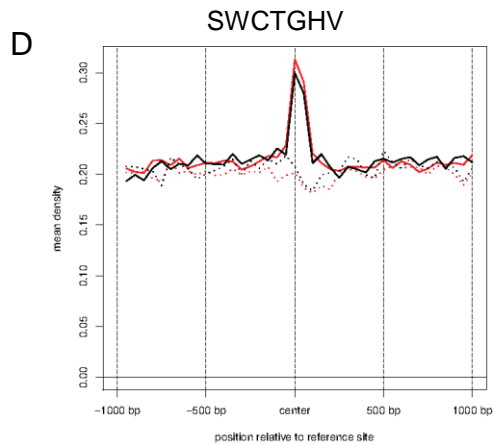
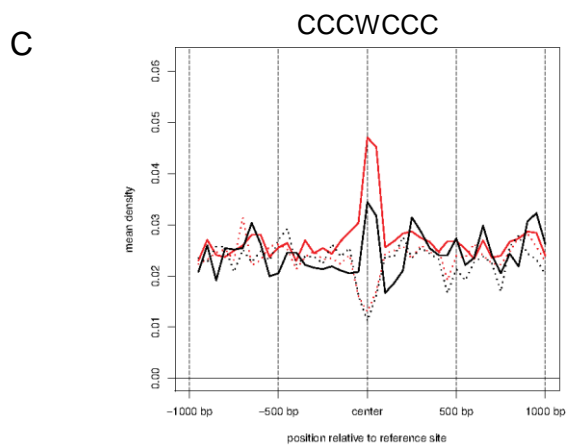
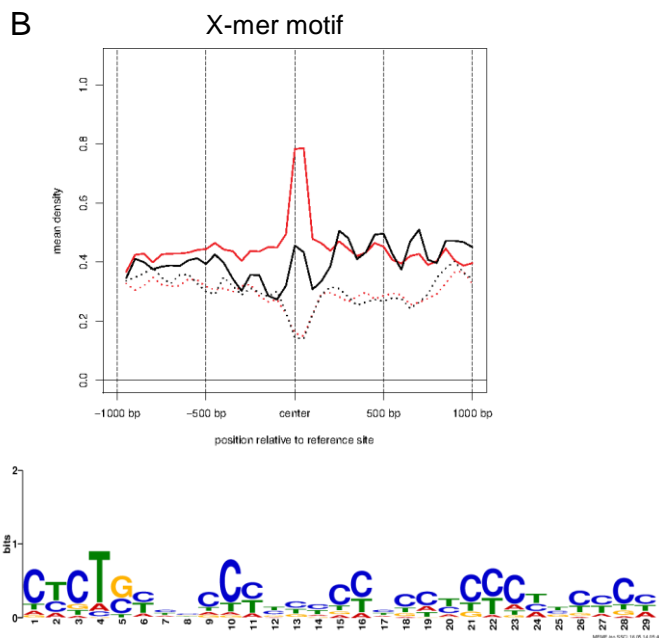
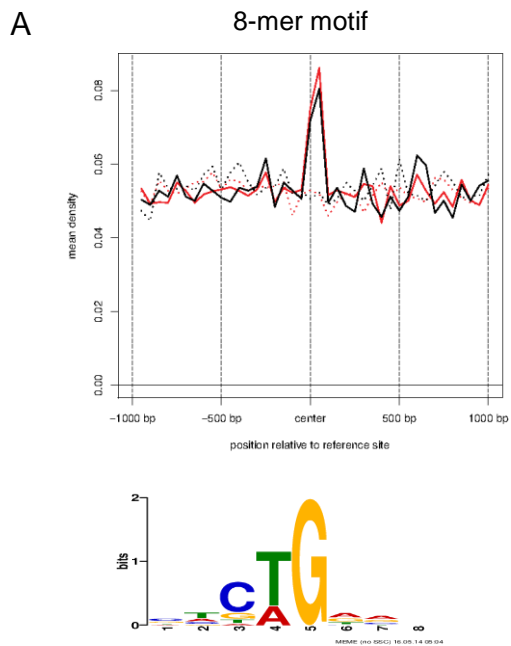


A



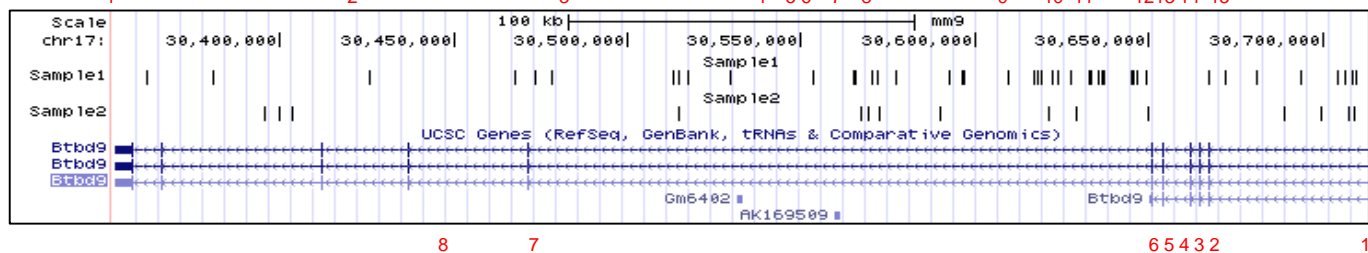
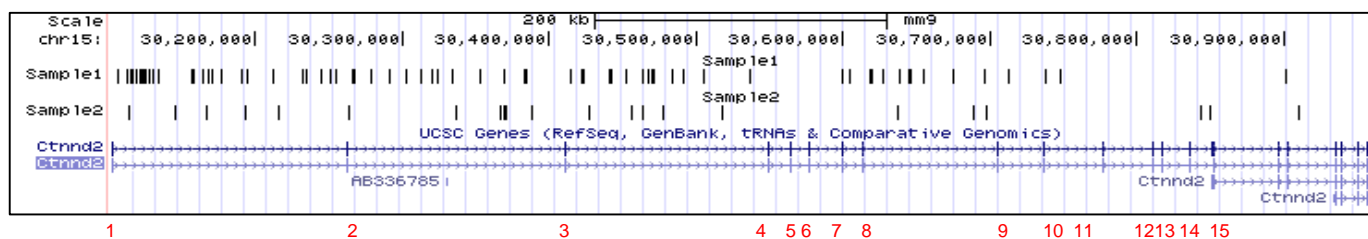
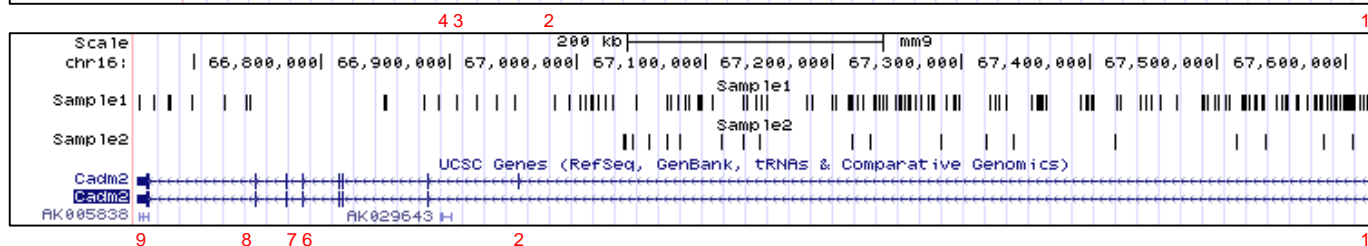
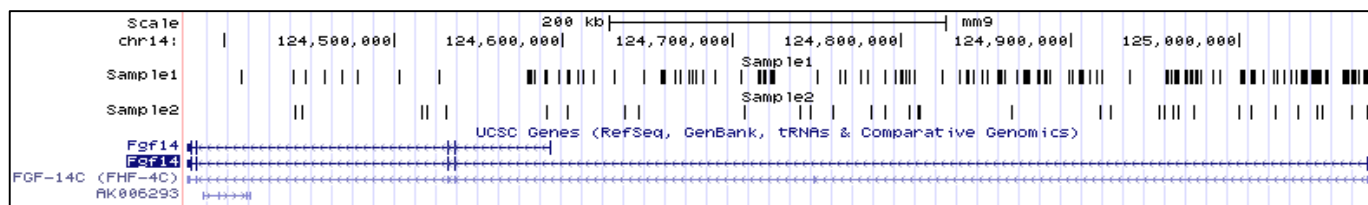
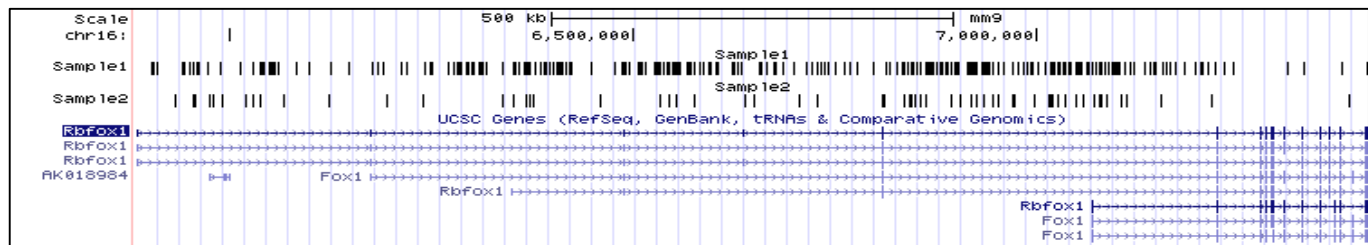
B



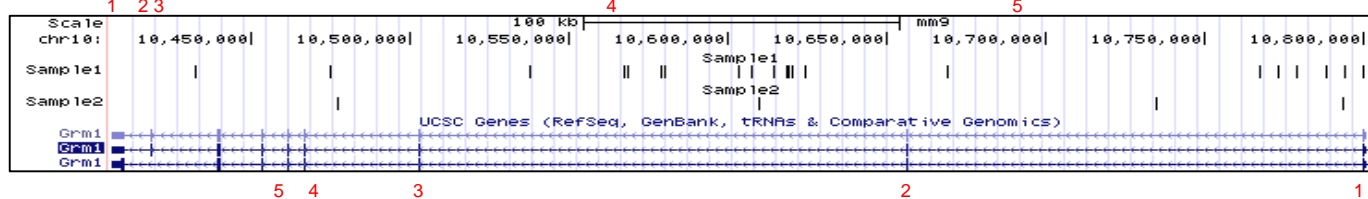
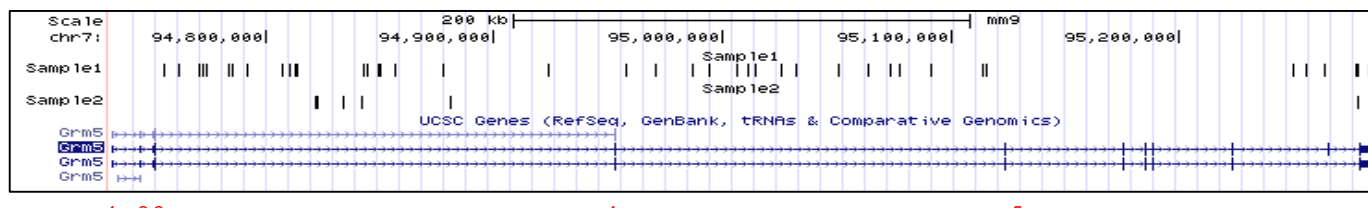


Martin_Suppl_Fig6

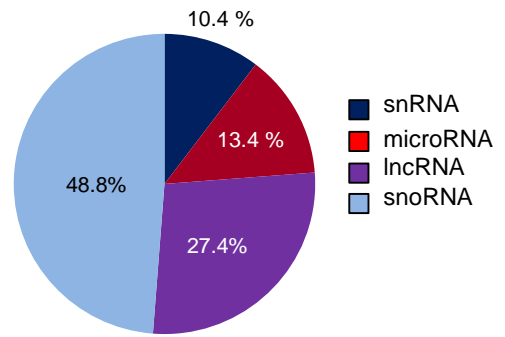
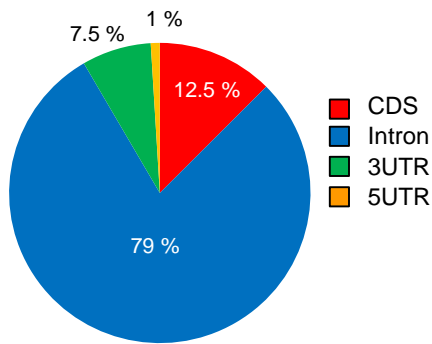
A



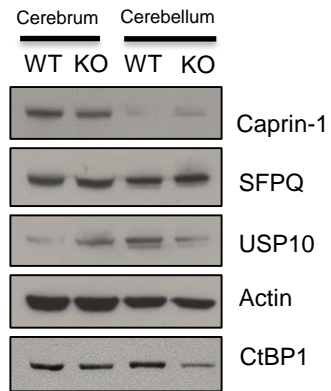
B



A



B



Martin_Suppl_Table1

Motif/kmer	Occurrence in dataset	Number of sequences with >0 occurrences	Enrichment (log2(odd-ratio), fisher test)	p-value (fisher test)	Corrected p-value
CACTCTG	79	79	2.03397352301535	8.36367078518999e-14	1.37030382144553e-09
CCTCCCC	117	109	1.48277400198375	4.28336426633652e-12	3.00239823448633e-08
CCCCACC	103	97	1.57239156067842	7.33007381466388e-12	3.00239823448633e-08
CTCTGAC	79	79	1.76375964057595	1.49842103529896e-11	4.91002604846763e-08
ACTCTGC	65	65	1.9906496684269	3.45484120941573e-11	9.43401972917789e-08
CCCTCCC	132	121	1.25080000558069	1.78107645540684e-10	3.70563925536815e-07
CCCACCC	108	102	1.3851537729765	1.80939416766023e-10	3.70563925536815e-07
CACCCCC	84	76	1.64673248276879	3.5785968015437e-10	6.51463666627688e-07
GACTCTG	74	73	1.67937289666419	4.54100214428132e-10	7.43997791319052e-07
ACTCTGT	66	64	1.76384087309488	1.23288781189943e-09	1.68330282584668e-06
AGAACTC	55	55	1.88481023118141	4.24776044338555e-09	4.97109336460206e-06
CCACCCC	95	89	1.3606822130546	4.71711369701264e-09	5.15234605412367e-06
CTCCCCC	85	81	1.39880639281943	1.06086864110913e-08	1.08632948849575e-05
CTCTGTC	86	82	1.37071012536049	1.5166117335089e-08	1.46165686128293e-05
TGGACAC	49	49	1.92360171644968	1.88389927588326e-08	1.62451609137217e-05
CTGGACA	51	51	1.87459322403525	2.22312578786484e-08	1.79045030758343e-05
TCTGACC	62	61	1.65618385975549	2.29488869990552e-08	1.79045030758343e-05
CCCCCAG	76	72	1.45394531501636	2.90082850189941e-08	2.16032609886908e-05
TCTCTGA	86	86	1.28982712574952	3.21685537316384e-08	2.29151993190941e-05
TCTGGAC	45	45	1.97698718113654	3.68738297254524e-08	2.32361856239159e-05
TGCACCC	45	45	1.97698718113654	3.68738297254524e-08	2.32361856239159e-05
CCCCCTC	76	70	1.41277622772998	7.74888662324093e-08	4.70213920130294e-05
CTCTGCA	79	75	1.3581328117457	8.95049943818321e-08	5.23732081411406e-05
CCCCTCC	97	92	1.17948278712954	1.03715284890212e-07	5.8955595738356e-05
ACTCTGA	62	62	1.50523224436894	1.23357008571972e-07	6.51961686594576e-05
GAACCTG	53	53	1.64984542023449	1.43285393920683e-07	7.33621216873897e-05
AGACTCT	51	50	1.69864489531004	1.79530511169155e-07	8.9134178632253e-05
CACCCCA	57	56	1.56908938997899	2.23674619256792e-07	0.000105313579080088
CCTGCCC	88	86	1.19100167809094	2.24974076403996e-07	0.000105313579080088
TCCACCC	48	46	1.72452686801238	3.98019904513845e-07	0.000175322791157168
CTGTCCC	73	69	1.33805768675863	4.06632450193627e-07	0.000175322791157168
CCCACTC	59	57	1.48487712820397	4.66973168391885e-07	0.00019617662540853
TGAGAAC	47	47	1.70506194156225	4.83862413239607e-07	0.000198190044462943
GAGGACT	37	37	1.96863261984354	5.92881408603587e-07	0.000225901604617701
ACCTCCC	49	47	1.65602429356152	6.77691919974897e-07	0.000246740098152638
CCTCCCA	67	66	1.32683526692484	7.13396692214799e-07	0.000254093291418419
CTGCCCC	72	71	1.27800864229547	7.80229588141143e-07	0.000271984714300095
CTCTGAA	72	71	1.25363273029603	9.16413493891877e-07	0.00031280247258176
TCCCTCC	81	80	1.16360324703028	9.54285918653778e-07	0.000319082050841296
ACTGTAG	56	56	1.45907951963319	1.1679009995712e-06	0.000382697799539491
CCTCTGC	90	88	1.07927595534817	1.29079764490828e-06	0.000414004449691914
CTGCACC	52	52	1.49972661427621	1.31397896630734e-06	0.000414004449691914
AACTCTG	54	54	1.44183964679332	1.60014098280295e-06	0.00047835663665765
TCTGAAC	51	51	1.51110654510082	1.68140080944381e-06	0.000491929836820131
TGAACAC	49	49	1.53558511562178	1.77615507559829e-06	0.000510535522080743
AAACTCT	44	44	1.65987237976682	1.8169558153044e-06	0.000513258690999091
CCTGAAC	40	40	1.744224536358	1.86994021874675e-06	0.000517743215449196
ACCCCA	52	47	1.56265888916071	1.92763282119146e-06	0.000517743215449196
CTCTGCC	90	88	1.06207399942437	2.00441458506353e-06	0.000529682718736788
AGGACTC	37	37	1.82404320585349	2.20711747633642e-06	0.00057399067829041
CTCTGGA	60	59	1.33734862622936	2.33269699300399e-06	0.00058798319282119
CTCCCTC	87	82	1.10224939663635	2.51369394392752e-06	0.000614691963840426
CCACCT	60	60	1.33135658816633	2.6760238405913e-06	0.000644764332415411
CCGAGAC	17	17	3.50663526061792	2.80647794292989e-06	0.000666396153869033
TCTGCAC	45	43	1.62644368101254	3.09316872866103e-06	0.000714158053901985
CCTCCCT	101	93	1.01231986908541	3.11999923482161e-06	0.000714158053901985
GAACTCT	52	52	1.4235217206299	3.26046135390866e-06	0.00073177258660876
CCCTGCC	83	76	1.12945117780143	3.43028829614746e-06	0.000756862194305637
CTGACCC	54	53	1.41461057292847	3.86235297056565e-06	0.000811292193201893
CCCTCTG	74	70	1.16213300532597	4.64202732126926e-06	0.0009340217394401
GAACACT	35	35	1.81382679028485	4.74459696854958e-06	0.000936572008827907
TCAGAAC	44	44	1.56008122439496	5.06519484062212e-06	0.000980858890266676
CTGAAAC	39	39	1.70743709821589	5.14855130389002e-06	0.000980858890266676

Oligonucleotides used for RT-PCRs and qRT-PCRs

Transcript	Forward (5'-3')	Reverse (5'-3')
<i>Rbfox1</i> e6-i6	AGCATGTTGGCGTCGCAAGGA	CTGTCTTTCTTTGACTACACGAGGGC
<i>Rbfox1</i> e7-i7	ACACAGACGCACTCGGAGCAG	AGATAAGGATGCCCAATCGCCAAGGA
<i>Rbfox1</i> 3'UTR	CGGAGCAGGGCCAGAGCAAC	CAGGTTGGCAGGTTACAGCGT
<i>Fgf14</i> e1-i1	AGGGCCTCTTCTTTCTCAGGGTGT	TGCATGTTCCCGGCCTCCT
<i>Fgf14</i> e4-i4	TGCTTTGCTCCATTTACTGCCTGGT	ACGTGTCAGAACCCATGCACACA
<i>Fgf14</i> 3'UTR	GTGGCGGAAAAGCCCCAGCA	GGGTAGCTGTCAGCAGGCAAGG
<i>Cadm2</i> e1-i1	CAGTGTCTGCGGGCTCCTGTT	GCGAAAGGTGCCCGCTAAGGA
<i>Cadm2</i> e8-i8	TGTTTTGTGGACGAAAGATGGTGCCG	TCCGCCACACTTTTCTTCCATTCATGT
<i>Cadm2</i> 3'UTR	TGATAGAGAACCCGTGCCGTGCT	TGGTGCAATGCTTGCAAATGGGAAAA
<i>Btbd9</i> e6-i6	CCAGCCCGTGTCTGCAGGTG	TCGTCACAGGTTTACATGTCCAC
<i>Btbd9</i> e9-i9	TGGTGGCCGACAGAACGAAAGTG	GAGGGAGGACTGCACAGGGAGT
<i>Btbd9</i> 3'UTR	GGCTGGGCTCTCTTCTGGTTTTGT	ACAAACAGGCCCAAGCTGGCA
<i>Ctnnd2</i> e1-i1	TTCGCCAGGAAGCAGTCGGG	AGACACTGACCCCGGACCCC
<i>Ctnnd2</i> i2-e3	TCTCCTGAGGCCAGTAACTGCACA	TCGCTCCAGCTGGCTGGCTA
<i>Ctnnd2</i> 3'UTR	GGGTGGTGGGCGTGTTATCG	AACCATTGATGAGGAAACACAGTGGCA
<i>Gria2</i> e2-i2	ACACACACACACAGGCACACCA	GGGGGCTATTTCCAAGGGGCG
<i>Gria2</i> i5-e6	CCCCAAGGCACAGCTGGGTT	GGTAGGCAAGAACAGCCTGGACG
<i>Gria2</i> 3'UTR	ACCACAGCAAATGAGATGACTGCACA	AGTGAGGGAGGGCATCAGTTTGTCA
<i>Gria2</i> e3-e4	TGGCACGCATCCATTTGTCATCCA	TCGCAGTCAAGGATTACACGCCG
<i>Grm1</i> i2-e3	CTGGCCTTGGAAGCCGCTC	CCCGTTCCCGCCCAACAAA
<i>Grm1</i> i3-e4	CGTGTTGGTGTCCAGCCTCAGC	GGCTGAAGCATGCTGGGGAGT
<i>Grm1</i> 3'UTR	ACAGCGGCACCGCATTGTA	ACGCACAGGGCCATGAACGAG
<i>Grm1</i> e4-e5	AGGCTGGACACCAACACGAGGA	AGGCCACGTAGCCAGGACA
<i>Grm5</i> i3	GGGCATGTGGGCCCTGTGTC	CTGCCACACCTGCAGCCTT
<i>Grm5</i> i4	GGGGAGTGGTGGGAAGGGT	TCCGACTGCCTCTGCCTCCC
<i>Grm5</i> 3'UTR	TCCCTACTGGGGTGCATTGTGT	TCCATAGGTGGGACTGGGGGC
<i>Grm5</i> e3-e4	TTCGGATGCCAGCAAGCCC	GGCTACCACCCGGGCTTTAGGT
<i>Gas5</i>	AGTCAGGGGCCTAACCGGGC	TCAAGGAAGCCCACCATCAGATAGAGC
<i>Hprt</i> 1i1	GCCTCCACGTGGGTTTGGGG	ACTGCCTGGTTAGGCCTCCGT
<i>Hprt1</i> e1-e2	CCGAGCGGTAGCACCTCCT	TCGAGCAAGTCTTTCAGTCCTGTCCA

Oligonucleotides used for RT-PCRs and qRT-PCRs

Transcript	Forward (5'-3')	Reverse (5'-3')
<i>Ghitm</i>	ACTGTTGGCTTCCTCCGTTCCCT	TCCGAGACTCTGCTGAGATGGACTG
<i>Eif3m</i>	AGCTGAGCTCTCAGAAGAAAACCGGA	GTCTCAGAGATGGGCGTTGCCTT
<i>Malat1</i>	GGCTGTGGCCTTGTGTTGGGT	CCAGTGGACCAGACCAACCCCC
<i>Batd2</i>	GCGGTTACAGGATGAGGAACGCC	TGGATGCTGGAACAGGAGCCGA
<i>Hexb</i>	TACCAAGACTCTTTCGGGACTTTCACC	AATTGTCTTACAGGCAGGAAGTGTCT
<i>Cacng7</i>	CCTTGGTGGTGGGCTTGGTCC	CGGGTACTTGATGGCCGGGG
<i>Dpysl</i>	GGTGATGGTGAGCCCATCACGG	GACATCCGCTCCTCAGTGCCG
<i>Sae1</i>	GTGCCTGACTTGCTGCTCCAGA	TGCTCTCTTGGCCTCAGGTCCATC
<i>Clk1</i>	AGGGCATCCCTATGGAGAACCTGG	ACTCCTCGATCTTTTCCTTCGGTGACT
<i>Kpna1</i>	ACCGCTTGACCATGACCCGGA	CCATCTGACAGGTATGAGAGAGCCCA
<i>Trpm7</i>	GCAGCACCCCTCAGTTGCGA	CCTGCCTTTCATAAGGCAAGCCC
<i>Adcy2</i>	TCCCACTGCCCTGGCCATCT	TGCTCCTTGGCCCCTGGTGT
<i>Akap9</i>	CAAGGAGCAACTGCAGCGAGACA	GCATCCCGCTCCACCTGCAA
<i>Bag4</i>	GCACCTCCTCTGAGGGGACAGG	GGCTGAGGGTGCAGCTGAAGG
<i>Cog7</i>	GAAGCCCTGAAGCAGGAGGCG	AGGTCTCCTCAATATCAGCACTCAGCG
<i>Gdi2</i>	TCGAGGGTGTTGATCCTAAGAAGACCT	ACGGGCTTTTACCGTATCTTGCCA
<i>Gria3</i>	ACCUCGUGACCCACAAAGCCC	ACCAGAGUCCAGGGUCCCCGU
<i>Kcnq5</i>	GCAGTTGTCTCTGCAAAAACCTCAGGGT	CCAGAGAGCATCCGCATATGTGCAAAA
<i>Lars2</i>	AGGGACAGACATTCCGCCTTCCA	TGTCCTTCTCAGGAGGGGCTGC
<i>Lig3</i>	ACCAGGAGAGAAGCTGGCTGTGA	CCAGGTCCCCGTGCAATGCC
<i>Picalm</i>	TGCAGTCTCTTCTTGGCAAGCACT	TGAGGTGGACACCGGAGAGGC
<i>Rpl13</i>	GCCCAGCCGGAATGGCATGA	AGCCCGGACCTTGGTGTGGT
<i>Ube3a</i>	GGAGTAGATGAGGGAGGCGTTTCCA	GAGTCTCCCAAGTCACGAAAGTTCC
<i>Vps35</i>	AGCATTAGCTGCTGGAGAAATTGGCT	CACGGCTCGGCCCTGATCTGG
<i>Meg3</i>	TGCAGGGAGAAGAGGGGAAGGG	GCAGCCTTGTTTGGGGGCCA
<i>Fcgr4</i>	AAGCTGGTCTCCAAAAGGCTGTGG	GGACCTCTAGTTGCACTGGGTCACT
<i>Calm1</i>	GTGACACCCTGGTGGCGAGC	TGTCCCCCTTCCAATCTCAGCTTTACA
<i>Stx1a</i>	AGCAGAGCATCGAGCAGGAGGA	TGCGCCCTTTGCAGCGTTCT
<i>Trpm3</i>	GCCAGCATGTTGGACTCACTCCC	AGTTCTGCAGCCCTCCGTGC
<i>Wsb1</i>	TGAGCAATGGTCTTTGCTGTGCCT	TCTTGGGTGGACATCACTCTTCGGA
<i>Gusb</i>	GCGGGACTGCATCGATCTGTGG	TTGACCCTGGTTCCCTGTCCCA