Supplementary material

18 tables, 10 figures, 17 spreadsheets and Supplementary Methods are in supplementary material.

Expression data is available in PainNetworks (http://www.painnetworks.org/) and raw and processed sequencing data is available in GEO (super series http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE107182. Rodent DRG data with all supplemental material http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE107180, IPSc neurons data http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE107181). Supplemental spreadsheets are available at http://doi.org/10.6084/m9.figshare.6508205.

Table of Contents

S.Figure Index

Index of S.Tables

Supplementary Methods

Single-cell mouse DRG data

Single cell mouse DRG data has been previously used and disseminated by (Li et al., 2016). We downloaded and reanalysed data in order to re-classify neuron subtypes and then looked for neuron subtype-specific expression.

Tissue specificity

Tissue specificity was calculated using the tau metric (Yanai et al., 2005):

$$\tau = \frac{\sum_{i=1}^{n}(1-\hat{x_i})}{n-1};\ \hat{x_i} = \frac{x_i}{\max\limits_{1\le i\le n}(x_i)}$$

applied on regularised transformed counts of ENSEMBL genes and novel LncRNAs. R implementation:

```
for (i in 1:N) {

expr_est[,i] <- rowMeans(eset[,eval(parse(text=paste("colData","$",condition, sep="")))==levels(eval(parse(text=paste("colData","$",condition, sep=""))))[i]])

expr_sd[,i] <- apply(eset[,eval(parse(text=paste("colData","$",condition, sep="")))==levels(eval(parse(text=paste("colData","$",condition, sep=""))))[i]], 1, function(x) std(x))

}

max_expr <- apply(expr_est, 1, function(x) max(x))

max_norm <- sweep(expr_est, 1, max_expr, "/")

t.index <- apply(max_norm, 1, which.max)

tau <- apply(max_norm, 1, function(x) sum(1-x) / (N-1))

tau_results <- data.frame(tau = tau, t.index = t.index)
```

```
for (i in 1:nrow(tau_results)) {
tau_results$mean_expr <- max_expr
tau_results$sem[i] <- expr_sd[i, t.index[i]]
}
return(tau_results)
}
```

IPS derived human neurons

Induced pluripotent stem cell generation

NHDF1 (from 44-year-old female) was reprogrammed with Yamanaka retroviruses SOX2, KLF4, OCT3/4, c-MYC and NANOG (Takahashi et al., 2006), and has previously been described (Hartfield et al., 2014). AD2-01 (from 51-year-old male) (Buskin et al., in preparation) and AD4-01 (from 68-year-old male) (Melguzo et al, in preparation) were reprogrammed using the CytoTune™-iPS Reprogramming Kit (ThermoFisher). The fibroblasts to generate AD2-01 and AD4-01 were obtained from a commercial source (Lonza, CC-2511).

CytoTune-iPS reprogramming was performed as directed by the manufacturer's instructions (ThermoFisher). The CytoTune reprogramming kit contains four Sendai virus-based reprogramming vectors each capable of expressing one of the four Yamanaka factors (KLF4, OCT3/4, SOX2 and c-MYC). Briefly, after fibroblast transduction with the four Sendai virus-based reprogramming vectors, cells are cultured for 5-6 days, with medium changes every other day (DMEM, high glucose (Sigma), 10% FBS (ThermoFisher), 1% Pen/Strep (100x, ThermoFisher), 200mM L-glutamine (Sigma), 1% non-essential amino acids (ThermoFisher)). The transduced fibroblasts are then passaged using 0.05% Trpsin-EDTA onto pre-prepared feeder layer plates containing mitotically inactivated mouse embryonic fibroblasts (MEF). 3-4 weeks after transduction, colonies should have grown to an appropriate size to allow for manual picking. Using an inverted microscope, a single colony displaying iPSC morphology is cut into 5-6 pieces using a 25 gauge needle, transferred into iPS media (KO-DMEM (ThermoFisher), 25% Knock Serum Replacement (ThermoFisher), 1% nonessential amino acids (100x, ThermoFisher), 200mM L- glutamine (Sigma), 1% Pen/ Strep (100x, ThermoFisher), 8 ng/ml human FGF2 (Miltenyi Biotec)) and plated onto pre-prepared MEF plates. Colonies are allowed to attach for 48 hours, and thereafter medium changes are performed daily. iPSCs were adapted to feeder-free conditions onto Matrigel (Scientific Laboratory Supplies)-coated plates in mTeSR1 medium (ScienCell). Bulk passaging was by 0.5 mM ethylenediaminetetraacetic acid (EDTA) to make large-scale, quality-controlled stocks that were cryopreserved in liquid nitrogen. The number of feeder-free passages was kept to a minimum. When selecting iPSCs from frozen stocks for differentiation, vials with the same passage number were selected for each cell line throughout all experiments performed in this study.

The iPSC lines AD2-01 and AD4-01 were obtained through the IMI/EU sponsored StemBANCC consortium via the Human Biomaterials Resource Centre, University of Birmingham, UK (http://www.birmingham.ac.uk/facilities/hbrc).

All iPSC lines were subject to strict quality control checks before the initiation of differentiation. Quality control checks of this line included: tests for Sendai virus clearance, fluorescence-activated cell sorting (FACS) for pluripotency markers, genomic integrity checks and embryoid body tri-lineage differentiation experiments. Cells are also confirmed as negative for Mycoplasma before cryopreservation.

Sensory neuron differentiation

For neuronal differentiation, iPSCs were passaged onto Matrigel®-coated six-well plates using TrypLE express (ThermoFisher Scientific) and maintained in mTeSR1 supplemented with 10 µM ROCK inhibitor (ScienCell). Twenty-four hours after plating, the medium was exchanged to mouse embryonic fibroblast (MEF) conditioned medium (ScienCell) supplemented with 10 ng/ml human recombinant FGF2. Cells were allowed to expand on MEF-conditioned medium until 50% confluent, at which time differentiation was started according to Chambers et al. (2012). Briefly, medium was exchanged to knockout serum replacement (KSR) medium containing; knockout-DMEM, 15% knockout-serum replacement, 1% GlutaMAX™, 1% non-essential amino acids, 100 µM β-mercaptoethanol, 1% antibiotic/antimycotic (ThermoFisher Scientific), supplemented with the SMAD inhibitors SB431542 (Sigma, 10 µM) and LDN-193189 (Stratech, 100 nM). The medium was gradually transitioned from KSR medium to N2 medium (Neurobasal® medium, 2% B27 supplement, 1% N2 supplement, 1% GlutaMAX™, 1% antibiotic/antimycotic) (ThermoFisher Scientific) over an 11-day period. On Day 2, the small molecules CHIR99021 (Apollo Scientific, 3 µM), SU5402 (R&D Systems, 10 µM) and DAPT (Sigma, 10 µM) were also added. SMAD inhibitors were removed from the media from Day 6 onwards. On Day 11, the now immature neurons were replated onto Matrigel®-coated coverslips (25 000 cells per 13 mm coverslip) in 100% N2 medium containing human recombinant NGF, GDNF, BDNF, NT3 (all at 25 ng/ml, PeproTech) and 10 µM ROCK inhibitor. CHIR99021 (3 µM) was included in the medium until Day 14, and laminin (1 µg/ml, ThermoFisher Scientific) was supplemented into the medium from Day 20 onwards. Medium changes were performed twice weekly after replating. Cytosine β-D-arabinofuranoside (AraC, 2 µM, Sigma) was included in the medium for 24 h following replating to remove the few non-neuronal dividing cells remaining in the culture. This differentiation resulted in a completely pure neuronal culture with extensive arborized neurites by 3 weeks after the end of the small inhibitor stage.

Ethics statement

Human iPSC lines used in this study were derived from human skin biopsy fibroblasts, following signed informed consent. Three control cells lines were used in this study – AD2-01, AD4-01 and NHDF1. NHDF1 were reprogrammed with approval from research

ethics committee: National Health Service, Health Research Authority, NRES Committee South Central, Berkshire, UK (REC 10/H0505/71).

Sequencing and mapping

Sequencing was performed at Oxford Genomics using the Illumina HiSeq2000 paired-end protocol with 100bp reads for rat and Illumina HiSeq4000 paired-end 100bp for mouse and 75bp for human IPSC and IPSC derived neurons. Oxford Genomics produced FastQ sequencing files which encode quality metrics following the Sanger standard, i.e. Sanger qualities, using the standard Phred score (Ewing and Green, 1998) to assess the probability that the corresponding base call is wrong. Sequencing was done in multiple sequencing lanes producing multiple technical replicates per sample. In general all these lanes gave high yield, consistent GC content, consistent and expected sequence insert between the paired-end adapters and high quality base calling. Conditions and strains were multiplexed in lanes and library batches.

Mapping to the genome was done using STAR aligner. Reads were mapped on the mm10 (GRCm38) mouse genome, rn6 (Rnor_6.0) rat genome and HG38 (GRCh38) human genome, all downloaded from ENSEMBL. STAR was run suing the following parameters (according to ENCODE guidelines):

--outFilterMultimapNmax 20

max number of multiple alignments allowed for a read: if exceeded, the read is considered

unmapped

--alignSJoverhangMin 8

minimum overhang for unannotated junctions

--alignSJDBoverhangMin 1

minimum overhang for annotated junctions

--alignIntronMin 20

minimum intron length

--alignIntronMax 1000000

maximum intron length

--alignMatesGapMax 1000000

maximum genomic distance between mates

Genome was generated with --sjdbOverhang ReadLength-1.

Conditions and strains were multiplexed in lanes and library batches. Lanes were merged as BAM files after mapping (Li et al., 2009).

Identification of novel LncRNAs

Our analysis pipeline is illustrated in Figure 1. Workflow similar to (Cabili et al., 2011; Gerstein et al., 2014; Ilott and Ponting, 2013). (Pertea et al., 2015) is similar but we directly reconstruct novel LncRNAs on the gene level and do not identify individual transcripts of the same gene. Doing this we get a non-redundant annotation of unique genes of LncRNAs suitable for count based DE analysis (Anders et al., 2015; Love et al., 2014). The concept of islands of expression is described in (Gerstein et al., 2014). These islands of expression outside annotated genes would include all previously unknown exons of novel LncRNAs and consequently we trimmed, clustered and grouped them together into gene models. Clustering and grouping into gene models was done by taking into account identified splicing events.

Steps of the method are as follows:

1. Map reads to the genome and identify splicing junctions using STAR. Fetch ENSEMBL (Harrow et al., 2012), RefgSeq (Pruitt et al., 2014) and XenoRefSeq annotations using BiomaRt (Durinck et al., 2009) (XenoRefSeq includes all known annotated gene models from other organisms which can be accurately aligned to the genome of the organism of the experiment.). Perform operations on genomic intervals (Aboyoun et al., 2013; Carlson et al., n.d.) : Group annotations on the gene level, unlist, collapse and concatenate annotated gene models. Extend gene models by 1000bp in both ends. Identify gaps in annotations, i.e. regions not covered by known gene models.

2. Read all BAM files (mapped RNA-seq reads) of the experiment in chunks to reduce memory footprint. Consider only reads which are properly paired (Lawrence et al., 2013). In parallel (Morgan et al., 2017), using all available processors, subset read-pairs overlapping regions not belonging to known gene models. From these reads calculate genomic coverage and create coverage vectors encoded by their run-length (Pagès et al., 2017). Slice islands of these vectors above coverage threshold and having length > 100, i.e. identify islands of expression (I.o.E). These operations are executed by the custom functions “BAM_to_IOE” and “findRegs”. All operations are strand-specific.

3. Use “awk” (Aho et al., 1978) to select only the novel splicing junctions (SJ) identified by STAR. Read these SJs in R and filter out the ones identified by =< 2 reads, span < 20bp or > 100000bp. SJs are represented as the genomic coordinates of predicted introns.

4. Collapse I.o.E. Calculate the intersect of regions covered by I.o.E. in all samples. Identify I.o.E. contained in annotated introns, grouped in the gene level. Identify I.o.E. overlapped by novel SJs and novel SJs overlapping I.o.E.

A. For I.o.E. not overlapped by SJs, filter out the ones contained within introns, not belonging in the intersect of I.o.E in all samples and have a width =< 200bp. Calculate their length normalised coverage by pooling reads from all samples and select the ones that are significantly expressed at an one-sided Pr (>|Z|) < 0.1. Feed coverage vectors in a smoothed z-score thresholding

signal processing algorithm (van Brakel, 2014) implemented in the custom function “dropDetect”. Rolling coverage was calculated over a smoothing window of 31bp, the minimum coverage drop threshold was set to 5 and the minimum intron length to 20bp. Trim ends by identifying sudden drops in coverage and identify introns not detected by the aligner. Remove identified introns from the I.o.E. to reconstruct gene models.

B. Collapse SJ and calculate a disjoined set of their genomic intervals, i.e. not overlapping subset keeping track of the original regions. Calculate a mapping of co-overlapping introns (SJs) and create a GrangesList (Aboyoun et al., 2013) which groups together co-overlapping SJs. On the top level of the list are sets of co-overlapping SJs.

C. Collapse I.o.E. overlapped by SJs. Keep track of overlaps by creating a network connectivity matrix that holds information about which I.o.E. are connected to each other. Create a GrangesList (Aboyoun et al., 2013) which has at the top level sets of co-overlapping I.o.E, represented as interconnected nodes. Find overlaps between these grouped I.o.E and grouped SJs. Update the connectivity matrix to keep track of collapsed I.o.E. grouped together by SJs.

D. Within each set of grouped SJs (set of interconnected nodes), calculate overlaps between the disjoined segments and the original SJs. In this way each original SJ votes for a disjoined segment. Within each group select putative intronic areas (i.e. disjoined SJs) overlapped by > *round(top(counts) – 2*sd(counts))*. In this way we identify consensus introns (SJs) within each group.

Groups of putative introns (SJs) and putative exons (I.o.E.) are stored in parallel lists with the same top levels, each represented as a GrangesList grouped together by the connectivity matrix. Subtract introns from I.o.E., trim the edges and reconstruct gene models in a customised annotation. Discard gene models with length < 200bp. Scan the annotation for overlaps between gene models and create a non-reduntant annotation suitable for counting features at the gene level. Export the customised annotation in the form of a gene transfer format (GTF) file.

5. Count features using HTSeq (Anders et al., 2015) and the intersection not empty strategy.

6. Read HTSeq’s output and create a table of counts across all samples (TOC). Remove gene models of putative LncRNAs not expressed in at least all samples of a condition or strain. Remove putative novel LncRNAs below an average expression cut-off threshold of > 0.5 fpkm for at least one condition.

7. Fetch the genomic sequences of the expressed gene models in FASTA format and calculate coding potential using CPAT (Wang et al., 2013). Discard gene models with positive coding potential and create a set of putative LncRNAs

8. Annotate LncRNAs according to their genomic context using BedTools (Quinlan and Hall, 2010). Discard models with retained introns or intronic LncRNAs not supported by evidence of novel SJs.

9. Count features of the full gene set of annotated genes and novel LncRNAs and continue with differential expression analysis.

The R script “identify_LncRNAs.R” in http://github.com/gbaskozos/Scripts_LncRNAs carries out the identification of novel genes based on RNA-seq coverage.

The custom function BAM_to_IOE <- function(bamfile, PATH, PATH_results, igRangesExt, param, len=100, dep=2, suffix=11) takes as input a BAM file list, genomic ranges outside gene models, length and depth of continuous coverage and identifies I.o.E. The custom function dropDetect <- function(coverage, start, seqnames, strand, lag,threshold, length, influence, intron_identification) uses a smoothed z-score thresholding algorithm, adapted from (van Brakel, 2014), to identify coverage drops and peaks. It takes as input a coverage vector, start of genomic ranges, strand, the smoothing window (lag), z threshold, minimum length of intron, influence of previous peaks/drops on current signal and intron identification TRUE or FALSE and identifies transcription ends and/or introns.

We only included putative LncRNAs in this novel annotation only if they were present in all replicates of the experiment.

The pipeline was scripted in R using bioconductor (Gentleman et al., 2004) packages and custom scripts. Other R packages used in identification of LncRNAs are: data.table (Dowle et al., 2015), rtracklayer (Lawrence et al., 2009), biovizBase (Yin et al., 2017), annotationDbi (Pagès et al., 2017), IRanges (Lawrence et al., 2013).

In ribozero libraries novel intronic gene models, fully contained in the intron of a gene model and covered >= 80% of its respective introns were considered retained introns coming from sequencing of non mature transcripts and were discarded.

Transcription Start Sites mapping to mm10

TSS data was downloaded from FANTOM 5 database (“FANTOM5 CAGE profiles of human and mouse samples | Scientific Data,” n.d.; Lizio et al., 2015). We downloaded TSS data that has been classified as “True TSS” by the “TSS classifier”. Then we used the UCSC LiftOver tool (Meyer et al., 2012) to translate genomic coordinates from the mm9 genome to the mm10. 51% of the true TSS were unambiguously mapped to mm10.

Differential expression analysis

DE analysis was done in R using DESeq2 (Love et al., 2014). PCA was always performed on regularized log transformed counts (Love et al., 2014) using the top 10000 (mouse and human) or 5000 (rat)genes and novel LncRNAs ranked by observed

variance. Hierarchical clustering was always performed on regularized log transformed counts using euclidean distances and complete linkage.

ENSEMBL gene annotations used: GRCh38.88 (HG38), GRCm38.87 (mm10) and Rnor_6.0.90 (rn6). To select all the annotated LncRNAs we programmatically downloaded from biomaRt all gene descriptions and gene biotypes, and then selected all genes with biotype “lincRNA”, “antisense”, “antisense RNA” “sense intronic” which denote either long intergenic non-coding RNAs, antisense long non-coding RNAs and intronic LncRNAs.

GO enrichment for DE ENSEMBL annotated genes was carried out using custom R scripts and methods developed in the topGO (Alexa and Rahnenfuhrer, 2010) R package. In the case of annotated genes we used as a background the total population of expressed genes and DE p.value < 0.05 as significance cut-off. We used the *elim* method introduced in (Alexa et al., 2006) to compute the significance of a node dependent on the significance of its children. We tested enrichments using the Kolmogorov-Smirnov test (Alexa et al., 2006; Alexa and Rahnenfuhrer, 2010).

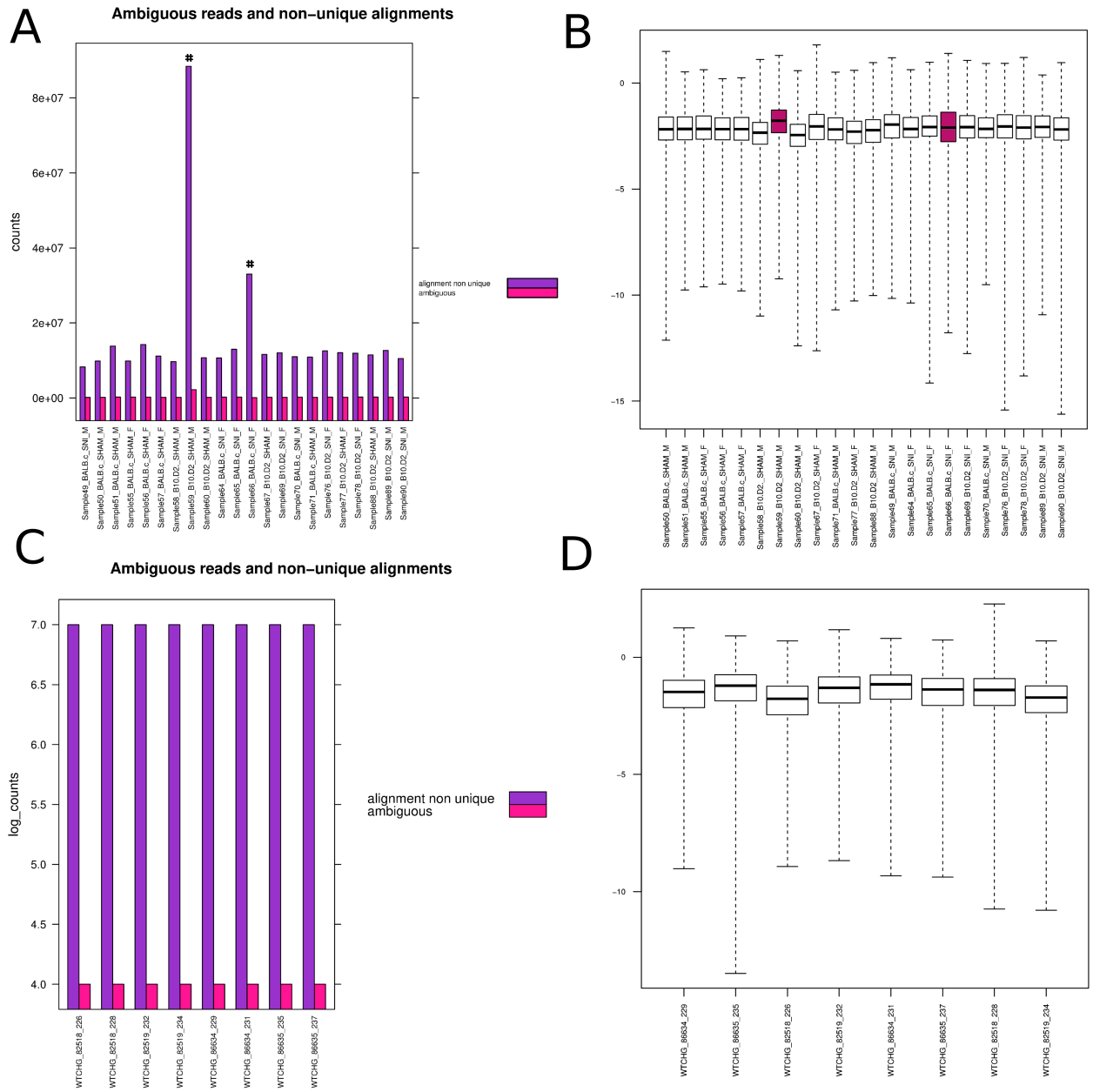
Gene set enrichment analysis and GO enrichment of genes belonging to WGCNA network modules was carried out using the Fisher’s Exact test and the hypergeometric distribution and calculated by custom R scripts and the GSEA (Morgan et al., 2017) and GO stats (Falcon and Gentleman, 2007) packages. Significance cut off was p.value < 0.05 and the minimum size of a GO BP term was 100 genes.

DE and counting features

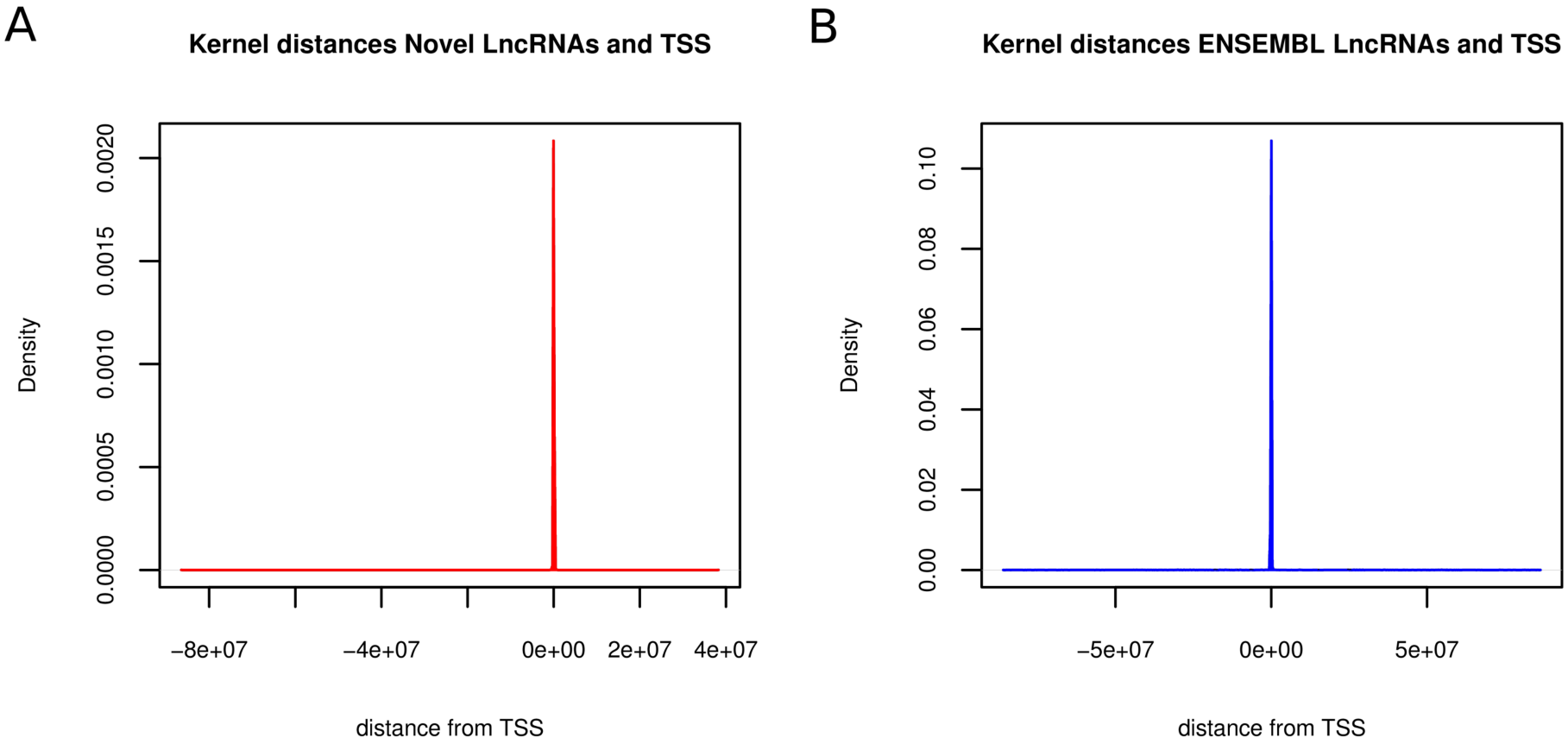
DE analysis was done using DESeq2 (Love et al., 2014) using default settings. Significance cut-off in all cases was FDR < 0.05. Counting of features was done using HTSeq (Anders et al., 2015) and the intersection not empty strategy to resolve ambiguously counted reads.

The generalised linear models fitted for each experiment were: mouse DRG: ~ sex + strain*condition, rat DRG: ~ condition, human IPSC: ~ cell_line*condition.

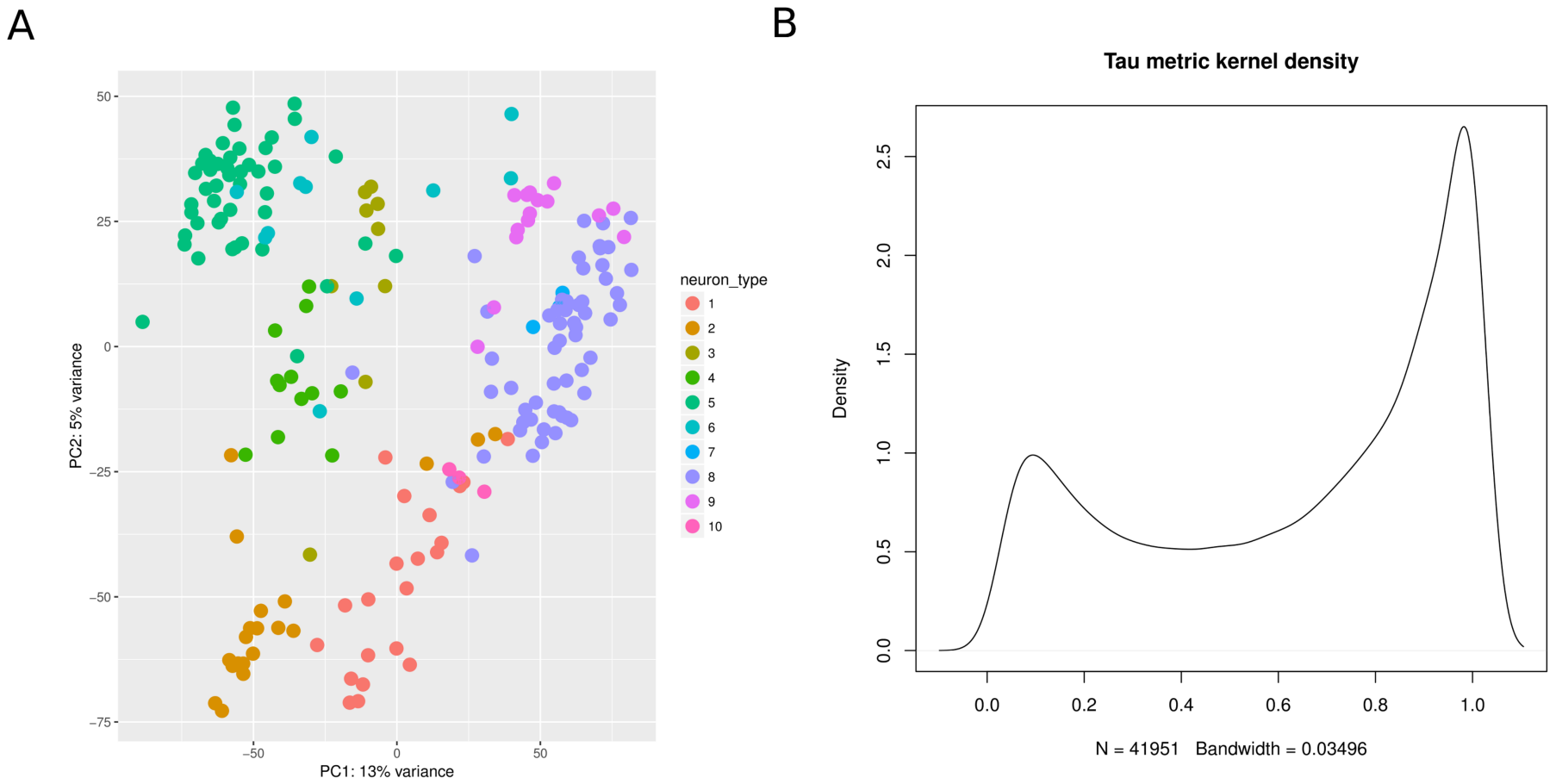
Supplementary figures



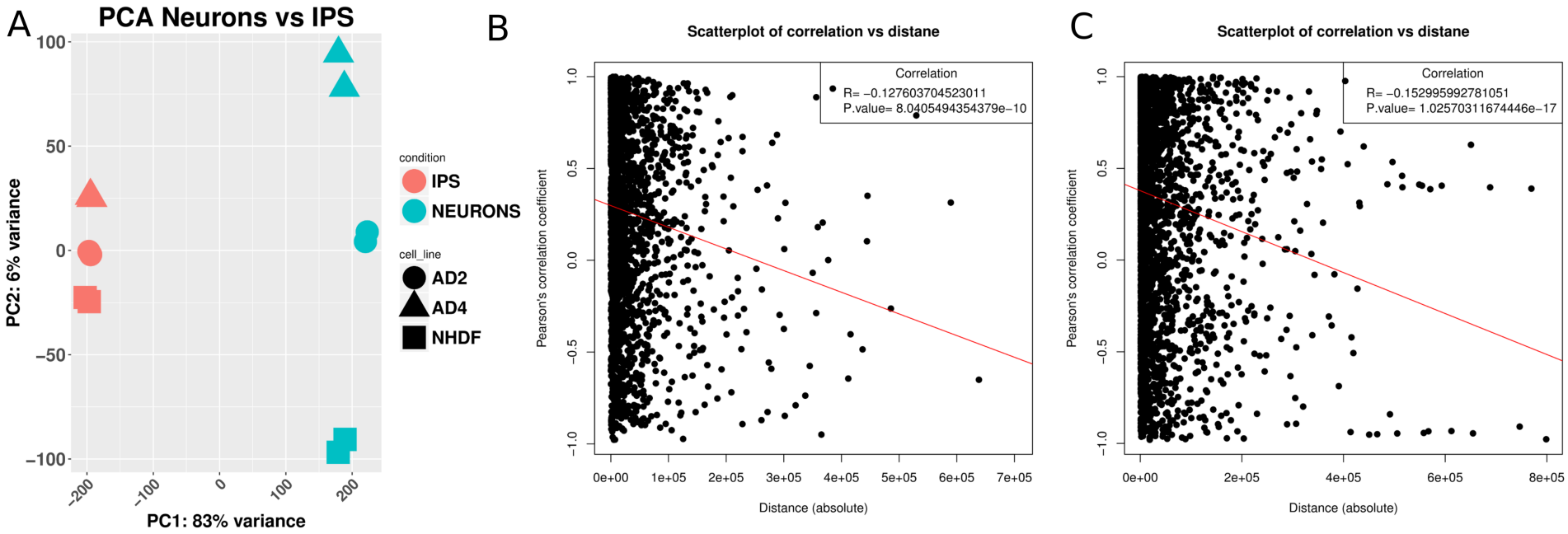
S.Figure 1: A,C: Ambiguously mapped reads per sample (pink bars) and read counts generated from reads mapped to multiple loci (purple bars) in mouse (A) and rat (C).# indicates excluded samples. B,D: Boxplot of Log10 Cook's distance for all ENSEMBL genes in all samples in mouse (B) and rat (D).



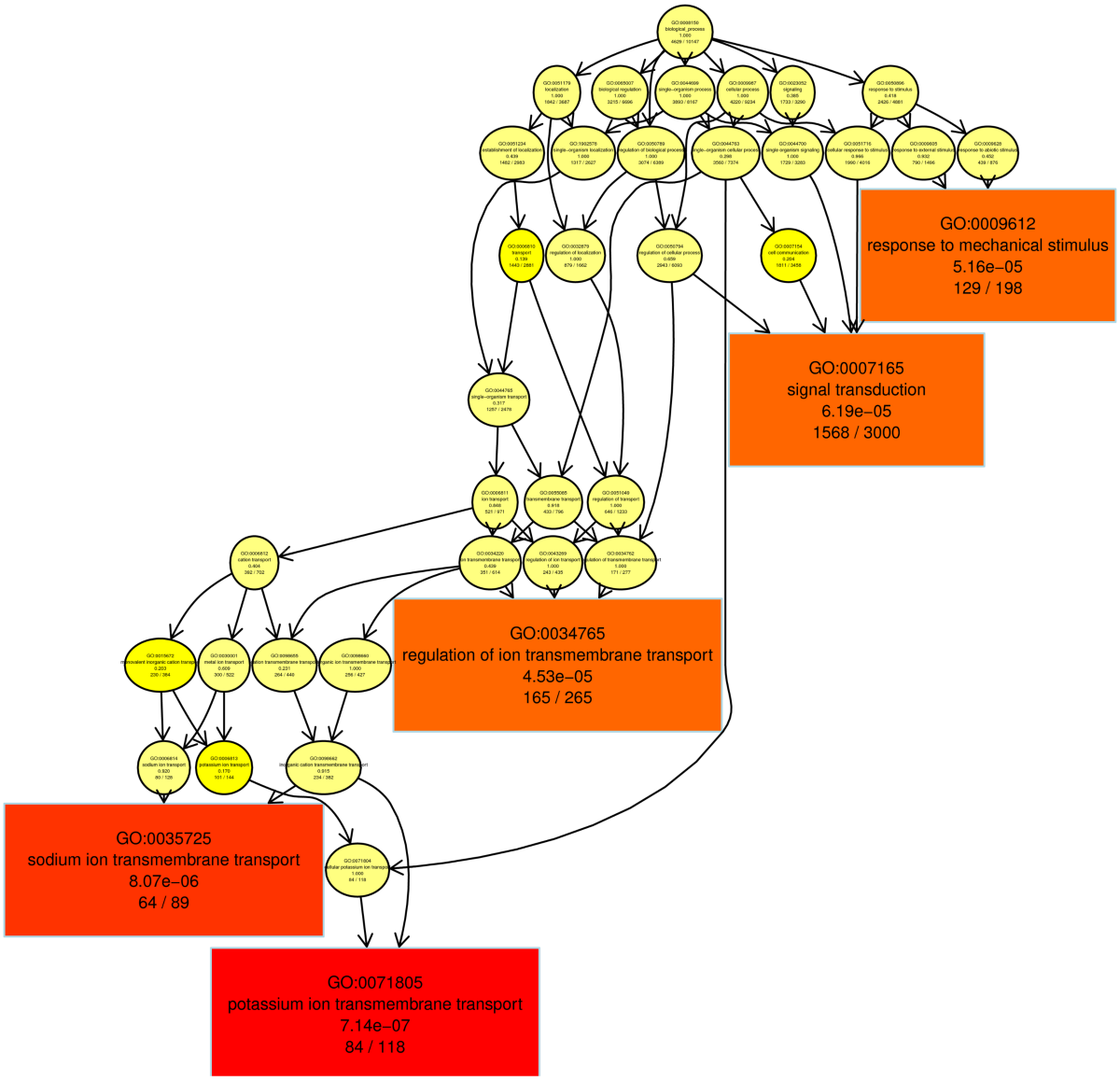
S.Figure 2: Distribution of distances between TSS and novel LncRNAs (A) and TSS and annotated LncRNAs (B).



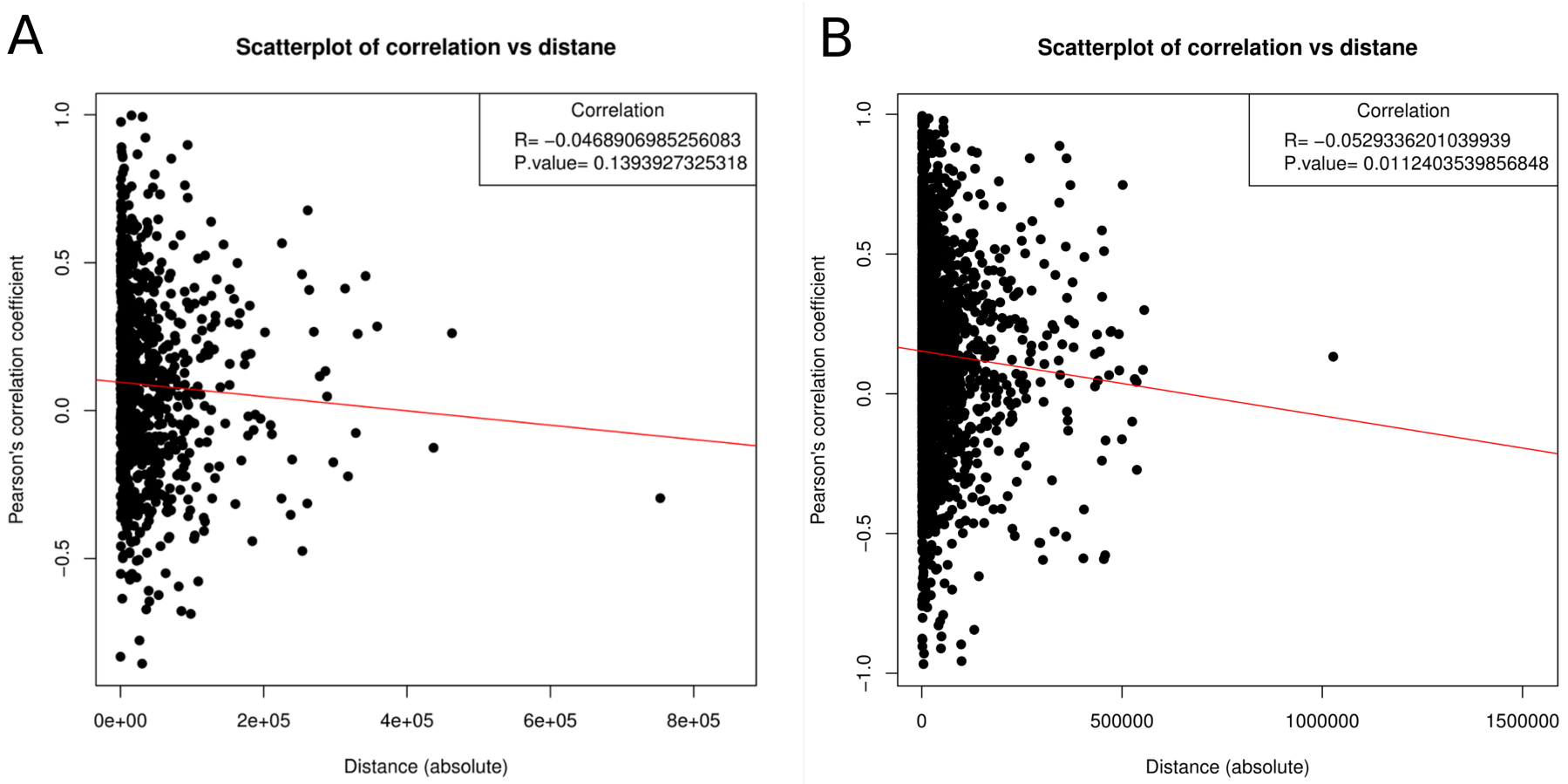
S.Figure 3: Neuron sub-type specificity of genes and LncRNAs (annotated and novel). A: PCA plot of the expression of ENSEMBL annotated genes only. Neuron sub-types are colour coded. B: Kernel density of the Tau specificity metric. The first peak of the bimodal distribution represents ubiquitous genes. The second peak represents neuron sub-type specific genes (tau > 0.8).



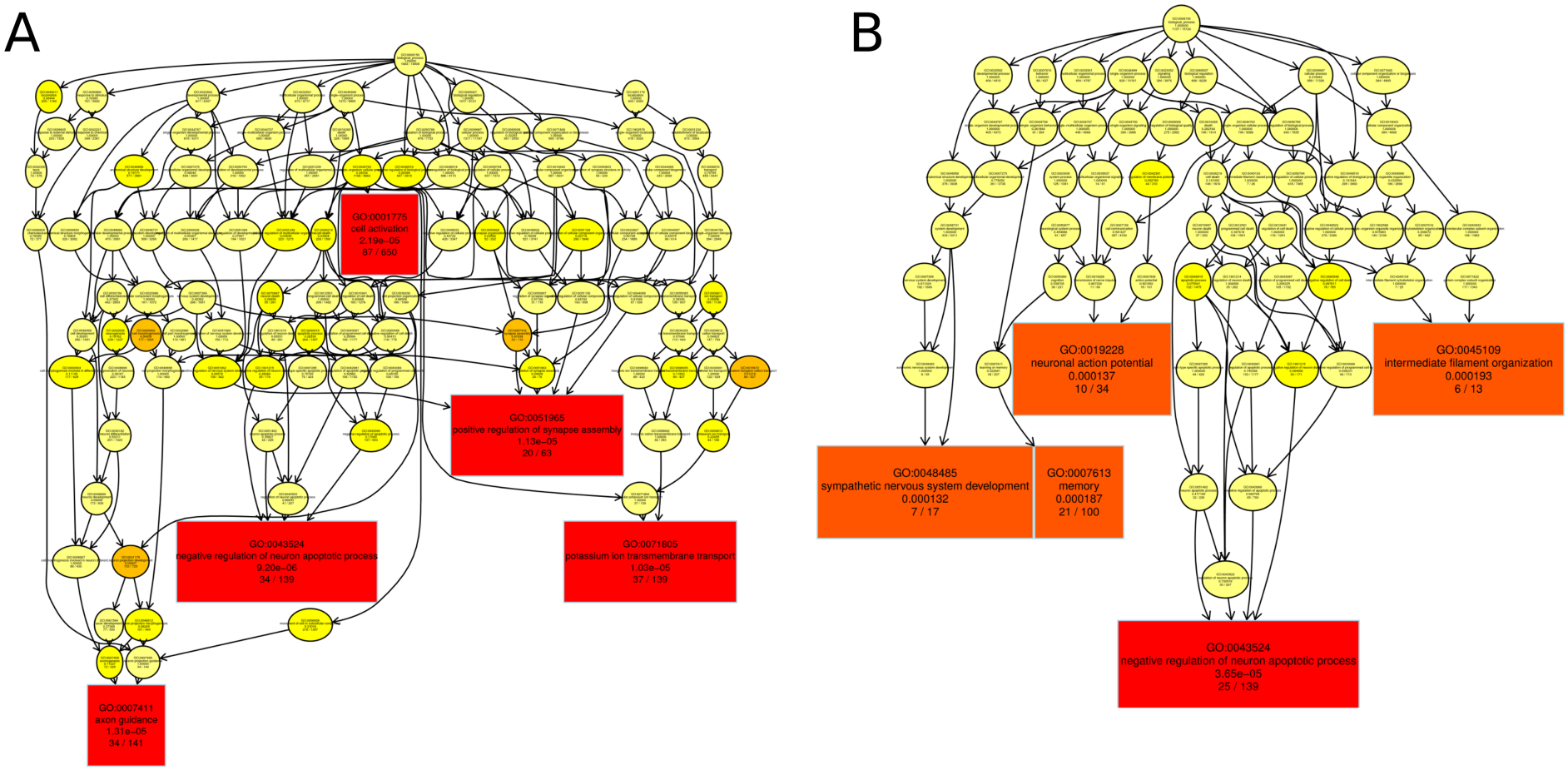
S.Figure 4: A: PCA plot of samples based on the expression of novel LncRNAs and ENSEMBL genes (1st 10000 novel LncRNAs and ENSEMBL genes, ranked by their SD) B, C: Correlation of expression vs distance for ENSEMBL annotated (A) and novel LncRNAs (B) in IPSC and IPSC derived sensory neurons.



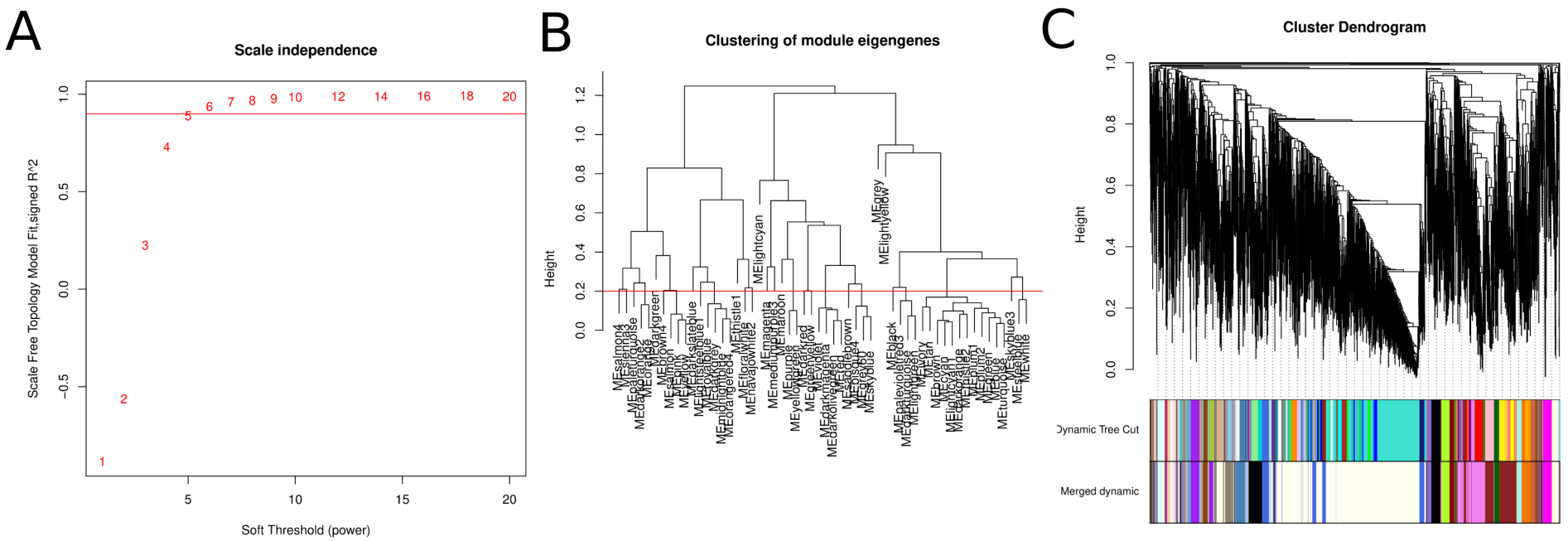
S.Figure 5: Over-representation analysis of GO biological process (BP) terms based on the significantly DE ENSEMBL genes (p.value < 0.05) SNT vs SHAM using the weighted Fisher exact test and the weighted KS test. GO sub-graph leading to the top 5 significantly enriched terms in Wistar rat.



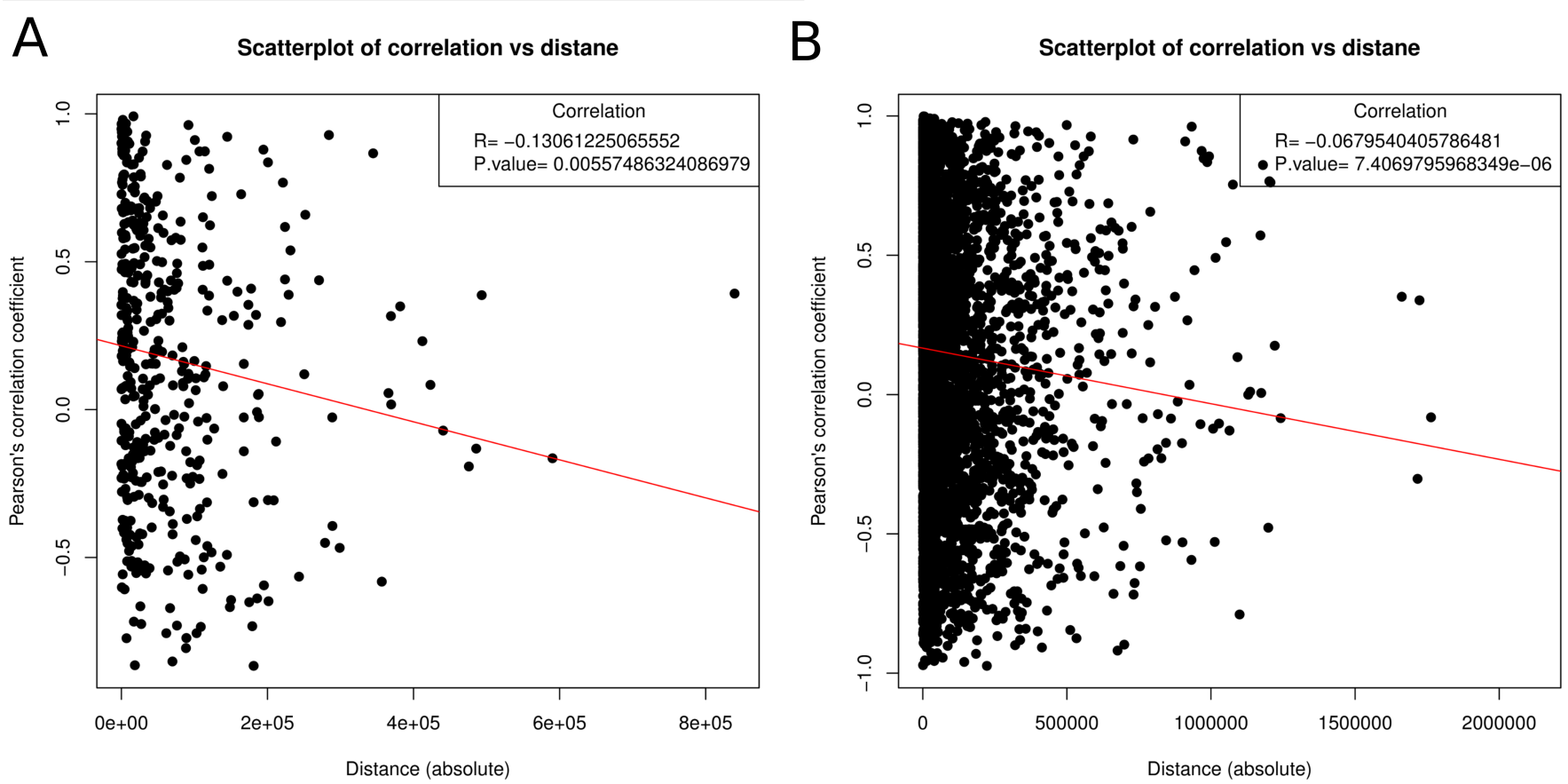
S.Figure 6: Correlation of expression vs distance for ENSEMBL annotated (A) and novel LncRNAs (B) in rat DRG.



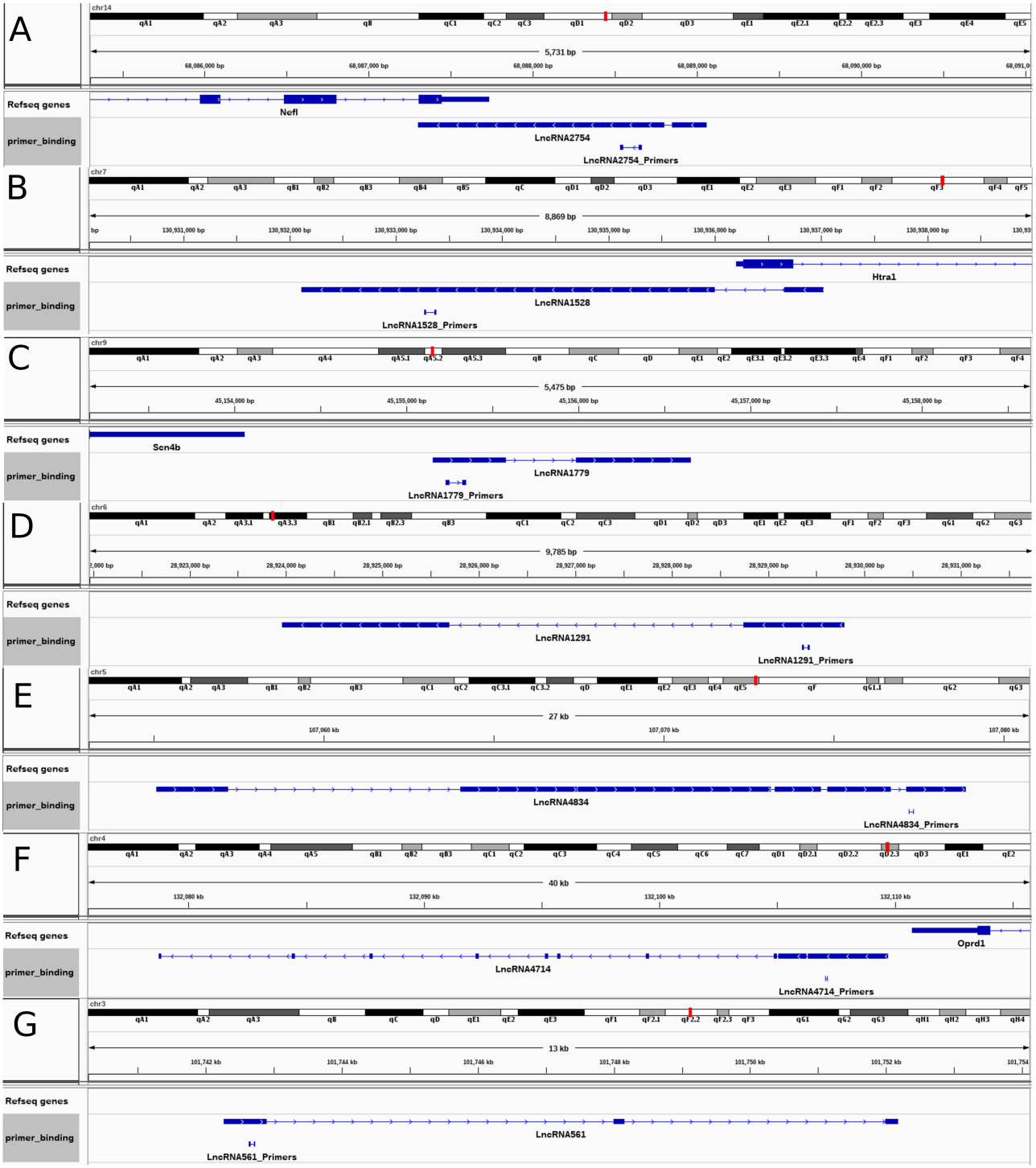
S.Figure 7: Over-representation analysis of GO biological process (BP) terms based on the significantly DE ENSEMBL genes (p.value < 0.05) SNI vs SHAM using the weighted Fisher exact test and the weighted KS test. GO sub-graph leading to the top 5 significantly enriched terms in BALB/c mice (A) and B10.D2 mice (B).



S.Figure 8: A: Soft threshold pick for WGCNA network. Topology was scaled-free with a soft threshold of 5 (A), bi-correlation matrix was raised in the power of 5 before clustering. B: Hierarchical clustering of un-merged modules, merging threshold of 0.2 is plotted with a red line. C: Gene dendrogram and module identification using dynamic tree followed by similar module clustering.



S.Figure 9: Correlation of expression vs distance for ENSEMBL annotated (A) and novel LncRNAs (B) in mouse DRG.



S.Figure 10: Gene models and primer binding location of validated LncRNAs. For every panel, track "Refseq genes" has RefSeq annotations for the respective genomic coordinates. A: LncRNA2754. B: LncRNA1528. C: LncRNA1779. D: LncRNA1291. E: LncRNA4834. F: LncRNA4714. G: LncRNA561.

Supplementary tables

Rat RNA-seq experiment										
Lane	% GC	% GCmapped	σpos(%GC)	insert ± MAD	% exonic	% exon cov'ge	maxpos %N	%lowQ	%lowQ end	avgQ
3.1	51.0 ± 10.7	50.7 ± 10.0	3.97	151 ± 42	20.4	89.9	0.1	0	0	35.1
3.2	50.9 ± 11.0	50.5 ± 10.6	2.7	150 ± 42	20.5	89.9	0	0	0	34.5
4.1	53.1 ± 11.1	52.8 ± 10.6	3.52	153 ± 42	18.5	89	0.1	0	0	35
4.2	52.9 ± 11.4	52.5 ± 11.0	2.41	153 ± 42	18.5	89.3	0	0	0	34.5
1.1	50.7 ± 10.7	50.6 ± 10.3	4.3	154 ± 42	19.9	89.7	0.2	0	0	34.9
1.2	50.7 ± 11.1	50.4 ± 10.9	2.84	152 ± 42	20	89.8	0.1	0	0	33.3
2.1	51.1 ± 10.4	51.0 ± 10.0	3.74	158 ± 46	23.3	90.1	0.3	0	0	34.8
2.2	51.3 ± 10.8	51.0 ± 10.5	2.81	156 ± 44	23.4	90.3	0	0	0	

S.Table 1: RNA-sequencing quality metrics for rat DRG. GC content and percentage, insert size, Q score (Sanger qualities).

Rat RNA-seq depth				
	condition	Uniquely mapped reads	Pairs of properly paired reads	Coverage based on uniquely mapped reads
1	SHAM	120422768	35315197	9.36
2	SHAM	139555954	41749413	10.84
3	SHAM	109587232	34360632	8.51
4	SHAM	103491410	33174806	8.04
5	D21_SNT	119119050	32544853	9.26
6	D21_SNT	135313916	41347957	10.52
7	D21_SNT	116232532	74032392	9.03
8	D21_SNT	119358556	35772291	9.28

S.Table 2: Overview of the experimental design and number of uniquely mapped reads and genome coverage per sample for rat DRG.

Mouse RNA-seq experiment										
Lane	% GC	% GCmapped	σpos(%GC)	insert ± MAD	% exonic	% exon cov'ge	maxpos %N	%lowQ	%lowQend	avgQ
1.1	48.3 ± 9.6	47.8 ± 9.4	4.45	197 ± 59	20.8	59.9	0.3	0	0	34
1.2	48.3 ± 10.3	47.8 ± 10.0	2.61	195 ± 59	21.8	61.6	0.6	0	0	31
2.1	48.1 ± 9.5	47.7 ± 9.4	4.46	191 ± 54	20.7	58.6	1.1	0	0.2	33.9
2.2	48.2 ± 10.2	47.6 ± 10.0	2.62	190 ± 53	21.7	60.3	1	0	0.1	30.8
3.1	48.2 ± 9.6	47.7 ± 9.4	4.45	192 ± 54	20.7	58.7	1	0	0.2	33.9
3.2	48.2 ± 10.3	47.6 ± 10.0	2.62	190 ± 54	21.7	60.4	1.1	0	0.1	30.9
4.1	48.2 ± 9.6	47.7 ± 9.4	4.45	192 ± 55	20.7	58.8	2.4	0	0.1	33.9
4.2	48.2 ± 10.3	47.7 ± 10.0	2.61	191 ± 55	21.6	60.5	1.3	0	0.1	30.7
5.1	48.1 ± 9.6	47.7 ± 9.4	4.46	192 ± 55	20.7	58.4	0.4	0	0	32.5
5.2	48.2 ± 10.3	47.5 ± 10.0	2.58	191 ± 55	21.6	59.9	0.1	0	0	30

S.Table 3: RNA-sequencing quality metrics for mouse DRG. GC content and percentage, insert size, Q score (Sanger qualities).

Mouse RNA-seq depth							
	strain	sex	condition	group	Uniquely mapped reads	Pairs of properly paired reads	Coverage based on uniquely mapped reads
1	BALB.c	M	SHAM	BALB.c_SHAM	128150964	58253817	7.25
2	BALB.c	M	SHAM	BALB.c_SHAM	157996400	70905687	8.83
3	BALB.c	F	SHAM	BALB.c_SHAM	153996898	71094756	8.70
4	BALB.c	F	SHAM	BALB.c_SHAM	141881266	62703319	8.02
5	BALB.c	F	SHAM	BALB.c_SHAM	133222250	60118606	7.54
6	B10.D2	M	SHAM	B10.D2_SHAM	129686932	59135611	7.33
7	B10.D2	M	SHAM	B10.D2_SHAM	131655420	59573126	7.44
8	B10.D2	F	SHAM	B10.D2_SHAM	143948956	65180670	8.14
9	BALB.c	M	SHAM	BALB.c_SHAM	137103922	62141859	7.75
10	B10.D2	F	SHAM	B10.D2_SHAM	148176972	67081783	8.37
11	B10.D2	M	SHAM	B10.D2_SHAM	155847954	71180013	8.81
12	BALB.c	M	SNI	BALB.c_SNI	116681670	53381176	6.60
13	BALB.c	F	SNI	BALB.c_SNI	152372720	69878368	8.61
14	BALB.c	F	SNI	BALB.c_SNI	149408206	67104230	8.45
15	B10.D2	F	SNI	B10.D2_SNI	140526254	63262815	7.94
16	BALB.c	M	SNI	BALB.c_SNI	135574098	61279835	7.66
17	B10.D2	F	SNI	B10.D2_SNI	153651302	69507207	8.69
18	B10.D2	F	SNI	B10.D2_SNI	158587850	72285022	8.97
19	B10.D2	M	SNI	B10.D2_SNI	146210692	65741153	8.27
20	B10.D2	M	SNI	B10.D2_SNI	138142970	62903003	7.81

S.Table 4: Overview of the experimental design and number of uniquely mapped reads and genome coverage per sample for mouse DRG.

Mouse DRG novel LncRNAs antisense of pain genes			
LncRNA name (coordinates)	LncRNA ID	Pain gene ENSEMBL ID	Pain Gene symbol
11:55394500-55395410(+)	LncRNA2153	ENSMUSG00000018593	Sparc
11:70240212-70242700(+)	LncRNA2170	ENSMUSG00000000320	Alox12
11:73297432-73307708(-)	LncRNA2276	ENSMUSG00000043029	Trpv3
12:113144963-113145499(-)	LncRNA2486	ENSMUSG00000006356	Crip2
14:63141461-63144107(-)	LncRNA2750	ENSMUSG00000021939	Ctsb
16:35299876-35310125(-)	LncRNA3043	ENSMUSG00000022840	Adcy5
2:131936420-131936829(-)	LncRNA467	ENSMUSG00000079037	Prnp
5:43868670-43869133(-)	LncRNA1094	ENSMUSG00000029084	Cd38
6:92158445-92168511(-)	LncRNA5190	ENSMUSG00000005893	Nr2c2
6:118167896-118169229(+)	LncRNA5075	ENSMUSG00000030110	Ret
7:114631428-114631907(+)	LncRNA1441	ENSMUSG00000030669	Calca
7:114635603-114715554(+)	LncRNA1442	ENSMUSG00000030669	Calca
9:119730828-119763742(+)	LncRNA5774	ENSMUSG00000034115	Scn11a

S.Table 5: Novel LncRNAs antisense of pain genes in mouse DRG.

Rat DRG novel LncRNAs antisense of pain genes			
LncRNA name (coordinates)	LncRNA ID	Pain gene ENSEMBL ID	Pain Gene symbol
1:36353933-36354161(-)	LncRNA623	ENSRNOG00000017601	Srd5a1
1:84239119-84313048(-)	LncRNA711	ENSRNOG00000018369	Prx
1:84321378-84322196(-)	LncRNA712	ENSRNOG00000018369	Prx
1:184188353-184188948(+)	LncRNA364	ENSRNOG00000011130	Calca
1:237907633-237910990(+)	LncRNA466	ENSRNOG00000017469	Anxa1
2:44293038-44314681(-)	LncRNA1595	ENSRNOG00000013963	Il6st
2:187160101-187160384(+)	LncRNA6942	ENSRNOG00000013953	Ntrk1
2:193900841-193901227(-)	LncRNA1853	ENSRNOG00000023226	S100a10
4:9610766-9626799(-)	LncRNA2925	ENSRNOG00000021441	Reln
4:9627458-9634043(-)	LncRNA2926	ENSRNOG00000021441	Reln
4:157375967-157495368(+)	LncRNA2856	ENSRNOG00000016294	Cd4
5:3737558-3783928(-)	LncRNA3507	ENSRNOG00000007354	Trpa1
6:103616022-103687750(+)	LncRNA8506	ENSRNOG00000006599	NA
7:41475712-41475995(-)	LncRNA4642	ENSRNOG00000023896	Dusp6
7:126680098-126683214(-)	LncRNA4757	ENSRNOG00000021463	Ppara
8:58755493-58755777(+)	LncRNA4864	ENSRNOG00000000196	Cyp19a1
8:75723255-75723599(-)	LncRNA5114	ENSRNOG00000010362	Anxa2
8:95968358-95969347(-)	LncRNA5148	ENSRNOG00000011071	Nt5e
8:96004232-96017662(-)	LncRNA5149	ENSRNOG00000011071	Nt5e

S.Table 6: Novel LncRNAs antisense of pain genes in rat DRG.

Novel intergenic LncRNAs in mouse DRG with a pain gene as their closest genomic neighbour				
LncRNA name (coordinates)	LncRNA ID	Pain gene ENSEMBL ID	Pain Gene symbol	Distance
11:63130153-63186753(+)	LncRNA2166	ENSMUSG00000018217	Pmp22	-1018
11:73310429-73310720(+)	LncRNA2172	ENSMUSG00000043029	Trpv3	-10066
11:97107759-97174736(-)	LncRNA6282	ENSMUSG00000001441	Npepps	-31106
12:104438172-104438855(+)	LncRNA2411	ENSMUSG00000021091	Serpina3n	-23843
13:55461928-55463245(+)	LncRNA6551	ENSMUSG00000074886	Grk6	-1001
15:72507982-72510044(-)	LncRNA7041	ENSMUSG00000036760	Kcnk9	-2075
16:93008313-93008946(-)	LncRNA3084	ENSMUSG00000022952	Runx1	182164
17:13019120-13032557(+)	LncRNA7242	ENSMUSG00000006818	Sod2	-1001
19:16389521-16404555(+)	LncRNA3423	ENSMUSG00000024639	Gnaq	-2058
19:22075141-22084363(+)	LncRNA3427	ENSMUSG00000052387	Trpm3	54756
19:22110149-22121663(+)	LncRNA3428	ENSMUSG00000052387	Trpm3	17456
19:59238371-59246400(+)	LncRNA3468	ENSMUSG00000040901	Kcnk18	-1001
19:60213784-60241092(-)	LncRNA7691	ENSMUSG00000045052	Prlhr	-225641
19:61096985-61110801(+)	LncRNA3471	ENSMUSG00000003228	Grk5	-9619
2:148398192-148412506(+)	LncRNA340	ENSMUSG00000037014	Sstr4	-13691
4:132078749-132109721(-)	LncRNA4714	ENSMUSG00000050511	Oprd1	-1005
5:35200106-35208079(+)	LncRNA970	ENSMUSG00000045318	Adra2c	67197
6:99652802-99691088(-)	LncRNA5199	ENSMUSG00000030069	Prok2	-20211
6:126602358-126635461(-)	LncRNA1339	ENSMUSG00000047976	Kcna1	-4936
7:100985647-100995563(-)	LncRNA5452	ENSMUSG00000032860	P2ry2	-1005
9:81600537-81627190(-)	LncRNA1916	ENSMUSG00000049511	Htr1b	-25730
9:119598332-119607455(-)	LncRNA1947	ENSMUSG00000034533	Scn10a	-1001
9:119740632-119746657(-)	LncRNA1950	ENSMUSG00000034115	Scn11a	-7102

S.Table 7: Novel intergenic LncRNAs (LincRNAs) with a pain gene as its closest genomic neighbour in mouse DRG. Distance between the LincRNA and the pain gene is given in genomic bases. Positive when downstream, negative when upstream.

Novel intergenic LncRNAs in rat DRG with a pain gene as their closest genomic neighbour				
LncRNA name (coordinates)	LncRNA ID	Pain gene ENSEMBL ID	Pain Gene symbol	Distance
LncRNA72	1:43357117-43357455(+)	ENSRNOG00000018191	Oprm1	97348
LncRNA6515	1:80611565-80611819(-)	ENSRNOG00000018454	Apoe	-1076
LncRNA218	1:104559137-104559721(+)	ENSRNOG00000014530	Nav2	16868
LncRNA6555	1:124041527-124042593(-)	ENSRNOG00000010853	Chrna7	2331
LncRNA254	1:126707509-126707906(+)	ENSRNOG00000011526	Pcsk6	41602
LncRNA255	1:126708538-126709260(+)	ENSRNOG00000011526	Pcsk6	40248
LncRNA6263	1:156532112-156551299(+)	ENSRNOG00000022635	Dlg2	1029
LncRNA6602	1:166038470-166043910(-)	ENSRNOG00000019283	P2ry2	1046
LncRNA6668	1:216972238-216972591(-)	ENSRNOG00000054917	Mrgpre	1055
LncRNA481	1:239947027-239950543(+)	ENSRNOG00000027770	Trpm3	404606
LncRNA482	1:239976336-239989419(+)	ENSRNOG00000027770	Trpm3	365730
LncRNA483	1:240230190-240236136(+)	ENSRNOG00000027770	Trpm3	119013
LncRNA6358	1:240351693-240354147(+)	ENSRNOG00000027770	Trpm3	1002
LncRNA6911	2:153776661-153802312(+)	ENSRNOG00000009514	Mme	1037
LncRNA1830	2:179580859-179582769(-)	ENSRNOG00000054204	Gria2	-1539
LncRNA1845	2:188962007-188968672(-)	ENSRNOG00000020778	Chrnb2	-119898
LncRNA1406	2:194005474-194073083(+)	ENSRNOG00000023226	S100a10	-104238
LncRNA6959	2:197654221-197654597(+)	ENSRNOG00000021157	NA	1189
LncRNA7285	2:240868804-240888490(-)	ENSRNOG00000023258	Nfkb1	2115
LncRNA7011	2:264042296-264114519(+)	ENSRNOG00000010325	Ptger3	-62598
LncRNA7535	3:11450542-11474804(-)	ENSRNOG00000013973	Lcn2	32996
LncRNA2321	3:12776875-12878720(-)	ENSRNOG00000017019	Lmx1b	90006
LncRNA7673	3:124082535-124082771(-)	ENSRNOG00000021256	Adra1d	-46787
LncRNA2203	3:127978865-127983805(+)	ENSRNOG00000004810	Plcb1	171264
LncRNA2204	3:128113557-128113971(+)	ENSRNOG00000004810	Plcb1	41098
LncRNA10015	3:128444601-128445050(+)	ENSRNOG00000004810	Plcb1	-25056
LncRNA7455	3:128602478-128688346(+)	ENSRNOG00000033119	Plcb4	68453
LncRNA2494	3:129383147-129383590(-)	ENSRNOG00000005509	Pak7	25656
LncRNA2300	3:177223778-177224015(+)	ENSRNOG00000016768	Oprl1	1722
LncRNA2301	3:177232664-177239901(+)	ENSRNOG00000016768	Oprl1	-1001
LncRNA3096	4:123860192-123866644(-)	ENSRNOG00000009019	Slc6a6	146728
LncRNA7857	4:146260088-146275790(+)	ENSRNOG00000006527	Slc6a1	1072
LncRNA2828	4:146468169-146468539(+)	ENSRNOG00000007420	Hrh1	-11095
LncRNA3127	4:150249008-150249315(-)	ENSRNOG00000014751	Ret	4636
LncRNA3172	4:169541564-169552161(-)	ENSRNOG00000008766	Grin2b	-8226
LncRNA3196	5:3754456-3761754(+)	ENSRNOG00000007354	Trpa1	21493
LncRNA3197	5:3763292-3763707(+)	ENSRNOG00000007354	Trpa1	19540
LncRNA3606	5:75546721-75556028(-)	ENSRNOG00000013656	Lpar1	-1014
LncRNA3390	5:123528297-123529199(+)	ENSRNOG00000007410	Dab1	375967
LncRNA3391	5:123824591-123826090(+)	ENSRNOG00000007410	Dab1	79076
LncRNA3788	6:9686274-9690983(+)	ENSRNOG00000015603	NA	99439
LncRNA3844	6:43815689-43817405(+)	ENSRNOG00000054259	Klf11	12540
LncRNA8727	6:128435212-128451897(-)	ENSRNOG00000010711	Dicer1	1029
LncRNA8804	7:41582475-41627947(+)	ENSRNOG00000023896	Dusp6	-103083
LncRNA4847	8:50186956-50328931(+)	ENSRNOG00000016847	Bace1	-24595
LncRNA5076	8:53966266-53983337(-)	ENSRNOG00000031890	Ncam1	64908
LncRNA5077	8:54059819-54065375(-)	ENSRNOG00000031890	Ncam1	158461
LncRNA4858	8:54990257-54990498(+)	ENSRNOG00000009848	Il18	3361
LncRNA5105	8:69165792-69169147(-)	ENSRNOG00000010176	Map2k1	1034
LncRNA5135	8:89144412-89149303(-)	ENSRNOG00000013042	Htr1b	13421

LncRNA5136	8:89165979-89170397(-)	ENSRNOG00000013042	Htr1b	34988
LncRNA5310	9:62102519-62106196(+)	ENSRNOG00000032659	Plcl1	185209
LncRNA5311	9:62220850-62223020(+)	ENSRNOG00000032659	Plcl1	68385
LncRNA9608	9:70788920-70796020(-)	ENSRNOG00000012961	NA	1007
LncRNA5357	9:98376133-98376670(+)	ENSRNOG00000019926	Ramp1	-11928
LncRNA5660	X:27864020-27864435(+)	ENSRNOG00000004118	Frmpd4	208391
LncRNA5790	X:107409312-107494462(+)	ENSRNOG00000002419	Plp1	1610

S.Table 8: Novel intergenic LncRNAs (LincRNAs) with a pain gene as its closest genomic neighbour in rat DRG. Distance between the LincRNA and the pain gene is given in genomic bases. Positive when downstream, negative when upstream.

Novel LncRNAs in mouse DRG antisense of voltage gated ion channels			
LncRNA name (coordinates)	LncRNA ID	Sense gene ENSEMBL ID	Sense Gene symbol
Potassium Channels			
1:87332176-87402632(+)	LncRNA48	ENSMUSG00000079436	Kcnj13
11:33844702-33880117(-)	LncRNA2252	ENSMUSG00000020155	Kcnmb1
15:99241659-99282134(-)	LncRNA7074	ENSMUSG00000037579	Kcnh3
3:32469532-32472694(+)	LncRNA522	ENSMUSG00000091091	Kcnmb3
3:65067032-65110222(-)	LncRNA633	ENSMUSG00000027827	Kcnab1
3:107111107-107111410(-)	LncRNA662	ENSMUSG00000040724	Kcna2
4:152391227-152391657(+)	LncRNA818	ENSMUSG00000028931	Kcnab2
7:143176517-143186434(-)	LncRNA1538	ENSMUSG00000009545	Kcnq1
Sodium Channels			
2:66665266-66676581(+)	LncRNA266	ENSMUSG00000034810	Scn7a
9:119730828-119763742(+)	LncRNA5774	ENSMUSG00000034115	Scn11a
TRP Channels			
11:73297432-73307708(-)	LncRNA2276	ENSMUSG00000043029	Trpv3

S.Table 9: Novel LncRNAs antisense of voltage-gated ion channels in mouse DRG.

Novel LncRNAs in rat DRG antisense of voltage gated ion channels			
LncRNA name (coordinates)	LncRNA ID	Sense gene ENSEMBL ID	Sense Gene symbol
Calcium channels			
4:151402902-151409369(-)	LncRNA3129	ENSRNOG00000008031	Cacna2d4
Potassium Channels			
1:216375194-216380918(-)	LncRNA916	ENSRNOG00000020532	Kcnq1
2:209895961-209921066(-)	LncRNA1883	ENSRNOG00000050416	Kcna10
4:159263373-159286693(+)	LncRNA2859	ENSRNOG00000052486	Kcna6
6:36819126-36821854(+)	LncRNA3837	ENSRNOG00000004899	Kcns3
Sodium Channels			
3:52509234-52510407(+)	LncRNA2066	ENSRNOG00000053122	Scn1a
3:52782292-52783549(+)	LncRNA7353	ENSRNOG00000029342	Scn7a
8:49415408-49434636(-)	LncRNA9305	ENSRNOG00000016221	Scn2b
Chloride Channels			
5:164802402-164871681(+)	LncRNA8228	ENSRNOG00000008345	Clcn6
TRP Channels			
5:3737558-3783928(-)	LncRNA3507	ENSRNOG00000007354	Trpa1

S.Table 10: Novel LncRNAs antisense of voltage-gated ion channels in rat DRG.

DE LncRNAs antisense of DE pain genes with opposite Log2 fold changes							
LncRNA ID	LncRNA name (coordinates)	Sense gene ID	Gene symbol	LncRNA Log2 fold change	LncRNA adj. p.value	Gene Log2 fold change	Gene adj. p.value
ENSG00000234377	RNF219-AS1	ENSG00000136160	EDNRB	5.92	< 0.001	-3.81	< 0.001
ENSG00000261762	NA	ENSG00000169684	CHRNA5	5.35	< 0.001	-2.43	< 0.001
ENSG00000215067	ALOX12-AS1	ENSG00000108839	ALOX12	1.77	< 0.001	-8.36	< 0.001
ENSG00000264107	NA	ENSG00000196712	NF1	-2.72	< 0.001	1.59	< 0.001
ENSG00000263766	NA	ENSG00000141279	NPEPPS	-0.99	0.049	1	< 0.001
ENSG00000204044	NA	ENSG00000124140	SLC12A5	-1.49	< 0.001	2.59	< 0.001
ENSG00000225756	DBH-AS1	ENSG00000123454	DBH	2.14	< 0.001	-5.31	< 0.001
LncRNA5804	8:90053955-90060834(+)	ENSG00000104327	CALB1	2.94	< 0.001	-4.78	< 0.001

S.Table 11: DE LncRNAs antisense of DE pain genes with opposite Log2 fold changes in IPSC vs IPSC derived sensory neurons. Log2 fold changes and p.values are for the comparison Neurons (AD2 parental line) vs IPSC AD2.

Expression changes of HAGLR LncRNA					
Organism and condition	Gene ID	baseMean	log2FoldChange	Adjusted p.value	symbol
Human neurons vs IPSc	ENSG00000224189	261.6	8.74	< 0.001	HAGLR
BALB/c mouse DRG SNI vs Sham	ENSMUSG00000075277	151.5	-0.46	0.004	Haglr
B10.D2 mouse DRG SNI vs Sham	ENSMUSG00000075277	151.5	-0.49	0.001	Haglr

S.Table 12: Expression changes of HAGLR LncRNA in mouse SNI vs Sham and human IPSC vs IPSC-derived sensory neurons.

DE LncRNAs antisense of DE pain genes with opposite Log2 fold changes in rat DRG							
LncRNA ID	LncRNA name (coordinates)	Sense gene ID	Gene symbol	LncRNA Log2 fold change	LncRNA adj. p.value	Gene Log2 fold change	Gene adj. p.value
LncRNA6781	2:26247635-26263258(+)	ENSRNOG00000025406	Iqgap2	-2.57	< 0.001	0.54	0.008
LncRNA7193	2:149445646-149481258(-)	ENSRNOG00000010680	Med12l	0.68	0.01	-0.75	< 0.001
LncRNA1982	3:5991884-5997326(+)	ENSRNOG00000007681	Brd3	-0.97	< 0.001	0.46	0.01
LncRNA2287	3:171273204-171276292(+)	ENSRNOG00000006314	Zbp1	-3.42	< 0.001	1.42	0.006
LncRNA3645	5:107733448-107751000(-)	ENSRNOG00000006615	Mtap	1.27	0.04	-0.51	0.005
LncRNA3709	5:144326425-144423068(-)	ENSRNOG00000010841	Col8a2	-0.61	< 0.001	0.43	0.002
LncRNA4282	7:13751062-13755737(+)	ENSRNOG00000007509	Slc1a6	-1.7	0.002	1.58	< 0.001
LncRNA9091	7:141680351-141760307(-)	ENSRNOG00000056106	Dip2b	-0.79	0.03	0.63	0.002
LncRNA5107	8:70169463-70191926(-)	ENSRNOG00000010634	Megf11	-0.88	< 0.001	2.16	< 0.001

S.Table 13: DE LncRNAs antisense of DE pain genes with opposite Log2 fold changes in rat DRG SNT vs Sham.

Intergenic LncRNAs close and highly correlated with pain genes									
LncRNA ID	LncRNA name (coordinates)	Closest gene symbol	Distance	LncRNA Log2 fold change	LncRNA adj. p.value	Gene Log2 fold change	Gene adj. p.value	Correlation	Cor. Adj. p.value
LncRNA1830	2:179580859-179582769(-)	Gria2	-1539	-2.31	< 0.001	-1.89	< 0.001	0.98	0.006
LncRNA2301	3:177232664-177239901(+)	Oprl1	-1001	-0.92	0.002	-0.68	0.009	0.98	0.006
LncRNA3172	4:169541564-169552161(-)	Grin2b	-8226	1.31	< 0.001	1.20	< 0.001	0.98	0.006
LncRNA3196	5:3754456-3761754(+)	Trpa1	21493	-2.43	< 0.001	-0.97	0.03	0.94	0.02
ENSRNOG00000062160	NA	Chrnb4	-555	-3.58	< 0.001	-2.59	0.24	0.94	0.02

S.Table 14: Intergenic LncRNAs DE in rat DRG SNT vs Sham with a pain genes as their closest genomic neighbour and a highly correlated expression. Pearson's R correlation coefficient has been calculated on regularised log2 transform counts.

Balb/c						
LncRNA ID	LncRNA symbol (coordinates)	Sense gene symbol	LncRNA Log2 Fold change	LncRNA adj. p.value	Gene Log2 Fold change	Gene adj. p.value
ENSMUSG00000097649	NA	Rftn2	0.49	0.04	0.48	< 0.001
ENSMUSG00000104677	NA	Nbea	-0.31	0.002	0.15	0.01
ENSMUSG00000097596	NA	Kcna6	-1.00	< 0.001	0.21	0.04
ENSMUSG00000100600	A230077H06Rik	Vstm2b	-1.83	< 0.001	-1.40	< 0.001
LncRNA6043	10:97669599-97680412(-)	Epyc	-0.75	< 0.001	0.89	0.01
LncRNA2252	11:33844702-33880117(-)	Kcnmb1	-0.44	< 0.001	-0.33	0.004
LncRNA2754	14:68087301-68089062(-)	Nefl	-0.53	0.002	-0.43	< 0.001
LncRNA531	3:56001321-56111248(+)	Nbea	-0.24	0.001	0.15	0.01
LncRNA4536	4:109395434-109413760(+)	NA	-0.33	0.03	-0.56	< 0.001
LncRNA963	5:30880183-30935677(+)	Cgref1	-1.2	0.04	0.24	0.002
B10.D2						
LncRNA ID	LncRNA symbol (coordinates)	Sense gene symbol	LncRNA Log2 Fold change	LncRNA adj. p.value	Gene Log2 Fold change	Gene adj. p.value
ENSMUSG00000100600	A230077H06Rik	Vstm2b	-1.74	< 0.001	-2.08	< 0.001
LncRNA203	1:171379256-171380555(-)	Nectin4	-0.73	0.02	-0.44	< 0.001
LncRNA6043	10:97669599-97680412(-)	Epyc	-0.73	< 0.001	1.07	0.001
LncRNA2754	14:68087301-68089062(-)	Nefl	-0.49	0.006	-0.47	< 0.001
LncRNA1528	7:130932111-130937027(-)	Htra1	-0.45	0.02	-0.23	0.001

S.Table 15: Antisense LncRNAs that were significantly DE on the opposite strand of a significantly DE gene in BALB/c and B10.D2 mouse strain DRG SNI vs Sham.

LncRNAs antisense of Potassium channels						
LncRNA ID	LncRNA symbol (coordinates)	Sense gene symbol	LncRNA Log2 Fold change	LncRNA adj. p.value	Gene Log2 Fold change	Gene adj. p.value
BALB/c Mouse DRG SNI vs Sham						
ENSMUSG00000097596	NA	Kcna6	-1.00	< 0.001	0.21	0.04
LncRNA2252	11:33844702-33880117(-)	Kcnmb1	-0.44	< 0.001	-0.33	0.004
B10.D2 Mouse DRG SNI vs Sham						
ENSMUSG00000097596	NA	Kcna6	-1.09	< 0.001	0.02	0.9
LncRNA2252	11:33844702-33880117(-)	Kcnmb1	-0.36	< 0.001	-0.2	0.18
Rat DRG SNT vs Sham						
LncRNA916	1:216375194-216380918(-)	Kcnq1	-0.82	0.03	-0.26	0.68
LncRNA2859	4:159263373-159286693(+)	Kcna6	-2.66	0.01	0.17	0.64
LncRNA3837	6:36819126-36821854(+)	Kcns3	-1.57	0.01	-2.66	< 0.001
Human IPSC-derived neurons vs IPSC						
ENSG00000226009	KCNIP2-AS1	KCNIP2	-1.7	0.009	3	< 0.001
ENSG00000269821	KCNQ1OT1	KCNQ1	1.39	< 0.001	-4.56	< 0.001
ENSG00000267365	KCNJ2-AS1	KCNJ2	-0.99	0.03	0.28	0.3
ENSG00000244558	NA	KCNK15	4.32	< 0.001	-0.11	0.8
ENSG00000177410	ZFAS1	KCNB1	-1.42	< 0.001	5.8	< 0.001
LncRNA2321	10:77248763-77249835(+)	KCNMA1	7.35	< 0.001	4.25	< 0.001
LncRNA1584	6:72614790-72622130(-)	KCNQ1	4.05	< 0.001	2.82	< 0.001
LncRNAs antisense of Sodium channels						
LncRNA ID	LncRNA symbol (coordinates)	Sense gene symbol	LncRNA Log2 Fold change	LncRNA adj. p.value	Gene Log2 Fold change	Gene adj. p.value
BALB/c Mouse DRG SNI vs Sham						
ENSMUSG00000087301	NA	Scn9a	-1.11	< 0.001	- 0.14	0.1
B10.D2 Mouse DRG SNI vs Sham						
ENSMUSG00000087301	NA	Scn9a	-1.36	< 0.001	- 0.02	0.9
Human IPSC-derived neurons vs IPSC						
ENSG00000269890	NA	OBSCN	3.12	< 0.001	-0.45	0.006
ENSG00000269934	NA	OBSCN	3.68	0.004	-0.45	0.006
ENSG00000236107	LOC101929680	SCN9A	5.54	< 0.001	6.14	< 0.001
LncRNA2728	12:51591357-51592521(-)	SCN8A	4.76	< 0.001	1.9	< 0.01

S.Table 16: LncRNAs antisense of voltage gated sodium and potassium channels that were significantly DE.

Intergenic LncRNAs close and highly correlated with pain genes									
LncRNA ID	LncRNA name (coordinates)	Gene symbol	Distance	LncRNA Log2 fold change	LncRNA adj. p.value	Gene Log2 fold change	Gene adj. p.value	Correl ation	Cor. Adj. p.value
LncRNA3468	19:59238371-59246400(+)	Kcnk18	-1001	-0.46	< 0.001	-0.28	0.008	0.97	< 0.001
				-0.19	0.19	-0.08	0.6		
LncRNA4714	4:132078749-132109721(-)	Oprd1	-1005	-0.77	< 0.001	-0.70	< 0.001	0.96	< 0.001
				-0.66	< 0.001	-0.8	< 0.001		

S.Table 17: Intergenic LncRNAs DE in mouse DRG SNI vs Sham with a pain genes as their closest genomic neighbour and a highly correlated expression. Pearson's R correlation coefficient has been calculated on regularised log2 transform counts.

LncRNA	Forward	Reverse	RT primer
Mouse			
LncRNA2754	GTAGTGCAAGCTTTGTCGTGG	TTGCGTGCTGCATTGGTATT	TTGCGTGCTGCATTGGTATT
LncRNA1528	TTAACTCCATGGCTCTCGGC	GCCTAAGGCAGGTCACACAT	-
LncRNA1779	CCTGGTGGCCATAAGGTGAG	CAGAGCATTGGGGGCTACAA	-
LncRNA1291	TCGCAGACCTCACTACCTTC	GGAGGGGTTATGTTTCCTGGAT	-
LncRNA4834	AGGCACGATGTCTGAAGCAA	TGGGAGGAGCAGTGTTAGGA	-
LncRNA4714	ATGCACAGCCAACAAACACTC	ATCCTCTCCCCTGAACCTCAT	-
LncRNA561	ATGAATGCAGCCTGACCACT	CATTCTCAGCAGGGCCAGTA	-
Housekeeping			
HPRT1	GTCCTGTGGCCATCTGCCTAG	TGGGGACGCAGCAACTGACA	TGGGGACGCAGCAACTGACA
Human			
HAGLR	CGCCCTTTCTGACCTGCTTA	TGGCAGTCGTCTGGACATTC	-
Housekeeping			
GADPH	AGGGCTGCTTTTAACTCTGGT	CCCCACTTGATTTTGGAGGGA	-
YWHAZ	CCTGCATGAAGTCTGTAACTGAG	GACCTACGGGCTCCTACAACA	-

S.Table 18: qPCR primers used for assessing the relative expression changes of novel LncRNAs and HAGLR in mouse DRG and human IPSC.

Supplementary data

http://doi.org/10.6084/m9.figshare.6508205

S. Data 1: Novel LncRNAs in mouse and rat, in syntenically conserved regions between human, mouse and rat.

S. Data 2: Novel LncRNAs in mouse DRG antisense of orthologous genes in mouse and rat.

S. Data 3: Novel LncRNAs in rat DRG antisense of orthologous genes in mouse and rat.

S. Data 4: Neuron sub-type specificity of ENSEMBL genes and novel LncRNAs. Neuron sub-type index is 1. MHN, 2. MHN (MI, IS), 3. C-LTMR, 4. MHN (IS), 5. MHN (IS), 6. MHN, 7. MHN (NS), 8. MR, 9. MHN, 10. MR. Average log2 expression holds the average expression in the neuron sub-type where the gene or LncRNAs was higher expressed. Expression SEM hold the standard error of the mean for the average expression estimator.

S. Data 5: DE analysis results and fpkm values of novel LncRNAs for IPSC-derived sensory neurons vs IPSC. Results are DESeq2 determined for the whole gene set of ENSEMBL genes and novel LncRNAs. Experimental design is ~ cell_line*condition.

S. Data 6: Expression data for all antisense LncRNAs in IPSC-derived sensory neurons vs IPSC.

S. Data 7: Expression data for all intergenic LncRNAs in IPSC-derived sensory neurons vs IPSC.

S. Data 8: Expression data for all LncRNAs antisense of pain genes in IPSC-derived sensory neurons vs IPSC.

S. Data 9: Expression data for all intergenic LncRNAs adjacent to pain genes in IPSC-derived sensory neurons vs IPSC.

S. Data 10: Novel LncRNAs in human IPSC and mouse, in syntenically conserved regions between human, mouse and rat.

S. Data 11: Novel LncRNAs in human IPSC antisense of orthologous genes in human and mouse.

S. Data 12: DE analysis results and fpkm values of novel LncRNAs for rat DRG SNT vs Sham. Results are DESeq2 determined for the whole gene set of ENSEMBL genes and novel LncRNAs. Experimental design is ~ condition.

S. Data 13: Expression data for all antisense LncRNAs in rat DRG SNT vs Sham.

S. Data 14: Expression data for all intergenic LncRNAs rat DRG SNT vs Sham.

S. Data 15: DE analysis results and fpkm values of novel LncRNAs for BALB/c and B10.D2 mouse DRG SNI vs Sham. Results are DESeq2 determined for the whole gene

set of ENSEMBL genes and novel LncRNAs. Experimental design is ~ sex + strain*condition.

S. Data 16: Expression data for all antisense LncRNAs in mouse DRG SNI vs Sham.

S. Data 17: Expression data for all intergenic LncRNAs mouse DRG SNI vs Sham.

Supplementary Bibliography

Aboyoun, P., Pages, H., Lawrence, M., 2013. GenomicRanges: Representation and manipulation of genomic intervals.

Aho, A.V., Kernighan, B.W., Weinberger, P.J., 1978. Awk – A Pattern Scanning and Processing Language (Second Edition).

Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169. https://doi.org/10.1093/bioinformatics/btu638

Carlson, M., Pages, H., Aboyoun, P., Falcon, S., Morgan, M., Sarkar, D., Lawrence, M., n.d. GenomicFeatures: Tools for making and manipulating transcript centric annotations.

Dowle, M., Srinivasan, A., Short, T., Saporta, S.L. with contributions from R., Antonyan, E., 2015. data.table: Extension of Data.frame.

Durinck, S., Spellman, P.T., Birney, E., Huber, W., 2009. Mapping Identifiers for the Integration of Genomic Datasets with the R/Bioconductor package biomaRt. Nat. Protoc. 4, 1184–1191. https://doi.org/10.1038/nprot.2009.97

Falcon, S., Gentleman, R., 2007. Using GOstats to test gene lists for GO term association. Bioinformatics 23, 257–8.

Lawrence, M., Gentleman, R., Carey, V., 2009. rtracklayer: an R package for interfacing with genome browsers. Bioinformatics 25, 1841–1842. https://doi.org/10.1093/bioinformatics/btp328

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M., Carey, V., 2013. Software for Computing and Annotating Genomic Ranges. PLoS Comput. Biol. 9. https://doi.org/10.1371/journal.pcbi.1003118

Morgan, M., Obenchain, V., Lang, M., Thompson, R., 2017a. BiocParallel: Bioconductor facilities for parallel evaluation.

Morgan, M., Pagès, H., Obenchain, V., Hayden, N., 2017b. Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import.

Morgan, M., Falcon, S., Gentleman, R., 2017. GSEABase: Gene set enrichment data structures and methods.

Pagès, H., Carlson, M., Falcon, S., Li, N., 2017. AnnotationDbi: Annotation Database Interface.

Pagès, H., Lawrence, M., Aboyoun, P., 2017. S4Vectors: S4 implementation of vector-like and list-like objects.

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033

van Brakel, J.-P., 2014. algorithm - Peak signal detection in realtime timeseries data - Stack Overflow [WWW Document]. URL https://stackoverflow.com/questions/22583391/peak-signal-detection-in-realtime-timeseries-data/22640362 (accessed 5.1.18).

Wang, L., Park, H.J., Dasari, S., Wang, S., Kocher, J.-P., Li, W., 2013. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res. 41, e74. https://doi.org/10.1093/nar/gkt006

Yin, T., Lawrence, M., Cook, D., 2017. biovizBase: Basic graphic utilities for visualization of genomic data.