Supplemental Material

Regulation of the ESC transcriptome by nuclear long noncoding RNAs

Jan H. Bergmann, Jingjing Li, Mélanie A. Eckersley-Maslin, Frank Rigo, Susan M. Freier

and David L. Spector

Contents:

Supplemental Methods

RNA Fluorescence *In Situ* Hybridization (FISH)

ESCs were grown on gelatinized, acid-cleaned #1.5 glass coverslips in the presence of feeders and fixed for 30 minutes in freshly-prepared 4% PFA (Electron Microscopy Sciences) diluted in D-PBS without $CaCl_2$ and $MgCl_2$ (Gibco, Life Technologies) and passed through 0.45 μm filter. Fixed cells were dehydrated and rehydrated through an ethanol gradient (50% - 75% - 100% - 75% - 50%- PBS) prior to permeabilization for 5 minutes in 0.5% Triton X-100. Protease QS treatment was performed at a 1:8,000 dilution. QuantiGene ViewRNA probe hybridizations were performed at 40°C in a gravity convection incubator (Precision Scientific), and incubation time of the pre-amplifier was extended to 2 hours. Nuclei were counter-stained with DAPI and coverslips mounted in p-Phenylenediamine-based anti-fade mounting medium (www.spectorlab.labsites.cshl.edu/protocols).

For standard RNA FISH of Firre, full-length cDNA (obtained by RACE and RT-PCR cloning, Supplemental Methods) was subject to nick translation in the presence of fluorescently labeled dUTP (Abbott Molecular). As negative control, linearized pGEM-T Easy vector (Promega) was used. Reactions were timed to generate probe fragments between 50-400 nt as determined by agarose gel electrophoresis, and RNA FISH was subsequently performed as described previously (Eckersley-Maslin et al. 2014), except that hybridization and post-hybridization washes were performed at 42°C.

Coverslips were imaged on a DeltaVision Core system (Applied Precision), based on an inverted IX-71 microscope stand (Olympus) equipped with a 60x U-PlanApo 1.40 NA oil immersion lens (Olympus). Images were captured at 1x1 binning using a CoolSNAP HQ CCD camera (Photometric) as z-stacks with a 0.2 μm spacing. Stage, shutter and exposure were controlled through SoftWorx (Applied Precision). Image deconvolution was performed in SoftWorx. Parameters for acquisition and post-acquisition processing for a given target across cell types and

hybridization controls were kept identical. For representation purposes, maximum intensity projections of image stacks are shown.

Nucleofection

For transfection of ASOs using nucleofection technology (Lonza), ESCs were harvested following soaking off of feeder cells for one hour, washed in D-PBS (Gibco, Life Technologies) and passed through a 70 μm nylon cell strainer (Corning). Cell count and viability was determined by trypan blue staining on a Countess automated cell counter (Life Technologies). For each reaction, $5x10^4$ viable cells were resuspended in P3 Primary Cell solution (Lonza), mixed with 2 μM control or 2 μM target-specific ASO and transferred to nucleocuvettes for nucleofection on a 4D-Nucleofector System (Lonza) using program code "DC-100". Cells were subsequently transferred onto gelatinized cell culture plates containing pre-warmed and supplemented growth medium. Growth medium was changed once after 16 hours.

RNA isolation, quality control and library preparation

For ESC and NPC transcriptome analyses, total RNA was isolated using TRIzol reagent (Ambion, Life Technologies) according to the manufacturer's instructions. Nuclear RNA fractions were obtained in parallel by resuspending cells to $2x10^7$ / ml in ice-cold nuclei buffer (10 mM Tris pH 7.6, 10 mM NaCl, 2 mM $MgCl_2$) supplemented with protease inhibitor cocktail (Sigma) and anti-RNAse (Ambion, Life Technologies). Cells were kept on ice for 10 minutes for hypotonic swelling prior to adding an equal volume of nuclei buffer containing 0.5% NP-40 for 5 minutes. Nuclei were collected, washed in nuclei buffer / 0.5% NP-40 and lysed in TRIzol. Stranded RNA-seq libraries for 76 bp paired-end sequencing on the Illumina GAIIx platform were prepared from poly(A)+ selected RNA as described previously (Eckersley-Maslin et al. 2014).

For ASO knock-down studies, total RNA was collected using TRIzol and precipitated in the presence of 7.5 µg GlycoBlue (Ambion, Life Technologies). RNA integrity was assessed on the Agilent 2100 Bioanalyzer instrument (Agilent Technologies) and only RNA samples with integrity scores > 9.5 (median 10.0) were used. Libraries were prepared from poly(A)+ selected RNA using the TruSeq RNA Sample Preparation Kit v2 (Illumina) according to the manufacturer's instructions, except that final libraries were subject to size selection followed by gel purification and validation on the Agilent 2100 Bioanalyzer. Barcoded libraries were quantified using real-time PCR-based quantification (KAPA Biosystems) and pooled at equal molarity. Sequencing was performed to the desired depth on the Illumina HiSeq2000 platform.

2'-O-Methoxyethyl (MOE) antisense oligonucleotides

Synthesis and purification of all 2'-MOE modified oligonucleotides was performed as previously described (Meng et al. 2014). These ASOs are 20-mer oligonucleotides containing a phosphorothioate backbone, 2'-*O*-methoxyethyl modifications on the first and last five nucleotides and a stretch of ten DNAs in the center. The sequences of the ASOs are: 5'-CCTTCCCTGAAGGTTCCTCC-3' (Control ASO); 5'-CCCCAGTGGTCTGGTCAGGC-3' (*Platr14* ASO #2); 5'-GGTCCTAATTCTTCATGCTG-3' (*Platr14* ASO #3); 5'-ATACTCTGAAGGGTCAGGTG-3' (*Firre* ASO #2); 5'-GTCAGAAGCATGCAATGCTG-3' (*Firre* ASO #5).

Identification of novel high-confidence transcription units

A*b initio* transcript assembly for each individual ESC or NPC sample was performed with Cufflinks2 (version 2.1.1) using default settings. Cufflinks2 transcript models were compared to those contained within the recently released GENCODE M3 annotation (April 2014, Ensembl v.76) using Cuffcompare. This data was specifically interrogated for novel intergenic transcripts, considering only loci yielding a spliced (≥ 2 exons) transcript of greater than 200 nt length.

Transcription units were required to be identified independently in AB2.2 cells and one or more of the Cast/BL6 clones to increase confidence and biological relevance.

Assessment of ASO specificity

The NCBI blastn algorithm was used to determine potential ASO targets at a given mismatch allowance. To this end, the 20-mer ASO sequence was aligned against a database of sequences comprising primary transcript sequences (exons + introns) plus all their splice forms annotated in Ensembl v. 76. Blastn was performed with highly relaxed parameters using a minimum seed (word side) of 7 and a penalty / reward score of -1/1. All hits within an E-value of 50,000 were retained.

Guttman lincRNA intersection

Chromosomal regions of lincRNAs described in Guttman et al. (Guttman et al. 2011) were converted to bed format under retention of strand information. The UCSC lift-over tool was used to convert feature coordinates from mm9 to mm10 with a minimal base ratio of 0.99. BEDtools (Quinlan and Hall 2010) was used to intersect these mm10-converted lincRNA regions with exon regions derived from the GENCODE M3 transcript annotation. A Guttman lincRNA was assigned to an intersecting GENCODE gene if it overlapped on the same strand.

ChIP-seq analysis

The following ESC and NPC ChIP-seq raw data were downloaded from the NCBI Gene Expression Omnibus: Pou5f1 and Nanog (GSM307137, GSM307140, GSM307141, GSM307154 and GSM307155 (Marson et al. 2008)); Sox2 and Pou3f2 (GSM1050286 – GSM1050290 (Lodato et al. 2013)); H3K4me3 (GSM307613, GSM307617, GSM307618 and GSM307625 (Mikkelsen et al. 2007)).

Reads were inspected using FastQC (http://www.bioinformatics.babraham.ac.uk/) and mapped to the mm10 reference genome using bowtie2 (version 2.2.3)(Langmead and Salzberg 2012), allowing 1 bp mismatch in seed alignment. Reads were trimmed within bowtie2, if necessary. Suspected PCR duplicates were removed using Picard Tools (http://picard.sourceforge.net). Creation of ChIP-seq tag directories from uniquely mapping reads and subsequent peak calling was performed using the Homer software package (Heinz et al. 2010). In brief, significant transcription factor binding peaks were called at FDR < 0.0001, requiring a minimum tag enrichment 4-fold over input and 6-fold over local background.

Real-time reverse transcription (RT)-PCR

TRIzol-extracted RNA was subject to treatment with RNAse-free DNAseI (Life Technologies) and subsequently reverse-transcribed into cDNA using TaqMan Reverse Transcription reagents and random hexamer oligonucleotides (Life Technologies). Real-time PCR reactions were prepared using Power SYBR Green Master Mix (Life Technologies) and performed on an ABI StepOnePlus Real-Time PCR system (Life Technologies) for 40 cycles of denaturation at 95°C for 15 seconds followed by annealing and extension at 60°C for 60 seconds. Primers were designed to anneal within an exon to detect both primary and processed transcripts. Primer specificity was initially tested by agarose gel electrophoresis and subsequently monitored by melting curve analysis. For all primers used, standard curves were prepared from serial dilutions of genomic DNA and primer efficiencies used to adjust delta-Ct calculations. For each sample, relative abundance was further normalized to the geometric mean of *Gapdh*, *Ppib* and *Pabpc1* housekeeping mRNA levels. Primer sequences are provided in Table S3.

RACE and lncRNA cloning

5' and 3' RACE of *Platr14* and *Firre* transcripts was performed on TRIzol-extracted RNA obtained from AB2.2 ESCs using the Ambion FirstChoice RLM-RACE kit (Life Technologies) according to the manufacturer's instructions. Fragments were amplified by nested PCR using AmpliTaq Polymerase (Life Technologies, extension time 45 seconds). Primer sequences for *Platr14* were: 5'-TTTCAGATGGTGTCCCCCTACAC-3' (5' RACE, outer) and 5'-GGGGTGTGTGTGTCTCAGAATC-3' (5' RACE, inner); 5'-GTACCAGCTAAAGAGGACACACAC-3' (3' RACE, outer) and 5'-GAGACAATGAGGTCTTGTGAATGC-3' (3' RACE, inner). Primer sequences for *Firre* were: 5'-CCCCCATTCCTCTCCAAGTCTC-3' (5' RACE, outer) and 5'-GTCTCTAAAACAACGAGGACGCAC-3' (5' RACE, inner); 5'-TTCATGAATGGCGAAATTTGGCC-3' (3' RACE, outer) and 5'-GGCATCACCTGACCCTTCAGAG-3' (3' RACE, inner). PCR products were separated on 2% agarose, bands excised, gel purified, sub-cloned into pGEM-T Easy (Promega) and 5 or more clones per fragment were sequenced using standard Sanger sequencing.

Transcript bodies were amplified from reverse-transcribed AB2.2 ESC RNA (with random hexamer primers as above) using PrimeStar polymerase (Clontech). Primer pairs for *Platr14* were: 5'-CCGAAGATTAACGGCAGGAG-3' and 5'-TGGTGGGTGGTGGACTAGAA-3'. Primer pairs for *Firre* were: 5'-AGGTGCGTCCTCGTTGTTTT-3' and 5'-AGCGCTGTTCAACCTCCAAT-3' (section 1); 5'-GATCCTAAGAGCAGAAGACAATCA-3' and 5'-TGCAGTGTCCACTAACTGTGTGA-3' (section 2); 5'-GGCCTCAAAGCTTCCTTTCC-3' and 5'-ATTACCCAAAGCAAGCACCA-3' (section 3); 5'-GACTGGCCCCTGGGATTTTC-3' and 5'-TCACAGCTCCGCTGTTGTCC-3' (section 4). All products were sub-cloned into pGEM-T Easy and insert sequences verified.

Full-length cDNAs were assembled using suitable restriction sites (*Platr14*) or by Gibson assembly (*Firre*)(Gibson et al. 2009).

Supplemental Figure Legends

Figure S1: Gene expression variability in ESCs and NPCs. A) Pearson correlation matrix of genes (mRNA + lncRNA) expressed across the 7 Cast/BL6 ESC and NPC clones (derivation 1 and 2). **B)** Boxplot of the pair-wise Pearson (left) and Spearman rank (right) correlations of genes expressed across the 7 Cast/BL6 ESC clones and the 5 Cast/BL6 NPC clones of the same derivation.

Figure S2: Identification of novel "intergenic" transcription units in ESCs and NPCs. A) Overview of intergenic transcription units not annotated in GENCODE M3 as identified by *ab initio* transcript assembly using Cufflinks in the indicated cell types and genetic background. For Cast/BL6, the union across the 7 clonal populations is summed. Only regions yielding a spliced transcript of greater than 200 nt length are considered. **B)** Example of a novel transcription unit identified in Cast/BL6 and AB2.2 ESCs. Read coverage from poly(A)+ RNA sequencing in ESCs is plotted. Oct4 ChIP-seq data re-analyzed from Marson et al., 2008 is shown in blue and indicates 3 discrete binding peaks in the vicinity of this unit. UCSC gene tracks depict the (silent) neighboring genes. **C)** Boxplot of FPKM values for all reconstructed transcript isoforms derived from novel loci identified in both genetic backgrounds of ESCs (left) and NPCs (right). **D)** Predicted biotypes of reconstructed transcripts based on the coding potential calculator support vector machine (Kong et al. 2007).

Figure S3: Expression profiles of ESC and NPC lncRNAs. A, B) Cumulative frequency plot of FPKM values for lncRNA (orange) and mRNA (grey) genes in ESCs (A) and NPCs (B) **C, D)** ChIP-seq profile of H3K4me3 centered on the annotated transcription start site (TSS) of lncRNA genes in ESCs (C) and NPCs (D). The union of lncRNAs detected in our ESC and NPC data sets is represented. For comparison, profiles for the same number of protein-coding mRNAs are provided. ChIP-seq raw data was obtained from Mikkelsen et al., 2007. **E)** Fractions of lncRNAs expressed in the given cell

type and TSS of which falls within 25 bp of a long terminal repeat (LTR) or LINE. (* $p < 0.002$; one-sided χ^2). **F)** Fraction of lncRNAs and mRNAs that are detected in the given cell type but not the other (FPKM ≥ 0.1; * $p < 10^{-15}$; one-sided χ^2).

Figure S4: Validation of gene expression changes between ESCs and NPCs. A, B) Scatterplots depicting the linear relationship between qRT-PCR-calculated log2 fold changes and FPKM-based log2 fold changes in AB2.2 (A) and Cast/BL6 (B) cells for 38 genes. Shaded area: 95% confidence for the linear model. Pearson correlation (*r*) is indicated. **C)** Linear relationship between DESeq2's counts-based log2 fold change and qRT-PCR as in (B).

Figure S5: Identification of nuclear-enriched lncRNAs (cont.). A) Nuclear to total log2 ratios by gene biotype as in Figure 2A for AB2.2 NPCs. **B)** Correlation of lncRNA FPKM values with the average number of detected RNA-FISH hybridization foci in AB2.2 ESCs and NPCs (representing 13 data points). Shaded area indicates 95 % confidence. Hybridization foci were counted in ImageJ (Abramoff et al. 2004) on maximum intensity projections above a manually adjusted background threshold. **C)** Single-molecule RNA FISH images for selected lncRNAs in AB2.2 ESCs and NPCs as in Figure 2B. Scale bars: 10 μm. Mitotic images are enlarged 2-fold.

Figure S6: Principal component analysis of knock-down libraries. A) *Platr14* and **B)** *Firre* knock-downs. Principle component analysis was performed for read counts after regularized log transformation and normalization for library sizes.

Figure S7: *Platr14* gene locus. A) Genome browser overview of the *Platr14* locus. Pou5f1 (blue) and Nanog (orange) ChIP-seq data from ESCs is plotted (data re-analyzed from Marson et al., 2008). Poly(A)+ RNA sequencing read coverage in ESCs is plotted in black. Unique sequences of 5' and 3'

RACE fragments generated from ESC RNA are indicated above the annotated transcript isoforms (UCSC genes). The main transcript structure as determined by cloning from RT-PCR and RACE is depicted. Target sites for ASOs are indicated. **B)** Zoom-in on the 5' end of *Platr14* indicating the position of cloned and annotated transcription start sites relative to the RLTR16B LTR. **C)** Zoom-in on the 3' end of *Platr14*.

Figure S8: Summary assessment of ASO specificity. A) Of all annotated genes expressed in ESCs (excluding the cognate ASO target) within the indicated threshold of basepair mismatches, blue bars correspond to those genes tested for differential expression. Of these, red bars correspond to genes that are significantly down-regulated following for *Platr14* ASO#2 (left) and ASO#3 (right). **B)** As in (A), for *Firre* ASO#2 (left) and ASO#5 (right). See Supplemental Methods for details.

Figure S9: *Firre* RNA FISH. A) Single-molecule RNA-FISH for *Firre* transcripts in female PGK12.1 ESCs. Maximum intensity projection of an ESC colony is shown. Dotted line marks nuclear outlines. **B)** RNA-FISH in AB2.2 ESCs using nick-translated *Firre* cDNA probes (top panel) or unspecific vector sequence (bottom). Arrows point to brighter hybridization foci likely corresponding to the site of *Firre* transcription on the single X-chromosome in male AB2.2 ESCs. Scale bars: 10 μm.

Figure S10: Gene expression changes after *Firre* knock-down. A) Heatmaps (gene-wise z-score) of genes significantly affected 24 hours after transfection of ASOs targeting *Firre* in AB2.2 ESCs in 4 biological replicates (FDR < 0.05). **B)** Expression of 5 genes previously reported to interact with the *Firre* locus in ESC (Hacisuleyman et al., 2014). Data represents mean and standard deviation of 4 biological replicates each of ESCs transfected with control ASO or *Firre*-specific ASOs. Only *Eef1a1* shows a response to *Firre* depletion, although only ASO#2 causes a change called as significant at the FDR threshold used.

Supplemental References

Abramoff MD, Magalhaes PJ, Ram SJ. 2004. Image Processing with ImageJ. *Biophotonics International* **11**: 36–42.

Eckersley-Maslin MA, Thybert D, Bergmann JH, Marioni JC, Flicek P, Spector DL. 2014. Random monoallelic gene expression increases upon embryonic stem cell differentiation. *Developmental Cell* **28**: 351–365.

Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Meth* **6**: 343–345.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell* **38**: 576–589.

Kong L, Zhang Y, Ye Z-Q, Liu X-Q, Zhao S-Q, Wei L, Gao G. 2007. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res* **35**: W345–9.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Meth* **9**: 357–359.

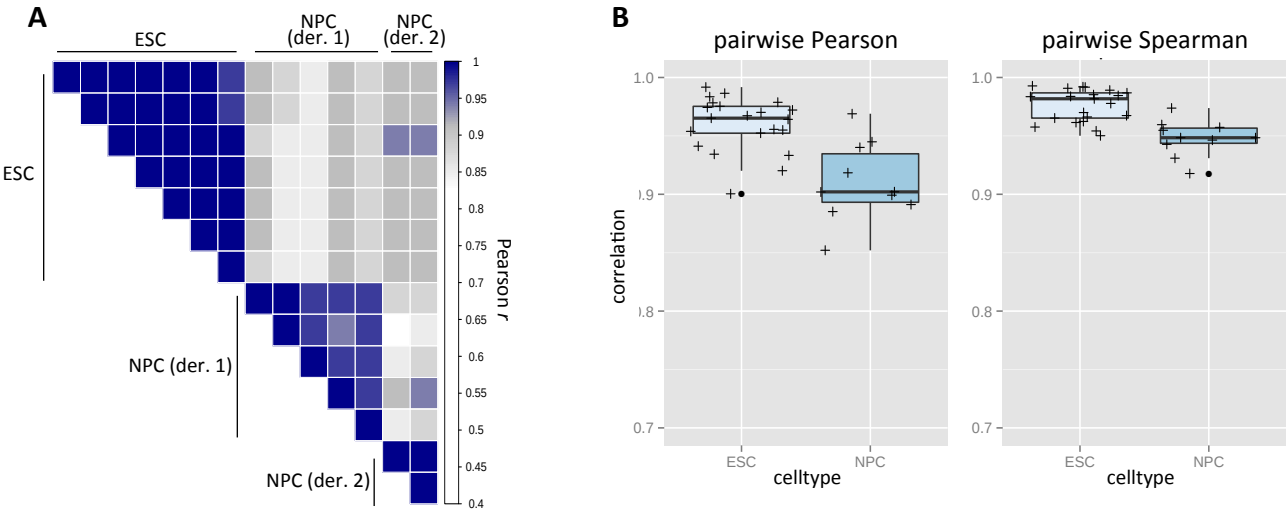
Lodato MA, Ng CW, Wamstad JA, Cheng AW, Thai KK, Fraenkel E, Jaenisch R, Boyer LA. 2013. SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state. *PLoS Genet* **9**: e1003288.

Marson A, Levine SS, Cole MF, Frampton GM, Brambrink T, Johnstone S, Guenther MG, Johnston WK, Wernig M, Newman J, et al. 2008. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**: 521–533.

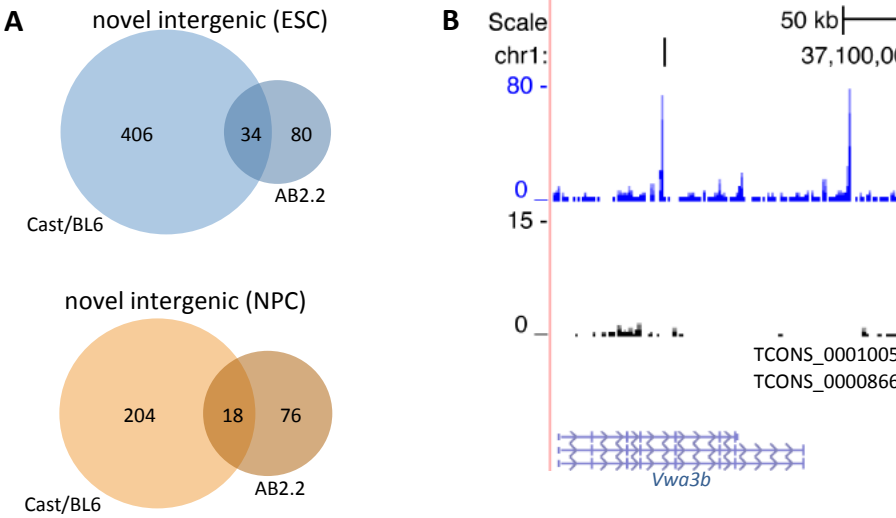
Meng L, Ward AJ, Chun S, Bennett CF, Beaudet AL, Rigo F. 2014. Towards a therapy for Angelman syndrome by targeting a long non-coding RNA. *Nature* 1–16.

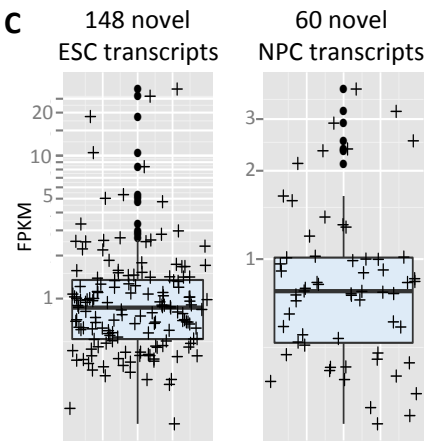
Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim T-K, Koche RP, et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**: 553–560.

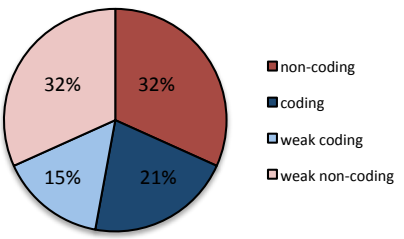
Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.



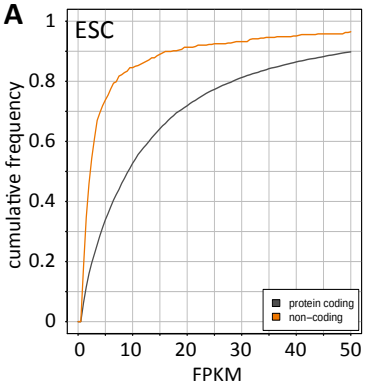
Bergmann et al., Figure S1

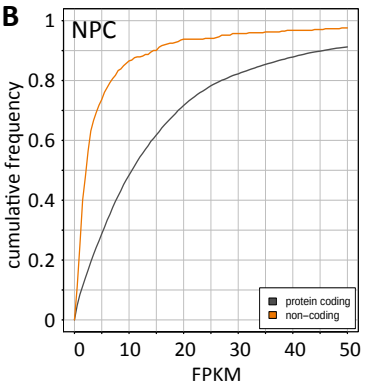


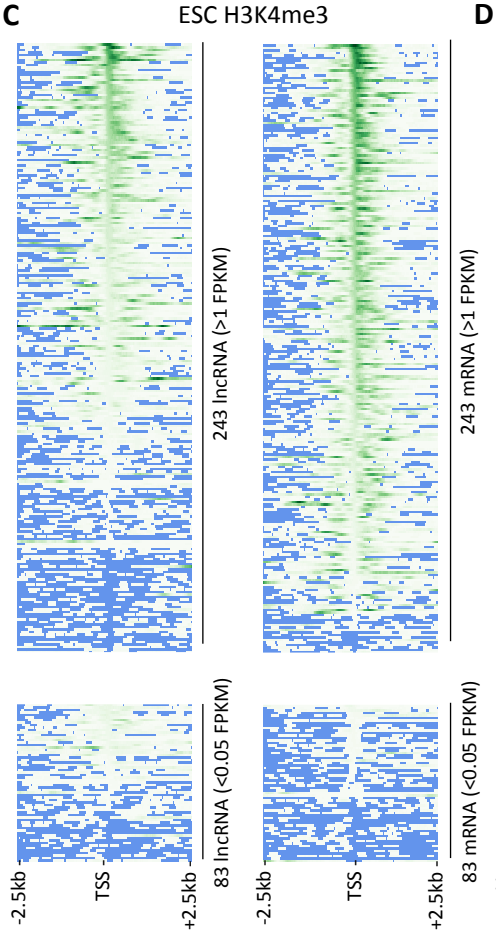


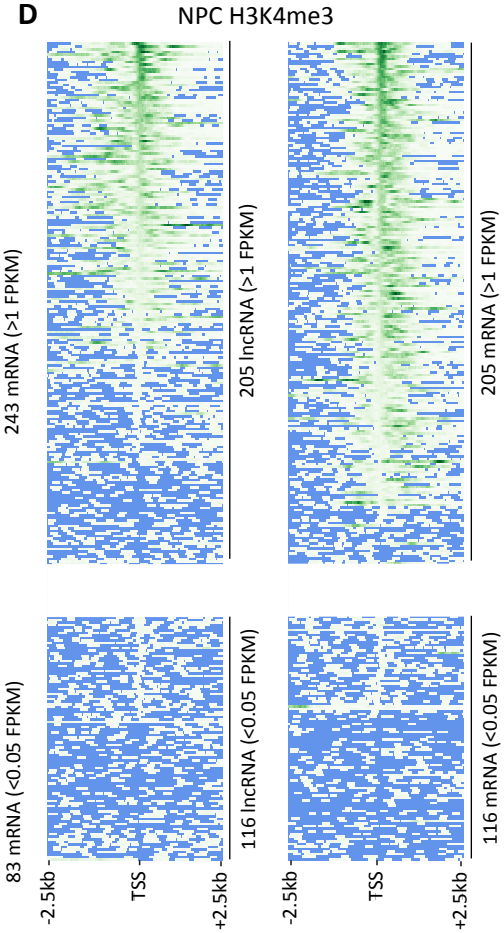


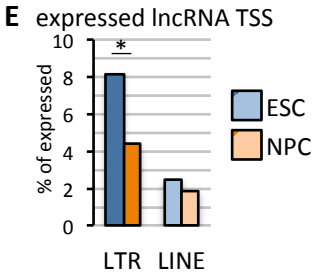
Bergmann et al., Figure S2

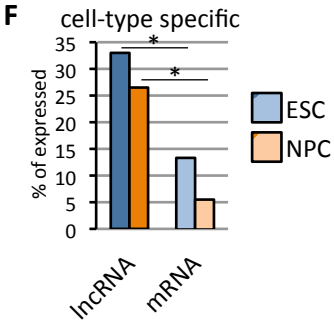

A
ESC
cumulative frequency
protein coding
non−coding
FPKM

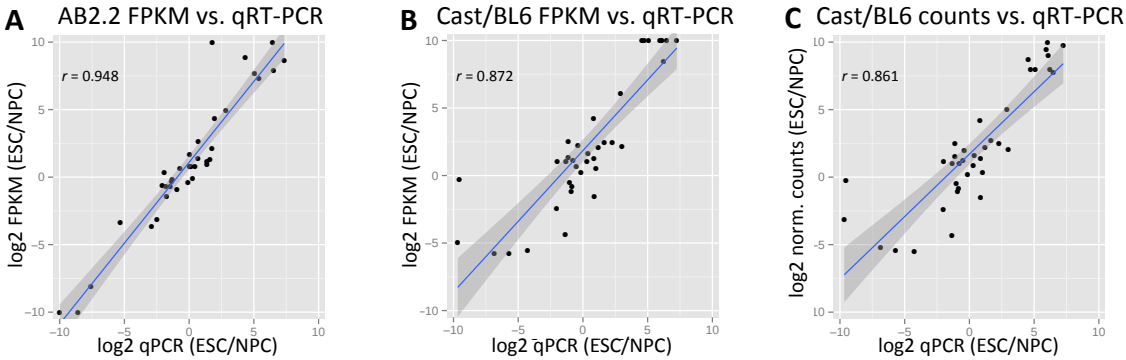

B
NPC
cumulative frequency
protein coding
non−coding
FPKM


C
ESC H3K4me3
243 lncRNA (>1 FPKM)
83 lncRNA (<0.05 FPKM)
243 mRNA (>1 FPKM)
83 mRNA (<0.05 FPKM)
-2.5kb
TSS
+2.5kb

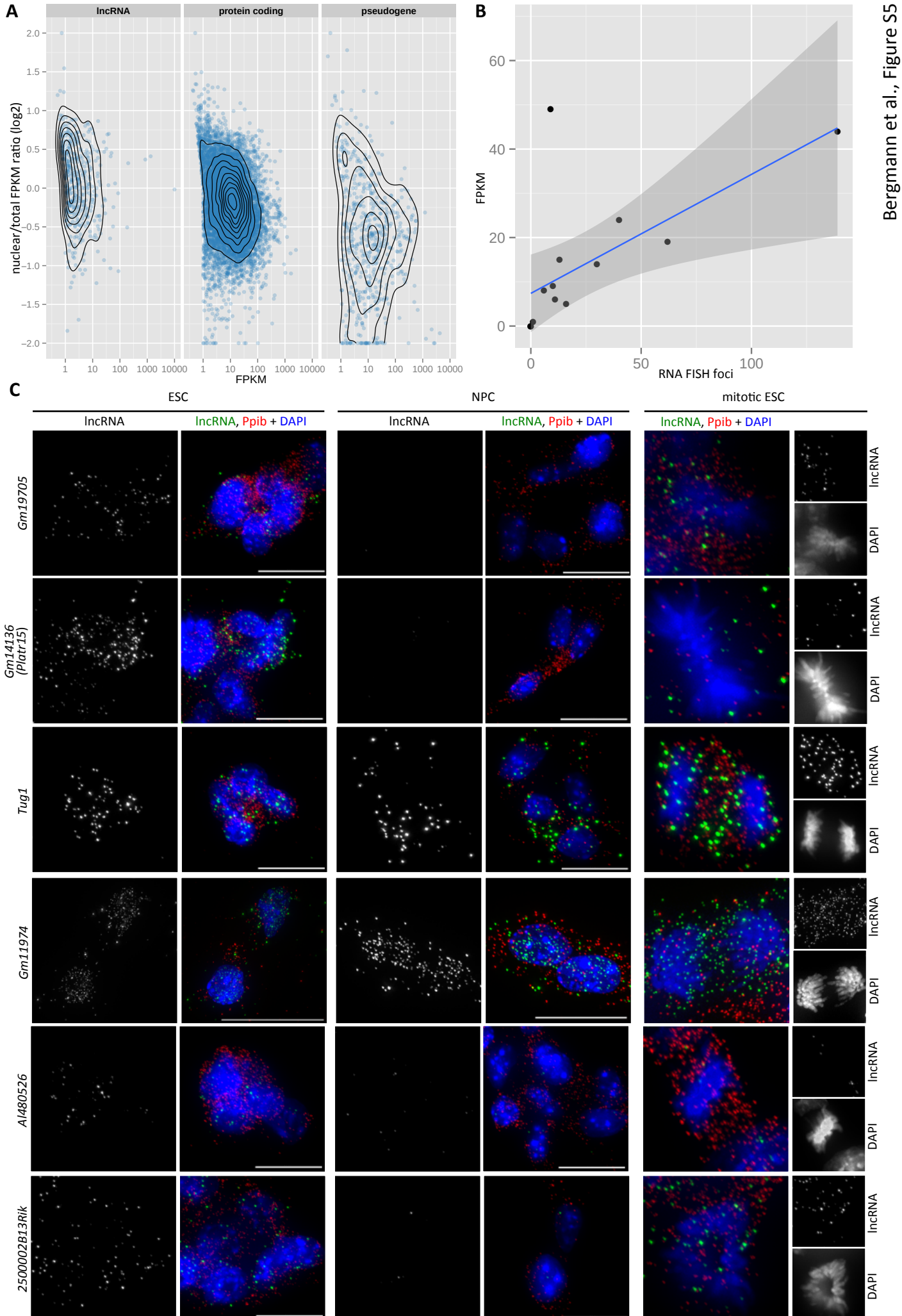

D
NPC H3K4me3
205 lncRNA (>1 FPKM)
116 lncRNA (<0.05 FPKM)
205 mRNA (>1 FPKM)
116 mRNA (<0.05 FPKM)
-2.5kb
TSS
+2.5kb


E
expressed lncRNA TSS
% of expressed
ESC
NPC
LTR
LINE

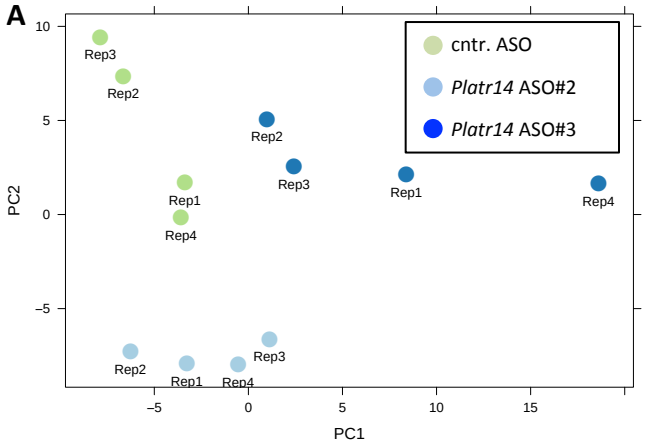

F
cell-type specific
% of expressed
ESC
NPC
lncRNA
mRNA

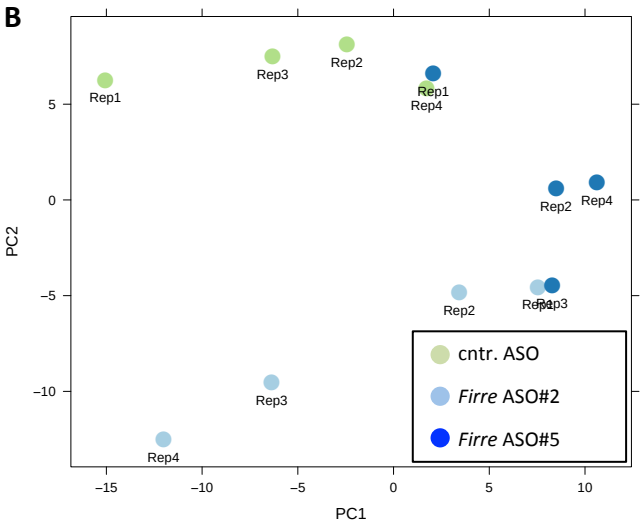


Bergmann et al., Figure S4

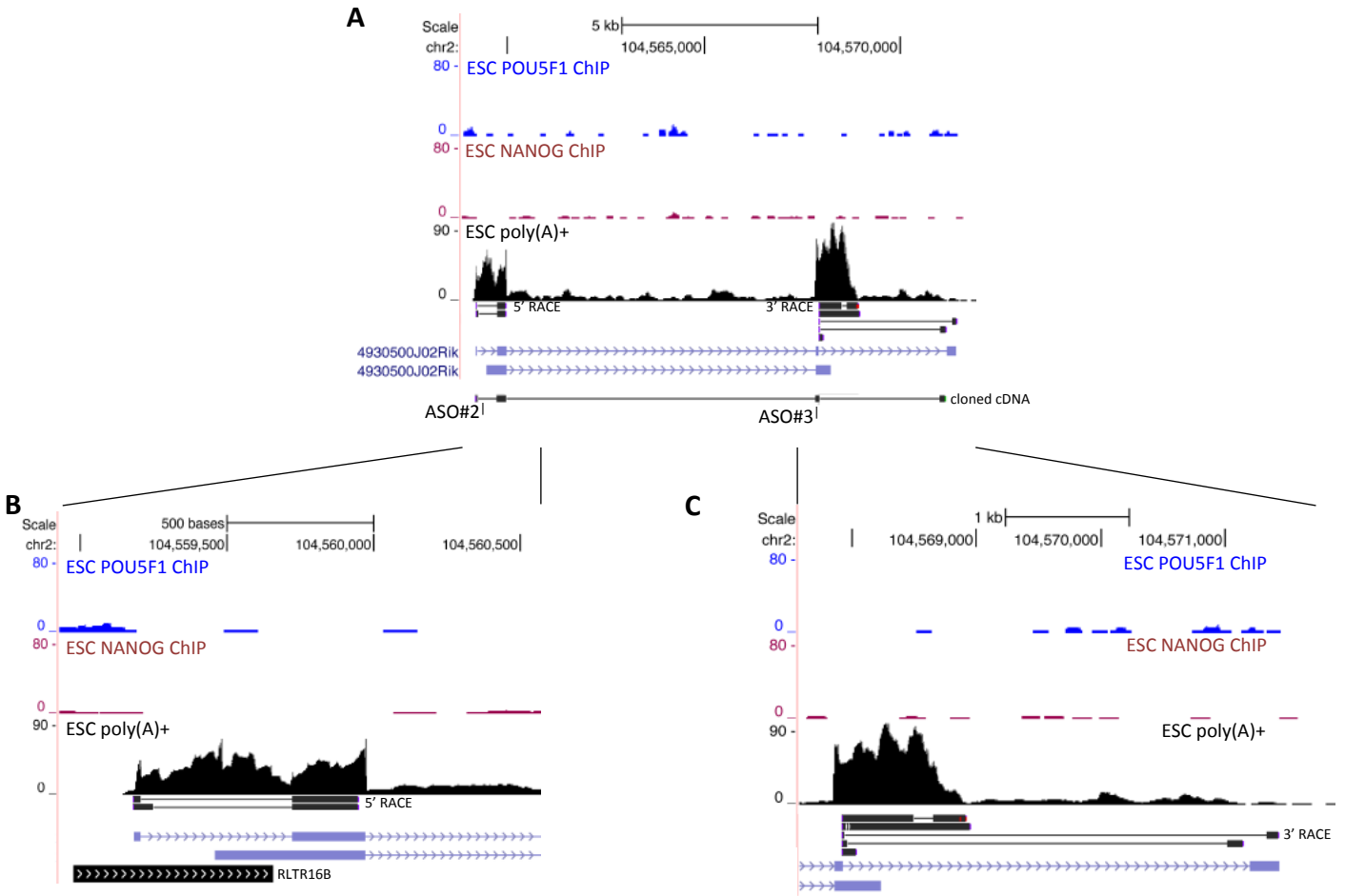


Bergmann et al., Figure S5

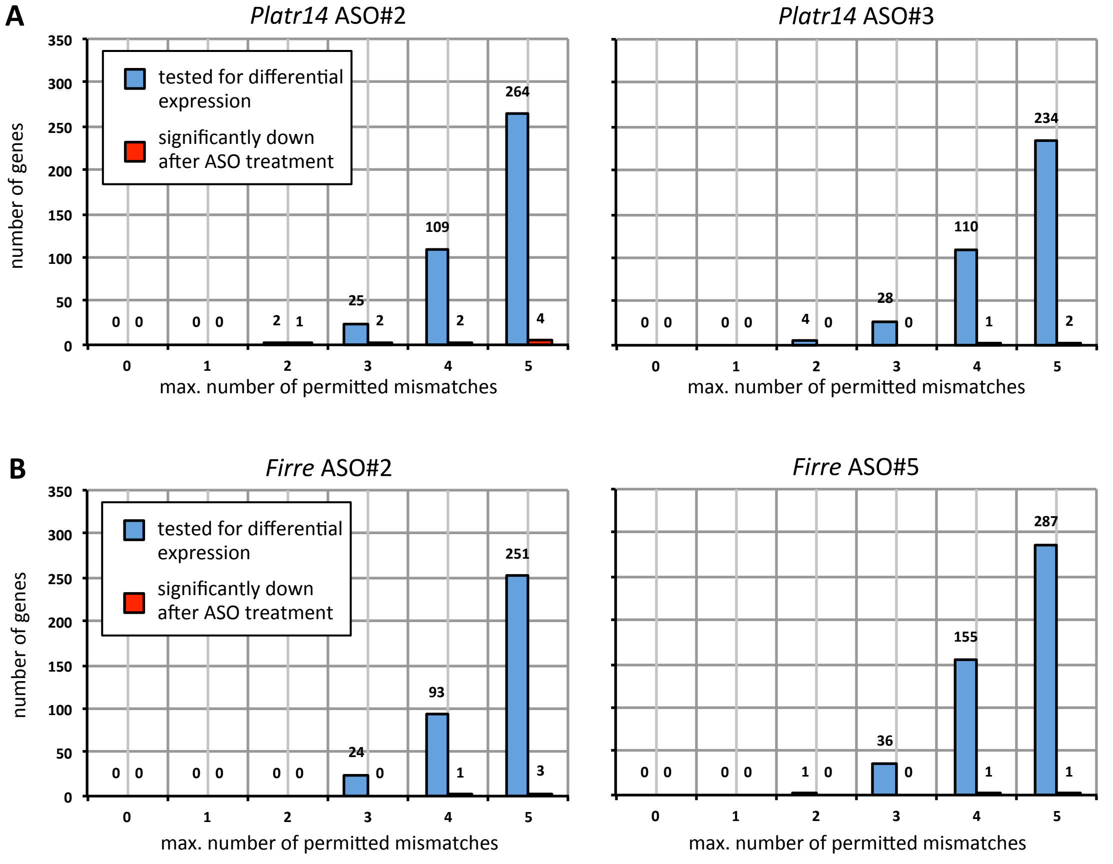




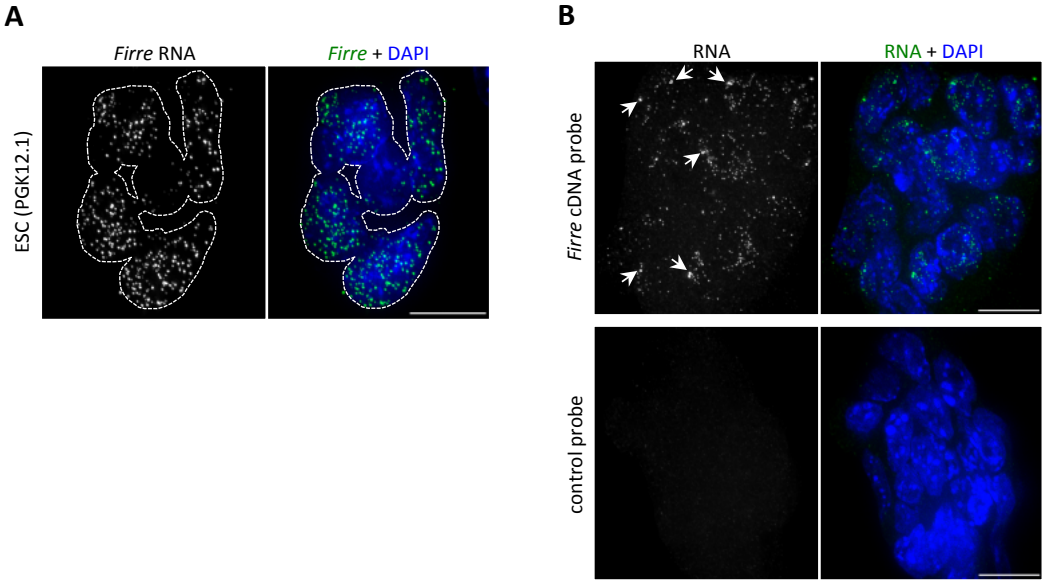
Bergmann et al., Figure S6

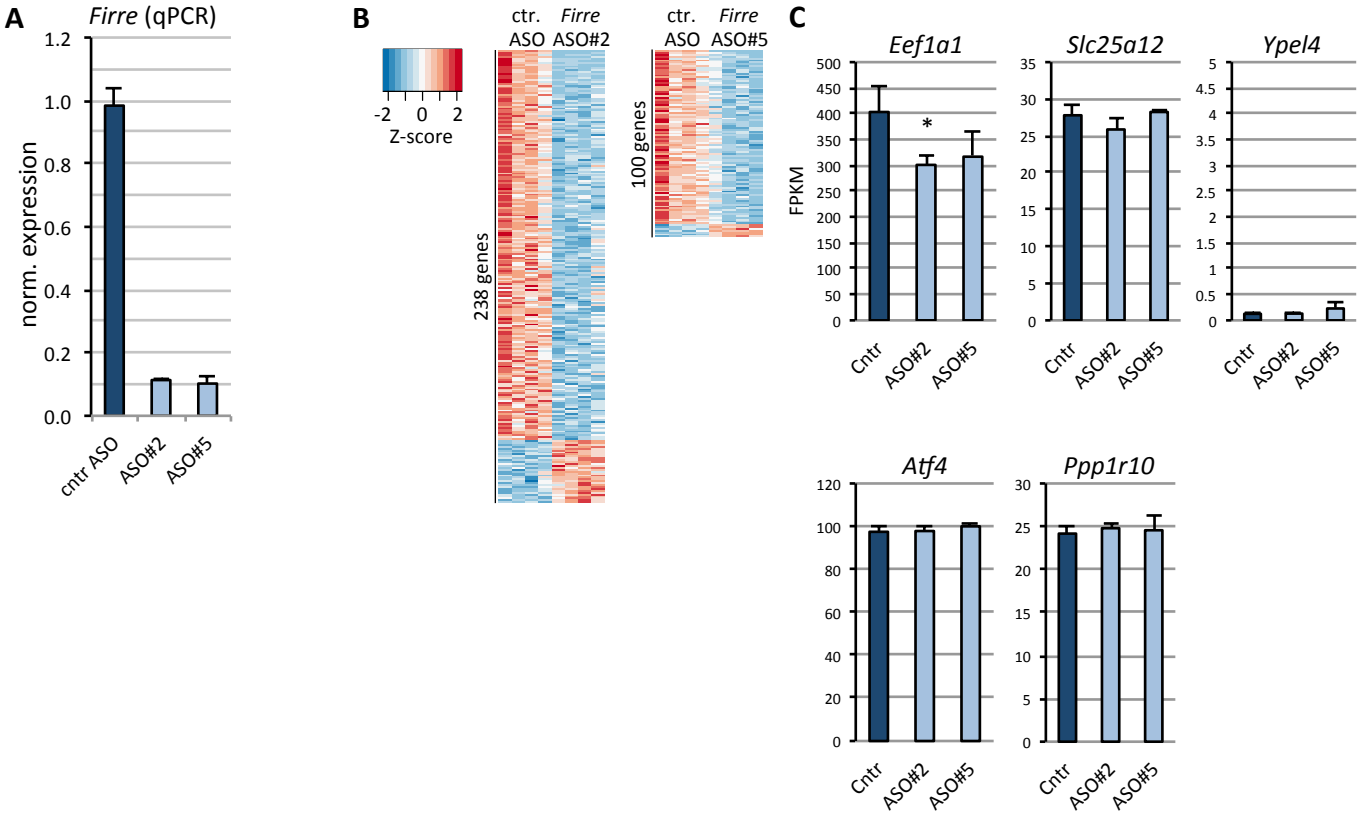


Bergmann et al., Figure S7



Bergmann et al., Figure S8


A
Firre RNA
Firre + DAPI
ESC (PGK12.1)
B
RNA
RNA + DAPI
Firre cDNA probe
control probe



Bergmann et al., Figure S10