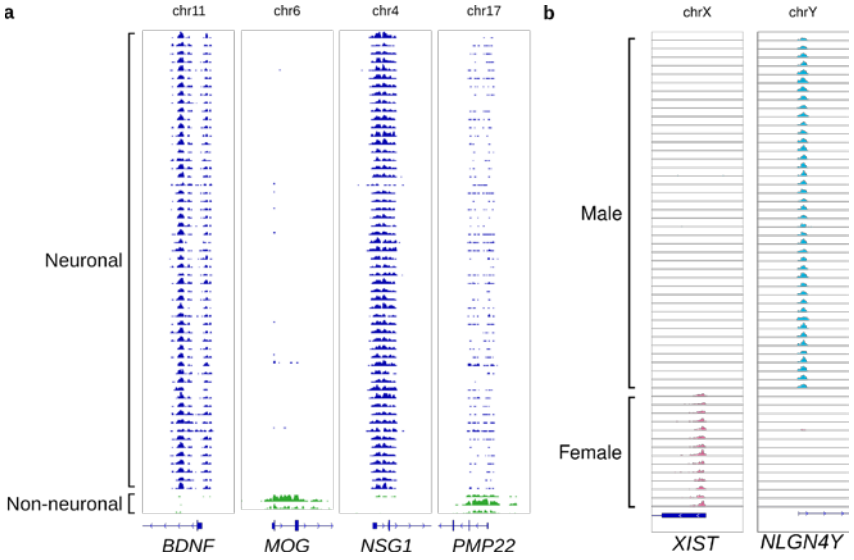
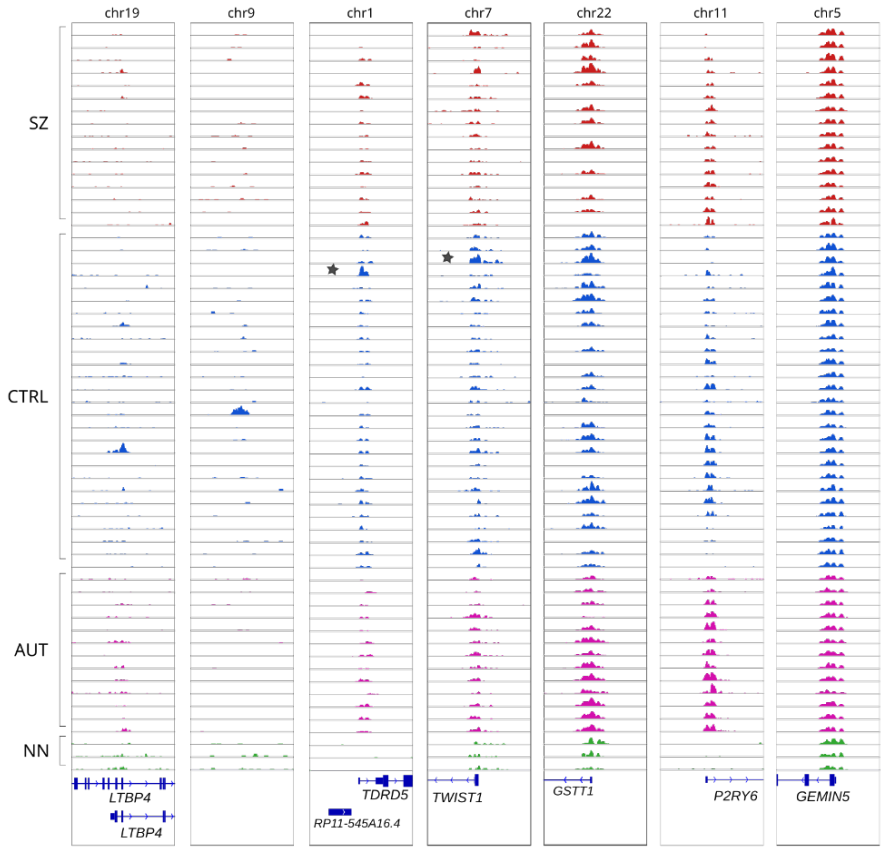
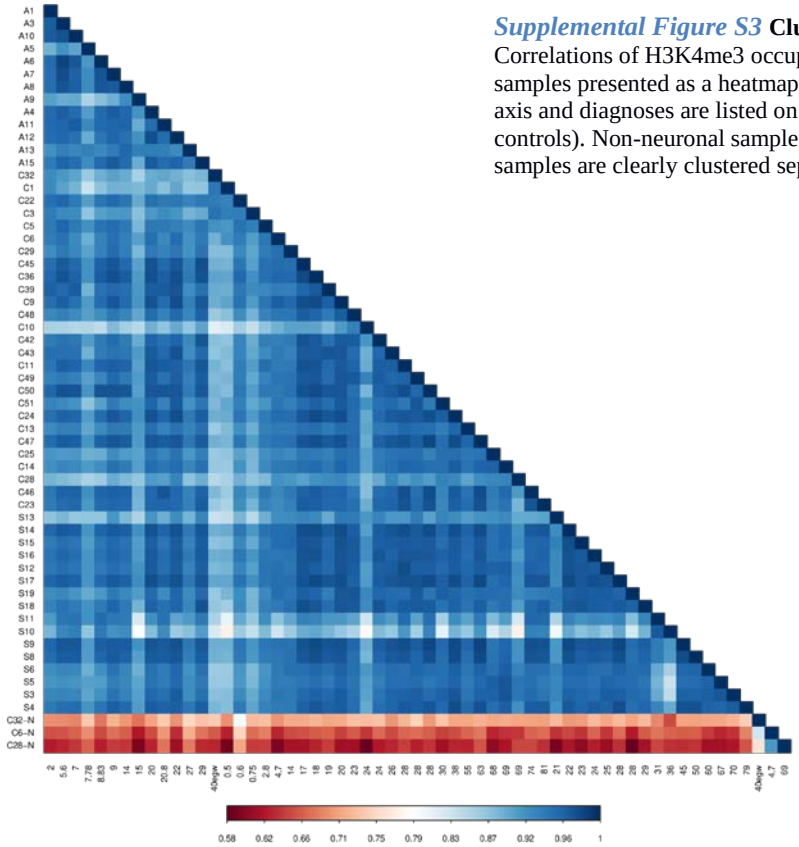
Supplemental Figure S1 **Controls for neuronal specificity of NeuN+ samples and matches for gender-specific genes in ChIP-seq data.** (**a**) Loci showing specific H3K4me3 signal in neuronal or non-neuronal samples: neuron specific *BDNF* and *NSG1* genes, glial *MOG* and *PMP22* genes. (**b**) As expected, analysis of neuronal cells shows the female-specific H3K4me3 peaks for non-coding RNA gene (*XIST*) on chromosome X and male-specific gene *NLGN4Y* on chromosome Y.

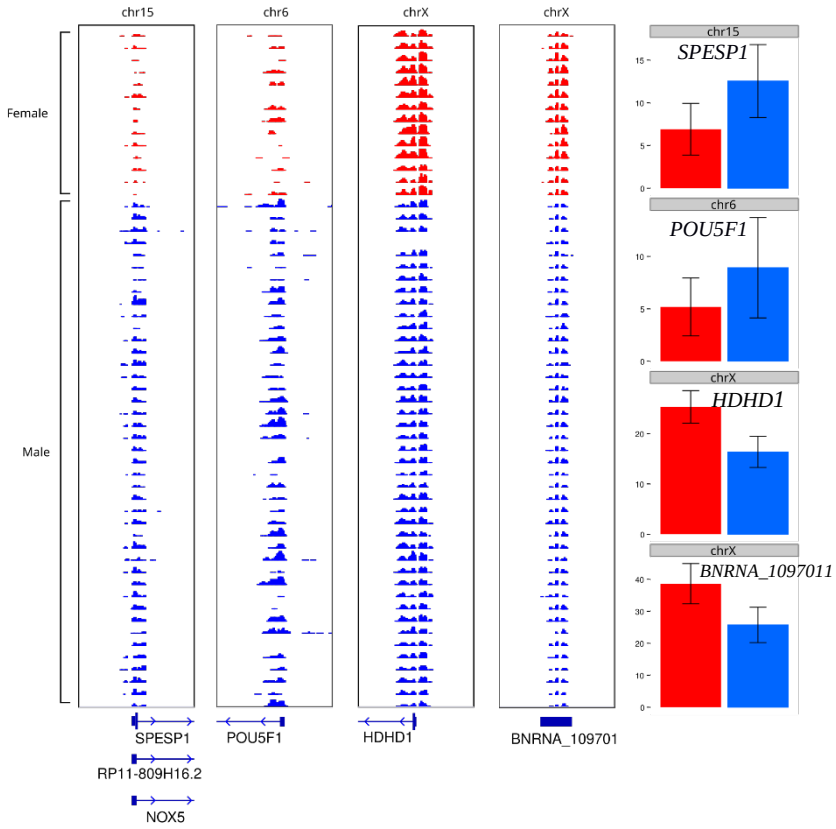


Supplemental Figure S2 **Examples of polymorphic H3K4me3 marked loci**. "Induced" peaks in single individuals (panels chr19, chr9); rare up-regulated peaks in single individuals (designated by star) in promoters of *TDRD5* and *TWIST1* genes (panels chr1, chr7); highly polymorphic peaks in promoter of *GSTT1* (known CNV for this region with global population frequency ~0.5) and *P2RY6* (no reported CNVs for this region) (panels chr22,chr11); locus with non-polymorphic H3K4me3 abundance across all samples (panel chr5).



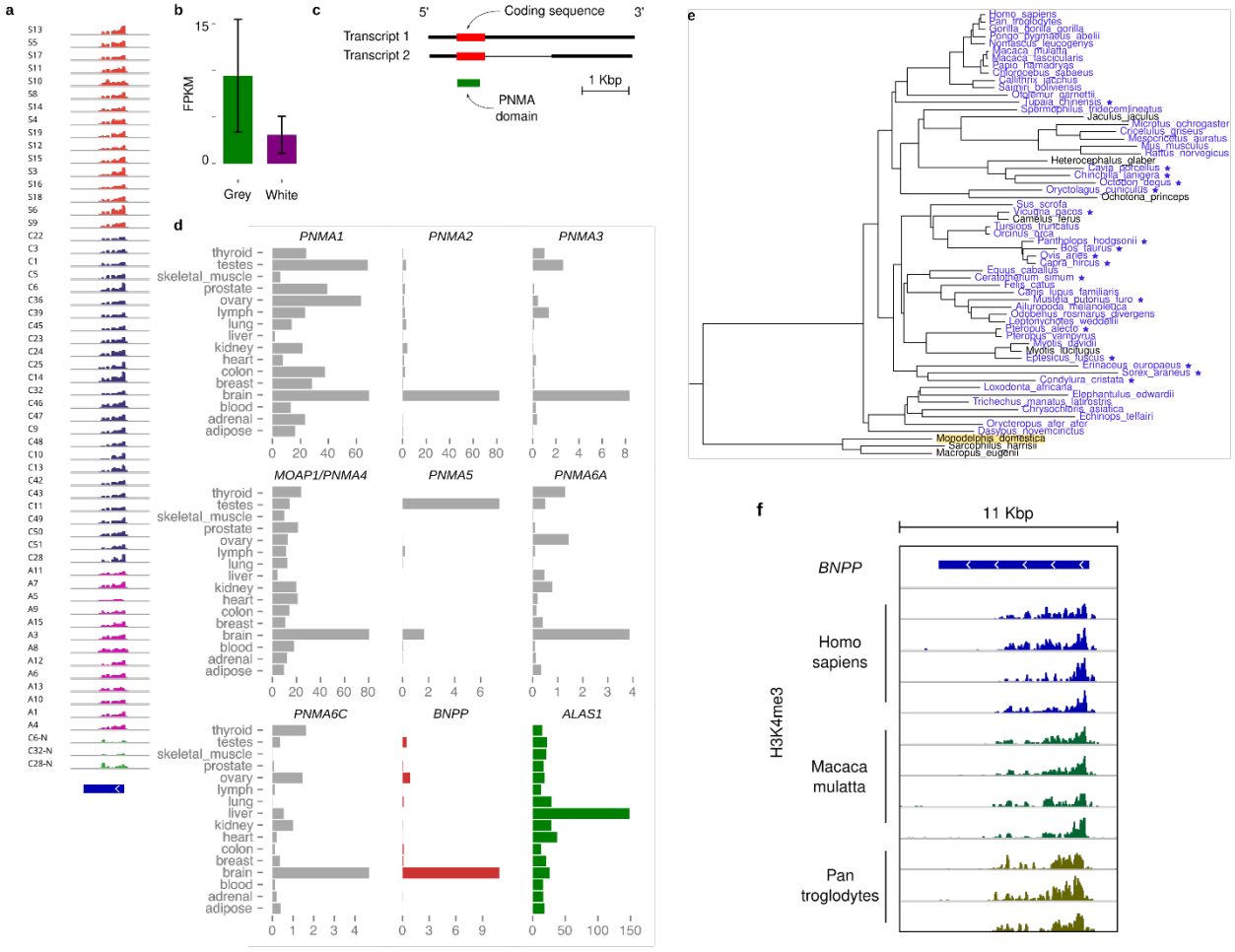


Supplemental Figure S3 **Cluster analysis of H3K4me3 profiles.** Correlations of H3K4me3 occupancy across genome between all pairs of samples presented as a heatmap. In the cluster analysis, ages are listed on x-axis and diagnoses are listed on y-axis (A-autism, S-schizophrenia, C-controls). Non-neuronal samples (C32-N, C6-N, C28-N) and all neuronal samples are clearly clustered separately.

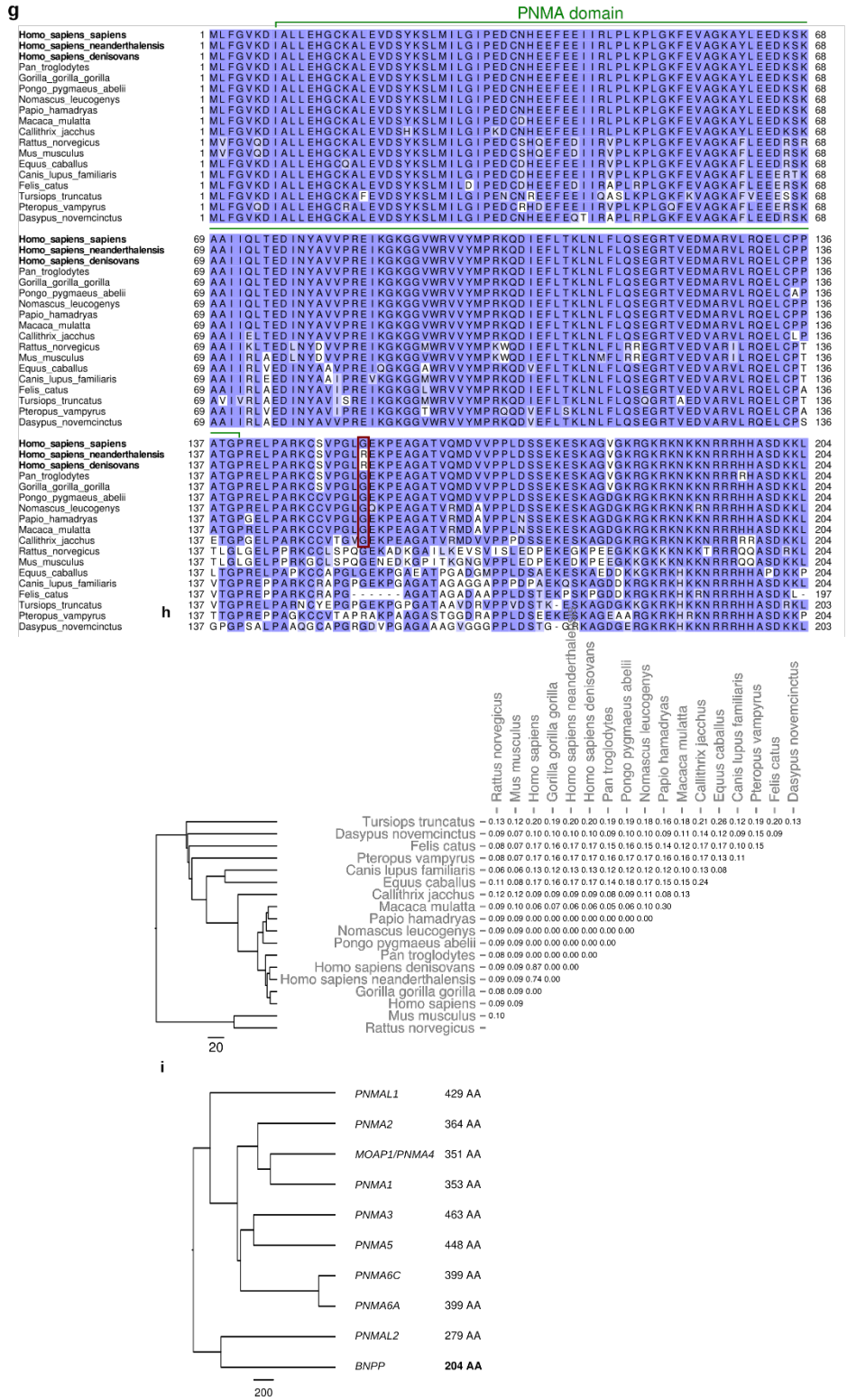


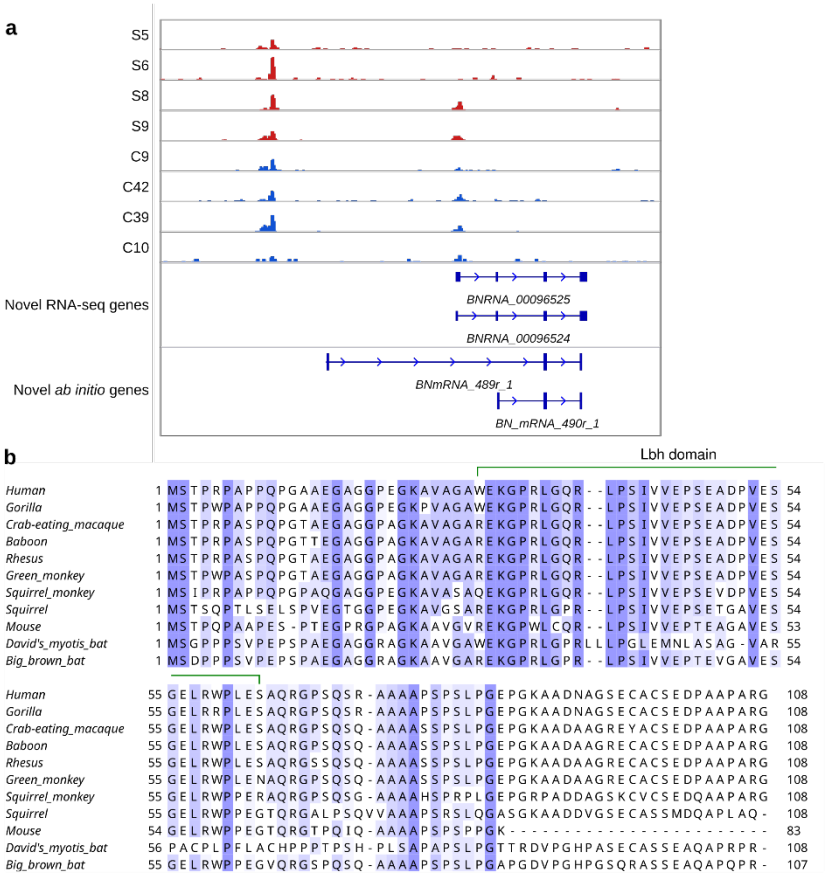
Supplemental Figure S4 **Sex-specific quantitative H3K4me3 peaks.** Examples of sex-specific alterations: promoter of *NOX5* shares H3K4me3 peak with a protein-coding gene *SPESP1* (Sperm Equatorial Segment Protein 1) and a non-coding *RP11-809H16.2*; *POUP5F1* translates to protein OCT4, an immunohistochemical marker for central nervous system germinomas, which dominantly develops in male subjects; gene *HDHD1* is known to partially escape X-inactivation; H3K4me3 abundance of a novel non-coding gene *BNRNA_109701* on chromosome X is different for male and female subjects. The histograms show the standard deviations in groups of male and female individuals for these loci.

Supplemental Figure S5 (a-f). **Example of the newly discovered protein-coded *BNPP* gene.** (**a**) H3K4me3 signal in human PFC neurons in a locus harboring *BNPP* gene. (**b**) RNA expression levels of *BNPP* in grey & white matter shown for 20 samples of each tissue. (**c**) Schematic representation of transcripts and localization of PNMA domain homology. (**d**) RNA expression in 16 tissues for genes of PNMA family and a new *BNPP* gene; *ALAS1* is a housekeeping gene. (**e**) Phylogenetic tree of species in which *BNPP* gene was found in this study. Among metatherians, we observed neither amino acid nor nucleotide homolog to human *BNPP* gene. Blue denotes species with *BNPP* gene; in species marked with star, only nucleotide (but not ORF) sequence was found possibly due to missassemblies in the reference genomes of these species. *Monodelphis domestica* is highlighted with yellow. (**f**) H3K4me3 signal for *Homo sapiens*, *Macaca mulatta* and *Pan troglodytes* in *BNPP* locus.

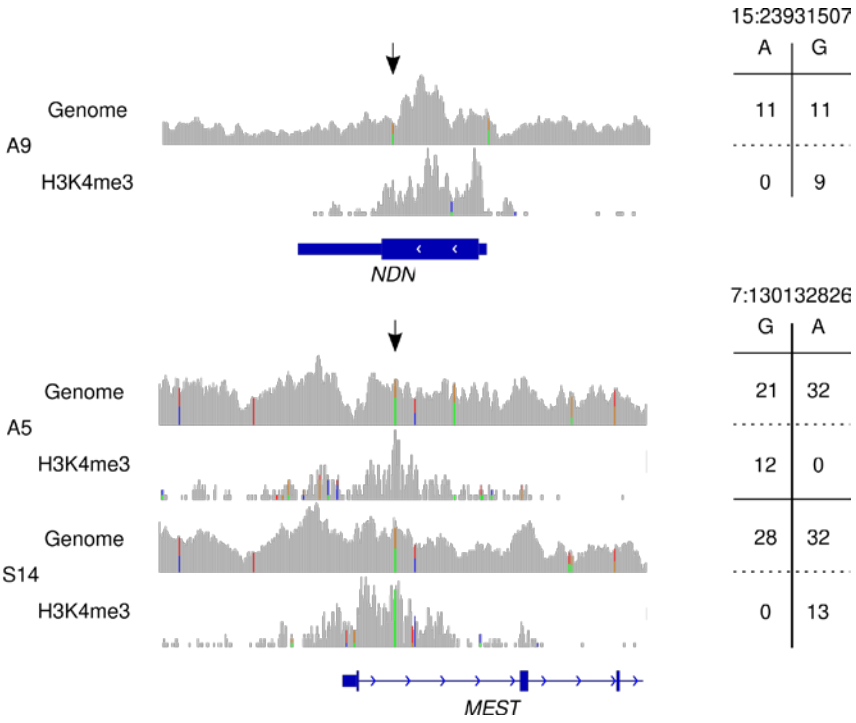


Supplemental Figure S5 (continuation, g-i) **Example of the newly discovered protein-coded *BNPP* gene.** (**g**) Multiple sequence alignment of *BNPP* protein sequence and members of PNMA family. Highlighted is position 154 bearing mutation specific for Neanderthal and Denisovan. (h) Table of synonymous to non-synonymous substitution ratios (Kn/Ks). (**i**) A family tree for PNMA-like genes; *PNMA6B* and *PNMA6D* are omitted, since they have 99% identity to *PNMA6A* and *PNMA6C*, respectively.





Supplemental Figure S6 **Example of the newly discovered protein coding *BN-LBH* gene.** (**a**) The gene for the *BN-LBH* (brain neuronal) protein was predicted by two independent methods: identification of brain RNA transcripts mapped at their 5'-UTR to H3K4me3 peak regions at chromosome 14:103550302-103551065; computational prediction of spliced transcripts for this H3K4me3-marked region *ab initio*. The H3K4me3 peak at 14:103550302-103551065 is relatively weak and occur in a few but not all individuals regardless the mental pathology. Another robust H3K4me3-marked locus with no predicted genes is located nearby (14:103541407-103542769). The transcripts correspond to four exons (5'-UTR non-coding exon and three coding exons) and encode the same amino acid sequence. (**b**) Predicted 106 amino-acid protein (BN-LBH) shows homology to protein domain for family of transcriptional regulators named Lbh (short for Limb, bud and heart) (PFAM search, 65-67% identity, 1.00E-23 – 5.00E-23). The Lbh-proteins are typically between 92 and 116 amino-acids in length and act as regulators of heart in embryonic development and have been suggested as transcriptional activators in MAPK signaling. The recovery of predicted amino-acid sequences from genomes of other mammals showed the high identity to open reading frame (ORF) sequences in other species as well (e.g., 99% to gorilla).



Supplemental Figure S7 **Epigenetic allelic imprinting.** The samples from subjects A5, A9 and S14 were used for both whole genome sequencing and ChIP-seq analysis. Genome sequencing analysis revealed heterozygous G/A SNP (11 reads with G and 11 reads with A) in promoter of imprinted NDN gene in A9 subject. ChIP-seq analysis for heterozygous SNP G/A in promoter of *NDN* gene reveals only single G-allele in A9 (0 reads with A and 9 reads with G). The subjects A5 and S14 have heterozygous SNP G/A at the same position of *MEST* gene, but H3K4me3 tags detect only single allele different (allele bearing G or A) for each individual.

Supplemental Figure S8 **Analysis of H3K4me3 signal and expression of *NUP210L* gene.** (**a**) H3K4me3 peak in *NUP210L* promoter was revealed in 5 individuals only. Individuals with SZ (S) are designated by red, control CTRL (C) – by blue tracks, autistic individuals (A)) – by purple. (**b**) Position for SNP rs114697636 in 4 individuals covered by reads from whole genomic sequencing data showing rare G variant in heterozygous G/A (S14) or homozygous G/G (A5) state and A/A (reference genome) genotype state common in human population (C13, A9). (**c**) H3K4me3 abundance in *NUP210L* locus in both non-neuronal (N-) and neuronal cells (N+) in the same individual. (**d**) *NUP210L* RT-PCR analysis of RNA transcripts from brain specimens on gel electrophoresis detected the products in individuals who carry G allele of rs114697636 SNP. UBC housekeeping gene was used as control. (**e**) Histograms of quantitative measurements of *NUP210L* transcripts by qPCR confirms the significant expression of *NUP210L* in brain of G-allele carrier (S14). (**f**) Evolutionary conserved site changed by G variant of rs114697636. **(g)** Gel-shift assay with P^{32}-labelled 21 bp oligonucleotide duplex probe harboring either C or G variants of rs114697636. The probes were bound to protein from cellular nuclear extracts, in particular, from HEPG2 cell line (with slightly more efficient binding for the G-variant) (left panel). The gel-shift does not depend on salt concentrations (middle panel). The cold competitor probes harboring either C (V1) or G (V2) abolish the binding and show that the upper band as a specific- and lower band as non-specific product. The cold probe with TATA sequence does not compete with the binding (bottom panel). **(h)** Population frequencies of G allele for rs114697636 (1000 Genomes phase 3); ALL = global, AFR = African, AMR = American, EAS = East Asian, EUR = European, SAS = South Asian. (**i**) Emerging (green) and disappearing (red) transcription factor binding sites (TFBS) as predicted by PERFECTOS-APE for two TFBS databases: HOCOMOCO and JASPAR.