Neurogenesis 2:1, e995524; January–December 2015; Published with license by Taylor & Francis Group, LLC

Identification and expression patterns of novel long non-coding RNAs in neural progenitors of the developing mammalian cortex

Julieta Aprea[1,†], Mathias Lesche[2,†], Simone Massalini[1], Silvia Prenninger[1], Dimitra Alexopoulou[2], Andreas Dahl[2], Michael Hiller[3,4,*], and Federico Calegari[1,*]

[1]DFG–Research Center and Cluster of Excellence for Regenerative Therapies; Dresden, Germany; [2]Deep Sequencing Group, Biotechnology Center; Dresden, Germany; [3]Max Planck Institute of Molecular Cell Biology and Genetics; Dresden, Germany; [4]Max Planck Institute for the Physics of Complex Systems; Dresden, Germany

[†]Authors are equal contributing joint-first authors.

Keywords: neurogenesis, cortical development, lncRNAs

Abbreviations: lncRNAs, long non-coding RNAs.

Long non-coding (lnc)RNAs play key roles in many biological processes. Elucidating the function of lncRNAs in cell type specification during organ development requires knowledge about their expression in individual progenitor types rather than in whole tissues. To achieve this during cortical development, we used a dual-reporter mouse line to isolate coexisting proliferating neural stem cells, differentiating neurogenic progenitors and newborn neurons and assessed the expression of lncRNAs by paired-end, high-throughput sequencing. We identified 379 genomic loci encoding novel lncRNAs and performed a comprehensive assessment of cell-specific expression patterns for all, annotated and novel, lncRNAs described to date. Our study provides a powerful new resource for studying these elusive transcripts during stem cell commitment and neurogenesis.

The rise of high-throughput sequencing has led to the identification of novel transcripts on a massive scale and this has fundamentally changed our perception about the prevalence and potential significance of non-coding RNAs.[1-5] In particular for long non-coding (lnc)RNAs, studies have reported that these transcripts compete with conventional protein-coding mRNAs for abundance and diversity.[6-10] Although the proportion of putative lncRNAs that are truly non-coding is being debated[11,12] it is clear that their various degrees of tissue specificity and, in some cases, evolutionary conservation make them prime candidates to play major roles in several biological functions.

In fact, a rapidly growing literature has addressed the role of a few lncRNAs pointing out their involvement in functions as diverse as chromatin remodeling, transcriptional coregulation, molecular decoying or splicing.[1-5,10] Yet, reflecting the novelty of this field, annotation of a large proportion of lncRNAs is still fragmentary and inconsistent. Moreover, while many studies have identified lncRNAs in species as diverse as plants to human and in a multitude of tissues from early embryos to senescent brains[6-9] very few have assessed their expression in specific cell types representative of a given biological process.

In particular for lncRNAs of the mammalian central nervous system, studies have focused on whole developing retinas,[13] developing, adult or senescent brains[14-16] or portion of tissues from different brain areas in order to bioinformatically infer the lineage of neural stem cells.[17] Although important, these studies could not assess the expression of lncRNAs in specific cell populations owning to the intrinsic difficulties associated with the identification and sorting of individual cell types that coexist as intermingled populations in complex tissues. Yet, overcoming this limitation is fundamental to identify lncRNAs differentially expressed in specific cell types as a prerequisite to study their putative function.

To identify specific cell types during mouse corticogenesis, our group has generated and characterized a double-reporter mouse line in which the 3 coexisting populations of neural progenitors undergoing proliferative versus differentiative division and neurons were identified by the combinatorial expression of 2 fluorescent reporters.[18] Specifically, our approach was based on the generation of a RFP line in which the expression of the red reporter was under the control of *Btg2*, a marker of neural progenitors undergoing differentiative divisions to generate either basal (intermediate) neurogenic progenitors or neurons.[19] This *Btg2*$^{\mathrm{RFP}}$ mouse line was then crossed to a previously described GFP line in which expression of the green reporter was under the control of *Tubb3,*[20] an early marker of postmitotic neurons. In essence, FAC-sorting of GFP– progenitors that were either RFP– or

*Correspondence to: Michael Hiller, Federico Calegari; E-mail: hiller@mpi-cbg.de; federico.calegari@crt-dresden.de

Submitted: 09/09/2014; Revised: 11/20/2014; Accepted: 12/02/2014
http://dx.doi.org/10.1080/23262133.2014.995524

RFP+ allowed us to separate the 2 types of proliferative progenitors (PP) and differentiating progenitors (DP), respectively. At the same time, we were able to identify GFP+ newborn neurons independently from the persistence of RFP fluorescence due to the inheritance of the red reporter from their neurogenic mother cells.[18]

Compared to previous transcriptome sequencing studies that analyzed randomly selected individual cells, portion of tissues, developmental stages or different species,[21-24] our approach was the first to compare 3 different cell types coexisting is space and time.[18] This allowed the identification of a small pool of transcripts, referred to as *switch* genes and representing ca. 2% of the whole transcriptome, that were specifically up-/down-regulated in the transient DP population while showing an *opposite* pattern in both PP and neurons (i.e. PP < DP > neuron or PP > DP< neuron for *on-* and *off-switch* genes, respectively). Identification of switch transcripts was particularly important because a remarkably high proportion of differentially expressed genes were up-/downregulated *consistently* in both DP and neurons (i.e., PP < DP < neuron or PP > DP > neuron) with most transcripts among this group being implicated in neuronal specification and maturation but not necessarily in the switch from PP to DP proper.[18] Hence, switch transcripts represent differentially expressed genes characterizing the signature of neurogenic commitment independently from the specification, migration and maturation of postmitotic neurons and differences among tissues, developmental stages or species.

Validating our approach, we found that essentially all known markers of neurogenic commitment were switch transcripts.[18] Moreover, in vivo manipulation of coding as well as non-coding switch transcripts that were either uncharacterized or had no reported function in stem cell commitment or cortical development led to evident phenotypes when overexpressed in neural progenitors of developing mouse embryos ([18] and [25]). In essence, these findings indicated that switch transcripts include a remarkably high proportion of genes functionally involved in neurogenic commitment despite the fact that several of them, particularly lncRNAs, were never studied in any context before.

Given the increasing interest toward understanding the cell biology and function of this elusive class of transcripts, we previously sought to identify novel switch lncRNAs and succeeded in validating at least 2 that we named <u>co</u>rtical <u>s</u>witch <u>l</u>ncRNA 1 (Cosl1) and Btg2-AS1, the latter an antisense, genic lncRNAs overlapping the DP marker Btg2.[18] However, our previous dataset was of limited use for the comprehensive, *de-novo* annotation of new lncRNAs due to the use of single-end sequencing, which makes it difficult to detect the exon-intron structure of novel genes.[26] Hence, we here decided to repeat our analysis by FAC-sorting PP (*Btg2*$^{\text{RFP}}$–/*Tubb3*$^{\text{GFP}}$–), DP (*Btg2*$^{\text{RFP}}$+/*Tubb3*$^{\text{GFP}}$–) and neurons (*Tubb3*$^{\text{GFP}}$+) at embryonic day (E) 14.5 followed by poly-A extraction and 100 bp paired-end, strand-specific, high-throughput sequencing. By this we aimed not only to assess the differential expression, including switch behavior, of already annotated lncRNAs whose profiles in specific neural cell types was unknown but also to identify novel lncRNAs that are unreported in any data set available to date.

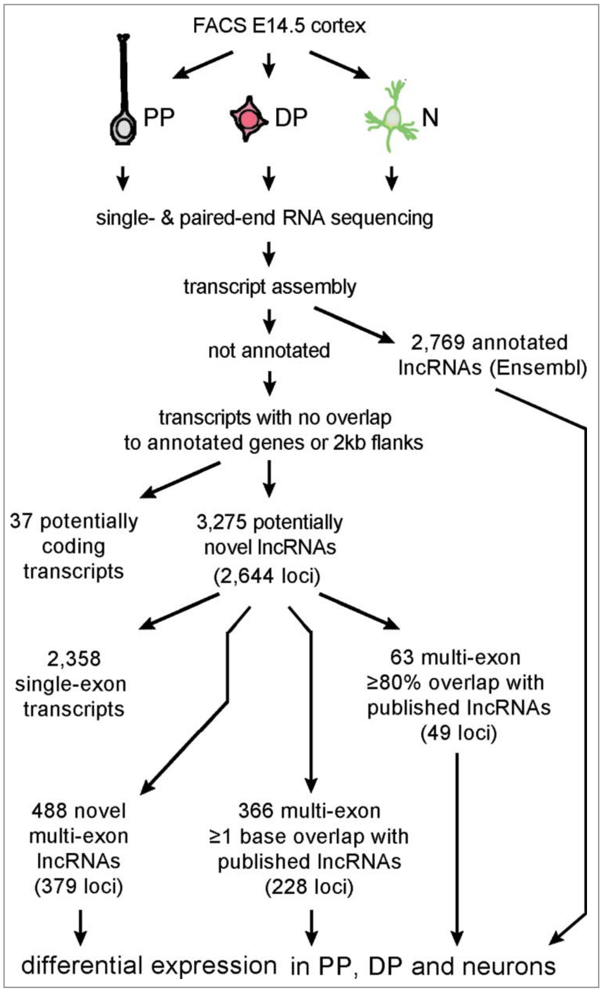
FACS E14.5 cortex
PP
DP
N
single- & paired-end RNA sequencing
transcript assembly
not annotated
2,769 annotated lncRNAs (Ensembl)
transcripts with no overlap to annotated genes or 2kb flanks
37 potentially coding transcripts
3,275 potentially novel lncRNAs (2,644 loci)
2,358 single-exon transcripts
63 multi-exon ≥80% overlap with published lncRNAs (49 loci)
488 novel multi-exon lncRNAs (379 loci)
366 multi-exon ≥1 base overlap with published lncRNAs (228 loci)
differential expression in PP, DP and neurons

Figure 1. Flowchart summarizing the key steps undertaken to identify novel lncRNAs including (from top to bottom) FAC-sorting of PP, DP and neurons (N) of the E14.5 lateral cortex, followed by sequencing, transcript assembly and separation between not annotated (center) and annotated (right) transcripts. Not annotated transcripts overlapping annotated genes or to their 2 kb flanks were excluded together with transcripts showing signatures of protein coding conservation or having one single exon. The remaining multi-exon transcripts were subsequently divided into 3 categories including (from left to right): novel or overlapping previously described lncRNAs by either less or more than 80% of their sequence. Numbers indicate transcripts or genomic loci (within parentheses) being identified. Differential expression analyses were performed for all categories except for not annotated, single-exon transcripts.

We assembled reads from PP, DP and neurons in 3 biological replicates independently for the new 100 bp paired-end as well as for our previous 75 bp single-end sequencing data[18] and analyzed the union of all resulting transcripts. To obtain a set of *bona fide* novel lncRNAs, we applied a series of filtering steps (**Fig. 1**) some of which are widely used to detect novel lncRNAs by sequencing.[27] First, we excluded all transcripts (2,769) already annotated in Ensembl as lncRNAs. Second, we excluded transcripts that overlapped, on the same strand, exons of known coding genes. Similarly, transcripts on the same strand and within 2 kb up-/downstream of an annotated coding gene were also excluded as these are likely 5′ or 3′ extensions of the corresponding coding gene. In doing so, we observed that splice sites of multi-exons transcripts allowed the unambiguous assignment of the transcript to one strand whereas the strand information of single-exon transcripts was often unreliable (data not shown). Hence, we also excluded single-exon transcripts overlapping coding genes, or their 2 kb flanking regions, irrespective of their strand orientation. Third, to increase the reliability of our analysis we kept transcripts that had a higher estimated abundance than their confidence interval[28] in at least 2 of the 3 cell types and, fourth, removed a small number of transcripts (37) that showed signatures of protein-coding conservation across mammalian species and, thus, might represent unannotated coding genes.

These steps ultimately provided us with a list of 3,275 lncRNAs (Supplementary File 1); 917 (28%) of which were multi-exonic. Clustering transcripts by exon overlap, we found that these 3,275 transcripts belonged to 2,644 loci and included isoforms derived from alternative splicing of the same lncRNA (examples in **Fig. S1**).

Next, we compared our list of transcripts to known lncRNAs that may not necessarily be annotated in Ensembl yet. To this end, we compiled a comprehensive list of all lncRNAs reported in 4 sequencing studies focusing on this class of transcripts.[8,12,17,29] To avoid matching transcripts that differ in their exon-intron structure, we required a stringent overlap (≥80 % in similarity). We found that 82 of our 3,275 (2.5%) transcripts overlapped previously reported lncRNAs (examples in **Fig. S1B and C**). This limited number of overlapping transcripts can be explained by the following 3 reasons. First, lncRNAs often have high cell- and tissue-specificity and those expressed in PP, DP and neurons might differ from lncRNAs obtained from the mixture of cell-types analyzed in previous studies of whole brains, other tissues or species. Second, alternative splicing produces high lncRNA transcript diversity (**Fig. S1B-E**), which makes it difficult to find highly similar transcript overlaps. Third, we found that prediction of single-exon transcripts, which comprise the majority (72%) of our 3,275 transcripts, is less reliable than that of multi-exon transcripts (**Fig. S1A**). In fact, while 6.8% (63) of our 917 multi-exon transcripts overlap previously reported lncRNAs, this was the case for only 0.8% (19 of 2,358) single-exon transcripts. To further support this, we counted overlaps down to even a single base and found overlaps for 46.7% (429 of 917) of multi-exon transcripts but only for 13.4% (317 of 2,358) of single-exon transcripts. This bias is most likely explained by the difficulties encountered by computational tools to correctly annotate the boundaries of single-exon transcripts from RNA-seq data. Supporting this, we noticed that a substantial proportion of single-exon transcripts were located in larger genomic regions that are transcriptionally active (example in **Fig. S1A**) possibly indicating transcriptional noise arising, for example, from open chromatin regions. Together with the fact that strand-annotation of single-exon transcripts was unreliable (see above), we decided to continue our analysis by focusing only on the 917 multi-exon transcripts identified thus far.

To provide a set of lncRNAs that are not in any gene annotation and were not detected by previous studies[8,12,17,29] we conservatively removed all transcripts with partial overlaps (even down to one individual base). Doing so led to 488 of our 917 multi-exon transcripts (379 distinct genomic loci) being truly novel and described in our current study for the first time (Supplementary File 2).

Next, we used RT-PCR to i) validate the presence of predicted new transcripts and ii) test the reliability in exon-intron structure annotations in those cases in which our prediction differed significantly from lncRNAs described in other studies, which might be due to alternative splicing and/or to the bioinformatics tools being used. To this aim, we selected 3 novel and 5 known lncRNAs (**Fig. 2**), the latter including transcripts significantly extending previously predicted lncRNAs as well as cases of high (≥80 %) or low (<80 %) overlap (**Fig. 2D**, **2E-G and 2H**, respectively). RT-PCR was performed on cDNAs obtained from the E14.5 brain using primers spanning across different exons (**Table S1**). In all 8 cases, sequence and size of the obtained PCR-amplicons were very closely resembling, if not entirely identical, to the structure predicted based on our sequencing data. In contrast, exon-intron structures differed, even substantially, from the lncRNA annotation of previous sequencing studies using different tissues and/or species[8,12,17,29] (**Fig. 2D-F**). This not only confirms the existence of previously unknown lncRNAs but also suggests that exon-usage of lncRNAs is highly diverse and tissue-specific. As a result, predictions of exon-intron structure based on sequencing analyses in one tissue may not necessarily be confirmed in other tissues or cell types.

An additional advantage of our approach is that transcriptomes of individual cell types likely provide a much higher resolution than whole organs to find lncRNAs with very specific expression patterns. In particular, our transcriptome sequencing in 3 cell populations allowed the classification of transcripts in 9 classes since 3 expression patterns (up-regulation, no change or down-regulation) were defined in the transition from PP to DP as well as in the transition from DP to neurons (**Fig. 3A**). (The third transition from PP to neurons was not considered biologically relevant given the fact that neurons are generated from DP and not PP.) LncRNA abundance was assessed by DESeq2 after pulling together those transcripts derived from the same locus. As a threshold for differential expression we considered an up- or down-regulation by > 50% (FDR 5%) in either PP or neurons relative to DP (i.e. DP/PP or DP/neurons being either < 0.67 or > 1.5). As a validation of this analysis we assessed differential expression of lncRNAs by using a genome-wide depository of *in situ* hybridizations on the developing mouse brain.[30] Among the inspected lncRNAs included in this resource and providing a robust signal above background, essentially all (20/20) displayed a signal distribution within the tissue that was fully consistent with sequencing data. In particular, transcripts highly expressed in PP, DP or neurons were found to be enriched in the ventricular zone, subventricular zone or intermediate zone/cortical plate, respectively (9 examples are shown in **Fig. S2**). This confirms and extends our previous study[18] in which differential expression

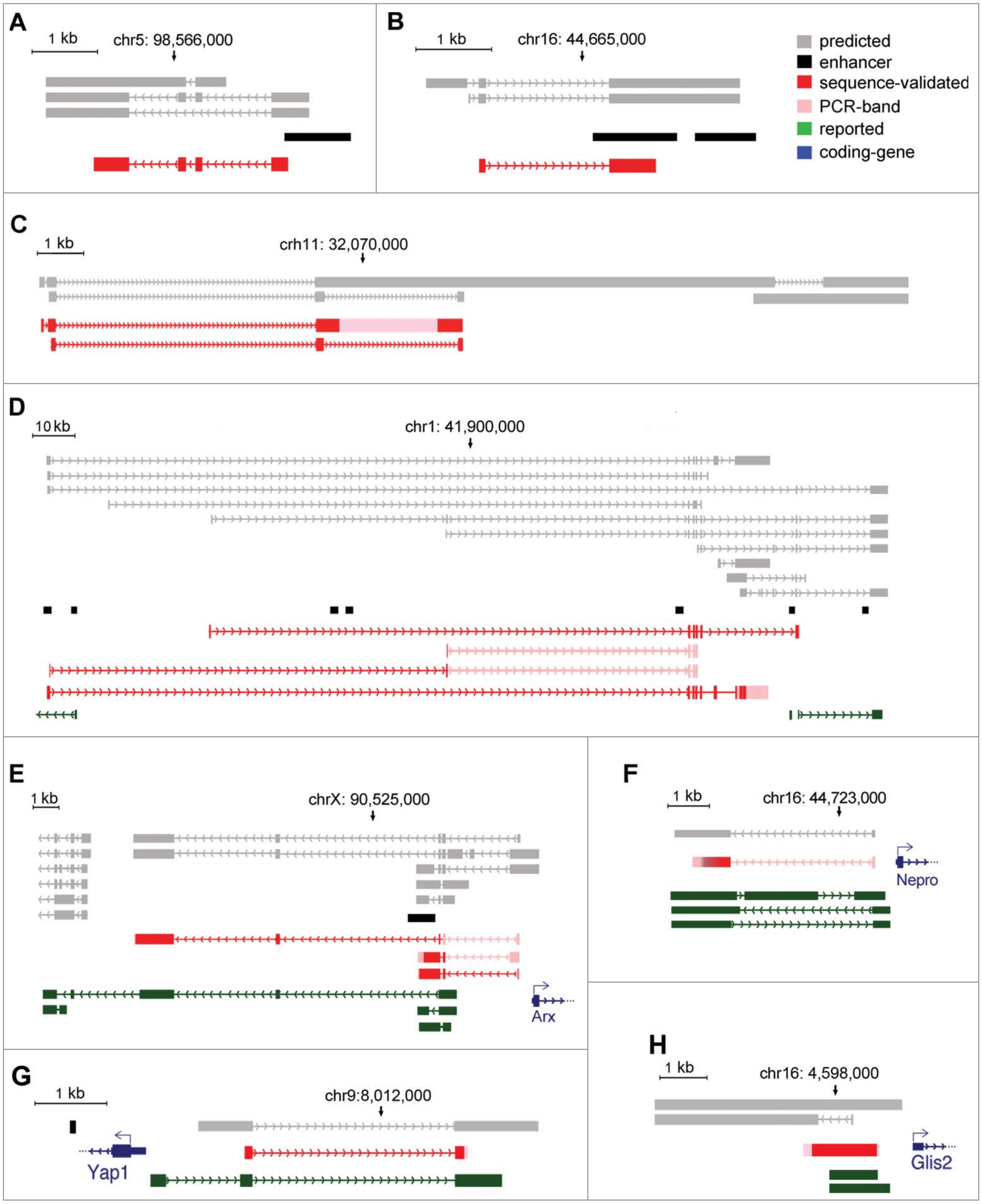

Figure 2. (**A-H**) RT-PCR validation of novel (**A-C**) and lncRNAs overlapping published transcripts (**D-H**) using E14.5 RNA brain extracts as templates and primers spanning across different exons (**Table S1**). Various types of predicted, reported and validated transcripts are depicted by colors as indicated in the legend (top-right), including regions containing neuronal enhancers (black) and/or coding genes (blue). Note the overall similarity in exon usage between predicted (gray) and validated (red) lncRNAs in our study. Arrows indicate the directionality of transcription.

assessed by sequencing was validated at the tissue level for more than 30 established markers of cell types and 24 uncharacterized switch genes. Next, lncRNAs displaying any of the 9 possible expression patterns were identified from the 3 groups of i) Ensembl-annotated lncRNAs (2,769) (**Fig. 3B**), ii) transcripts predicted in our study with any degree of overlap with reported, multi-exon lncRNAs (429) (**Fig. 3C**) and iii) novel, multi-exon lncRNAs (379) (**Fig. 3D**). The proportion of lncRNAs

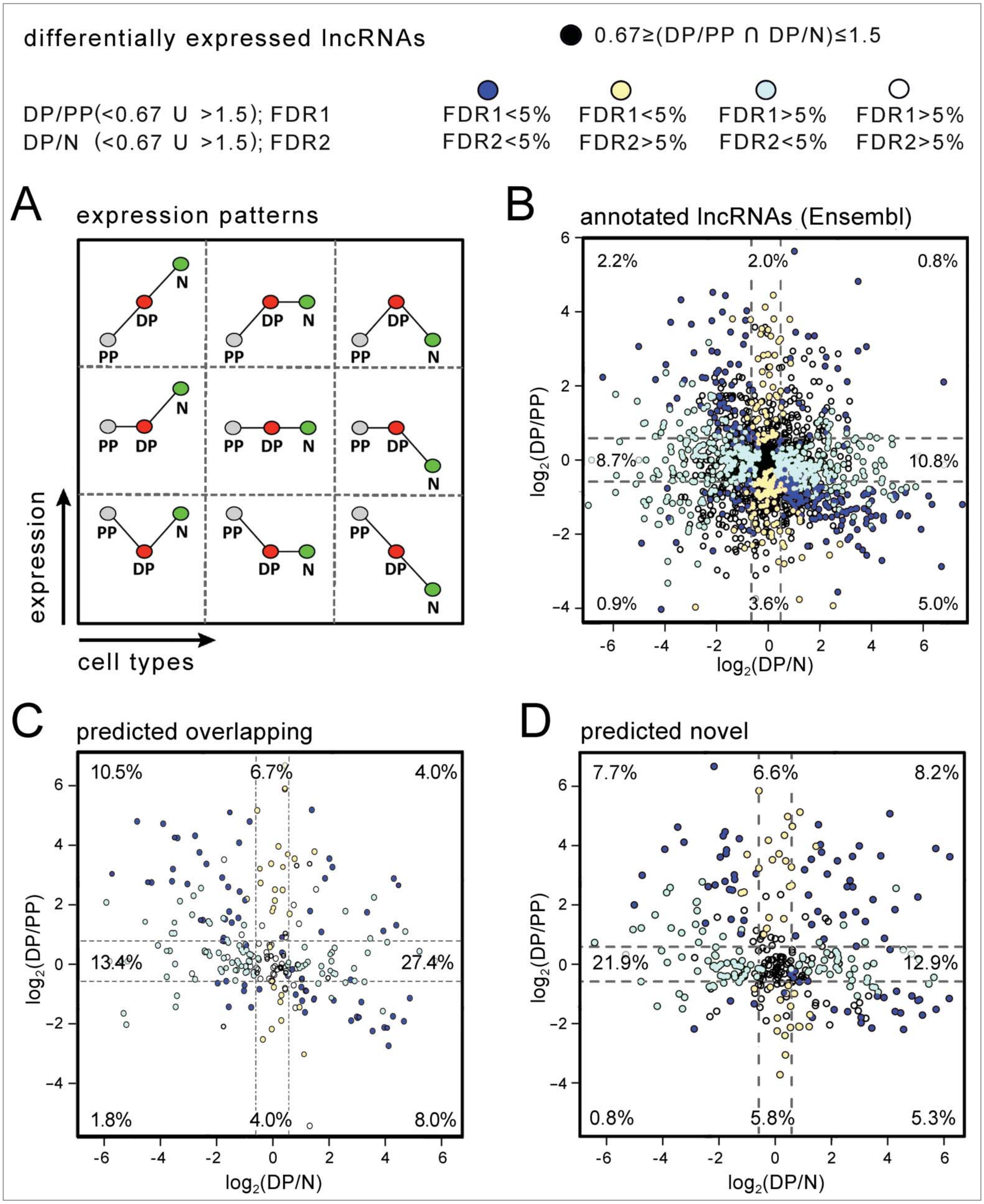

Figure 3. (**A**) Expression patterns of transcripts in the transition from PP (gray) to DP (red) and DP to neurons (green). Up-, constant- or down-regulation are represented by the slope of the lines connecting each cell type for a total of 9 combinations. (**B-D**) Distribution of transcripts according to their expression patterns as shown in (**A**) in the 3 categories of Ensembl-annotated (**B**), overlapping published (**C**), and novel (**D**) lncRNAs. Transcripts were represented by dots of different color according to the statistical significance (FDR, as indicated in the legend; top) of their fold-change ($\log_2$) in DP relative to either PP (DP/PP; *y*-axis) or neurons (DP/N; *x*-axis). Percentages within each quadrant indicate the abundance of a given expression pattern (**B-D**) relative to each category.

displaying any given expression pattern was then referred to the total number of lncRNAs belonging to each group.

For the first group, this yielded 402 (14.5%) annotated lncRNAs being differentially expressed between PP and DP. Among these, 207 continued to be differentially expressed also in the transition from DP to neurons with on- and off-switch lncRNAs representing 22 and 25 transcripts, respectively (together, 1.7% of the total). In addition, 540 (19.5%) lncRNAs were differentially expressed solely in the transition from DP to neurons but not from PP to DP (**Fig. 3B** and Supplementary File 3). Similar results were obtained with lncRNAs overlapping transcripts described in other studies corresponding to 97 (35.0%) lncRNAs being differentially expressed from PP to DP including 11 and 5 on- and off-switch, respectively (together, 5.8% of the total) (**Fig. 3C** and Supplementary File 4). Finally, analysis of novel lncRNAs yielded 130 (34.3%) loci being differentially expressed from PP to DP with 31 and 3 on- and off-switch lncRNAs, respectively (together, 9.0% of the total) (**Fig. 3D** and Supplementary File 4).

Notably, the proportion of differentially expressed lncRNAs, particularly on- and off-switch, was higher among the group of novel transcripts compared to overlapping or annotated genes (9.0%, 5.8% and 1.7%, respectively). This corroborates the conclusion that transcriptome analysis of individual cell types provides a better resolution than whole organs to identify differential expression. Given the assumption that differential expression provides a proxy for biological function, we conclude that our list of novel lncRNAs is more significantly enriched in transcripts functionally involved in neurogenic commitment than previous datasets.

Moreover, knowing that a significant proportion of lncRNAs overlap coding genes in antisense orientation we next examined to which extent the expression of neighboring coding and non-coding transcripts correlated. For this, we grouped all lncRNAs predicted by our study and used lncRNAs annotated in Ensembl as a second group of reference. In both groups, we observed a similar and strong correlation, either negative or positive, between the expression of adjacent coding- and lnc-transcripts (**Fig. 4A and B**). Interestingly, this correlation did not seem to be influenced by the distance of the loci in consideration since overlapping, genic and intergenic lncRNAs, even well over 100 kb apart, displayed very similar trends (**Fig. 4C**).

Moreover, the association between coding- and lnc-RNAs allowed us to assess gene ontology terms of the former to investigate whether terms related to corticogenesis were proportionally over-represented. Given that the correlation in the expression levels of coding- and lnc-RNAs was undistinguishable in the 2 groups of known and novel lncRNAs, we decided to pull these together and from this we selected all lncRNAs differentially expressed in any cell type. Consistent with a previous report on the developing brain,[29] gene ontology analysis of coding transcripts adjacent to differentially expressed lncRNAs revealed a consistent enrichment in GO terms related to regulation of neurogenesis, transcription and development (**Fig. 4D**). Again, this feature was independent from the distance between coding and non-coding loci (**Fig. 4D**).

All together, and to the best of our knowledge, our study provides the most comprehensive, cell type-specific description of differentially expressed lncRNAs during neurogenic commitment of mammalian cortical progenitors. Surprisingly, and despite the many exhaustive sequencing studies focusing on lncRNAs, we were here able to substantially increase the number of loci driving the expression of nearly 500 multi-exonic transcripts, several of which undergo alternative splicing. Importantly, our estimation of the number of novel transcripts represents a very conservative estimate given the fact that over 2 thousand single-exon transcripts could not be assessed reliably due to the limitations in sequencing and bioinformatics tools that are currently available. Moreover, with regard to the prediction of exon usage derived from deep-sequencing data, we observed that in all cases validated our assessment was closer to the true variants expressed in the developing mouse cortex than that of previous studies focusing on other tissues or species.[8,12,17,29] These differences in splicing of lncRNAs indicates the superiority of cell-specific analysis in identifying tissue-specific splice variants of these elusive transcripts. Supporting this, we found a substantial increase in the fraction of differentially expressed lncRNAs among the group of transcripts identified in our study than in those previously reported. Hence, we believe that our work provides the community with a highly specific and powerful resource to study the role of lncRNAs in stem cell commitment and brain development.

Materials and Methods

Animals, sorting and sequencing

The *Btg2*$^{\mathrm{RFP}}$/*Tubb3*$^{\mathrm{GFP}}$ line and sorting of E14.5 PP, DP and neurons was recently reported.[18] Briefly, cortices were dissociated using the papain-based neural dissociation kit (Milteney Biotec) and FACS performed at 4°C in the 4-way purity mode gating green (488 nm) and red (561 nm) fluorescence to collect ca. 1 × 10^6 cells from >3 embryos from different litters. Cells were immediately lysed in μMACS$^{\mathrm{TM}}$ mRNA Isolation Kit and lysates cleaned on LysateClear Colums (Miltenyi) resulting in ca. 1 μg of poly-A RNAs with an integrity number > 9.2. Libraries were prepared as described[18] using oligo(dT) for transcripts selection, first strand cDNA synthesis by random primers, second strand synthesis, end-repair, adaptor ligation, dUTP cleavage and enrichment with indexed primers. Sequencing was performed on an Illumina HiSeq 2000 platform pooling samples in 2 lanes for 100 bp paired-end sequencing resulting in ca. 90 million reads per sample. Raw data were deposited in NCBI–GO (accession Nr. GSE65487).

Bioinformatics and statistics

Sequencing data were obtained for PP, DP and neurons in 3 biological replicates and assembly performed independently for the new 100 bp paired-end and 75 bp single-end reads

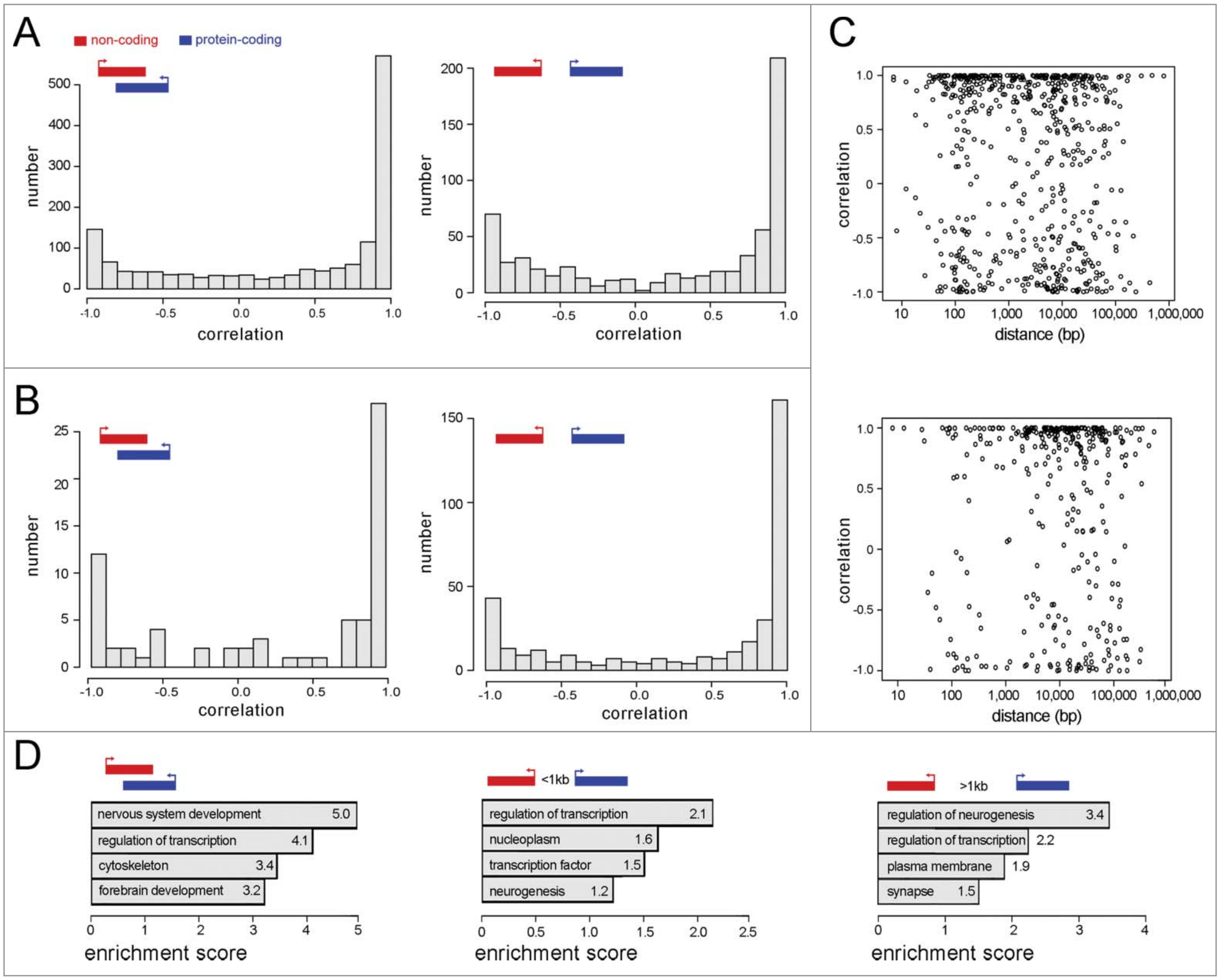

Figure 4. (**A and B**) Distribution of Pearson's correlation coefficients of expression levels for pairs of coding (blue) and non-coding (red) transcripts annotated in Enembl (**A**) or predicted by our study (**B**). Loci overlapping in antisense (left) or adjacent (right) are depicted separately as indicated by the respective legends (top-left). (**C**) Pearson's correlation coefficients (*y*-axis) of pairs of adjacent coding and non-coding transcripts annotated in Ensembl (top) or predicted by our study (bottom) and represented as a function of their distance (bp; *x*-axis). Note that the vast majority of pairs of transcripts have an almost perfect negative (-1.0) or positive (1.0) linear correlation that seems not to be influenced by their distance even up to 1 megabase. (**D**) Gene ontology terms (DAVID) of coding transcripts overlapping (left), within (middle) or beyond (right) 1 kb distance from a differentially expressed lncRNA.

previously reported.[18] The "Tuxedo Suite" of Bowtie, Tophat and Cufflinks[31-34] was used for transcript assembly. We aligned reads to the mm9 mouse genome using the splice junction mapper Tophat (version 2.0.10) that used Bowtie 2 (version 2.2.1) for mapping. Tophat was first called with the paired-end data and called again with the single-end data and junctions from the paired-end. We did not use any annotation file for each of the alignments. In the second step, we used Cufflinks on each Tophat run with the gene annotation of Ensembl version 67 to guide a reference annotation based transcript assembly. All transcript files were merged with Cuffmerge. To obtain lncRNAs, we excluded transcripts that overlap known gene annotations. Specifically, we merged gene annotations (UCSC genome browser tracks knownGene, refGene, vegaGene, exoniphy, ccdsGene, ensGene and mgcFullMrna) with pseudogene tracks (vegaPseudoGene, pseudoYale60, ucscRetroAli1). Bedtools (version 2.16.2) was used to exclude multi-exon transcripts that overlap any transcript or its 2 kb up- and down-stream flanks on the same strand. Overlapping single-exon transcripts were excluded regardless of strand since this information for single-exon transcripts was unreliable. Next, we removed transcripts that had a lower estimated abundance than their confidence interval.[27] Potentially coding genes were identified by RNAcode[35] on exons of novel transcripts using the mouse mm9 genome alignment with a total of 43 species[36] (window of at least 30 bp; $p<0.0001$). For lncRNA annotations from other studies,[8,12,17,29] we matched the accession numbers against the UCSC table mRNA to convert the loci from the mm5[29] to the mm9 assembly and overlapped our transcripts with known lncRNAs by requiring $\geq$ 80% similarity as: (2*exonic base overlap)/(exonic bases in our transcript + exonic bases in a known transcript).

Validation of lncRNAs

Validation of lncRNAs was performed by RT-PCR of E14.5 mouse brains using primers designed on Cufflinks prediction (Supplementary Data and **Table S1**). Amplicons were assessed by size and Sanger sequencing.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank the MPI-CBG and CRTD facilities for maintenance of the mouse lines and FACS. Authors' contributions: JA, ML, MH and FC conceived the project; JA performed the experiments supported by SM and SP; ML, MH and JA performed the bioinformatic analyses supported by DA and AD. FC wrote the manuscript. All authors approved the manuscript.

Funding

This work was supported by the CRTD, the TUD, the DFG Collaborative Research Center SFB655 (subprojects A20 and Z2), research grant CA 893/9–1 and the BMBF (project NGSgoesHPC).

Supplemental Material

Supplemental data for this article can be accessed on the publisher's website.

References

1. Ponting, CP, Oliver, PL, Reik, W. Evolution and functions of long noncoding RNAs. Cell 136, 629-41, (2009); PMID:19239885; http://dx.doi.org/10.1016/j.cell.2009.02.006
2. Ulitsky, I, Bartel, DP. lincRNAs: genomics, evolution, and mechanisms. Cell 154, 26-46, (2013); PMID:23827673; http://dx.doi.org/10.1016/j.cell.2013.06.020
3. Nagano, T, Fraser, P. No-nonsense functions for long noncoding RNAs. Cell 145, 178-81, (2011); PMID:21496640; http://dx.doi.org/10.1016/j.cell.2011.03.014
4. Mercer, TR, Mattick, JS. Structure and function of long noncoding RNAs in epigenetic regulation. Nat Struct Mol Biol 20, 300-7, (2013); PMID:23463315; http://dx.doi.org/10.1038/nsmb.2480
5. Mercer, TR, Dinger, ME, Mattick, JS. Long non-coding RNAs: insights into functions. Nat Rev Genet 10, 155-9, (2009); PMID:19188922; http://dx.doi.org/10.1038/nrg2521
6. Carninci, P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C, et al. The transcriptional landscape of the mammalian genome. Science 309, 1559-63, (2005); PMID:16141072; http://dx.doi.org/10.1126/science.1112014
7. Birney, E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447, 799-816, (2007); PMID:17571346; http://dx.doi.org/10.1038/nature05874
8. Necsulea, A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, Baker JC, Grützner F, Kaessmann H. The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature 505, 635-40, (2014); PMID:24463510; http://dx.doi.org/10.1038/nature12943
9. Katayama, S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, Nishida H, Yap CC, Suzuki M, Kawai J, et al. Antisense transcription in the mammalian transcriptome. Science 309, 1564-6, (2005); PMID:16141073; http://dx.doi.org/10.1126/science.1112009
10. Flynn, RA, Chang, HY, Long Noncoding RNAs in Cell-Fate programming and reprogramming. Cell Stem Cell 14, 752-61, (2014); PMID:24905165; http://dx.doi.org/10.1016/j.stem.2014.05.014
11. Ingolia, NT, Lareau, LF, Weissman, JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell 147, 789-802, (2011); PMID:22056041; http://dx.doi.org/10.1016/j.cell.2011.10.002
12. Guttman, M, Russell, P, Ingolia, NT, Weissman, JS, Lander, ES. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. Cell 154, 240-51, (2013); PMID:23810193; http://dx.doi.org/10.1016/j.cell.2013.06.009
13. Blackshaw, S, Harpavat S, Trimarchi J, Cai L, Huang H, Kuo WP, Weber G, Lee K, Fraioli RE, Cho SH, et al. Genomic analysis of mouse retinal development. PLoS Biol 2, E247, (2004); PMID:15226823; http://dx.doi.org/10.1371/journal.pbio.0020247
14. Lv, J, Cui W, Liu H, He H, Xiu Y, Guo J, Liu H, Liu Q, Zeng T, Chen Y, et al. Identification and characterization of long non-coding RNAs related to mouse embryonic brain development from available transcriptomic data. PLoS One 8, e71152, (2013); PMID:23967161; http://dx.doi.org/10.1371/journal.pone.0071152
15. Mercer, TR, Dinger, ME, Sunkin, SM, Mehler, MF, Mattick, JS. Specific expression of long noncoding RNAs in the mouse brain. Proc Natl Acad Sci U S A 105, 716-21, (2008); PMID:18184812; http://dx.doi.org/10.1073/pnas.0706729105
16. Wood, SH, Craig, T, Li, Y, Merry, B, de Magalhaes, JP. Whole transcriptome sequencing of the aging rat brain reveals dynamic RNA changes in the dark matter of the genome. Age 35, 763-76, (2013); PMID:22555619; http://dx.doi.org/10.1007/s11357-012-9410-1
17. Ramos, AD, Diaz A, Nellore A, Delgado RN, Park KY, Gonzales-Roybal G, Oldham MC, Song JS, Lim DA. et al. Integration of genome-wide approaches identifies lncRNAs of adult neural stem cells and their progeny in vivo. Cell Stem Cell 12, 616-28, (2013); PMID:23583100; http://dx.doi.org/10.1016/j.stem.2013.03.003
18. Aprea, J, Prenninger S, Dori M, Ghosh T, Monasor LS, Wessendorf E, Zocher S, Massalini S, Alexopoulou D, Lesche M., et al. Transcriptome sequencing during mouse brain development identifies long non-coding RNAs functionally involved in neurogenic commitment. EMBO J 32, 3145-60, (2013); PMID:24240175; http://dx.doi.org/10.1038/emboj.2013.245
19. Haubensak, W, Attardo, A, Denk, W, Huttner, WB. Neurons arise in the basal neuroepithelium of the early mammalian telencephalon: a major site of neurogenesis. Proc. Natl. Acad. Sci. USA 101, 3196-201 (2004); PMID:14963232
20. Attardo, A, Calegari, F, Haubensak, W, Wilsch-Brauninger, M, Huttner, WB. Live imaging at the onset of cortical neurogenesis reveals differential appearance of the neuronal phenotype in apical vs. basal progenitor progeny. PLoS ONE 3, e2388, (2008); PMID:18545663; http://dx.doi.org/10.1371/journal.pone.0002388
21. Ayoub, AE, Oh S, Xie Y, Leng J, Cotney J, Dominguez MH, Noonan JP, Rakic P. et al. Transcriptional programs in transient embryonic zones of the cerebral cortex defined by high-resolution mRNA sequencing. Proc Natl Acad Sci U S A 108, 14950-5, (2011); PMID:21873192; http://dx.doi.org/10.1073/pnas.1112213108
22. Han, X, Wu X, Chung WY, Li T, Nekrutenko A, Altman NS, Chen G, Ma H. Transcriptome of embryonic and neonatal mouse cortex by high-throughput RNA sequencing. Proc Natl Acad Sci U S A 106, 12741-6, (2009); PMID:19617558; http://dx.doi.org/10.1073/pnas.0902417106
23. Yao, MJ, Chen G, Zhao PP, Lu MH, Jian J, Liu MF, Yuan XB. Transcriptome analysis of microRNAs in developing cerebral cortex of rat. BMC genomics 13, 232, (2012); PMID:22691069; http://dx.doi.org/10.1186/1471-2164-13-232
24. Fietz, SA, Lachmann R, Brandl H, Kircher M, Samusik N, Schröder R, Lakshmanaperumal N, Henry I, Vogt J, Riehn A, et al. Transcriptomes of germinal zones of human and mouse fetal neocortex suggest a role of extracellular matrix in progenitor self-renewal. Proc Natl Acad Sci U S A 109, 11836-41, (2012); PMID:22753484; http://dx.doi.org/10.1073/pnas.1209647109
25. Artegiani B, de Jesus, Domingues AM, Bragado Alonso A, Brandl E, Massalini S, Dahl A Calegari F Tox.: A multifunctional transcription factor and novel regulator of mammalian corticogenesis. EMBO J (2014) PMID:25527292; DOI:10.15252/embj.201490061
26. Medvedev, P, Stanciu, M, Brudno, M. Computational methods for discovering structural variation with next-generation sequencing. Nat Methods 6, S13-20, (2009); PMID:19844226; http://dx.doi.org/10.1038/nmeth.1374
27. Ilott, NE, Ponting, CP. Predicting long non-coding RNAs using RNA sequencing. Methods 63, 50-9, (2013); PMID:23541739; http://dx.doi.org/10.1016/j.ymeth.2013.03.019

28. Nagaraj, N, Wisniewski JR, Geiger T, Cox J, Kircher M, Kelso J, Pääbo S, Mann M. Deep proteome and transcriptome mapping of a human cancer cell line. Mol Syst Biol 7, 548, (2011); PMID:22068331; http://dx.doi.org/10.1038/msb.2011.81
29. Ponjavic, J, Oliver, PL, Lunter, G, Ponting, CP, Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain. PLoS Genet 5, e1000617, (2009); PMID:19696892; http://dx.doi.org/10.1371/journal.pgen.1000617
30. Diez-Roux, G, Banfi S, Sultan M, Geffers L, Anand S, Rozado D, Magen A, Canidio E, Pagani M, Peluso I, et al. A high-resolution anatomical atlas of the transcriptome in the mouse embryo. PLoS Biol 9, e1000582, (2011); PMID:21267068; http://dx.doi.org/10.1371/journal.pbio.1000582
31. Langmead, B, Salzberg, SL. Fast gapped-read alignment with bowtie 2. Nat Methods 9, 357-9, (2012); PMID:22388286; http://dx.doi.org/10.1038/nmeth.1923
32. Kim, D. Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14, R36, (2013); PMID:23618408; http://dx.doi.org/10.1186/gb-2013-14-4-r36
33. Trapnell, C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28, 511-5, (2010); PMID:20436464; http://dx.doi.org/10.1038/nbt.1621
34. Liao, Y, Smyth, GK, Shi, W. Featurecounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923-30, (2014); PMID:24227677; http://dx.doi.org/10.1093/bioinformatics/btt656
35. Washietl, S, Findeiss S, Müller SA, Kalkhof S, von Bergen M, Hofacker IL, Stadler PF, Goldman N. et al. RNAcode: robust discrimination of coding and noncoding regions in comparative sequence data. RNA 17, 578-94, (2011); PMID:21357752; http://dx.doi.org/10.1261/rna.2536111
36. Hiller, M, Schaar BT, Indjeian VB, Kingsley DM, Hagey LR, Bejerano G. A "forward genomics" approach links genotype to phenotype using independent phenotypic losses among related species. Cell Rep 2, 817-23, (2012); PMID:23022484; http://dx.doi.org/10.1016/j.celrep.2012.08.032
37. Huang da, W, Sherman, BT, Lempicki, RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37, 1-13, (2009); PMID:19033363; http://dx.doi.org/10.1093/nar/gkn923