

RESEARCH REPORT

TECHNIQUES AND RESOURCES

The tissue-specific transcriptomic landscape of the mid-gestational mouse embryo

Martin Werber¹, Lars Wittler¹, Bernd Timmermann², Phillip Grote^{1,*} and Bernhard G. Herrmann^{1,3,*}

ABSTRACT

Differential gene expression is a prerequisite for the formation of multiple cell types from the fertilized egg during embryogenesis. Understanding the gene regulatory networks controlling cellular differentiation requires the identification of crucial differentially expressed control genes and, ideally, the determination of the complete transcriptomes of each individual cell type. Here, we have analyzed the transcriptomes of six major tissues dissected from mid-gestational (TS12) mouse embryos. Approximately one billion reads derived by RNA-seq analysis provided extended transcript lengths, novel first exons and alternative transcripts of known genes. We have identified 1375 genes showing tissue-specific expression, providing gene signatures for each of the six tissues. In addition, we have identified 1403 novel putative long noncoding RNA gene loci, 439 of which show differential expression. Our analysis provides the first complete transcriptome data for the mouse embryo. It offers a rich data source for the analysis of individual genes and gene regulatory networks controlling mid-gestational development.

KEY WORDS: RNA-seq, Embryogenesis, Long noncoding RNA, Mouse, Tissue-specific expression, Transcriptome

INTRODUCTION

The mouse is a well-established model organism for the investigation of mammalian development and human disease. The elucidation of the mouse genome sequence has provided access to any gene locus in the genome. Various approaches for gene function analysis *in vivo* have been described, making any gene amenable to loss-of-function studies. However, annotation of the transcriptome is still far from being finalized as the transcriptome differs from cell type to cell type and changes rapidly during development. Large-scale analyses of the mouse transcriptome have assessed temporal changes of 19,000 coding genes in the early mouse embryo [embryonic day (E) 6.5 to E9.0, corresponding to Theiler stage (TS) 9–14] using a custom microarray platform (Mitiku and Baker, 2007). In different approaches, single genes have been assayed for expression at various embryonic stages or in different tissues using *in situ* hybridization techniques (Diez-Roux et al., 2011; Gray et al., 2004). Whole-mount expression data have been obtained for 1912 genes of various functions by our group (<http://mamep.molgen.mpg.de>), and for 1561 transcription factors and TF-related genes (<http://embryos.jp>) analyzed in mid-gestational embryos (E8.5 to E11.5; reviewed by Geffers et al., 2012). All of these datasets provide invaluable information pertaining to

mainly protein-coding genes. However, none of these datasets has provided an accurate representation of the transcriptome of the mouse embryo. In particular, earlier studies using expression profiling did not cover the complete set of protein coding genes or alternative transcripts or noncoding RNA genes. The latter have come into focus in recent years, as noncoding genes are assumed to play important roles in gene regulation (for reviews, see Pauli et al., 2011; Rinn and Chang, 2012).

Among the highly diverse class of noncoding genes, long noncoding RNAs (lncRNAs) are thought to influence transcription by a wide range of mechanisms. For example, lncRNAs can interact with chromatin-modifying protein complexes involved in gene activation or silencing, and thus may be directly involved in epigenetic gene control (for a review, see Ulitsky and Bartel, 2013). Recent data generated from human cell lines by the ENCODE (Encyclopaedia of DNA Elements) Project Consortium provided a comprehensive view of the abundance of lncRNA genes in humans (Derrien et al., 2012; Dunham et al., 2012). However, despite recent functional analysis of a few lncRNA genes, not much is known about the functional role of lncRNAs in tissue differentiation and organogenesis in the mouse embryo (Grote et al., 2013; Li et al., 2013; Sauvageau et al., 2013).

RESULTS AND DISCUSSION

RNA deep sequencing of six tissues from the TS12 mouse embryo

We have determined the transcriptome of Theiler stage 12 (corresponding to E8.25, 3–6 somites) mouse embryos. To obtain a global view of the transcriptomic landscapes of different tissues at single-nucleotide resolution, we micro-dissected five prominent tissues comprising most of the embryo mass: the mesoderm caudal to the somites – predominantly containing presomitic and lateral mesoderm (caudal end mesoderm: CEM); the somites; the heart; the head (comprising the future brain, including the rhombencephalon and head mesoderm); and the presumptive spinal cord (PSC), as well as the remainder (carcass; extra-embryonic tissue was largely removed) from ten embryos (Fig. 1A,B). We generated cDNA from total RNA of each tissue and sequenced it on an Illumina HiSeq 2000 (RNA-seq).

We obtained 100 base paired-end reads and mapped approximately one billion reads in total against the latest mouse genome build mm10 (GRCm38) (supplementary material Table S1). The mapping results, coverage, and splice junction mappings are visualized and accessible via the UCSC genome browser (<http://overview.molgen.mpg.de/ucsc.html>; Fig. 1C). In each sample, the vast majority of reads mapped to coding exons (69.4–73.4%) and UTRs (19.6–21.1%), while the remaining reads mapped within introns (4.5–7.8%) or intergenic (1.3–2.0%) regions (supplementary material Table S2).

We calculated normalized transcript levels for each of the six samples on the basis of all annotated ENSEMBL transcripts with cufflinks, taking into account the default parameters for the alignment status (status=OK) and significance thresholds for FDR

¹Department of Developmental Genetics, Max-Planck-Institute for Molecular Genetics, Ihnestrasse 63-73, 14195 Berlin, Germany. ²Sequencing Core Facility, Max-Planck-Institute for Molecular Genetics, Ihnestrasse 63-73, 14195 Berlin, Germany. ³Institute for Medical Genetics, Charité – University Medicine Berlin, Campus Benjamin Franklin, Hindenburgdamm 30, 12203 Berlin, Germany.

*Authors for correspondence (grote@molgen.mpg.de; herrmann@molgen.mpg.de)

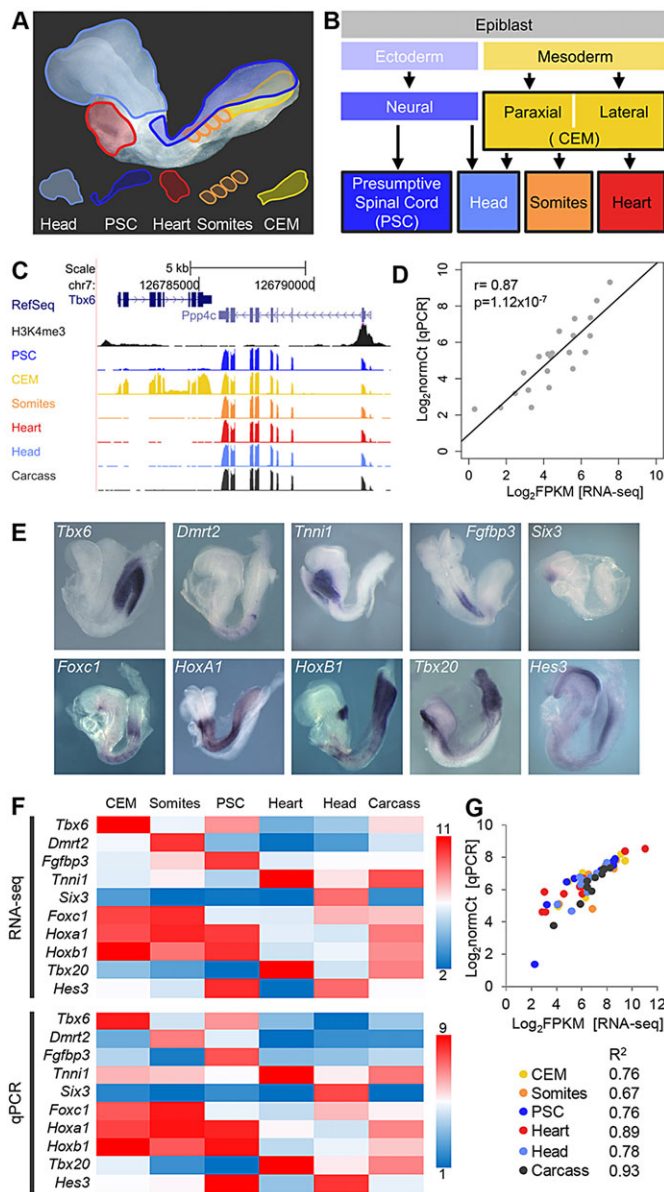


Fig. 1. Transcriptome analysis of six tissues from TS12 mouse embryos by RNA-seq provides qualitative and quantitative differential expression data. (A) Schematic of the tissues dissected from TS12 (E8.25, 3-6 somites) mouse embryos. PSC, presumptive spinal cord; CEM, caudal end mesoderm. The remains were combined in the carcass sample. (B) Schematic of cell lineages and organs originating from the epiblast; dissected tissues are framed. (C) Schematic representation of a browser window with RefSeq annotations, whole-embryo H3K4me3 ChIP-seq and RNA-seq tracks (y-axis of RNA-seq track is 2600 \times coverage). (D) Relative expression profiling of 22 housekeeping genes by RNA-seq (\log_2 FPKM) and by qPCR (\log_2 normCt) performed on two biological replicates. (E) Whole-mount *in situ* hybridization analysis of ten genes specifically expressed in one or more tissues. (F) Heatmap representation of expression levels detected by RNA-seq and qPCR on independent biological replicates. The colour scale bar indicates \log_2 values. (G) Scatterplot and the R^2 values for the correlation between the RNA-seq and the qPCR data.

(5%; significant=yes). To validate the accuracy of normalized transcript levels obtained from RNA-seq data, we compared relative transcript levels of 22 housekeeping genes obtained by quantitative real-time PCR (qPCR) to normalized transcript levels calculated from RNA-seq data and found that they correlate strongly ($r=0.87$

$p=1.12E-07$; Fig. 1D). To confirm that the RNA-seq data accurately reflect the tissue specificity of gene expression, we compared the expression data derived by RNA-seq with whole-mount *in situ* hybridization data for ten marker genes (Fig. 1E,F). The tissue-specific expression was well reflected by RNA-seq data. In addition, we compared RNA-seq data to quantitative RNA measurements by qPCR on an independent, biological replicate (Fig. 1F,G). The values correlated well between the two methods. Taken together, our transcriptome dataset provides accurate quantitative information of transcript levels in each of the six tissues.

For a comparison of the number of genes expressed in each tissue, we focused on fully read-covered transcripts derived from genes annotated in ENSEMBL [cufflinks alignment status (status=OK)]. Using this approach, we detected expression of 12,864 genes in the CEM, 13,384 in the head, 13,477 in the heart, 12,738 in the PSC, 12,674 genes in the somites and 14,158 in the carcass sample. Combining results over all sampled tissues, we detected significant transcription of 14,158 genes representing 38% of all 37,224 annotated genes.

We identified 1375 genes that are differentially expressed between the six tissue samples using cufflinks and cuffdiff with default settings (Fig. 2A and supplementary material Table S3). To extract co-expressed gene groups from this set, we applied hierarchical clustering to gene-normalized (Z-score-transformed) data for all tissues using Genesis (Sturn et al., 2002). We identified a specific gene signature for each of the six tissues. The CEM (52), head (85), PSC (37) and somites (68) revealed relatively small tissue-specific gene sets, while the heart (442) and the carcass sample (389) displayed significantly more specifically expressed genes (supplementary material Table S4). In addition to many key developmental genes, these clusters contain a number of genes that are not well annotated, and thus significantly extend the gene catalogues for studying fundamental developmental processes.

We visualized similarities in global expression and correlation between samples by applying Self Organizing Maps clustering and mosaic image graphing (GEDI-maps, Gene Expression Dynamics Inspector) (Eichler et al., 2003). The GEDI-maps illustrate that the differentially expressed gene sets are largely complementary between the six tissues, but also show significant overlap between head and PSC sharing neural tissue, between CEM and PSC sharing common progenitors, between CEM and one of its derivatives, the somites, and between the heart and carcass (Fig. 2B).

In order to characterize the functional significance of the tissue-specific gene signatures, we applied gene ontology enrichment analysis using DAVID for three categories: molecular function, biological process and cellular component (GOTERM_MF_FAT, GOTERM_BP_FAT, GOTERM_CC_FAT) (Huang et al., 2009). The top GO terms (cut off: $FDR \leq 0.05$), ranked by q-value, reflect the varied differentiation stages of the tissue samples (Fig. 2C).

Alternative splicing and extended UTRs

Alternative splicing greatly increases the variety of functionally distinct proteins in the cell. With the RNA-seq method employed here, the splicing of primary RNA can be observed by mapping of sequence read clusters to two genomic regions separated by intervening DNA fragments (introns). We compiled a unified database of splice junctions from all samples in our dataset using Tophat (Trapnell et al., 2009) and a second aligner, STAR (Dobin et al., 2013), and used this database with Tophat to prime additional junction alignments.

We evaluated the precision of splice junction mapping by comparing our mappings against annotated splice sites from public

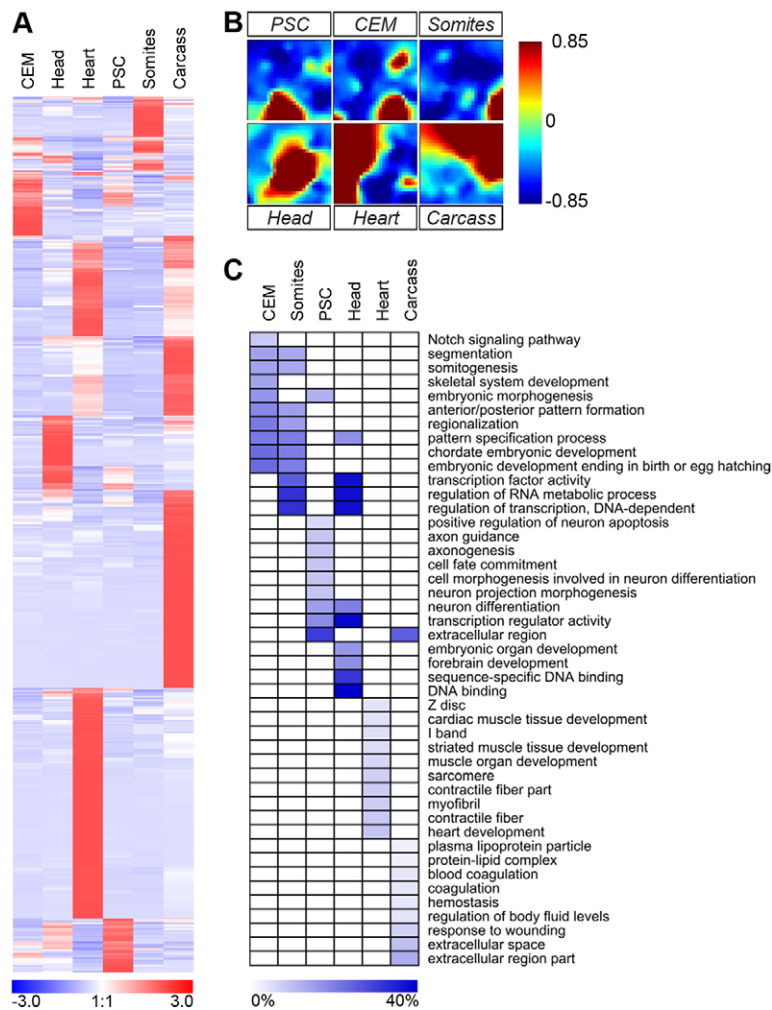


Fig. 2. Differential expression data of protein-coding genes identifies tissue-specific gene signatures. (A) Heatmap representation of 1375 differentially expressed protein-coding genes determined by pair-wise comparison. FPKM values were median normalized at the gene level. (B) GEDI maps illustrate divergent and overlapping expression among the six tissues. Each gene expression mosaic shows a heat-coloured self-organizing map (SOM) mosaic. Each tile represents a cluster of genes with similar expression across tissues. The colour indicates the expression strength of a gene cluster (blue, low expression; red, high expression). (C) Heatmap of gene function enrichment in various GO categories (DAVID). Results are displayed as a sorted heatmap depicting the percentage of preferentially expressed genes annotated to a particular function.

reference annotations (PRA) (ENSEMBL v.69). The majority of the splice junctions (75% on average) mapped exactly to the PRA positions, demonstrating that our approach reliably detects splice junctions (Fig. 3A). One quarter of the splice junctions fall into one of the following two groups: partially in agreement with PRA (partially novel), whereby either splice donor or splice acceptor site are different (17.4%); and completely novel (6.7%).

We determined the transcript isoforms from the six tissue samples and assigned them to seven transcript classes (Fig. 3B). In each tissue, the number of novel transcript isoforms exceeds the number of transcripts, which perfectly match PRA datasets. We classified alternative splice junctions and exon coverage with respect to the gene structure found in the PRA, into the following classes with the software ASprofile: exon skipping (SKIP), alternative transcript start or termination site (TSS, TTS), and alternative exon (AE) (Florea et al., 2013). The vast majority of novel events comprise alternative or extended first and last exons, most likely reflecting frequent tissue-specific differences in the choice of TSS and TTS [Fig. 3C and <http://dx.doi.org/10.6084/m9.figshare.993872> (see also http://overview.molgen.mpg.de/Table_alternative_exon_usage.zip)]. To evaluate whether the novel splicing events giving rise to alternative transcripts are embryo specific, we compared the junctions identified in our embryonic heart dataset against the results of a reanalysed adult heart dataset from ENCODE (using a cut-off of ten reads spanning a splice junction) (ENCODE Project Consortium, 2011). We find that out of 57,641 novel exons 96.1% (55,423) are embryonic heart

specific (3.9% also found in ENCODE). These data illustrate the excessive use of alternative promoters and alternative exons in embryonic tissue.

Novel long noncoding transcripts

We extended our analysis to include the class of long noncoding RNAs (lncRNAs). Out of 2240 lncRNA transcripts annotated in ENSEMBL, 1378 (61.5%) show significant expression in at least one of the six tissues, and 225 are differentially expressed (using cuffdiff; FDR 5%; Fig. 4B and supplementary material Table S5).

Based on comparison to the ENSEMBL (v.69) reference annotation, we additionally identified 4196 novel intergenic transcripts that could be condensed to 2438 loci with a minimum length of 200 bases. We characterized this heterogeneous set of transcripts in several steps. We identified 1143 (27.2%) transcripts with significant similarity to a protein using BLASTx against the NCBI non-redundant protein database (e-value threshold: e-30) (Fig. 4A). From the remaining 3053 transcripts we identified 449 (10.7%) as putative miRNAs in miRBASE (Griffiths-Jones, 2010). A comparison against RFAM revealed 474 (11.3%) transcripts with similarity to other types of small noncoding RNAs. Thus, excluding known classes of coding and noncoding RNA, the remaining 2130 (50.7%) transcripts represent high-quality putative novel lncRNAs. They comprise 2082 multi-exon transcripts belonging to 1403 expressed gene loci. Among these, 439 are differentially expressed between the six embryonic tissues (Fig. 4B and supplementary material Table S5). Thus, through this analysis we

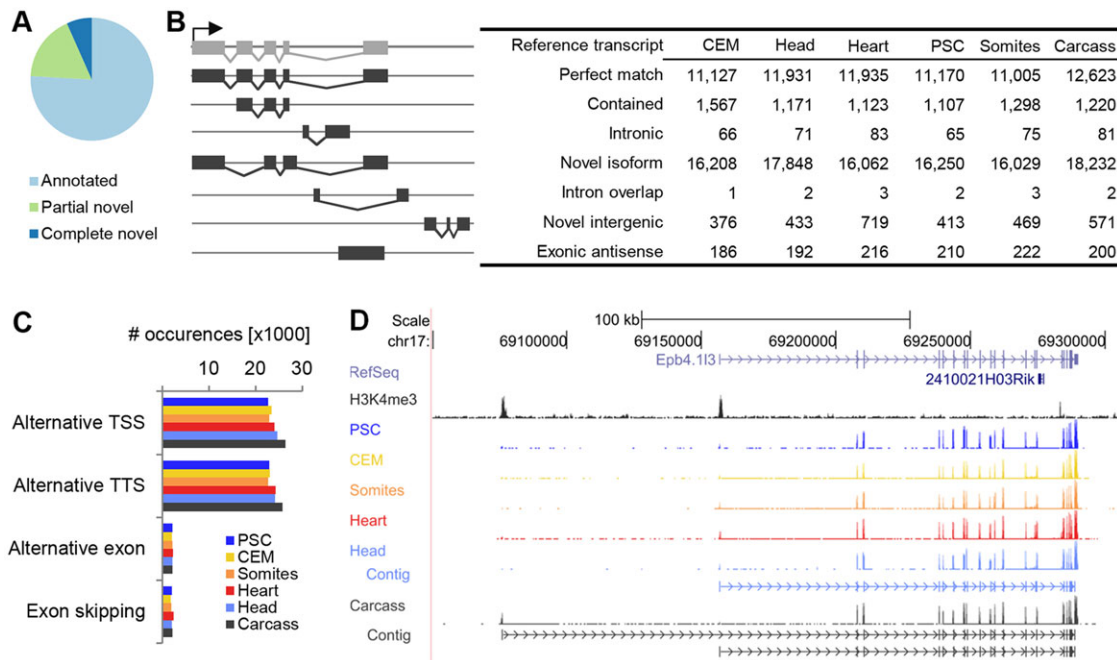


Fig. 3. RNA-seq data reveal extensive alternative transcript usage in mid-gestational embryos. (A) Distribution of previously annotated and novel splice junctions averaged over all six tissues analyzed. Approximately 25% of ~170,000 splice junctions are partially or completely novel. (B) Transcript categories and occurrences in each tissue. (C) Occurrences of alternative transcriptional start sites (TSS), alternative transcriptional termination sites (TTS), alternative exons and exon skipping in the six tissues analyzed (see <http://dx.doi.org/10.6084/m9.figshare.993872>). (D) Example (*Epb4.113*) of a transcript with tissue-specific use of an alternative TSS, marked by H3K4me3 enrichment (y -axis of RNA-seq track is 1000 \times coverage).

substantially extend the current annotation of lncRNAs in the mouse and provide the first set of differentially expressed lncRNAs for the mouse embryo.

Divergent coding-noncoding gene pairs (CNPs)

Visual inspection of putative novel lncRNAs in our transcriptome dataset suggested that many are transcribed in close proximity to protein-coding genes. Similar observations have already been described for mouse ES cells and human datasets (Khalil et al., 2009; Sigova et al., 2013). We identified 392 divergent coding-noncoding gene pairs (supplementary material Table S6). The majority are very close to each other, within less than 200 bp (Fig. 4C).

Divergent CNPs are likely to share the same promoter. We compiled the expression patterns for each gene pair across all samples and calculated the Spearman rank correlation coefficient as a measure of similarity. Using $r \geq 0.5$ as a cut-off, we identified a set of 150 CNPs, which show significantly correlated expression across tissue samples (Fig. 4D).

Alternative TSS and novel transcripts supported by H3K4me3 marks

Histone modifications reflect chromatin states and transcriptional activity of genomic regions. In particular, the tri-methylation of lysine 4 on histone H3 (H3K4me3) has been shown to reliably mark the transcriptional start sites of active and poised genes (Guenther et al., 2007). We reasoned that this mark would also provide support for low expression and novel transcripts, alternative first exons and novel long noncoding RNAs. Therefore, we determined the H3K4me3 histone modification marks on a genome-wide scale in whole TS12 embryos using ChIP-seq.

We mapped 13.3 million reads using bowtie and determined the regions significantly enriched for H3K4me3 using MACS (Zhang

et al., 2008). We detected 29,262 such regions ranging in size between 50 bp (lower detection threshold) and 3233 bp (mean=1004 bp; P -value cut-off 1.00E-05). Of these, 14,686 (50.2%) colocalized with TSS annotated in ENSEMBL (v.69), confirming that they reliably predict TSS on a genome-wide scale (supplementary material Table S7; data not shown). We found 1241 regions marking a distant (median=6.3 kb) novel TSS. An example is shown on Fig. 3D. Another 6442 H3K4me3 regions map inside gene bodies. Of these, 1703 mark putative enhancers identified in multiple mouse cell lines and tissues (Shen et al., 2012). We found 6893 H3K4me3 enrichment peaks outside of transcribed regions. Of those, 1187 overlapped with enhancers and 5706 regions could not be related to TSS, gene bodies or enhancers.

The H3K4me3 marks support the localization of the TSS of 888 putative novel lncRNAs, as illustrated by an example on Fig. 4E.

Our datasets significantly extend the number of lncRNA loci expressed in the mouse embryo and, in particular, the number of lncRNAs showing tissue-specific or differential expression. These data also aid in the identification of mouse orthologues of human lncRNAs involved in disease.

MATERIALS AND METHODS

Embryo preparation, RNA isolation and *in situ* hybridization

Mouse embryos of the C57BL/6J strain were dissected at E8.25 from their mothers and staged according to Theiler (Theiler, 1989). Embryos at TS12 (3- to 6-somite stage) were treated with Dispase (BD Biosciences) for 5 min at room temperature to facilitate the separation of the five tissues. Microdissection was performed in PBS using a flame-sharpened tungsten needle under a stereomicroscope. Tissue samples were pooled from ten embryos, which were lysed (RLT +2-mercaptoethanol, Qiagen), and RNA was isolated using the RNeasy Micro Kit (Qiagen) according to the manufacturer's protocol. The protocol for the whole-mount *in situ* hybridization and the probe sequence can be found on the MAMEP website (<http://mamep.molgen.mpg.de/protocol.html>).

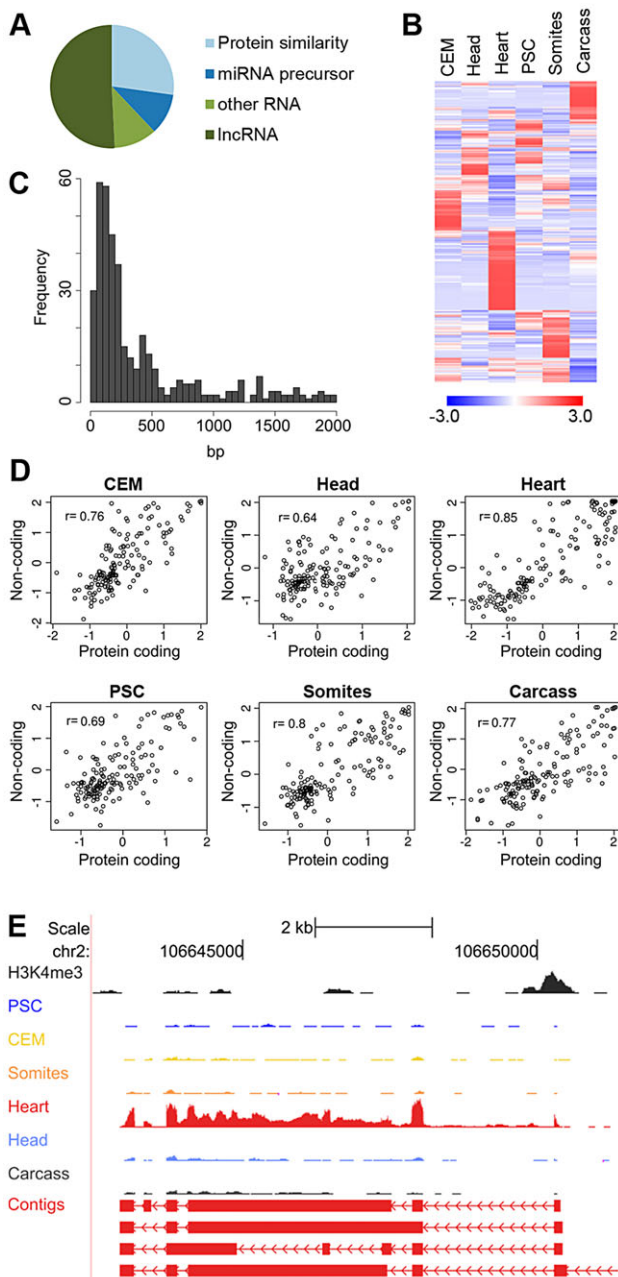


Fig. 4. Abundance and differential expression of lncRNA genes in mid-gestational mouse embryos. (A) Pie chart of novel noncoding transcripts detected by RNA-seq analysis. (B) Heat map representation of 664 differentially expressed lncRNAs. Differential expression was determined by pair-wise comparison (values in the colour bar are FPKM median normalized at the gene level). (C) Distance between the TSS of divergent coding/noncoding gene pairs (CNPs). (D) Scatter plot showing divergent significant correlation of expression of 150 CNPs in the six tissues. The Pearson correlation coefficient is given for each sample. (E) Representative example of a heart-specific lncRNA transcript with four splice variants. The H3K4me3 enrichment indicates the transcriptional start site (y -axis of RNA-seq track is 100 \times coverage).

H3K4me3 analysis

Whole embryos were stage matched (see above) and ten embryos pooled for H3K4me3 (Abcam, ab8580) ChIP, which was performed as described previously (Grote et al., 2013). Libraries were prepared according to the ChIP-Seq DNA Sample Prep Kit (Illumina). 13.3 million 36-base single reads were mapped with bowtie against the mouse mm10 (GRCm38)

reference assembly. Significantly enriched regions for H3K4me3 were determined with Model-based analysis of ChIP-seq (MACS) (Zhang et al., 2008). Overlaps with TSS and putative enhancer regions were calculated using BEDTools (Quinlan and Hall, 2010).

RNA-seq library preparation and sequencing

Sequencing libraries were prepared from total RNA (300 ng) according to the TruSeq RNA Sample Prep (Ver. A) (Illumina) library protocol and analyzed on a HiSeq 2000 genome analyzer with 100 base paired-end read setting. The paired-end libraries had an average fragment size of 310 bp; the paired-end sequences were 100 and 98 bases in length.

Transcriptome assembly and differential expression analysis with RNA-seq

The reads were mapped against the mouse mm10 (GRCm38) reference assembly using two different alignment programs, Tophat (with option $-b2$ -very-sensitive) (Trapnell et al., 2009) and STAR (Dobin et al., 2013), using ENSEMBL reference annotation for guided splice junction mapping. To exploit advantages of reference guided assembly without compromising the detection of novel transcripts, the mapped reads were assembled into transcripts using cufflinks with and without the RABT-guided assembly option (Trapnell et al., 2009). The resulting tissue-specific transcriptomes were merged using cuffcompare and custom PERL scripts re-assigning the gene identifiers. For each transcriptome the expression strength of individual transcripts and genes, and differential expression between tissues was estimated with cuffdiff.

Quantification of housekeeping genes by real-time PCR

For the biological replicates, five embryos were dissected and RNA isolated as described above. The following genes were used as housekeeping genes: *Prmt1*, *Srm*, *Csnk2b*, *Ube2m*, *Cd81*, *Atp6v0c*, *Rab1*, *Pgd*, *Rtm4*, *Erp29*, *Psm7*, *Gpx4*, *Lypla2*, *Rad9*, *B2m*, *Add1*, *Extl3*, *Usp11*, *Cd3eap*, *1810031K17Rik* (*Cndp1* – Mouse Genome Informatics), *Kif1c* and *Mxra8*. Predesigned TaqMan probe and primer sets were purchased (Life Technologies) and quantitative PCR was carried out on a StepOnePlus system (Life Technologies), as suggested by the manufacturer.

Accession number

The accession number at the ENA for the RNA-seq and the ChIP-seq data is PRJEB4513.

Acknowledgements

We thank Tracie Pennimpe and Arica Beisaw for comments on the manuscript. We are grateful to Sonia Paturej for help with preparing the sequencing libraries.

Competing interests

The authors declare no competing financial interests.

Author contributions

P.G., M.W. and L.W. conceived the project, L.W. and P.G. carried out experiments, B.T. supervised the sequencing, and M.W. performed all bioinformatics analyses. B.G.H., P.G. and M.W. designed the figures and wrote the manuscript.

Funding

This work was supported by the Max-Planck Society.

Supplementary material

Supplementary material available online at <http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.105858/-DC1>

References

- Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D. G. et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775-1789.
- Diez-Roux, G., Banfi, S., Sultan, M., Geffers, L., Anand, S., Rozado, D., Magen, A., Canidio, E., Pagani, M., Peluso, I. et al. (2011). A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biol.* **9**, e1000582.

- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21.
- Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., Epstein, C. B., Fritze, S., Harrow, J., Kaul, R. et al. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74.
- Eichler, G. S., Huang, S. and Ingber, D. E. (2003). Gene Expression Dynamics Inspector (GEDi): for integrative analysis of expression profiles. *Bioinformatics* **19**, 2321-2322.
- ENCODE Project and Consortium (2011). A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* **9**, e1001046.
- Florea, L., Song, L. and Salzberg, S. L. (2013). Thousands of exon skipping events differentiate among splicing patterns in sixteen human tissues. *F1000Res.* **2**, 188.
- Geffers, L., Herrmann, B. and Eichele, G. (2012). Web-based digital gene expression atlases for the mouse. *Mamm. Genome* **23**, 525-538.
- Gray, P. A., Fu, H., Luo, P., Zhao, Q., Yu, J., Ferrari, A., Tenzen, T., Yuk, D.-i., Tsung, E. F., Cai, Z. et al. (2004). Mouse brain organization revealed through direct genome-scale TF expression analysis. *Science* **306**, 2255-2257.
- Griffiths-Jones, S. (2010). miRBase: microRNA sequences and annotation. *Curr. Protoc. Bioinformatics* Chapter 12, Unit 12.9.1-10.
- Grote, P., Wittler, L., Hendrix, D., Koch, F., Währisch, S., Beisaw, A., Macura, K., Bläss, G., Kellis, M., Werber, M. et al. (2013). The tissue-specific lincRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Dev. Cell* **24**, 206-214.
- Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. and Young, R. A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77-88.
- Huang, D. W., Sherman, B. T. and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44-57.
- Khalil, A. M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B. E., van Oudenaarden, A. et al. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 11667-11672.
- Li, L., Liu, B., Wapinski, O. L., Tsai, M.-C., Qu, K., Zhang, J., Carlson, J. C., Lin, M., Fang, F., Gupta, R. A. et al. (2013). Targeted disruption of hotair leads to homeotic transformation and gene derepression. *Cell Rep.* **5**, 3-12.
- Mitiku, N. and Baker, J. C. (2007). Genomic analysis of gastrulation and organogenesis in the mouse. *Dev. Cell* **13**, 897-907.
- Pauli, A., Rinn, J. L. and Schier, A. F. (2011). Non-coding RNAs as regulators of embryogenesis. *Nat. Rev. Genet.* **12**, 136-149.
- Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842.
- Rinn, J. L. and Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145-166.
- Sauvageau, M., Goff, L. A., Lodato, S., Bonev, B., Groff, A. F., Gerhardinger, C., Sanchez-Gomez, D. B., Hacisuleyman, E., Li, E., Spence, M. et al. (2013). Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *Elife* **2**, e01749.
- Shen, Y., Yue, F., McCleary, D. F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanov, V. V. et al. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**, 116-120.
- Sigova, A. A., Mullen, A. C., Molinie, B., Gupta, S., Orlando, D. A., Guenther, M. G., Almada, A. E., Lin, C., Sharp, P. A., Giallourakis, C. C. et al. (2013). Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 2876-2881.
- Sturn, A., Quackenbush, J. and Trajanoski, Z. (2002). Genesis: cluster analysis of microarray data. *Bioinformatics* **18**, 207-208.
- Theiler, K. (1989). *The House Mouse: Atlas of Embryonic Development*. Heidelberg, Germany: Springer-Verlag. http://www.emouseatlas.org/Atlas/Theiler_book_download.html
- Trapnell, C., Pachter, L. and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111.
- Ulitsky, I. and Bartel, D. P. (2013). lincRNAs: genomics, evolution, and mechanisms. *Cell* **154**, 26-46.
- Zhang, Y., Liu, T., Meyer, C. A., Eeckhoutte, J., Johnson, D. S., Bernstein, B. E., Nussbaum, C., Myers, R. M., Brown, M., Li, W. et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137.

Table S1, related to Figure 1, Overall mapping statistics of the RNA-seq runs. Mapping statistics of the paired-end runs are derived from the flags of the bam alignment files: total mapped reads (Mapped); single read mappings (Read1/Read2); reads in mate mapped and in calculated distance and orientation (Proper paired); reads mapped and mate read mapped (Mate mapped); reads without mate mapped (Singleton); reads with mate on a different chromosome (Mate different chr).

Table S2, related to Figure 1, Sequencing coverage of annotated features in the genome. Read counts and read densities for different elements of the gene structure (coding exonic, UTR, intronic) and regions upstream and downstream (1kb, 5kb, 10kb) of the transcription unit in the public reference annotation.

Table S3, related to Figure 2, Differentially expressed genes in the midgestational embryo. Differentially expressed ($p \leq 0.05$) known (PRA) genes extracted from cuffdiff analysis results. Expression values are FPKM.

Table S4, related to Figure 2, Tissue-specific gene sets. Significantly differentially expressed genes were mean-normalized and grouped according to their tissue-specific expression. Tissue-specific gene sets are given either per tissue or per tissue combinations.

Table S5, related to Figure 4, Differentially expressed noncoding genes. Significantly expressed ($p \leq 0.05$) known (ENSEMBL) or novel (this study) noncoding genes extracted from cuffdiff analysis results.

Table S6, related to Figure 4, Differentially expressed divergent CNPs. Expression of protein coding (ENSEMBL) and divergent noncoding (ENSEMBL and novel) genes in the different tissues was analyzed using bedtools.

Table S7, related to Figures 3 and 4, H3K4me3 enriched regions in midgestational embryos. H3K4me3 enriched regions derived from TS12 (E8.25) mouse embryos were compiled by MACS.

Table, related to Figure 3, Differential expression of transcript isoforms in the midgestational embryo. Differential isoform transcripts in the different tissue samples were compiled using asprofile (<http://ccb.jhu.edu/software/ASprofile>, Prof. Liliana Florea) from the cufflinks results files in comparison to the public annotation. Categories extracted by the tool include exon skipping (SKIP), multiple exon skipping (MSKIP), alternative transcript start and termination sites (TSS, TTS), retention of single or multiple introns (IR, MIR), and alternative exon (AE). The same categories are also recorded with fuzziness at the feature boundaries; the names are then prefixed with an "X". The table can be downloaded from Figshare (<http://dx.doi.org/10.6084/m9.figshare.993872>) or http://overview.molgen.mpg.de/Table_alternative_exon_usage.zip.

Table S1

Sample	Mapped	Read1	Read2	Proper paired	Mate mapped	Singleton	Mate different chr
CEM	142,159,457	69,871,804	72,287,653	107,020,748	134,005,634	8,153,823	7,025,108
Head	162,320,439	79,228,006	83,092,433	127,363,510	150,852,084	11,468,355	7,083,862
Heart	167,229,639	82,291,982	84,937,657	130,908,150	157,818,756	9,410,883	5,786,474
PSC	154,056,589	75,381,539	78,675,050	122,647,528	144,868,144	9,188,445	4,411,416
Somites	173,254,259	84,519,563	88,734,696	135,307,460	161,931,564	11,322,695	7,198,742
Carcass	179,339,567	87,095,642	92,243,925	142,386,270	166,104,242	13,235,325	6,284,550
All	978,359,950	478,388,536	499,971,414	765,633,666	915,580,424	62,779,526	37,790,152

Table S2

	Total bases	CEM		Head		Heart	
		Read count	Reads/Kb	Read count	Reads/Kb	Read count	Reads/Kb
CDS Exon	36,033,768	115,905,053	3,216.60	138,992,771	3,857.29	147,681,172	4,098.41
5'UTR Exon	36,213,236	20,268,384	559.70	22,753,680	628.32	23,616,193	652.14
3'UTR Exon	15,971,145	15,078,743	944.12	17,032,905	1,066.48	16,600,471	1,039.40
Intronic region	946,361,203	13,054,996	13.79	12,810,538	13.54	12,371,857	13.07
TSS up 1kb	26,459,775	210,573	7.96	260,369	9.84	256,323	9.69
TSS up 5kb	120,275,748	256,807	2.14	217,425	1.81	172,907	1.44
TSS up 10kb	217,425,846	114,841	0.53	109,033	0.50	97,194	0.45
TES down 1kb	27,013,797	980,156	36.28	1,103,578	40.85	1,075,887	39.83
TES down 5kb	117,406,956	1,036,571	8.83	1,189,497	10.13	1,495,730	12.74
TES down 10kb	208,704,706	214,228	1.03	261,768	1.25	1,095,722	5.25

	Total bases	PSC		Somites		Carcass	
		Read count	Reads/Kb	Read count	Reads/Kb	Read count	Reads/Kb
CDS Exon	36,033,768	132,516,431	3,677.60	144,712,731	4,016.03	161,398,349	4,479.09
5'UTR Exon	36,213,236	22,145,730	611.54	24,572,367	678.55	26,240,778	724.62
3'UTR Exon	15,971,145	16,658,659	1,043.10	18,506,501	1,158.75	19,034,672	1,191.82
Intronic region	946,361,203	12,613,069	13.33	15,812,699	16.71	10,069,484	10.64
TSS up 1kb	26,459,775	248,279	9.38	268,802	10.16	271,792	10.27
TSS up 5kb	120,275,748	239,859	1.99	252,709	2.10	177,166	1.47
TSS up 10kb	217,425,846	115,689	0.53	124,340	0.57	95,638	0.44
TES down 1kb	27,013,797	1,110,929	41.12	1,201,520	44.48	1,091,599	40.41
TES down 5kb	117,406,956	1,196,605	10.19	1,289,533	10.98	1,097,091	9.34
TES down 10kb	208,704,706	260,135	1.25	285,500	1.37	227,224	1.09

[Download Table S3](#)[Download Table S4](#)[Download Table S5](#)[Download Table S6](#)[Download Table S7](#)