**Epigenetic regulation in early mammalian development**

Starting from fertilization and ending with implantation, preimplantation embryo development can be divided into several well-orchestrated stages: fertilization, cell cleavage, morula and blastocyst formation. During these stages, maternal and zygotic epigenetic factors play crucial roles. The gene expression profile is changed dramatically, chromatin is modified and core histone elements undergo significant changes. Each preimplantation embryo stage has its own characteristic epigenetic profile, consistent with the acquisition of the capacity to support development. Moreover, histone modifications such as methylation and acetylation as well as other epigenetic events can act as regulatory switches of gene transcription. Because the epigenetic profile is largely related to differentiation, epigenetic dysfunction can give rise to developmental abnormalities. Thus, epigenetic profiling of the embryo is of pivotal importance clinically. Furthermore, genome-wide association studies (GWAS) consistently show that a substantial proportion of genetic risk for common disease falls within regions exhibiting chromatin accessibility in disease-relevant tissues or cell types. Given the importance of these aspects, this platform eMED (Fig. 1) mainly focuses on the epigenetic profiles of transcription factors and chromatin states during preimplantation embryo development, as well as interactions between epigenetic and genetic regulation in these early developmental stages. Understanding the stages of preimplantation development and the underlying regulatory molecular mechanisms is of pivotal importance for basic reproductive biology and for practical applications including regenerative medicine.

Over a million DNA regulatory elements have been cataloged in the mammalian genome, but it remains challenging for linking these elements to the genes that they regulate in the process of early embryo development. We developed eMED, a statistical platform that connects regulatory elements and chromatin state to target genes using big data of high-throughput sequencing and state-of-the-art methods of deep learning. We curated 1,768 high-throughput sequencing datasets for mammalian embryonic development from Gene Expression Omnibus (GEO). These datasets include data from RNA-seq/miRNA-seq, ChIP-seq, ATAC-seq, Hi-C and Dnase-seq/FAIRE-seq. Moreover, we employ four deep learning methods to reanalyze the chromatin features based on the peaks called by MACS. These state-of-the-art methods of deep learning include DeepSEA (1), DeepBind (2), deltaSVM (3) and Epigram (4). We apply eMED to investigate how thousands of dynamically accessible elements orchestrate gene regulation in mammalian early embryo development. Overall, we spare no effort to interpret the epigenetic regulation involved in drastic epigenetic remodeling process in oocytes and sperm, and at early stages of embryonic development.

Gene expression changes drastically during the generation of oocytes and sperm in mammals, and halts completely by the time these cells are fully mature. In mice, expression resumes shortly after fertilization, with a minor wave of gene activation (dubbed zygotic genome activation; ZGA). A second, major wave of ZGA occurs at the late two-cell stage, marking deployment of the developmental gene-expression program. Four divisions later, a cell population called the inner cell mass develops. These cells will form the embryo proper, and can be extracted to derive embryonic stem (ES) cells in vitro. The epigenomic state of ES cells has been thoroughly investigated, but that of earlier developmental stages has remained elusive, mostly owing to the minimal amount of material available for study.

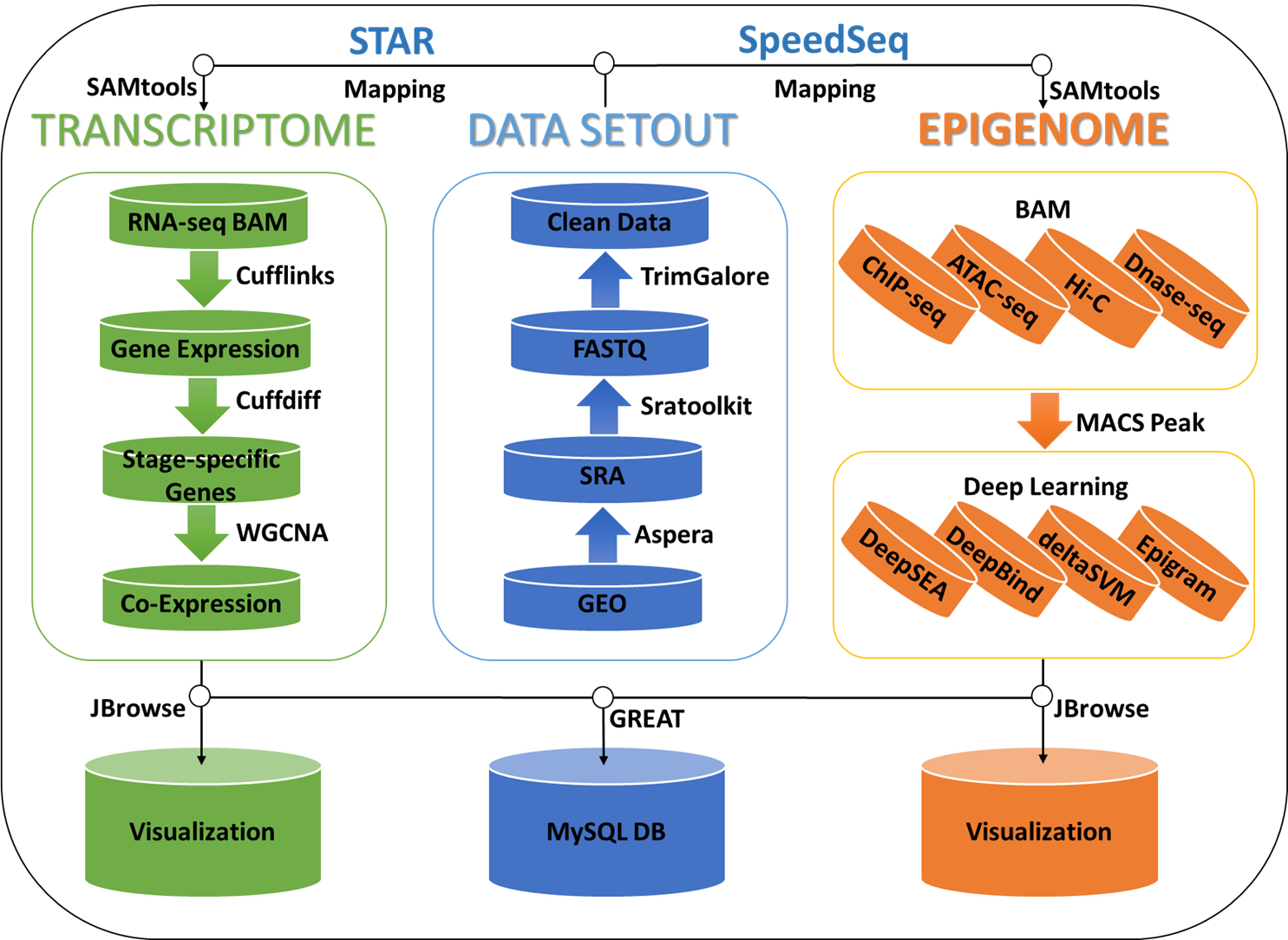


Fig. 1 Data processing workflow of eMED

Gene expression changes drastically during the generation of oocytes and sperm in mammals, and halts completely by the time these cells are fully mature. In mice, expression resumes shortly after fertilization, with a minor wave of gene activation (dubbed zygotic genome activation; ZGA). A second, major wave of ZGA occurs at the late two-cell stage, marking deployment of the developmental gene-expression program. Four divisions later, a cell population called the inner cell mass develops. These cells will form the embryo proper, and can be extracted to derive embryonic stem (ES) cells in vitro. The epigenomic state of ES cells has been thoroughly investigated, but that of earlier developmental stages has remained elusive, mostly owing to the minimal amount of material available for study.

Histone modification reprogramming is more complex than DNA methylation and no general model can be drawn from the accessible experimental data so far. However, similar to DNA methylation, histone modification changes dynamically during preimplantation development in stage- and cell type-specific manners, which are required for the precise regulation of gene expression. As participants in histone modification, the histone acetylases and deacetylases are also involved in chromatin remodeling and help to pave the path for various factors to the DNA.

The four current studies analysed the regions of the genome with which three histone modifications are associated in sperm and oocytes and in early mouse embryos. The authors adapted techniques to allow the analysis of just a few cells. First, Liu et al. (5), Dahl et al. (6) and Zhang et al. (7) studied modification of the amino-acid residue lysine 4 (K4) on histone H3 by three methyl groups (a modification referred to as H3K4me3). Second, Dahl et al. and Wu et al.4 examined modification of lysine 27 (K27) by an acetyl group (H3K27ac). Third, Liu et al. and Wu et al. analysed trimethylation of K27 (H3K27me3). The studies differed in the number of cells analysed and how the DNA and associated proteins (collectively called chromatin) were treated before analysis, but the groups all reached similar conclusions.

In ES cells and mature cell types, H3K4me3 is primarily clustered around small DNA regions at which gene transcription begins, and is associated with gene activity. One of the most striking findings of the current papers is that, in oocytes, H3K4me3 is enriched at low levels across large genomic regions, spanning more than 10 kilobases, and is mostly distant from transcription start sites. This pattern of 'non-canonical' H3K4me3 persists in the fertilized oocyte and in embryos at the early two-cell stage.

Liu et al. found that the number of regions that contain canonical H3K4me3 but not H3K27me3 increased sharply at the late two-cell stage. By contrast, the number of H3K27me3-only regions increased gradually (Fig. 2). This probably reflects different dynamics, and hence different mechanisms, in establishing these two epigenetic marks. H3K4me3 and H3K27me3 are mutually exclusive up to the 16-cell stage, possibly because of the low levels of H3K27me3. By contrast, ES cells contain many domains marked by both such histone modifications. Thus, bivalent domains of modification are established at later stages of development. By having both 'active' and 'repressive' modifications, bivalent domains are thought to be crucial for the efficient expression of lineage-specific developmental programs as cells start to differentiate into mature lineages.

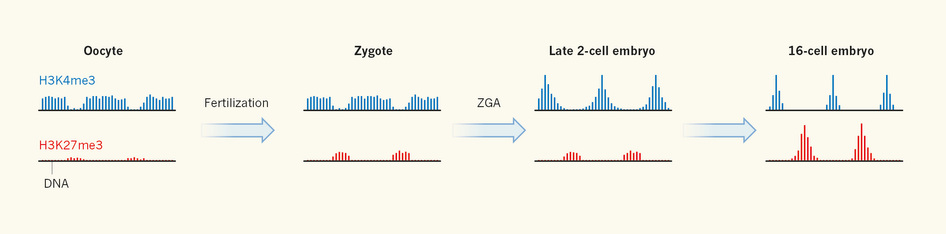


Fig.2 Dynamic landscape of histone modification

Finally, Dahl et al. found stage-specific H3K27ac domains (6), which are presumed to activate the expression of nearby genes. H3K27ac domains tended to be near genes associated with ZGA, and the authors used the domains to identify transcription factors that potentially bind to these nearby genes to regulate early, stage-specific developmental programs. Further work will be required to determine the specific details of the mechanisms by which these transcription factors drive development.

1. Zhou, J. and Troyanskaya, O.G. (2015) Predicting effects of noncoding variants with deep learning-based sequence model. *Nat Methods*, **12**, 931-934.

2. Alipanahi, B., Delong, A., Weirauch, M.T. and Frey, B.J. (2015) Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat Biotechnol*, **33**, 831-838.

3. Lee, D., Gorkin, D.U., Baker, M., Strober, B.J., Asoni, A.L., McCallion, A.S. and Beer, M.A. (2015) A method to predict the impact of regulatory variants from DNA sequence. *Nat Genet*, **47**, 955-961.

4. Whitaker, J.W., Chen, Z. and Wang, W. (2015) Predicting the human epigenome from DNA motifs. *Nat Methods*, **12**, 265-272, 267 p following 272.

5. Liu, X., Wang, C., Liu, W., Li, J., Li, C., Kou, X., Chen, J., Zhao, Y., Gao, H., Wang, H. *et al.* (2016) Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. *Nature*, **537**, 558-562.

6. Dahl, J.A., Jung, I., Aanes, H., Greggains, G.D., Manaf, A., Lerdrup, M., Li, G., Kuan, S., Li, B., Lee, A.Y. *et al.* (2016) Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature*, **537**, 548-552.

7. Zhang, B., Zheng, H., Huang, B., Li, W., Xiang, Y., Peng, X., Ming, J., Wu, X., Zhang, Y., Xu, Q. *et al.* (2016) Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature*, **537**, 553-557.