
DMP du projet "Multi-omics integrative analysis of the early human embryo"

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "Science Europe - DMP template (english)" fourni par Science Europe.

Plan Details

Plan title	DMP du projet "Multi-omics integrative analysis of the early human embryo"
Language	fra
Creation date	2021-03-12
Last modification date	2021-06-14
Identifier	

Project Details

Project title	Multi-omics integrative analysis of the early human embryo
Abstract	An increasing amount of multi-omics data, at the single cell level, allow investigating the early human embryo, from fertilization to day 14 of development, the legal time limit for <i>in vitro</i> culture. However, a comprehensive integrated analysis has not been conducted so far that might help to understand the interconnection of multiple molecular and cellular processes. Here, we propose an integrated approach, notably based on WGCNA, to investigate the correlation between DNA methylation and transcriptome changes during this early time window of human development.

Funding

- INSERM :

Research outputs :

1. single-cell Post-Bisulfite Adaptor-Tagging post-bisulfite DNA methylome sequencing of days 6-12 human embryos (Jeu de données)
2. single-cell RNA-Sequencing of days 6-12 human embryos (Jeu de données)

Contributors

Name	Affiliation	Roles
Gaël Castel - https://orcid.org/0000-0002-7563-1179		<ul style="list-style-type: none"> • Coordinateur du projet • Personne contact pour les données (scRNA-Seq embryo, scPBAT embryo) • Responsable du plan

Droits d'auteur :

Le(s) créateur(s) de ce plan accepte(nt) que tout ou partie de texte de ce plan soit réutilisé et personnalisé si nécessaire pour un autre plan. Vous n'avez pas besoin de citer le(s) créateur(s) en tant que source. L'utilisation de toute partie de texte de ce plan n'implique pas que le(s) créateur(s) soutien(nen)t ou aient une quelconque relation avec votre projet ou votre soumission.

DMP du projet "Multi-omics integrative analysis of the early human embryo"

1. Data description and collection or re-use of existing data

single-cell Post-Bisulfite Adaptor-Tagging post-bisulfite DNA methylome sequencing of days 6-12 human embryos

1a. How will new data be collected or produced and/or how will existing data be re-used?

Data used for this study were published by Zhou *et al.* 2019.

The single-cell post-bisulfite adaptor-tagging DNA methylome sequencing (scPBAT) dataset was collected at https://static-content.springer.com/esm/art%3A10.1038%2Fs41586-019-1500-0/MediaObjects/41586_2019_1500_MOESM3_ESM.zip as a supplementary table named Supplementary Table 10 Promoter_Methylation and selected TF networks of three lineages - XXXXXX(1).xlsx (Sheet 1)

1b. What data (for example the kind, formats, and volumes), will be collected or produced?

The scPBAT dataset contains 24628 gene promoter regions and 130 single cells, common to the scRNA-Seq dataset. These cells originate from day 6 to day 12 human embryos, and comprise epiblast (EPI), primitive endoderm (PE) and trophectoderm (TE) lineages.

Values range from 0 to 1, which means from complete demethylation to complete methylation of a given region. This represents a total amount of 25Mb data.

single-cell RNA-Sequencing of days 6-12 human embryos

1a. How will new data be collected or produced and/or how will existing data be re-used?

The scRNA-Seq dataset was collected from the GEO under accession number GSE109555 as GSE109555_TrioSeq TPM.txt.gz

1b. What data (for example the kind, formats, and volumes), will be collected or produced?

The scRNA-Seq dataset contains 19046 gene transcripts and 2544 single cells. For this study, we focus on the 130 cells common to the scPBAT dataset. These cells originate from day 6 to day 12 human embryos, and comprise epiblast (EPI), primitive endoderm (PE) and trophectoderm (TE) lineages.

The minimal value is 0, without a fixed maximum.
This represents a total amount of 146Mb data.

2. Documentation and data quality

2a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

A sample annotation table of the 130 single-cells is provided, which contains information on the developmental day, cell lineage, sex, and embryo of origin.

A flowchart describing the analysis workflow of this study is provided at
https://github.com/gaelcastel/multi_omics_human_embryo.git

2b. What data quality control measures will be used?

Library size control and histogram of count distribution are provided.

As input data have been pre-processed, we apply few additional normalization steps, consisting of filtering features based on variance and log-transforming the data.

3. Storage and backup during the research process

3a. How will data and metadata be stored and backed up during the research?

During this study, both datasets are stored as count matrices on the IFB cluster, at /shared/projects/dubii2021/gcastel/dubii_stage/.

These files are backed up locally on laboratory computer at C:\Users\E137833T\Desktop\dubii\stage\.

3b. How will data security and protection of sensitive data be taken care during the research?

All data used in this study are publicly available, so these do not require security and protection managing measures.

4. Legal and ethical requirements, codes of conduct

4a. If personal data are processed, how will compliance with legislation on personal data and on security be ensured?

All data used in this study have been collected according to patients' consent for research use of personal data and registered as described in Zhou *et al*, 2019:

"Research donors were recruited from Peking University Third Hospital. Before giving consent, donors have a suitable opportunity to receive proper counselling about the implications of the donation and potential risks. Gametes and embryos were collected with written informed consent from the donors in this study".

As the data have been anonymized and are publicly available, no additional security managing measures are required.

4b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

As the data used in this study originate from Zhou *et al*, 2019, the original publication is cited accordingly.
The code generated in this study is publicly available at https://github.com/gaelcastel/multi_omics_human_embryo.git

4c. What ethical issues and codes of conduct are there, and how will they be taken into account?

Codes of conduct include citing the original study of Zhou *et al*, 2019 when using their data.

5. Data sharing and long-term preservation

5a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

Input data matrices used in this study are publicly available. However, the access to FASTQ files is subjected to Genome Sequence Archive (GSA) authorization delivered on request.
As every request will be examined by a Chinese committee, this access might be strongly delayed (several weeks).
Regarding the code generated in this study, it is publicly available at
https://github.com/gaelcastel/multi_omics_human_embryo.git

5b. How will data for preservation be selected, and where data will be preserved long-term (for example a data repository or archive)?

The scRNA-Seq dataset is stored at GEO database (Gene Expression Omnibus, NCBI, USA), with the following accession number: GSE109555, and is publicly available.
The scPBAT dataset is stored as Supplemental Information on the Nature journal database at
<https://doi.org/10.1038/s41586-019-1500-0>
FASTQ files are stored at GSA (Genome Sequence Archive, China), and access is subjected to authorization upon request.

5c. What methods or software tools are needed to access and use data?

Conventional FTP or HTTPS protocols are sufficient to access data.
R is used to process and analyse the data.
Also, an access to a multi-core cluster is required to repeat some analyses described in this study.

5d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

The original study from Zhou *et al*, is registered under DOI <https://doi.org/10.1038/s41586-019-1500-0>

6. Data management responsibilities and resources

6a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

Laurent David, Principal Investigator at University of Nantes, is responsible for these data management.

6b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

The dedicated github repository of this study will be regularly maintained. Authors will answer e-mail queries from readers.