|  |
| --- |
| **0. Proposal name** |
| Transcription factor control of dynamic transitions within and beyond pluripotency |
| **1. Description of the data** |
| **1.1 Type of study**  Gene expression profiling data from RNA-seq experiments.  ChIP-seq epigenomic profiling data.  Analysis of protein expression in embryos, stem cell and primordial germ cell cultures.  **1.2 Types of data**  Quantitative gene expression and epigenomic data will be generated from cell lines, as will fluorescent images of protein levels (stored as .tifs), FACS plots (stored as FloJo files) and genotyping data (stored in lab books).  **1.3 Format and scale of the data**  Data will be stored in the appropriate format for the type of data, FASTQ format files for sequence data files (ChIP-seq, RNA-seq), tifs for image files and FloJo for FACS data. Data will typically be processed using tools from the Bioconductor suite for sequence data. The formats are accepted as standard by GEO database and the SRA/ENA sequence archives. This provides for both long term data sharing and long-term validity of the data. We estimate generating ~1Tb data per year. |
| **2. Data collection / generation** |
| **2.1 Methodologies for data collection / generation**  RNA will be prepared from embryos and from cell lines treated in precise ways to alter the level of a test inducer for a specified time period. Chromatin will be prepared using standard methodologies for ChIP-Seq analysis. Flourescent images and FACS data will be generated using antibodies that we validate as specific in house.  **2.2 Data quality and standards**  Data will be generated with three biological replicates for RNA-seq and gene expression and two replicates for chIP-seq. Single cell RNA-seq data will be generated from populations of >100 cells. |
| **3. Data management, documentation and curation** |
| **3.1 Managing, storing and curating data.**  Generated data will be stored using UoE DataStore file system. This is a high quality, enterprise-class storage with guaranteed backup and resilience. The intention is that data will be published in high impact journals and then data will be submitted to the appropriate public databases (SRA/ENA or GEO) with sample data and annotations as recommended by the public repositories.  **3.2 Metadata standards and data documentation**  Data once published will be available in the well-established SRA/ENA or GEO repositories using the data standards recommended by the repository provider. Prior to publication, data will be made available to a restricted set of collaborators from within UoE DataStore  **3.3 Data preservation strategy and standards**  We aim to publish the data in a timely manner and so make it available from within public repositories. Long-term data preservation for 10 years after the end of the grant will be achieved by the cost-effective UoE DataVault service.Imaging and FACS data will be stored on local servers. |
| **4. Data security and confidentiality of potentially disclosive information** |
| **4.1 Formal information/data security standards**  Not applicable.  **4.2 Main risks to data security**  Not applicable. |
| **5. Data sharing and access** |
| Identify any data repository (-ies) that are, or will be, entrusted with storing, curating and/or sharing data from your study, where they exist for particular disciplinary domains or data types. [Information on repositories is available here.](http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Guidance-for-researchers/WTX060360.htm)  European Nucleotide Archive <http://www.ebi.ac.uk/ena/>  Potentially also  Gene Expresion Omnibus <http://www.ncbi.nlm.nih.gov/geo/>  Sequence Read Archive <http://www.ncbi.nlm.nih.gov/sra>  **5.1 Suitability for sharing**  Yes; other researchers may wish to reanalyse data.  **5.2 Discovery by potential users of the research data**  Research presentations at conferences and publications aimed at open access journals. For public dissemination of our data and results we will use suitable public repositories (for example GEO) or when not available (for example for imaging data) we will use the University’s DataShare service, which provides various search options and stable public identifiers such as DOI for dataset citation. The identifiers from public repositories will be added to our manuscripts, which we will submit to journals that support open-access publication.  **5.3 Governance of access**  IC will make the decision as to whether to supply data to a potential new user.  **5.4 The study team’s exclusive use of the data**  Data will be exclusive until the first publication derived from the data.  **5.5 Restrictions or delays to sharing, with planned actions to limit such restrictions**  Some preliminary data will be presented prior to publication in order to seek further funding. We plan to share all of our research outcome within 3 years of the project. Our research outcomes will be available under CC-BY licence terms.  **5.6 Regulation of responsibilities of users**  Not applicable. |