4th Bioinformatics Stem Cells Satellite Workshop,

Stem Cell Network NRW - 6th International Meeting
Essen, April 7th, 2011

Sponsored by the DFG Priority Program SPP 1356 “Pluripotency and Cellular Reprogramming", by the Institute for Transfusion Medicine of the University Hospital Essen, by the Stem Cell Network NRW and by the Institute for Biostatistics and Informatics in Medicine and Ageing Research Rostock.

Program

9:20	Opening
9:30-10:15	Invited Lecture, Simon Tomlinson, Edinburgh, UK Mapping the molecular components of stem cell function using large-scale integrated analysis
10:15-10:45	Marcos Arauzo, Münster SILAC proteomics for disclosing reprogramming markers
10:45-11:05	Volkhard Helms, Saarbrücken Stochastic switching during cell differentiation/reprogramming (planned work)
Coffee	
11:30-12:30	Invited Lecture, Franz-Josef Müller / Bernhard Schuldt, Kiel / Aachen Pluripotent stem cells: one network to find them, one network to rule them?
12:30-13:00	Miguel Andrade, Berlin Searching for shortcuts on the road of induction of pluripotent stem cells from human fibroblasts
Lunch	
14:00-14:30	Andreas Kurtz, Berlin Cellfinder -- a stem cell data repository under the Open Source model
14:30-14:50	Arnoldo Muller-Molina, Münster Similarity Search of Gene Expression Micro Arrays
14:50-15:10	Lena Scheubert, Osnabrück/Rostock Learning biomarkers of pluripotent stem cells in mouse
15:10-15:30	Georg Fuellen, Rostock Evolution of Pluripotency: What's there to find? Are there rules?

Searching for shortcuts on the road of induction of pluripotent stem cells from human fibroblasts

Miguel A. Andrade-Navarro1 Nancy Mah1, Ying Wang2, Martin Schaefer1, and James Adjaye2
1Max Delbrück Center for Molecular Medicine, Berlin, Germany
2Max Planck Institute for Molecular Genetics, Berlin, Germany

Induction of pluripotent stem cells (iPS) from differentiated cells is a promising alternative to human embryonic stem cells (hESCs) as a source of stem cells for therapy. iPS can be derived by viral transduction of pluripotency related transcription factors. In particular, reprogramming of human foreskin fibroblast-HFF1 cells has been achieved by induction of Oct4, Sox2, Klf4 and c-Myc (OSKM). However, the time that the reprogramming requires and the low yield of the process suggests that there is room for improvement. Solutions are sought that include inducing alternative factors and supplementing drugs. To propose targets and to better understand the processes involved in reprogramming we produced microarray gene expression data following the first three days of a current protocol of generation of iPS from HFF1 by induction of OSKM. We integrated the analysis of gene expression data with public protein-protein interaction data and chromatin immunoprecipitation experiment results to propose pathways active during early reprogramming. Some of them contain candidates whose induction could shortcut the reprogramming procedure.

SILAC proteomics for disclosing reprogramming markers

Marcos J. Araúzo-Bravo[1], Nishant Singhal[2]
[1]Laboratory of Computational Biology and Bioinformatics, Max Planck Institute for Molecular Biomedicine, Munster, Germany
[2]Department of Cell and Developmental Biology, Max Planck Institute for Molecular Biomedicine, Munster, Germany

From Yamanakas's seminal work it is known that induction of a cocktail of four transcription factors (Oct4, Sox2, Klf4, and c-Myc) is enough to reprogram somatic cells into a pluripotent state. But this is a very low-efficient process. In order to understand better the reprogramming mechanism and to improve its efficiency we designed a high throughput proteomics strategy. We implemented an assay for screening nuclear fractions from extracts of pluripotent mouse cells based on Oct4 reactivation. Then we used SILAC (Stable Isotope Labeling with Amino acids in cell Culture) proteomics to identify proteins that were highly enriched in the Oct4 reactivation fraction with respect to the Oct4 non-reactivation fraction. The list of proteins in the reactivated fraction was sorted from lowest to highest H/L ratio of enrichment, and we picked up the cases (585 proteins) with H/L ratio lower than 0.02. That list was used to perform gene ontology (GO) significance enrichment analysis. We selected as a background dataset the 5111 proteins detected in embryonic stem cells (ESC) using a SILAC approach. We found that the chromatin remodeling terms are at the top rank of the significantly enriched terms (only after the transcriptions factor related terms). Among the highest enriched chromatin remodeling terms are those belonging to the NuRD complex (that is a transcriptional repressor complex) and to the SWI/SNF complex (that is a transcriptional activator ATP-dependent BAF chromatin-remodeling complex). Since we were interested in proteins that could be induced to improve the reprogramming efficiency, we focused our search on the transcriptional activator SWI/SNF complex. Among the top ranked SWI/SNF members with best H/L ratios we found to be Smarca4/Brg1 and Smarcc1/Baf155/Srg3. To validate till which extent these proteins could mediate reprogramming with higher efficiency we added Brg1 and Baf155 to the original Oct4, Sox2, Klf4, and c-Myc reprogramming cocktail of Yamanaka and we checked the reprogramming capabilities of the new cocktail. We found that it resulted in a significantly increased reprogramming efficiency. The reprogrammed cells could transmit to the germline exhibiting pluripotency and the global gene expression profile of the reprogrammed cells was very similar to the ESCs profile. Additionally, reprogramming was retained highly efficient when c-Myc was excluded, allowing us to dispose of the oncogenic features associated with c-Myc. Thus, the gene ontology analysis of high throughput proteomics data has allowed disclosing chromatin-remodeling molecules which increased the reprogramming efficiency of somatic cells.

Evolution of Pluripotency: What's there to find? Are there rules?

Georg Fuellen
Institute for Biostatistics and Informatics in Medicine and Ageing Research, University Rostock, Germany

We will showcase how evolutionary analyses may contribute towards understanding the cellular state of pluripotency. There are some success stories, but the overall impression is that "looking back in time" is a very difficult task. We demonstrate how frustrating evolutionary analyses can be in general, and we make a case that pluripotency has been a "playground of evolution", diminishing our ability to reconstruct past events. We present some in-silico studies on the evolution of gene regulation, and on the evolution of the network of gene/protein interaction and regulation that underlie pluripotency.

Stochastic switching during cell differentiation / reprogramming (planned work)

Volkhard Helms
Center for Bioinformatics, Saarbrücken, Germany

The Helms group studies pairwise biomolecular interactions and small to medium-scale molecular networks. For example, we have generated a stochastic dynamics model for the photophysical and -chemical processes taking place in chromatophore vesicles from Rhodobacter sphaeroides. Using Gillespie-type dynamics simulations, one can simulate the kinetics of the binding and dissociation reactions as well as the redox reactions taking place on a seconds time scale in one of these 50 - 100 nm large vesicles. Using an evolutionary algorithm for varying kinetic parameters and a set of 11 different time-dependent kinetic data probing the behavior of the entire vesicle, we were able to optimize 27 different kinetic and physicochemical

constants so that the in-silicio vesicle reproduces the non-equilibrium behavior of real vesicles when exposed to short light pulses. We now plan to apply this approach to study the stochastic nature of cellular reprogramming behavior. We will concentrate on a core set of reactions such as the Plurinet of Fuellen and co-workers around the magic cocktail of 4 transcription factors Oct4, Klf4, Myc and Sox2.
Motivated by our previous work on genomic imprinting together with Martina Paulsen (Saarland University), we will put an emphasis on accounting for the effects of DNA methylation on regulating gene expression.

Pluripotent stem cells: one network to find them, one network to rule them?

Franz-Josef Mueller1, Bernhard Schuldt2
1Zentrum für Integrative Psychiatrie, Kiel - ZIP gGmbH, Kiel, Germany
2Aachen Institute for Advanced Study in Computational Engineering Science (AICES), RWTH Aachen University, Aachen,Germany

Stem cell biologists are more and more not just studying isolated components of a cell type but rather begin to integrate systems wide interactions into more and more complex models.
Biological networks are at the center stage of these efforts, yet this type of model has been conceptualized and used in several, non-identical ways.

Theoretical approaches were influenced by ideas from electrical engineering, theoretical chemistry, physics, statistics and computer science. These methods are routinely scaled to integrate readouts from hundreds of million independent measurements.
Practical work at the bench mostly relies on the experimenters' knowledge, intuition, small pathways and back of the envelope concepts.
Given the diversity of network reconstruction approaches it is not surprising that there has been a lot of discussion about acceptable, state-of-the-art methods, which should "rule" the field.
Here we give an intuitive overview about the different uses of networks based on our experience in applying network concepts in pluripotent stem cell biology.
We will discuss when to focus on small-scale systems and when it might be advantageous to use on simplified models to stress general network structures.
Finally, we will give an outlook on how different approaches can be integrated into comprehensive mathematical models, which then might come full circle by establishing mental models of pluripotency which are necessary for planning novel and successful experimental interventions and measuring strategies.

Similary Search of Gene Expression Micro Arrays

Arnoldo Muller-Molina
Laboratory of Computational Biology and Bioinformatics, Max Planck Institute for Molecular Biomedicine, Munster, Germany

When searching for similarity of vectors of dimension larger than 20, the curse of dimensionality kicks in and it is virtually impossible to distinguish between objects. This problem is equivalent to what humans experience with farsightedness.
Since micro array data contains tens of thousands of dimensions, a traditional function like the euclidean distance is not effective for categorizing samples. In this work we propose a dimensional subset matching approach that attempts to avoid the curse of dimensionality by focusing in different parts of a vector for different classes of biological experimental conditions.

Learning biomarkers of pluripotent stem cells in mouse

Lena Scheubert
Institute for Computer Science, University Osnabrück, Germany

Pluripotent stem cells are able to self-renew, and to differentiate into all adult cell types. Many studies report data describing these cells, and characterize them in molecular terms. Machine learning yields classifiers that

can accurately identify pluripotent stem cells, but there is a lack of studies yielding minimal sets of best biomarkers (genes/features). We assembled gene expression data of pluripotent stem cells and non-pluripotent cells from mouse. After normalization and filtering, we applied machine learning, classifying samples into pluripotent and non-pluripotent with high cross-validated accuracy. Furthermore, to identify minimal sets of best biomarkers, we used three methods: information gain, random forests, and a wrapper of genetic algorithm and support vector machine (GA/SVM). We demonstrate that the GA/SVM biomarkers work best in combination with each other; pathway and enrichment analyses show that they cover the widest variety of processes implicated in pluripotency. The GA/SVM yields best biomarkers, no matter which classification method is used. The consensus best biomarker based on the three methods is Tet1, implicated in pluripotency just recently. The best biomarker based on the GA/SVM approach alone is Fam134b, possibly a missing link between pluripotency and some standard surface markers of unknown function processed by the Golgi.

CellFinder: A New Stem Cell and Developmental Biology platform and data repository

Stachelscheid H, Damaschun A, Lekschas F, Werner S, Neves M, Leser U, Nguyen-Dobinsky TN, Kurtz A
Charite Universitätsmedizin Berlin, Berlin-Brandenburg Center for Regenerative Therapies, Berlin, Germany

Cellular therapies are increasingly relevant for personalized medicine, requiring resources informing about the complex characteristics of cells. Large amounts of research data on stem cells are already available, but these are scattered, derived by diverse technologies, not standardized and are not available at the necessary integration level for cellome modelling. Consequently, the selection of cells, e.g. for therapeutic applications, is based on rather incomplete information.

CellFinder aims at mapping stem cell information to provide a basis for a global understanding of cells, to increase the comparability of data and to relate cells to more complex systems.

To organize the data stored in CellFinder we propose an ontology for the annotation of data from the organ down to the single cell level and mapping of homologous entities between species. Existing taxonomies such as the gene ontology (GO) or Foundational Model of Anatomy (FMA) are used for data annotation.

Based on this, the CellFinder platform provides the framework for comprehensive descriptions for human tissues, cells on molecular and functional levels, in vivo and in vitro. The descriptions can be attributed with experimental data, images, references to sources of the relevant materials.

Current work concentrates on the integration of existing cell biology datasets to represent molecular, functional, anatomical and cyto-histological levels. CellFinder is available at: http://www.cell-finder.org

How can we learn from large-scale integrated analysis of embryonic stem cell data?

Simon Tomlinson
Stem Cell Bioinformatics, MRC Centre for Regenerative Medicine, Edinburgh, UK

Embryonic stem (ES) cells are cells capable of self-renewal and differentiation into cell types of the three primary germ layers. ES cells are capable of expressing a rich gene expression repertoire in culture- partly due to spontaneous differentiation but also due to culture adaption and even clonal history of the cell line. Differences between lines under identical culture conditions may be partly genetic or epigenetic in origin. But even purified ES cells from a single clonal line are heterogeneous in culture and are capable of switching between multiple meta-stable states each with a distinct gene expression signature. In practice, it is very difficult to design laboratory experiments that adequately control for the apparent biological complexity of the system. As a result of the underlying biological complexity genome-scale ES functional genomic data can be difficult to interpret in detail.

Large-scale integrated analysis is a very attractive, powerful approach that can be used, for example to identify common gene expression signatures across sets of similar profiling experiments. In some way this approach controls for biological complexity and so allows general trends and patterns to be more readily observed. In this talk I will discuss integrated analysis approaches and highlight the remaining challenges presented by these methods.