Role of the Murine Reprogramming Factors in the Induction of Pluripotency

Rupa Sridharan, Jason Tchieu, Mike J. Mason, Robin Yachechko, Edward Kuoy, Steve Horvath, Qing Zhou, and Kathrin Plath

1David Geffen School of Medicine
2Department of Biological Chemistry
3Jonsson Comprehensive Cancer Center
4Molecular Biology Institute
5Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research
6Department of Statistics
7Department of Human Genetics
8Department of Biostatistics

University of California, Los Angeles, CA 90095, USA
9These authors contributed equally to this work

*Correspondence: kplath@mednet.ucla.edu
DOI 10.1615/j.cell.2009.01.001

Bradly Alicea
http://www.msu.edu/~aliceabrar/
Introduction

A) Why is this paper important?

1) Oct4, Sox2, and NANOG co-occupy many target genes (bind to same promoters during switch to pluripotency).

2) Serves to “parameterize” reprogramming (what are statistical signatures of a particular cell morphology, state, and dynamics?).

B) Authors use computational approach (with rodent model):

1) estimate binding probabilities, other quantitative indicators from ChIP-chip, microarray data (see next slide).

2) binding events: product of probabilities (see slide after next).
Prior work/basis for approach

Boyer et al., Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. Cell, 122, 947 (2005):

* sites occupied by Oct4 identified from peaks of ChiP-enriched DNA, 98% of known binding sites for human TFs = within 8kb of target genes.

* platform = false positive rate of < 1% and false negative rate of 20%.

* yeast TFs used to validate method (DNA microarray, cell culture protocols).

* used primary and secondary antibodies – incubated with cells in culture.

* in Sridharan paper, different cell lines created (ES, piPS, iPS) that expressed characteristic markers.
Goal: predicting cell state from relative, absolute TF activity

No studies (before this paper) to date that address:

* analysis of how four factors bind and function in iPS cells.

* combinations of factors fulfill transient roles during reprogramming.

* fail to activate transcriptional regulators.

* pluripotent via chromatin modifications or modulating signal transduction.
Probabilistic model

\[ p(OSK) = p(Oct4) \times p(\text{Sox2}) \times p(\text{Klf4}) \times (1-p(\text{cMyc})) \]

Binding group OSK

Presence of factors at prob(\(TF_n\)), contribute to pluripotency?

cMyc not included in OSK

P(OSK) is the particular binding group for this example. OSK includes Oct4 (O), Sox2 (S), and Klf4 (K). Does not include c-Myc (C).

\( p(binding\ group_n) \) is:

a) binding by a specific number of factors = sum of probabilities for a particular binding group (binding probability does not change with grouping).

\[ \text{* suppose } p(OSK) = p(0.8) \times p(0.85) \times p(0.7) \times p(1-0.65), \text{ then } p(OSK) = 0.167. \]

b) expected number for that condition = probability \( \times \) number of genes on array (1600 genes on microarray \( \times \) p(0.54) = 864).
Definition of Binding Groups

* 1 group where genes are bound by all four factors (OSCK). Includes Oct4, Sox2, c-Myc, and Klf4.

* 4 groups where genes are bound by different combinations of 3 factors (OSC, OSK, OCK, SCK). Example: OSC = Oct4, Sox2, and c-Myc.

* 6 groups where genes are bound by two factors (OS, OC, OK, SC, SK, CK). Example: OS = Oct4, Sox2.

* 4 groups where genes are bound by only one factor (O, S, C, K).

* 15 unique and non-redundant groups total (combinatorial).

* each group has a characteristic efficacy for reprogramming and results in particular cellular morphology (state).
**Measure: Hamming distance**

Hamming distance (borrowed from Information theory). Often used to quantify sequence distances in RNA evolution:

* compare two binary sequences (calculate numerical distance).

* each position in columns below = bit (can be either 0 or 1).

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Example:**

0 1 0 0 1 0
0 0 0 1 0 0 = h = 1

* sum of bottom column = h (or Hamming distance).

* h represents a functional difference between binding sites (each bit ~ binding site).
Example of how $h$ is used

Number of bases mismatched in Oct4, Sox2, c-Myc, and Klf4 binding between:

* ES and iPS cells (ES, iPS).

* ES and partially reprogrammed cells (ES, piPS).

$h = 0$, perfect match, $h = 4$, complete mismatch.

Example: $h = 1$ is a mismatch at Oct4, match at Sox2, c-Myc, and Klf4.

Number of factors differentially bound between iPS and partially reprogrammed cells:

* average for $(h(\text{iPS, ES}) - h(\text{piPS, ES}))$
Measure: $\log_2$ expression ratio

Log$_2$ (or signal log) ratio:
derived from microarray chip analysis.

* change in expression level of a transcript between 1) baseline and 2) experimental arrays.

* log transformed (base 2 – because you are concerned with two states).

Examples:

* if gene A increases two-fold from control to experimental condition (twice as much product), then log$_2$ ratio for gene = 1.

* if gene A decreases two-fold from control to experimental condition (half as much product), then log$_2$ ratio for gene = -1.
Results

a) Depicts genes bound by the same set of factors in ES and iPS cells.

b1) and b2) = genes bound by combinations (different) of the four factors in ES and iPS cells.

* binding strength for each gene may differ (see heat maps) between ES and iPS cells.
Exp column: log2 expression ratio (piPS/ES) for genes with > 2-fold [>1, <1] expression difference.

+binding column: binding in piPS but not in ES ($h$ parameter denotes distance between cell types for each gene).

-binding column: binding in ES but not in piPS cells.

Insets: most differentially expressed genes ($n$-fold$_{max}$).
Results (con’t)

Binding vs. binding strength:

ES = embryonic stem.

iPS = induced pluripotent.

piPS = partially reprogrammed.

* strong binding of O, S, and K in ES and iPS, not so much in piPS.

* O, S, and K bound in ES, less Uniform in iPS (with some C binding), pattern breaks down in piPS cells.
Results (con’t)

Frame a:
log_2 expression ratio (n-fold expression) for cell states across binding groups.

* mostly upregulated for ES-specific binding groups (except for O, SK, S, and K).

Frame c:
Venn diagram of OSK targets in iPS/ES cells (left), piPS cells(right), overlap (center).

* size of circles = frequency of activation for particular gene (e.g. Lefty2).
Results (con’t)

A) # of target genes enriched for H3K27me3 given binding group:
* piPS cells resemble control (MEF), save for OS and O.
* ES cells above control in most cases, differ from piPS in many cases as well.

B) gene expression data for target genes:
* enriched for H3K4me3, H3K27me3 for each cell state (ES, iPS, piPS, MEF).
Results (con’t)

a) FACS plot (2d, 4d – infected vs. uninfected fibroblasts).

b) expression differences between ES and fibroblasts (MEFs). Induced with doxycycline vs. uninduced controls.

c-Myc enhances early steps of reprogramming by:

* repressing fibroblast-specific expression.

* upregulating the metabolic program of embryonic state.
Conclusions

**Post-hoc hypothesis:** As fibroblast gene expression decreases and ES-specific gene expression increases, a stable region exists (on multidimensional profile of expression level data) near intersection of both activity types (e.g. partially reprogrammed state).
Conclusions (con’t)

TF binding:
c-Myc, Oct4, and combination of Sox2/Klf4 (OSC, OSK) pushes toward ES-state.

Later, no need for c-Myc, but need Oct4, Klf4, and Sox2 (OSK).

Methylation:
**Fibroblasts:** H3 methylated at K27 only.

**piPS:** bistable (methylated at both K4 and K27, weakly enriched).

**iPS/ES:** H3 methylated at K4 only (switch-like).