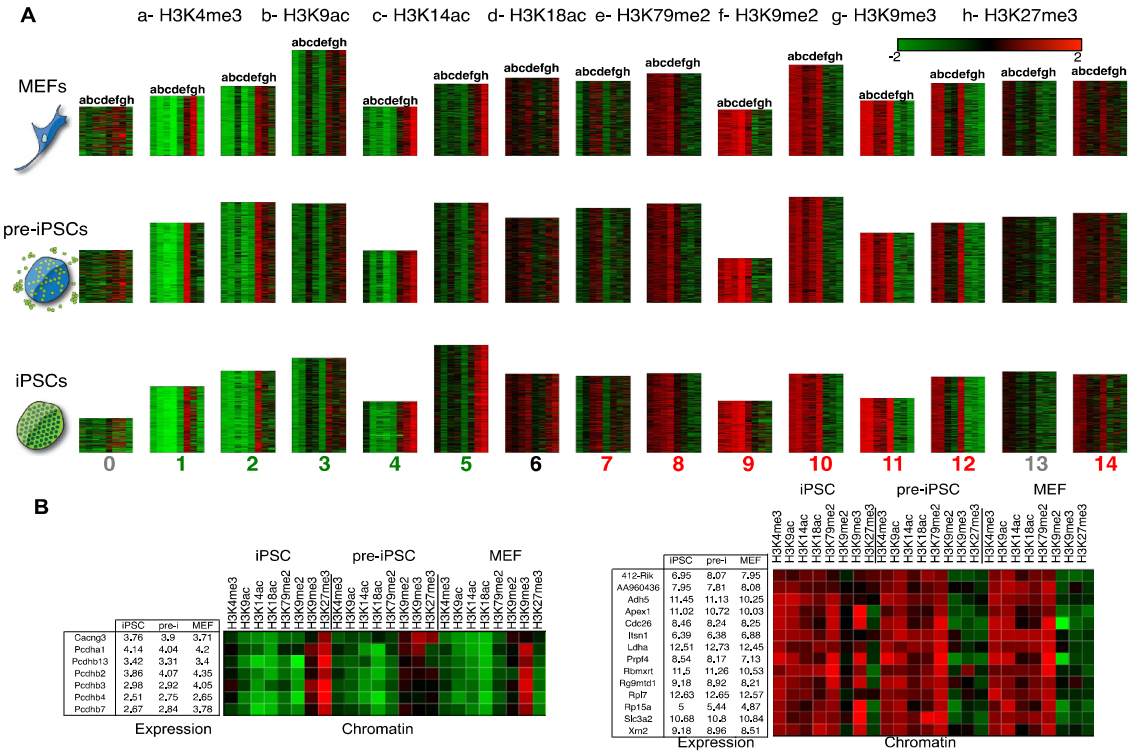
CMINT: Chromatin Module inference on cellular trajectories identifies poised epigenetic states in reprogramming to induced pluripotent stem cells.

Sushmita Roy[1,3], Rupa Sridharan[2,3]
[1.]Dept of. Biostatistics and Medical Informatics, [2.] Dept. of Cell and Regenerative biology,
[3.]Wisconsin Institute for Discovery, University of Wisconsin, Madison WI 53715, USA

During cell fate transitions the chromatin organization of the precursor cell changes from that of the endpoint cell. Current computational approaches to analyze chromatin modifications across multiple cell types do not model how the cell types are related on a lineage or a time course, which makes it difficult to reliably identify transitions. To overcome this limitation, we have developed a method called CMINT (Chromatin Module INference on Trees), a probabilistic clustering approach to systematically capture chromatin state dynamics across multiple cell types. We profiled eight histone modifications in three cell types capturing transition during transcription factor mediated reprogramming: the starting somatic cell (MEF), partial (pre-iPSC) and completely reprogrammed induced pluripotent stem cells (iPSCs). Compared to existing approaches for examining chromatin state across multiple cell types, CMINT captures higher quality clusters and reliably detects chromatin transitions between cell types. Using CMINT we found modules comprising distinct types of gene activation associated histone modifications within which transitions could occur without large gene expression changes. Novel multivalent states, comprising previously unknown



A. Heatmaps of 15 chromatin modules (0-14), obtained from CMINT application to a chromatin mark dataset measuring eight marks (a-f) in three cell types: MEFs, pre-iPSC, iPSC. Height of each heatmap is proportional to the number of genes. Module IDs are color-coded according to enrichment of repressive histone marks in green, activating histone marks in red, both (multivalent) in black. Module 0 and 13 lack specific enrichments and are shown in gray. B. Example gene sets that do not change greatly in expression but change in module membership. Left: gene names and log mRNA in the three cell types. Right: Heat map of histone modifications in each cell type.

combinations of activating and repressive histone modifications were identified: some contained pluripotency genes likely poising them for further activation in the iPSC state while others were enriched for Ctcf binding suggesting a chromatin organizational role for multivalency. Our method provides a systematic approach to model the relationships among cellular states to gain novel insights in gene regulation dynamics in diverse developmental and disease processes.