ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells.

Journal: Proc Natl Acad Sci U S A

Publication Year: 2009

Authors: Zhengqing Ouyang, Qing Zhou, Wing Hung Wong

PubMed link: 19995984

Funding Grants: Guiding the developmental program of human embryonic stem cells by isolated Wnt factors, Stanford CIRM Training Program

Public Summary:
A new sequencing technology was developed to increase the scope and the resolution of transcriptional regulation study. RNA sequencing (RNA-Seq) and ChIP-Seq experiments are now generating comprehensive data on transcript abundance and on regulator-DNA interactions. Compared with traditional methods, our approach not only offers higher power in predicting gene expression from ChIP-Seq data but also provides a way to capture cooperation among regulators. In mouse embryonic stem cells (ESCs), we find that a remarkably high proportion of variation in gene expression (65%) can be explained by the binding signals of 12 transcription factors (TFs). Two groups of TFs are identified. Whereas the first group (E2f1, Myc, Mycn, and Zfx) act as activators in general, the second group (Oct4, Nanog, Sox2, Smad1, Stat3, Tcfcp2l1, and Esrrb) may serve as either activator or repressor depending on the target. The two groups of TFs cooperate tightly to activate genes that are differentially up-regulated in ESCs. In the absence of binding by the first group, the binding of the second group is associated with genes that are repressed in ESCs and derepressed upon early differentiation.

Scientific Abstract:
Next-generation sequencing has greatly increased the scope and the resolution of transcriptional regulation study. RNA sequencing (RNA-Seq) and ChIP-Seq experiments are now generating comprehensive data on transcript abundance and on regulator-DNA interactions. We propose an approach for an integrated analysis of these data based on feature extraction of ChIP-Seq signals, principal component analysis, and regression-based component selection. Compared with traditional methods, our approach not only offers higher power in predicting gene expression from ChIP-Seq data but also provides a way to capture cooperation among regulators. In mouse embryonic stem cells (ESCs), we find that a remarkably high proportion of variation in gene expression (65%) can be explained by the binding signals of 12 transcription factors (TFs). Two groups of TFs are identified. Whereas the first group (E2f1, Myc, Mycn, and Zfx) act as activators in general, the second group (Oct4, Nanog, Sox2, Smad1, Stat3, Tcfcp2l1, and Esrrb) may serve as either activator or repressor depending on the target. The two groups of TFs cooperate tightly to activate genes that are differentially up-regulated in ESCs. In the absence of binding by the first group, the binding of the second group is associated with genes that are repressed in ESCs and derepressed upon early differentiation.

PNAS Lens Free Article Link: