# Predicting transcription factor target genes from ChIP-seq data

Weronika Sikora-Wohlfeld[a], Marit Ackermann[a], Eleni Christodoulou[a], Andreas Beyer[a,b]
a) Biotechnology Center, TU Dresden, Dresden, Germany
b) Center for Regenerative Therapy Dresden, Dresden, Germany

Immunoprecipitation of the chromatin fragments bound by a transcription factor (TF) coupled with DNA sequencing (ChIP-seq) yields precise information about genomic binding locations of TFs. Identification of target genes from the binding positions, however, remains an open problem. Here we present an evaluation of TF-target assignment algorithms that uses independent functional assays for scoring the performance of the methods.

Potential target genes are typically predicted from ChIP-seq data by correlating peak coordinates with the position and structure of proximal genes. In the simplest approach the TF targets are identified based on the presence of the binding event within a certain distance from the transcription start site (TSS). We compared this approach with more sophisticated methods, considering number of peaks and their distribution around the TSS. We applied those methods to 66 ChIP-seq studies focusing on the hematopoietic system and embryonic stem cells and quantitatively assessed their performance using 23 external datasets reporting expression changes upon changing the activity of a TF (either through engineered perturbations or during physiological transitions).

This analysis shows that considering both, the density of the peaks around the TSS and their distance to the TSS, gives best results. We propose a TF-target scoring method which is based on the TF-specific distribution of peak-TSS distances and thus does not require setting an arbitrary window. Hence, this method is adaptable to 'new' TFs without any prior knowledge. Our parameter-free method is ideal for the hundreds of TFs that are currently being studied in large consortia where prior knowledge about their biological functions and molecular details is scarce.