Cell

Transcription Factor Binding in Human Cells Occurs in Dense Clusters Formed around Cohesin Anchor Sites

Jian Yan,[1,4] Martin Enge,[1,4] Thomas Whitington,[1] Kashyap Dave,[1] Jianping Liu,[1] Inderpreet Sur,[1] Bernhard Schmierer,[1] Arttu Jolma,[1] Teemu Kivioja,[1,2,3] Minna Taipale,[1,3] and Jussi Taipale[1,3,*]
[1]Science for Life Laboratory, Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm 14183, Sweden
[2]Department of Computer Science, P.O. Box 68
[3]Genome-Scale Biology Program, P.O. Box 63
00014 University of Helsinki, Helsinki, Finland
[4]These authors contributed equally to this work
*Correspondence: jussi.taipale@ki.se
http://dx.doi.org/10.1016/j.cell.2013.07.034

SUMMARY

During cell division, transcription factors (TFs) are removed from chromatin twice, during DNA synthesis and during condensation of chromosomes. How TFs can efficiently find their sites following these stages has been unclear. Here, we have analyzed the binding pattern of expressed TFs in human colorectal cancer cells. We find that binding of TFs is highly clustered and that the clusters are enriched in binding motifs for several major TF classes. Strikingly, almost all clusters are formed around cohesin, and loss of cohesin decreases both DNA accessibility and binding of TFs to clusters. We show that cohesin remains bound in S phase, holding the nascent sister chromatids together at the TF cluster sites. Furthermore, cohesin remains bound to the cluster sites when TFs are evicted in early M phase. These results suggest that cohesin-binding functions as a cellular memory that promotes re-establishment of TF clusters after DNA replication and chromatin condensation.

INTRODUCTION

A large fraction of the human genome encodes information about when and where genes should be expressed. This information is embedded into at least three different types of gene regulatory elements: promoters, enhancers, and insulators (Ong and Corces, 2011). The interaction between enhancers and promoters is thought to involve DNA looping, mediated by two protein complexes, mediator and cohesin. The mediator complex links sequence-specific TFs with RNA polymerase II (pol II). Cohesin, in turn, is a large ring-shaped molecule, capable of encircling two DNA strands. Its name comes from its first identified function, the establishment of cohesion between sister chromosomes (Nasmyth, 2011; Sherwood et al., 2010). Later, it was found that cohesin also has a role in transcription (Rollins et al., 1999). The insulator protein CTCF (Wendt and Peters, 2009) recruits cohesin, and cohesin can also be loaded to promoter and enhancer elements in a CTCF-independent fashion (Kagey et al., 2010).

Although transcription is understood in broad conceptual terms, building predictive models has proven challenging. Even modeling where in the genome TFs bind has proven to be a formidable task. TF binding to DNA is a competitive reaction, in which the ensemble of all TFs in a cell compete against histones. Therefore, understanding where a single TF binds requires knowledge of the entire system. However, even the largest efforts so far, HT-ChIP (Garber et al., 2012) and the ENCODE project (Gerstein et al., 2012), have only analyzed less than 50 TFs in a single cell type, a number far below the estimated number of TFs active in a cell.

How the pattern of binding of hundreds of different TFs can be efficiently inherited after cell division is unclear, as even in *E. coli*, whose genome size is 0.15% of that of a human, it takes a single TF minutes to find its binding site (Hammar et al., 2012). Methods such as DNase I hypersensitivity have suggested that only a fraction of the human genome is accessible for TF binding (see, for example, Thurman et al., 2012), greatly increasing the speed by which TFs can find their target sites. However, the mechanisms by which such accessible regions could be inherited are unclear. They have been suggested to be marked by modified histones, which bind DNA more weakly than unmodified histones (Bode et al., 1980; Oliva et al., 1987) and facilitate TF binding (Lee et al., 1993). The modified histones could survive S phase by being backloaded after passage of the replication fork. However, the precision of backloading appears insufficient to mark short accessible regions (Radman-Livaja et al., 2011). Also, histone modifications at marked sites have been shown to be temporarily lost upon passage of the replication fork (Petruk et al., 2012), ruling out purely histone-based mechanisms of inheritance of accessible DNA. Most TFs are also evicted from DNA in early M phase (Martínez-Balbás et al., 1995; Zaidi et al., 2010), when chromatin is condensed. Although some accessible regions remain bound by specific TFs, a general mechanism by

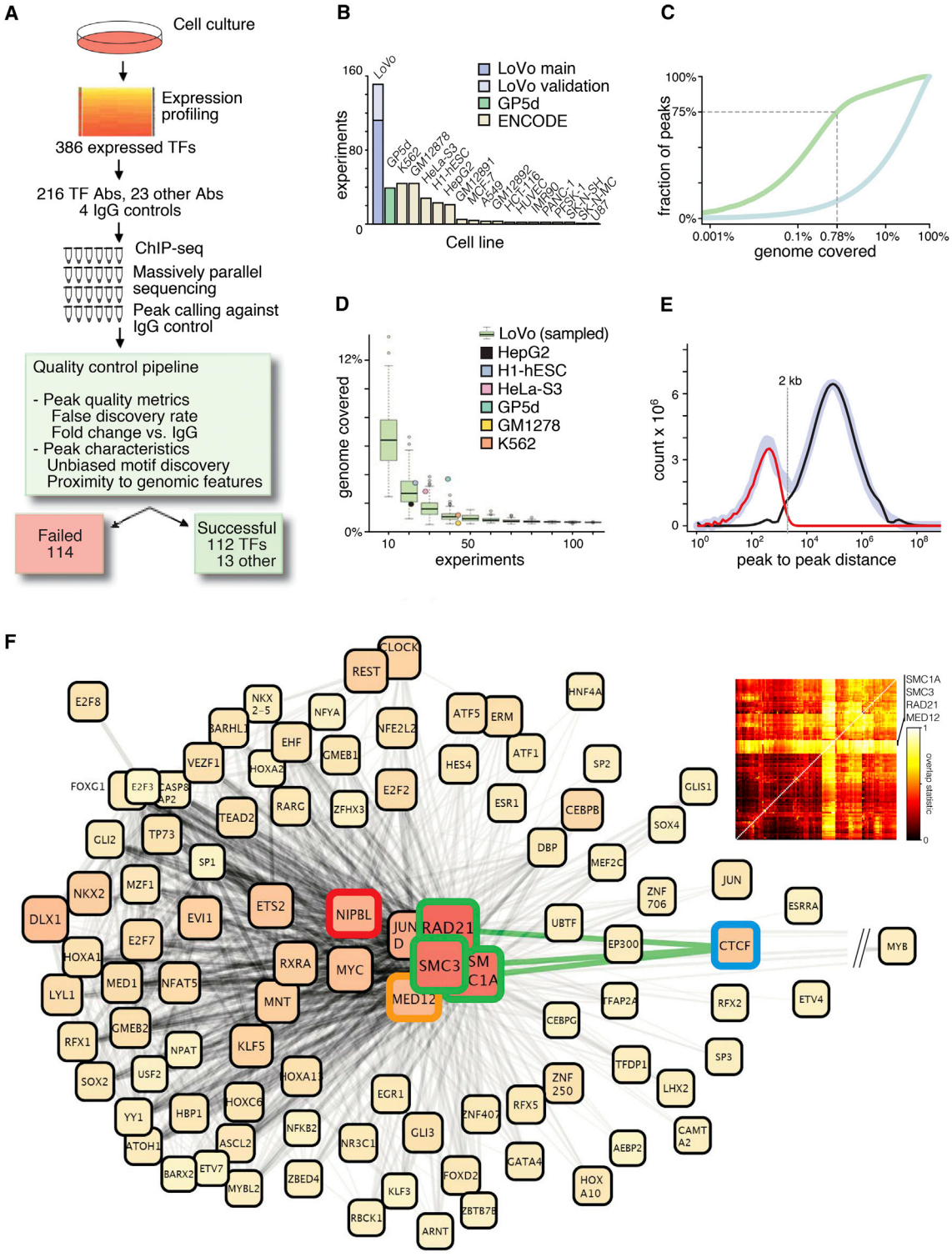

Figure 1. High-Throughput ChIP-seq

(A) Experimental setup and QC pipeline.

(B) Number of TF experiments in our and previous studies. Experiments that were deemed successful by the QC pipeline (A) are shown.

(C) Majority of all TF-binding sites are found in only ~0.8% of the genome. Fraction of peak-peak intervals (y axis) as a function of fraction of genome covered (x axis) by the same intervals (green line) indicates that 0.8% of genome contains more than 75% of all peaks. Distribution expected by random (blue) is shown for comparison.

(legend continued on next page)