Shared Transcription Factors Contribute to Stage-specific Transcriptional Programs during Blood Cell Differentiation

Felicia Ng, Fernando Calero-Nieto, Nicola Wilson, Rebecca Hannah, Evangelia Diamanti, Berthold Gottgens

Cambridge Institute for Medical Research, Wellcome Trust/MRC Building, Addenbrooke's Hospital, Hills Road, Cambridge, CB2 0XY United Kingdom

BACKGROUND

Transcription factors (TFs) have long been recognized as important regulators of haematopoietic cell type identity. Specific TFs have been shown to be critical for regulating pluripotency genes in haematopoietic stem cells while others drive differentiation to mature haematopoietic cell types (Orkin and Zon 2008). As a result, TFs have been extensively studied at all stages of haematopoietic development (Wilson et al. 2011). Furthermore, advances in the generation of TF binding maps by ChIP-seq permitted investigations at the genome level. While an abundance of ChIP-seq data exists for different haematopoietic cell types, not much is known about the genome-wide impact of TF binding in driving transcriptional programs of multiple cell types.

An observation from several independent ChIP-seq studies is the strong cell-type-specific binding pattern displayed by many haematopoietic TFs (Hannah et al. 2011; Wei et al. 2011; Pilon et al. 2011). These studies demonstrated that the binding profile of different TFs in the same cell type show stronger correlation than the binding profiles of the same TFs in different cell types. Interestingly, this cell-type-specific binding pattern was also observed for so-called 'master regulators' of haematopoietic stem cells, therefore, raising the question as to how 'master regulators' dictate cell type identity? Most importantly, is this observation an indication of 'functional' rather than 'opportunistic' binding events?

To address these questions, we have analysed the genome-wide binding maps of 10 key haematopoietic stem cell TFs in both primary mast cultures and a progenitor cell line. In addition, expression profiling by RNA-seq on both cell types were analysed in conjunction with the TF binding data to provide a more comprehensive view of gene expression regulation.

RESULTS

Gene expression profiling in mast cells and a progenitor cell line (HPC7) showed that many haematopoietic stem cells 'master regulators' were indeed expressed at similar levels in both cell

types. We also showed that HPC7 closely resembles common myeloid progenitors (precursors of mast cells) and recapitulates the gene expression profile of early blood stem/progenitor cells. Shared expression of key stem cell TFs, therefore, suggests that a more detailed comparative analysis of genome-wide binding patterns in both cell types may provide new insights into the transcriptional control of cell type identity.

A global comparison of HPC7 and mast ChIP-seq data for 10 stem cell TFs (Ctcf, E2a, Erg, Fli1, Gata2, Lmo2, Meis1, PU.1, Runx1, Scl) revealed very little overlap in binding sites (<30%). Moreover, pairwise correlation analysis of all 20 genome wide binding profiles followed by hierarchical clustering revealed clustering of all TFs by cell type, with the exception of Ctcf. These observations suggest that binding of the shared TFs are largely cell-type-specific for 2 closely related haematopoietic cell types. Having identified predominantly cell-type-specific binding patterns for key regulatory TFs raised the question as to whether TFs are passively recruited to cell-type-specific genomic regions of open chromatin with no major regulatory impact or actively participate in 2 different transcriptional programmes. To evaluate the extent to which cell-type-specific binding of shared TFs might be associated with gene expression, we developed multivariate linear regression models to correlate changes in TF binding (ΔTF) with changes in gene expression (ΔGE). Fitting in a simple linear regression model showed some correlation between ΔTF and ΔGE (R2 value ~22.7%). Further application of the linear model on subsets of the data – genes with at least 5TFs bound – increased the R2 value up to ~41.4%. Although higher variability was explained, this is not ideal since many genes were thrown out. We then sought an alternative approach by using generalized additive models (GAM) and by incorporating all pairwise interaction of shared TFs to account for cooperation between TFs. This approach allowed us to fit concordant pairs of TFs to differential gene expression in a non-linear fashion. GAM with interaction terms correlated more strongly with gene expression changes (R^2 ~ 41.8%) than GAM without interaction terms (R2 ~ 25.4%). We were also able to identify interesting TF pairs that co-operate to affect cell-type-specific gene expression.

The modelling approach suggested that cell type specific binding of shared TFs makes meaningful contributions to differential gene expression. However, it remained unclear whether cell-type-specific binding is largely mediated through direct or indirect binding to DNA. To do this, we carried out a comprehensive motif analysis of common as well as cell-type-specific TF-bound regions. We found that consensus sequence motifs of shared TFs were enriched across common and cell-type-specific regions indicating direct DNA binding of the shared TFs. Does this then suggest that cell-type-specific TFs are driving reorganization of shared TFs to cell-type-specific sites? Indeed, we observed specific enrichment and depletion of motifs in cell-type specific regions. From this

analysis, Mitf and c-Fos emerged as potential candidate regulators because their motifs were enriched only in mast-specific regions and our RNA-seq data showed significant over-expression of these genes in mast cells. We went on to generate ChIP-seq data for these 2 mast-specific factors and analysed overlapping binding sites with the 10 shared TFs in HPC7 and mast cells. We were able to show that Mitf and c-Fos binding co-occupy a substantial proportion of regions bound by shared TFs in both cell types but not HPC7-specific regions. Mitf and c-Fos also bind to mast-specific regions, and this 'new' binding is accompanied by relocation of shared TFs to these regions.

CONCLUSION

Taken together, these data are consistent with a model whereby mast cell specific and shared TFs contribute to gene regulation in mast cells by binding to both common and mast cell specific regulatory regions. A comprehensive understanding of how TFs interact with the genome will not only advance basic research but improves our mechanistic understanding of cellular reprogramming strategies developed within the stem cell and regenerative medicine area.

REFERENCES

Hannah R, Joshi A, Wilson NK, Kinston S, Göttgens B. 2011. A compendium of genome-wide hematopoietic transcription factor maps supports the identification of gene regulatory control mechanisms. *Experimental hematology* **39**: 531–41.

Orkin SH, Zon LI. 2008. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**: 631–44.

Pilon AM, Ajay SS, Kumar SA, Steiner LA, Cherukuri PF, Wincovitch S, Anderson SM, Mullikin JC, Gallagher PG, Hardison RC, et al. 2011. Genome-wide ChIP-Seq reveals a dramatic shift in the binding of the transcription factor erythroid Kruppel-like factor during erythrocyte differentiation. *Blood* **118**: e139–48.

Wei G, Abraham BJ, Yagi R, Jothi R, Cui K, Sharma S, Narlikar L, Northrup DL, Tang Q, Paul WE, et al. 2011. Genome-wide analyses of transcription factor GATA3-mediated gene regulation in distinct T cell types. *Immunity* **35**: 299–311.

Wilson NK, Tijssen MR, Göttgens B. 2011. Deciphering transcriptional control mechanisms in hematopoiesis:the impact of high-throughput sequencing technologies. *Experimental hematology* **39**: 961–8.