Methods 53 (2011) 151–162

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Review Article

A systems approach to analyze transcription factors in mammalian cells

Eric Soler [a,*], Charlotte Andrieu-Soler [a], Ernie de Boer [a], Jan Christian Bryne [b], Supat Thongjuea [b], Erikjan Rijkers [c], Jeroen Demmers [d], Wilfred van IJcken [e], Frank Grosveld [a]

[a] Department of Cell Biology, Erasmus Medical Center, Dr. Molewaterplein 50, Rotterdam, The Netherlands
[b] Computational Biology Unit, University of Bergen, Thormøhlensgate 55, N-5008 Bergen, Norway
[c] Department of Biochemistry and Center for Biomics, Erasmus Medical Center, Rotterdam, The Netherlands
[d] Mass Spectrometry & Proteomics Center, Erasmus Medical Center, Rotterdam, The Netherlands
[e] Center for Biomics, Erasmus Medical Center, Rotterdam, The Netherlands

ARTICLE INFO

Article history:
Accepted 6 August 2010
Available online 10 August 2010

Keywords:
Transcription factor
Mass spectrometry
High throughput
Next generation sequencing
ChIP-Seq
Network
Hematopoiesis
Erythroid cells

ABSTRACT

Transcription factors (TFs) play a central role in the development of multicellular organisms. The sequential actions of critical TFs direct cells to adopt defined differentiation pathways leading to functional, fully differentiated tissues. Here, we describe a generic experimental pipeline that integrates biochemistry, genetics and next generation sequencing with bioinformatics to characterize TF complexes composition, function and target genes at a genome-wide scale. We show an application of this experimental pipeline which aims to unravel the molecular events taking place during hematopoietic cell differentiation.

1. Introduction

Hematopoiesis is a complex process that leads to the production of all blood cell lineages. In the mouse embryo, hematopoiesis occurs in multiple distinct waves. During the first wave the yolk sac generates mainly primitive erythrocytes starting at embryonic day 7.5 (E7.5). The second and later wave leads to the production of definitive hematopoietic progenitors in the yolk sac, placenta and Aorta-Gonad-Mesonephros region. The first hematopoietic stem cells (HSCs) are autonomously generated starting at E10.5 in the dorsal aorta. HSCs then colonize the fetal liver which becomes the main hematopoietic tissue at mid-gestation, before migrating to and colonizing the bone marrow (BM) just before birth (for review see [1]). HSCs remain in the BM to maintain homeostasis during the entire adult life. Mature hematopoietic cells show dramatic differences in size, structure and function, but they all derive from the same pool of rare HSCs in response to extracellular and intracellular signals. Once engaged in a particular differentiation pathway, the cells will activate and/or repress lineage-specific gene expression programs in order to obtain their cellular identity and mature into fully differentiated cells (e.g. myeloid versus lymphoid). This process is highly dependent on the timely expression of highly conserved lineage-specific transcription factors (TF) that will determine cellular identity. Among such TF, Runx1, Tal1, Ldb1, Lmo2, Gata2, Gata1, Eto2 are known to control key steps of hematopoietic cell commitment [2–9].

While a growing number of these key transcription factors are being identified, the molecular mechanisms establishing their functions are poorly understood. Genes regulated by these transcription factors are largely unknown, but a rapid way to clarify this is:

(i) To identify TF interacting partners and to characterize the molecular complexes they form.
(ii) To map with a high resolution (all) the binding sites of these complexes in the genome in order to identify their target genes.

Here, we describe an experimental pipeline allowing characterization of TF complex composition, selection of critical complex components and their relevance *in vivo*, and genome-wide identification of their DNA binding sites in hematopoietic cells. This pipeline combines biochemical techniques (single-step protein complex capture coupled to mass spectrometry), genetic screens in zebrafish embryos and ChIP-Sequencing (ChIP-Seq) using massively parallel high throughput sequencing technologies.

* Corresponding author. Fax: +31 107044743.
E-mail address: e.soler@erasmusmc.nl (E. Soler).

1046-2023/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ymeth.2010.08.002

2. Description of the pipeline

The experimental pipeline is depicted in Fig. 1. Transcription factors of interest are fused to a composite affinity tag containing a biotinylable tag (Bio), a Flag tag and a V5 affinity tag separated by protease cleavage sites (PreScission and Tev). The presence of different epitope tags combined with protease cleavage sites allows multi-step protein purification to be performed when needed (not described in this article, and usually not necessary). Tagged TFs are expressed in (hematopoietic) cell lines at levels close to the endogenous expression level in order to avoid over-expression artifacts. Alternatively, tagged TF are knocked-in in mouse embryonic stem (ES) cells that are used either for *in vitro* differentiation experiments or to generate mouse expressing an endogenously tagged TF. The TF complexes are captured using single-step protein pull-down procedure [10] and bound proteins are identified by mass spectrometry. As a first quality check, the interactions with the newly identified binding partners are confirmed by co-immunoprecipitation and co-localization of the endogenous proteins in cell lines and mouse tissues in order to eliminate potential false positive hits.

Next, the expression profile of the novel interacting partners is determined by *in situ* hybridization in mouse and zebrafish embryos. Particular emphasis is given to factors predominantly expressed in hematopoietic tissues and during key steps of hematopoietic cells development in the embryo. For a rapid functional analysis, these factors are then subjected to morpholino-mediated knock-down in zebrafish to uncover potential hematopoietic defects *in vivo*. Genetic regulation of hematopoiesis is highly conserved between zebrafish and mammals. The use of zebrafish embryos offers a robust and efficient way to screen a significant number of factors in a short time (24–48 h) as compared to mouse knock-out.

Based on these assays, the most relevant candidates are conditionally targeted in mouse to further assess their role during mammalian development.

Finally, in order to identify target genes regulated by TFs and TF complexes on a genome-wide scale, chromatin immunoprecipitation (ChIP) experiments coupled to ultra high throughput sequencing (ChIP-Seq) are carried out. This technique allows fast mapping of virtually all binding sites of a TF in the genome. The massive amount of information generated by ChIP-Seq is integrated into a user-friendly interface (Solex software; Rijkers EJ, unpublished) allowing data sorting and visualisation. Ultimately, the most interesting new candidate proteins can be tagged for another round of complex determination and target gene identification to expand our knowledge on TF networks.

All the information generated during this workflow is integrated into a database of TF networks, which shows the co-localization and dynamics of TF binding to gene regulatory elements during the course of hematopoietic cells differentiation.

3. Protein tagging and single-step protein complex isolation

The single step purification of protein complexes has been described previously [10–11]. Briefly, the bacterial biotin ligase BirA is stably expressed in the cell line of interest, or in the mouse [12], together with the tagged transcription factor. The BirA enzyme catalyzes *in vivo* the addition of biotin to the Bio tag. The biotinylation of the target TF can be easily checked by Western blotting using streptavidin detection [10]. The biotin-tagged target transcription factor is then captured using streptavidin magnetic beads and pulled down from nuclear extract along with its interacting partners. Bio tag relies on the biotin–streptavidin interaction, which is one of the strongest non-covalent interactions known in nature (with a K_d 10^3–10^6 times greater than an antibody–antigen interaction). As a control, the same experiment is carried out in parallel with a cell or mouse line only expressing BirA in the absence of the tagged TF. Streptavidin beads containing the pulled down material are submitted to several washes to reduce non-specific interactions. Finally, on-bead trypsin digestion is performed overnight, and the resulting digested tryptic peptides are identified by nanoflow liquid chromatography tandem mass spectrometry (nanoLC–MS/MS). Specific components of TF complexes are identified by comparing background proteins (BirA cell line) to specific protein pull-downs.

Once new binding partners of a TF have been identified, the interactions between endogenous proteins are confirmed in cell lines or tissue homogenates by standard co-immunoprecipitation assay. We initially used this approach to characterize TF complexes

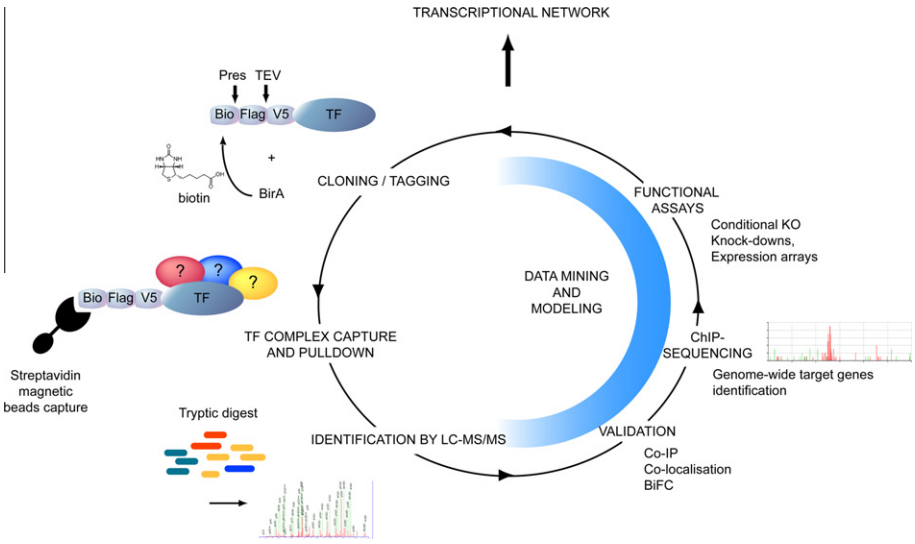

Fig. 1. Description of the experimental pipeline. Transcription factors (TF) are affinity tagged and expressed in cell lines or mouse tissues. TF complexes are captured using single-step protein pull-down and analyzed by mass spectrometry. Identified new components of TF complexes are confirmed by co-immunoprecipitation/co-localization and/or bimolecular fluorescence complementation (BiFC) assays [42]. ChIP-Sequencing (ChIP-Seq) is performed to identify genome-wide TF binding sites and target genes. TF and their binding partners are tested for their role *in vivo* during development by knock-down in zebrafish embryos and conditional inactivation in the mouse. Interesting new candidate proteins selected on the basis of this analysis enter the pipeline for a new round of detailed molecular and functional analysis.