

INNOVATION

## Novel ChIP-based strategies to uncover transcription factor target genes in the immune system

Amy S. Weinmann

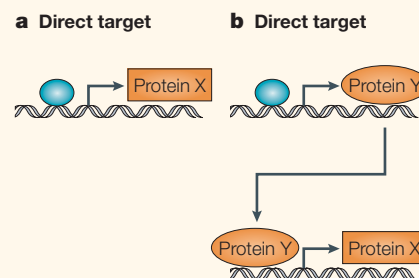
Transcription factors can have a marked effect on the fate of a cell by establishing the gene expression patterns that determine cellular function. Therefore, a great deal of effort has been invested in identifying and understanding the individual transcription factors that influence key activities. New strategies to identify transcription factor target genes based on their ability to bind to DNA in the nuclear environment have recently been developed, providing an opportunity to address many questions concerning the function of transcription factors. This article discusses the advantages and applications for these new strategies in reference to the developing immune response.

In the haematopoietic system, a large amount of work has established the identity of key transcription factors that are responsible for differentiation decisions and effector-cell functions. For example, the differentiation of CD4<sup>+</sup> T helper (T<sub>H</sub>) cells into either the T<sub>H</sub>1- or T<sub>H</sub>2-cell phenotype is influenced by transcription factors such as **T-bet**, signal transducer and activator of transcription 4 (**STAT4**), **GATA3** and **c-MAF**<sup>1,2</sup>. In another example, the activities of factors such as runt-related transcription factor 1 (**RUNX1**) and **GATA2** have been shown to be intimately involved in early haematopoietic-cell differentiation decisions<sup>3</sup>. To define the cellular and molecular mechanisms by which each factor acts, mice deficient in

these and other key factors have been used to identify genes that are functionally regulated by each factor. Complementary studies using overexpression in non-permissive cell types have also contributed to the hypotheses concerning the mechanisms by which each factor mediates its biological effects. However, the main disadvantage of using these genetic approaches to identify transcription factor target genes is that the differential gene expression patterns resulting from either the overexpression or the lack of expression of a transcription factor cannot delineate between the direct and indirect actions of that factor (FIG. 1). In other words, which genes are controlled by the factor directly binding to the regulatory elements for that gene, as opposed to those that are influenced through an indirect cascade of events in response to the control of other regulatory proteins? These difficulties in interpretation can lead to confusion over the order of events that occur in response to transcription-factor activity and make it difficult to accurately piece together the cascade of regulatory events involved in specific biological processes.

New approaches are being developed and refined to examine how transcription factors influence cellular decisions on the basis of their ability to regulate gene expression patterns directly. Here, I discuss chromatin immunoprecipitation (ChIP)-based target-gene approaches that localize transcription-factor binding in the context of the natural,

nuclear environment, and so avoid the main disadvantage of the classical target-gene approaches; most of which rely on interpreting altered mRNA levels and therefore cannot distinguish direct versus indirect target genes. These new ChIP-based approaches have been developed over the past few years, first in yeast<sup>4,5</sup> and then in the more complex human system<sup>6,7</sup>, and provide the exciting opportunity to address questions concerning the identity of the genes that have the potential to be directly regulated by an individual transcription factor. After outlining the procedure, I discuss applications for these techniques in studying the role transcription factors have in the immune response.



**Figure 1 | Classifying transcription factor target genes.** A simplified representation of direct and indirect transcription factor target genes. **a** | In the case of the direct target, the transcription factor (blue oval) directly interacts with regulatory regions of a gene (that is, promoter or enhancer) to influence transcription and ultimately protein expression. **b** | In the case of the indirect target, the transcription factor is still functionally responsible for regulating the downstream gene, but it is at least one step removed and does not directly interact with the regulatory regions of the gene. Instead, the transcription factor directly controls the expression of a second regulatory protein (protein Y) that is then directly responsible for regulating the expression of the final output protein. Both the direct and indirect targets are functionally regulated by the transcription factor, but, the varying molecular mechanisms require different strategies to manipulate the expression of the output protein with the smallest number of indirect effects.

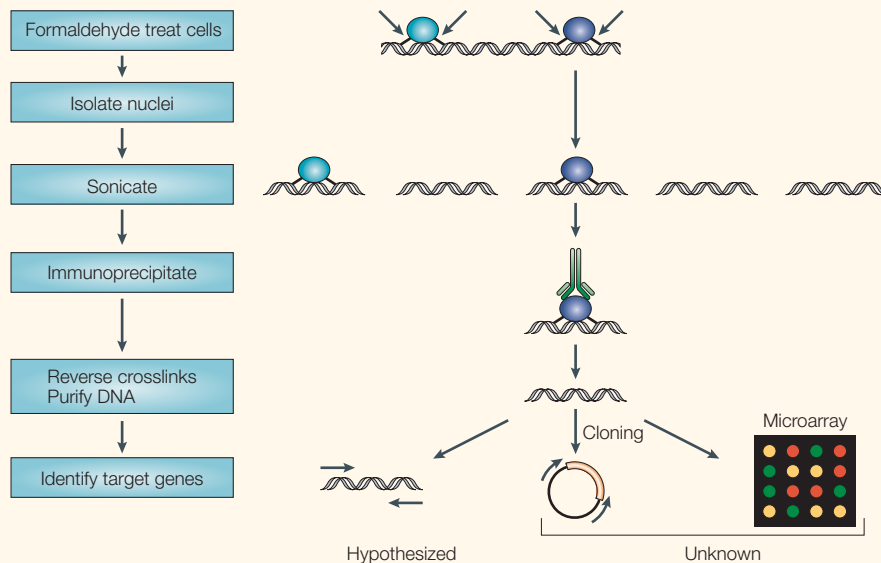


Figure 2 | **Schematic representation of ChIP-based target-gene identification strategies.** Cells are treated with formaldehyde to crosslink proteins in close association with DNA. After crosslinking, nuclei are isolated and sonicated to shear the genome randomly, allowing for a more refined localization of the protein–DNA interactions. Next, an immunoprecipitation step with a factor-specific antibody is carried out. After collecting the immunoprecipitated material, the crosslinks are reversed and the DNA is purified. The identity of the DNA fragments contained in the chromatin immunoprecipitation (ChIP) sample can then be determined by either the standard PCR method for hypothesized targets or, to identify unknown targets, by either cloning the fragments into a plasmid vector for sequencing, or by labelling the ChIP material with fluorophores followed by hybridization to a genomic microarray.

### ChIP-based assays

The standard ChIP assay allows the researcher to take a virtual molecular snapshot of protein–DNA interactions in the context of the living cell (FIG. 2). This is accomplished by treating the cells with formaldehyde to crosslink proteins in close association with DNA. Of note, formaldehyde can also crosslink protein–protein interactions. This is an important property allowing for the analysis of proteins that are recruited to the DNA through association with another DNA-binding factor. Immunoprecipitation of the sample with a transcription factor-specific antibody allows for the isolation of DNA fragments that are specifically associated with the protein of interest. In the standard ChIP assay, one must first have an idea of where a factor interacts with the genome, because the final analysis depends on PCR amplification of the DNA with gene-specific primers. If a fragment is specifically enriched in the immunoprecipitated material and is absent in a non-specific antibody control, this would indicate that the region is specifically bound by the protein of interest.

To adapt this procedure to identify unknown transcription factor target genes, strategies have been devised to identify the DNA fragments bound by the factor of interest without prior knowledge of the interacting

sequence (FIG. 2). One method to accomplish this goal is to clone the DNA from the immunoprecipitated material into a plasmid vector to allow for the sequencing of the fragments using common vector-specific primers<sup>8</sup>. After sequencing, genomic database searches can be carried out to determine the location of the immunoprecipitated DNA fragments in relation to known or predicted genes. Do they correspond to the promoter region of a gene or do they reside in other regions of the genome, such as distal or intronic enhancers? Although the individual cloning and sequencing of the ChIP material can be time consuming, this technique has the potential to identify any DNA fragment that is bound by a transcription factor. An alternative approach to provide a high throughput screening strategy is to use genomic microarray technology (BOX 1). So far, genomic microarrays that contain sets of selected promoters, a random representation of CpG islands and portions of continuous genomic sequence have all been used to screen ChIP samples<sup>6,7,9,10</sup>. Each has provided information about the localization of transcription factors in human cells, and at times, the results have been somewhat unexpected. In one study of nuclear factor- $\kappa$ B (NF- $\kappa$ B) binding, the results indicated that NF- $\kappa$ B could bind to many regions of the genome, not just to the promoters<sup>9</sup>. The data also indicated that

binding does not always correlate with a functional consequence in all cell types. For example, it was shown that tumour-necrosis factor (TNF) induced the binding of NF- $\kappa$ B to several genes that were not concomitantly altered in transcriptional activity<sup>9</sup>. This is an important point to keep in mind when interpreting data from ChIP-based target-gene studies, which can only provide information about the binding status of a transcription factor and not its functional effect on gene expression. This point is further discussed later.

### Advantages of ChIP-based assays

The modified ChIP assays have several advantages for identifying transcription factor target genes. First, by virtue of the properties of the ChIP assay, only DNA fragments specifically associated with the protein of interest will be identified. This eliminates the traditionally problematic mRNAs that are indirectly regulated by altered signal-transduction cascades. Second, this approach can be carried out with biologically relevant cell populations. Therefore, the transcription factor can be examined without altering its levels (that is, overexpression or gene knockout), leaving the stoichiometry of biologically relevant protein–protein interactions that occur in the context of a given cellular environment intact. This might be particularly important when examining issues concerning cell specificity of transcription-factor activity, as each cell type might have a different set of limiting cofactors. Another advantage of this technique is that it takes into account the chromatin environment of an individual cell type that might in fact be responsible for communicating a degree of cellular specificity. For example, repressive chromatin organization, as shown by a lack of characteristic DNase I hypersensitivity, is observed at the interleukin-4 (*IL-4*) locus in T<sub>H</sub>1 cells, correlates with the prevention of transcription-factor binding and gene inactivation<sup>11</sup>. By contrast, several DNase-I-hypersensitive regions that are associated with a more accessible chromatin structure are found at the *IL-4* locus in T<sub>H</sub>2 cells, where transcription factors can bind and induce transcription<sup>11</sup>. This example illustrates the importance of taking into account the endogenous chromatin organization, which can contribute to the regulation of factor binding and transcriptional activity in specific cell types. Finally, this technique also provides the advantage that prior knowledge of potential target genes or binding elements is not required. Instead, this technique provides an unbiased localization of transcription-factor binding in the

cellular environment. Therefore, new targets that might have been overlooked by current hypotheses can be uncovered.

Using these new target-gene identification approaches, many questions can be directly addressed to better understand the molecular details of the immune response. The molecular pathways that a transcription factor might regulate for functional maturation, cell-type-specific activity and family-member specificity are just a few of the questions that can be investigated. The following sections discuss these issues using the transcription factor T-bet as an example. T-bet represents a model transcription factor that is important in the development of many haematopoietic cell types and it highlights the insight that can be gained from these target-gene strategies.

### Using ChIP strategies to analyse T-bet Identification and characterization of T-bet.

The transcription factor T-bet is a T-box-family member that has been hypothesized to have a crucial role in the development of the immune response, as it was originally isolated due to its selective expression by the  $T_H1$ -cell lineage of  $CD4^+$   $T_H$  cells<sup>12</sup>. Transient overexpression of T-bet in naive  $CD4^+$   $T_H$  cells results in  $T_H1$ -cell lineage commitment, as shown by the ability of these cells to produce interferon- $\gamma$  (IFN- $\gamma$ ) and repress  $T_H2$ -type cytokine activity<sup>12</sup>. In addition, T-bet overexpression in  $T_H2$  cells

reverses their commitment and leads to IFN- $\gamma$  production. These experiments indicated that T-bet has a crucial role in the development of  $T_H1$  cells, a hypothesis that was confirmed by analysis of T-bet-deficient mice<sup>13</sup>.

Further studies of the expression pattern of T-bet have shown that it is also expressed and has specific roles in many other haematopoietic cell types, including natural killer (NK) cells, macrophages, dendritic cells, B cells and  $CD8^+$  cytotoxic T cells<sup>12,14–16</sup>. Surprisingly, in T-bet-deficient mice, although  $T_H1$  cells were unable to produce marked amounts of IFN- $\gamma$ ,  $CD8^+$  T cells could still produce high levels of the cytokine<sup>13</sup>, indicating a divergence of T-bet function in the different T-cell lineages. A related T-box-family member, Eomesodermin, has recently been identified and shown to be important in  $CD8^+$  T cells, which might explain why T-bet is less crucial for IFN- $\gamma$  production by this cell type<sup>17</sup>. Collectively, these data indicate that there is a degree of cell-type specificity to the functional activity of T-bet. Uncovering the molecular pathways that are regulated by T-bet in the different cell types will help to address the role that T-bet has in the developing immune response.

**Addressing molecular pathways.** The expression of a single transcription factor can have marked consequences on cell fate. In the example of T-bet, the developmental pathways of  $T_H1$  cells and other haematopoietic lineages

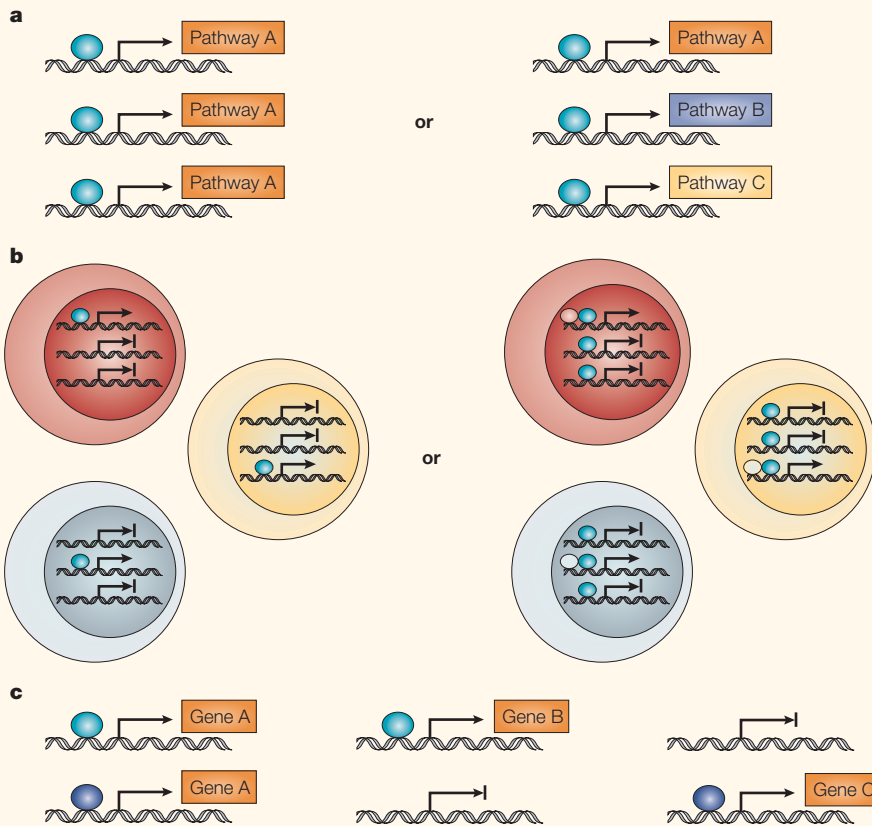
are impaired in its absence. A key question to address is the mechanism behind the role of T-bet in these processes. Does T-bet directly influence the transcriptional regulation of important cytokine and effector-cell genes or is it influencing cellular function through a more indirect pathway by regulating other crucial regulatory factors? Perhaps the most probable answer is that these options are not mutually exclusive, and a combination of direct and indirect effects will contribute to its function. Progress has been made using traditional methods to identify genes that are functionally regulated by T-bet in  $T_H1$  cells, and their identities (IFN- $\gamma$ , *HLX* and *IL-12R2*) indicate that a combination of effects is probable<sup>12,18,19</sup>. However, there has been debate about the nature of T-bet activity and where it fits into the regulation of  $T_H1$ -cell development relative to other crucial  $T_H1$ -cell transcription factors such as STAT1 and STAT4 (REFS 19,20). The new ChIP-based target-gene strategies might provide a degree of molecular clarity to this issue and identify which genes have the potential to be regulated directly by each factor during the sequential stages of development. These data will ultimately provide insight into the molecular series of events that occur during the differentiation of a  $T_H1$  cell and will be complementary to the genetic studies already carried out. Together, this information can then be used to piece together the series of gene expression events that occur during the commitment process.

Sometimes, our investigation into the molecular series of events that allow a cell to develop along a specific pathway can be limited due to the current knowledge that is available. In the case of a newly discovered factor, there might be little information known, whereas for a well-studied factor, the current data might impede the uncovering of new functions that do not fall within the established constraints. In the case of T-bet, the question remains whether T-bet has any additional, subtle roles in  $T_H1$ -cell differentiation that are independent of the regulation of cytokine and effector genes, which have been the focus of present research. Do specialized lineage-restricted transcription factors also have a role in the fine regulation of common cellular processes, such as mRNA stability or protein trafficking, to suit the individual needs of a particular cell type or developmental stage? To address these questions in an unbiased manner, the ChIP-based cloning method allows for the identification of any potential target gene, whereas the ChIP-microarray method is perhaps best suited to examine transcription factor target genes in a global manner, while keeping in mind that bias can

### Box 1 | Genomic microarrays

Several types of genomic microarray, including select promoters, CpG islands and continuous genomic sequences, have been used in chromatin immunoprecipitation (ChIP)-microarray studies to examine transcription-factor binding in human cells<sup>6,7,9,10</sup>. Each has advantages as well as inherent biases that need to be taken into account when interpreting experimental results.

Hypotheses can be rapidly tested by using microarrays that contain a large number of selected promoters for genes with functions that are thought to be regulated by a particular factor; however, such microarrays are unlikely to uncover new targets. New targets can be identified using randomized promoter microarrays, CpG-island microarrays or microarrays containing continuous genomic sequences. The rationale for using CpG-island microarrays is that CpG islands typically correspond to the promoter and first exon of a gene<sup>27</sup>, and it is estimated that more than 50% of promoters are in close proximity to a CpG island<sup>28,29</sup>. Therefore, CpG-island microarrays provide a promoter-enriched microarray that, although not comprehensive, presents a feasible way to sample target genes with numerous functions in a relatively unbiased manner. However, this type of analysis is limited to examining the binding largely at promoter regions, and it is important to note that non-CpG-island promoters will not be sampled. Finally, using microarrays that contain a continuous genomic sequence of a single chromosome has the advantage that any region of the genome that is bound by the transcription factor (that is, promoter, distal enhancer, intronic enhancer or a genomic region of unknown function) can be identified, as can any type of promoter (that is, CpG or non-CpG island). However, these microarrays are limited to examining the genes found on a single chromosome, making it difficult to assess the molecular pathways that are regulated by a single transcription factor. As genomic-microarray technology continues to develop and incorporate the genome-sequencing data, it will soon be possible to have more comprehensive microarrays that contain all promoters or possibly a greater number of chromosomes on a single microarray.



**Figure 3 | Addressing questions using new ChIP-based target-gene approaches.** **a** | Defining the molecular pathways that a transcription factor regulates — do the target genes have similar or diverse functions? Two opposing, simplified possibilities are shown. In the first, the transcription factor (blue oval) only regulates genes that are involved in a single cellular function (pathway A). In the second case, the transcription factor regulates genes whose products are involved in diverse cellular processes (pathways A, B and C). **b** | Is cell-type specificity of the transcription factor at the level of DNA binding or does it depend on the context of the association? Three cell types are represented by different colours. In one case, transcription-factor binding correlates with transcriptional activity, with binding being dependent on the nature of the chromatin structure in the individual cell type. In the opposing case, transcription-factor binding is not cell-type specific, but rather the transcriptional specificity is determined by the presence of other lineage-restricted binding proteins. **c** | Do closely related members of a transcription-factor family have the ability to bind to the same genes (gene A), or is their binding restricted in some settings (genes B and C)? The possibilities of overlapping and mutually exclusive family-member binding patterns at target genes are shown.

be introduced due to the content of the microarray (BOX 1). The identities of the gene-regulatory regions that are bound by the transcription factor might then provide clues as to which cellular processes have the potential to be directly affected by that factor (FIG. 3a).

**Addressing cell-type specificity.** T-bet, as with many other lineage-restricted transcription factors in the immune system, has a broader expression profile than originally anticipated. Not only is T-bet expressed by many haematopoietic lineages, but the T-bet-deficient phenotype has confirmed that it does in fact have a role in the fate of diverse cell populations. Cell-specific defects in T-bet-deficient  $T_H1$  cells, NK cells and B cells were observed<sup>13,15,21</sup>. In addition, defects in some antigen-specific cytotoxic T-cell

responses were also shown, as was impaired dendritic-cell development<sup>15,22</sup>. An important issue that needs to be addressed is the mechanism behind these cell-specific phenotypes. One possibility is that T-bet is targeted to and regulates different gene loci in each cell population (FIG. 3b). In this case, locus accessibility would probably account for the cell-type specificity of T-bet action. So, T-bet might have the same functional capacity in each cell type, but its action is restricted by cell-intrinsic chromatin structure that limits binding in different cell types. This finding would also imply that T-bet alone does not establish the chromatin structure of the target genes, but rather other cell-type-restricted factors are required to establish an accessible chromatin environment that exposes the binding sites for additional

factors. However, if T-bet is involved in chromatin-modification events, as is indicated by data at the *IFN- $\gamma$*  promoter<sup>23</sup>, one would then predict that T-bet has the ability to bind to chromatin templates. This raises the opposing possibility that T-bet is targeted to the same loci in each cell population, but that other cell-specific interactions at the target genes are responsible for the transcriptional specificity. In both cases, the loss of T-bet in a cell type that is competent for expression of a particular gene will result in a transcriptional defect in that gene, but T-bet expression in a cell type that is not competent for expression of that same gene will be inconsequential. There is likely to be a continuum of possibilities between these two opposing, simplified hypotheses. To start to address these questions, it is first necessary to identify the direct targets of T-bet in the different cell populations in which it is expressed. The ChIP-microarray target-gene approach again provides a particular advantage in addressing these questions from a global standpoint. Will factor binding be detected in the same pattern for all cell types? Alternatively, will there be no obvious overlap between the gene-regulatory regions that are bound in the diverse cell populations? Or will the results be somewhere in between? These data will provide valuable clues at the level of DNA binding. However, follow-up studies to determine whether binding has a functional consequence on transcriptional regulation in all or only a fraction of cases will then be crucial to address cellular specificity.

**Addressing family-member specificity.** One of the most problematic issues to address when examining the function of a specific transcription factor is the role of closely related family members in the same cell. When one family member is lost, as is the case in gene-knockout studies, another closely related family member might compensate in some situations. NF- $\kappa$ B-family member specificity at various target genes highlights this concept and was examined in studies using cells deficient in either a single or a combination of NF- $\kappa$ B-family members<sup>24</sup>. The data showed that there was a differential requirement for specific family members to functionally regulate certain endogenous promoters, but in several cases, this was only revealed when a combination of family members were lost. Other common methods to uncover the genes that are functionally regulated by a transcription factor involve the overexpression of that factor, which can cause the ratio of family members to be artificially skewed in a cell

Table 1 | **ChIP-based target-gene studies in mammalian cells**

Transcription factor studied	Method	Type of microarray	Findings	Refs
E2F1, E2F4, E2F6	ChIP cloning and ChIP microarray	CpG islands and selected promoters	Development of ChIP-based target-gene identification strategies in human cells using E2F4; family-member specificity also examined	6,7,8, 30,31
GATA1	ChIP microarray	Continuous genomic sequence from the $\beta$ -globin locus	GATA1 binding across the $\beta$ -globin locus examined	10
HNF1 $\alpha$ , HNF4 $\alpha$ , HNF6, RNA Pol II	ChIP microarray	Randomly selected promoters	Family-member and cell-type specificity of HNF transcription factors analysed	32
MBD1, MBD2, MBD3, MECP2	ChIP microarray	CpG islands	Several target genes bound multiple MBD-family members, but some targets had family-member specificity	33
MYC, MAX	ChIP microarray	CpG islands and randomly selected promoters	Numerous MYC target promoters were identified and examined for the role of MAX in their regulation; 15% of promoters on the arrays were bound by a MYC-MAX heterodimer in cancer-cell lines	34,35
MYC, SP1, p53	ChIP microarray	Continuous genomic sequence from chromosomes 21 and 22	Numerous MYC and SP1 binding sites were identified and many correlate with non-coding RNAs	36
NF- $\kappa$ B	ChIP microarray	Continuous genomic sequence from chromosome 22	NF- $\kappa$ B binding sites across chromosome 22 examined	9
RB	ChIP microarray	CpG islands	RB binding to target genes in different phases of the cell cycle examined	30
T-bet	ChIP microarray	CpG islands	Target genes were identified to examine the function of T-bet	26
BARX2	ChIP cloning	N.A.	Several BARX2 target genes were identified and confirmed with functional assays	37
EWS/ATF1	ChIP cloning	N.A.	As a modification to the procedure, ChIP fragments were directly cloned into a GFP-reporter vector to examine functional activity in an overexpression assay; identified EWS/ATF1 targets included POSH, ATM and ARNT2	38
EGR1	ChIP cloning	N.A.	Modified procedure to incorporate a multiplex PCR step to amplify ChIP samples before cloning; identified EGR1 target gene, TOE1	39
E2A	ChIP cloning	N.A.	Identified E2A target gene, BTL1	40
ORC1, ORC2	ChIP cloning	N.A.	Identified ORC target that might be a bidirectional origin of DNA replication	41
RUNX1	ChIP cloning	N.A.	Identified PKC- $\beta$ as a direct target of RUNX1	25

BARX2, BarH-like homeobox 2; ChIP, chromatin immunoprecipitation; EGR1, early growth response 1; GFP, green-fluorescent protein; HNF, hepatocyte nuclear factor; MBD, methyl-CpG-binding domain protein; MECP2, methyl-CpG-binding protein 2; N.A., not applicable; NF- $\kappa$ B, nuclear factor- $\kappa$ B; ORC, origin recognition complex; PKC- $\beta$ , protein kinase C- $\beta$ ; RB, retinoblastoma; RNA Pol II, RNA polymerase II; RUNX1, runt-related transcription factor 1; TOE1, target of EGR1.

population. In both of these cases, the experimental conditions can mask the normal activities of individual proteins and distort their natural function. Another powerful, but potentially problematic, experimental strategy to examine transcription-factor activity is the use of dominant-negative proteins. Dominant-negative proteins often repress the activity of more than just the intended protein. This can complicate the interpretation of results that characterize the role of an individual factor in a given cellular setting. This concept was highlighted by the discovery of Eomesodermin — a related T-box-family member that might have a complementary and/or independent role to T-bet in CD8<sup>+</sup> T cells<sup>17</sup>. In this case, dominant-negative constructs for either protein seem to repress the activity of both T-box-family members, resulting in a more marked phenotype than the loss of either transcription factor alone. Therefore, when interpreting the results of dominant-negative and

overexpression studies, it is important to exercise caution, as unidentified family members might be affected in addition to the intended protein.

One way to bypass these concerns and to examine the targets of a specific family member is to identify targets in the context of normal protein expression levels. The ChIP-based strategies provide a powerful method to accomplish this goal. Cells or tissues are directly analysed and the protein-DNA interactions are examined in the context of normal cellular levels without experimental manipulation. So, it is possible to test whether there truly is a redundancy for family members in the context of natural expression levels or whether there is, in fact, specificity (FIG. 3c). With the case of T-bet and Eomesodermin, differential expression levels in CD4<sup>+</sup> versus CD8<sup>+</sup> T cells imply that they might have different functional responsibilities in each cell population<sup>17</sup>. We can begin to test this hypothesis with a global analysis

of the binding pattern of each protein in the different cell types using the ChIP-microarray technique.

#### Caveats and further studies

Although this is an exciting time, as with all experimental strategies, caution must be taken in interpreting the results. As discussed, these techniques monitor the ability of a protein to associate with DNA, but before further analysis, it is not known whether there is a functional consequence to the detected interaction. It is entirely possible that not every interaction will result in a functional effect, as indicated by the study of NF- $\kappa$ B binding<sup>9</sup>. Therefore, follow-up expression analysis is required to determine whether the identified target genes are regulated by the factor in a specific cell population. This expression analysis is also required to provide insight into whether a factor is acting exclusively as an activator or a repressor, or can function as both in a context-dependent manner. In addition, follow-up,

detailed mechanistic studies will also need to be carried out to provide a more complete understanding of the role an individual factor has at a target gene in a given cell type. These follow-up studies will also independently validate the transcription factor target genes identified. To interpret the data accurately, the use of diverse experimental strategies for confirmation will provide a higher degree of confidence that the findings are relevant and not due to a caveat of an individual assay system, which despite our best efforts remain.

It is also worth noting that the premise of this technique requires a reliable antibody that works well in an immunoprecipitation assay. Especially if used to examine family-member specificity, care must be taken to ensure the characteristics and potential crossreactivity of the antibody to prevent misinterpretation of the results. It is also important to minimize the incorrect identification of targets (false positives) by carrying out comparisons between the specific antibody and a non-specific antibody control. Due to the highly repetitive nature of mammalian genomes, false positives can be problematic. As indicated earlier, independent verification of potential target genes will help to eliminate false positives, including those that are due to non-specific immunoprecipitation. The possibility of false negatives should also be kept in mind. An antibody might interact poorly, or not at all, with a protein if the epitope it recognizes is not easily accessible, which might be the case in some large, multiprotein complexes. So, specific enrichment in the ChIP assay cannot be used to determine relative binding affinity at diverse target genes, because epitope accessibility might influence the intensity of the ChIP signal. Finally, it is important to consider when interpreting the results of ChIP-microarray studies, that the identity of the sequences contained on the microarray might introduce bias into the types of target that can be identified (BOX 1).

## Conclusions

The ChIP-based target-gene approaches described here provide a new method to examine the role of crucial haematopoietic transcription factors in the developing immune response. Several studies have already been carried out and have provided unique insight into the roles of NF- $\kappa$ B, RUNX1 and T-bet<sup>9,25,26</sup>. As these and other factors continue to be studied (TABLE 1), the rapidly developing genomic microarray technology will provide an even more robust and complete platform to identify genomic binding sites. It is possible that in the future, as has already been accomplished in the yeast system<sup>4,5</sup>, complete genomic microarrays will have the potential to

show the binding pattern of a transcription factor across the entire genome. Comparing and contrasting the binding and expression data from multiple cell types, varying stimulation conditions and different family members will allow us to profile the series of molecular events that are responsible for the development of an immune response. These data can then provide a basis for functional gene-expression studies to allow for the specific therapeutic targeting of expression pathways that are dysregulated in disease states.

Amy S. Weinmann is at the Department of Immunology, University of Washington, Seattle, Washington 98195, USA.

e-mail: weinmann@u.washington.edu

doi:10.1038/nri1353

- Szabo, S. J., Sullivan, B. M., Peng, S. L. & Glimcher, L. H. Molecular mechanisms regulating T<sub>H</sub>1 immune responses. *Annu. Rev. Immunol.* **21**, 713–758 (2003).
- Glimcher, L. H. & Murphy, K. M. Lineage commitment in the immune system: the T helper lymphocyte grows up. *Genes Dev.* **14**, 1693–1711 (2000).
- Kondo, M., Scherer, D. C., King, A. G., Manz, M. G. & Weissman, I. L. Lymphocyte development from hematopoietic stem cells. *Curr. Opin. Genet. Dev.* **11**, 520–526 (2001).
- Ren, B. *et al.* Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306–2309 (2000).
- Iyer, V. R. *et al.* Genomic binding sites of the yeast cell-cycle transcription factor SBF and MBF. *Nature* **409**, 533–538 (2001).
- Ren, B. *et al.* E2F integrates cell cycle progression with DNA repair, replication, and G2/M checkpoints. *Genes Dev.* **16**, 245–256 (2002).
- Weinmann, A. S., Yan, P. S., Oberley, M. J., Huang, T. H.-M. & Farnham, P. J. Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. *Genes Dev.* **16**, 235–244 (2002).
- Weinmann, A. S., Bartley, S. M., Zhang, M. Q., Zhang, T. & Farnham, P. J. The use of chromatin immunoprecipitation to clone novel E2F target promoters. *Mol. Cell. Biol.* **21**, 6820–6832 (2001).
- Martone, R. *et al.* Distribution of NF- $\kappa$ B binding sites across human chromosome 22. *Proc. Natl Acad. Sci. USA* **100**, 12247–12252 (2003).
- Horak, C. E. *et al.* GATA-1 binding sites mapped in the  $\beta$ -globin locus by using mammalian chIP-chip analysis. *Proc. Natl Acad. Sci. USA* **99**, 2924–2929 (2002).
- Smale, S. T. & Fisher, A. G. Chromatin structure and gene regulation in the immune system. *Annu. Rev. Immunol.* **20**, 427–462 (2002).
- Szabo, S. J. *et al.* A novel transcription factor, T-bet, directs T<sub>H</sub>1 lineage commitment. *Cell* **100**, 655–669 (2000).
- Szabo, S. J. *et al.* Distinct effects of T-bet in T<sub>H</sub>1 lineage commitment and IFN $\gamma$  production in CD4 and CD8 T cells. *Science* **295**, 338–342 (2002).
- Lighvani, A. A. *et al.* T-bet is rapidly induced by interferon- $\gamma$  in lymphoid and myeloid cells. *Proc. Natl Acad. Sci. USA* **98**, 15137–15142 (2001).
- Lugo-Villarino, G., Maldonado-Lopez, R., Possemato, R., Penaranda, C. & Glimcher, L. H. T-bet is required for optimal production of IFN- $\gamma$  and antigen-specific T cell activation by dendritic cells. *Proc. Natl Acad. Sci. USA* **100**, 7749–7754 (2003).
- Liu, N., Ohnishi, N., Ni, L., Akira, S. & Bacon, K. B. CpG directly induces T-bet expression and inhibits IgG1 and IgE switching in B cells. *Nature Immunol.* **4**, 687–693 (2003).
- Pearce, E. L. *et al.* Control of effector CD8<sup>+</sup> T cell function by the transcription factor Eomesodermin. *Science* **302**, 1041–1043 (2003).
- Mullen, A. C. *et al.* Hlx is induced by and genetically interacts with T-bet to promote heritable T<sub>H</sub>1 gene induction. *Nature Immunol.* **3**, 652–658 (2002).
- Afkarin, M. *et al.* T-bet is a STAT1-induced regulator of IL-12R expression in naive CD4<sup>+</sup> T cells. *Nature Immunol.* **3**, 549–557 (2002).
- Usui, T., Nishikomori, R., Kitani, A. & Strober, W. GATA-3 suppresses T<sub>H</sub>1 development by downregulation of Stat4 and not through effects on IL-12R $\beta$ 2 chain or T-bet. *Immunity* **18**, 415–428 (2003).
- Peng, S. L., Szabo, S. J. & Glimcher, L. H. T-bet regulates IgG class switching and pathogenic autoantibody production. *Proc. Natl Acad. Sci. USA* **99**, 5545–5550 (2002).
- Sullivan, B. M., Juedes, A., Szabo, S. J., von Herrath, M. & Glimcher, L. H. Antigen-driven effector CD8<sup>+</sup> T cell function regulated by T-bet. *Proc. Natl Acad. Sci. USA* **100**, 15818–15823 (2003).
- Avni, O. *et al.* T<sub>H</sub> cell differentiation is accompanied by dynamic changes in histone acetylation of cytokine genes. *Nature Immunol.* **3**, 643–651 (2002).
- Hoffmann, A., Leung, T. H. & Baltimore, D. Genetic analysis of NF- $\kappa$ B/Rel transcription factors defines functional specificities. *EMBO J.* **22**, 5530–5539 (2003).
- Hug, B. A., Ahmed, N., Robbins, J. A. & Lazar, M. A. A chromatin immunoprecipitation screen reveals protein kinase C $\beta$  as a direct RUNX1 target gene. *J. Biol. Chem.* **279**, 825–830 (2004).
- Townsend, M. J. *et al.* T-bet regulates the terminal maturation and homeostasis of NK and V $\alpha$ 14i NKT cells. *Immunity* (in the press).
- Antequera, F. & Bird, A. CpG islands as genomic footprints of promoters that are associated with replication origins. *Current Biol.* **9**, R661–R667 (1999).
- Hannenhalli, S. & Levy, S. Promoter prediction in the human genome. *Bioinformatics* **17**, S90–S96 (2001).
- Ioshikhes, I. P. & Zhang, M. Q. Large-scale human promoter mapping using CpG islands. *Nature Genet.* **26**, 61–63 (2000).
- Wells, J., Graveel, C. R., Bartley, S. M., Madore, S. J. & Farnham, P. J. The identification of E2F1-specific genes. *Proc. Natl Acad. Sci. USA* **99**, 3890–3895 (2002).
- Oberley, M. J., Inman, D. R. & Farnham, P. J. E2F6 negatively regulates BRCA1 in human cancer cells without methylation of histone H3 on lysine 9. *J. Biol. Chem.* **278**, 42466–42476 (2003).
- Odum, D. T. *et al.* Control of pancreas and liver gene expression by HNF transcription factors. *Science* **303**, 1378–1381 (2004).
- Ballestar, E. *et al.* Methyl-CpG binding proteins identify novel sites of epigenetic inactivation in human cancer. *EMBO J.* **22**, 6335–6345 (2003).
- Mao, D. Y. L. *et al.* Analysis of Myc bound loci identified by CpG island arrays shows that Max is essential for Myc-dependent repression. *Current Biol.* **13**, 882–886 (2003).
- Li, Z. *et al.* A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. *Proc. Natl Acad. Sci. USA* **100**, 8164–8169 (2003).
- Cawley, S. *et al.* Unbiased mapping of transcription factor binding sites along human chromosome 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* **116**, 499–509 (2004).
- Stevens, T. A., Iacovoni, J. S., Edelman, D. B. & Meech, R. Identification of novel binding elements and gene targets for the homeodomain protein BARX2. *J. Biol. Chem.* **276**, 10104 (doi: 10.1074/jbc.M310259200).
- Jishage, M., Fujino, T., Yamazaki, Y., Kuroda, H. & Nakamura, T. Identification of target genes for EWS/ATF-1 chimeric transcription factor. *Oncogene* **22**, 41–49 (2003).
- de Belle, I., Wu, J. X., Sperandio, S., Mercola, D. & Adamson, E. D. *In vivo* cloning and characterization of a new growth suppressor protein TOE1 as a direct target gene of Egr1. *J. Biol. Chem.* **278**, 14306–14312 (2003).
- Greenbaum, S. & Zhuang, Y. Identification of E2A target genes in B lymphocyte development by using a gene tagging-based chromatin immunoprecipitation system. *Proc. Natl Acad. Sci. USA* **99**, 15030–15035 (2002).
- Ladenburger, E. M., Keller, C. & Knippers, R. Identification of a binding region for human origin recognition complex proteins 1 and 2 that coincides with an origin of DNA replication. *Mol. Cell. Biol.* **22**, 1036–1048 (2002).

## Acknowledgements

I thank S. Smale and P. Farnham for being very supportive mentors, and the members of their laboratories I worked so closely with in the past. I also thank K. Beima for helpful discussions and the Leukemia and Lymphoma Society for their generous support.

## Competing interests statement

The author declares that she has no competing financial interests.

## Online links

### DATABASES

The following terms in this article are linked online to: **LocusLink**: [http://www.ncbi.nlm.nih.gov/LocusLink/c-MAF|GATA2|GATA3|IFN- \$\gamma\$ |IL-4|RUNX1|STAT4|T-bet](http://www.ncbi.nlm.nih.gov/LocusLink/c-MAF|GATA2|GATA3|IFN-gamma|IL-4|RUNX1|STAT4|T-bet)

### FURTHER INFORMATION

Amy Weinmann's lab home page:

<http://depts.washington.edu/immunweb.faculty/profiles/weinmann.html>

Access to this interactive links box is free online.