PREDICTION OF REGULATORY TRANSCRIPTION FACTORS IN T HELPER CELL DIFFERENTIATION AND MAINTENANCE

YUE-HIEN LEE[1,2] yuehien@gmail.com
MANUELA BENARY[1] manuela.benary@cms.hu-berlin.de
RIA BAUMGRASS[2] baumgrass@drfz.de
HANSPETER HERZEL[1] h.herzel@biologie.hu-berlin.de

[1] *Institute for Theoretical Biology, Humboldt University of Berlin, Invalidenstr. 43, 10115 Berlin, Germany*
[2] *German Rheumatism Research Centre, Chariteplatz 1, 10117 Berlin, Germany*

Naive T-helper (Th) cells differentiate into distinct lineages including Th1, Th2, Th17 and regulatory T (Treg) cells. Each of these Th-lineages has specific functions in immune defense and T cell homeostasis. Th cell fate decisions and commitment are dependent on the kind and strength of T cell stimulation and the subsequent gene expression profiles.

Our analysis targeted the identification of new regulatory transcription factor binding sites (TFBSs) in the promoter regions of up- and down-regulated genes in Treg cell differentiation and lineage maintenance. For this approach we compared different gene groups from global gene expression studies with background models of randomly selected genes to identify significantly overrepresented TFBSs.

Results of our analysis suggest that Ets and IRF family members contribute to the regulation of the initial induction of Treg cells. Furthermore, we identified the overrepresented TFBS-pairs Runx-NFAT and GATA3-Foxp3 in Treg specific genes and Foxp3 dependent genes, respectively. Interestingly, previous studies have observed functional interactions of both TFBS-pairs in T cells. This study provides a starting point for further investigations to elucidate the transcriptional network in Treg cells.

Keywords: regulatory T cells; differentiation; promoter analysis; transcription factors.

1. Introduction

T-helper (Th) cells are the central players of adaptive immune responses in mammalians. There are different effector subsets of Th cells, such as Th1, Th2, and Th17 cells, as well as Treg cells. Each of these Th cell subsets has specific functions in pathogen defense and regulation of immune reaction. Dysregulation of T cell activation and differentiation can lead to autoimmune diseases and allergies. Therefore, it is important to understand how these processes are regulated. Transcription factors (TFs) and cytokines are crucial orchestrators of Th cell fate decision processes. Often, not only one master transcription factor, but a transcription factor network is essential for induction and maintenance of different Th-subpopulations.

Treg cells represent a unique population of Th cells that actively inhibit other

immune cells by different mechanisms. The lineage specifying transcription factor Foxp3 is a hallmark of the Treg cell lineage [13]. Ectopic overexpression of Foxp3 in resting Th cells successfully converts these into functional Treg cells [9]. Mutations or deletion of Foxp3 cause severe autoimmune disorders [6]. Peripheral naive Th cells can differentiate into Foxp3$^+$ Treg cells by simultaneous stimulation of the T cell receptor and the TGF-β receptor in vivo and ex vivo [13, 39]. In addition, TGF-β is necessary for the maintenance and in vivo expansion of peripheral Treg cells [15, 21]. However, the molecular mechanisms of TGF-β-dependent Treg cell generation and maintenance are largely unknown.

Few transcription factors have been reported to promote TGF-β-dependent de novo induction of Foxp3 expression. Among them are TIEG1 (also named Kfl10) [36], NFAT and SMAD3 [35]. Other transcription factors are supposed to act as negative regulators of Foxp3 expression, such as GATA3 [20] and STAT6 [33].

To discover more of such regulators of Foxp3 expression as well as Treg cell induction and maintenance, we analysed different gene groups for overrepresented binding sites of transcription factors and pairs of transcription factors.

2. Materials and Methods

Recent studies showed that many transcription factors bind symmetrically close to the TSS [3, 32]. Therefore, we decided to select the region -500 to +500 bp around the transcription start site (TSS) for each gene and call this region hereafter promoter region. Promoter regions were extracted from the Ensembl database [11]. All 610 vertebrate binding site motifs represented by positional weight matrices (PWMs) were taken from TRANSFAC 12.1 [24]. For prediction of single binding sites the TFBS perl module from Lenhard et al. [19] was used.

Selection of Gene Groups

To identify new regulating transcription factors in Treg cells we selected four gene groups (Table 1) from global gene expression studies and analyzed transcription factor binding sites in the genes within the depicted groups. Genes of the first two groups were extracted from our own global gene expression studies. The first gene group contains genes differentially expressed in naive T-helper cells after stimulation with anti-CD3/CD28 antibodies for three hours (Table 1, gene group STIM). This group comprises immediate early genes of Th cell activation. The second group constitutes genes which are differentially expressed after treatment of cells with low-dose TGF-β (0.2 ng/ml) and low-dose Cyclosporine A (CsA, 5 nM). These conditions promote the induction of Foxp3$^+$ regulatory Th cells. The third group contains Foxp3 dependent genes and the fourth group Treg specific genes. Both groups were extracted from four global gene expression and chromatin-immuno-precipitation (ChIP) studies [12, 25, 31, 40].

Table 1: Selected gene groups with number of genes in the group (# Genes) and the average GC-content of the promoter regions inside the group (GC %).

Gene group	Description	GC %	# Genes	Targets of our TFBS search
STIM	genes in Th cell activation	58.38	2871	regulatory TFs in stimulated naive Th cells
TGF-β /CsA	genes in Treg cell induction	56.86	101	regulatory TFs during initial Treg cell induction
TREG	Treg specific genes	57.12	543	regulatory TF pairs for Treg cell maintenance
FOXP3	Foxp3 dependent genes	57.6	601	regulatory TF pairs in Foxp3 dependent genes

Selection of Transcription Factors

To search for combinatorial control of TFs in Treg cells, we selected TFs with known regulatory functions in either T cell development or T cell function (Table 2).

Table 2: Selected transcription factors for combinatorial analyses.

Transcription factors	Function
AHR	activates differentiation of Treg and Th17 cells [26]
AP-1, NFAT, NF-κB	major TFs activated by T cell stimulation [30]
C/EBP	mediator of Il-17 signalling [29]
CREB, ATF	CREB/ATF binding sites are important for T cell receptor induced expression of Foxp3 [17]
EGR	EGR-2 and EGR-3 have negative effects on T cell activation [28]
Foxp3	master transcription factor in Treg cells [13]
GATA3	master transcription factor in Th2 cells [41]
p300, SMAD3, TGIF	components of TGF-β signalling pathway [23]
RAR-α/RXR-α	prevent the STAT6 inhibition of Foxp3 [33]
RORα	regulates the differentiation of Th17 cells [8]
RORγt	master transcription factor in Th17 cells [8]
Runx1	inhibits with Foxp3 and promotes with RORγt Th17 cell differentiation [38]
STAT3	main TF of Il-6 signalling pathway [8]
STAT5	regulates the Foxp3 expression directly [37]
TIEG	activates TGF-β-induced Foxp3 expression [36]

Determination of an Overrepresented Motif in a Gene Group

First, we generated an empirical distribution of predicted number of binding sites for the motif using randomly selected genes. Next, with this empirical distribution we calculated for each gene in the group the p-value to observe the predicted number of binding sites for the motif. P-values below 0.05 are considered as significant. Based on these p-values we applied two commonly used statistical hypothesis tests, Fisher's exact test and Z-test, to determine an overrepesentation of the motif in the group. For the Z-test the z-sores were calculated by comparing the sum of p-values of genes to a background model. We chose to use the sum of p-values in order to avoid single genes having to much influence. According to this approach, motifs with z-scores smaller than -3 are regarded as overrepresented in the group. We chose for Fisher's exact test the same significance level, corresponding to 0.0027.

Compensating GC-Rich Promoter Regions

Promoter regions of the selected gene groups are GC-rich (Figure 1). GC-rich promoters have also been observed in previous studies, e.g. by Bozek et al. [5] in clock-controlled genes. Potentially, GC-rich promoters have the ability to ease gene transcription, since CpG-island promoters are facilitating constitutive gene expression or rapid gene induction without requiring SWI/SNF nucleosome remodeling complexes [27].

If we quantify TFBS overrepresentation with our approach or with F-MATCH [16] using arbitrary mouse genes as a background model, we find many GC-rich motifs. This can be regarded as an artefact due to the high GC-content of our gene groups. Consequently, in this paper we apply GC-matched background models as in Bozek et al. [5].

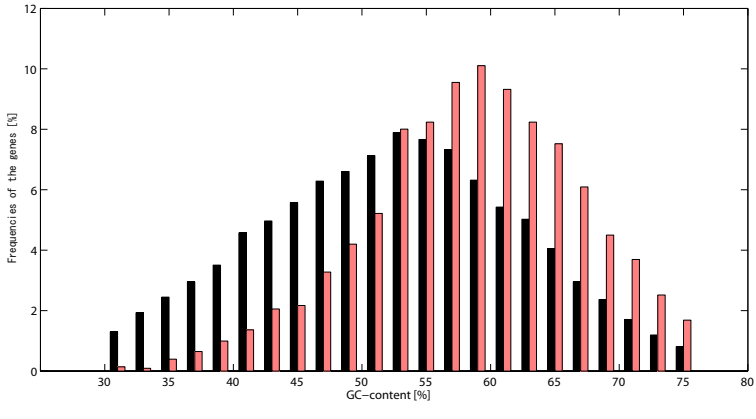


Fig. 1: GC-content distribution of analysed promoter regions (light) and promoter regions from the background model (black).

3. Results

Overrepresented Motifs in the Gene Group of Activated Naive Th Cells

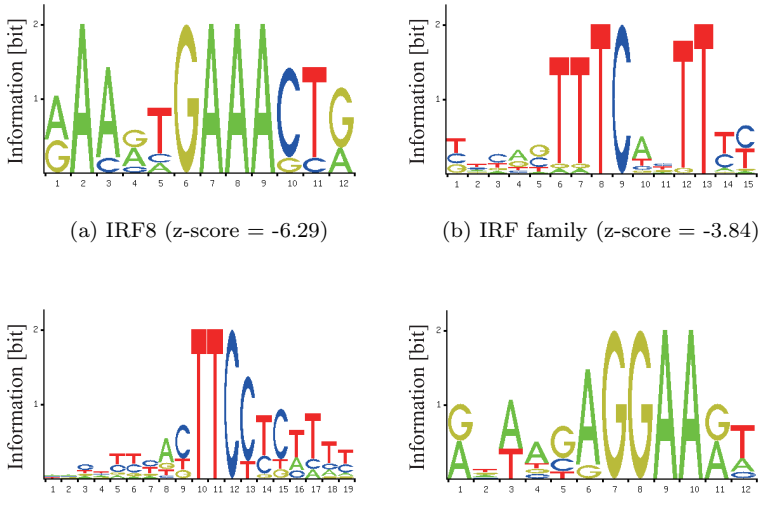
Overrepresented motifs were determined by Fisher's exact test and by the Z-test. By comparing results of Fisher's exact test to the results of the Z-test, we observed that all predictions by Fisher's exact test were obtained also by the Z-test. The lists contain overrepresented motifs with low and high GC-content. This indicates that we compensated the high GC-content successfully with our background model. We present in Table 3 overrepresented motifs detected by both tests. The transcription factors associated to the motifs point to links between T cell activation and cAMP and ERK signalling, cell cycle and differentiation. Interestingly, many of these transcription factors are also predicted to regulate clock-controlled genes [5].

Table 3: Overrepresented motifs in group STIM (p-values from Fisher's exact test). The comments are based on TRANSFAC and Gene Ontology [2].

Motif	TF	Comment	P-value
V$E2F_Q2	E2F family	cell cycle regulation	3.6271e-06
V$ELK1_02	Elk-1	ERK/MAPK target	7.9353e-06
V$NFY_01	NF-Y	regulates MHC II genes	2.1575e-05
V$NRF2_01	Nrf-2	stress response	2.5528e-05
V$CETS1P54_03	c-Ets-1	ERK/MAPK target	3.9109e-05
V$CREBATF_Q6	CREB, ATF	cAMP signalling	0.0001
V$NRF1_Q6	Nrf-1	induces MEF-2a	0.0002
V$HMGIY_Q3	HMG-I(Y)	lymphocyte differentiation	0.0003
V$ATF_01	ATF	cAMP signalling	0.0006
V$CDX_Q5	Cdx	cell differentiation	0.0013
V$CIZ_01	CIZ	myeloid cell differentiation	0.0016
V$MMEF2_Q6	MEF-2a	muscle specific	0.0016
V$CREB_02	CREB	cAMP signalling	0.0018
V$WHN_B	Foxn1	T cell proliferation	0.0019
V$AP2GAMMA_01	AP-2γ	retinoic acid target	0.0019
V$AHR_Q5	AHR	differentiation of T cells	0.0021
V$IRF_Q6_01	IRF family	differentiation of T cells	0.0022

Transcription Factors of the Ets Family and IRF Family Are Overrepresented in the Gene Group of Treg Cell Induction

We identified four motifs (IRF8, IRF family, PU.1 and Elf1) as overrepresented in the gene group TGF-β/CsA (see Figure 2). The binding site motif of IRF8 is quite similar to the one of IRF family. Note, that reverse complementary motifs represent the same binding site on the opposite DNA strand. The two similar motifs of PU.1 and Elf1 belong to the Ets family. In fact, it has been observed that PU.1 can bind to Elf1 binding sites [4]. PU.1 and Elf1 seem to differ from other Ets family members in their DNA binding specificity [4].



(a) IRF8 (z-score = -6.29)

(b) IRF family (z-score = -3.84)

(c) PU.1 (z-score = -4.16)

(d) Elf1 (z-score = -3.58)

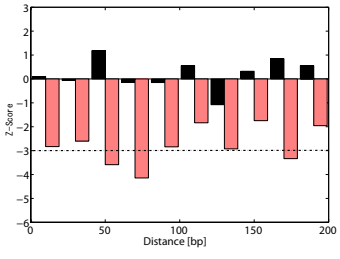
Fig. 2: Overrepresented motifs in the gene group TGF-β/CsA.

Combinatorial Analysis Revealed GATA3-Foxp3 and Runx-NFAT as Potential Regulatory Complexes

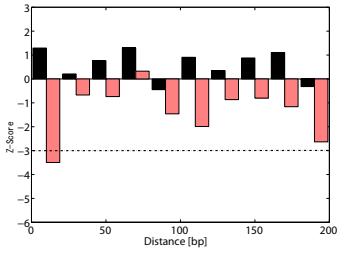
We used the transcription factors listed in Table 2 and the transcription factors found in the previous section. Altogether we had 25 motifs, 20 motifs from TRANSFAC, and 5 motifs based on sites from the literature such as Marson et al. [22]. Out of the 25 single motifs we created 325 motif pairs by combination. The distance between single binding sites in a motif pair was 1 to 15 bp in the initial screen.

The combinatorial analysis consisted of three steps. In the first step we searched in the gene groups TREG and FOXP3 for overrepresented motif pairs in the same

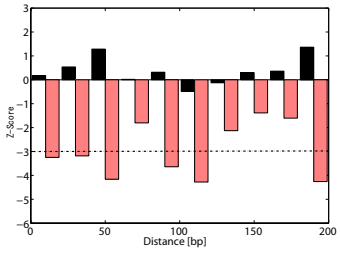
way as for the single motifs. We discovered 8 overrepresented motif pairs in the group TREG and 18 in the group FOXP3. In the next step the overrepresented motif pairs where filtered using a heuristic criterion to reduce the number of predictions. Only motif pairs with at least one motif having a z-score greater than zero are chosen for further analysis. Thus, we focused on motif pairs where at least one motif was not overrepresented. After this step we had 4 overrepresented motif pairs left. IRF8-STAT3, NFAT-STAT3 and GATA3-Foxp3 in the group FOXP3 and Runx1-NFAT in the group TREG. Since Runx1, Runx2 and Runx3 have almost identical DNA-binding domains [1], we refer from now on to Runx-NFAT. In the last step we conducted a distance analysis for the 4 remaining motif pairs (Figure 3). We assume that motif pairs with preferred short distances are particularly likely to interact. We divided the range of 0 to 200 bp into 20 bp intervals and repeated for each interval the overrepresentation analysis. As a result of the distance analysis only Runx-NFAT and GATA3-Foxp3 have a clear preference for short distances. Thus, our combinatorial analysis suggests that GATA3 and Foxp3 as well as Runx and NFAT interact in short distances with each other.



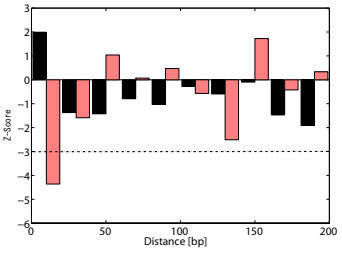
(a) NFAT-STAT3. Overrepresented in group FOXP3 (z-score = -3.35).



(b) GATA3-Foxp3. Overrepresented in group FOXP3 (z-score = -3.15).



(c) IRF8-STAT3. Overrepresented in group FOXP3 (z-score = -3.01).



(d) Runx-NFAT. Overrepresented in group TREG (z-score = -4.12).

Fig. 3: Distance analysis was applied for the overrepresented motif pairs in 20 bp intervals (red = target gene group, dark = control group). For NFAT-STAT3 and IRF8-STAT3 no distances were preferred over others, whereas Runx-NFAT and GATA3-Foxp3 have clear preferences for short distances.

4. Discussion

The transcriptional network regulating differentiation, function, and maintenance of Treg cells is largely unknown. In this study we used a bioinformatic approach to identify transcription factors involved in the regulation of these processes in Treg cells. For this purpose we analysed transcription factor binding sites in promoter regions to identify significantly overrepresented motifs in comparison to a background model. Importantly, in accordance with Bozek et al. [5], we compensated the high GC-content of the analysed promoter regions by using a GC-matched background model to reduce false positive results.

The analysis of immediate early genes in Treg cell induction suggests that transcription factors of the Ets family, PU.1 and Elf1, and of the IRF family are involved in Treg cell differentiation processes. Experimental data for IRF4, one IRF family member, support our results. It is already known that IRF4 is essential for the differentiation of Th17 cells [8]. These data do not rule out specific functions of other IRF family members in Treg cells. The role of Ets family transcription factors in peripheral Th cell differentiation is largely unknown. However, it has been reported that Elf1 regulates gene expression during T cell activation [34] and that transiently expressed PU.1 might inhibit Foxp3 during initial Th2 differentiation [10]. Interestingly, cell fate decisions of T cell progenitors in the thymus are controlled by PU.1 [14]. Therefore, an important role of PU.1 in peripheral Treg cell linage decision is possible, too.

Furthermore, we suppose from our results that the transcription factor pairs GATA3-Foxp3 and Runx-NFAT are important for Treg cells. Indeed, experimental data of Dardalhon et al. [7] showed already in Th2 cells that GATA3-Foxp3 interaction inhibited GATA3 mediated transactivation of Il-5. They suggested an inhibition of other GATA3 targets as well. Interestingly, there is already a hint that GATA3 and Foxp3 are transiently coexpressed during the second day of Treg cell induction [20]. Therefore, it would be interesting to investigate whether GATA3-Foxp3 interaction promotes Treg cell lineage decision by inhibition of Th2 cell development. This could be one mechanism how the master transcription factors help to shape Th cell lineage decisions.

Not only GATA3-Foxp3 but also Runx-NFAT was already described as an interaction pair in Th2 cells inhibiting IL-4 expression [18]. Our results suggest that transcription factors of the Runx family interact with NFAT to regulate Treg cell specific genes. Therefore, this interaction should be analysed in more detail in the context of Treg cell induction and maintenance.

Our bioinformatics analyses of global gene expression data revealed several new promising starting points for further investigations to elucidate the transcriptional networks regulating Treg cell differentiation and maintenance.

Acknowledgments

The work was supported by the IRTG and by a grant of the German Ministry of Education and Research within the Forsys Partner Network to H.H. and R.B..

References

[1] Alarcon-Riquelme, M.E., A RUNX trio with a taste for autoimmunity, *Nature Genetics*, 35:299–300, 2003.
[2] Ashburner, M., et al., Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, *Nature Genetics*, 25:25–29, 2000.
[3] Birney, E., et al., Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project, *Nature*, 447:799–816, 2007.
[4] Bockamp, E.O., et al., Transcriptional regulation of the stem cell leukemia gene by PU.1 and Elf-1, *J. Biol. Chem.*, 273:29032–29042, 1998.
[5] Bozek, K., et al., Regulation of clock-controlled genes in mammals, *PloS One*, 4:e4882, 2009.
[6] Brunkow, M.E., et al., Disruption of a new forkhead/winged-helix protein, scurfin, results in the fatal lymphoproliferative disorder of the scurfy mouse, *Nature Genetics*, 27:68–73, 2001.
[7] Dardalhon, V., et al., IL-4 inhibits TGF-beta-induced Foxp3+ T cells and, together with TGF-beta, generates IL-9+ IL-10+ Foxp3(-) effector T cells, *Nature Immunology*, 9:1347–1355, 2008.
[8] Dong, C., Th17 cells in development: an updated view of their molecular identity and genetic programming, *Nature Reviews Immunology*, 8:337–348, 2008.
[9] Fontenot, J.D., Gavin, M.A., Rudensky, A.Y., Foxp3 programs the development and function of CD4+CD25+ regulatory T cells, *Nature Immunology*, 4:330–336, 2003.
[10] Hadjur, S., et al., IL4 blockade of inducible regulatory T cell differentiation: the role of Th2 cells, Gata3 and PU.1, *Immunology Letters*, 122:37–43, 2009.
[11] Hammond, M.P., Birney, E., Genome information resources - developments at Ensembl, *Trends in Genetics: TIG*, 20:268–272, 2004.
[12] Hill, J.A., et al., Foxp3 transcription-factor-dependent and -independent regulation of the regulatory T cell transcriptional signature, *Immunity*, 27:786–800, 2007.
[13] Hori, S., Nomura, T., Sakaguchi, S., Control of regulatory T cell development by the transcription factor Foxp3, *Science*, 299:1057–1061, 2003.
[14] Huang, G., et al., PU.1 is a major downstream target of AML1 (RUNX1) in adult mouse hematopoiesis, *Nature Genetics*, 40:51–60, 2008.
[15] Huber, S., et al., Cutting edge: TGF-beta signaling is required for the in vivo expansion and immunosuppressive capacity of regulatory CD4+CD25+ T cells, *Immunology*, 173:6526–6531, 2004.
[16] Kel, A., et al., Beyond microarrays: Finding key transcription factors controlling signal transduction pathways, *BMC Bioinformatics*, 7 Suppl 2:S13, 2006.
[17] Kim, H., Leonard, W.J., CREB/ATF-dependent T cell receptor-induced FoxP3 gene expression: a role for DNA methylation, *Medicine*, 204:1543–1551, 2007.
[18] Lee, S.H., et al., Runx3 inhibits IL-4 production in T cells via physical interaction with NFAT, *Biochem. and Biophys. Research Communications*, 381:214–217, 2009.
[19] Lenhard, B., Wasserman, W.W., TFBS: computational framework for transcription factor binding site analysis, *Bioinformatics*, 18:1135–1136, 2002.
[20] Mantel, P., et al., GATA3-driven Th2 responses inhibit TGF-beta1-induced FOXP3 expression and the formation of regulatory T cells, *PLoS Biology*, 5:e329, 2007.

[21] Marie, J.C., et al., TGF-beta1 maintains suppressor function and Foxp3 expression in CD4+CD25+ regulatory T cells, *Medicine*, 201:1061–1067, 2005.
[22] Marson, A., et al., Foxp3 occupancy and regulation of key target genes during T-cell stimulation, *Nature*, 445:931–935, 2007.
[23] Massagu, J., Chen, Y.G., Controlling TGF-beta signaling, *Genes & Development*, 14:627–644, 2000.
[24] Matys, V., et al., TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes, *Nucleic Acids Research*, 34:D108–110, 2006.
[25] McHugh, R.S., et al., CD4(+)CD25(+) immunoregulatory T cells: gene expression analysis reveals a functional role for the glucocorticoid-induced TNF receptor, *Immunity*, 16:311–323, 2002.
[26] Quintana, F.J., et al., Control of Treg and Th17 cell differentiation by the aryl hydrocarbon receptor, *Nature*, 453:65–71, 2008.
[27] Ramirez-Carrozzi, V.R., et al., A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling, *Cell*, 138:114–128, 2009.
[28] Safford, M., et al., Egr-2 and Egr-3 are negative regulators of T cell activation, *Nature Immunology*, 6:472–480, 2005.
[29] Shen, F., et al., Identification of common transcriptional regulatory elements in interleukin-17 target genes, *J. Biol. Chem.*, 281:24138–24148, 2006.
[30] Smith-Garvin, J.E., Koretzky, G.A., Jordan, M.S., T cell activation, *Annual Review of Immunology*, 27:591–619, 2009.
[31] Sugimoto, N., et al., Foxp3-dependent and -independent molecules specific for CD25+CD4+ natural regulatory T cells revealed by DNA microarray analysis, *International Immunology*, 18:1197–1209, 2006.
[32] Suzuki, H., et al., The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line, *Nature Genetics*, 41:553–562, 2009.
[33] Takaki, H., et al., STAT6 inhibits TGF-beta1-mediated Foxp3 induction through direct binding to the Foxp3 promoter, which is reverted by retinoic acid receptor, *J. Biol. Chem.*, 283:14955–14962, 2008.
[34] Thompson, C.B., et al., Cis-acting sequences required for inducible interleukin-2 enhancer function bind a novel Ets-related protein, Elf-1, *Molecular and Cellular Biology*, 12:1043–1053, 1992.
[35] Tone, Y. et al., Smad3 and NFAT cooperate to induce Foxp3 expression through its enhancer, *Nature Immunology*, 9:194–202, 2008.
[36] Venuprasad, K., et al., The E3 ubiquitin ligase itch regulates expression of transcription factor Foxp3 and airway inflammation by enhancing the function of transcription factor TIEG1, *Nature Immunology*, 9:245–253, 2008.
[37] Yao, Z., et al., Nonredundant roles for Stat5a/b in directly regulating Foxp3, *Blood*, 109:4368–4375, 2007.
[38] Zhang, F., Meng, G., Strober, W., Interactions among the transcription factors Runx, RORgammat and Foxp3 regulate the differentiation of interleukin 17-producing T cells, *Nature Immunology*, 9:1297–1306, 2008.
[39] Zheng, S.G., et al., Generation ex vivo of TGF-beta-producing regulatory T cells from CD4+CD25- precursors, *Immunology* , 169:4183–4189, 2002.
[40] Zheng, Y., et al., Genome-wide analysis of Foxp3 target genes in developing and mature regulatory T cells, *Nature*, 445:936–940, 2007.
[41] Zhu, J., et al., GATA-3 promotes Th2 responses through three different mechanisms: induction of Th2 cytokine production, selective growth of Th2 cells and inhibition of

Th1 cell-specific factors, *Cell Research*, 16:3–10, 2006.