BMC
Genomics

An integrative computational systems biology approach identifies differentially regulated dynamic transcriptome signatures which drive the initiation of human T helper cell differentiation

Tarmo Äijö[1†], Sanna M Edelman[2†], Tapio Lönnberg[2], Antti Larjo[1,4], Henna Kallionpää[2,3], Soile Tuomela[2,3], Emilia Engström[2], Riitta Lahesmaa[2] and Harri Lähdesmäki[2,4*]

Abstract

Background: A proper balance between different T helper (Th) cell subsets is necessary for normal functioning of the adaptive immune system. Revealing key genes and pathways driving the differentiation to distinct Th cell lineages provides important insight into underlying molecular mechanisms and new opportunities for modulating the immune response. Previous computational methods to quantify and visualize kinetic differential expression data of three or more lineages to identify reciprocally regulated genes have relied on clustering approaches and regression methods which have time as a factor, but have lacked methods which explicitly model temporal behavior.

Results: We studied transcriptional dynamics of human umbilical cord blood T helper cells cultured in absence and presence of cytokines promoting Th1 or Th2 differentiation. To identify genes that exhibit distinct lineage commitment dynamics and are specific for initiating differentiation to different Th cell subsets, we developed a novel computational methodology (LIGAP) allowing integrative analysis and visualization of multiple lineages over whole time-course profiles. Applying LIGAP to time-course data from multiple Th cell lineages, we identified and experimentally validated several differentially regulated Th cell subset specific genes as well as reciprocally regulated genes. Combining differentially regulated transcriptional profiles with transcription factor binding site and pathway information, we identified previously known and new putative transcriptional mechanisms involved in Th cell subset differentiation. All differentially regulated genes among the lineages together with an implementation of LIGAP are provided as an open-source resource.

(Continued on next page)

* Correspondence: harri.lahdesmaki@aalto.fi
†Equal contributors
2Turku Centre for Biotechnology, University of Turku and Åbo Akademi University, Turku, Finland
4Department of Information and Computer Science, Aalto University, Helsinki, Finland
Full list of author information is available at the end of the article

(Continued from previous page)

Conclusions: The LIGAP method is widely applicable to quantify differential time-course dynamics of many types of datasets and generalizes to any number of conditions. It summarizes all the time-course measurements together with the associated uncertainty for visualization and manual assessment purposes. Here we identified novel human Th subset specific transcripts as well as regulatory mechanisms important for the initiation of the Th cell subset differentiation.

Keywords: Lineage commitment, Non-parametric analysis, Th cell differentiation, Time-course transcriptomics, Transcription factor binding

Background

T cells are key regulators of the adaptive immune system and have a central role in defense against pathogens and cancer as well as protection from autoimmune diseases. $CD4^+$ T lymphocytes can differentiate to functionally distinct effector subtypes, including T helper 1 (Th1), T helper 2 (Th2) and more recently described T helper 17 (Th17) cells [1]. Th1 cells secrete effector cytokine IFN-γ and regulate cell-mediated immunity and play a role in the pathogenesis of autoimmune diseases, such as multiple sclerosis. Th2 cells in turn produce IL-4, IL-5, and IL-13 cytokines, and mediate immunity against extracellular pathogens and allergic reactions. Th17 cells, characterized by the production of a proinflammatory cytokine IL-17, regulate inflammatory responses on the mucosal surfaces. For the overall health in humans and animals, the proper balance between different effector T cell types and T regulatory cells is crucial [2,3]. Aberrant activation of Th1 and Th17, or Th2 cells can trigger inflammatory autoimmune diseases as well as asthma and allergy. Previous studies utilizing genome-wide expression data and computational modeling have aimed at revealing the master regulators and regulatory networks within the differentiating Th1 and Th2 cells [4-9]. However, studies in human have been less extensive than in mouse due to the difficulty in collecting sufficient amount of samples to comprehensively profile T cell differentiation over time. In addition, lack of appropriate computational methods suitable for analyzing large-scale experimental data from multiple lineages over several time points spanning the lineage commitment process has limited the progress on revealing dynamics and molecular mechanisms underlying multiple lineage commitment.

A number of different time-series analysis approaches have been proposed to solve large-scale lineage commitment analysis problems. The general purpose *F*-test [10] can be used to test the difference between time-series data sets, but it does not extend to simultaneous comparison of multiple lineages and fails to take into account the correlation between the measurements at different time points. More recent approaches to analyze time-series data, including regression, differential expression, discriminant and clustering methods, are reviewed by Coffey and Hinde [11]. Methods for differential expression analysis include e.g. spline-based methods, generalized *F*-tests and hierarchical error and empirical Bayes models. Spline-based EDGE method by Storey *et al.* [12] is relevant for our problem because it provides comparisons for multiple conditions (lineage commitment profiles). Although EDGE computes a p-value for differential expression, it does not quantify the differential expression for all lineage comparisons, such as reciprocal genes (i.e., all lineages behave differently). ANOVA-based TANOVA method is based on the approach where different ANOVA structures are defined and the optimal one is found by evaluating the effects and significancies of the factors [13]. Recently, Stegle *et al.* [14] proposed an approach based on Gaussian processes (GP) to determine the time interval when a gene is differentially expressed. The methodology of Stegle *et al.* (2010) was limited to analyzing only two conditions. Moreover, it is often observed at transcriptional level that immediately after a treatment, such as activation of T cells by engagement of T cell receptor and CD28, genes are highly dynamic for some time but activity of gene expression decreases at later time points [15,16]. Thus, an ideal computational method – that does not exist at the moment – should take into account the temporal correlation, handle a non-uniform measurement grid, cope with non-stationary processes, and be able to do a well-defined analysis of multiple conditions.

Here we developed a computational methodology, LIGAP (Lineage commitment using Gaussian processes) which analyzes experimental data from any number of lineage commitment time-course profiles and analyzed genome-wide gene expression profiles of human umbilical cord blood T helper cells (Thp) activated through their CD3 and CD28 receptors and cultured in absence (Th0) or presence of cytokines promoting Th1 or Th2 differentiation. The results give insight into differences of the three lineages in the expression landscape and provide marker genes for lineage commitment identification. Key lineage specific, that is, differentially regulated, genes discovered computationally were validated either experimentally at protein level or based on the published literature. Using a module-based analysis, we identified known and putative regulatory control mechanisms by

overlaying highly coherent lineage profile clusters with genome-wide transcription factor (TF) binding predictions and pathway information. Consistent with the previously published results on IL-4/STAT6-mediated control of a large fraction of genes in Th2 program [17], our analysis revealed a comparable up-regulated and down-regulated modules, which are suggested to be controlled by STAT6 and other TFs. Interestingly, we also found that the genes which behave differently between all the lineages studied exhibit a consistent characteristic pattern, i.e., they are up-regulated in Th1 polarizing cells, down-regulated in Th2 polarizing cells, and in activated cells (Th0) the expression levels are between Th1 and Th2 cells. In addition, our analysis revealed a large set of novel genes, which are specific for different T cell subsets in human. All the gene expression data and differentially regulated genes as well as software implementing our computational analysis are made publicly available.
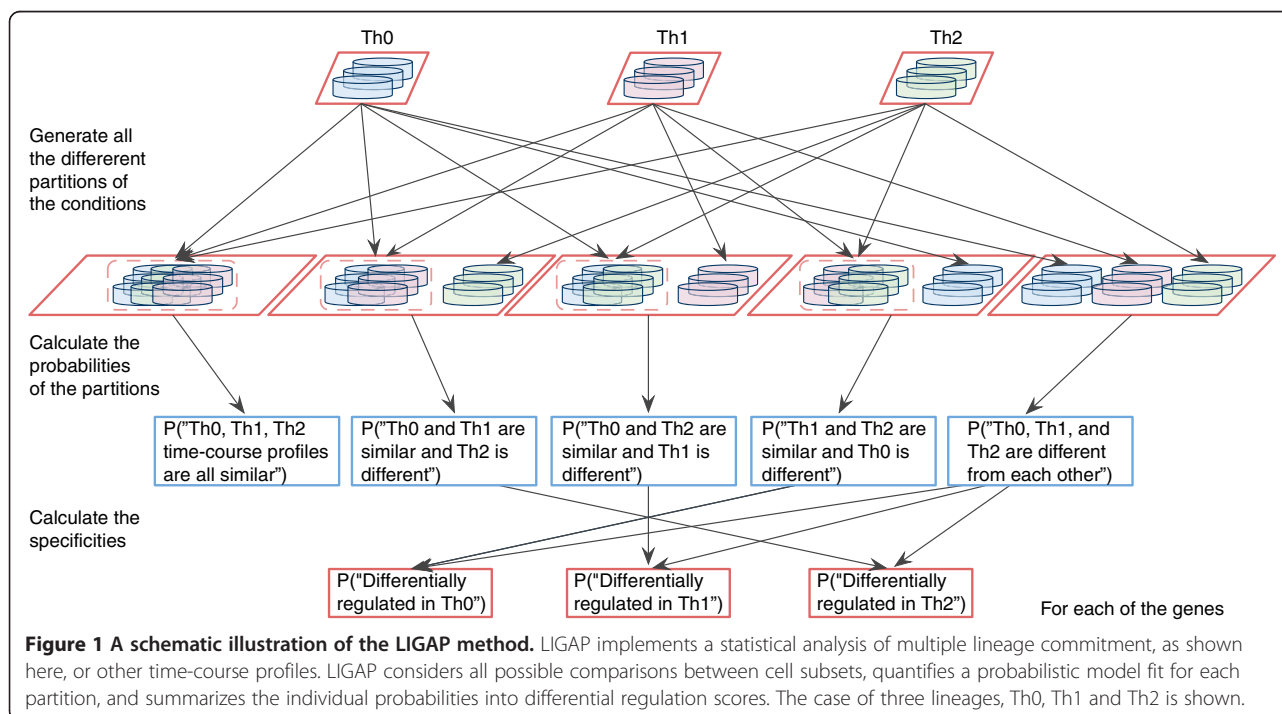
Results
Experimental data from primary human CD4+ T cells
We used previously published time-course gene expression measurements of activated primary human T cells (Th0) and cells polarized to differentiate to Th2 lineage [17] as well as previously unpublished data set representing Th1 polarizing cells originating from the same naïve Th precursor cells as the Th0 and Th2 cells. The gene expression of Th1 lineage was measured at time points 0, 12, 24, 48 and 72 hours. The measurements from Th0 and Th2 samples were available at the same time points.

LIGAP: A computational technique to identify condition specific time-course profiles
The discovery of condition specific genes at the level of gene expression is an important first step in systems biology studies. To capture temporal aspects of biological processes, such as cell differentiation, gene expression is commonly measured over time. We developed a novel model-based method LIGAP for detecting and visualizing changes between multiple lineage commitment time-course profiles. Briefly, for each gene at a time, our method carries out all comparisons between different cell subsets. In the case of Th0, Th1 and Th2 lineages, we assess all 5 alternatives; *(i)* "Th0, Th1, Th2 time-course profiles are all similar" (hypothesis H_1), *(ii)* "Th0 and Th1 are similar and Th2 is different" (hypothesis H_2), *(iii)* "Th0 and Th2 are similar and Th1 is different" (hypothesis H_3), *(iv)* "Th1 and Th2 are similar and Th0 is different" (hypothesis H_4), and *(v)* "Th0, Th1, and Th2 are all different from each other" (hypothesis H_5). LIGAP comparisons and quantifications are illustrated in Figure 1. The modeling is done using Gaussian processes, which provide a flexible and nonparametric approach for estimating smooth differentiation profiles. With the help of Bayesian statistics, we can quantify differences and similarities by assigning posterior probabilities for all the different profile comparisons between polarizing cell subsets. The problem can be seen as a model selection problem, where different comparisons are thought of as different model structures ($H_1, \ldots H_5$) and, given experimental lineage commitment profile data
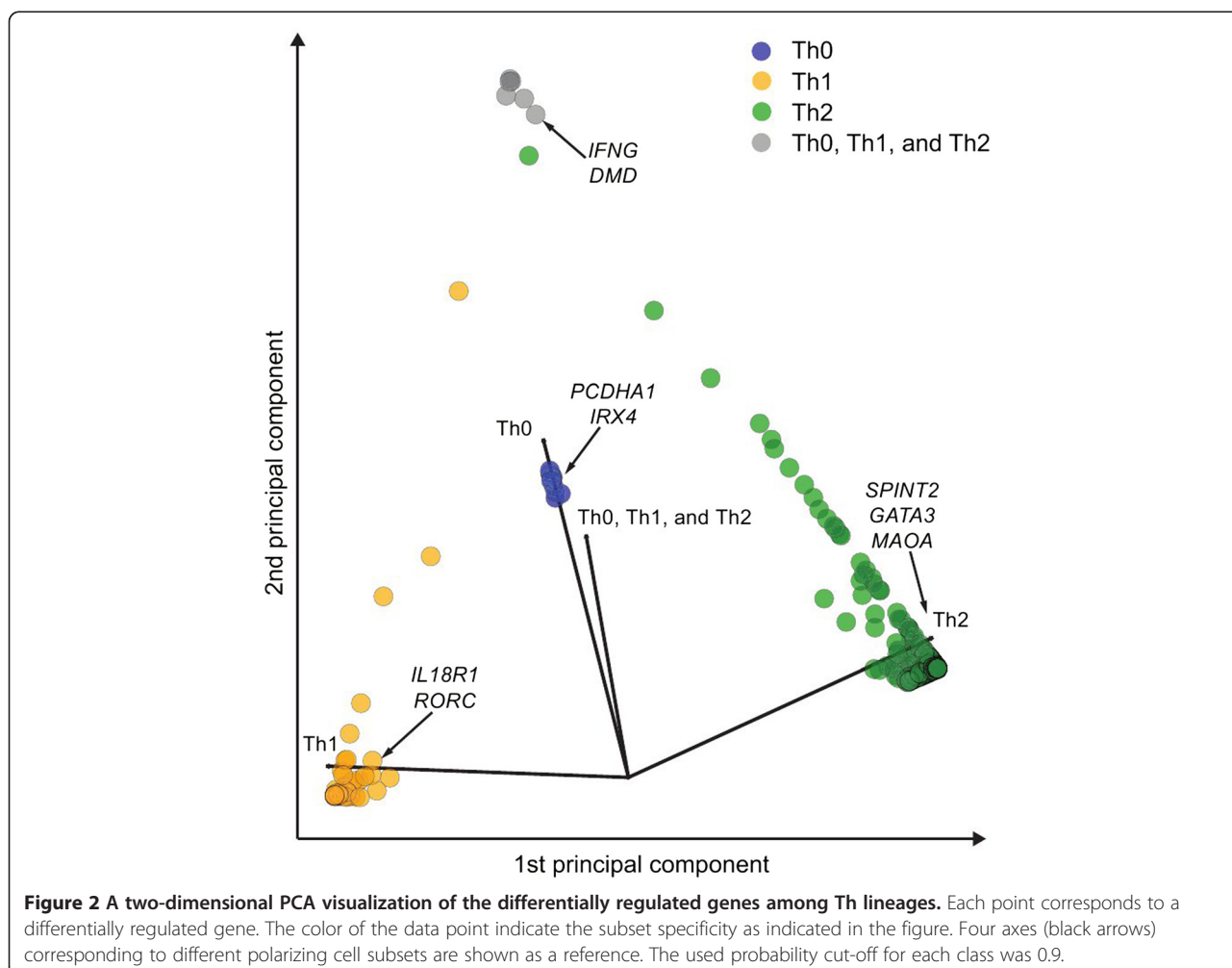

Figure 1 A schematic illustration of the LIGAP method. LIGAP implements a statistical analysis of multiple lineage commitment, as shown here, or other time-course profiles. LIGAP considers all possible comparisons between cell subsets, quantifies a probabilistic model fit for each partition, and summarizes the individual probabilities into differential regulation scores. The case of three lineages, Th0, Th1 and Th2 is shown.

D, the marginal likelihood $P(D \mid H_j)$, $j=1,\ldots,5$, is used to score different models. Using the Bayes' theorem, the marginal likelihoods can be converted into posterior probabilities of different hypothesis. These Bayesian model scores can be used further to quantify genes, which are specific for a certain lineage. For example, the probability of a gene being differentially regulated in Th2 lineage, i.e., score for Th2 is P("Gene is differentially regulated in Th2" | D) = P("Th0 and Th1 are similar and Th2 is different" | D) + P("Th0, Th1 and Th2 are all different" | D) = $P(H_2 \mid D) + P(H_5 \mid D)$. Genes which are differentially regulated in each of the conditions can be found by quantifying the probabilities P("Th0, Th1, and Th2 are all different from each other" | D) = $P(H_5 \mid D)$ or the three probabilities of differential regulation. Each score quantifies the amount of differential regulation, which refers to distinct temporal behavior from other lineages. The methodology generalizes to any number of lineages/conditions. Our method copes with non-uniform sampling, is able to model non-stationary biological processes (where e.g. changes are fast at the beginning of the experiment and slow at the end), can make comparisons for paired samples, and can carry out the analysis with different number of replicates and missing data. Importantly, the method affords comparison of more than two conditions of interest and is widely applicable to different experimental platforms.

LIGAP identifies signatures of Th0, Th1 and Th2 cell lineages

We analyzed the genome-wide gene expression time-course data from Th0, Th1 and Th2 lineages using LIGAP. For all genes, the method outputs the posterior probability values for each of the five hypotheses and also computes the scores for genes being differentially regulated in the Th subsets. An overview of the differentially regulated genes is shown in Figure 2, where the four-dimensional data points representing the condition specificities are projected into a plane using the principle component analysis (PCA). This demonstrates the convenience of the presented method as we are able to reduce highly complex data into a meaningful four-dimensional

Figure 2 A two-dimensional PCA visualization of the differentially regulated genes among Th lineages. Each point corresponds to a differentially regulated gene. The color of the data point indicate the subset specificity as indicated in the figure. Four axes (black arrows) corresponding to different polarizing cell subsets are shown as a reference. The used probability cut-off for each class was 0.9.

representation using a unified probabilistic framework. In Figure 2 individual points represent different genes and every gene is associated with four probabilities: P("Differentially regulated in Th0"), P("Differentially regulated in Th1"), P("Differentially regulated in Th2"), and P("Th0, Th1, and Th2 are all different from each other"). Note that IFNγ has the three probabilities P("Differentially regulated in Th0"), P("Differentially regulated in Th1"), and P("Differentially regulated in Th2") close to unity because the probability P("Th0, Th1, and Th2 are all different from each other") is close to unity. We set a criterion (P > 0.9) for the probabilities to call the differentially regulated probe sets; this threshold is in accordance with the Jeffrey's interpretation of "strong evidence" for the Bayes factor [18]. In addition, we required a minimum of two-fold change between a lineage and all other lineages at some time point during the differentiation for a gene to be called as differentially regulated. The top 49 and 50 gene symbols for Th1 and Th2 lineages, respectively, are listed in Table 1, whereas, the Th0 list includes only 18 genes. In a Additional file 1: Figure S1 are depicted two additional examples illustrating the advantage of considering temporal correlation in gene expression and thus improving the sensitivity of detecting consistent yet subtle changes. In addition, we repeated the analysis using EDGE [12] and TANOVA [13] methods using the default parameter values. TANOVA identified almost twice as many genes (~1,300) to be differentially regulated as LIGAP or TANOVA (~700). A comparison of the obtained ranked lists revealed a higher correspondence between the lists produced by LIGAP and EDGE than with the list produced by TANOVA (data not shown).

Our results of the Th subset specific genes agree well with known transcriptional changes during the human T cell differentiation. IFNγ, a hallmark molecule of Th1 lineage, was found to be one of the most significantly up-regulated Th1 specific transcripts (Table 1, Figure 3A, and Additional file 2: Table S1). Furthermore, *IL18R1* encoding the interleukin 18 receptor (IL18R), as well as IL-18 receptor accessory protein (IL18RAP) were among the top Th1 specific genes (Table 1, Figure 3B). Expression of IL18R is up-regulated specifically on Th1 cells but not on Th2 cells, thus, IL18R can be regarded as a differentiation marker for Th1 cells [15,19]. In fact, IL-12 and IL-18 can reciprocally up-regulate expression of each other's receptors in Th1 cells [15,20] and the IL-18 - IL18R system has a significant role in the synergistic effect of IL-12 and IL-18 in triggering efficient NF-κB signaling and enhancement of IFNγ production from human Th1 cells [21]. Intriguingly, in the absence of IL-12, IL-18 has also potential to induce Th2 differentiation and cytokine response [19,22]. The basic helix-loop-helix transcription repressor TWIST1 is also known to be expressed in Th1 cells in IL-12/STAT4, NF-κB and NFAT dependent way and its

role has been proposed to be linked to autoregulation of inflammatory cytokine production e.g. IFNγ [23]. Several studies have shown that CXCR6 is predominantly expressed in Th1 cells [24,25] and, inversely, in Th2 prone allergic conditions the expression of CXCR6 was reduced in allergic patients when compared to healthy individuals [26]. Also, an important Th1 linked function has been observed with MAP3K8 as it acts as an upstream activator of ERK via IL-12 and TCR-dependent signaling, promotes expression of T-bet and STAT4, and is actually a STAT4 target itself forming a feedback loop in the Th1 cells [27]. Deficiency in MAP3K8 leads to decreased IFNγ production in T cells and *in vivo* impaired host defense against *Toxoplasma gondii* [27].

Interestingly, the retinoic acid-related orphan receptor gamma (*RORC*) gene encoding RORγt, the key transcription factor in the differentiation program of Th17 cells, was also identified as a Th1 specific gene by the LIGAP analysis (Table 1) as its expression was up-regulated at 48 h time point (Figure 3C). In human, small numbers of T cells producing both IL-17 and IFNγ have been characterized in peripheral blood, in lamina propria of patients with Crohn's disease as well as in patients with psoriasis [28-30], but currently is it not known how such cells are derived from the naïve precursor cells. Other novel Th1 specific hits identified by the LIGAP include two cytoskeleton associated protein-coding genes dystrophin (*DMD*) and palladin (*PALLD*). *DMD* encodes actin-binding cytoskeletal structure molecule, which has been mostly studied in patients with Duchenne's muscular dystrophy [31]. These patients develop dystrophin specific autoreactive T cells [31], however, the biological role for dystrophin or palladin in differentiating Th cells is not known. Other genes novel in this context and putatively important for Th1 cell differentiation and/or function include *METRNL*, (meteorin, glial cell differentiation regulator-like), associated with rare cases of Mild ring 17 syndrome [32], *GLUL* encoding a glutamine synthetase, and associated with neuronal disorders and atherosclerotic carotid plaques [33,34], *MCTP2* (multiple C2 domains, trans membrane 2), *BBS12* (Bardet-Biedl syndrome 12), *STAG3* (stromal antigen 3), a meiotic gene, as well as *PGAP1* (post-GPI attachment to proteins 1). *NAPSB* coding for aspartic protease Napsin B is known to be expressed in human spleen and peripheral blood leucocytes, however, it is estimated to be only a transcribed pseudogene [35,36]. Similarly, *MIAT* (myocardial infarction associated transcript) is a non-protein coding gene [37], and the relevance of these transcripts in T cell differentiation is not understood, yet.

The top LIGAP hits of Th2 specific genes included several genes with very high probability values (Table 1, and Additional file 2: Table S1) and include a vast number of genes that are both specifically up-regulated and

Table 1 Differentially expressed genes in T cells polarized towards the Th0, Th1 and Th2 subsets

Top 18 Th0 specific genes			Top 49 Th1 specific genes			Top 50 Th2 specific genes		
Affymetrix probe ID	Gene symbol	P("Th0 specific")	Affymetrix probe ID	Gene symbol	P("Th1 specific")	Affymetrix probe ID	Gene symbol	P("Th2 specific")
203881_s_at	DMD(<2)	0.99967984	206618_at	IL18R1(+)	0.999997088	204388_s_at	MAOA(+)	1
223435_s_at	PCDHA1(<2)	0.999832654	203881_s_at	DMD(+)	0.999993101	227006_at	PPP1R14A(+)	1
205027_s_at	MAP3K8(<2)	0.996290375	210354_at	IFNG(+)	0.999987382	205419_at	GPR183(+)	1
228055_at	NAPSB(<2)	0.996196899	207072_at	IL18RAP(+)	0.9999811558	215172_at	PTPN20A(+)	1
220225_at	IRX4(<2)	0.991324709	213943_at	TWIST1(+)	0.999952924	205769_AT	SLC27A2(-)	1
205794_s_at	NOVA1(<2)	0.984972756	222547_at	MAP4K4(<2)	0.999933552	218976_at	DNAJC12(-)	1
221035_s_at	TEX14(<2)	0.981445323	242809_at	IL1RL1(<2)	0.999876515	208510_s_at	PPARG(+)	1
210354_at	IFNG	0.977204551	206007_at	PRG4(<2)	0.999828413	228708_at	RAB27B(+)	1
212012_s_at	KPNA6(<2)	0.976438534	228055_at	NAPSB(+)	0.99973265	45288_at	ABHD6(+)	1
219265_at	MOBKL2B(<2)	0.974857762	235458_at	HAVCR2(+)	0.99959007	229764_at	TPGR1(+)	1
212992_at	AHNAK2(<2)	0.971558242	225955_at	METRIL(+)	0.99954195	235199_at	RNF125(+)	1
223727_at	KCNIP2(<2)	0.968952518	206974_at	CXCR6(+)	0.999495469	203153_at	IFIT1(-)	1
200907_s_at	PALLD(<2)	0.967031439	206785_s_at	KLRC2(<2)	0.99882105	203097_s_at	RAPGEF2(-)	1
1570169_at	CSMD2(<2)	0.950549266	200648_s_at	GLUL(+)	0.997745987	208891_at	DUSP6(+)	1
200906_s_at	PALLD(<2)	0.940149574	205027_s_at	MAP3K8(+)	0.996774094	223159_s_at	NEK6(+)	1
216341_s_at	GNRHR(<2)	0.933169413	225142_at	JHDM1D(<2)	0.996653927	210715_s_at	SPINT2(+)	1
222890_at	CCDC113(<2)	0.921542966	215672_s_at	AHCYL2(<2)	0.996237492	208158_s_at	OSBPL1A(+)	1
201283_s_at	TRAK1(<2)	0.914512473	219383_at	PRR5L(<2)	0.995895316	225752_at	NIPA1(+)	1
			220603_s_at	MCTP2(+)	0.993459782	206638_at	HTR2B(+)	1
			239533_at	GPR155(<2)	0.992909201	205579_at	HRH1(+)	1
			229603_at	BBS12(+)	0.989517632	226508_s_at	TNFSF13B(-)	0.999999999
			237559_at	GPR55(<2)	0.988653639	244413_at	CLECL1(+)	0.999999999
			204284_at	PPP1R3C(<2)	0.988109742	203708_at	PDE4B(-)	0.999999999
			202625_at	LYN(<2)	0.986582272	227438_at	ALPK1(-)	0.999999998
			223767_at	GPR84(<2)	0.984948052	210762_s_at	DLC1(+)	0.999999998
			209348_s_at	MAF(+)	0.983885642	235570_at	RBMS3(+)	0.999999997
			210448_s_at	P2RX5(+)	0.9830704	212077_at	CALD1(+)	0.999999997
			228057_at	DDIT4L(<2)	0.978509028	235301_at	KIAA1324L(+)	0.999999997
			203129_s_at	KIF5C(<2)	0.978078282	209576_at	GNAI1(+)	0.999999997
			1554190_s_at	C10orf81(<2)	0.976976156	1554878_a_at	ABCD3(+)	0.999999996
			228806_at	RORC(+)	0.976176574	205569_at	LAMP3(+)	0.999999996
			227568_at	HECTD2(<2)	0.975089149	205900_at	KRT1(+)	0.999999996
			219753_at	STAG3(+)	0.972731568	210145_at	PLA2G4A(+)	0.999999996
			223374_s_at	B3GALNT1(<2)	0.971777376	227529_s_at	AKAP12(+)	0.999999995
			205098_at	CCR1(<2)	0.969847411	209602_s_at	GATA3(+)	0.999999995
			200907_s_at	PALLD(+)	0.967031845	225566_at	NRP2(+)	0.999999995
			211122_s_at	CXCL11(<2)	0.965307014	1552807_a_at	SIGLEC10(+)	0.999999995
			227697_at	SOCS3(+)	0.965003456	220307_at	CD244(<2)	0.999999994
			212683_at	SLC23A44(<2)	0.960505919	239151_at	BMS1P1(+)	0.999999994
			213572_s_at	SERPINB1(<2)	0.953874496	221241_s_at	BCL214(-)	0.999999993
			244321_at	PGAP1(+)	0.953326576	229625_at	GBP5(-)	0.999999992
			217127_at	CTH(+)	0.947970476	204912_at	IL10RA(+)	0.999999999

Table 1 Differentially expressed genes in T cells polarized towards the Th0, Th1 and Th2 subsets *(Continued)*

225962_at	ZNRF1(+)	0.944840091	214974_x_at	CXCL5(-)	0.999999989
219532_at	ELOVL4(<2)	0.937454137	221971_x_at	CTGLF10P(+)	0.999999989
222838_at	SLAMF7(+)	0.934337899	203853_s_at	GAB2(+)	0.999999989
228658	MIAT(+)	0.934243107	222457_s_at	LIMA1(+)	0.999999986
206948_s_at	PMCH(<2)	0.928324396	213385_at	CHN2(+)	0.999999983
232030_at	KIAA1632(<2)	0.914780073	205630_at	CRH(+)	0.999999979
213596_at	CASP4(<2)	0.907680384	207861_at	CCL22(+)	0.999999979
			219209_at	IFIH1(-)	0.999999977

Probe sets that fulfill the two-fold change criterion are marked based on the direction of the expression (+ denotes up-regulation and - denotes down-regulation) in the given condition. For example, *IFNγ* expression is enhanced in Th1 compared to Th0 and Th2, whereas expression of *SLC27A2* is decreased in Th2 compared to Th0 and Th1. In addition, probe sets that do not fulfill the fold change criterion are marked in the lists with "<2". All genes from Th0 and Th1 conditions as well as the top 50 of the Th2 specific genes are shown.

down-regulated in Th2 conditions compared to other Th subsets. Therefore, the list of Th2 specific genes with highest probability is consistent with the previously published results obtained with other computational methods [17]. Importantly, GATA3, the well characterized master transcription regulator of Th2 polarization [38] was identified among the top Th2 hits (Table 1). The transcriptional expression profile of GATA3 was observed to be

Figure 3 A detailed visualization of six differentially regulated genes. X-axis corresponds to time and y-axis shows the normalized log₂-transformed gene expression values. Solid line denotes the estimated mean expression profile and the shaded area shows the 95% confidence interval. Measurements are shown with individual points. Different lineages are color-coded as shown in the figure. (**A**) *IFNG*, (**B**) *IL18R1*, (**C**) *RORC*, (**D**) *GATA3*, (**E**) *MAOA*, (**F**) *SPINT2*, (**G**) *PPP1R14A*, and (**H**) *DUSP6*.

highly up-regulated at all time points among the cells cultured in Th2 polarizing conditions, whereas the expression profiles in Th0 and Th1 cells exhibited down-regulation (Figure 3D). In addition to well-known subset signature molecules, the analysis identified also a number of poorly characterized molecules in relation to their function in polarized Th cells. Among the highly expressed top 50 Th2 hits, the specificity of these transcripts relative to Th0, but not to Th1, has already been identified at different time points with the standard LIMMA methods (Smyth, 2005) in the past [17]. One of these Th2 specific top hits was *MAOA*, a gene encoding monoamine oxidase A, whose expression was increasingly up-regulated during the time course (Table 1, Figure 3E). This enzyme degrades amine neurotransmitters, (e.g. dopamine, norepinephrine, and serotonin) and was previously found to be up-regulated in human peripheral blood monocytes after IL-4 and IL-13 stimulation [39] as well as in Th2 cells derived from cord blood naïve CD4+ T cells and, importantly, being indirectly controlled by STAT6 [15,17]. It has been hypothesized that MAOA may act as an anti-inflammatory mediator by degrading serotonin which inhibits generation of TNFα from macrophages and up-regulates phagocytosis [40]. The biological significance of MAOA in Th2 cells, however, remains to be elucidated. Another interesting Th2 specific top hit was *SPINT2* (Table 1, Figure 3F) encoding a transmembrane serine

peptidase inhibitor Kunitz type 2 (also called HAI-2 and placental bikunin). SPINT2 was originally named after its homology to hepatocyte growth factor activator inhibitor 1 and its first isolation from human placenta [41,42]. The Kunitz inhibitory domains display potent inhibitory activity towards several trypsin-like serine proteases [43] and mutations in the human *SPINT2* gene cause a broad spectrum of abnormalities in organogenesis [44]. In addition, SPINT2 may function as a tumor suppressor gene, as its mRNA levels are down-regulated in several human cancers (e.g. gliomas, colorectal cancers and liver cancer) and a deficiency in SPINT2 expression is linked with poor prognosis of breast cancer [45]. There are no previous studies where the possible functional role of SPINT2 in human lymphocytes is unraveled, however, *SPINT2* was recently found to be a STAT6 target in human macrophages as well as in human Th2 cells [17,46]. We, hence, chose to experimentally validate the specificity of SPINT2 in primary human Th2-polarizing cells. We tested the specificity of SPINT2 expression at protein level on the cell surface of the Th cells with flow cytometry. At 24 hours after activation and induction of polarization the Th2 cells were found to express significantly more SPINT2 than the Th1 polarizing cells or the activated Th0 cells (Figure 4A). As some of the human SPINT2 transcripts do not harbor the coding signal for the transmembrane domain [47], we therefore also investigated if SPINT2 would be secreted
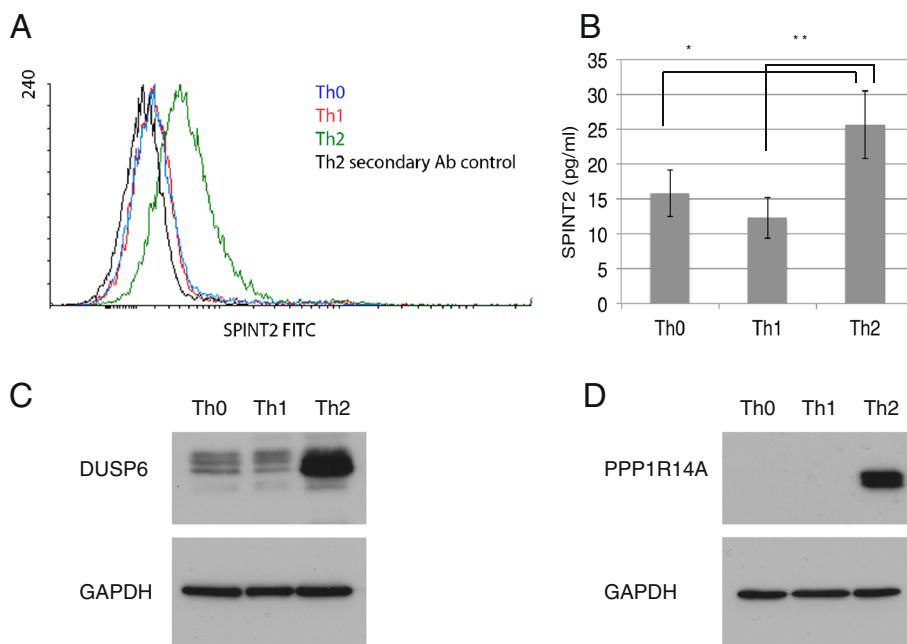
Figure 4 Experimental validation of characteristic expression of SPINT2, DUSP6 and PPP1R14A in primary human T helper 2 cells. (**A**) A histogram overlay showing SPINT2 expression at protein level on the cell surface of the Th cells measured with flow cytometer at 24 hours after activation. (**B**) SPINT2 secretion from different T helper cells measured with ELISA at 48 hours after activation. * P<0.05, ** P<0.01. A western-blot images showing (**C**) DUSP6 and (**D**) PPP1R14A expression on Th cells at 72 h post activation. Representatives of three biological replicates are presented. A house keeping protein GAPDH is shown as a loading control.

from the Th subsets. The SPINT2 concentrations were measured from the culture supernatants by enzyme-linked immunosorbent assay (ELISA) at 48 hours after activation and polarization, and the Th2 cells were observed to secrete significantly more SPINT2 than Th0 or Th1 cells (Figure 4B). The Th2 specific hits included also PPP1R14A, a phosphorylation-dependent inhibitor of smooth muscle myosin phosphatase, involved in regulation of smooth muscle contraction [48] as well as DUSP6 (dual specificity phosphatase 6), responsible for dephosphorylation of ERK1/2 [49]. Recently, IL-4 induced RNA expression of signaling molecules PPP1R14A and DUSP6 have been reported [15,17,50]. As the regulation of phosphorylation of the signaling intermediates is known to be highly important in defining the cell differentiation, we wanted to experimentally validate the subset specific expression of these two signaling molecules at protein level. We detected a clear Th2 specific PPP1R14A and DUSP6 protein expression at 72 hours time point post activation and initiation of the polarization, and very little or no expression in Th0 and Th1 lineages (Figures 4C and 4D).

Reciprocal regulators of lineage commitment

In context of determination of T cell subset identity, a key group of genes is the one where the expression kinetics differ between all the lineages. The list of these significantly different genes is shown in Table 2. An illustrative example gene from this list is the well-known Th1 signature cytokine gene IFNG (Figure 3A) as well as TBX21 encoding T-bet, a hallmark transcription factor in Th1 differentiated cells, both of which are also known to suppress Th2 activity [51]. In addition, MAP3K8, FAS, IL12RB2, and IL-26, have been identified to play role in Th1 polarized cells (cf. Table 2). Moreover, Table 2 and Additional file 2: Table S1 contain numerous differentially regulated transcripts which are only poorly characterized or their role in CD4+ Th cells has not been studied. The novel Th1 specific genes DMD and PALLD, encoding cytoskeletal associated proteins dystrophin and palladin, fall into the reciprocally regulated genes in the Th subsets studied here. Also, Th1 specific putative pseudogene NAPSB and non-coding transcript MIAT show reciprocal transcript profiles. Other novel genes include PRR5L, which has been identified to interact with a highly conserved protein kinase TOR (target of rapamycin), a central controller of cell growth and apoptosis [52]. OSBPL10 encodes oxysterol binding protein-like 10, an intracellular lipid receptor that regulates cellular lipid metabolism [53]. P2RY14 (purinergic receptor P2Y, G-protein coupled, 14) is a membrane receptor for UDP-glucose and plays a role in immune responses in human airway as well as female reproductive track epithelial cells by stimulating cytokine and chemokine production and recruitment of neutrophils [54-56]. P2RY14 has also been

identified to function in mouse splenic T cells as a regulator of IL-2 induced proliferation, however, no specific link to Th1 cells has been observed [57]. Also, the significance of ATP9A (ATPase, class II, type 9A), LPAR3 (lysophosphatidic acid receptor 3) functioning in G-protein coupled receptor signaling, XRN1 (5'-3' exoribonuclease 1), BSPRY (B-box and SPRY domain containing), MCTP2 (multiple C2 domains, transmembrane 2) or PTPRO (protein tyrosine phosphatase, receptor type, O) in Th1 cells is yet to be studied. Recent data indicate that in B cells, PTPRO dephosphorylates Syk, a kinase that is critical in signal transduction of B-cell receptor [58].

The Th2 up-regulated genes, PDE7B, SETBP1, C9orf135, TPRG1, IGSF3, or PPP1R14A have not been linked to CD4+ Th cell function, although their IL-4 mediated up-regulation has been published, and furthermore, SETBP1, TPRG1 and PPP1R14A have been identified as direct targets of STAT6 [17]. Interestingly, we observed that most of the genes whose expression differs between all the three lineages behave in a similar manner, i.e., they are up-regulated in Th1 and down-regulated in Th2.

Among the reciprocally regulated genes we found 34 genes up-regulated in Th1 condition and only six genes behaved in the opposite manner. The hierarchical clustering of the kinetic profiles is depicted in Figure 5A. This suggests that there are common mechanisms that induce reverse regulatory behavior. For example, the genes up-regulated in Th1 condition might be controlled downstream of IFNγ. This hypothesis is supported by the clear similarity between the profiles of IFNγ and the profiles of the clustered genes. We prepared a similar figure showing the differences in the kinetics of all the LIGAP identified genes. These results are depicted in Figure 5B and they show the similarity between the Th0 and Th1 lineages and their dissimilarity between the Th2 lineage.

Transcription factor binding sites in Th2 lineage

To extend our transcriptional analysis into transcriptional regulation, we decided to systematically analyze both genome-wide transcription factor (TF) binding site predictions made in silico and comprehensive literature-derived information about target genes of selected TFs. First, we predicted which of the transcription factors have binding sites in the RefSeq gene promoters (defined as [−1000,500] bp around TSS) using the ProbTF tool [76] combined with an empirical p-value computation. We focused on genes that were identified by the previous LIGAP analysis and considered all transcription factors that had known binding specificities (position specific frequency matrices, PSFMs) in TRANSFAC [77] (version 2009.3). We did not restrict our analysis only to those TFs whose transcripts are differentially expressed because, e.g., STAT6 is not differentially expressed

Table 2 The genes whose expression time-courses differ between all the lineages

Affymetrix probe ID	Gene symbol	Functional annotation*	Known characteristics in CD4+ T helper cells
203881_s_at	DMD(+)	other	NR
205027_s_at	MAP3K8(+)	kinase	required for proper IFNg production [27]
228055_at	NAPSB(+)	unknown	NR
210354_at	IFNG(+)	cytokine	positive regulation on T cell proliferation, tyrosine phosphorylation of STAT1 and production of IL-12 [59]
200907_s_at	PALLD(+)	other	NR
222838_at	SLAMF7(+)	other	expression upregulated in Th1 cells[16]
237322_at	MIAT(+)	other	NR
1555486_a_at	PRR5L(+)	other	NR
209369_at	ANXA3(+)	enzyme	expression upregulated in Th1 cells [15]
230109_at	PDE7B(-)	enzyme	expression upregulated in Th2 cells [17]
205933_at	SETBP1(-)	other	expression upregulated in Th2 cells [17]
216252_x_at	FAS(+)	transmembrane receptor	expression regulated by IFNg, important for T cell activation [60,61]
219073_s_at	OSBPL10(+)	other	NR
221271_at	IL21(<2)	cytokine	important for Th17 and follicular Th cell differentiation, increases activation of STAT3 [62-64]
223475_at	CRISPLD1(+)	other	expression upregulated in Th1 cells [16]
206999_at	IL12RB2(+)	transmembrane receptor	expression regulated by IFNg, plays a central role in Th1 differentiation [15,65]
200878_at	EPAS1(+)	transcription regulator	expression regulated by STAT6 in Th2 cells [17]
206974_at	CXCR6(+)	G-protein coupled receptor	predominantly expressed in Th1 cells [24,66,67]
206637_at	P2RY14(+)	G-protein coupled 1receptor	NR
212062_at	ATP9A(+)	transporter	NR
210029_at	IDO1(+)	enzyme	NR
236519_at	C9orf135(-)	other	expression upregulated in Th2 cells [17]
214038_at	CCL8(+)	cytokine	selectively expressed on Th2 cells, important for trafficking of Th2 cells, required for allergic immune response by Th2 cells [68]
229764_at	TPRG1(-)	other	expression upregulated in Th2 cells [17]
231192_at	LPAR3(+)	G-protein coupled receptor	NR
202421_at	IGSF3(-)	other	expression upregulated in Th2 cells [17]
221111_at	IL26(<2)	cytokine	expressed on human Th1 cells, increases activation if human STAT1 and STAT3 [69,70]
227006_at	PPP1R14A(-)	other	expression upregulated in Th2 cells [17]
1555785_a_at	XRN1(+)	enzyme	NR
230110_at	MCOLN2(+)	ion channel	expression upregulated in Th1 cells [16]
222746_s_at	BSPRY(+)	other	NR
220684_at	TBX21(+)	transcription regulator	expression up-regulated by IL-12, regulates the production of Th1 hallmark cytokine IFNg [71]
220603_s_at	MCTP2(+)	other	NR
210839_s_at	ENPP2(+)	enzyme	used in homing to secondary lymphoid organs [72]
210164_at	GZMB(+)	peptidase	important for TCR-induced cell death in Th2 but not Th1 cells [73]
206126_at	CXCR5(+)	G-protein coupled receptor	Il-4 increases expression of CXCR5, important for homing and sensitivity of T cells [74,75]
208121_s_at	PTPRO(+)	phosphatase	NR

Probe sets that fulfill the two-fold change criterion are marked with the + and - signs following the gene symbols based on the direction of the expression where + denotes up-regulation and - denotes down-regulation of Th1 specific genes. In addition, probe sets that do not fulfill the fold change criterion are marked in the lists with "<2". The known associations of genes in Th cell functions or subset specific gene expressions are listed in the table. *) Annotation based on Ingenuity Pathway Analysis ® by Ingenuity Systems. NR, not reported.
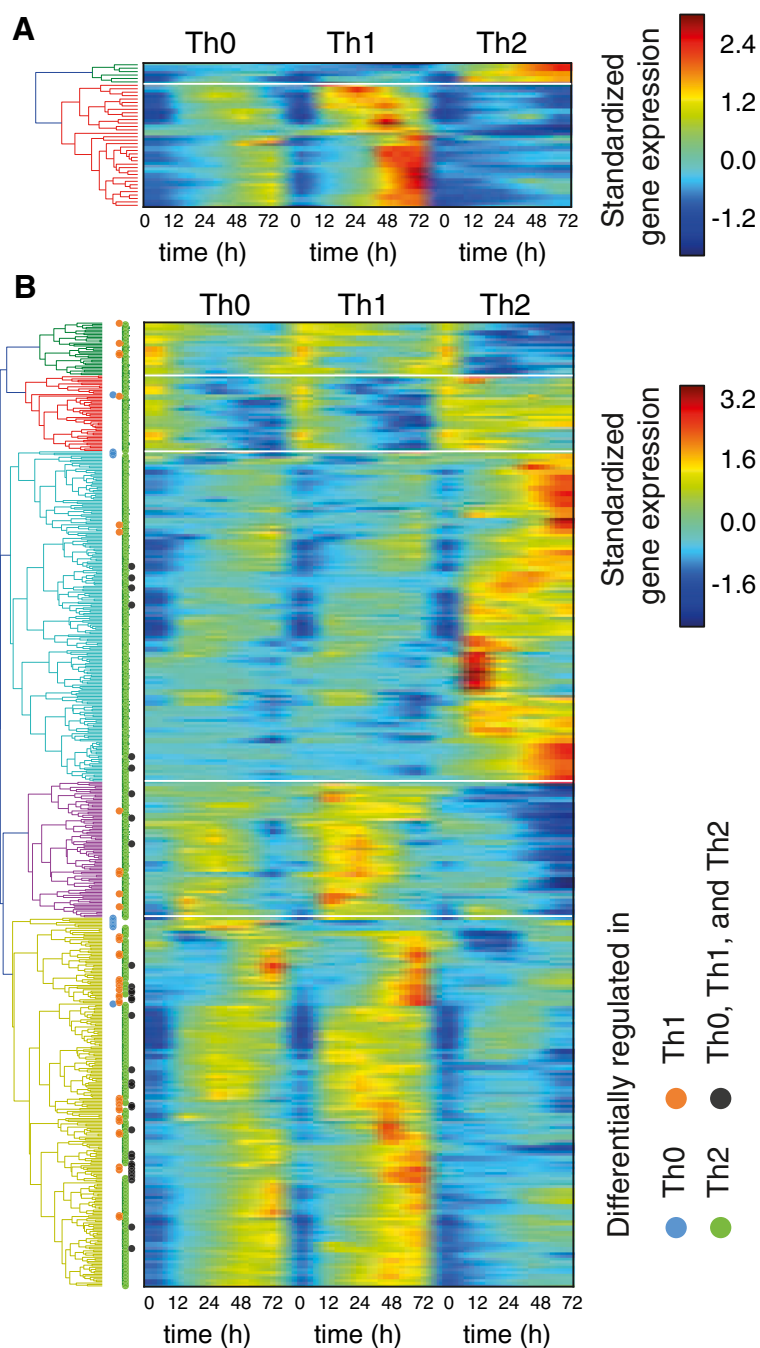
Figure 5 A global-view of the time-course profiles of reverse regulators of T helper cell differentiation and the time-course profiles of differentially regulated genes. (**A**) The set of genes that are specific in Th0, Th1, and Th2 are clustered in two clusters. The lower cluster holds altogether 34 unfiltered genes and the upper cluster contains only six genes. Most of the Th0, Th1 and Th2 specific genes are preferentially expressed in Th1 cells and have a lower expression level in Th2 cells. (**B**) The kinetics of the genes LIGAP identified to be differentially regulated are clustered in five clusters. Information about differential regulation is shown with colored dots. Consistent with (**A**), majority of the 34 genes specific in Th0, Th1, and Th2 are assigned to the fifth cluster, whereas the six genes are assigned to the third cluster. (**A and B**) Clusters are indicated with different colors on the branches in the dendrogram and with horizontal white lines in the heat maps. Standardized expression values are shown. The probe set data was standardized across the time points and lineages.

during the early differentiation although it is a master regulator in the early differentiation of Th2 cells [17].

An important goal is to identify master regulators of the lineage commitment processes. Recently, it was found out that most of the direct targets of STAT6, an important regulator of Th2 differentiation, were up-regulated in Th2 cells [17]. Here we were interested in identifying TFs whose binding sites are enriched in the promoter regions of the genes which are differentially regulated in Th2 conditions, both among the up-regulated and down-regulated genes. Instead of looking at individual TF binding predictions that are prone to contain false positives, we used the Fisher's exact test to search for enrichment of binding sites, in comparison to randomly selected gene set. The same analysis was carried out separately for all the differentially regulated gene sets and by taking into account the direction of regulation (repressed or activated).

Using a p-value cut-off of 0.01 for TF binding, we identified three hits from the enrichment analysis among Th2 specific up-regulated genes and three among the Th2 specific down-regulated genes. The results are depicted in Figure 6. The different enriched IRF family

motifs were combined and their targets were pooled. In accordance to our previously published results [17], the strongest hit within the Th2 up-regulated genes was STAT6 (p-value = 6.4 e-4), followed by NKX3A (p-value = 4.4 e-3), and CDP (p-value = 9.7 e-3). *NKX3A* is a member of the NKX family of homeobox genes that is expressed in prostate epithelium and functions as a potential prostate tumor suppressor [78]. Recently, in a study focusing on Jurkat cells, a GATA3 binding site on the promoter of *NKX3* gene was identified [79]. Furthermore, in mouse increased expression of *Nkx3a* was observed to be regulated by IL-4 independently of STAT6 [80]. CDP (CCAAT displacement protein) is highly conserved homeodomain transcription factor involved in many cellular processes, including differentiation, development and proliferation [81,82]. Interestingly, CDP has been identified as a repressive regulator of *CD8α* silencer region and *TCRβ* enhancer region and plays a role in promoting repressive chromatin modifications via association with histone deacetylase 1 and histone 3 methyltransferase [83-86]. It is important to notice that the binding motif analysis does not give information about the possible direction of regulation, e.g., it is
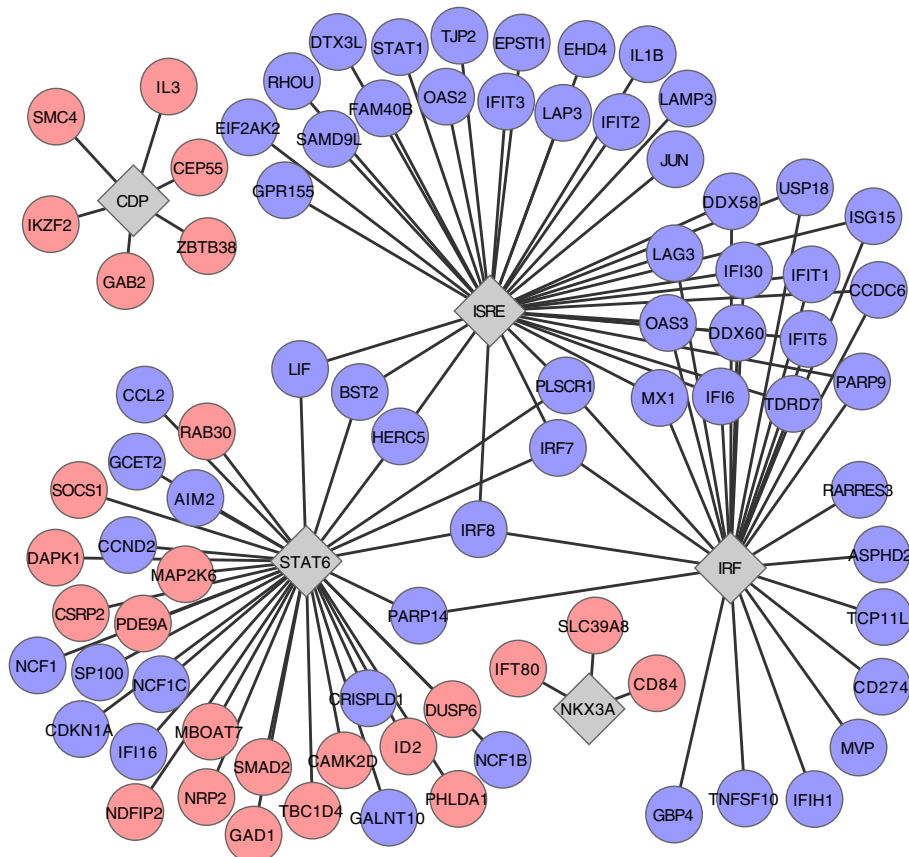
Figure 6 The network spanned by the enriched transcription factors and their predicted targets. Altogether five different TF binding family motifs were enriched either among the genes that were either down- or upregulated in Th2 cells. The TFs are depicted using gray tilted squares and the predicted target genes that are up- or down-regulated in Th2 cells are depicted using red or blue circles, respectively.

an open question whether CDP might up-regulate Th2 specific genes or down-regulate the genes in Th0 and Th1 lineages.

The three TF hits having enriched predicted binding sites among the Th2 down-regulated genes were the interferon regulatory factor (IRF) family of TFs (p-value = 2.5 e-6), IFN-stimulated genes factor 3 (ISGF3) (p-value = 1.8 e-4) and STAT6 (p-value = 3.5 e-3). IRF family consists of IRF1 to IRF10 and has been shown to be essential in expression of type I interferon genes, IFN-stimulated genes (ISG) and other pro-inflammatory response related cytokines [87]. These genes are maintained down-regulated during Th2 proliferation and therefore, the results are in line with the Th2 effector cells characteristics [88]. Moreover, IFNγ-induced expression of IRF1 and IRF2 has been shown to directly down-regulate IL-4 production by repressing *IL-4* promoter sites [89]. Opposing to other IRF family proteins, IRF4 has been shown to directly activate IL-4 promoter and IL-10 regulatory elements and be essential in Th2 cell differentiation by influencing the expression of GFI1, a transcriptional repressor in Th2 cells [90-92]. However, the analysis relying on known TF binding specificities will not allow segregation of individual members of the IRF family. Further, an essential regulator of most ISGs is ISGF3 that is composed of STAT1, STAT2 and IRF9 complex and works in conjunction with IRFs [93]. Identification of STAT6 as a regulator among the Th2 down-regulated genes is well in line with our previously published results, although its effect was observed to be less profound within Th2 down-regulated genes than among Th2 up-regulated target genes [17]. Comparison analysis of the predicted STAT6 target genes and Th2 up-regulated and down-regulated genes gave 16 and 19 overlapping genes, respectively. The full lists of overlapping genes are in Additional file 3: Table S2. We further analyzed the correlation between predicted STAT6 target promoters and experimentally observed promoter associated binding sites (Elo *et al.*, 2010), and observed significant correlation (p<0.05) between the target sites. The full list of predicted STAT6 target genes and promoter associated STAT6 binding sites identified by ChIP-seq as well as the overlapping genes are listed in the Additional file 3: Table S2. The overlapping binding sites included promoters for *C14orf177, CISH, HMMR, INO80, MGAT1, NUDCD2, SOCS1, SPINT2* and *ZNF570* genes.

Discussion

Identification of the key T helper cell regulators provides possible targets for modulation of immune response. To reveal T cell subset specific genes and their often subtle differences in expression, we developed a novel computational method, LIGAP. Traditional ways of identifying differentially expressed genes, such as the *t*-test, are problematic in studying time-series data since there is a

need to carry out hypothesis tests on individual time points. On the other hand, commonly used statistical tests for whole time-course, including e.g. F-test, do not account for the inherent correlation between measurement time points. LIGAP overcomes many problems that have previously prevented quantitative comparisons of multiple differentiation profiles, with or without replicates. Among several beneficial features, LIGAP models correlation between time points and can cope with non-stationarities and non-uniform measurement grid. Other methods, such as EDGE, uses splines to estimate smooth time-course profiles but does not quantify the differential expression for all lineage comparisons. TANOVA uses standard regression framework and lacks explicit correlation structure between time points. Our study highlights the validity of the method by identifying known and novel differentially regulated genes and their kinetic differences during T helper cell differentiation. In addition, the non-parametric computational analysis automatically provides informative illustrations of time-course profiles together with associated uncertainty.

LIGAP calculated Th0 specific gene set contains only 18 genes and Th1 specific 49 genes compared to 466 genes that are specific to Th2 conditions. Activation of Thp cells through TCR and CD28 results in induction of IFNγ, which in turn leads to activation of Th1 signature genes. Addition of IL-12, however, results in enhanced induction of these genes and Th1 programming. Consistent with our previous results genes differentially regulated in response to Th1 programming are much more limited than those detected in response to initiation of Th2 response [16,94].

Most of the Th1 specific genes encode well-known Th1 signature molecules. However, also genes new in this context were discovered. Interestingly, we identified *RORC* as one of the Th1 specific genes. Up-regulation of *RORC* in Th1 cells and existence of Th17/Th1 cells, however, remain conflicting as the master regulator of Th1 differentiation, T-bet, is known to inhibit transcription of *RORC* through RUNX1 [95], and expression of IL12Rβ2 is down-regulated by IL-17 [96]. It has been suggested that the high concentration of TGFβ required for *in vitro* Th17 polarization would inhibit IFNγ production [97], hence, it remains an open question whether some conditions would drive the differentiation of IL-17 and IFNγ producing cells from same naïve precursor T cell. Notably, *ex vivo* Th17 cells could be induced to develop further into Th17/Th1 cells by the combined actions of IFNγ and IL-12, and such conditions resulted in permissive chromatin remodeling at the *IL12RB2* locus and loss of repressive histone modification at the *TBX21* locus [29,98].

As an example of previously uncharacterized differentially regulated genes, we validated the expression of

Th2-associated phosphatases DUSP6 and PPP1R14A on protein level. PPP1R14A was shown in human pancreatic and melanoma tumor cell lines to positively regulate Ras/MAPK signaling [99], which are also involved in IL-4 induced signaling cascades. In T cells, the ERKs are activated though TCR stimulation and a TCR-mediated activation of Ras/MAPK signaling is required in differentiating murine Th2 but not in Th1 cells [100]. Furthermore, the Ras/MAPK cascade was shown to enhance the stability of GATA3 protein [101] as well as STAT6 independent CD3 and CD28 induced initial IL4 production [102]. DUSP6 on other hand is known to negatively regulate members of the mitogen-activated protein (MAP) kinase superfamily associated with cellular proliferation and differentiation [103]. More specifically, DUSP6 expression was shown to be induced by ERK1/2 signaling in differentiating mouse embryonic cell line and in human retinal pigment epithelial cells [104,105] and it was hypothesized that DUSP6 is an essential part of a negative feedback loop of ERK1/2 signaling [106]. However, the T cell associated functions of both PPP1R14A and DUSP6 are completely unknown. Therefore, their significance in the signaling cascades of differentiating Th2 cells remains a highly interesting area of future research.

SPINT2 was recently identified as a direct STAT6 target in differentiating human Th2 cells [17] and in this study we are the first to show that SPINT2 is upregulated in Th2 cells at protein level as compared to other Th cell subsets. We found SPINT2 to be specifically expressed on Th2 cell surface as well as secreted into the culture medium, suggesting presence of a multiple transcripts of which some may lack the anchoring transmembrane domain [47]. Human SPINT2 (HAI-2) is a physiological inhibitor of matrix cleaving proteases and decreased expression of SPINT2 has been linked to progression of several cancers [107-109]. Up-regulated expression of extracellular proteases is crucial for procancerous pathways as this enables efficient remodeling of the extracellular matrix as well as cleavage and activation of growth factors and their receptors. Interestingly, a truncated and secreted SPINT2 may act as an inhibitor for the activator of hepatocyte growth factor (HGF) and HGF is prominently expressed in lung tissue and is linked to reduced expression of Th2 cytokines and TGFβ, reduction of allergic airway inflammation, airway hyperresponsiveness and remodeling as well as reduced recruitment of eosinophils to the site of allergic inflammation in vivo [110,111]. This suggests that SPINT2 might enhance Th2 response in allergic airway inflammation by inhibiting HGF signaling.

The LIGAP method elegantly identified the reciprocally regulated genes within the Th0, Th1 and Th2 conditions. Essentially, the list included genes encoding the hallmark Th1 specific transcription factor T-bet and cytokine IFNγ as well as the transmembrane receptor for IL-12. This list also included few cytoskeleton associated proteins, such as dystrophin (DMD), and palladin (PALLD), of which there is no current knowledge for their function in differentiating T helper cells. The observation suggests differences in cellular structures or putatively in the interaction of APC with the Th cell subsets as rearrangement of the cytoskeleton in T cells plays an important role in the organization of the immunological synapse (IS) and Th1 and Th2 cells are known to form morphologically distinct ISs [112,113]. In addition to MAP3K8, molecules that participate in phosphorylation signaling cascades e.g. P2RY14, LPAR3, PPP1R14A, and PTPRO suggest their potential role for initiation or regulation of differentiation cascades. Importantly, the results presented here enable opportunities for further data mining and follow-up studies addressing the functions and importance of the novel Th subset specific genes.

The identification of STAT6 as the most significant TF regulating Th2 specific enhancement of transcription by the TF binding analysis is well in line with our previous STAT6 ChIP results [17]. Furthermore, the analysis between the predicted STAT6 target gene promoters and experimentally observed promoter associated binding sites showed statistically significant correlation. Interestingly, the overlapping STAT6 targets included INO80, which has been identifies as a part of a chromatin remodeling complex [114] and may hence, be involved in Th2 specific epigenetical regulation of Th cell differentiation. STAT6 specific regulation of Mannosyl (alpha-1,3-)-glycoprotein beta-1,2-N-acetylglucosaminyltransferase (MGAT1), a N-glycan-processing enzyme [115], may on one hand be involved in modifying the Th2 cell specific surface glycoprotein structures [116]. The overlapping target sites included also the promoter for SPINT2. The number of predicted STAT6 binding sites, however, was much larger than the experimentally observed binding sites, which may reflect the typically observed high false positive rate of computational binding predictions and the cell type specific state of chromatin as well as other competing factors affecting binding in vitro. The data created here also further suggests novel control mechanisms involving GATA3 regulated NKX3A as well as chromatin modification associated CDP. Only less than 10% of the Th2 down-regulated genes were reported to be direct targets of STAT6 by Elo et al., (2010) suggesting other major regulatory mechanisms play role among the IL-4 induced down-regulated genes. We found enrichment of IRF family and ISGF3 binding motifs in promoter regions of genes that are repressed in Th2 polarizing conditions, indicating that these TFs may play a significant role in the suppressing undesired gene expression in differentiating Th2 cells. Indeed, several IRF family members have been identified

as differentially expressed during Th cell differentiation and necessary for both Th1 and Th2 polarization. As the IRF family proteins, excluding IRF1, share the same binding specificity model in TRANSFAC, the individual regulatory role for these factors is, however, difficult to postulate based on *in silico* TF binding site analysis.

Conclusions

The proposed LIGAP method can quantify a well-defined probabilistic specificity score for each gene and for each condition promoting a certain lineage commitment. In addition to grouping and ranking genes based on their dynamics, LIGAP summarizes all time-course measurements, together with the associated uncertainty, in an illustrative summary plot for visualization and manual assessment purposes. While here we have demonstrated the utility of LIGAP in analysis of gene expression dynamics, the LIGAP method is widely applicable to many types of datasets including quantitative time-course experiments and generalizes to any number of conditions.

Methods

Human CD4+ T cell purification and culturing. The human naïve umbilical cord blood CD4+ T cells were isolated as previously described [17]. Briefly, umbilical cord blood was collected from healthy neonates born in Turku University Hospital, Finland. Mononuclear cells were separated with Ficoll-Paque gradient centrifugation (#GEHE17-1440-3, Amersham Biosciences) and CD4+ T cells were then isolated with magnetic beads (Dynal CD4 Positive Isolation Kit, #113-31D, Invitrogen). After isolation the CD4+ cells were pooled to prepare cell cultures consisting cells from several neonates. The same pooled cells as utilized for Th0 (activated) and Th2 (activated and IL-4 stimulated) culture conditions by Elo et al. (2010) were used parallel for Th1 polarizing cultures. For activation, the cells were treated with plate-bound anti-CD3 (500 ng/24-well culture plate well, #IM1304, Immunotech) and soluble anti-CD28 (500 ng/ml, #IM1376, Immunotech) in density of $2\text{-}4 \times 10^6$ cells/ml of Yssel's medium (Iscove modified Dulbecco medium, #31980-048, Invitrogen) supplemented with Yssel medium concentrate [117], 1% human AB serum (#C11-011, PAA) and 100 U/ml Penicillin and 100 μg/ml Streptomycin (#P0781, Sigma) at 37°C in 5% CO2. For induction of Th1 cell polarization, IL-12 (2.5 ng/ml, # 219-IL, R&D Systems) was added to the cultures. At 48h after activation, IL-2 was added (17 ng/ml, #202-IL, R&D Systems) to all the cells and the polarizing conditions were maintained throughout the culture. The polarizing Th cells were harvested at time points 0, 12, 24, 48 hours in three replicates and at 72 hours in two replicates.

All the data included in this manuscript has been acquired under the permission from the Ethics Committee of the Hospital District of Southwest Finland approving the anonymous collection of cord blood samples after a parental consent, and the permission being in compliance with the Helsinki Declaration

Microarray studies. The preparation of samples for microarray detections was done as described in [17]. Essentially, total RNA (RNeasy Mini Kit, Qiagen) was extracted from the cultured cells and cRNA hybridized on Affymetrix GeneChip HG-U133 Plus 2.0 arrays (Affymetrix, Santa Clara, USA). All the microarray samples included in this study have been prepared at Finnish DNA Microarray Centre, Turku. The raw microarray data were processed using robust multi-array average normalization and log2-transformed in R (version 2.12.0) using the Bioconductor affy package (version 1.28.0).

Flow cytometry. The Th0, Th1 and Th2 condition cells at 24 hours were stained for SPINT2 expression studies. Purified anti-SPINT2 (8.7 μg/ml, #HPA 011101-100UL, Sigma-Aldrich) was used as primary antibody followed by staining with FITC-conjugated F(ab')2 anti-rabbit IgG secondary antibody (1:1000 dilution, #11-4839-81, eBioscience). The stainings were analyzed with LSR II flow cytometer (BD Biosciences) and Flowing Software (www.flowingsoftware.com).

ELISA. The cell culture supernatants (at 48 hours) from Th0, Th1 and Th2 conditions were assayed for SPINT2/HAI-2 secretion by ELISA (# DY1106, R&D) according to the manufacturer instructions.

LIGAP. We construct our model-based lineage commitment comparison and visualization methodology, called LIGAP, using non-parametric GP regression similar to that in [14], extend the methodology to any number of conditions and propose to use a non-stationary neural network (NN) covariance function $k(x^p, x^q) = \sigma^*\text{asin}$ $(x^{p\ '}\text{diag}(l^{-2})^*x^q\ /\ \text{sqrt}[(1+x^{p'}\text{diag}(l^{-2})^*x^p)^*(1+x^{q'}\text{diag}(l^{-2})^*x^q)])$. The vectors x^p and x^q are augmented by an extra bias unit value entry and the parameter l defines the length-scale and σ controls the signal variance [118]. A non-stationary covariance function is chosen because often after cell activation or other stimulation the effects on temporal behavior of gene expression are very active and dynamic right after the stimulation but they mellow down over time and, thus, the observed behavior is non-stationary. For each gene at a time, LIGAP makes all comparisons between different cell subsets over the whole time-course data sets. In our application, the multiple hypotheses H_j are defined by the different partitions of the cell lineages. For example, if there are only two different lineages, then there are two different partitions (or hypothesis): H_1 denotes that lineages are similar and H_2 denotes that lineages are different. In our application consisting of three lineages, Th0, Th1 and Th2, we have 5 alternative hypotheses; *(i)* "Th0, Th1, Th2 time-course profiles are all similar" (hypothesis H_1), *(ii)* "Th0 and Th1

are similar and Th2 is different" (hypothesis H_2), *(iii)* "Th0 and Th2 are similar and Th1 is different" (hypothesis H_3), *(iv)* "Th1 and Th2 are similar and Th0 is different" (hypothesis H_4), and *(v)* "Th0, Th1, and Th2 are different from each other" (hypothesis H_5). LIGAP comparisons and quantifications are illustrated in Figure 1. In general, the total number of different partitions of N lineages is known in literature as the Bell number B_n (e.g., $B_1 = 1$, $B_2 = 2$, $B_3 = 5$, $B_4 = 15$, etc.) [119].

Bayes factor is commonly used to see the evidence of the two alternative hypotheses; differentially expressed or not within a given time interval. To extend this to multiple lineages, we use the marginal likelihood $p(D_i \mid H_j)$ to define the posterior probabilities of the different hypotheses H_j. For each of the hypothesis H_j, the data D_i for the ith gene is split according to the partitioning. For example, for our application containing three lineages, hypothesis H_1 corresponds to grouping data from all lineages, hypothesis H_2 corresponds to splitting the data so that Th0 and Th1 time-course profiles are grouped together and time-course profiles from Th2 forms its own subset of data, hypothesis H_3 corresponds to splitting the data so that Th0 and Th2 time-course profiles are grouped together and Th1 forms its own subset of data, etc.

For each hypothesis, non-parametric regression is carried out separately for each subset of the data. For example, for the hypothesis H_3 we fit a GP to the combination of Th0 and Th2 time-course profiles and another GP to the Th1 time-course profiles. Following the standard GP regression methodology [118], the marginalization is done over the latent regression function and the hyperparameters are estimated using type II maximum likelihood estimation with a conjugate gradient based optimization algorithm initiated with ten randomly chosen hyperparameter values. Under the assumption of Gaussian likelihood and noise, the marginal likelihood can be written out analytically, and thus its value can be easily evaluated [118]. The marginal likelihood of a certain hypothesis (i.e., partitioning) is the product of the marginal likelihood of the separate subsets. The key idea behind the modeling is to find the marginal likelihood of the data under different hypotheses and thus have a probabilistic score to objectively compare different hypotheses.

Using the Bayes' theorem and assuming unbiased, equal prior probabilities for different hypotheses (i.e., $P(H_k) = P(H_l)$ for all k and l), we can write the posterior probabilities for the ith gene as $P(H_j \mid D_i) = P(D_i \mid H_j)P(H_j)/C$, where $C = \Sigma_j\ P(D_i \mid H_j)P(H_j)$ is a normalizing constant. Finally, these quantities can be combined to quantify the score of differential regulation for each gene. For example, the probability of the ith gene being differentially regulated in Th2 lineage can be quantified as P("Gene is differentially regulated in Th2" $\mid D_i$) = $P(H_2 \mid D_i) + P(H_5 \mid D_i)$.

ProbTF. ProbTF [76] method is used to make TF binding predictions on promoters of all RefSeq genes. Sequence specificities of TFs are taken from the TRANSFAC database [77] version 2009.3. All non-redundant PSFMs associated to human were taken, totaling 248 matrices. Promoters are defined as the [−1000,500] bp region around TSS. To assess statistical significance, we construct a TF specific null distribution by randomly sampling 50000 genomic locations of size 1501 nucleotides, against which the p-values of TF binding are computed.

Hierarchical clustering. The hierarchical clustering in Figure 5 was done using complete linkage and Euclidean distance metric.

Data access. The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus [120] and are accessible through GEO Series accession number GSE 32959 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE32959).

Additional files

Additional file 1: Figure S1. Additional data file 1 is a PDF containing two panels illustrating the sensitivity of Sorad to identify changes in time series by integrating modest changes between time series over time.

Additional file 2: Table S1. The following additional data are available with the online version of this paper. Additional data file 2 is a table showing the full results of the LIGAP analysis.

Additional file 3: Table S2. Additional data file 3 is a table listing the enriched transcription factor binding motifs in the set of Th2 specific genes and the overlapping genes, and the full listing of predicted STAT6 binding sites and the overlapping genes with the published promoter associated STAT6 targets.

Author details
[1]Department of Signal Processing, Tampere University of Technology, Tampere, Finland. [2]Turku Centre for Biotechnology, University of Turku and Åbo Akademi University, Turku, Finland. [3]Turku Doctoral Programme of Biomedical Sciences, Turku, Finland. [4]Department of Information and Computer Science, Aalto University, Helsinki, Finland.

References

1. Weaver CT, Harrington LE, Mangan PR, Gavrieli M, Murphy KM: **Th17: an effector CD4 T cell lineage with regulatory T cell ties.** *Immunity* 2006, **24**:677–688.
2. Neurath MF, Finotto S, Glimcher LH: **The role of Th1/Th2 polarization in mucosal immunity.** *Nat Med* 2002, **8**:567–573.
3. Sheikh A, Strachan DP: **The hygiene theory: fact or fiction?** *Curr Opin Otolaryngol Head Neck Surg* 2004, **12**:232–236.
4. Hutton JJ, Jegga AG, Kong S, Gupta A, Ebert C, Williams S, Katz JD, Aronow BJ: **Microarray and comparative genomics-based identification of genes and gene regulatory regions of the mouse immune system.** *BMC Genomics* 2004, **5**:82.
5. Aghajanova L, Skottman H, Strömberg A-M, Inzunza J, Lahesmaa R, Hovatta O: **Expression of leukemia inhibitory factor and its receptors is increased during differentiation of human embryonic stem cells.** *Fertil Steril* 2006, **86**:1193–1209.
6. Mariani L, Löhning M, Radbruch A, Höfer T: **Transcriptional control networks of cell differentiation: insights from helper T lymphocytes.** *Prog Biophys Mol Biol* 2004, **86**:45–76.
7. van den Ham H-J, de Boer RJ: **From the two-dimensional Th1 and Th2 phenotypes to high-dimensional models for gene regulation.** *Int Immunol* 2008, **20**:1269–1277.
8. van den Ham HJ, De Waal L, Andeweg AC, De Boer RJ: **Identification of helper T cell master regulator candidates using the polar score method.** *J Immunol Methods* 2010, **361**:98–109.
9. Pedicini M, Barrenäs F, Clancy T, Castiglione F, Hovig E, Kanduri K, Santoni D, Benson M: **Combining network modeling and gene expression microarray analysis to explore the dynamics of Th1 and Th2 cell regulation.** *PLoS Computational Biol* 2010, **6**:e1001032.
10. Smyth G, limma: **Linear Models for Microarray Data Bioinformatics and Computational Biology Solutions Using R and Bioconductor.** In Edited by Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S. Springer: New York; 2005:397–420. Gail M, Krickeberg K, Samet JM, Tsiatis A, Wong W (Series Editor): *Statistics for Biology and Health*, Springer, New York, pages 397–420.
11. Coffey N, Hinde J: **Analyzing Time-Course Microarray Data Using Functional Data Analysis - A Review.** *Stat Appl Genet Mol Biol* 2011, **10**:10–32.
12. Storey JD, Xiao W, Leek JT, Tompkins RG, Davis RW: **Significance analysis of time course microarray experiments.** *Proc Natl Acad Sci U S A* 2005, **102**:12837–12842.
13. Zhou BY, Xu WH, Herndon D, Tompkins R, Davis R, Xiao WZ, Wong WH: **Inflammation Host Response I: Analysis of factorial time-course microarrays with application to a clinical study of burn injury.** *Proc Natl Acad Sci U S A* 2010, **107**:9923–9928.
14. Stegle O, Denby KJ, Cooke EJ, Wild DL, Ghahramani Z, Borgwardt KM: **A robust Bayesian two-sample test for detecting intervals of differential gene expression in microarray time series.** *J Comput Biol: J Comput Mol Biol* 2010, **17**:355–367.
15. Lund RJ, Ylikoski EK, Aittokallio T, Nevalainen O, Lahesmaa R: **Kinetics and STAT4- or STAT6-mediated regulation of genes involved in lymphocyte polarization to Th1 and Th2 cells.** *Eur J Immunol* 2003, **33**:1105–1116.
16. Lund RJ, Löytömäki M, Naumanen T, Dixon C, Chen Z, Ahlfors H, Tuomela S, Tahvanainen J, Scheinin J, Henttinen T, et al: **Genome-wide identification of novel genes involved in early Th1 and Th2 cell differentiation.** *J Immunology (Baltimore, Md: 1950)* 2007, **178**:3648–3660.
17. Elo LL, Järvenpää H, Tuomela S, Raghav S, Ahlfors H, Laurila K, Gupta B, Lund RJ, Tahvanainen J, Hawkins RD, et al: **Genome-wide profiling of interleukin-4 and STAT6 transcription factor regulation of human Th2 cell programming.** *Immunity* 2010, **32**:852–862.
18. Jeffreys H: *Theory of probability.* 3rd edition. USA: Oxford University Press; 1998.
19. Nakanishi K, Yoshimoto T, Tsutsui H, Okamura H: **Interleukin-18 regulates both Th1 and Th2 responses.** *Annu Rev Immunol* 2001, **19**:423–474.
20. Tominaga K, Yoshimoto T, Torigoe K, Kurimoto M, Matsui K, Hada T, Okamura H, Nakanishi K: **IL-12 synergizes with IL-18 or IL-1beta for IFN-gamma production from human T cells.** *Int Immunol* 2000, **12**:151–160.
21. Micallef MJ, Ohtsuki T, Kohno K, Tanabe F, Ushio S, Namba M, Tanimoto T, Torigoe K, Fujii M, Ikeda M, et al: **Interferon-gamma-inducing factor enhances T helper 1 cytokine production by stimulated human T cells: synergism with interleukin-12 for interferon-gamma production.** *Eur J Immunol* 1996, **26**:1647–1651.
22. Xu DM, Trajkovic V, Hunter D, Leung BP, Schulz K, Gracie JA, McInnes IB, Liew FY: **IL-18 induces the differentiation of Th1 or Th2 cells depending upon cytokine milieu and genetic background.** *Eur J Immunol* 2000, **30**:3147–3156.
23. Niesner U, Albrecht I, Janke M, Doebis C, Loddenkemper C, Lexberg MH, Eulenburg K, Kreher S, Koeck J, Baumgrass R, et al: **Autoregulation of Th1-mediated inflammation by twist1.** *J Exp Med* 2008, **205**:1889–1901.
24. Kim CH, Kunkel EJ, Boisvert J, Johnston B, Campbell JJ, Genovese MC, Greenberg HB, Butcher EC: **Bonzo/CXCR6 expression defines type 1-polarized T-cell subsets with extralymphoid tissue homing potential.** *J Clin Invest* 2001, **107**:595–601.
25. Langenkamp A, Nagata K, Murphy K, Wu L, Lanzavecchia A, Sallusto F: **Kinetics and expression patterns of chemokine receptors in human CD4 + T lymphocytes primed by myeloid or plasmacytoid dendritic cells.** *Eur J Immunol* 2003, **33**:474–482.
26. Casas R, Lindau C, Zetterström O, Duchén K: **Downregulation of CXCR6 and CXCR3 in lymphocytes from birch-allergic patients.** *Scand J Immunol* 2008, **68**:351–361.
27. Watford WT, Hissong BD, Durant LR, Yamane H, Muul LM, Kanno Y, Tato CM, Ramos HL, Berger AE, Mielke L, et al: **Tpl2 kinase regulates T cell interferon-gamma production and host resistance to Toxoplasma gondii.** *J Exp Med* 2008, **205**:2803–2812.
28. Annunziato F, Cosmi L, Santarlasci V, Maggi L, Liotta F, Mazzinghi B, Parente E, Filì L, Ferri S, Frosali F, et al: **Phenotypic and functional features of human Th17 cells.** *J Exp Med* 2007, **204**:1849–1861.
29. Lexberg MH, Taubner A, Albrecht I, Lepenies I, Richter A, Kamradt T, Radbruch A, Chang H-D: **IFN-γ and IL-12 synergize to convert in vivo generated Th17 into Th1/Th17 cells.** *Eur J Immunol* 2010, **40**:3017–3027.
30. Zaba LC, Suárez-Fariñas M, Fuentes-Duculan J, Nograles KE, Guttman-Yassky E, Cardinale I, Lowes MA, Krueger JG: **Effective treatment of psoriasis with etanercept is linked to suppression of IL-17 signaling, not immediate response TNF genes.** *J Allergy Clin Immunol* 2009, **124**:1022–1010. e1021-1395.
31. Mendell JR, Campbell K, Rodino-Klapac L, Sahenk Z, Shilling C, Lewis S, Bowles D, Gray S, Li C, Galloway G, et al: **Dystrophin immunity in Duchenne's muscular dystrophy.** *N Engl J Med* 2010, **363**:1429–1437.
32. Surace C, Piazzolla S, Sirleto P, Digilio MC, Roberti MC, Lombardo A, D'Elia G, Tomaiuolo AC, Petrocchi S, Capolino R, et al: **Mild ring 17 syndrome shares common phenotypic features irrespective of the chromosomal breakpoints location.** *Clin Genet* 2009, **76**:256–262.
33. Arai S, Shibata H, Sakai M, Ninomiya H, Iwata N, Ozaki N, Fukumaki Y: **Association analysis of the glutamic acid decarboxylase 2 and the glutamine synthetase genes (GAD2, GLUL) with schizophrenia.** *Psychiatr Genet* 2009, **19**:6–13.
34. Saksi J, Ijäs P, Nuotio K, Sonninen R, Soinne L, Salonen O, Saimanen E, Tuimala J, Lehtonen-Smeds EM, Kaste M, et al: **Gene expression differences between stroke-associated and asymptomatic carotid plaques.** *J Mol Med (Berlin, Germany)* 2011, **89**:1015–1026.
35. Tatnell PJ, Powell DJ, Hill J, Smith TS, Tew DG, Kay J: **Napsins: new human aspartic proteinases. Distinction between two closely related genes.** *FEBS Lett* 1998, **441**:43–48.
36. Puente XS, Gutiérrez-Fernández A, Ordóñez GR, Hillier LW, López-Otín C: **Comparative genomic analysis of human and chimpanzee proteases.** *Genomics* 2005, **86**:638–647.
37. Ishii N, Ozaki K, Sato H, Mizuno H, Saito S, Takahashi A, Miyamoto Y, Ikegawa S, Kamatani N, Hori M, et al: **Identification of a novel non-coding RNA, MIAT, that confers risk of myocardial infarction.** *J Hum Genet* 2006, **51**:1087–1099.
38. Ouyang W, Löhning M, Gao Z, Assenmacher M, Ranganath S, Radbruch A, Murphy KM: **Stat6-independent GATA-3 autoactivation directs IL-4-**

independent Th2 development and commitment. *Immunity* 2000,
12:27–37.

39. Chaitidis P, Billett EE, O'Donnell VB, Fajardo AB, Fitzgerald J, Kuban RJ,
Ungethuem U, Kühn H: **Th2 response of human peripheral monocytes
involves isoform-specific induction of monoamine oxidase-A.**
J Immunology (Baltimore, Md: 1950) 2004, **173**:4821–4827.

40. Kubera M, Maes M, Kenis G, Kim Y-K, Lasoń W: **Effects of serotonin and
serotonergic agonists and antagonists on the production of tumor
necrosis factor alpha and interleukin-6.** *Psychiatry Res* 2005, **134**:251–258.

41. Kawaguchi T, Qin L, Shimomura T, Kondo J, Matsumoto K, Denda K,
Kitamura N: **Purification and cloning of hepatocyte growth factor
activator inhibitor type 2, a Kunitz-type serine protease inhibitor.**
J Biol Chem 1997, **272**:27558–27564.

42. Marlor CW, Delaria KA, Davis G, Muller DK, Greve JM, Tamburini PP:
**Identification and cloning of human placental bikunin, a novel serine
protease inhibitor containing two Kunitz domains.** *J Biol Chem* 1997,
272:12202–12208.

43. Xu Y, Carr PD, Guss JM, Ollis DL: **The crystal structure of bikunin from the
inter-alpha-inhibitor complex: a serine protease inhibitor with two
Kunitz domains.** *J Mol Biol* 1998, **276**:955–966.

44. Szabo R, Hobson JP, Christoph K, Kosa P, List K, Bugge TH: **Regulation of
cell surface protease matriptase by HAI2 is essential for placental
development, neural tube closure and embryonic survival in mice.**
Development (Cambridge, England) 2009, **136**:2653–2663.

45. Parr C, Watkins G, Mansel RE, Jiang WG: **The hepatocyte growth factor
regulatory factors in human breast cancer.** *Clin Cancer Res: An Official J
Am Assoc for Cancer Res* 2004, **10**:202–211.

46. Ahn H-J, Kim JY, Ryu K-J, Nam H-W: **STAT6 activation by Toxoplasma
gondii infection induces the expression of Th2 C-C chemokine ligands
and B clade serine protease inhibitors in macrophage.** *Parasitol Res* 2009,
105:1445–1453.

47. Naganuma S, Itoh H, Uchiyama S, Tanaka H, Nagaike K, Miyata S,
Uchinokura S, Nuki Y, Akiyama Y, Chijiiwa K, Kataoka H: **Characterization
of transcripts generated from mouse hepatocyte growth factor
activator inhibitor type 2 (HAI-2) and HAI-2-related small peptide
(H2RSP) genes: chimeric mRNA transcribed from both HAI-2 and H2RSP
genes is detected in human but not in mouse.** *Biochem Biophys Res
Commun* 2003, **302**:345–353.

48. Eto M, Senba S, Morita F, Yazawa M: **Molecular cloning of a novel
phosphorylation-dependent inhibitory protein of protein phosphatase-1
(CPI17) in smooth muscle: its specific localization in smooth muscle.** *FEBS
Lett* 1997, **410**:356–360.

49. Groom LA, Sneddon AA, Alessi DR, Dowd S, Keyse SM: **Differential
regulation of the MAP, SAP and RK/p38 kinases by Pyst1, a novel
cytosolic dual-specificity phosphatase.** *EMBO J* 1996, **15**:3621–3632.

50. Lund R, Ahlfors H, Kainonen E, Lahesmaa A-M, Dixon C, Lahesmaa R:
**Identification of genes involved in the initiation of human Th1 or Th2
cell commitment.** *Eur J Immunol* 2005, **35**:3307–3319.

51. Usui T, Preiss JC, Kanno Y, Yao ZJ, Bream JH, O'Shea JJ, Strober W: **T-bet
regulates Th1 responses through essential effects on GATA-3 function
rather than on IFNG gene acetylation and transcription.** *J Exp Med* 2006,
203:755–766.

52. Thedieck K, Polak P, Kim ML, Molle KD, Cohen A, Jenö P, Arrieumerlou C,
Hall MN: **PRAS40 and PRR5-like protein are new mTOR interactors that
regulate apoptosis.** *PLoS One* 2007, **2**:e1217.

53. Perttilä J, Merikanto K, Naukkarinen J, Surakka I, Martin NW, Tanhuanpää K,
Grimard V, Taskinen M-R, Thiele C, Salomaa V, *et al*: **OSBPL10, a novel
candidate gene for high triglyceride trait in dyslipidemic Finnish
subjects, regulates cellular lipid metabolism.** *J Mol Med (Berlin, Germany)*
2009, **87**:825–835.

54. Müller T, Bayer H, Myrtek D, Ferrari D, Sorichter S, Ziegenhagen MW, Zissel
G, Virchow JC, Luttmann W, Norgauer J, *et al*: **The P2Y14 receptor of
airway epithelial cells: coupling to intracellular Ca2+ and IL-8 secretion.**
Am J Respir Cell Mol Biol 2005, **33**:601–609.

55. Arase T, Uchida H, Kajitani T, Ono M, Tamaki K, Oda H, Nishikawa S,
Kagami M, Nagashima T, Masuda H, *et al*: **The UDP-glucose receptor
P2RY14 triggers innate mucosal immunity in the female reproductive
tract by inducing IL-8.** *J Immunology (Baltimore, Md: 1950)* 2009,
182:7074–7084.

56. Skelton L, Cooper M, Murphy M, Platt A: **Human immature monocyte-
derived dendritic cells express the G protein-coupled receptor GPR105**

(KIAA0001, P2Y14) **and increase intracellular calcium in response to its
agonist, uridine diphosphoglucose.** *J Immunology (Baltimore, Md: 1950)*
2003, **171**:1941–1949.

57. Scrivens M, Dickenson JM: **Functional expression of the P2Y14 receptor in
murine T-lymphocytes.** *Br J Pharmacol* 2005, **146**:435–444.

58. Chen L, Juszczynski P, Takeyama K, Aguiar RCT, Shipp MA: **Protein tyrosine
phosphatase receptor-type O truncated (PTPROt) regulates SYK
phosphorylation, proximal B-cell-receptor signaling, and cellular
proliferation.** *Blood* 2006, **108**:3428–3433.

59. Agnello D, Lankford CSR, Bream J, Morinobu A, Gadina M, O'Shea JJ,
Frucht DM: **Cytokines and transcription factors that regulate T helper cell
differentiation: new players and new insights.** *J Clin Immunol* 2003,
23:147–161.

60. Moulian N, Bidault J, Planché C, Berrih-Aknin S: **Two signaling
pathways can increase fas expression in human thymocytes.**
Blood 1998, **92**:1297–1307.

61. Akimzhanov AM, Wang X, Sun J, Boehning D: **T-cell receptor complex is
essential for Fas signal transduction.** *Proc Natl Acad Sci U S A* 2010,
107:15105–15110.

62. Nurieva RI, Chung Y, Hwang D, Yang XO, Kang HS, Ma L, Wang YH,
Watowich SS, Jetten AM, Tian Q, Dong C: **Generation of T follicular helper
cells is mediated by interleukin-21 but independent of T helper 1, 2, or
17 cell lineages.** *Immunity* 2008, **29**:138–149.

63. Torchinsky MB, Blander JM: **T helper 17 cells: discovery, function, and
physiological trigger.** *Cellular and Mole Life Sci* 2010, **67**:1407–1421.

64. Strengell M, Matikainen S, Siren J, Lehtonen A, Foster D, Julkunen I,
Sareneva T: **IL-21 in synergy with IL-15 or IL-18 enhances IFN-
gamma production in human NK and T cells.** *J Immunol* 2003,
170:5464–5469.

65. Jacobson NG, Szabo SJ, Weber-Nordt RM, Zhong Z, Schreiber RD, Darnell JE,
Murphy KM: **Interleukin 12 signaling in T helper type 1 (Th1) cells
involves tyrosine phosphorylation of signal transducer and activator
of transcription (Stat)3 and Stat4.** *The J Experimental Med* 1995,
181:1755–1762.

66. Sallusto F, Lanzavecchia A, Mackay CR: **Chemokines and chemokine
receptors in T-cell priming and Th1/Th2-mediated responses.** *Immunol
Today* 1998, **19**:568–574.

67. Bonecchi R, Bianchi G, Bordignon PP, D'Ambrosio D, Lang R, Borsatti A,
Sozzani S, Allavena P, Gray PA, Mantovani A, Sinigaglia F: **Differential
expression of chemokine receptors and chemotactic responsiveness of
type 1 T helper cells (Th1s) and Th2s.** *The J Experimental Med* 1998,
187:129–134.

68. Debes GF, Diehl MC: **CCL8 and skin T cells–an allergic attraction.** *Nat
Immunol* 2011, **12**:111–112.

69. Sheikh F, Baurin VV, Lewis-Antes A, Shah NK, Smirnov SV, Anantha S,
Dickensheets H, Dumoutier L, Renauld J-C, Zdanov A, *et al*: **Cutting edge:
IL-26 signals through a novel receptor complex composed of IL-20
receptor 1 and IL-10 receptor 2.** *J Immunology (Baltimore, Md: 1950)* 2004,
172:2006–2010.

70. Wolk K, Kunz S, Asadullah K, Sabat R: **Cutting edge: immune cells as
sources and targets of the IL-10 family members?** *J Immunology
(Baltimore, Md: 1950)* 2002, **168**:5397–5402.

71. Miller SA, Weinmann AS: **Molecular mechanisms by which T-bet regulates
T-helper cell commitment.** *Immunol Rev* 2010, **238**:233–246.

72. Kanda H, Newton R, Klein R, Morita Y, Gunn MD, Rosen SD: **Autotaxin,
an ectoenzyme that produces lysophosphatidic acid, promotes the
entry of lymphocytes into secondary lymphoid organs.** *Nat Immunol*
2008, **9**:415–423.

73. Devadas S, Das J, Liu C, Zhang L, Roberts AI, Pan Z, Moore PA, Das G, Shi Y:
**Granzyme B is critical for T cell receptor-induced cell death of type 2
helper T cells.** *Immunity* 2006, **25**:237–247.

74. Kim CH, Nagata K, Butcher EC: **Dendritic cells support sequential
reprogramming of chemoattractant receptor profiles during naive to
effector T cell differentiation.** *J Immunology (Baltimore, Md: 1950)* 2003,
171:152–158.

75. Sallusto F, Mackay CR, Lanzavecchia A: **The role of chemokine receptors in
primary, effector, and memory immune responses.** *Annu Rev Immunol*
2000, **18**:593–620.

76. Lähdesmäki H, Rust AG, Shmulevich I: **Probabilistic inference of
transcription factor binding from multiple data sources.** *PLoS One*
2008, **3**:e1820.

77. Wingender E, Chen X, Hehl R, Karas H, Liebich I, Matys V, Meinhardt T, Prüss M, Reuter I, Schacherer F: **TRANSFAC: an integrated system for gene expression regulation.** *Nucleic Acids Res* 2000, **28**:316–319.

78. Chen H, Bieberich CJ: **Structural and functional analysis of domains mediating interaction between NKX-3.1 and PDEF.** *J Cell Biochem* 2005, **94**:168–177.

79. Kusy S, Gerby B, Goardon N, Gault N, Ferri F, Gérard D, Armstrong F, Ballerini P, Cayuela J-M, Baruchel A, et al: **NKX3.1 is a direct TAL1 target gene that mediates proliferation of TAL1-expressing human T cell acute lymphoblastic leukemia.** *J Exp Med* 2010, **207**:2141–2156.

80. Chen Z, Lund R, Aittokallio T, Kosonen M, Nevalainen O, Lahesmaa R: **Identification of novel IL-4/Stat6-regulated genes in T lymphocytes.** *J Immunology (Baltimore, Md: 1950)* 2003, **171**:3627–3635.

81. Ellis T, Gambardella L, Horcher M, Tschanz S, Capol J, Bertram P, Jochum W, Barrandon Y, Busslinger M: **The transcriptional repressor CDP (Cutl1) is essential for epithelial cell differentiation of the lung and the hair follicle.** *Genes Dev* 2001, **15**:2307–2319.

82. Sansregret L, Goulet B, Harada R, Wilson B, Leduy L, Bertoglio J, Nepveu A: **The p110 isoform of the CDP/Cux transcription factor accelerates entry into S phase.** *Mol Cell Biol* 2006, **26**:2441–2455.

83. Yao X, Nie H, Rojas IC, Harriss JV, Maika SD, Gottlieb PD, Rathbun G, Tucker PW: **The L2a element is a mouse CD8 silencer that interacts with MAR-binding proteins SATB1 and CDP.** *Mol Immunol* 2010, **48**:153–163.

84. Kaul-Ghanekar R, Jalota A, Pavithra L, Tucker P, Chattopadhyay S: **SMAR1 and Cux/CDP modulate chromatin and act as negative regulators of the TCRbeta enhancer (Ebeta).** *Nucleic Acids Res* 2004, **32**:4862–4875.

85. Li S, Moy L, Pittman N, Shue G, Aufiero B, Neufeld EJ, LeLeiko NS, Walsh MJ: **Transcriptional repression of the cystic fibrosis transmembrane conductance regulator gene, mediated by CCAAT displacement protein/cut homolog, is associated with histone deacetylation.** *J Biol Chem* 1999, **274**:7803–7815.

86. Nishio H, Walsh MJ: **CCAAT displacement protein/cut homolog recruits G9a histone lysine methyltransferase to repress transcription.** *Proc Natl Acad Sci U S A* 2004, **101**:11257–11262.

87. Barnes B, Lubyova B, Pitha PM: **On the role of IRF in host defense.** *J Interferon Cytokine Res: the official journal of the International Society for Interferon and Cytokine Research* 2002, **22**:59–71.

88. Zhu J, Paul WE: **Peripheral CD4+ T-cell differentiation regulated by networks of cytokines and transcription factors.** *Immunol Rev* 2010, **238**:247–262.

89. Elser B, Lohoff M, Kock S, Giaisi M, Kirchhoff S, Krammer PH, Li-Weber M: **IFN-gamma represses IL-4 expression via IRF-1 and IRF-2.** *Immunity* 2002, **17**:703–712.

90. Hu C-M, Jang SY, Fanzo JC, Pernis AB: **Modulation of T cell cytokine production by interferon regulatory factor-4.** *J Biol Chem* 2002, **277**:49238–49246.

91. Tominaga N, Ohkusu-Tsukada K, Udono H, Abe R, Matsuyama T, Yui K: **Development of Th1 and not Th2 immune responses in mice lacking IFN-regulatory factor-4.** *Int Immunol* 2003, **15**:1–10.

92. Lee C-G, Hwang W, Maeng K-E, Kwon H-K, So J-S, Sahoo A, Lee SH, Park ZY, Im S-H: **IRF4 regulates IL-10 gene expression in CD4(+) T cells through differential nuclear translocation.** *Cell Immunol* 2011, **268**:97–104.

93. Laver T, Nozell SE, Benveniste EN: **IFN-beta-mediated inhibition of IL-8 expression requires the ISGF3 components Stat1, Stat2, and IRF-9.** *J Interferon Cytokine Res: the official journal of the International Society for Interferon and Cytokine Research* 2008, **28**:13–23.

94. Lund RJ, Chen Z, Scheinin J, Lahesmaa R: **Early target genes of IL-12 and STAT4 signaling in Th cells.** *J Immunol* 2004, **172**:6775–6782.

95. Lazarevic V, Chen X, Shim J-H, Hwang E-S, Jang E, Bolm AN, Oukka M, Kuchroo VK, Glimcher LH: **T-bet represses T(H)17 differentiation by preventing Runx1-mediated activation of the gene encoding RORγt.** *Nat Immunol* 2011, **12**:96–104.

96. Toh M-L, Kawashima M, Hot A, Miossec P, Miossec P: **Role of IL-17 in the Th1 systemic defects in rheumatoid arthritis through selective IL-12Rbeta2 inhibition.** *Ann Rheum Dis* 2010, **69**:1562–1567.

97. Manel N, Unutmaz D, Littman DR: **The differentiation of human T(H)-17 cells requires transforming growth factor-beta and induction of the nuclear receptor RORgammat.** *Nat Immunol* 2008, **9**:641–649.

98. Bending D, Newland S, Krejcí A, Phillips JM, Bray S, Cooke A: **Epigenetic changes at Il12rb2 and Tbx21 in relation to plasticity behavior of Th17 cells.** *J Immunology (Baltimore, Md: 1950)* 2011, **186**:3373–3382.

99. Jin H, Sperka T, Herrlich P, Morrison H: **Tumorigenic transformation by CPI-17 through inhibition of a merlin phosphatase.** *Nature* 2006, **442**:576–579.

100. Yamashita M, Kimura M, Kubo M, Shimizu C, Tada T, Perlmutter RM, Nakayama T: **T cell antigen receptor-mediated activation of the Ras/mitogen-activated protein kinase pathway controls interleukin 4 receptor function and type-2 helper T cell differentiation.** *Proc Natl Acad Sci U S A* 1999, **96**:1024–1029.

101. Yamashita M, Shinnakasu R, Asou H, Kimura M, Hasegawa A, Hashimoto K, Hatano N, Ogata M, Nakayama T: **Ras-ERK MAPK cascade regulates GATA3 stability and Th2 differentiation through ubiquitin-proteasome pathway.** *J Biol Chem* 2005, **280**:29409–29419.

102. Tripathi P, Sahoo N, Ullah U, Kallionpaa H, Suneja A, Lahesmaa R, Rao KV: **A novel mechanism for ERK-dependent regulation of IL4 transcription during human Th2-cell differentiation.** *Immunol Cell Biol* 2011, .

103. Bermudez O, Pagès G, Gimond C: **The dual-specificity MAP kinase phosphatases: critical roles in development and cancer.** *Am J Physiol Cell Physiol* 2010, **299**:C189–202.

104. Reffas S, Schlegel W: **Compartment-specific regulation of extracellular signal-regulated kinase (ERK) and c-Jun N-terminal kinase (JNK) mitogen-activated protein kinases (MAPKs) by ERK-dependent and non-ERK-dependent inductions of MAPK phosphatase (MKP)-3 and MKP-1 in differentiating P19 cells.** *Biochem J* 2000, **352**:701–708.

105. Schweppe RE, Cheung TH, Ahn NG: **Global gene expression analysis of ERK5 and ERK1/2 signaling reveals a role for HIF-1 in ERK5-mediated responses.** *J Biol Chem* 2006, **281**:20993–21003.

106. Marchetti S, Gimond C, Chambard JC, Touboul T, Roux D, Pouyssegur J, Pages G: **Extracellular signal-regulated kinases phosphorylate mitogen-activated protein kinase phosphatase 3/DUSP6 at serines 159 and 197, two sites critical for its proteasomal degradation.** *Mol Cell Biol* 2005, **25**:854–864.

107. Nakamura K, Abarzua F, Kodama J, Hongo A, Nasu Y, Kumon H, Hiramatsu Y: **Expression of hepatocyte growth factor activator inhibitors (HAI-1 and HAI-2) in ovarian cancer.** *Int J Oncol* 2009, **34**:345–353.

108. Bergum C, List K: **Loss of the matriptase inhibitor HAI-2 during prostate cancer progression.** *Prostate* 2010, **70**:1422–1428.

109. Dong W, Chen X, Xie J, Sun P, Wu Y: **Epigenetic inactivation and tumor suppressor activity of HAI-2/SPINT2 in gastric cancer.** *Int J Cancer J Int du Cancer* 2010, **127**:1526–1534.

110. Ito W, Kanehiro A, Matsumoto K, Hirano A, Ono K, Maruyama H, Kataoka M, Nakamura T, Gelfand EW, Tanimoto M: **Hepatocyte growth factor attenuates airway hyperresponsiveness, inflammation, and remodeling.** *Am J Respir Cell Mol Biol* 2005, **32**:268–280.

111. Ito W, Takeda M, Tanabe M, Kihara J, Kato H, Chiba T, Yamaguchi K, Ueki S, Kanehiro A, Kayaba H, Chihara J: **Anti-allergic inflammatory effects of hepatocyte growth factor.** *Int Arch Allergy Immunol* 2008, **146**(1):82–87.

112. Beemiller P, Krummel MF: **Mediation of T-cell activation by actin meshworks.** *Cold Spring Harbor perspectives in biology* 2010, **2**:a002444.

113. Thauland TJ, Koguchi Y, Wetzel SA, Dustin ML, Parker DC: **Th1 and Th2 cells form morphologically distinct immunological synapses.** *J Immunology (Baltimore, Md: 1950)* 2008, **181**:393–399.

114. Bakshi R, Mehta AK, Sharma R, Maiti S, Pasha S, Brahmachari V: **Characterization of a human SWI2/SNF2 like protein hINO80: Demonstration of catalytic and DNA binding activity.** *Biochem Biophys Res Commun* 2006, **339**:313–320.

115. Chen HL, Li CF, Grigorian A, Tian WQ, Demetriou M: **T Cell Receptor Signaling Co-regulates Multiple Golgi Genes to Enhance N-Glycan Branching.** *J Biol Chem* 2009, **284**:32454–32461.

116. Toscano MA, Bianco GA, Ilarregui JM, Croci DO, Correale J, Hernandez JD, Zwirner NW, Poirier F, Riley EM, Baum LG, Rabinovich GA: **Differential glycosylation of T(H)1, T(H)2 and T-H-17 effector cells selectively regulates susceptibility to cell death.** *Nat Immunol* 2007, **8**:825–834.

117. Yssel H, de Vries JE, Koken M, Van Blitterswijk W, Spits H: **Serum-free medium for generation and propagation of functional human cytotoxic and helper T cell clones.** *J Immunol Methods* 1984, **72**:219–227.

118. Rasmussen CE, Williams CKI, Books24x7 Inc: **Gaussian processes for machine learning.** In *Adaptive computation and machine learning.* Cambridge, Mass: MIT Press; 2006.
119. Becker HW, Riordan J: **THE ARITHMETIC OF BELL AND STIRLING NUMBERS.** *Am J Math* 1948, **70**:385–394.
120. Edgar R, Domrachev M, Lash AE: **Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.** *Nucleic Acids Res* 2002, **30**:207–210.