



SUBJECT AREAS:
DIFFERENTIATION
GENE REGULATION
SYSTEMS BIOLOGY
CELL SIGNALLING

Received
27 April 2012

Accepted
1 June 2012

Published
18 June 2012

Correspondence and requests for materials should be addressed to K.V.S.R. (kanury@icgeb.res.in or kanury.rao@gmail.com)

Gene Set Enrichment Analysis Identifies LIF as a Negative Regulator of Human Th2 Cell Differentiation

Ubaid Ullah¹, Parul Tripathi¹, Riitta Lahesmaa² & Kanury V. S. Rao¹

¹Immunology Group International Centre for Genetic Engineering and Biotechnology Aruna Asaf Ali Marg New Delhi-110067 INDIA, ²Turku Centre for Biotechnology University of Turku and Abo Academy University Turku Graduate School of Biomedical Sciences FI-20520 Turku, FINLAND.

In this study we show that IL-4 is crucial during reinforcement window of human Th2 differentiation for optimal Th2 development. We have also shown here that during this stage, IL-4 helps in cellular decision-making process of differentiation versus proliferation. We have combined computational and experimental methods to analyze Th2 transcription network to name novel players of the process of Th2 differentiation. Here we report that LIF through STAT3 negatively regulates Th2 differentiation. This approach can be generalized to analyze “omics” data to identify key regulatory modules.

The process of Th2 differentiation is composed of three separate stages namely, initiation, reinforcement and maintenance¹. During initiation signaling downstream of TCR and IL-4R cooperate to activate proper signaling network required for Th2 differentiation to begin. This activates many transcription factors including STAT6 and GATA3 that then regulate *IL4* transcription. In the second stage, this endogenous IL-4 further feeds back through IL-4R to reinforce its own expression. And finally the last stage is marked by the culmination of the process through epigenetic changes in the *IL4* locus. Th2 cell differentiation is one of the most studied systems of cellular differentiation and the contributions many of the important players of the process are well documented particularly in mouse system²⁻¹². We have previously attempted at building a holistic model of human Th2 differentiation process¹³⁻¹⁵ and were the first to show gene expression kinetics and identify STAT6 targets during early human Th2 differentiation¹⁶. However, systems level understanding of later stages of Th2 differentiation is yet to be achieved.

In recent years, knowledge based methods such as Gene Set Enrichment Analysis at Broad Institute¹⁷ and Explain analysis system¹⁸ of Biobase database have been developed to aid in extracting biologically interesting information using the genome wide transcriptome data.

Although the requirement of IL-4 for Th2 differentiation was known, we wanted to investigate the effects of IL-4R signaling during reinforcement window of the process. An unbiased global transcriptome profiling along with GSEA was used to identify all the relevant pathways over represented during the process. Using a network approach we also identified hubs of Th2 transcription network that may be crucial for Th2 development. We further experimentally characterized one of the master regulator STAT3 and its upstream protein Leukemia inhibitory factor (LIF). LIF is a pleiotropic glycoprotein of IL-6 family. Its role has been well characterized in maintaining pluripotency in stem cells and in the context of Th cell differentiation it has been shown to suppress Th17 phenotype¹⁹. Here we begin to characterize the role this protein in human Th2 cell development.

Results

Continuous presence of IL-4 is required for optimal Th2 development. IL-4 drives both the initiation and the reinforcement stages of the Th2 polarization response^{1,20}. In view of the kinetics of the process though, we speculated that endogenous IL-4 would be specifically involved during the reinforcement/commitment of Th2 lineage specification. To test this we generated parallel cultures of either Th activation, or Th2 differentiation and added to these a combination of neutralizing antibodies against IL-4 and IL-4R. Antibodies were first added at day three after initiation of the cultures, and then supplemented every 24 h up to day 5. This timing of addition was based on the fact that IL-4 transcription peaks at around day3 post stimulation⁷. Neutralization of exogenous IL-4 significantly reduced the extent of Th2 cell generation ($P \leq 0.05$, Figure 1a). Further, this inhibition was also

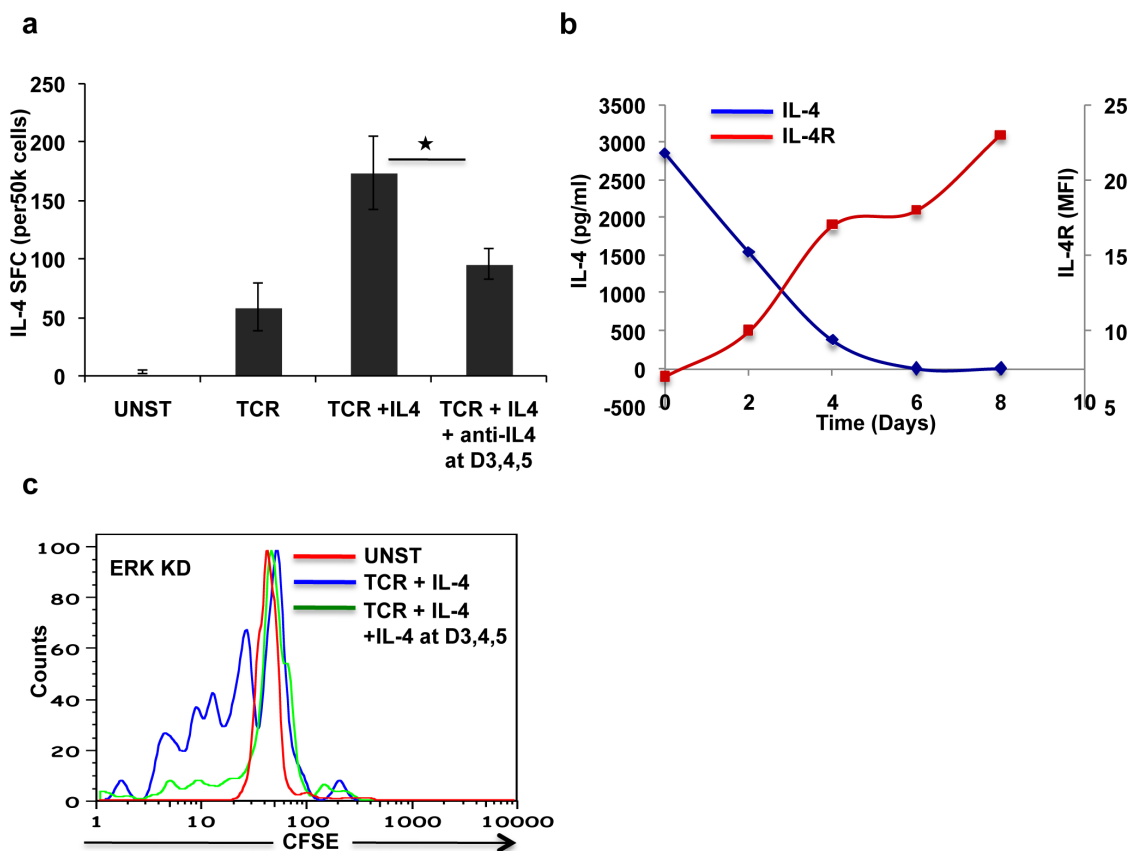


Figure 1 | Continuous IL-4 is required for optimal Th2 development. **Panel a** shows results of IL-4 ELISPOT experiments. Culture conditions are mentioned at the bottom. Error bars represent standard error of mean and * denotes significant (p value 0.017) difference between the two sample (Two tailed paired sample t test). In **panel b** blue line represents IL-4 level detected in the supernatant at given time points and red line is a measure of IL-4R expression level at corresponding time points. **Panel c** shows a result obtained in a CFSE labeling experiment in ERK silenced cells. Red, blue and green colors represent not activated, TCR + IL-4 stimulated and TCR + IL-4 stimulated at day 1 followed by additional doses of IL-4 at day 3,4 and 5. Less fluorescence indicates more cell divisions because as the cells divide fluorescent label is diluted.

observed in TCR activated T cells (“Data not shown”), supporting that this was likely due to depletion of the endogenously produced cytokine (Figure 1a). The partial attenuation seen may well derive from the possibility that the added antibodies were only partly effective at neutralizing the autocrine/paracrine action of the endogenously produced IL-4.

We also probed the puzzling question of why endogenous IL-4 was necessary even under Th2 polarizing conditions, where a saturating amount of exogenous IL-4 was added at the beginning of the culture. For this, we collected supernatants at regular intervals from cultures of naive cells subjected to Th2 polarization, and estimated amount of IL-4 present. In parallel, we also took aliquots of the cells at select time points to determine the levels of surface expression of the IL-4 receptor (IL-4R). We found that the concentration of the externally added IL-4 was continuously declining such that IL-4 levels were barely detectable by 120 hours. Interestingly, this decline correlated well with a reciprocal increase in cell surface expression of IL-4R that was evident from the 48 hours time point onwards (Figure 1b). Thus, the observed decrease in IL-4 concentration in the culture supernatants likely derives from its consumption by IL-4R, whose surface levels were rapidly being upregulated over the same time span. This would then rationalize the requirement for endogenous IL-4, during the later stages of polarization.

Endogenous IL-4 helps in differentiation versus proliferation decision making process. We recently showed that ERK regulates human Th2 differentiation and it does so by activating IL-4 transcription and that siRNA mediated silencing of ERK leads to

significant reduction in number of IL-4 producing cells⁷. Therefore we wanted to test whether endogenous IL-4 helps in decision of proliferation versus differentiation processes of the cell. For this, GFP- or ERK-silenced cells were labeled with CFSE and their proliferation rates were compared under conditions of Th2 polarization. Relative to that of GFP-silenced cells, depletion of ERK resulted in a significant increase in the number of cell doublings by day 6 of the culture. The addition of IL-4 at and after day 3 to day 5, however, substantially reversed this effect (Figure 1c) suggesting that continuous presence of endogenous IL-4 during the process of Th2 differentiation may act by limiting the proliferative capacity of cells and, presumably, thereby biasing towards differentiation.

At this juncture, it was of interest to investigate the global effects of IL-4 during reinforcement stage of Th2 differentiation. We deployed similar neutralization strategy as in Figure 1a. Whole genome transcription profiling was done to identify the target of IL-4 and eventually novel players of Th2 transcription network.

Identification of differentially regulated genes. To recapitulate the methodology, Stimulation of cells was done at day 0 through TCR in the presence of IL-4 and at day 3 half the cells were treated with anti IL-4 and anti-IL-4R neutralizing antibodies. RNA was isolated and hybridized to a human genome array. Feature extraction and normalization was done as described in the methods section. Differentially expressed genes thus identified are shown in Table 1. More than 3000 genes were upregulated in both the samples as compared to unstimulated cells, and similar statistics were obtained for down regulated genes. The complete list of differentially regulated genes



Table 1 | Differentially regulated genes in different samples

	Upregulated	Downregulated
TCR + IL-4	3638	3300
TCR + IL-4 + ABS	3686	3680
IL-4 Dependent Genes	128	151

can be accessed through Geo database at NCBI (Accession Number GSE 16022).

Identification of IL-4 dependent genes their GO classes. IL-4 dependent genes were identified using the methodology detailed in the method section. The exercise resulted in identification of 128 IL-4 dependent upregulated genes, and 151 IL-4 dependent down regulated genes. The list of these up and down regulated genes along with their functions, as described in Biobase Knowledge Library, (BKL) (www.biobase-international.com) are given as Supplementary Table 2 and 3 respectively. To identify if any Gene Ontology (GO) classes are enriched in any of these two differentially expressed gene sets, GO analysis was performed on “Explain Analysis System” of Biobase (TRANSFAC) database. Many GO classes were enriched in the upregulated gene set including cell activation, cell differentiation and regulation of T cell anergy at the p value threshold of 0.01 as shown in Figure 2. However at this cut off only two GO classes were enriched in down regulated gene set (Developmental process and Multi-cellular organismal process) and the overall low enrichment was retained even at p value threshold of 0.1. The observed difference in the enrichment in the two classes may partly be attributed to the ability of IL-4 to specifically upregulate target genes.

MATCH analysis identifies key TFBS in the promoter of IL-4 dependent genes. Further to identify which transcription factors might be important in regulating the expression of these genes, MATCH analysis was done again on “Explain Analysis System” of Biobase database. MATCH analysis was done on three test-background set combinations. In the first two cases up and down regulated gene sets were each used as test set against the background of human house keeping gene promoter sets. In the third analysis, TFBS search was performed in the upregulated set down regulated sets using downregulated set as background. Results are displayed in Table 2. Enriched matrices are shown along with Yes/No ratio and p value. “Yes/No” is the ratio of the average number of putative binding sites per 1000 base pair for the query and the background sets. Value greater than 1 indicates an over representation of the motif in the query set.

Only two matrices were enriched in Up versus Down MATCH (Table 4) analysis pointing towards the fact that most of transcription factors (TFs) required for up and down regulation are probably the same in case of IL-4. However, two matrices unique to upregulated group are V\$HTF_01 (“V\$”) is matrix nomenclature adapted by TRANSFAC database) and V\$NFKB_C. Although NFKB is a well studied molecule and known to function in IL-4 dependent Th2 differentiation process²¹, role of Xbp-1 (TF represented by V\$HTF) in T cell activation or differentiation is not investigated. Xbp-1 is known to be necessary for plasma B cell differentiation and this analysis suggest future experiment for investigating the role of this molecule in T cell differentiation.

A closer look at Table 2 and 3 also suggests that there are two matrices that are present in Up versus housekeeping comparison but absent in Down versus housekeeping comparison and these are NFKB and MYOD suggesting that MYOD, which is critical for muscle cell differentiation, could also play a prominent role in T cell differentiation.

Gene Set Enrichment Analysis of whole dataset reveals significant enriched pathways. Analysis of gene expression data by identifying differentially expressed genes and interpreting biological information independently for each gene has several limitation as outlined elsewhere¹⁷. These include low signal to noise ratio owing to high noise in microarray data, and the possibility of missing crucial effects on pathways. GSEA, on the contrary, analyzes whole data at biological pathway level by performing unbiased global searches for genes that are coordinately up or down regulated in predefined biological gene sets. GSEA first ranks the original microarray data on the basis of expression profile in two groups test and control. And then in the ranked list it asks whether a significant number of genes from predefined biological sets occur towards the top or bottom of ranked list.

To identify pathways enriched in TIL4 treated samples over the samples treated with neutralizing antibodies, GSEA was performed on day 6 transcription data. Results are shown in Table 5. Overview of GSEA analysis is illustrated in Figure 3a and b. Panel a shows distribution of enrichment score (ES) of all the gene sets enriched. On the X axis, gene sets that are present before “0” mark were overrepresented in the background set and those that are present after it were the sets enriched in the test set (TIL4 as opposed to TIL4ABS group). A dot plot of significance and Normalized Enrichment Score NES for is given in Figure 3b. This Figure (Figure 3b) shows that gene sets bearing NES, either less than -1 or more than +1 are significant at the cut off FDR q value of less than 25 %. Although in TIL4ABS samples, there were many gene sets with nominal p value less than 0.25, or even 0.05, but none of them was enriched at FDR q value of less than 0.25. The nominal p value is less relevant for comparing across the gene sets than FDR q value. Parameters used for the analysis are detailed in Methods section. Details on statistics used are described elsewhere¹⁷. As shown in Table 5 many pathways known to be associated with cytokine signaling were enriched in TCR + IL-4 treated sample. Enriched pathways can be classified under the heads of signaling pathways and metabolic pathways.

Signaling pathways associated with IL-4. Many of the signaling pathways known to be essential for T cell activation and differentiation are enriched in TCR + IL-4 treated samples. Although the genes in these pathways were marginally upregulated and would have been missed these in a single gene analysis, GSEA successfully identifies pathways known to be central in transducing cytokine signal owing to its ability to analyze the whole transcriptome data at pathway level. The most significant hit was JAK-STAT pathway followed by “Cytokine and cytokine receptor interaction”. None of these pathways came as a surprise because these were expected to be enriched.

Gene set “Differentiation pathway in PC12 cells” from STKE (Signal Transduction Knowledge Environment) is significantly enriched at FDR q value of less than 0.04 indicating that the pathway is enriched in TIL4 samples at the highest level of significance (Less significant but acceptable levels for FDR q value being 0.1 and 0.25). It may, therefore, be of critical importance to study some of the enriched genes from this pathway for their involvement in T helper cell differentiation. Some of the core-enriched genes from this pathway such as Egr1/2/3 are already known to be important for Th1/2 differentiation^{22,23}. Other signaling pathways enriched were Calcineurine NF-AT signaling and ATR BRCA pathway. Because Calcineurine NFAT signaling plays a crucial role in T cell activation, enrichment of this pathway along with other pathway of known implications in Th2 differentiation strengthen the reliability of GSEA results.

Although IL-4 is known to have both positive and negative effects in various types of cancers^{24–27} its association to leukemia’s of Fanconi anemia (FA) type has not been investigated in detail. Enrichment of ATR BRCA pathway that has many genes common



Figure 2 | Gene Ontology of IL-4 dependent upregulated genes. Enriched gene ontology (GO) categories in IL-4 dependent upregulated genes using human housekeeping genes as a background set. Blue bar represents actual enriched number of genes while red bar represents expected enrichment by chance.

with FA suggests a possible role of IL-4 in these lymphocyte malignancies. Cell cycle Pathway is also enriched, albeit at low significance (FDR q value 0.21). Core enriched genes of this pathway are pro-proliferative and probably signal for “go ahead” to the cell cycle at various stages. These genes include CCND2 (required for G1/S transition), CDC25C (triggers mitosis and suppresses growth arrest by p53) and CCNA (required both at the start and G2/M transition). The whole list of coordinately upregulated genes in all the enriched pathways is given as Supple Table II. A closer inspection of the ranked list reveals many noteworthy points. For example, DUSP2 and DUSP9, which are known to dephosphorylate ERK, are at the

bottom of the ranked list while other dual specificity phosphatase, DUSP13, which is not specific for ERK1/2, is towards the top indicating the possibility that ERK1/2 plays a dominant role in transmitting signals from IL-4R.

Metabolic Pathways associated with IL-4. Metabolic pathways enriched were Folate biosynthesis, amino acid and prostaglandin metabolism as outlined in Table 5. Folate plays a critical role in protein and nucleic acid synthesis²⁸, and low Folate level has been associated with decreased expression of immune related proteins especially those involved in inflammation and complement



Table 2 | Overrepresented matrices in IL-4 dependent upregulated genes. Promoters of human house keeping genes were taken as background set. Yes/No is the ratio of occurrence of these matrices in test and background set

Matrix name	Yes/No	P-value
V\$AP1_C	1.1373	0.000
V\$CEBP_Q2_01	1.1052	0.005
V\$CEBP_Q3	1.1191	0.001
V\$CEBPB_01	1.1518	0.001
V\$CEBPB_02	1.156	0.001
V\$CEBPGAMMA_Q6	1.3287	0.000
V\$E12_Q6	1.3096	0.001
V\$E2A_Q2	1.2657	0.007
V\$E2A_Q6	1.2712	0.001
V\$E47_01	1.3222	0.000
V\$E47_02	1.2744	0.001
V\$EBOX_Q6_01	1.2207	0.000
V\$FOXO4_01	1.2718	0.003
V\$FOXO4_02	1.1895	0.001
V\$GATA3_03	1.1256	0.004
V\$HMGYI_Q3	1.5328	0.000
V\$IRF2_01	1.4039	0.009
V\$IRF_Q6_01	1.2397	0.000
V\$LEF1_Q2	1.1034	0.003
V\$LEF1_Q2_01	1.2437	0.009
V\$LEF1TCF1_Q4	1.1359	0.004
V\$LMO2COM_01	1.2777	0.001
V\$MYOD_Q6_01	1.2051	0.005
V\$NFKAPPAB_01	1.1103	0.006
V\$OCT1_01	1.3849	0.000
V\$OCT1_02	1.1441	0.001
V\$OCT1_04	1.5574	0.000
V\$OCT1_05	1.3651	0.003
V\$OCT1_06	1.3025	0.000
V\$OCT1_07	1.422	0.000
V\$OCT1_B	1.3286	0.000
V\$OCT1_Q5_01	1.3137	0.000
V\$OCT1_Q6	1.2044	0.003
V\$OCT_C	1.2071	0.000
V\$OCT_Q6	1.3224	0.000
V\$SOX_Q6	1.1702	0.000
V\$STAT_01	1.2474	0.001
V\$TATA_01	1.4071	0.000
V\$TBP_Q6	1.2767	0.000
V\$YY1_01	1.1021	0.002

activation^{29,30}. Given its functions it was cogent for it to be crucial in the pro-proliferative function of IL-4, which explains its enrichment in TIL4 group over TIL4ABS group. Genes from enriched metabolism pathways listed in Table 5 suggest that differentiation supports or requires synthesis of more amino acids of acidic nature (ASP GLU) while at the same time degrading hydrophobic residue (ALA) indicating the possibility of existence of preference for amino acids of a kind for a particular cellular process. Although this may be exciting, it requires further experimental validation.

Network of core enriched genes follow a power law degree distribution. Although analysis of individual enriched pathways gave fascinating insights into the pleiotropic functions of IL-4, it was pertinent to study whether these individual pathways integrate to form a functional network. For this, a functional association network was made out of all core enriched genes from all enriched pathways using STRING database. No additional white nodes (Additional interactors in the literature but not in the input data) were allowed to ensure that only associations among this set of proteins be picked up. The functional association networks thus generated (Figure 3c) followed scale free topology (Figure 3d), which is said to be existing in a network if the degree distribution

Table 3 | Overrepresented matrices in IL-4 dependent downregulated genes. Promoters of human house keeping genes were taken as background set. Yes/No is the ratio of occurrence of these matrices in test and background set

Matrix name	Yes/No	P-value
V\$AP1_01	1.205	0.000
V\$AP1_C	1.3405	0.000
V\$CEBP_Q2_01	1.2731	0.000
V\$CEBP_Q3	1.2553	0.000
V\$CEBPB_01	1.3246	0.000
V\$CEBPB_02	1.2979	0.000
V\$CEBPDDELTA_Q6	1.3411	0.000
V\$CEBPGAMMA_Q6	1.4482	0.000
V\$E2A_Q6	1.2531	0.001
V\$E47_01	1.2284	0.006
V\$E47_02	1.2884	0.001
V\$EBOX_Q6_01	1.2466	0.000
V\$FOXO4_01	1.5627	0.000
V\$FOXO4_02	1.4478	0.000
V\$GATA_C	1.2076	0.000
V\$GATA_Q6	1.3056	0.000
V\$GATA3_02	1.3429	0.000
V\$GATA3_03	1.2502	0.000
V\$HMGYI_Q3	1.6228	0.000
V\$IRF_Q6	1.6027	0.005
V\$IRF_Q6_01	1.4153	0.000
V\$IRF1_01	1.5781	0.003
V\$IRF1_Q6	1.3482	0.000
V\$LEF1_Q2_01	1.3735	0.000
V\$LEF1TCF1_Q4	1.2864	0.000
V\$LMO2COM_01	1.2579	0.002
V\$LMO2COM_02	1.3388	0.009
V\$OCT_C	1.3966	0.000
V\$OCT_Q6	1.5614	0.000
V\$OCT1_01	1.6683	0.000
V\$OCT1_02	1.4038	0.000
V\$OCT1_04	1.7378	0.000
V\$OCT1_05	1.6232	0.000
V\$OCT1_06	1.5706	0.000
V\$OCT1_07	1.9252	0.000
V\$OCT1_B	1.648	0.000
V\$OCT1_Q5_01	1.6103	0.000
V\$OCT1_Q6	1.4729	0.000
V\$SOX_Q6	1.2634	0.000
V\$STAT_01	1.277	0.000
V\$STAT1_05	1.216	0.003
V\$TATA_01	1.5219	0.000
V\$TBP_Q6	1.4944	0.000
V\$YY1_01	1.2633	0.000

follows the power law. Power law in this context would mean that there are extremely few nodes with exceedingly high degree while a majority of the nodes only have few edges. This hub topological feature of biological networks make them robust against random node failure.

STAT3 and JUN are key nodes of Th2 network at day 6. The network thus generated was analyzed using “network analyzer” plug-in of Cytoscape. This plug-in calculated all the relevant parameters such as degree, clustering coefficient and other

Table 4 | Overrepresented matrices in IL-4 dependent upregulated gene set with downregulated genes as background set

Matrix name	Yes/No	P-value
V\$HTF_01	1.2341	0.00043443
V\$NFKB_C	1.2448	0.0025



Table 5 | GSEA enriched pathways at day 6. Only pathways that are enriched at p value threshold of 0.05 and FDR q value of 0.25 are shown

Name of the enriched pathway	NES	NOM p value	FDR q value
JAK-STAT pathway	1.94	0.00	0.04
Differentiation pathway in PC12 cells	1.91	0.00	0.04
Cytokine-cytokine receptor interaction	1.88	0.00	0.05
Folate biosynthesis	1.86	0.00	0.05
Hematopoietic cell lineage	1.83	0.00	0.06
DNA replication reactome	1.74	0.00	0.12
Valine, Leucine and Isoleucine degradation	1.70	0.01	0.17
Prostaglandin and leukotriene metabolism	1.69	0.01	0.17
Calcineurine NF-AT signaling	1.66	0.01	0.19
Alanine and Aspartate metabolism	1.66	0.02	0.18
ATR BRCA pathway	1.64	0.02	0.19
Glutamate metabolism	1.64	0.02	0.17
Cell cycle	1.61	0.00	0.21
GPCRR DB	1.58	0.05	0.25

centrality measures. Betweenness centrality (BC) is one such parameter, which, for a vertex, is a measure of how frequently the vertex lies on a randomly chosen shortest path. It shows how much is the load share of this vertex in a network³¹ meaning the extent of

information flow through this node. Both node size and intensity of color represent Betweenness centrality. As is evident PCNA, STAT3 and JUN as the nodes of highest “importance”(Figure 3b). The observation that STAT3 is coming as a key regulatory node as

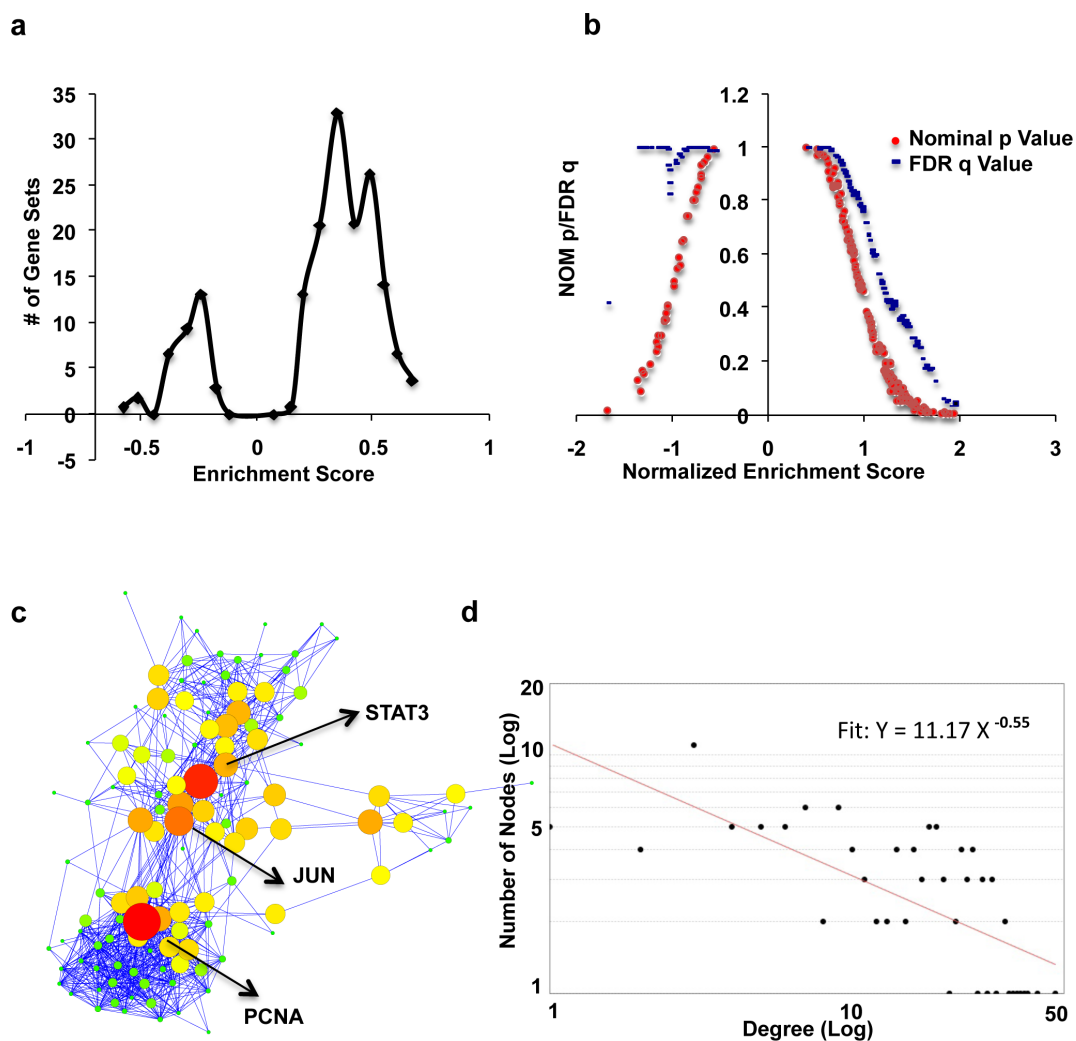


Figure 3 | Gene Set Enrichment Analysis (GSEA). Panel a shows enrichment score (ES) distribution of all the enriched gene sets. **Panel b** shows variation of p value (Red) and q value (Blue) with Normalized Enrichment Score (NES). Please see text for detail. **Panel c** is a network of all core enriched genes of all the pathways taken together. Node size is a measure of Betweenness Centrality (BC). Larger is the size higher is the BC. Nodes with highest BC values are highlighted in yellow. **Panel d** shows degree distribution of network shown in **panel c** with fitted power law line.

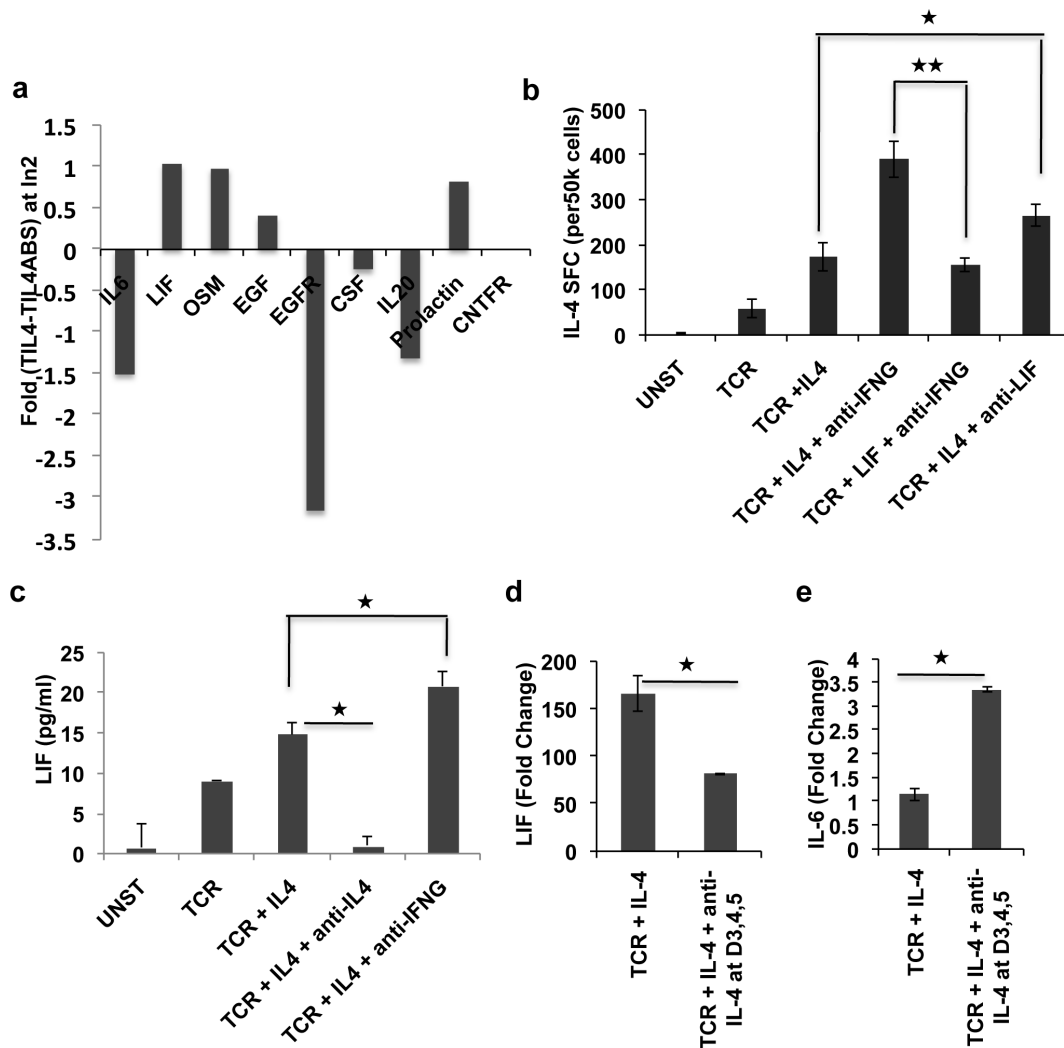


Figure 4 | LIF is secreted by Th2 cells and it negatively regulates IL-4 production. **Panel a** shows mRNA expression profile of STAT3 upstream molecules. **Panel b** shows ELISPOT results for IL-4 producing cells in differentially activated populations. **Panel c** shows results obtained in LIF ELISA experiments. **Panel d** and **e** are mRNA expression data adapted from microarray of LIF and IL-6 respectively. Fold change is calculated with respect to untreated sample. In all cases mean of three biological and two technical replicates are shown. Error bars represent standard deviation and * represents significance at p value of less than 0.05 (see text).

opposed STAT6 was initially surprising because STAT6 and not STAT3, along with GATA3, was known to be the lineage specific transcription factor for Th2 development^{4,6,9}. However, a recent report showed that STAT3 synergizes with STAT6 to give optimal Th2 development⁵.

LIF-IL-6-STAT3 axis may regulate outcome of Th2 differentiation.

Next we sought to investigate which upstream molecules control STAT3 activity during Th2 development. All the known upstream molecules that target STAT3 were searched in the microarray data. We found that LIF was the only molecule that was going up in TCR + IL-4 treated sample as compared to the unstimulated and this upregulation was reversed upon addition of neutralization antibody against IL-4 and IL-4R (Figure 4a). Given the role of LIF as a negative regulator of Th17 differentiation¹⁹ and the fact that STAT3 is a known target of LIF and that STAT3 is a positive regulator of Th2 differentiation⁵, we hypothesized that LIF may positively regulate Th2 differentiation.

To investigate the role of LIF in Th2 development we decided to test: A) whether LIF is secreted by Th2 cells, B) can LIF substitute for IL-4 in Th2 development and C) what is the effect of LIF on phosphorylation of STAT3.

LIF is secreted by Th2 cells. LIF is known to be secreted by activated T cells^{32,33}, so here we sought to investigate if LIF is secreted in the supernatant of T cell cultures which are differentiated in Th2 environment for 7 days. Results are shown in Figure 4c. Although the overall expression is very low (in pg only), a definite trend can be seen. TCR + IL-4 + anti-IFN γ sample had highest LIF secretion followed by TCR + IL-4 and TCR alone. Interestingly in the sample where IL-4 signaling was blocked at the later stage, the level of LIF was comparable to that what is produced in the unstimulated sample suggesting that possibly Th2 cells produce more LIF than other subsets. At the mRNA level also, the TCR + IL-4 samples had significantly higher level of LIF over TCR + IL-4 + anti-IL-4 samples. Given the known antagonism of LIF and IL-6 we also checked the expression of IL-6 in the whole genome transcriptome data and found that indeed there exists antipathy between the two (Figure 4d-e).

LIF negatively regulates Th2 differentiation. To address whether LIF has positive or negative effect on Th2 development, cells were stimulated through TCR in the presence of cytokine and or neutralizing antibody as mentioned in the Figure 4b. Stimulation media were supplemented again at day 3, and after 6 days cells

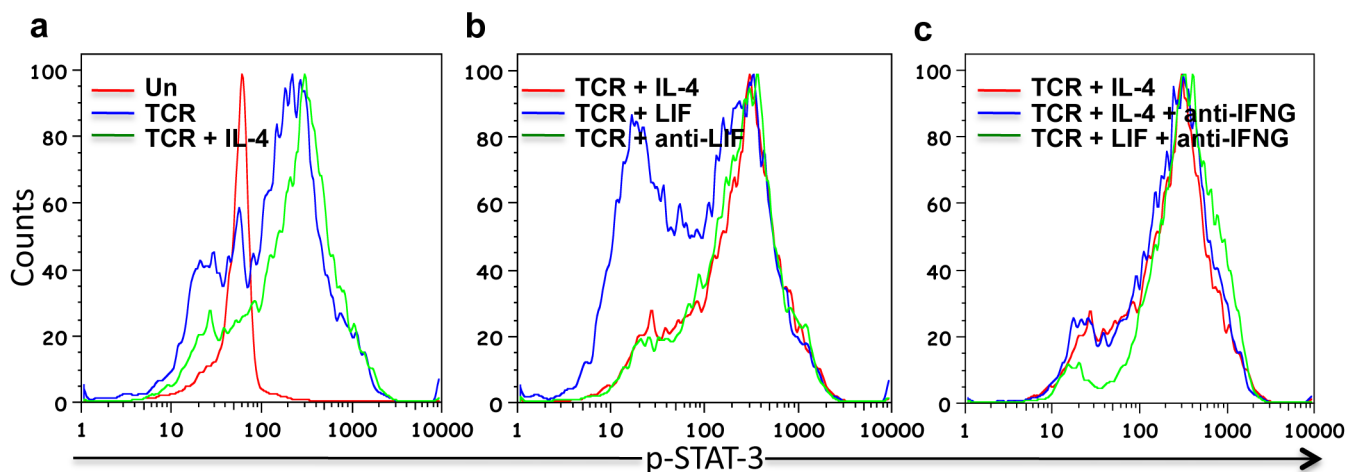


Figure 5 | STAT3 Phosphorylation in response to different stimuli. Panel a shows pSTAT3 profile obtained in a flow experiment. Unstimulated, TCR triggered and TCR + IL4 treated samples are shown in Red, Blue or Green color respectively. In panel b Red is TCR + IL-4 treated, Blue is TCR + LIF treated, green is TCR + anti-LIF treated. In panel c TCR + IL-4 is Red, TCR + IL-4 + anti-IFNG is Blue and TCR + LIF + anti-IFNG is green. Results represent experiment done at least in duplicate.

were washed and restimulated with PMA for 24 hours and ELISPOT assay was done to determine the fraction of IL-4 secreting cells. As expected more IL-4 producing cells were produced where IL-4 was present (TCR + IL-4) during trigger, as opposed to those where no IL-4 was added initially. Most notably though, addition of anti-LIF neutralizing antibody yielded exactly the opposite result and IL-4 producing cell population grew significantly over TCR + IL-4 sample. Even in the presence of anti-IFN γ , LIF appears to down-grade development of Th2 differentiation (Figure 4b). Thus, overall interpretation of Figure 4b supports the notion that LIF is a negative regulator of Th2 differentiation process. This role of LIF is in line with its role in promoting stemness in stem cells³⁴ and preventing it from differentiating to any lineages.

LIF down modulates STAT3 phosphorylation. A recent study reported the role of STAT3 in Th2 differentiation and there it was observed that STAT3 phosphorylation peaks at day 3 followed by a steady decline. We, therefore, studied STAT3 phosphorylation by LIF and IL-4 at day 3 post stimulation. Figure 5a shows pSTAT3 profile either in unstimulated cells (Red) or after TCR triggering in presence (Green) or absence (Blue) of IL-4. We observed that both TCR and TCR + IL-4 show similar profile with TCR + IL-4 having slightly more phosphorylation than TCR alone. TCR + LIF treated cells, on the other hand, had much less phosphorylation, as compared to TCR or TCR + IL-4 treated cells suggesting that LIF may inhibit sustained STAT3 phosphorylation (Figure 5b). We also observed pSTAT3 level in the presence of anti-IFN γ and found that the profile for TCR + IL-4 in the presence (Blue) or absence (Red) of anti-IFN γ were similar (Figure 5c). Thus, taken together results in Figure 5 suggest that although LIF is secreted by Th2 cells it inhibits the differentiation of naïve cells into Th2 by reducing sustained phosphorylation of STAT3.

Discussion

IL-4 is not only a signature cytokine for Th2 but also elementally required for Th2 development. It has been shown that even TCR stimulus is sufficient to drive naïve CD4 positive cells toward Th2 suggesting that during initial trigger, presence of IL-4 was not essential^{35,36}. However, we report here that during reinforcement window it is required for optimal Th2 development as blocking IL-4 signaling during this period resulted in significant decline in number of Th2 cells generated at the end of the culture period.

In order to get a global understanding of the process and to identify all the targets of IL-4 during this window, complete

transcriptome analysis was done. Transcription factor binding site search of IL-4 dependent upregulated genes led to identification of IRF2 binding site in these genes, which competitively inhibits IRF1 mediated transcriptional activation of interferon genes. GSEA is a knowledge based recent method extensively used (2476 citations as of 9th April 2012) for analyzing transcription data¹⁷. Here in this study GSEA has been used to identify IL-4 dependent pathways. Core enriched genes of each of these pathways were then separately analyzed using a network based approach. On one hand, this study gives additional insight into the process of Th2 differentiation and on the other it represents an approach that can be adapted to analyze high throughput “omics” data for key regulatory modules.

The point that STAT3 is emerging as a crucial node in day 6 Th2 transcription network goes hand in hand with the observation that STAT3 binding site is present in *IL4* promoter, and also the recent report that STAT3 is indeed required for optimal Th2 development. It may also be worth mentioning that a similar analysis of day 2 data suggested a significant role for SHC-1 adapter molecule during early Th2 development (“Unpublished observation”). Since STAT3 was already a known player of Th2 differentiation, we hypothesized that there may be upstream molecules that regulate STAT3 activity and thereby Th2 differentiation. LIF, Oncostatin M (OSM) and Prolactin were upregulated in differentiated Th2 cells, but only LIF was selected for detailed analysis because LIF and OSM belong to the same family of cytokines and the two also shared a receptor chain to transmit the signal. LIF is known to function both in favoring differentiation in some cell types, and inhibiting differentiation in stem cells where it helps to maintain self-renewing phenotypes. Recently it was shown to inhibit Th17 cell differentiation, and therefore we thought that it might promote the Th2 phenotype. At the same time given its opposite role in different situations it was appropriate to rationalize that it may inhibit Th2 development. Thus in these settings, as shown, LIF inhibited Th2 differentiation by preventing sustained phosphorylation of STAT3. It is also fascinating to note that although stimulation with LIF alone led to substantial reduction in STAT3 phosphorylation, this effect was reversed when anti-IFN γ was present during stimulation. Thus, while LIF is secreted by Th2 cells, albeit in low amounts, it negatively feeds back into IL-4 by down-regulating Th2 differentiation.

Methods

Reagents and cell culture. All cell culture experiments were performed in RPMI 1640 media supplemented with penicillin (100 U/ml)/streptomycin (100 μ g/ml) and 10% FCS (Life Technologies). CD4⁺ T cell positive isolation kits were obtained from Dynal



Biotech. Recombinant human IL-4, IL-2 cytokines, anti-human IL-4 and IL-4R neutralizing antibodies, were procured from R & D Systems. Phycolpaque was obtained from Amersham Pharmacia Biotech. Anti-human CD3 and CD28 were obtained from BD Biosciences.

Isolation of human CD4⁺ lymphocytes from umbilical cord blood and their stimulation. We obtained approval for this work by the institutional ethics committee of International Centre for Genetic Engineering and Biotechnology, and human blood samples were obtained with written informed consent. Purified CD4⁺T cells were isolated from cord blood mononuclear cells (CBMCs) using positive CD4⁺T cell isolation kits from Dynal following manufacturers protocol. The resulting populations were >98% viable, >98% CD3⁺, CD4⁺ and CD45RA⁺T cells. These cells were stimulated with plate bound anti-CD3 (2.5 µg/ml in PBS) and 500 ng/ml soluble anti-CD28 Monoclonal Ab in the presence recombinant human IL-4 (Referred as TIL4 stimulation in the text) or TCR stimulation in presence of anti IL-4 and anti IL-4R neutralizing antibodies (Referred as ABS stimulation in the text). rhIL-2 (17 ng/ml) was added 48 hours after the initiation of polarization. Cell concentration was maintained in the range of 0.5–2 × 10⁶ cells/ml.

Transfection of cells with siRNA. Nucleofections were carried out in 100 µl Opti-MEM I (Invitrogen, Carlsbad, CA, USA) using Nucleofector™ device (Amaxa, Cologne, Germany)^{16,37}. After electric shock cuvette was left unattended for 10 minutes following which cells were transferred to culture plates in 10% complete RPMI medium. After 36 hours the cells were harvested, lysed in lysis buffer and the efficiency of knock down was determined by Western Blotting⁷.

ELISA. R & D duo set ELISA kits (catalogue number DY204) were used for detection of IL-4 in culture supernatant following manufacturer's protocol. Briefly, plates were coated with capture antibody and incubated overnight at 4°C. Next day plates were washed with wash buffer (supplied) and samples were added after blocking the plate with block buffer (supplied). Plates were washed and biotin coated detection antibody added for 2 hours at room temperature. Signal was developed by adding streptavidin-HRP followed by addition of substrate (R & D). Reaction was stopped using 2N H₂SO₄. Optical density of each well standard and sample was determined at 450 nm.

ELISPOT assay. ELISPOT assays for IL-4 were performed with human IL-4 ELISPOT kits (R & D systems). For this, polarized cells were re-stimulated after 7 days of culture with PMA (5 ng/ml) for 18–20 h. Manufacturer's instructions were followed. Briefly a total of 50,000 cells were spotted per well on a plate precoated with respective antibodies. Plate incubated overnight at 37°C in CO₂ incubator. Cells were removed, biotin labeled detection antibody was followed by washing and addition of streptavidin labeled alkaline phosphatase. BCIP/NBT substrate was then added that generated black brown spot at respective locus where cells secreting that cytokine was found. Spot forming units were counted and analyzed by using an immunospot reader (CTL Analyzers, LLC, Cleveland, OH).

Flow Cytometry. For surface staining cells were harvested and washed in PBS. Approximately one million cells per tube were taken, blocked in 2% FCS in PBS or 0.5% BSA in PBS for 10–20 minutes on ice followed by staining as recommended by antibody provider, one hour on ice or RT. Cells were then washed in FACS wash buffer (0.5% BSA in PBS). Appropriate single stain control was taken for compensation when doing multicolor flow cytometry.

For intracellular staining cells were fixed in 3–4% Formaldehyde for 10 minutes at room temperature followed by permeabilization in 90% methanol for 30 minutes on ice. Non specific blocking was done by 0.5% BSA in PBS for 10 min at RT followed by surface staining, if any, followed by washing and intracellular staining by either labeled primary antibody or unlabeled primary with labeled secondary antibody. After each staining step cells were washed prior to addition of new antibody with FACS wash buffer. After last staining step cells were washed and resuspended in 500 µl of PBS/wash buffer for acquisition.

For CFSE labeling cells were washed with RPMI to remove any FCS, 10 million cells were labeled per ml of RPMI in final concentration of 5 µM at 37 degrees in incubator for 10 minutes. An equal volume of 100% FCS was used to stop labeling and cells were washed three times prior to culture. In all cases data were acquired on FACSCalibur and analyzed using FlowJo.

RNA Isolation, cDNA preparation, microarray hybridization and scanning. After 6 day of culture cells were harvested and kept in RNALater until RNA isolation. RNA was purified using Qiagen's RNeasy minikit (Cat#74104) and quality was checked by Agilent Bioanalyzer. The RNA Integrity score for all samples was at least greater than 9. Labeling was done using Agilent's Quick-Amp labeling Kit (p/n:5190-0444) with T7 promoter based-linear amplification to generate labeled complementary RNA (One-Color Microarray-Based Gene Expression Analysis). Hybridization was done using Agilent's In situ Hybridization kit 5188-5242. Slides were scanned using Agilent Microarray Scanner (G2565BA) followed by data extraction by Agilent Feature Extraction software version 9.5.3. Scanned images were overall clean, uniform intensity and with very low background noise. Percentile Shift Normalization was done using GeneSpring GX 10 Software.

Microarray study design and identification of differentially expressed (DE) genes. There were altogether three groups namely, (i) Unstimulated (UNST), (ii) Stimulated by plate bound CD3, CD28 monoclonal antibodies in presence of rhIL-4 (TIL4) for 6

days and (iii) TIL4 stimulation with IL-4/R neutralizing antibodies added on day 3, day 4 and day 5 (ABS). IL-4 dependent genes were selected from Day 6 data. Only genes significantly differentially expressed (p<0.05) from TIL4 and ABS as compared to unstimulated groups were chosen for further analysis. IL-4 dependent upregulated genes are those which are upregulated (Fold above +1) after triggering TCR in presence of IL-4 and whose expression go down (Fold below -1) when IL-4/R neutralizing antibodies are added. Similarly IL-4 dependent down-regulated genes are those whose expression level go down (Fold below -1) on triggering TCR in presence of IL-4 and whose expression go up (Fold above +1) after antibody treatment. Microarray data has been submitted in Geo repository at NCBI (Accession Number: GSE116022).

Gene Ontology (GO). GO analysis was done on Explain analysis system interface. Biobase knowledge Library (BKL) manual curation was used for the analysis. P value cut off used was 0.01. All three GO categories; Biological Process, Molecular Function and Cellular Component were included.

MATCH Analysis. Putative transcription factor binding site search on the promoters of DE genes have been performed using professional version of MATCH program¹⁸. MATCH Identifies binding sites of a given transcription factor on the promoter of given set of genes using collections of known transcription factor binding site (TFBS) and positional weight matrices. We have used promoter region from -500 to 100 relative to TSS unless otherwise stated. Promoters from human housekeeping genes were taken as a negative set unless otherwise stated. Only high quality matrices were used for the analysis. P value cut off of 0.01 was used to extract significantly enriched binding sites in test set.

Gene Set Enrichment Analysis. GSEA is a method of analyzing and interpreting microarray and such data using biological knowledge. The data in question is analyzed in terms of their differential enrichment in a predefined biological set of genes¹⁷. These predefined biological sets can be published information about biochemical pathway or coexpression in a previous experiment.

GSEA was performed using GSEA version 2.07 from the Broad Institute at MIT. Parameters used for the analysis were as follows. The data set had 13474 features (Agilent probes), which were collapsed to 7396 genes. c2.cp.v2.5.symbols.gmt gene set was used for running GSEA and 1000 permutations were used to calculate p value and permutation type was set to gene_set. All basic and advanced fields were set to default except for "Metric for ranking genes" and that was set to Diff_of_Classes.

Network analysis. Network visualization and analysis was done using cytoscape³⁸ and its plugin "Network Analyzer". Power law fitting of degree distribution was also done using same cytoscape plugin.

Statistical analysis. Microsoft Excel T test was used to compare means. Paired two tailed test was used to calculate p value. 95% confidence was used to reject null hypothesis.

1. Ansel, K. M., Djuretic, I., Tanasa, B. & Rao, A. Regulation of Th2 differentiation and Il4 locus accessibility. *Annu Rev Immunol* **24**, 607–656 (2006).
2. Amsen, D. *et al.* Instruction of distinct CD4 T helper cell fates by different notch ligands on antigen-presenting cells. *Cell* **117**, 515–526 (2004).
3. Ho, I. C., Hodge, M. R., Rooney, J. W. & Glimcher, L. H. The proto-oncogene c-maf is responsible for tissue-specific expression of interleukin-4. *Cell* **85**, 973–983 (1996).
4. Kaplan, M. H., Schindler, U., Smiley, S. T. & Grusby, M. J. Stat6 is required for mediating responses to IL-4 and for development of Th2 cells. *Immunity* **4**, 313–319 (1996).
5. Stritesky, G. L. *et al.* The transcription factor STAT3 is required for T helper 2 cell development. *Immunity* **34**, 39–49.
6. Takeda, K. *et al.* Essential role of Stat6 in IL-4 signalling. *Nature* **380**, 627–630 (1996).
7. Tripathi, P. *et al.* A novel mechanism for ERK-dependent regulation of IL4 transcription during human Th2-cell differentiation. *Immunity Cell Biol.*
8. Yamashita, M. *et al.* 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* **96**, 1024–1029 (1999).
9. Zheng, W. & Flavell, R. A. The transcription factor GATA-3 is necessary and sufficient for Th2 cytokine gene expression in CD4 T cells. *Cell* **89**, 587–596 (1997).
10. Zhu, J., Cote-Sierra, J., Guo, L. & Paul, W. E. Stat5 activation plays a critical role in Th2 differentiation. *Immunity* **19**, 739–748 (2003).
11. Zhu, J., Guo, L., Watson, C. J., Hu-Li, J. & Paul, W. E. Stat6 is necessary and sufficient for IL-4's role in Th2 differentiation and cell expansion. *J Immunol* **166**, 7276–7281 (2001).
12. Chen, Z. *et al.* Identification of novel IL-4/Stat6-regulated genes in T lymphocytes. *J Immunol* **171**, 3627–3635 (2003).
13. Hamalainen, H. *et al.* Distinct gene expression profiles of human type 1 and type 2 T helper cells. *Genome biology* **2**, RESEARCH0022 (2001).



14. Lund, R., Aittokallio, T., Nevalainen, O. & Lahesmaa, R. Identification of novel genes regulated by IL-12, IL-4, or TGF-beta during the early polarization of CD4+ lymphocytes. *J Immunol* **171**, 5328–5336 (2003).
15. Lund, R. J. *et al.* Genome-wide identification of novel genes involved in early Th1 and Th2 cell differentiation. *J Immunol* **178**, 3648–3660 (2007).
16. Elo, L. L. *et al.* Genome-wide profiling of interleukin-4 and STAT6 transcription factor regulation of human Th2 cell programming. *Immunity* **32**, 852–862.
17. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545–15550 (2005).
18. Kel, A. E. *et al.* MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res* **31**, 3576–3579 (2003).
19. Cao, W. *et al.* Leukemia inhibitory factor inhibits T helper 17 cell differentiation and confers treatment effects of neural progenitor cell therapy in autoimmune disease. *Immunity* **35**, 273–284.
20. Swain, S. L., Weinberg, A. D., English, M. & Huston, G. IL-4 directs the development of Th2-like helper effectors. *J Immunol* **145**, 3796–3806 (1990).
21. Das, J. *et al.* A critical role for NF-kappa B in GATA3 expression and TH2 differentiation in allergic airway inflammation. *Nat Immunol* **2**, 45–50 (2001).
22. Boyle, K. B. *et al.* The transcription factors Egr1 and Egr2 have opposing influences on adipocyte differentiation. *Cell Death Differ* **16**, 782–789 (2009).
23. Collins, S., Wolfrum, L. A., Drake, C. G., Horton, M. R. & Powell, J. D. Cutting Edge: TCR-induced NAB2 enhances T cell function by coactivating IL-2 transcription. *J Immunol* **177**, 8301–8305 (2006).
24. Gooch, J. L., Christy, B. & Yee, D. STAT6 mediates interleukin-4 growth inhibition in human breast cancer cells. *Neoplasia (New York, N.Y.)* **4**, 324–331 (2002).
25. Kawakami, M. *et al.* Interleukin 4 receptor on human lung cancer: a molecular target for cytotoxin therapy. *Clin Cancer Res* **8**, 3503–3511 (2002).
26. Li, Z. *et al.* Endogenous interleukin-4 promotes tumor development by increasing tumor cell resistance to apoptosis. *Cancer research* **68**, 8687–8694 (2008).
27. Prokopchuk, O., Liu, Y., Henne-Bruns, D. & Kornmann, M. Interleukin-4 enhances proliferation of human pancreatic cancer cells: evidence for autocrine and paracrine actions. *British journal of cancer* **92**, 921–928 (2005).
28. Dhur, A., Galan, P. & Herberg, S. Folate status and the immune system. *Prog Food Nutr Sci* **15**, 43–60 (1991).
29. Duthie, S. J. *et al.* Blood folate status and expression of proteins involved in immune function, inflammation, and coagulation: biochemical and proteomic changes in the plasma of humans in response to long-term synthetic folic acid supplementation. *J Proteome Res* **9**, 1941–1950.
30. Piyathilake, C. J. *et al.* Indian women with higher serum concentrations of folate and vitamin B12 are significantly less likely to be infected with carcinogenic or high-risk (HR) types of human papillomaviruses (HPVs). *Int J Womens Health* **2**, 7–12.
31. Goh, K. I., Kahng, B. & Kim, D. Universal behavior of load distribution in scale-free networks. *Phys Rev Lett* **87**, 278701 (2001).
32. Mahic, M., Kalland, M. E., Aandahl, E. M., Torgersen, K. M. & Tasken, K. Human naturally occurring and adaptive regulatory T cells secrete high levels of leukaemia inhibitory factor upon activation. *Scand J Immunol* **68**, 391–396 (2008).
33. Vanderlocht, J. *et al.* Leukemia inhibitory factor is produced by myelin-reactive T cells from multiple sclerosis patients and protects against tumor necrosis factor-alpha-induced oligodendrocyte apoptosis. *J Neurosci Res* **83**, 763–774 (2006).
34. Bauer, S. & Patterson, P. H. Leukemia inhibitory factor promotes neural stem cell self-renewal in the adult brain. *J Neurosci* **26**, 12089–12099 (2006).
35. Jankovic, D. *et al.* Single cell analysis reveals that IL-4 receptor/Stat6 signaling is not required for the in vivo or in vitro development of CD4+ lymphocytes with a Th2 cytokine profile. *J Immunol* **164**, 3047–3055 (2000).
36. Ouyang, W. *et al.* Stat6-independent GATA-3 autoactivation directs IL-4-independent Th2 development and commitment. *Immunity* **12**, 27–37 (2000).
37. Tahvanainen, J. *et al.* PRELI is a mitochondrial regulator of human primary T-helper cell apoptosis, STAT6, and Th2-cell differentiation. *Blood* **113**, 1268–1277 (2009).
38. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498–2504 (2003).

Acknowledgements

This study was supported by a grant from the Department of Biotechnology, Government of India (KVSR), a grant from the Academy of Finland (Centre of Excellence in Molecular Systems Immunology and Physiology Research, 2012–2017, Decision No. 250114, and grants 116639, 115939, 140409, 140019), The Sigrid Juselius Foundation (KVSR and RL), The National Technology Agency of Finland and The European Seventh Commission Framework Grant EC-FP7-SYBILLA-201106 (RL). UU is the recipient of a Senior Research Fellowship from the Council of Scientific and Industrial Research, Government of India.

Author contribution

UU, PT and KVSR conceived the idea UU and PT performed experiment and analyzed data, UU and KVSR wrote manuscript and all authors reviewed the manuscript.

Additional information

Supplementary information accompanies this paper at <http://www.nature.com/scientificreports>

Competing financial interests: The authors declare no competing financial interests.

License: This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>

How to cite this article: Ullah, U., Tripathi, P., Sahoo N., Lahesmaa, R. & Rao, K.V.S. Gene Set Enrichment Analysis Identifies LIF as a Negative Regulator of Human Th2 Cell Differentiation. *Sci. Rep.* **2**, 464; DOI:10.1038/srep00464 (2012).