Transcriptomes of the B and T Lineages Compared by Multiplatform Microarray Profiling

Michio W. Painter,* Scott Davis,* Richard R. Hardy,† Diane Mathis,* Christophe Benoist,* and The Immunological Genome Project Consortium[1]

T and B lymphocytes are developmentally and functionally related cells of the immune system, representing the two major branches of adaptive immunity. Although originating from a common precursor, they play very different roles: T cells contribute to and drive cell-mediated immunity, whereas B cells secrete Abs. Because of their functional importance and well-characterized differentiation pathways, T and B lymphocytes are ideal cell types with which to understand how functional differences are encoded at the transcriptional level. Although there has been a great deal of interest in defining regulatory factors that distinguish T and B cells, a truly genomewide view of the transcriptional differences between these two cells types has not yet been taken. To obtain a more global perspective of the transcriptional differences underlying T and B cells, we exploited the statistical power of combinatorial profiling on different microarray platforms, and the breadth of the Immunological Genome Project gene expression database, to generate robust differential signatures. We find that differential expression in T and B cells is pervasive, with the majority of transcripts showing statistically significant differences. These distinguishing characteristics are acquired gradually, through all stages of B and T differentiation. In contrast, very few T versus B signature genes are uniquely expressed in these lineages, but are shared throughout immune cells. ***The Journal of Immunology*****, 2011, 186: 000–000.**

T and B lymphocytes are closely related cell lineages of the immune system, having the unique ability to somatically rearrange gene segments encoding receptors for Ag, the key molecules of the adaptive immune system. Both lineages are thought to arise from the same bone marrow precursors, the nature of which is somewhat debated at present. They complete remarkably parallel stages of differentiation and selection before reaching morphologically similar mature states, as naive lymphocytes resting in secondary lymphoid organs, from which activation by cognate Ag will provoke their terminal differentiation to effector or memory states.

Although T and B lymphocytes broadly share a role in the adaptive immune system, their functions within this responsive structure are entirely different: T cells participate primarily in cell-mediated immunity and in orchestrating cellular responses, whereas B cell production of Abs is the hallmark of humoral immunity. As these functional differences are usually assumed to be underpinned by differences in the basic cell biology of these lymphocytes, there has been some interest in determining what, beyond the Ag receptors and their ancillary factors, distinguishes B and T lymphocytes. In particular, how differently B and T lymphocytes use the blueprint of genes encoded in the genome.

A notable early study used cDNA subtractive hybridization, in which cDNA from T and B cells was isolated and subjected to exhaustive subtraction, to estimate that T and B cells differ by only 2% of their mRNA (1, 2), among which TCR-encoding genes were eventually isolated. Since then, several key regulators have been found, through knockout studies, to be necessary for the differentiation of either the T or B lineages: Pax5, Ebf1, or Sfpi1 (PU.1) for B cells and Notch1 and Gata3 for T cells (3–7). Although identifying such lineage-specification factors is of course essential, viewing the differences between lineages solely through the lens of a few control factors necessarily overlooks the complex transcriptional programs present in any given cell. The development of microarray technologies and the continued improvements in microarray platforms and their annotations have allowed a perspective on the transcriptome that is global and also more quantitatively nuanced. A few early studies used this approach to compare T and B lymphocytes (8–11), identifying sets of genes that are differentially expressed in B and T cells, as well as more generally shared sets; as might be expected, transcripts that varied during T or B lymphocyte differentiation showed more interlineage differential than invariant housekeeping genes (8).

Although generating such data for transcripts that are strongly expressed and/or clearly differential is straightforward, there is difficulty in arriving at more general conclusions for the entire transcriptome in such comparisons. These problems lie in the confidence one can have in calls that a transcript is present or absent in a given dataset, given the difficulty in distinguishing true signals from noise due to false negatives (nonperforming features on a microarray, subthreshold detection) or false positives (cross-hybridizing microarray features), both of which are poorly controlled on any one microarray (12, 13). In addition, the use of arbitrary thresholds to define expression differentials tends to create overly simplistic distinctions. In the current study, we

*Department of Pathology, Harvard Medical School, Boston, MA 02215; and †Fox Chase Cancer Center, Philadelphia, PA 19111

[1]All authors and their affiliations appear at the end of this article.

Received for publication August 10, 2010. Accepted for publication November 30, 2010.

This work was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health (R24 AI072073 to C.B., D.M., J. Collins, and D. Laidlaw).

The datasets presented in this article have been submitted to the National Center for Biotechnology Information/Gene Expression Omnibus under accession number GSE15907.

Address correspondence and reprint requests to Dr. Christophe Benoist and Diane Mathis, Department of Pathology, Harvard Medical School, 77 Avenue Louis Pasteur, NRB 10, Boston, MA 02115. E-mail address: cbdm@hms.harvard.edu

The online version of this article contains supplemental material.

Abbreviations used in this article: FC, fold change; GMM, Gaussian Mixture Model; ImmGen, Immunological Genome Project.

Table I. Summary of multiplatform gene expression data: part I

Sample	Expressed Genes (%)	False Positives (%)	False Negatives (%)	Overall Concordance (%)
Affymetrix CD19	51	8	2	84
Affymetrix CD4	50	8	2	84
Agilent CD19	43	4	7	92
Agilent CD4	43	4	7	92
Illumina CD19	47	9	8	84
Illumina CD4	47	10	8	83
Nimblegen CD19	46	5	4	89
Nimblegen CD4	46	4	4	89

Splenic $CD4^+$ T cells and $CD19^+$ B cells were profiled on Affymetrix, Agilent, Nimbelgen, and Illumina whole-genome microarrays. Resulting gene-expression data from each platform were analyzed to yield the percentage of expressed probes, percentage of false positives (defined as a probe being expressed on one platform, but not the other three), percentage of false negatives (defined as the absence of a probe's expression in one platform but present in the other three), and overall concordance (defined as the overall percentage of probes for which expression or absence is in agreement with the majority of platforms).

have attempted to robustly define the transcriptome differences underlying T and B lymphocytes by exploiting the unique datasets generated in the pilot phases of the Immunological Genome Project (ImmGen). ImmGen is a collaborative group of immunology and computational biology laboratories aiming to decipher, on a broad scale, the patterns of gene expression and genetic regulatory networks of the immune system of the mouse (14). We used the cross-verifying power of expression profiling on independent microarray platforms, as well as the breadth of gene-expression datasets available in the ImmGen database, to robustly explore what distinguishes T and B lymphocytes at the transcriptional level and to analyze when these distinctions are acquired during T and B lineage differentiation.

Materials and Methods

Mice

Six-week-old C57BL/6J mice were bred in specific pathogen-free conditions under Institutional Animal Care and Use Committee protocol (protocol 02954).

Cell sorting and flow cytometry

All cells were purified using the sorting protocol and mAbs listed on http://www.ImmGen.org.

Microarray analysis

For multiplatform microarrray profiling, RNA was prepared from sorted $CD4^+$ T cell and $CD19^+$ B cell populations from C57BL/6J mice using TRIzol reagent as described (15). RNA was amplified and hybridized on the Affymetrix Mouse Gene 1.0 ST, Agilent Mouse GE 1-Color, Illumina Mouse-6 v1.1 BeadChip, and Nimblegen Mouse X12 arrays according to the procedures specific to each platform. Raw data were preprocessed using software compatible for each platform and all normalized using the RMA algorithm. Thresholds on expression values above which a gene was considered expressed were derived for each platform by one of two distribution-based approaches. For platforms with well-defined negative control probe sets (Illumina Mouse-6 and Nimblegen X12), the threshold for greater-than-chance expression was defined as expression values greater than or equal to the 95% quantile of expression values in the negative controls. The negative controls for Agilent and Affymetrix arrays, however, exhibited notably different behavior in relation to noncontrol probes (likely due to the inclusion of intronic probes with some degree of expression) and thus did not allow for the same type of control-based analysis as Illumina and Nimblegen. For these samples, a Gaussian Mixture Model (GMM) was used to arrive at thresholds consistent with a controls-based approach. GMM is an Expectation-Maximization algorithm, the aim of which is to optimize the likelihood that a set of data points is generated by a mixture of Gaussian distributions. In this case, the MATLAB software "fit" function with parameter "gauss3" was used to the model the observed chipwide expression distribution profile of all noncontrol probe sets, such that each Gaussian component of the mixture corresponded to a different source of signal (i.e., background and genuine expression). Thresholds for greater-than-chance expression were then empirically defined as the value above which there is an equal probability that the signal is part of either distribution. This setting was validated on the Illumina and Nimblegen arrays by a good fit with thresholds derived from true negative controls. Specifically, the average percentage of genes in the four-platform common genome expressed above the GMM-derived thresholds for Affymetrix and Agilent were 50.5 and 42.7%, respectively, which is concordant with the controls-derived thresholds used for Nimblegen and Illumina (47.7–46.4%). Conversely, the equivalent controls-derived thresholds for Affymetrix and Agilent were highly discordant, with averages of 15.5 and 84.8%, respectively (data not shown).

Table II. Summary of multiplatform gene expression data: part II

Concordant Chips	Expressed in CD4 (%)	Expressed in CD19 (%)
2 of 4	49.74	49.67
3 of 4	43.26	43.35
4 of 4	32.41	32.06

The overall expression of the genome in T and B cells was calculated based on the number of genes registering as significantly expressed for each platform with concordance being defined as a given gene's expression or absence in two, three, or four out of four platforms (rows).

For data analysis using ImmGen datasets, raw data for all populations were normalized using the RMA algorithm (16) implemented in the "Expression File Creator" module in the GenePattern suite (17). Differential signatures were visualized using the "Multiplot" module. Signature transcripts were clustered using the "Hierarchical Clustering" module, using Pearson's correlation as a metric, and visualized using the "Hierarchical Clustering Viewer" heat map module.

To display the expression of transcripts during differentiation, a modified K-means algorithm was used to cluster the B and T cell signatures to represent the developmental activation of their respective genes. Unlike the traditional K-means approach of clustering observations around randomly determined centroids, this analysis used predefined, theoretical centroids, each characterized by a stepwise expression profile corresponding to successive stages of activation. Consequently, n-1 centroids were used to cluster a signature comprised of n stages of development. Pearson's correlation coefficient was used as the distance metric. This results in the clustering of probe sets around the single-stage activation exemplar to which it is most correlated.

The "Population Plots" position cell populations in a two-dimensional frame of reference, created using the expression values of sets of genes that most distinguish two reference populations. The x- and y-axes (B-ness and T-ness, respectively, in Fig. 4) were defined by expression values for the signature genes overexpressed in one reference population relative to the other: expression values of these genes were normalized relative to the reference populations (scaled to 0 and 1, where 0 is the expression value in the "low" population and 1 the value in the "high" population); scaled values for all signature genes were then averaged to yield the x and y coordinates of the populations tested.

For cluster analysis, expression values were normalized to the mean expression for each gene, and a partition-clustering algorithm (pam, S-Plus) was applied to the expression values in the T cell differentiation series. This cluster composition was then applied to expression values within non-T/non-B datasets within ImmGen (precursors, myeloid, and NK cells).

All datasets have been deposited at National Center for Biotechnology Information/Gene Expression Omnibus under accession number GSE15907 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE15907).

Results

Defining gene expression in T and B cells from the four-platform data

As part of the evaluation process to select a microarray platform most compatible with the ImmGen project, bulk $CD4^+$ T cells and $CD19^+$ B cells were sorted from spleen suspensions of 6-wk-old C57BL/6J mice for RNA preparations that were used to probe microarrays from four different commercial sources (Affymetrix Mouse Gene 1.0 ST array, Agilent Mouse GE 1-Color Array, Illumina Mouse-6 v1.1 Expression Beadchip Array, and Nimblegen Mouse X12 array). Three replicate datasets were generated for each cell type and each array (except one technical failure for Agilent), and the data were used for a comparative assessment of reproducibility and noise of importance in the context of the ImmGen program (data not shown). Relevant to the present project, we used the combined datasets to address the depth and variation of gene expression in B and T lymphocytes, under the assumption that comparable signals obtained in independent microarrays would be highly confirmatory, particularly because the various arrays use fundamentally different oligonucleotide probes (multiple 22-mers for Affymetrix, single long nucleotides for others) and probe/label chemistries (cDNA or cRNA). We generated a "Common Gene Table," which included 12,299 genes represented in at least three out of four arrays (full data listed in Supplemental Table I). We then defined, for each array, threshold expression values above which a probe was scored as showing significant expression (at a probability of $p < 0.05$, as detailed in Supplemental Material; because reliable negative controls are only present on two of the arrays, these thresholds for significant expression were based on those negative controls when present and on a Gaussian deconvolution of expression profiles similarly applied to all four platforms). This analysis showed excellent agreement between the platforms: the expression patterns in either T or B cells proved quite reproducible overall, being between 43 and 50% of the genes represented (Table I), with only a low proportion of false positives (signals detected on one array but absent on all others and thus likely to represent spurious noise) and false negatives (signals absent on a given array but present on at least two others). Combining the results from all four arrays and scoring those genes found to be expressed in at least two of the platforms showed that a very similar proportion of the genome (49.7%) is active in both B and T cells (Table II).

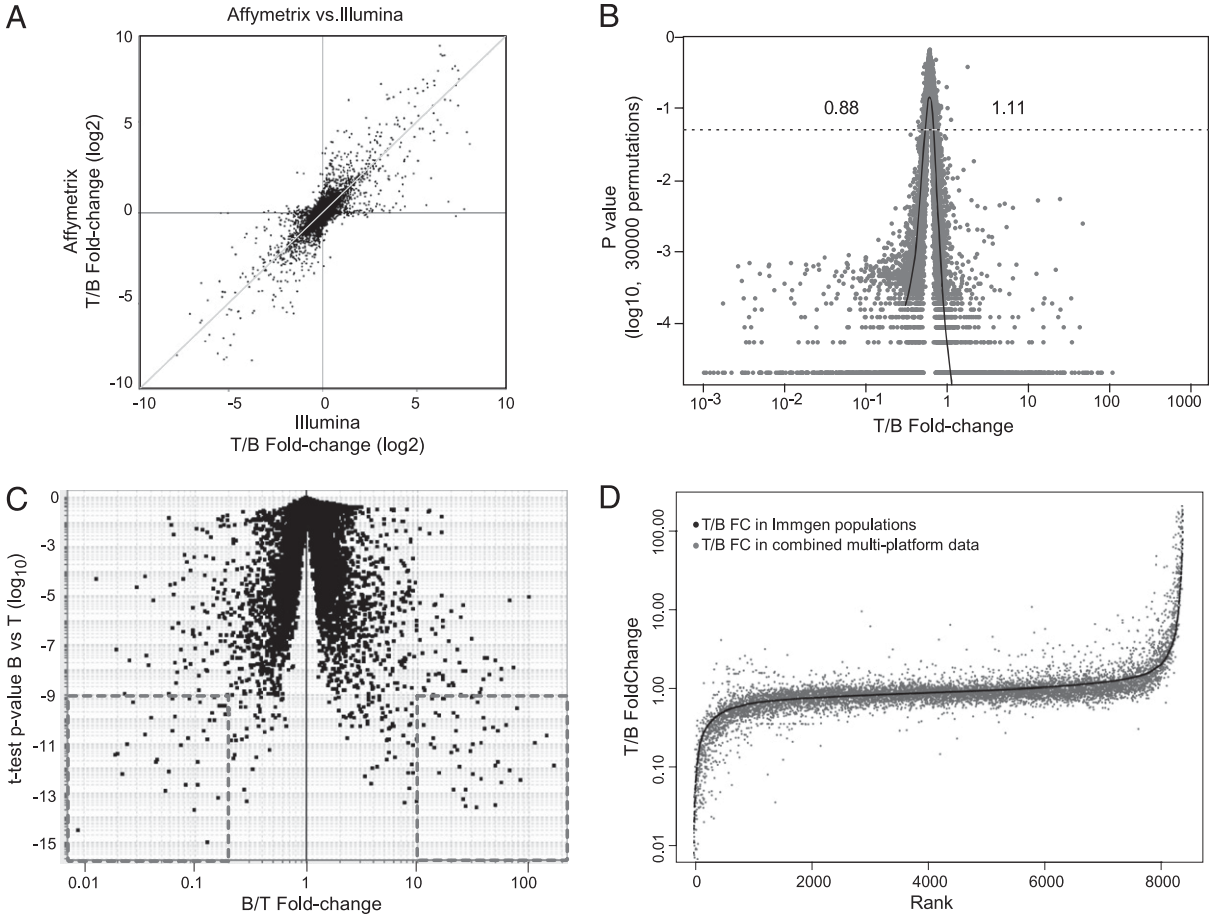

FIGURE 1. Defining T versus B differential signatures. *A*, RNA preparations from $CD4^+$ cells and CD19 B cells were profiled on Affymetrix and Illumina whole-genome microarrays, and the T versus B FC was calculated for the same genes on both microarrays. *B*, Consensus T versus B cell expression ratios were calculated by combining information from four different microarray platforms, and a false discovery rate on these FC values was estimated by repeated randomization of the datasets, testing how often the FC observed for a given gene could be observed by chance. The threshold FC values that reached statistical significance were estimated at <0.88 and >1.11, for a genomewide $p = 0.05$. *C*, Datasets from several populations of mature T cells (whole $CD3^+CD4^+$ splenocytes, naive $CD4^+$ and $CD8^+$ cells from spleen and lymph node, $CD44^{hi}$ $CD4^+$ and $CD8^+$ splenocytes) and B cells (whole $CD19^+$ splenocytes, mature bone marrow Fraction F cells, T3 splenic subset, follicular B from spleen and peritoneal cavity, marginal zone B), all profiled on the Affymetrix MuGeneST1.0 platform, were analyzed in combination to generate consensus measures of differential expression. The aggregate T versus B expression ratios are plotted against the Student *t* test *p* value. "Top 100" signature genes for B and T are outlined. *D*, Comparison of T/B FC determined from the multiplatform data (black dots) or from the combined ImmGen datasets (gray dots).

Next, we generated a robust signature of differential T versus B expression, again harnessing the combinatorial power of the multiplatform measurement to determine with a high degree of confidence the differences in transcript abundance. The data in the Common Gene Table described above were filtered for transcripts scoring positively in at least one cell type (8411 genes) and subsequently used to generate fold change (FC) estimates of the T/B ratio of expression for each of the four microarrays (calculated from the mean of the triplicate expression values). There was, for the most part, very good concordance between the FC values on different platforms, consistent with results from previous microarray comparison projects (18), as illustrated for one comparison in Fig. 1*A* (all comparisons are shown in Supplemental Fig. 1, and all data is listed in Supplemental Table II). We then generated consensus FCs by averaging the FCs measured on each microarray (the most differential transcripts are listed in Table III and all data in Supplemental Table II). To avoid spurious effects due to aberrant values on any one microarray platform, an outlier elimination procedure was implemented in which the FC value from one platform was disregarded if it fell >3 SD away from the mean of the other three platforms. T versus B differential expression ranged up to 633-fold (for an Ig V region), with 174 out of 8411 transcripts showing a differential of 20-fold or greater and 1364 out of 8411 a differential of 2-fold or greater.

We estimated the significance of these aggregate FCs by a data randomization procedure: triplicate expression values for $CD4^+$ T cells and $CD19^+$ B cells were scrambled for each gene and each platform, and the aggregate FC was recalculated from this randomized data as before (again applying the outlier elimination procedure). The procedure was repeated 30,000 times, counting the number of times the mock FC value for a given gene was equal or greater to that observed, yielding an estimate of the probability that the observed FC could be due to chance. As shown in Fig. 1*B*, most of the changes were highly significant. The range of FC values that reached significance at $p < 0.05$ was estimated from the FC versus p value scatter plot with a locally smoothed regression (loess; dark line on Fig. 1*B*). Significance was observed at very low FC values (>1.11 or <0.86) involving 5671 of the 8411 commonly expressed genes analyzed. From a technical standpoint, these data confirm the notion that combinatorial microarray profiling can reliably report on minute differences in expression (19). Overall, these data indicate that the difference between T and B lymphocytes involves a relative minority of transcripts with large differences in expression, but that a large fraction (at least 65%) of transcripts are subtly but significantly different in B and T cells.

Table III. Multiplatform T versus B differential signature genes

Gene Symbol	Combined Multiplatform T/B Ratio	FDR
Igl-V1	0.002	<0.00003
H2-Ab1	0.002	<0.00003
Ly6d	0.002	<0.00003
Ms4a1	0.002	<0.00003
H2-Aa	0.002	<0.00003
H2-Eb1	0.003	<0.00003
Scd1	0.003	0.000166667
Cd74	0.003	<0.00003
Blnk	0.004	<0.00003
H2-Dmb2	0.004	0.0006
Ly86	0.005	0.000366667
Cr2	0.005	<0.00003
H2-Dmb1	0.005	<0.00003
Lyn	0.005	0.0002
Plac8	0.005	<0.00003
Stk23	0.005	6.66667E-05
Fcer2a	0.005	<0.00003
Napsa	0.005	3.33333E-05
Rasgrp3	0.006	<0.00003
Faim3	0.006	0.0001
2010001m09rik	0.006	3.33333E-05
Cd79b	0.006	0.000666667
Hhex	0.006	6.66667E-05
Bank1	0.007	<0.00003
Tnfrsf13c	0.007	3.33333E-05
Cd3g	177.559	<0.00003
Cd247	131.154	<0.00003
Cd3d	125.911	<0.00003
Il7r	117.127	<0.00003
Tcra	98.672	<0.00003
Trat1	96.180	<0.00003
Igfbp4	88.251	<0.00003
2610019f03rik	84.180	<0.00003
E430004n04rik	80.586	<0.00003
A530021j07	76.378	<0.00003
Prkcq	76.298	0.002433333
2310032f03rik	70.026	6.66667E-05
Itk	68.390	<0.00003
Prkch	60.929	<0.00003
Tcf7	56.097	3.33333E-05
Bcl11b	55.890	<0.00003
Lat	55.061	0.0002
Tcrb-V13	45.987	<0.00003
Thy1	44.725	<0.00003
1700025g04rik	44.512	6.66667E-05
Tnfrsf7	43.149	<0.00003
Fyb	43.011	<0.00003
Bc021614	40.585	0.000133333
Cd6	40.556	<0.00003
Ampd1	40.043	<0.00003

Consensus T versus B FC values (calculated as the average of all four platforms, eliminating outliers) along with FDR for the top 25 most differentially expressed genes for $CD4^+$ T and $CD19^+$ B cells.

FDR, false discovery rate.

Defining a T versus B consensus signature from the broader ImmGen data

Although using multiplatform microarray profiling provided a technically robust T versus B signature, it was limited to bulk $CD4^+$ and $CD19^+$ splenocytes, which do not necessarily represent the broader range of T and B lymphocytes. Thus, to complement this signature, we thought it worthwhile to create a T versus B signature that would encompass a wider range of T and B cell subpopulations, but on a single microarray platform. The datasets of mature B and T lymphocytes available on the ImmGen database should enable the definition of differential signatures of T-ness and B-ness across more subpopulations. We selected datasets from a wide range of mature T and B cells, including $CD4^+$ and $CD8^+$ T cells from the spleen, lymph node, and thymus as well as B cells of different subtypes (follicular, marginal zone, B1) from the spleen, peritoneal cavity, and bone marrow. A composite T versus B signature was calculated by averaging across the two groups of populations, and the significance of these FC values was estimated with a simple Welsh's t test (the most differential transcripts are listed in Table IV). As shown in Fig. 1*C*, many genes were differentially expressed to a highly significant degree: 1078 genes, or 3% of the genes on the microarray, attained significance at a p value $<10^{-5}$ (a conservative threshold for corrected genomewide significance) for FC values ranging from 1.2–180 (given the increased variance, this comparison is less effective at ascribing significance to the numerous but subtle differences described above).

We then asked whether this second signature derived from multiple B and T cell populations within the ImmGen datasets would compare with that derived above by multiplatform analysis of $CD4^+$ and $CD19^+$ splenocytes. The majority of each signature's

Table IV. ImmGen T versus B differential signature genes

	B Cells							T Cells									
Gene Symbol	Follicular Peritoneal Cavity	Follicular Spleen	Bone Marrow Fraction F	Marginal Zone Spleen	T3 Spleen	CD19 Spleen	CD19 Spleen	CD4 Spleen	CD4 Spleen	CD4 Memory Spleen	CD4 Naive Lymph Node	CD4 Naive Spleen	CD4 Naive Thymus	CD8 Naive Mesenteric Lymph Node	CD8 Naive Spleen	Average FC B Versus T	*t* Test *p* Value B Versus T
Cd3g	42	47	48	57	59	59	39	4826	5382	5933	5459	4946	6261	5875	5249	0.0091	1.55×10^{-13}
prkch	51	47	43	49	41	45	39	2317	2523	2840	2135	1824	2578	2867	2162	0.0187	1.10×10^{-10}
Fyb	28	25	39	35	30	44	24	1918	2463	2079	1424	1006	1314	1691	991	0.0201	1.11×10^{-6}
Prkcq	47	43	44	60	45	38	46	2756	2862	2056	2365	1951	2188	2194	1958	0.0202	1.22×10^{-10}
Tcf7	104	71	92	87	119	89	71	3611	4515	2959	4419	4051	4272	4149	3606	0.0229	2.19×10^{-11}
Il7r	45	38	46	45	39	57	34	2431	2586	1797	1876	1564	1432	1645	1517	0.0233	2.25×10^{-8}
Itk	89	71	81	85	71	88	66	3515	3967	2359	3560	2775	3221	3644	2966	0.0243	2.81×10^{-10}
Cd96	41	38	34	35	35	45	33	1162	1684	1890	1355	783	1612	2072	1104	0.0256	4.36×10^{-7}
Ms4a4b	35	35	37	31	39	44	33	1274	1627	1631	1267	921	1116	1455	947	0.0284	1.11×10^{-8}
Cd3d	69	62	62	53	53	60	46	1637	1300	1704	1985	1339	2022	1535	1405	0.0357	6.66×10^{-10}
Themis	45	28	32	33	38	35	31	1237	1428	816	894	637	691	1132	710	0.0366	7.12×10^{-7}
Lcp2	70	57	82	70	68	79	60	1761	2278	2487	1637	1321	1744	1858	1270	0.0388	3.25×10^{-8}
Thy1	163	119	136	132	146	109	126	2164	2770	3308	3606	3169	3341	4165	3952	0.0402	3.06×10^{-9}
Slfn1	77	62	96	60	70	69	59	1531	1580	1386	2378	1996	2516	1331	1208	0.0404	3.13×10^{-7}
Lat	107	96	109	95	101	91	85	2759	2230	2003	2337	1981	2538	2301	1796	0.0436	5.96×10^{-11}
Cd3e	106	101	110	104	100	82	89	1509	1875	2152	2282	2202	2437	2351	2269	0.0463	7.19×10^{-11}
Emb	79	26	104	188	19	56	62	1392	1804	1170	1583	1316	1359	1675	1143	0.0533	8.13×10^{-10}
Skap1	65	80	75	85	63	79	68	1394	1552	1345	1378	1217	1309	1374	1035	0.0555	4.88×10^{-12}
Actn1	93	99	146	113	104	161	164	2141	2402	963	2334	2205	2207	2411	2146	0.0598	2.54×10^{-8}
Txk	77	84	92	83	80	86	86	1598	1589	1136	1526	1279	1395	1509	1188	0.0599	2.78×10^{-11}
Camk4	33	37	40	31	40	43	33	781	739	635	573	508	602	477	408	0.0621	1.58×10^{-8}
Apol7e	49	51	50	54	45	39	48	411	673	568	823	539	986	1272	693	0.0643	7.59×10^{-6}
Cd6	98	89	101	72	83	63	72	1128	1233	1517	1350	1287	1646	995	962	0.0653	3.39×10^{-9}
H2-Eb1	4404	5567	5931	5923	5822	5021	6015	131	141	143	152	142	129	148	175	38.1028	4.42×10^{-12}
Ebf1	1654	2138	1525	1610	1690	1394	2251	45	36	33	49	58	39	36	60	39.4510	2.86×10^{-9}
Cd22	3533	4785	4131	4396	5360	3507	5135	84	123	216	87	97	86	93	99	39.8824	9.48×10^{-10}
Kmo	1084	1366	1251	1075	1872	1290	1597	33	26	30	28	45	31	38	39	40.5334	1.61×10^{-8}
Faim3	3656	4425	4157	3211	5937	4007	4470	98	91	112	81	121	82	87	112	43.6234	9.21×10^{-9}
Lrrk2	2620	2360	2195	2015	1979	1762	2500	43	43	43	45	52	46	52	55	46.3868	9.68×10^{-11}
Cd180	1294	2790	2054	3633	2793	2016	3434	51	50	48	49	65	45	59	60	48.2906	2.26×10^{-6}
lgk	1289	1969	1429	1284	1396	1210	1737	32	24	24	35	40	24	24	38	49.2119	4.00×10^{-9}
Cd19	2630	2615	2690	2918	3084	2478	3043	45	45	52	50	75	48	52	66	51.2828	1.58×10^{-13}
H2-DMb2	3814	4477	4339	4410	4189	4520	4545	73	79	99	79	73	91	85	76	52.7524	1.78×10^{-15}
Rasgrp[3]	1289	2267	1782	1062	2031	1471	2135	32	27	29	32	40	31	28	41	52.9939	2.36×10^{-7}
Pax5	4321	4312	4243	3696	5143	4253	5237	55	81	71	84	94	94	76	101	54.3416	3.34×10^{-8}
Cd79a	6593	6616	7781	5661	7289	5562	6138	99	105	105	131	121	105	117	165	55.0504	2.39×10^{-11}
Lyn	2265	2725	2538	2978	2750	2644	2928	36	31	84	64	41	43	33	46	57.0073	3.56×10^{-13}
Ly86	1773	2578	2088	3852	2338	3170	3130	49	42	42	45	59	48	43	50	57.0689	2.39×10^{-7}
Ebf1	2438	2476	2199	2291	2561	2058	2667	34	29	30	39	45	34	32	44	66.5712	3.09×10^{-13}
Bank1	3567	4202	3493	3483	3633	3104	4467	41	40	51	45	59	49	44	61	76.1673	2.42×10^{-11}
Scd1	3163	3947	3987	3679	3843	3073	3637	44	36	37	34	55	47	42	51	83.8459	1.46×10^{-12}
Cd74	8929	8999	9023	7887	9559	6753	8279	70	80	208	85	94	127	57	59	86.9893	4.71×10^{-12}
H2-Ab1	5135	6548	6600	6091	6837	4060	6350	67	46	42	62	55	57	46	66	108.2105	8.45×10^{-10}
Ms4a1	3235	4144	3393	5387	4218	3659	4861	36	25	24	31	25	24	21	35	149.2743	3.72×10^{-9}
H2-Aa	5899	6380	7007	6401	6485	4747	6502	43	28	28	35	45	32	32	31	180.5529	7.57×10^{-12}

Expression value, population FC value (defined as the average FC between all B and T populations above), and p values (Student t test) for these FC are shown for the top 25 most differentially expressed genes for T and B populations. All cell types were profiled on the Affymetrix 1.0 ST array.

"Top 100" most distinguishing transcripts are shared, with 64% of T cell transcripts and 52% of B cell transcripts being present in both the multiplatform and ImmGen determinations. A ranked plot of the T versus B FCs in the two signatures reveals good overall matching across the differential ranking (Fig. 1*D*, Supplemental Table II). Some differences between the two signatures were observed, however, which are to be expected, as the ImmGen determination used a broad array of T and B populations, whereas the multiplatform determination used solely $CD4^+$ and $CD19^+$ splenocytes (for instance, CD4 itself ranks differently in the two signatures).

Are the transcripts that distinguish T and B cells specific to these lymphoid lineages?

Having generated these robust T versus B differential signatures, we next asked whether the transcripts that most distinguish T and B cells are unique to these cells or whether their expression is also shared with cells of other non-T/non-B lineages. Because in most schemas of hematopoietic cell differentiation, B and T lymphocytes represent terminal splits of the same lymphocyte branch, one might expect that the transcripts that sharply distinguish them may be uniquely expressed, solely present there and not in any other lineage (as are TCR and Ig transcripts, for instance). More generally, it is of interest to ask how many transcripts uniquely define a particular cell type and how many truly T- or B-specific genes actually exist, other than the Ag-specific receptors that defined these cells. To address this question, we mapped the expression of the 100 genes that most strongly differentiate T or B cells across the other immune cell populations of the ImmGen database (dendritic cells and macrophages, NK cells, stem cells; γδT cells were not considered because they were too similar to αβT cells). As shown in the heat map representations of Figs. 2 and 3, T and B signature transcripts were shared extensively with other lineages. As might be expected, T cell transcripts were more frequently shared with NK cells and B cell transcripts with dendritic or other myeloid cells, but this was not an absolute rule, and there were significant clusters of T signature transcripts present in myeloid cells and B signature transcripts in NK cells. Even stromal cells and monocytes expressed some B or T cell genes. These data indicate that the transcripts that most distinguish

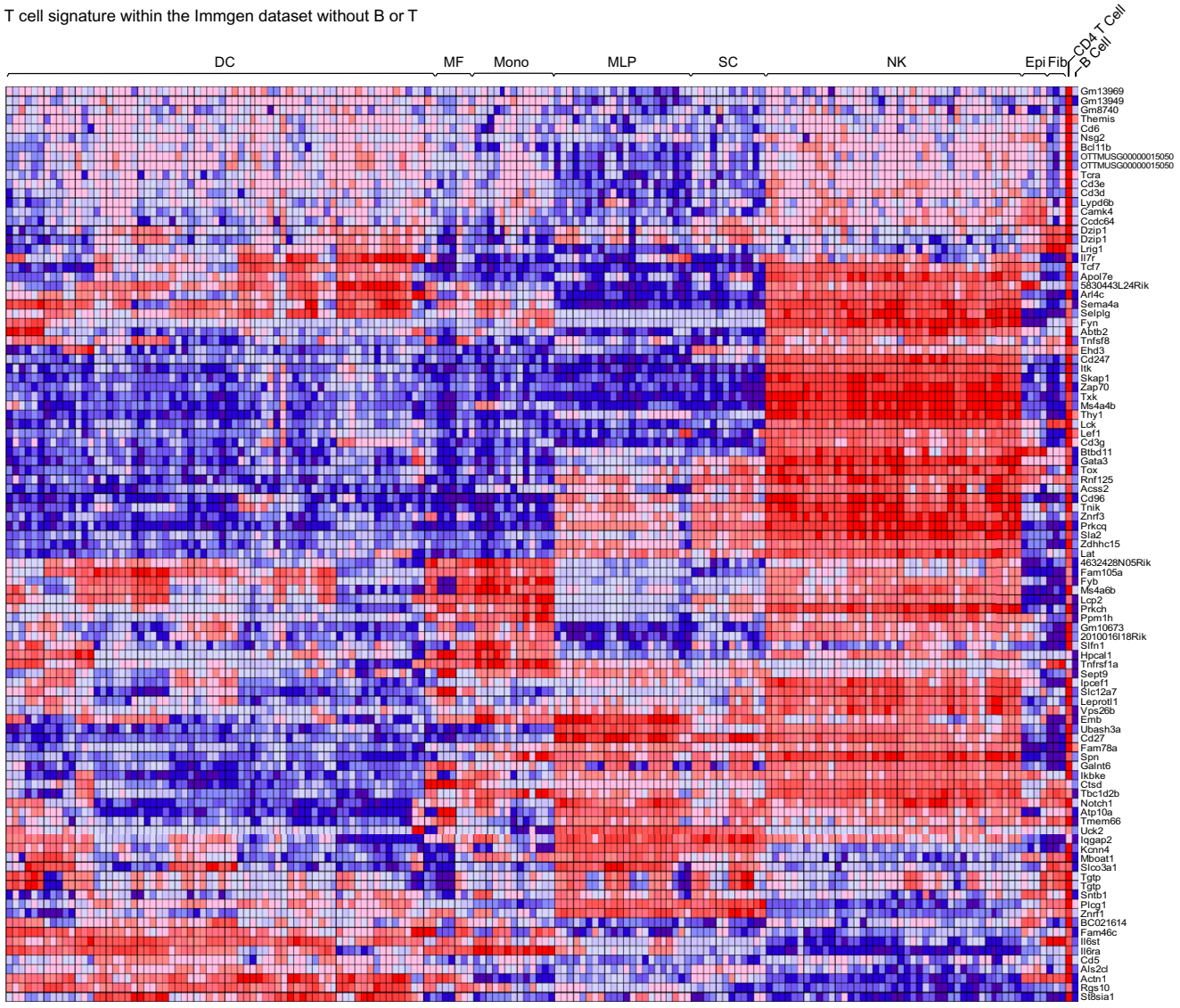

FIGURE 2. The transcripts that most distinguish T and B cells are expressed throughout immune cells. Heat map representations of the expression of the "Top 100" T cell signature genes across the immune cell populations contained in the ImmGen database. Genes are arranged by hierarchical clustering.

T and B lymphocytes are broadly expressed in other immune cells, and hardly any transcripts fall into the category of being absolutely specific to B or T lymphocytes.

We cannot completely rule out the possibility that this conclusion is influenced by spurious lymphocyte contamination in some datasets, but this seems unlikely because if a given dataset were contaminated with T or B lymphocytes, one would expect that all of the T- or B-specific signature would appear expressed. It is clear from Figs. 2 and 3, however, that only distinct modules of the T or B signatures are expressed within a given population.

How are transcriptional characteristics of mature T and B cells acquired during differentiation?

The differentiation of T and B lymphocytes is a well-characterized process marked by distinct stages that can be tracked by the expression of various cell-surface molecules (20, 21). As such, T and B cells are attractive lineages with which to ask how the identity of mature cells is acquired. Although a good deal is known about the timing of expression of various transcription factors during the differentiation of these two cell types (3, 22, 23), differentiation along the T and B lineages involves many other transcripts (24). We thus asked how the identity of mature T and B cells, as reflected in their above-defined distinguishing transcripts, is acquired during differentiation. In other words, when does a B cell become a B cell or a T cell become a T cell? To address this question, we used an ordering algorithm to arrange T and B signature transcripts according to the stage at which they are induced during differentiation. As shown in the heat map representations of Fig. 4*A* and 4*B*, we found that signature transcripts are acquired in a sequential manner, evenly through several steps of differentiation rather than being coordinately turned on at one particular stage. These steps do not particularly coincide with the rearrangement of Ag receptor genes, but occur through the double-negative and double-positive stages for thymic T cell precursors and through the transitions of pro- and pre-B cells in the bone marrow. In this respect, the full identity of T and B cells is realized gradually and not fully attained until maturity. This finding goes against the notion that expressing a TCR is what makes a T cell or a BCR a B cell.

Conversely, we asked when signature transcripts of the other lineages were switched off, plotting the expression of T cell sig-

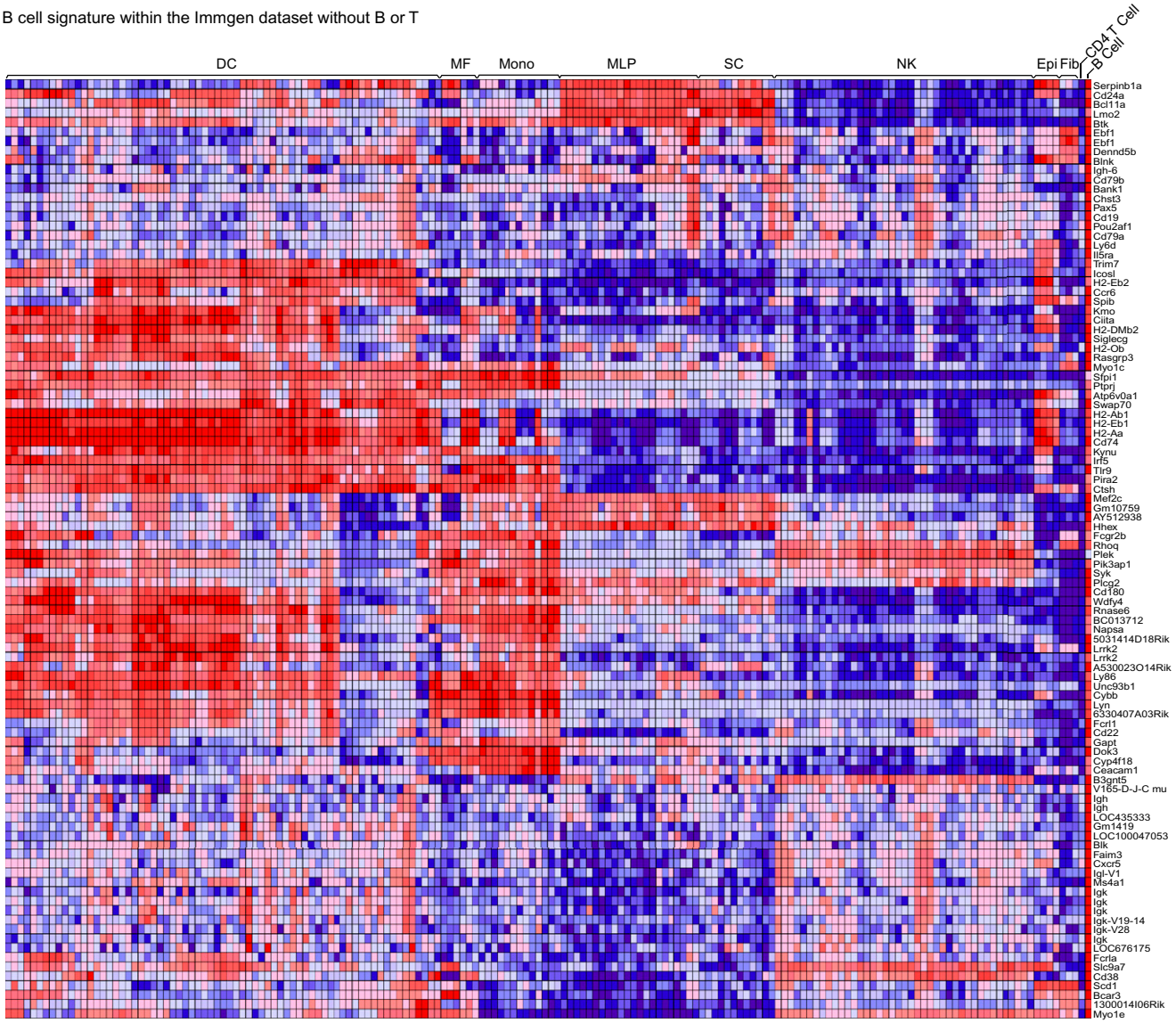

FIGURE 3. The transcripts that most distinguish T and B cells, continued. Heat map representations of the expression of the "Top 100" B cell signature genes across the immune cell populations contained in the ImmGen database. Genes are arranged by hierarchical clustering.

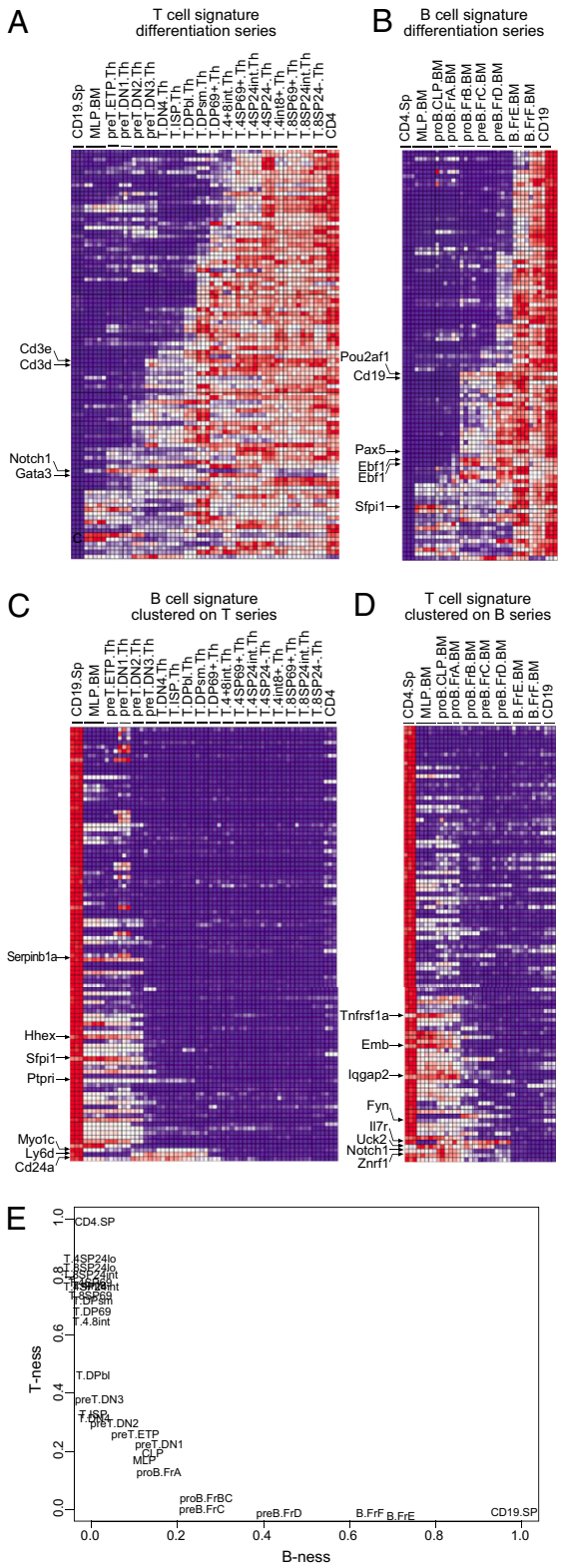

FIGURE 4. The transcripts that most distinguish T and B cells are acquired, or lost, in stages throughout differentiation. Heat map representations of the expression of the "Top 100" T cell of B cell genes during T cell differentiation in the thymus (*A*, *C*) or during B cell differentiation in the bone marrow (*B*, *D*). Cell types have been arranged according to their sequence during differentiation, and genes were clustered using an ordering algorithm according to the stage at which they are expressed. *E*, Population plot in which cell types have been positioned according to their T-ness and B-ness, defined from the aggregate expression values of genes most differentially expressed in mature B and T cells (see *Materials and Methods*).

nature genes during B cell differentiation and vice versa. As illustrated in Fig. 4*C* and 4*D*, signature genes of the other lineage are turned off quite early during differentiation, faster than the defining signature transcripts are acquired. In T cells, most B cell signature transcripts are turned off by the double-negative 2 stage, whereas in B cells most T cell signature transcripts are turned off by the fraction B, pro-B cell stage.

This progression of identity acquisition through the early lineages is reflected in the population plots of Fig. 4*E*, in which populations are positioned according to their expression of T- and B-defining transcripts and where the sequence of differentiation is clearly delineated.

Do the same regulatory modules control signature genes in T or B lineages and in non-T/non-B cells?

The expression signatures that distinguish T cells from B cells are acquired through distinct steps of T or B cell differentiation, and their expression is also shared with other non-T/non-B lineages along distinctive patterns (Figs. 2–4). It was thus of interest to ask whether the same regulatory influences operate in both contexts or whether transcripts obey different regulators (or combinations thereof) during T cell differentiation and when they are active outside the T lineage. Transcriptional regulation operates on modules of coregulated transcripts, which are similarly controlled by shared regulators; strongly correlated expression throughout a panel of cell populations is an indicator of such coregulation. By extension, common regulatory influences (transcription factors, microRNAs) operating within stages of T differentiation and through non-T/non-B lineages should be reflected as pairwise correlations that exist in both contexts. To address this question, we measured the pairwise correlation coefficients between transcripts of the "Top 200" T signature, across both the T-differentiation and non-T/non-B data groups. A Pearson correlation coefficient was used as a metric. As a reference, pairwise correlation coefficients across the same two data groups were also computed for a randomly selected set of transcripts. As illustrated in Fig. 5*A*, correlations between T signature transcripts within the T-differentiation data group showed a skewed distribution, with a much greater proportion of high correlation coefficients than within the reference gene set. In contrast, this bias was far more modest within the non-T/non-B data group. The different distribution of pairwise correlations for T signature genes within the T and non-T/non-B data groups was compared directly in the scatter plot of Fig. 5*B* (after transformation to a *z*-score to normalize against the distributions of correlation coefficients within the reference gene set). As expected, most pairs of transcripts correlated strongly within the T lineage, but showed little or no correlation within non-T/non-B lineages. In contrast, some transcript pairs did show strong correlation across both data groups (mapping to the top right quadrant of Fig. 5*B*). This distribution suggests that the majority of coregulatory relationships that operate within stages of T cell differentiation are not maintained in other lineages, although a few are.

To investigate this point further, we used a simple sequential clustering algorithm to parse the T-signature transcripts into distinct coregulated clusters, according to their expression patterns through T cell differentiation, and identifying the subclusters that did or did not show correlation within the non-T/non-B data group. As shown in Fig. 5*C*, some subclusters did show good homogeneity of expression in both data groups (e.g., cluster 1, which corresponded to a set of genes predominantly activated in the late stages of thymic T cell differentiation and quite uniquely coexpressed in NK cells), whereas others showed no preserved pattern of expression in non-T/non-B cells (e.g., cluster 2, also

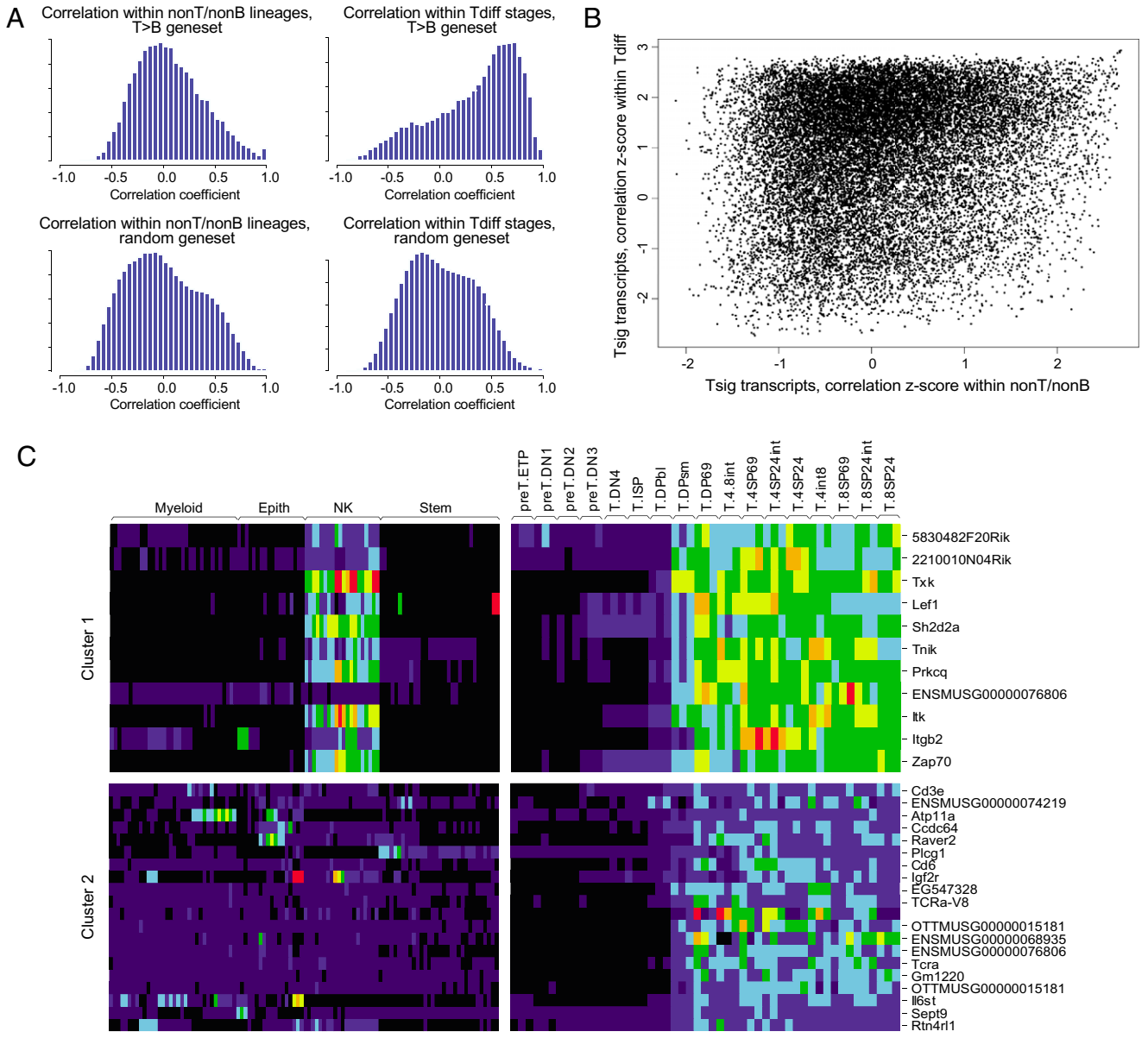

FIGURE 5. Partial sharing of coregulated gene clusters within T cell differentiation and outside the T cell lineage. To determine which transcripts exhibit coordinated expression, as a reflection of possible shared regulatory mechanisms, pairwise correlation coefficients were calculated for all transcripts of the "Top 200" T cell signature genes within all ImmGen datasets except for T and B cells (nonT/nonB) or within the T cell differentiation datasets. As a reference, the same coefficients were calculated on a set of 2000 transcripts picked at random. *A*, Distribution of the correlation coefficients; note that there is a very significant skewing of the distribution of correlation coefficients between T signature genes in the T-differentiation data group (*top left panel*) and far less marked within the non-T/non-B data group (*top right panel*). *B*, Scatter plot comparison of all pairwise correlations between T signature genes within the non-T/non-B (*x*-axis) or T-differentiation (*y*-axis) data groups; to avoid artifacts due to the different sizes and composition of the non-T/non-B and T-differentiation datasets, the primary correlation coefficients were transformed to a *z*-score by reference to the mean and SD of the correlation coefficients for the randomly picked reference gene set. Note that the majority of transcript pairs that show strong correlation within the T-differentiation data group (*z*-score >2) show no correlation within the non-T/non-B populations (*z*-scores distributed around 0), although there is a distinct shoulder of gene pairs that do show some correlation across both conditions (top right of the plot). *C*, A k-means clustering algorithm was used to partition T-signature genes into distinct clusters based on their correlation within the T-differentiation data group. Transcript levels for representative clusters are shown as a heat map for the non-T/non-B (*left panels*) and T-differentiation (*right panels*) data groups. A few clusters showed consistent expression across both data groups (e.g., Cluster 1, *top panel*, primarily reflecting shared expression with NK cells), whereas many were only coregulated within the T-differentiation data group.

activated late in T differentiation but that showed no consistent expression pattern outside the T lineage). Thus, only a minority of the transcripts that characterize T lymphocytes belong to coregulated gene clusters that are reused in different cell types.

Discussion

A central goal of this work was to define, from a genomewide perspective, the transcriptional differences that underlie T and B lymphocytes. We used the power of combinatorial microarray profiling as well as the breadth of cell populations available from the ImmGen project to explore the transcripts that provide their identities to T and B lymphocytes, in a more robust and in-depth perspective than could be provided in the comparisons preformed previously (8–11). The results show that transcriptional differences between B and T cells are very broad, not solely limited to a few specific markers commonly used to distinguish them by flow

cytometry. In contrast, there are very few transcripts uniquely specific to B and T cells, most being shared with other cell types in the immune system.

Combinatorial microarray profiling to describe the transcriptome of a cell has several distinct advantages over gene expression profiling with a single array. First, this approach eliminates any probe biases inherent to a particular chip's design. It is likely that this cross-checking resulted in our finding no difference in the overall number of genes expressed in T cells compared with B cells, which had been suggested by Hoffman et al. (8). In addition, combining platforms avoids the false positives and false negatives that commonly affect 5–10% of the probe sets on any one microarray support. Finally, combinatorial profiling allows for discovery of differential gene expression at greater depth and confidence. Thus, in contrast to previous studies, we estimate that at least 65% of the transcripts expressed in T and B cells are differential, most of which at very subtle FC values. In fact, had we compared even more datasets, it is plausible that every single gene expressed in T and B cells would be found to be significantly different.

Although this breadth is impressive, what does it mean that such a large percentage of genes is differentially expressed in such subtle manner when thinking of the physiology of T and B lymphocytes? One perspective is that these broadly distributed but subtle levels of differential expression actually have little or no functional impact on the cell. One can imagine that a transcriptional regulator activates or represses the expression of a particular gene or module that specifies an important function in either T or B cells but that, in doing so, it also creates transcriptional or posttranscriptional perturbations that ripple at low levels throughout the genetic regulatory network of the cell. These small expression variations across the genome would essentially be an unavoidable reverberation accompanying a larger and more meaningful variation, but have no functional consequences in themselves, if the key networks that regulate metabolic homeostasis or cell proliferation and survival are sufficiently robust in the context of such variation. There would thus be no need to guard against such changes. A similar argument has been made for the impact of microRNAs, each of which can have mild but widespread effects, but with perhaps only a few truly meaningful and evolutionarily selected targets. In contrast, these variations between B and T cells are so pervasive that it is difficult to believe that they are not meaningful in some way. In addition, microarrays tend to compress and underrepresent differences in transcript abundance relative to quantitative PCR. Differences of 1.2–1.3-fold by microarray are often closer to 2-fold when measured by real-time PCR. Such differences may thus be in a range that influences many genetic or molecular systems (e.g., copy number dependence in heterozygous mutations, metabolic regulation, etc.). Of course, testing the significance of many minor variations is not experimentally tractable today.

We also found that the vast majority of these T/B differential transcripts are not specific to either of these lineages, but are widely represented throughout immune system cell types. Some of this shared expression might have been expected based on known physiology (e.g., Ag presentation pathways active in both B cells and dendritic cells, cytotoxic effector molecules in NK and T cells), but other elements were less predictable. Again, some of these shared expression patterns may be unintended side effects of transcriptional control pathways, but these data suggest that there is much reutilization of functional proteins across cell types. There is precedent for cross-lineage sharing of gene products, even if their activity varies with context. For instance, the transcription factor *Tbx21* (also known as T-bet) controls different specialized functions in different cells, favoring Th1 effector functions in T cells, promoting class switching to IgG2a in B cells, and necessary for induction of type I IFNs in dendritic cells by TLR9 ligands (25). Similarly, Blimp-1 was originally discovered as a transcriptional repressor of IFN-β in human HeLa cells, then found to be required for the differentiation and maintenance of Ig-secreting B cells and plasma cells, and later identified as impacting T cell differentiation at several stages (in the thymus during Th1/2 specification and in regulatory T cells) (26).

Overall, the picture painted by these studies of the relationship between T and B lymphocytes departs somewhat from prior notions, with very few transcripts that are exquisitely specific of either cell, but with differences in transcriptome distributions that are very broad but also quite nuanced.

Acknowledgments

We thank Drs. Vladimir Jojic and Mark Davis for comments and eBioscience, Affymetrix, and Expression Analysis for support of the ImmGen Project. We also thank the members of the ImmGen Consortium.

ImmGen Consortium

Yan Zhou, Susan Shinton, and Richard Hardy (Division of Basic Science, Fox Chase Cancer Center, Philadelphia, PA 19111)

Natasha Asinovski, Scott Davis, Ayla Ergun, Jeff Ericson, Tracy Heng, Jonathan Hill, Gordon Hyatt, Daniel Gray, Michio Painter, Catherine Laplace, Adriana Ortiz-Lopez, Diane Mathis, and Christophe Benoist (Department of Pathology, Harvard Medical School, Boston, MA 02115)

Angelique Bellemare-Pelletier, Kutlu Elpek, and Shannon Turley (Department of Cancer Immunology and AIDS, Dana Farber Cancer Institute, Boston, MA 02115)

Adam Best, Jamie Knell, and Ananda Goldrath (Division of Biology, University of California, San Diego, La Jolla, CA 92093)

Joseph Sun, Natalie Bezman, and Lewis Lanier (Department of Microbiology and Immunology and the Cancer Research Institute, University of California, San Francisco, San Francisco, CA 94143)

Milena Bogunovic, Julie Helft, Ravi Sachidanandam, and Miriam Merad (Department of Gene and Cell Medicine and the Immunology Institute, Mount Sinai School of Medicine, New York, NY 10029)

Claudia Jakubzick, Emmanuel Gautier, and Gwendalyn Randolph (Department of Gene and Cell Medicine and the Immunology Institute, Mount Sinai School of Medicine, New York, NY 10029)

Nadia Cohen and Michael Brenner (Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115)

Jim Collins and James Costello (Center for Biodynamics, Boston University, Boston, MA 02215)

Radu Jianu and David Laidlaw (Department of Computer Science, Brown University, Providence, RI 02912)

Vladimir Jojic and Daphne Koller (Department of Computer Science, Stanford University, Stanford, CA 94305)

Nidhi Malhotra, Katelyn Sylvia, Kavitha Narayan, and Joonsoo Kang (Department of Pathology, University of Massachusetts Medical School, Worcester, MA 01655)

Tal Shay and Aviv Regev (Broad Institute and Massachusetts Institute of Technology, Cambridge, MA 02142)

Disclosures

The authors have no financial conflicts of interest.

References

1. Davis, M. M., D. I. Cohen, A. L. DeFranco, and W. E. Paul. 1982. The isolation of B and T cell-specific genes. In *B and T Cell Tumors: Biological and Clinical Aspects*, Vol 24. E. Vitetta, ed. Academic Press, New York. p. 215–220.
2. Hedrick, S. M., D. I. Cohen, E. A. Nielsen, and M. M. Davis. 1984. Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. *Nature* 308: 149–153.
3. Rothenberg, E. V. 2007. Cell lineage regulators in B and T cell development. *Nat. Immunol.* 8: 441–444.
4. Tanigaki, K., and T. Honjo. 2007. Regulation of lymphocyte development by Notch signaling. *Nat. Immunol.* 8: 451–456.

5. Pai, S. Y., M. L. Truitt, C. N. Ting, J. M. Leiden, L. H. Glimcher, and I. C. Ho. 2003. Critical roles for transcription factor GATA-3 in thymocyte development. *Immunity* 19: 863–875.
6. Busslinger, M. 2004. Transcriptional control of early B cell development. *Annu. Rev. Immunol.* 22: 55–79.
7. Hagman, J., and K. Lukin. 2006. Transcription factors drive B cell development. *Curr. Opin. Immunol.* 18: 127–134.
8. Hoffmann, R., L. Bruno, T. Seidl, A. Rolink, and F. Melchers. 2003. Rules for gene usage inferred from a comparison of large-scale gene expression profiles of T and B lymphocyte development. *J. Immunol.* 170: 1339–1353.
9. Kluger, Y., D. P. Tuck, J. T. Chang, Y. Nakayama, R. Poddar, N. Kohya, Z. Lian, A. Ben Nasr, H. R. Halaban, D. S. Krause, et al. 2004. Lineage specificity of gene expression patterns. *Proc. Natl. Acad. Sci. USA* 101: 6508–6513.
10. Hutton, J. J., A. G. Jegga, S. Kong, A. Gupta, C. Ebert, S. Williams, J. D. Katz, and B. J. Aronow. 2004. Microarray and comparative genomics-based identification of genes and gene regulatory regions of the mouse immune system. *BMC Genomics* 5: 82.
11. Abbas, A. R., D. Baldwin, Y. Ma, W. Ouyang, A. Gurney, F. Martin, S. Fong, M. van Lookeren Campagne, P. Godowski, P. M. Williams, et al. 2005. Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. *Genes Immun.* 6: 319–331.
12. Kothapalli, R., S. J. Yoder, S. Mane, and T. P. Loughran, Jr. 2002. Microarray results: how accurate are they? *BMC Bioinformatics* 3: 22.
13. Wu, C., R. Carta, and L. Zhang. 2005. Sequence dependence of cross-hybridization on short oligo microarrays. *Nucleic Acids Res.* 33: e84.
14. Heng, T. S., M. W. Painter; Immunological Genome Project Consortium. 2008. The Immunological Genome Project: networks of gene expression in immune cells. *Nat. Immunol.* 9: 1091–1094.
15. Yamagata, T., D. Mathis, and C. Benoist. 2004. Self-reactivity in thymic double-positive cells commits cells to a CD8 alpha alpha lineage with characteristics of innate immune cells. *Nat. Immunol.* 5: 597–605.
16. Irizarry, R. A., B. M. Bolstad, F. Collin, L. M. Cope, B. Hobbs, and T. P. Speed. 2003. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* 31: e15.
17. Reich, M., T. Liefeld, J. Gould, J. Lerner, P. Tamayo, and J. P. Mesirov. 2006. GenePattern 2.0. *Nat. Genet.* 38: 500–501.
18. MAQC Consortium, L. Shi, L. H. Reid, W. D. Jones, R. Shippy, J. A. Warrington, S. C. Baker, P. J. Collins, F. de Longueville, E. S. Kawasaki, et al. 2006. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat. Biotechnol.* 24: 1151–1161.
19. Venanzi, E. S., R. Melamed, D. Mathis, and C. Benoist. 2008. The variable immunological self: genetic variation and nongenetic noise in Aire-regulated transcription. *Proc. Natl. Acad. Sci. USA* 105: 15860–15865.
20. Fowlkes, B. J., and D. M. Pardoll. 1989. Molecular and cellular events of T cell development. *Adv. Immunol.* 44: 207–264.
21. Hardy, R. R., and K. Hayakawa. 2001. B cell development pathways. *Annu. Rev. Immunol.* 19: 595–621.
22. Rothenberg, E. V., J. E. Moore, and M. A. Yui. 2008. Launching the T-cell-lineage developmental programme. *Nat. Rev. Immunol.* 8: 9–21.
23. Northrup, D. L., and D. Allman. 2008. Transcriptional regulation of early B cell development. *Immunol. Res.* 42: 106–117.
24. Mick, V. E., T. K. Starr, T. M. McCaughtry, L. K. McNeil, and K. A. Hogquist. 2004. The regulated expression of a diverse set of genes during thymocyte positive selection in vivo. *J. Immunol.* 173: 5434–5444.
25. Peng, S. L. 2006. The T-box transcription factor T-bet in immunity and autoimmunity. *Cell. Mol. Immunol.* 3: 87–95.
26. Martins, G., and K. Calame. 2008. Regulation and functions of Blimp-1 in T and B lymphocytes. *Annu. Rev. Immunol.* 26: 133–169.