Gene Coexpression Analysis in Single Cells Indicates Lymphomyeloid Copriming in Short-Term Hematopoietic Stem Cells and Multipotent Progenitors

Laetitia Gautreau, Amine Boudil, Valérie Pasqualetto, Lamia Skhiri, Laure Grandin, Marta Monteiro, Jean-Philippe Jais and Sophie Ezine

J Immunol 2010; 184:4907-4917; Prepublished online 5 April 2010; doi: 10.4049/jimmunol.0902184
http://www.jimmunol.org/content/184/9/4907

Supplementary Material http://www.jimmunol.org/content/suppl/2010/04/06/jimmunol.0902184.DC1

Why The JI?

- Rapid Reviews! 30 days* from submission to initial decision
- No Triage! Every submission reviewed by practicing scientists
- Speedy Publication! 4 weeks from acceptance to publication

*average

References This article cites 55 articles, 23 of which you can access for free at:
http://www.jimmunol.org/content/184/9/4907.full#ref-list-1

Subscription Information about subscribing to The Journal of Immunology is online at:
http://jimmunol.org/subscription

Permissions Submit copyright permission requests at:
http://www.aai.org/About/Publications/JI/copyright.html

Email Alerts Receive free email-alerts when new articles cite this article. Sign up at:
http://jimmunol.org/alerts
Gene Coexpression Analysis in Single Cells Indicates Lymphomyeloid Copriming in Short-Term Hematopoietic Stem Cells and Multipotent Progenitors

Laetitia Gautreau,*+†,1 Amine Boudil,*+†,1 Valérie Pasqualetto,*+† Lamia Skhiri,*+† Laure Grandin,‡ Marta Monteiro,*+† Jean-Philippe Jais,‡ and Sophie Ezine*+†

Progressive restriction to a differentiation pathway results from both activation and silencing of particular gene expression programs. To identify the coexpression and the expression levels of regulatory genes during hematopoietic stem cell (HSC) differentiation toward the T cell branch, we applied a new single-cell RT-PCR technique to analyze the simultaneous expression of 13 genes in 9 functionally purified populations from the bone marrow and the thymus. We report in this paper that Lin−Sca1+c-kit+ HSCs display, at the single-cell level, a homogeneous and high transcriptional activity as do early thymic progenitors. Moreover, the coexpression of lymphoid and myeloid genes is an early event detected in ∼30% of short-term HSC and most multipotent progenitors, suggesting novel sources for the generation of early thymic progenitors, common lymphoid progenitors (CLPs), and common myeloid progenitors. Loss of multipotency in Lin−Sca1+c-kit+ cells directed to the lymphoid branch is characterized by Lmo2 and Gata2 gene expression downregulation. Indeed, highest levels of Gata2 expression are detected only in long-term and short-term HSC populations. Complete shutdown of Pu1 gene expression in all triple-negative (TN3) stage thymic pre-T cells is indicative of total T cell commitment. Interestingly, this is also observed in 30% of TN2 cells and 25% of CLP in the bone marrow, suggesting a possible initiation of T cell engagement in TN2 and CLP. Also, our strategy highlights similar gene patterns among HSCs and intrathymic progenitors, proposing, therefore, that identical activation signals are maintained until further maturation and generation of CD4 and CD8 coreceptors bearing thymocytes. The Journal of Immunology, 2010, 184: 4907–4917.

C

Do these processes occur progressively or abruptly? When are lineage-specific genes turned on? When is the final decision made? Although some of the transcriptional regulators involved in early stages of lymphocyte development have been identified, the molecular mechanisms by which HSCs and lymphoid-restricted progenitors are instructed toward a T cell fate have yet to be defined. HSCs are clonogenic cells that possess both self-renewal properties and a multilineage potential (and can thus give rise to all types of mature blood cells). In adult mice, HSCs reside in the bone marrow (BM) and differentiate (via many several intermediate stages) into lymphoid-restricted progenitors (common lymphoid progenitors [CLPs]), which in turn develop into B, NK, T, and dendritic cells (1, 2). The progressive restriction of cell fate from the HSC stage to the CLP stage is characterized by the sequential loss of megakaryocyte/erythroid and granulocyte-macrophage (GM) potentials (3). The key molecular determinant in T lineage restriction is signaling by the membrane protein Notch, which inhibits B cell development in progenitors (4, 5). The thymus expresses an abundance of Notch-activating ligands, including the Delta-like family member DL-4 in particular (6). This environment is thus conducive to high-intensity Notch signaling in progenitor cells that arrive from the BM and are destined for T cell differentiation. The earliest thymic progenitors lack CD3, CD4, and CD8 molecules and are thus referred to as triple-negative (TN) cells. These TN cells can be further divided into four subsets on the basis of CD44, c-Kit, and CD25 expression (7). Within the heterogeneous CD44+CD25+TN1 population, the early T lineage progenitors (ETPs) are c-Kit+ multipotent cells (8). T cell specification occurs as ETPs differentiate into c-Kit+CD44+CD25+TN2 cells, which have lost B cell and most myeloid potentials. The TN2-TN3 transition (during which both CD44 and c-kit are downregulated) correlates with full commitment to the T cell lineage (9). It is at this point that pre-TCR signaling (through

*Institut National de la Santé et de la Recherche Médicale, Unité 591; 1Université Paris Descartes, Unité Mixte de Recherche S591; and 2Assistance Publique-Hôpitaux de Paris, Site Necker-Enfants-Malades, Service de Biostatistique et Informatique Médicale, Paris, France

†L.Ga. and A.B. contributed equally to this work.

Received for publication July 13, 2009. Accepted for publication February 26, 2010.

This work was supported by grants from the Ministère de la Recherche et de la Technologie and the Association de la Recherche sur le Cancer (to L.Ga.) and La Ligue Contre le Cancer and the Société Française d’Hématologie (to A.B.) and a postdoctoral fellowship from the Fondation pour la Recherche Médicale (to L.S.). This project was also funded in part by grants from the Institut National de la Santé et de la Recherche Médicale, Association de la Recherche sur le cancer, la Fondation pour la Recherche Médicale, l’Association Française contre les Myopathies, and the Action Concertée Thématique sur les Cellules Souches Adultes (sponsored by the Institut National de la Santé et de la Recherche Médicale, Association Française contre les Myopathies, Vaincre la Mucoviscidose et le Ministère de la Jeunesse, de l’Éducation et de la Recherche) (to S.E.).

Address correspondence and reprint requests to Dr. Sophie Ezine, Institut National de la Santé et de la Recherche Médicale, Unité 591, 156 rue de Vaugirard, F-75730 Paris 15, France. E-mail address: sophie.ezine@inserm.fr

The online version of this article contains supplemental material.

Abbreviations used in this paper: BM, bone marrow; CI, confidence interval; CLP, common lymphoid progenitor; DP, double-positive; ETP, early thymic progenitor; GM, granulocyte-macrophage; HSC, hematopoietic stem cell; ISP, immature single-positive; Lin, lineage; LSK, Lin−Sca1+c-kit+; LT, long-term; MPP, multipotent progenitor; NQ, not quantified; SP, single-positive; ST, short-term; TN, triple-negative.

Copyright © 2010 by The American Association of Immunologists, Inc. 0022-1767/10S16.00

www.jimmunol.org/cgi/doi/10.4049/jimmunol.0902184
a rearranged β-chain bound to a surrogate α-chain) is required for cells to survive and progress through β-selection. Pre-TCR signaling results in CD25 downregulation, which in turn induces the transition to the TN4 stage. Cells that have undergone β-selection exhibit a burst of proliferation, followed by upregulation of first CD8 (the immature single-positive [ISP] stage) and then CD4 (the double-positive [DP] stage). The latter cells rearrange their TCRα gene locus and undergo positive and negative selection. Positively selected DP thymocytes that escape negative selection then differentiate into single-positive (SP) CD4 or CD8 cells.

Functional analysis is fundamental in understanding how genomic expression profiles influence cell fate. Several recent reports have described gene expression analysis in hematopoietic cells in humans (10, 11) and mice (12, 13). This type of study is usually performed by using microarrays. Despite the fact that microarrays can determine the expression of hundreds or even thousands of genes simultaneously, it has major limitations. First, microarrays can identify so many new genes that identification of truly interesting ones becomes haphazard. Second, microarrays only report on average gene expression at the cell population level and can thus overlook possible cell-to-cell differences in expression patterns and/or cell fate. Because the events occurring in each individual cell are unknown, current methods may fail to identify the gene expression balance that ultimately establishes the T lineage identity. Indeed, each cell subset has a known potential and a specific position on the differentiation pathway. Cell markers are hallmarks for the various differentiation pathways and have given rise to the paradigm whereby cells with the same phenotypic profile are identical. Hematopoietic subsets and lineage commitment have been studied using single-cell RT-PCR (14, 15). However, lineage choice and differentiation are greatly influenced not only by the presence or absence of a specific transcription factor but also by its concentration (16). Thus, depiction of the molecular events that occur between each developmental stage must use a strategy that evaluates the quantitative heterogeneity of populations by analyzing gene expression modulations for each step. To this end, we applied a single-cell multiplex RT-PCR technique (17) for the si-}

Materials and Methods

Mice

All analyzed cell populations were sorted from 2-mo-old C57BL/6 (B6) wild-type mice (both males and females). B6 mice were purchased from Centre d’Elevage R. Janvier (Le Genest St. Isl, France).

Abs

The following mAbs were used for cell sorting and were obtained from BD Pharmingen (San Diego, CA): anti-CD3 (145-2C11), anti-CD4 (RM4-5), anti-CD8α (53-6.7), anti-CD8β (H35-172), anti-CD11b/Mac-1 (M1/70), anti-CD19 (1D3), anti-CD24 (HSA and M1/69), anti-CD25 (IL-2Rx and PC61), anti-CD44 (H-CAM and 1M781), anti-CD117/c-kit (SCF receptor and 2B8), anti-CD127/IL-7Rx (A7R34), anti-NK1.1 (PK136), anti-Sc-1 (stem cell Ag-1 and E13-161.7), anti-Ly-5/Gr1 (RB6-8C5), anti-TCRβ (HAM or H5-579), and anti-TCRα (GL3). They were directly coupled to FITC, allophycocyanin, PE, and PerCP or conjugated with biotin (the latter being revealed by streptavidin-allophycocyanin or streptavidin-PECy7 [BD Pharmingen]).

For cell sorting, BM cells were first incubated with unconjugated TER119 and Ly-6C/Gr1 (RB6-8C5) rat Abs, which are specific for erythroid and myeloid cell subsets, respectively. Positive cells were magnetically depleted with sheep anti-rat IgG-conjugated beads and sheep anti-mouse IgG-conjugated beads (Dynabeads M-450; Dynal Biotech, Oslo, Norway). The remaining cells (Lineage-negative cells, Lin−, enriched) were labeled with Abs against c-kit, Sca-1 and lineage (Lin) Ags (CD3, CD19, Mac-1, NK1.1, and TCRβ). The anti-IL-7Rx mAb was additionally used to isolate CLPs from LSK cells, and LSK subsets purification. Lin−-enriched cells were stained after incubation with B220, Gr-1, CD8, CD4, CD5, TER119 uncoupled Ab, and magnetic depletion; the remaining cells were stained with Ab against c-Kit, Sca-1, Flt-3, CD34, and lineage Ags (CD3, CD19, Mac-1, NK1.1, and DX-5). Cells were sorted on a BD FACSAria I for the indicated phenotypes: long-term (LT)-HSC (LSK CD34+FLT3−), short-term (ST)-HSC (LSK CD34−FLT3−), and multipotent progenitor (MPP) (LSK CD34+FLT3+).

For cell sorting of TN CD3+ CD4− CD8− cells in the thymus, cell suspensions were first incubated with unconjugated TER119, CD5 (53-7.3), and CD8α (LyT2) rat Abs. Positive cells were removed magnetically with sheep anti-rat IgG-conjugated beads (Dynabeads M-450; Dynal Biotech) and sheep anti-mouse IgG-conjugated beads. The negative fraction was labeled with Abs against CD25, CD44 and lineage Ags (Mac-1, 9C5, NK1.1, TCRβ, TRKβ, CD8β, and CD19). The anti–c-kit mAb was additionally used to isolate ETPs from total TN1 cells. ISP, DP, and SP cells were sorted after depletion of erythroid and B cells. The phenotypes used to sort these populations were CD8α+CD4−TCRβ+ HSA−, CD8α+CD4+TCRβ− B and CD8α+CD4+TCRβ− B mixed with CD4+CD8α+TCRβ− B, respectively.

FACS

Cell sorting was performed on a FACSVantage upgraded with DIVA software and equipped with an automatic cell deposition unit (BD Biosciences, San Jose, CA). Cells were collected in individual PCR tubes containing 5 µL PBS-diethyl pyrocarbonate 0.1% and stored at −80°C.

Primers design

Gene sequence data and exon/intro boundaries were obtained from the Ensembl (www.ensembl.org) and NCBI nucleotide databases (www.ncbi.nlm.nih.gov/entrez). Primers were designed manually, according to the strict rules described in Ref. 17. Briefly, 3′ and 5′ primers were chosen in intron or exon to avoid genomic encompass. We designed 20-bp primers with similar melting temperatures and a guanine-cytosine content close to 50% to obtain similar amplification efficiencies. Moreover, amplified fragments from the different genes were similar in length. To prevent primer competition, we selected primers and potential amplicons that did not cross-hybridize. Primer dimers were excluded. Primer compatibility and the size of the amplified fragments were assessed using the freely available Amplify1.2 software (http://engels.genetics.wisc.edu/amplify). The selected primers are listed in Supplemental Table I.

cDNA from BM and thymus progenitors

Control cDNAs used for optimization of the technique were isolated from BM and thymus progenitors. First, BM and thymus cell suspensions were enriched in lineage-negative cells with the SpinSep Mouse Progenitor Pre-Enrichment Cocktail (StemCell Technologies, Vancouver, British Columbia, Canada). RNAs were extracted from these cells using the μMACS mRNA Isolation Kit. cDNAs were synthesized by incubation for 1 h at 37°C with 1.1 mM poly(T) (Applied Biosystems, Foster City, CA) in a 360-µL volume reaction containing 50 mM KCl and 10 mM Tris HCl at pH 8.3 (Applied Biosystems), 3.3 mM MgCl2 (Applied Biosystems), 2.5 mM 2′-deoxycytidine 5′-triphosphates (Applied Biosystems), 960 U RNase block (Strategene, La Jolla, CA), and 840 U murine leukemia virus reverse transcription (Applied Biosystems). The reaction was stopped by a 10-min incubation at 95°C. cDNAs from BM and thymus progenitors were mixed, diluted, aliquoted, and stored at −20°C.

RT-PCR

Cells were lysed by freezing at −80°C, followed by heating to 65°C for 2 min. After cooling at 4°C, RNA was specifically reverse transcribed for 1 h at 37°C and then incubated for 3 min at 95°C to stop the reaction. Next, cDNAs generated by the reverse transcription reaction were amplified by

Downloaded from http://www.jimmunol.org/ by guest on October 30, 2017
a seminested PCR. The first round of PCR consisted of a denaturation step at 95˚C for 10 min, 15 amplification cycles (45 s at 95˚C, 60 s at 60˚C, and 90 s at 72˚C) and a final step at 72˚C for 10 min. This simultaneous amplification of all cDNAs was followed by a second round of specific PCRs: the first-round PCR products were separated and amplified with specific primers. The second round of PCR consisted of a denaturation step at 95˚C for 10 min and then 54 amplification cycles (30 s at 95˚C, 45 s at 70˚C, and 60 s at 72˚C, with the hybridization temperature decreased from 70–60˚C every other cycle). The types and amounts of reagents used for reverse transcription and PCR amplifications were described in Ref. 17. PCR products were detected on a 1.5% agarose ethidium bromide gel.

Standard preparation

The E2a gene was chosen from the multiplex gene list as the standard for the quantitative RT-PCR. LSK cells were lysed and RNA was reverse transcribed by a specific reverse transcription. The resulting cDNAs were amplified in a seminested PCR using specific 3’ and 5’ E2a primers. E2a cDNA was extracted from a 1.5% agarose ethidium bromide gel and purified using the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI). The purified DNA was then quantified by incorporation of Picogreen dsDNA Quantitation Reagent (Molecular Probes, Eugene, OR). Absorbance was assessed using the ABI PRISM 7900 HT Sequence Detection System (Applied Biosystems). Next, the concentration of the purified cDNA was determined by using the absorbance values of a serial dilution with Picogreen dsDNA Quantitation Reagent (Molecular Probes). Last, the standard was determined by including six different concentrations of E2a cDNA.

Real-time quantitative PCR

Real-time quantitative PCR was performed by adding 12 µl of SYBR Green PCR Master Mix (Applied Biosystems) to each well containing 4 µl of first PCR products and 8 µl of a primer mix with 0.25 µM of each specific primer in a 24-µl reaction volume, using the ABI Prism 7900 HT Sequence Detection System (Applied Biosystems). After a denaturation step at 95˚C for 10 min, 60 amplification cycles (30 s at 95˚C, 30 s at 60˚C, and 45 s at 72˚C) were performed. Slope values were determined for the exponential PCR phase by using the Sequence Detector System software (version 2.2; Applied Biosystems). The PCR efficiency was assessed for the linear phase of the reaction using LinRegPCR software (version 7.0). Statistical tests of the technique’s validity (t-test and an ANOVA) were performed using GraphPad Prism for Windows software (version 4; GraphPad, San Diego, CA).

Statistical analysis

Statistical analysis of gene expression data were performed in the R statistical computing environment (www.R-project.org). Limiting dilution analyses were performed with the StatMod software package (18, 19). Gene expression frequencies in the cell populations were estimated using a complementary log-log binomial, generalized, linear model. Two-sided 95% Wald confidence intervals (CIs) were computed (except in the case of zero outgrowths, when 95% Wilson intervals were used). It was assumed that only one limiting cell subset was necessary and sufficient for generating a positive response (the single-hit kinetics model). Any deviation from the single-hit kinetics assumption was tested using the statistical slope test from the complementary log-log binomial, generalized, linear model. Gene expression frequencies in different cell populations were compared by introducing a group variable into the model. A heat map representation was generated by clustering expression frequencies profiles according to genes or cell populations. We defined the distance with the nonparametric Spearman correlation coefficient both overall and for any of the studied genes (t-test; p > 0.05) (Supplemental Fig. 1 C) other than Ets1 and Pu1, for which amplification was more efficient in the multiplex condition than separately.

Gene expression frequencies among HSCs, committed lymphoid progenitors, and intrathymic progenitors support their combination in a multiplex

We analyzed nine hematopoietic populations located along the T cell pathway. HSCs Lin−Sca-1hi-KIThi (LSK) and Lin−IL-7Rα−Sca-1−KIt− cells (CLPs) were purified from the BM. Within the thymus, we isolated ETP (Lin−CD44−CD25−KItb), TN2 (Lin−CD44+CD25−), TN3 (Lin−CD44−CD25+), TN4 (Lin−CD44−CD25−), ISP (CD8α+CD4+TCRB+HSA+), DP (CD8α+CD4−TCRB+), and SP (CD8α+CD4−TCRB− mixed with CD4+CD8α−TCRB−) populations.

The cell sorting data are shown in Supplemental Fig. 2. Between two and four sorts were performed for each population. Cells were double-sorted into individual PCR tubes at different cell dilutions (from 1–500 cells/well) (Supplemental Table 2). For each gene frequency, between one and four different cell dilutions were performed. Next, frequencies were calculated using a binomial law. For each cell dilution, >24 samples were analyzed. Single-cell analysis was performed with ~40 cells. CIs were also evaluated for each calculated frequency.

The list of genes constituting the multiplex is given in Table I and represents genes involved in HSC development (Cmyb, Gata2, Am11, and Puf5) (19, 39, 40, 41), in lymphoid development (Puf5, Ets1, E2a, Il7ra, Ifi1, Id3, and Lmo2), in myeloid development (Mpo, Gm-csf, and Puf5), and in megakaryocyte and erythrocyte development (Fig1) (42). The first step involved performing a gene expression analysis on the various subsets, to determine the expression pattern for each gene along the T cell differentiation pathway (Fig. 1). This was obtained from the limiting dilution analysis shown in Supplemental Table II. For the developmental reference gene Il7ra, the expression results were consistent with cell analysis and sorting; the highest frequency was found in the CLP population (1). In the thymus, only the TN2, TN3, and SP populations showed significant levels of Il7ra expression (Fig. 1C), and this was confirmed at the protein level (Supplemental Fig. 3).

From the frequency of expression of a given gene within a specific subset, four major groups of genes were constituted according to their differential representation within the selected subsets (Fig. 1). The first group was continuously expressed at a high frequency in LSK+ HSC cells through to ISP cells and...
characterized by genes involved in HSC and lymphoid development, such as C-myb, E2a, Aml1, and Ifi1 (Fig. 1A–D). The second group of genes (Fog1 and Lmo2) was also expressed in all subsets but not at the same frequency (Fig. 1E, 1F). Notably, the frequency of cells expressing Lmo2 is abundant in HSC and decline thereafter. The third group contained genes expressed in HSC, myeloid, and lymphoid lineages (Ets1, Pu1, Mpo, Gata2, and Gm-csfr) and was detected in early-stage populations only— from LSK through to TN2 (Fig. 1G–J, 1M). In contrast, Id3 was the only gene for which the frequency of positive cells was significantly upregulated in thymocytes, after the TN3 pre-T stage (Fig. 1K). The Id3 gene is known to be involved in both αβ and γδ T cell development (39, 40).

Hence, this analysis evaluates for the first time, the frequency of the indicated genes among HSCs, lymphoid committed progenitors, and intrathyrmic and mature thymocytes. It shows their differential expression at various developmental stages sustaining their combination within a multiplex. To analyze the levels of expression, we proceed with the quantitative analysis of each gene in single cells.

Quantitative gene expression analysis in single cells reveals the intrinsic heterogeneity of lymphoid committed and intrathyrmic progenitors compared with HSCs

Fig. 2 shows the number of mRNA molecules detected for each gene within 30–47 individual cells in the BM (LSK and CLP) and the thymus (from ETP to SP populations). Each graph refers to one gene and each point in a graph represents a single cell.

The data confirmed that C-myb, E2a, Aml1, and Ifi1 genes are abundantly and permanently expressed by all analyzed populations. Levels of expression are the highest in LSK HSC and CLP compared with intrathyrmic progenitors. Interestingly, these data revealed that the lymphoid genes E2a and Aml1 are expressed in HSCs. More mature thymic populations (DP and SP) present a higher heterogeneity, on a per cell basis, compared with previous immature stages, pointing to a differential gene usage in the thymus. Within the second group of genes, Fog1 and Lmo2 were also highly transcribed in the LSK population and abruptly shut off (Fog1) or downregulated (Lmo2) in CLPs. When detected, Fog1 maintained a constant level of expression in the thymus until the ISP stage. In contrast, the level of expression of Lmo2 decreased significantly upon thymic entry and then remained low. Indeed, ectopic expression of Lmo2 enhances T cell tumors (41). Therefore, in normal T cell development, expression of this gene is progressively shut down from the CLP stage.

In the third group, it is noteworthy that the myeloid gene Mpo (transiently expressed in the thymus) was mostly detected in the thymus, among ETPs and only rare cells were scored positive among HSCs. It might indicates that Mpo+ LSK cells seed the thymic ETP compartment where they proliferate. In the human thymus, Mpo+ cells are reported among the earliest progenitors (42). Involved in HSC survival and proliferation (28), Gata2 gene expression levels revealed substantial heterogeneity with high expressors present only in LT- and ST-HSC populations compared with MPP (Supplemental Figs. 2F, 4).The rare low Gata2-expressing ETPs might well represent circulating MPP and/or recent thymic seeding cells (43). Last, Id3 expression is detectable at the TN3 stage, before pre-TCR selection, at the highest level and in TN4 cells. Recent studies suggested its involvement in the restriction of the developmental potential of the γδ lineage (44).

Hence, quantitative gene expression analysis in single cell revealed the strong molecular heterogeneity within unique cell surface marker profiles. It showed, also, that intrathyrmic progenitors can express certain genes with the same intensity as BM LSK and CLP progenitors do. Moreover, some genes are progressively modulated, whereas others are definitively shut down to mark the abrogation of a specific potentiality and/or of a signaling pathway. Therefore, on the basis of this gene network, the lymphoid branch conserve most HSC genes; in contrast, these are lost in immature (DP) and mature (SP) T cells.

<table>
<thead>
<tr>
<th>Genes</th>
<th>Description</th>
<th>Known Role(s)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aml1</td>
<td>Other name: Runx-1</td>
<td>Intra-thymic differentiation</td>
<td>20</td>
</tr>
<tr>
<td>C-myb</td>
<td>Proto-oncogene</td>
<td>HSC homeostasis and T cell development</td>
<td>21, 22</td>
</tr>
<tr>
<td>Ets1</td>
<td>A member of the Ets family</td>
<td>NK and early thymocytes differentiation</td>
<td>23, 24</td>
</tr>
<tr>
<td>E2a</td>
<td>Codes for an E protein</td>
<td>T and B cell development</td>
<td>25, 26</td>
</tr>
<tr>
<td>Fog1</td>
<td>Friend of Gata1</td>
<td>Erythroid and megakaryocyte differentiation</td>
<td>27</td>
</tr>
<tr>
<td>Gata2</td>
<td>Encodes for zinc finger GATA-2 protein</td>
<td>Survival or proliferation of HSCs</td>
<td>28</td>
</tr>
<tr>
<td>Gm-csfr</td>
<td>GM-CSF receptor</td>
<td>Growth factor essential for granulocyte and macrophage differentiation</td>
<td>29</td>
</tr>
<tr>
<td>Id3</td>
<td>A member of the Id protein family</td>
<td>Inhibition of E proteins</td>
<td>30</td>
</tr>
<tr>
<td>Il7ra</td>
<td>The IL-7Ra-chain</td>
<td>T and B cell development</td>
<td>31</td>
</tr>
<tr>
<td>Ifi1</td>
<td>IFN regulatory factor-1</td>
<td>NK and T cell development</td>
<td>32, 33</td>
</tr>
<tr>
<td>Lmo2</td>
<td>Oncogene (rhombotin-2)</td>
<td>T cell differentiation</td>
<td>34</td>
</tr>
<tr>
<td>Mpo</td>
<td>Codes for myeloperoxidase</td>
<td>Catalyses reactions forming cytotoxic radicals</td>
<td>35</td>
</tr>
<tr>
<td>Mrp-S21</td>
<td>Codes for the S21 protein of the 28S subunit of the 55S ribosome, expressed by all mammal cells = housekeeping gene</td>
<td>Protein translation</td>
<td>36</td>
</tr>
<tr>
<td>Pu1</td>
<td>A member of the Ets family</td>
<td>Involvement in myeloid, B and T cell development</td>
<td>37, 38</td>
</tr>
</tbody>
</table>
FIGURE 1. Modulation of gene expression frequencies during T cell differentiation. Each graph represents the expression frequencies of a single gene at nine different hematopoietic stages. BM and thymus populations were isolated and gene expression was assessed using a qualitative RT-PCR multiplex. The expression profile of each population was determined for all genes (A–M) by using either a complementary log-log binomial, generalized, linear model based on three to four cell dilutions or single-cell data. The y-axis represents the percentages of gene expression. The 95% CIs are indicated.
homogeneous. Importantly, lymphoid and myeloid lineage-specific genes are expressed in all cells analyzed and this reveals that expression precedes commitment. The committed lymphoid progenitor CLP (Fig. 3B) is sorted by the IL-7Rα cell surface expression, marker of the population and thus expressed this gene in all cells (Fig. 3B). This subset is characterized by the total shut down of Gata2 gene expression. Indeed, transduction of Gata2 in CLPs immediately induces a significant reduction in the expression of lymphoid-affiliated transcription factors (45). Hence, an extremely well-regulated sequence of gene expression activation or inactivation steps (of which Gata2 shutoff might be the initial event) controls lineage specification.

Compared with LSK, Lmo2, C-myc, Aml1, and Pu1 gene expressions are partially shut down in few CLP cells but remained highly expressed in other cells (Fig. 3B). Thus, only ∼40% of CLP express the five coexpressors of the stem cell cluster; in contrast, in absence of Lmo2, a new cluster is detected representing 20% of CLP and characterized by four coexpressors (Fig. 4B). Therefore, new gene patterns and new clusters are identified following lymphoid commitment. Hence, this multiple lymphoid subset revealed more diverse cluster combinations than the multipotent LSK populations (Fig. 4A, 4B), suggesting that new signaling pathways had been activated.

Within the thymus, the multipotent ETP subset can still generate myeloid, B, NK, dendritic, and T cells. As such, characterization of multipotency is revealed by the fact that ∼50% expressed the stem cell cluster (Fig. 4C) and that all cells express Pu1 with a high intensity as LSK (data not shown); in addition, 30% of ETPs express the reduced cluster, identified in CLP, with four coexpressors (Fig. 4C). Therefore, this keeps open all potentialities with a bias toward the lymphoid branch. Interestingly, a typical feature of the ETP population is Mpo expression in almost all cells, rarely detected in preceding populations (Fig. 3C).

The next thymic differentiation step, TN2, defined as CD44−CD25+, conserved myeloid, NK, and DC potentials. Less than 20% of the population expressed the stem cell cluster identified in LSK (Fig. 4D). However, the reduced cluster, lacking Lmo2 expression, with four coexpressors C-myc, E2a, Irf1, and Aml1, is present in ∼70% of TN2 cells; it was thus named the “thymic cluster” (Fig. 4D). Overall, levels of expression are slightly reduced (Fig. 3D). Although these changes are related to modulations of gene expression, other modifications concerned gene shut down and new expressions. Indeed, Pu1-negative cells (∼30%) are detected, indicating progressive loss of B, myeloid, and dendritic cell potentialities (Fig. 3D). In contrast, high IL-7Rα expressers are present (in ∼50% TN2), at the same intensity that CLP; they are also present at the protein level (Supplemental Fig. 3). Altogether, at this step of differentiation, T cell specialization started with the abrupt shutoff of Pu1 in few TN2 cells, with only rare cells expressing high levels of Pu1 gene (data not shown). The transition to the TN3 pre-T stage, marks the onset of restricted thymic precursors, with full pre-T cell potential in all cells. At the cellular level, it is characterized by the loss of CD44, and cells are identified as CD44−CD25+. This full commitment is depicted by the total loss of Pu1 expression characterizing the pre-T cell stage (Fig. 3E). As for TN2 cells, the thymic cluster C-myc, E2a, Irf1, and Aml1 is represented in ∼70% of TN3 (Fig. 4E), and all cells coexpress high levels of Irf1 gene, warrant of survival and proliferation (Fig. 3E). In addition, part of this population coexpresses also high intensity of Id3. TN3 cells are on the way to become αβ T cells, although a small fraction is still capable of adopting the γδ fate (46), and Id3 expression seems to control γδ T cell production (44).

At the TN4 stage (Fig. 3F), two subsets can be distinguished according to either the total shutdown or maintenance of most of the genes studied here. One subset conserves only Cmyb and few Id3 expressers, at significant levels, whereas the second conserves the expression of most of the genes. The ISP population coexpresses most of the genes and seems to stem from the related TN4 subset, suggesting the death and/or the exit of the Cmyb+ (E2a Irf1 Aml1 Fog1)− TN4 subset (Fig. 4G).

Cells in TN4 (Fig. 4F) and ISP (Fig. 4G) express the thymic cluster (∼30 and 50%, respectively). However, it is clear that the levels of expression of most genes are diminished compared with the TN2 stage (Fig. 3).

FIGURE 2. Quantitative assessment of mRNA expression in single cells. Individual cells were sorted from BM (LSK and CLP) and thymus (ETP, TN2, TN3, TN4, ISP, DP, and SP) populations. Positive cells were scored for the expression of a particular gene (notably 28S rRNA expression) and further analyzed in a real-time quantitative RT-PCR. The logarithmic y-axis represents the absolute number of mRNA molecules per cell. Each graph corresponds to one gene, and each point in a graph corresponds to a single cell. Pairwise comparisons between the cell populations were performed with a Wilcoxon rank-sum test. *The statistical significance threshold was set to p < 0.05.
FIGURE 3. Quantitative coexpression pattern of 14 hematopoietic genes at the single-cell level. Individual cells were sorted either from BM [LSK and CLP (A, B)] or thymus [ETP, TN2, TN3, TN4, ISP, DP, and SP (C–I)] populations. Expression of the 14 genes was determined simultaneously in each individual cell using a real-time quantitative RT-PCR. Positive cells were scored for the expression of a particular gene (notably 28S rRNA expression). Each horizontal row depicts the same individual cell. Each vertical row represents a different gene. The colors indicate the range of the absolute number of mRNA molecules per cell (see the legend below the figure). Empty squares correspond to the absence of gene expression. Brown squares correspond to positive cells in which mRNA levels were not quantifiable. NQ, not quantified.
The further maturation steps lead to more mature, CD4- and CD8-coexpressing thymocytes and most of the previously expressed genes have now been shut down (Fig. 3H,3I). The initial network of genes is disrupted, reflecting probably, new regulations. Several combinations of coexpressed genes have been firmly changed, reflecting the multiple signaling pathways triggered by the CD4 and CD8 coreceptors (Fig. 4H,4I).

**Lymphomyeloid coexpressors are detected in ST-HSC populations**

An important question is to decipher how early, in the development of LSK subsets, lymphomyeloid priming takes place. To investigate this point, we evaluated among LT-HSCs, ST-HSCs, and MPPs the coexpression profile of the various genes within the multiplex. Data in Fig. 5 show the molecular signature of these populations. It clearly reveals that lymphoid genes (E2a, Aml1, Lmo2, and Pu1) are detected very early in LT-HSC population. In contrast, myeloid genes defined by Mpo and Gmcsfr are activated later, in ~35% of ST-HSC population. Altogether, these data point to the initiation of lymphoid-myeloid copriming in ST-HSC. This coexpression is increased in MPP and few IL-7Rα expressers are detected. Thus, ST-HSC and MPP represented the earliest subsets for lymphomyeloid copriming, and could be, potentially, new source for ETPs, CLPs, and common myeloid progenitors.

**Discussion**

We report in this paper, the quantitative coexpression analysis, at the single-cell level, of 13 genes in 9 hematopoietic progenitor populations isolated from the BM and the thymus. Our results show that LSK BM cells coexpress high levels of lineage-specific genes. Analysis of the coexpression profiles revealed that most regulatory genes expressed in LSK cells are also expressed in intrathymic progenitors (albeit at lower levels), suggesting that similar activation pathways are triggered. Importantly, we detected lymphomyeloid copriming in ST-HSC and MPP subsets, indicating that they might contribute to ETP, CLP, and common myeloid progenitor generation.

Our original strategy allowed to detect that BM LSK population is very active: it has a very homogeneous expression pattern and an extremely high transcriptional rate on a single-cell basis. This high transcriptional activity might explain the population’s general availability for all developmental options. Indeed, all programs are still available: mRNA transcripts of genes expressed in myeloid
(Pu1, Mpo, and Gmcsfr) and lymphoid lineages (Aml1, Pu1, Lmo2, and E2a) are detected among one-third of ST-HSC population. This last observation is in accordance with recent data on chromatin modifications (47) but had been missed in previous studies (15), which focused on the expression of late lymphoid differentiation factors (Cd3 and Pax5) rather than early ones. Therefore, combined with earlier studies by Adolfsson et al. (3), a fraction of ST-HSC coexpress megakaryocyte-erythroid-lymphoid-myeloid genes. Thus, in HSCs, high levels of expression of all lineage-affiliated genes precede commitment and multiple gene programs collaborate.

This study reveals that LSKs, CLPs, and thymic TN progenitors conserve the coexpression of several genes, identified in this paper as the stem cell cluster; this may account for some of the plasticity observed in CLPs and TN subsets. Indeed, with appropriate stimulation, HuIL2Rb-transduced CLPs can revert to GM populations via GM-CSF expression (48). However, this latter study did not mention the percentage of CLPs in which the GM program had been turned on. Furthermore, Hu-IL2Rb expression is still able to direct TN2 pro-T cells (but not TN3) toward GM and dendritic pathways. Hence, the full differentiation program is still permissive among intrathymic progenitors.

Also emerging from these data, highly expressed genes in LSK cells are progressively downregulated in intrathymic progenitors and turned off thereafter, with the steepest downregulation observed before the DP stage. Indeed, within the thymus, progenitors develop during their migration from the corticomedullary junction to the outer cortex; at this stage they reach the TN3 stage and rapidly process to the DP stage to start an inverse pathway toward the medulla (49). At the DP stage, they cross, again, the deep cortex and express the CD4 and CD8 coreceptors; these molecules drive diverse cellular connections and stimuli, disrupting the gene network established initially in LSK+ HSCs. Thus, the thymic environment is able to maintain the expression of stem cell genes and to further proceed to more specific T cell pathways.

Lymphoid commitment in the BM, represented by the analysis of the CLPs, is characterized by the abrupt and total abrogation of Gata2 expression and is indicative of the last phase of a multipotent stage (28). Most importantly, Lmo2 gene expression downregulation and heterogeneous expression on a single-cell basis is detected: some CLPs maintain the high level of Lmo2 transcripts detected in LSK cells, whereas others shut off its expression (as in all thymocytes). Indeed, retroviral overexpression of Lmo2 in CD34+ cells caused severe abnormalities in T cell development (50). Furthermore, Lmo2 transgenic mouse revealed dramatic perturbations of thymus differentiation prior to tumor formation (41); therefore, its downregulation must be sharply controlled for normal development. Altogether, absence of Gata2 and Lmo2 molecular markers could be used to identify and select subsets, within the Lin− BM progenitor fraction, directed toward a lymphoid cell fate.

Several studies have shown that loss of Pu1 expression is a hallmark of T cell specialization, characterizing the TN2 pro-T (Pu1+) to the TN3 pre-T cells (Pu1−) (51). We identify, in this study by the single-cell analysis approach, for the first time, Pu1− cells at the TN2 stage. This suggests that T cell specialization is a progressive process that started at the pro-T cell stage; some cells express also the Il7ra gene as TN3 pre-T cells. However, such Pu1− Il7ra− cells are also present in the BM at the CLP stage. Therefore, this might indicate that, first, some CLP cells can support further maturation steps toward T cell specialization within the BM, indicating T cell specialization before phenotypic changes (23, 37). Second, the identification of similar populations (Pu1− Il7ra−) among CLP and TN2 subsets could indicate that CLP enters the thymus and localizes in the TN2 niche. Indeed, it was reported that they could rapidly adopt the phenotype of TN2 thymocytes when cultured on Delta-like 1-expressing stroma cells.
Fe´ de´ ratif de Recherche 94 for technical support in single-cell sorting.

References

Development pathway. These profiles could be a powerful tool to fully express in HSCs from one differentiation stage to another. Hence, this strategy should be considered as an adjunct to all evaluating changes in the activity and contributions of known genes lineages they characterize). Nevertheless, our strategy established of each gene product remains the same. Our experiments only addressed the expression of certain regulatory genes (chosen according to their potentialities and the lineages they characterize). Nevertheless, our strategy established new criteria for learning about molecular T cell hematopoiesis and evaluated changes in the activity and contributions of known genes fully expressed in HSCs from one differentiation stage to another. Hence, this strategy should be considered as an adjunct to all microarray studies. For the first time, we have provided a molecular profile of related and defined multiple subsets along the T cell development pathway. These profiles could be a powerful tool to modulate production of cell types for therapeutic purposes. Inhibition of Gata2 and Lmo2 gene products may be useful to accelerate lymphoid commitment and/or production. Our work reveals the transcriptional heterogeneity behind cell surface Ags that are otherwise used to describe individual subsets. This analysis highlights the facts of a continuum of differentiating cells among each step, under the coordinated activities of multiple transcription factors. Therefore, both cellular and molecular markers must be considered and used to track normal and pathological development. This is the first study to analyze gene transcription status from HSCs to normal T cell differentiation, to provide a molecular signature, and to suggest molecular markers for isolation of specialized populations.

In conclusion, molecular and cellular markers are valuable tools for the determination of differentiation process. The strategy we used in this study provides for the first time the signatures of HSCs, CLPs, and intrathymic and mature thymocytes, using the same gene network. We believe that these studies will provide better understanding of hematopoiesis in a normal and pathological context. Moreover, this study provides new tools useful for characterizing stem and progenitors cells when limited in numbers and/or when their functional characterization in vivo is not possible.

Acknowledgments
We thank the members of Unite 591 for support. We also thank E. Schneider for reading the manuscript and J. M´egret and C. Cordier from the Institut F´ed´eratif de Recherche 94 for technical support in single-cell sorting.

Disclosures
The authors have no financial conflicts of interest.

References
