



**HAL**  
open science

## Temporal gene expression profiles reflect the dynamics of lymphoid differentiation

Smahane Chalabi, Agnes Legrand, Victoria Michaels, Marie-Ange Palomares, Robert Olaso, Anne Boland, Jean-François Deleuze, Sophie Ezine, Christophe Battail, Diana Tronik-Le Roux

### ► To cite this version:

Smahane Chalabi, Agnes Legrand, Victoria Michaels, Marie-Ange Palomares, Robert Olaso, et al.. Temporal gene expression profiles reflect the dynamics of lymphoid differentiation. International Journal of Molecular Sciences, 2022, 23 (3), pp.1115. 10.3390/ijms23031115 . cea-04346534

**HAL Id: cea-04346534**

**<https://cea.hal.science/cea-04346534v1>**

Submitted on 15 Dec 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Article

# Temporal Gene Expression Profiles Reflect the Dynamics of Lymphoid Differentiation

Smahane Chalabi <sup>1,2</sup>, Agnes Legrand <sup>3</sup>, Victoria Michaels <sup>3</sup>, Marie-Ange Palomares <sup>1</sup>, Robert Olasso <sup>1</sup>, Anne Boland <sup>1</sup>, Jean-François Deleuze <sup>1</sup>, Sophie Ezine <sup>3,\*</sup>, Christophe Battail <sup>1,4,†</sup> and Diana Tronik-Le Roux <sup>5,6,†</sup>

- <sup>1</sup> Centre National de Recherche en Génomique Humaine, CEA, Université Paris-Saclay, 91057 Evry, France; smahane.chalabi@gmail.com (S.C.); palomares@cng.fr (M.-A.P.); olasso@cng.fr (R.O.); boland@cng.fr (A.B.); deleuze@cng.fr (J.-F.D.); christophe.battail@cea.fr (C.B.)
- <sup>2</sup> Laboratoire Européen de Recherche pour la Polyarthrite Rhumatoïde—Genhotel, Université Paris-Saclay, Univ Evry, 91042 Evry, France
- <sup>3</sup> INSERM, Unité 1151, Université De Paris, Unité Mixte de Recherche 8253, 75015 Paris, France; agnes.legrand@inserm.fr (A.L.); vicky0678@gmail.com (V.M.)
- <sup>4</sup> Laboratoire Biologie et Biotechnologies pour la Santé, IRIG, UMR 1292 INSERM-CEA-UGA, Univ. Grenoble Alpes, 38000 Grenoble, France
- <sup>5</sup> Atomic Energy and Alternative Energies Agency (CEA), Department of Research in Hemato-Immunology (SRHI), Saint-Louis Hospital, 75010 Paris, France; diana.le-roux@cea.fr
- <sup>6</sup> IRSL, HIPI-UMRS 976, University of Paris, 75006 Paris, France
- \* Correspondence: sophie.ezine@inserm.fr
- † These authors contributed equally to this work.



**Citation:** Chalabi, S.; Legrand, A.; Michaels, V.; Palomares, M.-A.; Olasso, R.; Boland, A.; Deleuze, J.-F.; Ezine, S.; Battail, C.; Tronik-Le Roux, D. Temporal Gene Expression Profiles Reflect the Dynamics of Lymphoid Differentiation. *Int. J. Mol. Sci.* **2022**, *23*, 1115. <https://doi.org/10.3390/ijms23031115>

Academic Editor: Alessandro Poggi

Received: 14 December 2021

Accepted: 16 January 2022

Published: 20 January 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Understanding the emergence of lymphoid committed cells from multipotent progenitors (MPP) is a great challenge in hematopoiesis. To gain deeper insight into the dynamic expression changes associated with these transitions, we report the quantitative transcriptome of two MPP subsets and the common lymphoid progenitor (CLP). While the transcriptome is rather stable between MPP2 and MPP3, expression changes increase with differentiation. Among those, we found that pioneer lymphoid genes such as *Rag1*, *Mpeg1*, and *Dntt* are expressed continuously from MPP2. Others, such as CD93, are CLP specific, suggesting their potential use as new markers to improve purification of lymphoid populations. Notably, a six-transcription factor network orchestrates the lymphoid differentiation program. Additionally, we pinpointed 24 long intergenic-non-coding RNA (lincRNA) differentially expressed through commitment and further identified seven novel forms. Collectively, our approach provides a comprehensive landscape of coding and non-coding transcriptomes expressed during lymphoid commitment.

**Keywords:** lymphopoiesis; hematopoietic differentiation; RNAseq; long non-coding RNA; transcriptome; gene networks

## 1. Introduction

Hematopoietic stem cell (HSC) differentiation is a continuous process in which stem cells gradually lose their self-renewal capacity to generate hierarchically organized subpopulations of heterogeneous multipotent progenitors (MPPs) with decreasing self-renewal potential, which ultimately give rise to all mature blood cells. All these populations display different signatures that are mainly defined according to the presence or absence of a limited number of surface markers. The scarcity of the known markers prevents a fully homogeneous purification.

In mouse bone marrow (BM), all HSCs and MPPs reside in the lineage negative (Lin<sup>-</sup>) Sca-1<sup>+</sup>c-Kit<sup>hi</sup> (LSK) compartment (about 0.1% of all BM cells). Within this population, MPP subsets are described according to Slam antigens [1] or the expression of VCAM1 and Flt3 surface markers [2]. According to the latter classification, three subsets of MPPs

can be identified. MPP1 (VCAM1<sup>+</sup>Flt3<sup>-</sup>) the most immature, can generate all blood cell lineages. MPP2 (VCAM1<sup>+</sup>Flt3<sup>+</sup>) has lost megakaryocyte/erythroid potential and MPP3 (VCAM1<sup>-</sup>Flt3<sup>+</sup>) has lost myeloid potential and gives rise to the common lymphoid progenitors (CLPs) [3]. This more restricted progenitor (Lin<sup>-</sup>Sca-1<sup>low</sup>c-Kit<sup>low</sup>IL7Ralpha<sup>+</sup>) has lost the ability to produce myeloid cells *in vivo* but is able to generate mainly B cells, the most abundant lymphoid subset produced in the BM but also T cells, dendritic cells (DC), Natural Killers (NK), and Innate Lymphoid Cells (ILCs) [3,4].

Microarrays studies have provided useful insights into HSC and MPP biology [5–7]. Similarly, single-cell RNA (scRNA) sequencing has helped advance our understanding of the fate of hematopoietic cells [8–12]. However, the limited sensitivity of currently available scRNA approaches, capturing only the expression of a few thousand genes per cell, does not allow quantification of the entire protein-coding and non-coding transcriptomes. Despite these experimental improvements, the trajectories of MPPs to more committed populations and particularly to lymphoid cell populations, are still unresolved [13].

Several key molecules, such as transcription factors (TFs), are reported to orchestrate complex and fine-tune regulatory networks involved in the hematopoietic differentiation. TFs instruct lineage fate decisions and control cell transcriptomes as cells proliferate, develop, and respond to their environments. They trigger the activation or silencing of transcriptional programs through their binding to specific sequences in regulatory regions, called TF binding sites (TFBSs). Noteworthy, deregulation or chromosomal translocations of TFs that act on the hematopoietic system disturb the homeostatic balance and prompt human malignancies [14] emphasizing the need to extensively identify their dynamic patterns of expression. Several reviews provide insights into the functions of some critical TFs involved in different hematopoietic lineages [15–18]. Some of these such as the C/EBP family, PU.1, E2A, IKAROS, and FOXO1 are general TFs [19]; whereas TCF7, E2A, IRF8, BCL11B are more specifically involved in the development of lymphoid cells [20]. Elucidation of the sequential emergence of these factors and their organization into gene regulatory networks would be a major step towards gaining a mechanistic understanding of cellular decision-making processes.

Another class of gene expression regulators, not extensively studied yet, include the long non-coding RNAs (lncRNA), which are RNA transcripts without protein-coding capacity that act at the transcriptional and post-transcriptional level. They have multifaceted roles in the field of physiology and the development of disease and have emerged as important constituents of regulatory networks [21,22].

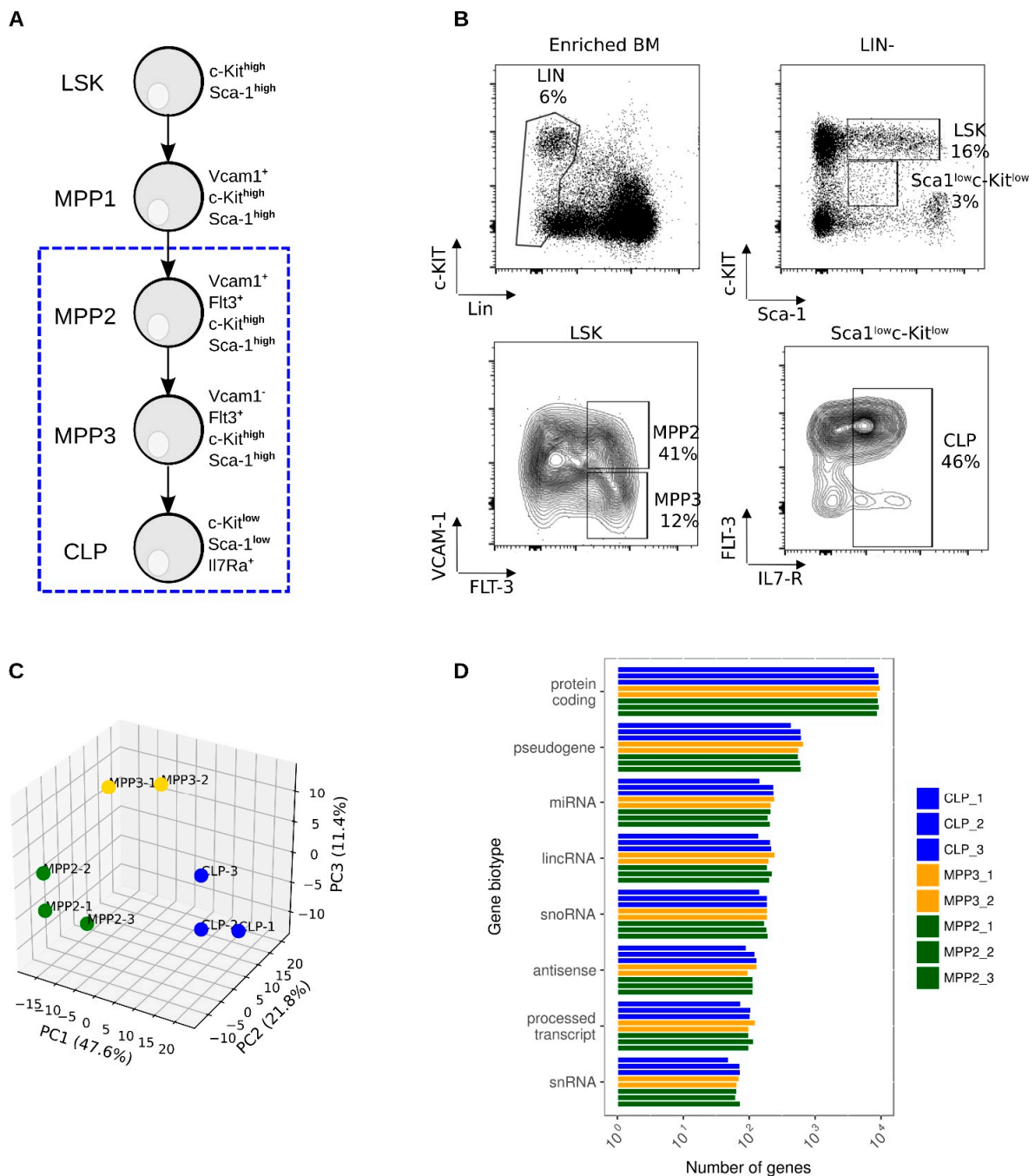
In the hematopoietic system, some lncRNAs control progenitor differentiation [23–27]. Other lncRNAs are involved in erythroid maturation and erythro-megakaryocyte development [28,29], or associated with myeloid committed progenitors such as *H19*, *Meg3*, and *Malat1* [30]. However, the full analysis of lncRNAs involved in the commitment of lymphoid progenitors is still to be performed. Hence, a current challenge is to perform the full count of lncRNAs and to translate their temporal expression modifications into molecular functions.

In this context, we carried out deep RNA-sequencing of the whole transcriptome of three different populations at specific stages of the lymphoid differentiation trajectory to order lymphoid gene expression through commitment from multipotent progenitors (MPP2 and MPP3) to the next committed subset, CLP. Our findings revealed specific and dynamic expression profiles for the different cell populations as the result of the action of TFs regulatory networks. Some of the newly identified markers may help to refine the purification of progenitors. Finally, we expanded the current repertoire of annotated and yet unidentified lncRNAs expressed in lymphoid progenitors. Altogether, the data provide novel gene expression information that may help a better understanding of the programs involved in the commitment of hematopoietic progenitor cells to the lymphoid lineage.

## 2. Results

### 2.1. Whole-Transcriptome Profiling of Hematopoietic Progenitors

Transcriptome analysis was performed to gain deeper insight into the temporal expression changes associated with the commitment of multipotent progenitors (MPPs) to the lymphoid lineage. To this end, MPP2, MPP3, and CLP populations were isolated by fluorescence-activated cell sorting (FACS) from the BM of B6 mice based on the surface expression of VCAM1, FLT3, and IL7Ralpha within the LSK population (Figure 1A,B). We aimed to analyze the total population of CLP to be able to have access to the different potentialities of this progenitor.



**Figure 1.** Isolation and total transcriptome sequencing of hematopoietic progenitor cells MPP2, MPP3, and CLP (A) Diagrammatic representation of the differentiation of LSK to MPPs and CLPs. The phenotype of each population is indicated. (B) FACS gates used to isolate BM progenitors: LSK

(Lin<sup>-</sup>Sca1<sup>+</sup>c-Kit<sup>+</sup>), MPP2 (LSK, VCAM1<sup>+</sup>Flt3<sup>+</sup>), MPP3 (LSK, VCAM1<sup>-</sup>Flt3<sup>+</sup>), CLP (Lin<sup>-</sup>Sca1<sup>low</sup>c-Kit<sup>low</sup>IL7R<sup>+</sup>). Lin<sup>-</sup> stands for lineage negative. One representative experiment out of nine is presented. The percentage (mean  $\pm$  SEM) of each population obtained from nine experiments with about 20 mice each are: Lin<sup>-</sup>: 10.1  $\pm$  1.6; LSK, within Lin<sup>-</sup> cells: 10.6  $\pm$  2.1; CLP: 41.9  $\pm$  3.0 within Lin<sup>-</sup>c-Kit<sup>low</sup>Sca1<sup>low</sup>. Within LSK: MPP1: 17.6  $\pm$  1.8; MPP2: 34.6  $\pm$  2.3; MPP3: 15.7  $\pm$  2.0. (C) Principal component analysis (PCA) of gene expression values generated from biological replicates of MPP2, MPP3, and CLP. Expression data were normalized in FPKM (Fragments Per gene Kilobase and per Million reads), adjusted using ComBat algorithm before their clustering by PCA. (D) Bar plot showing the distribution of gene biotype quantification from gene expression values.

On average 100 K cells were analyzed for each population in three independent biological experiments (Table S1). We generated sequence libraries from these low RNA inputs using the SMARTer Stranded RNAseq and sequenced at high depth of coverage to generate  $\sim 2 \times 120$  M reads per biological replicate.

Quality controls performed on raw and aligned reads were in line with what is expected for total RNA sequencing (Table S1). In particular, the distribution of aligned reads on the mouse genome indicated that  $\sim 25\%$  of alignments were localized in gene intronic regions. We observed a higher proportion of duplicated reads (62% on average), compared to classical RNA sequencing data. This is due to the low quantity of total RNA engaged in sequence library preparation. However, our high depth of coverage achieved per biological replicate allowed us to compensate for this read duplication rate. For further analysis, we only eliminated one replicate (MPP3 replicate 3) because of its very low read alignment rate (31% of aligned reads on mouse genome with only 5% of unique reads).

Transcriptome quantification (as described in the gene quantification panel, Figure S1), identified on average 10,742, 11,019, and 10,480 expressed genes (FPKM > 0.1 in at least one biological replicate before ComBat batch effect correction [31]) for MPP2, MPP3, and CLP populations, respectively (Table S2). Therefore, nearly the same proportion of genes is expressed in each progenitor. Principal component analysis (PCA) performed on this set of corrected genes expressed in each biological replicates revealed that the transcriptomes of the three hematopoietic progenitors were clearly separated (Figure 1C). Furthermore, PCA showed the nearby clustering of cell replicates. We also assessed reproducibility of transcriptomes generated from biological replicates by unsupervised clustering of expressed genes before and after batch effect correction (Figure S2). We found that the removal of batch effects, due to the technical variability associated with the different batches of mice, significantly improved the clustering.

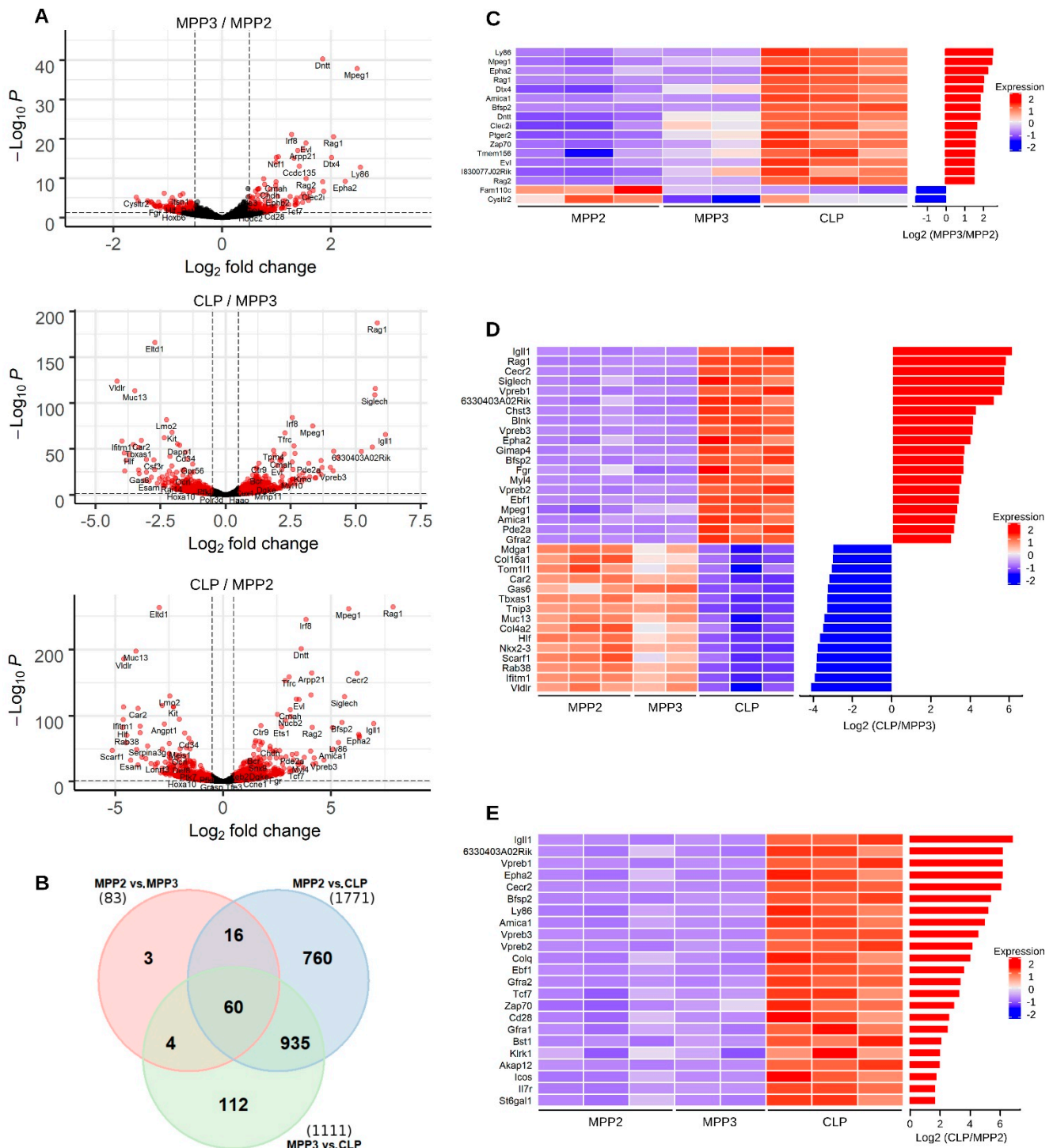
Classification of expressed genes into RNA categories indicated that on average 85% and 2% of expressed genes are protein-coding and long non-coding genes (lncRNA), respectively (Figure 1D). Moreover, distribution of expressed genes into RNA categories is highly reproducible within biological replicates and among progenitors, revealing the high suitability of the RNAseq data to be interrogated to obtain a comprehensive landscape of protein-coding and lncRNA genes involved in lymphoid commitment.

## 2.2. Identification of Novel Temporal Genes Expression Profiles Associated with Lymphoid Commitment

To identify gene expression changes that occur during the transition MPP2-MPP3-CLP, we first performed pairwise comparisons of protein-coding gene abundances. We detected in total 1890 differentially expressed genes (DEGs) (Table S3). The distribution of DEGs among MPP2 vs. MPP3, MPP3 vs. CLP, and MPP2 vs. CLP comparisons was 83, 1111, and 1771 DEGs, respectively ( $|\text{Log}_2$  Fold Change| > 0.5, FDR = 0.05). This reveals the stability of transcriptional programs expressed by MPP2 and MPP3, and the strong shift in expression at the MPP3-CLP transition (Figure S3).

A detailed analysis of the few DEGs identified at the MPP2-MPP3 transition showed that they were in majority over-expressed in MPP3 (Figure 2A top panel). As for the DEGs identified between the MPP2/MPP3 and CLP stages, they were distributed in a balanced manner between over- and under-expressed genes (Figure 2A middle and bottom panels,

respectively). The comparison of the DEGs revealed that 91% of the MPP2 vs. MPP3 DEGs (76 out of 83 genes) and 90% of the MPP3 vs. CLP DEGs (995 out of 1111 genes) were also identified by the MPP2 vs. CLP transition (Figure 2B).



**Figure 2.** Differential gene expression analysis between progenitor cells. (A) Volcano plots showing for each gene (indicated by dots) the logarithm of the ratio expression levels between MPP2 and MPP3 (top panel), MPP3 and CLP (middle panel), MPP2 and CLP (bottom panel) according to the logarithm of the adjusted *p*-value generated from differential gene expression analysis. Red dots depict significantly under- or over-expressed genes for each comparison of progenitor cells. (B) Venn diagram illustrating the overlap of differentially expressed protein-coding genes identified from

pairwise comparison between progenitor populations. (C,D) Heatmaps representing the abundance values of the most differentially expressed genes between MPP2 and MPP3, MPP3 and CLP cells, respectively. These genes were selected according to the logarithm of the ratio between their expression levels in the pairs of cell populations (fold change). The bar plots (right panel) illustrate the  $\text{Log}_2$  fold change per gene, for each pairwise comparison. (E) Heatmap showing the abundance values of the most differentially expressed genes between MPP2 and CLPs, associated with a very weak expression in MPP2. The bar plot (right panel) illustrates the  $\text{Log}_2$  fold change per gene.

To better understand the differentiation process, we analyzed the main DEGs associated with each stage. We first selected DEGs between MPP2 and MPP3 with  $|\text{Log}_2(\text{MPP3}/\text{MPP2})| > 1.5$  (Figure 2C). Among the top 17 genes, nine were associated with the lymphoid immune system pathway (*Ly86*, *Dntt*, *Dtx4*, *Rag1*, *Rag2*, *Arpp21*, *Zap70*, *Clec2i*, *Bfsp2*) suggesting that these nine genes represent the earliest lymphoid genes already expressed in MPP.

Regarding the top DEGs between MPP3 and CLP (Figure 2D,  $|\text{Log}_2(\text{CLP}/\text{MPP3})| > 3$ ), we found that the cytoskeleton and adhesion genes were the most highly modulated. These categories include upregulated genes such as *Myl4*, *Epha2*, *Fgr*, *Amica1* and downregulated genes such as *Col4a2*, *Col16a1*, *Scarf1*. This probably reflects the circulation/move of CLP towards the lymphoid niche, as recently suggested [32]. In addition, CLP highly express several genes of the immune system, such as the B cell specific genes *Blnk*, *Siglech*, *Ebf1*, *Rag1*, *Igll1*, *Vpreb1*, *Vpreb2*, *Vpreb3*, *Gfra2*, and *Mpeg1*. Moreover, *Gimap4*, a specific target gene of *Pu1* within T-cell-progenitors is also detected [33].

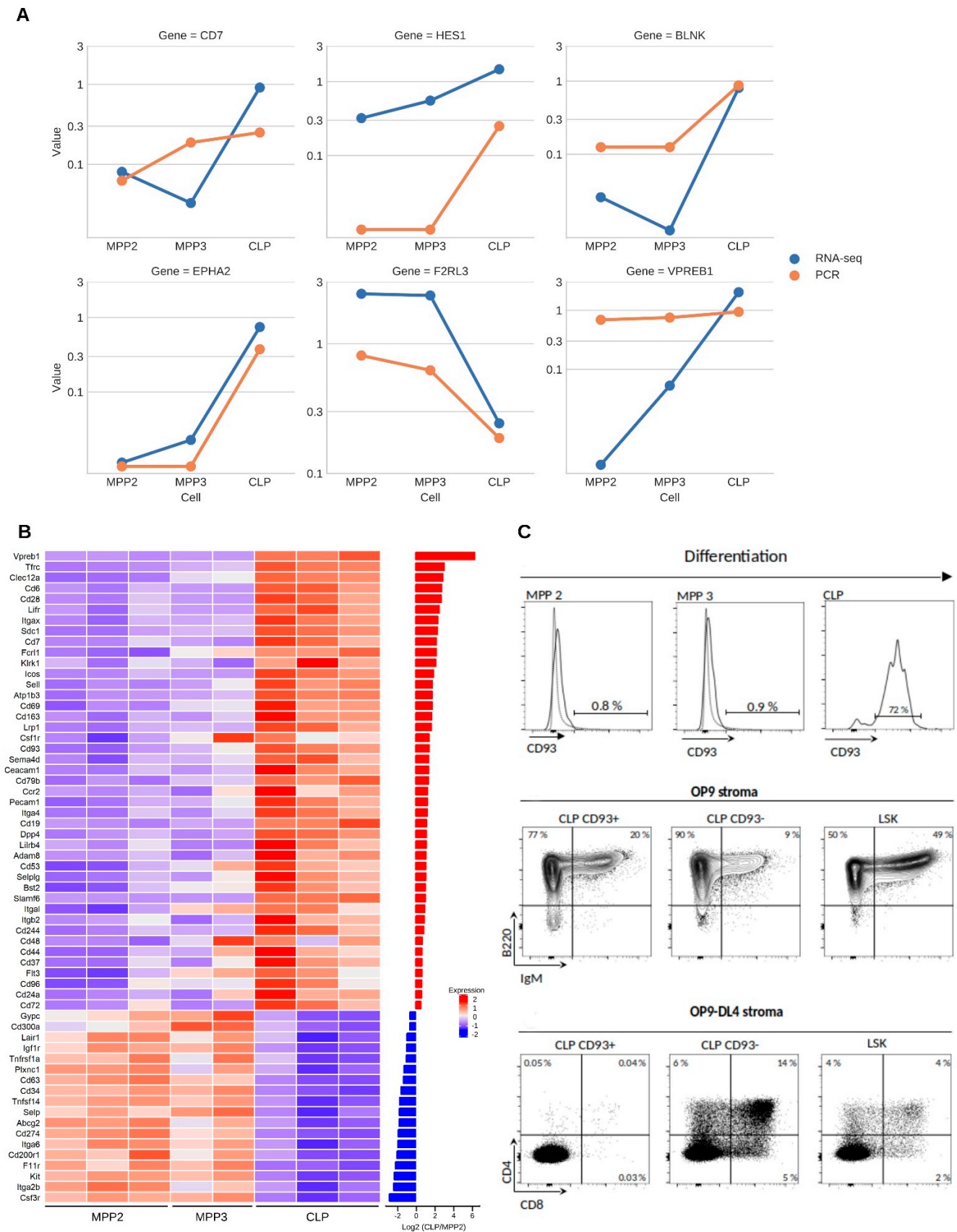
To identify markers that emerge specifically in CLP, we look for genes very weakly expressed in MPP2 and over-expressed in CLP (mean ( $\text{Log}_2$  MPP2)  $< 10^{-2}$  FPKM and  $|\text{Log}_2(\text{CLP}/\text{MPP2})| > 1.5$ ). Among the 23 genes detected, 19 were B and T cell specific genes (Figure 2E), demonstrating the presence of both programs in CLP. Applying an additional filter that removes undetectable genes in the MPP3 population reveals a short list of 10 genes (*Igll1*, *6330403A02Rik*, *Vpreb1*, *Cecr2*, *Vpreb3*, *Ebf1*, *Gfra1*, *Bst1*, *Klrk1*, *Akap12*) that are specifically expressed in CLP.

To test further the robustness of the RNAseq results, we selected a set of six genes (*CD7*, *HES1*, *BLNK*, *EPHA2*, *VPREB1*, and *F2RL3*) to perform the RT-PCR validation. To this end, all three progenitors were sorted at a concentration of 10 cells/well. For each population, eight of these wells were independently amplified by RT-PCR. The results of the PCR amplification are shown in Figure 3A. The *y*-axis corresponds to the percent of positive wells (0.1 correspond to 10% whereas 1 correspond to 100% of positive wells). Therefore, the temporal expression profiles obtained for these genes are in accordance with their RNAseq quantification (Figure 3A, Table S7). Moreover, specific transcripts encoding well-known markers were also preferentially expressed in the expected population, showing further the robustness of the RNAseq results (Figure S3). In particular, *Flt3*, *Notch1*, *Ccr9*, *Cxcr4*, *Il7ralpha*, *Ets1*, and *Gata3* were over-expressed in CLP cells compared to MPPs, and *c-Kit* was over-expressed in MPPs compared to CLP, consistent with its downregulation in lymphoid cells [34].

Altogether, the data demonstrate that the expression of lymphoid genes is initiated in MPP2 and remains active along the lymphoid differentiation ending up in a fully lymphoid CLP population expressing a new set of specific genes.

### 2.3. Exploratory Analysis of MPP-to-CLP-Lineage Specific Surface Markers

To further identify specific cell surface markers (SM) that ultimately might improve the purification of progenitor cells, we performed hierarchical clustering on the temporal gene expression profiles of the 677 SMs differentially expressed in the three progenitors (MPP2-MPP3-CLP). Genes were separated into two groups depending on whether their expression values increase or decrease from MPP2 to CLP. Both categories included comparable numbers of deregulated genes (341 upregulated versus 336 downregulated) (Table S5).



**Figure 3.** Exploration of surface markers and validation of CD93. (A) Line plots visualizing the temporal gene expression profiles generated from RNaseq (blue line) and RT-PCR (red line) experiments. The y-axis describes RNaseq data expressed as FPKM values, whereas RT-PCR corresponds to the



percent of positive wells (0.1 corresponds to 10% whereas 1 corresponds to 100% of positive wells). (B) Heatmap showing the abundance values of the surface markers (CD) differentially expressed between MPP2 and CLP cells. The bar plot (right panel) illustrates the Log<sub>2</sub> fold change per gene. (C) (Top panel) CD93 surface expression in BM progenitors (dotted lines represent the negative population). (Middle panel) In vitro differentiation of CD93<sup>+</sup> and CD93<sup>-</sup> CLP subsets compared with LSK. Sorted CD93<sup>+</sup> and CD93<sup>-</sup> CLP were seeded on OP9 (B cell competent) stroma. (Bottom panel) In vitro differentiation of CD93<sup>+</sup> and CD93<sup>-</sup> CLP subsets compared with LSK. Sorted CD93<sup>+</sup> and CD93<sup>-</sup> CLP were seeded on OP9-DL4 (T cell competent) stroma. CD93<sup>-</sup> CLP are able to generate B and T cells; in contrast, CD93<sup>+</sup> CLP generate only B cells. Dot plots show the cells 21 days after co-culture.

Among the SMs over-expressed in CLP and almost undetectable in MPP2, we found B cell specific such as *Vpreb1*, *Bfsp2*, *Gfra2*; T/NK cell specific such as *Klrk1*, *Zap70*, *Icos*, *Cd28*, *Tcf7*; and pan-lymphoid such as *Slamf6* (Figure 2E).

Notably, the expression of CD34, *c-Kit* and *Csf3r*, the granulocyte colony stimulating factor receptor are downregulated in MPP2 suggesting that one important aspect of lymphoid engagement is the downregulation of receptors that are important for myeloid lineages [35].

We then restricted our analysis to the 87 SMs corresponding to known clusters of differentiation (CD) (<https://www.genenames.org/data/genegroup/#!/group/471> (accessed on 4 March 2019) (Figure 3B). From these, FACS analysis pinpointed CD93 (also known as AA4.1) as a good candidate. Consistent with the RNAseq results, this marker was not significantly detected in the MPP populations (MPP2 = 0.9% and MPP3 = 1.6%) whereas the expression in CLP was high (about 72%) (Figure 3C top panel).

To get a better insight into the role of CD93, the CLP population was split into two subsets according to the CD93 expression (CD93<sup>+</sup> and CD93<sup>-</sup>) and their capacity to generate B or T cells was assessed on OP9 and OP9-DL4 stroma, respectively. The sorted CD93<sup>+</sup> and CD93<sup>-</sup> CLP were seeded and cultured for 21 days. After this period, FACS analysis showed that both subsets do generate the full B lineage (assessed with anti-B220) whereas only CD93<sup>-</sup> cells generate T cells that develop into CD4/CD8 populations in OP9-DL4 stroma (Figure 3C middle and bottom panels). Altogether, the results show that the CLP is composed of two different populations based on the expression of CD93 and experimentally confirm the specific expression of CD93 in CLP compared to MPP2 and MPP3, showing that it can be used as an additional marker to improve the purification of these populations.

#### 2.4. Identification of Biological Pathways through Lymphoid Commitment

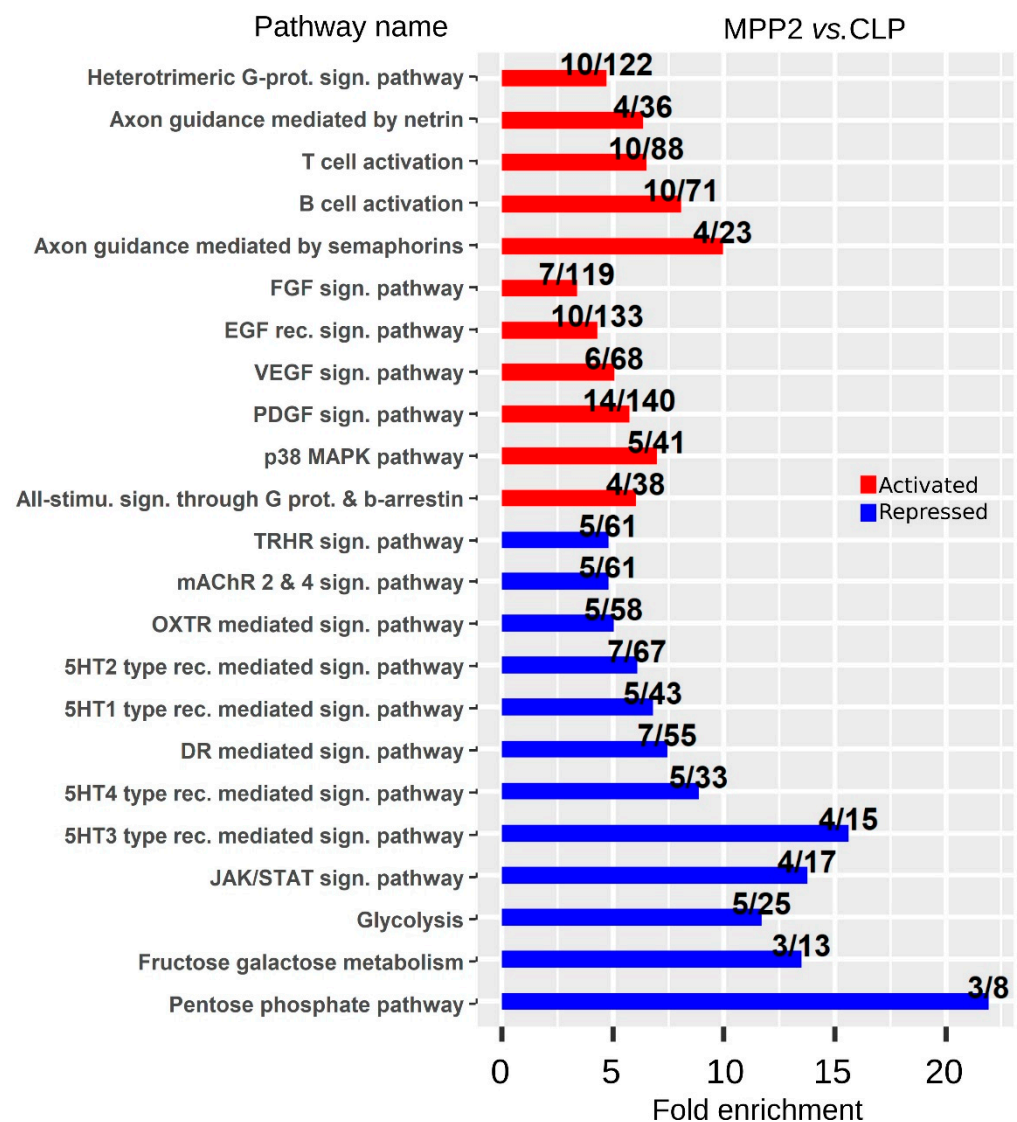
We continued to explore the DEGs by searching for biological pathways activated or repressed during lymphoid commitment in order to identify the key pathways that trigger the commitment to CLP.

Biological pathways activated or repressed in CLPs, identified by ontological enrichment analysis performed using the PantherDB tool (version 14.1 Released 2019-03-12), are represented in Figure 4 and Table S6. Top activated categories include genes involved in the establishment of the immune system, such as B and T cell activation. In addition, we found the activation of EGF receptor, FGF, VEGF, PDGF, and p38 MAPK signaling pathways.

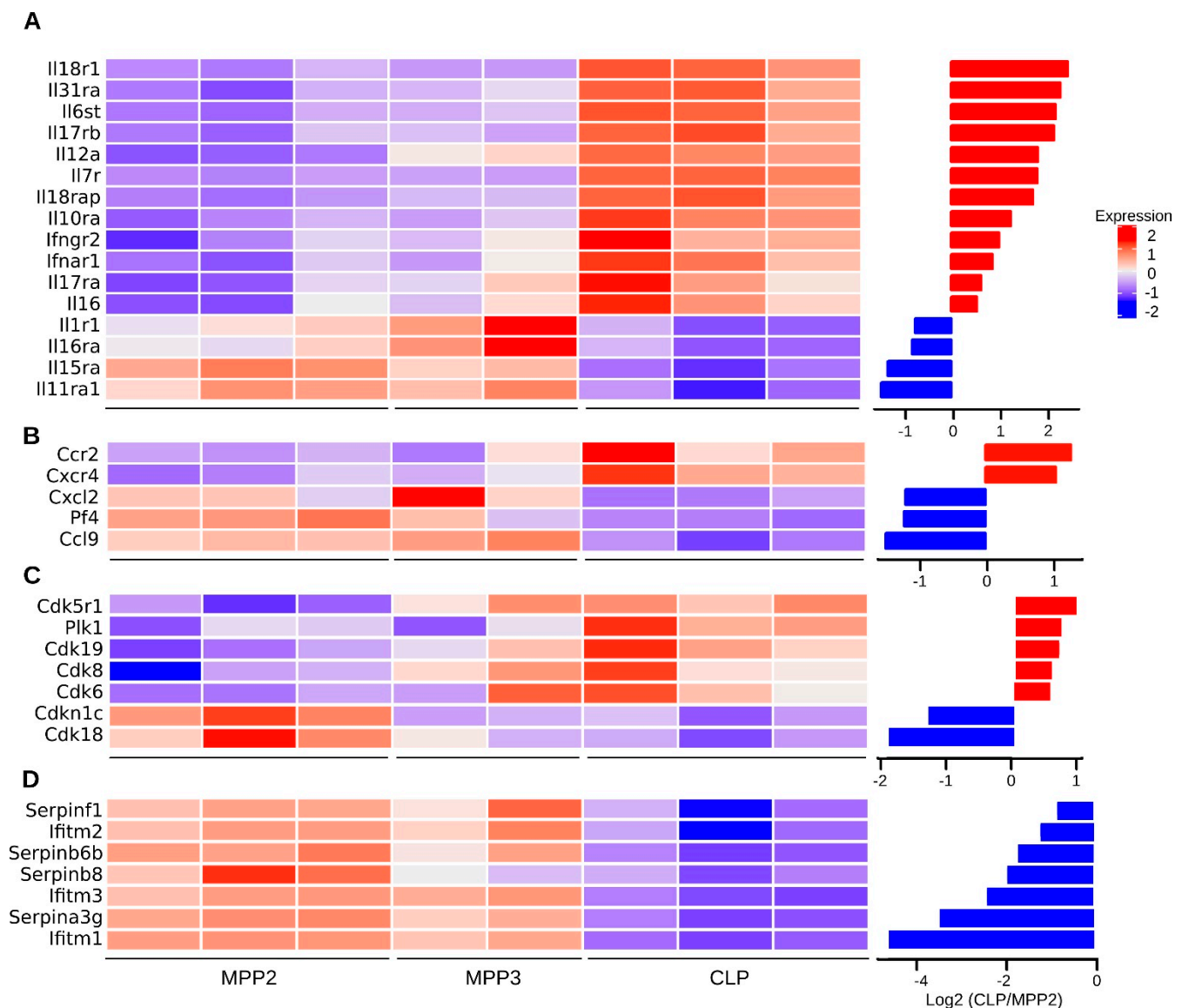
In contrast, the major biological pathways repressed in CLPs were involved in metabolism (pentose phosphate, fructose, glycolysis). Notably, as glycolysis is a key player in the maintenance of stemness through provision of energy [36], its repression is logically associated with the loss of stemness property during the differentiation of MPP cells into CLP.

Some biological pathways such as those associated with inflammation (chemokine, cytokine, and interleukin signaling pathways) or cell differentiation (integrin, endothelin, and insulin-like growth factor-1 signaling pathways) were enriched in both, up- and down-regulated genes, indicating the complex regulation of these pathways during progenitor differentiation (Figure S6). Notably, comparison of our transcriptome analysis with two publicly available datasets (Cabezas-Wallscheid et al., 2014 [37]; Klimmeck et al., 2014 [30])

reveal that the most enriched pathways were associated with cytokine regulation, cell cycle regulators, and stress-associated genes (Tables S14 and S16). Among cytokines, it was interesting to highlight IL18 and *Il18r1*, the cytokine receptor associated with the strongest over-expression (Figure 5A). *Il18r1* is expressed in a variety of immune cells including thymic ETP, NK cells, macrophages, neutrophils, B cells, and fully differentiated Th1. Furthermore, it was also detected in B cell disorders [38]. Indeed, IL-18 can positively impact BM lymphopoiesis and T cell development, presumably via interaction with the c-Kit and IL-7 signaling axis. Furthermore, IL-18 can substitute for IL-7 in early thymic development processes and can synergize with IL-7 in promoting proliferation [39]. Two chemokines, *Ccr2* and *Cxcr4*, over-expressed in CLP were also highlighted (Figure 5B), reflecting trafficking and de-attachment of newly generated progenitors from the stroma to reach the circulation [40–42].



**Figure 4.** Biological pathways associated with lymphoid commitment. Bar plots illustrating pathway enrichment analysis from surface markers and transcription factors upregulated or downregulated between MPP2 and CLP cells. The pathways identified as specifically enriched from the group of genes upregulated between MPP2 and CLP were shown in red, those from downregulated genes were shown in blue. For each pathway, the bar plots illustrate the fold enrichment and the ratio describes the number of activated/repressed genes in the pathway compared to the total number of genes annotated in the pathway.



**Figure 5.** Expression analysis of functional group of genes. Heatmaps showing the abundance values of genes differentially expressed between MPP2 and CLP cells and associated with (A) interleukins and their receptors, (B) chemokines, (C) cell cycle proteins, (D) cellular integrity and stress. The bar plots (right panels) illustrate the Log<sub>2</sub> fold change per gene.

In addition, consistent with an active cell cycle status, the quiescence-enforcer Cdkn1c (p57, a regulator of HSC quiescence) is found mostly expressed in MPP2 compared to CLP (Figure 5C) [43]. Conversely, CLPs clearly showed robust expression levels of the cell-cycle driver *Cdk6*, *Cdk8*, *Cdk19*, *Cdk5r1* compared to MPPs (which express *Cdk18*, closely linked to actin [44]).

Concerning the mechanisms protecting cellular integrity, we found a novel set of Serpins (*Serpib8*, *Serpib6b*, *Serpib9*, *Serpinf1*, *Serpina3g*) over-expressed in MPPs compared to CLP (Figure 5D).

### 2.5. Prediction of a Gene Regulatory Network Related to Lymphoid Differentiation

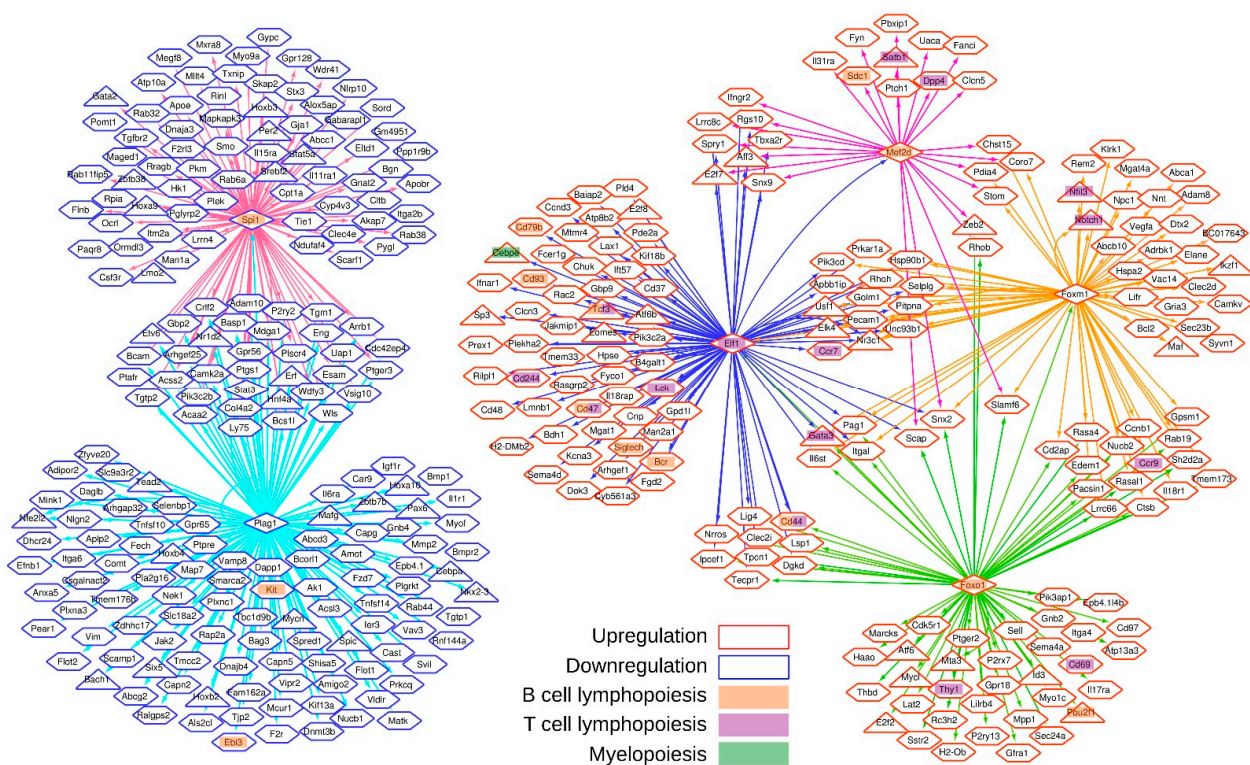
To identify the regulatory networks controlling lymphoid differentiation, we first analyzed the expression profiles of TFs across the MPP2-CLP transition.

We detected 28 TFs over-expressed in CLP ( $\text{Log}_2(\text{CLP}/\text{MPP2}) > 1$ ) (Figure S5), which include TF specific for the T and NK lineages such as *Gata3*, *Hes1*, *Tcf7*, *Eomes*; TFs specific

for DC and B cells such as *Ebf1*, *Foxo1*, *Klf4*, and *Irf8*; and TF such as *Ets1*, *Tcf3*, *Ikzf1-Ikaros*, *Pou2f1/Oct1*, *Sp3* that act on several lymphoid lineages. This confirms the multilineage developmental potential of CLP [8].

Regarding the 25 downregulated TFs in CLP ( $\text{Log}_2(\text{CLP}/\text{MPP2}) < -1$ ) (Figure S5), we notably identified genes specific to non-lymphoid lineages, such as those involved in erythroid and myeloid/megakaryocytes development (*Gata2*, *Hoxa10*, *Hlf*, *Nfe2*, *Nfix*) and genes important for stem cell differentiation (*Lmo2*, *Tal1*) (Figure S5).

To obtain a more dynamic view of the differentiation, we inferred networks of co-regulated genes that have common TFBS. These were searched in promoter regions of the 758 co-expressed SMs and TFs. For this motif enrichment analysis, we used the AME method from MEME suite and the collection of TFBS position frequency matrices referenced in the HOCOMOCO mouse database [45,46]. TFBS motifs for SPI1, PLAG1, ELF1, FOXO1, FOXM1, MEF2D, and STAT3 were predicted as significantly enriched in promoter regions of co-expressed genes (Table S4). Since SPI1 and PLAG1 are downregulated TFs, we focus on their under-expressed target genes that are associated with the repression of gene expression programs. Consistent with previous literature reports, in this network we found *c-Kit* and *Ebi3*. Notably, 35 genes are regulated simultaneously by both TFs (Figure 6). The other genes included in this network are newly described. The network of upregulated genes controlled by *Elf1*, *Foxo1*, *Foxm1*, and *Mef2d* is involved in the positive regulation. In this case, such factors control the expression of multiple essential regulators involved in T and B cell biology.



**Figure 6.** Gene regulatory network governing surface marker expression during lymphoid differentiation. Gene regulatory network inferred using gene co-expression, TF motif enrichment, and scanning algorithms. The node shapes represent transcription factors (triangles) or surface markers (hexagons). The node border colors annotate genes that are upregulated (red) or downregulated (blue) during lymphoid differentiation. The node background colors illustrate the genes known to be involved in B cell lymphopoiesis (orange), in T cell lymphopoiesis (purple), or in myelopoiesis (green). The edges indicate predicted regulatory interactions between core TFs and target genes. An arbitrary color was assigned to outgoing edges connected to the same TF in order to facilitate network visualization.

In conclusion, six main TFs orchestrate the network of genes whose expression varies during commitment of MPP2 to CLP (Figure 6). The robustness of these results is supported by an extensive literature analysis showing that 61 TF–target gene associations that correspond to 12% of the total network connections identified here are already published in other models (Table S8). Altogether, the integrated analysis of co-expression and regulatory binding site prediction allowed us to identify a broad number of co-regulated genes associated with lymphoid differentiation, which highlights the complexity and fine regulation of the process.

#### 2.6. Annotated and Novel Long Non-Coding RNAs Associated with Lymphoid Progenitor Differentiation

To further characterize molecular events involved in the differentiation of the lymphoid lineage, we explored the temporal expression of lncRNAs.

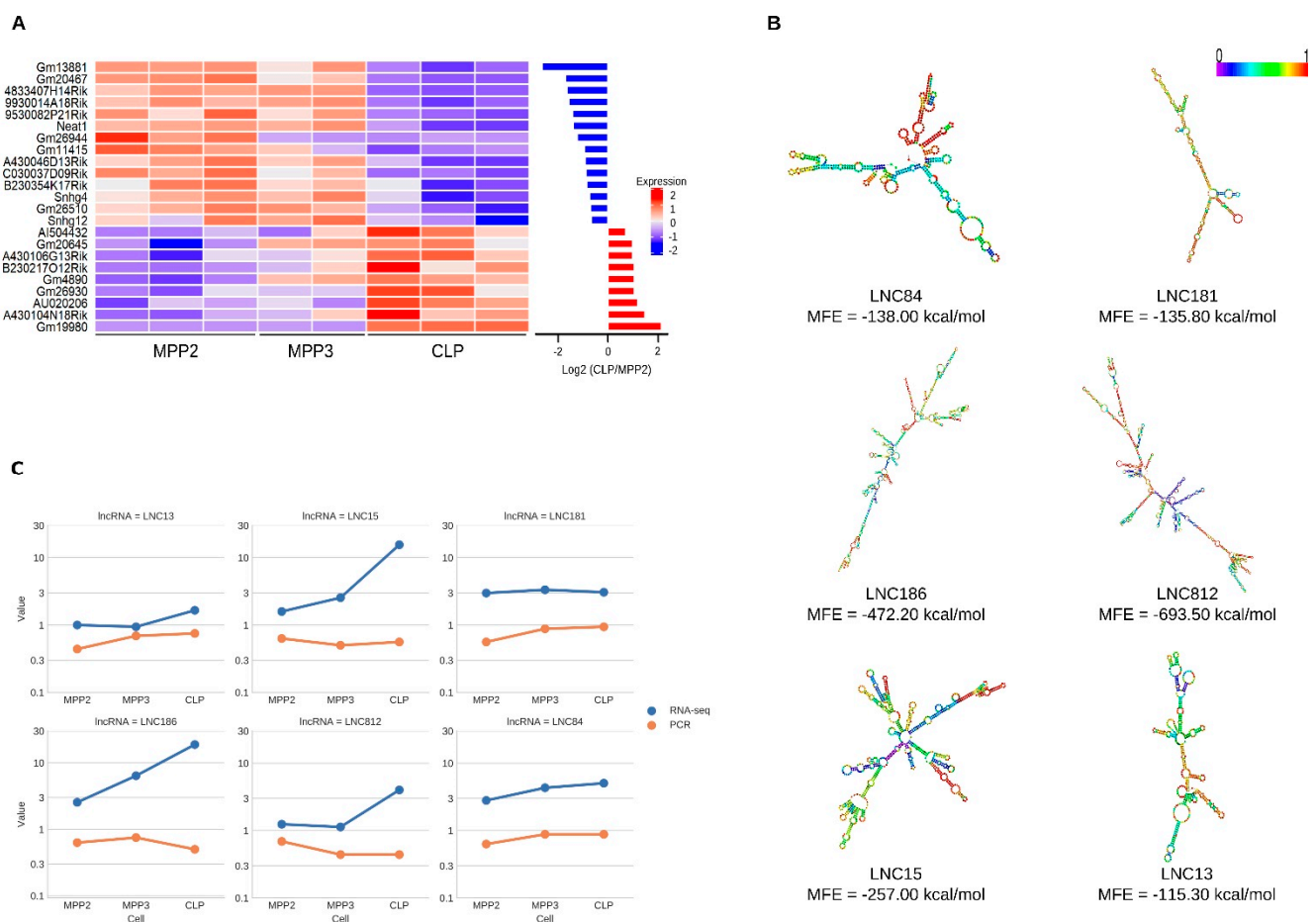
We first focused on lincRNA with known genomic annotation. From the total RNA sequencing data generated, we detected 353 lincRNAs expressed in the three progenitors MPP2, MPP3, and CLP (Table S2), from which 23 were differentially expressed between MPP2 and CLP cells (Figure 7A and Table S10). Nine of them were also differentially expressed by both MPP2 vs. CLP and MPP3 vs. CLP comparisons, but not between MPP2 and MPP3 (Table S9). Again, this shows a certain MPP2–MPP3 stability followed by significant expression changes during the transition of MPP3 to CLP. Their precise function in the emergence of the CLP merits further investigation. Interestingly, *Snhg7* is the only lincRNA that has a different expression profile: its expression first increases from MPP2 to MPP3 before decreasing significantly at the CLP stage.

Another set of 14 differentially expressed lincRNAs were only found in the transition from MPP2 to CLP, but not from MPP3 to CLP (Figure 7A and Table S10), suggesting a role in other hematopoietic lineages.

To go beyond the analysis of annotated lincRNAs, we searched for novel lncRNAs. To this end, we carried out genome-guided transcript assembly using Cufflinks tool and selected assembled transcripts which did not overlap annotated mouse genes on the same strand (GRCM38 Ensembl release 75). Given that more than 96% of novel assembled transcripts were mono-exonic, we decided to focus our analysis on this category. We assessed the coding potential of each novel transcript using the CPAT tool [47] and kept putative novel long non-coding RNA greater than 200 bp in length. We then extracted raw read counts for these transcripts using HTSeq-count tool and identified differentially expressed transcripts with DESeq2 software. A short list of 45 novel lncRNAs predicted by this methodology were equally distributed between intergenic and intronic regions (Table S11).

To obtain an estimate of the biological pertinence of these novel lncRNAs, we examined their predicted secondary structure. The calculation and visualization of the MFE (minimum free energy) structures were obtained using the RNAfold tool. The results indicated that these novel lncRNAs could be modeled with stable secondary structures (MFE < 80 kcal/mol) (Figure 7B), consistent with a potential biological function [48].

Finally, to assess the relevance of our predictions, we carried out the experimental validation of six novel lncRNAs (three in intergenic and three in intronic regions) by RT-PCR (Figures 7C and S7 and Table S12). The results show that these novel lncRNAs are indeed expressed in these progenitors and validate, for two of them (LNC13 and LNC84), their upregulation during the differentiation towards the CLP.



**Figure 7.** Annotated and novel long non-coding RNAs associated with lymphoid progenitor differentiation. **(A)** Heatmap showing the abundance values of the annotated long non-coding RNAs (lncRNA) differentially expressed between MPP2 and CLP cells. The bar plot (right panel) illustrates the Log<sub>2</sub> fold change per gene. **(B)** Visualization of the predicted secondary structures of novel lncRNAs. The stability of lncRNA structures is calculated using MFE (minimum free energy). **(C)** Line plots visualizing the temporal gene expression profiles generated from RNAseq (blue line) and RT-PCR (red line) experiments. The y-axis describes RNAseq data expressed as FPKM values, whereas RT-PCR correspond to the percent of positive wells (0.1 corresponds to 10% whereas 1 corresponds to 100% of positive wells).

### 3. Discussion

A major step towards gaining a mechanistic understanding of cellular decision-making processes is to determine what the phenotypes of cells are, at different transition stages, and to establish cell type-specific regulatory networks that control the spatiotemporal expression of lineage-specific genes.

In this study, we tracked lymphoid commitment through a comprehensive RNAseq analysis on three purified progenitor populations: MPP2-MPP3-CLP, to highlight transcriptional modifications occurring through the gradual loss of multilineage potential to reach the fully committed lymphoid progenitor.

The RNAseq data reveal that the transcriptional program is rather stable between MPP2 and MPP3 cells followed by a strong shift in expression at the MPP3-CLP transition. Notably, while lymphoid gene expression is barely detectable in HSC [49], our analysis reveals that MPP2 cells express pioneer lymphoid lineage restricted genes such as *Rag1*, *Mpeg1*, and *Dntt* which remain active along the lymphoid differentiation towards the CLP population. These results are consistent with previously published expression data showing that the abundance of these three genes is barely detectable in the HSC and MPP1

stages and then increases from MPP2 [37]. *Rag1* and *Dnnt* are recombinases and transferases, respectively, involved in the B and T-cell receptor recombination [50]. *Mpeg1* is a myeloid gene whose expression is detected on zebra fish B cells [51]. However, for a complete lymphoid priming the upregulation of key lymphoid TF such as *Ebf1*, *Tcf7*, *Spib*, *Hes1*, *Eomes*, and genes associated with both B cell (*VpreB1/3*, *CD93*) and T cell receptor signaling (*Zap70*) is mandatory.

The transcriptome analysis reveals a pluri-lymphoid (T, B, DC, NK) signature of CLP, even though the main fate of this progenitor is to generate the B cell lineage, and sequentially, the other lineages [4]. Our analysis is consistent with the plasticity of such progenitors that have diverse open options immediately available. The final decision might be undeniably dependent on the microenvironment [52]. For example, CLP in the BM generate B cell progenitors, however, injected intra thymically, they produce mainly T cells [4]. Collectively, the data presented here illustrate coexistence of multiple competing lineage-specific transcriptional programs and reveal novel signaling programs not previously identified as operating during lymphoid commitment.

We also identified genes, such as *CD93*, that are highly expressed in CLP cells compared to MPP2 and MPP3. *CD93* is a cell-surface glycoprotein and type I membrane protein that is involved in intercellular adhesion and in the clearance of apoptotic cells [53]. It is required for maintenance of antibody secretion and persistence of plasma cells in the bone marrow niche [54]. The differential expression of *CD93* suggests that it can be used as additional marker to improve the purification of progenitors. Notably, we observed differences within the CLP population: the *CD93<sup>+</sup>* and *CD93<sup>-</sup>* cells generate the full B lineage whereas only *CD93<sup>-</sup>* cells generate T cells that develop into *CD4/CD8* populations in OP9-DL4 stroma.

In this study, bioinformatic analyses enable to identify networks of interacting TFs which in the case of multilineage differentiation may give rise to multiple distinct outcomes. In particular, we found that TFs such as *Spi1*, *Plag1*, *Elf1*, *Foxo1*, *Foxm1*, and *Mef2d* regulate sets of co-expressed related genes most of them reported for the first time. The fact that very well known transcription factors, such as *Gata3*, *Notch1*, or *c-kit* are well classified in each category adds more validity to the work. More importantly, we identified a broad number of co-regulated genes associated with lymphoid differentiation, which might help unravel new aspects of this complex process.

In addition, computational screening identified 353 lincRNAs expressed in lymphoid progenitors with known genomic annotation, of which 24 were DEGs. The level of expression of nine of those was similar in MPP2 and MPP3 cells but highly modulated during the MPP3-CLP transition (Figure 7) suggesting a role of these lincRNAs in the establishment of the CLP population. A set of 14 differentially expressed lincRNA (Figure 7A) were modulated at the MPP2-CLP transition, but not through the MPP3-CLP transition, suggesting that they might also be implicated in the differentiation of other hematopoietic lineages.

Even though the role of many of the identified lincRNAs is not yet known and requires further systematic investigations, most of them are involved in cancer progression or used as markers for cancer detection, emphasizing the importance of these molecules in the homeostatic balance. In particular, several members of the small nucleolar RNA host gene (SNHG) family are oncogenes [55] that play a crucial role in leukemogenesis and have prognostic value in acute myeloid leukemia [26]. In addition, high expression of SNHG7 promotes proliferation, metastasis, predicts poor prognosis and poor survival for patients.

Consistent with our results, *Gm20645* and *Neat1* were found up- and downregulated, respectively, in the course of differentiation from HSC to MPP4 [37].

Furthermore, we identified 45 novel lincRNAs equally distributed between intergenic and intronic regions. The high expression of six of them in lymphoid progenitors as well as their predicted stable secondary structure, correlate with a potential function. Their identification constitutes a great breakthrough resulting from this work that would certainly require further deep investigation to measure specifically their biological role.

Altogether, our publicly available data represent a comprehensive resource for the lymphoid lineage commitment field that definitely will improve our understanding of the transcriptional organization underlying lineage specification.

## 4. Materials and Methods

### 4.1. Mice

C57Bl/6 Ly5.1 Thy1.2 (B6) mice were used at 10–18 weeks old (both males and females). They were purchased from CDTA (Orléans, France) and were kept in a specific pathogen-free facility of SFR Necker (Agreement No. 75-1026). All experimental procedures using animals were approved by the “Comité d'éthique en experimentation animale” of the Paris Descartes University and the French Ministry of Research, Innovation and Education under the following reference APAFIS #26599-2019071812345604.

### 4.2. Antibodies

For cell sorting, antibodies were obtained from BD Biosciences (San Diego, CA, USA), eBioscience (San Diego, CA, USA), or SONY (Weybridge, UK): anti-NK1.1 (PK136), anti-TCRb (H57-597), anti-CD3 (145-2-C11), anti-Mac 1/CD11b (M1/70), anti-CD19 (1D3), anti-GR 1 (RB6-8C5), anti-Ter 119 (Ly76), anti-Flt 3/CD135 (Flk2; A2F 10.1), anti-c-KIT/CD117 (SCF receptor; 2B8), anti-Sca 1 (E13-161.7), anti-IL7Ralpha/CD127 (A7R34), anti-VCAM1/CD106 (429), anti-CD138/Sdc1 (MI15), anti-CD93 (AA4.1). They were directly coupled to FITC, APC, APC-eF780, PE, PerCP-Cy5.5, BV421 or conjugated with biotin, the latter being revealed by streptavidin-PE-Cy7. All staining was performed for 20 min, at 4 °C with agitation.

### 4.3. Isolation of Bone Marrow Progenitors

BM cells from six mice were collected by crushing femurs, tibia, hips, and spine bones and pooled. The cells were incubated with unconjugated rat antibodies against B220, Ter119, Gr1, and Mac1. Positive cells were magnetically depleted with sheep anti-rat IgG-conjugated beads and sheep anti-mouse IgG-conjugated beads (Dynabeads M-450; Dynal Thermos Fisher Scientific, Villebon, France). The remaining cells (Lin<sup>-</sup>, enriched) were labeled with fluorescent conjugated antibodies against c-Kit, Sca-1, and Lin antigens (NK1.1, TCRβ, CD3, Mac1, CD19, GR1, and Ter119). Lin<sup>-</sup>Sca-1<sup>+</sup>c-Kit<sup>+</sup> cells are called LSK. Anti-VCAM1 and anti-Flt3 were additionally used to isolate the MPP2 (LSK-Flt3<sup>+</sup>VCAM1<sup>+</sup>) and the MPP3 (LSK-Flt3<sup>+</sup>VCAM1<sup>-</sup>) subsets. The CLP population is characterized by the addition of anti-IL7Ralpha and their phenotype is Lin<sup>-</sup>Sca-1<sup>lo</sup>c-Kit<sup>lo</sup>IL7Ralpha<sup>+</sup> (Figure 1B). Cell sorting was performed on a FACS Aria III upgraded with DIVA software (BD Biosciences, San Diego, CA USA). Three independent biological replicates of MPP2, MPP3, and CLP were considered for transcriptome analysis. Pelleted cells were stored at -80 °C in QIAzol lysis reagent until RNA extraction.

### 4.4. Total RNA Isolation and Sequencing

Total RNA was extracted using miRNeasy Micro Kit (QIAGEN, Les Uvils, France) with DNase digestion. NanoDrop spectrophotometer was used to determine total RNA quantity and the 2100 Bioanalyser (Agilent) to qualify total RNA and subsequent libraries. Nuclear and mitochondrial ribosomal RNA was depleted using Ribo-Zero Gold rRNA Removal Kit (Epicentre). Whole transcriptome strand-specific libraries from three progenitors (MPP2, MPP3, and CLP) were performed using SMARTer Stranded RNaseq Kit (Clontech) with a total input of (0.11–0.33) µg of total RNA per cell population, knowing that rRNA can reach 90% of transcriptome according to cells. This library preparation allows the sequencing of RNAs coding for proteins and long non-coding RNAs (>~200 bp) polyA<sup>+</sup> and polyA<sup>-</sup> (lncRNA, miRNA precursors, snRNA, snoRNA, etc.). Three biological replicates, per cell population, were sequenced on Illumina HiSeq2000 as 100 bp pair-end reads at high depth of coverage (2 × 120 million reads per replicate).



#### 4.5. RT-PCR Analysis

Ten cells were collected in individual PCR tubes containing 5  $\mu$ L PBS 1 $\times$  following cell sorting on a FACSAria III U equipped with an automatic cell deposition unit (ACDU, BD Biosciences, San Diego, USA). Cells were lysed by freezing at  $-80$   $^{\circ}$ C, followed by heating to 65  $^{\circ}$ C for 2 min.

After cooling at 4  $^{\circ}$ C, RNA was specifically reverse transcribed by adding 10  $\mu$ L of a mix containing 0.13  $\mu$ M RT primer, 1 $\times$  PCR Buffer (Thermo), 3.3 mM MgCl<sub>2</sub> (Thermo), 1 mM dNTPs (GE Healthcare), 40 units of RNaseOUT (Thermo), and 35 units of MuLV Reverse Transcriptase (Thermo), in a 15  $\mu$ L reaction volume for 1 h at 37  $^{\circ}$ C and then incubated for 3 min at 95  $^{\circ}$ C for inactivation of reverse transcriptase. cDNA generated were diluted with 30  $\mu$ L of 1 $\times$  PCR Buffer.

A 4  $\mu$ L aliquot of diluted cDNA was then amplified: first, a denaturation step at 95  $^{\circ}$ C for 10 min and then 54 amplification cycles (30 s at 95  $^{\circ}$ C, 45 s at 70  $^{\circ}$ C, and 60 s at 72  $^{\circ}$ C, with the hybridization temperature decreased from 70 to 60  $^{\circ}$ C during the six first cycles). The PCR mix contained 0.25  $\mu$ M of each primer (Reverse and Forward), 1 $\times$  PCR Buffer (Thermo), 2 mM MgCl<sub>2</sub> (Thermo), 0.25 mM dNTPs (Thermo), and 0.5 units of AmpliTaq Gold DNA Polymerase (Thermo), in 20  $\mu$ L reaction volume. PCR products were detected on a 1.5% agarose ethidium bromide gel. Primer sequences are described in Table S7.

#### 4.6. Culture of Lymphoid Precursors

After sorting, 1100 cells (CLP CD93<sup>+</sup> or CLP CD93<sup>-</sup>) were seeded on a 24-well plate coated with 6000 OP9 or OP9-DL4 cells per well the day before. The co-culture was completed in Opti-MEM medium (Thermo) supplemented with 10% of FBS (Thermo), 100 U/mL of Penicillin-Streptomycin (Thermo), 50  $\mu$ M  $\beta$ 2mercaptoethanol, 1 ng/mL IL7 (Immunotools, Friesoythe, Germany), 5 ng/mL Flt3-L. Acquisition was performed 12 days later with a FACSCanto II (BD Biosciences). All cytometry analyses were performed with FlowJo™ Software V10.2 (BD Biosciences).

#### 4.7. Comparison with Public Expression Data

Transcriptome data showing genes that increase in expression from MPP2 to CLP cells was intersected, with those of Cabezas et al. (2014) [37], whose expression increased in MPP4 cells compared to all the other populations (HSC, MPP1, MPP2, and MPP3) (Table S13) and those of Klimmeck et al. (2014) [30] whose expression increased in LS + K cells compared to LS – K cells (Table S15). These gene lists were explored by the PantherDB tool (Tables S14 and S16).

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms23031115/s1>.

**Author Contributions:** C.B., D.T.-L.R. and S.E. designed and coordinated the study. A.L., S.E. and V.M. prepared all biological samples, performed the culture, the cytometry analysis and PCR. M.-A.P. optimized and performed the RNAseq library preparation. A.B., R.O. and J.-F.D. coordinated the RNAseq sequencing. C.B. coordinated the bioinformatic analysis. S.C. performed the bioinformatic analysis. C.B., D.T.-L.R., S.C. and S.E. performed the data interpretation. C.B., D.T.-L.R., S.C. and S.E. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** We gratefully acknowledge funding from the Fondation de la Recherche Médicale (FRM) Bioinformatics Analysis for Biology Research program (DBI20131228570), the Inserm, the CEA, the CNRS, the University Paris Descartes and the Secular Society (TSS).

**Data Availability Statement:** The RNAseq data sets from MPP2-MPP3-CLP cells can be accessed through ArrayExpress (<https://www.ebi.ac.uk/arrayexpress> (accessed on 13 December 2021); accession number E-MTAB-11217).

**Acknowledgments:** We thank the CNRGH (CEA) sequencing facilities and Nicolas Wiart for his support on the use of the high-performance computing clusters at CNRGH. We gratefully acknowledge the experimental assistance provided by Lionel Mockel and Julie Renard, SHRI-CEA.

**Conflicts of Interest:** The authors declare that there is no conflict of interest.

## References

1. Yilmaz, Ö.H.; Valdez, R.; Theisen, B.K.; Guo, W.; Ferguson, D.O.; Wu, H.; Morrison, S.J. Pten Dependence Distinguishes Haematopoietic Stem Cells from Leukaemia-Initiating Cells. *Nature* **2006**, *441*, 475–482. [[CrossRef](#)]
2. Lai, A.Y.; Kondo, M. Asymmetrical Lymphoid and Myeloid Lineage Commitment in Multipotent Hematopoietic Progenitors. *J. Exp. Med.* **2006**, *203*, 1867–1873. [[CrossRef](#)] [[PubMed](#)]
3. Kondo, M.; Weissman, I.L.; Akashi, K. Identification of Clonogenic Common Lymphoid Progenitors in Mouse Bone Marrow. *Cell* **1997**, *91*, 661–672. [[CrossRef](#)]
4. Karsunky, H.; Inlay, M.A.; Serwold, T.; Bhattacharya, D.; Weissman, I.L. Flk2+ Common Lymphoid Progenitors Possess Equivalent Differentiation Potential for the B and T Lineages. *Blood* **2008**, *111*, 5562–5570. [[CrossRef](#)]
5. Chambers, S.M.; Shaw, C.A.; Gatza, C.; Fisk, C.J.; Donehower, L.A.; Goodell, M.A. Aging Hematopoietic Stem Cells Decline in Function and Exhibit Epigenetic Dysregulation. *PLoS Biol.* **2007**, *5*, e201. [[CrossRef](#)]
6. Forsberg, E.C.; Passegué, E.; Prohaska, S.S.; Wagers, A.J.; Koeva, M.; Stuart, J.M.; Weissman, I.L. Molecular Signatures of Quiescent, Mobilized and Leukemia-Initiating Hematopoietic Stem Cells. *PLoS ONE* **2010**, *5*, e8785. [[CrossRef](#)] [[PubMed](#)]
7. Gazit, R.; Garrison, B.S.; Rao, T.N.; Shay, T.; Costello, J.; Ericson, J.; Kim, F.; Collins, J.J.; Regev, A.; Wagers, A.J.; et al. Transcriptome Analysis Identifies Regulators of Hematopoietic Stem and Progenitor Cells. *Stem Cell Rep.* **2013**, *1*, 266–280. [[CrossRef](#)]
8. Alberti-Servera, L.; von Muenchow, L.; Tsapogas, P.; Capoferri, G.; Eschbach, K.; Beisel, C.; Ceredig, R.; Ivanek, R.; Rolink, A. Single-Cell RNA Sequencing Reveals Developmental Heterogeneity among Early Lymphoid Progenitors. *EMBO J.* **2017**, *36*, 3619–3633. [[CrossRef](#)]
9. Amann-Zalcenstein, D.; Tian, L.; Schreuder, J.; Tomei, S.; Lin, D.S.; Fairfax, K.A.; Bolden, J.E.; McKenzie, M.D.; Jarratt, A.; Hilton, A.; et al. A New Lymphoid-Primed Progenitor Marked by Dach1 Downregulation Identified with Single Cell Multi-Omics. *Nat. Immunol.* **2020**, *21*, 1574–1584. [[CrossRef](#)]
10. Grover, A.; Sanjuan-Pla, A.; Thongjuea, S.; Carrelha, J.; Giustacchini, A.; Gambardella, A.; Macaulay, I.; Mancini, E.; Luis, T.C.; Mead, A.; et al. Single-Cell RNA Sequencing Reveals Molecular and Functional Platelet Bias of Aged Haematopoietic Stem Cells. *Nat. Commun.* **2016**, *7*, 11075. [[CrossRef](#)] [[PubMed](#)]
11. Moignard, V.; Göttgens, B. Dissecting Stem Cell Differentiation Using Single Cell Expression Profiling. *Curr. Opin. Cell Biol.* **2016**, *43*, 78–86. [[CrossRef](#)] [[PubMed](#)]
12. Zhou, W.; Yui, M.A.; Williams, B.A.; Yun, J.; Wold, B.J.; Cai, L.; Rothenberg, E.V. Single-Cell Analysis Reveals Regulatory Gene Expression Dynamics Leading to Lineage Commitment in Early T Cell Development. *Cell Syst.* **2019**, *9*, 321–337.e9. [[CrossRef](#)]
13. Cheng, H.; Zheng, Z.; Cheng, T. New Paradigms on Hematopoietic Stem Cell Differentiation. *Protein Cell* **2020**, *11*, 34–44. [[CrossRef](#)]
14. Orkin, S.H.; Zon, L.I. Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. *Cell* **2008**, *132*, 631–644. [[CrossRef](#)] [[PubMed](#)]
15. Iwasaki, H.; Akashi, K. Myeloid Lineage Commitment from the Hematopoietic Stem Cell. *Immunity* **2007**, *26*, 726–740. [[CrossRef](#)]
16. Kim, S.-I.; Bresnick, E.H. Transcriptional Control of Erythropoiesis: Emerging Mechanisms and Principles. *Oncogene* **2007**, *26*, 6777–6794. [[CrossRef](#)] [[PubMed](#)]
17. Nutt, S.L.; Kee, B.L. The Transcriptional Regulation of B Cell Lineage Commitment. *Immunity* **2007**, *26*, 715–725. [[CrossRef](#)]
18. Rothenberg, E.V. Negotiation of the T Lineage Fate Decision by Transcription-Factor Interplay and Microenvironmental Signals. *Immunity* **2007**, *26*, 690–702. [[CrossRef](#)]
19. Kato, H.; Igarashi, K. To Be Red or White: Lineage Commitment and Maintenance of the Hematopoietic System by the “Inner Myeloid”. *Haematologica* **2019**, *104*, 1919–1927. [[CrossRef](#)] [[PubMed](#)]
20. De Obaldia, M.E.; Bhandoola, A. Transcriptional Regulation of Innate and Adaptive Lymphocyte Lineages. *Annu. Rev. Immunol.* **2015**, *33*, 607–642. [[CrossRef](#)]
21. Liu, L.; Zhang, Y.; Lu, J. The Roles of Long Noncoding RNAs in Breast Cancer Metastasis. *Cell Death Dis.* **2020**, *11*, 749. [[CrossRef](#)]
22. Slack, F.J.; Chinnaiyan, A.M. The Role of Non-Coding RNAs in Oncology. *Cell* **2019**, *179*, 1033–1055. [[CrossRef](#)] [[PubMed](#)]
23. Delás, M.J.; Jackson, B.T.; Kovacevic, T.; Vangelisti, S.; Munera Maravilla, E.; Wild, S.A.; Stork, E.M.; Erard, N.; Knott, S.R.V.; Hannon, G.J. LncRNA Spehd Regulates Hematopoietic Stem and Progenitor Cells and Is Required for Multilineage Differentiation. *Cell Rep.* **2019**, *27*, 719–729.e6. [[CrossRef](#)]
24. Luo, M.; Jeong, M.; Sun, D.; Park, H.J.; Rodriguez, B.A.T.; Xia, Z.; Yang, L.; Zhang, X.; Sheng, K.; Darlington, G.J.; et al. Long Non-Coding RNAs Control Hematopoietic Stem Cell Function. *Cell Stem Cell* **2015**, *16*, 426–438. [[CrossRef](#)]
25. Morceau, F.; Chateauvieux, S.; Gaigneaux, A.; Dicato, M.; Diederich, M. Long and Short Non-Coding RNAs as Regulators of Hematopoietic Differentiation. *Int. J. Mol. Sci.* **2013**, *14*, 14744–14770. [[CrossRef](#)]
26. Shi, J.; Ding, W.; Lu, H. Identification of Long Non-Coding RNA SNHG Family as Promising Prognostic Biomarkers in Acute Myeloid Leukemia. *OncoTargets Ther.* **2020**, *13*, 8441–8450. [[CrossRef](#)] [[PubMed](#)]
27. Wu, Z.; Gao, S.; Zhao, X.; Chen, J.; Keyvanfar, K.; Feng, X.; Kajigaya, S.; Young, N.S. Long Noncoding RNAs of Single Hematopoietic Stem and Progenitor Cells in Healthy and Dysplastic Human Bone Marrow. *Haematologica* **2019**, *104*, 894–906. [[CrossRef](#)]

28. Alvarez-Dominguez, J.R.; Hu, W.; Yuan, B.; Shi, J.; Park, S.S.; Gromatzky, A.A.; van Oudenaarden, A.; Lodish, H.F. Global Discovery of Erythroid Long Noncoding RNAs Reveals Novel Regulators of Red Cell Maturation. *Blood* **2014**, *123*, 570–581. [[CrossRef](#)]
29. Paralkar, V.R.; Mishra, T.; Luan, J.; Yao, Y.; Kossenkov, A.V.; Anderson, S.M.; Dunagin, M.; Pimkin, M.; Gore, M.; Sun, D.; et al. Lineage and Species-Specific Long Noncoding RNAs during Erythro-Megakaryocytic Development. *Blood* **2014**, *123*, 1927–1937. [[CrossRef](#)] [[PubMed](#)]
30. Klimmeck, D.; Cabezas-Wallscheid, N.; Reyes, A.; von Paleske, L.; Renders, S.; Hansson, J.; Krijgsveld, J.; Huber, W.; Trumpp, A. Transcriptome-Wide Profiling and Posttranscriptional Analysis of Hematopoietic Stem/Progenitor Cell Differentiation toward Myeloid Commitment. *Stem Cell Rep.* **2014**, *3*, 858–875. [[CrossRef](#)] [[PubMed](#)]
31. Johnson, W.E.; Li, C.; Rabinovic, A. Adjusting Batch Effects in Microarray Expression Data Using Empirical Bayes Methods. *Biostatistics* **2007**, *8*, 118–127. [[CrossRef](#)] [[PubMed](#)]
32. Wei, Q.; Frenette, P.S. Niches for Hematopoietic Stem Cells and Their Progeny. *Immunity* **2018**, *48*, 632–648. [[CrossRef](#)]
33. Hosokawa, H.; Romero-Wolf, M.; Yui, M.A.; Ungerback, J.; Quiloan, M.L.G.; Matsumoto, M.; Nakayama, K.I.; Tanaka, T.; Rothenberg, E.V. Bcl11b Sets Pro-T Cell Fate by Site-Specific Cofactor Recruitment and by Repressing Id2 and Zbtb16. *Nat. Immunol.* **2018**, *19*, 1427–1440. [[CrossRef](#)]
34. Katayama, N.; Shih, J.-P.; Nishikawa, S.; Kina, T.; Clark, S.C.; Ogawa, M. Stage-Specific Expression of c-Kit Protein by Murine Hematopoietic Progenitors. *Blood* **1993**, *82*, 2353–2360. [[CrossRef](#)] [[PubMed](#)]
35. Kondo, M.; Scherer, D.C.; Miyamoto, T.; King, A.G.; Akashi, K.; Sugamura, K.; Weissman, I.L. Cell-Fate Conversion of Lymphoid-Committed Progenitors by Instructive Actions of Cytokines. *Nature* **2000**, *407*, 383–386. [[CrossRef](#)]
36. Folmes, C.D.L.; Nelson, T.J.; Martinez-Fernandez, A.; Arrell, D.K.; Lindor, J.Z.; Dzeja, P.P.; Ikeda, Y.; Perez-Terzic, C.; Terzic, A. Somatic Oxidative Bioenergetics Transitions into Pluripotency-Dependent Glycolysis to Facilitate Nuclear Reprogramming. *Cell Metab.* **2011**, *14*, 264–271. [[CrossRef](#)]
37. Cabezas-Wallscheid, N.; Klimmeck, D.; Hansson, J.; Lipka, D.B.; Reyes, A.; Wang, Q.; Weichenhan, D.; Lier, A.; von Paleske, L.; Renders, S.; et al. Identification of Regulatory Networks in HSCs and Their Immediate Progeny via Integrated Proteome, Transcriptome, and DNA Methylome Analysis. *Cell Stem Cell* **2014**, *15*, 507–522. [[CrossRef](#)]
38. Airoidi, I.; Raffaghello, L.; Cocco, C.; Guglielmino, R.; Roncella, S.; Fedeli, F.; Gambini, C.; Pistoia, V. Heterogeneous Expression of Interleukin-18 and Its Receptor in B-Cell Lymphoproliferative Disorders Deriving from Naive, Germinal Center, and Memory B Lymphocytes. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **2004**, *10*, 144–154. [[CrossRef](#)]
39. Gandhapudi, S.K.; Tan, C.; Marino, J.H.; Taylor, A.A.; Pack, C.C.; Gaikwad, J.; Van De Wiele, C.J.; Wren, J.D.; Teague, T.K. IL-18 Acts in Synergy with IL-7 to Promote Ex Vivo Expansion of T Lymphoid Progenitor Cells. *J. Immunol.* **2015**, *194*, 3820–3828. [[CrossRef](#)] [[PubMed](#)]
40. Cheng, M.; Zhou, J.; Wu, M.; Boriboun, C.; Thorne, T.; Liu, T.; Xiang, Z.; Zeng, Q.; Tanaka, T.; Tang, Y.L.; et al. CXCR4-Mediated Bone Marrow Progenitor Cell Maintenance and Mobilization Are Modulated by c-Kit Activity. *Circ. Res.* **2010**, *107*, 1083–1093. [[CrossRef](#)]
41. Si, Y.; Tsou, C.-L.; Croft, K.; Charo, I.F. CCR2 Mediates Hematopoietic Stem and Progenitor Cell Trafficking to Sites of Inflammation in Mice. *J. Clin. Investig.* **2010**, *120*, 1192–1203. [[CrossRef](#)]
42. Singh, P.; Mohammad, K.S.; Pelus, L.M. CXCR4 Expression in the Bone Marrow Microenvironment Is Required for Hematopoietic Stem and Progenitor Cell Maintenance and Early Hematopoietic Regeneration after Myeloablation. *Stem Cells* **2020**, *38*, 849–859. [[CrossRef](#)] [[PubMed](#)]
43. Matsumoto, A.; Takeishi, S.; Kanie, T.; Susaki, E.; Onoyama, I.; Tateishi, Y.; Nakayama, K.; Nakayama, K.I. P57 Is Required for Quiescence and Maintenance of Adult Hematopoietic Stem Cells. *Cell Stem Cell* **2011**, *9*, 262–271. [[CrossRef](#)]
44. Matsuda, S.; Kominato, K.; Koide-Yoshida, S.; Miyamoto, K.; Isshiki, K.; Tsuji, A.; Yuasa, K. PCTAIRE Kinase 3/Cyclin-Dependent Kinase 18 Is Activated through Association with Cyclin A and/or Phosphorylation by Protein Kinase A. *J. Biol. Chem.* **2014**, *289*, 18387–18400. [[CrossRef](#)] [[PubMed](#)]
45. Kulakovskiy, I.V.; Vorontsov, I.E.; Yevshin, I.S.; Soboleva, A.V.; Kasianov, A.S.; Ashoor, H.; Ba-alawi, W.; Bajic, V.B.; Medvedeva, Y.A.; Kolpakov, F.A.; et al. HOCOMOCO: Expansion and Enhancement of the Collection of Transcription Factor Binding Sites Models. *Nucleic Acids Res.* **2016**, *44*, D116–D125. [[CrossRef](#)]
46. McLeay, R.C.; Bailey, T.L. Motif Enrichment Analysis: A Unified Framework and an Evaluation on ChIP Data. *BMC Bioinformatics* **2010**, *11*, 165. [[CrossRef](#)]
47. Wang, L.; Park, H.J.; Dasari, S.; Wang, S.; Kocher, J.-P.; Li, W. CPAT: Coding-Potential Assessment Tool Using an Alignment-Free Logistic Regression Model. *Nucleic Acids Res.* **2013**, *41*, e74. [[CrossRef](#)] [[PubMed](#)]
48. Zuker, M.; Stiegler, P. Optimal Computer Folding of Large RNA Sequences Using Thermodynamics and Auxiliary Information. *Nucleic Acids Res.* **1981**, *9*, 133–148. [[CrossRef](#)]
49. Beerman, I.; Bock, C.; Garrison, B.S.; Smith, Z.D.; Gu, H.; Meissner, A.; Rossi, D.J. Proliferation-Dependent Alterations of the DNA Methylation Landscape Underlie Hematopoietic Stem Cell Aging. *Cell Stem Cell* **2013**, *12*, 413–425. [[CrossRef](#)]
50. Schatz, D.G.; Ji, Y. Recombination Centres and the Orchestration of V(D)J Recombination. *Nat. Rev. Immunol.* **2011**, *11*, 251–263. [[CrossRef](#)]

51. Ferrero, G.; Gomez, E.; Lyer, S.; Rovira, M.; Miserocchi, M.; Langenau, D.M.; Bertrand, J.Y.; Wittamer, V. The macrophage-expressed gene (mpeg) 1 identifies a subpopulation of B cells in the adult zebrafish. *J. Leukoc. Biol.* **2020**, *107*, 431–443. [[CrossRef](#)] [[PubMed](#)]
52. Prohaska, S.S.; Scherer, D.C.; Weissman, I.L.; Kondo, M. Developmental Plasticity of Lymphoid Progenitors. *Semin. Immunol.* **2002**, *14*, 377–384. [[CrossRef](#)] [[PubMed](#)]
53. McKearn, J.P.; Baum, C.; Davie, J.M. Cell Surface Antigens Expressed by Subsets of Pre-B Cells and B Cells. *J. Immunol.* **1984**, *132*, 332–339.
54. Chevrier, S.; Genton, C.; Kallies, A.; Karnowski, A.; Otten, L.A.; Malissen, B.; Malissen, M.; Botto, M.; Corcoran, L.M.; Nutt, S.L.; et al. CD93 Is Required for Maintenance of Antibody Secretion and Persistence of Plasma Cells in the Bone Marrow Niche. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 3895–3900. [[CrossRef](#)] [[PubMed](#)]
55. Zhou, Y.; Tian, B.; Tang, J.; Wu, J.; Wang, H.; Wu, Z.; Li, X.; Yang, D.; Zhang, B.; Xiao, Y.; et al. SNHG7: A Novel Vital Oncogenic LncRNA in Human Cancers. *Biomed. Pharmacother. Biomed. Pharmacother.* **2020**, *124*, 109921. [[CrossRef](#)] [[PubMed](#)]