bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Deep deconvolution of the hematopoietic stem cell regulatory microenvironment reveals a high degree of specialization and conservation between mouse and human.

Jin Ye[*,1], Isabel A. Calvo[*,2,3], Itziar Cenzano[*,2], Amaia Vilas[2,3], Xabier Martinez-de-Morentin[4], Miren Lasaga[4], Diego Alignani[2,3], Bruno Paiva[2,3], Ana C. Viñado[2,3], Patxi San Martin-Uriz[2], Juan P. Romero[2,#], Delia Quilez Agreda[5], Marta Miñana Barrios[5], Ignacio Sancho González[5], Gabriele Todisco[6,7], Luca Malcovati[6], Nuria Planell[4], Borja Saez [&,2,3], Jesper Tegner[&,1,8,9], Felipe Prosper[&,2,3,10], David Gomez-Cabrero[&,1,4,8,9,11]

[1]Biological and Environmental Sciences and Engineering Division (BESE), King Abdullah University of Science and Technology KAUST, Thuwal, 23955, Saudi Arabia.

[2]Universidad de Navarra, CIMA, Hematology-Oncology Program, Instituto de Investigación Sanitaria de Navarra (IdiSNA), Navarra, Spain.

[3]Centro de Investigación Biomédica en Red de Cáncer, CIBERONC.

[4]Navarrabiomed, Complejo Hospitalario de Navarra (CHN), Universidad Pública de Navarra (UPNA), IdiSNA, Pamplona, Spain.

[5]Hospital Reina Sofía de Tudela, Navarra, Spain.

[6]Department of Molecular Medicine, University of Pavia & Unit of Precision Hematology Oncology, IRCCS S. Matteo Hospital Foundation, Pavia, Italy.

[7]Department of Medicine, Center for Hematology and Regenerative Medicine, Karolinska Institutet, Stockholm, Sweden.

[8]Department of Medicine, Centre for Molecular Medicine, Karolinska Institutet, Stockholm, Stockholm, Sweden.

[9]Computer, Electrical, and Mathematical Sciences and Engineering Division (CEMSE), King Abdullah University of Science and Technology KAUST, Thuwal, 23955, Saudi Arabia.

[10]Service of Hematology and Cell Therapy, Clínica Universidad de Navarra (CUN), Pamplona, Spain.

[11]Centre for Host Microbiome Interactions, Faculty of Dentistry, Oral & Craniofacial Sciences, King's College, London, UK.

*Equal first author contribution

[&]Senior authors with equal contribution.

Current address: 10x Genomics Inc., 6230 Stoneridge Mall Road, Pleasanton, CA 94588, USA.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ABSTRACT

Understanding the regulation of normal and malignant human hematopoiesis requires comprehensive cell atlas of the hematopoietic stem cell (HSC) regulatory microenvironment. Here, we develop a tailored bioinformatic pipeline to integrate public and proprietary single-cell RNA sequencing (scRNA-seq) datasets. As a result, we robustly identify for the first time 14 intermediate cell states and 11 stages of differentiation in the endothelial and mesenchymal BM compartments, respectively. Our data provide the most comprehensive description to date of the murine HSC-regulatory microenvironment and suggests a higher level of specialization of the cellular circuits than previously anticipated. Furthermore, this deep characterization allows to infer conserved features in human, suggesting that the layers of microenvironmental regulation of hematopoiesis may also be shared between species. Our resource and methodology are a steppingstone towards a comprehensive cell atlas of the BM microenvironment.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

INTRODUCTION

Uncovering pathogenetic mechanisms requires identifying the corresponding major groups of genes in the disease-relevant tissues[1,2]. To this end, collective efforts such as the Human Single-Cell Atlas have been launched, aiming at providing a single-cell map of human tissues and organs. A core case is the hematopoietic system, where single-cell RNA sequencing (scRNA-seq) has allowed to refine our understanding of hematopoiesis in mouse and human[3–5]. Furthermore, these studies have challenged the classical view of hematopoiesis differentiation as a compendium of discrete cellular states with decreased differentiation potential towards a more dynamic view in which hematopoietic stem and progenitor cells (HSPC) gradually pass through a continuum of differentiation states[6–9]. Moreover, recent studies using scRNA-seq technologies have shed light on the organization of the hematopoietic regulatory microenvironment in the mouse[10–15]. These studies have resolved some of the controversies regarding the overlap of stromal populations previously described and the description of certain discrete stromal cells as professional, hematopoietic cytokine-producing populations[12]. Moreover, further combination with *in-situ* technologies helped to delineate the relationship between specific stromal cell types in the murine BM[12]. This data has provided a wider and more dynamic picture of hematopoiesis and their regulatory microenvironment, allowing for a provocative hypothesis to rise, such as whether their specific association with given niches controls transcriptional states in hematopoietic stem cells and whether these states are reversible upon occupying alternative niches.

Nevertheless, these studies are limited by the number of cells sequenced, potentially hampering our ability to resolve the full spectrum of cellular states and differentiation stages that define the stromal BM microenvironment. Further, knowledge on the conservation of the cellular composition in the human BM stroma is in its infancy due to the difficulty of obtaining high-quality samples with sufficient stromal cell numbers from healthy individuals. This leaves us with two outstanding challenges; how to piece together such different fragments towards a comprehensive molecular atlas and to what extent such an atlas in mice is conserved in the human bone marrow.

Here, we integrate three scRNA-seq datasets (two publicly available[10,11] and one in-house) separately targeting two well-defined populations (endothelial and mesenchymal cells). The integration of distinct data sets required developing tailored bioinformatics pipelines to ensure the robust identification of cell types and stages. We identify 14 endothelial subclusters and 11 subpopulations defining different stages of differentiation in the mesenchyme. Our analysis provides the most comprehensive atlas of the cellular composition in the mouse bone marrow.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Last, we asked to what extent such an atlas could provide insight into the less accessible human BM microenvironment. To this end, we made the first pilot study, profiling the human bone marrow using scRNA-seq, which was integrated with our mouse BM atlas. This analysis demonstrated substantial conservation between species.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

RESULTS

Data integration and high-resolution clustering strategy

Figure 1 provides a graphical summary of the experimental design and the analysis flow. We integrated selected subsets of cells from three distinct mouse datasets: two recently published ("Tikhonova et al."[11], 6626 cells, and "Baryawno et al."[10], 38443 cells) and an independent dataset ("*In-house*" dataset, 13402). These datasets differ in the procedures for isolation of cells within the BM microenvironment. This includes unbiased isolation of cells lacking hematopoietic markers ("*Baryawno*"[10] and "*In-house*") **(Fig. S1a)** versus targeted isolation of populations of interest as in Tikhonova et al.[11]. Furthermore, not every cell type identified in one study is present in the other datasets.

We decided to focus our analysis on those bona fide niche populations such endothelial (EC) and mesenchymal (MSC) cells due to their presence in the three studies and their relevance in controlling HSC maintenance. For the integrative analysis, we used the Tikhonova study as a reference to facilitate the integration, considering that their cells were isolated based on fluorescence reporter expression driven by cell-specific gene promoters: VE-Cad for endothelial cells and LEPR for mesenchymal cells. Therefore, to identify and label the cells of interest, we integrated separately "In house + Tikhonova" and "Baryawno + Tikhonova" **(Fig. 2a)** using the cell labels from the Tikhonova study. As a result, and after quality filters (see Methods), we labeled in each dataset endothelial cells (N=9587) and mesenchymal cells (N=5291) that were used for integration of the three datasets for each cell-type.

Following the integration, endothelial and mesenchymal groups were used to identify cellular subtypes and stage-specific cells. However, current state-of-the-art clustering methodologies, including Louvain clustering[16,17], could not discriminate robustly among different cell subtypes **(Fig. 2b left panel and Fig. S3a)**, in part because there is a high degree of cell-to-cell similarity when considering cells of the same origin[18]. To enable robust sub-clustering, we customized an existing bootstrapping-based approach. In brief, first, a divide-and-conquer strategy is applied, where the first level of robust clusters is identified (see Methods). Next, we proceeded with another round of clustering, yielding the second level of robust sub-clusters. As a termination criterion, no sub-clustering was considered if a cluster was found to have no sub-divisions in the Louvain high-resolution clustering[17,19] **(Fig. 2b left panel and Fig. S3a)**. As a result, the cells are grouped into clusters; then, in the second step, we applied a bootstrapping-based methodology adapted from the Bosiljka study[18] (see Methods and **Fig. S2a)** to quantify the robustness of each cluster. We formulated two evaluation metrics: for each cell, we computed "*how many times it has been correctly assigned to the cluster proposed*" (e.g. recall per cell in

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig.S2d), and for each cluster, we quantified "*how many cells are correctly assigned*" (e.g. #correct in Fig.S2e); see Methods for a detailed explanation. If a non-robust cluster was identified, the cells of such cluster were then assigned to the neighboring clusters repeating the Random-Forest based strategy. For instance, in the analysis of ECs, three levels of sub clustering were conducted (**Fig. 2b** middle and right panels). In Level 1, two clusters were identified (A and B), and considering that possible sub-clusters were identified in the Louvain high-resolution clustering **(Fig.2b,** (second panel from the left**)**, they were sub-clustered resulting in Level 2. After Level 2 only three subclusters were further investigated (A1, A2, B3) in Level 3 **(Fig.2b, third panel)**. Level 3 was identified as the final level, and a robustness analysis was conducted for the Level 3 clusters **(Fig. 2b fourth panel**, **c-e and Fig. S2c-h)**. Non-robust clusters were eliminated and their cells were reassigned to neighboring clusters (see Methods). Similarly, we applied the same robust clustering to the mesenchymal stromal cells **(Fig. S3a-h)**. By using this approach, we were able to describe 14 subclusters in the endothelium **(Fig. 2b fourth panel and Fig. 3a)** and 11 in the mesenchyme **(Fig. S3a fourth panel and Fig. 3c)**. Encouragingly, we observed that the final subclusters were not biased to any study or any specific cell stage **(Fig. 3b,d and Fig. S2b)** and only clusters B2 and D3 did not have cells from the three datasets**.**

A fundamental tenet of our analysis is that the identification of distinct cellular states in the microenvironment and the description of the gene markers defining those subtypes could only be resolved by integrating the three datasets. The rationale being that despite the partial overlap between the datasets, a larger number of cells would contribute to a deeper characterization of the subpopulations. To directly address this and using the markers derived from the integrative analysis, we conducted a series of analyses to assess whether the integration provided additional insights compared to each dataset.

First, we investigated what percentage of final integrated-based markers were recovered by each dataset by itself **(Fig. 3e,f upper panel)**, what percentage of false negatives **(Fig. 3e,f middle panel)**, and false positives **(Fig. 3e,f lower panel)**. For some of the subclusters, such as B3.4, A2.1, and A2.6 (endothelium) or C2.1 and C3 (mesenchyme), over 50% of the markers could not be detected by each dataset separately. In a second analysis, we quantified the robustness of the defined clusters using data from each study separately **(Fig. 3g,h)**. Only the Baryawno dataset allows for the robust identification of all the subclusters except for B3.4 in the endothelium. However, performing the same clustering strategy using only the Baryawno's dataset in the endothelium, could not identify all the subclusters with the same level of resolution compared to those observed with the integrated dataset **(Fig. S4a)**.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Taken together, these data demonstrate that our customized approach for the integration of multiple datasets allows for a robust deconvolution of cell states when there is a high degree of similarity in between cells of the same origin.

Deep characterization of the BM endothelial cell compartment

Next, we aimed to characterize each of those 14 endothelial subclusters **(Fig. 3a)** based on the identified markers **(Table S1)**. Using the expression of those molecular markers, we could discriminate between arteries and sinusoids **(Fig. 4a and Table S2)** in agreement with previous reports[20–22]. Arterial clusters showed high expression of specific arterial genes such as *Ly6a, Ly6c1, Igfbp3,* and *Vim* **(Fig. 4b upper panels)**. At the same time, sinusoidal cells were defined by their characteristic signature expressing *Adamts5, Stab2, Il6st, and Ubd* **(Fig. 4b lower panels)**. Importantly, besides these already known markers, differential expression analysis of the integrative datasets revealed some novel genes specific of each endothelial population, such as *Igfbp7* and *Ppia* for arteries and *Cd164* or *Blrv* for sinusoids **(Fig. 4c)**. The expression of these genes would be consistent with the role *Igfbp7* and *Blvrb* in the maintenance of endothelial vasculature homeostasis[23,24].

Beyond characterizing arteries and sinusoids, we annotated their respective cell states using clusters markers based on genes **(Table S1, Fig. 4d)** and gene sets (derived from gene set enrichment analysis; **Table S3, Fig. 4e)**. Our final annotation is described in **Fig. 4f**. Gene sets related to vasculature development and remodeling were identified within the endothelial subclusters, confirming the identity of this cell population **(Table S3)**. We uncovered two subclusters (A1.1 and A2.5, respectively) within the arteries and sinusoids groups, which were enriched in gene sets involved in wounding. This finding is in agreement with the role of EC in pro- and anti-thrombotic processes[25]. Gene sets involved in extracellular matrix assembly, cell adhesion, and migration processes were specifically enriched in arteries, in line with the importance of these biological processes for vascular morphogenesis[26]. In relation to the structural support provided by arterial cells, subcluster B3.2 (actin, endocytosis) implicated in matrix remodeling was defined by the expression of *RhoC, Apln,* and *Anxa2*. Other arterial subclusters such as ROS and Immune (A1.2 and B3.1, respectively) include highly expressed gene sets involved in the regulation of reactive oxygen species metabolic process and cytokine-mediated signaling pathway. These findings are in line with the role of ECs in maintaining the REDOX balance and leukocyte regulation[27,28]. Sinusoidal-endothelial subclusters such as A2.1 and A2.6 showed enrichment in GO terms related to artery development and endothelial

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

proliferation processes, two critical steps within the process of angiogenesis[29]. Furthermore, the sinusoidal subcluster A2.7 expressed gene sets involved in ion transport and signaling-related signatures. This is in concordance with the need of EC to constantly sense and adapt to alterations in response to microenvironmental cues[30,31]. Of note, ion channels play a role in EC functions controlled by intracellular Ca^{2+} signals, such as the production and release of many vasoactive factors such as nitric oxide. In addition, these channels are involved in the regulation of the traffic of macromolecules, controlling intercellular permeability, EC proliferation, and angiogenesis[32]. Importantly, several markers that were found only with the dataset integration correspond to genes within the GO categories used to label the clusters, hence, revealing their important role in defining the function of these cell states. For example, in subcluster B2, genes such as, *Gja1*, *Tgfb3,* and *Ablim2* are involved in regulating cell junctions and cytoskeletal organization[33,34].

Taken together, these results suggest a remarkable level of specialization of the bone marrow endothelial cells. The specificity of the distinct functional states indicates that the endothelial compartment is a more dynamic and flexible tissue with a richer intrinsic repertoire than previously anticipated.

Deep characterization of the BM mesenchymal cell compartment

Applying the same robust clustering to mesenchymal stromal cells, we identified 11 subclusters and proceeded with the annotation **(Fig. 3c, Fig. S3, and Table S4-S6)**. Based on the expression of canonical markers, we first discriminated clusters between early mesenchymal (MSC), and cells already committed to the osteolineage (OLN-primed) **(Fig. 5a and Table S5)**. The high expression of *Cxcl12*, *LepR*, *Adipoq,* or *Vcam1,* among others, confirmed the identity of the early MSC group **(Fig. 5b left column)**; whereas *Bglap*, *Cd200*, *Alpl,* or *Col1a1* expression revealed the presence of osteolineage committed cells within the mesenchymal compartment **(Fig. 5b right column)**. Furthermore, we identified a number of previously unrecognized, differentially expressed genes between the MSC and OLN-primed clusters such as *Sbds* and *Itgb1* for MSCs and *Enpp1* and *Vkorc1*, for OLN-primed cell type population **(Fig. 5c)**. *Itgb1*, highly expressed in MSC is implicated in human chondrogenic differentiation of mesenchymal cells[35]. Among OLN-primed specific markers, *Enpp1* and *Vkorc1* have been shown to regulate bone development by regulating bone calcification[36–39].

Besides MSCs and OLN-primed MSCs, we identified additional subpopulations. Through the marker identification and gene set analysis of the 11 sub-clusters **(Fig. 5d,e and Table S4,6)** we were able to characterize and label each of the clusters as shown in **Fig. 5f**. GO terms such

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

as adipogenesis, assembly and organization, immune response, cell migration or muscle differentiation were enriched in the C2.3, C2.1, C1, C3, C4.2 and C4.1 subsets respectively, confirming the identity of this MSC cell group. Furthermore, terms related to extracellular matrix, chondrocyte differentiation and bone development, including bone formation, ossification, or epithelial migration, among others, were identified in the OLN-primed subclusters (C5, C6, and D respectively), verifying the identity of these more mature cells (OLN-primed cells) within the mesenchymal stromal cells.

Taken together, these results demonstrate that the newly identified mesenchymal subpopulations could not be properly characterized without the multi-dataset integration and a novel clustering approach. Further, our data provide evidence of the heterogeneity of the mesenchymal compartment in the BM.

Composition of the human endothelial and mesenchymal BM microenvironment

While our data reveal a previously unrecognized heterogeneity in the murine BM endothelial and mesenchymal compartments, information about the composition of the human microenvironment and how much of this heterogeneity is observed in humans remains unanswered. To address this issue, we performed scRNA-seq analysis in prospectively isolated EC (TO-PRO-3^{-}, CD45^{-}, CD235^{-}, Lin^{-}, CD31^{+} and CD9^{+}[40,41]) and MSC-OLN (TO-PRO-3^{-}, CD45^{-}, CD235^{-},Lin^{-},CD271^{+}[42–44], CD146$^{+/-}$) **(Fig. S5a)** from iliac crest bone marrow aspirates from four healthy young adults (20-30 years of age) (**Fig. 6a**). As described in **Fig. 6b**, we added additional filtering steps in the bioinformatic analysis to identify the two populations of interest, EC and MSC. As an additional quality control, we estimated the contribution of each human sample to EC and MSC subsets and cell cycle status (**Fig. S5b,c)**. The EC (907 and 658 cells, clusters 1 and 6 respectively) (**Fig. 6c and Fig. S5b middle panel**) identity was confirmed based on the expression of canonical endothelial markers such as *PECAM1* (coding for *CD31*), *CD9, ICAM2, VLC,* and *ITGB1* (**Fig. 6d and Table S7**). In addition, examining functional pathways in clusters 1 and 6 revealed enrichment in GO terms associated with blood coagulation and hemostasis **(Table S8).** The MSC identity (249 cells, cluster 11, **Fig. 6c and Fig. S5b middle panel**) was confirmed by the expression of the mesenchymal specific genes (*CXCL12 and LEPR*) and the OLN-Primed specific genes *ANGPT1, COL1A1, and VCAM1,* among others (**Fig. 6e and Table S7**). Further, enrichment in functions associated with extracellular matrix organization and response to the mechanical stimulus was demonstrated in osteolineage cells **(Table S8)**. In summary, the generated human data suggests that single-cell RNA sequencing from iliac crest aspirates can aid in describing the complexity of the human BM

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

microenvironment. Nevertheless, the limited number of EC and MSC, as well as the presence of contaminating populations did not allow a fine-grained clustering as the one performed in the mouse data.

Significant conservation of the EC and MSC population in the human BM microenvironment

Based on the limitations of the human data, we next investigated to what extend the knowledge uncovered in mouse could be applied to identify subpopulations/cell states in the human BM microenvironment. As a first step, we used the markers identifying the different clusters in mouse to annotate the human cells using SingleR[45], separately for EC **(Fig. S6a,b)** and MSC **(Fig. S6c,d)**. We observed that markers identified in mouse allowed us to separate the cells into clusters for both human EC and MSC **(Fig. S6a,c)**, although the scoring associated with the cells was not very strong **(Fig. S6b,d)** (there was not a perfect match between mouse and human EC and MSC populations). Nevertheless, this analysis suggests that part of the biological mechanisms defining the BM microenvironment may be shared between species.

Therefore, we decided to investigate the conserved features between mouse and human per cluster. For endothelial cells, we analyzed the enrichment in mouse EC markers within the human EC most variable genes (MVG), and we observed that for most of the subclusters, around 20-30 markers genes were shared between species, which account for less than 20% of the total mice markers for each subcluster **(Fig. 7a and Table S8-S9)**. However, we also observed that such numbers were associated with an enrichment score (ES) (by comparing expected vs. observed, see Methods) of up to two-fold for some of the clusters: wounding (A1.1) with 2.15-ES, the junction (B2) 2.26-ES, arteriogenesis (A2.1) 2.02-ES and Signaling (A2.7) 2.5-ES **(Fig. 7a)**. Importantly, for some of the subclusters as sinusoidal signaling (A2.7) and the arterial of angiogenesis (B3.3), these shared genes are critical for defining each of those EC functional states **(Table S9)**. Both human and mouse ECs express *DDIT4*, *JUN*, *CITED2*, *GADD45G*, *DUSP1*, *FOS,* and *CLDN5,* which are part of a wide array of transcription factors, growth factors, and signaling pathways that have been described to regulate the maintenance of vascular homeostasis under physiological conditions[46–48]. Similarly, ECs subclusters involved in angiogenesis in both species shared the expression of *RGCC*, *GATA2*, *KLF2,* and *CAV2* genes, which are known to be implicated in angiogenic related processes[49,50].

In the case of mesenchymal cells, using a similar analysis, we identified a high percentage of shared genes in all the subclusters, with up to 35% shared genes in ten out of eleven subclusters. It should be noted that we identified >3.5-ES for three subclusters, such as RNA-

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Myogenesis (C4.1), Factors-Immune (C4.2), and ossification (D2). Importantly, some of the shared genes correspond to genes that allowed the subcluster labelling through GO categories **(Fig. 7b and table S8 and S10)**. Genes such as *CXCL12, APOE,* or *IGFBP3* are associated with cell migration and lipid transport pathways among others[51,52], and characterized the murine adipogenesis subcluster (C2.3). Other genes, such as *IGFBP5*, are involved in actin filament assembly and organization[53] and defined the cell adhesion subcluster (C2.1). *IFIT3, MIDN, and ILR1* belong to pathways associated to interferon regulation or autoimmunity[54,55] and were identified in the immune response subclusters (C3 and C4.2). Additionally, the expression of *COL5A1* and *CADM1* genes, previously related to collagen fibril organization and bone mineralization processes[56,57], defined the bone formation subcluster (D1). Moreover, genes such as *SPP1* or *CLEC11A*, which are related to osteoblasts function and mineralization[58–60] defined the mouse OLN-primed MSCs subcluster associated to ossification (D2) and are also expressed in human MSC-OLN cluster. Altogether, this data indicates the conservation of the osteogenic microenvironment between both species.

Together, our analysis suggests that deep characterization of cellular states in mice can be used to infer conserved features in the human BM microenvironment. Importantly, our data reveal a high degree of conservation regarding the complexity and heterogeneity of the EC and MSC compartment in the BM between mouse and human, suggesting that the layers of microenvironmental regulation of hematopoiesis and the identified plasticity in mice may also be shared between species.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

DISCUSSION

Our study dissects the intrinsic organization and the heterogeneity within the endothelial (EC) and mesenchymal cell populations (MSC) governing the BM microenvironment. This was accomplished through customized bioinformatics integration of multiple datasets along with the inclusion of over 50.000 murine bone marrow stromal cells. We were able to identify new subsets of MSC and EC but, but more importantly, to define new molecular markers for the identification of highly specialized subpopulations of cells in the BM microenvironment. Pathway enrichment analysis unveiled multiple, potentially transient cell states defined by differential gene expression and the enrichment of specific functional characteristics. Importantly, EC subsets were characterized by enrichment in pathways known to be essential for endothelial homeostasis maintenance, demonstrating a high degree of specialization in the endothelium. Similarly, multiple transient cell states in the MSC compartment were defined and characterized by their differentiation capacity. Importantly, our deep deconvolution of the heterogenous mesenchymal and endothelial compartments became feasible only by integrating multiple datasets. Of note, our analysis showcases that a research paradigm aiming for the generation of a detailed comprehensive molecular atlas of an organism requires both multi-omic data and computational integration. Here, we have relied on what is referred to as unpaired unimodal (scRNA-seq) data. Clearly, a natural next step is to use and further develop new computational tools that permit the integration[61] of unpaired multi-omics datasets such as scRNA-seq, scATAC-seq, and other data modalities. Recent technological developments enable several multiple omics recorded from the same cell, i.e. paired data, which leverages our ability to dissect and molecularly characterize the intrinsic organization of the bone marrow niche environment. Advances in computational biology have started to develop such tools[62–65].

While our study did not directly address the influence between stromal cells in the hematopoietic stem cell niche and the HSCs, the deep resolution of our study allows for some inferences to be made. Of note, we detected the expression of vascular endothelial growth factor-C (*Vegf-c*) in mouse endothelial and mesenchymal cells **(Table S1, S4, S5, S7, and S10)**. *Vegf-c* has recently been implicated in the maintenance of the perivascular niche and the recovery of hematopoiesis upon injury[66]. *Vegf-c* is specifically expressed in the endothelial B1 and mesenchymal C1 subpopulations, suggesting an important role of these specialized endothelial and mesenchymal cells in the preservation of the integrity of the perivascular niche. In addition, Apelin+ (Apln) endothelial cells have been recently implicated in HSC maintenance and regeneration upon injury[67]. Importantly, two endothelial subclusters (B3.2-actin endocytosis and

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A1.1-wounding arteries) demonstrated expression of Apln, suggesting that these EC states represent specific sources of hematopoietic support and vascular regeneration upon injury. Osteolectin+ LepR+ - mechanosensitive peri-arteriolar mesenchymal cells with osteogenic potential - are implicated in lymphoid, but not HSC maintenance[59]. Importantly, osteolectin expression defines murine cluster D2 (ossification) and shows conservation in human MSCs, suggesting the preservation of a specialized lymphoid niche between species.

Detailed characterization of the human BM niche has not yet been addressed. Approaches undertaken in a mouse system cannot readily be transferred to the human system. Furthermore, differences in sample processing can also impact the results. In that sense, our results, despite the low number of cells, may represent the first dataset that includes scRNAseq from the human endothelial and mesenchymal BM microenvironment. While we were able to identify mesenchymal and endothelial cells based on canonical markers shared with mice[10,11,68–75], our human scRNAseq did not possess enough resolution to elucidate the heterogeneity of the human BM stroma to the same level as with the mouse data. Based on the extensive knowledge generated in the mouse, we therefore focused on characterizing how much of the information and targets from the mouse can be of interest in human characterization. This analysis allowed us to identify the expression of the human orthologs to the murine cluster-defining genes with different degrees of enrichment in the endothelium and mesenchyme. Moreover, some of these shared genes in mice and human stromal cells corresponded to the GO-defining genes of the different clusters identified in the mouse. These findings suggest a significant degree of conservation regarding the cellular states that define the stromal microenvironment in mouse and human. Although additional studies and improved processing of human samples will be required for deep characterization of the human BM microenvironment, these preliminary results validate our integrative cross-species approach.

As an example of the added value, the current study identifies candidates of relevance in the study of BM related diseases. *Sbds*, a ribosome maturation protein associated with the Shwachman-Diamond syndrome, represent a previously unrecognized marker of immature MSCs based on the dataset integration. Sbds deficiency has been implicated in ossification defects and metabolic changes in HSPCs[76,77], potentially contributing to myelodysplasia and AML onset in Shwachman-Diamond syndrome patients. On a broader note, deep molecular analysis of the BM microenvironment set the stage for computational disease modelling[78].

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Taken together, our study provides a deeper understanding of the composition and specialization of the BM microenvironment and point towards a significant degree of conservation between species. Moreover, we demonstrate the usefulness of the multi-dataset integration and the customized clustering approach used in our study to improve the resolution of complex tissues and organs. This approach promises to aid in the construction of cell atlases by reducing the resources associated with sequencing that a single lab will need to invest in order to obtain meaningful depth in single-cell analysis.

Future studies integrating genome, transcriptome, epigenome, proteome, and anatomical positioning together with functional assays to correlate descriptive phenotypes with functional data will help fully resolve the composition, regulation, and connectivity in the BM microenvironment in health and disease.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MATERIALS AND METHODS

Mouse and human samples

Female C57BL/6J mice (CD45.2, Jackson Laboratory #000664) at age 20 weeks were used for scRNA-seq experiments. All animal experiments were performed in accordance with national and institutional guidelines and procedures were approved by the Ethical Committee for Animal Testing of the University of Navarra.

The human sample collection and research conducted in this study were approved by the Research Ethics Committee of the University of Navarra. All the protocols used in this study were in strict compliance with the legal and ethical regulations. After informed consent, a total volume of approximately 60 mL bone marrow was obtained by aspiration from the posterior iliac crest from healthy young volunteers (20-30 years of age).

Isolation and FACS sorting of murine bone marrow microenvironment cells

Mice (x6) were euthanized via CO_2 asphyxiation. Bones from humerus, radius, iliac crests, femurs and tibia were harvested in PBS 1X containing 2 % FBS and 2 mM EDTA (modPBS). All steps were performed on ice to preserve cell viability and RNA integrity. Muscles and soft tissue were thoroughly removed from the bones and BM cells were obtained by crushing in modPBS. Cells were then filtered through a 70 µm cell strainer and red blood lysed with ACK buffer (NH_4Cl 150 mM, $KHCO_3$ 10 mM, and Na_2EDTA 0,1 mM) for 10 minutes at room temperature (RT). Remaining calcified bone fragments were collected on a 50 mL conical tube and digested with the appropriate volume of PBS with 0.3% collagenase I and dispase (5 U/ml) during 15 min at 37ºC and shaking at 200 rpm. FBS representing 10% of the digestion volume was added to stop the collagenase digestion. After digestion, the calcified and crushed fractions were filtered through a 70 µm filter into a collection tube and pooled into one sample. Cells were subsequently stained for 20 minutes on ice first in the appropriate volume of modified PBS 1X (3 ml/mouse) with 160 ul/mouse of biotinylated lineage cocktail (Mac1, CD3, Gr1, B220 and TER119) followed by incubation with streptavidin magnetics microbeads (100 µl/mouse). Negative selection was performed using Miltenyi LD columns according to manufacturers' protocol.

After selection, the sample was stained with the following combination of conjugated antibodies at a concentration of 1/200: APC-Cy7 labeled streptavidin, BV510 labeled anti-CD45, APC labeled anti-CD45 and APC labeled anti-Ter119. Samples were then stained with 0.05 µM of Vybrant dye orange (VDO) at 37ºC for 30 minutes to label living cells. Annexin V was also

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

added, in combination with 7AAD to discard apoptotic and dead cells from the sample respectively. For annexin V staining, cells were stained with 1 µl/mouse of Annexin V-FITC on an appropriate volume of 1X Annexin V binding buffer in the dark for 15 min at RT. Samples were then resuspended in 1X Annexin V buffer and 5 µl of 7AAD dye (up to $1x10^6$ cells) were added. BM non-hematopoietic cells were FACS sorted using BD FACSAria II sorter collected in PBS 1X supplemented with 0.05% UltraPure BSA and cell viability was assessed using Nexcelom Cellometer. Data were analyzed by FACSDiva (BD) or FlowJo (version 10.7.1) software.

Isolation and FACS sorting of human bone marrow endothelial and and mesenchymal-osteolineage cells

All sample processing steps were performed on ice to preserve cell viability and RNA integrity. A total volume of approximately 60 ml of bone marrow was obtained by aspiration from the posterior iliac crest. Red blood cells were lysed twice with 45 ml of ACK buffer per 5 ml of human sample during 15 minutes at RT with rotation. Sample was then filtered through a 70 µm cell strainer, centrifuged, and stained for 30 min on ice with the following combination of conjugated antibodies at a concentration of 1/100 except anti-Lin (3ul/test- test $25x10^6$cells): BV510 labeled anti-Lin (including CD3, CD10, CD19, CD45 and CD64), BV421 labeled anti-CD235, BV421 labeled anti-CD45, FITC labeled anti-CD31, APC-Cy7 labeled anti-CD9, PE labeled anti-CD146 and PerCP-Cy5.5 labeled anti-CD271. Dead cells and debris were firstly excluded by FSC, SSC and adding 10 µl of TO-PRO-3. BM niche populations were prospectively isolated based on the following immunophenotype: ECs: TO-PRO-3$^-$/Lin$^-$/CD45$^-$/CD235$^-$/CD9$^+$/CD31$^+$ and MSCs: TO-PRO-3$^-$/Lin$^-$/CD45$^-$/CD235$^-$/CD31$^-$/CD271$^+$/CD146$^{+/-}$. FACS sorting was performed on a BD FACSAria II sorter, sorted BM niche cells were collected in PBS 1x and 0.05% UltraPure BSA and cell viability was determined using Nexcelom Cellometer. Data were analyzed by FACSDiva (BD) or FlowJo (version 10.7.1) software.

Profiling by Single-cell RNA-sequencing (scRNA-seq)

scRNA-seq was performed using the Single Cell 3' Reagent Kits v3.1 (10X Genomics) according to the manufacturer's instructions. For human samples, endothelial and mesenchymal cells were pooled before scRNA-seq was performed. Approximately 15,000 cells were loaded at a concentration of 1,000 cells/µL on a Chromium Controller instrument (10X Genomics) to generate single-cell gel bead-in-emulsions (GEMs). In this step, each cell was encapsulated with primers containing a fixed Illumina Read 1 sequence, followed by a cell-identifying 16 bp

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

10X barcode, a 10 bp Unique Molecular Identifier (UMI) and a poly-dT sequence. A subsequent reverse transcription yielded full-length, barcoded cDNA. This cDNA was then released from the GEMs, PCR-amplified and purified with magnetic beads (SPRIselect, Beckman Coulter). Enzymatic Fragmentation and Size Selection was used to optimize cDNA size prior to library construction. Illumina adaptor sequences were added, and the resulting library was amplified via end repair, A-tailing, adaptor ligation and PCR. Libraries quality control and quantification was performed using Qubit 3.0 Fluorometer (Life Technologies) and Agilent's 4200 TapeStation System (Agilent), respectively. Sequencing was performed in a NextSeq500 (Illumina) (Read 1: 26 cycles, i7 Index: 8 cycles, Read 2: 49 cycles) at an average depth of 60,000 reads/cell in mice and 30,000 reads/cell in human.

Single-cell RNA-seq Analysis of mouse samples

Sample selection: Sample GSM3674224, GSM3674225, GSM3674226, GSM3674227, GSM3674228, GSM3674229 from GSE128423 by Baryawno, sample GSM2915575, GSM2915576, GSM2915577 from GSE108891 by Tikhonova and one in-house mouse bone marrow niche sample was included in this analysis.

Filtering: the single cell analysis of mice samples analysis was performed using R (version 4.0.3, 3.6.3) and Seurat (version 4.0.0, 3.2.3)[79]. Three bone marrow niche samples were filtered individually based on the 10th and 90th quantile of number of features and counts. Cells with more than 5% mitochondrial genes were also removed. Each dataset was normalized using SCTransform[80] separately.

Pairwise integration and selection of the target population: In-house dataset and Baryawno were integrated with Tikhonova separately using Seurat (version 3.2.3.). Using as a reference the annotation from Tikhonova dataset, cells that aligned with LEPR+ cells and VE-Cad+ cells were annotated as MSC and EC respectively. MSC-like cells and EC-like cells from different datasets were normalized again and integrated using Seurat without further filtering.

Clustering: After filtering and quality control, a divide-and-conquer strategy was applied to the clustering of mouse ECs and MSCs separately. Firstly, following integration, dimension reduction with principal component analysis (PCA), data visualization with Uniform Manifold Approximation and Projection (UMAP), computation of K-nearest neighbors and clustering using resolution of 1 were performed as a reference of high-resolution limit. Secondly, IKAP[81] was

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

applied to each integrated dataset as "level 1" clustering (Fig. 2,b). Each cluster from level 1 was then compare with the high-resolution reference. The cluster from level 1 was further divided using IKAP to level 2 if the cluster were far from cluster limit. The process would be repeated until at least one cluster reach cluster limit.

Cluster evaluation: To evaluate the stability of these clusters, a bootstrapping strategy was adopted[18]. The strategy was conducted in a pairwise manner with basic steps as follow:

1. Select two clusters, randomly split the clusters to five equal groups and use one group of cells (20%) as testing dataset.
2. Identify the set of differentially expressed genes (DEGs) between the 2 clusters using the Wilconxon Rank Sum test.
3. Train a random forest classifier with 20 genes selecting the top 10 DEGs from each cluster based on average log2 fold change.
4. Applied the classifier to the 20% testing dataset.
5. Repeat step1-4 for five times for different groups such that each cell in these two clusters was classified once.
6. Repeat step1-5 ten times.
7. Repeat step1-6 for all cluster pairs

There are three types of results that can be summarized from this bootstrapping strategy:

1) For every pair of clusters, a cell pertains to which one most of the time? If a cell has been assigned to a cluster more than five times in 10 runs, then this cluster was considered as dominant cluster for this cell.
2) For every cell: how many times it was assigned to its original cluster? We defined *recall per cell* as the proportion of assignment to its original cluster in all runs from all comparisons made between clusters. See Fig. S2d.
3) For each cluster: how many times out of all pair comparisons, it was the dominant cluster for cells in this cluster? The times a cluster has been dominant cluster for a cell ranges from 0 to the total number of comparisons made between clusters. We represent this type of result with a bar plot shown the distribution of cells with different times that this cluster has been its dominant cluster in all comparisons (Fig. S2e).

Clusters where more than 50% of the cells has been "incorrectly" assigned robustly at least once to another dominant cluster will be considered unstable (Fig. S2, cluster A2.2). The cells of

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

such cluster will be assigned to other clusters (see dominant plots, **Fig. S7** as an example) using – to that end - a random forest classifier as described before.

Gene Set analysis: After clustering, DEGs of each cluster were identified using a Wilconxon Rank Sum test. For each cluster, the relevant gene sets were identified using the top 50 DEGs using clusterProfiler (version 3.18.1)[82].

Added value analysis

Added value 1: Comparing the DEGs defined by individual dataset and integrated dataset. The individual datasets were normalized with SCTransform and DEGs were identified within top 3000 most variable genes using a Wilconxon Rank Sum test. For integrated dataset, DEGs were identified within top 3000 most variable genes from integrated assay using a Wilconxon Rank Sum test. False negative and false positive rate were calculated by comparing the DEGs identified by integrated dataset and individual dataset (Fig. 3e,f).

Added value 2: Cluster stability evaluation for individual dataset. To understand if the clusters identified from three datasets can remain stable within a single dataset or not, the bootstrapping strategy was applied to each dataset with the annotation identified by the integrated dataset.

Added value 3: Comparing cluster identified by a single dataset. To further understand if the clusters can be identified by one dataset only, Baryawno dataset was used as an example considering its large cell populations. The same pipeline from normalization to bootstrapping was applied to this dataset and the clusters identified from this single dataset was compared with the clusters identified by three datasets using Jaccard index.

Single-cell RNA-seq Analysis of Human samples

Preprocessing of sequencing data: Preprocessing of single-cell RNA-seq data for each in house sample were conducted by CellRanger count from Cell Ranger (version 6.0.1) using reference genome GRCh38.

Sample filtering: The single cell analysis of human analysis was performed as described in before except for human cells with more than 10% mitochondrial genes were also removed. Because the exploratory data analysis revealed potential contamination of B cells, we applied

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

an additional filter: cells with more than 10% reads mapped to immunoglobin genes were excluded from downstream analysis.

Integration: After filtering, each sample was normalized using SCTransform and integrated using 3000 most variable features using Seurat. Following integration, dimension reduction with PCA, data visualization with UMAP, computation of K-nearest neighbors with 20 dimensions and clustering using resolution of 0.4 were performed.

Select EC and MSC (without further integration): Clusters were annotated based on biological insights on markers. Cluster 11 were identified as *mesenchymal* and cluster 1, 6 were identified as *endothelial*. During the exploratory analysis, human EC cells were subclustered at resolution 0.4 and one of the clusters identified was further filtered for the downstream analysis because the cells in the cluster were not expressing EC marker genes. Several outliers from human MSC cells were also removed.

Compare human MVGs with mouse DEGs: Human EC specific MVGs (932) and human MSC specific MVGs (976) identified from "RNA" assay were used. The MVGs from human EC or MSC were compared with DEGs from each mouse cluster. The enrichment score for a given cluster *i* was defined as the ratio between "*the number of genes shared between human MVGs and mouse DEGs from cluster i*" and "*the expected number of genes*", where the later was computed as follows:

$\frac{MVG_{human} \cap MVG_{mouse}}{MVG_{mouse}} \times l_i$, l_i: the number of DEGs from mouse cluster i.

SingleR analysis between mouse and human

To annotate human cells using mouse clusters as reference the singleR tool (version 1.4.1)[45] was utilized for ECs and MSCs separately. Cell type specific MVGs with expression values "integrated" assay from Seurat object were used for this analysis.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURES

Figure 1. Overview of the paper. Graphical brief description of the paper.

Figure 2. Data integration and high-resolution clustering strategy. (a) Integrated analysis of the bone marrow niche datasets (two publicly available, Tikhonova et al. and Baryawno et al., and one generated in our lab, *in-house*) separately for two well-defined populations (endothelial and mesenchymal cells). Tikhonova et al. data set are used as a reference considering their separated cell profiling strategy for COL2.3+, LEPR+ and VE-Cad+ populations. In the top row, a UMAP projection is depicted for each single-cell RNA dataset. In the left-lower ("In-house + Tikhonova") and in the right-lower ("Baryawno + Tikhonova") data sets are integrated to identify endothelial and mesenchymal populations. **(b)** Clustering strategy: the analysis of endothelial cells as an example. An upper limit to the cluster is set for the clustering (left panel) using Louvain high-resolution clustering. Then, an iterative divide-and-conquer strategy identifies the optimal level of clusters at different levels: Level 1 (second panel from the left), Level 2 (third panel), and Level 3 (fourth panel). **(c), (d), (e)** The robustness analysis for sub-clustering B3 (from Level 2 to Level 3). Specifically: **(c)** sub-clusters identified, **(d)** the fraction of assignments to its original cluster *using a random-forest + bootstrapping strategy* and **(e)** summary of the results (d) per cluster, #correct indicates the times a cluster is a dominant cluster (see Methods) for a cell inside in all pairwise comparisons (see Fig.S2 for the sub-clustering analysis of A1 and A2).

Figure 3. Quality control and added value of the clustering analysis. (a) Representation of the final endothelial clusters. **(b)** Left panel: proportion of cells per dataset in each cluster in the final Endothelial clustering analysis. Right panel: proportion of cells per cell cycle stage using Seurat in each cluster in the final Endothelial clustering analysis. **(c,d)** Similar as (a,b) for MSC. **(e,f)** The added value of the integrated approach. Upper panel: every cell depicts the % of markers identified per cluster using only one dataset when compared with the markers identified in the total dataset. Middle panel: every cell depicts the % of False Negatives. Bottom panel: every cell depicts the % of False Positives when comparing the analysis conducted within each dataset with the integrated analysis (considered as the correct result). (e) and (f) are respectively associated with endothelial and mesenchymal cells. **(g,h)** Robustness of the cluster characterization using only cells from a single dataset but maintaining the same cluster structure. (g) and (h) are respectively associated with endothelial and mesenchymal cells.

Figure 4. Deep characterization of the endothelial cell compartment in the BM. (a) UMAP representation of arteries (red) and sinusoids (blue) within the endothelial cell population. The right-bottom subpanel depicts the final endothelial clusters identified. **(b)** Violin plot of gene expression for known markers of arteries and sinusoids cell sub-types. **(c)** Violin plot of gene expression for new marker candidates separating arteries and sinusoids cell sub-types. **(d)** Dot plot of the top 5 markers for each endothelial subcluster. Dot size corresponds to the proportion of cells within the group expressing each gene, and dot color correspond to the average expression level. **(e)** Selected set of gene-sets derived from the gene-set analysis conducted

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

with top 50 markers per cluster (see Methods). **(f)** Final clustering of the endothelial cell population and the labeling assigned based on marker genes and gene-set analysis.

Figure 5. Deep characterization of the mesenchymal cell compartment in the BM. (a) UMAP representation of mesenchymal (red) and Osteolineage-primed (OLN-primed) (blue) within the mesenchymal compartment. The right-upper subpanel depicts the final mesenchymal clusters identified. **(b)** Violin plot of gene expression for known markers of Mesenchymal (red) and Osteolineage-primed (blue) cells. **(c)** Violin plot of gene expression for new marker candidates separating Mesenchymal (red) and Osteolineage-primed (blue) cells. **(d)** Dot plot of the top 5 markers for each mesenchymal subcluster. Dot size corresponds to the proportion of cells within the group expressing each gene, and dot color correspond to the average expression level. **(e)** Selected set of gene-sets derived from the gene-set analysis conducted with top 50 markers per cluster (see Methods). (*) *transmembrane receptor protein serine/threonine kinase signaling pathway*. **(f)** Final clustering of the mesenchymal cell population and the annotation based on marker genes and gene-set analysis.

Figure 6. Composition of the human endothelial and mesenchymal BM microenvironment. (a) Experimental design for the human BMM characterization. **(b)** Scheme of customized bioinformatics pipeline filtering the cells with a large number of Immunoglobulin genes. **(c)** UMAP visualization of color-coded clustering of the human BM microenvironment after filtering cells. **(d,e)** Expression of representative markers for endothelial population (d) and mesenchymal-osteolineage cells (e) using an UMAP representation.

Figure 7. Conservation analysis of the EC and MSC population in the human BM microenvironment. (a,b) Quantification of the conservation for EC (a) and MSC-OLN (b) cells for each cluster. The enrichment of those genes that are cluster markers in mouse and observed in most variable genes (MVG) of EC and MSC human cells respectively. The right column shows, among the genes identified in human, those that are part of the gene-sets used to label the cluster.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ACKNOWLEDGMENTS

We would like to thank the staff of the flow cytometry core, the advances genomic lab and the animal facility at CIMA Universidad de Navarra for their invaluable technical and intellectual assistance. We are particularly grateful to the healthy volunteers who donated bone marrow tissue for this study.

This research was funded by grants from The Spanish Government, through project PID2019-111192GA-I00 (MICINN) to DGC. Instituto de Salud Carlos III (ISCIII) and co-financed by FEDER: PI16/02024, PI17/00701 and PI19/01352, TRANSCAN EPICA AC16/00041, CIBERONC CB16/12/00489; Redes de Investigación Cooperativa (TERCEL RD16/0011/0005); Spanish Ministry of Economy, Industry and Competitivity (RTHALMY SAF2017-92632-EXP); Departamento de Salud-Gobierno de Navarra 40/2016 and Departamento de Desarrollo Económico y Empresarial (AGATA 0011-1411-2020-000010 and 0011-1411-2020-000013). The study was also supported by Cancer Research UK [C355/A26819] and Cancer Research UK, FCAECC and AIRC under the Accelerator Award Program (EDITOR), the Multiple Myeloma Research Foundation Networks of excellence 2017 Immunotherapy Program Grant Award, the International Myeloma Foundation (Brian van Novis), and Paula and Rodger Riney Foundation to FP. Instituto de Salud Carlos III (ISCIII) (PI17/01346 and PI20/00152), co-funded by the ERDF *(A way to make Europe)*; FC-AECC (AIO16163636SAEZ); Gobierno de Navarra (0011-3638-2020-000011) co-funded by the ERDF through the *Operative Program 2014-2020 of Navarra* and Gobierno de Navarra (0011-3597-2020-000005) to BS. G.T. was supported by the Associazione Italiana per la Ricerca sul Cancro (AIRC), 5x1000 Program (Project #21267). IC was supported by Juan de la Cierva grant from Ministerio de Ciencia, Innovación y Universidades, Gobierno de España, Sara Borrell award from Instituto de Salud Carlos III (ISCIII) and Marie Curie grant (H2020-MSCA-IF-837491) from European Commission. ICA was supported by AECC predoctoral fellowship.

Data Availability

Public data was gathered from the following accesing numbers: samples GSM3674224, GSM3674225, GSM3674226, GSM3674227, GSM3674228, and GSM3674229 from GSE12842; samples GSM2915575, GSM2915576, and GSM2915577 from GSE108891. Additionally, we profiled an in-house mouse bone marrow niche sample; our raw sequencing data and expression-count will be deposited in public repositories before the publication of the manuscript. Currently, we are setting a GEO submission.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

REFERENCES

1. Wolkenhauer, O., Auffray, C., Jaster, R., Steinhoff, G. & Dammann, O. The road from systems biology to systems medicine. *Pediatric Research* vol. 73 502–507 (2013).
2. Gomez-Cabrero, D. *et al.* Systems Medicine: From molecular features and models to the clinic in COPD. *J. Transl. Med.* **12**, 1–11 (2014).
3. Giladi, A. *et al.* Single-cell characterization of haematopoietic progenitors and their trajectories in homeostasis and perturbed haematopoiesis. *Nat. Cell Biol. 2018 207* **20**, 836–846 (2018).
4. Nestorowa, S. *et al.* A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood* **128**, e20–e31 (2016).
5. Rodriguez-Fraticelli, A. E. *et al.* Clonal analysis of lineage fate in native haematopoiesis. *Nat. 2018 5537687* **553**, 212–216 (2018).
6. Velten, L. *et al.* Human haematopoietic stem cell lineage commitment is a continuous process. *Nat. Cell Biol. 2017 194* **19**, 271–281 (2017).
7. Karamitros, D. *et al.* Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. *Nat. Immunol. 2017 191* **19**, 85–97 (2017).
8. Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F. D. & Klein, A. M. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science (80-.).* **367**, (2020).
9. E, L. & B, G. From haematopoietic stem cells to complex differentiation landscapes. *Nature* **553**, 418–426 (2018).
10. Baryawno, N. *et al.* A Cellular Taxonomy of the Bone Marrow Stroma in Homeostasis and Leukemia. *Cell* **177**, 1915-1932.e16 (2019).
11. Tikhonova, A. N. *et al.* The bone marrow microenvironment at single-cell resolution. *Nature* **569**, 222–228 (2019).
12. Baccin, C. *et al.* Combined single-cell and spatial transcriptomics reveal the molecular, cellular and spatial bone marrow niche organization. *Nat. Cell Biol.* **22**, 38–48 (2020).
13. Wolock, S. L. *et al.* Mapping Distinct Bone Marrow Niche Populations and Their Differentiation Paths. *Cell Rep.* **28**, 302-311.e5 (2019).
14. Zhong, L. *et al.* Single cell transcriptomics identifies a unique adipose lineage cell population that regulates bone marrow environment. *Elife* **9**, 1–28 (2020).
15. Matsushita, Y. *et al.* A Wnt-mediated transformation of the bone marrow stromal cell identity orchestrates skeletal regeneration. *Nat. Commun.* **11**, (2020).
16. Traag, V. A., Waltman, L. & Eck, N. J. van. From Louvain to Leiden: guaranteeing well-

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

connected communities. *Sci. Reports 2019 91* **9**, 1–12 (2019).

17. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, P10008 (2008).
18. Tasic, B. *et al.* Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19**, 335–346 (2016).
19. EZ, M. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).
20. Rafii, S., Butler, J. M. & Ding, B. Sen. Angiocrine functions of organ-specific endothelial cells. *Nature* **529**, 316–325 (2016).
21. Itkin, T. *et al.* Distinct bone marrow blood vessels differentially regulate haematopoiesis. *Nature* **532**, 323–328 (2016).
22. Hooper, A. T. *et al.* Engraftment and Reconstitution of Hematopoiesis Is Dependent on VEGFR2-Mediated Regeneration of Sinusoidal Endothelial Cells. *Cell Stem Cell* **4**, (2009).
23. Tamura, K. *et al.* Insulin-like growth factor binding protein-7 (IGFBP7) blocks vascular endothelial cell growth factor (VEGF)-induced angiogenesis in human vascular endothelial cells. *Eur. J. Pharmacol.* **610**, 61–67 (2009).
24. Klóska, D. *et al.* Biliverdin reductase deficiency triggers an endothelial-to-mesenchymal transition in human endothelial cells. *Arch. Biochem. Biophys.* **678**, 108182 (2019).
25. Yau, J. W., Teoh, H. & Verma, S. Endothelial cell control of thrombosis. *BMC Cardiovascular Disorders* vol. 15 (2015).
26. Davis, G. E. & Senger, D. R. Endothelial extracellular matrix: Biosynthesis, remodeling, and functions during vascular morphogenesis and neovessel stabilization. *Circulation Research* vol. 97 1093–1107 (2005).
27. Testa, U., Labbaye, C., Castelli, G. & Pelosi, E. Oxidative stress and hypoxia in normal and leukemic stem cells. *Experimental Hematology* vol. 44 540–560 (2016).
28. Zhao, E. *et al.* Bone marrow and the control of immunity. *Cellular and Molecular Immunology* vol. 9 11–19 (2012).
29. Naito, H., Iba, T. & Takakura, N. Mechanisms of new blood-vessel formation and proliferative heterogeneity of endothelial cells. *Int. Immunol.* **32**, 295–305 (2021).
30. Quillon, A., Fromy, B. & Debret, R. Endothelium microenvironment sensing leading to nitric oxide mediated vasodilation: A review of nervous and biomechanical signals. *Nitric Oxide - Biology and Chemistry* vol. 45 20–26 (2015).
31. March, S., Hui, E. E., Underhill, G. H., Khetani, S. & Bhatia, S. N. Microenvironmental

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

regulation of the sinusoidal endothelial cell phenotype in vitro. *Hepatology* **50**, 920–928 (2009).

32. Nilius, B. & Droogmans, G. Ion channels and their functional role in vascular endothelium. *Physiological Reviews* vol. 81 1415–1459 (2001).
33. Barrientos, T. *et al.* Two novel members of the ABLIM protein family, ABLIM-2 and -3, associate with STARS and directly bind F-actin. *J. Biol. Chem.* **282**, 8393–8403 (2007).
34. Okamoto, T. & Suzuki, K. The role of gap junction-mediated endothelial cell-cell interaction in the crosstalk between inflammation and blood coagulation. *International Journal of Molecular Sciences* vol. 18 (2017).
35. Hamidouche, Z. *et al.* Priming integrin α5 promotes human mesenchymal stromal cell osteoblast differentiation and osteogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 18587–18591 (2009).
36. Hajjawi, M. O. R. *et al.* Mineralisation of collagen rich soft tissues and osteocyte lacunae in Enpp1-/- mice. *Bone* **69**, 139–147 (2014).
37. Mackenzie, N. C. W. *et al.* Altered bone development and an increase in FGF-23 expression in Enpp1 -/- mice. *PLoS One* **7**, 32177 (2012).
38. Spohn, G. *et al.* VKORC1 deficiency in mice causes early postnatal lethality due to severe bleeding. *Thromb. Haemost.* **101**, 1044–1050 (2009).
39. Price, P. A. Vitamin K-Dependent Formation of Bone Gla Protein (Osteocalcin) and Its Function. *Vitam. Horm.* **42**, 65–108 (1985).
40. Kenswil, K. J. G. *et al.* Characterization of Endothelial Cells Associated with Hematopoietic Niche Formation in Humans Identifies IL-33 As an Anabolic Factor. *Cell Rep.* **22**, 666–678 (2018).
41. Barreiro, O. *et al.* Endothelial tetraspanin microdomains regulate leukocyte firm adhesion during extravasation. *Blood* **105**, 2852–2861 (2005).
42. Hashemi, S. *et al.* Characterization of novel platelet and endothelial cell target antigens in a family with genetic susceptibility to autoimmunity. *Am. J. Hematol.* **38**, 293–303 (1991).
43. Quirici, N. *et al.* Isolation of bone marrow mesenchymal stem cells by anti-nerve growth factor receptor antibodies. *Exp. Hematol.* **30**, 783–791 (2002).
44. Ghazanfari, R., Li, H., Zacharaki, D., Lim, H. C. & Scheding, S. Human Non-Hematopoietic CD271pos/CD140alow/neg Bone Marrow Stroma Cells Fulfill Stringent Stem Cell Criteria in Serial Transplantations. *Stem Cells Dev.* **25**, 1652–1658 (2016).
45. Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol. 2019 202* **20**, 163–172 (2019).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

46. Echavarria, R. & Hussain, S. N. A. Regulation of angiopoietin-1/Tie-2 receptor signaling in endothelial cells by dual-specificity phosphatases 1, 4, and 5. *J. Am. Heart Assoc.* **2**, (2013).

47. Jia, J. *et al.* AP-1 transcription factor mediates VEGF-induced endothelial cell migration and proliferation. *Microvasc. Res.* **105**, 103–108 (2016).

48. Escudero-Esparza, A., Jiang, W. G. & Martin, T. A. Claudin-5 participates in the regulation of endothelial cell motility. *Mol. Cell. Biochem.* **362**, 71–85 (2012).

49. Linnemann, A. K., O'Geen, H., Keles, S., Farnham, P. J. & Bresnick, E. H. Genetic framework for GATA factor function in vascular biology. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 13641–13646 (2011).

50. Lee, J. S. *et al.* Klf2 Is an Essential Regulator of Vascular Hemodynamic Forces In Vivo. *Dev. Cell* **11**, 845–857 (2006).

51. Robert, A. W., Marcon, B. H., Dallagiovanna, B. & Shigunov, P. Adipogenesis, Osteogenesis, and Chondrogenesis of Human Mesenchymal Stem/Stromal Cells: A Comparative Transcriptome Approach. *Frontiers in Cell and Developmental Biology* vol. 8 (2020).

52. Amable, P. R., Teixeira, M. V. T., Carias, R. B. V., Granjeiro, J. M. & Borojevic, R. Gene expression and protein secretion during human mesenchymal cell differentiation into adipogenic cells. *BMC Cell Biol.* **15**, (2014).

53. Sureshbabu, A. *et al.* IGFBP5 induces cell adhesion, increases cell survival and inhibits cell migration in MCF-7 human breast cancer cells. *J. Cell Sci.* **125**, 1693–1705 (2012).

54. Kim, S. H. *et al.* Epigenetic regulation of IFITM1 expression in lipopolysaccharide-stimulated human mesenchymal stromal cells. *Stem Cell Res. Ther.* **11**, 1–12 (2020).

55. Guo, W. *et al.* In vivo immune interactions of multipotent stromal cells underlie their long-lasting pain-relieving effect. *Sci. Rep.* **7**, 1–2 (2017).

56. Kahai, S., Vary, C. P. H., Gao, Y. & Seth, A. Collagen, type V, α1 (COL5A1) is regulated by TGF-β in osteoblasts. *Matrix Biol.* **23**, 445–455 (2004).

57. Nakamura, S. *et al.* Negative feedback loop of bone resorption by NFATc1-dependent induction of Cadm1. *PLoS One* **12**, 1–14 (2017).

58. Yue, R., Shen, B. & Morrison, S. J. Clec11a/osteolectin is an osteogenic growth factor that promotes the maintenance of the adult skeleton. *Elife* **5**, 27 (2016).

59. Shen, B. *et al.* A mechanosensitive peri-arteriolar niche for osteogenesis and lymphopoiesis. *Nature* **591**, 438–444 (2021).

60. Huang, W., Yang, S., Shao, J. & Li, Y. P. Signaling and transcriptional regulation in

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

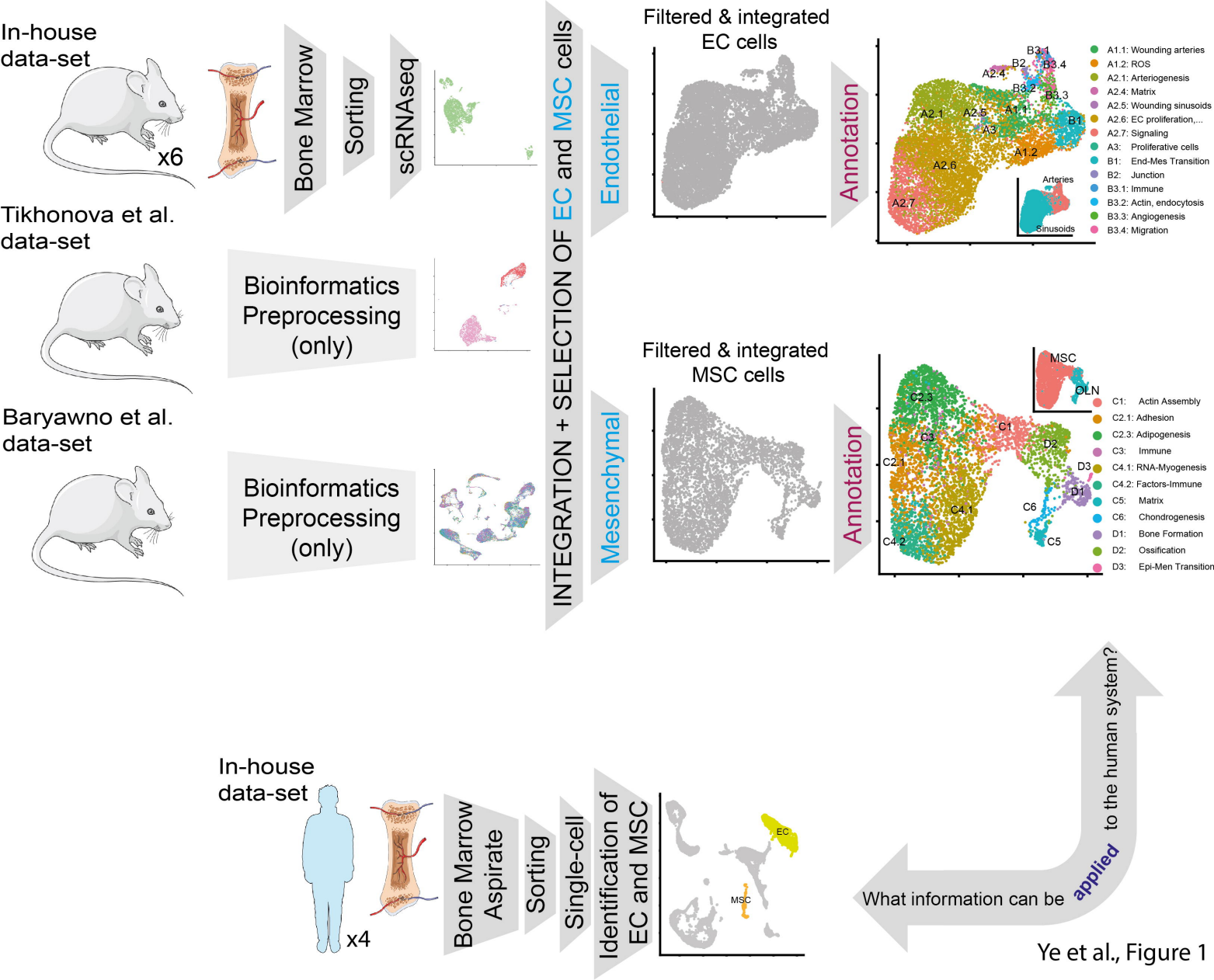
osteoblast commitment and differentiation. *Front. Biosci.* **12**, 3068–3092 (2007).

61. Argelaguet, R., Cuomo, A. S. E., Stegle, O. & Marioni, J. C. Computational principles and challenges in single-cell data integration. *Nat. Biotechnol. 2021* 1–14 (2021) doi:10.1038/s41587-021-00895-7.
62. Wu, K. E., Yost, K. E., Chang, H. Y. & Zou, J. BABEL enables cross-modality translation between multiomic profiles at single-cell resolution. doi:10.1073/pnas.2023070118/-/DCSupplemental.
63. Martinez-de-Morentin, X., Khan, S. A., Lehmann, R., Tegner, J. & Gomez-Cabrero, D. Machine Translation between paired Single Cell Multi Omics Data. *bioRxiv* 2021.01.27.428400 (2021) doi:10.1101/2021.01.27.428400.
64. Ashhurst, T. M. *et al.* Integration, exploration, and analysis of high-dimensional single-cell cytometry data using Spectre. *Cytom. Part A* (2021) doi:10.1002/CYTO.A.24350.
65. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).
66. Fang, S. *et al.* VEGF-C protects the integrity of the bone marrow perivascular niche in mice. *Blood* **136**, 1871–1883 (2020).
67. Chen, Q. *et al.* Apelin+ Endothelial Niche Cells Control Hematopoiesis and Mediate Vascular Regeneration after Myeloablative Injury. *Cell Stem Cell* **25**, 768-783.e6 (2019).
68. Kalucka, J. *et al.* Single-Cell Transcriptome Atlas of Murine Endothelial Cells. *Cell* **180**, 764-779.e20 (2020).
69. Severe, N. *et al.* Stress-Induced Changes in Bone Marrow Stromal Cell Populations Revealed through Single-Cell Protein Expression Mapping. *Cell Stem Cell* **25**, 570-583.e7 (2019).
70. Wang, Z. *et al.* Single-cell RNA sequencing deconvolutes the in vivo heterogeneity of human bone marrow-derived mesenchymal stem cells. *bioRxiv* 2020.04.06.027904 (2020) doi:10.1101/2020.04.06.027904.
71. Matthews, E. *et al.* Single cell RNA sequence analysis of human bone marrow samples reveals new targets for isolation of skeletal stem cells using DNA-coated gold nanoparticles. *bioRxiv* (2020) doi:10.1101/2020.06.17.156836.
72. Aoki, K. *et al.* Identification of CXCL12-abundant reticular cells in human adult bone marrow. *Br. J. Haematol.* **193**, 659–668 (2021).
73. Stumpf, P. S. *et al.* Transfer learning efficiently maps bone marrow cell types from mouse to human using single-cell RNA sequencing. *Commun. Biol.* **3**, (2020).
74. Xie, J. *et al.* Single-Cell Atlas Reveals Fatty Acid Metabolites Regulate the Functional

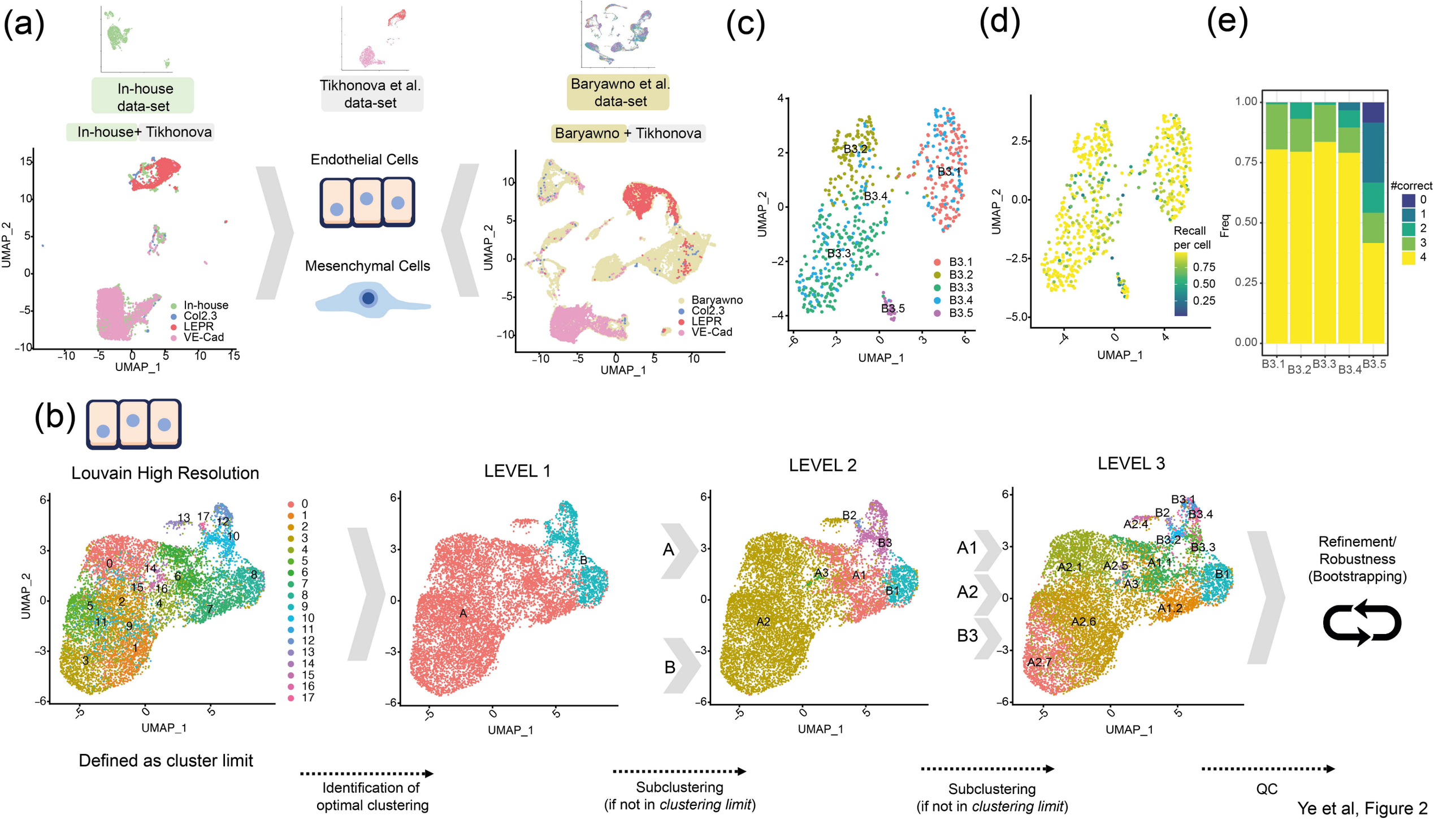
bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Heterogeneity of Mesenchymal Stem Cells. *Front. Cell Dev. Biol.* **9**, 1–15 (2021).

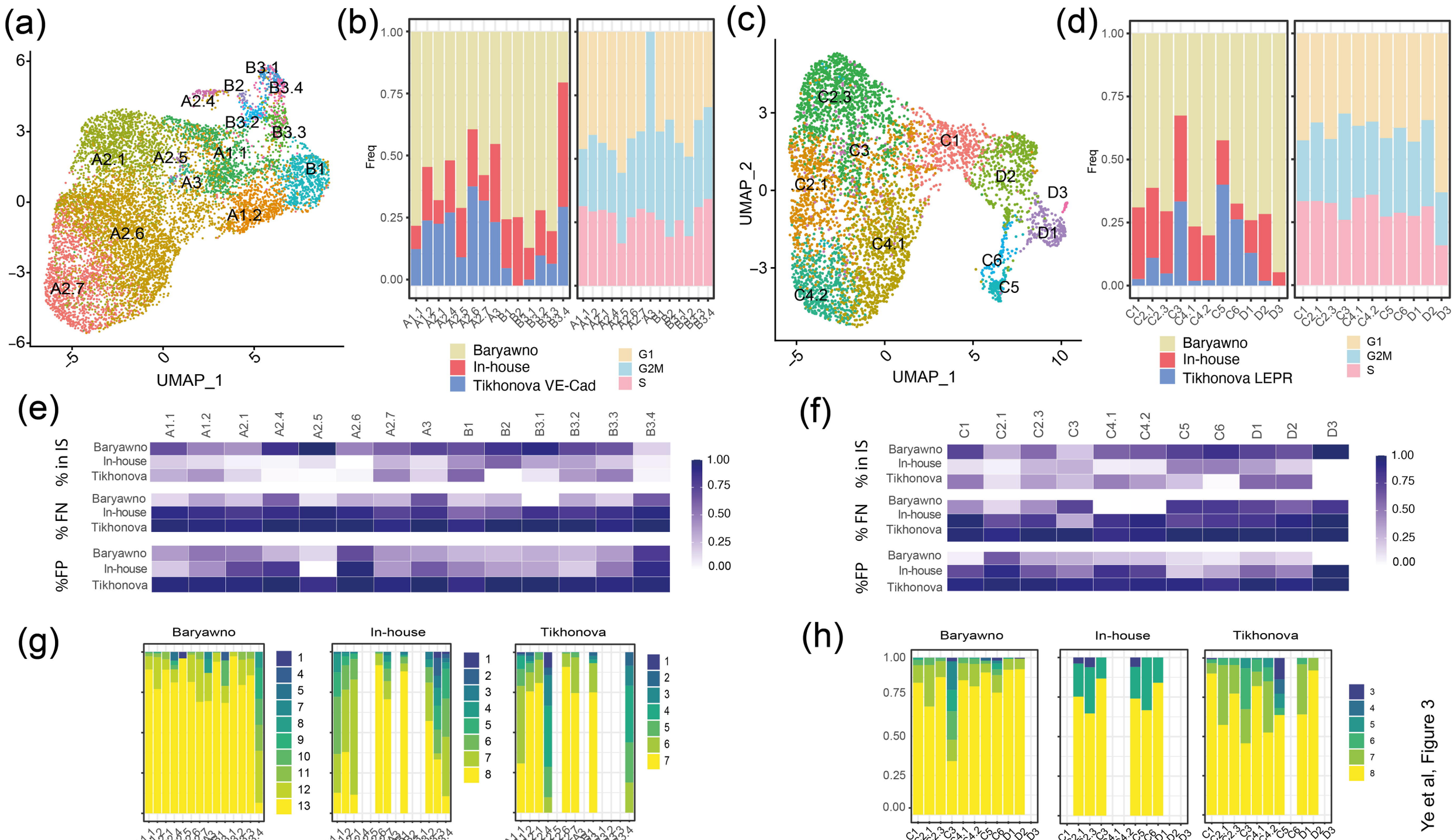
75. Leimkühler, N. B. *et al.* Heterogeneous bone-marrow stromal progenitors drive myelofibrosis via a druggable alarmin axis. *Cell Stem Cell* **28**, 637-652.e8 (2021).
76. Raaijmakers, M. H. G. P. *et al.* Bone progenitor dysfunction induces myelodysplasia and secondary leukaemia. *Nature* **464**, 852–857 (2010).
77. Zambetti, N. A. *et al.* Mesenchymal Inflammation Drives Genotoxic Stress in Hematopoietic Stem Cells and Predicts Disease Evolution in Human Pre-leukemia. *Cell Stem Cell* **19**, 613–627 (2016).
78. Tegnér, J. N. *et al.* Computational disease modeling – fact or fiction? *BMC Syst. Biol. 2009 31* **3**, 1–3 (2009).
79. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
80. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *bioRxiv* 576827 (2019) doi:10.1101/576827.
81. Chen, Y. C. *et al.* IKAP-Identifying K mAjor cell Population groups in single-cell RNA-sequencing analysis. *Gigascience* **8**, (2019).
82. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omi. A J. Integr. Biol.* **16**, 284–287 (2012).



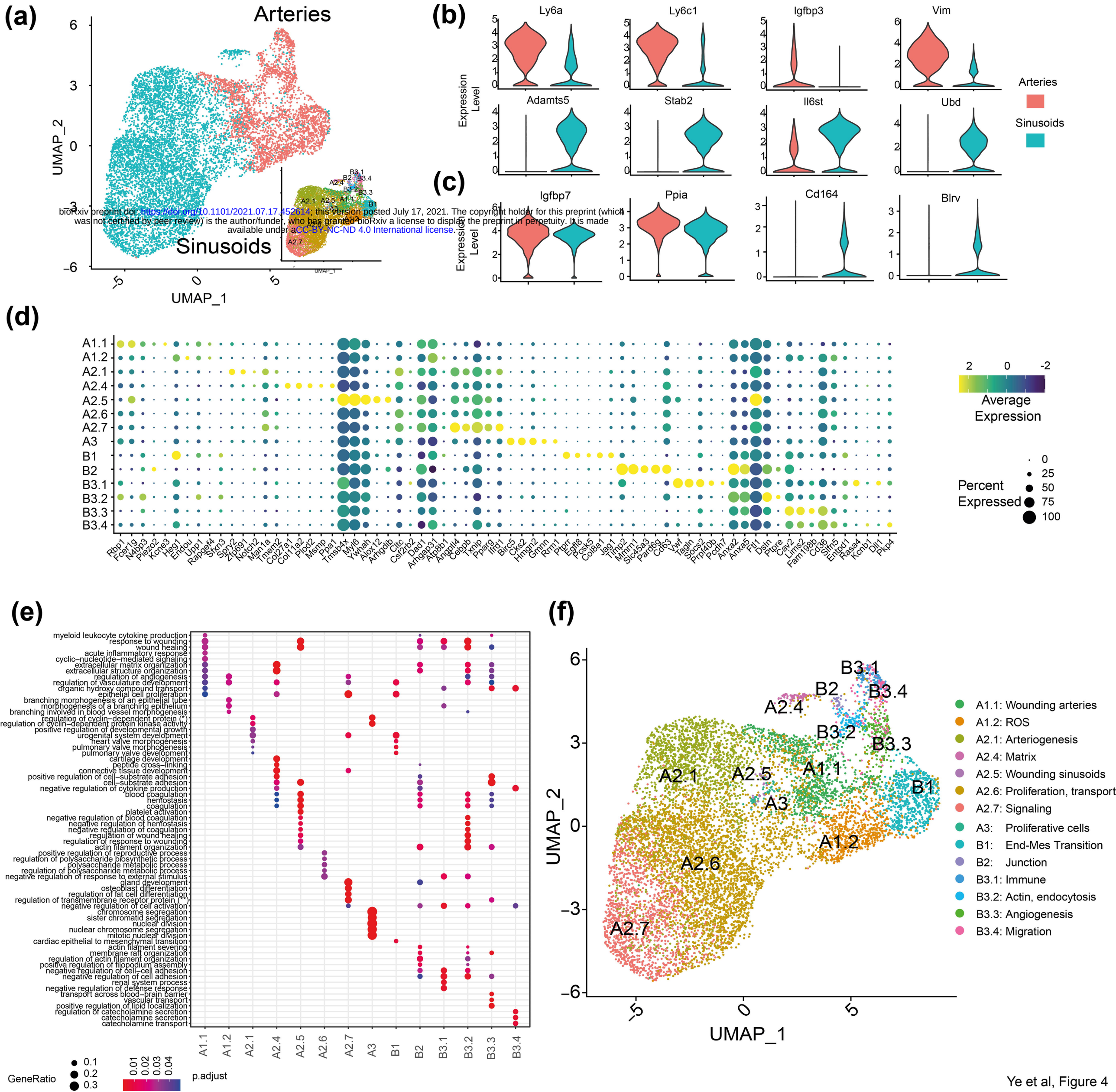
Ye et al., Figure 1



Ye et al, Figure 2

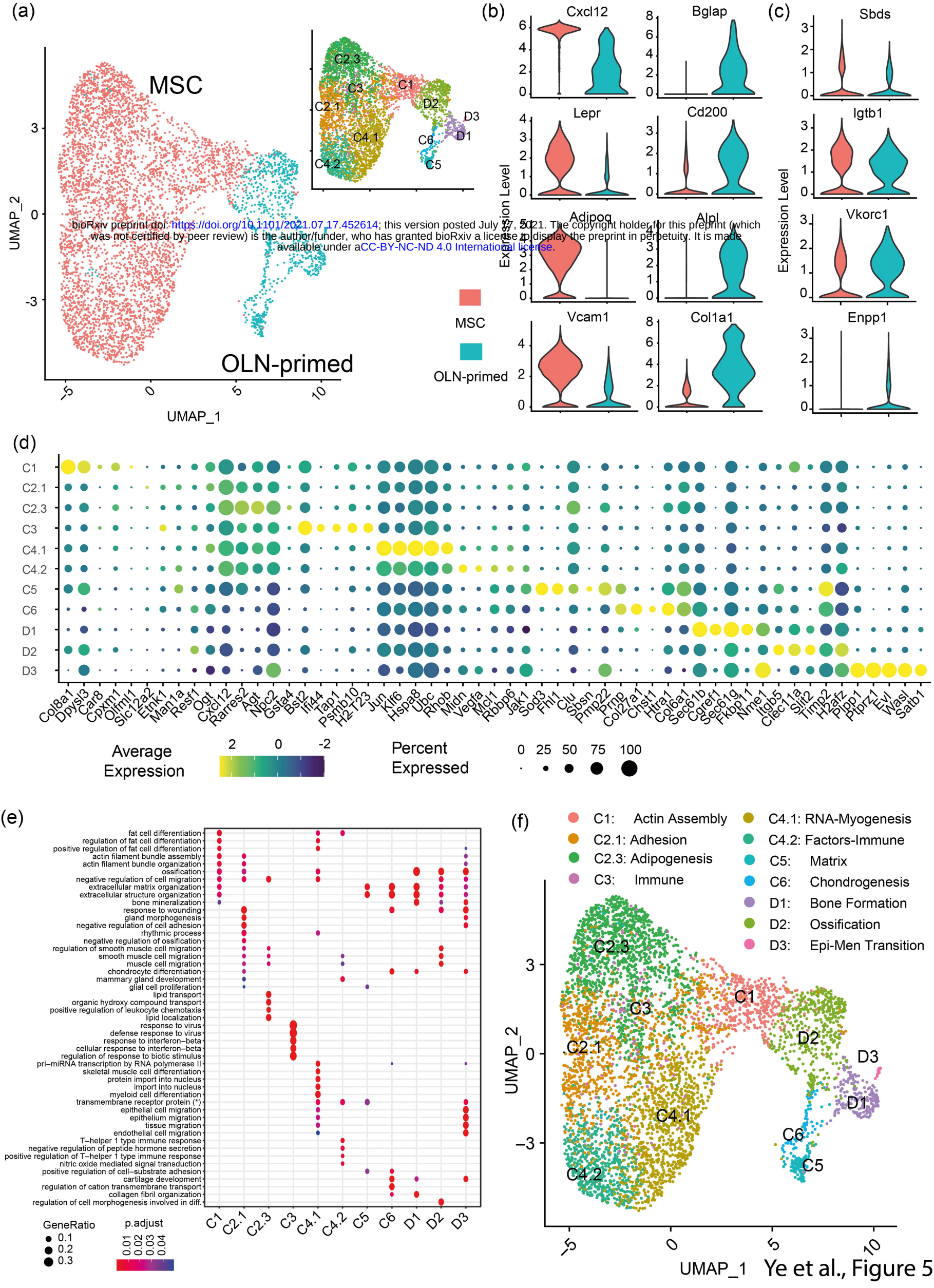


Ye et al, Figure 3



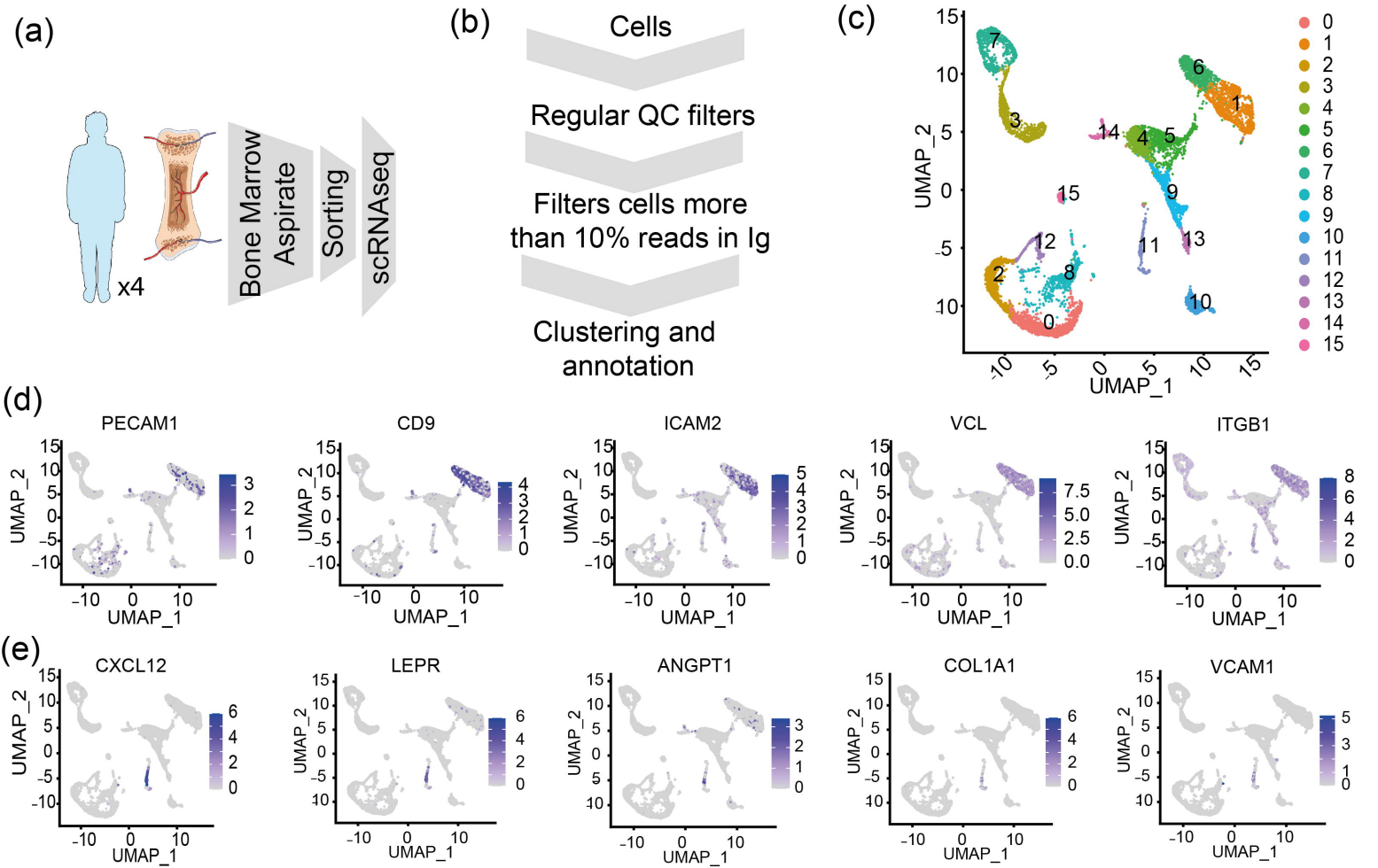
bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ye et al, Figure 4



bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452614; this version posted July 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ye et al., Figure 5



Ye et al, Figure 6

(a)

	Number of Mice Differentially expressed genes	Number of Genes Shared in human MVG	Percentage of Genes Shared in human MVG (%)	Enrichment Score	Human orthologs identified in Gene-sets used to label mice clusters
ARTERIES					
A1.1 Wounding	180	33	18.33	2.15	
A1.2 ROS	248	28	11.29	1.32	*CDKN1C, CLDN5*
B1 EndMes Transition	347	54	15.56	1.82	*PIK3R3, GLUL*
B2 Junction	109	21	19.27	2.26	*S100A10*
B3.1 Immune	134	22	16.42	1.92	*FGL2, IFITM2, APOE*
B3.2 Actin, Endocytosi:	193	33	17.10	2.00	*ANXA5, SERPINE1, CXCR4*
B3.3 Angiogenesis	201	32	15.92	1.87	*CAV2, GATA2, KLF2, RGCC*
B3.4 Migration	99	8	8.08	0.95	
SINUSOIDS					
A2.1 Arteriogenesis	174	30	17.24	2.02	*PIM1*
A2.4 Matrix	104	12	11.54	1.35	*CCDC80, LGALS3*
A2.5 Wounding Sin	54	10	18.52	1.95	
A2.6 EC prol, Transp	78	10	12.82	1.50	*DAB2*
A2.7Signaling	169	36	21.30	2.50	*DDIT4, JUN, CITED2, GADD45G, DUSP1, FOS, CLDN5*
A3 Proliferative	114	7	6.14	0.72	*UBE2C, TACC3*

(b)

	Number of Mice Differentially expressed genes	Number of Genes Shared in human MVG	Percentage of Genes Shared in human MVG (%)	Enrichment Score	Human orthologs identified in Gene-sets used to label mice clusters
MESENCHYMAL					
C1 Actin assembly	118	37	31.36	3.31	*LIMCH1, IGF1, WIF1, MARCKS*
C2.1 Adhesion	144	32	22.22	2.35	*IGFBP5, TNC*
C2.3 Adipogenesis	128	32	25.00	2.64	*NPC2, CXCL12, IGFBP3, PLTP, APOE, CYGB*
C3 Immune	229	24	10.48	1.11	*BST2, MNDAL, GBP2, ISG15, RSAD2*
C4.1 RNA-Myogenesis	209	74	35.41	3.74	*JUN, KLF2, JUNB, FOS, CITED2*
C4.2 Factors - Immune	232	79	34.05	3.60	*MIDN, IRF1, EGR1, RORA*
OSTEOLINEAGE PRIMED					
C5 Matrix	157	32	20.38	2.15	*CLU, LGALS3, COL3A1, CRISPLD2, NINJ1, COL6A1, FGFR1, FN1, CRISPLD2, CCDC80*
C6 Chondrogenesis	181	53	29.28	3.09	
D1 Bone Formation	159	39	24.53	2.59	*BGLAP3, COL5A1, CADM1*
D2 Ossification	214	75	35.05	3.70	*CLEC11A, SEMA7A, ALPL, MGLL, SPP1, ANK, SKIL, PDGFD, PTH1R, MEF2C, SERPINF1*
D3 Epi-Mes. Transition	101	23	22.77	2.41	