nature portfolio

https://doi.org/10.1038/s41586-024-07946-4

Supplementary information

Single-cell multi-omics map of human fetal blood in Down syndrome

In the format provided by the authors and unedited

Supplementary Information

Single-cell multi-omics map of human foetal blood in Down's Syndrome

Andrew R. Marderstein, Marco De Zuani, Rebecca Moeller, Jon Bezney, Evin M. Padhi, Shuo Wong, Tim H. H. Coorens, Yilin Xie, Haoliang Xue, Stephen B. Montgomery, Ana Cvejic

Supplementary Notes

Supplementary Figures

Supplementary Notes

Supplementary Note 1. Detecting *GATA1* mutations in foetal samples using RNA-seq

Samples were not tested experimentally for *GATA1* mutation. However, mutations in *GATA1* occur at the later stages of foetal development compared to the ones used in this study and are not observed before 20 pcw[1]. In line with this, the larger screening studies in newborns have shown that the incidence of *GATA1* mutations is around 4%[2,3]. Furthermore, Haasart et al.[4] did not observe a mutation in *GATA1* in any of the foetal stem and progenitor cells following WGS of HSPCs from 9 independent human foetuses gestational age week 12–17[4].

While our Down syndrome samples have not been tested for *GATA1* mutations, we set out to detect putative somatic variants from the single cell RNA sequencing data. There are, however, several limitations to this approach:

- The mutant copy of *GATA1* needs to be expressed to be detected in single cell RNA sequencing.
- Sequencing reads need to span the sites of the mutations. Since we are using the 10x Chromium 3' platform, the vast majority of reads will cover the 3' UTR and terminal exons of GATA1 but will not extend to the earlier exons. The majority of *GATA1* mutations that have been described in the Down syndrome context are truncating mutations in the first exons, towards the 5' end of the gene.
- The RNA editing machinery can cause sequence variation without an underlying DNA mutation.
- The coverage of *GATA1* per cell is very sparse.

These limitations notwithstanding, we used the variant calling framework based on deepSNV/Shearwater[5,6] to detect variants in the *GATA1* locus.

- Extract reads mapping to the *GATA1* locus and generate a per position pileup of nucleotide counts. These counts are aggregated pseudo-bulks per sample.
- Estimate a site-specific error profile per position estimating from the single-cell RNA sequencing data of the non-trisomy 21 samples.
- For every site, in every pseudobulk, use a likelihood ratio test to determine whether the observed variant depth and total depth is likely to come from the background error distribution. We correct for multiple hypothesis testing using the Benjamini-Hochberg approach.

With this approach, we detected seven variants in the *GATA1* locus that passed the significance threshold of $q<0.01$. Of these, two variants are inherited mutations, consistent with high variant allele frequency (VAFs) across all samples of the same individuals (chrX:48787974 G>A in TS21_6 and chrX:48791310 G>A in TS21_13). The remaining five variants are all confined to single sample and while likely somatic, are unlikely to impact the function of GATA1. Of the five variants, three are intronic, one is synonymous and one missense (G352A), which is predicted to be a benign variant. In addition, even though none of the variants listed here were previously identified as the consequences of RNA editing, we cannot definitively rule out the A>G and C>T may be caused by A>I and C>U RNA editing, respectively. Overall, none of the variants detected in the *GATA1* locus from the single-cell RNA sequencing data are likely to contribute to any preleukaemic phenotype.

Supplementary Note 2. Cellular taxonomy of Down's Syndrome and disomic foetal liver and bone marrow

Based on differential expression analysis and marker genes we manually annotated distinct blood populations in the foetal liver and BM. Within the hematopoietic progenitor compartment, we annotated clusters as HSC/MPPs (*CD34*, *SPINK2*, *MLLT3*), megakaryocyte-erythroid-mast progenitors (MEMPs - expressing *GATA1*, *KLF1*, *TESPA1*), granulocytic progenitors (GPs - expressing *MPO*, *AZU1*, *SPI1*, *LYZ*, *CD34*), as well as numerous mature blood cell types.

We identified clear transcriptional signatures of early and late erythroid cells (with graded expression of *AHSP*, *HBA1*, *ALAS2*, and *GYPA*), megakaryocytes (MKs - expressing *PLEK*, *ITGA2B*, and *GP9*), mast cells (*GATA2*, *CPA3* and *HDC*), monocyte precursors (*LYZ* and *SPI1*), plasmacytoid dendritic cells (pDCs - expressing *IL3RA*, *IRF8*, *CLEC4C*, and *JCHAIN*), and conventional DC2 (cDC2 - expressing *CD1C*, *CLEC4A*, *CLEC10A*), monocytes (*CD68, S100A9*, *MNDA*, *FCN1*), liver-specific Kupffer cells and femur-specific tolerogenic macrophages (expressing *CD163*, *MS4A7*, *CTSB*) and osteoclasts (*ACP5, MMP9, CTSK*).

In the lymphoid compartment, we identified NK progenitors and NK cells (expressing varying levels of *NKG7*, *PRF1*, and *GZMA*); *SPI1*+ NK cells that were exclusively present in the Ts21 BM and had gene expression signature of both NK cells (*NKG7*, *PRF1*, and *GZMA)* and myeloid cells (*SPI1, CD14*). The B cell lineage included pre-pro B cells, which showed expression of *IL7R*; pro-B cells, expressing higher levels of B-cell genes *VPREB1, MME*, *PAX5, RAG1, RAG2* compared with pre-pro B progenitors and cluster of mature B cells expressing high levels of *CD79B, IGHM* and decreased levels of *IGLL1* compared with pre-pro/pro-B cell clusters. In addition, we identified clusters of highly cycling cells expressing lineage specific genes and proliferation markers e.g., *MKI67*.

Supplementary Note 3. Cell populations in the foetal liver niche

Hepatocytes are the main and the most abundant parenchymal cell type in the liver. Cells in this cluster expressed well known markers of hepatocytes such as *ALB*, *AFP*[7]. Liver sinusoidal endothelial cells (LSECs) are the most abundant non-parenchymal cell type in the liver that form a barrier between the circulation, and the underlying hepatocytes and hepatic stellate cells within the space of Disse[8]. In line with this we identified a large cluster of LSEC cells that expressed typical endothelial markers (*KDR*, *CDH5*), lymphatic vessel endothelial hyaluronan receptor 1 *(LYVE1)*[9] but also scavenger receptors such as stabilin 1 and 2 (*STAB1* and *STAB2*)[10].

In addition, we identified non-LSEC endothelial cells i.e. arterial endothelial cells expressing *KDR*, *CDH5* but also calcitonin-receptor-like receptor (*CALCRL)* and *RAMP2* suggesting their sensitivity to adrenomedullin signalling[11]. RAMP2 is required to transport CALCRL to the plasma membrane where it acts as adrenomedullin receptor. LSECs maintain close interactions with hepatic stellate cells which are liver-specific mesenchymal cells.

We identified a population of hepatic stellate cells based on the expression of a vitamin A-associated transcript *RBP1*[12] as well as *DCN, COL1A1* and a cluster of activated hepatic stellate cells that expressed α smooth muscle actin (*ACTA2*), *TAGLN* and *MYL9*[7].

Supplementary Note 4. Cell populations in the bone marrow niche

In total we analysed 79,512 niche cells from Ts21 and 32,501 niche cells from disomic BM. We thereby identified, pericytes (*PDGFRB*, *TAGLN*, *ACTA2, RGS5, NOTCH3, MYH11, MCAM*), myofibroblasts (*PDGFRB, PDGFRA, TAGLN*, *ACTA2, NOTCH3)*, non-myelinating Schwann cells (*MPZ*, *NRXN1)*[13], *PDGFRA*-positive mesenchymal population that expressed high level of *CXCL12* therefore representing CXCL12-abundant reticular (CAR) cells as well as LepR+ CAR cells[14]. LepR+ CAR expressed *LEPR, CXCL12, PDGFRA, PDGFRB* and low level of nestin (*NES),* but not key HSC niche factors such as stem cell factor (*KITLG)* and Angiopoietin-1 (*ANGPT1*). disomic BM had additional clusters of mesenchymal cells (annotated here as 1 and 2).

In the osteo-lineage, we identified two clusters, osteoprogenitors (present in both disomic and Ts21) and osteoblasts (present only in disomic). Osteoprogenitors expressed *RUNX2,* the master regulator transcription factor (TF) known to control the commitment of mesenchymal stem cells (MSCs) to osteo-lineage and osterix (*SP7*), its downstream target TF[15]. Other osteogenic genes included osteopontin (*SPP1*) and integrin-bone sialoprotein (*IBSP*) which together are required for early and late osteo-lineage differentiation[16]. Cells in osteoprogenitor cluster co-expressed *FOXC1* which is essential TF for development of adipo-osteogenic progenitors and which up-regulates *CXCL12* and inhibits adipogenic processes in CAR cell progenitors[17]; *SOX9*, a TF essential for chondrocyte differentiation[18] and, aggrecan *(ACAN*) a chondrocyte marker gene[19]. Immature osteoblasts (present only in disomic BM) further upregulated *SP7* and bone gamma-carboxyglutamate protein (*BGLAP),* an osteoblast maturation marker[20]. In contrast, in Ts21 a smaller proportion of cells up-regulated osteo-lineage related genes such as *IBSP, BGLAP* and *SP7* (Supp. Figure 3).

Finally, we identified three main endothelial cell (EC) clusters all expressing *CDH5* and *KDR*. The endothelial population included *STAB2*, *STAB1*, *LYVE1, TSPAN7* - expressing liver sinusoidal ECs (LSECs)*,* arteriolar endothelial cells (high level of *CD34, KITLG*[21] and *CXCL12*), and transitioning endothelial cells that co-expressed endothelial as well as stromal markers (such as *DCN*, *COL1A1, PDGFRA)*. Specifically in disomic samples we found an additional cluster of immature endothelial cells which had a low expression of *CDH5*, *KDR and CD34.*

Supplementary Note 5. Osteo-lineage development is altered in Ts21 bone marrow

Osteo-lineage is an important regulator of the HSC's activity in the BM. Therefore, we examined if Ts21 affects differentiation and gene expression of osteoblasts in BM (Supp. Figure 4, Supp. Table 6-7). Interestingly, the osteo-lineage appeared perturbed in Ts21 with many of the key osteo-lineage genes (such as *BGLAP, IBSP*, or *SP7*) being sparsely expressed (Supp. Figure 3). To examine this further, we generated a differentiation trajectory within the BM niche and calculated dynamically expressed genes along the osteo-lineage trajectory[22–24] (from CXCL12-abundant reticular (CAR) cells to leptin receptor positive CAR (*LepR*+ CAR) cells to osteoprogenitors to osteoblasts, the latter being present only in disomic BM) (see Methods; Extended Data Figure 1a,b; Supp. Figure 5a). The 25 genes whose expression had the strongest positive correlation with osteoblast differentiation in disomic samples included *IBSP*, *RUNX2*, *SP7*, and *SPP1* (Extended Data Figure 1b), which are all related to "ossification". In contrast, genes related with "extracellular matrix organisation" and "epithelial cell proliferation" were negatively correlated. In Ts21, however, there was a weaker correlation between terminal state probability (a quantitative measure of osteo-lineage

differentiation) and gene expression (both positive and negative) (Extended Data Figure 1b) suggesting a less coordinated regulation of osteo-lineage related genes (Supp. Figure 5b) that partially impeded osteo-lineage development.

We further observed that a population of endothelial cells (here termed transitioning endothelial cells) formed a continuum that connected clusters from endothelial cells with the osteo-lineage (Extended Data Figure 1a,c; Supp. Figure 5c), suggesting that endothelial cells can go through endothelial-to-osteoblast transition in disomic foetuses. It has been recently reported that endothelial cells can give rise to BM stromal niche cells through endothelial to mesenchymal transition (EndoMT)[25]. Along the differentiation trajectory transitioning endothelial cells upregulated mesenchymal genes such as *PDGFRA*, *PDGFRB* and *PRRX1* (Supp. Figure 5d) as well as "ossification" related genes such as *IBSP*, *ALPL* and osteoblast cadherin *CDH11*, and downregulated "endothelium development" and "angiogenesis" genes such as *KDR* and *CDH5* (Extended Data Figure 1c, Supp. Figure 3, Supp. Figure 5c). In Ts21, however, transitioning endothelial cells showed less prominent upregulation of mesenchymal genes pointing again towards a defect in osteo-lineage differentiation in Ts21 BM (Extended Data Figure 1a, Supp. Figure 3, Supp. Figure 5d). Our data supports the notion that perturbed osteo-lineage differentiation in DS foetuses starts early during foetal development, and may underlie previous clinical observations such as shortened bone lengths and increased risk of osteoporosis[26–29].

Supplementary Note 6. Evaluating cell-type abundances following data-set integration and label transfer

Cell types exist on a continuum, and clear and distinct clusters on a UMAP are not expected during haematopoiesis. Thus, to test whether technical biases that arise from separate annotation will lead to confounding in differential cell-type abundance tests, we re-clustered our data using 1) integration (with both Harmony and scVi methods) and 2) reference-query mapping (for disomic samples with both Popescu et al. Nature 2019[30] and our Ts21 data). Using the new clusters, we repeated the cell type abundance comparison.

First, we integrated the full dataset using Harmony (Supp. Figure 8a, b) using 5000 highly variable genes to construct the PCA space. We estimated 107 principal components (PCs) as the optimal number of PCs based on random matrix theory (implemented in pegasus' pegasus.elbowplot function). Next, we applied 3 iterations of Harmony to integrate data across samples, correcting for sample/biological replicate, environment, and diseases. The output from Harmony was used as input for calculating the K-nearest neighbour graph and computing clusters with Leiden clustering (resolution = 3.5). We annotated the clusters based on their top marker genes. Overall, the Harmony integration-identified cell types expressed key marker genes across all the anticipated cell types (Supp. Figure 8c), validating accurate annotation of cell types.

Next, we compared the Harmony integration-identified cell type abundance analysis to our separate cell type abundance analysis shown in Figure 1c. The results were remarkably consistent between the two approaches (Supp. Figure 8d, e). The only difference being that, while the trend remained consistent, mast cells were not anymore significantly differentially abundant (FDR = 0.103).

Given that benchmarking studies[31] have shown that Harmony is not the preferred integration method for complex integration tasks such as our dataset, which combines two environments (liver and femur) with two highly different genetic backgrounds (disomic and trisomic), we repeated the integration and clustering using scVI (Supp. Figure 9). We ran scVI on the raw counts for the top 5000 highly variable genes with sample, environment, and disease used as batch covariates and a *max_epochs=400* (as the default falls to *max_epochs*=7 when there are 1.1 million cells such as our dataset). The embedding space from scVI was used as input for calculating the K-nearest neighbour graph, computing Leiden clusters, and finally annotating clusters.

The scVI-integrated proportions across samples correlated closely with separate dataset proportions (r = 0.957) (Supp. Figure 9d), a slight improvement compared to the Harmony-based integration (r = 0.949). Furthermore, the statistical results from our differential cell abundance analysis confirmed all results from the separate analysis, including a significant increase in mast cells found in Ts21 (Supp. Figure 9e).

Next, we performed reference query mapping from Popescu et al.[30] to our disomic liver dataset. This allowed us to compare the cell type abundances based on annotations from a published dataset to what we observed in both our original separate-dataset annotations. First, we performed the reference query mapping using scArches. To improve the reference query mapping accuracy and rejection rate, we considered reference mapping for only the 45,497 haematopoietic cells identified through our separate annotations for disomic liver. We combined all scRNA datasets into a single matrix and identified the top 5000 highly variable genes. We used scVI to embed both datasets into the same latent dimension using a variational autoencoder. Finally, we predicted cell-type labels using a KNN classifier (scHPL) with default settings (50 neighbours for knn, a rejection threshold of 0.5, and a match threshold of 0.25). We mapped the reference query cell type labels to the broad cell type groups and compared the group abundances to those found using the separate-dataset annotation labels. Given that the Ts21 liver cells were vastly different (due to presence of trisomy) and much larger than the Popescu et al.[30] dataset, we did not transfer labels for the Ts21 data for this analysis and perform a cell-type abundance test.

Overall, in the three disomic samples that were tested for cell-type abundance differences, we found that the new reference query labels correlated strongly with the previous annotations, helping validate our prior annotations (r=0.998; Supp. Figure 10a).

Next, we performed a reference query mapping from Ts21 liver to disomic liver and compared the Ts21-transferred annotations to our separate-dataset annotations. Compared to the original separate-dataset annotations, we found that the estimated cell-type abundances were highly correlated with abundances from the Ts21-based reference query mapping in the three stage-matched disomic liver samples (r=0.998; Supp. Figure 10a). Furthermore, we repeated the Ts21 *versus* disomic cell abundance analysis using the original Ts21 liver annotations and Ts21-label transfer for disomic liver annotations, and replicated our original separate-dataset differential abundance results. Overall, our results demonstrate that the cell-type annotations are consistent across all annotation strategies and the differential abundance results are robust.

Finally, we projected our disomic BM dataset onto the Jardine et al. data set[32]. Since we mostly used a CD235 depletion strategy, late erythroid cells were largely depleted from our samples as evident from the UMAP (Supp. Figure 10b). In this analysis, we filtered our disomic BM to only include blood cells, and then embedded our blood cell-only disomic BM and Jardine et al.[32] disomic BM into a single latent representation using scVI integration. Considering that cells from these two studies were collected from very different developmental stages and sorting strategies, our cells mapped to similar cell types in the Jardine et al.[32] data, reflecting that similar cell populations were captured in both studies.

Supplementary Note 7. Examining cell2location deconvolution performance using different scRNA-Seq references

To examine how scRNA-seq-identified cell types are spatially arranged in disomic and DS foetal liver, we performed 10X Visium on two consecutive 10 µm sections from 11 Ts21 foetal livers (matched to the original samples used for the scRNA-seq) and five disomic foetal livers (Supp. Table 9). We estimated the abundance of cell types within each Visium spot by leveraging cell-type expression profiles from the scRNA-seq[33] and quantified the colocalization of cell types residing in the same Visium spot (see Methods, Supp. Table 10).

To assess how the choice of reference affects model training and prediction of the cell abundances on Visium spots, we performed cell2location-based deconvolution on different reference scRNA-seq datasets and Ts21 and/or disomic Visium sections. First, we separately trained cell2location reference models on disomy and Ts21 datasets, estimated cell abundances on disomy or Ts21 sections using the corresponding models, and computed Pearson correlation for only disomy, and then only for Ts21 sections. (Supp. Figure 11). We matched cell types found in both references, and observed moderately strong correlations (median $r > 0.7$) for different cell types. Furthermore, there was no significant difference in the correlations. This suggested that the general concordance of cell-type abundances between Ts21 and disomy references are similar when applied to either the Ts21 or the disomy sections.

Next, we partitioned our disomy dataset into two subsets with comparable cell number (60,327 vs 50,344), trained two separate cell2location reference models on these two scRNA-Seq subsets, estimated cell abundances on disomy sections by these two models, and computed Pearson correlation on cell type abundances deconvoluted from the two references. We found that the two estimated abundances exhibited greater concordance relative to the disomy versus trisomy reference, suggesting that the context-specificity of the reference plays a significant role in the estimated abundances. We repeated the analysis, but using the Ts21 dataset (Supp. Figure 11). Again the two subsets had a comparable cell number (391,572 vs 388,727). There was even larger concordance, which supports our previous finding and also suggests that the number of cells captured in the reference (10 fold higher in Ts21 compared to disomy) improves the robustness of the results.

Finally, we used a publicly available foetal liver scRNA-Seq dataset, (113,063 cells)[30] as a reference to train cell2location, estimated cell abundances on our disomy Visium spots, and computed Pearson correlation on cell type abundances between the estimations from the public reference and from our (matched) disomy scRNA-seq reference (108,795 cells).

Our results suggest the estimation of the cell abundance is affected by the choice of reference. The results were most consistent when partitioning the reference into two sets, but differ to a greater extent when new references are used instead (such as using the trisomy reference for disomy sections and vice versa, or using a public reference instead).

Finally, we investigated whether the spatial organisation of cells in Ts21 *versus* disomic liver was affected. To this end, we quantified the colocalization of cell types residing in the same Visium spot. Our analysis showed that haematopoietic populations with close lineage relationships tended to co-localise in the same spot in foetal liver (Supp. Figure 12). For example, relative abundance of HSC/MPPs was highly correlated with megakaryocyte-erythroid-mast progenitors (MEMPs), and early erythroid cells; monocyte progenitors were correlated with monocytes and NK progenitors with NK cells (Supp. Figure 12a). In addition, Kupffer cells and liver sinusoidal endothelial cells (LSECs), known to reside in close vicinity in foetal liver[34], were also highly correlated in our Visium data (Supp. Figure 12).

Supplementary Note 8. Examining Ligand-Receptor Interactions in HSCs

We calculated, based on the co-expression of genes in different cell types clusters[35], which ligand-receptor pairs (L-Rs) are potentially involved in HSC/MPP cell interactions in Ts21 and the disomic foetal liver (see Methods). HSC/MPPs shared the most L-Rs with arterial endothelial cells and the least with hepatocytes (Supp. Figure 12b). We identified a number of well-known L-Rs involved in HSC/MPPs development and function which were present in both Ts21 and disomic samples, such as VEGFB-FLT1[36], ANGPT1-TEK[37], and NOTCH1-JAG1[38] (Supp. Figure 12c). Gene-set enrichment analysis of L-Rs expressed in HSC/MPPs (used for their interaction with arterial endothelial cells) revealed engagement of ERK1/2 cascade (Supp. Figure 12d) which is a common target of growth factors regulating haematopoiesis[39]. Activation of ERK1/2 pathways promotes HSC proliferation, a hallmark of foetal haematopoiesis, at the expense of quiescence. The interaction of HSCs with hepatocytes (both in Ts21 and disomic) included L-Rs relevant for iron ion uptake such as TFRC-TF (Supp. Figure 12c).

Supplementary Note 9. Evaluating the robustness of differential expression results to analysis decisions

Our main analysis did not account for sex and age or any genetic markers as covariates in analysis. We only pseudobulked cells on a per-sample level. Considering that analysis decisions such as sample-level pseudobulking or omission of covariates have the possibility of obscuring true differences, we first aimed to evaluate the robustness of our differential expression analysis results to these decisions. In all analyses, we ignored individual genetic polymorphisms and genetic ancestry, as we did not have access to individual genotype data. Furthermore, common regulatory variants should have similar frequency between trisomy 21 samples and disomic samples, as germline trisomy 21 is approximately independent from genome-wide SNP frequencies. Therefore, it is not expected that regulatory variants which majorly impact expression would cluster within Ts21 cases or disomic controls; ultimately, impacting statistical results as a confounder that hasn't been accounted for. We conducted the HSC Liver *versus* Femur differential expression analysis using four different statistical approaches for both disomic and Ts21 datasets (Supp. Figure 13).

1. Sample-level pseudobulks with the addition of sorting accounted for as a fixed effect (the final analysis)

2. Sample-level pseudobulks with the addition of sorting and age (PCW) accounted for as fixed effects
3. Foetus-level pseudobulks with no other covariates besides Liver vs Femur
4. Foetus-level pseudobulks with the addition of age (PCW) accounted for as a fixed effect

We additionally repeated analyses 2 and 4 using sex instead of age.

For age, we did not find any impact of these analysis decisions on the differential expression results, as the correlation between log-fold changes across all analyses had a Pearson correlation $r > 0.98$. Therefore, our differential expression results are consistent regardless of pseudobulking on sample or foetus. Furthermore, PCW had little to no impact on differential expression results (correlation between log-fold changes for all genes before and after including age as a covariate: r=0.999 in the disomic cells and r=1 in the Ts21 cells), (Supp. Figure 13a). To address whether inclusion of chr 21 genes masks any differences, we removed chr 21 and repeated our analyses in the Ts21 dataset. We observed similar results to previously, where $r > 0.98$ of the log-fold changes across non-chr 21 genes between all analyses. We further evaluated whether inclusion of non-differentially expressed genes masks any results, by further subsetting non-chr 21 genes to those with nominal $P < 0.05$ and then $FDR < 0.05$. We again observed strong correlations ($r > 0.97$), suggesting that inclusion of neither chr 21 genes nor non-differentially expressed genes mask differences. However, we noted that pseudobulk strategy (sample vs foetus level) had a larger influence on the top 200 differentially expressed genes, where log-fold changes had correlation $r = 0.90$ with strong visible concordance in scatterplots. Thus, we found our results to be statistically robust regardless of these analysis choices.

Next, we repeated the above analysis except for sex as a covariate in place of age. Using the correlation of log-fold changes across genes as a metric of the sensitivity of our results, we found very high correlations between adjusting for sex and not adjusting for sex ($r > 0.96$ for disomic HSCs and $r > 0.99$ for Ts21 HSCs). Furthermore, 99.1% of genes that were significantly differentially expressed ($FDR < 0.1$) in the original Ts21 HSC analysis were recovered after accounting for sex as a covariate ($FDR < 0.1$, Supp. Figure 13a). Therefore, sex did not have a large impact on differential expression results.

We additionally conducted an exploratory analysis of developmental age association with cell-type composition (Supp. Figure 13c). We first estimated cell type composition for the 8 major cell types detailed in Figure 1c across disomic and Ts21 samples. Next, we performed: (1) a correlation test between PCW and abundance, and (2) a non-parametric test (Kruskal-Wallis) to discern differences between PCW groups. For both tests, the 8 major cell types were not significantly associated with cell composition (nominal $P > 0.05$ for all tests). We visualised the association between cell type group and age, and saw potential outliers existing in PCW 12. Thus, we performed a Mann Whitney U test of PCW 12 abundances versus PCW 13 + 14 abundances. Again, we found no significant association (nominal $P > 0.05$ for all tests). As a result and in addition to the identical differential expression analysis results above, we did not consider PCW further in cell abundance results.

Furthermore, we note that the observed biological differences in Liver vs Femur are supported by similar differential expression analysis results in the Ts21 and disomic datasets. We

compared the correlation for all genes between the log-fold changes in the Ts21 dataset and in the disomic dataset (Supp. Figure 13d). Overall, while the correlation within a single dataset was $r > 0.98$, the correlation between the two datasets ranged from $0.87 < r < 0.93$. This high correlation reflects that not only are our results robust to statistical decisions, our results pick up on consistent biological differences between Liver and Femur environments that are present regardless of trisomy. Furthermore, the lower correlation reflects inherent biological differences that exist in Ts21 that do not exist in the disomic dataset, and vice versa. This motivates the analysis related to Figure 2a, since there is a need to distinguish purely environment-driven expression differences from Ts21-induced expression differences. Finally, we found that these high correlations are not expected given random chance; for example, the correlation between log-fold changes from a Liver vs Femur analysis in NK cells versus the analyses above HSCs/MPPs was minimal, with a Pearson correlation $r = 0.03$ ($P = 7.7 \times 10^{-6}$).

Finally, we assessed the impact of integration and cell type annotation on our differential expression analysis. We conducted the HSC Liver *versus* Femur differential expression analysis again using cluster annotations from a Harmony-based integration across all datasets. We observed highly similar results between the new annotations and the original analysis, with $r = 0.95$ in the disomic dataset and $r = 0.96$ in the Ts21 dataset when comparing log-fold changes across genes between the two analyses. This indicates the robustness of our findings, as differential expression results were consistent across the various annotation processes.

Supplementary Note 10. Evaluating the impact of ambient RNA on differential expression results

As we observed differentiation bias in our Ts21 samples, we tested the robustness of our differential expression results that we highlighted in Figure 2a-b after ambient RNA removal. Using CellBender, we reconstructed cell x gene expression matrices after ambient RNA removal in cell-containing droplets. We ran CellBender using default settings, but re-ran samples at half the learning rate if the CellBender HTML report recommended doing so (Suppl. Table 12). After calculating sample-level pseudobulks from the new post-ambient-removal scRNA matrix, we repeated the differential expression analysis of liver versus femur in Ts21 HSCs and then in disomy HSCs with limma (Suppl. Table 13).

First, we compared the new differential expression results (post-ambient) with the previous results that do not take into account ambient RNAs (pre-ambient). We subset to genes with nominal $P < 0.05$ in the pre-ambient Ts21 or disomy analyses, and calculated the correlation between log-fold changes between the pre-ambient and post-ambient analyses. Overall, we found a strong correlation of $r = 0.76$ in the Ts21 analysis and $r = 0.88$ in the disomy analysis. We note that the two analyses show similar concordance at more stringent significance thresholds; when limiting to genes with $FDR < 0.1$, we found an $r = 0.80$ in the Ts21 analysis and $r = 0.86$ in the disomy analysis. The strong correlations were not driven by chr 21, as subsequent removal of chr 21 genes and calculation of Pearson correlation between non-chr 21 genes in the Ts21 and disomy analyses resulted in $r = 0.79$ and $r = 0.87$ respectively.

Overall, between Ts21 liver and femur HSCs, we identified 967 differentially expressed genes in the pre-ambient analysis and 653 genes differentially expressed in the post-ambient analysis. While there are several examples of genes being differentially expressed in only one

of the two analyses, we observed strong association between a gene being differentially expressed at $P < 0.05$ in the pre-ambient analysis and also differentially expressed in the post-ambient analysis (OR > 9; $P < 10^{-234}$ in Ts21 and $P < 10^{-52}$ in disomy). To evaluate the robustness for several of the gene sets that we highlighted in the main text, we repeated the original analysis with an additional filter based on the post-ambient one. In the original analysis, we performed gene set enrichment analyses using "Ts21-induced" genes with logFC > 0 (upregulated in liver). In the new analysis, we required genes to also have post-ambient logFC > 0 and nominal $P < 0.05$ in Ts21 HSCs. In this way, any genes whose differential expression could be potentially explained by the ambient RNA are ignored in downstream enrichment analyses (leading to removal of 22% of Ts21-induced, liver-upregulated HSC genes). We repeated our gene set enrichment analyses with this new gene set, and found our highlighted gene sets to be robust to ambient RNA (*FDR* < 0.1). The top 5 TFs, highlighted in Figure 2b (GATA1, CEBPD, GATA2, FOXA1, TAF7), remained significant, as well as GO terms related to myeloid and erythrocyte differentiation, reactive oxidative stress, transcriptional and regulatory binding processes, and apoptosis.

Thus, in conclusion, we found our differential expression results to be minorly impacted by potential ambient RNA (given strong but imperfect correlations of log-fold changes and examples of genes no longer differentially expressed), but that ambient RNA contamination has a negligible impact on our key biological conclusions (as the highlighted pathways are retained after ambient RNA removal).

Supplementary Note 11. Evaluating cross-dependent dysregulation of chr21 genes versus other genes

To assess microenvironmental effects in DS, we examined gene expression across all cell types in foetal liver vs BM (Methods). We observed around 50% increase in expression of chr21 genes across all cell types in Ts21 samples compared to control (Extended Data Figure 2a,b; Supp. Figure 14; Supp. Table 11, 14). Changes in non-chr21 gene expression appeared to be cell type-dependent but also influenced by the environment from which cells were isolated (liver or BM) (Extended Data Figure 2a,b). We further observed an exponential relationship in dysregulation of genes on chr21 *versus* non-chr21 genes (in contrast to an approximate linear relationship of chr1 versus non-chr1 genes), suggesting that a "critical mass" of around 30%-50% dysregulated genes on chr21 was necessary before widespread (>10%) dysregulation occurred (Extended Data Figure 2c). This widespread dysregulation was not driven by individual genes, such as particular chromatin modifiers or TFs, indicating a potential contribution of feedback loops or synergistic regulation.

Since the cell types that show high dysregulation of chr21 genes are different in the two environments (liver and femur), our analysis demonstrates that the environment plays a role in driving this phenomena, however additional factors might also be relevant.

To explore this further, we first tested whether the overexpression of specific genes on chromosome 21 significantly correlates with the overall extent of dysregulation on both chr21 or other chromosomes. For each chr21 gene and for all cell types, we tested the Pearson correlation between the log2-fold change of gene expression (Ts21 vs disomic samples) and the % of DEG on (i) chr21 or (ii) non-chr21. After FDR correction, we did not find any significant correlation.

Next, we considered the alternative hypothesis that the expression level of genes on chr21 could influence their dysregulating effect. The rationale behind this idea is that important chromatin modifiers or TFs might consistently exhibit 50% overexpression in Ts21 cells, yet their impact would become critical only when their baseline expression is naturally sufficiently high. We tested this hypothesis for each chr21 gene and each cell type. We tested the Pearson correlation between the average cell-type-specific gene expression (from Ts21 or disomic samples) and the % of DEG on (i) chr21 or (ii) non-chr21. After FDR correction, we did not find any significant correlation.

Therefore, with our data in these analyses, we were unable to identify individual genes (such as particular chromatin modifiers or transcription factors) that drive genome-wide dysregulation of gene expression. We speculate that alternative hypotheses that we were unable to address include the existence of feedback loops that modulate chr21 gene expression (which would hinder analyses from a single time point) or synergistic regulation (where overexpressed genes lying in close proximity synergistically contribute exponential regulatory effects to other genes).

Supplementary Note 12. Differential expression across cell types

Using our scRNA-seq data, we further looked into Ts21 expression differences across cell types. Generally, we found that HSC differential expression (between Ts21 and disomy) was an effective predictor of differential expression in other cell types, as the log-fold changes of differential expression strongly correlated at non-chr21 genes across most blood types (Supp. Figure 14a; Suppl. Table 11, 14). Similarly, if genes are differentially expressed in HSCs, they were significantly more likely to be differentially expressed in nearly all other blood cell types as well.

Next, we evaluated pathways upregulated in Ts21 across different haematopoietic cell types. We looked at 12 blood cell types that were abundant in both datasets and for which we detected upregulated gene sets in Ts21 (using enrichR Biological Process gene set with genes which have FDR < 0.05 and log-fold change > 0 in the liver Ts21 vs disomy analyses). Overall, there was notable sharing of enriched gene sets across cell types in the dataset (Supp. Figure 14b; Suppl. Table 14), such that a gene set detected in HSCs was detected in a median of three other cell types as well, suggesting pervasive effects on haematopoiesis in Ts21 across differentiation stages. For example, gene sets related to hemostasis were found across the majority of cell types in the Ts21 liver. Many cell types in addition to HSCs also upregulated genes involved in the immune response, potentially contributing to dysregulated pathogen sensing, immune response initiation, and adaptive immunity in Ts21[40,41]; HSCs, mast cells, pDCs, pre pro B cells, and pro B cells upregulate genes involved in the pattern recognition receptor signalling pathway. Various MHC class II genes were also downregulated in Ts21 (*CD74*, *IGHM*, *HLA-DPB1*, among others), which reflects broad immune system dysregulation. Three gene sets were specifically found in HSCs: peptide hormone secretion, supramolecular fiber organisation, and epithelial cell apoptosis, which can be interconnected processes contributing to tissue homeostasis.

We also noted that, while expression differences can be overall similar across cell types, the magnitude of expression differences can vary. *GATA1* expression is upregulated generally across cell types, but the expression differences between Ts21 and Disomy cells decrease in later differentiation stages. In HSCs and MEMPs, *GATA1* has a greater than 1.5-log-fold

change in higher expression in Ts21; but in early erythroid cells and megakaryocytes, the log-fold change is below 0.5. In contrast, *RUNX1* expression differences increase after HSC differentiation, with non-significant differential expression in HSCs (LFC=0.22) but LFC = 0.74, 0.55, and 0.69 in MEMPs, early erythroid, and megakaryocytes respectively ($P < 4 \times 10\text{-}6$). We found that RUNX1 motifs were enriched in HSC peaks that were more accessible in Ts21, suggesting that it is likely that RUNX1 itself is not upregulated in Ts21 HSCs but its effect on downstream gene expression (through increased binding activity) may be increased in Ts21. In later progenitor stages, RUNX1 is DE suggesting misregulation in later differentiation stages. There are also genes with significantly increased expression in Ts21 across most cell types in the data, such as *SOD1*, which plays a crucial role in cellular defence against oxidative stress. *SOD1* is the most significantly differentially expressed gene in Ts21 HSCs (compared to Disomy), and it is also upregulated across most other blood types. However, it is not differentially expressed in stromal populations, with the exception being its differential expression in liver sinusoidal endothelial cells (LSECs).

Supplementary Note 13. Impact of chromosome 21 on differential expression analysis

We developed a simulation framework to carefully consider the statistical impact of the additional chromosome 21 on the differential expression analysis (DEA) results, specifically for genes not on chromosome 21. To do so, we simulated trisomy-like samples from disomic samples by increasing the reads counts of all genes on chromosome 21 by 50%. For example, if there are 100 reads counted for the chromosome 21 gene *ABCG1* in a pseudobulk HSC sample, we increased the reads by 50% such that there are 150 *ABCG1* reads in total (which was done for all chromosome 21 genes). We then assessed how the differential expression results of disomic samples were affected by this manipulation.

First, we performed DEA between the HSCs (pseudobulk) of two disomic foetuses (foetus #15633 and #15781). We then simulated a trisomy-like sample from one of the foetuses and ran DEA again. Finally, we compared the results by computing the median log fold change for genes on each chromosome (Extended Data Figure 2d), and comparing log fold change between the two analyses for all genes (Extended Data Figure 2e).

Overall, we found that, in the real data, the median log fold change on each chromosome was roughly centered around zero (Extended Data Figure 2d, left). When looking at the chromosome 21 spiked analysis, there was roughly a 50% increase on chromosome 21, while the median log-fold changes on other chromosomes were similar to the real data (Extended Data Figure 2d, right). As a further check of the impact of the chromosome 21 dosage effect on analysis, we found that the estimated log-fold change for each gene was nearly identical between both analyses, except for chromosome 21 where there is an induced increase in expression (Extended Data Figure 2e). Therefore, the increased expression on chromosome 21 has minimal impact on the differential expression results on other chromosomes.

Supplementary Note 14. Linking observed differences to altered methylation in Ts21

DNA methylation analysis of 196 DS and 439 non-DS newborn blood samples identified 1052 differentially methylated regions (DMRs) associated with DS[42]. They further demonstrated that differences in genome-wide DNA methylation found in neonatal Ts21 blood cells correlated with differences in gene expression observed in DS (N = 3) and non-DS (N = 3) foetal liver CD34+ blood population.

First, we contrasted our Ts21 vs disomy differential expression results with Muskens et al.[42] Despite differences in the cell type tested in Muskens et al.[42] compared to our study, we found strong concordance of results, such that genes that were differentially expressed (DE) in CD34+ foetal liver cells (i.e. HSPCs) in the Muskens et al.[42] datasets were highly correlated with genes that were DE in foetal liver HSCs from our dataset (r = 0.6 between the log-fold changes of the genes expressed in both datasets and significant in the Muskens et al.[42] analysis). Furthermore, the sign concordance of log-fold changes were quite strong (OR = 9.6; Fisher's exact test *P*-value = 3.8×10^{-31}), such that genes upregulated in Ts21 CD34+ in Muskens et al.[42] were typically also upregulated in Ts21 liver HSCs in our dataset (Supp. Figure 15a).

Muskens et al.[42] demonstrated that Ts21-associated DMRs at promoter or enhancer regions in neonatal blood are correlated with genes that were differentially expressed in Ts21 CD34+ cells (Supp. Figure 15b; Fisher's exact test *P*-value = 1.4×10^{-6}). Using our own differential expression results, we examined the relationship between DMRs and differential gene expression in Ts21 versus disomic foetal liver HSCs. We replicated the published results, showing that hypomethylation in promoter/enhancer regions is correlated with increased expression in Ts21 HSCs (Fisher's exact test *P*-value = 1.3×10^{-3}) (Supp. Figure 15c). Similar to Muskens et al.[42], we did not observe a statistically significant association between DMRs and expression outside of annotated promoter/enhancer regions.

The multiome data from our study enabled us to extend Muskens et al.[42] analysis by exploring the chromatin accessibility landscape. Thus, we tested whether the direction of differential methylation was associated with the direction of chromatin accessibility differences observed in Ts21 vs disomy foetal liver HSCs. We observed a strong trend between increased chromatin accessibility and hypomethylation in both promoter/enhancer regions (Fisher's exact test *P* = 5.4×10^{-27}) and outside promoter/enhancer regions (*P* = 6.1×10^{-6}) (where promoter/enhancer regions were annotated according to Muskens et al.[42]). This demonstrates the value of our resource in exploring the link between altered DNA methylation and accessibility in Ts21, particularly throughout the genome (Supp. Figure 15d).

Supplementary Note 15. Comparison of cell-type frequencies from 10X multiome versus scRNA-seq

Two out of three foetal Ts21 samples were used for both snRNA-Seq and scRNA-Seq. We find that cell type proportions of broad cell type classes are correlated (average Pearson *r* across 4 samples, made up of 2 foetuses with 2 replicates each, *r* = 0.50), but there are differences. We observed clear differences, as HSCs/Progenitor populations were identified at nearly twice the frequency in either of the 15582 multiome replicates (>47%) compared to the 15582 scRNA sample (<26%), and similar trends were found in foetus 15669 between multiome and scRNA. As a result, we did not observe significant differences in HSC abundances between disomy and Ts21 in the multiome data, but we did detect significant differences in erythroid, mast cell, and B cell population abundances. Thus, we observed similar frequencies of cell types in the multiome and scRNA-seq data, while acknowledging that the frequency of cell types in the multiome does not fully replicate the results observed in the scRNA-Seq experiment (Supp. Figure 18). There are a few rational explanations for this:

1. The multiome sample size was smaller than for scRNA-Seq, both in terms of the number of cells sequenced and the number of biological replicates, leading to lower statistical power.

2. The samples used in the scRNA-Seq and multiome experiments were partially matching - two out of three foetal Ts21 samples were used for both experiments, and none of the disomic samples used in multiome were used in scRNA-seq (Supp. Table 1, 2 and 8). This discrepancy in sample overlap, combined with relatively lower number of samples used for multiome, may potentially account for the observed differences in cell abundances between the two technologies.
3. Moreover, previous studies have demonstrated that different technologies are likely to produce differences in cell abundances. For example, a previous study by Massier et al.[43] shows clearly that two technologies can sometimes recover cell types at different abundances. While abundances of three of the four major cell classes within Massier et al.[43] were consistent between single-cell and single-nuclei technologies, vascular cells were found at nearly three-times higher abundance in single-nuclei data compared to single-cell data (Figure 1c from Massier et al.[43]).

Supplementary Note 16. TF regulation of different HSC subpopulations

We next identified TFs that regulate the expression of genes specific to each of the three HSC branches by simulating "computational knockouts" of each TF (see Methods). The key TFs regulating HSCs in branch 1 included *STAT2*, *ETV1/4*, and *IKZF1* (Supp. Table 17); in branch 2 (defined by topic 9): the *GATA* family, *ZNF460*, and *EWSR-FLI1* (FDR < 0.05; Extended Data Figure 4b, Supp. Table 17); and in branch 3 (defined by topic 8), there were no topics significantly driven by particular TFs (FDR < 0.05), but there was suggestive evidence of regulation by TFs such as *LHX4*, *CUX2*, *EMX2*, and the *HOX* family (e.g. *HOXA2* and *HOXB5*) (nominal P < 0.05), (Extended Data Figure 4b, Supp. Table 17). Several genes regulated by the *GATA* family[44] TFs relevant to branch 2 are known regulators of erythropoiesis and were specifically expressed in Ts21 cells (Extended Data Figure 4b). Thus overall, we observed systemic evidence, on both chromatin accessibility and transcriptional levels, that a subpopulation of Ts21 HSCs is "primed" to differentiate towards erythroid lineage.

Supplementary Note 17. Additional signalling is necessary to enact transcription of erythroid lineage related genes in Ts21 HSCs

We next asked if the expression of genes upregulated in each of the HSCs branches were best predicted by a regulatory potential (RP) model based on local (cis) accessibility alone (Local chromatin accessibility-Influenced Transcriptional Expression (LITE) model) or if their expression is better predicted by a model additionally considering genome-wide accessibility (Non-local chromatin accessibility-Influenced Transcriptional Expression (NITE) model). To address this question, we used MIRA to compute LITE and NITE models for each gene and then a likelihood ratio test to contrast the two models. We found that the top 300 genes upregulated in Ts21 HSCs had a much higher NITE score than those downregulated in Ts21 HSCs (Extended Data Figure 4c, d), indicating that expression of Ts21-specific genes cannot be fully modelled by local chromatin accessibility and that additional factors, such as additional signalling, are necessary to enact transcription of these genes. The genes with the highest NITE score were mainly expressed in branch 2 (Extended Data Figure 4e). One such example is *TFR2,* which is required for efficient development of red blood cells[45,46]. Based on the observed *cis* accessibility, *TFR2* was predicted to be expressed in both branch 2 and 3, but it was only expressed in branch 2, causing an overestimation of its expression in branch 3 by the LITE model (Extended Data Figure 4f). Its expression specifically in branch 2, despite much broader accessibility, suggests that *TFR2* expression in Ts21 HSCs requires additional signalling that does not primarily affect transcription via remodelling local accessibility. Genes

with the high NITE score were associated with “interleukin 2 signalling pathway”, which plays an important role in erythropoiesis[47], and “myelodysplastic syndrome” (Extended Data Figure 4g). This is in line with the model in which subpopulation of Ts21 HSCs attained a “primed” chromatin state which in response to signalling (i.e. additional factor(s)) activated the expression of erythroid lineage related genes.

Supplementary Note 18. Motif enrichment in differentially accessible regions of HSCs

To search for the key transcription factors (TFs) that contribute to haematopoietic differences in Ts21 compared to disomic foetal liver HSCs, we first determined which TF motifs are enriched in peaks with significantly higher accessibility and in peaks with significantly reduced accessibility in Ts21 HSCs using monaLisa[48]. To do so, we first stratified peaks with significant differential accessibility (Padj < .1) based on their direction (higher in Ts21 vs higher in disomy). Then, utilising the JASPAR2022 database[49] and monaLisa, we identified motifs that were enriched in Ts21 peaks compared to disomy (and vice versa) while accounting for peak length in the model. We employed a hypergeometric test for hypothesis testing and allowed zero or one occurrences of each motif per sequence to assess significance.

We found a greater number of motifs in peaks with higher accessibility in disomic HSCs, compared to peaks with higher accessibility in Ts21 HSCs. Our previous investigation revealed a higher number of peak-gene links in Ts21 HSCs, attributable to higher chromatin accessibility and gene expression in Ts21 samples. There was only one example (*RAPGEF2*[50]) where the peak-gene link was directionally discordant. Deficiency of *RAPGEF2* has been shown to lead to defective foetal liver erythropoiesis[50] and it was upregulated in the Ts21 HSCs ($P = 0.04$). Taken together, the analyses of peak-gene links and TF motifs suggest the following:

1) There appears to be less coordination between chromatin accessibility and gene expression in disomic HSCs overall, leading to an observation of more TF motifs in accessible chromatin regions that are not necessarily associated with gene expression;

2) HSCs in Ts21 exhibit lineage priming, resulting in a smaller number of TF motif occurrences that are, however, linked with gene expression. For instance, increased GATA and RUNX1 TF activities were noted in Ts21 peaks as well as MECOM which is often overexpressed in acute myeloid leukaemia[51]. In contrast, TFs such as BACH and NFE2, as well as the AP1 complex, exhibited increased activity in disomic peaks (Extended Data Figure 5a). Next, we aimed to explore the enrichment of TFs in either 1) promoters or 2) enhancers within Ts21 and disomic HSCs.

1) For the enrichment analysis of TFs in promoters, we utilised the Eukaryotic Promoter Database (EPD)[52], which offers a curated collection of experimentally validated promoters. Notably, EPD promoters were well-represented in our dataset, with approximately 60% of EPD promoters (13,414) overlapping with ATAC peaks in HSCs. In addition, a large portion of genes expressed in HSCs (80%, totaling 11,810) contained one or more EPD promoters; 95.8% (i.e. 11,309) of these expressed genes had at least one accessible EPD promoter (Extended Data Figure 5b). Crucially, genes linked with promoters displaying differential accessibility between Ts21 and disomic HSCs were also enriched for differentially expressed (DE) genes in the separate scRNA-seq dataset (Extended Data Figure 5c), validating the utility of EPD promoters for TF motif enrichment analysis.

To perform a promoter centric motif analysis, we followed a similar procedure as described above where we stratified promoters based on the direction of accessibility. Given the low

overall number of differentially accessible promoters we also included 4000 randomly sampled non-differentially accessible sequences to boost statistical power in our analyses. In the motif enrichment analysis within EPD promoters overlapping with ATAC peaks, we found that GATA family members and FLI1 were more abundant in promoters exhibiting increased accessibility in Ts21 compared to disomic HSCs. However, while GATA TF motifs were present across both non-promoters and promoters, the signal was much stronger in non-promoters (Supp. Table 18; Figure 4a). The GO term enrichment analysis of genes linked to promoters with greater accessibility in Ts21 revealed associations with interferon (IFN) alpha and beta, as well as IFN-mediated signalling pathways (Extended Data Figure 5d). Notably, IFN has been shown to stimulate the exit of HSCs from quiescence and their activation in the cell cycle *in vivo*.

Additionally, we found varied utilisation of AP1 motifs among DA promoters (Figure 4b). AP-1 is a generic name for different sets of homo- or heterodimers made up of members of the JUN, FOS, ATF, and MAF families and is considered to be a master integrator of extracellular signals, thus allowing cells to adapt to changes in their environment. Depending on cell type specific co-factors and chromatin context *in vivo*[53,54], AP-1 dimers bind to either 12-O-tetradecanoylphorbol-13-acetate (TPA)-responsive element (TRE motif) or cAMP-responsive element (CRE). Our analysis revealed that the TRE motif was more prevalent in DA promoters with greater accessibility in disomic HSCs, whereas the CRE motif was more frequent in DA promoters with greater accessibility in Ts21 (Figure 4b). The GO term enrichment analysis of genes linked with Ts21 DA promoters containing CRE motifs revealed terms related to the autophagy of mitochondria and mitochondrial organization (Extended Data Figure 5e). This is in line with our previous analysis of cycling HSCs in Ts21 liver that showed upregulation of genes involved in mitochondrial dysfunction and macroautophagy as well as findings from previous studies that mitochondrial dysfunction contributes to the number of DS-associated pathologies. Disomic DA promoters with TRE motifs exhibited enrichment in granulocyte-macrophage colony-stimulating factor (GM-SCF), a critical regulator of myelopoiesis which was impaired in Ts21 (Extended Data Figure 5f).

2) To examine TF motif enrichment in enhancers in Ts21 and disomic HSCs, we used the activity-by-contact (ABC) model[55] to prioritize functional enhancers. ABC predicts which enhancers regulate which genes based on read counts of ATAC-seq and a cell-type-averaged Hi-C map, and tests 5 Mb cis windows of candidate elements and genes on the same chromosome. ABC was performed on pseudobulk fragment files collated from all HSCs, and we removed EPD promoters from the ABC peak set to identify the final set of putatively functional enhancers. Around 80% of expressed genes in HSCs had at least one ABC enhancer (Extended Data Figure 6a). Specifically, around 1700 genes had at least one differentially accessible (DA) enhancer and around 500 genes had more than one DA enhancer (Extended Data Figure 6b). To confirm the relevance of these enhancers in driving differential expression of HSC genes in Ts21 compared to control, we calculated enrichment by performing two Fisher exact tests with the differential expression results from the scRNA-seq dataset (Extended Data Figure 6c):

1. Assessing overlap between genes that had at least one differentially accessible ABC enhancer and those that are differentially expressed.
2. Assessing overlap between genes with differentially accessible ABC enhancers and those that are differentially expressed.

The strong enrichment confirmed that identified ABC enhancers are indeed relevant in regulating genes that are differentially expressed in Ts21 compared to disomic HSCs. Furthermore, ABC enhancers were significantly more likely to be DA (FDR < 0.1) compared to promoters (18.1% ABC enhancers versus 9.1% promoters were DA; binomial test *P*-value = 10^{-219}), and we overall identified 2.1-times the number of DA ABC enhancers (2650) compared to DA promoters (1281). Interestingly, we found that DA enhancers exhibited much larger accessibility changes compared to DA promoters (Extended Data Figure 6d), particularly at enhancers with greater accessibility in trisomy.

We next examined which TF motifs are enriched in DA ABC enhancers. To accomplish this, we again conducted a stratified motif analysis on significant DA ABC enhancers, stratifying them based on the direction of differential accessibility. Analogous to promoters, we observed a strong enrichment of GATA family transcription factors and MECOM in peaks exhibiting higher accessibility in Ts21 compared to disomy (Figure 4a). Conversely, in peaks with higher accessibility in disomy, we observed significant enrichment of various transcription factors, including those from the ETS family and IRF/STAT families and NFE2, as well as transcription factors on chromosome 21 such as BACH1, ERG, and ETS2. Furthermore, we identified enrichment of TRE AP-1 motifs in disomy-biased peaks; however, we do not identify the CRE AP-1 motif enriched in trisomy biased ABC enhancers suggesting that the CRE motif is specifically employed by Ts21 DA promoters (Figure 4b).

Finally, we aimed to rank the contribution of each TF motif to differential accessibility of promoters and enhancers, while accounting for the differential TF usage between promoters and enhancers. We scanned all differentially accessible ABC enhancers and EPD promoters (2039 sequences total) for significant motif matches using motifmatchR for each motif in JASPAR2022 database. For each motif, we then fit the following model: *LogFC ATAC ~ motif containing + promoter + motif containing:promoter*, where motif containing and promoter are binary covariates indicating respectively whether the tested peak is an enhancer or promoter and whether it contains a specific motif. We report all TF where the model p-values were FDR < 0.1. Based on multiple-R^2, we identified that the GATA family motifs were the most predictive of differential accessibility of all regulatory elements, explaining more than 15% of the variance in differential accessibility (as estimated by the proportion of variance explained in the logFC value). Additionally, we found that GATA motifs have a significant interaction with enhancer versus promoter usage, which suggested that GATA TFs drive much stronger differential accessibility in Ts21 enhancers (Extended Data Figure 6e).
For example, the GATA motif was the top TF that affected accessibility. The GATA motif particularly influenced the degree of accessibility changes at Ts21 ABC enhancers (as estimated by logFC) (Extended Data Figure 6f). This demonstrates that GATA TFs have an outsized role in altering accessibility in Ts21 compared to other regulatory TFs, and this effect was much stronger in enhancers than promoters.

Finally, we explicitly looked at chromosome 21 TFs and their motif enrichments in our promoters and enhancers. For ABC enhancers, we find that BACH1, ERG, GABPA, and ETS2 are all enriched in enhancers that have significantly higher accessibility in disomy, where as for promoters we observe that BACH1 is also enriched in disomy biased promoters but ETS2 is enriched in trisomy biased promoters.

In summary, TF motifs such as GATA, RUNX1, and MECOM are enriched in higher accessibility Ts21 foetal liver HSC regions, while reduced accessibility regions in Ts21 are enriched for a diversity of motifs. This suggests that a small number of lineage TFs become more active in Ts21, increase coordination between chromatin accessibility and gene expression, and drive lineage bias. Notably, differential TF usage between Ts21 and disomic HSCs was distinct between promoters and enhancers. For example, analysis of TF enrichment in promoters revealed differential utilisation of AP-1 motifs between Ts21 and disomic HSCs. Ts21 HSCs exhibited enrichment of CRE motifs in differentially accessible (DA) promoters, involved in mitochondrial dysfunction and autophagy, while disomic HSCs showed enrichment in TRE motifs, which were associated with granulopoiesis. At enhancers, the GATA family of TFs played a significant role in altering enhancer accessibility; this was also observed at promoters, but to a smaller extent.

Therefore, overall, our findings suggest a distinct regulatory landscape in Ts21 HSCs, characterised by altered TF motif enrichment and chromatin accessibility patterns, in promoters and enhancers, which may contribute to the haematopoietic differences observed in Down Syndrome.

Supplementary Note 19. Analysis of publicly-catalogued somatic mutations

We contrasted whether publicly-catalogued Ts21 or disomic intronic mutations from Hasaart et al.[4] were more likely to reside within cis-regulatory elements (CREs) of Ts21 HSC-expressed genes. The result showed a greater enrichment of Ts21 mutations compared to disomic mutations within the CREs of Ts21 HSC-expressed genes. We indicated that our finding is in line with the hypothesis that “altered chromatin accessibility landscape in Ts21 HSCs allows differential mutation acquisition in regulatory regions of genes”.

To increase evidence for the aforementioned hypothesis, we repeated our analysis of CREs within Ts21 HSC-expressed genes, by using Ts21 cycling HSC-expressed genes or disomy HSC-expressed genes. Using Ts21 HSC-expressed and cycling HSC-expressed genes, the Ts21 intronic mutations are significantly more likely to be found in cis-regulatory elements compared to disomy intronic mutations ($P < 0.005$). In disomy HSC-expressed genes, Ts21 intronic mutations were not overly concentrated in cis-regulatory elements compared to disomy intronic mutations ($P = 0.096$) (Figure 4e). Further, the percentage of intronic somatic mutations overlapping CREs for disomy is greatest in disomy HSC-expressed genes. For Ts21, it is greatest in Ts21 cycling HSC-expressed genes, followed by less-cycling HSC-expressed genes, and then finally disomy (Figure 4e). Thus, the association between Ts21 mutations and CREs relative to disomic mutations suggest that the regulatory landscape impacts where somatic mutations fall.

As we observed different associations depending on whether the gene was expressed in Ts21 versus disomy HSCs, we directly evaluated differentially expressed genes (DEGs) and their enrichment of Ts21 versus disomy somatic mutations. We performed the analysis above with DEGs from (i) disomy vs Ts21 HSC, and (ii) Ts21 cycling vs less-cycling HSCs; overall, assessing whether Ts21 intronic mutations are more likely to be in cis regulatory elements of differentially expressed genes (compared to disomy somatics). We evaluated the robustness of our results across different *P*-value thresholds for DEGs. Overall, we found that Ts21 intronic somatics were much more likely than disomy somatics to fall within CREs of DEGs for Ts21 vs disomy DEGs and for cycling vs less-cycling HSCs. This trend prevailed across DEG

significance thresholds, and the enrichments strengthened with a more stringent significance threshold (Extended Data Figure 7b, c). The higher concentration of Ts21 mutations within the regulatory regions of differentially expressed genes indicate either that Ts21 mutations are more likely to appear because of a changed regulatory landscape (i.e. more accessible chromatin regions), or the possibility that the samples from our study harbour regulatory somatics in these same regions, which cause differential expression. However, further evaluating this would require whole genome and ATAC sequencing in the same and a considerably larger set of samples.

Overall, an altered regulatory landscape is associated with Ts21 somatic mutation acquisition, as evidenced by:
a) Compared to control/disomy mutations, Ts21 intronic mutations are enriched in the cis regulatory elements of Ts21-expressed genes but not disomy genes.
b) CREs of genes differentially expressed in Ts21 v disomy and Ts21 cycling v less cycling also show enrichment of Ts21 intronic mutations relative to disomy intronic mutations.

Supplementary Note 20. Enrichment of leukaemia and blood cell traits GWAS SNPs in multiome peaks

Genome-wide association studies (GWAS) have identified a myriad of genetic variants associated with a range of blood cell traits and diseases, many of which are non-coding and impact causal genes through *cis*-regulatory elements. We used variants associated with blood cell variation because many have been implicated in leukaemia risk despite limited knowledge of the genetic architecture of leukaemia risk[56].

We evaluated the overall accessibility of regulatory elements important to blood differentiation across HSC branches. By tagging key non-coding elements using fine-mapped GWAS loci, we identified individual cells with increased accessibility for red blood cell (RBC), white blood cell (WBC), and lymphoid cell (Lymph) traits[57,58]. As expected, we found significant enrichment of GWAS SNPs in open chromatin regions for Lymph counts, WBC counts, and RBC counts within their corresponding cell types (Supp. Figure 19a), thus confirming the identity of annotated cell types and the utility of the approach.

Notably, integration of the Multiome dataset with the scRNA data also informed likely causal genes in blood cell development at GWAS-harbouring enhancers. We found that the *C11orf21* intronic peak carrying the fine-mapped RBC SNP rs2077078 had increased accessibility in Ts21 HSCs (P = 1.4x10^{-3}) and its accessibility was associated with both *C11orf21* and *TSPAN32* expression (P < 0.03) (Extended Data Figure 8a-right). However, based on our differential expression analysis of Ts21 and disomic liver HSCs in the scRNA dataset, only *TSPAN32* was dysregulated in Ts21 HSCs (P = 1.4x10^{-3}) (Extended Data Figure 8a, b). This suggests that *TSPAN32* can contribute to the observed lineage bias and that *TSPAN32* is the likely causal gene of the rs7775698-linked enhancer and its role in red blood cell development. Another example is intronic variant rs2075672 which lies in a peak primarily accessible in Ts21 HSCs (P = 5.0x10^{-3}) and its accessibility correlates with *TFR2* expression.

To investigate the effect of overexpression of *TFR2* in erythroid differentiation, we used the HUDEP-2 cell line. This cell line is derived from CD34+ umbilical cord blood cells and expresses erythroid-specific markers such as CD235, CD71 and CD36. When cultured in differentiation medium, HUDEP-2 cells start to differentiate and upregulate CD235, a marker

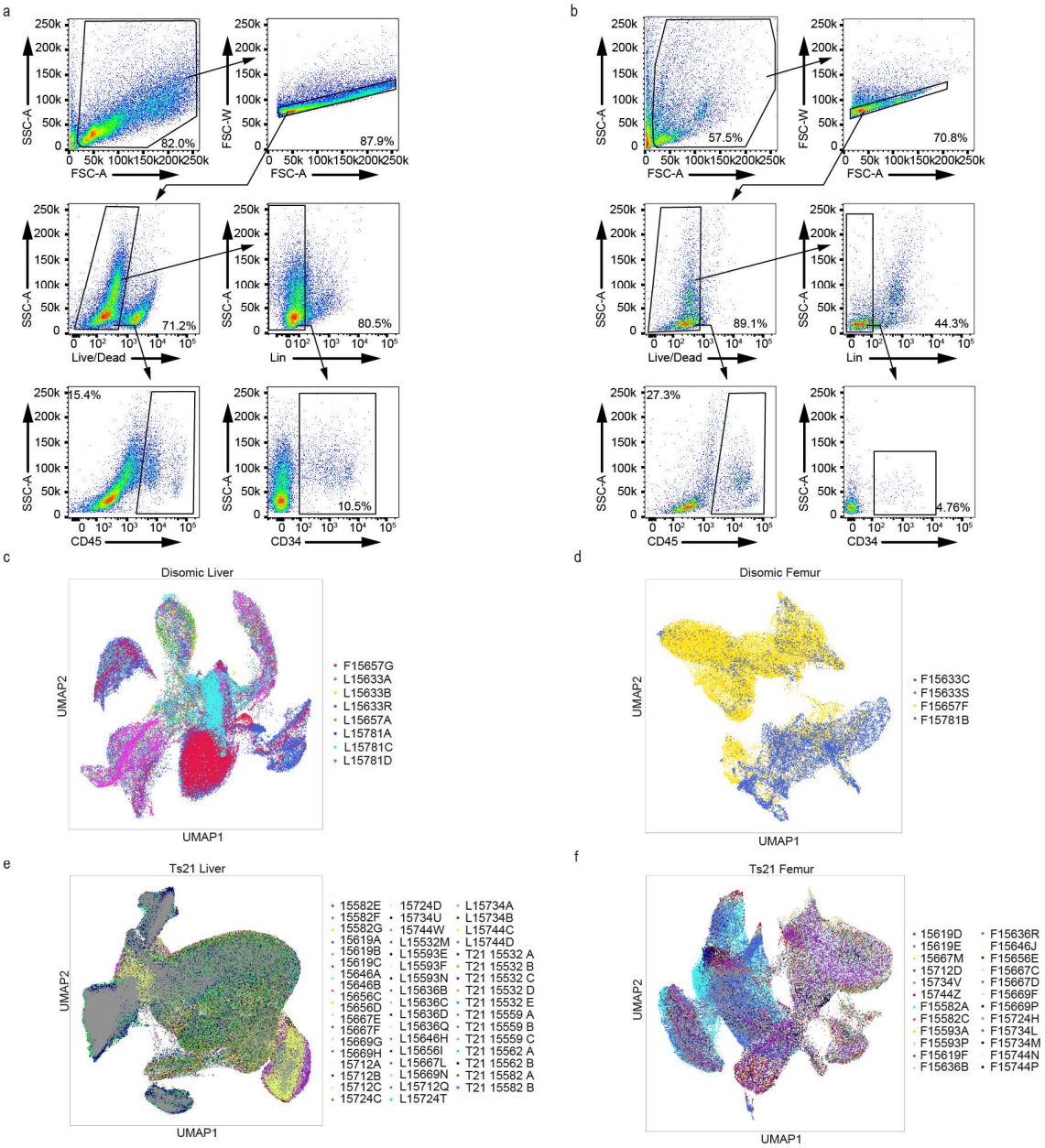
of late erythroid cells[59]. To validate our in-vitro model, we first measured the expression level of CD235a in HUDEP-2 cells in the expansion medium (undifferentiated state) and then four days after culturing in the differentiation medium. The median fluorescence intensity (MFI) of CD235a increased 26 fold following differentiation (Supp. Figure 20 a, b). Similarly, to test the expression of *TFR2* during erythroid differentiation, we first compared the surface expression of TFR2 in HUDEP-2 cells in the undifferentiated state with the expression upon differentiation. Our analysis showed that TFR2 expression increased 2.5-fold in differentiated HUDEP-2 cells compared to undifferentiated HUDEP-2 (Supp. Figure 20 b), suggesting that expression of TFR2 goes up with differentiation.

Next, we examined the effects of *TFR2* overexpression on HUDEP-2 cells differentiation. We introduced *TFR2* overexpression constructs into undifferentiated HUDEP-2 cells using lentiviruses and incubated the cells in an expansion medium for three days. Subsequently, we assessed TFR2 expression using flow cytometry. We observed around a 3-fold increase in TFR2 expression (Supp. Figure 20 c). This increased expression of TFR2 also led to 2.1 fold increase of CD235a (Extended Data Figure 8c), a 2.2-fold increase in CD71 and a 3.5 fold increase in CD36 expression in the highly TFR2 expressing HUDEP-2 cells (Supp. Figure 20 d). Taken together, our data suggest that overexpression of TFR2 can induce differentiation of erythroid progenitors.

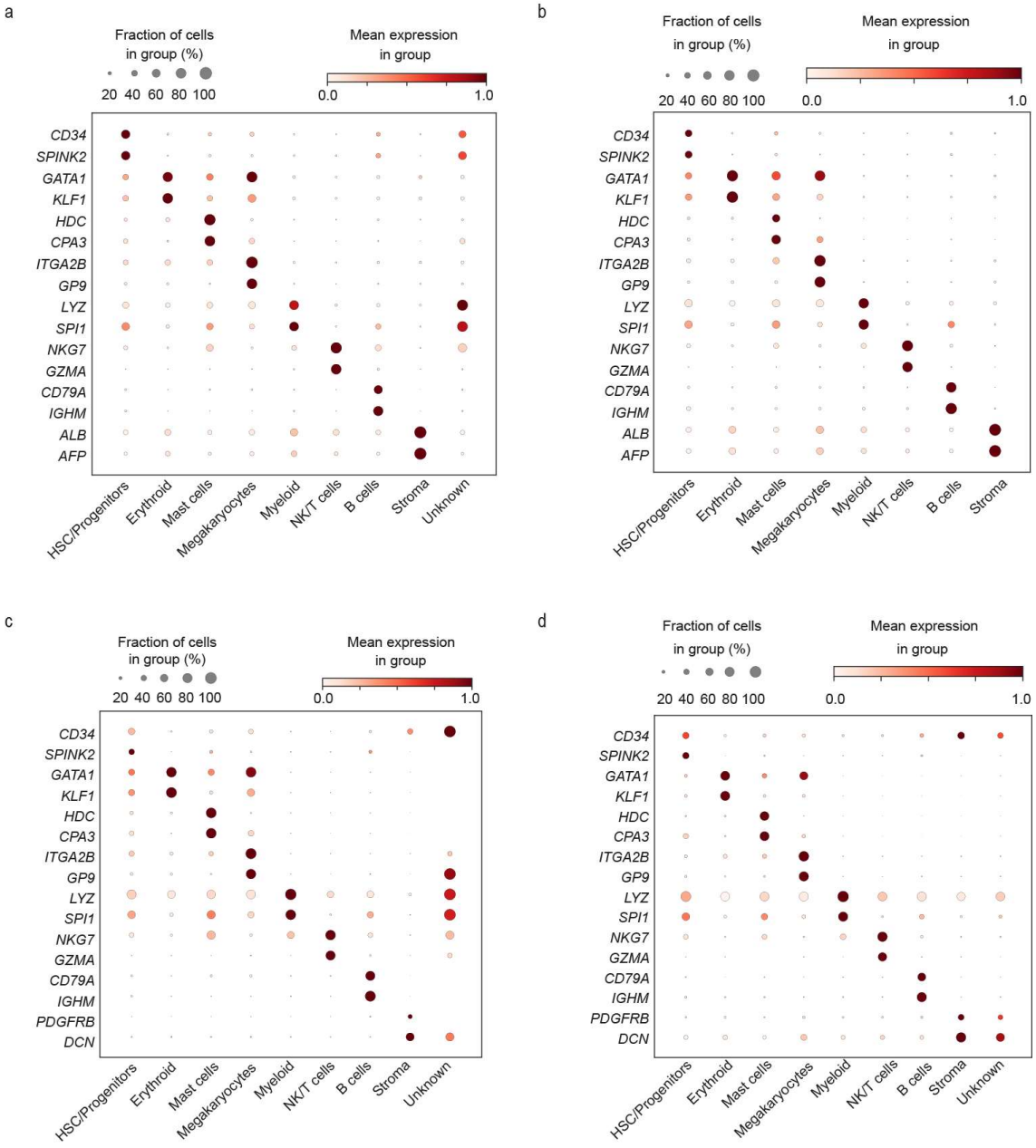
To address whether leukaemia GWAS SNPs showed accessibility differences in Ts21, we used the largest published childhood acute lymphoblastic leukaemia (ALL) GWAS study[60] and performed SCAVENGE analysis similar to the analysis we presented for RBC, WBC, and lymphocyte count. We found a significant enrichment in B cell populations (Supp. Figure 19b), which confirms the utility of the SCAVENGE approach, but no significant difference was observed between Ts21 and disomic HSCs or B cells (Mann-Whitney U test P-value > 0.05). This might be the case because any potential differences that might exist are currently undetectable due to the limited number of identified GWAS signals associated with childhood ALL risk.

Supplementary Figures

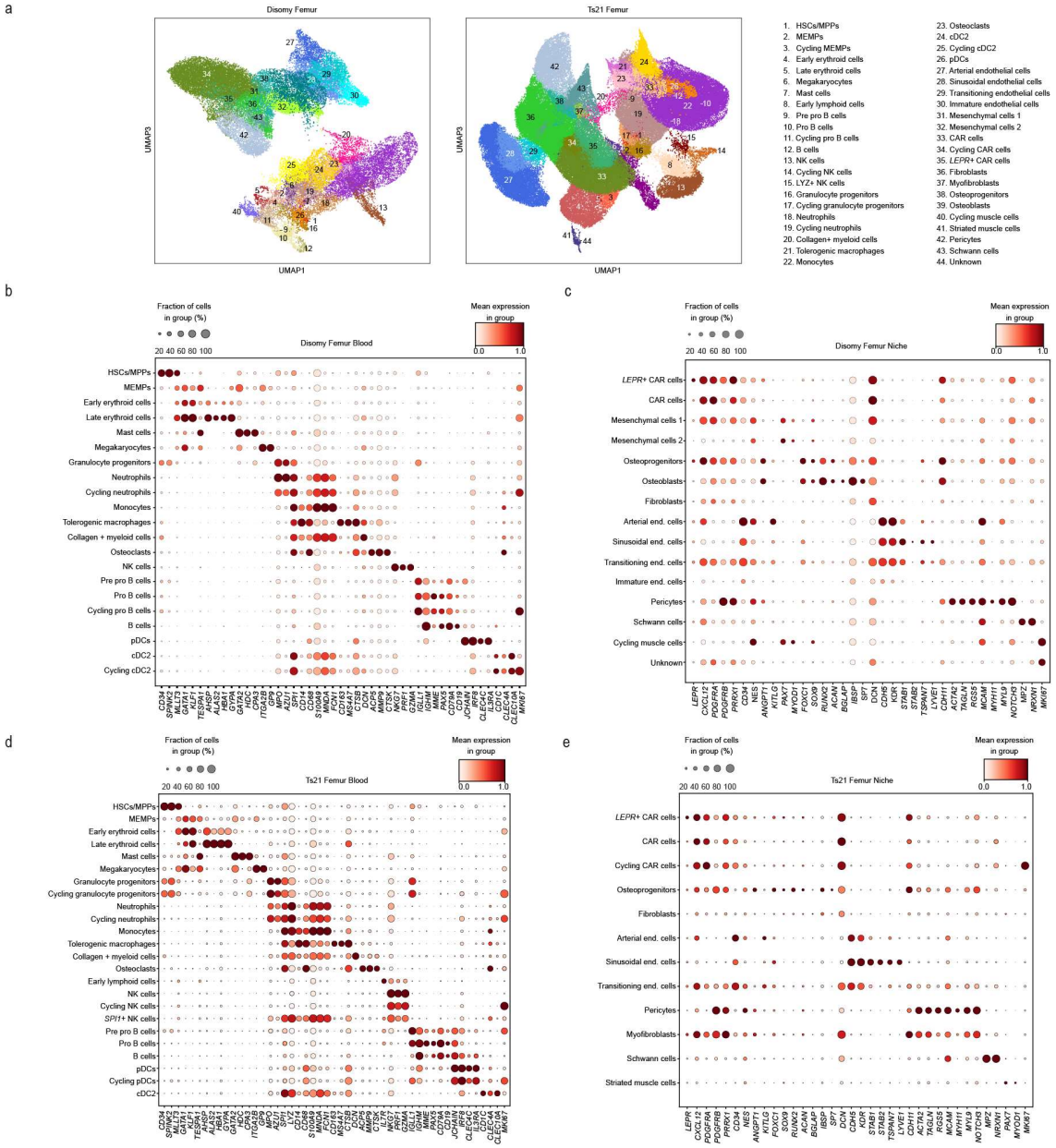
Supplementary Figure 1. Sorting strategy and UMAP visualisation of batch effect corrected scRNA-Seq datasets. **a-b,** FACS sorting panel and gating strategy for the isolation of CD45+ and CD34+Lin- cells from foetal liver (**a**) and foetal BM (**b**). **c-f**, UMAP visualisation of the foetal scRNA-Seq samples after the batch effect correction with Harmony. Each colour represents a different sample. **c,** disomic liver (n=110,671, k=3, k*=8). **d**, disomic femur (n=53,807, k=3, k*=4). **e**, Ts21 liver (n=780,299, k=15, k*=50). **f**, Ts21 femur (n=162,775, k=12 , k*=24). n refers to the total number of cells, k indicates the number of biologically independent samples, k* indicates the total number of replicates (both biological and technical).



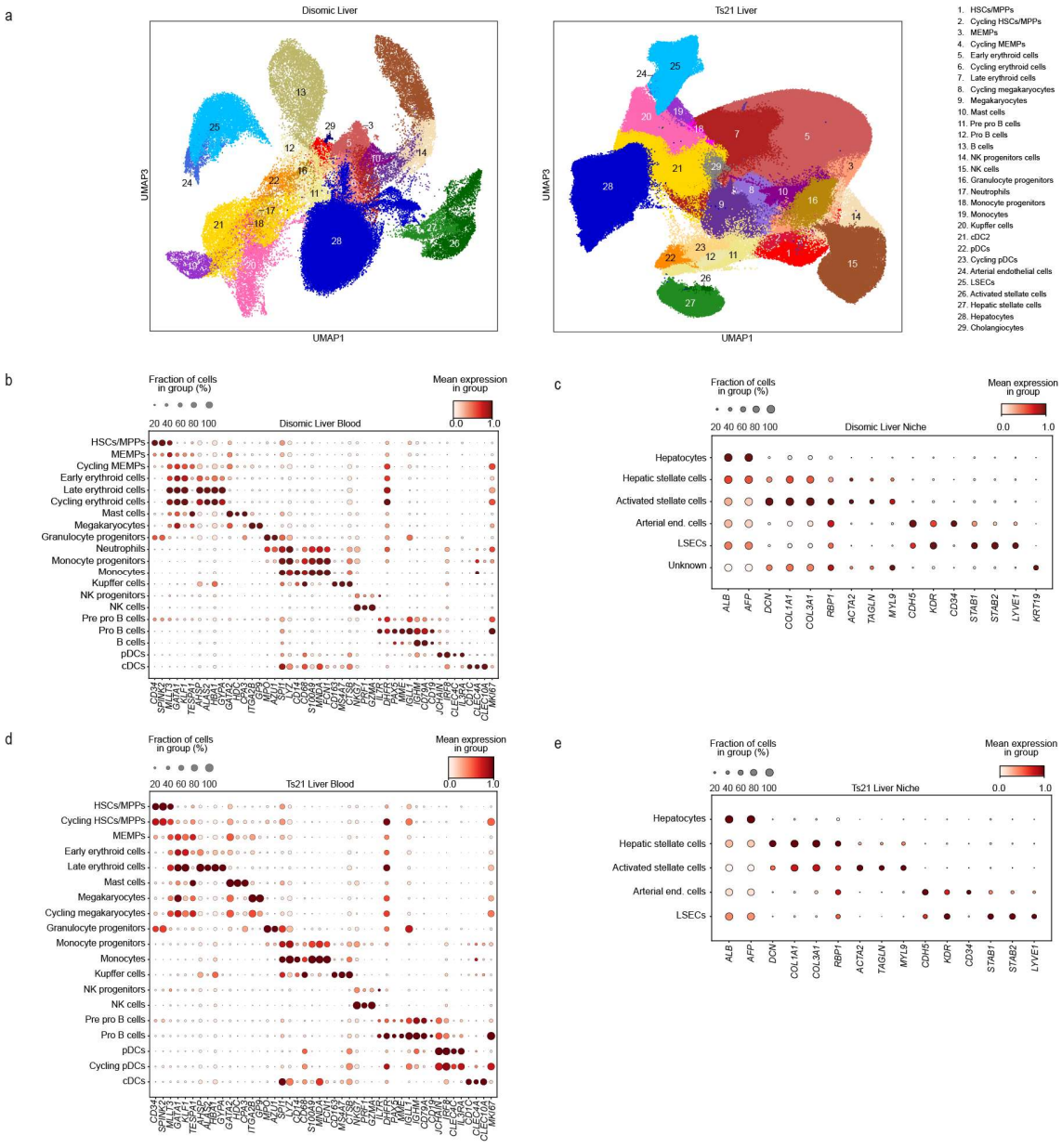
Supplementary Figure 2. Single-cell transcriptome analysis of disomic foetal liver and bone marrow. a-d, Dot plot of the standardised gene expression of manually selected marker genes in annotated broad cell types in Ts21 foetal liver (**a**), Ts21 foetal BM (**b**), disomic foetal liver (**c**), and disomic foetal BM (**d**). For each gene, the minimum value of its expression is subtracted and the result is divided by the maximum value of its expression. The dot size indicates the percentage of cells that express the gene of interest within each cell type.



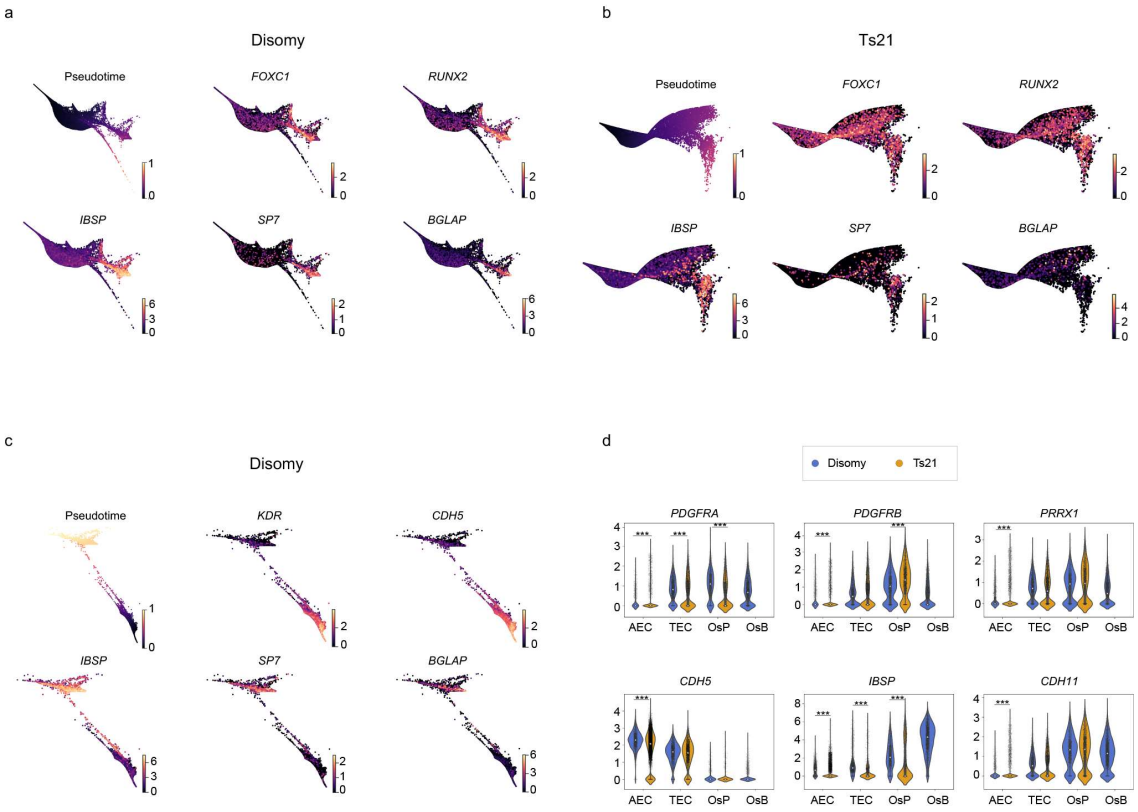
Supplementary Figure 3. A single-cell atlas of disomic and Down's Syndrome foetal bone marrow. a. UMAP visualisation of cells from disomic foetal BM (left; n=53,807, k=3) and Ts21 (right; n=162,775, k=12) coloured by cell type/state. **b-e,** Dot plot of the standardised gene expression of manually selected marker genes in the identified cell types. For each gene, the minimum value of its expression is subtracted and the result is divided by the maximum value of its expression. The dot size indicates the percentage of cells that express the gene of interest within each cell type. **b**, disomic BM haematopoietic cells. **c**, disomic BM niche cells. **d**, Ts21 BM haematopoietic cells. **e**, Ts21 liver BM cells; n refers to the total number of cells, k indicates the number of biologically independent samples.



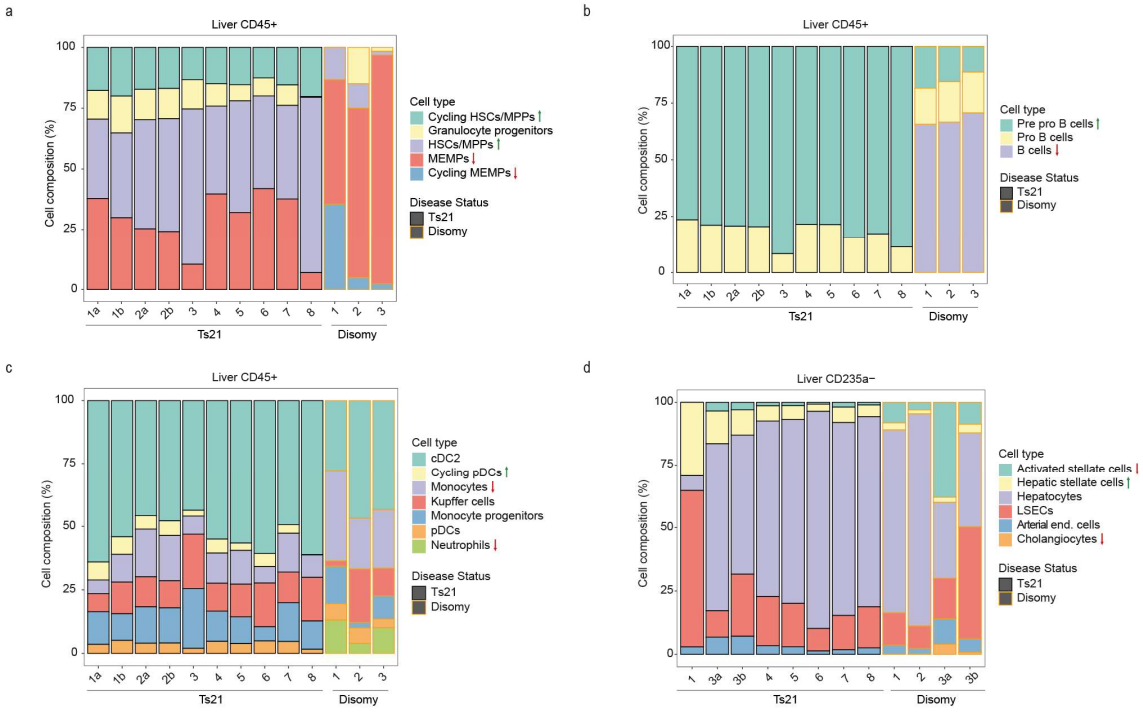
Supplementary Figure 4. A single-cell atlas of disomic and Down's Syndrome foetal liver. a. UMAP visualisation of single cells from disomic foetal liver (left; n=110,671, k=3) and Ts21 (right; n=780,299, k=15) coloured by cell type/state. **b-e,** Dot plot of the standardised gene expression of manually selected marker genes in the identified cell types. For each gene, the minimum value of its expression is subtracted and the result is divided by the maximum value of its expression. The dot size indicates the percentage of cells that express the gene of interest within each cell type. **b**, disomic liver haematopoietic cells. **c**, disomic liver niche cells. **d**, Ts21 liver haematopoietic cells. **e**, Ts21 liver niche cells. n refers to the total number of cells, k indicates the number of biologically independent samples.



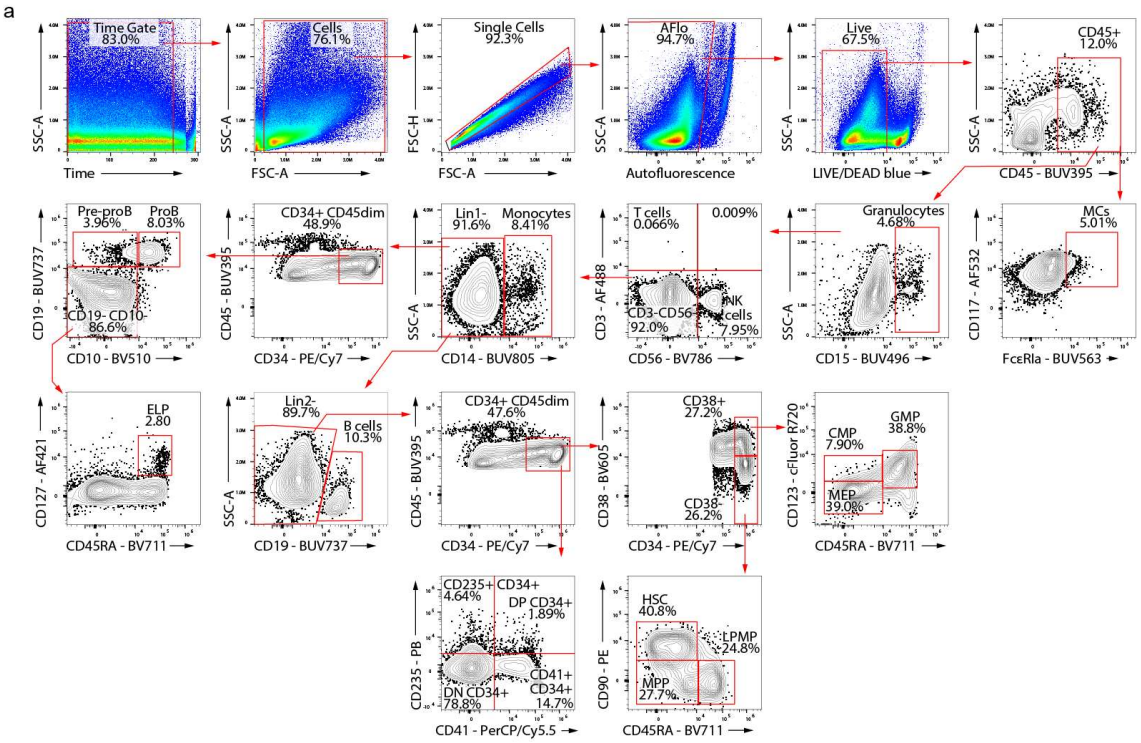
Supplementary Figure 5. Inferring differentiation trajectory of osteoblasts. a-c, Pseudotime and gene expressions (normalised and log1p transformed) on the diffusion map in: **a,** disomic CAR cells, LepR+ CAR cells, osteoprogenitors and osteoblasts**; b,** Ts21 CAR cells, LepR+ CAR cells, osteoprogenitors**; c,** disomic arterial endothelial cells, transitioning endothelial cells and osteoblasts. **d,** Violin plots showing expression of genes in arterial endothelial cells (AEC), transitioning endothelial cells (TEC), osteoprogenitors (OsP), and osteoblasts (OsB). Normalised and log1p transformed gene expression was compared between disomic and Ts21 cells for each cell type; asterisk marks p-values < 0.001, computed by two-sided Wilcoxon rank sum test followed by Bonferroni correction.



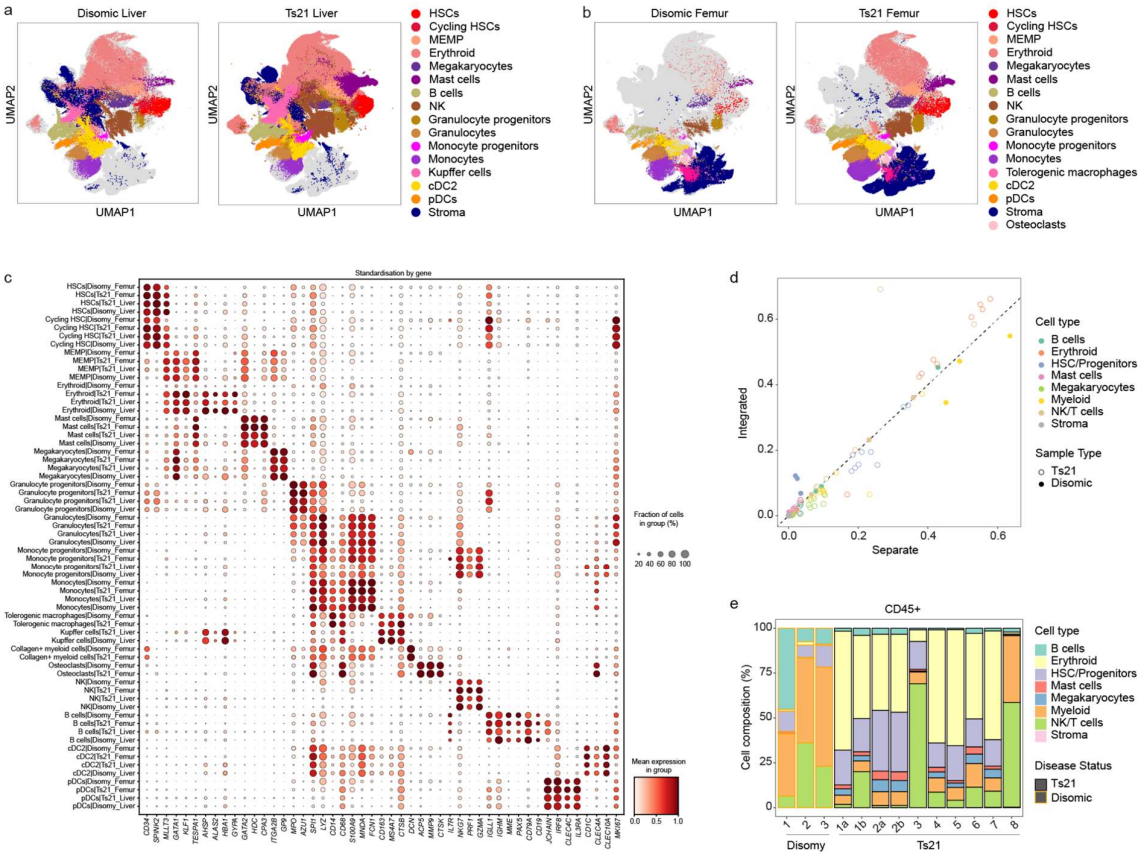
Supplementary Figure 6. Relative abundances of blood and stroma cells in Ts21 and disomic liver. a-d. Stacked bar plots of relative abundances of different cell types/states in stage-matched Ts21 (black outline) and disomic (orange outline) foetal liver. Arrows at the side of the labels indicate a significant increase (↑) or decrease (↓) in the proportion of cells in Ts21 compared to disomic liver. Difference in cell proportion was tested by a two-sided Wilcoxon Rank Sum test followed by Bonferroni correction (FDR < 0.1). **a**, Haematopoietic progenitors subsets identified within CD45+ cells. **b**, B cells and B cell progenitors identified within CD45+ cells. **c**, Myeloid cells identified within CD45+ cells. **f**, Non-haematopoietic cells (niche) identified within CD235- depleted cells.



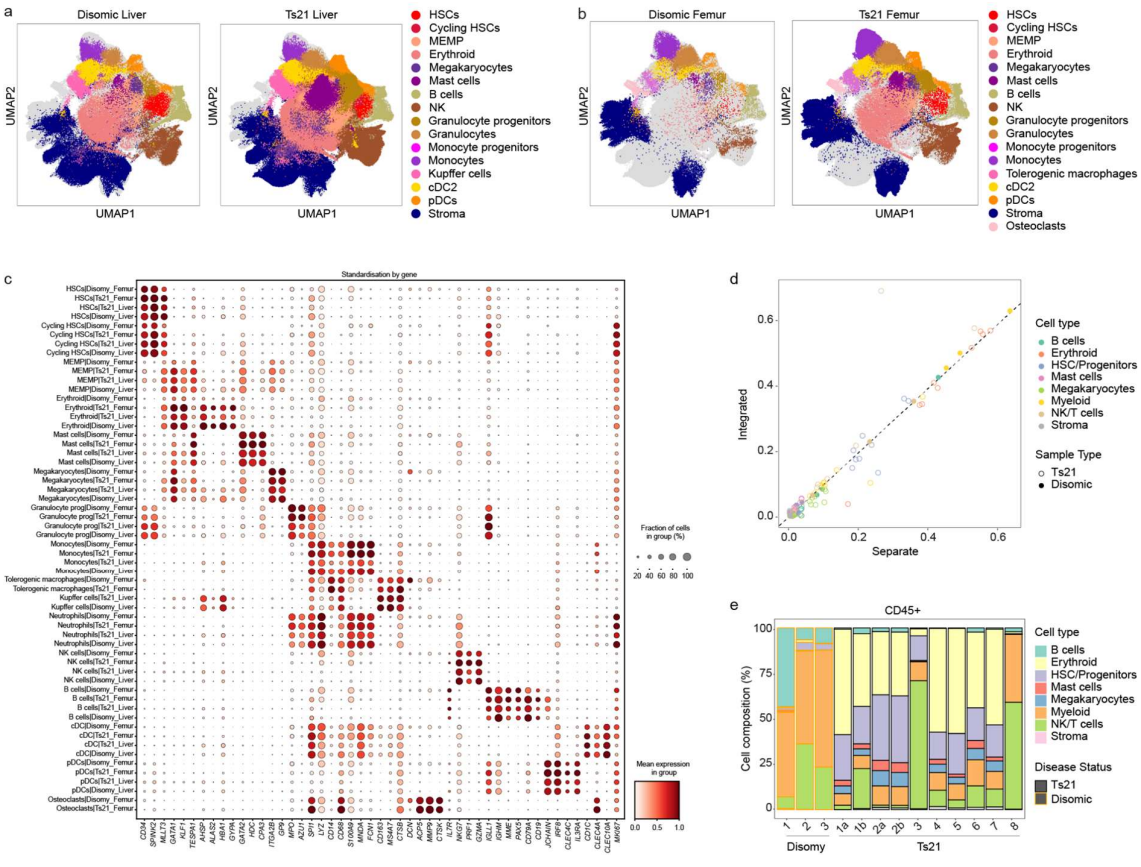
Supplementary Figure 7. Sorting strategy and frequency of immunophenotypic blood cells. a, Representative gating strategy used to identify stem cell subsets, progenitors and mature cells in foetal liver. Gates are based on the disomic foetal liver sample. **b**, Frequency of progenitor and stem cell subsets in T21 and disomic foetal liver. The values on the boxplot indicate the results of a two-sided Welch-corrected t-test comparing the percentage of cells in disomic and T21 foetal livers.



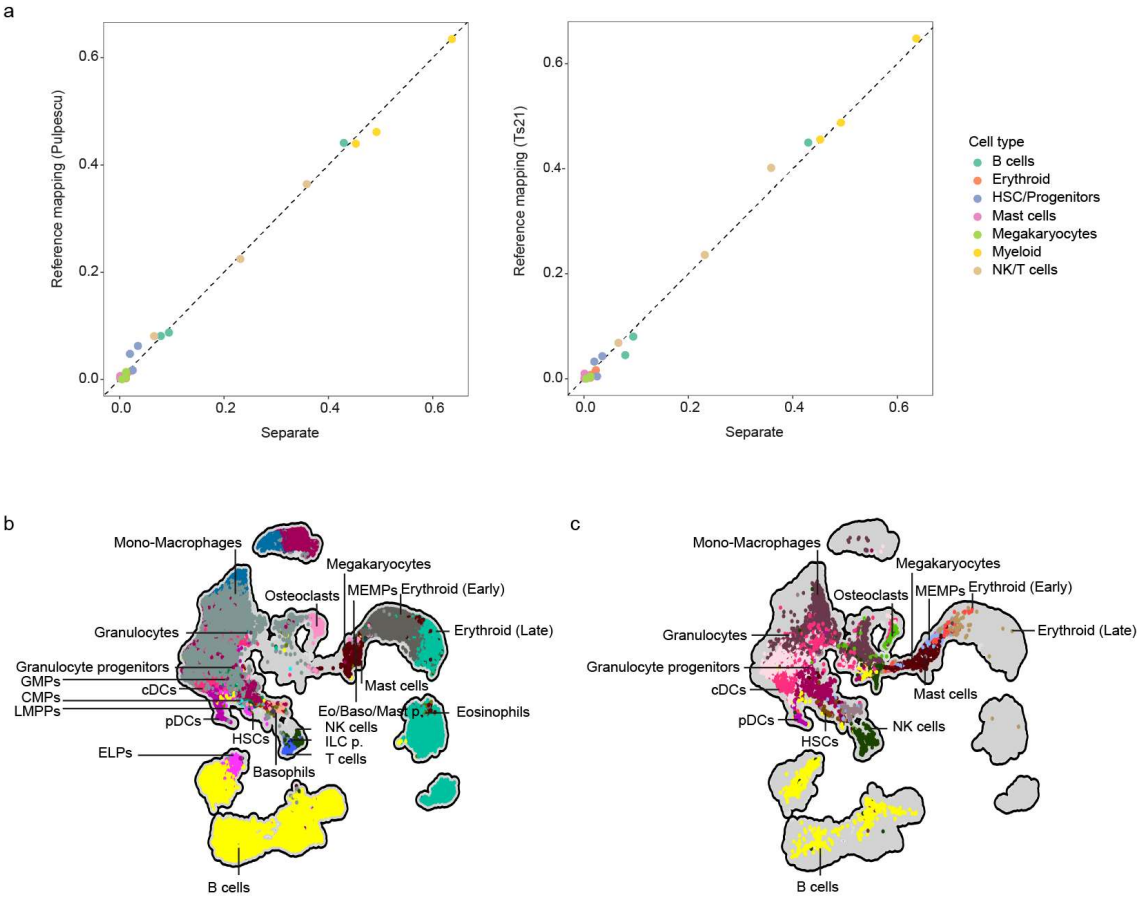
Supplementary Figure 8. Integration of liver, femur, Ts21 and disomic cells using Harmony. 3D UMAP visualisation of cells from (**a**) Ts21 liver (n=780,299, k=15), and disomic liver (n=110,671, k=3) and (**b**) Ts21 femur (n=162,775, k=12), and disomic femur (n=53,807, k=3) following integration of datasets using Harmony. Cells are coloured by broad cell type categories. n refers to the total number of cells, k indicates the number of biologically independent samples (foetuses). **c,** Gene expression dot plot of marker genes from Supp. Figure 3 and 4 across cell types identified from Harmony-based dataset integration. **d,** Using the Harmony-integrated cell type labels in stage-matched Ts21 and disomic foetuses, we compared cell type proportions between separate and integrated. Each point represents the estimated abundance in a different sample. **e,** A stacked bar plot of relative abundances of different broad cell types based on annotations after integration.



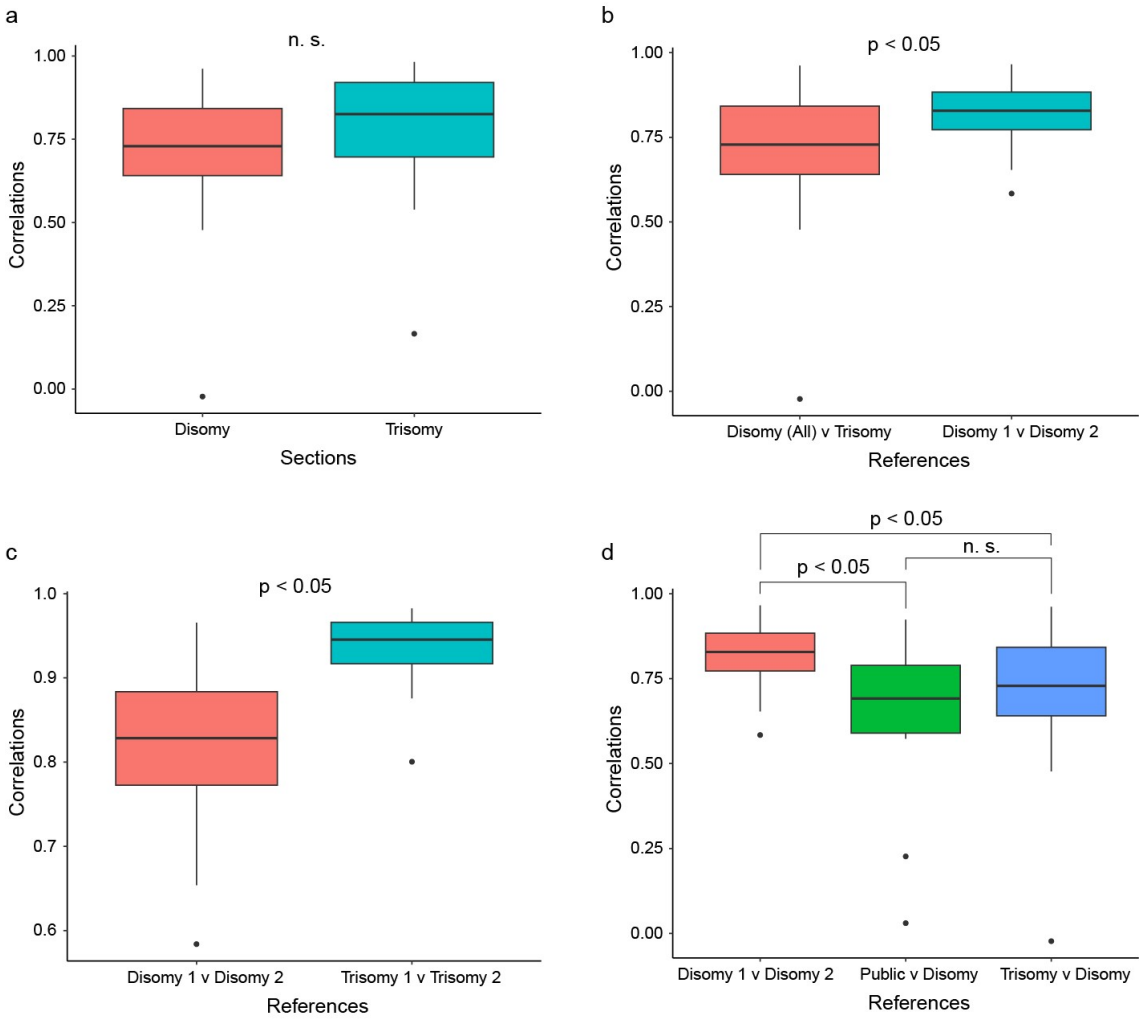
Supplementary Figure 9. Integration of liver, femur, Ts21 and disomic cells using scVI tools. 3D Uniform manifold approximation and projection (UMAP) visualisation of cells from (**a**) Ts21 liver (n=780,299, k=15), and disomic liver (n=110,671, k=3) and (**b**) Ts21 femur (n=162,775, k=12), and disomic femur (n=53,807, k=3) following integration of datasets using Harmony. Cells are coloured by broad cell type categories. n refers to the total number of cells, k indicates the number of biologically independent samples (foetuses). **c.** Dot plot showing expression of marker genes from Supp. Figure 3 and 4 across cell types identified following dataset integration using scVI tools. **d.** Using the scVI-integrated cell type labels in stage-matched Ts21 and disomic foetuses, we compared cell type proportions between separate and integrated. Each point represents the estimated abundance in a different sample **e.** A stacked bar plot of relative abundances of different broad cell types based on annotations after integration.



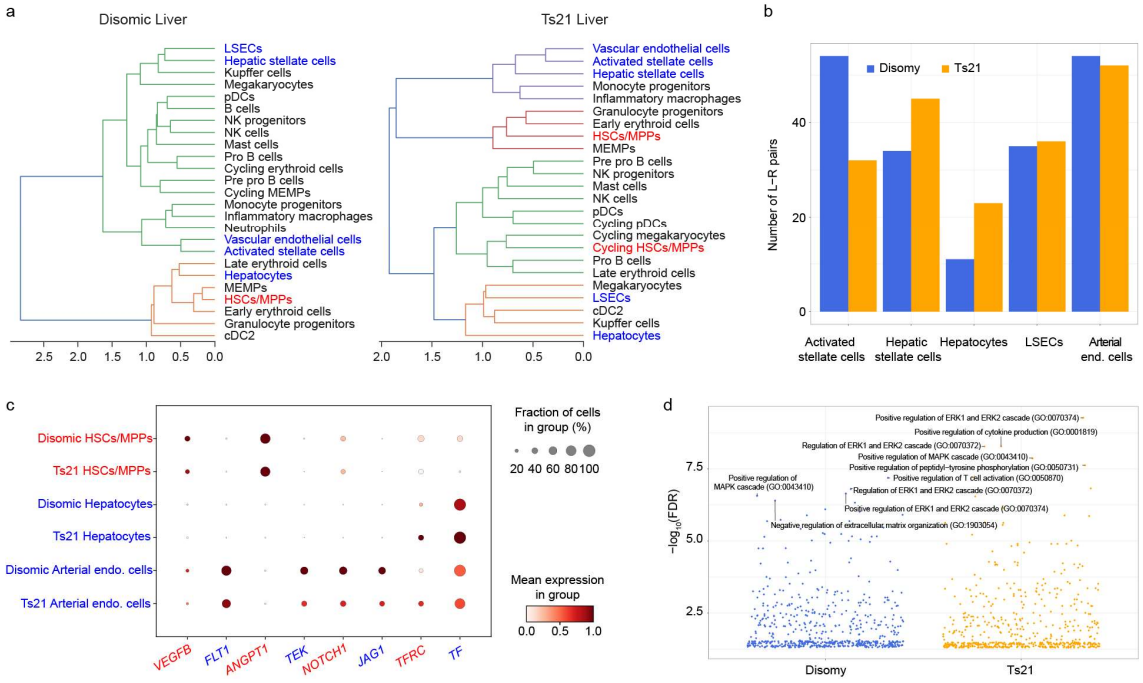
Supplementary Figure 10. Integration of liver and disomic cells and label transfer. a, Comparison of cell type proportions for stage-matched disomic liver samples from the separate-dataset annotations to the reference query label transfer annotations using Popescu et al.[30] (left) or the Ts21 dataset as a reference (right). **b**, Integration of Jardine et al.[32] and Marderstein et al. cells from disomic BM using scVI. The Jardine et al.[32] cells are shown on the left, and the Marderstein et al. cells are shown on the right. Cells are coloured by the annotated population from their original studies.



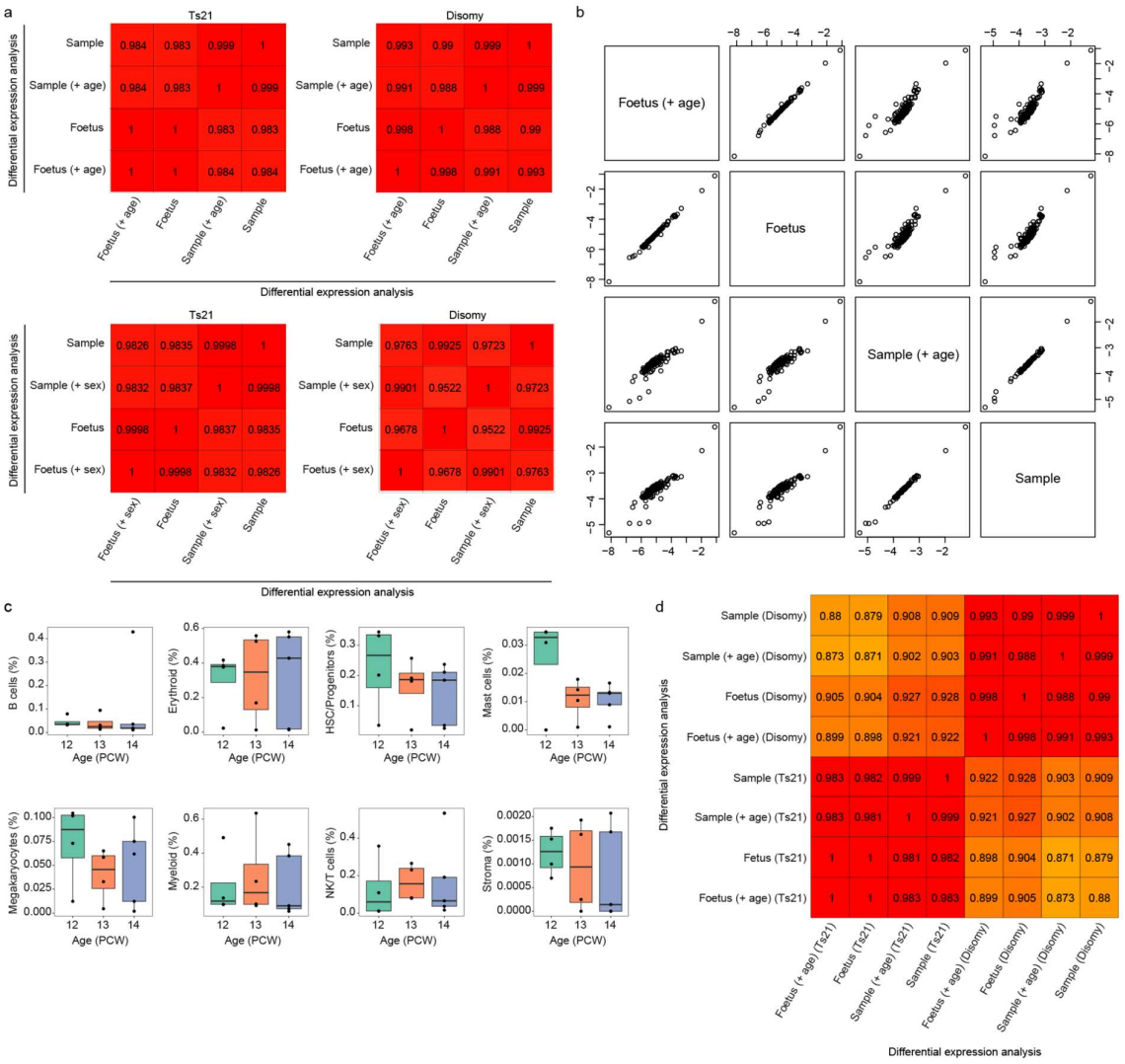
Supplementary Figure 11. Impact of reference selection on the prediction of the cell abundances on Visium spots. **a**, Cell types were subset to those present in both disomy and Ts21 references, and the correlation between abundances of corresponding cell types were computed (using both disomy- and trisomy-based cell2location deconvolutions). Each point in the boxplots displays the correlations. **b**, The correlation between abundances of corresponding cell types are shown across disomy sections. Cell2location was performed using two disomy subset references (Disomy 1 and Disomy 2), all disomy- and trisomy-based cell2location deconvolutions (4 models total). For comparing disomy versus trisomy, only cell types were subset to those present in both disomy and Ts21 references. Each point in the boxplots displays the correlations. **c,** The correlation between abundances of corresponding cell types are shown and compared between disomy and trisomy. Cell2location was performed using two disomy subset references (Disomy 1 and Disomy 2) on disomy samples or two trisomy subset references (Trisomy 1 and 2) on trisomy samples. Each point in the boxplots displays a different cell type. **d,** The correlation between abundances of corresponding cell types are shown. Cell2location was performed using two disomy subset references (Disomy 1 and Disomy 2), all disomy samples, Popescu et al.[30] (public), and Trisomy as a reference for deconvoluting disomy samples. Each point in the boxplots displays a different cell type.



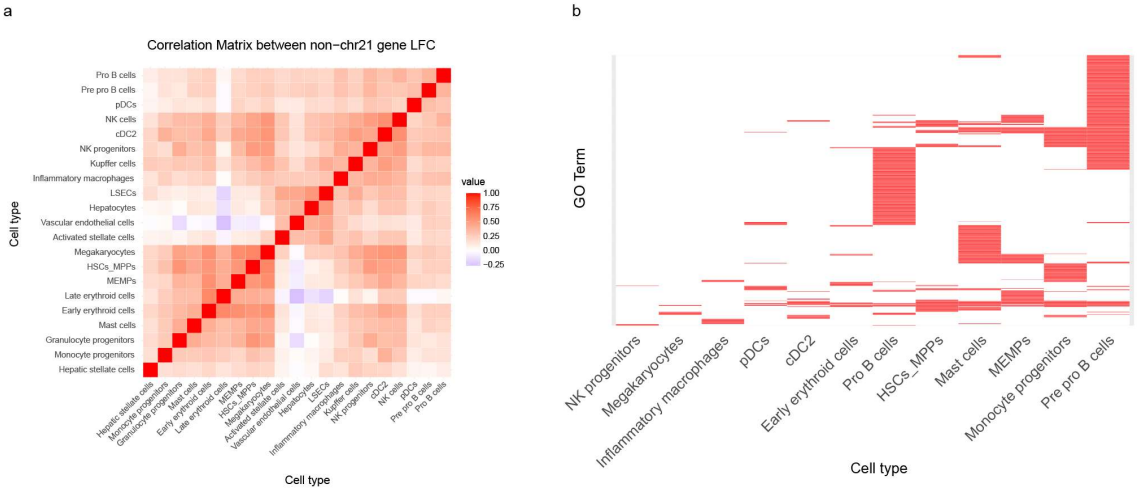
Supplementary Figure 12. Colocalisation and L-R analysis from Visium data. a, Dendrogram from hierarchical clustering results based on correlation distance, with inter-cluster distance estimated by Ward variance minimization algorithm. HSCs/cycling HSCs are coloured in red, stromal cells are coloured in blue, all other blood cells are coloured in black. **b,** The number of L-Rs that HSC/MPPs use to interact with different niche cells. **c**, Dotplot of L-Rs expressed in HSC/MPPs, hepatocytes and arterial endothelial cells. Dot size represents the fraction of cells expressing the gene, and colour indicates the mean expression level, both computed within the cell type cluster. The genes coloured in red are L/Rs associated with HSCs/MPPs, and the genes coloured in blue are those associated with hepatocytes and arterial endothelial cells. **d,** Gene-set enrichment analysis of L-Rs expressed in HSC/MPPs used for their interaction with arterial endothelial cells. Top 5 terms in each group were labelled. The presented results include all sections from Ts21 (n=20 sections; 2 consecutive sections from 10 biological replicates) and disomic (n=10 sections; 2 consecutive sections from 5 biological replicates) tissues.



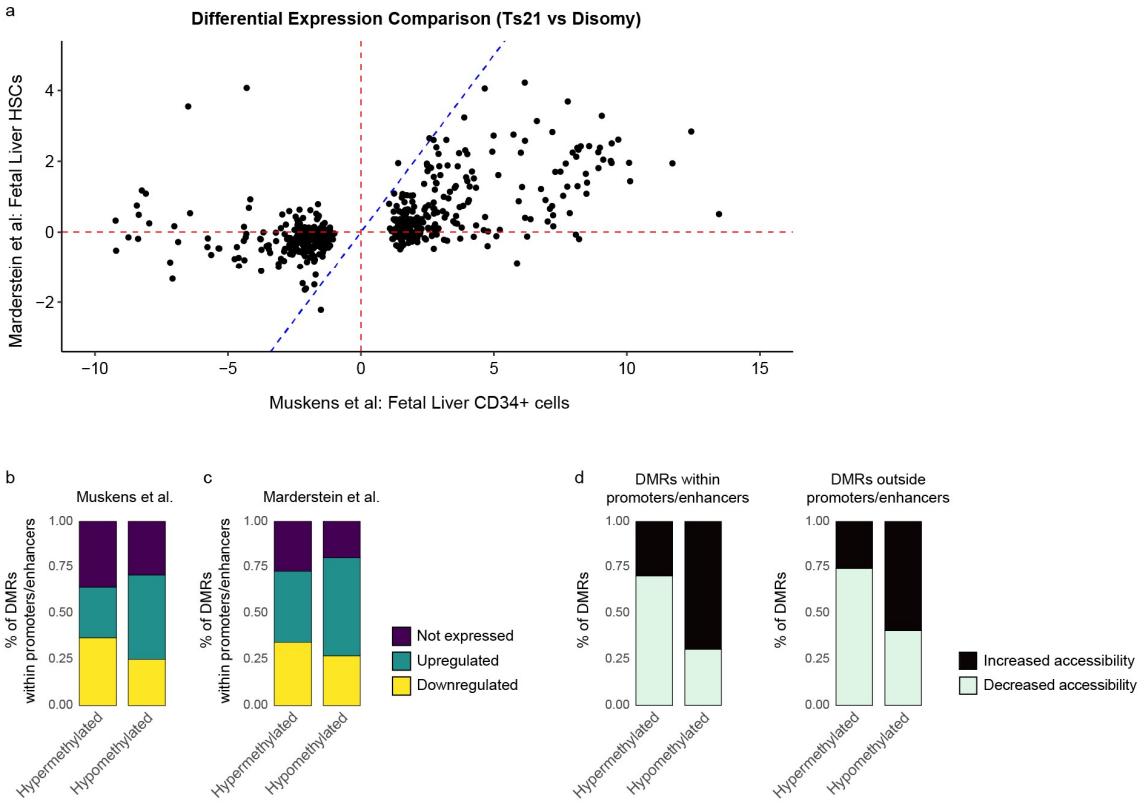
Supplementary Figure 13. Evaluating the robustness of differential expression results. **a,** Pearson's correlation coefficient between Liver vs Femur log-fold changes across all genes for the four analysis types, using age (**top**) or sex (**bottom**) as covariates. **b**, Log-fold changes for top 200 differentially expressed genes not located on chr 21 across the 4 analysis choices in Ts21 Liver versus Femur. **c**, The relationship between post-conceptual weeks and cell type abundance (estimated as proportions) for the 8 broad cell type groups. **d**, Pearson's correlation coefficient between Liver vs Femur log-fold changes across all genes for the four analysis types in the two datasets (8 distinct analyses in all).



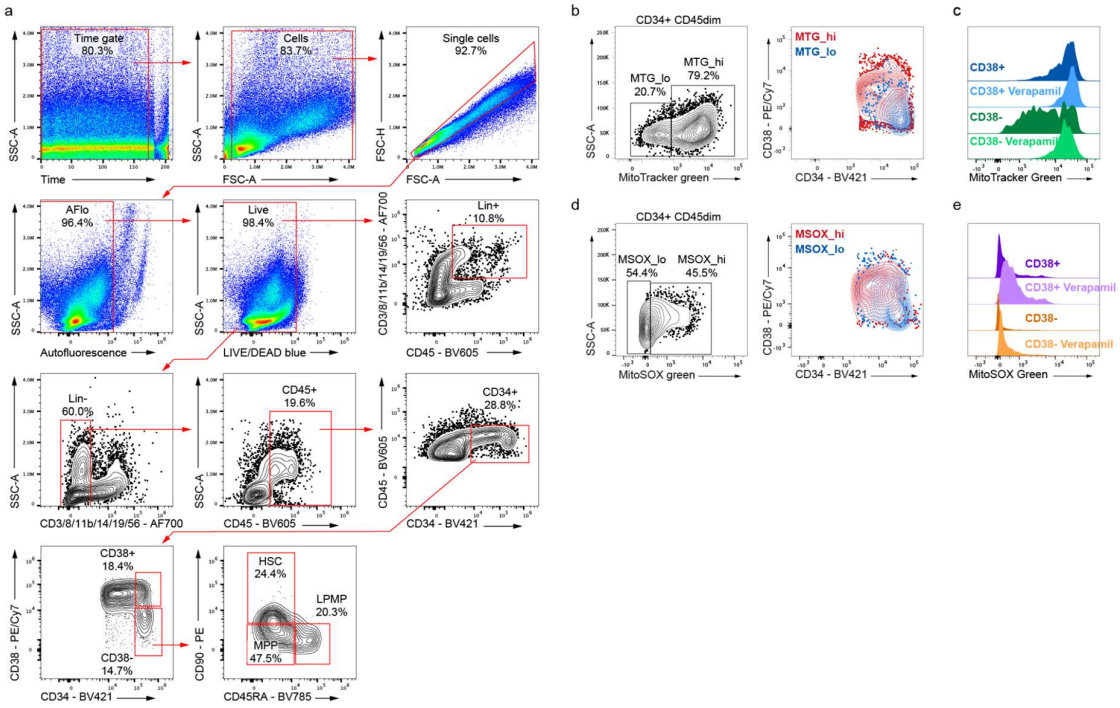
Supplementary Figure 14. Gene expression differences across cell types in Ts21 versus disomic. a, Correlation matrix between log-fold changes across non-chromosome 21 genes, with cell types ordered after performing hierarchical clustering. **b**, Presence of enriched GO terms (FDR < 0.1) across 12 blood cell types based on genes significantly upregulated in Ts21 foetal liver.



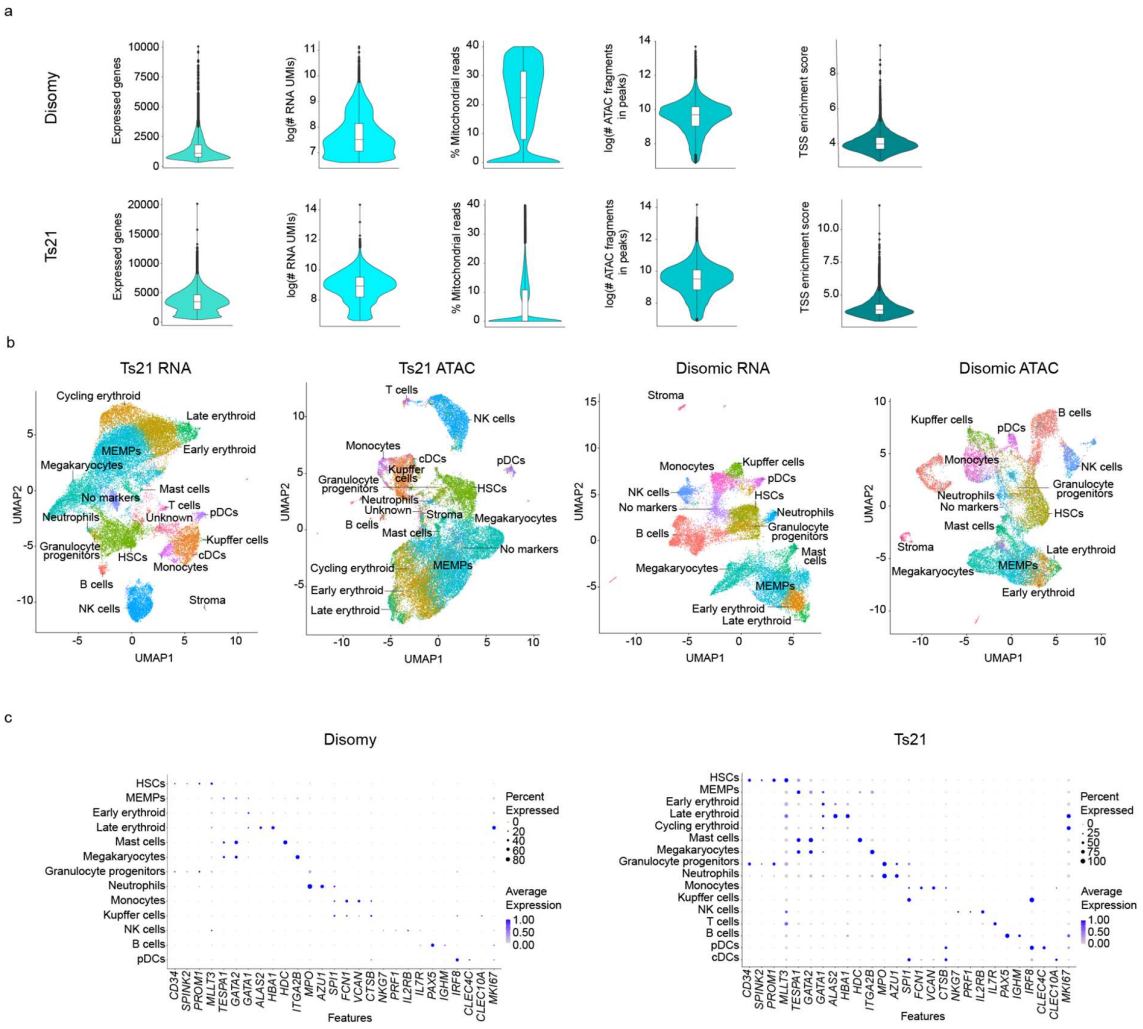
Supplementary Figure 15. Link between Ts21 alterations in the epigenome and transcriptome with DNA methylation. a, Comparison of differential expression results between Muskens et al.[42] and our study. Differential expression results are from Supplementary Data 9 by Muskens et al.[42], which include genes significantly differentially expressed between Ts21 foetal liver versus disomy foetal liver CD34+ cells (x-axis), and the results from our analysis in Ts21 versus disomy foetal liver HSCs (y-axis). Log-fold changes for each gene are shown along the axes. The dashed red line represents x=0 and y=0, and the blue line represents x=y. **b-c**, Stacked barplot showing the association between (i) hypo- and hypermethylation at DMRs within gene enhancers and promoters, and (ii) differential expression of the associated gene, for both (**b**) Muskens et al.[42] and (**c**) our study. **d**, Stacked barplot showing the association between (i) hypo- and hypermethylation at DMRs within gene enhancers and promoters in CD34+ cells from Muskens et al.[42], and (ii) differential accessibility of the region in liver HSCs from our study. Left shows DMRs within promoter or enhancer regions, while right shows DMRs outside promoter or enhancer regions (annotated as in the gene body or intergenic).



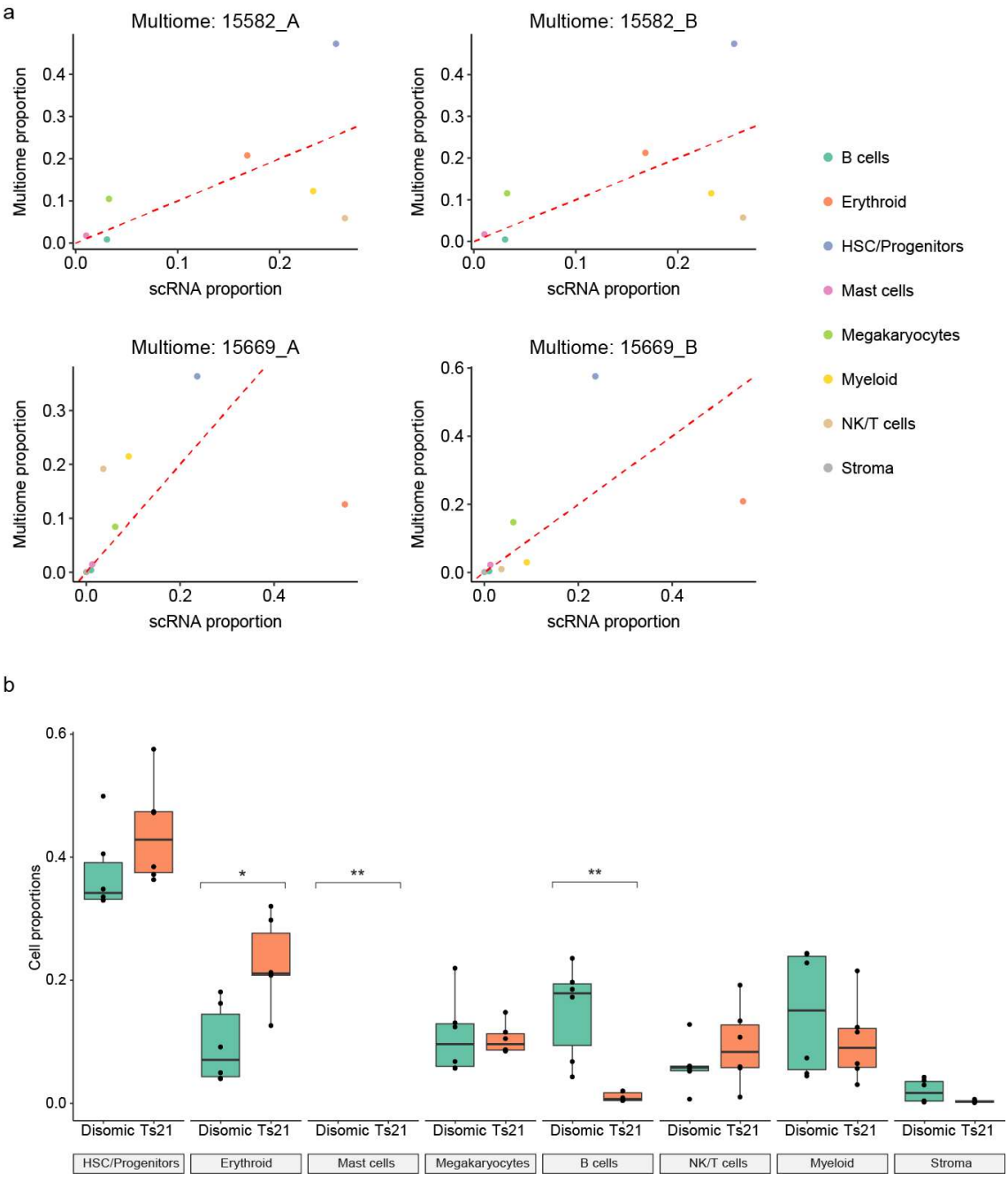
Supplementary Figure 16. Sorting strategy and measuring mitochondrial mass and oxidative stress in haematopoietic cells. a, Representative gating strategy used for the mitochondrial probe experiments. Both MitoSOX and MitoTracker stainings shared the same antibody panel. **b**, Left: Representative gating strategy used to identify MTG_hi and MTG_lo populations. Right: MTG_hi and MTG_lo cells were plotted against CD34 and CD38, showing that MTG_lo cells are enriched in the CD34+ CD38- fraction. **c**, Density plot of the MitoTracker Green FM fluorescence in CD38+ and CD38- cells treated or not with Verapamil 50μM during the MitoTracker incubation. Cells were gated on Time gate/Cells/Single cells/AFlo/Live/Lin-/CD45+/CD34+. **d**, Left: Representative gating strategy used to identify MSOX_hi and MSOX_lo populations. Right: MSOX_hi and MSOX_lo cells were plotted against CD34 and CD38, showing that MSOX_lo cells are enriched in the CD34+ CD38- fraction. **e**, Density plot of the MitoSOX Green fluorescence in CD38+ and CD38- cells treated or not with Verapamil 50μM during the MitoSOX incubation. All cells were gated on Time gate/Cells/Single cells/AFlo/Live/Lin-/CD45+/CD34+.



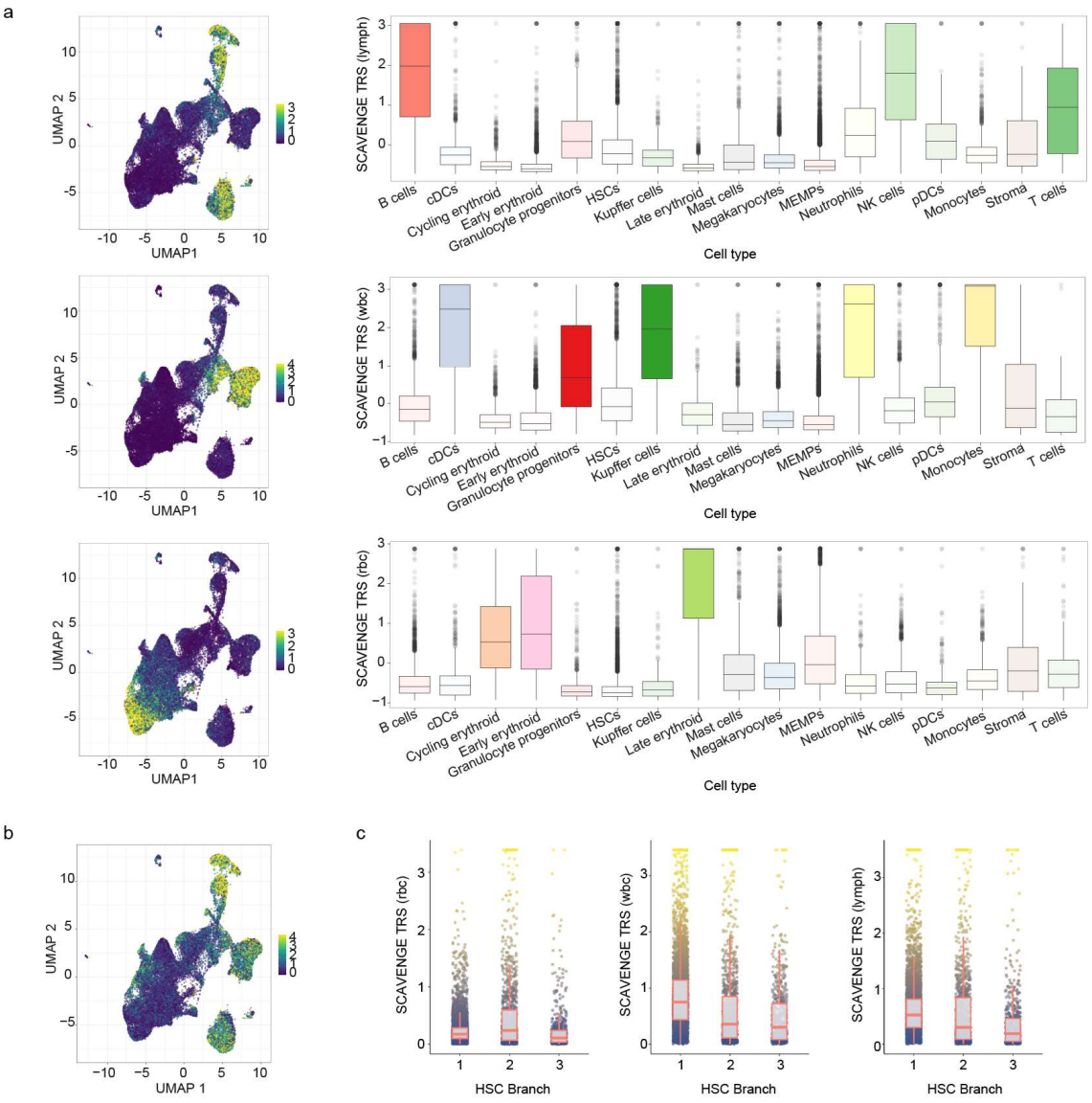
Supplementary Figure 17. Multiome atlas. a, Violin and boxplots for quality control metrics from multiome data. The bottom, middle, and top lines of the boxplot reflect the 25th, 50th (median), and 75th TRS percentiles, and the boxplot whiskers reflect 1.5x the interquartile range. **b,** UMAP visualisation of snRNA-seq (left) and corresponding scATAC-seq (right) of CD45+ cells isolated from Ts21 foetal liver (n=35,633, k=3, median PCW=13, left) and disomic foetal liver (n=21,257, k=3, median PCW=13, right). n- number of cells; k- number of biological replicates. **c,** Dot plot of the standardised gene expression of manually selected marker genes in the identified cell types. For each gene, the minimum value of its expression is subtracted and the result is divided by the maximum value of its expression. The dot size indicates the percentage of cells that express the gene of interest within each cell type.



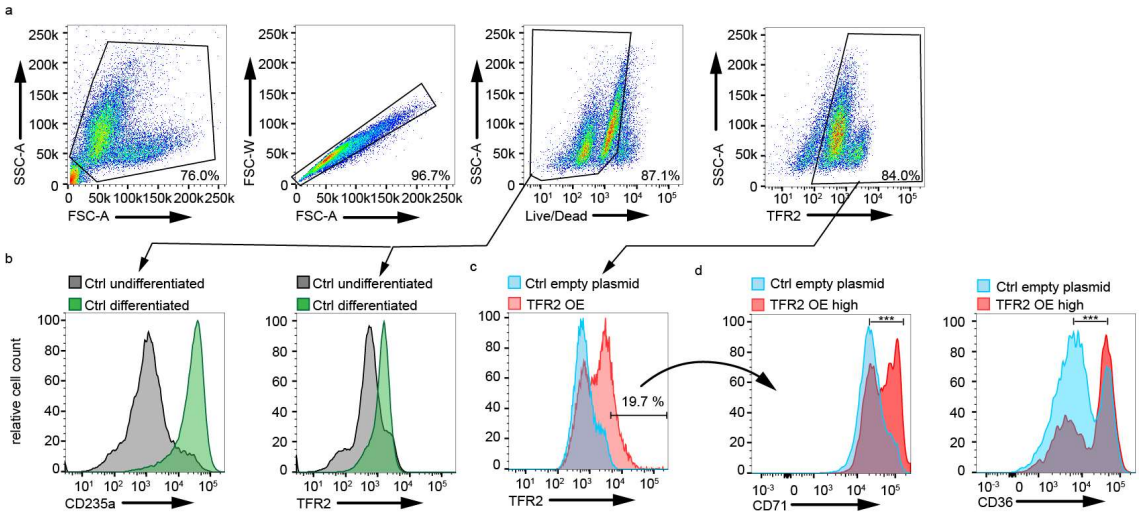
Supplementary Figure 18. Comparison of cell-type frequencies in Ts21 and disomy using 10X multiome. a, Correlation between estimated cell type proportions within scRNA (x) and multiome samples (y) in two multiome replicates (left, right subplots) for sample 15582 (top, $r > 0.55$) and sample 15669 (bottom, $r > 0.37$). Dashed red line is $x = y$. **b**, Two-sided Mann-Whitney U test of differences in cell-type proportions between Ts21 and disomic samples. Significance was determined after FDR correction across cell types. P<0.05(*); P<0.01(**).



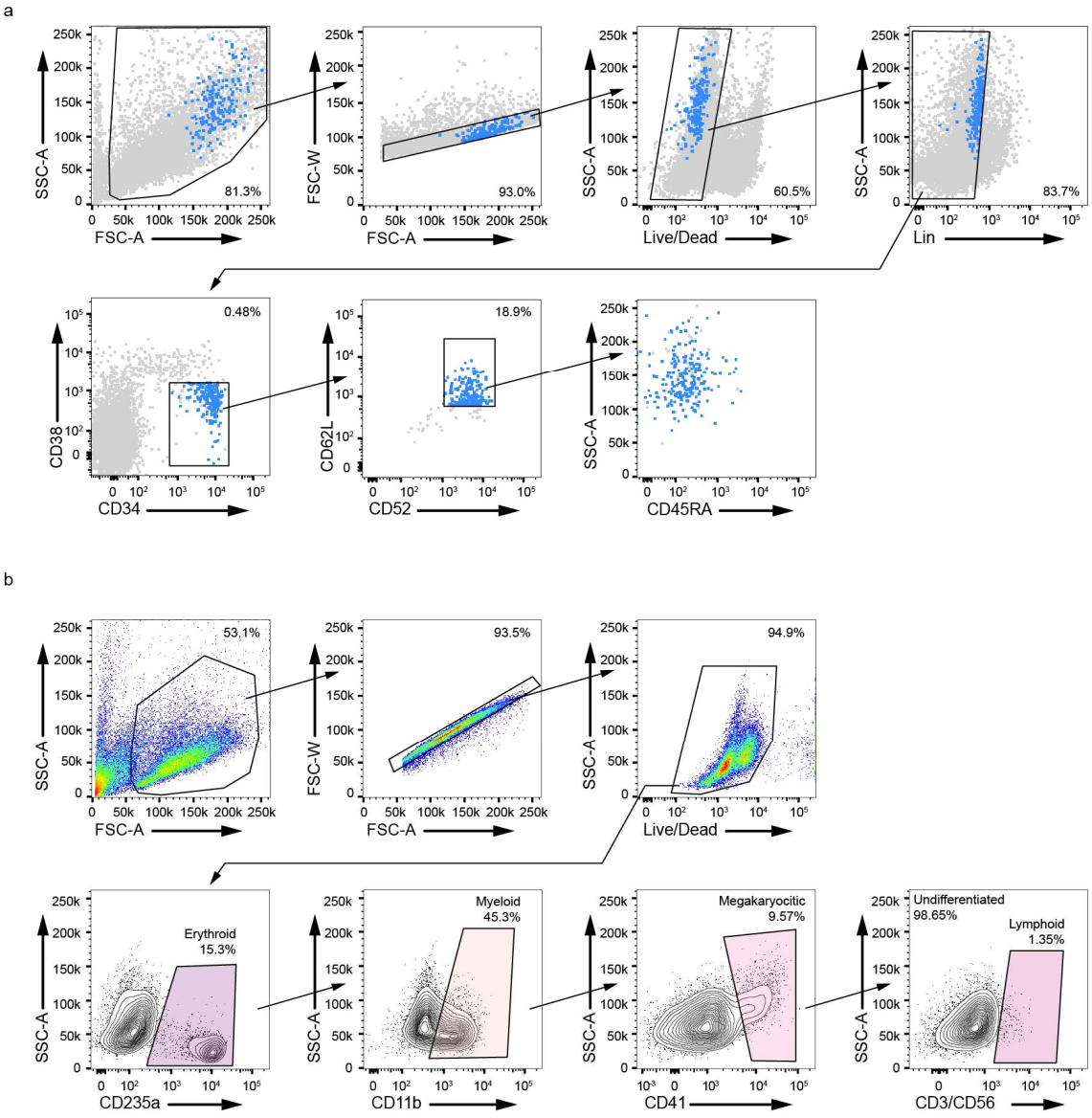
Supplementary Figure 19. SCAVENGE application to scATAC. a, Left**,** UMAP of scATAC data in the combined data (disomic and Ts21)**.** Each point is a different cell, coloured by SCAVENGE trait relevance scores (TRS) for lymphocyte counts (top; lymph), white blood cell counts (middle; wbc), and red blood cell counts (bottom; rbc). Right, boxplots display SCAVENGE TRS across cell populations. The bottom, middle, and top lines of the boxplot reflect the 25th, 50th (median), and 75th TRS percentiles, and the boxplot whiskers reflect 1.5x the interquartile range. **b.** SCAVENGE TRS for acute lymphocytic leukaemia (ALL) GWAS displayed on the UMAP. **c**, SCAVENGE TRS for 3 blood cell traits (red blood cell counts (rbc), white blood cell counts (wbc), and lymphocyte counts (lymph)) across the 3 identified HSC branches. Each point is a different cell. The bottom, middle, and top lines of the boxplot reflect the 25th, 50th (median), and 75th TRS percentiles, and the boxplot whiskers reflect 1.5x the interquartile range.



Supplementary Figure 20. Gating strategy and surface expression levels of erythroid markers in HUDEP-2 cells. **a,** Gating strategy used to determine surface expression of TFR2, CD235a, CD71 and CD36. **b**, Histograms showing CD235a expression (left) and TFR2 expression (right) in HUDEP-2 cells in undifferentiated state (gray) and upon differentiation (green). **c,** TFR2 expression in HUDEP-2 transduced with an empty vector (Ctrl, blue) in comparison with TFR2 overexpressing HUDEP-2 (light red). **d**, CD71 and CD36 expression in Ctrl HUDEP-2 (blue) and in the population of highly expressing TFR2 cells (dark red). Differences in median fluorescence intensity were tested by Mann-Whitney U test, *** p-value < 0.001.



Supplementary Figure 21. Single-cell CFU assay. a, Representative FACS sorting panel and gating strategy for the isolation of HSC/MPPs (CD34+CD38-CD52+CD62+Lin-). All HSC/MPPs were CD45RA-. **b,** Representative FACS plot showing the lineage composition of a quadri-lineage colony.



List of Supplementary Tables
(provided as spreadsheet)

Supplementary Table 1 - Disomic foetal cells passing quality control

Supplementary Table 2 - Ts21 foetal cells passing quality control

Supplementary Table 3 - Marker genes for each cell population. Column descriptions: type = DownSyndrome or Disomic, organ = Femur or Liver, numerical_labels = cluster number, cell_type = labelled cell type population for the cluster, Rank = gene rank for the cell type, logfoldchanges = log-fold change of differential expression between cells in cluster and all other cells, pvals = p-values, pvals_adj = FDR-adjusted p-values.

Supplementary Table 4 - Cell types per replicate, Ts21 liver scRNA-seq

Supplementary Table 5 - Cell types per replicate, Disomy liver scRNA-seq

Supplementary Table 6 - Cell types per replicate, Ts21 femur scRNA-seq

Supplementary Table 7 - Cell types per replicate, Disomy femur scRNA-seq

Supplementary Table 8 - Cell type abundance comparison. Median abundance in Ts21 samples and disomic samples, Mann-Whitney U test p-values and their FDR corrected values.

Supplementary Table 9 - 10X Visium summary

Supplementary Table 10 - Sample-level cell-type abundances estimated using cell2location. The normalised cell-type abundances were computed from all spots for each section, which were used for sample-level comparisons of disomy versus trisomy samples. “Percent_ref” indicates the relative abundances (%) for each cell type in the sample using either the “t21”, “disomy”, or “public” scRNA-seq references. “Ref_comp” is the computed difference in estimated abundances calculated from two different references, which is a measure of inherent variability due to reference.

Supplementary Table 11 - Differential expression between Ts21 and disomic cells. Column descriptions: organ = Femur or Liver, cell = cell type population, names = gene name, logFC = log-fold change of differential expression between Ts21 and disomic samples, AveExpr = average expression in all samples, t = test statistic, P.Value = p-values, adj.P.Val = FDR-adjusted p-values.

Supplementary Table 12 - Differential expression between Ts21 and disomic HSCs with and without cellbender

Supplementary Table 13 - Differential expression between liver and femur HSCs with and without cellbender

Supplementary Table 14 - Enriched GO terms across blood cell types. The “set” reflects down- and up- regulated genes.

Supplementary Table 15 - Multiome summary

Supplementary Table 16 - Key characteristics of RNA and ATAC topics. Top 200 genes and enriched TFs relevant to those genes in RNA topics, and top 10% peaks and enriched TFs relevant to ATAC topics. For calculating enriched TFs in RNA topics, we used regulatory potential modelling to calculate association scores between each TF and gene. A Wilcoxon test is performed to test for differences in TF association scores between the top genes for each topic and all other genes, thus identifying which TFs are "regulating" that topic. For calculating enriched TFs in ATAC topics, we tested the enrichment of TF motifs in the top peaks for the topic. Only TFs with P-value < 0.05 are shown. Column descriptions: datatype=RNA or ATAC; feature = Top 200 Genes, Top 10% Peaks, TF enrichment of Top 200 Genes, or TF enrichment of Top 10% Peaks; variable = topic number; Rank = rank of value in feature set; value = either TF label, gene, or peak; pval = p-value for TF enrichment (otherwise NA).

Supplementary Table 17 - TF regulator results from MIRA. Source (src) specifies which branches are the TF regulator results related to. The columns id, name, parsed_name, pval, and test_statistic detail motif identifier, TF name, a computationally parsed TF identifier, enrichment significance (nominal P-value), and enrichment test statistic from MIRA.

Supplementary Table 18 - Transcription factor motif enrichment in Ts21 and disomic HSCs. MonaLisa enrichments (log2enr) and significance (negLog10Padj). Peak_type represents "all", "Enhancer", or "promoter" for different analyses conducted.

Supplementary Table 19 - Peak-gene links tested by SCENT. Gene_peak, gene, and peak describe the peak and gene tested. chr = chromosome, snp_pos = basepair position, rsid, and pip = finemap posterior inclusion probability are filled if the peak overlaps with a fine-mapped GWAS SNP (PIP > 0.2) for red blood cell counts. beta_H, se_h, z_H, p_H, boot_basic_p, fdr_H contain the estimated effect, standard error, z-score, non-bootstrapped P-value, bootstrapped P-value, and Benjamini-Hochberg corrected P-value of chromatin accessibility for gene expression within disomic HSCs. *_t21 contains the same information for Ts21 HSCs, *_t21_dn contains the same information for Ts21 HSCs after downsampling to the disomic dataset size, and *_int contains the same information for the interaction term from the SCENT interaction model. Prefixes logFC, AveExpr, t, P.value, adj.P.Value, and B (which respectively represent log2 fold change, average expression across all samples in log2 counts per million, test statistic, nominal P-value, Benjamini-Hochberg corrected P-value, and log-odds of differential expression) are limma outputs from differential analyses of Ts21. The suffixes smATAC, smRNA, and lgRNA represent the limma analysis for multiome ATAC, multiome RNA, and large scRNA HSCs, where logFC > 0 is an increase in Ts21. The differential analyses were performed using pseudobulks.

Supplementary Table 20 - Genes dysregulated on chromosome 21

Supplementary Table 21 - Antibodies used in the study

Supplementary References

1. Chou, S. T. *et al.* Trisomy 21 enhances human fetal erythro-megakaryocytic development. *Blood* **112**, 4503–4506 (2008).
2. Pine, S. R. *et al.* Incidence and clinical implications of GATA1 mutations in newborns with Down syndrome. *Blood* **110**, 2128–2131 (2007).
3. Taub, J. W. *et al.* Prenatal origin of GATA1 mutations may be an initiating step in the development of megakaryocytic leukemia in Down syndrome. *Blood* **104**, 1588–1589 (2004).
4. Hasaart, K. A. L. *et al.* Mutation accumulation and developmental lineages in normal and Down syndrome human fetal haematopoiesis. *Sci. Rep.* **10**, 12991 (2020).
5. Gerstung, M., Papaemmanuil, E. & Campbell, P. J. Subclonal variant calling with multiple samples and prior knowledge. *Bioinforma. Oxf. Engl.* **30**, 1198–1204 (2014).
6. Martincorena, I. *et al.* Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880–886 (2015).
7. MacParland, S. A. *et al.* Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat. Commun.* **9**, 4383 (2018).
8. Poisson, J. *et al.* Liver sinusoidal endothelial cells: Physiology and role in liver diseases. *J. Hepatol.* **66**, 212–227 (2017).
9. Mouta Carreira, C. *et al.* LYVE-1 is not restricted to the lymph vessels: expression in normal liver blood sinusoids and down-regulation in human liver cancer and cirrhosis. *Cancer Res.* **61**, 8079–8084 (2001).
10. Qian, H. *et al.* Stabilins are expressed in bone marrow sinusoidal endothelial cells and mediate scavenging and cell adhesive functions. *Biochem. Biophys. Res. Commun.* **390**, 883–886 (2009).
11. Kamitani, S. *et al.* The RAMP2/CRLR complex is a functional adrenomedullin receptor in human endothelial and vascular smooth muscle cells. *FEBS Lett.* **448**, 111–114 (1999).
12. Lepreux, S. *et al.* Cellular retinol-binding protein-1 expression in normal and

fibrotic/cirrhotic human liver: different patterns of expression in hepatic stellate cells and (myo)fibroblast subpopulations. *J. Hepatol.* **40**, 774–780 (2004).

13. Yamazaki, S. *et al.* Nonmyelinating Schwann cells maintain hematopoietic stem cell hibernation in the bone marrow niche. *Cell* **147**, 1146–1158 (2011).
14. Sugiyama, T., Kohara, H., Noda, M. & Nagasawa, T. Maintenance of the hematopoietic stem cell pool by CXCL12-CXCR4 chemokine signaling in bone marrow stromal cell niches. *Immunity* **25**, 977–988 (2006).
15. Nakashima, K. *et al.* The novel zinc finger-containing transcription factor osterix is required for osteoblast differentiation and bone formation. *Cell* **108**, 17–29 (2002).
16. Aubin, J. E. & Triffitt, J. T. Chapter 4 - Mesenchymal Stem Cells and Osteoblast Differentiation. in *Principles of Bone Biology (Second Edition)* (eds. Bilezikian, J. P., Raisz, L. G. & Rodan, G. A.) 59–81 (Academic Press, San Diego, 2002). doi:10.1016/B978-012098652-1.50106-2.
17. Omatsu, Y., Seike, M., Sugiyama, T., Kume, T. & Nagasawa, T. Foxc1 is a critical regulator of haematopoietic stem/progenitor cell niche formation. *Nature* **508**, 536–540 (2014).
18. Wagner, T. *et al.* Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. *Cell* **79**, 1111–1120 (1994).
19. Doege, K. J., Sasaki, M., Kimura, T. & Yamada, Y. Complete coding sequence and deduced primary structure of the human cartilage large aggregating proteoglycan, aggrecan. Human-specific repeats, and additional alternatively spliced forms. *J. Biol. Chem.* **266**, 894–902 (1991).
20. Weinreb, M., Shinar, D. & Rodan, G. A. Different pattern of alkaline phosphatase, osteopontin, and osteocalcin expression in developing rat bone visualized by in situ hybridization. *J. Bone Miner. Res. Off. J. Am. Soc. Bone Miner. Res.* **5**, 831–842 (1990).
21. Xu, C. *et al.* Stem cell factor is selectively secreted by arterial endothelial cells in bone marrow. *Nat. Commun.* **9**, 2449 (2018).
22. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to

transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).

23. Wolf, F. A. *et al.* PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59 (2019).
24. Lange, M. *et al.* CellRank for directed single-cell fate mapping. *Nat. Methods* **19**, 159–170 (2022).
25. Kenswil, K. J. G. *et al.* Endothelium-derived stromal cells contribute to hematopoietic bone marrow niche formation. *Cell Stem Cell* **28**, 653-670.e11 (2021).
26. Gray, D. L., Dicke, J. M., Dickerson, R., McCourt, C. & Odibo, A. O. Reevaluating humeral length for the detection of fetal trisomy 21. *J. Ultrasound Med. Off. J. Am. Inst. Ultrasound Med.* **28**, 1325–1330 (2009).
27. Longo, D., DeFigueiredo, D., Cicero, S., Sacchini, C. & Nicolaides, K. H. Femur and humerus length in trisomy 21 fetuses at 11-14 weeks of gestation. *Ultrasound Obstet. Gynecol. Off. J. Int. Soc. Ultrasound Obstet. Gynecol.* **23**, 143–147 (2004).
28. Weisz, B. *et al.* Association of isolated short femur in the mid-trimester fetus with perinatal outcome. *Ultrasound Obstet. Gynecol. Off. J. Int. Soc. Ultrasound Obstet. Gynecol.* **31**, 512–516 (2008).
29. Gavris, M. *et al.* Bone Tissue in Down Syndrome Patients Deteriorates Following Aging: A Study Based on Bone Ultrasound Analysis. *J. Med. Ultrasound* **22**, 29–36 (2014).
30. Popescu, D.-M. *et al.* Decoding human fetal liver haematopoiesis. *Nature* **574**, 365–371 (2019).
31. Luecken, M. D. *et al.* Benchmarking atlas-level data integration in single-cell genomics. *Nat. Methods* **19**, 41–50 (2022).
32. Jardine, L. *et al.* Blood and immune development in human fetal bone marrow and Down syndrome. *Nature* **598**, 327–331 (2021).
33. Kleshchevnikov, V. *et al.* Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* **40**, 661–671 (2022).
34. Bonnardel, J. *et al.* Stellate Cells, Hepatocytes, and Endothelial Cells Imprint the Kupffer

Cell Identity on Monocytes Colonizing the Liver Macrophage Niche. *Immunity* **51**, 638-654.e9 (2019).

35. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).
36. Gerber, H.-P. *et al.* VEGF regulates haematopoietic stem cell survival by an internal autocrine loop mechanism. *Nature* **417**, 954–958 (2002).
37. Zhou, B. O., Ding, L. & Morrison, S. J. Hematopoietic stem and progenitor cells regulate the regeneration of their niche by secreting Angiopoietin-1. *eLife* **4**, e05521 (2015).
38. Pajcini, K., Shao, L., Paik, N. & Sottoriva, K. Notch Ligand Jagged1 Is a Fetal Liver Niche Factor for the Function of Embryonic Hematopoietic Stem Cells. *Blood* **138**, 203–203 (2021).
39. Chung, E. & Kondo, M. Role of Ras/Raf/MEK/ERK signaling in physiological hematopoiesis and leukemia development. *Immunol. Res.* **49**, 248–268 (2011).
40. Waugh, K. A. *et al.* Triplication of the interferon receptor locus contributes to hallmarks of Down syndrome in a mouse model. *Nat. Genet.* **55**, 1034–1047 (2023).
41. Malle, L. *et al.* Autoimmunity in Down's syndrome via cytokines, CD4 T cells and CD11c+ B cells. *Nature* **615**, 305–314 (2023).
42. Muskens, I. S. *et al.* The genome-wide impact of trisomy 21 on DNA methylation and its implications for hematopoiesis. *Nat. Commun.* **12**, 821 (2021).
43. Massier, L. *et al.* An integrated single cell and spatial transcriptomic map of human white adipose tissue. *Nat. Commun.* **14**, 1438 (2023).
44. Katsumura, K. R., Bresnick, E. H., & GATA Factor Mechanisms Group. The GATA factor revolution in hematology. *Blood* **129**, 2092–2102 (2017).
45. Richard, C. & Verdier, F. Transferrin Receptors in Erythropoiesis. *Int. J. Mol. Sci.* **21**, 9713 (2020).
46. Forejtnikovà, H. *et al.* Transferrin receptor 2 is a component of the erythropoietin receptor complex and is required for efficient erythropoiesis. *Blood* **116**, 5357–5367

(2010).

47. Chopra, M., Langenhorst, D., Beilhack, A., Serfling, E. & Patra, A. K. Interleukin-2 critically regulates bone marrow erythropoiesis and prevents anemia development. *Eur. J. Immunol.* **45**, 3362–3374 (2015).
48. Machlab, D. *et al.* monaLisa: an R/Bioconductor package for identifying regulatory motifs. *Bioinforma. Oxf. Engl.* **38**, 2624–2625 (2022).
49. Castro-Mondragon, J. A. *et al.* JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **50**, D165–D173 (2022).
50. Satyanarayana, A. *et al.* RapGEF2 is essential for embryonic hematopoiesis but dispensable for adult hematopoiesis. *Blood* **116**, 2921–2931 (2010).
51. Balgobind, B. V. *et al.* EVI1 overexpression in distinct subtypes of pediatric acute myeloid leukemia. *Leukemia* **24**, 942–949 (2010).
52. Meylan, P., Dreos, R., Ambrosini, G., Groux, R. & Bucher, P. EPD in 2020: enhanced data visualization and extension to ncRNA promoters. *Nucleic Acids Res.* **48**, D65–D69 (2020).
53. Isakova, A. *et al.* SMiLE-seq identifies binding motifs of single and dimeric transcription factors. *Nat. Methods* **14**, 316–322 (2017).
54. Eferl, R. & Wagner, E. F. AP-1: a double-edged sword in tumorigenesis. *Nat. Rev. Cancer* **3**, 859–868 (2003).
55. Fulco, C. P. *et al.* Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).
56. Kachuri, L. *et al.* Genetic determinants of blood-cell traits influence susceptibility to childhood acute lymphoblastic leukemia. *Am. J. Hum. Genet.* **108**, 1823–1835 (2021).
57. Vuckovic, D. *et al.* The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell* **182**, 1214-1231.e11 (2020).
58. Yu, F. *et al.* Variant to function mapping at single-cell resolution through network propagation. *Nat. Biotechnol.* **40**, 1644–1653 (2022).

59. Kurita, R. *et al.* Establishment of immortalized human erythroid progenitor cell lines able to produce enucleated red blood cells. *PloS One* **8**, e59890 (2013).
60. Jeon, S. *et al.* Genome-wide trans-ethnic meta-analysis identifies novel susceptibility loci for childhood acute lymphoblastic leukemia. *Leukemia* **36**, 865–868 (2022).