**Peer Review File**

**Manuscript Title:** Single-cell multi-omics map of human foetal blood in Down's Syndrome

## Reviewer Comments & Author Rebuttals

### Reviewer Reports on the Initial Version:

Referee #1 (Remarks to the Author):

In this manuscript (Marderstein et al.), the authors conducted a large-scale single-cell profiling of human fetal liver and bone marrow from healthy and Down Syndrome (DS) samples. The primary biological question of the study, which was presented by the authors in the Introduction, was the mechanisms of increased risk of leukemia, especially of myeloid origin, in DS individuals. The aneuploidy of chromsome 21 is known to cause expansion of specific hematopoietic cell populations in fetal liver and the authors hypothesized that such populations could survive and become a seed (or support) for the progression from the preleukemic state to full leukemia.

Thus they generated sinlge-cell mRNA (with three gating strategies), multimodal RNA+ATAC and Visium datasets for liver and bone marrow of healthy and DS fetuses. The RNA-seq and subsequent analyses recapitulated the known increase in HSC/MPP, erythroid and megakaryocyte populations and decrease in myeloid population. The Visium confirmed these trends, and identified heterogeneity in spatial colocalization between healthy and DS. The genes on the chromosome 21 were on average 1.5 fold highly expressed in the DS, and the expression of genes on the other chromosomes was affected in a cell-type-specific manner. Genes are also differentially expressed between liver and BM. The authors generated multimodal data and conducted several analyses including topic modeling and enhancer-gene linkage. The HSC in DS seemed to be more primed toward erythroid lineage. GWAS fine-mapped variants related to RBC count were enriched in the enhancers of cells from erythroid lineage. Genes targeted by putative enhancers in DS HSCs harboring fine-mapped RBC GWAS variants were more likely to be differentially expressed in DS, but it was not clear if these enhancers are specifically accessible in DS or generally accessible in HSCs.

Major comments:
1. I think the dataset from fetus samples of DS and healthy, which included multiple novel single-cell technologies, will be valuable resource to the field to test biological hypotheses about DS as well as hematopoiesis in general. However, the analyses presented in this manuscript are less focused, sometimes difficult to follow, and did not directly address the primary biological questions that the authors mentioned in the introductory section. Some of the basic analyses do not reflect the standard practice accepted in the field (as I elaborate below). Moreover, it was difficult to find solid and novel biological insights into normal hematopoiesis during development or to the pathogenesis of DS by reading the manuscript.

For example, the visium-based differential abundance test could be significantly affected by sampling bias, and the difference in spatial colocalization between healthy and DS was not presented in the context of these differential abundance or other biological hypothesis. Similarly, significant heterogeneity between healthy and DS was continuously presented in the results from

e.g., ligand-receptor analyses, trajectory analyses by PAGA and the dynamically expressed genes according to the trajectory, correlation between RNA and ATAC (by MIRA), concentration of cells in each topic modeling pathway, and peak-gene linkage correlation (by SCENT). However, in all instances, it was not clear to me what was their biological implication in the context of leukemic development (or other biology) in DS. In each of the analyses, I was not sure if the authors would like to show the physiological developmental processes/mechanism of hematopoiesis, or would like to address the question of leukemic development in the DS.

The comparison between liver and bone marrow was conducted multiple times, but it was not clear how these analyses are relevant to their primary question.

Cherry-picked gene names were sometimes presented (e.g., in the section of gene-environment interactions or DEG) and connected to the known biology of DS or hematopoiesis (oxidative stress/mitochondrial dysfunction/autophagy), but this speculation was not generally validated by experiments.

The putative causal variants and genes for RBC count were identified by using DS multiome data, but it was again unclear how this should be interpreted to understand the biology of DS.

In summary, while I agree with the authors that the datasets are very useful and valuable to the field, the analyses presented in the manuscript did not clearly and convincingly show the novel biological mechanism of the DS.

2. The authors conducted the clustering and cell type annotation in each tissue and each clinical category (DS and healthy) separately, while accounting for batches within each category, for the main part of the analyses (e.g., Figure 1). The clustering did not seem to result in a clear separation of each cluster (Figure 1b). Given this situation, and maybe the continuous nature of these cell types, there might have been a possibility of technical bias in the subsequent analyses such as differential abundance test (Figure 1c). While the authors might have been aware of this possibility and conducted the integrated analyses in Supplementary Figure 14, the systematic comparison between the pre- and post- integration was not performed. Because the data integration is usually conducted in such analyses, I would like to confirm if the results were robust after the data integration, such as the concordance of cell type annotation and the differential abundance test between DS and healthy. Also, it could be considered to perform reference-query mapping, since relevant datasets are available from previous publications such as Jardine et al 2021, Human Cell Atlas, and Suo et al 2022.

3. Given the large magnitude of differences in the expression profiles of genes on the chromosome 21, I wonder how the log-normalization (for each cell) of genes on the other chromosomes was affected just by having the larger denominator from chrom 21 in normalization. Are the DEGs expected due to having the additional dosage of genes on the chromosome 21, or is it the biological truth beyond the expectation (e.g., Figure 2a)? Are they in line with the recent publication Malle et al. Nature 2023?

4. I noticed an outlier sample in the composition of major cell types in Figure 1c (Ts21 sample 8). More generally, I wonder how the post conceptual weeks or different gating strategies of the fetuses were accounted for in the major analyses in the manuscript.

5. It was not clear why the authors only focused on RBC GWAS for the enrichment of causal variants within HSCs and identification of causal variant-gene pairs. As for the technical aspect of the analysis, I would like to confirm how the authors conducted the Fisher's exact test for the enrichment. How were the background regions or variants selected to be compared against SCENT peaks and RBC fine-mapped variants?

Minor:
1. I noticed that the reference numbers might not be correctly assigned in the main text (at least for the last several papers), which made reading of the paper challenging.
2. I am afraid there are figure mislables in the main text. Figure 4e appeared three times in different context.
3. The main text says "Indeed, cell cycle analysis confirmed that HSC/MPPs, both in Ts21 (G2/M+S=29.8%) and healthy liver (21.8%), were cycling more compared to their counterparts in BM (26.9% in Ts21 and 18.3% in healthy) (Figure 2d)." but Figure 2d shows the comparison between healthy vs DS.

Referee #2 (Remarks to the Author):

In this study, Marderstein et al. reported an in-depth characterisation of human prenatal hematopoiesis in both fetal liver and bone marrow using cutting-edge single cell approaches (scRNA-seq, scATAC-seq and Spatial transcriptomics), and extend our understanding of the molecular and cellular features perturbed in Down syndrome during fetal life. In a series of elegant bioinformatic analyses, the Authors showed that:
- Trisomy 21 significantly perturbed cellular composition of the hematopoietic system in the fetal liver and bone marrow,
- Perturbation of the hematopoietic system in Down syndrome is a result of both cell-intrinsic Ts21-induced and environmentally driven mechanisms,
- Spatial organisation and cell to cell interactions are perturbed in trisomic fetal livers,
- There is a higher concordance between chromatin accessibility and gene expression in trisomic hematopoietic stem cells,
- GWAS SNP analyses strongly suggest that trisomic HSC are `primed` toward the erythroid lineage in the fetal liver.
Altogether, this study describes the largest atlas of altered fetal hematopoiesis in Down syndrome realised at the single cell level to date, thus represents an outstanding resource of datasets for future work in the field of normal and pathologic hematopoietic development. This study also provides novel insights into the mechanisms of leukemia predisposition seen in this population.

However, several limitations remain and need to be addressed:

1- Some variability is seen in the cell composition in fetal livers and bone marrows between biological and technical replicates (supp Table). This is particularly visible for sample 8 in Figure 1c. How can this be explained? Could the cell composition in bone marrow be represented as in figure 1c (CD45+), and in Supp figure 5. Are the differences seen between samples in the liver conserved in the bone marrow? This would also facilitate the direct comparison between fetal liver and bone marrow from the same primary sample (15619, 15582, 15712…). Cellular composition of Fetal bone marrow in both normal and Down syndrome context have been described previously (Jardine et al.), are the cellular difference seen in this new cohort concordant? Are there significant differences in the B cell composition between Ts21 and Healthy, and between FL and FBM? It is currently hard to see as they are in separated supp figures 3&4.

2- The exponential cross dependency is a very interesting observation. How can this be explained mechanistically? Can this be only due to environment as suggested? Could this also be driven by chromosome 21 genes that regulate chromatin accessibility or gene expression? Some Ts21 genes have been shown to be implicated in chromatin modification: HMGN1 and CHAF1B, along with many transcription factors ERG/ETS2/BACH1…. Can the Authors comment on this?

3- The Authors show that in Ts21 there is a greater concordance between chromatin accessibility and RNA expression, and speculate that this is due to signalling specific to Ts21? Are there arguments against a Ts21 driven clonal selection of the most adapted Ts21 cells, with a specific open chromatin and RNA expression, in this fetal liver environment? Something like a prenatal clonal hematopoiesis?

4- Stromal cell heterogeneity and cell to cell interaction within the bone marrow has been reported previously. Are these perturbed interactions also seen in the fetal liver. Can the Authors provide evidence of these perturbed interactions at the protein level using immunohistochemistry/immunofluorescence?

5- One key observation is the HSC/MPP heterogeneity. The Authors describe stem cell pattern composed of 3 branches with branch 1 being the `healthy` one, and branch 2/3 only seen in Down syndrome fetal liver; branch 2 being the one `primed` toward the erythroid lineage and may contain the clone that predispose to myeloid leukemia. Branch 3 is less well defined. Several questions remain here: what are the of them? Can the key characteristics (RNA genes, genesets, regulatory element…) of the TOPICS that define these branches be reported in supp? Can cell to cell interactions be estimated for each of these branches? Do these branches exist in Fetal bone marrow? Can these different branches of HSC sorted to compare their progeny at the single cell level as in figure 4f-h?

6- Branch 3 seems to be a more `by default` branch in Ts21 fetal liver with no clear association with transcription factors or strong geneset enrichment, apart from a transcriptional signature of HDAC inhibition… What does this means? Is this branch less prone to transformation in fetal liver? Where is this branch coming from? Can Velocity be used to estimate its origin? If it exists in the bone marrow, could this branch become in this different environment?

7- Since Down syndrome children are also predisposed to lymphoid leukemia (almost exclusively of B phenotype), one may wonder if some GWAS SNP B-cell specific are also over-represented. Has this been assessed? In fetal liver or fetal bone marrow? or even in HSC/CLP? Can Scavenge be used to determine B cell v. T cell traits?

8- This study is mostly focused on fetal liver hematopoiesis in Down syndrome, and strongly suggest that the most probable origin of TAM/ML-DS is within the branch 2 of Ts21 HSC/MPP. However, nothing is mentioned about GATA1. Have these Down syndrome samples (whole fetal livers or HSC/MPP) been tested for GATA1 mutations? Is the GATA1 locus particularly accessible?
9- Finally, most of the results are daescriptive, apart from the functional validation of the `primed` erythroid phenotype shown in Figure 4f-h, which correlates with earlier study from Jardine et al. There is a lot of speculative arguments about the potential impact of TFR2, TSPAN32 but not C11orf21, ERK1/2 pathway activation, oxidative phosphorylation, mitochondrial dysfunction, in fetal Ts21 HSC but no functional validation. This limits the impact of the manuscript as it has not been tested in this specific context (fetal liver HSC/hematopoiesis). Can some of these be tested functionally to strengthen the conclusions?

Minor comments:

1- Technical: population collected for sequencing were sorted and/or depleted, how does this affect the sequencing? And readout (cell composition)? Has this been tested? Has the purity of sorted cells been confirmed?
2- Genetic: are all Ts21 used in this study have a complete trisomy 21? No mosaic?
3- Age differences between healthy and Ts21 samples: can this explain some of the difference seen in cell composition? In fetal liver and in bone marrow?
4- Please show where is located the SNPs in Figure 4d-e?
5- Integration of healthy and Ts21 datasets has been performed in figure 3a, why not integrating the scRNA-seq data also?
6- How representative are spatial transcriptomic results shown in supp figure 6b/c? does supp figure 6c/d/e this include all samples?
7- Number of replicates (and cells) used for bioinformatic analyses should be added to the figure legends (Figure 2-4, and supp).
8- Has the multi-omic analyses has been performed in bone marrow as suggested by the title of the manuscript.

Referee #3 (Remarks to the Author):

In this paper the authors analyze single cell data from the liver and bone marrow of Down Syndrome and “healthy” fetuses. This is a bioinformatic intensive paper with all conclusions drawn from the analysis of gene expression and ATAC datasets.

A strength of this paper is the large number (>1 million) of cells analyzed that were obtained by sorting liver and marrow from a reasonable cohort of Down Syndrome fetuses. Importantly, both haematopoietic and stromal/microenvironmental cells were analyzed.

A significant finding (Figure 1c) is the marked differences in the proportions of CD45+ HSPCs, erythroid, and megakaryocyte cells vs. myeloid cells and B-cells in the fetal liver of Down Syndrome vs. ‘healthy” fetuses, respectively. Spatial analysis reveals differences in the cellular architecture of the liver, though it was difficult to understand what is being shown in the spatial analysis presented in Fig. 1e. Is the increase in hepatocyte frequency possibly due to a difference in hematopoietic elements in the spatial analysis or was the density of hepatocytes higher? If so, was there a significant difference in the overall size of the liver in Down Syndrome fetuses vs. ‘healthy’ fetuses?

Trajectory analysis of stromal and osteolineage cells showed marked differences between Ts21 and ‘healthy’ marrow but it is unclear what the different trajectories actually indicate in Fig. 1f/g. It is interesting that the differences in gene expression during osteolineage differentiation might be connected to skeletal abnormalities in adults with Down Syndrome. The trajectory analysis potentially linking endothelial cells to osteolineage cells, with differences noted in Ts21, is an intriguing finding.

The overexpression of chromosome 21 genes in all cell types examined is significant (supplemental Fig. 8). Did which genes that were overexpressed vary by cell type? The marked down-regulation of expressed genes did not occur in late erythroid cells (supplemental fig. 8). Could this contribute to alterations in newborns with Down Syndrome- polycythemia, macrocytosis?

It is an intriguing conjecture that Ts21 and ‘healthy’ HSC/MPP may have not only cell intrinsic differences but also different microenvironments and different responses to liver and marrow microenvironments. The analysis of both compartments is a strength of this paper.

The multiome analysis provides evidence that Ts21 HSCs have an erythroid bias (Fig. 3b) consistent with the polycythemia seen in Down Syndrome neonates. I did not understand the MIRA analysis with HSC ‘branches’ and ‘topics’ and look forward to a clarified text.

Importantly, “HSCs” were sorted from Ts21 and ‘healthy’ livers and cultured in vitro. These experiments provide functional evidence for the erythroid bias of these cells proposed by multiple bioinformatic analyses.

While the authors bring up the marked predisposition of young children with Down Syndrome to leukemia (ML-DS) based on mutations in GATA1 resulting in a shortened form of the protein (GATA1s) it is not clear how the findings in this paper relate to this at all, since there was no mention

in the findings of cells harboring GATA1s expression or mutations. Apparently, ~20-30% of Down Syndrome infants have detectable GATA1s cells. Is it possible from the transcriptomic data that were obtained to identify GATA1s hematopoietic cells? If not, then could some of the fetuses have GATA1s cells in various proportions effecting the results? Newborns with Down Syndrome without GATA1s have characteristic changes in their circulating red cells (polycythemia, macrocytosis) and platelets (many references). It is these changes that are reflected in the biological characteristics of hematopoiesis that are delineated in this paper.

Minor:

1.Page 2, bottom paragraph: CD45, despite its billing as a pan-haematopoietic marker, does not capture all haematopoietic cells (or, I suspect, even most of them).

2. Page 3, middle paragraph: it is stated that the "number" of myeloid cells was decreased in the Ts21 liver. Is it the 'number' or the 'proportion' of myeloid cells that are decreased?

3.I'm not sure if "healthy" is the most appropriate term for the control fetuses. How was their 'health' assessed? Might 'normal' or 'disomic', or some other term, be more appropriate?

4.Page 4, middle paragraph: "We next assessed which ligand receptor pairs are involved…". Did you actually assess which are involved or did you look for the presence of putative/potential/possible interactions based on mRNA expression?

5.Page 4, middle paragraph: The ligand-receptor pair analysis "revealed activation of the ERK1/2 cascade." I assume that activation requires phosphorylation. Might 'engagement' be a more appropriate term for 'activation'?

6. In Figure 3a it is curious that 2 types of macrophages but no monocytes are identified.

7. Page 10, third paragraph: the 2 references to 'Figure 4e' should be 'Figure 4g'.

Referee #4 (Remarks to the Author):

Here, Marderstein and colleagues perform a comprehensive RNA and ATAC survey of an early time point of human fetal development to resolve molecular, niche and gene regulatory differences among comparable hematopoietic populations in down syndrome and controls. The authors should be commended on the depth and number samples considered in these analyses, including the use of orthogonal single-cell and spatial profiling methods. This includes in depth profiling of the hematopoietic niche which other fetal studies have largely ignored. The results presented, if true, represent a potentially valuable dataset and analytical results to inform future studies of leukemia development in children. Unfortunately, the manner in which data is presented and analyzed raise significant concerns that hinder the ability of this reviewer to confidently believe the single-cell and molecular findings. In particular, statistical rigor is lacking in the single-cell comparisons due to imprecise inference and comparison of cell populations which likely biases the reported transcriptional, DNA accessibility and spatial differences. The authors further miss an opportunity to explore and validate cellular pathway impacts in each of the identified cell-states, beyond simply separating them into chr21 versus non-chromosome 21 and focusing on HSC. Further, with the exception of Figure 1d, it does not appear that the samples are considered as independent biological replicates. More importantly, there is insufficient experimental validation of the author generated hypotheses (niche, ligand-receptor, gene regulatory) based on these data. Finally, it is not clear what the novel verified findings are or mechanisms of action beyond prior literature (increased cell-cycle gene expression induction Ts21 HSC, mitochondrial dysfunction, etc.).

Major concerns

A particular problem common to single-cell atlas-level disease association papers is that individual specimens and donors are often not considered as covariates in the analysis, but rather a mix of cells that ignores their inherent heterogeneity (race, genetics, sex, developmental age, etc.). Are the results statistically robust considering replicate samples as pseudobulks (e.g., not driven by a few samples of a certain sex)?

There are significant concerns regarding the computation of cell cluster frequency, cell identity and gene expression differences, when separate versus joint-integrated clustering is performed on independent aggregated datasets (Ts21 liver, Ts21 marrow, healthy liver, healthy marrow alone). Hematopoiesis has been described as a cloud, with cluster subdivisions within that cloud remaining highly inconsistent between different sample sets and clustering methods. Here, the authors infer similarity between healthy and Ts21 based on similar markers. While this apples-to-oranges comparison is unfortunately often reported in literature, it is subject to inherent biases in clustering of different datasets. This is most notable for cycling HSCs/MPP, which are frequently observed in human adult marrow by scRNA-Seq (PMC6296228), but are reported to be absent in all but Ts21 liver (a major claim of the paper). The authors attempt to address this in Supplementary Figure 14 through a joint analysis of samples, but it appears that the cell frequency results reported throughout are on the inferred analogous clusters independently determined versus using this integrated tissue/disease combined dataset, without confirming results in the clusters derived from all samples. The authors should not only verify the frequency differences using these two orthogonal methods but include similar analyses of label projection to wild-type only cells (e.g., using the

combined joint clusters since only a few wild-type cells should be present for all clusters), in addition to flow cytometry validation, to ensure the accuracy of these predictions. It would be expected that a robust integration (e.g., Figure 3a) should enable the identification of all liver and marrow populations (healthy and disease).

Are the cell frequencies of populations matching between the snRNA-Seq and scRNA-Seq for the same samples? While snRNA-Seq will capture some larger cells, overall, the cell frequencies should be nearly identical. The presence of matching samples with two technologies further enables internal verification of whether the integration between samples and technologies worked. Specifically, the authors should visualize marker genes for all cells in a few matching samples for all clusters (high resolution clusters) using heatmaps, which show results for all cells and selected genes rather than an average, to illustrate whether Harmony is accurately matching clusters and cells (gene expression markers are consistent and coherent across technologies).

Similar to the above concern, if cell2location is only trained with single-cell references from the matching single-cell samples (e.g., Ts21), it will not be able to assess the presence of cell profiles such as cycling HSCs/MPP in the normal liver samples which are missing in this reference. The authors should provide the H&E images side-by-side with the Visium profiles to confirm that they are comparing similar anatomic structures. Orthogonal experimental validation of Ts21 differences in niche organization and perturbed receptor-ligand interactions are further required to support the author claims.

What signaling pathways are perturbed in the distinct cell-types (beyond simply chr21 versus non) that are likely to impact leukemia predisposition? The authors appear to really only look at HSC in any depth. Will some of the samples at this time point have GATA1 mutations? Did the authors look and if so, are the associations driven by fetuses with these mutations?

The datasets and code should be made available upon review (reviewer access or other tokens provided), to ensure completeness of the resource and code descriptions.

Corrections for ambient RNA are needed to ensure accurate assessment of scRNA-Seq and snRNA-Seq gene expression differences, which contribute substantially to false positive gene expression changes when cell frequency differences are significant among captures or disease. In most cases, >40% of multiome snRNA-Seq counts are due to ambient contamination from lysed nuclei (PMC7335186).

The authors need to clarify differences in their Ts21 observations and those of Jardine et al. 2021 (PMC7612688), which apply a similar single-cell approach. What accounts for the inconsistent findings and do the authors' findings hold up if they supervise their clusters onto Jardine’s.

It is possible the single-cell skewing observed in this study is associated with developmental immaturity which can be addressed through the reanalysis of existing single-cell atlases. In particular, prior studies that perform deep temporal and spatial profiling of fetal hematopoiesis, spanning skin, liver, marrow and aorta, using these same technologies.

It is unclear how environmental impacts (e.g., if a gene is commonly or uniquely upregulated in HSC in liver and marrow in Ts21 versus controls) are determined, since different clusters are annotated in the different tissues and samples (apples-to-oranges comparison). How are the authors aligning these cell-states and confirming their accuracy, beyond HSC?

The inferred role of SNPs with C11orf21 and TSPAN32 are too inferential and thus require orthogonal experimental validation.

Minor Concerns

“Iterative clustering” is not clearly described or cited.

Cell population marker genes and differentially expressed genes for all cell population comparisons (with statistics) are not provided along with cell-barcode to cluster associations (Supplementary Tables).

In Supplementary Figures 3 and 4 different colors are used for the same cell annotations which is confusing. Further, cell type annotations are displayed in different orders in most plots. The reviewer recommends the authors reorder cell annotations consistently and by lineage and maturity. What is label 30X? I cannot find this in the UMAP.

Referee #5 (Remarks to the Author):

The provenance and ethical use of the human foetal bone and liver samples has been documented in the manuscript's ethics statement. Per the authors' statement, these samples have been obtained from the Joint MRC/Wellcome Trust (Grant MR/R006237/1) Human Developmental Biology Resource (HDBR), with maternal informed consent, in accordance with ethical approval by the National Health Service (NHS) Research Health Authority, REC Ref: 18/LO/0822. HDBR is regulated by the UK Human Tissue Authority and operates in accordance with the relevant HTA Codes of Practice. The HBDR has obtained ethical approval to collect and distribute embryonic and fetal material to all UK based research projects, without the need for individual researchers to obtain their own project specific ethics, providing their project fits within the remit of the HDBR approval (as documented in the HBDR website which contains the ethics approval letters, https://www.hdbr.org/ethical-approvals). The study, which is funded by the European Research Council, seems to fit within this scope.

Given the HBDR authorization, the investigators/authors were not required to make project-based applications for ethical approval. They will be deemed to have ethical approval from the HBDR ethics committee and are able to rely on the attestation pertaining the ethical provenance of the tissues, including the obtention of donor informed consent (maternal). Of note the ethical approval of the HBDR is set to expire in June and December 2023 (Newcastle and London). I hence do not have particular concerns pertaining the provenance and ethical research use of the samples, and therefore - from an ethics perspective, recommend the article for publication.

**Author Rebuttals to Initial Comments:**

We would like to thank the Reviewers for their useful questions and positive comments. All Reviewers acknowledged that we generated a comprehensive dataset and performed compelling data analysis but suggested that further functional validation, additional analysis, and rewriting parts of the text would strengthen our manuscript. We recognise that these are fair and important points and, therefore, we have designed and performed additional experiments that would allow us to address Reviewers' main points while being mindful that human foetal samples are not easy to acquire and that the amount of obtained material is limited.

To summarise we have performed the following experiments and analysis:

## A. Experiments

1) Performed dose response experiment to determine optimal Verapamil concentration for its use in point 2) and 3).

2) Measured mitochondrial mass in Ts21 (n=3) and disomic (n=3) foetal haematopoietic cells.

3) Measured oxidative stress in Ts21 (n=3) and disomic (n=3) foetal liver hematopoietic cells.

4) Compared the abundance of immunophenotypic HSCs, progenitor cells and mature cells in Ts21 (n=3) and disomic (n=3) using 22 antibody panels for flow cytometry.

## B. Analysis

1) Compared the validity of our original cell type annotations within our scRNA-seq data using new annotations derived from multiple other methods:

(i) Harmony-based integration and clustering,

(ii) scVI-based integration and clustering,

(iii) reference-query mapping of disomic liver using Pulpescu et al. Nature 2019, and

(iv) reference-query mapping of disomic liver using the Ts21 liver separate-dataset annotations,

(v) reference-query mapping of disomic femur samples using Jardine et al. Nature 2021

2) Assessed the robustness of differential abundance and differential expression analyses to the choice of annotations outlined in point (1).

3) Assessed the impact of covariates, pseudobulking, additional chromosome 21 dosage, and single-cell sequencing technology on the statistical results.

4) Expanded analysis regarding the integration of fine-mapped GWAS SNPs and multiome data.

5) Analyzed the cis-regulatory activity of somatic mutations from Hasaart et al, 2019 from foetal Ts21 and disomic HSCs in relation to HSC gene expression profiles and ENCODE data.

6) Investigated further why cross-dependency exists between chromosome 21 and other chromosomes.

7) Contrasted GATA1 accessibility between Ts21 and disomic HSCs.

8) Examined somatic mutation acquisition using scRNA-Seq data

We further rewrote parts of the manuscript to improve its clarity, better describe our key findings, and make it more accessible to a broader audience. In particular, we have rewritten the Abstract, Introduction and Discussion sections, and modified major sections of Results such as those related to the multiome analysis. Please find below an outline of our point-by-point response to Reviewers comments together with additional details of the experiments and bioinformatics analysis that we did.

Reviewer #1

In this manuscript (Marderstein et al.), the authors conducted a large-scale single-cell profiling of human fetal liver and bone marrow from healthy and Down Syndrome (DS) samples. The primary biological question of the study, which was presented by the authors in the Introduction, was the mechanisms of increased risk of leukemia, especially of myeloid origin, in DS individuals. The aneuploidy of chromsome 21 is known to cause expansion of specific hematopoietic cell populations in fetal liver and the authors hypothesized that such populations could survive and become a seed (or support) for the progression from the preleukemic state to full leukemia.

Thus they generated sinlge-cell mRNA (with three gating strategies), multimodal RNA+ATAC and Visium datasets for liver and bone marrow of healthy and DS fetuses. The RNA-seq and subsequent analyses recapitulated the known increase in HSC/MPP, erythroid and megakaryocyte populations and decrease in myeloid population. The Visium confirmed these trends, and identified heterogeneity in spatial colocalization between healthy and DS. The genes on the chromosome 21 were on average 1.5 fold highly expressed in the DS, and the expression of genes on the other chromosomes was affected in a cell-type-specific manner. Genes are also differentially expressed between liver and BM. The authors generated multimodal data and conducted several analyses including topic modeling and enhancer-gene linkage. The HSC in DS seemed to be more primed toward erythroid lineage. GWAS fine-mapped variants related to RBC count were enriched in the enhancers of cells from erythroid lineage. Genes targeted by putative enhancers in DS HSCs harboring fine-mapped RBC GWAS variants were more likely to be differentially expressed in DS, but it was not clear if these enhancers are specifically accessible in DS or generally accessible in HSCs.

Major comments:
1. I think the dataset from fetus samples of DS and healthy, which included multiple novel single-cell technologies, will be valuable resource to the field to test biological hypotheses about DS as well as hematopoiesis in general. However, the analyses presented in this manuscript are less focused, sometimes difficult to follow, and did not directly address the primary biological questions that the authors mentioned in the introductory section. Some of the basic analyses do not reflect the standard practice accepted in the field (as I elaborate below). Moreover, it was difficult to find solid and novel biological insights into normal hematopoiesis during development or to the pathogenesis of DS by reading the manuscript.

The novel biological insights include:

1) Gene-environment interactions impact phenotypic differences between liver and BM haematopoietic cells, as evidenced by environment-dependent gene expression changes in Ts21 foetuses. (Figure 2)

2) We discovered an exponential relationship between dysregulated genes on chr21 and non-chr21 genes which is not driven by a single TF or chromatin regulator on chr21. Instead, our

data suggests feedback loops and/or synergistic regulation among overexpressed genes in close proximity. (Figure 2a)

3) Ts21 liver HSCs exhibited higher levels of cycling behaviour, oxidative stress, and mitochondrial dysfunction. Furthermore, while oxidative stress has been proposed as a mutational process involved in DS blood samples, we experimentally validated increased mitochondrial mass as a phenotype of Ts21 haematopoietic cells. (Figure 2b-2g)

4) Ts21 HSCs, while multipotent, exhibited an erythroid lineage bias. This may underlie the common clinical red blood cell phenotypes found in Down Syndrome, such as polycythaemia (increased number of erythrocytes) or macrocytosis (large erythrocytes). First, erythroid cell counts are higher in Ts21 fetal liver compared to disomic foetal liver. Ts21 HSCs were more likely to be predicted to differentiate towards the erythroid lineage, and had higher concordance between RNA and ATAC which further suggests early lineage priming in Ts21 HSCs. Finally, we validated HSC's lineage bias in vitro. (Figure 3, Figure 4h-j)

5) Expression of top DE Ts21-specific genes cannot be fully modelled by local (i.e. cis) chromatin accessibility suggesting that additional signalling is necessary to enact transcription of erythroid lineage related genes in Ts21 HSCs. (Supplementary Figure 16)

6) Ts21 significantly perturbed interactions between enhancers and their target genes (E-G). Many regulatory relationships present in disomic cells suffered widespread loss in Ts21, while new enhancer-gene relationships emerged in Ts21 due to general higher accessibility and gene expression. (Figure 4a-f)

7) We pinpointed active regulatory elements that affect erythroid lineage bias through the use of fine-mapped GWAS loci. Accessibility of these important regulatory elements differ in Ts21 HSCs compared to disomic HSCs, and ultimately impact expression of key genes. (Figure 4a-f)

8) We found that Ts21 mutations in the introns of Ts21 HSC-expressing genes are more likely to reside within cis-regulatory elements compared to intronic mutations identified in disomic HSCs.

Together, our findings allow us to hypothesize that abundant enhancer-gene restructuring gives rise to co-occurring haematological conditions. Further, we can now hypothesize that the interplay between altered chromatin organisation, gene expression dynamics, and the strain of oxidative stress coalesces to create an environment potentially conducive to pre-

leukemic and leukaemic changes. These findings illuminate key roles of microenvironment and genetic background in hematopoiesis, revealing new insights into DS pathogenesis. We have modified the Abstract, Introduction, and particularly the Discussion to reflect these findings.

For example, the visium-based differential abundance test could be significantly affected by sampling bias, and the difference in spatial colocalization between healthy and DS was not presented in the context of these differential abundance or other biological hypothesis.

We agree with the Reviewer that the Visium-based differential abundance test could be significantly affected by sampling bias. However, both scRNA-Seq and Visium could be impacted by sampling biases - scRNA-Seq due to the different sensitivity of cell types to digestion and Visium due to differences in parts of tissues that were sectioned and used for analysis. For that reason, we used Visium as an additional approach to verify differences in cell abundances observed in scRNA-Seq. We reasoned that since the same trends in terms of cell type abundances were detected using two independent approaches, there is a higher likelihood that the observed differences are true and not a result of technical biases.

It should be noted that sections for both disomic and Down's Syndrome Visum were taken from the remaining tissue after the part of the liver was used for the scRNA-Seq experiment i.e. tissues used for scRNA-Seq and Visium are matched. Considering that the foetal liver is 15-20 mm at 12-14 pcw (pea size) it is not feasible to make a specific choice of the area that will be used in Visum. Therefore, parts of the liver used for scRNA-Seq and for Visium were chosen randomly and no systematic biases in sample collection were possible. Furthermore, the presented results in Figure 1 are the average of all sections taken in Ts21 (n=20 sections; 2 consecutive sections from 10 biological replicates) and disomic (n=10 sections; 2 consecutive sections from 5 biological replicates) respectively, which together with the previous point makes it unlikely that the results were affected due to the systematic bias in sampling of tissue.

The Reviewer rightly pointed out that our study did not offer any hypothesis as to how differences in spatial colocalization of different cell types could affect cell abundances or their potential biological effects. Unfortunately, due to the nature of Visium technology, which does not offer single-cell spatial resolution of cells, there was a limitation in terms of the analysis we could perform. Also, there is a limitation in the availability of material we can obtain (both in terms of the amount of tissue and the number of foetal samples) and therefore we were constrained in the follow-up analysis we can perform. For those reasons we used Visium as a validation rather than hypothesis generating tool.

Similarly, significant heterogeneity between healthy and DS was continuously presented in the results from e.g., ligand-receptor analyses, trajectory analyses by PAGA and the dynamically expressed genes according to the trajectory, correlation between RNA and ATAC (by MIRA), concentration of cells in each topic modeling pathway, and peak-gene linkage correlation (by SCENT). However, in all instances, it was not clear to me what was their biological implication in the context of leukemic development (or other biology) in DS. In each of the analyses, I was not sure if the authors would like to show the physiological developmental processes/mechanism of hematopoiesis, or would like to address the question of leukemic development in the DS.

In this paper, our primary aim was to investigate the mechanisms underlying the dysregulation of haematopoiesis in Ts21. We believe that understanding these mechanisms is crucially linked to understanding the increased risk of leukaemia development. Independent from any acquired mutations and leukaemia development, newborn babies with DS frequently suffer from haematological abnormalities such as polycythemia (elevated number of erythrocytes) (REF). Therefore, the trisomy itself modulates gene expression, which leads to erythroid skewing. We were able to provide insights into how additional chr21 leads to observed haematological abnormalities by changing the interactions between enhancers and erythroid lineage genes. More specifically, we found that Ts21 HSCs have 4.1 times more peak-gene links and that the vast majority of these were significantly more accessible and upregulated in Ts21 compared to control. Furthermore, we found that RBC GWAS-harboring enhancers were more accessible in a subpopulation of Ts21 HSCs, and were enriched for association with genes that are differentially expressed between Ts21 and control HSCs. These results have two important implications:

1) they suggest a key role for peak-gene links in biasing differentiation of Ts21 HSCs towards the erythroid lineage.

2) they open up a possibility to link altered chromatin organisation in Ts21 with somatic mutations accrual thus providing a novel perspective on the etiology of leukaemia in DS. Notably, chromatin organization is well known to impact somatic mutation rates in different cancers.

To test our hypothesis that Down's syndrome impacts chromatin accessibility potentially leading to higher mutagenesis due to increased exposure to oxidative stress, we conducted an analysis using somatic mutations from Hasaart et al in foetal DS and disomic HSCs. We found that DS mutations located in the introns of Ts21 HSC-expressing genes are 60% more likely to reside within *cis*-regulatory elements compared to intronic mutations identified in disomic HSCs ($P = 3x10^{-4}$; Figure 1).

This result is in line with the hypothesis that altered chromatin accessibility landscape in Ts21 HSCs provides a milieu for mutation acquisition (possible due to increased exposure to

oxidative stress) in regulatory regions of genes. This potentially increases propensity of Ts21 HSCs to preleukaemia and leukaemia development.

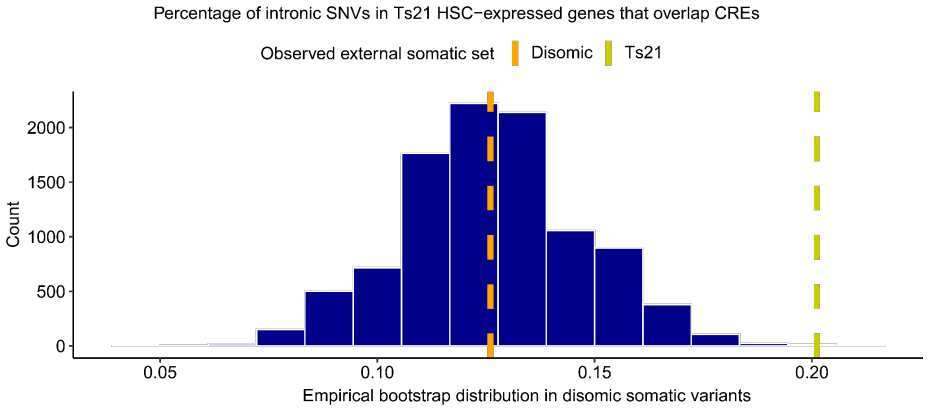



***Figure 1.*** *Cataloged DS mutations (by Hasaart et al.) in the introns of DS HSC-expressed genes (from our scRNA-seq) are more likely to overlap candidate cis-regulatory elements (CREs) from the ENCODE project than cataloged intronic mutations from disomic cells (also by Hasaart et al.). The histogram represents the empirical distribution of this overlap from 10,000 bootstrapped sets of somatic variants found in disomic HSCs.*

We have included this analysis in the Main Text Results.

The comparison between liver and bone marrow was conducted multiple times, but it was not clear how these analyses are relevant to their primary question.

The increase in somatic mutation rate in Ts21 foetal HSPCs compared to disomic is not sufficient to explain the rate of leukaemia development in DS children[1]. As most of the symptoms that occur in DS, including haematological abnormalities, are highly variable in both severity and penetrance across individuals, it is likely that genetic background and environment are significant factors[2]. One underappreciated factor is the haematopoietic microenvironment which together with the cell–cell competition and selection, is likely to play a role in the development of leukaemia in children with DS[1]. As Ts21 leukaemia is believed to originate from fetal liver HSCs[3,4], we reasoned that molecular changes (e.g. differential expression) that differ between the liver and femur and which are specific to Ts21 would be the most relevant to study in depth.

Therefore, understanding gene-environment interactions is important to fully understand the mechanisms underlying the observed haematological abnormalities. To assess the effect that

environment/niche has on the gene expression and phenotype of HSCs in Ts21 we compared foetal liver and bone marrow, two environments that HSCs inhabit during development.

We edited the manuscript to make this more clear.

Cherry-picked gene names were sometimes presented (e.g., in the section of gene-environment interactions or DEG) and connected to the known biology of DS or hematopoiesis (oxidative stress/mitochondrial dysfunction/autophagy), but this speculation was not generally validated by experiments.

To experimentally validate that DS foetal haematopoietic cells exhibit higher oxidative stress and changes in mitochondrial mass compared to disomic cells (Figure 2), we have now measured mtROS levels, using MitoSOX probe, and mitochondrial mass using the MitoTracker green FM dye. MitoSOX is a fluorogenic dye specifically targeting mitochondria in live cells. Oxidation of the MitoSOX reagent by mitochondrial superoxide produces bright green fluorescence. MitoTracker probes are cell-permeant mitochondria dyes that following incubation, passively diffuse across the plasma membrane and accumulate in the mitochondria of live cells. Compared to other dyes of the same family, MitoTracker green FM accumulates in the mitochondria regardless of mitochondrial membrane potential and is not sensitive to oxidative stress, making it a valuable tool to measure mitochondrial mass[5,6].

It has been reported that HSCs and progenitor cells (murine bone marrow and human cord blood) but not lineage positive cells can efflux the dye through xenobiotic efflux pumps (*Bcrp1* and *Mdr1a/b)* and therefore the level of fluorescence does not faithfully represent the level of mitochondrial mass. To circumvent the bias caused by preferential dye efflux from immature hematopoietic cells, we treated the cells with Verapamil to block the xenobiotic efflux pumps[7].

To validate the feasibility of the assays, we first stained disomic foetal haematopoietic cells with MitoTracker and assessed the mitochondrial mass of progenitor cells by flow cytometry. Within the foetal liver Lin- CD34+ CD45dim population, our data demonstrated the existence of two subpopulations stained differently for MitoTracker, named MTG_hi and MTG_lo. The MTG_lo population was enriched especially in immature CD38- cells (Figure 3, panel A). This is in line with the expected preferential dye efflux of more immature hematopoietic cells (de Almeida et al., Cell Stem Cell, 2017). Following treatment with Verapamil, which blocks efflux pumps, we observed an increase in MitoTracker dye retention, confirming the effectiveness of the assay in human foetal haematopoietic cells (Figure 3, panel B). Regarding mtROS levels, within the foetal liver Lin- C34+ CD45dim population, we also found a subpopulation that exhibited higher staining intensity for MitoSOX green (MSOX_hi) and a population that displayed lower staining intensity (MSOX_lo). Similarly to what we described for MitoTracker, the MSOX_lo population was enriched in CD38- cells (Figure 3, panel C), and treatment with 50 µM Verapamil increased the retention of MitoSOX dye in both CD38+ and CD38- populations (Figure 3, panel D).

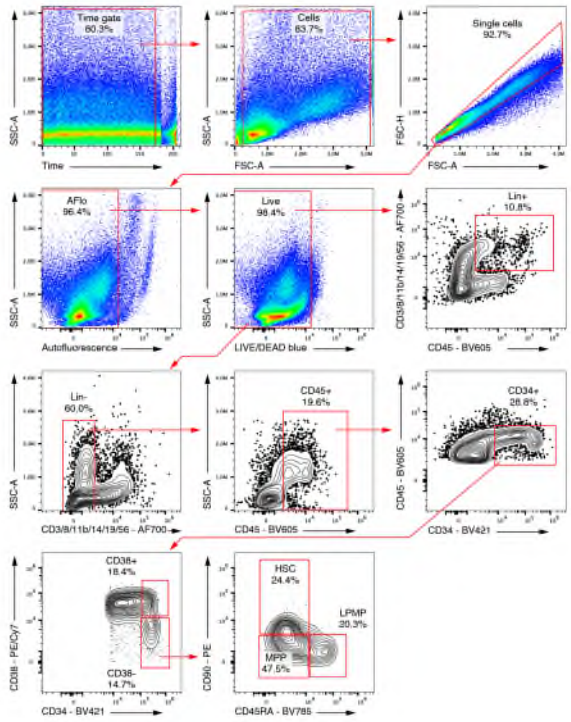



***Figure 2.*** *Representative gating strategy used for the mitochondrial probe experiments. Both MitoSOX and MitoTracker stainings shared the same antibody panel.*

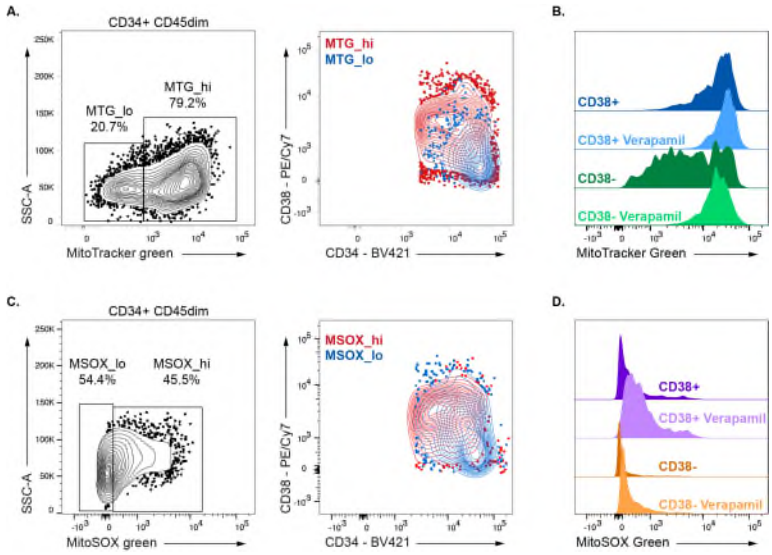



***Figure 3.*** *A. Left: Representative gating strategy used to identify MTG_hi and MTG_lo populations. Right: MTG_hi and MTG_lo cells were plotted against CD34 and CD38, showing that MTG_lo cells are enriched in the CD34+ CD38- fraction. B. Density plot of the MitoTracker Green FM fluorescence in CD38+ 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+. C. 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. D. 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+.*

Next, we compared the mitochondrial mass (using MitoTracker) and mtROS levels (using MitoSOX) in 3 Ts21 and 3 disomic foetal liver samples stained in the presence of Verapamil (Figure 4 and 5). We observed that Ts21 CD34+CD38+ cells had higher mitochondrial mass and mtROS levels compared to disomic progenitors, whereas CD34+CD38- and HSCs were comparable between the two conditions. Previous studies have shown that disomic HSCs exhibit very low electron transport chain activity and minimal mROS production as they mostly rely on glycolysis. In contrast, committed progenitors have much higher oxygen consumption, mitochondrial ATP production, and maximal respiratory capacity[8]. Taken together, our results support the hypothesis that as the HSCs and immature CD34+CD38- cells start to differentiate to committed CD34+CD38+ progenitors, they ramp-up oxidative phosphorylation. This leads to higher mtROS production in Ts21 cells compared to discomic cells, possibly due to their higher mitochondrial mass and dysfunction in the electron transport chain activity.

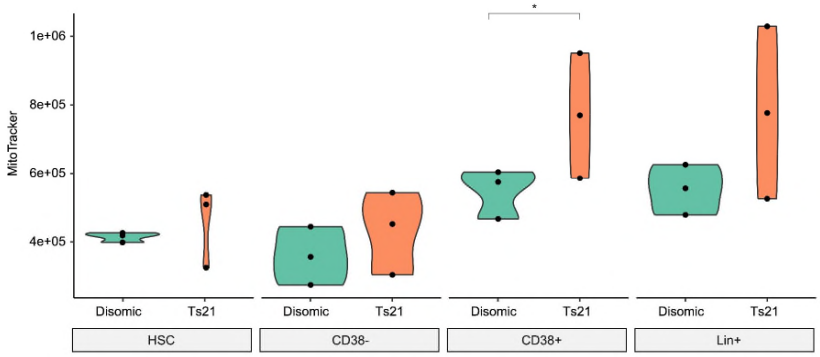



***Figure 4.*** *MitoTracker staining in disomic and Ts21 haematopoietic populations. HSC-haematopoietic stem cells (Lin-/CD34+/Cd38-/CD90+/CD45RA-); CD38- (Lin-/CD45+/CD34+/Cd38- population); CD38+ (Lin-/CD45+/CD34+/CD38+ population); Lin+ (CD3/8/11b/56/14/19, lineage positive population).*

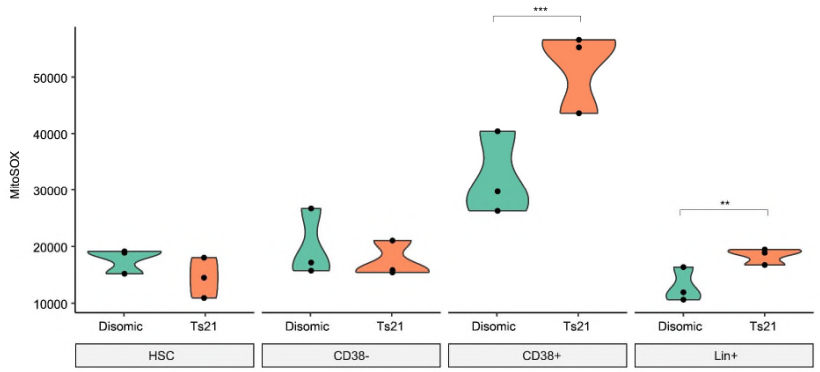


***Figure 5.*** *MitoSOX staining in disomic and Ts21 haematopoietic populations. HSC-haematopoietic stem cells (Lin-/CD34+/Cd38-/CD90+/CD45RA-); CD38- (Lin-/CD45+/CD34+/Cd38- population); CD38+ (Lin-/CD45+/CD34+/CD38+ population); Lin+ (*CD3/8/11b/56/14/19, *lineage positive population); P<0.05(*); P<0.01(**); P<0.001 (***).*

Experimental details are included in the Methods section of the revised manuscript within the section titled “MitoTracker and MitoSOX Data analysis”:

To test whether Ts21 cells have significantly different mitochondrial mass or mtROS from disomic cells, we fit a Gaussian generalized linear mixed model. Using single-cell values, we transformed the MitoSOX and mitochondrial mass values using a rank inverse normal transformation and used the transformed values as the response variable. While accounting for age as a fixed effect and sample as a random intercept, we tested for the effect of disease status (a fixed effect) within our model and determined significance using FDR < 0.1 across the 8 fitted models (4 cell types by 2 response variables).

The putative causal variants and genes for RBC count were identified by using DS multiome data, but it was again unclear how this should be interpreted to understand the biology of DS.

Our analysis suggested that HSCs in Ts21 showed higher probability to differentiate toward erythroid lineage, however the underlying mechanism was not clear. Here, we analyzed putative causal variants and genes for RBC count in the context of single-cell multi-omic data to understand how gene regulation could contribute to the observed erythroid bias.

First, we showed that a subpopulation of Ts21 HSCs have increased chromatin accessibility of fine-mapped GWAS SNPs relevant to RBC count compared to disomic HSCs. These SNPs are likely located in regulatory elements that are important for determining red blood cell count, and this result suggests that the regulatory element activity relevant for erythroid lineage/commitment differs between Ts21 and disomic HSCs.

To understand how these regulatory elements might contribute to the observed erythroid bias in Ts21, we linked them to their candidate target genes. Our analysis showed that the presence of trisomy alters the relationship between chromatin accessibility and gene expression in two ways: 1) many of the peak-gene links present in disomic HSCs were lost in Ts21 cells (and vice versa); and 2) there were many more peak-gene links in Ts21, due to the enhanced concordance between peak accessibility and gene expression in Ts21 HSCs compared to disomic HSCs (Fig 4c-d in the main text). This higher coordination of chromatin accessibility and gene expression could be triggered by external signaling that is unique to Ts21 HSCs. Finally, we found that the candidate target genes of regulatory elements relevant for erythroid bias (as determined by harboring a fine-mapped GWAS variant) were more likely

to be differentially expressed between Ts21 and disomic HSCs, suggesting that these regulatory elements indeed contribute to the Ts21 erythroid bias by altering expression of key genes.

Overall, we used putative causal variants and genes for RBC count to identify important regulatory elements, and ultimately distinguish how these regulatory elements contribute to gene dysregulation between Ts21 and disomic HSCs.

We made this more clear in the relevant Results sections of the manuscript.

In summary, while I agree with the authors that the datasets are very useful and valuable to the field, the analyses presented in the manuscript did not clearly and convincingly show the novel biological mechanism of the DS.

2. The authors conducted the clustering and cell type annotation in each tissue and each clinical category (DS and healthy) separately, while accounting for batches within each category, for the main part of the analyses (e.g., Figure 1). The clustering did not seem to result in a clear separation of each cluster (Figure 1b). Given this situation, and maybe the continuous nature of these cell types, there might have been a possibility of technical bias in the subsequent analyses such as differential abundance test (Figure 1c). While the authors might have been aware of this possibility and conducted the integrated analyses in Supplementary Figure 14, the systematic comparison between the pre- and post- integration was not performed. Because the data integration is usually conducted in such analyses, I would like to confirm if the results were robust after the data integration, such as the concordance of cell type annotation and the differential abundance test between DS and healthy. Also, it could be considered to perform reference-query mapping, since relevant datasets are available from previous publications such as Jardine et al 2021, Human Cell Atlas, and Suo et al 2022.

It is true that cell types exist on a continuum, and that 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[9] and scVi[10] methods) and 2) reference-query mapping (for disomic samples with both Pulpescu et al. Nature 2019 and our Ts21 data). Using the new clusters, we reperformed the cell type abundance comparison.

First, we integrated the full dataset using Harmony (Figure 6) 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 neighbor 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 (Figure 7), validating accurate annotation of cell types.

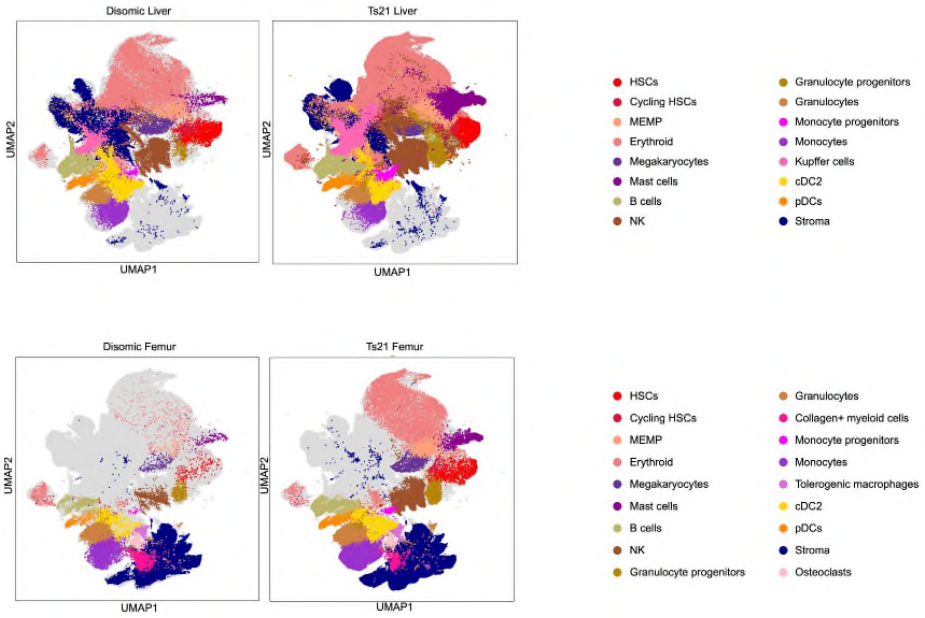



***Figure 6.*** *3D Uniform manifold approximation and projection (UMAP) visualisation of cells from Ts21 liver (n=780,299, k=15), Ts21 femur (n=162,775, k=12), disomic liver (n=110,671, k=3) and disomic femur (n=53,807, k=3) following integration of datasets using Harmony. Cells are coloured by broad cell type categories. n refers to the total number of cells, k indicates the number of biologically independent samples (foetuses).*

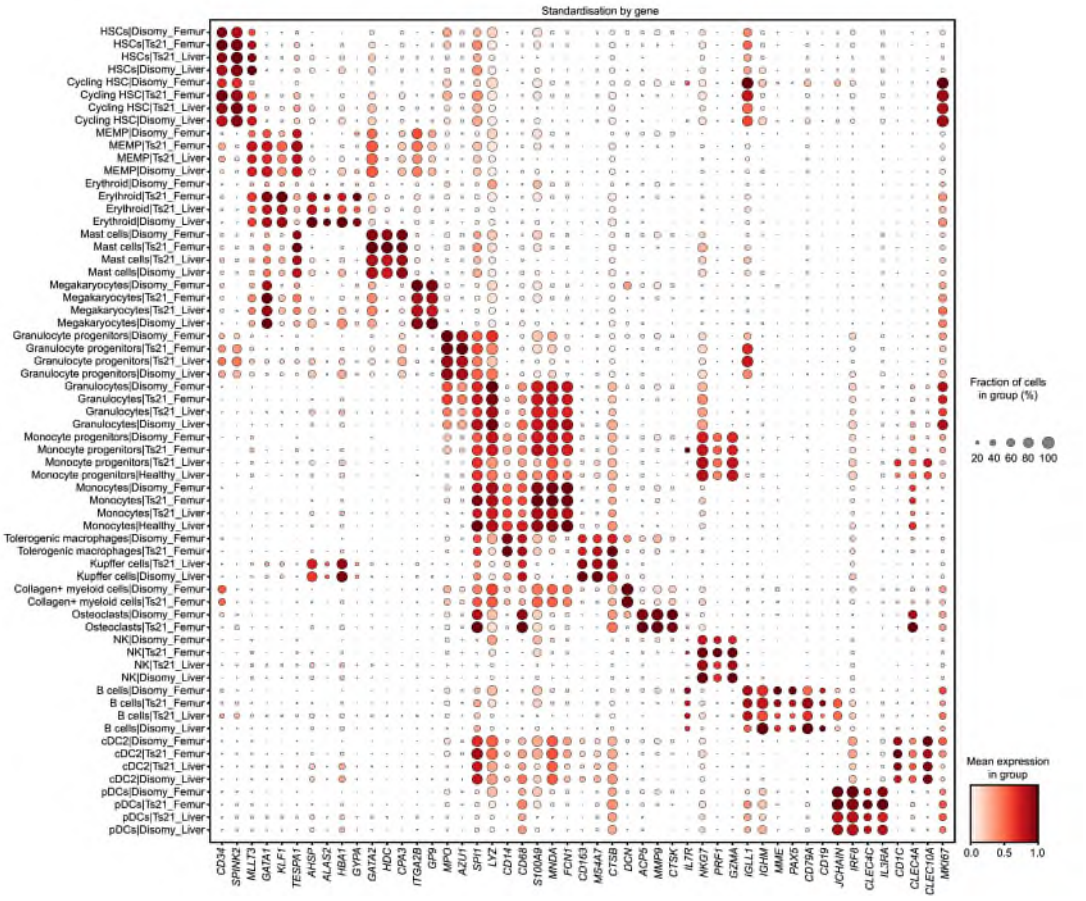
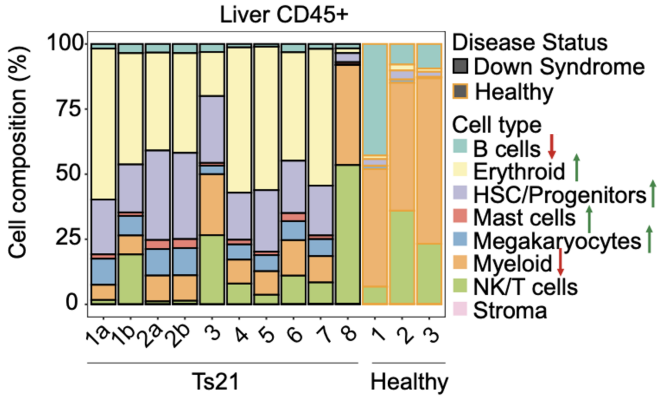



***Figure 7.*** *Gene expression dot plot of marker genes from Supp. Figure 3 and 4 across cell types identified from Harmony-based dataset integration.*

Next, we compared the Harmony integration-identified cell type abundance analysis to our separate cell type abundance analysis shown in Main Text Figure 1c.




***Main Text Figure 1c (copied from original submission).*** *Stacked bar plot of relative abundances of different broad cell types in stage-matched Ts21 (black outline) and disomic foetuses (orange outline). Arrows at the side of the labels indicate a significant increase (↑) or decrease (↓) in the number of cells in Ts21 liver, compared to disomic liver. Difference in cell*

*proportion was tested by Mann-Whitney U test and significance was determined by FDR < 0.1.*

The results were remarkably consistent between the two approaches (Figure 8). The only difference being that, while the trend remained consistent, mast cells were not anymore significantly differentially abundant (FDR = 0.103).

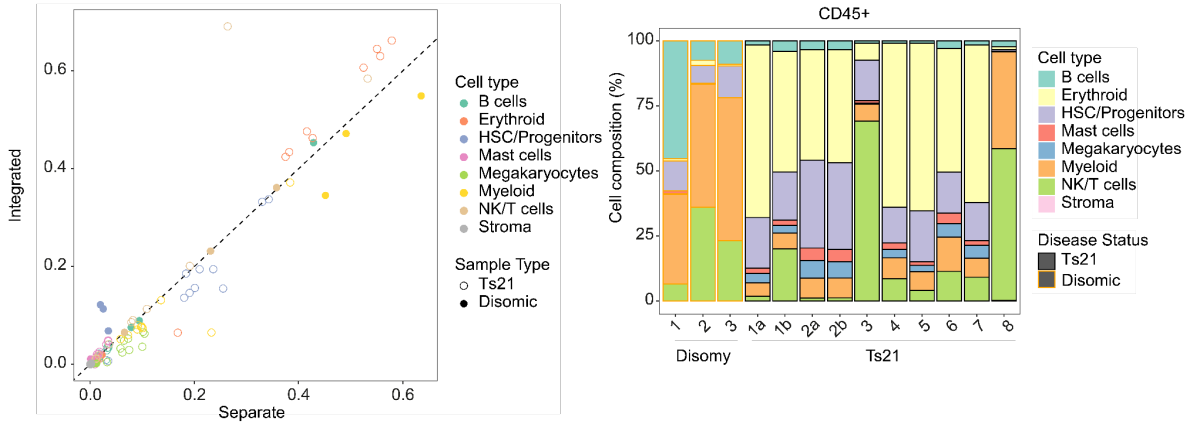



***Figure 8**. Using the Harmony-integrated cell type labels in stage-matched Ts21 and disomic foetuses, we compared cell type proportions between separate and integrated (left), and (right) made a stacked bar plot of relative abundances of different broad cell types similar to Main Text Figure 1c. In (left), each point represents the estimated abundance in a different sample. Panel (right) can be compared to Main Text Figure 1c, which has been copied above.*

Given 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) (Luecken et al. 2021 *Nat Methods*), we repeated the integration and clustering using scVI (Figure 9 and 10). 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 neighbor graph, computing Leiden clusters, and finally annotating clusters.

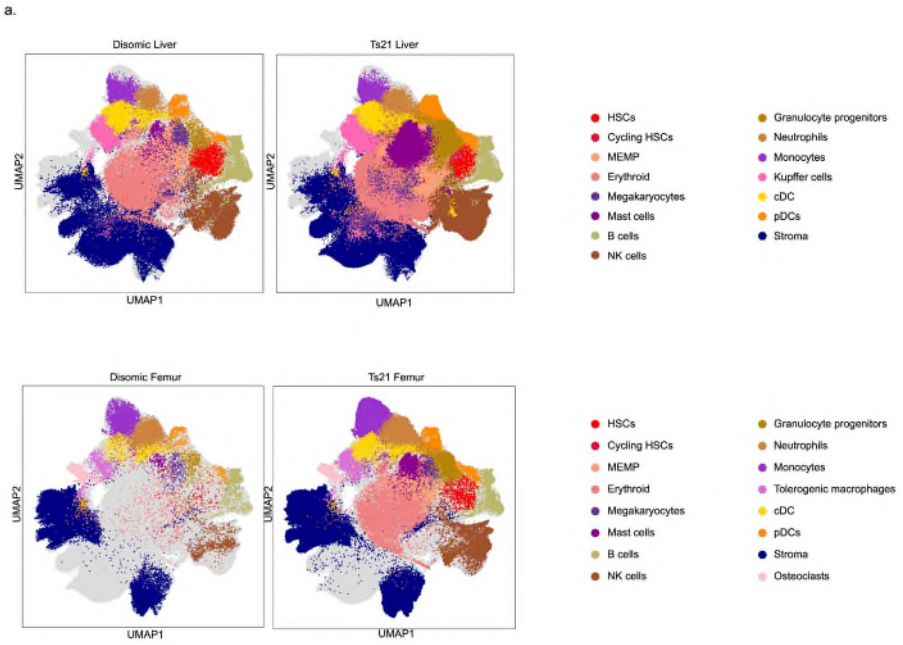



***Figure 9.*** *3D Uniform manifold approximation and projection (UMAP) visualisation of cells from Ts21 liver (n=780,299, k=15), Ts21 femur (n=162,775, k=12), disomic liver (n=110,671, k=3) and disomic femur (n=53,807, k=3) following integration of datasets using scVI tools. Cells are coloured by broad cell type categories. n refers to the total number of cells, k indicates the number of biologically independent samples (foetuses).*

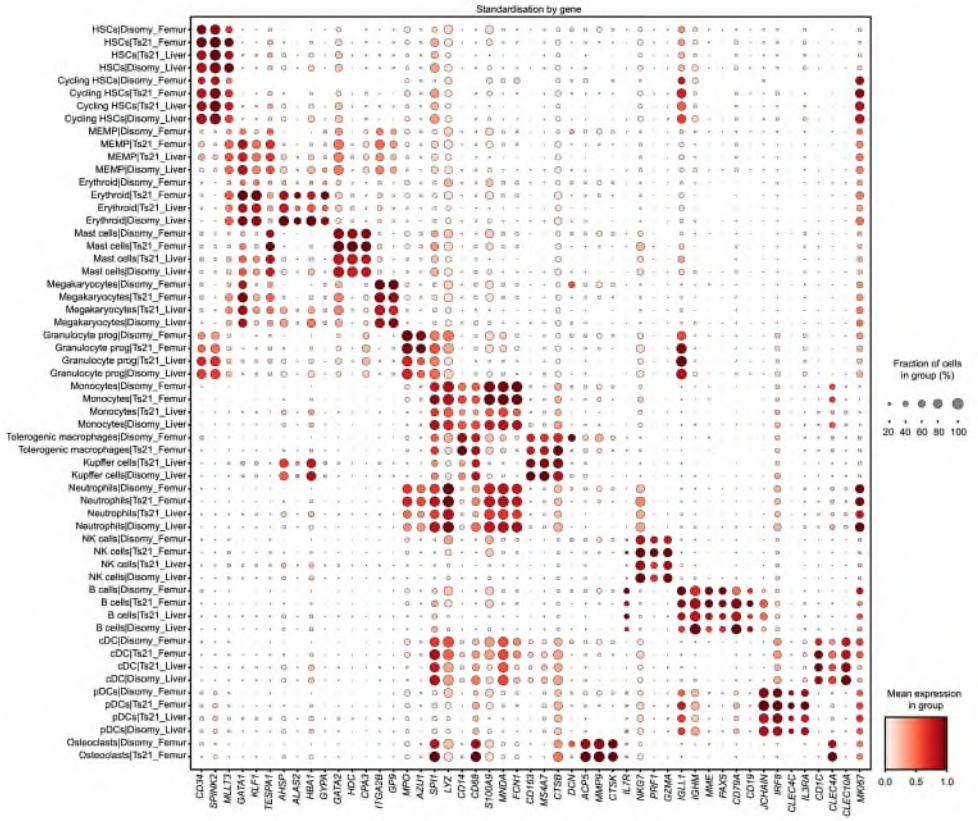


***Figure 10.*** *Dot plot showing expression of marker genes from Supp. Figure 3 and 4 across cell types identified following dataset integration using scVI tools.*

The scVI-integrated proportions across samples correlated closely with separate dataset proportions ($r$ = 0.957) (Figure 11 left), 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 (Figure 11 right).

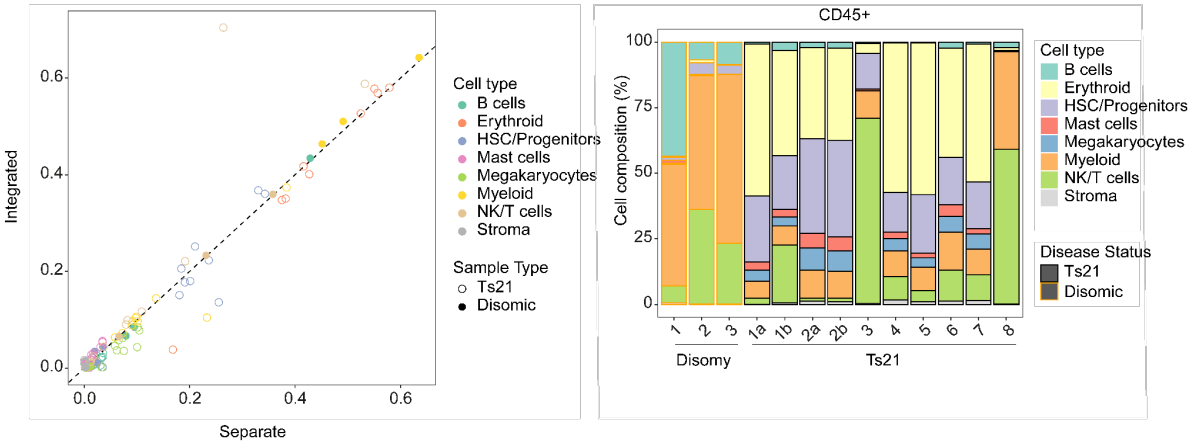



***Figure 11***. *Using the scVI-integrated cell type labels in stage-matched Ts21 and disomic foetuses, we compared cell type proportions between separate and integrated (left), and (right) made a stacked bar plot of relative abundances of different broad cell types similar to Main Text Figure 1c. In (a), each point represents the estimated abundance in a different sample.*

Next, we performed reference query mapping from Popescu et al. 2019 Nature 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. Given that the Ts21 liver cells were vastly different (due to presence of trisomy) and much larger than the Pulpescu et al. dataset, we did not transfer labels for the Ts21 data for this analysis and perform a cell-type abundance test. After using scVI to embed both our disomic liver dataset and the Pulpescu et al. liver dataset into a single matrix and predicting new cell-type labels for disomic liver cells using a KNN classifer (scHPL), we found that the new reference query labels correlated strongly with the previous annotations in the three disomic samples that were tested for cell-type abundance differences, helping validate our prior annotations ($r$=0.998; Figure 12, left).

(Detailed methods: 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 neighbors 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.)

Finally, 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; Figure 12, right). 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.

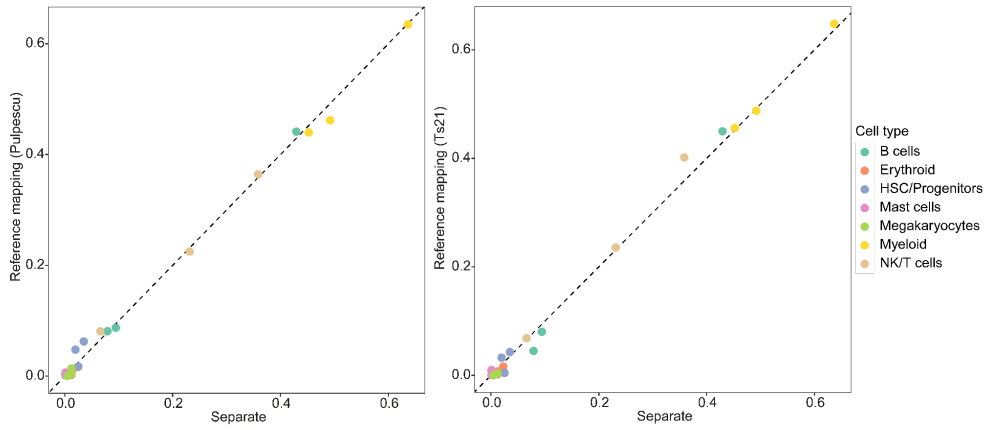



***Figure 12.*** *Comparison of cell type proportions for stage-matched disomic liver samples from the separate-dataset annotations to the reference query label transfer annotations using Pulpescu et al. (left) or the Ts21 dataset as a reference (right).*

Overall, our results demonstrate that the cell-type annotations are consistent across all annotation strategies and the differential abundance results are robust. We include these results within the Supplementary Results section of the manuscript, under “Evaluating cell-type abundances following data-set integration and label transfer”, and mention in the Main Text:

“We confirmed the robustness of our cell-type annotations and differences in frequency of cell types between Ts21 and disomic liver samples after integration using either Harmony or scVI,

as well as reference-query mapping from a published disomic foetal liver atlas and our Ts21 liver cells to disomic liver cells (Supplementary Results; Supp. Figure 7-9)."

3. Given the large magnitude of differences in the expression profiles of genes on the chromosome 21, I wonder how the log-normalization (for each cell) of genes on the other chromosomes was affected just by having the larger denominator from chrom 21 in normalization. Are the DEGs expected due to having the additional dosage of genes on the chromosome 21, or is it the biological truth beyond the expectation (e.g., Figure 2a)?

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 (fetus #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 (Figure 13), and comparing log fold change between the two analyses for all genes (Figure 14).

Overall, we found that, in the real data, the median log fold change on each chromosome was roughly centered around zero (Figure 13, 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 (Figure 13, 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 (Figure 14). Therefore, the increased expression on chromosome 21 has minimal impact on the differential expression results on other chromosomes. We include this analysis in the Supplementary Results.

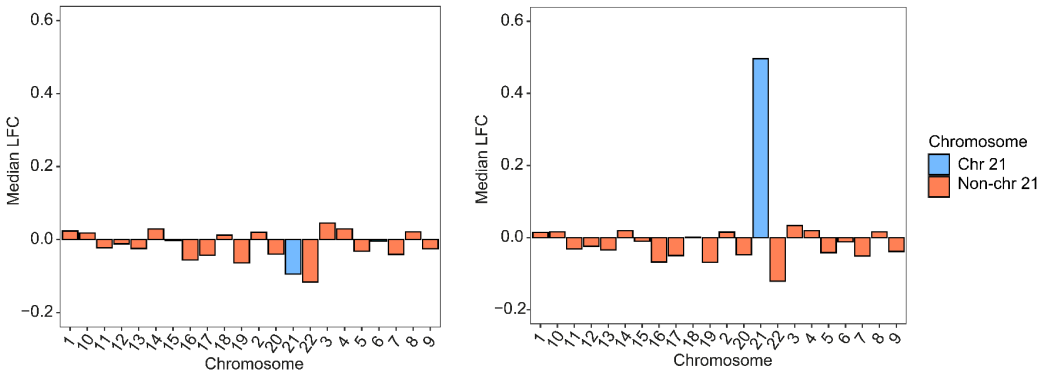



***Figure 13.*** *Left, median log-fold changes for genes on each chromosome between fetuses #15633 and #15781. Right, median log-fold changes for genes on each chromosome between fetuses #15633 and #15781, after increasing mapped reads to genes on chromosome 21 by 50% in two #15781 pseudobulk samples.*

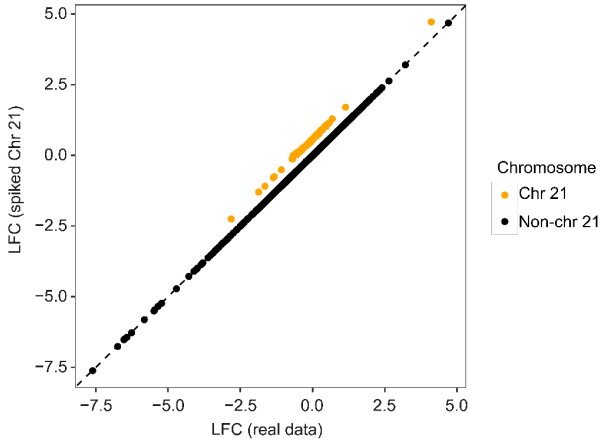



***Figure 14.*** *Comparison of log-fold changes for all genes between the two analyses. Orange points represent the genes on chromosome 21, which have increased expression within the two #15781 pseudobulk samples in the spiked chr21 analysis (y-axis). The log-fold changes of the underlying real data is along the x-axis.*

Are they in line with the recent publication Malle et al. Nature 2023?

We would like to thank the reviewer for drawing our attention to Malle et al. 2023 manuscript - “Autoimmunity in Down’s syndrome via cytokines, CD4 T cells and CD11c+ B cells“. However, in short, it is difficult to compare the two studies because while our study focused on gene regulation in trisomy during foetal development, Malle et al. focused on assessing immune cell function in adult DS patients. In the aforementioned study, the authors studied blood plasma (i.e. liquid, non-cellular part of blood) of adult DS patients and B and T cell

repertoire. They used a diverse set of immunophenotyping techniques such as BCR analysis, multiplex cytokine analysis, mass cytometry, flow cytometry, B cell stimulation, B and T cell co-culture, and auto-antibody analysis. This is very different from our study of the transcriptome and chromatin accessibility of blood cells along differentiation trajectory collected from foetal liver and bone marrow. Our results point to a highly cell-type and context-specific Ts21 effect on differential expression. If the reviewer has a specific suggestion in terms of analysis/comparison they would like to see, we would be happy to explore this further.

4. I noticed an outlier sample in the composition of major cell types in Figure 1c (Ts21 sample 8). More generally, I wonder how the post conceptual weeks or different gating strategies of the fetuses were accounted for in the major analyses in the manuscript.

In Main Text Figure 1c, Ts21 sample #8 has a unique cell type composition that is dissimilar from the other Ts21 samples. As the reviewer mentions, this could be due to technical or biology variability, and it does raise the question of whether such variables (post-conceptual weeks or different gating strategies) impact the results in the manuscript.

We explicitly accounted for different gating strategies in differential abundance and differential expression analyses. In the abundance analyses, we only analyzed samples that were sorted with the same procedure, which means that there should be no technical impact from different sorting strategies. For example, when calculating differences in blood cell abundance in Ts21 and disomic, only samples obtained with the same sorting strategy (CD45+) were used. In differential expression analyses, we accounted for sorting as a fixed effect within our statistical models. Therefore, gating strategy is already accounted for in the existing analyses.

In response to the comment about post conceptual weeks, we conducted an exploratory analysis of its associations in our data (Figure 15). We first estimated cell type composition for the 8 major cell types detailed in Figure 1c across disomic and Ts21 samples (Figure 15). 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 the cell distribution across PCW (nominal $P > 0.05$ for all tests). We visualized 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, we did not consider PCW further in cell abundance results.

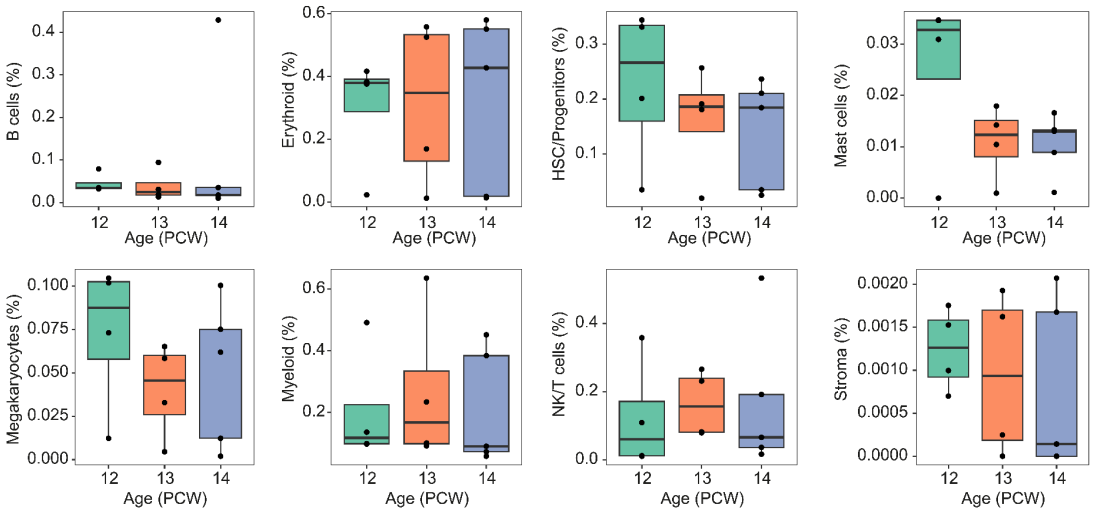



***Figure 15.*** *The relationship between post-conceptual weeks and cell type abundance (estimated as proportions) for the 8 broad cell type groups.*

To determine the role of age (PCW) in expression analyses, we incorporated PCW as a covariate in the Liver *vs* Femur differential expression analyses. We found that this did not impact the results, as across all genes, the correlation between log-fold changes with and without age included in the models was r=0.999 in the disomic dataset and r=1 in the Ts21 dataset.

Overall, selected samples were all within a narrow developmental window, between 12-14 pcw, with little effect of PCW on statistical results, and sorting strategy was accounted for within the existing statistical analyses. Therefore, we don't think that sorting strategy or stage of development confounds the observed findings in the manuscript, and we discuss further aspects related to this in the Supplementary Results section titled “Evaluating the robustness of differential expression results to analysis decisions”.

5. It was not clear why the authors only focused on RBC GWAS for the enrichment of causal variants within HSCs and identification of causal variant-gene pairs.

We as well as others have reported an increase in the number of erythroid cells in the Ts21 foetal liver compared to disomic samples. We further confirmed erythroid bias of Ts21 HSCs in our single-cell in vitro differentiation assays. This raises an important question - can the increase in the number of erythroid cells be explained by a bias of HSCs to differentiate towards the erythroid lineage? To assess this, we used our multiome data.

Overall, there is significant heterogeneity within the HSC population. Specifically (using MIRA), we found evidence that a subpopulation of Ts21 HSCs belonging to branch 2 is "primed" to differentiate towards the erythroid lineage on both a chromatin accessibility and transcriptional level. In addition, Ts21 HSCs in branch 2 were significantly enriched for cells with increased accessibility at RBC GWAS SNPs (which implicate important regulatory elements) and not to WBC or LYMPH counts (Manuscript/Supplementary Figure 17). Therefore, we further focused on understanding the mechanism driving associations of non-coding RBC SNPs with erythroid bias in Ts21 by correlating enhancer accessibility with gene expression and thus identifying candidate target genes.

Furthermore, the SCAVENGE analyses revealed GWAS SNPs enrichments only in particular HSC branches: RBC counts in branch 2 and WBC counts in branch 1 (FDR < 0.1). As we observed the erythroid bias in Ts21 and branch 2 was composed of mostly Ts21 HSCs (while branch 1 was mixed Ts21/disomic), this further motivated a focus within downstream analyses.

We have improved our methods and results to be more clear of our SCAVENGE analysis and results in the manuscript.

Within the Methods section:

SCAVENGE TRS were contrasted between different lineage branches using a pseudobulk approach similar to the one we employed for differential expression analysis. On a per-sample-and-branch level, we pseudobulked the SCAVENGE scores by summing SCAVENGE scores across all cells of the same branch and of the same sample. We then applied three linear models to assess the impact of each of the three branches on SCAVENGE TRS. In each model, we accounted for the number of cells within the pseudobulk as a covariate. We determined significance by multiple test corrections at FDR < 0.1 across the 9 analyses (3 branches (1, 2, 3) and 3 traits (red blood cell, white blood cell, and lymphocyte counts)).

Within the Main text Results:

We next explored enrichment of blood cell traits in different HSC branches (identified by MIRA) from Ts21 and disomic foetuses. We found that HSCs from branch 1 were significantly enriched for cells with increased accessibility of GWAS SNPs relevant to WBC count, while HSCs from branch 2 were significantly enriched for cells with increased accessibility of RBC count GWAS SNPs (Figure 17b). Considering that the branch 2 was mostly composed of Ts21 HSCs, our analysis suggests that RBC GWAS-harbouring enhancers were more accessible in a subset of Ts21 HSCs.

As for the technical aspect of the analysis, I would like to confirm how the authors conducted the Fisher's exact test for the enrichment. How were the background regions or variants selected to be compared against SCENT peaks and RBC fine-mapped variants?

We considered a total of 37,999 peak-gene links in this analysis. These are peak-gene links tested in Ts21, which are:

- peak-gene combinations within 500 kb from each other
- >5% cells were accessible at the peak
- >5% cells expressed the gene

Using the association results from this analysis, we constructed a 2x2 contingency table based on two factors:

(1) Does the peak have a fine-mapped GWAS SNP for RBC?
(2) Is the peak associated with gene expression?

For criteria 1, we considered a GWAS SNPs with a posterior inclusion probability of 0.2. For criteria 2, we analyzed three different P-value thresholds (FDR < 0.1, FDR < 0.2, nominal P < 0.05). Therefore, we are effectively comparing whether peaks with a GWAS variant are more likely to be associated with gene expression compared to peaks without a GWAS variant, provided that the peaks are accessible (in >5% cells) and close to a gene (<500 kb) that is expressed (in >5% cells).

As an example of this contingency table at nominal P < 0.05:

| | No fine-mapped variants | fine-mapped variants |
|---|---|---|
| Nominal P > 0.05 | 30521 | 63 |
| Nominal P < 0.05 | 7389 | 26 |

Application of a two-sided Fisher’s exact test results in an OR = 1.7 [CI = (1.08, 2.74)], with a P-value of 0.031.

We note that another way of analyzing this data could be a binomial test. We treat the peaks without a GWAS variant as background, and test whether the peaks with a GWAS variant have a greater proportion of associations with expression. This results in 29.2% of GWAS-harboring peaks with an expression association, versus 19.2% in the background (peaks with no GWAS variants), for a 1.5-fold enrichment (P = 0.031).

We have revised the full methods section for this as such:

"We performed two enrichment tests with regard to our Ts21 peak-gene links and RBC GWAS. We defined SCENT peaks and SCENT genes according to three significance thresholds from the SCENT peak-gene analysis: FDR < 0.1, FDR < 0.2, and nominal P < 0.05, and identified GWAS-related peaks and genes using fine-mapped RBC GWAS SNPs at PIP > 0.2. In the first analysis, we assessed the enrichment of fine-mapped RBC GWAS SNPs (PIP > 0.2) within Ts21 SCENT peaks (also known as active cis-regulatory regions) using Fisher's exact test. Here, we are comparing whether peaks with a GWAS variant are more likely to be associated with gene expression compared to a background set of peaks without a GWAS variant, which include peaks accessible (in >5% cells) and close to a gene (< 500 kb) that is expressed (in >5% cells). In the second analysis, we assessed the enrichment of differential expression at target GWAS genes defined by Ts21 SCENT peak-gene links. We defined differential expression as FDR < 0.1 in the differential expression analysis of HSCs using the large scRNA-seq data. We subsetted to significant peak-gene links, and calculated a 2x2 contingency table reflecting (i) whether the peak contained a fine-mapped variant, and (ii) whether the target gene is differentially expressed. A Fisher's exact test assessed the hypothesis that important enhancers (containing fine-mapped GWAS SNPs) are having a role in differential expression within HSCs (affecting their target genes). "

Minor:
1. I noticed that the reference numbers might not be correctly assigned in the main text (at least for the last several papers), which made reading of the paper challenging.

We have now corrected this within the revised manuscript.

2. I am afraid there are figure mislables in the main text. Figure 4e appeared three times in different context.

We have corrected this in the revised manuscript.

3. The main text says "Indeed, cell cycle analysis confirmed that HSC/MPPs, both in Ts21 (G2/M+S=29.8%) and healthy liver (21.8%), were cycling more compared to their counterparts in BM (26.9% in Ts21 and 18.3% in healthy) (Figure 2d)." but Figure 2d shows the comparison between healthy vs DS.

The figure 2d shows the percentage of HSCs that are in G2/M+S in Ts21 vs disomic but also in liver (left plot) vs femur (right plot). To make this more clear, we have amended the sentence in the revised manuscript.

Resubmitted text:

“Indeed, cell cycle analysis confirmed that HSC/MPPs, both in Ts21 (G2/M+S=29.8%) and disomic liver (21.8%) (Figure 2d, left panel), were cycling more compared to their counterparts in BM (26.9% in Ts21 and 18.3% in disomic, Figure 2d, right panel).”

Reviewer #2

In this study, Marderstein et al. reported an in-depth characterisation of human prenatal hematopoiesis in both fetal liver and bone marrow using cutting-edge single cell approaches (scRNA-seq, scATAC-seq and Spatial transcriptomics), and extend our understanding of the molecular and cellular features perturbed in Down syndrome during fetal life. In a series of elegant bioinformatic analyses, the Authors showed that:
- Trisomy 21 significantly perturbed cellular composition of the hematopoietic system in the fetal liver and bone marrow,
- Perturbation of the hematopoietic system in Down syndrome is a result of both cell-intrinsic Ts21-induced and environmentally driven mechanisms,
- Spatial organisation and cell to cell interactions are perturbed in trisomic fetal livers,
- There is a higher concordance between chromatin accessibility and gene expression in trisomic hematopoietic stem cells,
- GWAS SNP analyses strongly suggest that trisomic HSC are `primed` toward the erythroid lineage in the fetal liver.
Altogether, this study describes the largest atlas of altered fetal hematopoiesis in Down syndrome realised at the single cell level to date, thus represents an outstanding resource of datasets for future work in the field of normal and pathologic hematopoietic development. This study also provides novel insights into the mechanisms of leukemia predisposition seen in this population.

However, several limitations remain and need to be addressed:

1- Some variability is seen in the cell composition in fetal livers and bone marrows between biological and technical replicates (supp Table). This is particularly visible for sample 8 in Figure 1c. How can this be explained?

Sample #8 indeed has a unique cell type composition that is dissimilar from the other Ts21 samples. We don’t have an explanation for this as there could be unknown technical and/or biological variability. However, we don't think that sorting strategy, stage of development, or sex confounds the observed findings in the manuscript. We evaluated this in the following analysis.

We explicitly accounted for different gating strategies in differential abundance and differential expression analyses. In the abundance analyses, we only analyzed samples that were sorted with the same procedure, which means that there should be no technical impact from different sorting strategies. For example, when calculating differences in blood cell abundance in Ts21 and disomic, only samples obtained with the same sorting strategy (CD45+) were used. In differential expression analyses, we accounted for sorting as a fixed effect within our statistical models. Therefore, gating strategy is already accounted for in the existing analyses.

In response to the comment about post conceptual weeks, we conducted an exploratory analysis of its associations in our data (Figure 1). We first estimated cell type composition for

the 8 major cell types detailed in Figure 1c across disomic and Ts21 samples (Figure 1). 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 the cell distribution across PCW (nominal P > 0.05 for all tests). We visualized 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, we did not consider PCW further in cell abundance results.

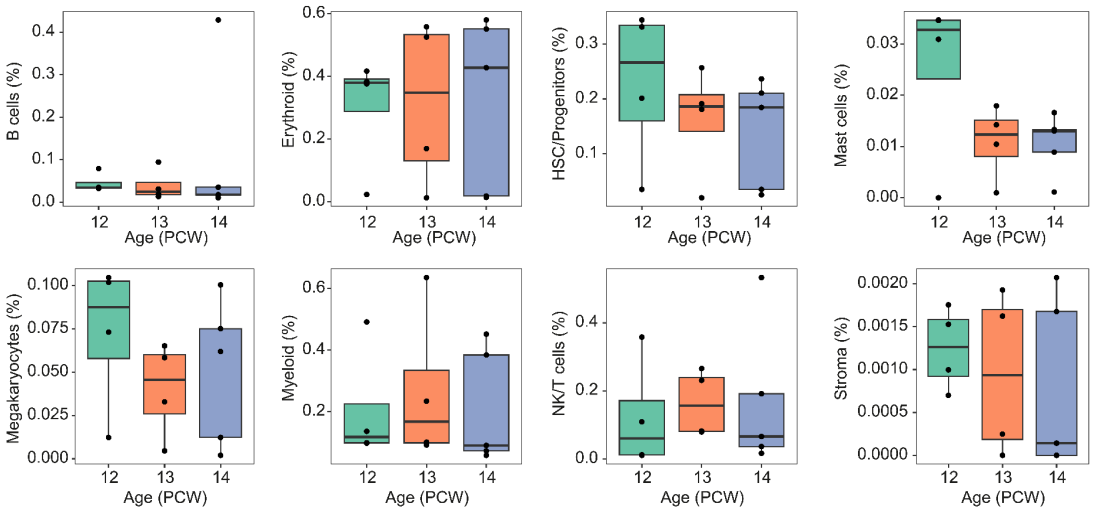



***Figure 1.*** *The relationship between post-conceptual weeks and cell type abundance (estimated as proportions) for the 8 broad cell type groups.*

To determine the role of age (PCW) in expression analyses, we incorporated PCW as a covariate in the Liver *vs* Femur differential expression analyses. We found that this did not impact the results, as across all genes, the correlation between log-fold changes with and without age included in the models was r=0.999 in the disomic dataset and r=1 in the Ts21 dataset.

Finally, we evaluated the role of sex as a covariate in differential expression analyses using the same framework as age above. 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). Therefore, sex did not have a large impact on differential expression results.

Overall, selected samples were all within a narrow developmental window, between 12-14 pcw, with little effect of PCW or sex on statistical results, and sorting strategy was accounted for within the existing statistical analyses. Therefore, we don't think that sorting strategy, sex or stage of development confounds the observed findings in the manuscript. This analysis is now part of the Supplementary Results section “Evaluating the robustness of differential expression results to analysis decisions”.

Could the cell composition in bone marrow be represented as in figure 1c (CD45+), and in Supp figure 5. Are the differences seen between samples in the liver conserved in the bone marrow? This would also facilitate the direct comparison between fetal liver and bone marrow from the same primary sample (15619, 15582, 15712…).

As the Reviewer correctly noted we have primary samples from liver and BM from the same foetuses. However, we utilised different sorting strategies in disomic and Ts21 bone marrow samples. The majority of cells in our Ts21 samples came from CD235- cells (erythroid depletion) and as such they are not suitable for assessing blood cell abundances (whereas disomic cells were sorted using both CD45+ and CD235- strategy). As a result, we were not able to quantify any potential differences in the frequency of cell types in Ts21 vs disomic bone marrow.

Cellular composition of Fetal bone marrow in both normal and Down syndrome context have been described previously (Jardine et al.), are the cellular difference seen in this new cohort concordant? Are there significant differences in the B cell composition between Ts21 and Healthy, and between FL and FBM? It is currently hard to see as they are in separated supp figures 3&4.

For the reasons outlined above, we were not able to quantify any potential differences in the frequency of cell types in Ts21 vs disomic bone marrow and compare with results presented in Jardine et al. and with our results from foetal liver.

However, we would like to point out some important differences that need to be taken into account when comparing the results between the two studies. Jardine et al. analysed a total of 16,743 foetal BM cells from Ts21 (n=4 foetuses). Of these, 12,914 were erythroid cells and the remaining **3,829** were all other cells that included 32 cell types (Supp.Table 10). As a result, the smallest cluster (CAR cells) in Jardine et al. has only 4 cells. The most likely reason for such skewing of cells toward erythroid lineage is related with the sorting strategy used by Jardine et al. (live cells i.e. without any enrichment or depletion, Supp. Table 1). This is very different from the strategy Jardine et al. used when sorting cells from disomic foetal BM that included CD45+, CD45-, and “TOTAL” (probably referring to no sorting or enrichment strategy) cells.

Our data set included 161,728 foetal BM cells from Ts21 (n=10 foetuses), which were sorted mostly using CD235- (erythroid cells depletion) strategy. As a result, only 6,768 cells were of the erythroid lineage and the remaining **154,960** were all other cells that included 36 cell types. This sorting strategy and the number of sorted cells allowed us to capture cells i) that are present at different frequencies in the foetal bone marrow and ii) to have high confidence about the cell type identity because of the sufficient number of cells in each cluster. However, this sorting strategy does not allow us to contrast any cellular composition differences in femur to those that we observed in the liver.

2- The exponential cross dependency is a very interesting observation. How can this be explained mechanistically? Can this be only due to environment as suggested? Could this also be driven by chromosome 21 genes that regulate chromatin accessibility or gene expression? Some Ts21 genes have been shown to be implicated in chromatin modification: HMGN1 and CHAF1B, along with many transcription factors ERG/ETS2/BACH1…. Can the Authors comment on this?

This is an interesting question. Our analysis suggests that when greater than 50% of genes on chromosome 21 are dysregulated in a cell type, there is dysregulation of more than 10% of genes on the other chromosomes within an exponential non-linear relationship. This trend was observed in both femur and liver.

First, since the cell types that show high dysregulation of chr21 genes are different in the two environments, 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).

In the main text, we made the following addition:

“We were unable to identify individual genes (such as particular chromatin modifiers or TFs) that drive dysregulation, suggesting the possible contribution of feedback loops or synergistic regulation.”

And added the analysis to the methods section, “**Analysis of Ts21 versus disomic differentially expressed genes”** and Supplementary section **“Evaluating cross-dependent dysregulation of chr21 genes versus other genes”.**

3- The Authors show that in Ts21 there is a greater concordance between chromatin accessibility and RNA expression, and speculate that this is due to signalling specific to Ts21? Are there arguments against a Ts21 driven clonal selection of the most adapted Ts21 cells, with a specific open chromatin and RNA expression, in this fetal liver environment? Something like a prenatal clonal hematopoiesis?

This is a very interesting point. Previous work by Hasaart KAL et al. (doi: 10.1038/s41598-020-69822-1) showed that HSPCs from Ts21 foetuses have higher somatic mutation load and a higher variance in somatic mutation load compared to disomic HSPCs, which was apparent before gastrulation. This increased somatic mutation rate in Ts21 foetal HSPCs might increase the chance to acquire an oncogenic driver mutation but was not sufficient to explain the rate of leukaemia development in DS children. They further speculate that cell–cell competition and selection as well as composition of the haematopoietic microenvironment are likely to also play a role in the development of leukaemia in children with DS. In line with this, they further showed that clones which had a highest mutational burden are not necessarily the cells which undergo clonal expansion and give rise to the DS-associated myeloid preleukemia. These Ts21 foetal HSPC with a significantly higher mutation load compared to other Ts21 foetal cells did show contribution of an additional signature SBS18 which has previously been associated with oxidative stress-induced mutagenesis.

In our study we observed a population of HSCs which was highly cycling and was primed to differentiate towards erythroid lineage. We would speculate that many of these cycling HSCs undergo differentiation and never engraft to bone marrow and therefore are lost from the HSC’s pool. However, some of these HSCs might acquire *GATA1* mutations which would lead to pre-leukaemia.

4- Stromal cell heterogeneity and cell to cell interaction within the bone marrow has been reported previously. Are these perturbed interactions also seen in the fetal liver. Can the Authors provide evidence of these perturbed interactions at the protein level using immunohistochemistry/immunofluorescence?

Experimental validation of the spatial transcriptomics results would require collecting additional material which we were not able to do. This analysis could be part of future investigations which would specifically focus on the question of detailed characterisation of niche and ligand-receptor interactions in Ts21 *vs* disomic samples.

5- One key observation is the HSC/MPP heterogeneity. The Authors describe stem cell pattern composed of 3 branches with branch 1 being the `healthy` one, and branch 2/3 only seen in Down syndrome fetal liver; branch 2 being the one `primed` toward the erythroid lineage and may contain the clone that predispose to myeloid leukemia. Branch 3 is less well defined. Several questions remain here: what are the of them? Can the key characteristics (RNA genes, genesets, regulatory element…) of the TOPICS that define these branches be reported in supp? Can cell to cell interactions be estimated for each of these branches? Do these branches exist in Fetal bone marrow? Can these different branches of HSC sorted to compare their progeny at the single cell level as in figure 4f-h?

HSC branches observed in the multiome data should be understood as different states that HSCs can occupy, of which some are present in both disomic and Ts21 HSCs and some are Ts21 specific. For each of these states we can identify key characteristics, like expressed genes and important TFs. In terms of the cell-cell interactions, the multiome data were performed on CD45+ cells and therefore we don't have information about niche cells and can't look at the interaction of these different HSC subpopulations with the niche cells. Also, we only performed multiome analysis on foetal liver so we can't conclusively say if the same level of heterogeneity exists within the HSC population in the bone marrow. Unfortunately, we were not able to identify specific cell surface markers that would allow us to confidently sort these different subpopulations of HSCs.

We have added TFs, peaks, genes for each of the ATAC and RNA topics in Supplementary Table 11.

6- Branch 3 seems to be a more `by default` branch in Ts21 fetal liver with no clear association with transcription factors or strong gene set enrichment, apart from a transcriptional signature of HDAC inhibition… What does this means? Is this branch less prone to transformation in fetal liver? Where is this branch coming from? Can Velocity be used to estimate its origin? If it exists in the bone marrow, could this branch become in this different environment?

These are all very interesting questions. All HSCs in the multiome data have a common "origin" in a sense that both branch 2 and 3 diverge from branch 1, which is the "earliest" HSC population based on the Velocity analysis in Figure 3d. During development, HSCs need to migrate from foetal liver to bone marrow; and thus the more successful particular HSC clones are in engraftment, the more likely we are able to detect them in the bone marrow. Considering that HSCs in branch 2 are highly cycling and that they show priming towards erythroid lineage, we can speculate that these HSCs would be less likely to engraft in the bone marrow as they would differentiate, unless they acquire mutations. However, since we don't have multiome

data for the foetal BM, we can't test our hypothesis.

7- Since Down syndrome children are also predisposed to lymphoid leukemia (almost exclusively of B phenotype), one may wonder if some GWAS SNP B-cell specific are also over-represented. Has this been assessed? In fetal liver or fetal bone marrow? or even in HSC/CLP? Can Scavenge be used to determine B cell v. T cell traits?

To address this question we used the largest published childhood acute lymphoblastic leukemia (ALL) GWAS study (Jeon et al. 2022 *Leukemia)* 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 (Figure 2), 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).

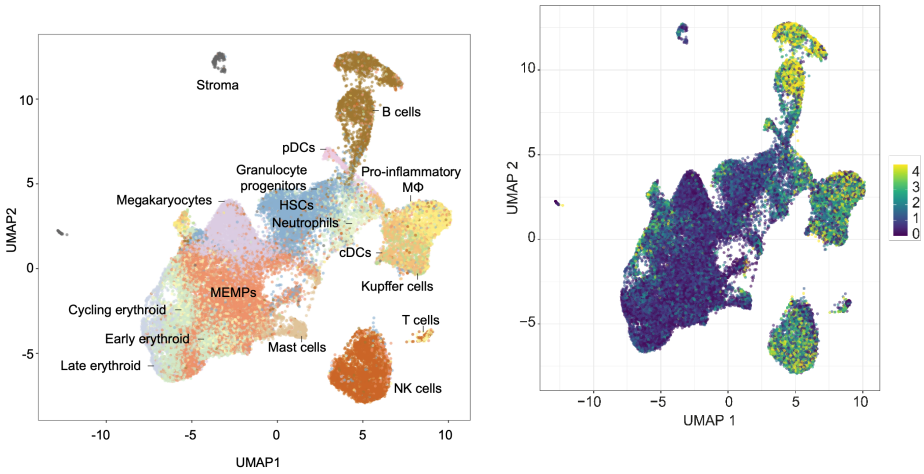



***Figure 2.*** *Left is Figure 3a from the main text, which shows the cell type labels positioning in the UMAP. Right describes the SCAVENGE scores for childhood ALL, with the strongest enrichments seen in B cell populations (top right).*

We believe that any potential differences that might exist are currently undetectable due to the limited number of identified GWAS signals associated with childhood ALL risk (Jeon et al, 2022). This significantly reduces the statistical power to distinguish between Ts21 and disomic individuals. In comparison, GWAS for hematological traits (such as red blood cell or lymphocyte counts) have much greater statistical power. Over 3,000 loci have been associated with blood cell traits, with a range of 157-692 per individual trait, and heritable variation is linked to childhood ALL (Kachuri et al. 2021 AJHG).

While of significant interest, we are unable to directly evaluate hypotheses related to B cell vs T cell predisposition and fetal liver vs bone marrow with the existing data. First, multiome data were only processed from foetal liver, and not from foetal BM. Second, we face constraints

regarding the statistical power and data availability. Indeed, the childhood B-cell ALL GWAS is included in the Jeon et al. ALL GWAS, with even lower power. Furthermore, the childhood B-cell ALL GWAS reports the top SNPs at each locus, but the full summary statistics are not publicly available (or at least, we could not find them) which prevents repeating statistical fine-mapping necessary for a SCAVENGE analysis from that study.

To our knowledge, there are no larger GWAS datasets for B cell lymphocyte or T cell lymphocyte counts. For example, UK Biobank only measured an overall lymphocyte count measure with no further splitting into B/T cell compartments. This is partly a key reason why our understanding of blood cell counts and cancer risk is limited to broad cell type classes instead of subpopulations, despite likely distinct etiological mechanisms (as further discussed in Kachuri et al 2021 AJHG or Williams et al 2019 Cancer Epidemiol Biomarkers Prev).

8- This study is mostly focused on fetal liver hematopoiesis in Down syndrome, and strongly suggest that the most probable origin of TAM/ML-DS is within the branch 2 of Ts21 HSC/MPP. However, nothing is mentioned about GATA1. Have these Down syndrome samples (whole fetal livers or HSC/MPP) been tested for GATA1 mutations?

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[11]. In line with this, the larger screening studies in newborns have shown that the incidence of GATA1 mutations is around 4%[12–14] . Furthermore, Haasart et al. did not observe a mutation in *GATA1* in any of the foetal stem and progenitor cells following WGS of HSPCs from 9 independent human fetuses gestational age week 12–17.

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.

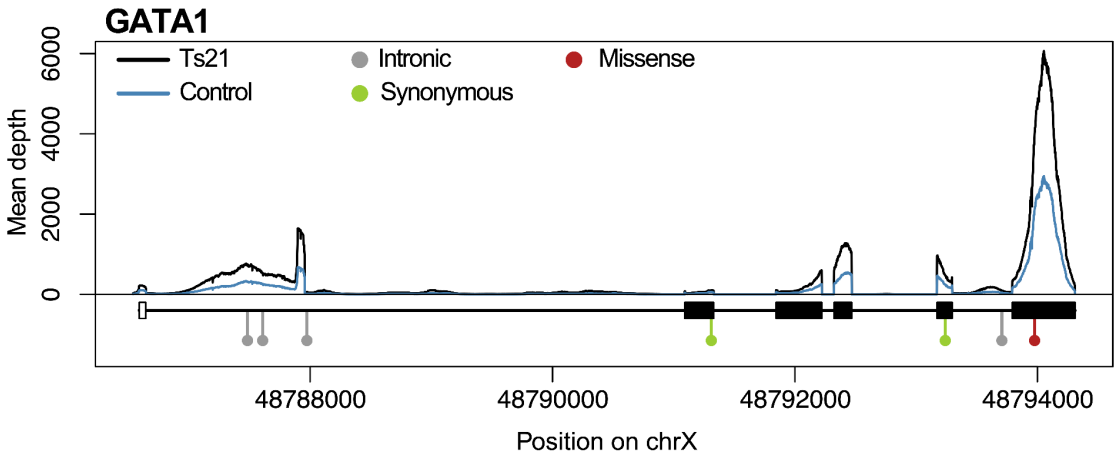



***Figure 3.*** *Mean number of reads for each site spanning the GATA1 locus, black line for Ts21 samples and blue for control samples. Boxes indicate exons of GATA1 (black =coding, white=non-coding). Detected mutations are annotated by their positions, as well as their predicted effect (grey=intronic, green=synonymous, red=missense).*

These limitations notwithstanding, we used the variant calling framework based on deepSNV/Shearwater[15,16] 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 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.

We have included this analysis in the manuscript within the Supp. Results section titled “Detecting GATA1 mutations in foetal samples using RNA-seq”.

Is the GATA1 locus particularly accessible?

Yes, our analysis shows that GATA1 gene body and promoter accessibility is higher in Ts21 HSCs *versus* disomic HSCs (19.3% versus 6.7% of cells).

We have added this to the manuscript.

9- Finally, most of the results are descriptive, apart from the functional validation of the `primed` erythroid phenotype shown in Figure 4f-h, which correlates with earlier study from Jardine et al. There is a lot of speculative arguments about the potential impact of TFR2, TSPAN32 but not C11orf21, ERK1/2 pathway activation, oxidative phosphorylation, mitochondrial dysfunction, in fetal Ts21 HSC but no functional validation. This limits the impact of the manuscript as it has not been tested in this specific context (fetal liver HSC/hematopoiesis). Can some of these be tested functionally to strengthen the conclusions?

Due to the limited availability of human foetal tissue, we focused on performing experiments that would significantly strengthen some of the important conclusions of the manuscript. We did the following experiments:

1) Performed dose response experiment to determine optimal Verapamil concentration for its use in point 2) and 3).

2) Measured mitochondrial mass in Ts21 (n=3) and disomic (n=3) foetal haematopietic cells.

3) Measured oxidative stress level in Ts21 (n=3) and disomic (n=3) foetal liver haematopietic cells.

4) Compared the abundance of immunophenotypic HSCs, progenitor cells and mature cells in Ts21 (n=3) and disomic (n=3) using flow cytometry.

To experimentally validate that DS foetal haematopoietic cells exhibit higher oxidative stress and changes in mitochondrial mass compared to disomic cells (Figure 4), we have now measured mtROS levels, using MitoSOX probe, and mitochondrial mass using the

MitoTracker green FM dye. MitoSOX is a fluorogenic dye specifically targeting mitochondria in live cells. Oxidation of the MitoSOX reagent by mitochondrial superoxide produces bright green fluorescence. MitoTracker probes are cell-permeant mitochondria dyes that following incubation, passively diffuse across the plasma membrane and accumulate in the mitochondria of live cells. Compared to other dyes of the same family, MitoTracker green FM accumulates in the mitochondria regardless of mitochondrial membrane potential and is not sensitive to oxidative stress, making it a making it a valuable tool to measure mitochondrial mass[5,6].

It has been reported that HSCs and progenitor cells (murine bone marrow and human cord blood) but not lineage positive cells can efflux the dye through xenobiotic efflux pumps (*Bcrp1* and *Mdr1a/b)* and therefore the level of fluorescence does not faithfully represent the level of mitochondrial mass. To circumvent the bias caused by preferential dye efflux from immature hematopoietic cells, we treated the cells with Verapamil to block the xenobiotic efflux pumps, (DOI: 10.1016/j.stem.2020.09.018).

To validate the feasibility of the assays, we first stained disomic foetal haematopoietic cells with MitoTracker and assessed the mitochondrial mass of progenitor cells by flow cytometry. Within the foetal liver Lin- CD34+ CD45dim population, our data demonstrated the existence of two sub-populations stained differently for MitoTracker, named MTG_hi and MTG_lo. The MTG_lo population was enriched especially in immature CD38- cells (Figure 5, panel A). This is in line with the expected preferential dye efflux of more immature hematopoietic cells (de Almeida et al., Cell Stem Cell, 2017). Following treatment with Verapamil, which blocks efflux pumps, we observed an increase in MitoTracker dye retention, confirming the effectiveness of the assay in human foetal haematopoietic cells (Figure 5, panel B). Regarding mtROS levels, within the foetal liver Lin- C34+ CD45dim population, we also found a subpopulation that exhibited higher staining intensity for MitoSOX green (MSOX_hi) and a population that displayed lower staining intensity (MSOX_lo). Similarly to what we described for MitoTracker, the MSOX_lo population was enriched in CD38- cells (Figure 5, panel C), and treatment with 50 µM Verapamil increased the retention of MitoSOX dye in both CD38+ and CD38- populations (Figure 5, panel D).

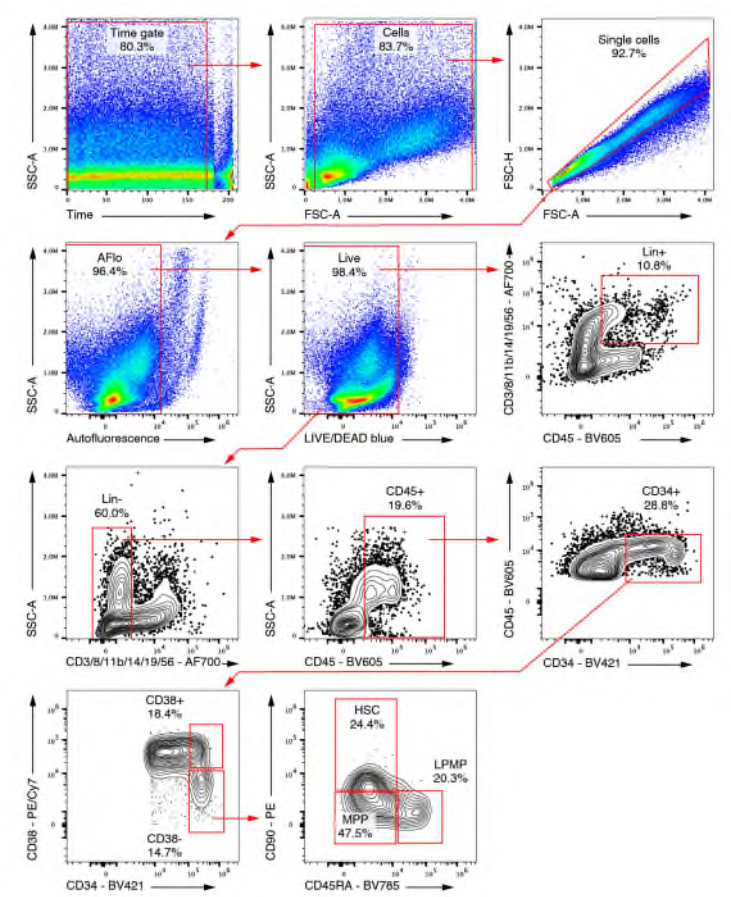



***Figure 4*** *Representative gating strategy used for the mitochondrial probe experiments. Both MitoSOX and MitoTracker stainings shared the same antibody panel.*

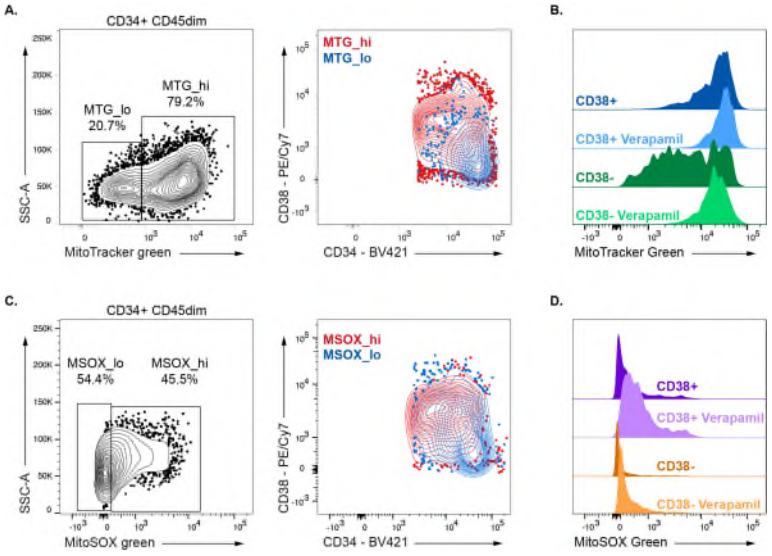



***Figure 5.*** *A. Left: Representative gating strategy used to identify MTG_hi and MTG_lo populations. Right: MTG_hi and MTG_lo cells were plotted against CD34 and CD38, showing that MTG_lo cells are enriched in the CD34+ CD38- fraction. B. Density plot of the MitoTracker Green FM fluorescence in CD38+ 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+. C. 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. D. 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+.*

Next, we compared the mitochondrial mass (using MitoTracker) and mtROS levels (using MitoSOX) in 3 Ts21 and 3 disomic foetal liver samples stained in the presence of Verapamil (Figure 6 and 7). We observed that Ts21 CD34+CD38+ cells had higher mitochondrial mass and mtROS levels compared to disomic progenitors, whereas CD34+CD38- and HSCs were comparable between the two conditions. Previous studies have shown that disomic HSCs exhibit very low electron transport chain activity and minimal mROS production as they mostly rely on glycolysis. In contrast, committed progenitors have much higher oxygen consumption, mitochondrial ATP production, and maximal respiratory capacity. Taken together, our results support the hypothesis that as the HSCs and immature CD34+CD38- cells start to differentiate to committed CD34+CD38+ progenitors, they ramp-up oxidative phosphorylation. This leads to higher mtROS production in Ts21 cells compared to discomic cells, possibly due to their higher mitochondrial mass and dysfunction in the electron transport chain activity.

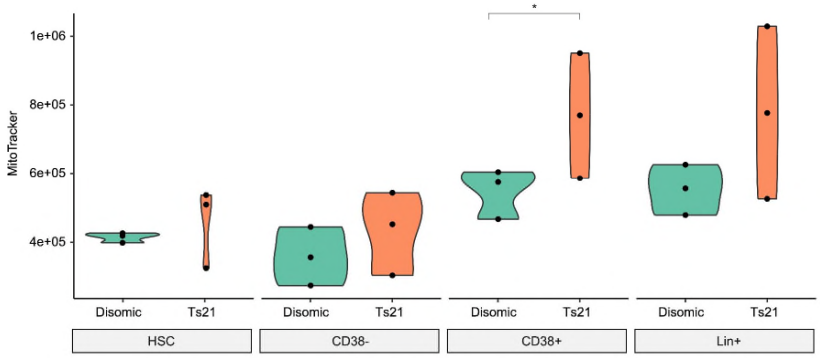



***Figure 6.*** *MitoTracker staining in disomic and Ts21 haematopoietic populations. HSC-haematopoietic stem cells (Lin-/CD34+/Cd38-/CD90+/CD45RA-); CD38- (Lin-/CD45+/CD34+/Cd38- population); CD38+ (Lin-/CD45+/CD34+/CD38+ population); Lin+ (*CD3/8/11b/56/14/19, *lineage positive population).*

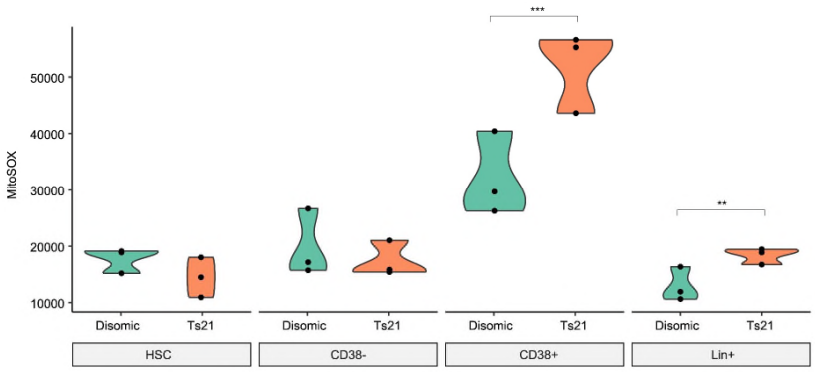


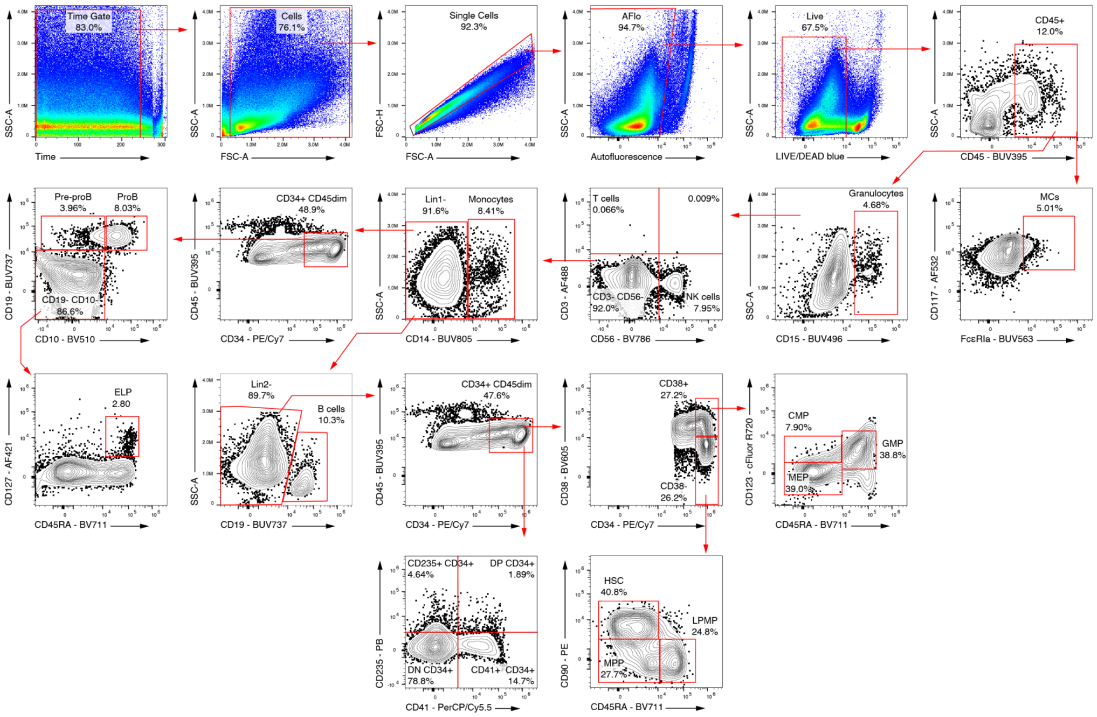

***Figure 7.*** *MitoSOX staining in disomic and Ts21 haematopoietic populations. HSC- haematopoietic stem cells (Lin-/CD34+/Cd38-/CD90+/CD45RA-); CD38- (Lin-/CD45+/CD34+/Cd38- population); CD38+ (Lin-/CD45+/CD34+/CD38+ population); Lin+ (*CD3/8/11b/56/14/19, *lineage positive population). ; P<0.05(*); P<0.01(**); P<0.001 (***).*

Experimental details are included in the Methods section of the revised manuscript.

To further confirm that the Ts21 causes wide perturbation in foetal haematopoiesis we compared Ts21 with the disomic immunophenotypic stem and progenitor compartment using flow cytometry (Figure 8). In line with previous reports we observed an increase in HSCs (CD34+CD38-CD90+CD45RA-) and megakaryocyte-erythroid progenitor cells (MEPs, CD45+ Lin-(CD15/3/56/14/19) CD34+ CD38+ CD123- CD45RA-) but a decrease in multipotent progenitors (MPPs, CD45+ Lin-(CD15/3/56/14/19) CD34+ CD38- CD90- CD45RA-) and granulocyte-monocyte progenitors (GMPs, CD45+ Lin-(CD15/3/56/14/19) CD34+ CD38+ CD123dim/+ CD45RA+) in Ts21 compared to disomic cells.

Within the CD34+ compartment we observed an increase in erythroid CD235+ and megakaryocytic CD41+ biased cells. No change was observed in the frequency of common-myeloid progenitors (CMPs, CD45+ Lin-(CD15/3/56/14/19) CD34+ CD38+ CD123dim/+ CD45RA-).

In the B cell compartment we observed a decrease in early lymphoid progenitors (ELPs, CD45+ CD15- CD3- CD56- CD14- CD19- CD10- CD127+ CD45RA+), Pre-pro-B cells (CD45+ CD15- CD3- CD56- CD14- CD19+ CD10-), Pro-B cells (CD45+ CD15- CD3- CD56- CD14- CD19+ CD10+), and B cells (CD45+ CD15- CD3- CD56- CD14- CD19+), confirming a defect in B cell development. Furthermore, we saw a decrease in the number of granulocytes (CD45+ CD15+) and an increase in the number of NK cells (CD45+ CD15- CD3- CD56+) in Ts21 compared to discomic samples and no difference in Myelo-Monocytes (CD45+ CD15- CD3- CD56- CD14+) and T cells (CD45+ CD15- CD3+ CD56-).


Time Gate
83.0%
Cells
76.1%
Single Cells
92.3%
AFlo
94.7%
Live
67.5%
CD45+
12.0%
CD45 - BUV395
Pre-proB
3.96%
ProB
8.03%
CD19- CD10-
86.6%
CD10 - BV510
CD34+ CD45dim
48.9%
CD34 - PE/Cy7
Lin1-
91.6%
Monocytes
8.41%
CD14 - BUV805
T cells
0.066%
0.009%
CD3- CD56-
92.0%
NK cells
7.95%
CD56 - BV786
Granulocytes
4.68%
CD15 - BUV496
MCs
5.01%
FcεRIa - BUV563
ELP
2.80
CD45RA - BV711
Lin2-
89.7%
B cells
10.3%
CD19 - BUV737
CD34+ CD45dim
47.6%
CD38+
27.2%
CD38-
26.2%
CMP
7.90%
GMP
38.8%
MEP
39.0%
CD123 - cFluor R720
CD235+ CD34+
4.64%
DP CD34+
1.89%
DN CD34+
78.8%
CD41+ CD34+
14.7%
CD41 - PerCP/Cy5.5
HSC
40.8%
LPMP
24.8%
MPP
27.7%
CD90 - PE

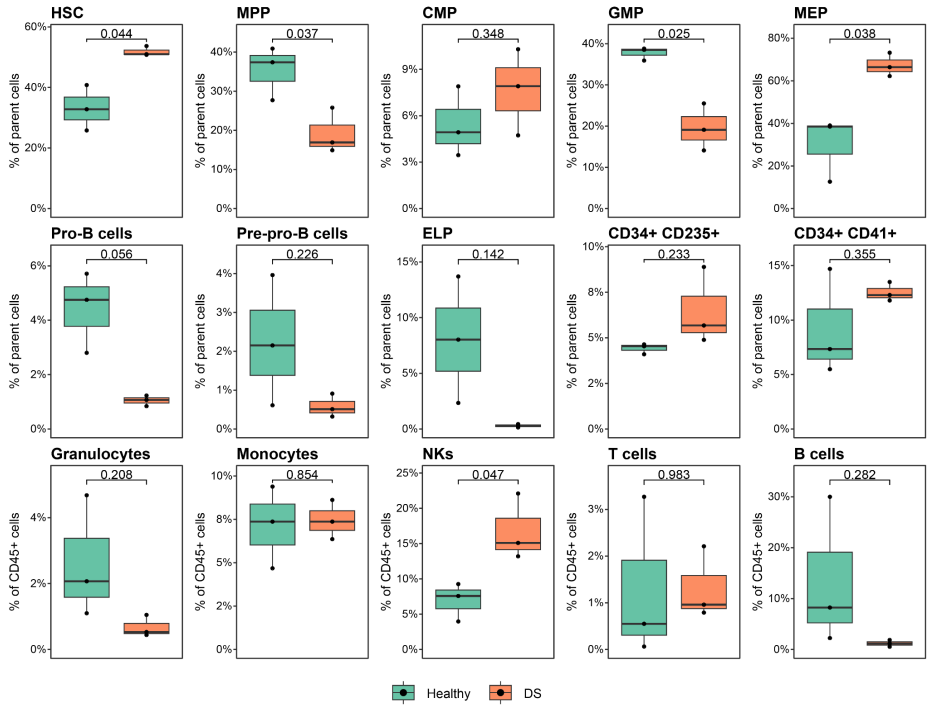

HSC
0.044
MPP
0.037
CMP
0.348
GMP
0.025
MEP
0.038
Pro-B cells
0.056
Pre-pro-B cells
0.226
ELP
0.142
CD34+ CD235+
0.233
CD34+ CD41+
0.355
Granulocytes
0.208
Monocytes
0.854
NKs
0.047
T cells
0.983
B cells
0.282
% of parent cells
% of CD45+ cells
Healthy
DS

***Figure 8.*** *Top - 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. Bottom - Frequency of progenitor and stem cell subsets in T21 and disomic foetal liver. The values on the boxplot indicate the results of a Welch-corrected t-test comparing the percentage of cells in disomic and T21 foetal livers.*

Minor comments:

1- Technical: population collected for sequencing were sorted and/or depleted, how does this affect the sequencing? And readout (cell composition)? Has this been tested? Has the purity of sorted cells been confirmed?

We explicitly accounted for different gating strategies in the differential abundance and differential expression analyses. In the abundance analyses, we only analyzed samples that were sorted with the same procedure, which means that there should be no technical impact from different sorting strategies. For example, when calculating differences in blood cell abundance in Ts21 and disomic, only samples obtained with the same sorting strategy (CD45+) were used. In differential expression analyses, we accounted for sorting as a fixed effect within our statistical models. Therefore, gating strategy is already accounted for in the existing analyses.

In response to the comment about post conceptual weeks, we conducted an exploratory analysis of its associations in our data (Figure 9). We first estimated cell type composition for the 8 major cell types detailed in Figure 1c across disomic and Ts21 samples (Figure 9). 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 the cell distribution across PCW (nominal $P > 0.05$ for all tests). We visualized 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, we did not consider PCW further in cell abundance results.

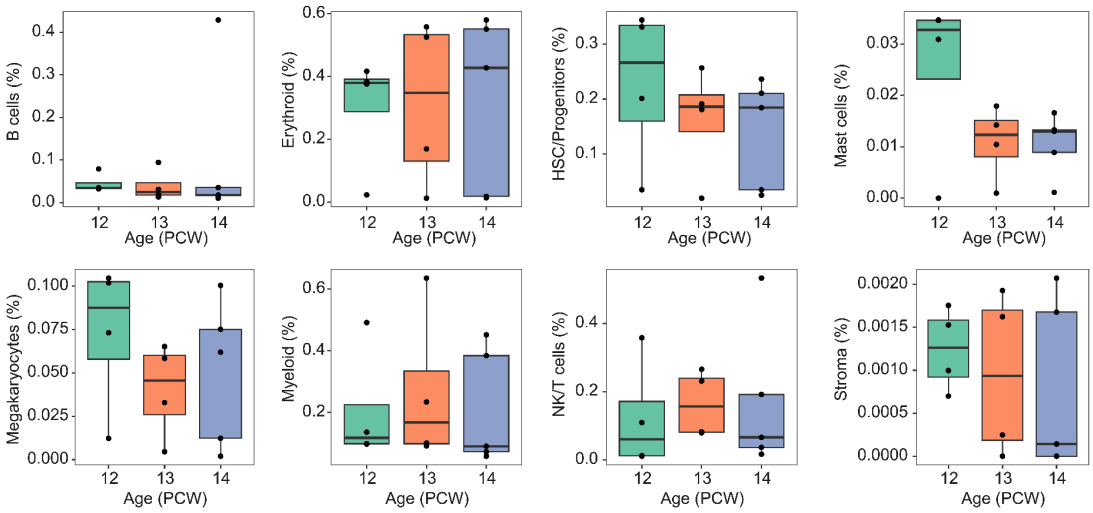



***Figure 9.*** *The relationship between post-conceptual weeks and cell type abundance (estimated as proportions) for the 8 broad cell type groups.*

To determine the role of age (PCW) in expression analyses, we incorporated PCW as a covariate in the Liver *versus* Femur differential expression analyses. We found that this did not impact the results, as across all genes, the correlation between log-fold changes with and without age included in the models was r=0.999 in the disomic dataset and r=1 in the Ts21 dataset.

Overall, selected samples were all within a narrow developmental window, between 12-14 pcw, with little effect of PCW on statistical results, and sorting strategy was accounted for within the existing statistical analyses. Therefore, we don't think that sorting strategy or stage of development confounds the observed findings in the manuscript, and we discuss this in the Supplementary Results section titled “Evaluating the robustness of differential expression results to analysis decisions”.

2- Genetic: are all Ts21 used in this study have a complete trisomy 21? No mosaic?

We do not think that the samples analysed in this study were mosaics for trisomy of chr21. The karyotype for each sample used in this study was determined by Quantitative fluorescent PCR (QF-PCR). The analysis was performed by the tissue bank from which the samples were obtained. QF-PCR has been used by many laboratories for prenatal diagnosis of the most common aneuploidies. QF-PCR was performed using chromosome specific microsatellite markers. It is a rapid test that can detect most abnormalities diagnosed by conventional karyotyping. QFPCR showed a normal result with an apparently normal diploid complement for chromosomes 13, 15, 16, 18, 21, 22 and the sex chromosomes in our disomic samples

and trisomy for chr 21 in DS samples. The karyotype assays used by the tissue bank are able to pick up mosaicisms. The tissue bank generally sends another organ sample for karyotyping for confirmation and it tests whether it was tissue specific. If the samples turned out to be mosaic, the tissue bank would have informed us.

3- Age differences between healthy and Ts21 samples: can this explain some of the difference seen in cell composition? In fetal liver and in bone marrow?

Overall, we did not see age contributing significantly to cell composition within our data.

For detailed analysis please see our response to comment 1 (technical).

4- Please show where is located the SNPs in Figure 4d-e?

We have included this information in Figure 4f (updated numbering) in the revised manuscript.

5- Integration of healthy and Ts21 datasets has been performed in figure 3a, why not integrating the scRNA-seq data also?

We have completed integrated analysis in the revised manuscript. First, we integrated the full dataset using Harmony (Figure 10) 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 neighbor 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 (Figure 11), validating accurate annotation of cell types.

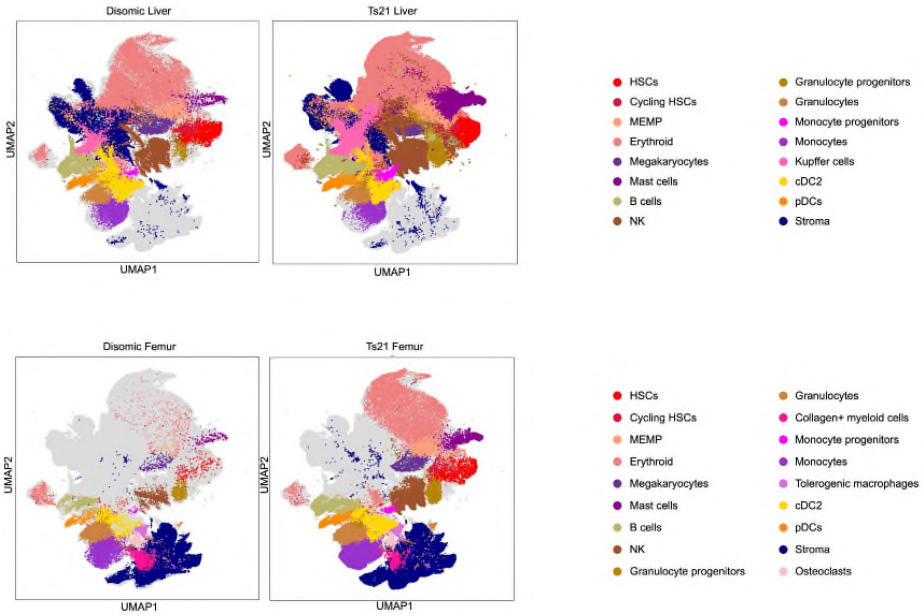



***Figure 10.*** *3D Uniform manifold approximation and projection (UMAP) visualisation of cells from Ts21 liver (n=780,299, k=15), Ts21 femur (n=162,775, k=12), disomic liver (n=110,671, k=3) and disomic femur (n=53,807, k=3) following integration of datasets using Harmony. Cells are coloured by broad cell type categories. n refers to the total number of cells, k indicates the number of biologically independent samples (foetuses).*

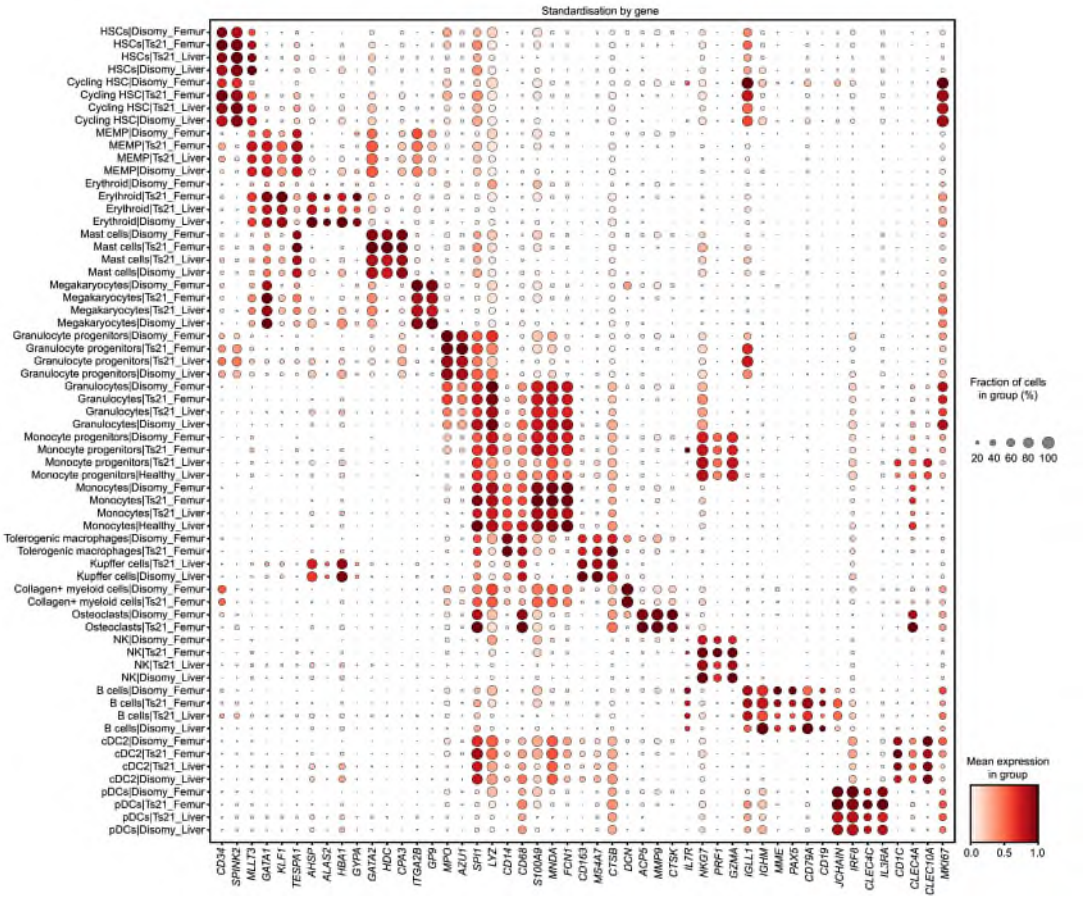
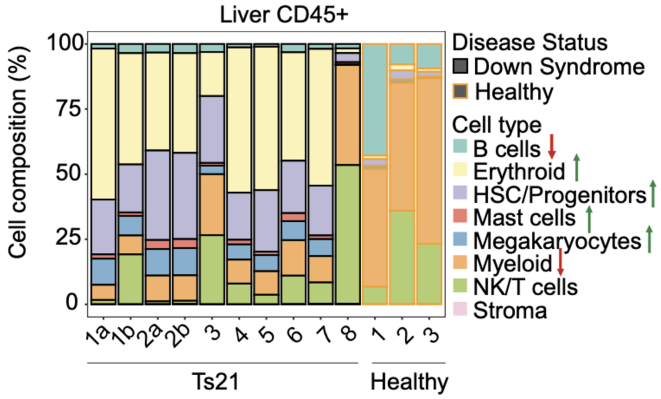



***Figure 11.*** *Gene expression dot plot of marker genes from Supp. Figure 3 and 4 across cell types identified from Harmony-based dataset integration.*

Next, we compared the Harmony integration-identified cell type abundance analysis to our separate cell type abundance analysis shown in Main Text Figure 1c.




***Main Text Figure 1c (copied from original submission).*** *Stacked bar plot of relative abundances of different broad cell types in stage-matched Ts21 (black outline) and disomic foetuses (orange outline). Arrows at the side of the labels indicate a significant increase (↑) or decrease (↓) in the number of cells in Ts21 liver, compared to disomic liver. Difference in cell*

*proportion was tested by Mann-Whitney U test and significance was determined by FDR < 0.1.*

The results were remarkably consistent between the two approaches (Figure 12). The only difference being that, while the trend remained consistent, mast cells were not anymore significantly differentially abundant (FDR = 0.103).

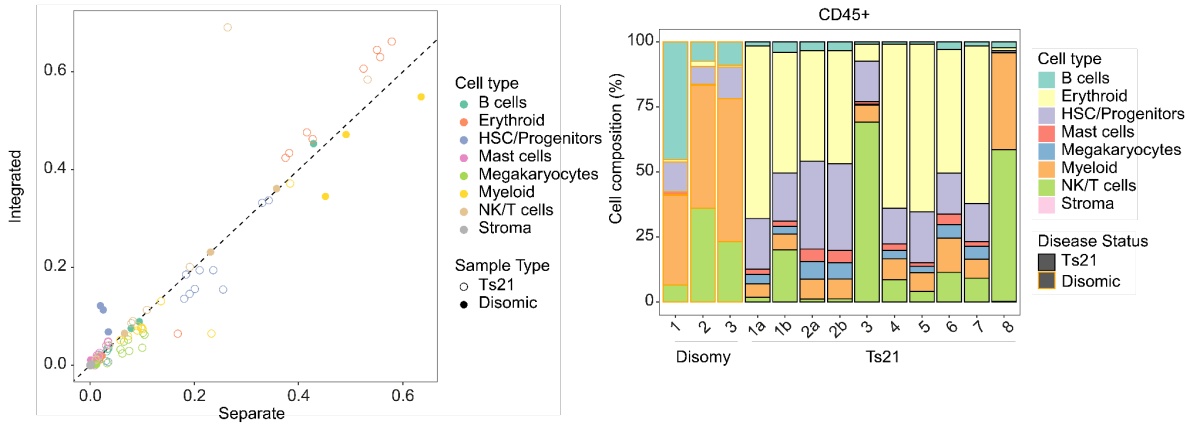



***Figure 12**. Using the Harmony-integrated cell type labels in stage-matched Ts21 and disomic foetuses, we compared cell type proportions between separate and integrated (left), and (right) made a stacked bar plot of relative abundances of different broad cell types similar to Main Text Figure 1c. In (left), each point represents the estimated abundance in a different sample. Panel (right) can be compared to Main Text Figure 1c, which has been copied above.*

Given 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) (Luecken et al. 2021 *Nat Methods*), we repeated the integration and clustering using scVI (Figure 13 and 14). 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 neighbor graph, computing Leiden clusters, and finally annotating clusters.

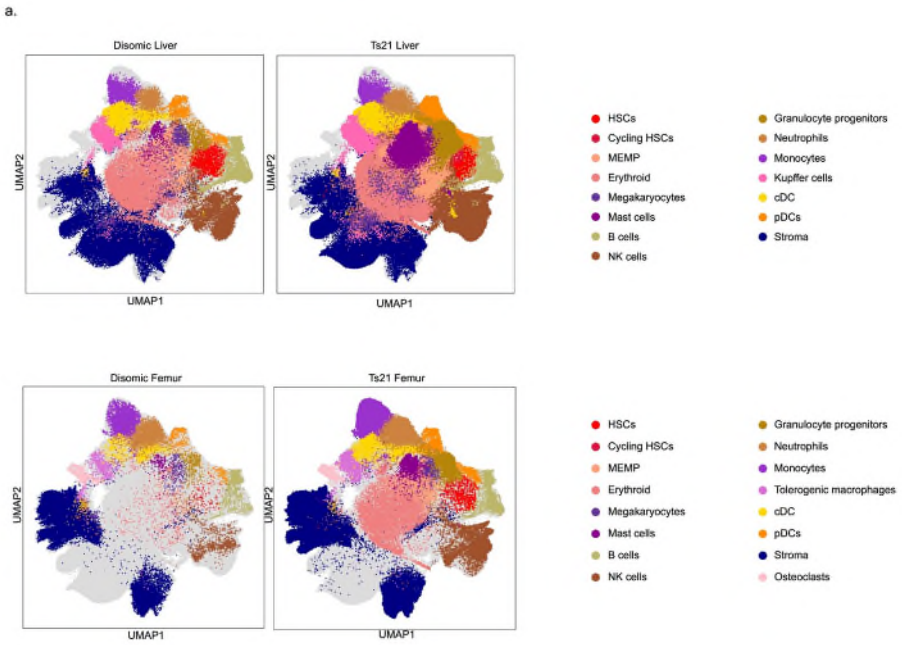



***Figure 13.*** *3D Uniform manifold approximation and projection (UMAP) visualisation of cells from Ts21 liver (n=780,299, k=15), Ts21 femur (n=162,775, k=12), disomic liver (n=110,671, k=3) and disomic femur (n=53,807, k=3) following integration of datasets using scVI tools. Cells are coloured by broad cell type categories. n refers to the total number of cells, k indicates the number of biologically independent samples (foetuses).*

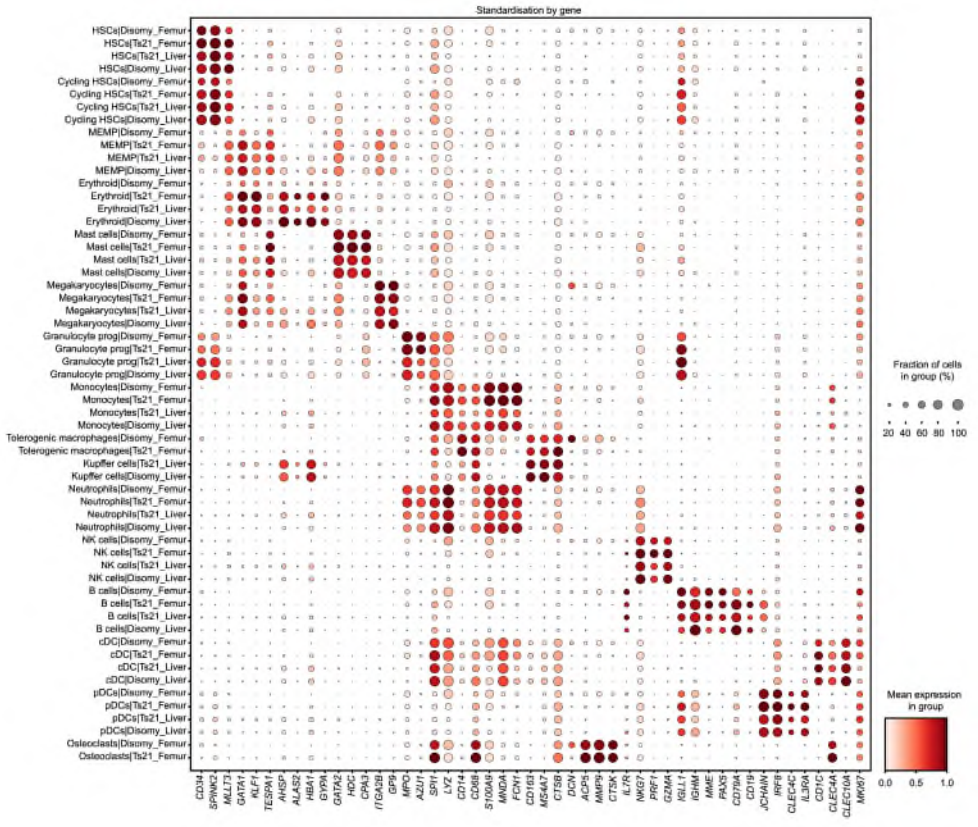



***Figure 14.*** *Dot plot showing expression of marker genes from Supp. Figure 3 and 4 across cell types identified following dataset integration using scVI tools.*

The scVI-integrated proportions across samples correlated closely with separate dataset proportions ($r$ = 0.957) (Figure 15 left), 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 (Figure 15 right).

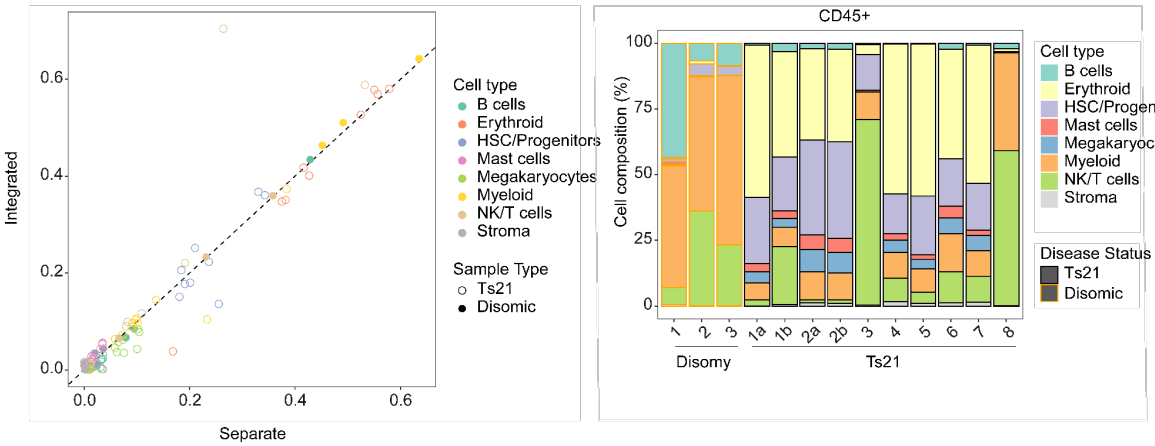


***Figure 15***. *Using the scVI-integrated cell type labels in stage-matched Ts21 and disomic foetuses, we compared cell type proportions between separate and integrated (left), and (right) made a stacked bar plot of relative abundances of different broad cell types similar to Main Text Figure 1c. In (a), each point represents the estimated abundance in a different sample.*

Overall, our results demonstrate that the cell-type annotations are consistent across all annotation strategies and the differential abundance results are robust. We include these results within the Supplementary Results section of the manuscript, under “Evaluating cell-type abundances following data-set integration and label transfer”, and mention in the Main Text:

“We confirmed the robustness of our cell-type annotations and differences in frequency of cell types between Ts21 and disomic liver samples after integration using either Harmony or scVI, as well as reference-query mapping from a published disomic foetal liver atlas and our Ts21 liver cells to disomic liver cells (Supplementary Results; Supp. Figure 7-9).”

6- How representative are spatial transcriptomic results shown in supp figure 6b/c? does supp figure 6c/d/e this include all samples?

The spatial transcriptomic results presented in Figure 1 and Supp. Figure 10 (updated numbering) 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, and therefore they are a good summary of the findings we obtained in this analysis.

7- Number of replicates (and cells) used for bioinformatic analyses should be added to the figure legends (Figure 2-4, and supp).

We have added this information to the figure legends for Figures 2-4.

8- Has the multi-omic analyses has been performed in bone marrow as suggested by the title of the manuscript.

We have not performed multiome analysis on the samples from the bone marrow due the low number of cells we could retrieve. We have now changed the title of the study to: “Single-cell multi-omics map of human foetal blood in Down's Syndrome”

Reviewer #3 (Remarks to the Author):

In this paper the authors analyze single cell data from the liver and bone marrow of Down Syndrome and “healthy” fetuses. This is a bioinformatic intensive paper with all conclusions drawn from the analysis of gene expression and ATAC datasets.

A strength of this paper is the large number (>1 million) of cells analyzed that were obtained by sorting liver and marrow from a reasonable cohort of Down Syndrome fetuses. Importantly, both haematopoietic and stromal/microenvironmental cells were analyzed.

A significant finding (Figure 1c) is the marked differences in the proportions of CD45+ HSPCs, erythroid, and megakaryocyte cells vs. myeloid cells and B-cells in the fetal liver of Down Syndrome vs. ’healthy” fetuses, respectively. Spatial analysis reveals differences in the cellular architecture of the liver, though it was difficult to understand what is being shown in the spatial analysis presented in Fig. 1e. Is the increase in hepatocyte frequency possibly due to a difference in hematopoietic elements in the spatial analysis or was the density of hepatocytes higher? If so, was there a significant difference in the overall size of the liver in Down Syndrome fetuses vs. ‘healthy’ fetuses?

Visium is a spot-based method, where theoretically on average 1-10 cells occupy each spot. We used cell2location to assess the number of hepatocytes in each Visium spot. Next, we summed-up the total number of hepatocytes across all spots to determine the overall count in each tissue section. Finally, this count was normalized by the total number of cells present in the same section to calculate the “hepatocyte frequency”. Due to the normalization step by the total number of cells, the size of the liver doesn’t influence the calculation of cell-type frequency. Therefore, higher frequency implies higher density of hepatocytes. Figure 1d summarizes the cell type frequency for each section from each fetus (represented as dots in the boxes).

Having said this, we did not measure the size of the liver in Ts21 *versus* disomic samples. However, previous reports suggest that the liver of newborns with DS is significantly larger compared to disomic livers. So, we can not exclude the possibility of size differences in addition to the compositional differences that we observed.

Trajectory analysis of stromal and osteolineage cells showed marked differences between Ts21 and ‘healthy’ marrow but it is unclear what the different trajectories actually indicate in Fig. 1f/g.

In Figure 1f we showed PAGA graphs overlaid on the diffusion maps, computed from the 11 (Ts21) or 13 (disomic) cell types representing the BM niche population. This graph shows two key points:
1) It visually demonstrates which cell types are connected, thereby showing their involvement/position in the differentiation trajectory

2) Osteoblasts appear to be connected to endothelial cells, through “transitioning endothelial cells”. This provides evidence that endothelial cells can go through endothelial-to-osteoblast transition through transcriptional changes in disomic fetuses. While we also detected these “transitioning endothelial cells” in Ts21, they appeared to be connected to multiple cell types of the mesenchymal lineage. This can suggest that these transitioning endothelial cells in Ts21 have difficulty executing proper transcriptional programs.
The trajectory in Figure 1g displays the directionality of the differentiation process from CAR cells to osteoprogenitors and osteoblasts in disomic and Ts21 samples. This trajectory is then used to calculate genes that are dynamically expressed along the osteo-lineage trajectory.

We edited the manuscript to make this more clear.

It is interesting that the differences in gene expression during osteolineage differentiation might be connected to skeletal abnormalities in adults with Down Syndrome. The trajectory analysis potentially linking endothelial cells to osteolineage cells, with differences noted in Ts21, is an intriguing finding.

The overexpression of chromosome 21 genes in all cell types examined is significant (supplemental Fig. 8). Did which genes that were overexpressed vary by cell type? The marked down-regulation of expressed genes did not occur in late erythroid cells (supplemental fig. 8). Could this contribute to alterations in newborns with Down Syndrome-polycythemia, macrocytosis?

The genes overexpressed did vary by cell type and environment. While every cell type showed overexpression of some chr21 genes, individual chr21 genes were differentially expressed in a mean 5.1 and a median 3 cell types in liver samples.

To test if some specific chr21 genes have higher dysregulation potential, 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).

In the main text, we made the following addition:

"We were unable to identify individual genes (such as particular chromatin modifiers or TFs) that drive dysregulation, suggesting the possible contribution of feedback loops or synergistic regulation."

And added the analysis to methods section, "**Analysis of Ts21 versus disomic differentially expressed genes"** and Supplementary section **"Evaluating cross-dependent dysregulation of chr21 genes versus other genes".**

Related to the second comment, we saw down regulation of expressed non-chromosome 21 genes in late erythroid cells. This could suggest a more severe phenotype for late erythroid cells within the femur that is not observed in the liver; and as suggested by the reviewer, this interplay could be one factor contributing to polycythemia and macrocytosis.

It is an intriguing conjecture that Ts21 and 'healthy' HSC/MPP may have not only cell intrinsic differences but also different microenvironments and different responses to liver and marrow microenvironments. The analysis of both compartments is a strength of this paper.

The multiome analysis provides evidence that Ts21 HSCs have an erythroid bias (Fig. 3b) consistent with the polycythemia seen in Down Syndrome neonates. I did not understand the MIRA analysis with HSC 'branches' and 'topics' and look forward to a clarified text.

We edited the manuscript to make this more clear. Please see the analyses within the following Main Text Results section "**Multiome analysis reveals erythroid priming in Ts21 HSCs in foetal liver".**

Importantly, "HSCs" were sorted from Ts21 and 'healthy' livers and cultured in vitro. These experiments provide functional evidence for the erythroid bias of these cells proposed by multiple bioinformatic analyses.

While the authors bring up the marked predisposition of young children with Down Syndrome to leukemia (ML-DS) based on mutations in GATA1 resulting in a shortened form of the protein (GATA1s) it is not clear how the findings in this paper relate to this at all, since there was no mention in the findings of cells harboring GATA1s expression or mutations.

Apparently, ~20-30% of Down Syndrome infants have detectable GATA1s cells. Is it possible from the transcriptomic data that were obtained to identify GATA1s hematopoietic cells? If not, then could some of the fetuses have GATA1s cells in various proportions effecting the results? Newborns with Down Syndrome without GATA1s have characteristic changes in their circulating red cells (polycythemia, macrocytosis) and platelets (many references). It is these changes that are reflected in the biological characteristics of hematopoiesis that are delineated in this paper.

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[11]. In line with this, the larger screening studies in newborns have shown that the incidence of GATA1 mutations is around 4%[12–14] . Furthermore, Haasart et al. did not observe a mutation in *GATA1* in any of the foetal stem and progenitor cells following WGS of HSPCs from 9 independent human fetuses gestational age week 12–17.

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.

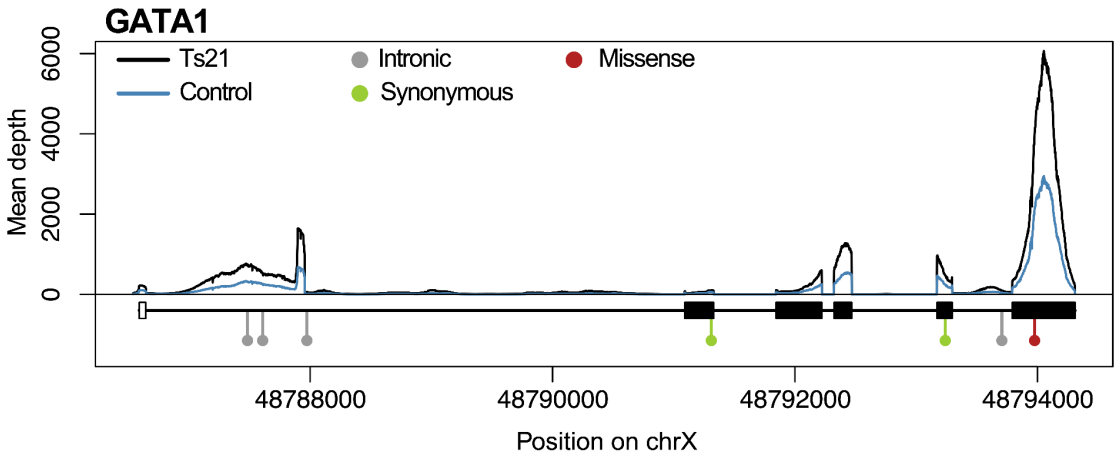



***Figure 3.*** *Mean number of reads for each site spanning the GATA1 locus, black line for Ts21 samples and blue for control samples. Boxes indicate exons of GATA1 (black =coding, white=non-coding). Detected mutations are annotated by their positions, as well as their predicted effect (grey=intronic, green=synonymous, red=missense).*

These limitations notwithstanding, we used the variant calling framework based on deepSNV/Shearwater[15,16] 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 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.

We have included this analysis in the manuscript within the Supp. Results section titled “Detecting GATA1 mutations in foetal samples using RNA-seq”.

Minor:

1.Page 2, bottom paragraph: CD45, despite its billing as a pan-haematopoietic marker, does not capture all haematopoietic cells (or, I suspect, even most of them).

We have amended this sentence.

2. Page 3, middle paragraph: it is stated that the “number” of myeloid cells was decreased in the Ts21 liver. Is it the ‘number’ or the ‘proportion’ of myeloid cells that are decreased?

We have changed the “number” into “proportion” in the main text.

3.I’m not sure if “healthy” is the most appropriate term for the control fetuses. How was their ‘health’ assessed? Might ‘normal’ or ‘disomic’, or some other term, be more appropriate?

The reviewer raised an important point here. We felt that using the word ”normal” might not be appropriate and might sound insensitive, and for that reason we had previously opted for the term disomic even though it is not ideal. The term “disomic” might be the best of all options, so we have changed the text and figures accordingly throughout the manuscript.

4.Page 4, middle paragraph: “We next assessed which ligand receptor pairs are involved…“. Did you actually assess which are involved or did you look for the presence of putative/potential/possible interactions based on mRNA expression?

The reviewer is correct, we looked at the possible interactions based on the mRNA expression. We have now corrected the text: “We next calculated, based on the co-expression of genes in different cell types clusters, which ligand-receptor pairs (L-Rs) are potentially involved in HSC/MPP cell interactions in Ts21 and the disomic foetal liver”

5.Page 4, middle paragraph: The ligand-receptor pair analysis “revealed activation of the ERK1/2 cascade.” I assume that activation requires phosphorylation. Might ‘engagement’ be a more appropriate term for ‘activation’?

We would like to thank the reviewer for pointing out these inconsistencies in the use of terminology. We have now corrected the text to “engagement of the ERK1/2 cascade”.

6. In Figure 3a it is curious that 2 types of macrophages but no monocytes are identified.

We have now corrected this. The populations that we observed in the foetal liver are monocytes and Kupffer cells (macrophages).

7. Page 10, third paragraph: the 2 references to ‘Figure 4e’ should be ‘Figure 4g’.

We have now corrected this omission.

Reviewer #4

Here, Marderstein and colleagues perform a comprehensive RNA and ATAC survey of an early time point of human fetal development to resolve molecular, niche and gene regulatory differences among comparable hematopoietic populations in down syndrome and controls. The authors should be commended on the depth and number samples considered in these analyses, including the use of orthogonal single-cell and spatial profiling methods. This includes in depth profiling of the hematopoietic niche which other fetal studies have largely ignored. The results presented, if true, represent a potentially valuable dataset and analytical results to inform future studies of leukemia development in children. Unfortunately, the manner in which data is presented and analyzed raise significant concerns that hinder the ability of this reviewer to confidently believe the single-cell and molecular findings. In particular, statistical rigor is lacking in the single-cell comparisons due to imprecise inference and comparison of cell populations which likely biases the reported transcriptional, DNA accessibility and spatial differences. The authors further miss an opportunity to explore and validate cellular pathway impacts in each of the identified cell-states, beyond simply separating them into chr21 versus non-chromosome 21 and focusing on HSC. Further, with the exception of Figure 1d, it does not appear that the samples are considered as independent biological replicates. More importantly, there is insufficient experimental validation of the author generated hypotheses (niche, ligand-receptor, gene regulatory) based on these data. Finally, it is not clear what the novel verified findings are or mechanisms of action beyond prior literature (increased cell-cycle gene expression induction Ts21 HSC, mitochondrial dysfunction, etc.).

Major concerns

A particular problem common to single-cell atlas-level disease association papers is that individual specimens and donors are often not considered as covariates in the analysis, but rather a mix of cells that ignores their inherent heterogeneity (race, genetics, sex, developmental age, etc.). Are the results statistically robust considering replicate samples as pseudobulks (e.g., not driven by a few samples of a certain sex)?

Our original 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 (Figure 1).

1. Sample-level pseudobulks with the addition of sorting accounted for as a fixed effect (the original analysis in the manuscript)
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 fetus. 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). Thus, we found our results to be statistically robust regardless of these analysis choices.

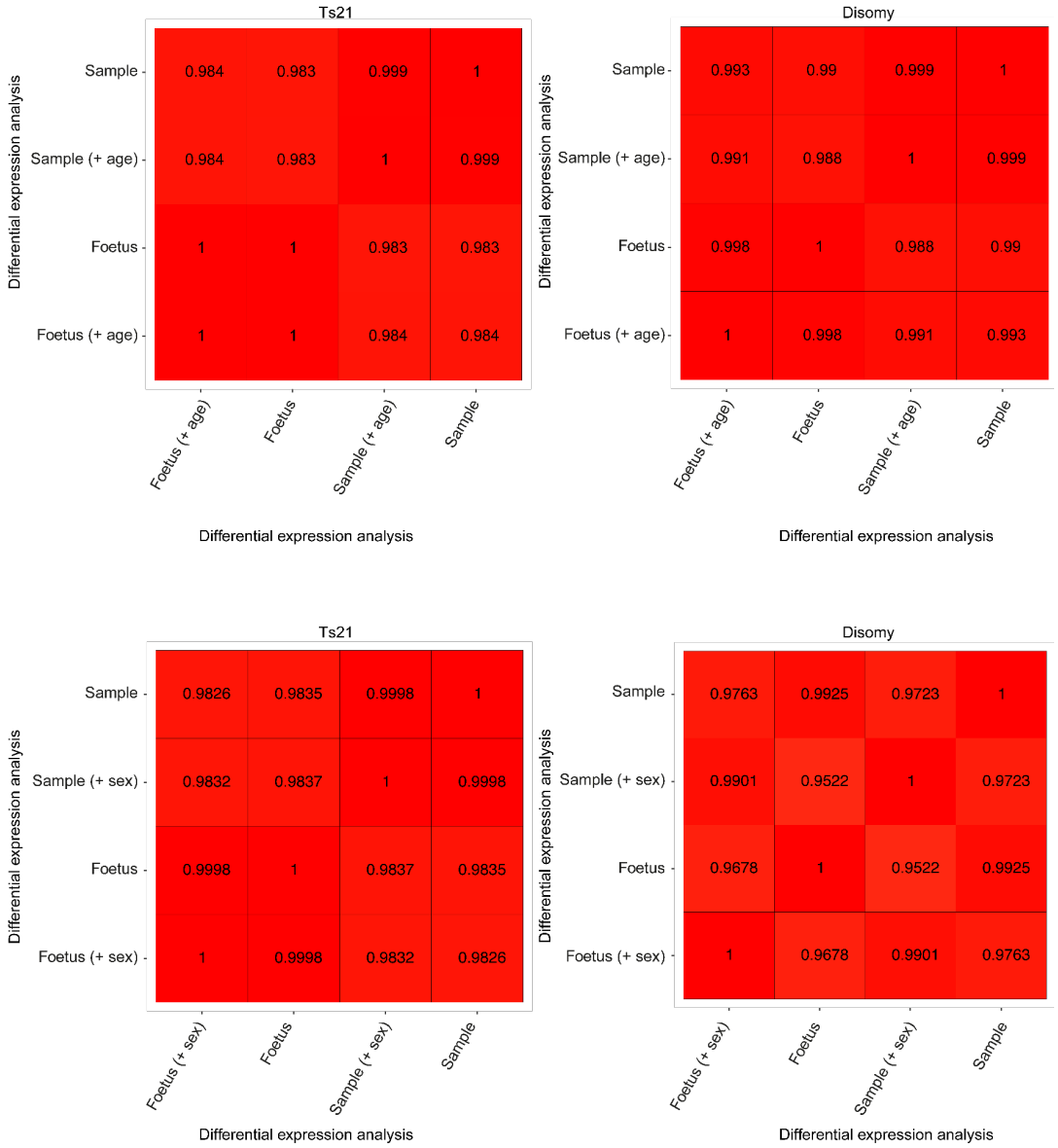



***Figure 1.*** *Pearson's correlation coefficient between Liver vs Femur log-fold changes across all genes for the four analysis types, using age (top row) or sex (bottom row) as covariates.*

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 ($FDR < 0.1$). Therefore, sex did not have a large impact on differential expression results (Figure 1).

In response to the comment about developmental age, we additionally conducted an exploratory analysis of its association with cell-type composition (Figure 2). 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 visualized 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.

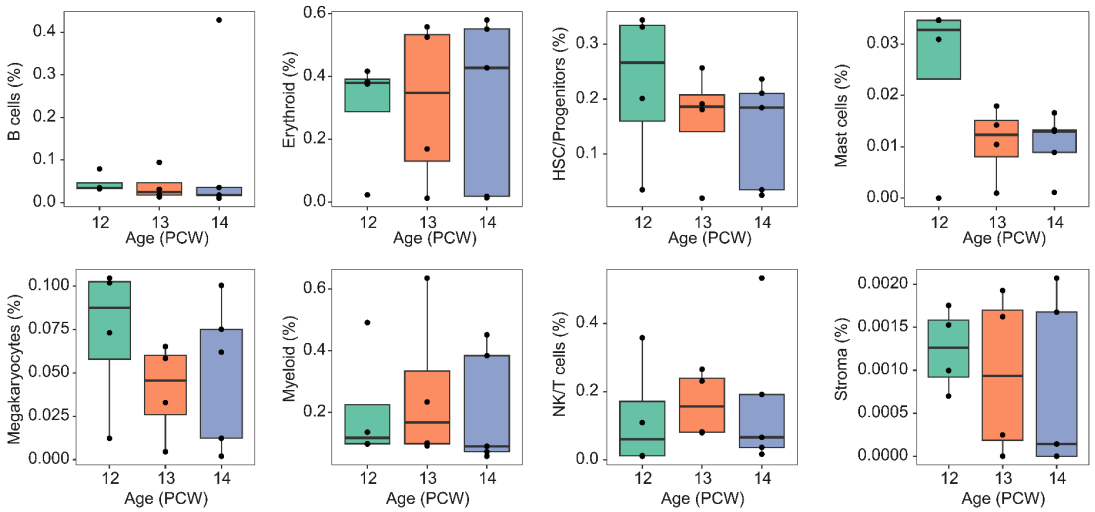



***Figure 2.*** *The relationship between post-conceptual weeks and cell type abundance (estimated as proportions) for the 8 broad cell type groups.*

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 (Figure 3). 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 2b in the original manuscript, 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 $r = 0.03$ ($P = 7.7 \times 10^{-6}$) when compared to the original analysis from the manuscript.

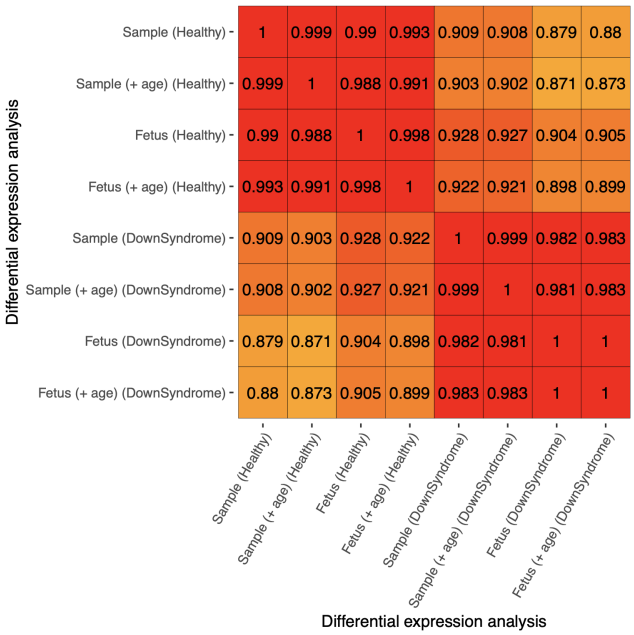



***Figure 3.*** *Pearson's correlation coefficient between Liver vs Femur log-fold changes across all genes for the four analysis types in the two datasets (8 distinct analyses in all).*

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 (see the details below in response to a latter reviewer comment regarding the validity of cell-type annotations within and without integration). 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.

We included this analysis in the Supplementary Results section **"Evaluating the robustness of differential expression results to analysis decisions".**

There are significant concerns regarding the computation of cell cluster frequency, cell identity and gene expression differences, when separate versus joint-integrated clustering is performed on independent aggregated datasets (Ts21 liver, Ts21 marrow, healthy liver, healthy marrow alone). Hematopoiesis has been described as a cloud, with cluster subdivisions within that cloud remaining highly inconsistent between different sample sets and clustering methods. Here, the authors infer similarity between healthy and Ts21 based on similar markers. While this apples-to-oranges comparison is unfortunately often reported in literature, it is subject to inherent biases in clustering of different datasets. This is most notable for cycling HSCs/MPP, which are frequently observed in human adult marrow by

scRNA-Seq (PMC6296228), but are reported to be absent in all but Ts21 liver (a major claim of the paper).

It is well documented that 90-95% of adult HSCs in the bone marrow are quiescent[17] i.e. in G0. So, the population of cells annotated as “cycling HSCs” in adult bone marrow likely represents cells cycling at a different rate compared to the “cycling HSCs” observed in foetal liver Ts21. To support this, we plotted the expression of *MKI67* across multiple cell types annotated in PMC6296228 (Figure 4). The plot clearly illustrates that the adult “cycling HSCs” have a limited expression of *MKI67*. We would like to point out that in our study, the “HSCs” cluster also contained cycling cells in all conditions: 29.8% of HSCs in DS liver, 21.8% in disomic liver, 26.9% in Ts21 BM, and 18.3% in disomic BM. In contrast, 99% of the cells in the “cycling HSCs” cluster were in G2M/S phase. Overall, our findings lead us to conclude that there are significant differences in the cycling properties of HSCs identified in Ts21 foetal liver and those identified in adult bone marrow.

In the following response (please see below), we directly address the concerns regarding cell cluster frequency and identity, and in the previous response we specifically addressed the concerns regarding gene expression differences.

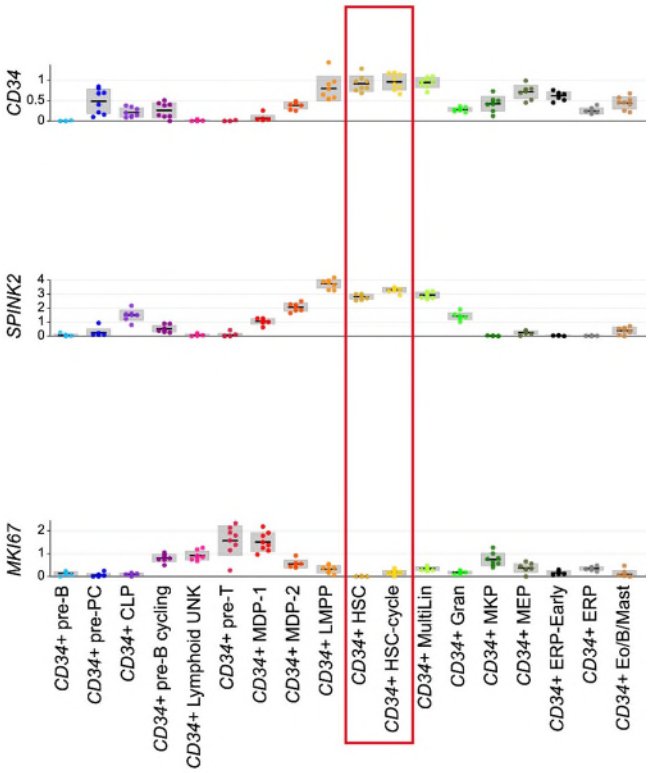



***Figure 4.*** *Box plot showing expression of stem cell markers (CD34 and SPINK2) and MKI67 across progenitor populations identified in Hay et al., 2018. Box in red is framing expression of relevant genes in HSCs and cycling HSCs.*

The authors attempt to address this in Supplementary Figure 14 through a joint analysis of samples, but it appears that the cell frequency results reported throughout are on the inferred analogous clusters independently determined versus using this integrated tissue/disease combined dataset, without confirming results in the clusters derived from all samples. The authors should not only verify the frequency differences using these two orthogonal methods but include similar analyses of label projection to wild-type only cells (e.g., using the combined joint clusters since only a few wild-type cells should be present for all clusters), in addition to flow cytometry validation, to ensure the accuracy of these predictions. It would be expected that a robust integration (e.g., Figure 3a) should enable the identification of all liver and marrow populations (healthy and disease).

To address this question we have performed the following analysis:

1) Compared the validity of our original cell type annotations within our scRNA-seq data using new annotations derived from multiple other methods:

(i) Harmony-based integration and clustering,

(ii) scVI-based integration and clustering,

(iii) reference-query mapping of disomic liver using Pulpescu et al. Nature 2019, and

(iv) reference-query mapping of disomic liver using the Ts21 liver separate-dataset annotations,

(v) reference-query mapping of disomic femur samples using Jardine et al. Nature 2021

2) Assessed the robustness of differential abundance and differential expression analyses to the choice of annotations from the point (1).

3) Compared the abundance of immunophenotypic HSCs, progenitor cells and mature cells in Ts21 (n=3) and disomic (n=3) using the 22-antibody panel for flow cytometry.

First, we integrated the full dataset using Harmony (Figure 5) 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 neighbor 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 (Figure 6), validating accurate annotation of cell types.

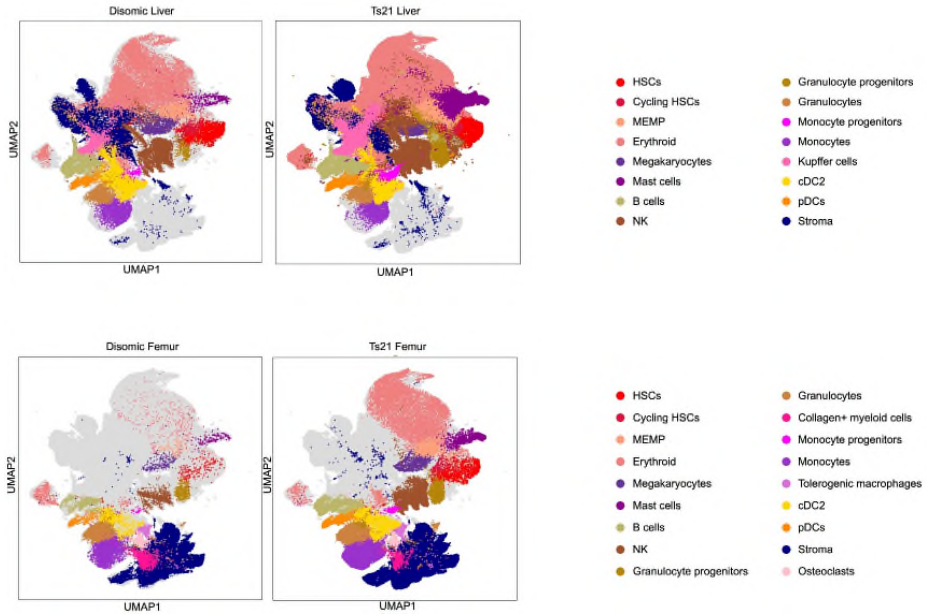



***Figure 5.*** *3D Uniform manifold approximation and projection (UMAP) visualisation of cells from Ts21 liver (n=780,299, k=15), Ts21 femur (n=162,775, k=12), disomic liver (n=110,671, k=3) and disomic femur (n=53,807, k=3) following integration of datasets using Harmony. Cells are coloured by broad cell type categories. n refers to the total number of cells, k indicates the number of biologically independent samples (foetuses).*

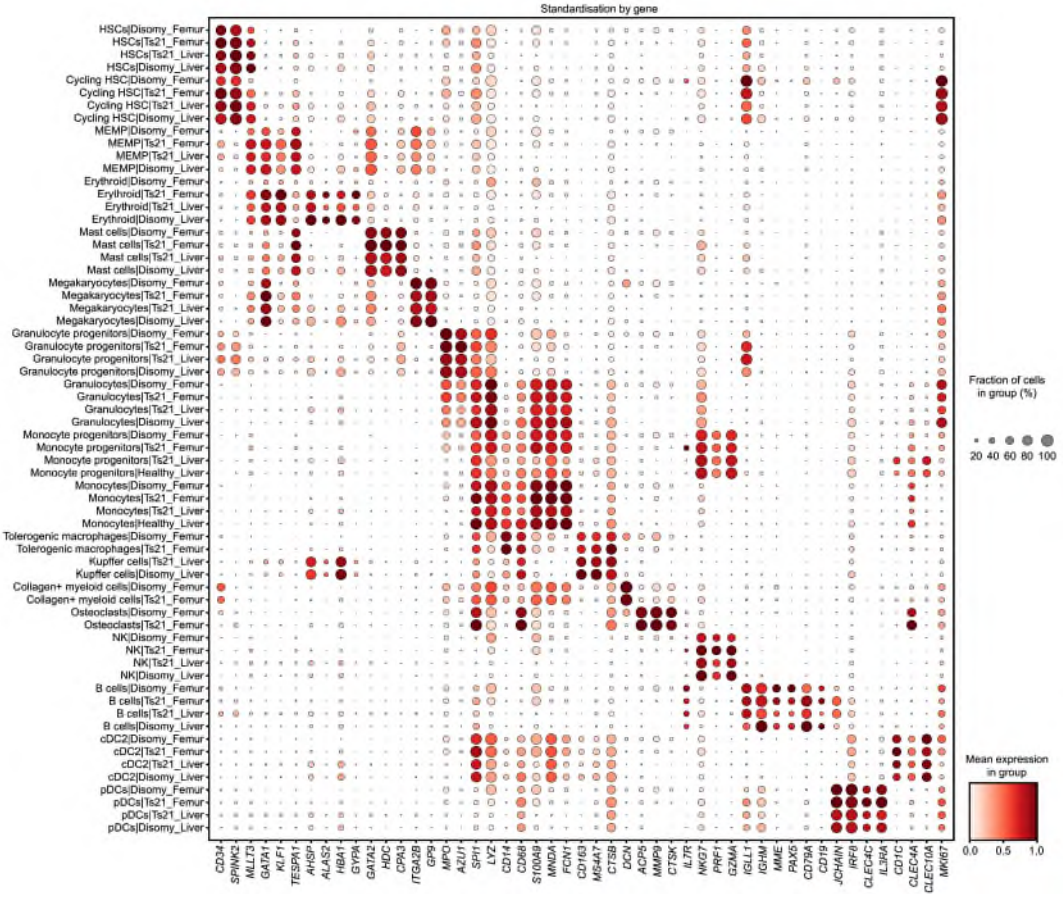
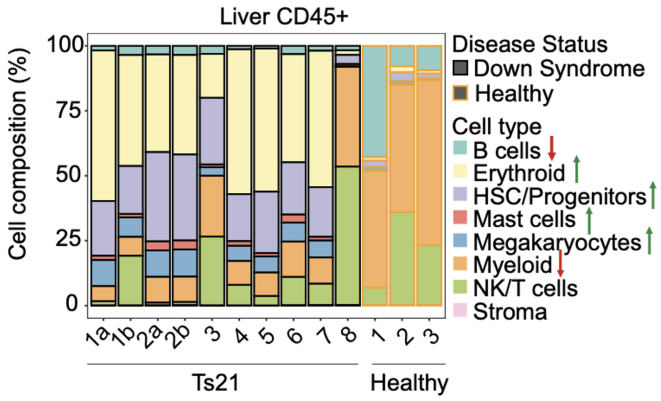


***Figure 6.*** *Gene expression dot plot of marker genes from Supp. Figure 3 and 4 across cell types identified from Harmony-based dataset integration.*

Next, we compared the Harmony integration-identified cell type abundance analysis to our separate cell type abundance analysis shown in Main Text Figure 1c.




***Main Text Figure 1c (copied from original submission).*** *Stacked bar plot of relative abundances of different broad cell types in stage-matched Ts21 (black outline) and disomic foetuses (orange outline). Arrows at the side of the labels indicate a significant increase (↑) or decrease (↓) in the number of cells in Ts21 liver, compared to disomic liver. Difference in cell proportion was tested by Mann-Whitney U test and significance was determined by FDR < 0.1.*

The results were remarkably consistent between the two approaches (Figure 7). The only difference being that, while the trend remained consistent, mast cells were not anymore significantly differentially abundant (FDR = 0.103).

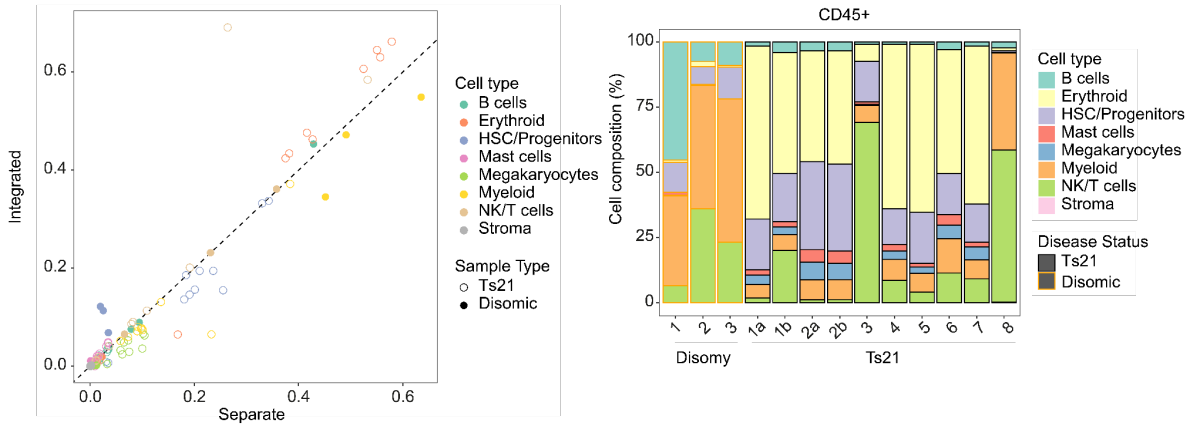


***Figure 7**. Using the Harmony-integrated cell type labels in stage-matched Ts21 and disomic foetuses, we compared cell type proportions between separate and integrated (left), and (right) made a stacked bar plot of relative abundances of different broad cell types similar to Main Text Figure 1c. In (left), each point represents the estimated abundance in a different sample. Panel (right) can be compared to Main Text Figure 1c, which has been copied above.*

Given 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) (Luecken et al. 2021 *Nat Methods*), we repeated the integration and clustering using scVI (Figure 8 and 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 neighbor graph, computing Leiden clusters, and finally annotating clusters.

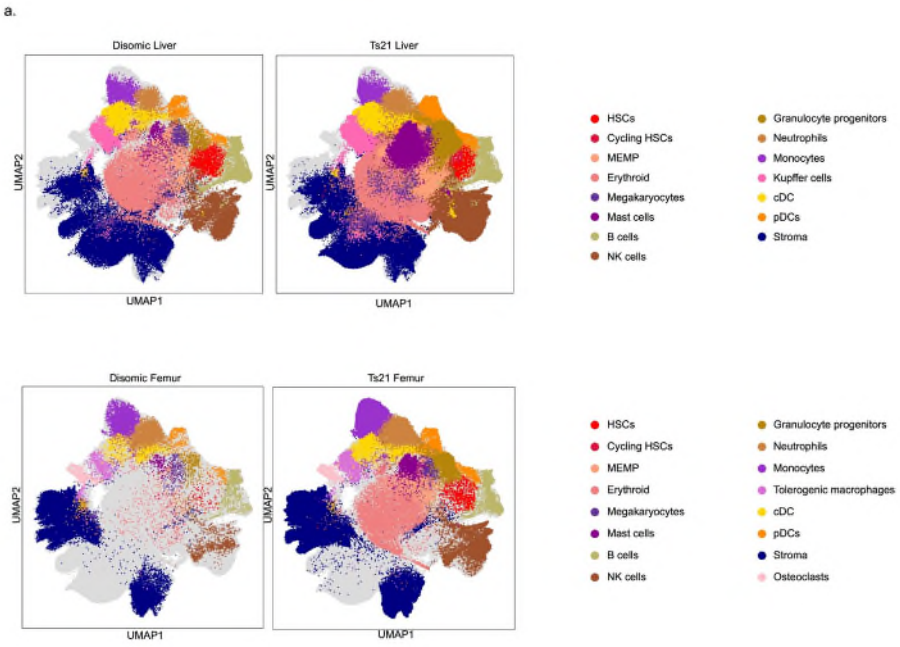


***Figure 8.*** *3D Uniform manifold approximation and projection (UMAP) visualisation of cells from Ts21 liver (n=780,299, k=15), Ts21 femur (n=162,775, k=12), disomic liver (n=110,671, k=3) and disomic femur (n=53,807, k=3) following integration of datasets using scVI tools. Cells are coloured by broad cell type categories. n refers to the total number of cells, k indicates the number of biologically independent samples (foetuses).*

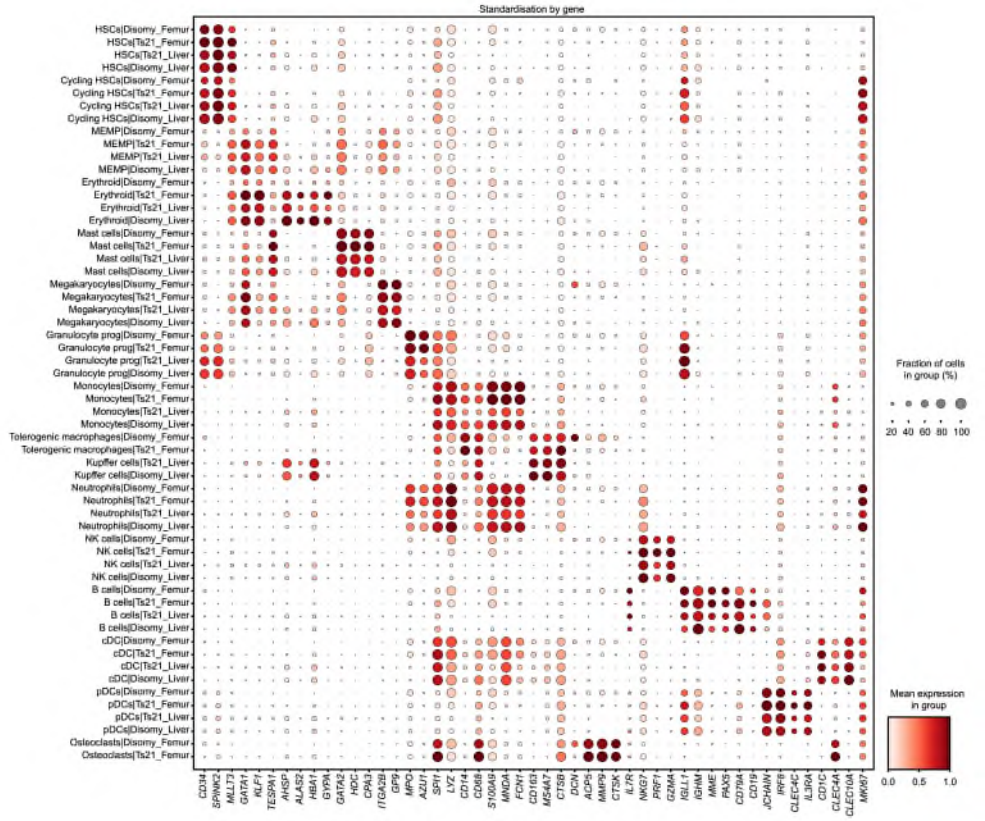



***Figure 9.*** *Dot plot showing expression of marker genes from Supp. Figure 3 and 4 across cell types identified following dataset integration using scVI tools.*

The scVI-integrated proportions across samples correlated closely with separate dataset proportions ($r$ = 0.957) (Figure 10 left), 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 (Figure 10 right).

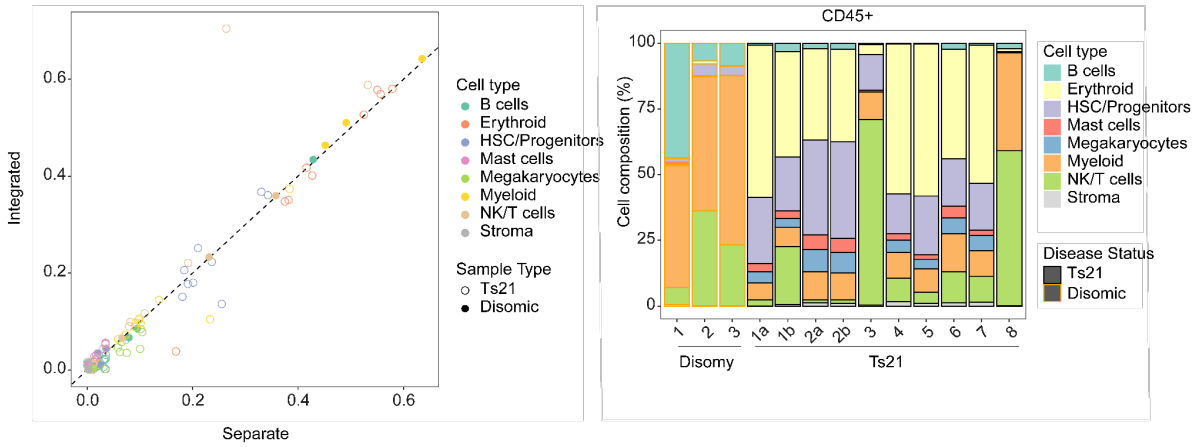



***Figure 10***. *Using the scVI-integrated cell type labels in stage-matched Ts21 and disomic foetuses, we compared cell type proportions between separate and integrated (left), and (right) made a stacked bar plot of relative abundances of different broad cell types similar to Main Text Figure 1c. In (a), each point represents the estimated abundance in a different sample.*

Next, we performed reference query mapping from Popescu et al. 2019 Nature 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 neighbors 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 Pulpescu et al. 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; Figure 11, left).

Finally, 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; Figure 11, right). 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.

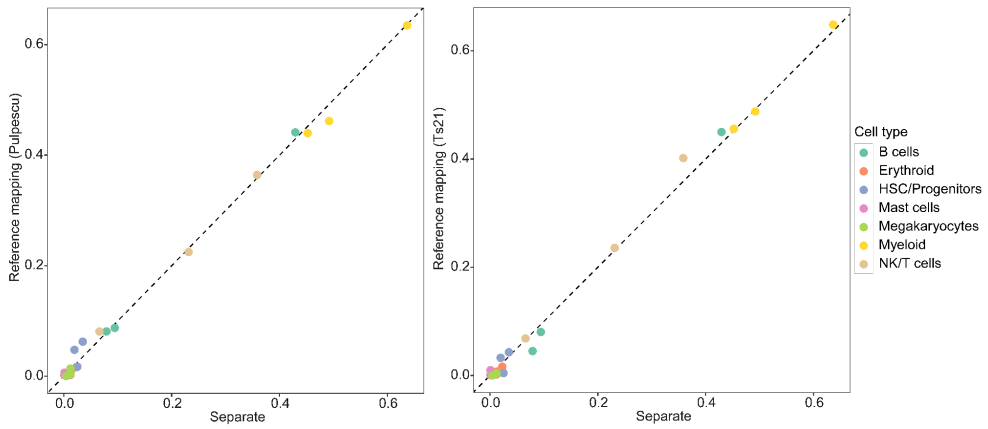



***Figure 11.*** *Comparison of cell type proportions for stage-matched disomic liver samples from the separate-dataset annotations to the reference query label transfer annotations using Pulpescu et al. (left) or the Ts21 dataset as a reference (right).*

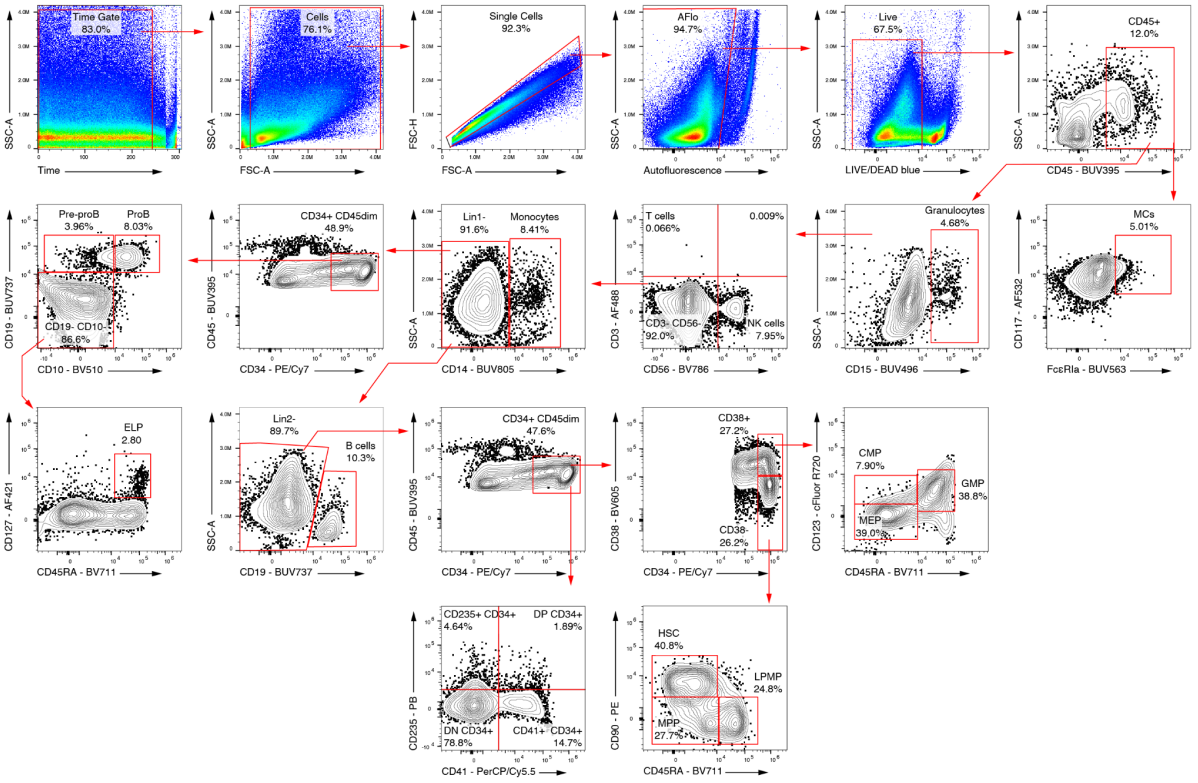
Overall, our results demonstrate that the cell-type annotations are consistent across all annotation strategies and the differential abundance results are robust. We include these results within the Supplementary Results section of the manuscript, under “Evaluating cell-type abundances following data-set integration and label transfer”, and mention in the Main Text:

“We confirmed the robustness of our cell-type annotations and differences in frequency of cell types between Ts21 and disomic liver samples after integration using either Harmony or scVI, as well as reference-query mapping from a published disomic foetal liver atlas and our Ts21 liver cells to disomic liver cells (Supplementary Results; Supp. Figure 7-9).”

To further confirm that the Ts21 causes wide perturbation in foetal haematopoiesis, we compared Ts21 with the disomic immunophenotypic stem and progenitor compartment using flow cytometry (Figure 12). In line with previous reports, we observed an increase in HSCs (CD34+CD38-CD90+CD45RA-) and megakaryocyte-erythroid progenitor cells (MEPs, CD45+ Lin-(CD15/3/56/14/19) CD34+ CD38+ CD123- CD45RA-) but a decrease in multipotent progenitors (MPPs, CD45+ Lin-(CD15/3/56/14/19) CD34+ CD38- CD90- CD45RA-) and granulocyte-monocyte progenitors (GMPs, CD45+ Lin-(CD15/3/56/14/19) CD34+ CD38+ CD123dim/+ CD45RA+) in Ts21 compared to disomic cells.

Within the CD34+ compartment we observed an increase in erythroid CD235+ and megakaryocytic CD41+ biased cells. No change was observed in the frequency of common-myeloid progenitors (CMPs, CD45+ Lin-(CD15/3/56/14/19) CD34+ CD38+ CD123dim/+ CD45RA-).

In the B cell compartment, we observed a decrease in early lymphoid progenitors (ELPs, CD45+ CD15- CD3- CD56- CD14- CD19- CD10- CD127+ CD45RA+), Pre-pro-B cells (CD45+ CD15- CD3- CD56- CD14- CD19+ CD10-), Pro-B cells (CD45+ CD15- CD3- CD56- CD14- CD19+ CD10+), and B cells (CD45+ CD15- CD3- CD56- CD14- CD19+), confirming a defect in B cell development. Furthermore, we saw a decrease in the number of granulocytes (CD45+ CD15+) and an increase in the number of NK cells (CD45+ CD15- CD3- CD56+) in Ts21 compared to discomic samples and no difference in Myelo-Monocytes (CD45+ CD15- CD3- CD56- CD14+) and T cells (CD45+ CD15- CD3+ CD56-).



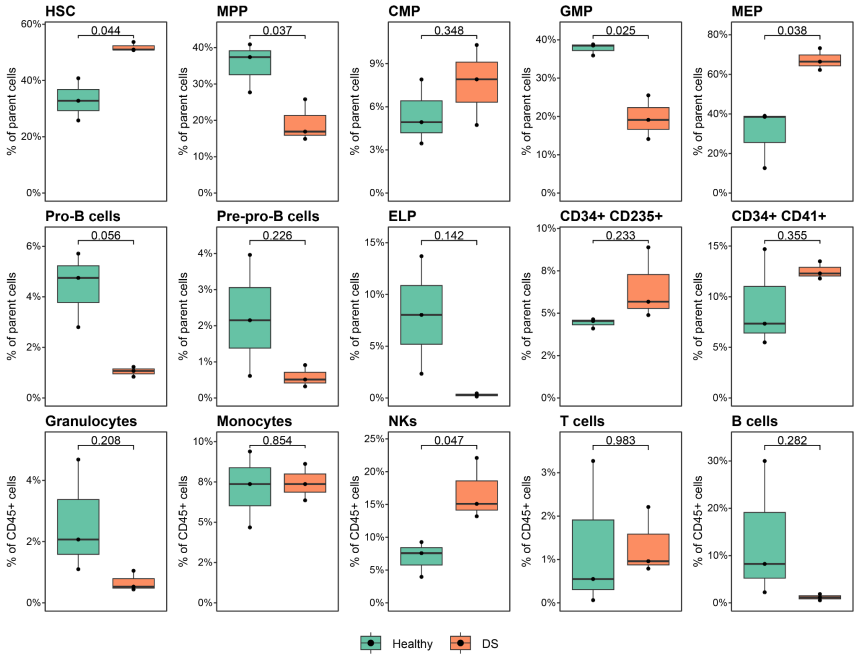



***Figure 12.*** *Top - representative gating strategy used to identify stem cell subsets, progenitors and mature cells in foetal liver. Gates are based on the disomic foetal liver sample. Bottom - Frequency of progenitor and stem cell subsets in T21 and disomic foetal liver. The values on the boxplot indicate the results of a Welch-corrected t-test comparing the percentage of cells in disomic and T21 foetal livers.*

Are the cell frequencies of populations matching between the snRNA-Seq and scRNA-Seq for the same samples? While snRNA-Seq will capture some larger cells, overall, the cell frequencies should be nearly identical. The presence of matching samples with two technologies further enables internal verification of whether the integration between samples and technologies worked.

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 (Figure 13). 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, Massier et al. Nat Comm 2023 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. 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.).

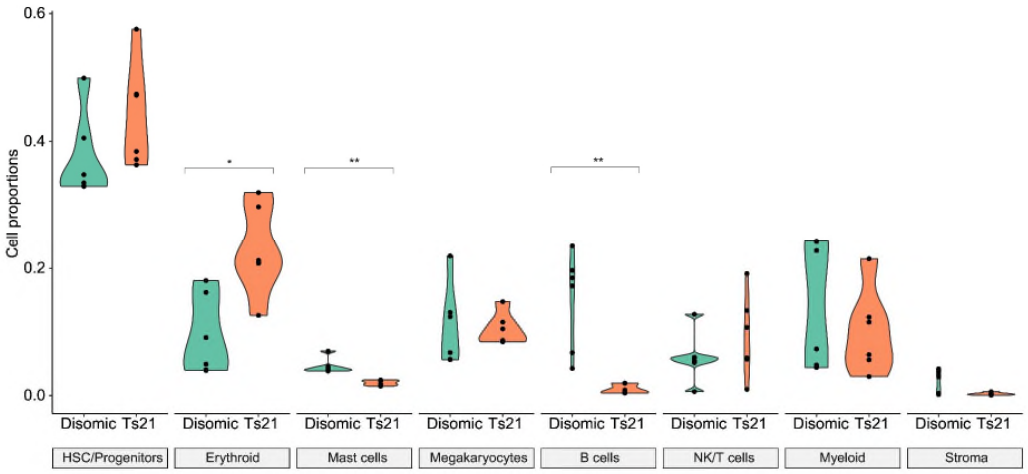
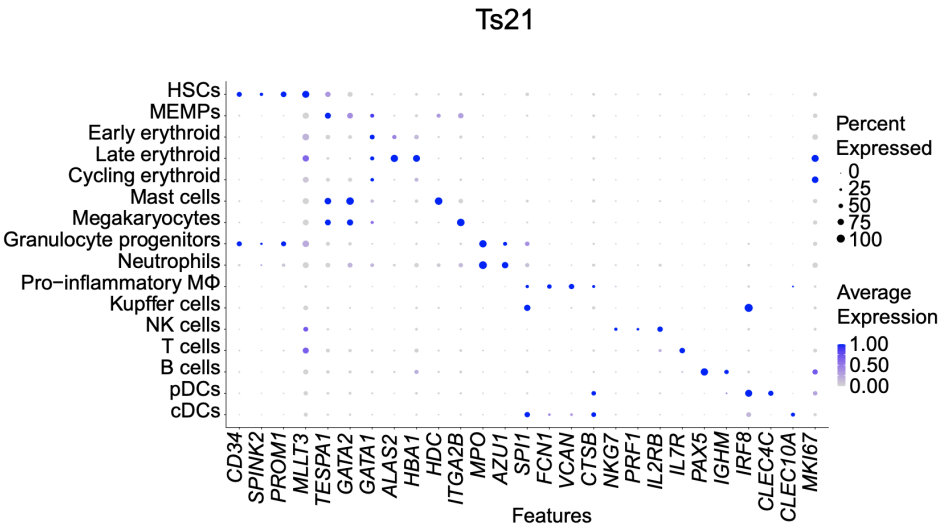



***Figure 13.*** Mann-Whitney U test of differences in cell-type proportions between Ts21 and disomic samples. Significance was determined after FDR correction across cell types. *P<0.05(*); P<0.01(**).*

We have included this analysis in the Supplementary Results.

Specifically, the authors should visualize marker genes for all cells in a few matching samples for all clusters (high resolution clusters) using heatmaps, which show results for all cells and selected genes rather than an average, to illustrate whether Harmony is accurately matching clusters and cells (gene expression markers are consistent and coherent across technologies).

To ensure matching of cell type populations across samples and technologies, we display marker genes for all cells in sample 15582 across clusters. We selected the marker genes that we showed in the Ts21 multiome dotplot (Supp Fig 9c, copied below).




***A copy of Supplementary Figure 14c from the manuscript.***

First, we would like to clarify that we did not perform Harmony integration across technologies, instead treating the scRNA-seq and multiome datasets as separate entities.

Second, we opted for the use of dotplots for exploring expression markers within visualizations. There are graphical challenges with the use of heatmaps due to the inherent sparsity in single-cell data and the large number of cells, which provides challenges for visualizing genes that are lowly expressed and cell populations at low abundance.

Next, we show two technical replicates of 15582 in the multiome data. Overall, the samples are similar in abundances (as specified by the white lines distinguishing each cell type) and expression (as specified by the marker genes for each cell type), indicating that Harmony accurately integrated the data (Figure 14A).

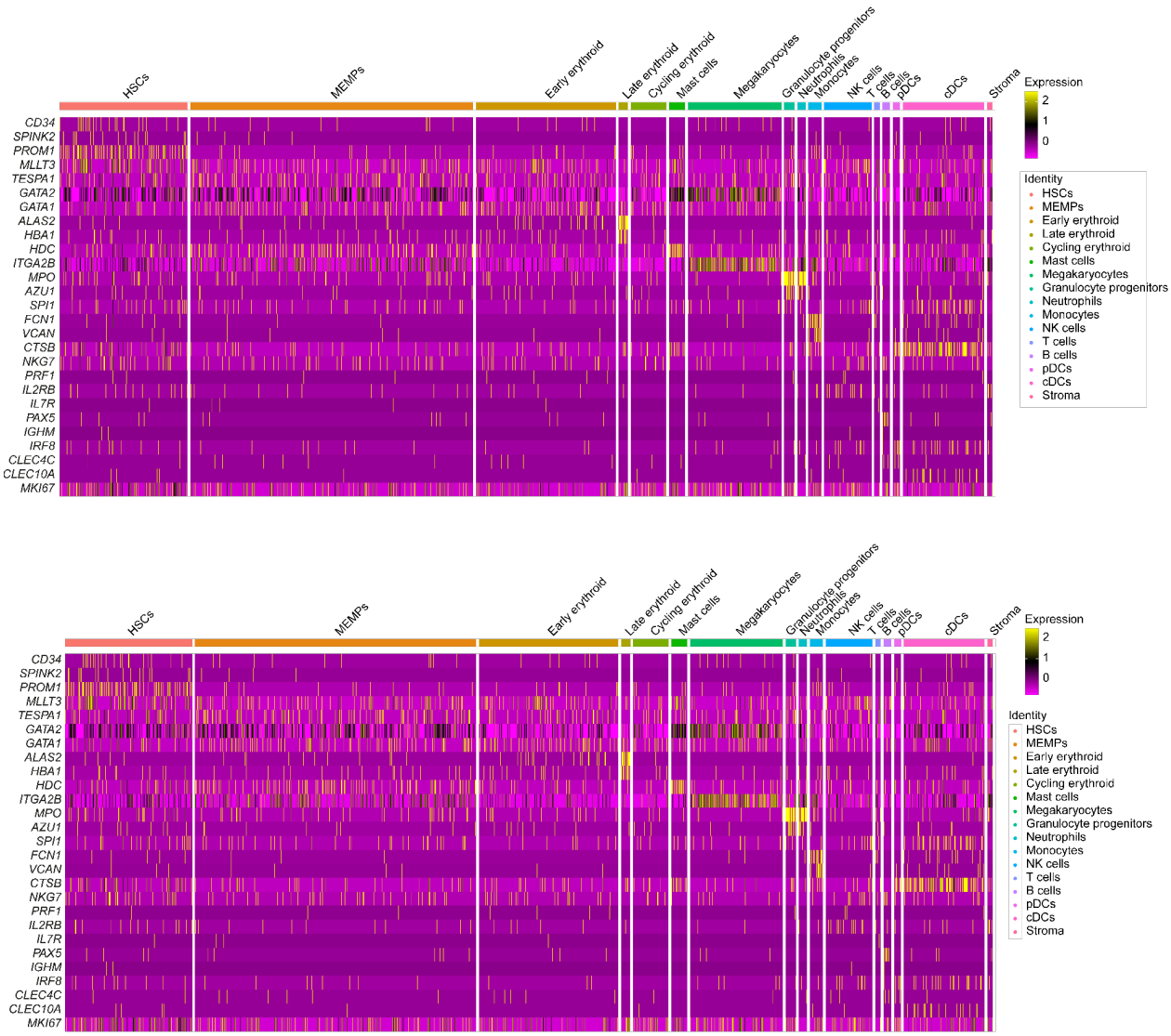



***Figure 14A.*** *Heatmap displaying the expression of marker genes for two technical multiome replicates from the same Ts21 sample. Rows are genes, columns are cells (sorted by population), and each tile of the heatmap represents scaled expression.*

We generated the heatmap using the matching 15582 sample from the scRNA-seq that underwent CD45+ sorting. In this sample, as described in the previous response, we see differences in cell-type frequencies compared to the multiome, but the expression of marker genes are relatively consistent with expression from the same populations in the multiome. We also see general differences in technologies in terms of gene capture; for example, genes commonly expressed in NK cells are much better captured using scRNA-seq technology compared to multiome technology, or *ALAS2* (erythroid marker) is well-expressed in snRNA from multiome but not scRNA.

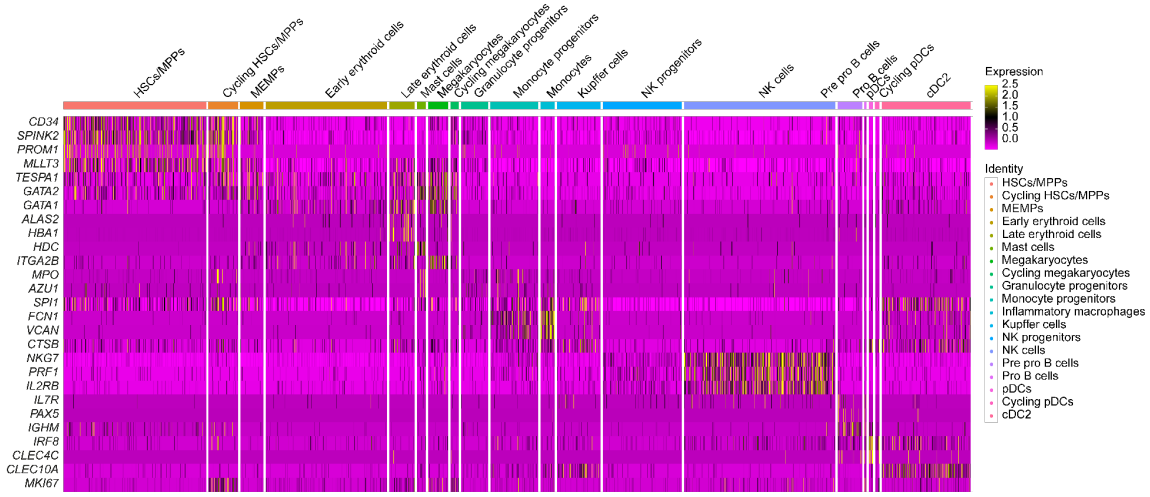



***Figure 14B.*** *Heatmap displaying the expression of marker genes for two technical multiome replicates from the same Ts21 sample. Rows are genes, columns are cells (sorted by population), and each tile of the heatmap represents scaled expression.*

Similar to the above concern, if cell2location is only trained with single-cell references from the matching single-cell samples (e.g., Ts21), it will not be able to assess the presence of cell profiles such as cycling HSCs/MPP in the normal liver samples which are missing in this reference. The authors should provide the H&E images side-by-side with the Visium profiles to confirm that they are comparing similar anatomic structures. Orthogonal experimental validation of Ts21 differences in niche organization and perturbed receptor-ligand interactions are further required to support the author claims.

To analyse the spatial data, we applied cell2location using the matching single-cell as reference data. This approach was chosen to ensure the accuracy of cell type inference, as attempting to infer cell types not present in the sample (e.g. cycling HSCs in disomic liver, or more generally Ts21-driven differences in expression and cell state) could lead to inaccurate results. Because of these key factors, we opted to use the matching dataset instead.

Second, the H&E images from the Visium slides were of low resolution and therefore are not very informative in terms of the morphology. However, we also took higher resolution images of consecutive sections after those that were used for the Visium experiment (Figure 15). It is clear from these H&E images that the morphology of the foetal liver is not very complex at this stage of development. It should be noted that sections for both disomic and DS Visum were taken from the remaining of the tissue after the part of the liver was used for the scRNA-Seq experiment. Considering that the foetal liver is 15-20 mm in lenght at 12-14 pcw (pea size), it is not feasible to make a specific choice of the area that will be used in Visium. Therefore, parts of the liver used for scRNA-Seq and for Visium were chosen randomly and no systematic biases in sample collection were possible. Furthermore, the presented results in Figure 1 are the average of all sections taken in Ts21 (n=20 sections; 2 consecutive sections from 10

biological replicates) and disomic (n=10 sections; 2 consecutive sections from 5 biological replicates) respectively, which together with the previous point makes it unlikely that the results were affected due to the systematic bias in sampling of tissue.

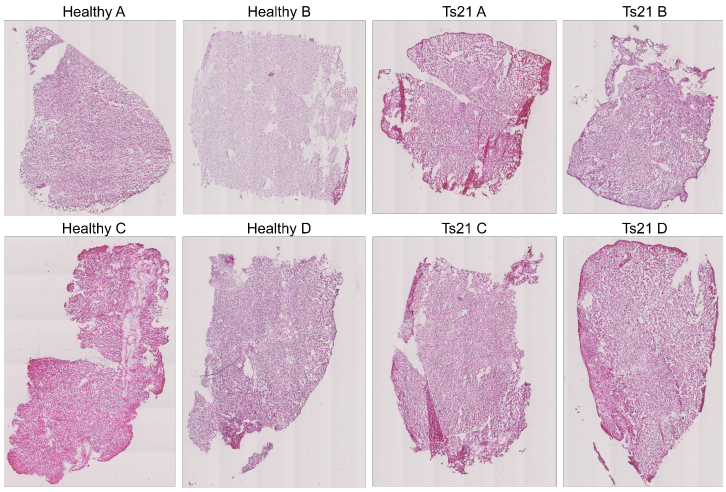



***Figure 15.*** *Haematoxilin and eosin staining of Ts21 and disomic foetal liver sections.*

Experimental validation of the spatial transcriptomics results would require collecting additional material which we were not able to do. This analysis could be part of future investigations which would specifically focus on the question of detailed characterisation of niche and ligand-receptor interactions in Ts21 vs disomic samples.

What signaling pathways are perturbed in the distinct cell-types (beyond simply chr21 versus non) that are likely to impact leukemia predisposition? The authors appear to really only look at HSC in any depth.

Previous work by Wagenblast et al., Science, 2021 suggested that the initial *GATA1* mutation driven preleukemia occurs in HSCs and not downstream progenitors. We have therefore focused on HSCs in our analysis as a cell of origin of preleukaemia. Our DEA suggested that processes related with the oxidative stress are upregulated in Ts21 *versus* disomic HSCs. In line with these findings, previous work by Hasaart KAL et al. (doi: 10.1038/s41598-020-69822-1) showed that HSPCs from Ts21 foetuses have higher somatic mutation load and the presence of signature SBS18 which has previously been associated with oxidative stress-induced mutagenesis. Finally, our analysis of multiome data revealed that Ts21 HSCs have 4.1 times more peak-gene links and that the vast majority of these were significantly more accessible and upregulated in Ts21 compared to control. This opens up a possibility to link altered chromatin organisation in Ts21 and oxidative stress with somatic mutations accrual thus providing an alternative perspective on the etiology of leukaemia in DS (rather than particular signalling in distinct cell types). Notably, chromatin organization is well known to impact somatic mutation rates in different cancers.

To test our hypothesis that Down's syndrome impacts chromatin accessibility potentially leading to higher mutagenesis due to increased exposure to oxidative stress, we conducted an analysis using somatic mutations from Hasaart et al in foetal DS and disomic HSCs. We found that DS mutations located in the introns of Ts21 HSC-expressing genes are 60% more likely to reside within *cis*-regulatory elements compared to intronic mutations identified in disomic HSCs ($P$ = 3x10$^{-4}$; Figure 16).

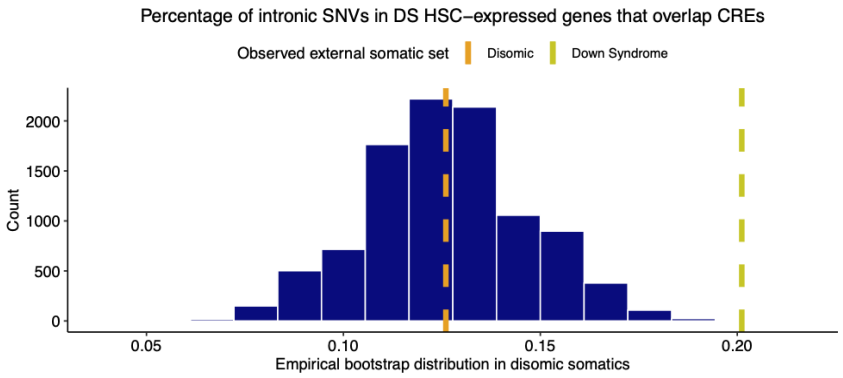



***Figure 16.*** *Cataloged DS mutations (by Hasaart et al.) in the introns of DS HSC-expressed genes (from our scRNA-seq) are more likely to overlap candidate cis-regulatory elements (CREs) from the ENCODE project than cataloged intronic mutations from disomic cells (also by Hasaart et al.). The histogram represents the empirical distribution of this overlap from 10,000 bootstrapped sets of somatic variants found in disomic HSCs.*

This result is in line with the hypothesis that altered chromatin accessibility landscape in Ts21 HSCs provides a milieu for mutation acquisition (possible due to increased exposure to oxidative stress) in regulatory regions of genes. This potentially increases propensity of Ts21 HSCs to preleukaemia and leukaemia development.

We include this analysis in the Main Text.

Will some of the samples at this time point have GATA1 mutations? Did the authors look and if so, are the associations driven by fetuses with these mutations?

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[11]. In line with this, the larger screening studies in newborns have shown that the incidence of GATA1 mutations is around 4%[12–14] . Furthermore, Haasart et al. did not observe a mutation in *GATA1* in any of the foetal stem and progenitor cells following WGS of HSPCs from 9 independent human fetuses gestational age week 12–17.

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[11]. In line with this, the larger screening studies in newborns have shown that the incidence of GATA1 mutations is around 4%[12–14] . Furthermore, Haasart et al. did not observe a mutation in *GATA1* in any of the foetal stem and progenitor cells following WGS of HSPCs from 9 independent human fetuses gestational age week 12–17.

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.

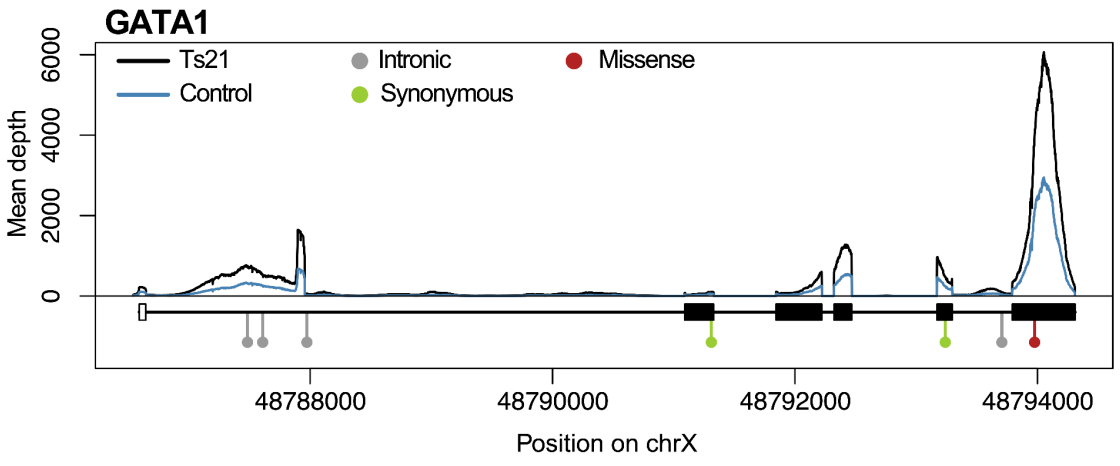



***Figure 3.*** *Mean number of reads for each site spanning the GATA1 locus, black line for Ts21 samples and blue for control samples. Boxes indicate exons of GATA1 (black =coding, white=non-coding). Detected mutations are annotated by their positions, as well as their predicted effect (grey=intronic, green=synonymous, red=missense).*

These limitations notwithstanding, we used the variant calling framework based on deepSNV/Shearwater[15,16] 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 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.

We have included this analysis in the manuscript within the Supp. Results section titled “Detecting GATA1 mutations in foetal samples using RNA-seq”.

The datasets and code should be made available upon review (reviewer access or other tokens provided), to ensure completeness of the resource and code descriptions.

We have uploaded the fastq files for scRNA-seq, 10X Multiome, and 10X Visium to ArrayExpress. We can provide reviewer access to the data and our code upon request, and we will work with the editor to grant this request and maintain anonymity.

Corrections for ambient RNA are needed to ensure accurate assessment of scRNA-Seq and snRNA-Seq gene expression differences, which contribute substantially to false positive gene expression changes when cell frequency differences are significant among captures or

disease. In most cases, >40% of multiome snRNA-Seq counts are due to ambient contamination from lysed nuclei (PMC7335186).

Respectfully, we hold a different viewpoint regarding the need to perform ambient RNA correction. As detailed in the textbook “Orchestrating Single-Cell Analysis with Bioconductor” by Amezquita et al. 2023 (manuscript presenting the textbook found in Amezquita et al. *Nature Methods* 2019), ambient contamination is generally consistent across samples such that spurious heterogeneity should not impact results. Additionally, we observed an average of roughly 50% higher expression on chromosome 21 in Ts21 which aligns with the expectation from an additional chromosome dosage, reflecting that expression differences are true expression differences and not majorly impacted by ambient RNA.

Direct link to the relevant textbook section can be found below.

http://bioconductor.org/books/3.16/OSCA.advanced/droplet-processing.html#removing-ambient-contamination

The authors need to clarify differences in their Ts21 observations and those of Jardine et al. 2021 (PMC7612688), which apply a similar single-cell approach. What accounts for the inconsistent findings and do the authors' findings hold up if they supervise their clusters onto Jardine’s.

It is not clear to us which differences in findings the Reviewer is referring to. To clarify, we did not perform abundance testing in the bone marrow samples as we utilised different sorting strategies in disomic and Ts21 samples. The majority of cells in our BM samples came from CD235- cells (erythroid depletion) and as such they are not suitable for assessing blood cell abundances. As a result we were not able to quantify any potential differences in the frequency of cell types in Ts21 vs disomic BM. Furthermore, Jardine et al. only analysed Ts21 *versus* disomic BM and not Ts21 *versus* disomic foetal liver. Comparing our Ts21/disomic foetal liver results with their BM results would not be suitable. This is an important point as earlier work by others[3] has shown that in DS foetal liver, ***but not marrow (BM)***, frequency of immunophenotypic megakaryocyte-erythroid progenitors was increased and granulocyte-monocyte progenitors was reduced.

More generally speaking there are many important differences that need to be taken into account when comparing the results between our BM study and Jardin et al.

Jardine et al. study was primarily designed as a disomic foetal BM atlas, and therefore they collected BM from foetuses from 12-19 pcw. One of the conclusions from their study is that *cell type abundances change with foetal development*. Furthermore, the analysis of the DS was a small part of their study as they analysed the total 16,743 foetal BM cells from Ts21

(n=4 foetuses). Of these, 12,914 were erythroid cells and the remaining **3,829** were all other cells that included 32 cell types (Supp.Table 10 in Jardin et al.). As a result, the smallest cluster (CAR cells) in Jardine et al. has only 4 cells. The most likely reason for such skewing of cells toward erythroid lineage is related with the sorting strategy used by Jardine et al. (live cells i.e. without any enrichment or depletion, Supp. Table 1). This is very different from the strategy Jardine et al. used when sorting cells from disomic foetal BM that included CD45+, CD45-, and "TOTAL" (probably referring to no sorting or enrichment strategy) cells.

Our study primarily focused on Ts21 and therefore our disomic and Ts21 samples were mostly from 12-14 pcw. Our data set included 161,728 foetal BM cells from Ts21 (n=10 foetuses), which were sorted using CD235- strategy (erythroid cell depletion) and to a lesser extent CD45+. As a result, only 6,768 cells were of the erythroid lineage and the remaining **154,960** were all other cells that included 36 cell types. Our disomic dataset is comparatively small and includes **49,194** cells (n=3 foetuses).

Since we mostly used CD235 depletion strategy, late erythroid cells were largely depleted from our samples as evident from the UMAP (Figure 18) where we projected our disomic BM dataset onto the Jardine et al. data set. 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. 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. data, reflecting that similar cell populations were captured in both studies.

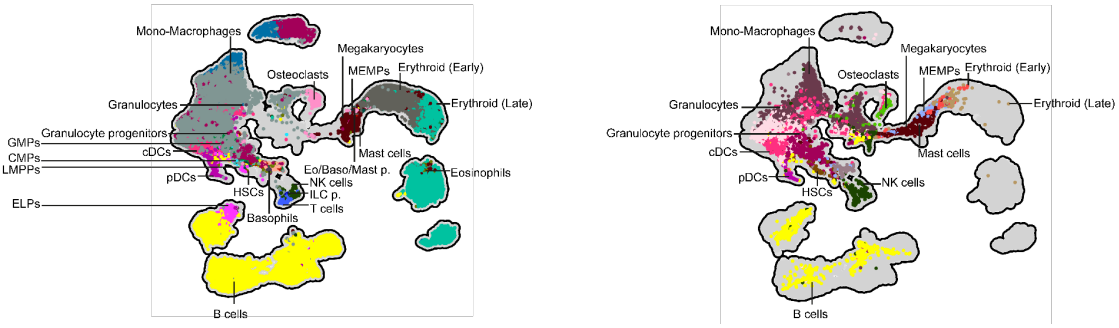



***Figure 18.*** *Integration of Jardine et al. and Marderstein et al. cells from disomic BM using scVI. The Jardine et al. cells are shown on the left, and the Marderstein et al. cells are shown on the right. Cells are colored by the annotated population from their original studies.*

It is possible the single-cell skewing observed in this study is associated with developmental immaturity which can be addressed through the reanalysis of existing single-cell atlases. In particular, prior studies that perform deep temporal and spatial profiling of fetal hematopoiesis, spanning skin, liver, marrow and aorta, using these same technologies.

In response to the comment about developmental immaturity/post conceptual weeks (PCW), we conducted an exploratory analysis of its associations in our data (Figure 2). We found no significant association (nominal $P > 0.05$ for all tests) between cell type group and age. Overall, selected samples were all within a narrow developmental window, between 12-14 pcw, with little effect of PCW on statistical results. Therefore, we don't think that stage of development confounds the observed findings in the manuscript. Importantly, we performed flow cytometry analysis that confirmed observed differences in cell abundances, giving further strength to our conclusions.

We would also like to point out that foetal liver and BM are the only haematopoietic sites during human foetal development and as such the frequency of cell types in these sites compared to skin, and aorta would differ. Therefore, we do not think that re-analysing previously published data-sets of disomic foetal haematopoiesis spanning skin, liver, marrow and aorta would improve our findings.

It is unclear how environmental impacts (e.g., if a gene is commonly or uniquely upregulated in HSC in liver and marrow in Ts21 versus controls) are determined, since different clusters are annotated in the different tissues and samples (apples-to-oranges comparison). How are the authors aligning these cell-states and confirming their accuracy, beyond HSC?

First, it is difficult to ascertain whether a gene is commonly or uniquely upregulated across cell types because each cell type has a different population size. As a result of differences in population size, the presence of a differentially expressed gene in one cell type and the absence in another may be purely statistical. This is the key reason why we adopted a complex downsampling procedure for HSCs, which we used to identify Ts21-induced expression changes between Liver and Femur for HSCs (as we could not simply take the Ts21 differentially expressed genes as the Ts21 population was much larger). If we were to repeat the same procedure across additional cell populations to conclude whether genes are differentially expressed or not in specific cell populations, we would be required to downsample to the smallest population sizes. This would erode statistical power and further be a computationally expensive task. Furthermore, there are several cell types in liver and femur environments and Ts21 and disomic samples that vary widely in frequency such that some cell populations are specific for a particular niche/disease state. This would hinder such an analysis. Due to these concerns, we decided not to pursue the identification of genes widely dysregulated or uniquely dysregulated, as it would be difficult to conduct this computational and statistical task robustly. Instead, as HSCs was our primary cell type of interest, we focused the manuscript and analysis in depth on the HSC population but we shared code, results, and data that would allow other researchers to explore additional cell types in depth.

We now mention this in the methods section titled “Identification of context-specific DEGs in HSC/MPPs”.

The inferred role of SNPs with C11orf21 and TSPAN32 are too inferential and thus require orthogonal experimental validation.

To clarify, we inferred the role of enhancers in the Ts21 context and not directly the SNP’s role/mechanism.

Secondly, we performed the peak-gene analysis for a large number of genes and the highlighted genes are just some of the examples of this large set. Given the large number of genes, a functional analysis of any one gene would be expected to be only a small piece of the effect here—and because of this, functional analysis of a specific gene(s) we feel is not in the scope of the current study. We showed these examples to provide greater detail to the observed trends in the data.

Minor Concerns

“Iterative clustering” is not clearly described or cited.

To address this confusion, we have rephrased the methods as follows:

“We then performed an iterative clustering procedure to identify clusters in the single-cell data. Broadly, our iterative clustering procedure first finds initial clusters using the Leiden algorithm, next merges clusters from seemingly identical cell populations, and finally subclusters into further refined populations using K-means clustering. Thus, the iterative clustering allowed us to further refine initial clustering, such that initial clusters containing multiple cell types can be further split into lower-level cell types. This is particularly useful for broad cell types such as erythroid cells that were further split into early and late erythroid cells."

Cell population marker genes and differentially expressed genes for all cell population comparisons (with statistics) are not provided along with cell-barcode to cluster associations (Supplementary Tables).

We have added the additional information within three supplementary tables in the revised manuscript:

- A cell-barcode by cluster table
- A marker gene table (differentially expressed genes between cells from one cluster versus all other cells)
- A differential expression table of Ts21 versus disomic samples for each cell type in both liver and femur

These are Supp. Tables 2, 15, and 16.

In Supplementary Figures 3 and 4 different colors are used for the same cell annotations which is confusing.

We have corrected this in the revised version of the manuscript.

Further, cell type annotations are displayed in different orders in most plots. The reviewer recommends the authors reorder cell annotations consistently and by lineage and maturity. What is label 30X? I cannot find this in the UMAP.

We have now changed the order of cell types so that is consistent in all Figure and Supp. Figures. Label X was marking a cell type that we were not able to annotate. We have changed the label to “unknown”.

Reviewer #5

The provenance and ethical use of the human foetal bone and liver samples has been documented in the manuscript's ethics statement. Per the authors' statement, these samples have been obtained from the Joint MRC/Wellcome Trust (Grant MR/R006237/1) Human Developmental Biology Resource (HDBR), with maternal informed consent, in accordance with ethical approval by the National Health Service (NHS) Research Health Authority, REC Ref: 18/LO/0822. HDBR is regulated by the UK Human Tissue Authority and operates in accordance with the relevant HTA Codes of Practice.
The HBDR has obtained ethical approval to collect and distribute embryonic and fetal material to all UK based research projects, without the need for individual researchers to obtain their own project specific ethics, providing their project fits within the remit of the HDBR approval (as documented in the HBDR website which contains the ethics approval letters, https://www.hdbr.org/ethical-approvals). The study, which is funded by the European Research Council, seems to fit within this scope.

Given the HBDR authorization, the investigators/authors were not required to make project-based applications for ethical approval. They will be deemed to have ethical approval from the HBDR ethics committee and are able to rely on the attestation pertaining the ethical provenance of the tissues, including the obtention of donor informed consent (maternal). Of note the ethical approval of the HBDR is set to expire in June and December 2023 (Newcastle and London). I hence do not have particular concerns pertaining the provenance and ethical research use of the samples, and therefore - from an ethics perspective, recommend the article for publication.

## References


1. Hasaart, K. A. L. *et al.* Mutation accumulation and developmental lineages in normal and Down syndrome human fetal haematopoiesis. *Sci. Rep.* **10**, 12991 (2020).

2. Antonarakis, S. E. *et al.* Down syndrome. *Nat. Rev. Dis. Primer* **6**, 9 (2020).

3. Tunstall-Pedoe, O. *et al.* Abnormalities in the myeloid progenitor compartment in Down syndrome fetal liver precede acquisition of GATA1 mutations. *Blood* **112**, 4507–4511 (2008).

4. Wagenblast, E. *et al.* Mapping the cellular origin and early evolution of leukemia in Down syndrome. *Science* **373**, eabf6202 (2021).

5. Pendergrass, W., Wolf, N. & Poot, M. Efficacy of MitoTracker Green and CMXrosamine to measure changes in mitochondrial membrane potentials in living cells and tissues. *Cytom. Part J. Int. Soc. Anal. Cytol.* **61**, 162–169 (2004).

6. Probes for Mitochondria—Section 12.2 - UK. https://www.thermofisher.com/uk/en/home/references/molecular-probes-the-handbook/probes-for-organelles/probes-for-mitochondria.html.

7. Mansell, E. *et al.* Mitochondrial Potentiation Ameliorates Age-Related Heterogeneity in Hematopoietic Stem Cell Function. *Cell Stem Cell* **28**, 241-256.e6 (2021).

8. de Almeida, M. J., Luchsinger, L. L., Corrigan, D. J., Williams, L. J. & Snoeck, H.-W. Dye-Independent Methods Reveal Elevated Mitochondrial Mass in Hematopoietic Stem Cells. *Cell Stem Cell* **21**, 725-729.e4 (2017).

9. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

10. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-

cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).

11. Chou, S. T. *et al.* Trisomy 21 enhances human fetal erythro-megakaryocytic development. *Blood* **112**, 4503–4506 (2008).

12. Pine, S. R. *et al.* Incidence and clinical implications of GATA1 mutations in newborns with Down syndrome. *Blood* **110**, 2128–2131 (2007).

13. 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).

14. Hoeller, S. *et al.* Morphologic and GATA1 sequencing analysis of hematopoiesis in fetuses with trisomy 21. *Hum. Pathol.* **45**, 1003–1009 (2014).

15. Martincorena, I. *et al.* Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880–886 (2015).

16. Gerstung, M., Papaemmanuil, E. & Campbell, P. J. Subclonal variant calling with multiple samples and prior knowledge. *Bioinforma. Oxf. Engl.* **30**, 1198–1204 (2014).

17. Pietras, E. M., Warr, M. R. & Passegué, E. Cell cycle regulation in hematopoietic stem cells. *J. Cell Biol.* **195**, 709–720 (2011).

**Reviewer Reports on the First Revision:**

Referee #1 (Remarks to the Author):

I appreciate the authors' response to my comments and additional analyses.

My major comments:
1. "DS mutations located in the introns of Ts21 HSC-expressing genes are 60% more likely to reside within cis-regulatory elements compared to intronic mutations identified in disomic HSCs": I was confused with the authors' claim that altered chromatin accessibility landscape in Ts21 HSCs provides a milieu for mutation acquisition in regulatory regions of genes. The greater overlap of intronic mutations with cCREs than chance does not prove the hypothesis of causal effect of greater accessibility on mutational load. I am even not sure if this greater overlap is a specific finding for DS or a general finding for intronic somatic mutation in general.

2. As I commented in the previous review, the genes highlighted from DEG seemed to be arbitrarily chosen, especially ROS genes (FOXO3, APOE, DUSP1) that the authors later tried to functionally validate, but not based on their significance or pathway's significance. Although authors performed experimental studies on those oxidative stress effects, the rationale for specifically choosing this pathway for validation is still unclear.

3. On the additional experiment on mitochondrial mass and oxidative stress, the violin plot is not appropriate since there are only three samples in each category.

4. As the reviewer #4 also pointed out, can the authors provide the direct comparison of the cell frequencies of each cell population between the snRNA-Seq and scRNA-Seq for each sample who was profiled by both technologies, given the limited replication of comparison between Ts21 and disomic?

Referee #2 (Remarks to the Author):

Marderstein et al. have addressed all comments in a clear point-by-point response. This review process strengthened the conclusions of the study.
Notably also, this manuscript is much clearer, providing insights into the procedures/analyses and a better description of the results.

Altogether, as mentioned in the revised abstract, the study now provides a molecular map of fetal haematopoiesis in Down syndrome. This is the strength of the multi-omic work: building a valuable resource to be used for future research questions in the field of leukemia development in children with Down syndrome.

More specifically, a substantial amount of data have been added:
- New bioinformatic analyses:
o Assess annotations by several new methods
o Integration of already published datasets for validation/comparison
o Challenging the data (role of each chr21 genes, GWAS SNP…)
- New experiment using new primary samples (n=6; 3 DS and 3 non-DS fetal livers) to mitochondrial dysfunction and confirm earlier observation about the perturbed proportion of haematopoietic progenitors in DS fetal livers (PMID: 18812473, PMID: 23045701).

While these new analyses extend our understanding of the potential intrinsic mechanisms, the study is still lacking functional validations. Independent models such as induced Pluripotent cells or animal models could have been used to assess some of the very interesting but speculative comments obtained from the multi-omic data.
As an example, while it has been nicely shown that no chr21 genes, alone, can explain some of the phenotype (expression), genetic engineering approaches to knock out several ch21 genes could be performed to address this comment. This has been done in the past to demonstrate the role of several ch21 genes in leukaemia development in iPS cells (PMID: 27134169).
Notably, a novel ch21 player has been recently described: RUNX1 isoforms, with RUNX1 transcription factor being a well-known master regulator of early hematopoiesis (PMID: 36493345). Has this been assessed? Have RUNX1 binding sites been identified in the enhancer/promoter regions. Any other ch21 TF? ERG? ETS2?

The authors should mention/discuss the known altered pattern of DNA methylation linked to trisomy 21 and how could this be correlated with their expression data, genomic accessibility, erythroid-primed HSC and progenitors and the observed hematopoietic phenotypes.

Moreover, very nice observation highlights the potential impact of the overexpression of TRF2 and TSPAN32 in the erythroid-primed phenotype of the HSC. That could be reproduced and tested in an in vitro setting.

As reported here (Figure 4i) and other, there is a bias toward erythroid and megakaryocytic colonies. While the erythroid primed HSC are well characterized in the study, it is less clear for the megakaryocytic lineage?

For example, have the Authors checked the HSC differentiating toward megakaryocytes like in figure.3b. Also, the megakaryocytic bias is not particularly seen in figure.1d. Please comment.

Finally, because the sample used here allow to better dissect the impact of the liver environment in comparison of previous studies (Jardine et al, 2021), validation of the communications between hematopoietic stem cell and progenitors and the surrounding populations (osteo…) should be validated in independent experiments to better understand the biology and the impact of these cellular perturbation in predisposing leukemia development.

Minor comments:

- Figure.1e: increase size of the text for the pathways negatively and positively correlated.

- Same comment for Figure.2c.

- Figure.2f/g: not sure violin plots are the best way to represent only 3 replicates. Bar graph with average and error bars would be more appropriate.

- Please clarify the calculation of the peaks that are disomic and Ts21 specific. 191 – 60 = 122. But it is mentioned that `62.1% (72/116) are specific to disomic HSC`. 116? or 122? Also, please clarify where the 839 Ts21 specific peak genes come from?

Referee #3 (Remarks to the Author):

The authors have responded in detail to the critiques of the reviewers', providing a 69-page response.

Some new data have been added that strengthen the manuscript. Importantly, much of the text has been now more clearly written with the addition of introductory and conclusion sentences. This has markedly improved the manuscript. However, the middle paragraph of the section involving the interactions between enhancers and the target genes (lines 386-402) continues to remain quite inscrutable to me.

Strengths of the paper continue to be 1) the analysis of a large number of primary human fetal cells taken from liver and marrow of disomic and trisomic fetuses, 2) the investigation of both hematopoietic, as well as microenvironmental (e.g., osteolineage) cells from these locations, 3) extensive bioinformatic analyses of the transcriptomic and epigenetic data obtained, and 4) the inclusion of functional assays of disomic and trisomic 'HSCs' to assess their lineage bias. These latter studies confirmed the erythroid lineage bias present in T21 HSPCs that was inferred from the bioinformatic analysis, 5) a focus on the microenvironmental changes induced by Ts21 and a confirmation that ROS is increased in hematopoietic cells within the fetal liver but not bone marrow, and 6) evidence by gene expression of altered osteolineage differentiation, which is an intriguing finding.

I agree with the authors that 'disomic' is a more appropriate term than 'healthy/normal'.

Based on the response to reviewers, a good faith effort was made to analyze for the presence of GATA1s mutation. These data did not appear to make it into the paper. The shifting emphasis in the paper away from TAM/ML-DS to the erythroid perturbations found in infants with DS and a potential role of the fetal liver microenvironment in contributing to the leukemia risk makes sense given the data presented.
Some minor issues still remain.

1.Line 90: CD45 is still described as a 'pan-haematopoietic' marker. I would again argue that since it does not recognize erythroid cells, it is far from pan-haematopoietic.

2.Lines 112-115: Since this sentence describes newly added and 'important' information, the results should be better described. E.g., what experimental approach was used? The results should be more detailed than 'widely perturbed'.

3.Fig.1E. The results are more clearly evident. Are hepatocytes not evenly distributed in the liver, but rather are concentrated on the periphery?

4. Lines 224, 311 (and other places): "Supplementary Results" are generically referred to. Which Supplemental Results?

5. What are the axes in Fig. 3c and 3d?

Referee #4 (Remarks to the Author):

The authors have addressed several of the prior reviewer questions with new bioinformatics analyses that clarify the strengths and limitations of the data, including orthogonal single-cell population-level integration/supervised analyses and reporting of cell barcode to cluster associations. However, some important prior issues raised were not adequately addressed, which are likely to have profound impacts on the rigor of the results (gene expression, cell-type deconvolution, placeholder code and data availability). While the authors perform new experiments, the results appear to be somewhat contradictory, with claims regarding links to DS leukemia mutations which remain only hypotheses at this point.

The authors' comment that ambient RNA differences should not vary between samples and should not impact the results. While I agree that ambient RNA fractional abundances typically do not vary much between samples, the specific and important concern is that when there are large shifts in the cell lineages present in a sample (as there are between Ts21 and disomic samples), the biased lineages will result in an increased abundance of non-native transcripts in other cell populations (e.g., erythroid progenitor transcripts in HSC). The authors cite Amezquita et al. 2023 as evidence that ambient RNAs will not impact analysis results, however, I see no such evidence in the cited paper, and rather a large body of literature to support that the alternative conclusion: (https://bioconductor.org/books/3.13/OSCA.multisample/ambient-problems.html, PMC10366499, PMC7763177). Simply, ambient RNAs from cell populations that are disproportionate in one sample but not others, will result in aberrant differential expression changes in addition to those that are biologically valid.

The request for the authors to consider gene expression differences outside the HSC were for multiple reasons. First, if as the authors claim that GATA1 mutation driven preleukemia occurs selectively in HSCs, the alternative hypothesis needs to be explored, specifically that these differences observed are unique to HSC and cannot be attributed to something like ambient RNAs or altered microenvironment. Second, if large gene expression differences are occurring in other populations, the authors purported model may be more complex. Performing differential expression analyses is easy and should be performed for other cell types to ensure comprehension.

The argument for not performing cell2location with a broader cell population reference does not appear to be valid. A common single-cell reference (versus precision) is the typical approach for cell2location analyses, in which matched single-cell references are typically not available. The claim that such analyses could lead to inaccurate results, would seem to suggest that cell2location can’t effectively predict cell-type associations with a broader reference due to an increase in false positives. If this is the case, then how can the performed analysis be considered reliable? Including additional references (some cell types won’t be effectively captured in some samples due to the enrichment strategy used), should allow the reviewers to address.

For the correlation analyses with pseudobulk fold changes, if the authors were to restrict their analyses to only differentially expressed genes (rather than all genes which will mask the main differences), excluding those on chromosome 21, do the differences between disomic and ts21 still

persist? For these analyses, ambient RNAs and Chromosome 22 genes are expected to significantly contribute to differences (e.g., erythroid progenitor bias leading to more erythroid differences in non-erythroid progenitor cells). Increased Ts21 chromosome 22 gene expression levels should persist after correction (e.g., SoupX).

Figure 1f-h trajectory analysis indicates two mesenchymal and osteoblast populations are absent in Ts21 Femur. The authors then claim that "Osteo-lineage development is altered in Ts21 bone marrow", however, are a sufficient depth of stromal cells captured to make this conclusion? It is debatable whether most stromal cell types can be obtained from CD235a depletion without any enrichment of stromal cells markers e.g., CD271. Since the bones of the extremities of fetuses with Ts21 are only slightly shorter (PMID: 2531547), it is shocking to see Ts21 samples do not contain any osteoblasts (line 181). To further evidence this claim, the authors would need to measure bone mass/femur length and examine the histology of the femur, preferably IHC staining of osteoblasts and osteoprogenitors.

While the new mitochondrial mass and oxidative stress analyses provide a link between gene expression associated differences and DS, any links between such functional associations and specific mutations not yet identified in these subjects are premature. While the authors report that Ts21 HSC/MPP exhibit increased expression of oxidative stress and mitochondrial dysfunction genes, this result appears to be is contradictory to results shown in Figure 2f-g and later on line 280-284, where higher mtROS is only observed in CD34+CD38+ more mature progenitors and mature Lin+ cells, but not in CD34+CD38- HSC/MPP (title 267 is also inaccurate). This again would typically suggest an ambient RNA artifact due to increased erythroid progenitor transcript contamination. Further, the authors' experimental results indicate that oxidative stress does not impact Ts21 stem and early progenitor cells. It is possible that GATA1 mutations occur in later committed progenitors due to oxidative stress (not measured due to CD235 depletion)?

Any relevant scripts or code should be made available for review, not after, as previously requested. Data accession numbers should be included, even if restricted now.

**Author Rebuttals to First Revision:**

We would like to thank the Reviewers for reading the revised manuscript and for their further questions and suggestions. To address these outstanding questions we performed additional functional validation, analysis, and rewrote parts of the text.

To summarise, we have performed the following experiments and analysis:

**A. Experiments**

- Performed overexpression of TFR2 in HUDEP-2 erythroid cell line and confirmed its role in erythroid cells differentiation.

**B. Analysis**

- Extended somatic mutation analysis to strengthen support that altered Ts21 regulatory landscape is linked to somatic mutation accrual.

- Performed extensive analysis of transcription factor motifs enriched in Ts21 HSC promoters and enhancers, revealing diverse regulatory changes (the largest being GATA) and differential AP-1 motif usage.

- Linked our Ts21 HSC transcriptomic and epigenomic differences to previously observed Ts21 differences in DNA methylation using a publicly available dataset from Muskens et al.

- Evaluated gene expression changes across all blood cell types

- Modified figures, such as converting violin plots to barplots with error bars (Fig 2)

- Provided direct comparison of scRNA versus multiome composition

- Reported the megakaryocytic bias observed in Ts21 HSCs

- Tested consistency of cell2location results using a publicly available scRNA-Seq dataset[1] as a reference

- Confirmed robustness of our initial findings (DE analysis and subsequent metrics), by accounting for ambient RNA, removing chr21 genes, and non-DE genes

Our revision of the manuscript confirmed and further strengthened our initial findings.

Referee #1 (Remarks to the Author):

I appreciate the authors' response to my comments and additional analyses.

Thank you.

My major comments:

1. "DS mutations located in the introns of Ts21 HSC-expressing genes are 60% more likely to reside within cis-regulatory elements compared to intronic mutations identified in disomic HSCs": I was confused with the authors' claim that altered chromatin accessibility landscape in Ts21 HSCs provides a milieu for mutation acquisition in regulatory regions of genes. The greater overlap of intronic mutations with cCREs than chance does not prove the hypothesis of causal effect of greater accessibility on mutational load. I am even not sure if this greater overlap is a specific finding for DS or a general finding for intronic somatic mutation in general.

The analysis we performed used the same cis-regulatory elements (CREs) and genes, and contrasted whether Ts21 or disomic intronic mutations were more likely to reside within the identified CREs. The result showed a greater enrichment of Ts21 mutations compared to disomic mutations within these regions. 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 original analysis of CREs within Ts21 HSC-expressed genes, by also comparing within Ts21 cycling HSC-expressed genes and disomy HSC-expressed genes (Figure 1). Using Ts21 HSC-expressed and cycling HSC-expressed genes, the Ts21 intronic mutations are significantly more likely to be found in the matched set of 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 1a-c). 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 A1a-c). Thus, the association between Ts21 mutations and CREs relative to disomic mutations indicates that the regulatory landscape impacts where somatic mutations fall.

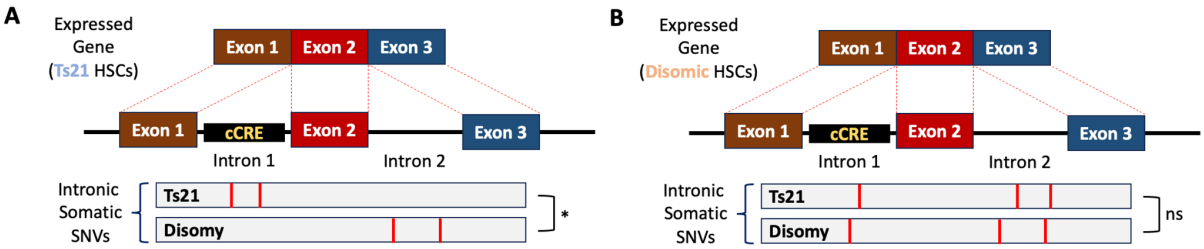



***Figure 1:*** Schematic overview of testing whether intronic somatic mutations are enriched in expressed genes of (A) Ts21 HSCs or (B) disomic HSCs.

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 consistent results 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 (Figure 2d-e). 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 harbor regulatory somatics in these same regions, which cause differential expression. However, we mention in the Discussion that 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 matched cis regulatory elements of Ts21-expressed genes but not disomy genes.

b) Matched 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.

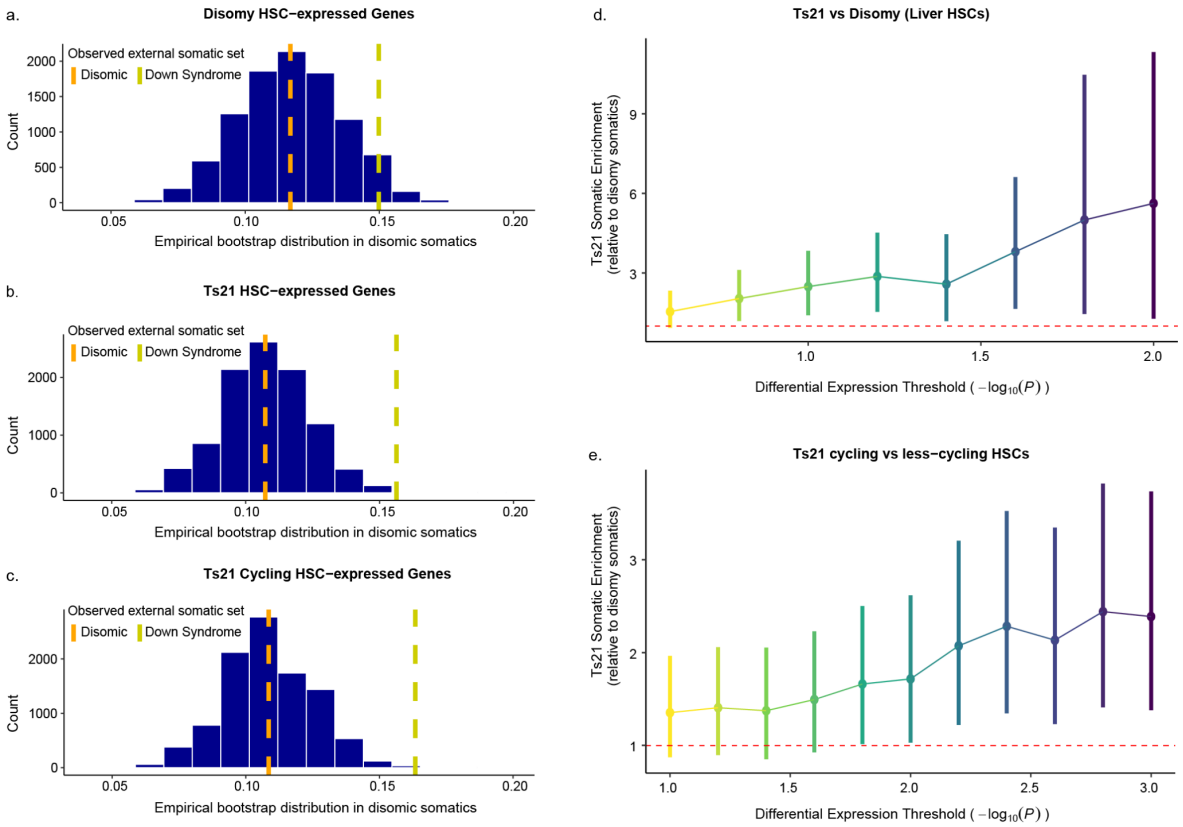



***Figure 2:*** *(a-c) Cataloged Ts21 mutations in the introns of (a) disomy HSC-expressed genes, (b) Ts21 (less-cycling) HSC-expressed genes, and (c) Ts21 Cycling HSC-expressed genes are more likely to overlap candidate cis-regulatory elements (CREs) from the ENCODE project than cataloged intronic mutations from disomic cells. The histograms represent the empirical distribution of this overlap from 10,000 bootstrapped sets of somatic variants found in disomic HSCs. The orange and yellow lines represent the observed overlap of intronic mutations in CREs for disomic and Ts21 mutations respectively. (d-e) The enrichment of Ts21 somatics relative to disomic somatics within CREs of differentially expressed genes from (d) Ts21 vs Disomy HSCs and (e) Ts21 cycling vs less-cycling HSCs, across different* P*-value thresholds.*

2. As I commented in the previous review, the genes highlighted from DEG seemed to be arbitrarily chosen, especially ROS genes (FOXO3, APOE, DUSP1) that the authors later tried to functionally validate, but not based on their significance or pathway's significance. Although authors performed experimental studies on those oxidative stress effects, the rationale for specifically choosing this pathway for validation is still unclear.

ROS was one of the most significant pathways to come up in our gene set enrichment analyses of Ts21 induced genes upregulated in Ts21 fetal liver HSCs (Figure 2c in the main manuscript). In addition, oxidative stress has been previously[2] described as a potential cause of mutagenesis in the analysis of Ts21 somatic mutations (which show an oxidative stress signature). Furthermore (which we did not originally mention in the manuscript), the oxidative stress-related gene *SOD1* is the most significantly differentially expressed gene in Ts21 HSCs (compared to Disomy). Therefore, based on our own results and previously reported results, we decided to focus on this pathway and validate it in experiments. We also validated experimentally other findings from our single-cell analyses, such as increased cycling and erythroid bias of HSCs, which were originally detected in our single-cell analyses (such as the presence of a highly-cycling HSC cluster in Figure 2d in the main manuscript and the enrichment of genes related to erythroid differentiation in the Ts21 induced gene set in Figure 2c in the main manuscript).

3. On the additional experiment on mitochondrial mass and oxidative stress, the violin plot is not appropriate since there are only three samples in each category.

Thank you for this suggestion. We have replaced violin plots with barplots (mean) with error bars (representing 95% confidence interval), Figure 3.

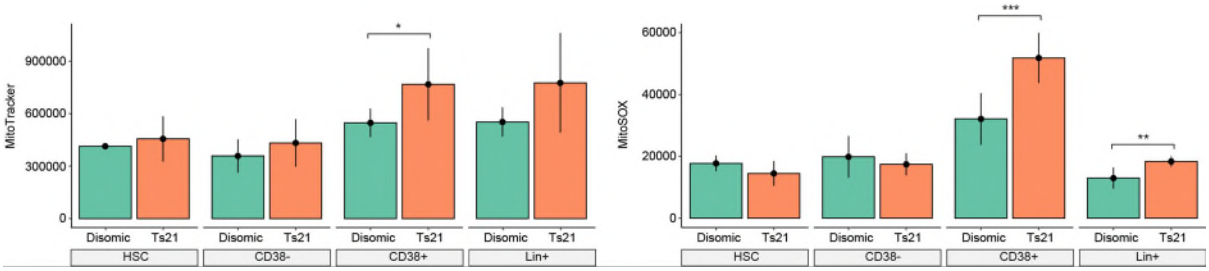



***Figure 3. Quantitative comparison of*** *MitoTracker staining (left) and MitoSOX staining (**right**) in disomic and Ts21 haematopoietic populations; HSC-haematopoietic stem cells (Lin-/CD34+/Cd38-/CD90+/CD45RA-); CD38- (Lin-/CD45+/CD34+/Cd38- population); CD38+ (Lin-/CD45+/CD34+/CD38+ population); Lin+ (CD3/8/11b/56/14/19, lineage positive population).*

4. As the reviewer #4 also pointed out, can the authors provide the direct comparison of the cell frequencies of each cell population between the snRNA-Seq and scRNA-Seq for each sample who was profiled by both technologies, given the limited replication of comparison between Ts21 and disomic?

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 fetuses with 2 replicates each, $r = 0.50$; Figure 4), but there are differences, as discussed in Massier et al. Nat Comm 2023. For example, in their study, vascular cells were found at nearly three-times higher abundance in single-nuclei data compared to single-cell data. In our study, we see 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 (Supplementary Figure 19b in the main manuscript).

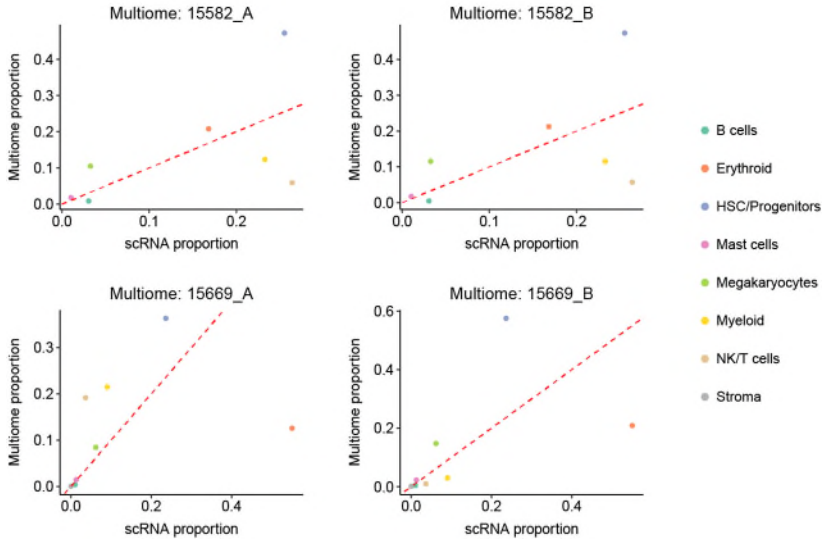



***Figure 4:*** Correlation between estimated cell type proportions within scRNA (x) and multiome samples (y) in two multiome replicates (left, right subplots) for sample 15582 (top, $r > 0.55$) and sample 15669 (bottom, $r > 0.37$). Dashed red line is $x = y$.

Reviewer #2

Marderstein et al. have addressed all comments in a clear point-by-point response. This review process strengthened the conclusions of the study.

Notably also, this manuscript is much clearer, providing insights into the procedures/analyses and a better description of the results.

Altogether, as mentioned in the revised abstract, the study now provides a molecular map of fetal haematopoiesis in Down syndrome. This is the strength of the multi-omic work: building a valuable resource to be used for future research questions in the field of leukemia development in children with Down syndrome.

Thank you.

More specifically, a substantial amount of data have been added:

- New bioinformatic analyses:

o Assess annotations by several new methods

o Integration of already published datasets for validation/comparison

o Challenging the data (role of each chr21 genes, GWAS SNP…)

- New experiment using new primary samples (n=6; 3 DS and 3 non-DS fetal livers) to mitochondrial dysfunction and confirm earlier observation about the perturbed proportion of haematopoietic progenitors in DS fetal livers (PMID: 18812473, PMID: 23045701).

While these new analyses extend our understanding of the potential intrinsic mechanisms, the study is still lacking functional validations. Independent models such as induced Pluripotent cells or animal models could have been used to assess some of the very interesting but speculative comments obtained from the multi-omic data. As an example, while it has been nicely shown that no chr21 genes, alone, can explain some of the phenotype (expression), genetic engineering approaches to knock out several ch21 genes could be performed to address this comment. This has been done in the past to demonstrate the role of several ch21 genes in leukaemia development in iPS cells (PMID: 27134169).

We have spent considerable time thinking how to address this question. The main issues that we found are the following - in the aforementioned paper (PMID: 27134169), the authors deleted a 4-Mb region from a single copy of chr 21, resulting in reversion to disomy in this deleted region alone. The deleted region contained 20 protein-coding and several non-coding RNA genes. In this case the deleted region was chosen based on the higher than expected expression of these genes and importantly these genes made a cluster. Therefore, deletion of the single 4-Mb region allowed the authors to knock out multiple genes from one of the chr 21 copies. In the subsequent analysis, the authors performed KO of RUNX1 or their cofactors ERG2 and ETS2 alone or in combination of two. RUNX1 is a well known regulator of haematopoiesis which possibly influenced the choice of this particular gene for KO out of 20+ in the 4-Mb region.

In our study we identified thousands of misregulated genes across cellular contexts. These genes collectively contribute to the phenotype and are dispersed across chr 21 as well as other chromosomes. Therefore, to rescue the Ts21 phenotype, we would need to knock out many individual genes on a single copy of chr 21 or other chromosomes that are not clustered together. This is very difficult to achieve and is further complicated by the fact that we do not have a specific set of genes for which we would expect, if knocked-out in combination, to rescue the Ts21 haematopoietic phenotype. Our analysis (for further evidence please see our comment below) suggests that complex regulatory mechanisms are at play. Dissecting on a more mechanistic level which of the dysregulated genes and their combinations contribute to the different aspects of the complex phenotype in DS would not be feasible for us as a part of this revision and possibly would be more appropriate as a part of an independent study like the one cited by the Reviewer. For those reasons we have not been able to provide any further results related to this question.

Notably, a novel ch21 player has been recently described: RUNX1 isoforms, with RUNX1 transcription factor being a well-known master regulator of early hematopoiesis (PMID: 36493345). Has this been assessed? Have RUNX1 binding sites been identified in the enhancer/promoter regions. Any other ch21 TF? ERG? ETS2?

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 (Machlab et al. 2022 Bioinformatics). 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, utilizing the JASPAR2022 database (Castro-Mondragon et al 2022) 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.

Overall, 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. 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[3]. In contrast, TFs such as BACH and NFE2, as well as the AP1 complex, exhibited increased activity in disomic peaks (Figure 5).

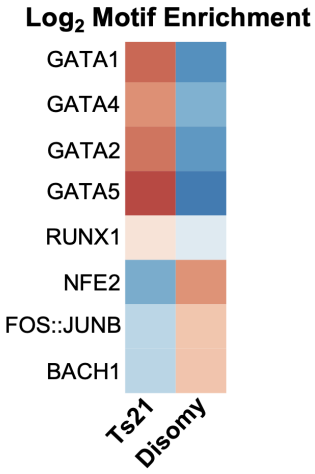



***Figure 5:*** *Enrichment of selected motifs by monaLisa in peaks with greater accessibility in Ts21 (left) and reduced accessibility in disomy (right).*

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 utilized the Eukaryotic Promoter Database (EPD), 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 (Figure 6a). 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 (Figure 6b), validating the utility of EPD promoters for TF motif enrichment analysis.

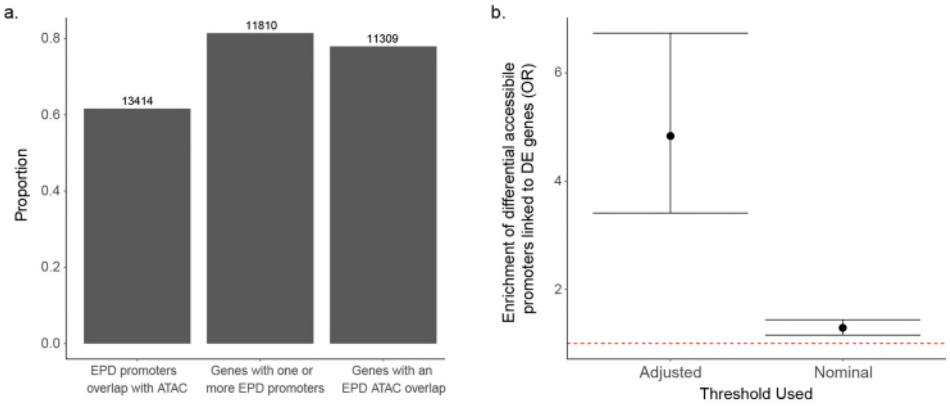



***Figure 6: (****a) Barplots describing the number and proportion of (i) EPD promoters that overlap with peaks accessible in HSCs, (ii) expressed genes with at least one EPD promoter, and (iii) expressed genes with at least one EPD promoter that overlaps with peaks accessible in HSCs. (b) The association between differentially accessible EPD promoters and differentially expressed (DE) linked genes (odds ratios with standard errors). Left used DE genes at FDR < 0.05 (adjusted) and right used DE genes at nominal P < 0.05.*

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 15 in the main manuscript, Figure 7a). 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 signaling pathways (Figure 7c). Notably, IFN has been shown to stimulate the exit of HSCs from quiescence and their activation in the cell cycle *in vivo*.

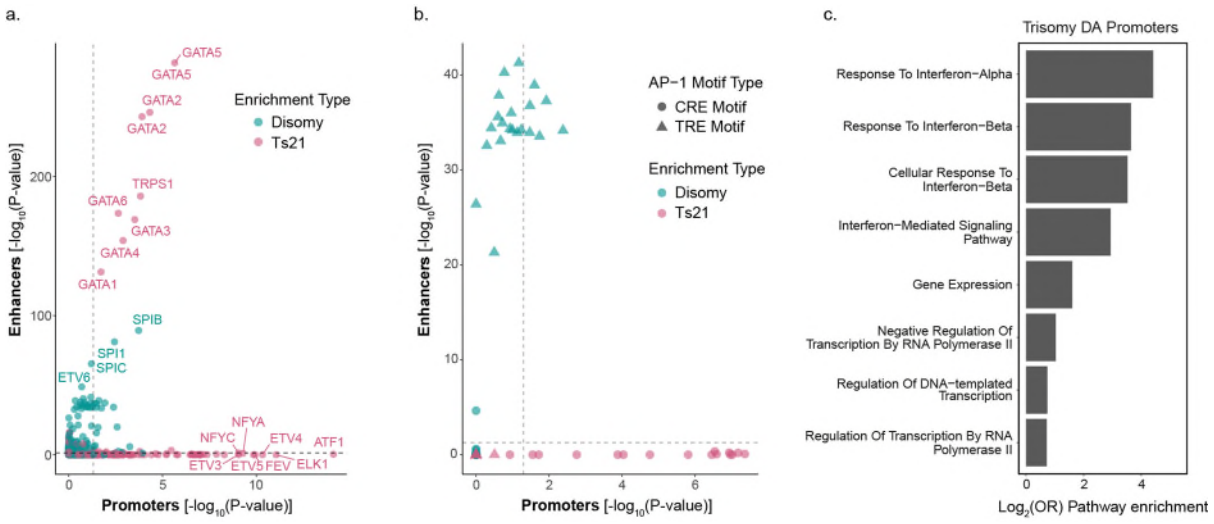



***Figure 7. (a)*** *All motif enrichments and (b) AP-1 motif related enrichments in promoters (x-axis) and enhancers (y-axis) for elements with greater accessibility in Ts21 (red) and disomic (blue) HSCs. (c) Gene set enrichment analysis of genes linked to differentially accessible (DA) promoters with greater accessibility in Ts21.*

Additionally, we found varied utilization of AP1 motifs among DA promoters (Figure 7b). 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*[4], 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 7b). 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 (Figure 8a). 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 (Figure 8b).

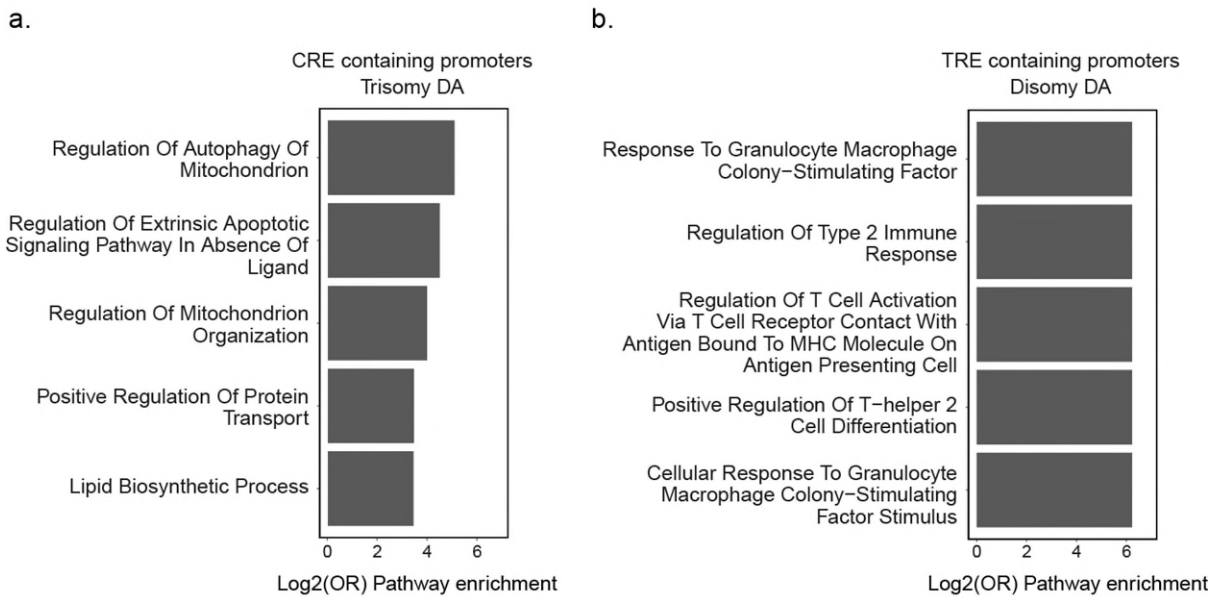



***Figure 8.*** *Gene set enrichment analysis results for (a) genes linked to CRE-containing promoters with greater accessibility in trisomy and (b) genes linked to TRE-containing promoters with greater accessibility in disomy.*

2) To examine TF motif enrichment in enhancers in Ts21 and disomic HSCs, we used the activity-by-contact (ABC) model[5] 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 (Figure 9a). Specifically, around 1700 genes had at least one differentially accessible (DA) enhancer and around 500 genes had more than one DA enhancer (Figure 9b). 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 (Figure 9c):

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.

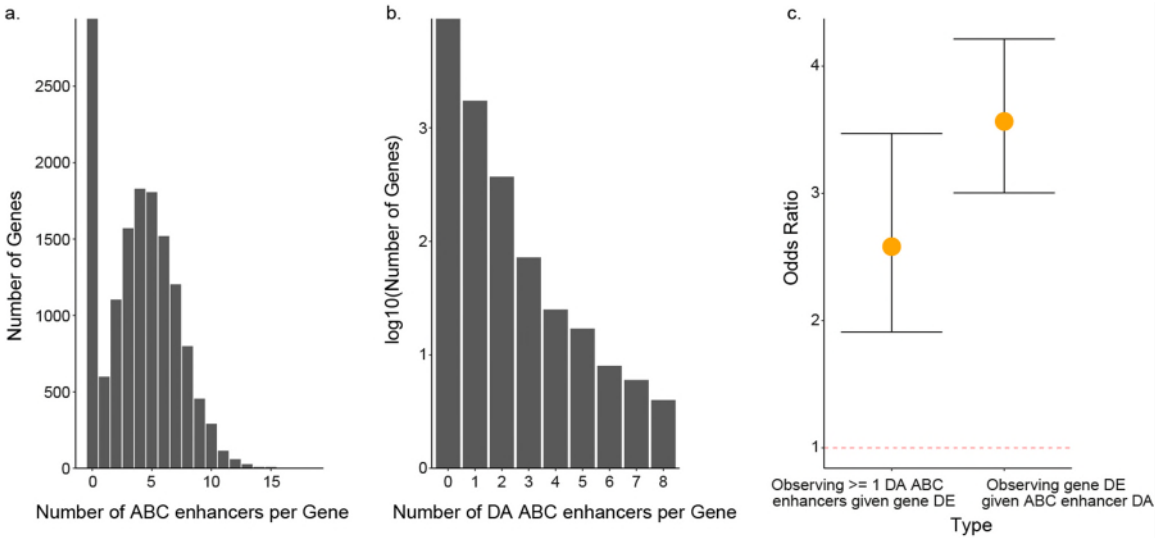



***Figure 9:*** *Histogram of the number of ABC enhancers per gene (a), differentially accessibility ABC enhancers per gene (b), and association between differentially accessible ABC enhancers and differential expression of their target genes (c).*

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 (Figure 10), particularly at enhancers with greater accessibility in trisomy.

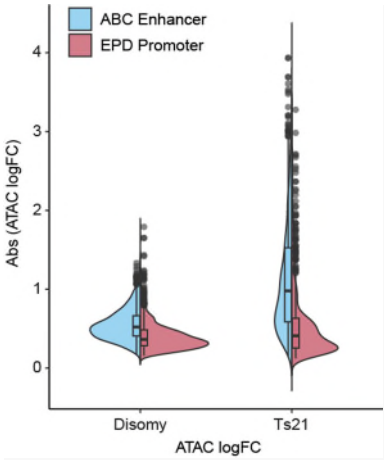



***Figure 10****. The absolute value of log-fold changes for differentially accessible peaks, stratified by peaks with significantly higher accessibility in disomy (left) and Ts21 (right), and ABC enhancers (blue) or EPD promoters (red).*

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 7a). 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 7b).

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 (Figure 11).

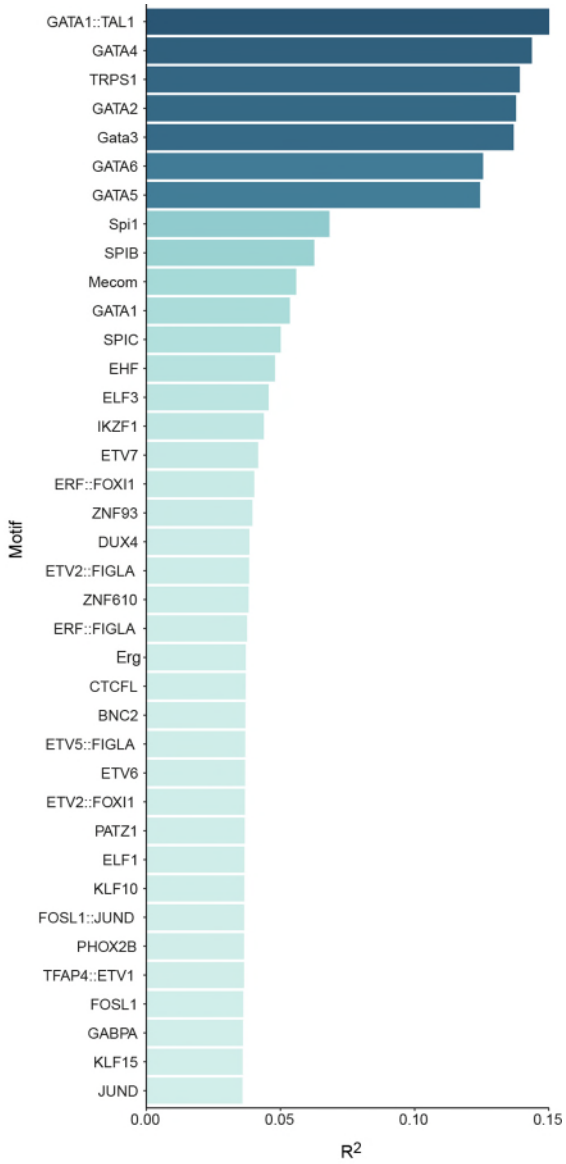



***Figure 11.*** *Proportion of variance explained in the differential accessibility of trisomy 21 HSCs for each TF motif instance.*

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) (Figure 12). 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.

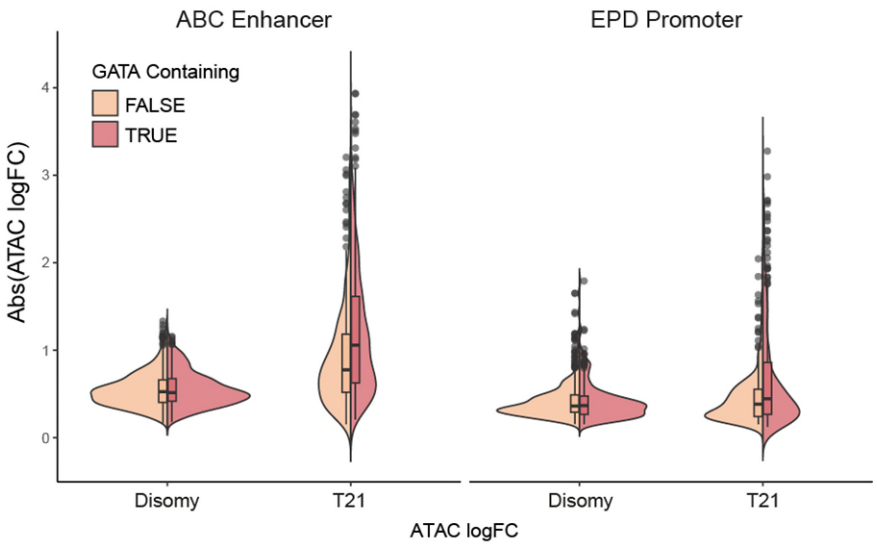



***Figure 12.*** *The differential accessibility, estimated by |logFC| between Ts21 and disomic HSCs, between different peak sets and stratified by GATA-containing elements or not.*

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 fetal 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 utilization 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, characterized by altered TF motif enrichment and chromatin accessibility patterns, in promoters and enhancers, which may contribute to the haematopoietic differences observed in Down Syndrome.

The authors should mention/discuss the known altered pattern of DNA methylation linked to trisomy 21 and how could this be correlated with their expression data, genomic accessibility, erythroid-primed HSC and progenitors and the observed hematopoietic phenotypes.

We would like to thank the reviewer for this interesting suggestion. To address the reviewer comment we used publicly available data from the study titled "The genome-wide impact of trisomy 21 on DNA methylation and its implications for hematopoiesis". DNA methylation analysis (by Muskens et al, 2021, https://doi.org/10.1038/s41467-021-21064-z) of 196 DS and 439 non-DS newborn blood samples identified 1052 differentially methylated regions (DMRs) associated with DS. 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. Despite differences in the cell type tested in Muskens et al. 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. 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. analysis). Furthermore, the sign concordance of log-fold changes were quite strong (OR = 9.6; Fisher's exact test *P*-value = $3.8 \times 10^{-31}$), such that genes upregulated in Ts21 CD34+ in Muskens et al. were typically also upregulated in Ts21 liver HSCs in our dataset (Figure 13).

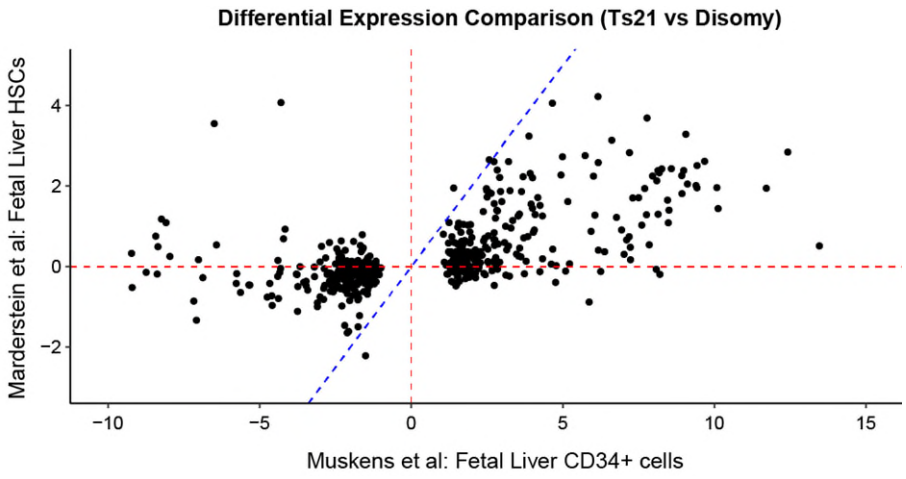



***Figure 13:*** *Comparison of differential expression results between Muskens et al. and our study (Marderstein et al.). Differential expression results are from Supplementary Data 9 by Muskens et al., which include genes significantly differentially expressed between Ts21 foetal liver versus disomy 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.*

Muskens et al. 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 (Figure 14a; Fisher's exact test *P*-value = $1.4 \times 10^{-6}$). 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 = $1.3 \times 10^{-3}$) (Figure 14b). Similar to Muskens et al., we did not observe a statistically significant association between DMRs and expression outside of annotated promoter/enhancer regions.

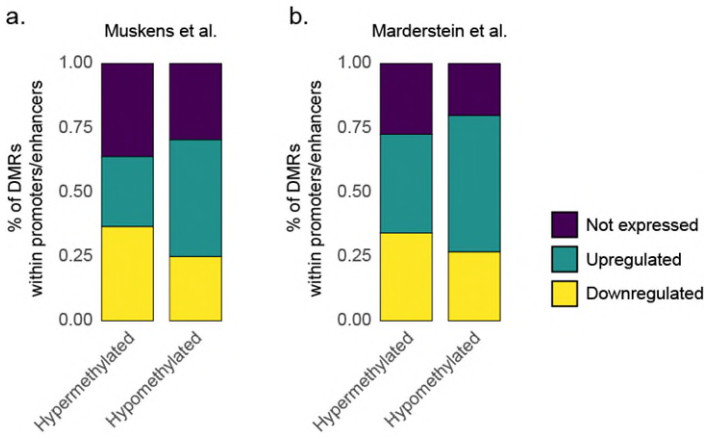



***Figure 14:*** *Stacked barplot showing the association between (i) hypo- and hypermethylation at DMRs within gene enhancers and promoters, and (ii) differential expression of the associated gene, for both (a) Muskens et al. and (b) Marderstein et al.*

The multiome data from our study enabled us to extend Muskens et al. 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 fetal liver HSCs. We observed a strong trend between increased chromatin accessibility and hypomethylation in both promoter/enhancer regions (Fisher's exact test $P = 5.4 \times 10^{-27}$) and outside promoter/enhancer regions ($P = 6.1 \times 10^{-6}$) (where promoter/enhancer regions were annotated according to Muskens et al.). This demonstrates the value of our resource in exploring the link between altered DNA methylation and accessibility in Ts21, particularly throughout the genome (Figure 15).

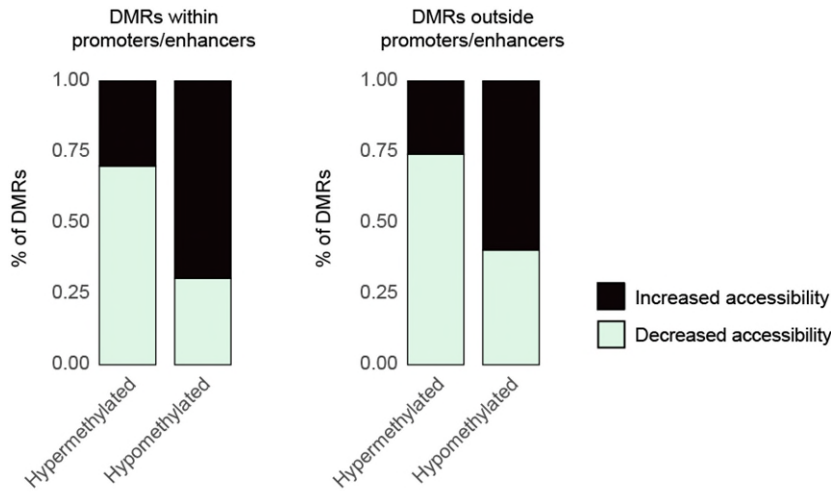



***Figure 15:*** *Stacked barplot showing the association between (i) hypo- and hypermethylation at DMRs within gene enhancers and promoters in CD34+ cells from Muskens et al., and (ii) differential accessibility of the region in liver HSCs from Marderstein et al. Left shows DMRs within promoter or enhancer regions, while right shows DMRs outside promoter or enhancer regions (annotated as in the gene body or intergenic).*

Moreover, very nice observation highlights the potential impact of the overexpression of TRF2 and TSPAN32 in the erythroid-primed phenotype of the HSC. That could be reproduced and tested in an in vitro setting.

Thank you. To address the reviewer's comment, we sought to investigate the effect of overexpressing *TFR2* in disomic foetal liver HSCs samples. We isolated LT-HSCs from foetal livers and overexpressed TFR2 using lentiviral transduction. The efficacy of the lentiviral particles was tested on the producer cells (HEK293T), which increased surface expression of TFR2 around 80% after transduction. We repeated the overexpression experiment in the LT-HSCs with three different donors using varying concentrations of lentiviruses and transduction enhancers, but could not achieve sufficient overexpression of *TFR2* in the foetal LT-HSCs. As our source of primary foetal cells is limited, we continued our experiments with an immortalized human erythroid progenitor cell line, HUDEP-2[6].

HUDEP-2 cells are derived from CD34+ umbilical cord blood cells and express erythroid-specific markers such as CD235, CD71 and CD36. When cultured in *expansion* media HUDEP-2 remain in undifferentiated state but when cultured in *differentiation* media they start to differentiate and upregulate CD235, a marker of late erythroid cells[7]. To validate our in-vitro model, we first measured the expression level of CD235a in HUDEP-2 cells in the *expansion* media (undifferentiated state) and then four days after culturing in the *differentiation* media. The median fluorescence intensity (MFI) of CD235a increased 26 fold following differentiation (Fig 16b). This is similar to the expression levels seen in differentiated erythroid cells from fresh CD34+ cord blood cells[8]. Similarly, to test the

expression of TFR2 during differentiation, we compared the surface expression of TFR2 in HUDEP-2 cells in the undifferentiated state with TFR2 expression after differentiation. Our analysis showed that TFR2 expression increased 2.5 fold in differentiated compared to undifferentiated HUDEP-2 (Fig 16b), 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 *expansion* media for three days. Subsequently, we assessed TFR2 expression using flow cytometry. We observed around a 3-fold increase in TFR2 expression (Fig 16, c). This increased expression of TFR2 also led to 2.1 fold increase of CD235a, a 2.2 fold increase in CD71 and a 3.5 fold increase in CD36 expression in the HUDEP-2 cells expressing the highest levels of TFR2 (Fig 16, d). Taken together, our data suggest that overexpression of TFR2 can induce differentiation of erythroid progenitors.

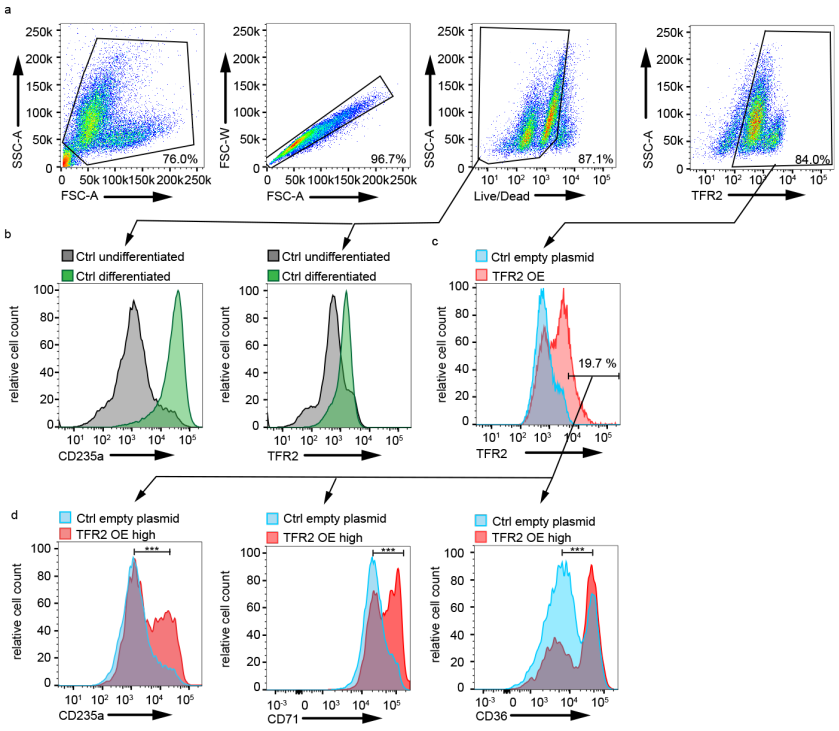



***Figure 16: Gating strategy and surface expression levels of erythroid markers in HUDEP-2 cells****. (a) Gating strategy used to determine surface expression of TFR2, CD235a, CD71 and CD36. (b) Histograms showing CD235a expression (left) and TFR2 expression (right) in HUDEP-2 cells in undifferentiated 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) CD235a, CD71 and CD36 expression in Ctrl HUDEP-2 (blue) and in the population of highly expressing TFR2 cells (dark red).*

As reported here (Figure 4i) and other, there is a bias toward erythroid and megakaryocytic colonies. While the erythroid primed HSC are well characterized in the study, it is less clear for the

megakaryocytic lineage?For example, have the Authors checked the HSC differentiating toward megakaryocytes like in figure.3b.

We would like to thank the reviewer for drawing our attention to this question. We have now checked the probability of HSCs to differentiate towards megakaryocytic lineage and we do see bias of Ts21 HSCs to differentiate towards megakaryocytes (Figure 17). We have now included this result in the main manuscript (Figure 3b in the main manuscript).

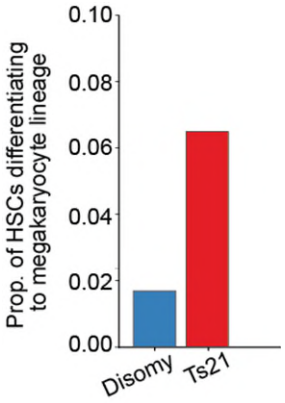



***Figure 17.*** Proportion of HSCs differentiating toward the erythroid lineage indicated by an absorption probability > 0.5.

Also, the megakaryocytic bias is not particularly seen in figure.1d. Please comment.

Following analysis suggested by Reviewer 4 we examined how choice of scRNA-Seq reference affects model training and prediction of the cell abundances on Visium spots. Our results suggest the estimation of the cell abundance is affected by the choice of reference (for details please see analysis we did for reviewer 4). In our initial analysis we used matched scRNA-Seq and Visium tissues (i.e. cells and tissues were from the same foetus). However, when we repeated the same analysis using previously published scRNA-Seq dataset as a reference to estimate cell abundances on our Visium sections some of the results were not reproducible. While we still think that the use of matched samples should be the preferred way of performing this type of analysis we have now removed these results from the manuscript.

Finally, because the sample used here allow to better dissect the impact of the liver environment in comparison of previous studies (Jardine et al, 2021), validation of the communications between

hematopoietic stem cell and progenitors and the surrounding populations (osteo…) should be validated in independent experiments to better understand the biology and the impact of these cellular perturbation in predisposing leukemia development.

We appreciate reviewers suggestions and agree that our dataset allows more comprehensive analysis of cell-cell interactions compared to published studies/datasets. There are several ways to achieve it that were not tractable in this study. One line of investigation could include a series of in vitro experiments in which foetal HSCs were co-cultured with different populations of niche cells present in the foetal bone marrow. The HSCs proliferation and lineage output could be used as a read-out to assess the relevance of different niche components. This would be then further assessed in the absence of specific ligands and receptors through CRISPR perturbation of disomic and Ts21 HSCs and niche components. Such experiments would be a substantial undertaking and not readily feasible. The disomic and in particular Ts21 foetal material is not readily available and is limited in quantity hindering our ability to promptly acquire samples and perform such experiments.

Minor comments:

- Figure.1e: increase size of the text for the pathways negatively and positively correlated.

This is corrected now.

- Same comment for Figure.2c.

This is corrected now.

- Figure.2f/g: not sure violin plots are the best way to represent only 3 replicates. Bar graph with average and error bars would be more appropriate.

We have replaced violin plots with barplots (mean) with error bars (representing 95% confidence interval), Figure 18.

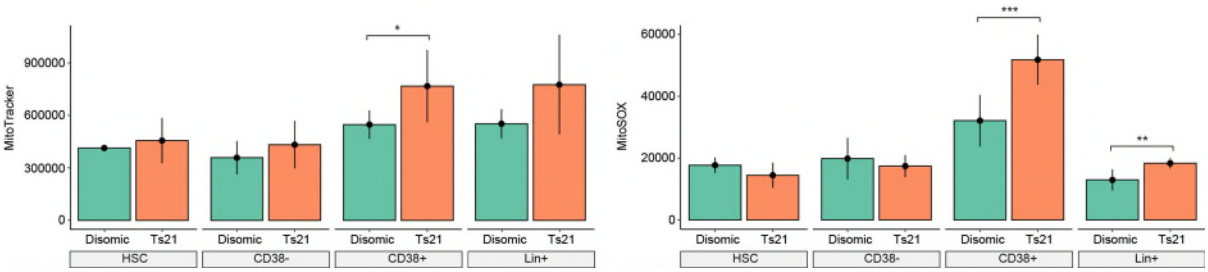



***Figure 18.*** *Quantitative comparison of MitoTracker staining (left) and MitoSOX staining (right) in disomic and Ts21 haematopoietic populations; HSC-haematopoietic stem cells (Lin-/CD34+/Cd38-/CD90+/CD45RA-); CD38- (Lin-/CD45+/CD34+/Cd38- population); CD38+ (Lin-/CD45+/CD34+/CD38+ population); Lin+ (CD3/8/11b/56/14/19, lineage positive population).*

- Please clarify the calculation of the peaks that are disomic and Ts21 specific. 191 – 60 = 122. But it is mentioned that `62.1% (72/116) are specific to disomic HSC`. 116? or 122? Also, please clarify where the 839 Ts21 specific peak genes come from?

We thank the reviewer for noticing this discrepancy. There were 122 disomic-specific peak-gene links, and 76 had a significant interaction. Additionally, we incorrectly stated there were 839 Ts21-specific links because we ignored peaks that were not accessible in disomy. Considering the full set, there were 1495 Ts21-specific links, of which 59 had a significant interaction. This has all been corrected in the text and we apologize for the confusion.

Referee #3 (Remarks to the Author):

The authors have responded in detail to the critiques of the reviewers', providing a 69-page response.

Some new data have been added that strengthen the manuscript. Importantly, much of the text has been now more clearly written with the addition of introductory and conclusion sentences. This has markedly improved the manuscript. However, the middle paragraph of the section involving the interactions between enhancers and the target genes (lines 386-402) continues to remain quite inscrutable to me.

Thank you for the supportive comments. We appreciate the reviewer mentioning the lack of clarity regarding the analysis of multiome data for discovering interactions between regulatory elements and their target genes. We have taken this into account while revising this section in an effort to improve its clarity.

Strengths of the paper continue to be 1) the analysis of a large number of primary human fetal cells taken from liver and marrow of disomic and trisomic fetuses, 2) the investigation of both hematopoietic, as well as microenvironmental (e.g., osteolineage) cells from these locations, 3) extensive bioinformatic analyses of the transcriptomic and epigenetic data obtained, and 4) the inclusion of functional assays of disomic and trisomic 'HSCs' to assess their lineage bias. These latter studies confirmed the erythroid lineage bias present in T21 HSPCs that was inferred from the bioinformatic analysis, 5) a focus on the microenvironmental changes induced by Ts21 and a confirmation that ROS is increased in hematopoietic cells within the fetal liver but not bone marrow, and 6) evidence by gene expression of altered osteolineage differentiation, which is an intriguing finding.

I agree with the authors that 'disomic' is a more appropriate term than 'healthy/normal'.

Based on the response to reviewers, a good faith effort was made to analyze for the presence of GATA1s mutation. These data did not appear to make it into the paper. The shifting emphasis in the paper away from TAM/ML-DS to the erythroid perturbations found in infants with DS and a potential role of the fetal liver microenvironment in contributing to the leukemia risk makes sense given the data presented.

Some minor issues still remain.

1.Line 90: CD45 is still described as a ‘pan-haematopoietic’ marker. I would again argue that since it does not recognize erythroid cells, it is far from pan-haematopoietic.

We have now corrected that.

2.Lines 112-115: Since this sentence describes newly added and ‘important’ information, the results should be better described. E.g., what experimental approach was used? The results should be more detailed than ‘widely perturbed’.

We have now added additional information.

The text in the manuscript reads:

“To further confirm that the Ts21 causes wide perturbation in foetal haematopoiesis we compared Ts21 with the disomic immunophenotypic stem and progenitor compartment using flow cytometry. We validated that the Ts21 causes increase in the HSCs and MEMPs and decrease in the B cell populations in the foetal liver using flow cytometry (Supp. Figure 6), which underscores the robustness of our computational classification of cell types in scRNA-seq.”

3.Fig.1E. The results are more clearly evident. Are hepatocytes not evenly distributed in the liver, but rather are concentrated on the periphery?

Following analysis suggested by Review 4 we examined how choice of scRNA-Seq reference affects model training and prediction of the cell abundances on Visium spots. Our results suggest the estimation of the cell abundance is affected by the choice of reference (for details please see analysis we did for reviewer 4). In our initial analysis we used matched scRNA-Seq and Visium tissues (i.e. cells and tissues were from the same foetus). However, when we repeated the same analysis using previously published scRNA-Seq dataset as a reference to estimate cell abundances on our Visium sections some of the results were not reproducible. While we still think that the use of matched samples should be the preferred way of performing this type of analysis we have now removed these results from the manuscript.

4. Lines 224, 311 (and other places): “Supplementary Results” are generically referred to. Which Supplemental Results?

We have now added additional information.

5. What are the axes in Fig. 3c and 3d?

It is UMAP axes 1 and 2 (e.g. UMAP-1 and UMAP-2). We have now added axes on the plots.

Reviewer #4

The authors have addressed several of the prior reviewer questions with new bioinformatics analyses that clarify the strengths and limitations of the data, including orthogonal single-cell population-level integration/supervised analyses and reporting of cell barcode to cluster associations. However, some important prior issues raised were not adequately addressed, which are likely to have profound impacts on the rigor of the results (gene expression, cell-type deconvolution, placeholder code and data availability). While the authors perform new experiments, the results appear to be somewhat contradictory, with claims regarding links to DS leukemia mutations which remain only hypotheses at this point.

The authors' comment that ambient RNA differences should not vary between samples and should not impact the results. While I agree that ambient RNA fractional abundances typically do not vary much between samples, the specific and important concern is that when there are large shifts in the cell lineages present in a sample (as there are between Ts21 and disomic samples), the biased lineages will result in an increased abundance of non-native transcripts in other cell populations (e.g., erythroid progenitor transcripts in HSC). The authors cite Amezquita et al. 2023 as evidence that ambient RNAs will not impact analysis results, however, I see no such evidence in the cited paper, and rather a large body of literature to support that the alternative conclusion:

(https://bioconductor.org/books/3.13/OSCA.multisample/ambient-problems.html, PMC10366499, PMC7763177). Simply, ambient RNAs from cell populations that are disproportionate in one sample but not others, will result in aberrant differential expression changes in addition to those that are biologically valid.

Thank you. As we observed differentiation bias in our Ts21 samples, we tested the robustness of our differential expression results that we highlighted in Fig 2b-c in the main manuscript 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. 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.

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.

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 2c in the main manuscript (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).

The request for the authors to consider gene expression differences outside the HSC were for multiple reasons. First, if as the authors claim that GATA1 mutation driven preleukemia occurs selectively in HSCs, the alternative hypothesis needs to be explored, specifically that these differences observed are unique to HSC and cannot be attributed to something like ambient RNAs or altered microenvironment. Second, if large gene expression differences are occurring in other populations, the authors purported model may be more complex. Performing differential expression analyses is easy and should be performed for other cell types to ensure comprehension.

We would like to emphasise that the claim that GATA1 mutation driven preleukaemia occurs in LT-HSCs does not come from our study, but rather was convincingly demonstrated in a study led by John Dick's and Eric Lechman's groups (Science, 2021, DOI: 10.1126/science.abf6202). They were able to show that progeny downstream of LT-HSCs are not able to initiate preleukemic transformation, as no consistent human CD45+ engraftment was detected after 12 weeks in mice transplanted with control or GATA1s cells from either control or Ts21 foetal liver progenitors. This is consistent with the limited

self-renewal and repopulation potential of progenitors but also highlights the inability of Ts21 foetal liver GATA1s progenitor cells compared with LT-HSCs to initiate preleukemia.

Having said this, we agree with the reviewer that we should demonstrate that any differences between disomic and Ts21 populations reported in our study are not driven by ambient RNA. Therefore, we performed ambient RNA removal using CellBender. As described in detail in the previous response, we found that ambient RNA does not explain the dysregulated pathways that we observed in Ts21.

Looking at the differences in gene expression in the progenitor population is an interesting question as it will inform on the effect of the trisomy on different populations and might highlight those that are more affected. This analysis however will not change the fact that progenitor populations have a limited self-renewal potential and are not able to initiate pre-leukaemia. Thus, the results of this analysis will provide further comprehension of the effect of trisomy on blood production, as the reviewer also pointed out, but will not necessarily change the model of the disease progression. In other words, big changes in gene expression in progenitor populations will have immediate effect on the foetal blood production and the functional properties of generated mature cells, but acquisition of GATA1 mutation in progenitor populations will not have long term consequences such as preleukaemia development.

Using our scRNA-seq data, we further looked into Ts21 expression differences across cell types. We found that HSC differential expression (between Ts21 and disomy) was an effective predictor of differential expression in multiple other cell types, as the log-fold changes of differential expression strongly correlated at non-chr21 genes across most blood types (Figure 19). 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.

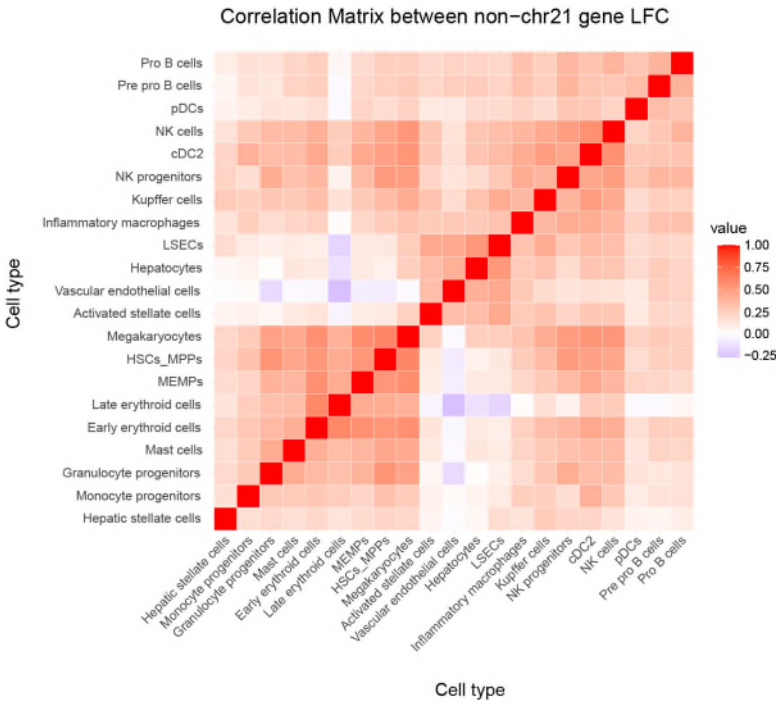



***Figure 19:*** *Correlation matrix between log-fold changes across non-chromosome 21 genes, with cell types ordered after performing hierarchical clustering.*

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 (Figure 20), such that a gene set detected in HSCs was detected in a median of 3 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 (Waugh et al. Nat Genet 2023; Malle et al. Nature 2023). For example, HSCs, mast cells, pDCs, pre pro B cells, and pro B cells upregulate genes involved in the pattern recognition receptor signaling 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 organization, and epithelial cell apoptosis, which can be interconnected processes contributing to tissue homeostasis.

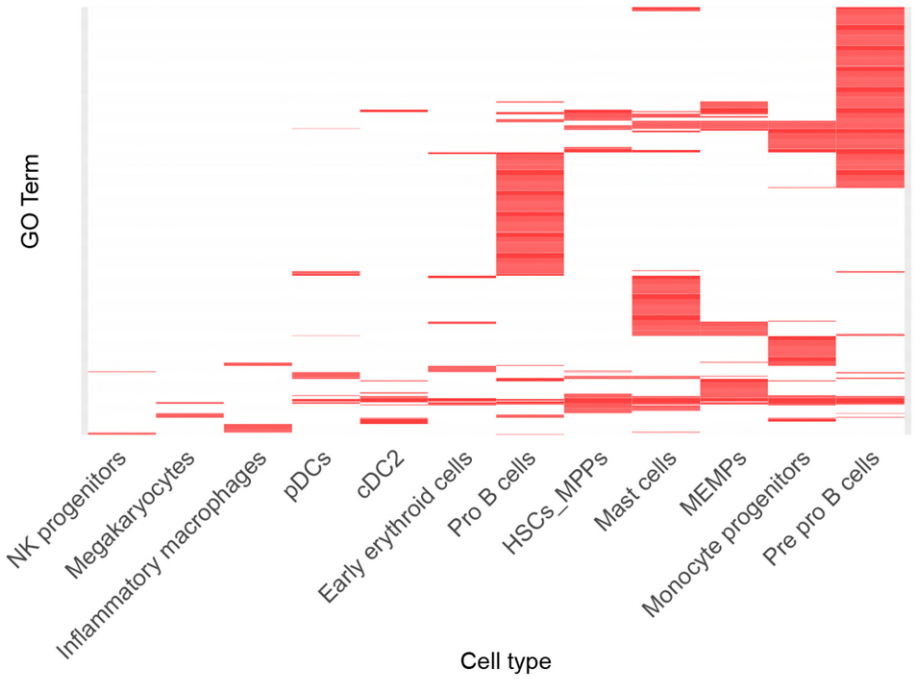



***Figure 20:*** Presence of enriched GO terms (*FDR* < 0.1) across 12 blood cell types based on genes significantly upregulated in Ts21 fetal liver.

We also noted that, while expression differences may be overall similar across cell types, the magnitude of expression differences can greatly 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 ($P < 4 \times 10\text{-}6$). Since we found that RUNX1 motifs were enriched in HSC peaks that were more accessible in Ts21 (Figure 5), 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 defense 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).

We report our Ts21-upregulated and Ts21-downregulated gene set enrichment results in Supplementary Table 11, which allows readers of our manuscript to explore pathway-based expression results in greater detail.

The argument for not performing cell2location with a broader cell population reference does not appear to be valid. A common single-cell reference (versus precision) is the typical approach for cell2location analyses, in which matched single-cell references are typically not available. The claim that such analyses could lead to inaccurate results, would seem to suggest that cell2location can't effectively predict cell-type associations with a broader reference due to an increase in false positives. If this is the case, then how can the performed analysis be considered reliable? Including additional references (some cell types won't be effectively captured in some samples due to the enrichment strategy used), should allow the reviewers to address.

Here the reviewer suggested that a common single-cell reference is a typical approach in cell2location analysis. However, to the best of our knowledge, it is quite common for published Visium data to be matched with scRNA-Seq. As an example, please see Li et al. (https://doi.org/10.1038/s41592-022-01480-9), where 45 paired datasets (containing both spatial transcriptomics data and scRNA-seq data) were analysed.

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. Across all spots and all sections, we calculated the correlation between the abundance estimates for any pair of two cell types.

1. 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. (Figure 21). 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.

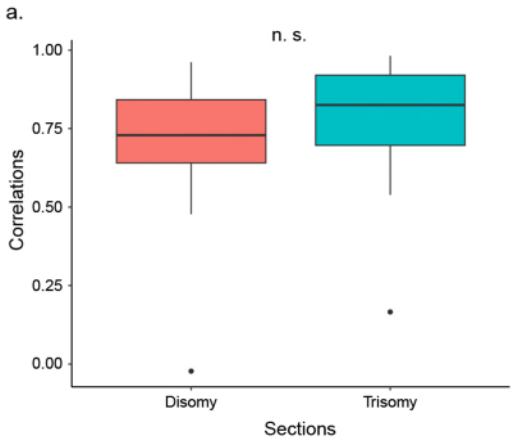



***Figure 21.*** *Cell types were subset to those present in both disomy and Ts21 references, and the correlation between abundances for cell types of the same type were computed (using both disomy- and trisomy-based cell2location deconvolutions). Each point in the boxplots displays the correlations. There is no significant difference between the two sections.*

2. As a comparison, we partitioned our disomy dataset into two subsets with comparable cell number (60,327 vs 50,344). We 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 (Figure 22). We found that the two estimated abundances exhibited significantly 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.

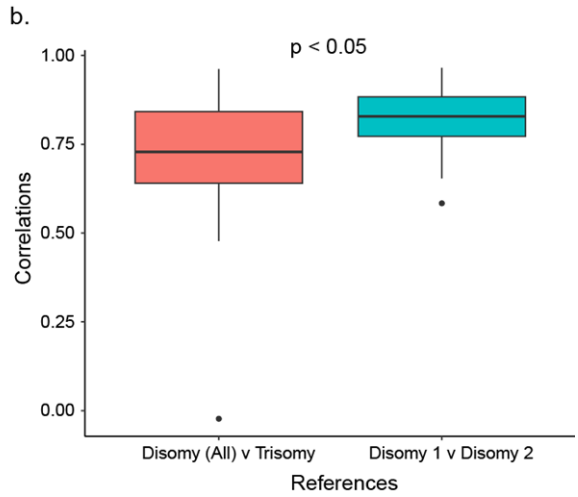



***Figure 22.*** *The correlation between abundances for cell types of the same type are shown across disomy sections. Cell2location was performed using two disomy subset references (Disomy 1 and*

*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. There is greater concordance between the two disomy references, compared to disomy versus trisomy references.*

3. We repeated the analysis in 2, but using the Ts21 dataset (Figure 23). 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.

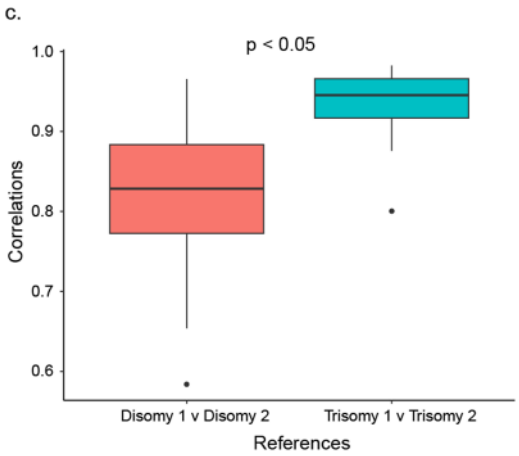



***Figure 23**. The correlation between abundances for cell types of the same type are shown and compared between disomy and trisomy. Cell2location was performed using two disomy subset references (Disomy 1 and Disomy 2) on disomy samples or two trisomy subset references (Trisomy 1 and 2) on trisomy samples. Each point in the boxplots displays a different cell type. There is greater concordance between the two trisomy references, compared to disomy references.*

4. Finally, we used publicly available foetal liver scRNA-Seq dataset from Popescu et al. (https://doi.org/10.1038/s41586-019-1652-y) as a reference to train cell2location (113,063 cells), 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) (Figure 24). Here, we observed that the disomy versus public correlations are not significantly different from disomy versus trisomy, but significantly worse than the two disomy subset comparison. This suggests that using a matched reference may be critically important for accurate and robust conclusions from spatial transcriptomics data.

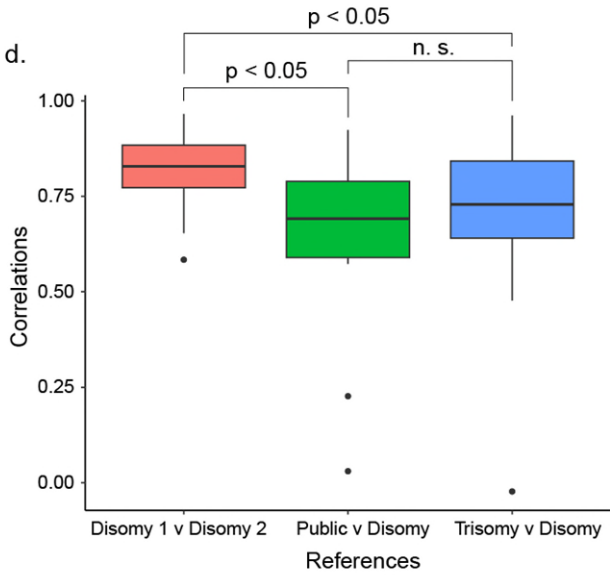



***Figure 24****. The correlation between abundances for cell types of the same type are shown. Cell2location was performed using two disomy subset references (Disomy 1 and Disomy 2), all disomy samples, Popescu et al. (public), and Trisomy as a reference for deconvoluting disomy samples. Each point in the boxplots displays a different cell type. There is greater concordance between the two disomy subset references, compared to disomy versus other references (trisomy, public).*

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). For these reasons, we have excluded from the manuscript any analyses related to calculating cell type abundances using Visium data.

For the correlation analyses with pseudobulk fold changes, if the authors were to restrict their analyses to only differentially expressed genes (rather than all genes which will mask the main differences), excluding those on chromosome 21, do the differences between disomic and ts21 still persist? For these analyses, ambient RNAs and Chromosome 22 genes are expected to significantly contribute to differences (e.g., erythroid progenitor bias leading to more erythroid differences in non-erythroid progenitor cells). Increased Ts21 chromosome 22 gene expression levels should persist after correction (e.g., SoupX).

In the previous reviewer response, we conducted the HSC Liver *versus* Femur differential expression analysis using four different statistical approaches for both disomic and Ts21 datasets (Figure 1 in the rebuttal letter).

1. Sample-level pseudobulks with the addition of sorting accounted for as a fixed effect (the original analysis in the manuscript)
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 found the results strongly concordant across these analysis decisions, such that log-fold changes had a Pearson correlation $r > 0.98$ and differentially expressed genes were consistent (99.1% overlap).

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 next evaluated whether inclusion of non-differentially expressed genes affects 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 (Figure 25).

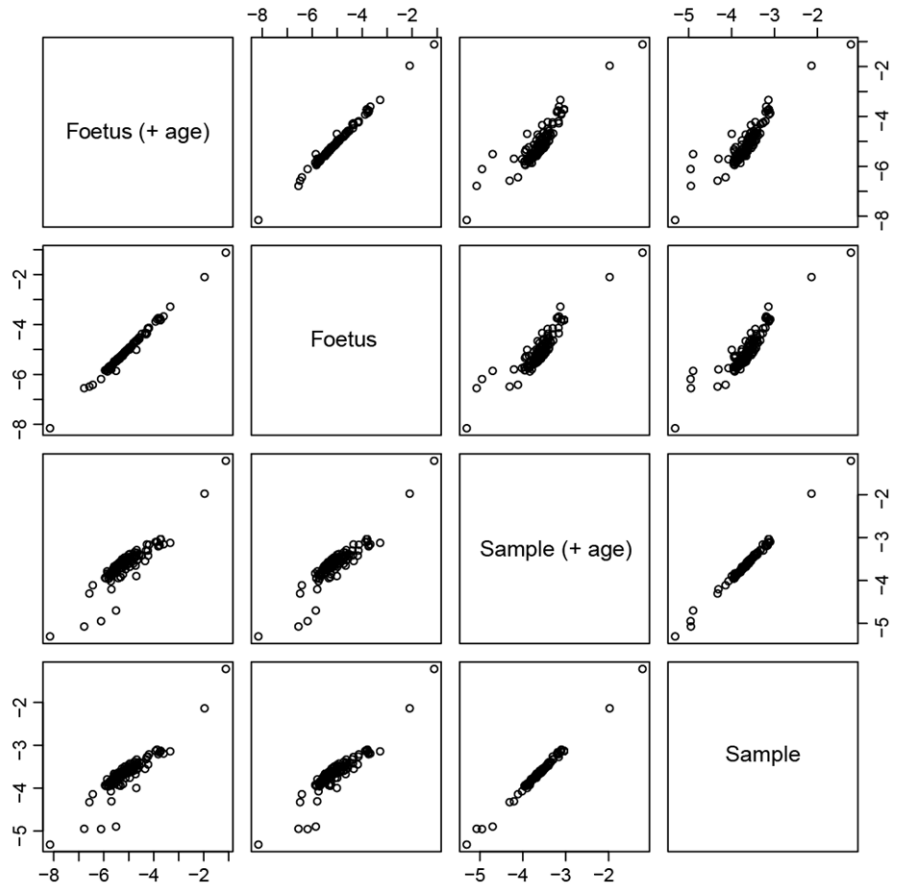



***Figure 25.*** *Log-fold changes for top 200 differentially expressed genes not located on chr 21 across the 4 analysis choices in Ts21 Liver versus Femur.*

Finally, as we noted in response to an above reviewer comment, ambient RNA does not explain the observed differences in Ts21 liver versus femur. We removed ambient RNA using CellBender. We compared the new liver versus femur HSC 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. When limiting our gene set enrichment analyses to those only significant in the post-CellBender analyses, we recapitulated the major gene set enrichments observed in the original pre-CellBender analysis.

Therefore, chr 21 genes, non-differentially expressed genes, and ambient RNA contamination do not majorly confound our original results.

Figure 1f-h trajectory analysis indicates two mesenchymal and osteoblast populations are absent in Ts21 Femur. The authors then claim that “Osteo-lineage development is altered in Ts21 bone marrow”, however, are a sufficient depth of stromal cells captured to make this conclusion? It is debatable whether most stromal cell types can be obtained from CD235a depletion without any enrichment of stromal cells markers e.g., CD271.

We would like to reiterate that the bone marrow from Ts21 and disomic samples were processed the same way, in that we used CD235 depletion to remove mature erythrocytes. The remaining cells were then used for library preparation and sequenced. We had a total of around 160,000 cells in Ts21 and 49,000 cells in disomic BM samples. Therefore, the depth of stromal cell capture can not explain depletion of some cell types/states in Ts21 and not in disomic cells, as the Ts21 data set was three times bigger.

The reviewer is correct that having a more specific sorting strategy, like CD271, would potentially enrich some rare stromal cell populations from BM that we might have not captured. Therefore, it is possible that there are other cell types, in addition to the osteo lineage, in the BM that are affected by trisomy. Of note, other studies that examined foetal BM did not capture cell types other than those that we reported. (https://doi.org/10.1016/j.stem.2022.10.005; Jardine et al, Nature).

Since the bones of the extremities of fetuses with Ts21 are only slightly shorter (PMID: 2531547), it is shocking to see Ts21 samples do not contain any osteoblasts (line 181). To further evidence this claim, the authors would need to measure bone mass/femur length and examine the histology of the femur, preferably IHC staining of osteoblasts and osteoprogenitors.

We would like to clarify that in our Ts21 dataset we couldn’t identify a clear cluster of osteoblasts but we did not state that there are no osteoblasts. It is always difficult to define distinct clusters and precise boundaries between cell populations which are along differentiation continuum. For that reason we assessed osteo-lineage development in the context of its differentiation trajectory. We noted that the osteoblasts differentiation was altered and that there was less coordinated regulation of osteo-lineage related genes in Ts21. Unfortunately obtaining a large number of long bones from Ts21 foetuses is not feasible for us, and therefore we can not measure bone mass/femur length across dozens of foetuses (in order to get sufficient statistical power) and examine the histology of the femur.

While the new mitochondrial mass and oxidative stress analyses provide a link between gene expression associated differences and DS, any links between such functional associations and specific mutations not yet identified in these subjects are premature. While the authors report that Ts21 HSC/MPP exhibit increased expression of oxidative stress and mitochondrial dysfunction genes,

this result appears to be is contradictory to results shown in Figure 2f-g and later on line 280-284, where higher mtROS is only observed in CD34+CD38+ more mature progenitors and mature Lin+ cells, but not in CD34+CD38- HSC/MPP (title 267 is also inaccurate). This again would typically suggest an ambient RNA artifact due to increased erythroid progenitor transcript contamination. Further, the authors' experimental results indicate that oxidative stress does not impact Ts21 stem and early progenitor cells. It is possible that GATA1 mutations occur in later committed progenitors due to oxidative stress (not measured due to CD235 depletion)?

In our MitoSox experiments we measured reactive oxygen species whereas our DE analysis measured changes in gene expression. These are two different methodologies with different levels of sensitivity. In general, although widely used, MitoSOX is considered a "blunt tool" and not very sensitive. The title "Haematopoietic cells in Ts21 have higher mitochondrial mass and oxidative stress" is not incorrect as both MitoSox and MitoTracker showed that haematopoietic cells in Ts21 do have higher mitochondrial mass and increased ROS.

We did not measure ROS in erythroid progenitors nor in erythroid cells. For clarity, Lineage+ cells are CD3/8/11b/56/14/19 (as stated in the manuscript) and thus include T cells, NK cells, monocytes/macrophages, granulocytes and B cells, many of which are actually decreased in number in DS compared to disomic samples. Therefore, we do not know if there is an increase in the ROS levels in erythroid cells as we did not measure it and can not argue with certainty that general increase in the number of erythrocytes can explain observed oxidative stress in HSCs due to the contamination from erythroid progenitor transcripts. More importantly, we have further evidence of mitochondrial dysfunction in HSCs through orthogonal analysis and results that can not be attributed to transcript contamination from erythroid progenitors. Specifically, we examined the enrichment of TF motifs in promoters within Ts21 and disomic HSCs using our multiome data.

For this purpose, we utilized the Eukaryotic Promoter Database (EPD), 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 (Figure 26a). 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 (Figure 26b), validating the utility of EPD promoters for TF motif enrichment analysis.

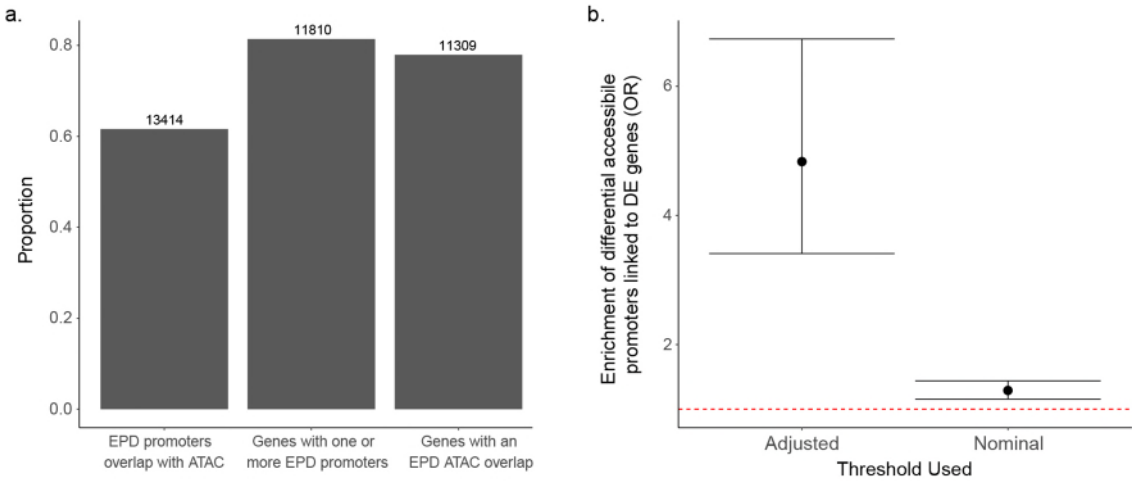



***Figure 26:*** *(a) Barplots describing the number and proportion of (i) EPD promoters that overlap with peaks accessible in HSCs, (ii) expressed genes with at least one EPD promoter, and (iii) expressed genes with at least one EPD promoter that overlaps with peaks accessible in HSCs. (b) The association between differentially accessible EPD promoters and differentially expressed (DE) linked genes (odds ratios with standard errors). Left used DE genes at FDR < 0.05 (adjusted) and right used DE genes at nominal P < 0.05.*

To perform a promoter centric motif analysis, 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 (Figure 27a). We also found varied utilization of AP1 motifs among DA promoters (Figure 27b). 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*[4], 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 27b). AP-1 has been shown to have redox sensitive DNA binding activity to CRE motif[9]. 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 (Figure 28a). This is in line with 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 (Figure 28b).

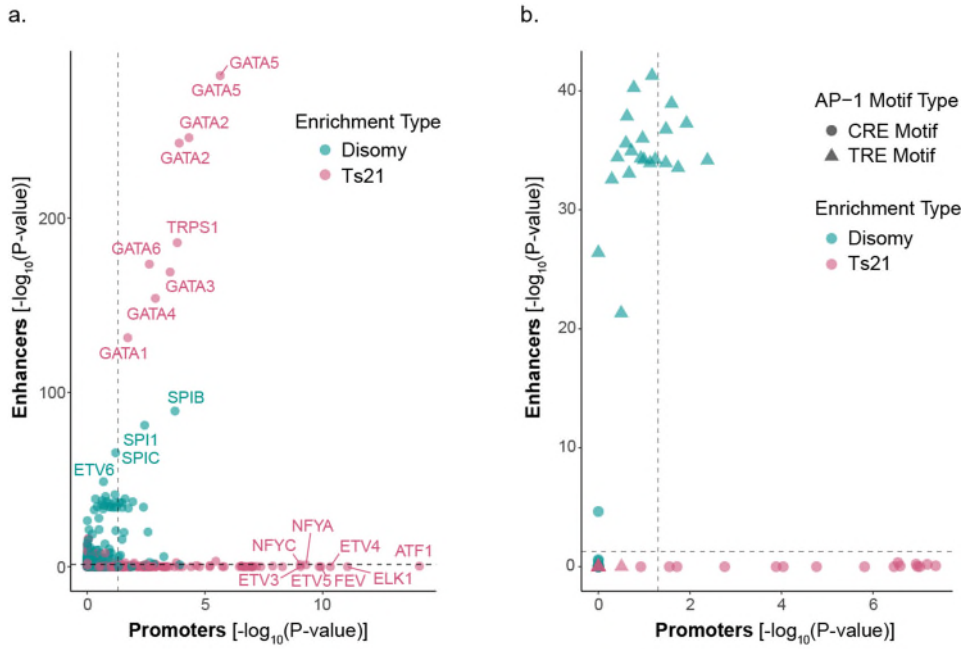



***Figure 27. (a)*** *All motif enrichments and (b) AP-1 motif related enrichments in promoters (x-axis) and enhancers (y-axis) for elements with greater accessibility in Ts21 (red) and disomic (blue) HSCs.*

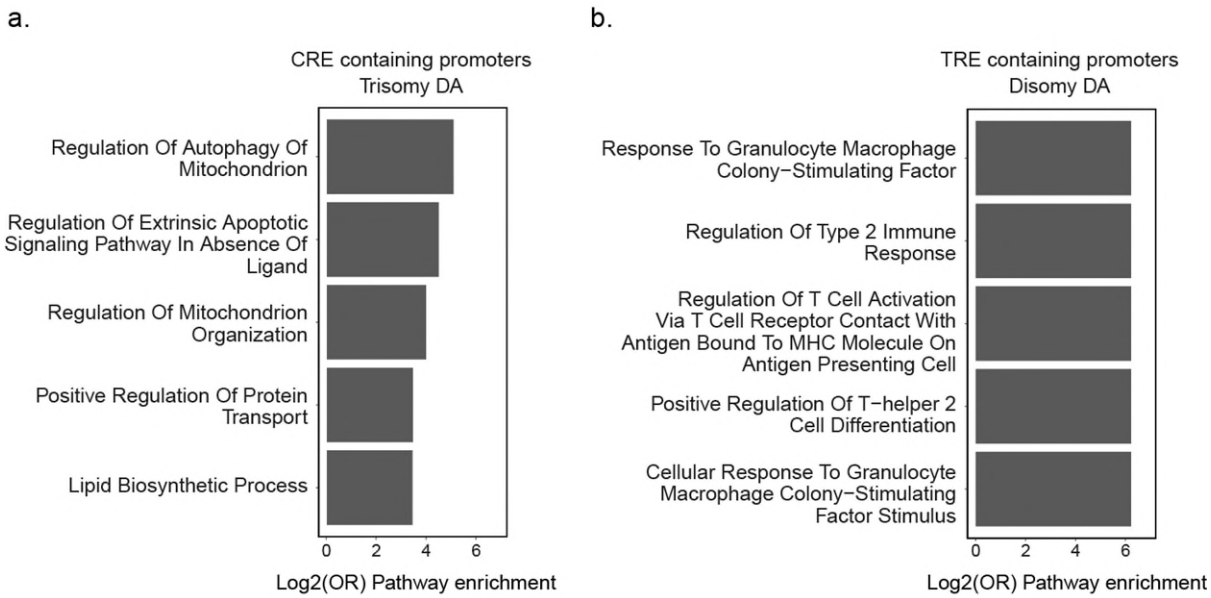



***Figure 28.*** *Gene set enrichment analysis results for (a) genes linked to CRE-containing promoters with greater accessibility in trisomy and (b) genes linked to TRE-containing promoters with greater accessibility in disomy.*

Finally, as reviewer suggested, it is possible that GATA1 mutations occur in later committed progenitors. However, acquisition of GATA1 mutation in committed progenitors is not able to trigger preleukaemia (please see our earlier comment).

We would like to point out that CD235 is not a marker for committed progenitors but for erythrocytes.

Any relevant scripts or code should be made available for review, not after, as previously requested. Data accession numbers should be included, even if restricted now.

We previously wrote under the Code and data availability portion of the manuscript:

"The code will be organised and released at: https://github.com/drewmard/t21-proj,https://github.com/drewmard/t21_multiome, and https://gitlab.com/cvejic-group/downsyndrome/.

The data has been uploaded to ArrayExpress and will be released upon publication."

The three github links containing all scripts and code have already been made available, and therefore we have updated the manuscript to write "The code is available at: https://github.com/drewmard/t21-proj,https://github.com/drewmard/t21_multiome, and https://gitlab.com/cvejic-group/downsyndrome/". We will include the ArrayExpress accession numbers in the new submission, which are:

E-MTAB-13070

E-MTAB-13062

E-MTAB-13067

**References**


1. Popescu, D.-M. *et al.* Decoding human fetal liver haematopoiesis. *Nature* **574**, 365–371 (2019).

2. Hasaart, K. A. L. *et al.* Mutation accumulation and developmental lineages in normal and Down syndrome human fetal haematopoiesis. *Sci. Rep.* **10**, 12991 (2020).

3. Balgobind, B. V. *et al.* EVI1 overexpression in distinct subtypes of pediatric acute myeloid leukemia. *Leukemia* **24**, 942–949 (2010).

4. Isakova, A. *et al.* SMiLE-seq identifies binding motifs of single and dimeric transcription factors. *Nat. Methods* **14**, 316–322 (2017).

5. Fulco, C. P. *et al.* Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).

6. Kurita, R. *et al.* Establishment of immortalized human erythroid progenitor cell lines able to produce enucleated red blood cells. *PloS One* **8**, e59890 (2013).

7. Rogers, C. E., Bradley, M. S., Palsson, B. O. & Koller, M. R. Flow cytometric analysis of human bone marrow perfusion cultures: erythroid development and relationship with burst-forming units-erythroid. *Exp. Hematol.* **24**, 597–604 (1996).

8. Mao, B. *et al.* Early Development of Definitive Erythroblasts from Human Pluripotent Stem Cells Defined by Expression of Glycophorin A/CD235a, CD34, and CD36. *Stem Cell Rep.* **7**, 869–883 (2016).

9. Abate, C., Patel, L., Rauscher, F. J. & Curran, T. Redox regulation of fos and jun DNA-binding activity in vitro. *Science* **249**, 1157–1161 (1990).

**Reviewer Reports on the Second Revision:**

Referee #1 (Remarks to the Author):

I appreciate the author's efforts on addressing my previous comments. I have only two additional comments.

Response #3. I'd suggest the authors to plot individual points for each replicate in this figure, given the limited number of samples.
Response #4. I'd suggest the authors to discuss this in the main text as one of the limitations of the analyses.

Referee #2 (Remarks to the Author):

The Authors have thoroughly addressed the concerns, and the revised manuscript now integrates a substantial amount of new data of high significance that support the conclusion of the study.

This includes the comparison with DNA methylation datasets, the bias towards Megakaryocytic lineage and the functional assessment of TRF2 in erythroid progenitor cells.
For the latter, could the Authors rephrase the new statement on line 505 to 508. Indeed, noting that these experiments have been performed in proliferation media only, it is a big statement to say they have an `increased maturation` as they only show an increased expression of CD235 surface marker. Have the Authors tested the kinetic of expression of CD235 in TRF2-overexpressing HUDEP-2 cells under differentiation?

The Authors are now providing a robust description of the increased accessibility for GATA factors in promotors and enhancers in T21 HSCs that correlate to some extent with priming and potentially predisposition to TAM/ML-DS.
These revisions are also now fully integrated into the manuscript, emphasizing the functional impact the GATA factors in these phenotypes.
Although I understand the Authors` comments regarding the difficulty of knocking down several chromosome 21 genes to demonstrate their true impact, I only have one comment:
Isn't it surprising that apart from RUNX1, none of the chromosome 21 transcription factors known to drive leukaemogenesis in children with DS (i.e ERG & ETS2) are rather more accessible in disomic cells? Same observation for BACH1, a regulator of erythropoiesis.
One would ask which specific chromosome 21 genes are enriched/overexpressed/more functional in T21 HSCs (or in other key progenitors such as MEP, ErP, MkP).
Can the lineage bias toward the erythroid/megakaryocytic lineage be solely associated with the chromosome 21 RUNX1? And obviously its potential interaction with GATA1?
This should be commented somewhere in the final manuscript.

Altogether, this review process emphasized the novelty and strengthened the conclusions of the study.

Referee #2 (Remarks on code availability):

I do not have enough expertise in coding.

Referee #4 (Remarks to the Author):

The reviewer thanks the authors for performing the requested and important evaluations, in addition to ensuring that the code is available for review. While many of the original concerns remain regarding interpretation of the results and conflicting cell conclusions (e.g., oxidative stress), the authors have addressed this primary reviewer concerns through discussion in the manuscript or supplementary results.

The remaining concerns are minor but important to address:

1. For the performed evaluations, the authors would ideally report the direct quantitative impacts (specific number of differentially expressed genes, deconvolution predicted cell type frequencies) as opposed to high-level correlation and pathway analyses. However, the specific requested analyses appear to be rigorous. For these evaluations, please include these quantitative differences in an unbiased manner (total number of DEGs with and without cellbender for HSC, variance in specific cell population cell frequency for cell2location predictions for the compiled all cell population reference versus the matched donor reference), as the existing evaluations (correlation) are not biologically informative. Such numbers rather than correlation are needed to assess impact. Please also clarify that none of the reported DEGs are from cellbender.

2. Based on the broad cell population DEG comparisons for Ts21 vs. controls, I would infer there is an upregulation of oxidative stress genes within stem cells and more mature progenitors, although MitoSOX and MitoTracker indicate these changes are only significant in more mature CD34+CD38+ progenitors. This may just be due to sampling issues (n=3), but confirming whether this indeed is the case (do the induction of oxidative stress genes correlate with MitoSOX by cell state) will be important for clarifying the proposed mechanism for why this occurs. Related, the called out oxidative stress genes are not denoted as differentially expressed in Supplemental Table 10 in HSC/MPP (DUSP1, FOXO3, APOE). Why is this?

3. Supplemental Table 5 and 6 are referenced when discussing bone marrow cell population frequency. This needs to be corrected to Supplemental Tables 6,7.

4. Supplemental Table 3 - ensure gene symbols are not converted to dates.

Referee #4 (Remarks on code availability):

The code is now provided in a manner enabling reproducibility. No new software packages or specific novel resources are described, rather re-use of existing tools.

**Author Rebuttals to Second Revision:**

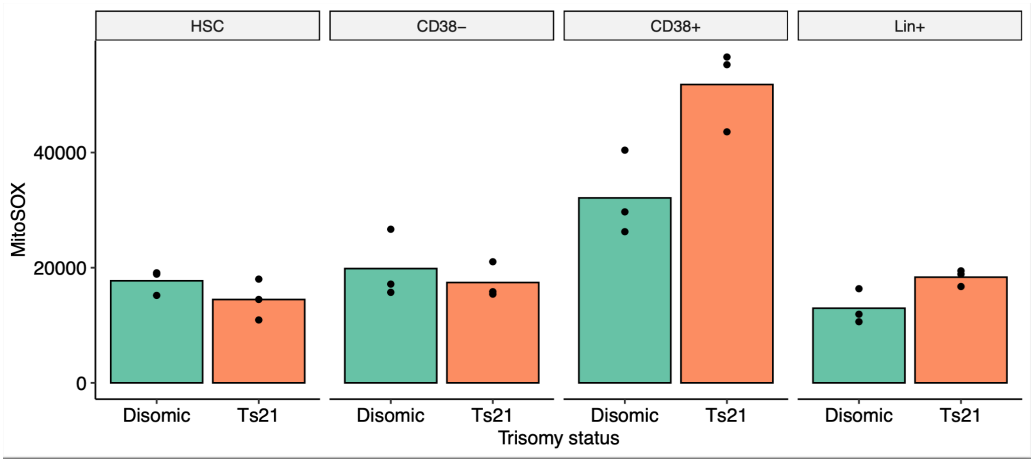
We would like to thank the reviewers for carefully reading our comments and assessing the analysis we performed. We appreciate the effort they have put to improve the quality of the work and robustness of our findings. Below we provide additional information and experiments that reviewers requested.

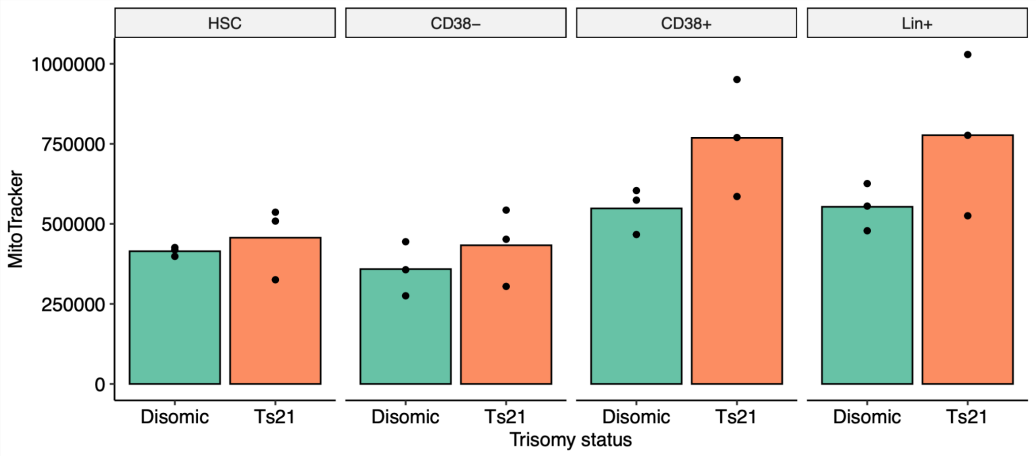
**Referee #1 (Remarks to the Author):**

I appreciate the author's efforts on addressing my previous comments. I have only two additional comments.

Response #3. I'd suggest the authors to plot individual points for each replicate in this figure, given the limited number of samples.

We have modified the figures as recommended:






Response #4. I'd suggest the authors to discuss this in the main text as one of the limitations of the analyses.

Thank you for this suggestion. We have now entered a following comment in the main text:

"In the Multiome data, we noted comparable frequencies of cell types as observed in the scRNA-seq data. However, it's important to acknowledge that the frequency of cell types in the Multiome dataset doesn't entirely mirror the results seen in the scRNA-seq experiment, possibly due to the differences in technologies and smaller sample size of our multiome dataset compared to scRNA-Seq. (Supplementary Results - Comparison of cell-type frequencies from 10X multiome versus scRNA-seq; Supp. Figure 19). In line with our analysis of scRNA-seq, we detected higher erythroid cell proportions."

More comprehensive analysis remains in the Supplementary Results - Comparison of cell-type frequencies from 10X multiome versus scRNA-seq.

**Referee #2 (Remarks to the Author):**

The Authors have thoroughly addressed the concerns, and the revised manuscript now integrates a substantial amount of new data of high significance that support the conclusion of the study.

This includes the comparison with DNA methylation datasets, the bias towards Megakaryocytic lineage and the functional assessment of TRF2 in erythroid progenitor cells.

For the latter, could the Authors rephrase the new statement on line 505 to 508. Indeed, noting that these experiments have been performed in proliferation media only, it is a big statement to say they have an `increased maturation` as they only show an increased expression of CD235 surface marker.

We have now corrected the sentence in the manuscript to the following:

"Overexpression of *TFR2* in erythroid-like HUDEP-2 cells resulted in increased expression of the maturation marker, CD235"

Have the Authors tested the kinetic of expression of CD235 in TRF2-overexpressing HUDEP-2 cells under differentiation?

Yes, we have performed these experiments (Supp. Figure 26 c-g, main manuscript). Specifically, we introduced *TFR2* overexpression constructs into undifferentiated HUDEP-2 cells using lentiviruses and incubated the cells in *differentiation media* for four days. Subsequently, we assessed TFR2, CD235a, CD71 and CD36 expression after 2 and 4 days using flow cytometry. Our analysis showed that TFR2 expression continuously increased over time in both ctrl and *TFR2* overexpressed HUDEP-2 cells. We observed a statistically significant difference between the level of expression of CD235a (1.3-fold) and CD71 (1.4-fold) but not for CD36 in *TFR2* overexpressed *versus* ctrl HUDEP-2 cells at 2 day, but no significant difference at 4 days (Figure 1). This outcome is expected considering our initial experiments (depicted in Supplementary Figure 26 a, b in the manuscript) revealed that HUDEP-2 cells exhibit up-regulation of TFR2, CD235a, CD71, and CD36 starting from day 2 when cultured in *differentiation media*. Consequently, while overexpression of TFR2 initially accelerates HUDEP-2 cell differentiation in *differentiation media* (where differentiation typically occurs), the expression of differentiation markers plateaus by 4 days as cells are fully differentiated.

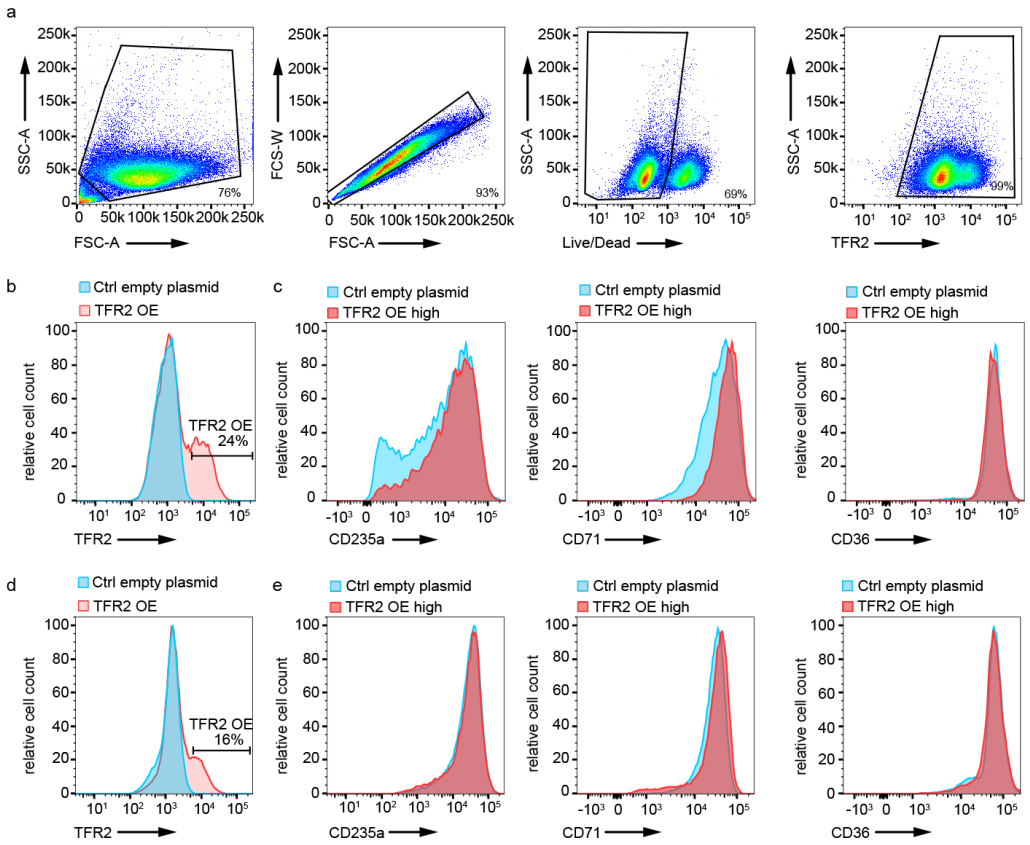



***Figure 1.*** *(a) Gating strategy used to determine surface expression of TFR2 (b, d) Histograms showing TFR2 expression in HUDEP-2 transduced with an empty vector (Ctrl, blue) in comparison with TFR2 overexpressing HUDEP-2 (light red), cultured two days (b) or four days (d) in differentiation medium. (c, e) Expression of CD235a, CD71 and CD36 in TFR2 highly expressing HUDEP-2 cells (TFR2 OE, red) in comparison with control cells cultured in differentiation media for two (b) and four (c) days.*

The Authors are now providing a robust description of the increased accessibility for GATA factors in promotors and enhancers in T21 HSCs that correlate to some extent with priming and potentially predisposition to TAM/ML-DS.

These revisions are also now fully integrated into the manuscript, emphasizing the functional impact the GATA factors in these phenotypes.

Although I understand the Authors` comments regarding the difficulty of knocking down several chromosome 21 genes to demonstrate their true impact, I only have one comment:

Isn’t it surprising that apart from RUNX1, none of the chromosome 21 transcription factors known to drive leukaemogenesis in children with DS (i.e ERG & ETS2) are rather more accessible in disomic cells? Same observation for BACH1, a regulator of erythropoiesis.

Thank you for this interesting comment. We would like to emphasise that ETS2 was enriched in *enhancers* that have significantly higher accessibility in disomic HSCs but it was enriched in *promoters* with higher accessibility in Ts21 HSCs (Supp Table 18). Recent work by Dudnyk et al. (DOI: 10.1126/science.adj011) has shown that disruption of ETS motif in *promoters* leads to decrease in the transcription initiation signal, highlighting the relevance of ETS motifs in *promoters* in gene expression. Overall, phenotypes observed in Ts21 HSCs are not solely shaped by broad alterations in the accessibility of regulatory regions and their motifs, but rather the sequence and regulatory context (e.g. promoters versus enhancers) shape gene regulation and contribute to the observed phenotypes. Therefore, although our data suggest that GATA TFs certainly have an outsized role in altering accessibility in Ts21 compared to other regulatory TFs, other TFs both on chr 21 and on other chromosomes play an important role as well (Supp Table 18, Supp. Figure 23e).

One would ask which specific chromosome 21 genes are enriched/overexpressed/more functional in T21 HSCs (or in other key progenitors such as MEP, ErP, MkP).

Can the lineage bias toward the erythroid/megakaryocytic lineage be solely associated with the chromosome 21 RUNX1? And obviously its potential interaction with GATA1?

This should be commented somewhere in the final manuscript.

We believe that the lineage bias is not solely driven by RUNX1 and GATA1. Our analysis of peak-gene links in HSCs suggests that there are hundreds of misregulated peak-gene links that are enriched for SNPs for RBC that possibly contribute to lineage bias. To address the reviewer's question about specific genes that are dysregulated on chr 21, we compiled a table with chr 21 encoded genes that are differentially expressed in Ts21, TFs with enriched motifs in differentially accessible peaks and genes that are linked with peaks in Ts21 (Supp. Table 20 in the manuscript).

We have now extended the following paragraph in the main text to clarify our findings:

"Therefore, our results indicate that trisomy restructured the regulatory element-gene landscape by altering the correlation between peak accessibility and gene expression, which led to a substantial increase in accessibility and upregulation of these elements in Ts21 compared to the control; overall indicating that a small number of lineage TFs become more active in Ts21, increase coordination between chromatin accessibility and gene expression, and drive lineage bias. It is important to note that while our data indicate a significant impact of GATA TFs on altering accessibility in Ts21 compared to other regulatory TFs, other TFs and genes, both located on chromosome 21 (Supp. Table 20) and on other chromosomes, also play a crucial role in this process."

We have further emphasized this point in the discussion, writing:

"Compared to disomic HSCs, they exhibited greater concordance between RNA and ATAC profiles, which is a feature typical of mature cell states[34,57], with broad molecular dysregulation in Ts21 that extended beyond chromosome 21."

Altogether, this review process emphasized the novelty and strengthened the conclusions of the study.

Referee #2 (Remarks on code availability):

I do not have enough expertise in coding.

**Referee #4 (Remarks to the Author):**

The reviewer thanks the authors for performing the requested and important evaluations, in addition to ensuring that the code is available for review. While many of the original concerns remain regarding interpretation of the results and conflicting cell conclusions (e.g., oxidative stress), the authors have addressed this primary reviewer concerns through discussion in the manuscript or supplementary results.

The remaining concerns are minor but important to address:

1. For the performed evaluations, the authors would ideally report the direct quantitative impacts (specific number of differentially expressed genes, deconvolution predicted cell type frequencies) as opposed to high-level correlation and pathway analyses. However, the specific requested analyses appear to be rigorous. For these evaluations, please include these quantitative differences in an unbiased manner (total number of DEGs with and without cellbender for HSC, variance in specific cell population cell frequency for cell2location predictions for the compiled all cell population reference versus the matched donor reference), as the existing evaluations (correlation) are not biologically informative. Such numbers rather than correlation are needed to assess impact. Please also clarify that none of the reported DEGs are from cellbender.

As the reviewer requested, we included Supp. Table 12 and 13 in the manuscript showing DEG analysis results both with and without cellbender. Supp. Table 12 describes the HSC Ts21 *versus* disomy DE analysis, while Supp. Table 13 describes the liver *versus* femur analysis. With the exception of APOE, all originally reported genes of the liver versus femur analysis are significant after performing ambient RNA removal using cellbender. We have revised this in the main text, writing:

"Notably, genes involved in response to ROS (*FOXO3, SOD1, DUSP1, JUN, FOS*) were upregulated in HSC/MPPs in Ts21 foetal liver compared to BM". We have added in the Supplemental Results section **Evaluating the impact of ambient RNA on differential expression results**: "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."

In addition, as the reviewer suggested, we have included a Supp. Table 10 indicating cell abundances predicted by cell2location using different references, as described in detail in the supplementary section titled "Examining cell2location deconvolution performance using different scRNA-Seq references". We have included columns specifying the difference in predicted abundances between different references.

2. Based on the broad cell population DEG comparisons for Ts21 vs. controls, I would infer there is an upregulation of oxidative stress genes within stem cells and more mature progenitors, although MitoSOX and MitoTracker indicate these changes are only significant in more mature CD34+CD38+ progenitors. This may just be due to sampling issues (n=3), but confirming whether this indeed is the case (do the induction of oxidative stress genes correlate with MitoSOX by cell state) will be important for clarifying the proposed mechanism for why this occurs. Related, the called out oxidative stress genes are not denoted as differentially expressed in Supplemental Table 10 in HSC/MPP (DUSP1, FOXO3, APOE). Why is this?

To further investigate the impact of oxidative stress in HSCs and their progenitors, we utilized all 228 genes associated with GO:0000302 "response to reactive oxygen species" and calculated their mean log-fold changes in Ts21 versus disomy across all cell types tested in the scRNA-seq data (Figure 2). The mean log-fold changes for these oxidative stress-related genes in HSCs was found to be 1 standard deviation (SD) above the calculated mean in other cell types, ranking as the 2nd highest among all cell types. Monocyte progenitors exhibited the highest differential expression, with a mean DE of +2.45 SD, making them the biggest outlier in terms of oxidative stress gene upregulation. Finally, MEMPs showed a mean DE of -0.71 SD, indicating a lower impact of oxidative stress gene upregulation compared to other cell types. These results support the notion that HSCs exhibit upregulation of oxidative stress genes, which intensifies in more mature progenitors like CD34+CD38+ committed progenitors (specifically monocyte progenitors). Here we systematically evaluated many oxidative stress related genes and their upregulation in Ts21 cells. Our new findings are in line with the initial hypothesis that leukemia risk and blood differentiation abnormalities in Down Syndrome (DS) are closely linked to oxidative stress mechanisms. This provides a compelling direction for further experimental validation and analysis to elucidate these pathways more comprehensively.

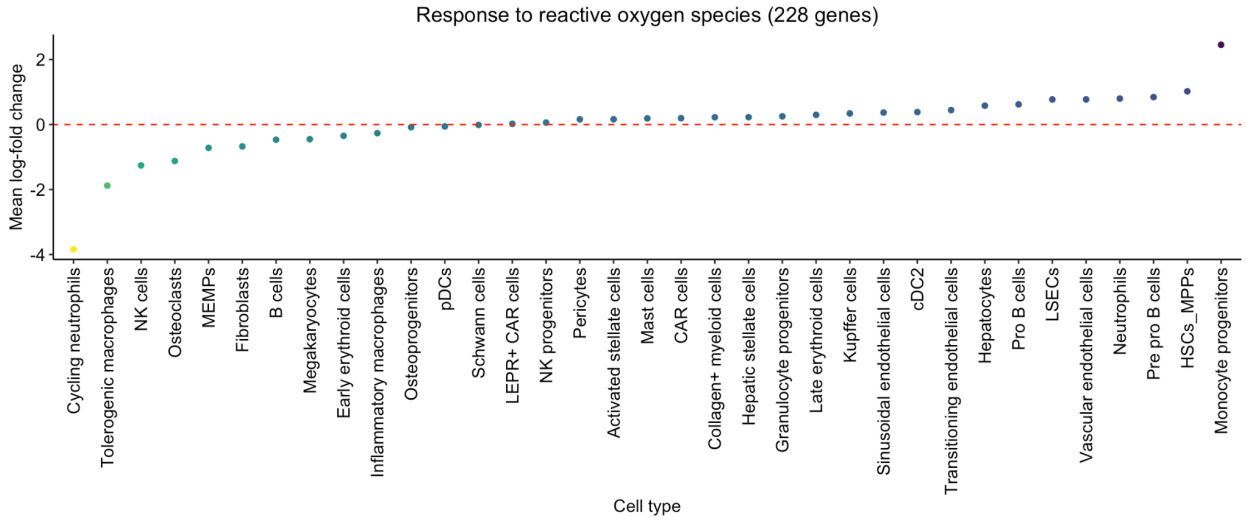



***Figure 2.*** *Mean log-fold change values across 228 genes in the GO:0000302 "response to reactive oxygen species" for each cell type.*

To answer the reviewer's question about the content of the Supp. Table 10 (in previous revision of the manuscript, now Supp. Table 11) - oxidative stress was originally observed in the DEA of liver versus femur HSCs (Figure 2b, c). Supp. Table 11 is showing DE genes in Ts21 vs disomic HSCs in the liver. The lack of significant differential expression of specific oxidative stress genes (e.g. *DUSP1*, *FOXO3*, *APOE*) in the HSC/MPP population in Supp. Table 11 could be attributed to cell-type-specific expression patterns, statistical power, and the sensitivity of our differential expression analysis in these subpopulations. However, the top upregulated gene in Ts21 compared to disomic HSCs was SOD1 (superoxide dismutase type 1) which regulates expression of genes involved in oxidative stress response.

We are committed to addressing these complexities in future studies and appreciate the reviewer's suggestions for deeper investigation into the correlation between oxidative stress gene induction and MitoSOX by cell state. This will indeed be an important step to clarify the proposed mechanisms. Thank you again for your valuable feedback.

3. Supplemental Table 5 and 6 are referenced when discussing bone marrow cell population frequency. This needs to be corrected to Supplemental Tables 6,7.

Thank you, we have now corrected this.

4. Supplemental Table 3 - ensure gene symbols are not converted to dates.

Thank you, we have now corrected this.

Referee #4 (Remarks on code availability):

The code is now provided in a manner enabling reproducibility. No new software packages or specific novel resources are described, rather re-use of existing tools.

**Reviewer Reports on the Third Revision:**

Referee #2 (Remarks to the Author):

I appreciate the thorough responses from the Authors.
This has significantly strengthened the description of the results and the conclusions of this study.

As a minor comment, could the `Disomic` and `Ts21` be swapped for consistency throughout the figures (left/right). Also, `quantitative comparison of ` should not be in bold (line 842).

Other than that, I have not further concerns to be addressed.

Referee #2 (Remarks on code availability):

I am not computational biologist but all 3 links work.

Referee #4 (Remarks to the Author):

The reviewer thanks the authors for the additional clarifications and thorough access to supplemental results. The authors have addressed all major outstanding concerns of this reviewer.

Referee #4 (Remarks on code availability):

All reported software packages and methods appear to be appropriate and run in an acceptable manner, as per the author descriptions.

**Author Rebuttals to Third Revision:**

Referees' comments:

Referee #2 (Remarks to the Author):

I appreciate the thorough responses from the Authors.
This has significantly strengthened the description of the results and the conclusions of this study.

As a minor comment, could the `Disomic` and `Ts21` be swapped for consistency throughout the figures (left/right).

We appreciate this suggestion and have strived for consistency in this matter for the most part. In the few instances that a variant structure is used in the figures, we felt that this made it easier to comprehend the data.

Also, `quantitative comparison of ` should not be in bold (line 842).

The font was changed accordingly.

Other than that, I have not further concerns to be addressed.

Referee #2 (Remarks on code availability):

I am not computational biologist but all 3 links work.

Referee #4 (Remarks to the Author):

The reviewer thanks the authors for the additional clarifications and thorough access to supplemental results. The authors have addressed all major outstanding concerns of this reviewer.

Referee #4 (Remarks on code availability):

All reported software packages and methods appear to be appropriate and run in an acceptable manner, as per the author descriptions.