

## Single-cell gene and isoform expression analysis reveals signatures of ageing in haematopoietic stem and progenitor cells

Laura Mincarelli <sup>1✉</sup>, Vladimir Uzun<sup>1</sup>, David Wright<sup>1</sup>, Anita Scoones<sup>1</sup>, Stuart A. Rushworth <sup>2</sup>, Wilfried Haerty <sup>1✉</sup> & Iain C. Macaulay <sup>1✉</sup>

Single-cell approaches have revealed that the haematopoietic hierarchy is a continuum of differentiation, from stem cell to committed progenitor, marked by changes in gene expression. However, many of these approaches neglect isoform-level information and thus do not capture the extent of alternative splicing within the system. Here, we present an integrated short- and long-read single-cell RNA-seq analysis of haematopoietic stem and progenitor cells. We demonstrate that over half of genes detected in standard short-read single-cell analyses are expressed as multiple, often functionally distinct, isoforms, including many transcription factors and key cytokine receptors. We observe global and HSC-specific changes in gene expression with ageing but limited impact of ageing on isoform usage. Integrating single-cell and cell-type-specific isoform landscape in haematopoiesis thus provides a new reference for comprehensive molecular profiling of heterogeneous tissues, as well as novel insights into transcriptional complexity, cell-type-specific splicing events and consequences of ageing.

<sup>1</sup>Earlham Institute, Norwich Research Park, Norwich NR4 7UZ, United Kingdom. <sup>2</sup>Norwich Medical School, The University of East Anglia, Norwich Research Park, Norwich, United Kingdom. ✉email: [lm26@sanger.ac.uk](mailto:lm26@sanger.ac.uk); [wilfried.haerty@earlham.ac.uk](mailto:wilfried.haerty@earlham.ac.uk); [iain.macaulay@earlham.ac.uk](mailto:iain.macaulay@earlham.ac.uk)

Single-cell RNA-seq (scRNA-seq) technologies are now applied to a broad spectrum of biological systems<sup>1</sup> with particular impact in the study of stem cell and developmental biology<sup>2–4</sup>. With the advance of short-read technologies capable of analysing many thousands of cells in a single experiment, it has become possible to identify cell types and state transitions in complex biological systems. In particular, scRNA-seq has highlighted the continuous nature of haematopoietic hierarchy. Cell types, previously thought of as discrete entities, have been shown to exist in a continuum of states from stem cells to mature progenitors<sup>5</sup>. Investigating the regulatory events that occur in these state transitions and how they change with age is central to the understanding of the regenerative potential of stem cells in health and disease.

Alternative splicing (AS) of mRNA transcripts is a mechanism by which several isoforms can be generated from individual genomic loci, enabling significant increases in transcriptomic and proteomic complexity. AS can affect many aspects of gene expression, including transcript export from the nucleus, transcript stability, and, critically, the production of functionally distinct protein isoforms. AS is thought to occur in at least 62% of multi-exonic genes in mouse<sup>6</sup> and up to 95% of multi-exonic genes in human<sup>7</sup>. Increasingly, there is an understanding that isoform (co-)expression in tissues and cells can reveal previously unseen complexities in cell signalling responses, as was demonstrated recently for G-protein coupled receptors<sup>8</sup>.

In haematopoiesis, substantial levels of alternative splicing have been observed in sorted populations of stem and progenitor cells<sup>9,10</sup>, but in general, it remains overlooked in transcriptional studies of haematopoiesis, including scRNA-seq studies. Because the vast majority of scRNA-seq studies of haematopoiesis have used 3' cDNA sequencing, AS events are unlikely to be captured, and thus an entire class of biologically important information about isoform usage is lost. Advances in long-read sequencing technologies have enabled unequivocal detection of AS isoforms<sup>11</sup>, and recently, these approaches have been adapted to explore full-length transcript sequences from single-cell experiments<sup>12,13</sup>. Here, we have applied an integrated approach for parallel short- (Illumina) and long-read (Pacific Biosciences; PacBio) single-cell sequencing of Fluorescence Activated Cell Sorting (FACS) enriched populations, using the 10X Genomics Chromium, to enable comprehensive profiling of cellular diversity, gene expression and alternative splicing events in the mouse haematopoietic system.

We generated cell-barcoded cDNA from haematopoietic stem and progenitor cells isolated from young (8 weeks old) and aged (72+ weeks old) C57BL/6J mice. We then undertook conventional Illumina sequencing of this cDNA to reveal haematopoietic cell states, gene expression, and cell frequency changes associated with ageing. Parallel PacBio sequencing of this full-length cDNA and integration of the cell barcodes annotated the single-cell short-read data with isoform-level information, enabled a survey of isoform expression in the haematopoietic stem cells (HSCs) and their progeny (Fig. 1A). We demonstrate that AS is readily detectable by long-read sequencing of scRNA-seq libraries and that many genes, including known regulators of HSCs and their progeny, are expressed as diverse transcripts, often encoding functionally distinct proteins. These functionally divergent isoforms, undetectable by short-read sequencing alone, indicate that isoform-level analysis is critical for the understanding of cellular systems and states.

## Results

**Annotation of short-read scRNA-seq data with isoform-level information.** Using fluorescence-activated cell sorting (FACS), we isolated the Lineage-negative, cKit (Cd117) positive (LK) cell

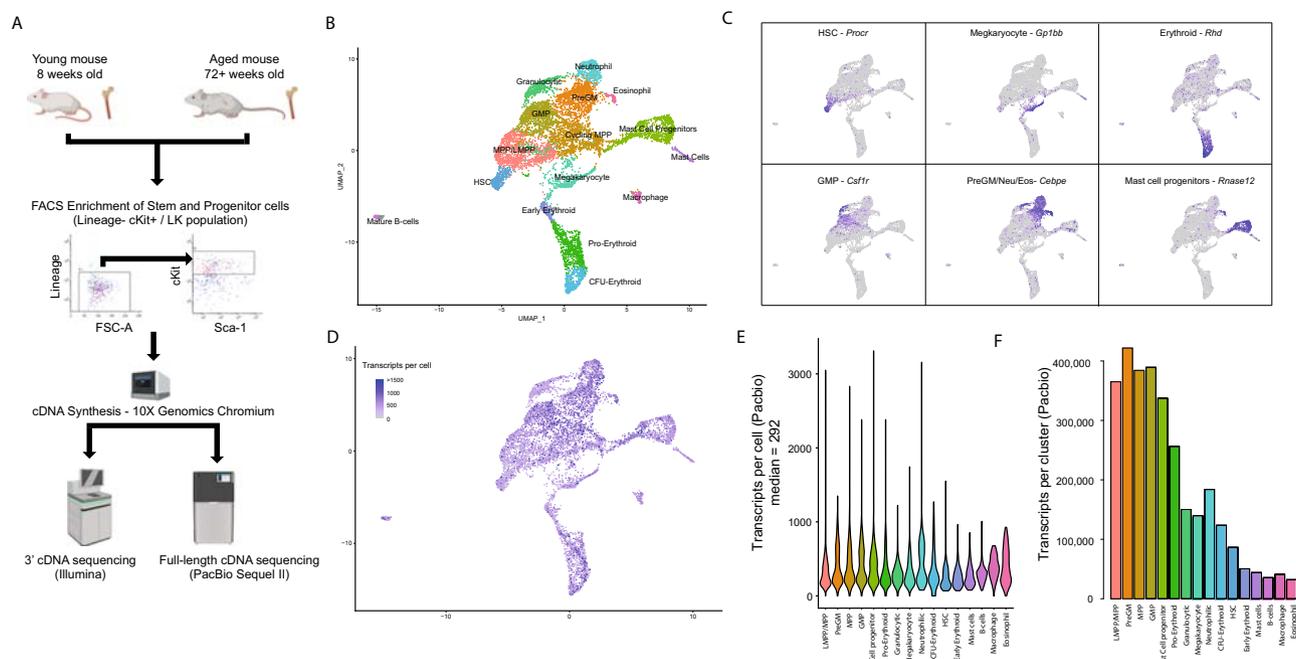
fraction of mouse bone marrow cells, a population containing stem and progenitor cells<sup>14</sup> from young (8 weeks old,  $n = 3$ ) and aged (72+ weeks old,  $n = 3$ ) mice. We generated standard 10×3' scRNA-seq libraries from these populations from each mouse, revealing the diversity of cell types present within the 8000 LK cells passing quality control (Fig. 1B).

Analysis of this data identified 15 subclusters within a largely continuous LK population, which could be manually annotated based on classical marker gene expression, including haematopoietic stem cell (HSCs) here associated with *Procr* (*Cd201*) expression<sup>15</sup>, as well as intermediate and committed progenitor cells (Fig. 1B, C, Supplementary Data 1). This includes myeloid, megakaryocytic and erythroid lineages, matching the diversity expected from FACS-based phenotypic analysis of the same population<sup>14</sup>. Small numbers of mature B-cells, myeloid cells and mast cells were also observed but were transcriptionally very distinct from the main stem and progenitor cell cluster and most likely represent low-level contamination with mature cells (Supplementary Fig. 1). In order to examine transcriptional diversity in the haematopoietic system at isoform level, we performed long-read PacBio sequencing (IsoSeq) on each of the six cDNA pools generated from the 10X Genomics platform, similar to an approach recently applied in cerebellar cells<sup>12</sup>. This approach, taking advantage of the cell barcoding technology used in 10X Genomics library preparation, enables isoform identification and association with cell populations and individual cells through the integration of long- and short-read data.

A detailed breakdown of the PacBio sequencing statistics is presented in Supplementary Table 1. In brief, PacBio sequencing yielded a total of 17.9 million circular consensus sequencing (CCS) reads with a median read length of 1471 bases (Supplementary Fig. 2A). These reads mapped to an average of 16,427 genes per sample, representing an average of 33,345 transcripts per sample and an average of 31 reads per transcript. Transcript coverage averaged 74% (Supplementary Fig. 2B), and alternative isoforms were detected for 52.3% of genes (Supplementary Fig. 2C), with the majority of transcripts being protein-coding and spread across a variety of gene categories (Supplementary Fig. 2D, E).

Demultiplexing of the long-reads using the short-read cell barcodes enabled 5.8 million CCS reads (32% of the total) to be assigned to individual cells, reflecting other similar studies<sup>12</sup> with many reads omitted due to incomplete barcode sequencing/detection in the long-read data, as well as long-reads that could not be assigned due to the QC filtration of the short-read data. Using the multi-modal capabilities of Seurat<sup>16</sup>, we integrated short- and long-read datasets enabling annotation of the short-read dataset with isoform-level transcript expression (Fig. 1D) and allowing side-by-side comparison of gene and isoform expression from the respective platforms. A median of 411 reads, corresponding to a median of 292 transcripts, could be assigned per cell (Supplementary Fig. 3, Fig. 1E), with the number of isoforms detected being too low to allow meaningful comparisons at single-cell resolution. However, with 50,000–500,000 long-reads per cluster, scaling with the number of cells per cluster (Fig. 1F), it is possible to visualise isoform usage across the analysed populations and to associate isoform expression with cell-type clusters.

**Alternative splicing in haematopoietic transcription factor networks.** We first screened the dataset to identify AS events in key transcription factors (TFs) that regulate cell fate decisions across haematopoiesis using a regulatory network derived from relevant single-cell studies<sup>17</sup>. Our long-read sequencing detected 28 (of 31) TFs in that network (Fig. 2A), including three



**Fig. 1** Integrated short- and long-read single-cell RNA-seq of murine haematopoietic stem and progenitor cells. **A** Method overview (created using elements from BioRender) **B** UMAP plot of haematopoietic stem and progenitor cells obtained from 3' cDNA sequencing of the Lin<sup>-</sup> cKit<sup>+</sup> population. Clusters have been annotated based on known marker gene expression. **C** Selected marker gene expression demonstrating the presence of HSCs, megakaryocytes and myeloid/erythroid progenitors. **D** Numbers of transcripts, derived from long-read sequencing, mapping to individual cells. **E** Distribution of transcripts detected in cells from each of the cell-type clusters. **F** Total number of transcripts detected in each of the cell-type clusters.

predominant isoforms of *Lmo2* - *Lmo2-202*, *Lmo2-203* and *Lmo2-208*, each of which encodes a protein differing in length (228, 220 and 158 amino acids, respectively) with progressive truncation from the N-terminus (Fig. 2B). The major isoform, *Lmo2-208* is ubiquitously expressed, while *Lmo2-202* and *-203* show limited expression in the megakaryocyte and erythroid lineages. Additionally, by screening the data for novel exons, we identified a novel, in-frame variant of *Lmo2* with a 297-bp exon supported by 27 long-reads (Supplementary Fig. 4). In human cell lines, long- and short-protein isoforms of *Lmo2* (equivalent to the proteins encoded by *Lmo2-202* and *-203*) have been shown to have distinct functional roles in the formation of TF complexes and regulation of gene expression<sup>18</sup>.

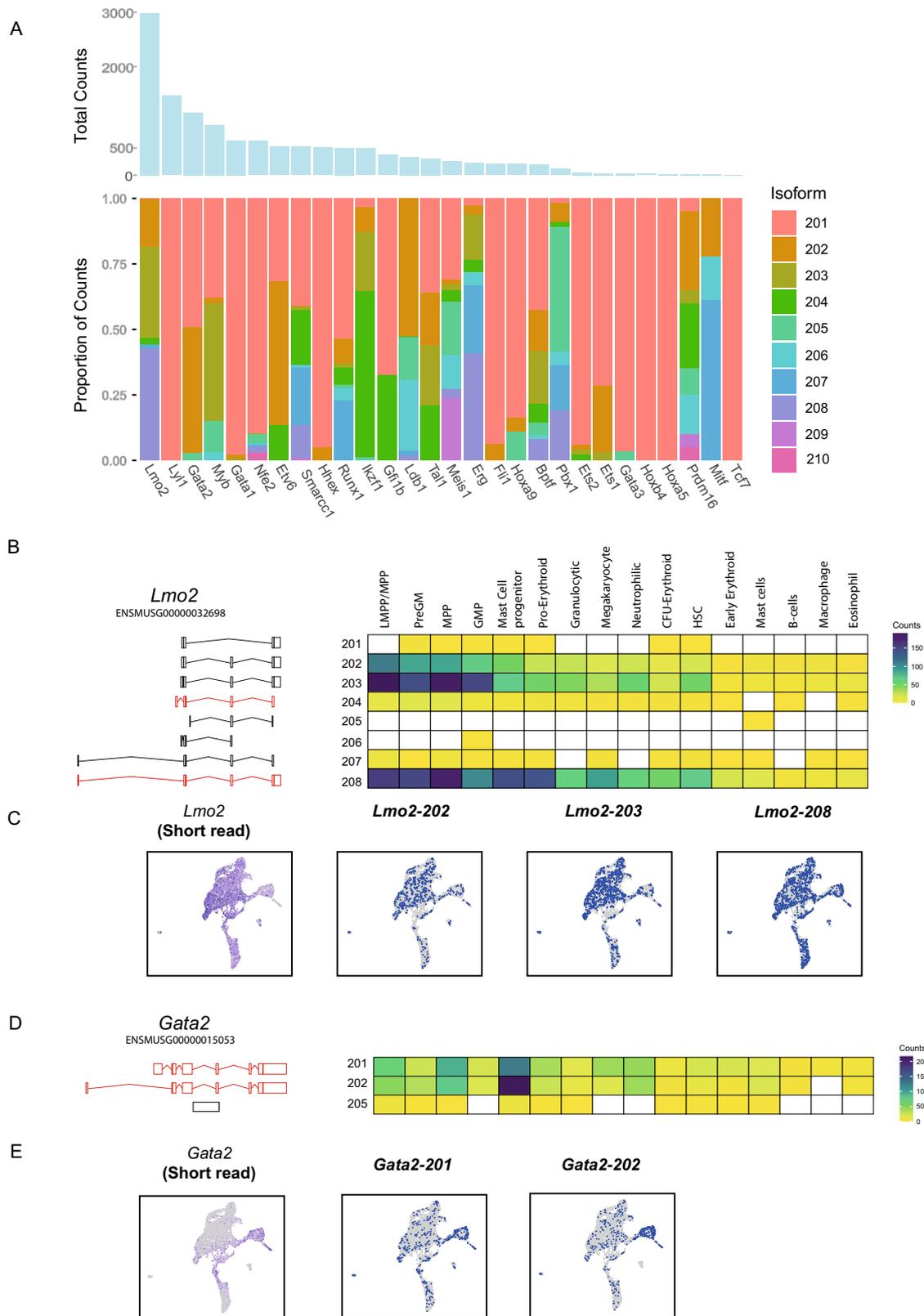
*Gata2* expression, which in our short-read data is restricted to stem and early progenitor cells, mast cell and MkP and erythroid populations, consists of two main isoforms, *Gata2-201* and *Gata2-202* (Fig. 2C). They are translated into the same protein but are differentiated by the usage of distinct distal exons, and have previously been shown to exhibit some lineage-specific expression<sup>19</sup>. Here, both isoforms are most abundant in the mast cell population, with *Gata2-202* showing more restricted expression in this cell type (Fig. 2D, E). We also observe the expression of multiple isoforms of *Ldb1* and *Tal1* (Supplementary Fig. 5A, B) and even in transcripts with relatively low long-read counts, such as *Meis1*, we can observe an exon skipping isoform (exon 11), which encodes a functionally distinct protein, equivalent to *MEIS1D* in human<sup>20</sup> (Supplementary Fig. 5C).

**Alternative splicing and cytokine receptors.** Transmembrane proteins can determine the cellular response to its environment, particularly regulatory cytokines. We screened the long-read data for isoform expression of a panel of cytokine receptors (Fig. 3A). Two predominant isoforms of the stem cell factor (SCF) receptor *Kit* (*Kit-201* and *Kit-204*) were observed to be expressed in virtually all cell types (Fig. 3B). These isoforms encode proteins

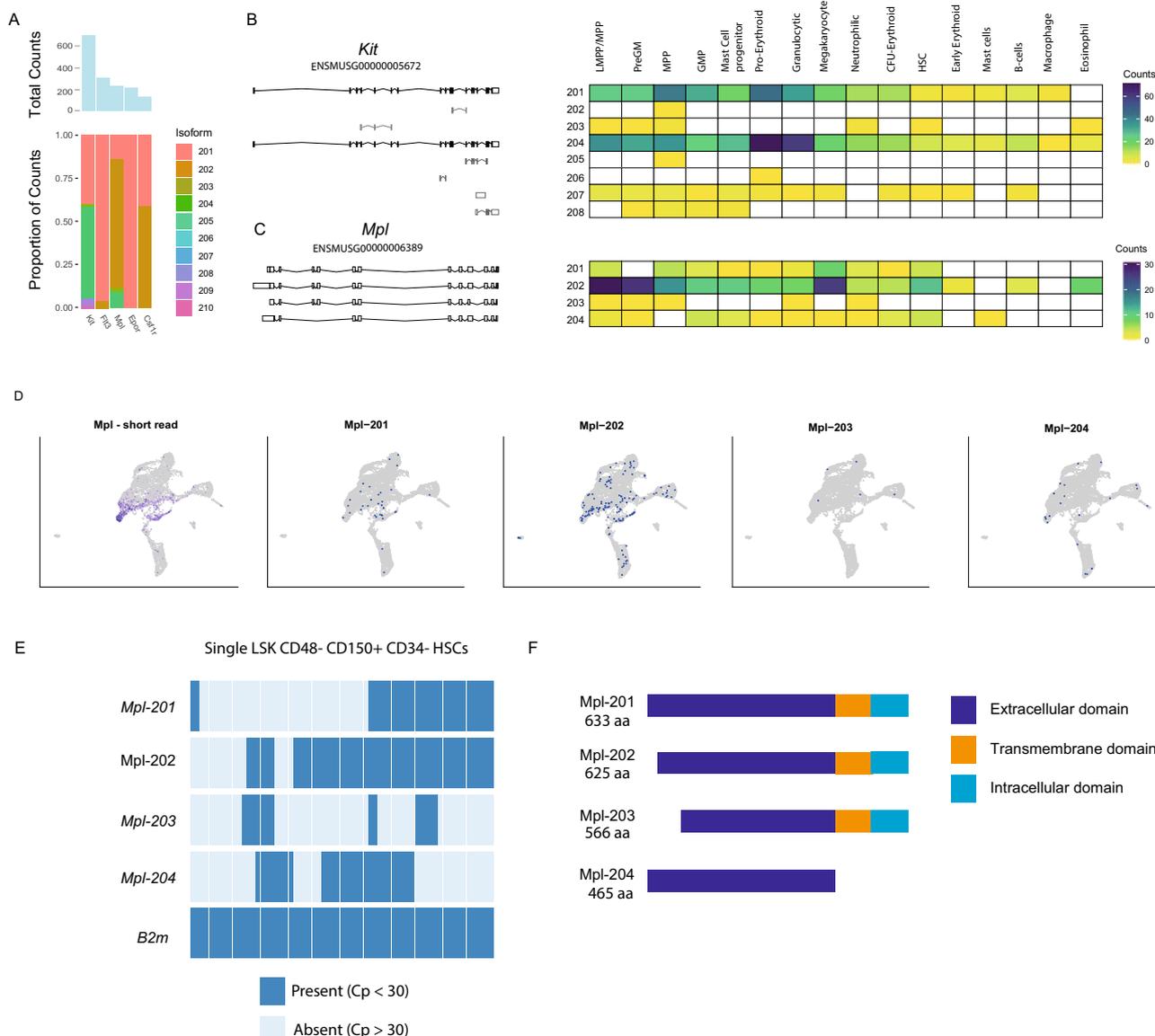
differing in length by 4 amino acids (KGNN) which, when absent, results in the loss of a low-complexity region in the juxtamembrane extracellular region of the protein. The human equivalent of these variants has been shown to display distinct signalling activities<sup>21–23</sup>. Four isoforms of *Mpl*, the gene encoding the thrombopoietin (ThPO) receptor<sup>24</sup>, were detectable in both stem cells and megakaryocytes (Fig. 3C, D). The primary isoform, *Mpl-201*, encodes the transmembrane Mpl receptor. *Mpl-202* encodes a shortened version of the protein, lacking eight amino acids in the extracellular domain. *Mpl-204*, which lacks exons 9 and 10 completely, encodes a truncated protein with no transmembrane domain. This isoform is detectable in the stem cell population, and functional studies have indicated that it has an inhibitory role on normal Mpl signalling<sup>24</sup>.

Using junction-targeting qPCR, we examined the distribution of Mpl isoforms in single FACS-isolated HSCs (LSK Cd150+ Cd48<sup>-</sup> Cd34<sup>-</sup>) (Fig. 3E). This demonstrated that individual HSCs frequently express more than one isoform, and in some cases, three or four isoforms could be detected in the same single cell. *Mpl-202* is the most common isoform but is frequently co-expressed with *Mpl-204* and also occasionally with *Mpl-203*. The observation that individual HSCs can express more than one transcript encoding distinct protein isoforms (Fig. 3F) of this key cytokine receptor suggests a way in which ThPO signalling could regulate in the functionally heterogeneous or lineage-primed stem cell pool. This highlights the critical need to understand not just gene but also isoform expression in single-cell studies.

**Signatures of ageing.** We integrated short- and long-read data to reveal age-associated signatures in cell type abundance, gene and isoform expression. While the majority of cell type abundances were unchanged between 8- and 72-week-old mice, there was an increased abundance of phenotypic LT-HSCs (Fig. 4A–C) in keeping with previous findings based on stem cells defined by



**Fig. 2 Alternative splicing of haematopoietic transcription factors. A** Total long-read counts and isoform frequency plot for haematopoietic transcription factors. **B** Cell-type expression of isoforms of *Lmo2*. Transcripts are numbered based on their Ensembl isoform suffix. Transcripts in red encode the same protein sequence, while transcripts in black encode distinct proteins. Transcripts in grey are noncoding. **C** UMAP projection of gene and isoform expression for *Lmo2* in haematopoietic cells. **D** Cell-type expression of isoforms of *Gata2*. Both major transcripts encode the same protein, although the transcript has an alternative 5' UTR. **E** UMAP projection of gene and isoform expression for *Gata2* in haematopoietic cells.



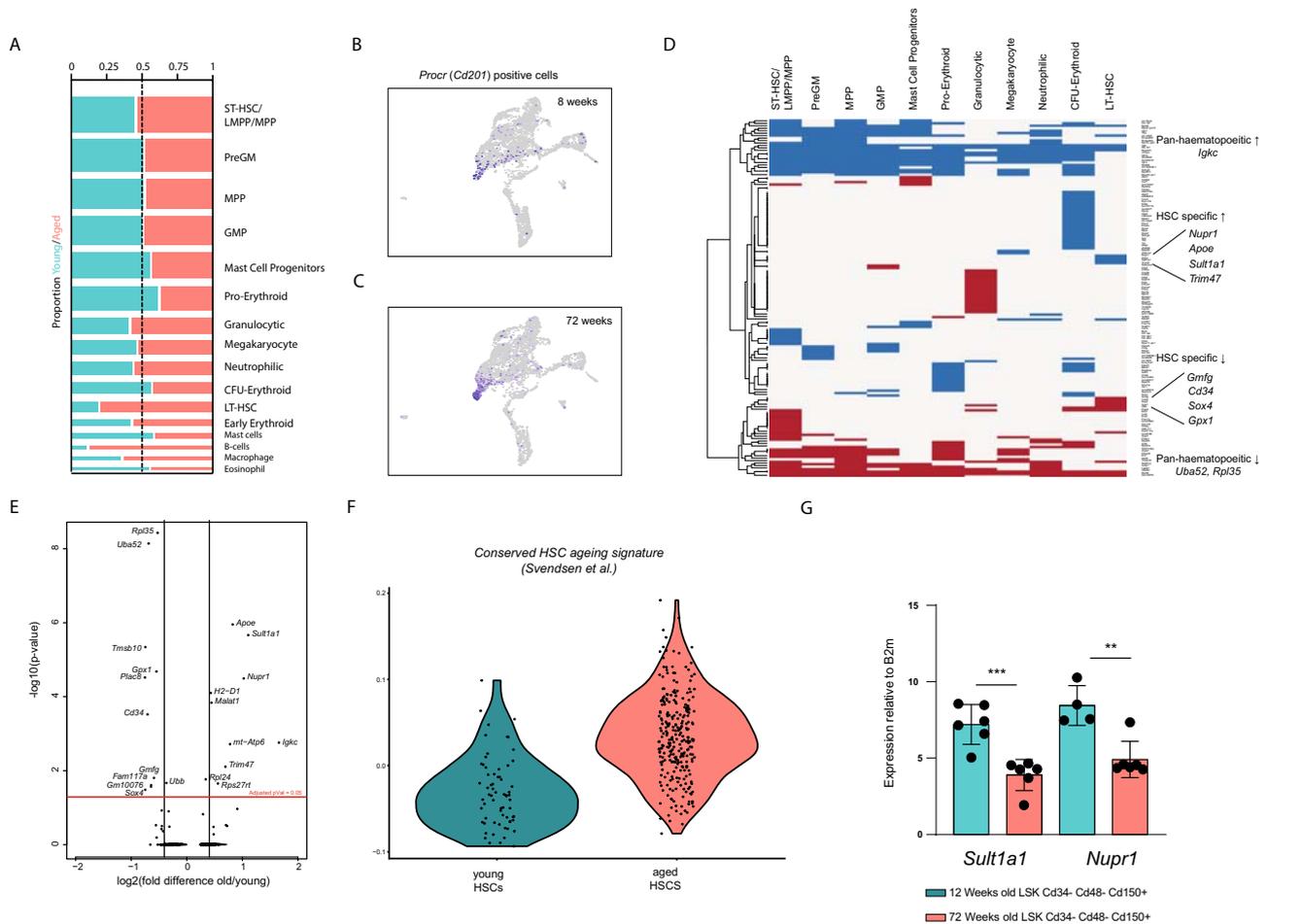
**Fig. 3 Isoform diversity in haematopoietic cytokine receptors. A** Total long-read counts and isoform frequency plot for cytokine receptors regulating haematopoiesis. **B** Cell-type expression of isoforms of *Kit*. Transcripts in black encode distinct proteins from the same gene. Those in grey are noncoding. **C** Cell-type expression of isoforms of *Mpl*. Transcripts in black encode distinct proteins from the same gene. Those in grey are noncoding. **D** UMAP projection of gene and isoform expression for *Mpl* in haematopoietic cells. **E** Detection of *Mpl* isoforms by junction-specific qPCR in individual phenotypic (LSK Cd48<sup>-</sup> Cd150<sup>+</sup> Cd34<sup>-</sup>) HSCs. Heatmap shows isoform expression in individual cells (columns) of four isoforms of *Mpl*. *B2m* is presented as a positive control. **F** Diagram of the distinct protein structures encoded by the distinct isoforms of *Mpl*.

protein marker expression (Lineage negative, Sca-1 positive, cKit positive (LSK) Cd34<sup>-</sup> Cd48<sup>-</sup> Cd150<sup>+</sup> cells)<sup>25</sup>.

To identify transcriptional signatures associated with ageing across the haematopoietic hierarchy, we performed differential gene expression analysis in each subpopulation using the short-read sequencing data (Fig. 4D). We observed downregulation of genes encoding the cytosolic ribosomal components *Rpl35* and *Uba52* in stem cells and all progenitors and upregulation of the immunoglobulin kappa chain (*Igkc*) throughout the haematopoietic system. To further explore the transcriptional response to ageing, we performed a comparison of the gene expression signature of young and aged HSCs with that from a curated database of HSC ageing genes<sup>26</sup> showed that the signature from 220 ageing-associated genes was enriched for in our aged HSCs (Fig. 4F) with a general trend of the proportional increase of cells expressing the top 12 most consistent aged HSC marker genes

(Supplementary Fig. 6). We observed age-associated upregulation of HSC-specific genes including *Sult1a1* and *Nupr1* in addition to pan-haematopoietic upregulated genes including *Igkc*. We also observed downregulation of *Rpl35*, *Uba52*, *Tmsb10*, *Gpx1*, *Plac8* and *Cd34* (Fig. 4E, Supplementary Data 3). *Sult1a1* and *Nupr1* are highly restricted to the LT-HSC population and are indeed almost exclusively expressed in aged LT-HSCs (Fig. 4E–G), and this increase in expression was confirmed in FACS-purified HSCs (LSK Cd34<sup>-</sup> Cd48<sup>-</sup> Cd150<sup>+</sup>) from young and aged mice by qPCR (Fig. 4G). While *Nupr1* was not detected in the long-read data, the *Sult1a1-203* isoform was the predominant isoform detected, encoding a 188 aa variant of the *Sult1a1* protein.

A global comparison of the long-read data from young and aged mice showed an age-associated increase in the expression of noncoding transcripts, including IgV pseudogenes, lncRNAs and transcripts with retained introns (Supplementary Fig. 7). This is



**Fig. 4 Signatures of ageing in haematopoietic stem and progenitor cells.** **A** Frequency of cells per cluster between young and aged mice. **B, C** UMAP showing *Procr* expressing cells (HSCs) in young and aged mice, respectively. **D** Differential expression for each cluster, based on short-read data. Data are presented as a heatmap of fold-change values, where blue is upregulated with age and red is downregulated with age. **E** Differential expression in the HSC cluster. **F** Upregulation of aged HSC gene expression signature (from Svendsen et al.<sup>26</sup>) in cells from the aged HSC cluster. **G** qPCR confirmation of *Sult1a1* and *Nupr1* upregulation in individual aged phenotypic (LSK Cd48– Cd150+ Cd34–) HSCs (\*\*\*) indicates *p*-value < 0.001, \*\* indicates *p*-value < 0.005).

consistent with observations that an increased frequency of intron retention has been identified as a signature of ageing in fruitfly, mouse and human<sup>27–29</sup>.

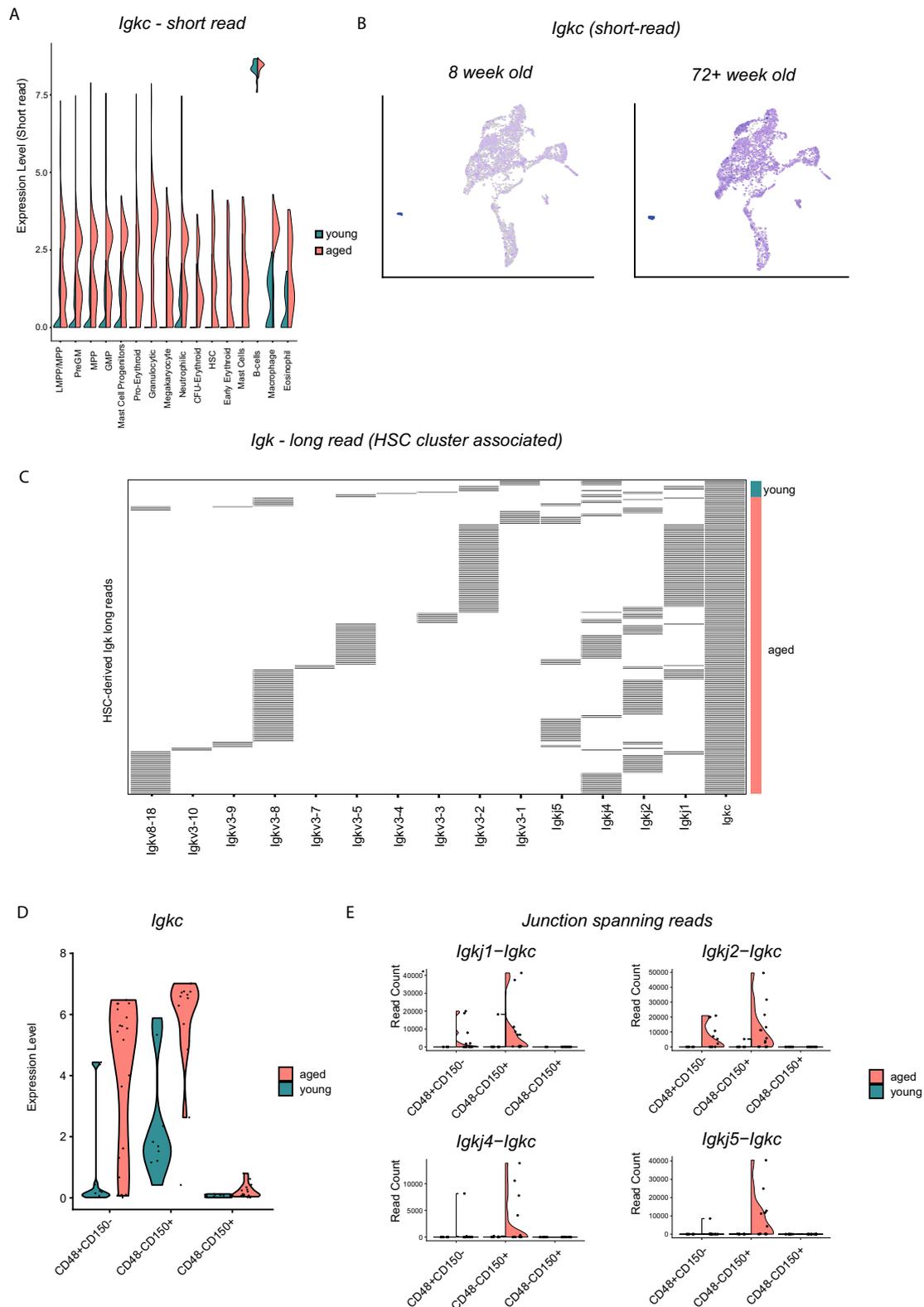
Following the observation that *Igkc* transcripts were upregulated throughout the myeloid progenitor populations and even transcriptionally phenotypic HSCs (Fig. 5A, B), we used the long-read data to determine that these molecules are VJ-recombined *Igkc* transcripts (Fig. 5C, Supplementary Fig. 8), and distinct from IgV pseudogenes (which account for only 1% of immunoglobulin reads sequenced).

To further investigate this, we confirmed the expression of *Igkc* in FACS-isolated pools of LSK Cd48– Cd150+ HSCs, LSK Cd48+ Cd150– cells (predominantly LMPPs) and LSK Cd48+ Cd150+ (early multipotent progenitors) (25 cells per pool) using Smart-seq2 and short-read Illumina sequencing. Here we observed upregulation in cells from aged mice, with slightly increased expression in phenotypic HSCs but more dramatic increases in the progenitor populations (Fig. 5D) with a similar pattern for reads spanning *Igkc* and J-unctions (Fig. 5E), although the detection of junction spanning reads using short-read sequencing is challenging—in contrast to the long-read sequences where the junction detection is unequivocal. A similar pattern was observed for the *Igh* heavy chain (Supplementary Fig. 9C, D). Representative FACS plots for sorted populations are shown in Supplementary Fig. 10.

As non-lymphoid immunoglobulin transcript expression could have arisen from contaminating cells or molecules, we investigated B-cell lineage genes in the short-read data (Supplementary Fig. 11), and Cd45 isoform expression in the long-read data (Supplementary Fig. 11G) confirmed that no B-cell markers were promiscuously expressed throughout the data, nor increase in Cd45r (typically associated with B-cells) detection in cDNA from the aged mice. This indicates that the immunoglobulin molecules detected in these experiments are highly likely to have originated from non-lymphoid committed cells, including myeloid progenitor populations and potentially even transcriptionally phenotypic HSCs.

**Discussion**

A complete understanding of transcriptional cellular heterogeneity requires an understanding of the isoform-level expression of genes. The integration of short- and long-read sequencing approaches for single-cell RNA-seq enables the annotation of cells and cell types with isoform-level expression information and thus reveals an additional layer of complexity in cellular transcriptional and functional phenotypes. Here, we integrate these approaches to annotate heterogeneous populations of haematopoietic cells with isoform-level information, revealing that many genes, including key regulators of haematopoiesis, are expressed



**Fig. 5 Expression of rearranged Immunoglobulin light chain Kappa (*Igk*) in aged stem and progenitor cells. **A** *Igk* expression in short-read profiling of Lin<sup>-</sup> cKit<sup>+</sup> haematopoietic subclusters, split by age. **B** UMAP projection of *Igk* expression in short-read profiling of Lin<sup>-</sup> cKit<sup>+</sup> haematopoietic cells, split by age. **C** Observation of full-length, VJ-recombined *Igk* transcripts in long-reads derived from the HSC cluster alone. Data are shown as a heatmap of long-reads (rows) with the presence of V and J regions (columns) indicated in black. **D** Expression of *Igk* in phenotypic haematopoietic and stem and progenitor cells from young and aged mice. **E** *Igk*-J region junctions in phenotypic haematopoietic stem and progenitor cells from young and aged mice.**

as multiple—often functionally distinct—isoforms. This approach also enables the integration of previously unannotated isoforms into reference transcriptomes to allow a more accurate annotation of cell-type-specific isoform expression and a better understanding of the contribution of isoform dynamics to cellular function.

Short-read analysis of a typical 10X Genomics experiment measures just 1–3000 genes per cell (~15,000 molecules per cell) representing about 10% of an average cell's transcriptome (assuming 200,000 mRNA molecules comprising 5–15,000 genes<sup>30</sup>, and thus ~10% of the cell's transcriptome is “read” in a standard 10X genomics experiment. Our long-read sequencing measures a median of 292 transcripts per cell (from a median of 411 reads per cell), which is approximately 1.5% of the transcripts present within a hypothetical cell. In the course of our and other<sup>12</sup> analyses of single-cell long-read data, many reads are discarded due to incomplete capture of the cell barcode, and thus, only ~30% of reads can be assigned to single cells, which is a significant limitation of the approach as it currently exists.

We would estimate that without further technical improvement, 100 times more sequencing would reveal the true extent of transcript diversity per cell. This is presently not practical, and methods for more detailed and focussed analysis of smaller numbers of (e.g., FACS isolated) cells are required. In the present case, we use this approach to annotate cell populations with isoform information and to inform single-cell resolution analysis by qPCR of a single gene, but methods which capture global isoform diversity at single-cell resolution would be highly desirable.

With this approach, we demonstrate that over half the genes detected in a standard single-cell analysis of the haematopoietic hierarchy are actually present as multiple isoforms, including transcription factors and transmembrane receptors with key roles in haematopoietic differentiation. Complexes such as Lmo2/Gata2/Ldb1/Tal1 may be readily subject to regulation through isoform expression, with distinct protein isoforms having distinct roles within the overall complex. The finding that many of the transcription factors comprising established networks are present as multiple isoforms suggests an additional layer of complexity that should be taken into account, for example, when building gene regulatory networks from single-cell expression data.

Similarly, it may be that alternative splicing events underlie how phenotypically highly similar cells respond differently to extrinsic signals. We detected multiple isoforms of the transcript encoding the ThPO receptor *Mpl*, a key regulator of stem cell maintenance and megakaryocyte differentiation, and identified that these isoforms, with functionally distinct protein products, are heterogeneously expressed in phenotypic HSCs. We demonstrate that single HSCs often express more than one isoform of *Mpl*, with each isoform potentially having antagonistic responses to ligand binding. Functional heterogeneity in isoform expression/co-expression could potentially have a critical role in how individual HSCs respond to ThPO, and it would be of considerable interest to understand the role this might play in both steady-state and stress haematopoiesis.

The integrated analysis of short- and long-read data further allowed the identification of multiple signatures of haematopoietic ageing. Here, we have used 7-week-old mice as young and >72-week-old mice as aged, which most likely represents an early rather than an extreme ageing phenotype. Consistent with previous reports, we observe an expansion of a transcriptionally phenotypic HSC population with ageing, while other cell populations remain largely unchanged, and using a compilation of existing studies of HSC ageing, we confirm that the aged cells here express a signature of HSC ageing and that this signature is restricted to HSCs within the LK population. Although *Sult1a1*

and *Nupr1* have previously been observed to be upregulated in aged HSCs, we here demonstrate that their expression is exclusive to HSCs and not seen elsewhere in the haematopoietic hierarchy. The genes encoding *Nupr1* and *Sult1a1* are co-localised in a 50 kb region of chromosome 7, and their elevated expression has previously been shown to be associated with an age-related increase in the H3K4me3 chromatin mark in HSCs<sup>31</sup>. *Sult1a1* encodes for a sulfotransferase which acts on substrates including hormones and neurotransmitters, and *Nupr1* has a regulatory role in cell proliferation and apoptosis, but neither has a described functional role in haematopoiesis<sup>31,32</sup>.

Although there was a clear transcriptional response to ageing, both in HSCs and the wider haematopoietic hierarchy, the usage of isoform was remarkably stable at the resolution analysed here. We observed no specific changes in isoform expression between young and aged mice but the long-read data did show marginally increased expression of lncRNAs. Highly tissue-specific changes in lncRNA expression in aged human tissues have been reported<sup>33</sup>, and abnormally elevated expression of some lncRNAs seems to relate to telomere shortening and senescence<sup>34</sup>. Similarly, an overall increase in intron retention was observed; however, neither signature was specific to any particular gene or lncRNA, indicative of global dysregulation rather than a specific functional response.

The combined use of short- and long-read sequencing enabled the detection of *Igkc* upregulation in aged haematopoietic cells, including transcriptionally phenotypic myeloid progenitors and even HSCs. This was highly unexpected but perhaps not without precedent. Previous work has shown that unrecombined *Igkc* transcripts are expressed in aged LT-HSCs possibly as a result of epigenetic dysregulation<sup>35</sup>, and a recent scRNA-seq study of HSC ageing also detected *Igkc* as the most upregulated transcript with ageing<sup>36</sup> (Supplementary Fig. 12A). Interestingly, a reanalysis of the Chambers et al.<sup>35</sup> data in the ageing database described in Svendsen et al.<sup>26</sup> showed that many *Igkv* loci were upregulated in aged HSCs (Supplementary Fig. 12B), and *Igkc* and *Igvl* expression was also observed to be upregulated in microarray analysis of gene expression in vWF+ platelet-biased HSCs when compared with vWF- HSCs<sup>37</sup> (Supplementary Fig. 12C). Thus, there is precedent for the aberrant expression of *Igkc* in ageing HSCs.

The expression of these transcripts is dependent on genomic rearrangements, typically seen in committed lymphoid progenitors downstream of HSCs, but there is an increasing body of evidence that the expression of recombined Ig molecules in non-lymphoid cells is possible<sup>38–40</sup>, including human CD34+ cord blood stem and progenitor cells<sup>41</sup> and acute myeloid leukaemia<sup>42</sup>. Indeed, the expression of IgG has been confirmed at both the transcriptional and protein levels in human epithelial cancer cells<sup>43–46</sup>, with a restricted variable region repertoire which exerts profound pro-tumorigenic effects<sup>47</sup>.

Further studies of the causative or associated genomic and epigenetic events underlying this aberrant expression will be essential to confirm the genomic origin of immunoglobulin expression in non-lymphoid cell types, including multipotent stem and progenitor cell populations. Currently available methods to sequence immunoglobulin recombination at the genome level regard “low input” as 100,000–200,000 cells (Chovanec et al.<sup>48</sup>), making rare cell-type specific measurements immensely challenging. At present, methods that enable high-confidence single-cell *genomic* recombination analysis—using targeted, high-fidelity amplification and sequencing—have not been described. These approaches, ideally coupled with capture of the cell's functional or transcriptional phenotype, will be essential to explore the origins of non-lymphoid recombinant immunoglobulin expression and, if confirmed, address the mechanisms by which expression of these recombinant molecules occurs.

Overall, our study not only provides a comprehensive picture of gene and isoform expression in a variety of haematopoietic cell types but also offers novel insights into transcriptional signatures of ageing. Our results highlight the need to further characterise isoform diversity at the single-cell level and to build an isoform atlas for different cell types to reveal the full extent of transcriptional heterogeneity in development, ageing and disease. With continued improvements in the throughput of long-read sequencing library preparation and platforms and the development of methods targeting specific cells or transcripts, we envision that long-read sequencing will enable detailed characterisation of the total landscape of isoform diversity at the single-cell level at a scale comparable to current short-read based methods. This will be applicable to large-scale atlasing studies of cellular heterogeneity and in haematological malignancies where mutations in splicing factors are common<sup>49</sup>. Furthermore, advances in mass spectrometry-based single-cell proteomic analysis<sup>50</sup>, combined with long-read transcript discovery, will become a critical tool enabling the confirmation of protein isoform expression.

## Methods

**Stem and progenitor cell isolation.** All animal work in this study was carried out in accordance with regulations set by the United Kingdom Home Office and the Animal Scientific Procedures Act of 1986. Bone marrow was isolated from the spine, femora, tibiae, and ilia of 8 weeks and 72 weeks old C57BL/6J mice. Red blood cell depletion was performed with ammonium chloride lysis (STEMCELL Technologies), and lineage-negative cells were isolated using the EasySep Mouse Hematopoietic Progenitor Cell Isolation Kit (STEMCELL Technologies).

The lineage-depleted cells were stained with the following fluorophore-conjugated monoclonal antibodies: Cd105-PE, clone MJ7/18, Miltenyi; Cd4-Vioblu, clone REA604, Miltenyi; Cd11b-Vioblu, clone REA592, Miltenyi; Cd117-Pe Vio770, clone REA791, Miltenyi; Cd8a-Vioblu, clone 53-6.7, Miltenyi; Cd50-Vioblu, clone REA421, Miltenyi; Cd45R-Vioblu, clone REA755, Miltenyi; GR1-Vioblu, clone REA810, Miltenyi; Sca-APC, clone REA422, Miltenyi; Cd48-APC Cy7, clone HM48-1, Miltenyi; Cd150-BV510, clone TC15-12F12, Cd34-PeCy5, MEC147, Miltenyi. Approximately 10,000 LK (Lin<sup>-</sup>, Cd117<sup>+</sup>) cells per sample were sorted using the BD FACSMelody cell sorter (BD Biosciences, San Jose, California) into 1× PBS containing 4% BSA. For low-input and single-cell qPCR, a pool of 25 cells and single LSK Cd150+ Cd48- Cd34- HSCs respectively were sorted directly into Smart-seq2 lysis buffer<sup>51</sup>.

**Sequencing of single-cell cDNA libraries.** Sorted cells were processed by 3' end single-cell RNA-Seq using the 10X Genomics Chromium (V2 Kit) according to the manufacturer's protocol (10X Genomics, Pleasanton, CA) with an increase to 16 cycles for the cDNA PCR amplification. Six libraries were prepared, each from cells from an individual mouse—three aged and three young. Libraries were sequenced on a NextSeq 500 or NovaSeq 6000 (Illumina, San Diego) in paired-end, single index mode as per the 10X Genomics recommended metrics.

Raw Illumina sequencing data were analysed with the 10X Genomics Cell Ranger pipeline (version 3.0.2) to obtain a single-cell expression matrix object. Subsequent analysis was performed in R using Seurat version 3<sup>16</sup>. Cells showing gene counts lower than 1000 and a mitochondrial gene expression percentage higher than 5% were excluded from further analysis. Within Seurat, data were normalised using `NormalizeData` (`normalisation.method = "LogNormalize"`, `scale.factor = 10000`) and data from multiple samples were merged using the `FindIntegrationAnchors` and `IntegrateData` commands.

**Pacific Biosciences Sequel sequencing of single-cell cDNA libraries.** Libraries compatible with the Pacific Biosciences Sequel/Sequel II systems were prepared from 800 ng input cDNA generated from each of the six individual mice following the “no size selection” Iso-Seq library preparation method according to the manufacturer's instructions (IsoSeq Template Preparation for Sequel System V05), with the following modifications: the elution incubation time during AMPure beads purification was increased to 10 min and the second AMPure bead purification step, following the exonuclease reaction, was omitted to optimise library concentration. In total, six libraries were sequenced, each originating from the LK population of an individual mouse.

**Pacific Biosciences long-read analysis.** Circular Consensus reads (CCS) were generated using the following parameters: maximum subread length 20,000, minimum subread length 50 and minimum number of passes 3. Reads with identified polyA or polydT were demultiplexed using `bbduk` <https://sourceforge>.

`net/projects/bbmap/`) ( $k = 16$ ,  $hdist = 3$ ) using the 10X genomics barcodes identified from the short-read analysis. Long reads were mapped to the mouse genome (mm10) using Minimap 2 (v2.17) and to the gencode (vM19) transcriptome.

Novel exons were identified by investigating the alignment of the reads to the transcriptome identifying inserts of at least 21 nucleotides located at exon junctions. To confirm the existence of these exons, the alignment of the reads to the genome was parsed, with exonic sequences located within the previously identified intron and supported by at least two reads retained for further analysis. We further removed any exonic regions overlapping RefSeq annotations (GRCm38, last accessed February 19 2019). To identify reads supporting V(D)J recombination events, we used `IgBlast` v1.14.0<sup>52</sup> using default parameters (`-min_V_length 9 -min_J_length 0 -min_D_match 5 -D_penalty -2 -J_penalty -2`).

**Custom transcriptome annotation.** We took advantage of the long-read PacBio data to annotate and explore alternative splicing events using TALON<sup>53</sup>. We identified a total of 11,013 novel transcripts supported by at least five reads and identified in three or more of the samples (44,993 novel transcripts if at least two reads and two samples). Those annotations also enabled the identification of 910 novel cassette exons, and 4118 and 3465 novel alternative 5' and 3' splicing sites, respectively. We also identified 143 novel junctions between previously annotated splice sites. We investigated the impact of the novel splicing events on the coding potential of the transcripts. Among the 10,576 novel transcripts arising from protein-coding genes, a total of 6747 transcripts were identified as coding, whereas 3830 were deemed noncoding (Supplementary Data 2).

TALON v.5.0 was used to construct a custom transcriptome annotation. First, 11 genome-aligned sam files were passed to `TranscriptClean` v.2.0.2 [2] for correction of read microindels (<5 bp) and mismatches, though any non-canonical splice junctions were retained for downstream analyses, and clean-up was not variant-aware. Internal priming artefacts are a known issue with oligo-dT selection methods [3], and this was assessed using a T-window size of 20 bp (equivalent to the primer T sequence), with reads labelled accordingly using TALON functions. Read annotation was performed using the *Mus musculus* Gencode v.M24 reference annotation `gtf` and `GRCm38.p6` genome with minimum alignment identity = 0.9 and coverage = 0.8 on the samples. Identified transcripts were subsequently filtered using a minimum count threshold of  $N = 5$  reads in  $K = 3$  samples. Thresholds were selected to balance sensitivity with accuracy. An updated annotation was produced using this filtered set of transcripts. Novel antisense transcripts that perfectly matched existing gene models were removed as unreliable mappings. The TALON custom `gtf` contains only features detected with reads present in the dataset, so a complete custom transcriptome annotation was compiled by merging the Gencode v.M24 reference and custom TALON `gtfs`. Transcript coding potential was assessed through frame preservation and applying `CPAT2.0`<sup>54</sup> using Gencode long noncoding RNAs (<https://www.gencodegenes.org/>) and CDS as training sets.

**Data integration in Seurat.** In order to reduce the batch effect, count matrices produced by short-read sequencing for individual libraries were combined in Seurat using `FindIntegrationAnchors` and `IntegrateData` functions (`dims = 1:20`). Illumina and PacBio reads were integrated into Seurat using the `CreateAssayObject` command to add the long-read data to an existing Seurat object already containing the short-read data. This links the demultiplexed long-reads with the short-read data through the cell barcodes present in both.

**Single-cell and low-input RNA-seq and qPCR.** Amplified cDNA was generated from sorted cells using the Smart-seq2 protocol<sup>45</sup>. This material was then used as input for qPCR reactions using assays targeting *Sult1a1* and *Nupr1*. Low-volume qPCR reactions were set up using the Mosquito HV instrument (STP Labtech) and analysed using a LightCycler (Roche). For relative abundance, data are presented as expression relative to a housekeeping gene. For single-cell isoform junction detection PCRs, reactions were performed as above using junction-spanning primers. Data are presented as a presence/absence heatmap, where analyses with Ct values < 30 were considered to be expressed.

For low-input RNA-seq, sequencing libraries were prepared from cDNA using the Nextera protocol (Illumina) at a reduced volume using the Mosquito HV instrument (STP Labtech). Libraries were pooled and sequenced on an Illumina NovaSeq 6000 SP Lane.

**Statistics and reproducibility.** Single-cell long-read experiments were carried out on samples from each of three young (8 weeks) and three aged (72+ weeks) mice. qPCR data were generated from samples from six individual young and six individual aged mice. Data are presented as expression relative to *B2m* in the same sample ( $-1/\Delta Ct$ ), and statistical significance was measured using an unpaired *t*-test comparing normalised expression levels in young and aged stem cells. Low-input RNA-seq data (for immunoglobulin junction detection) were generated from three young and five aged mice.

**Reporting summary.** Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

Sequencing data can be accessed on the NCBI-GEO archive, accession number GSE166709. Source data for the graphs and charts are given in Supplementary Data 4, and any remaining information can be obtained from the corresponding authors upon reasonable request.

Received: 13 March 2023; Accepted: 12 May 2023;

Published online: 24 May 2023

## References

- Svensson, V., Vento-Tormo, R. & Teichmann, S. A. Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* **13**, 599–604 (2018).
- Dahlin, J. S. et al. A single-cell hematopoietic landscape resolves 8 lineage trajectories and defects in Kit mutant mice. *Blood* **131**, e1–e11 (2018).
- Pijuan-Sala, B. et al. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495 (2019).
- Tabula Muris Consortium et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018).
- Laurenti, E. & Göttgens, B. From haematopoietic stem cells to complex differentiation landscapes. *Nature* **553**, 418–426 (2018).
- Frankish, A. et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).
- Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**, 1413–1415 (2008).
- Marti-Solano, M. et al. Combinatorial expression of GPCR isoforms affects signalling and drug responses. *Nature* **587**, 650–656 (2020).
- Goldstein, O. et al. Mapping whole-transcriptome splicing in mouse hematopoietic stem cells. *Stem Cell Rep.* **8**, 163–176 (2017).
- Chen, L. et al. Transcriptional diversity during lineage commitment of human blood progenitors. *Science* **345**, 1251033 (2014).
- Hardwick, S. A., Joglekar, A., Flicek, P., Frankish, A. & Tilgner, H. U. Getting the entire message: progress in isoform sequencing. *Front. Genet.* **10**, 709 (2019).
- Gupta, I. et al. Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. *Nat. Biotechnol.* **36**, 1197–1202 (2018).
- Philpott, M. et al. Nanopore sequencing of single-cell transcriptomes with scCOLOR-seq. *Nat. Biotechnol.* **39**, 1517–1520 (2021).
- Pronk, C. J. H. et al. Elucidation of the phenotypic, functional, and molecular topography of a myeloerythroid progenitor cell hierarchy. *Cell Stem Cell* **1**, 428–442 (2007).
- Balazs, A. B., Fabian, A. J., Esmon, C. T. & Mulligan, R. C. Endothelial protein C receptor (CD201) explicitly identifies hematopoietic stem cells in murine bone marrow. *Blood* **107**, 2317–2321 (2006).
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
- Hamey, F. K. et al. Reconstructing blood stem cell regulatory network models from single-cell molecular profiles. *Proc. Natl Acad. Sci. USA* **114**, 5822–5829 (2017).
- Sun, W. et al. Homo-binding character of LMO2 isoforms and their both synergic and antagonistic functions in regulating hematopoietic-related target genes. *J. Biomed. Sci.* **17**, 22 (2010).
- Rodrigues, N. P., Tipping, A. J., Wang, Z. & Enver, T. GATA-2 mediated regulation of normal hematopoietic stem/progenitor cell function, myelodysplasia and myeloid leukemia. *Int. J. Biochem. Cell Biol.* **44**, 457–460 (2012).
- Zeddies, S. et al. MEIS1 regulates early erythroid and megakaryocytic cell fate. *Haematologica* **99**, 1555–1564 (2014).
- Montero, J. C., López-Pérez, R., San Miguel, J. F. & Pandiella, A. Expression of c-Kit isoforms in multiple myeloma: differences in signaling and drug sensitivity. *Haematologica* **93**, 851–859 (2008).
- Lebedev, T. D. et al. Two receptors, two isoforms, two cancers: comprehensive analysis of KIT and TrkA expression in neuroblastoma and acute myeloid leukemia. *Front. Oncol.* **9**, 1046 (2019).
- Young, S. M., Cambareri, A. C., Odell, A., Geary, S. M. & Ashman, L. K. Early myeloid cells expressing c-KIT isoforms differ in signal transduction, survival and chemotactic responses to Stem Cell Factor. *Cell. Signal.* **19**, 2572–2581 (2007).
- Coers, J., Ranft, C. & Skoda, R. C. A truncated isoform of c-Mpl with an essential C-terminal peptide targets the full-length receptor for degradation. *J. Biol. Chem.* **279**, 36397–36404 (2004).
- Dykstra, B. et al. Long-term propagation of distinct hematopoietic differentiation programs in vivo. *Cell Stem Cell* **1**, 218–229 (2007).
- Svensden, A. F. et al. A comprehensive transcriptome signature of murine hematopoietic stem cell aging. *Blood* **138**, 439–451 (2021).
- Adusumalli, S., Ngjan, Z.-K., Lin, W.-Q., Benoukraf, T. & Ong, C.-T. Increased intron retention is a post-transcriptional signature associated with progressive aging and Alzheimer's disease. *Aging Cell* **18**, e12928 (2019).
- Ong, C.-T. & Adusumalli, S. Increased intron retention is linked to Alzheimer's disease. *Neural Regen. Res.* **15**, 259–260 (2020).
- Mariotti, M., Kerepesi, C., Oliveros, W., Mele, M. & Gladyshev, V. N. Deterioration of the human transcriptome with age due to increasing intron retention and spurious splicing. Preprint at *bioRxiv* <https://doi.org/10.1101/2022.03.14.484341> (2022).
- Shapiro, E., Biezuner, T. & Linnarsson, S. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* **14**, 618–630 (2013).
- Sun, D. et al. Epigenomic profiling of young and aged HSCs reveals concerted changes during aging that reinforce self-renewal. *Cell Stem Cell* **14**, 673–688 (2014).
- Gazit, R. et al. Fgd5 identifies hematopoietic stem cells in the murine bone marrow. *J. Exp. Med.* **211**, 1315–1331 (2014).
- Marttila, S., Chatsirisupachai, K., Palmer, D. & de Magalhães, J. P. Ageing-associated changes in the expression of lncRNAs in human tissues reflect a transcriptional modulation in ageing pathways. *Mech. Ageing Dev.* **185**, 111177 (2020).
- Jin, L., Song, Q., Zhang, W., Geng, B. & Cai, J. Roles of long noncoding RNAs in aging and aging complications. *Biochim. Biophys. Acta Mol. Basis Dis.* **1865**, 1763–1771 (2019).
- Chambers, S. M. et al. Aging hematopoietic stem cells decline in function and exhibit epigenetic dysregulation. *PLoS Biol.* **5**, e201 (2007).
- Hérault, L. et al. Single-cell RNA-seq reveals a concomitant delay in differentiation and cell cycle of aged hematopoietic stem cells. *BMC Biol.* **19**, 19 (2021).
- Sanjuan-Pla, A. et al. Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature* **502**, 232–236 (2013).
- Chen, Z., Qiu, X. & Gu, J. Immunoglobulin expression in non-lymphoid lineage and neoplastic cells. *Am. J. Pathol.* **174**, 1139–1148 (2009).
- Huang, J. et al. Rearrangement and expression of the immunoglobulin  $\mu$ -chain gene in human myeloid cells. *Cell. Mol. Immunol.* **11**, 94–104 (2014).
- Fuchs, T. et al. Expression of combinatorial immunoglobulins in macrophages in the tumor microenvironment. *PLoS ONE* **13**, e0204108 (2018).
- Liu, J. et al. Immunoglobulin gene expression in umbilical cord blood-derived CD34<sup>+</sup> hematopoietic stem/progenitor cells. *Gene* **575**, 108–117 (2016).
- Qiu, X. et al. Immunoglobulin gamma heavy chain gene with somatic hypermutation is frequently expressed in acute myeloid leukemia. *Leukemia* **27**, 92–99 (2013).
- Zhao, J. et al. Current insights into the expression and functions of tumor-derived immunoglobulins. *Cell Death Discov.* **7**, 148 (2021).
- Babbage, G. et al. Immunoglobulin heavy chain locus events and expression of activation-induced cytidine deaminase in epithelial breast cancer cell lines. *Cancer Res.* **66**, 3996–4000 (2006).
- Qiu, X. et al. Human epithelial cancers secrete immunoglobulin G with unidentified specificity to promote growth and survival of tumor cells. *Cancer Res.* **63**, 6488–6495 (2003).
- Zheng, H. et al. Immunoglobulin alpha heavy chain derived from human epithelial cancer cells promotes the access of S phase and growth of cancer cells. *Cell Biol. Int.* **31**, 82–87 (2007).
- Cui, M. et al. Immunoglobulin expression in cancer cells and its critical roles in tumorigenesis. *Front. Immunol.* **12**, 613530 (2021).
- Chovanec, P. et al. Unbiased quantification of immunoglobulin diversity at the DNA level with VDJ-seq. *Nat. Protoc.* **13**, 1232–1252 (2018).
- Saez, B., Walter, M. J. & Graubert, T. A. Splicing factor gene mutations in hematologic malignancies. *Blood* **129**, 1260–1269 (2017).
- Specht, H. et al. Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCOPE2. *Genome Biol.* **22**, 50 (2021).
- Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
- Ye, J., Ma, N., Madden, T. L. & Ostell, J. M. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res.* **41**, W34–W40 (2013).
- Wyman, D. et al. A technology-agnostic long-read analysis pipeline for transcriptome discovery and quantification. Preprint at *bioRxiv* <https://doi.org/10.1101/672931> (2019).
- Wang, L. et al. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res.* **41**, e74 (2013).

## Acknowledgements

I.C.M. is supported by a BBSRC New Investigator Grant [BB/P022073/1] and the BBSRC National Capability in Genomics and Single Cell Analysis at Earlham Institute [BB/CCG1720/1]. W.H., L.M. and D.W. are supported by the BBSRC Core Strategic Programme Grant [BB/P016774/1], and W.H. by a UK Medical Research Council [MR/P026028/1] award. S.R. is funded by the Rosetrees Trust, The Big C and the Medical Research Council [MR/T02934X/1]. Next-generation sequencing was delivered via the BBSRC National Capability in Genomics and Single Cell Analysis [BB/CCG1720/1] at Earlham Institute.

## Author contributions

L.M. and I.C.M. designed the experiments. L.M., S.A.R., A.S. and I.C.M. performed the experiments. L.M., V.U., D.W., W.H. and I.C.M. analysed data. L.M., W.H. and I.C.M. wrote the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-023-04936-6>.

**Correspondence** and requests for materials should be addressed to Laura Mincarelli, Wilfried Haerty or Iain C. Macaulay.

**Peer review information** This manuscript has been previously reviewed in another Nature Portfolio journal. The manuscript was considered suitable for publication without further review at *Communications Biology*. *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary handling editors: Anam Akhtar and Joao Manuel de Sousa Valente.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023