Single-cell isoform RNA sequencing (ScISOr-Seq) across thousands of cells reveals isoforms of cerebellar cell types.

Ishaan Gupta[1,*], Paul G Collier[1,*], Bettina Haase[2], Ahmed Mahfouz[1,3,4], Anoushka Joglekar[1], Taylor Floyd[1], Frank Koopmans[5], Ben Barres[6,&], August B Smit[5], Steven Sloan[6,] Wenjie Luo[7], Olivier Fedrigo[2], M Elizabeth Ross[1], Hagen U Tilgner[1,+]

[1] Brain and Mind Research Institute and Center for Neurogenetics, Weill Cornell Medicine, 413 east 69th Street, New York, NY 10021
[2] The Rockefeller University, 1230 York Avenue, New York, NY, 10065
[3] Leiden Computational Biology Center, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden
[4] Delft bioinformatics Lab, Delft University of Technology, van Mourik Broekmanweg 6, 2628 XE Delft
[5] Dept. Molecular and Celullar Neurobiology, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University, Amsterdam, The Netherlands
[6] Department of Neurobiology, Stanford University, 299 Campus Dr, Stanford, CA 94305-5125
[7] Brain and Mind Research Institute and Appel Alzheimer's research institute, Weill Cornell Medicine

*equal contribution
+ corresponding author
& author, who tragically passed away and could not see the result of this work.

Abstract

Full-length isoform sequencing has advanced our knowledge of isoform biology[1–11]. However, apart from applying full-length isoform sequencing to very few single cells[12,13], isoform sequencing has been limited to bulk tissue, cell lines, or sorted cells. Single splicing events have been described for <=200 single cells with great statistical success[14,15], but these methods do not describe full-length mRNAs. Single cell short-read 3' sequencing has allowed identification of many cell sub-types[16–23], but full-length isoforms for these cell types have not been profiled. Using

our new method of single-cell-isoform-RNA-sequencing (ScISOr-Seq) we determine isoform-expression in thousands of individual cells from a heterogeneous bulk tissue (cerebellum), without specific antibody-fluorescence activated cell sorting. We elucidate isoform usage in high-level cell types such as neurons, astrocytes and microglia and finer sub-types, such as Purkinje cells and Granule cells, including the combination patterns of distant splice sites[6–9,24,25], which for individual molecules requires long reads. We produce an enhanced genome annotation revealing cell-type specific expression of known and 16,872 novel (with respect to mouse Gencode version 10) isoforms (see isoformatlas.com).

ScISOr-Seq describes isoforms from >1,000 single cells from bulk tissue without cell sorting by leveraging two technologies in three steps: In step one, we employ microfluidics to produce amplified full-length cDNAs barcoded for their cell of origin. This cDNA is split into two pools: one pool for 3' sequencing to measure gene expression (step 2) and another pool for long-read sequencing and isoform expression (step 3). In step two, short-read 3'-sequencing provides molecular counts for each gene and cell, which allows clustering cells and assigning a cell type using cell-type specific markers. In step three, an aliquot of the same cDNAs (each barcoded for the individual cell of origin) is sequenced using Pacific Biosciences ("PacBio")[1,2,4,5,26] or Oxford Nanopore[3]. Since these long reads carry the single-cell barcodes identified in step two, one can determine the individual cell from which each long read originates. Since most single cells are assigned to a named cluster, we can also assign the cell's cluster name (e.g. "Purkinje cell" or "astrocyte") to the long read in question (Fig 1A) – without losing the cell of origin of each long read.

Results

Detection of cell types

We apply ScISOr-Seq to describe cell-type specific isoforms in mouse cerebellum at postnatal day 1 (P1). We sequence a mean of 17,885 reads per cell (as given by 10xGenomics' summary statistics). After filtering cells and considering only reads confidently mapped to genes, we have 3,875 unique molecular identifiers (UMIs) and 1,448 genes per cell during 3'end sequencing. In step 2, 6,627 cells were clustered into 17 groups (Figs. 1B,D). High expression of well-established cell-type specific markers identifies many clusters as cell types: High expression of *Pdgfra*, *Olig1* and *Olig2* identified a cluster of oligodendrocyte precursors (OPCs, Fig 1B,C). *Clu* and *Apoe* identified two clusters of astrocytes and *Gdf10*[27,28] identified a cluster of Bergmann Glia (BG). We also identified three large clusters of neuronal subtypes: namely (i) cells with high expression of *Neurod1* and *Ccdn2*, which we refer to as external granular layer (EGL) cells in several stages of differentiation. These give rise primarily to granule neurons that migrate into the internal granular layer (IGL) over the first weeks of mouse postnatal development; (ii) Purkinje-like (marked by *Pcp4, Gad1* and *Gad2*) in the Purkinje cell layer (PCL) and (iii) other neurons known to be present in the deep cerebellar nuclei and internal granular layer (collectively referred to as "IGL" heareafter), in which cells display high expression of *Pnoc*, *Snhg11*, *Tcf7l2*, *Gad1*, *Gad2* and *Lhx9*. This cluster was clearly composed of at least two clusters: One expressing *Gad1* and *Gad2* and the other expressing *Lhx9* and *Tcf7l2* (Figure 1B). These cell-type specific expression patterns exhibit specific anatomical localization within the developing cerebellum (Figure 1C, Allen Brain Atlas). Three further clusters expressed genes associated with neural progenitor cells: one expressing *Ccnd2* (which is highly expressed in the postnatal EGL), another

Atoh1 (glutamatergic neuron precursors from the rhombic lip and EGL) and the third *Ptf1a* (GABAergic neuron precursors from the ventricular zone) (Fig 1B). We could also identify other cell populations, such as microglia, which highly expressed well-known myeloid-associated genes (e.g. *C1qa*, *C1qb*, *C1qc* and *Tmem119)*. Alongside neuroglial subtypes and progenitors, we furthermore identified clusters expressing genes specific to endothelial and circulatory-system cells. In summary, our clustering recapitulates a large proportion of cell types classically observed in P1 cerebellum. Overall EGL, IGL cells and astrocytic cells were the largest cluster and blood cells the smallest. Detected reads, short-read UMIs and genes per cell showed slight differences between cell types but were of similar orders of magnitude. Consistent with their large size, Purkinje cells had the highest number of read, UMI and gene counts, while blood cells had the lowest gene count.

Reliability and replication of cell-type detection

Sequencing of a second replicate (rep2) and within-replicate analysis showed that most distinct clusters were highly dissimilar to any other clusters in the same replicate. To assess stability of clusters, we tripled Illumina sequencing depth for rep2. In all clusters (with one exception) based on shallower sequencing depth, 95-100% of cells were still attributed to the same cluster, even with three-fold deeper sequencing. Analysis of comparability of marker gene between clusters of the two replicates using the Jaccard index identified highly similar clusters with one exception: The smallest cluster (blood cells) in replicate 1 (rep1), was missing in rep2. Cell-type abundance was reproducible between replicates and highly correlated (Pearson correlation = 0.91, correlation- test p-value = 4.5×10^{-5}).

Detection of single-cell barcodes in full-length cDNAs

We then employed 850ng of full-length cDNAs, tagged for their cell of origin, for isoform sequencing to generate ~5.2 million PacBio circular consensus reads ("CCS"). These CCS showed mean full passes per SMRT cell of 16-34 and thus favor a lower error rate compared to earlier ISO-Seq publications[1,4]. Since cellular barcodes are located close to the polyA-tail, we first searched for polyA-tails. Aiming at detecting polyT-sequences, even with a hypothetical 10% error rate, we located the first nine consecutive Ts ("T9") in the first 200bp of each read and of its reverse complement. 61.6% of CCS contained such a T9, broadly consistent with our previous estimation (67%)[1,4]. Reads with and without T9s showed similar lengths, apart from CCS <=200bp accumulating in non-T9 CCS. 1.4% of T9-CCS had a T9 in the read start and the complement's start. These may include chimeras, which were removed from further analysis, introduced during reverse transcription, PCR or blunt-end PacBio library preparation. In total, for 58.0% (compared to 74.0% for 10x-3'seq) of the polyA-tail-containing CCS, we identified a perfect-match 16mer cellular barcode (each corresponding to one of the 6,627 single cells) and therefore the exact single cell, in which the RNA isoform was transcribed. As a theoretical foundation, we determined for all 6,627 barcodes, the minimal editing distance to any other barcode: For 92.7% of barcodes, this minimal ("Levenshtein") distance was 3 or greater, and for the remaining barcodes it was two. This shows that for most barcodes, there is one specific error pattern of three errors that would lead to a wrongly identified cell. However, in most cases three random errors would only discard the read because none of the 6,627 known barcodes is detected. Both experiments and simulations show that our single-cell barcode-detection procedure is extremely specific. Overall, we detected a median of 270 long reads, 260 UMIs and

129 genes per single cell. 3.8% of UMIs are observed twice, with a theoretical prediction of 3.4% (Methods). 99.3% (6,581/6,627) of clustered cells were detected with CCS (Figure 2A-D). 97.4% (6,459/6,627) of clustered cells had >100 CCS (Figure 2D). Detected short-read and long-read UMIs per single cell correlated highly (Pearson correlation = 0.95, correlation-test $p < 2.2 \times 10^{-16}$, Figure 2E). Long-read statistics (reads, UMIs, genes, Figure 2F-H) per cell cluster mirrored those in short-reads, with lower long-read numbers.

ScISOr-Seq using Nanopore sequencing

Using 1µg of barcoded cDNA on a Nanopore MinIon, we searched for cellular barcodes in 2.3 million Nanopore reads[29]. We found lower relative numbers of Nanopore 1D reads with a T9, supposedly due to problems with homopolymers in Nanopore data[29]. However, ~31.4% (1D) and ~35.2% (passed 1D^2) of Nanopore reads have a 30bp window with >=25 Ts. Although the variation from the expected position in Nanopore reads is larger than for CCS (90bp vs. 3bp), accumulation around the expected position is observed and exact barcode matches reveal unique barcodes in 6.0% (43,948/732,590, 1D) and 32.7% (9,454/28,931, 1D^2). Therefore, we can expect ~50,000 cluster-specific long reads per MinIon flow cell. With each current MinIon flow cell requiring 1µg of cDNA, further PCR (with associated biases) is needed to carry out large-scale ScISOr-Seq on Nanopore, whereas the employed 16-cycle PCR is sufficient to run 20- 50 SMRTcells on PacBio yielding up to 5 million long reads assigned to single cells.

A cell-type resolved isoform annotation

We aligned PacBio CCS to the mouse genome[30] (version mm10) using STAR[31] and carried out mapping quality control as previously performed[1,4,6]. We analyzed

novel isoforms with respect to mouse Gencode version 10, as outlined previously[1,6,32] to produce a long- read enhanced and cell-type resolved annotation. We considered 10,691 unique novel (with respect to mouse Gencode version 10) isoforms that affected 4,859 genes. For these isoforms, we required all splice sites to be known in Gencode[33] (version 10) and each junction and internal exon to be either annotated or observed at least twice in ScISOr-Seq. The unique novel isoforms contain new exon-exon junctions linking previously known splice sites, such as the skipping of exons annotated as constitutive. Artifacts in next-generation sequencing have been demonstrated[34]. To assess whether the long-range 16-cycle PCR in ScISOr-Seq generates chimeric transcripts, we obtained 164 million 150bp-paired-end reads on bulk RNA from P1 cerebella only employing a 6-cycle short-range PCR after RNA fragmentation. Based on this experiment, we confirmed 91.6-97.6% of the novel ScISOr-Seq junctions across different cell types (Figure 3A). To reduce the influence of PCR artifacts on the enhanced annotation to a minimum and to allow for adding lowly expressed transcripts, we generated a final enhanced cell-type resolved annotation with strong 6-cycle-PCR short-read support. In this enhanced annotation, for each added isoform, each intron and internal exon was required to be annotated in Gencode or to be supported by two or more 6-cycle-PCR short reads, resulting in 16,872 isoforms for 6,927 genes (Figure 3B). For each of these isoforms we know the single cell of origin and therefore the cell type that produced this isoform. 42.8% (7,219/16,872) employed at least one splice site not annotated in Gencode. With respect to the UCSC[35] and RefSeq[36] annotations, 94.0% and 70.9% respectively of added isoforms were novel We performed ScISOr-Seq for rep2, albeit at a lower sequencing depth (6 SMRT cells, compared to 23 for rep1). New rep2-isoforms replicated in rep1 in 65.7%

(microglia) to 76.2% (NPCs)(Figure 3C) of the cases (irrespectively of the cell type they were observed in rep1). Given replication of an isoform in any cell type, cell-type specific replication of a rep2-isoform in the same cell type in rep1 reached 70-80% in larger clusters, but lower percentages in smaller clusters with dramatically fewer long reads (Figure 3D). To validate the correct calling of the individual cell of origin for each isoform, we performed immunopanning to specifically isolate microglia in P1 cerebella followed by short- read RNAseq. This data was compared to all isoforms originating from a single microglial cell (and then to isoforms of single cells belonging to other cell types). This confirmed the microglial origin of long-read junctions exclusively observed in microglial single-cell long reads as compared to junctions observed exclusively in non-microglial single-cell long reads (Figure 3E). Similarly, immunopanning for astrocytes, Bergmann glia (both marked by *Hepacam)* and OPCs (which are known to be enriched in *Hepacam*-sorting) and short-read sequencing showed the highest coverage for junctions observed exclusively in astrocyte, Bergmann Glia and OPC ScISOr-Seq isoforms This was more pronounced for junctions observed three or more times in ScISOr-Seq data in one cell type. These data suggest that junctions observed only in astrocytes, Bergmann Glia and OPCs are also expressed at a lower level in other cell types originating from the same stem cell.

Database of cell-type specific isoform expression in the cerebellum

We first looked at alternative splicing in *Tpm1* gene that is expressed across multiple cell types and is known to have extensive alternative splicing according to Gencode[33], UCSC[35] and RefSeq[36] annotations. This gene contains five alternatively spliced blocks of exons namely AS1-AS5 (observable in >=3 reads). AS1 and AS5

represent alternatively spliced blocks of single or multiple 5' and 3' exons along with the associated untranslated regions while AS2-AS4 represent single alternatively spliced exons within the coding region of the gene. We observed 4 novel isoforms (Figure 4, black and bold) of *Tpm1* as compared to the observed 21 isoforms according to the Gencode annotation. Out of these "Novel Isoform 4", where AS4 is spliced out while other alternatively spliced exons are included, was the major isoform expressed in Astrocytes with 10 UMIs. Out of the annotated transcripts, only the OPCs express the ENSMUST00000113685.9 (Figure 4, red) transcript with 15 UMIs which was also the most abundantly expressed isoform in OPCs. Other annotated transcripts ENSMUST00000113686.7 (Figure 4, orange) and ENSMUST00000113690.7 (Figure 4, green) were the most abundant isoforms in EGL and IGL respectively. In order to make this data accessible to the research community, we have created a fully searchable database (see isoformatlas.com) of isoforms for every gene showing their cell type of origin (as shown in Figure 4) and their single-cell of origin projected onto the TSNE-plot shown in Figure 1B.

Discussion

Brain disorders (e.g. Alzheimer's disease) are highly associated with risk genes including *MAPT* and *APOE*. Interestingly, these genes are expressed in multiple cell (sub-)types. Therefore, cell-type specific isoform expression is critical and may decipher the action of disease-associated SNPs. Here, we (i) describe isoform expression across heterogeneous cell types and (ii) enhance genome annotation with cell-type specific isoform expression. A drawback is that employing multiple deeply sequenced replicates is for now very expensive with long-reads, making precise quantification of abundance changes between cell types as outlined

in rMATS[37] more difficult. However, our full-length RNAs from single cells cover all single nucleotide polymorphisms in the coding region of mature RNA and may help attributing single cells to a specific individual[38] in pooled samples.

Acknowledgements:

This work was supported by start-up funds (Weill Cornell Medicine) and a Leon Levy Fellowship in Neuroscience to HUT. This work used the Genomics Resources Core Facility and owes special thanks to Dr. Jenny Xiang and Angela Wan.

Author contributions:

Devised the experiments: PGC, IG, SS, HUT. Performed experiments: PGC, BH, IG, SS, OF, WL; Devised analysis: IG, AS, HUT. Performed analysis: IG, AM, AJ, TF, FK, HUT; Discussed and interpreted results throughout the project: all authors; Wrote the paper: IG, HUT with inputs from all authors. Supervised research: BB, MER, HUT.

References:

1. Sharon, D., Tilgner, H., Grubert, F. & Snyder, M. A single-molecule long-read survey of the human transcriptome. *Nat. Biotechnol.* **31,** 1009–14 (2013).
2. Au, K. F. *et al.* Characterization of the human ESC transcriptome by hybrid sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **110,** E4821-30 (2013).
3. Oikonomopoulos, S., Wang, Y. C., Djambazian, H., Badescu, D. & Ragoussis, J. Benchmarking of the Oxford Nanopore MinION sequencing for quantitative and qualitative assessment of cDNA populations. *Sci. Rep.* **6,** 31602 (2016).
4. Tilgner, H., Grubert, F., Sharon, D. & Snyder, M. P. Defining a personal, allele-specific, and single-molecule long-read transcriptome. *Proc. Natl. Acad. Sci. U. S. A.* **111,** (2014).
5. Koren, S. *et al.* Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat. Biotechnol.* **30,** 693–700 (2012).
6. Tilgner, H. *et al.* Comprehensive transcriptome analysis using synthetic long-read sequencing reveals molecular co-association of distant splicing events. *Nat. Biotechnol.* **33,** 736–42 (2015).
7. Tilgner, H. *et al.* Microfluidic isoform sequencing shows widespread splicing coordination in the human transcriptome. *Genome Res.* gr.230516.117 (2017). doi:10.1101/gr.230516.117
8. Bolisetty, M. T., Rajadinakaran, G. & Graveley, B. R. Determining exon connectivity in complex mRNAs by nanopore sequencing. *Genome Biol.* **16,** 204 (2015).
9. Roy, C. K., Olson, S., Graveley, B. R., Zamore, P. D. & Moore, M. J. Assessing long-distance RNA sequence connectivity via RNA-templated DNA–DNA ligation. *Elife* **4,** (2015).
10. Treutlein, B., Gokce, O., Quake, S. R. & Südhof, T. C. Cartography of neurexin alternative splicing mapped by single-molecule long-read mRNA sequencing. *Proc. Natl. Acad. Sci.* **111,** E1291-9 (2014).
11. Schreiner, D. *et al.* Targeted combinatorial alternative splicing generates brain region-specific repertoires of neurexins. *Neuron* **84,** 386–98 (2014).
12. Karlsson, K. & Linnarsson, S. Single-cell mRNA isoform diversity in the mouse brain. *BMC Genomics* **18,** 126 (2017).
13. Byrne, A. *et al.* Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. *Nat. Commun.* **8,** 1–11 (2017).
14. Song, Y. *et al.* Single-Cell Alternative Splicing Analysis with Expedition Reveals Splicing Dynamics during Neuron Differentiation. *Mol. Cell* **67,** 148–161.e5 (2017).
15. Shalek, A. K. *et al.* Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498,** 236–40 (2013).
16. Zeisel, A. *et al.* Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science (80-.).* **347,** 1138–42 (2015).
17. Lake, B. B. *et al.* Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science (80-.).* **352,** 1586–1590 (2016).
18. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single

cells. *Nat. Commun.* **8,** 14049 (2017).
19. Darmanis, S. *et al.* A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl. Acad. Sci.* **112,** 201507125 (2015).
20. Molyneaux, B. J. *et al.* DeCoN: Genome-wide analysis of invivo transcriptional dynamics during pyramidal neuron fate selection in neocortex. *Neuron* **85,** 275–288 (2015).
21. Deng, Q., Ramsköld, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343,** 193–6 (2014).
22. Jaitin, D. A. *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343,** 776–9 (2014).
23. Pollen, A. A. *et al.* Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat. Biotechnol.* **32,** 1053–8 (2014).
24. Fededa, J. P. *et al.* A polar mechanism coordinates different regions of alternative splicing within a single gene. *Mol. Cell* **19,** 393–404 (2005).
25. Fagnani, M. *et al.* Functional coordination of alternative splicing in the mammalian central nervous system. *Genome Biol.* **8,** R108 (2007).
26. Eid, J. *et al.* Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* **323,** 133–138 (2009).
27. Mecklenburg, N. *et al.* Growth and differentiation factor 10 (Gdf10) is involved in bergmann glial cell development under Shh regulation. *Glia* **62,** 1713–1723 (2014).
28. Koirala, S. & Corfas, G. Identification of novel glial genes by single-cell transcriptional profiling of Bergmann glial cells from mouse cerebellum. *PLoS One* **5,** (2010).
29. Jain, M. *et al.* Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* **36,** 338–345 (2018).
30. Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420,** 520–562 (2002).
31. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
32. Tilgner, H. *et al.* Accurate identification and analysis of human mRNA isoforms using deep long read sequencing. *G3 (Bethesda).* **3,** 387–97 (2013).
33. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22,** 1760–74 (2012).
34. Mcmanus, C. J. *et al. Correction for McManus et al., Global analysis of trans-splicing in Drosophila*. *Proceedings of the National Academy of Sciences* **110,** (2013).
35. Karolchik, D. *et al.* The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res.* **42,** D764-70 (2014).
36. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44,** D733–D745 (2016).
37. Shen, S. *et al.* rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci.* **111,** E5593–E5601 (2014).
38. Kang, H. M. *et al.* Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* (2017). doi:10.1038/nbt.4042.

Figures

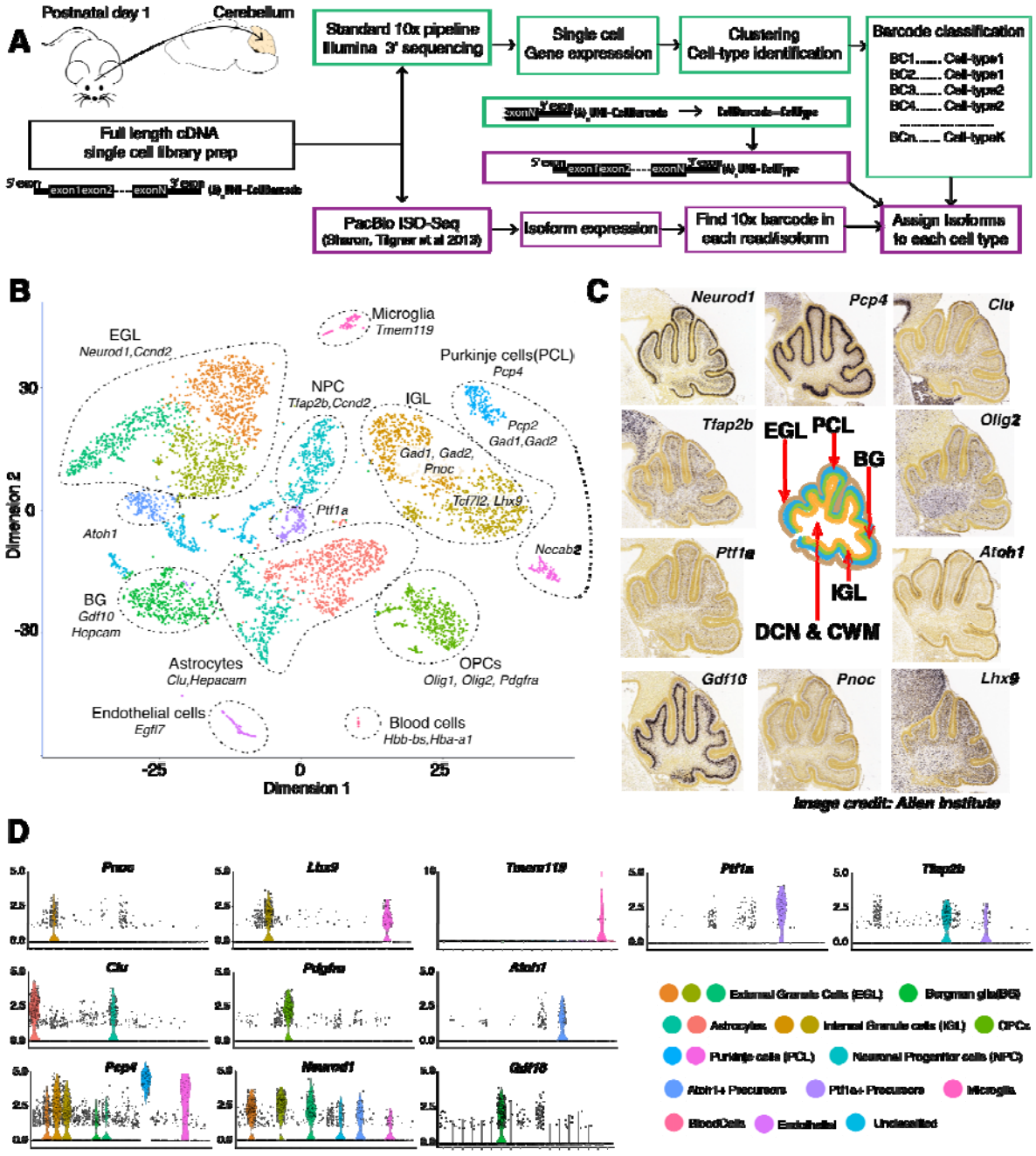

Figure 1: (A) Outline of our ScISOr-Seq approach. (B) TSNE-plot depicting cell clusters, marker genes and names given to clusters, including: Bergman glia (BG), External granule cell layer neurons (EGL), Internal granule cell layer and other neurons in the interior of the cerebellum (IGL), two clusters of Purkinje cell layer neurons (PCL), oligodendrocyte progenitor cells (OPCs), Atoh1+ neuronal progenitors, Ptf1a+ neuronal progenitors and other neuronal progenitors (NPCs)
(C) In-situ hybridization images from the Allen Brain Atlas depicting expression of marker genes in specific layers. (D) Expression patterns of selected marker genes across cell types.

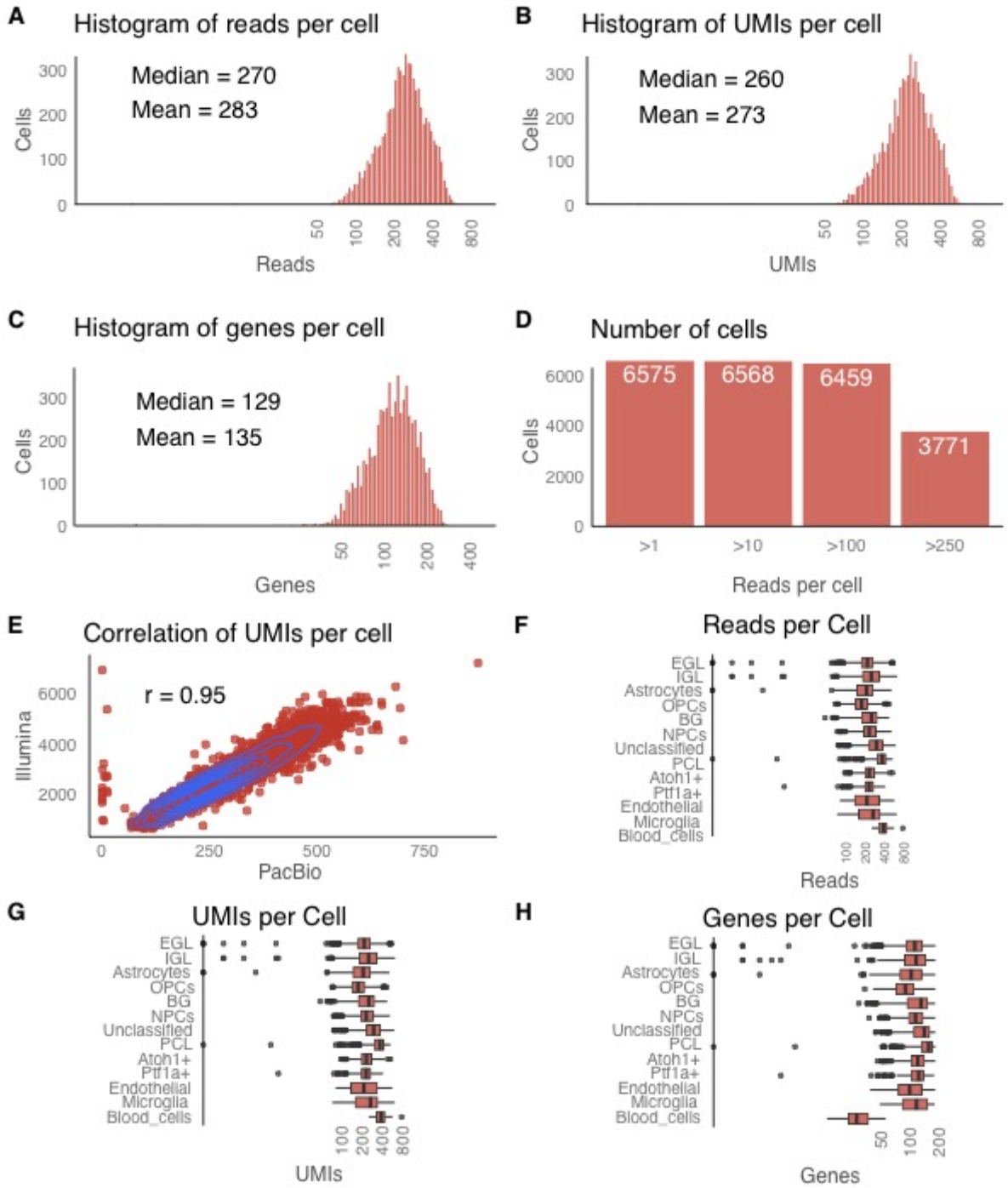

Figure 2: Long-read statistics. Distribution of (A) long reads (B) long-read UMIs and (C) genes per cell. (D) Number of cells >1,>10,>100,>250 long-reads (E) Dotplot and correlation between long-read UMIs and short-read UMIs per cell. Distribution of (F) long reads (G) long-read UMIs and (H) long-read detected genes per cell.

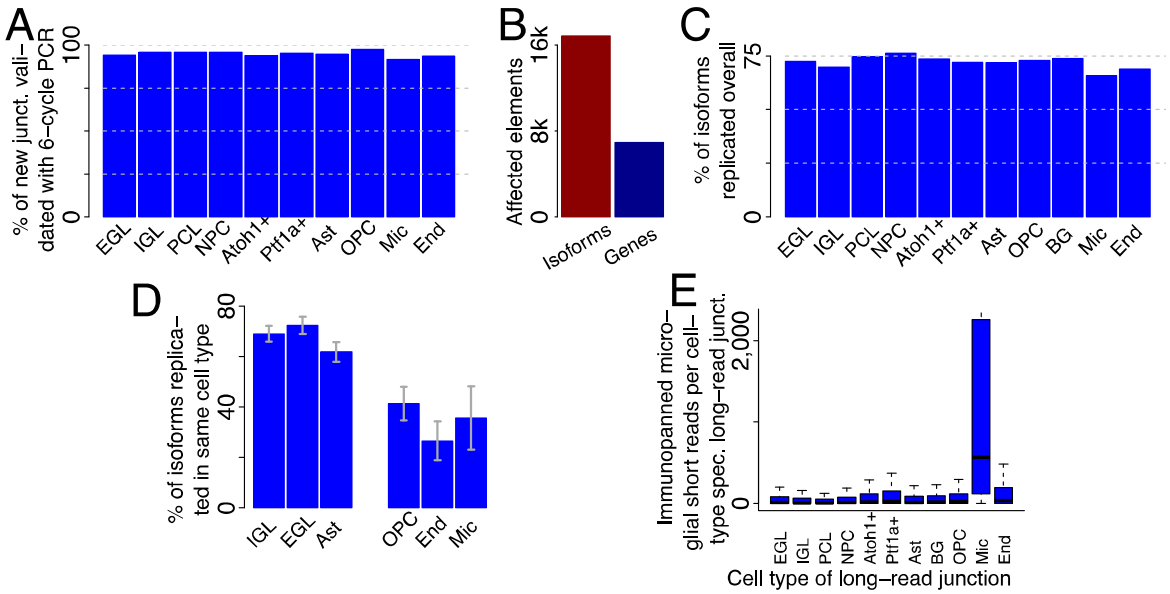

Figure 3: An enhanced and cell-type specific annotation. (A) Percentage of long- read derived junctions that could be validated using low-cycle PCR from bulk P1 cerebellum. (B) Number of isoforms added to the annotation and number of affected genes. (C) Percentage of complete unique isoforms from replicate 2 that could also be observed in replicate 1 (in any cell type) broken up by cell type of origin from replicate 1. (D) Percentage of complete unique replicated isoforms from replicate 2 that could also be observed in replicate 1 (in the same cell type) broken up by cell type of origin from replicate 1 (E) Distributions of coverage with microglial short reads for introns in the enhanced annotation that were exclusively observed in one cell type (indicated by name under the x-axis).

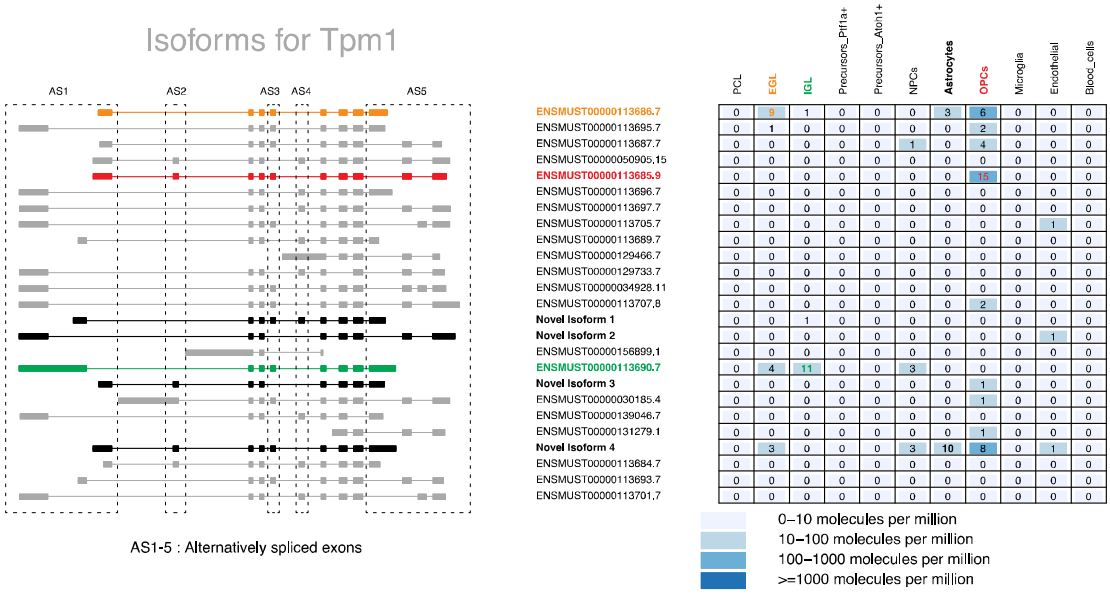

Figure 4: Single gene view for the *Tpm1* gene (from isoformatlas.com), Left: Isoforms of the gene, where each exon is a rectangular block joined by an intervening line representing introns. Alternatively spliced exon blocks identified by AS1-5 are enclosed by dashed-lined boxes Right: Table representing the distribution of UMI counts per isoform (in rows) and cell type of origin as identified in Figure 1 (in columns). Major isoforms in each cell type are colored orange (EGL), red (OPCs) and green (IGL). Novel isoforms are colored in black.