Genome Medicine

Cryopreservation of human cancers conserves tumour heterogeneity for single-cell multi-omics analysis

Sunny Z. Wu[1,2], Daniel L. Roden[1,2], Ghamdan Al-Eryani[1,2], Nenad Bartonicek[1,2], Kate Harvey[1], Aurélie S. Cazet[1,2], Chia-Ling Chan[1,3], Simon Junankar[1,2], Mun N. Hui[1,4], Ewan A. Millar[5,6,7], Julia Beretov[5,8], Lisa Horvath[1,4,9], Anthony M. Joshua[1,10], Phillip Stricker[10], James S. Wilmott[11,12], Camelia Quek[11,12], Georgina V. Long[11,12,13], Richard A. Scolyer[11,12,14], Bertrand Z. Yeung[15], Davendra Segara[10], Cindy Mak[4], Sanjay Warrier[16,17], Joseph E. Powell[3,18], Sandra O'Toole[1,2], Elgene Lim[1,2,10] and Alexander Swarbrick[1,2*]

Abstract

Background: High throughput single-cell RNA sequencing (scRNA-Seq) has emerged as a powerful tool for exploring cellular heterogeneity among complex human cancers. scRNA-Seq studies using fresh human surgical tissue are logistically difficult, preclude histopathological triage of samples, and limit the ability to perform batch processing. This hindrance can often introduce technical biases when integrating patient datasets and increase experimental costs. Although tissue preservation methods have been previously explored to address such issues, it is yet to be examined on complex human tissues, such as solid cancers and on high throughput scRNA-Seq platforms.

Methods: Using the Chromium 10X platform, we sequenced a total of ~ 120,000 cells from fresh and cryopreserved replicates across three primary breast cancers, two primary prostate cancers and a cutaneous melanoma. We performed detailed analyses between cells from each condition to assess the effects of cryopreservation on cellular heterogeneity, cell quality, clustering and the identification of gene ontologies. In addition, we performed single-cell immunophenotyping using CITE-Seq on a single breast cancer sample cryopreserved as solid tissue fragments.

Results: Tumour heterogeneity identified from fresh tissues was largely conserved in cryopreserved replicates. We show that sequencing of single cells prepared from cryopreserved tissue fragments or from cryopreserved cell suspensions is comparable to sequenced cells prepared from fresh tissue, with cryopreserved cell suspensions displaying higher correlations with fresh tissue in gene expression. We showed that cryopreservation had minimal impacts on the results of downstream analyses such as biological pathway enrichment. For some tumours, cryopreservation modestly increased cell stress signatures compared to freshly analysed tissue. Further, we demonstrate the advantage of cryopreserving whole-cells for detecting cell-surface proteins using CITE-Seq, which is impossible using other preservation methods such as single nuclei-sequencing.

(Continued on next page)

* Correspondence: a.swarbrick@garvan.org.au
[1]The Kinghorn Cancer Centre and Cancer Research Division, Garvan Institute of Medical Research, Darlinghurst, NSW, Australia
[2]St Vincent's Clinical School, Faculty of Medicine UNSW, Sydney, NSW, Australia
Full list of author information is available at the end of the article

Wu *et al. Genome Medicine*   (2021) 13:81

Page 2 of 17

(Continued from previous page)

Conclusions: We show that the viable cryopreservation of human cancers provides high-quality single-cells for multi-omics analysis. Our study guides new experimental designs for tissue biobanking for future clinical single-cell RNA sequencing studies.

Keywords: Single-cell RNA sequencing, Cryopreservation, Tumour heterogeneity, CITE-Seq, Breast cancer, Prostate cancer, Melanoma

Background

The tumour microenvironment (TME) is composed of neoplastic cells, mesenchymal and immune cells that interact to shape cancer progression and therapeutic response [1]. Advances in high-throughput single-cell RNA sequencing (scRNA-seq) technologies have rapidly developed in recent years, providing a powerful platform to resolve the aetiology of the TME in solid cancers [2]. Performing scRNA-seq on clinical samples remains logistically and technically challenging, mainly due to the transport of patient tissue from operation rooms to laboratories for processing, which are often complicated by short notices and core-facility access hours. The need to process fresh tissue specimens at the time of tissue availability, often as a single specimen, often introduces large experimental costs and confounding batch effects upon studies with large numbers of patients and prevents the selection and triage of cases for analysis based on histopathological analysis.

Several approaches have been developed to address such issues. Madissoon et al. benchmarked short-term cold preservation of tissue prior to scRNA-Seq, which showed little impact on transcriptome integrity within the first 24 h [3]. Cell type-specific transcriptional changes can emerge after longer cold preservation periods (> 24 h), particularly affecting immune subpopulations in normal tissues [3]. Cold preservation is yet to be evaluated for complex tissues such as solid tumours, which possess distinct features in tissue viability. Factors including tissue necrosis, hypoxia and therapeutic treatments often result in poor viability of cells in solid tumour tissues. Regardless, such short-term storage periods still limit the ability to perform simultaneous sample processing. In particular for microfluidic droplet-based scRNA-Seq, this can result in batch effects for larger studies processed over several time points [4], where cell fixation and cryopreservation methods can minimise such issues. Cell fixation methods using agents such as methanol [5] or dithio-bis (succinimidyl propionate) (DSP) [6] are effective and can be applied to overcome several barriers of cold preservation. However, scRNA-Seq studies have shown that fixation methods can elevate ambient background RNA and do not maintain cell integrity as effectively as cryopreservation using DMSO [7]. Furthermore, fixation methods are not always

practical with solid cancers which require lengthy dissociation protocols and also preclude certain downstream procedures such as antibody staining or cell culture. Although sequencing of nuclei from snap frozen tissue can be applied to avoid dissociation methods [8], this approach is not compatible with powerful cell surface immunophenotyping methods with DNA-barcoded antibodies such as CITE-Seq [9]. Sequencing of nuclei also does not permit the selection of cell subsets of interest or the removal of low-quality cells prior to capture.

While cryopreservation of cells using DMSO has been applied to study cell lines, PBMCs and other model organisms by scRNA-Seq [7, 10, 11], it is unclear how more complex solid tissues may be impacted by this freezing process. An important study by Guillaumet-Adkins et al. showed that the cryopreservation of whole-cells and tissues can be used to conserve transcriptional profiles from experimental systems such as human cell lines and mouse tissues [12]. These models represent fairly homogeneous systems, and it is unclear whether the highly heterogeneous nature of the TME is also conserved following cryopreservation. In addition, studies have only benchmarked cryopreservation of intact tissues using low-throughput plate-based scRNA-seq technology [12, 13], where highly viable cells are selected by FACS for immediate lysis and mRNA hybridisation [14]. It is yet to be determined if cryopreservation of tissues can be applied to more recent high throughput scRNA-Seq platforms such as the Chromium 10X platform. These platforms are fundamentally different to FACS-based scRNA-Seq methods, as single-cells are captured through droplet-based microfluidics, where viability selection is not simultaneously performed. More recent studies using droplet-based microfluidics have analysed circulating T-regulatory cells [15] and have shown that heat shock and cell stress pathways can emerge as transcriptional artefacts from cryopreservation. It is unclear how these stress responses affect the diversity of cell types within intact tissues, where the efficiency of the cryoprotectant may be affected, and further require tissue dissociation.

In this study, we aimed to examine the effect of cryopreserving dissociated cells and solid tissue fragments from human tumours prior to scRNA-Seq on the 10X Chromium platform. We tested this across three

Wu *et al. Genome Medicine*   (2021) 13:81

Page 3 of 17

common cancer types: breast, prostate and melanoma. Following cryopreservation, we demonstrated a strong conservation of the heterogeneous neoplastic, mesenchymal and immune subpopulations. We show that scRNA-Seq results of cells from cryopreserved solid tissue and from cryopreserved dissociated cell suspensions are comparable to those from cells prepared from fresh tissue, with minimal impact on downstream analysis methods. In some tumours sequenced from cryopreserved solid tissues and after overnight cold storage conditions, we observe some minor gene expression changes associated with cell stress responses. Lastly, we show that cryopreserving whole-cells allows for powerful immunophenotyping methods such as CITE-Seq, which is not possible using nuclei-based sequencing methods. Our findings allow a simple biobanking protocol to process patient samples, significantly decreasing technical variation among larger patient cohorts and serial time-points analyses. Our biobanking protocol unlocks patient cohorts previously collected in such a manner and serves as a guide for the sample collection in future clinical scRNA-Seq studies.

Methods

Primary tissue dissociation and sample preparation

We examined tissue from three primary breast cancers, two primary prostate cancers and a lymph node metastasis from a patient with a cutaneous melanoma (Additional file 1). Fresh surgically resected tissues were washed with RPMI 1640 (ThermoFisher Scientific) and diced into $1-2\,mm^3$ pieces. Tissue pieces were mixed and approximately one third were viably frozen in cryogenic vials in 5% dimethyl sulfoxide (DMSO) and 95% fetal bovine serum (FBS) at 1 °C/minute in −80 °C using Mr. Frosty™ Freezing Containers (ThermoFisher). This was classified as the solid cryopreserved tissue (CT) sample. The remaining tissue was further minced with scissors and enzymatically dissociated using the Human Tumour Dissociation Kit (Miltenyi Biotec) following the manufacturer's protocol. Following incubation with the enzymes, the sample was resuspended in media (80% RPMI 1640, 20% FBS) and filtered through MACS® SmartStrainers (70 μM; Miltenyi Biotec). The resulting single cell suspension was centrifuged at 300×*g* for 5 min. At this stage, a proportion of the dissociated cell suspension was frozen in cryogenic vials in 10% DMSO, 50% FBS, and 40% RPMI 1640 at 1 °C/minute in −80 °C using Mr. Frosty™ Freezing Containers (ThermoFisher). This was classified as the dissociated cryopreserved cell suspension (CCS) sample. For the dissociated fresh tissue (FT) sample, red blood cells were lysed with Lysing Buffer (Becton Dickinson) for 5 min and neutralised with media (80% RPMI 1640, 20% FBS). Cells were further filtered through a 40-μm filter and centrifuged at 300×*g* for 5 min. Viability was assessed using Trypan Blue (ThermoFisher). For samples with a viability score of < 80%, enrichment was

performed using the EasySep Dead Cell Removal (Annexin V) Kit (StemCell Technologies) following the manufacturer's protocol. This was not performed for tumours that had too low of a total cell yield (less than 1×10^5 cells) following dissociation and splitting across cryopreservation conditions. Enriched cell suspensions were resuspended in a final solution of PBS with 10% FBS solution prior to loading on the 10X Chromium platform.

For the processing of cryopreserved replicates, samples were frozen at −80 °C for ~ 1 week followed by ~ 5 weeks at −196 °C for prior to scRNA-Seq. For obvious logistical reasons (freezing storage time), FT samples were run on the 10X Chromium platform immediately whilst CT and CSS samples were processed simultaneously at a later date. Following cryopreservation, samples were thawed in a 37 °C water bath and washed multiple times with RPMI 1640. CT samples were dissociated in the same manner as the FT samples, as described above. In general, we observed a lower cell viability in cryopreserved samples compared to their respective FT sample, as measured using Trypan Blue. Cryopreserved samples were enriched for live cells if viability was assessed to be < 80%, as described above. Viability enrichment was not performed or repeated multiple times for samples that had less than 1×10^5 cells. For both cryopreserved replicates from breast tumours, the mouse cell line NIH3T3 was thawed and spiked in at 2% of the total cell number prior to cell loading on the 10X Chromium. The CCS sample for PC-P2 resulted in a very low cell number (less than 400) after sequencing and was excluded from subsequent cluster and cell annotation analyses. This sample had a very low emulsion volume and was due to an instrumental failure on the Chromium 10X platform, likely due to a microfluidic clogging issue.

Single-cell RNA sequencing on the 10X chromium platform

High throughput scRNA-Seq was performed using the Chromium Single Cell 3′ v2 and 5′ chemistry (10X Genomics) according the to the manufacturer's instructions. All replicates within a case were captured using the same chemistry. A total of 6000 cells were targeted per lane. The scRNA-Seq libraries were sequenced on the Illumina NextSeq 500 platform with pair-end sequencing and dual indexing according to the recommended Chromium platform protocol; 26 cycles for Read 1, 8 cycles for i7 index and 98 cycles for Read 2.

Data processing

Read demultiplexing and alignment to the GRCh38 human reference genome was performed using the Cell Ranger Single Cell Software v2.0 (10X Genomics) with the cellranger mkfastq and count functions, respectively. For cryopreserved replicates from breast tumours with

mouse cell line spike in (NIH3T3), the above steps were performed using the GRCh38 human and mm10 mouse reference genomes. Raw count matrices were filtered for 'real' barcodes using the EmptyDrops package in *R* which calculates deviations against a generated ambient background RNA profile [16]. Additional conservative cutoffs were further applied based on the number of genes detected per cell (greater than 200) and the percentage of mitochondrial unique molecular identifier (UMI) counts (less than 20%). Filtered barcodes from matched replicates were then processed and integrated using the Seurat v3 package in *R* as per the developers' vignettes [17]. For the comparison of transcript metrics across cryopreserved replicates, including the number of genes, UMIs and gene correlations, we performed down-sampling of sequencing libraries by the total number of mapped reads using the cellranger aggr function. For comparison of clusters across cryopreservation conditions, cells were randomly downsampled to the lowest replicate size using the data.table package in *R*.

Silhouette scores, mixing metric and local structure metric

We applied clustering and mixability metrics from Stuart et al. to quantitative measure the robustness of the cryopreserved replicates to reflect good technical replicates with the FT [17]. Stratified random downsampling was first applied to each case to generate clusters with equal sizes across all three conditions. This was performed using data.table package in *R*. As a positive control, FT datasets were randomly downsampled to generate two pseudo-replicates. Three comparisons were computed per case: FT-1 vs FT-2, FT-1 vs CCS and FT-1 vs CT. For the melanoma case, the comparisons were FT-1 vs FT-2, FT-1 vs CCS and FT-1 vs CO. Silhouette scores, mixing metrics and local structure metrics were all computed using code adopted from the Seurat v3 package [17].

Bulk and cluster level gene correlations

Adjusted R^2 correlation values were calculated using linear regression, implemented in *R*. Sequencing libraries normalised by the number of mapped reads using Cell-Ranger were used. Pseudo-replicate bulks and cluster-level bulks were generated from log-normalised gene expression values. FT bulk and cluster level replicates were compared to cryopreserved replicates (CCS/CT/CO).

Differential gene expression and pathway enrichment

Integrated cases were split by replicate. Differential gene expression was then performed between integrated cluster IDs across each of the replicates using the MAST method through the *FindAllMarkers* function in Seurat (log fold change threshold of 0.25, *p value* threshold of 1×10^{-5} and FDR threshold of 0.05) [18]. All DEGs from

each cluster were then passed on to the ClusterProfiler package for functional enrichment [19]. The *compareCluster* function was used with the enrichGO default CC sub-ontology under the human org.Hs.eg.db database. The overlaps of detected GO pathways across each replicate were computed and visualised using the euler and ggplot2 packages in R. The same parameters described above were used for differential gene expression and GO pathway comparisons between conditions across all cells and cluster/cell type level.

CITE-Seq staining and data processing

Samples were stained with 10X Chromium 3′ mRNA capture compatible TotalSeq-A antibodies (BioLegend, USA). Staining was performed as previously described by Stoeckius et al. (2017) with a few modifications [9]. Briefly, a maximum of 2 million cells per sample was resuspended in 100 µl of cell staining buffer (BioLegend, USA) with 5 µl of Fc receptor Block (TrueStain FcX, BioLegend, USA) for 15 min followed by a 30 min staining of the antibodies at 4 °C. A concentration of 1 µg/100 µl was used for all antibody markers used in this study. The cells were then washed 3x with PBS containing 10% FBS media followed by centrifugation (300x*g* for 5 min at 4 °C) and expungement of supernatant. The sample was then resuspended in PBS with 10% FBS for 10X Chromium capture. Indexed CITESeq libraries were spiked in to 10X scRNA-Seq libraries for sequencing on the NextSeq500 platform (Illumina). Reads were demultiplexed using CellRanger v2.0. Cell counts of CITE antibodies were calculated from sequenced CITE libraries with CITE-seq-Count v.1.4.3 using default parameters recommended by developers. Counts were integrated with scRNA-seq data using Seurat (v.3.1.4), scaled and normalised.

Results

Cryopreservation allows for robust conservation of cellular heterogeneity in human breast cancers

The preservation of cellular heterogeneity is an important factor for analysing solid cancers. We first investigated this in primary human breast cancers collected from three patients. To minimise spatial biases from regional sampling, fresh surgical specimens were initially cut in to 1–2 mm^3 pieces and thoroughly mixed. One third of the mix was immediately cryopreserved with DMSO at −80 °C (designated as the cryopreserved tissue—CT), and the remaining mix was dissociated into a single-cell suspension using a commercial kit-based method (see the "Methods" section). A fraction of this cell suspension was immediately cryopreserved with DMSO at −80 °C (designated as the cryopreserved cell suspension—CCS), and the remaining of this cell suspension was processed immediately for scRNA-Seq using the

Chromium 10X platform (designated as fresh tissue—FT). After storage of the cryopreserved samples, both CT and CCS, at −80 °C for about 1 week, they were stored in liquid nitrogen at −196 °C for up to 5 weeks to mimic standard tissue biobanking procedures. Following cryopreservation, CT and CCS samples were thawed and processed for scRNA-Seq in the same manner as the FT sample. For cryopreserved replicates, we spiked in the mouse NIH3T3 fibroblast cell line as a positive control (∼ 2%) for the scRNA-Seq experimental workflow. In total, we sequenced 23,805, 29,865 and 24,250 cells from breast cancer patients 1–3, (assigned as BC-P1, BC-P2 and BC-P3), respectively.

A detailed comparison was performed between samples processed as FT, CCS or CT (Fig. 1a). We performed batch correction and integration of all matched fresh and cryopreserved replicates using the anchoring based method in Seurat v3 (Fig. 1b) [17]. This revealed consistent 'mixability' across the three conditions, where a strong overlap was also observed in Uniform Manifold Approximation and Projection (UMAP) space. This was also observed in the non-batch corrected data (Additional file 2: Figure S1a), reflecting good technical replicates on the 10X Chromium platform. To account for variation in cell-type proportions, all matched conditions were downsampled to the lowest replicate cell number to examine the composition of cells in each cluster (Fig. 1c). Only three clusters across all three datasets were not comprised of cells from all three conditions (Fig. 1c). These differential clusters were all detected in the BC-P2 dataset, including clusters c11 (737 cells), c18 (191 cells) and c23 (27 cells). Clusters c11 and c18 were only detected in the FT sample and resembled cell doublets captured from a varying number of cells sequenced per replicate, which ultimately contributes to differences in the expected doublet rate. These clusters showed characteristics of cell doublets, including the expression of markers from multiple cell lineages such as *EPCAM*, *PTPRC*, *PECAM1* and *PDGFRB* (Additional file 2: Figure S1b). Cluster c23 was comprised of smaller cell numbers, and may be a result of sampling rarer cell types, rather than from the cryopreservation process. To our surprise, the mouse NIH3T3 fibroblast spike-ins could also be detected in all cryopreserved replicates following the mapping of reads to the human GRCh38 reference genome alone (c19 in BC-P1, c17 in BC-P2 and c14 in BC-P3). These were confirmed as mouse cells by remapping reads to both human and mouse reference genomes, suggesting that mouse reads were assigned to their human orthologs when mapping to a single reference genome using CellRanger. NIH3T3 fibroblast spike-ins captured from different cryopreserved replicates and independent experiments mixed well (Additional file 2: Figure S1c), indicating high reproducibility on the 10X

Genomics platform. As expected, NIH3T3 fibroblasts highly expressed markers *Dlk1, Acta2, Vim, Actg1, Col1a1* and *Col1a2* (Additional file 2: Figure S1d).

From investigating the expression of canonical cell type markers, we identified a strong preservation of major cell lineages in cryopreserved replicates (Fig. 1d). As observed in the representative case BC-P1 (Fig. 1d), we identified a strong conservation of the housekeeping gene *ACTB*, cancer/epithelial cells (*EPCAM*; clusters c1, c5, c13, c20 and c14), myoepithelial cells (*KRT14*; c6), T cells (*CD3D*; c3, c7 and c17), B-cells (*MS4A1*; c16), plasmablasts (*JCHAIN*; c18), myeloid cells (*CD68*; c12 and c21), endothelial (*PECAM1*; c0, c8, c9, c11, c15 and c22), perivascular cells (*PDGFRB*; c2) and cancer-associated fibroblasts (CAFs; *PDGFRA*; c4 and c10) (Fig. 1b, d; Additional file 2: Figure S2a; Additional file 3). Similar trends in the preservation of the TME were observed in all three breast cancer cases (Additional file 2: Figure S1b; Additional file 2: Figure S2b-c; Additional file 3). Consistent with the even distribution of cells from each condition across all the breast tumour clusters (Fig. 1c), these cluster annotations were also detected when analysing each sample individually using unbiased clustering and UMAP (Additional file 2: Figure S3a-c). In summary, cryopreservation of human breast cancers as either solid tissue or single cell suspension maintains the heterogeneity of major cell lineages detected from processing fresh tissue.

Cryopreserved replicates resemble good technical replicates with the fresh tissue data

Although visual inspection of the dimensional reduction UMAP plots indicated good mixability and minimal technical variation emerging from cryopreservation, we applied several metrics adopted from Stuart et al. [17] to quantitatively measure the impact on downstream clustering. We examined silhouette coefficient scores, mixing metric and local structure metric to measure the robustness of cryopreservation to reflect good technical replicates with the FT. As described in the previous section, we performed stratified downsampling of cells to account for differences emerging from the total number of cells sequenced. We compared cells from FT against cells from matched cryopreserved replicates independently in the following comparison conditions: FT vs CCS and FT vs CT. As a positive control, we compared two sets of FT cells downsampled from the same dataset to reflect perfect technical replicates (FT-1 vs FT-2).

Silhouette coefficient scores, which range from − 1 to + 1, measure how similar a cell is to cells from its own cluster in dimensional reduction space. We applied this to measure the mixability of the cryopreserved replicates, where scores closer to 0 indicate a higher mixability between replicates irrespective of cryopreservation

Fig. 1 (See legend on next page.)
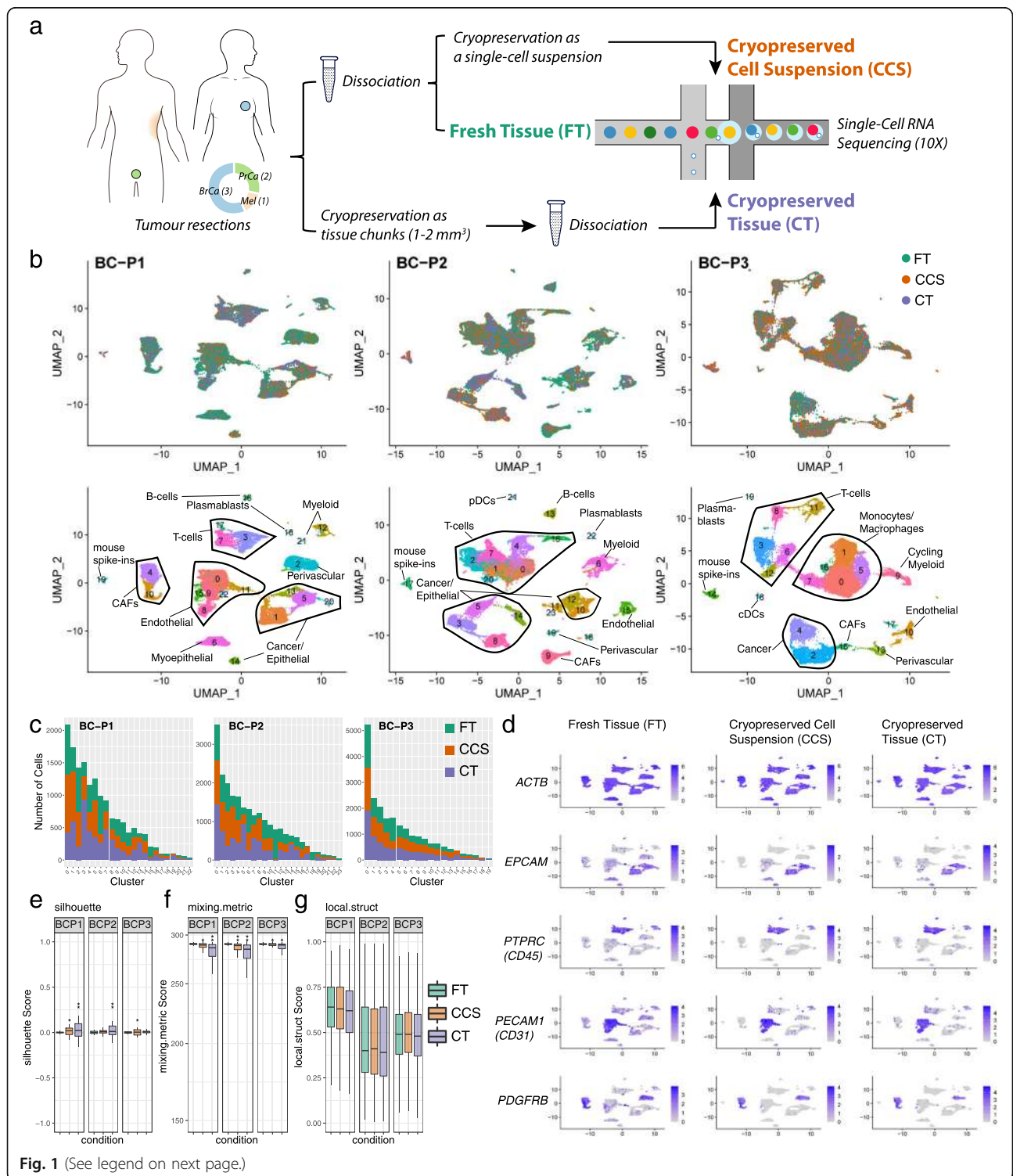
(See figure on previous page.)

Fig. 1 Cryopreservation allows for robust cell-type detection in clinical breast cancer samples. **a** Experimental workflow. **b** UMAP visualisation of 23,803, 29,828 and 24,250 cells sequenced across dissociated fresh tissue (FT; green), dissociated cryopreserved cell suspensions (CCS; orange) and solid cryopreserved tissue (CT; purple) replicates from three primary breast cancer cases (BC-P1, BC-P2 and BC-P3). UMAPs are coloured by cryopreserved replicate (top) and by cluster ID (bottom) with cell types annotations overlayed. Matched replicates were integrated using the Seurat v3 method. **c** Number of cells detected per cluster. Cells were downsampled to the lowest replicate size. **d** FeaturePlot visualisations of gene expression from BC-P1 fresh and cryopreserved replicates, showing the conservation of the housekeeping gene *ACTB* and heterogeneous cancer/epithelial (*EPCAM*), immune (*PTPRC*/CD45), endothelial (*PECAM1*/CD31) and fibroblast/perivascular (*PDGFRB*) clusters. **e, g** Distribution of silhouette scores (range −1 to + 1) (**e**), mixing metric (**f**) and local structure metrics (**g**) of clustering following cryopreservation. Samples were downsampled by replicate and cluster sizes and compared to the respective FT samples. Cell comparisons were performed across downsampled FT-1 vs FT-2 cells (positive control), FT vs CCS cells and FT vs CT cells. Stars represent standard deviations: **e** silhouette scores s.d. 0.02–0.05* and s.d. > 0.05**; **f** mixing metrics s.d. 2–10* and s.d. > 10**; **g** local structure metrics s.d. > 0.05*
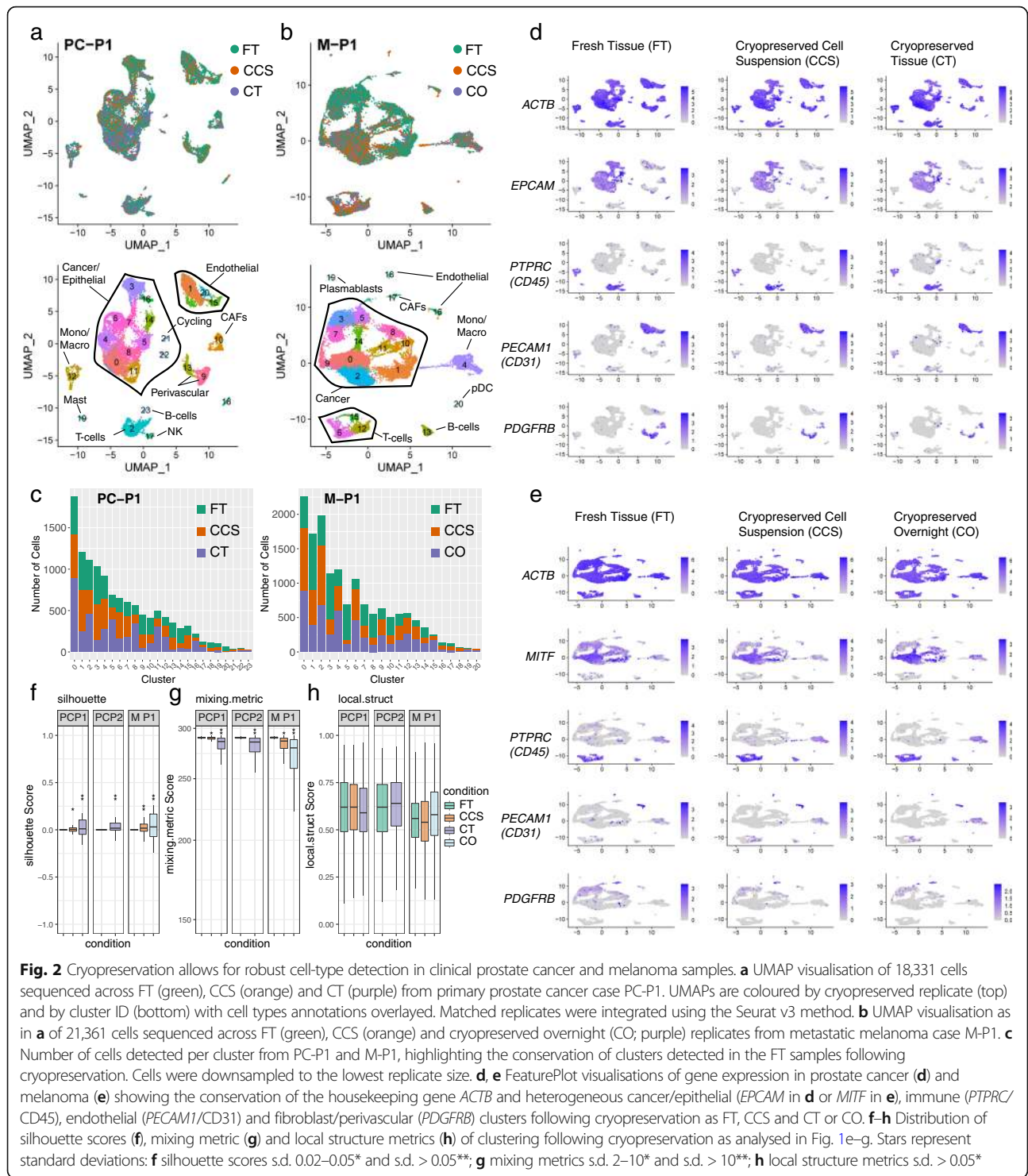
condition. As expected from our positive control comparisons (FT-1 vs FT-2), this yielded average silhouette scores close to 0 for all three breast cancer cases (Fig. 1e). In general, we observed values close to 0 for all cryopreserved replicate comparisons, with no silhouette scores outside of the −0.25 to 0.25 range (Fig. 1e). Minor variations, as indicated through increased standard deviations, were observed in the CCS replicates of two cases: BC-P1 and BC-P3 (Fig. 1e; Additional file 4). Similarly, increased standard deviations were observed when comparing CT replicates in two cases: BC-P1 and BC-P2 (Fig. 1e; Additional file 4).

We next applied the mixing metric to assess how well cryopreserved replicates 'mixed' with the FT data after integration (Fig. 1f). The mixing metric examines the distribution of replicates in a cell's neighbourhood ($k = 5$ and k.max = 300), where values closer to 300 resemble a high 'mixability' [17] (Fig. 1f). Overall, very high mixing metric scores were observed across the comparison conditions from all three breast cancer cases; however, slightly lower values and higher standard deviations were consistently detected in cells cryopreserved as CT compared to CCS (Fig. 1f; Additional file 4). Finally, we assessed how local cell clusters ($k = 100$) detected in individual replicates were preserved upon data integration using the local structure metric [17]. In all three cases, this revealed no significant differences in the standard deviations from our positive FT control comparisons and the cryopreserved replicates (Fig. 1g; Additional file 4), indicating that the clusters identified in individual replicates were largely consistent upon integration with the FT data. Overall, we conclude that cryopreservation can yield good quality technical replicates. Only minor variations in clustering, as determined by silhouette coefficients and mixing metrics, arose from processing as dissociated CCS and solid CT, with the latter resulting in slightly more variable data.

Cryopreservation yields high-quality data in prostate cancers and a metastatic melanoma

Tissue architectures differ across cancer sites and metastatic lesions. To assess the impact of cryopreservation across different tissue sites, we repeated our benchmarking on primary prostate cancer tissue collected from two patients (PC-P1 and PC-P2), and a regional lymph node metastasis collected from one patient with a known cutaneous melanoma (M-P1). For the metastatic melanoma sample, we benchmarked cell suspensions cryopreserved immediately (CCS sample) as well as after overnight cold storage of the tissue at 4 °C in media (designated as cryopreserved overnight—CO). The CO replicate mimics conventional biobanking procedures where tissue is collected from late patient procedures, stored at 4 °C and processed the following day. In total, we sequenced 18,333, 18,327 and 21,363 cells from PC-P1, PC-P2 and M-P1, respectively (Fig. 2a). Here, the CCS replicate from PC-P2 resulted in low cell number and was excluded from subsequent comparisons (see the "Methods" section). Similar to the breast cancer data, comparisons of the non-batch corrected data revealed a good mixture of cells from all conditions, reflecting that of good technical replicates (Additional file 2: Figure S1a). Batch correction and data integration revealed consistent mixability across the three conditions in UMAP space (Fig. 2a, b; Additional file 2: Figure S1e). Only one very small cluster in PC-P1 (c2 - 64 cells) was not comprised of cells from all three conditions (Fig. 2c) and is again likely a result of cell sampling rather than cryopreservation. All clusters detected in M-P1 were comprised of cells from all conditions (Fig. 2c). Similar to our benchmarking in breast cancers, we observed a strong conservation of the housekeeping gene *ACTB* and markers for cancer clusters (*EPCAM* in prostate and *MITF* in melanoma), immune subsets (*PTPRC*), endothelial cells (*PECAM1*/CD31) and fibroblast/perivascular (*PDGFRB*) cells in prostate cancers and the metastatic melanoma (Fig. 2d, e; Additional file 2: Figure S1f; Additional file 2: Figure S2d-f). These cluster annotations could also be detected when analysing each condition individually using unbiased clustering and UMAP (Additional file 2: Figure S3d-e). Upon examining clustering metrics, we found similar trends with slightly higher variation in silhouette scores and mixing metrics

Fig. 2 Cryopreservation allows for robust cell-type detection in clinical prostate cancer and melanoma samples. **a** UMAP visualisation of 18,331 cells sequenced across FT (green), CCS (orange) and CT (purple) from primary prostate cancer case PC-P1. UMAPs are coloured by cryopreserved replicate (top) and by cluster ID (bottom) with cell types annotations overlayed. Matched replicates were integrated using the Seurat v3 method. **b** UMAP visualisation as in **a** of 21,361 cells sequenced across FT (green), CCS (orange) and cryopreserved overnight (CO; purple) replicates from metastatic melanoma case M-P1. **c** Number of cells detected per cluster from PC-P1 and M-P1, highlighting the conservation of clusters detected in the FT samples following cryopreservation. Cells were downsampled to the lowest replicate size. **d**, **e** FeaturePlot visualisations of gene expression in prostate cancer (**d**) and melanoma (**e**) showing the conservation of the housekeeping gene *ACTB* and heterogeneous cancer/epithelial (*EPCAM* in **d** or *MITF* in **e**), immune (*PTPRC/ CD45*), endothelial (*PECAM1/CD31*) and fibroblast/perivascular (*PDGFRB*) clusters following cryopreservation as FT, CCS and CT or CO. **f**–**h** Distribution of silhouette scores (**f**), mixing metric (**g**) and local structure metrics (**h**) of clustering following cryopreservation as analysed in Fig. 1e–g. Stars represent standard deviations: **f** silhouette scores s.d. 0.02–0.05* and s.d. > 0.05**; **g** mixing metrics s.d. 2–10* and s.d. > 10**; **h** local structure metrics s.d. > 0.05*

emerging from cells cryopreserved as CT compared to CCS (Fig. 2f, g; Additional file 4). For the melanoma comparisons, the CO replicate exhibited a much higher variation in silhouette scores and mixing metric compared to CCS, indicating potential transcriptional artefacts arising from overnight cold preservation prior

to cryopreservation (Fig. 2f, g; Additional file 4). No major differences were observed in the local structure metric of both prostate and melanoma cases (Fig. 2h), indicating that clustering neighbourhoods in individual replicates were consistently detected upon integration with the FT data.

Wu *et al. Genome Medicine* (2021) 13:81

Page 9 of 17
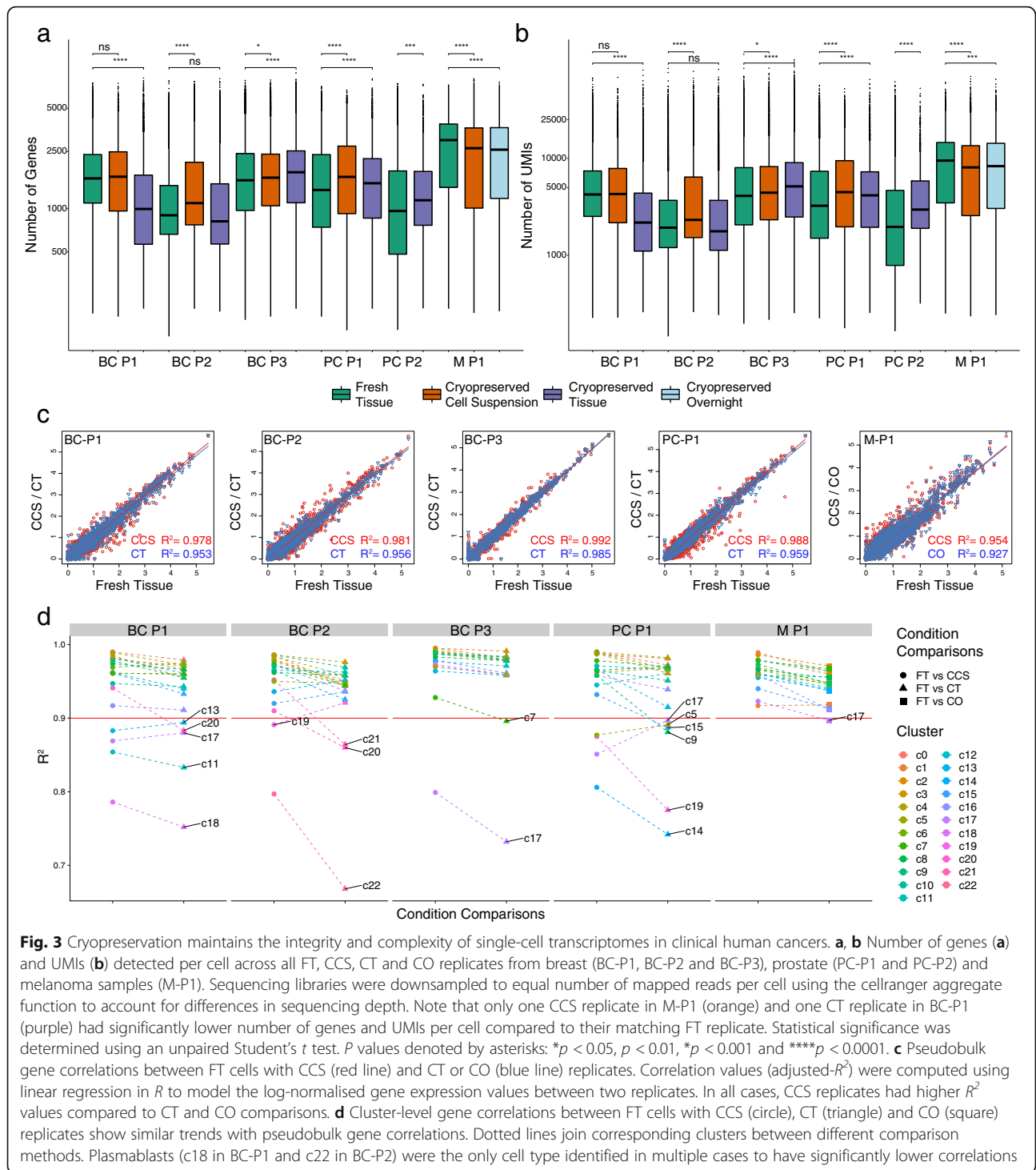
Impact of cryopreservation on cellular stress

As we observed small decreases in cell viability following cryopreservation, we next examined the annotated scRNA-Seq datasets from all tumour types to see if this was a broad effect or rather impacted a specific cell type. This did not reveal any significant trends for the seven common cell types annotated across all tumours analysed (B-cells, CAFs, cancer/epithelial, monocyte/macrophage, plasmablasts, perivascular cells and T-cells), suggesting there are no cell-type specific impacts following cryopreservation ($p > 0.05$, paired t test; data not shown). We next assessed the percentage of mitochondrial transcripts per cell, which is a commonly assessed metric in scRNA-Seq data to determine cell viability [20]. Cells undergoing stress have higher percentages of mitochondrial transcripts from either permeable membranes, causing loss of cytoplasmic mRNA, genome degradation or increased metabolic demand [20]. In the general comparisons of all cells, two of five samples analysed as CCS (BC-P3 and M-P1) and two of five samples analysed as CT (PC-P1 and PC-P2) had significantly higher mitochondrial percentages compared to FT (Additional file 2: Figure S4a). These differences were dispersed across multiple cell types in each respective sample (Additional file 2: Figure S4b-f), where cancer/epithelial cells were the only consistent cell type to have significantly increased mitochondrial percentages across more than one case (CCS condition from BC-P2, BC-P3 and M-P1). Despite these minor differences, all cell types sequenced across all conditions and tumours had a mean mitochondrial percentage of less than 10% (Additional file 2: Figure S4b-f), suggesting that the survival and viability of specific cell types were not markedly impacted by cryopreservation. Taken together, our benchmarking across multiple tissue sites indicates that cryopreservation preserves the cellular heterogeneity of the TME and acts as good quality technical replicates.

Tumour cryopreservation maintains the integrity and complexity of single-cell transcriptomes

We next investigated whether gene expression and transcriptome integrity were affected through the cryopreservation process. We first examined the number of genes and unique molecular identifiers (UMIs) detected per cell across cryopreserved replicates. For this comparison, libraries were first downsampled by the number of mapped sequencing reads to account for differences emerging from varying sequencing depths. This revealed that an average of 1809, 1842 and 1694 genes and 6149, 6525 and 5851 UMIs per cell were detected across all FT, CCS and CT replicates, respectively (Fig. 3a, b). Within matched cases, only cryopreserved cell suspension conditions from

M-P1 (from both CCS and CO) yielded a lower average number of genes and UMIs per cell compared to the FT (Fig. 3a, b). Similarly, only one CT replicate (BC-P1) had a significantly lower number of genes and UMIs detected per cell compared to the FT (Fig. 3a, b). Although this was not observed across multiple cases, a lower detection rate from CT may reflect a minor impact on transcript abundance and quality from the cryopreservation process. In addition, cell type and cell size can be an important factor determining transcript abundance. To determine that these subtle changes were not due to differences in cell abundance across cryopreserved replicates, we confirmed that these changes were also present at the cluster level (Additional file 2: Figure S5a-f). For example, although cancer cells (clusters c1, c5 and c14 in BC-P1) generally hold more transcripts compared to T cells (clusters c3, c7 and c17 in BC-P1), less genes and UMIs were also found in these respective cell types captured in CT replicate, as per the bulk comparisons (Additional file 2: Figure S5a).

We next investigated the gene correlation between FT samples and their respective cryopreserved replicates. Bulk gene correlations revealed high R^2 values between FT and all cryopreserved replicates ($R^2 > 0.90$; Fig. 3c) where on average, CCS replicates had higher R^2 values with the FT sample (mean $R^2 = 0.98$, min = 0.95 and max = 0.99) compared to the CT replicates (mean $R^2 = 0.96$, min = 0.93 and max = 0.99) (Fig. 3c). Similarly, we examined if this trend was unique to particular cell types on the cluster level (Fig. 3d). Only clusters containing cells from all three replicates with a minimum cluster size 100 and at least 20 cells per replicate were examined for representative gene correlations, in order to not be skewed by low cell numbers. Cluster correlations revealed consistent trends with the bulk comparisons, where CCS replicates consistently showed slightly higher R^2 correlation than with FT replicates (Fig. 3d). Although a majority of clusters displayed high correlations ($R^2 > 0.90$; indicated by the red line in Fig. 3d), several smaller clusters showed significantly lower correlations than the bulk ($R^2 < 0.90$; Fig. 3d) including five clusters in BC-P1 (c13—cancer/epithelial, c20—cancer/epithelial, c17—T cells, c11—endothelial and c18—plasmablasts), four clusters in BC-P2 (c19—perivascular, c21—pDCs, c20—T cells and c22—plasmablasts), two clusters in BC-P3 (c7—monocyte/macrophage and c17—unassigned cluster), six clusters in PC-P1 (c17—NK cells, c5—cancer/epithelial, c15—endothelial, c9 perivascular, c19—mast cells and c14—cancer/epithelial) and one cluster in M-P1 (c17—CAFs). The majority of these poorly correlated clusters were comprised of small cell numbers. The only cell type consistently found to have very poor

Fig. 3 Cryopreservation maintains the integrity and complexity of single-cell transcriptomes in clinical human cancers. **a**, **b** Number of genes (**a**) and UMIs (**b**) detected per cell across all FT, CCS, CT and CO replicates from breast (BC-P1, BC-P2 and BC-P3), prostate (PC-P1 and PC-P2) and melanoma samples (M-P1). Sequencing libraries were downsampled to equal number of mapped reads per cell using the cellranger aggregate function to account for differences in sequencing depth. Note that only one CCS replicate in M-P1 (orange) and one CT replicate in BC-P1 (purple) had significantly lower number of genes and UMIs per cell compared to their matching FT replicate. Statistical significance was determined using an unpaired Student's *t* test. *P* values denoted by asterisks: *$p < 0.05$, $p < 0.01$, *$p < 0.001$ and ****$p < 0.0001$. **c** Pseudobulk gene correlations between FT cells with CCS (red line) and CT or CO (blue line) replicates. Correlation values (adjusted-R^2) were computed using linear regression in R to model the log-normalised gene expression values between two replicates. In all cases, CCS replicates had higher R^2 values compared to CT and CO comparisons. **d** Cluster-level gene correlations between FT cells with CCS (circle), CT (triangle) and CO (square) replicates show similar trends with pseudobulk gene correlations. Dotted lines join corresponding clusters between different comparison methods. Plasmablasts (c18 in BC-P1 and c22 in BC-P2) were the only cell type identified in multiple cases to have significantly lower correlations

correlation values across multiple cases ($R^2 < 0.80$) was plasmablasts (c18 in BC-P1 and c22 in BC-P2), suggesting that this cell type is more prone to transcriptional changes due to cryopreservation (Fig. 3d). Taken together, we find that cryopreservation can conserve high-quality transcriptomes for scRNA-Seq. These data

suggest that processing scRNA-Seq from CCS yields slightly higher quality data than from CT. Although the sample number was small, we found that cryopreservation induced changes in transcriptome integrity of plasmablasts identified in breast tumours, warranting some caution for studying this cell type using this method.

Tumour cryopreservation maintains biological pathways detected in fresh tissue samples

Biological and functional findings from scRNA-Seq experiments are often interpreted through gene ontology (GO) analysis for pathway enrichments across unique cell clusters. To assess if such downstream analyses are impacted by cryopreservation, we first separated our integrated clusters by their cryopreservation conditions. We then performed differential gene expression and GO pathway enrichment to assess how pathways detected across FT clusters were detected in their respective cryopreserved replicates. This analysis revealed a good overlap of total detected pathways in all cancer cases, with over 64% of all FT pathways consistently detected in both cryopreserved replicates in all cases (min = 64% and max = 77%; Fig. 4a). For pathways that were unique to FT replicates and not detected in the matching cryopreserved replicates, no common pathways were shared across the FT replicates from all six cases; however, a total of seven pathways were shared across three cases. Though this may reflect gene expression programs that might be affected by the cryopreservation

process, these pathways were mostly detected across different cell types, with the exception of the gene sets GO: 0016628 ('oxidoreductase activity') and GO:0016791 ('phosphatase activity'), which were unique to cancer/epithelial cells and T-cells from three FT replicates, respectively (Additional file 4). From the high concordance of GO pathways detected in cryopreserved replicates, we concluded that these minor differences were likely due to the variations in the scRNA-Seq platform or false discovery rather than the cryopreservation process.

We next assessed the variability of pathway enrichment scores for cryopreserved cells from each cluster (Fig. 4b–d). This analysis revealed minimal variability across clusters from all six cases of breast cancers, prostate cancers and melanoma, represented by the small range of -log10 q value enrichment scores for cells across FT and cryopreserved replicates (Fig. 4b–d; Additional file 2: Figure S6a-c). Taken together, these data indicate that the minor variations emerging from cryopreservation, as shown previously through clustering metrics (Fig. 1e–g; Fig. 2f–h), transcript detection (Fig.
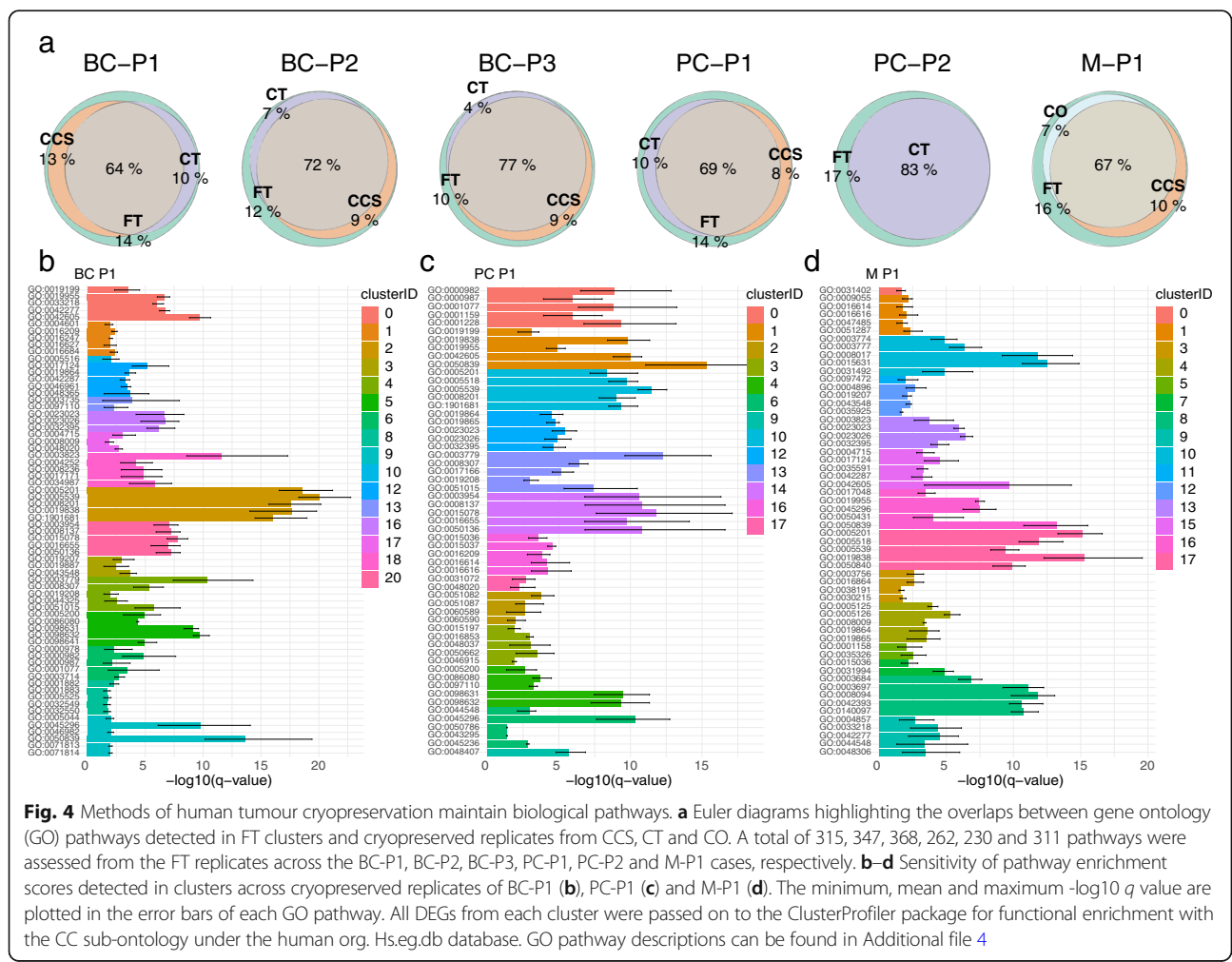

Fig. 4 Methods of human tumour cryopreservation maintain biological pathways. **a** Euler diagrams highlighting the overlaps between gene ontology (GO) pathways detected in FT clusters and cryopreserved replicates from CCS, CT and CO. A total of 315, 347, 368, 262, 230 and 311 pathways were assessed from the FT replicates across the BC-P1, BC-P2, BC-P3, PC-P1, PC-P2 and M-P1 cases, respectively. **b–d** Sensitivity of pathway enrichment scores detected in clusters across cryopreserved replicates of BC-P1 (**b**), PC-P1 (**c**) and M-P1 (**d**). The minimum, mean and maximum -log10 q value are plotted in the error bars of each GO pathway. All DEGs from each cluster were passed on to the ClusterProfiler package for functional enrichment with the CC sub-ontology under the human org. Hs.eg.db database. GO pathway descriptions can be found in Additional file 4

3a, b) and gene correlations (Fig. 3c, d), only have minor impacts on the conservation of key biological pathways.

Gene expression artefacts arising from tumour cryopreservation

While the previous analysis examined the conservation of biological pathways, we next examined if distinct transcriptional artefacts emerged from the cryopreservation process i.e. pathways that were enriched in cryopreserved conditions compared to their respective FT sample. We first examined this on the bulk level comparing all cells in FT vs all cells in either the CCS or CT samples using differential gene expression (Additional file 5) and GO pathway analysis (Additional file 6). We
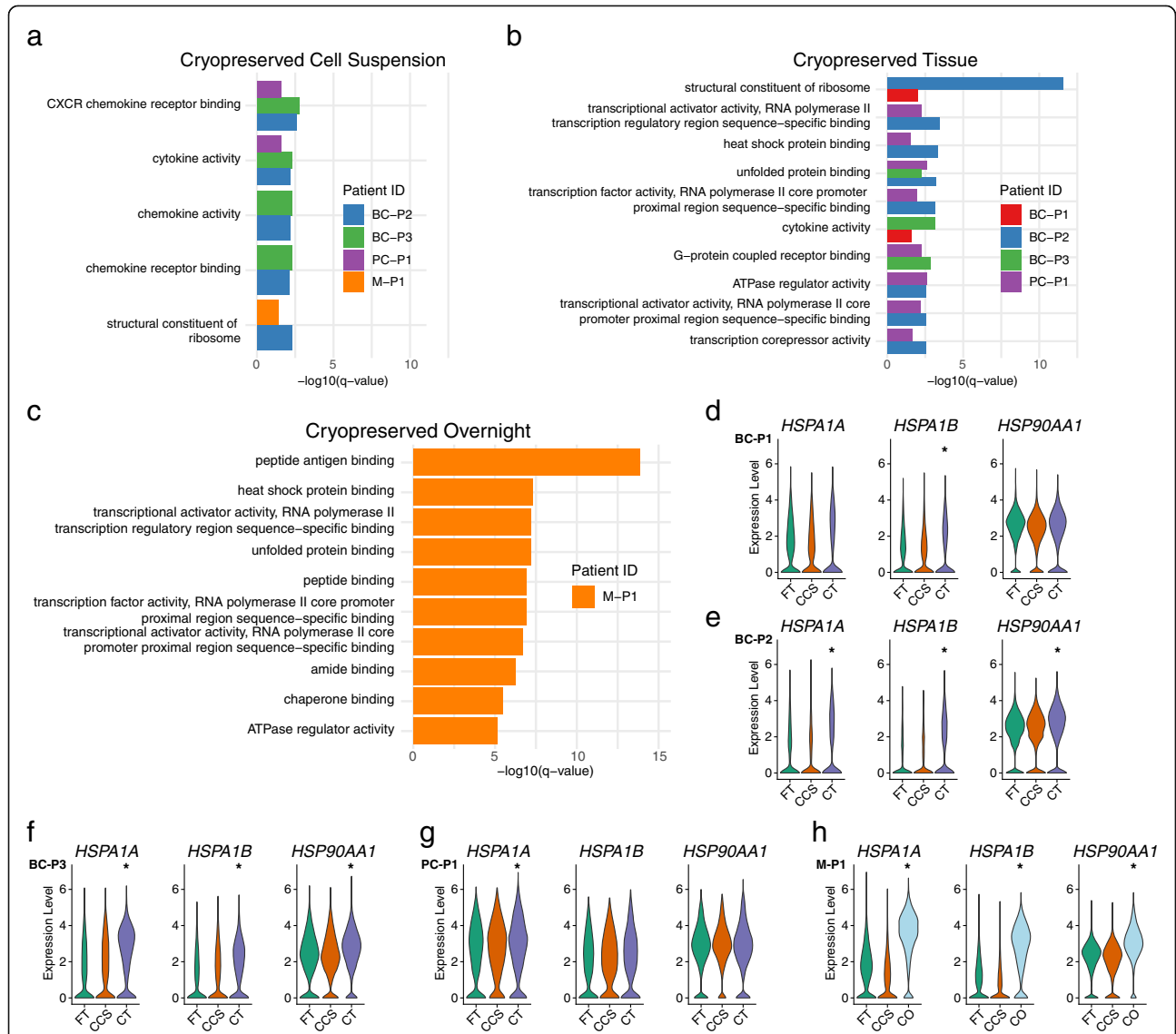
Fig. 5 Gene expression artefacts arising from cryopreservation. **a–c** Enrichment scores for gene ontology pathways that are unique to cryopreservation conditions: cryopreserved cell suspension (CCS; **a**), cryopreserved tissue (CT; **b**) and cryopreserved after overnight cold storage (CO; **c**). Comparisons were performed between all cells from each matched condition, which were first downsampled by total cell number and total number of sequencing reads. For the CCS (**a**) and CT (**b**) conditions, only pathways that were shared across multiple cases were analysed, which led to a total of 5 and 21 pathways for each condition, respectively. A total of 54 pathways were enriched in the CO (**c**) condition. Only the top 10 pathways based on enrichment scores are plotted for CT (**b**) and CO (**c**) conditions. DEGs from each condition (Additional file 5) were passed on to the ClusterProfiler package for functional enrichment with the CC sub-ontology under the human org.Hs.eg.db database. GO pathway descriptions can be found in Additional file 6. **d–h** Expression violin plots of the genes *HSPA1A*, *HSPA1B* and *HSP90AA1* from cell stress response pathways (heat shock protein binding GO:0031072 and unfolded protein binding GO:0051082) that were commonly enriched across CT and CO conditions. Tumours for BC-P1 (**d**), BC-P2 (**e**), BC-P3 (**f**), PC-P1 (**g**) and M-P1 (**h**) are grouped by their cryopreservation conditions: fresh tissue (FT), CCS, CT or CO. Asterisk indicates significance values where adjusted *p* values are less than 0.05, as calculated using the MAST method

focused on pathways that were detected in the same condition across two or more tumours. This revealed a total of 23 pathways that were unique to cells cryopreserved as either CCS or CT. Five pathways were enriched in cells cryopreserved as CCS (Fig. 5a), which were related to chemokine receptor activity, cytokine activity and ribosomal constituents however, these gene sets and enrichment scores were relatively small, consisting of 2–4 genes (Additional file 6). A total of 21 pathways were enriched in cells cryopreserved as CT (Fig. 5b), which included ribosomal constituents, transcriptional/RNA polymerase activity, cellular stress response (heat shock protein binding and unfolded protein binding) and ubiquitin ligase binding (Fig. 5b; Additional file 6). These pathways were not identified across all tumours analysed as CT and were mostly detected in two of the four tumours (Fig. 5b). For the melanoma sample (M-P1) that was uniquely dissociated and cryopreserved after overnight cold storage (CO), we identified a total of 54 pathways that were enriched in comparison to the FT and CCS (Fig. 5c). This revealed an enrichment of pathways related to peptide and antigen binding (MHC), cellular stress response (heat shock protein binding and unfolded protein binding) and transcriptional activity/RNA polymerase activity (Fig. 5c; Additional file 6).
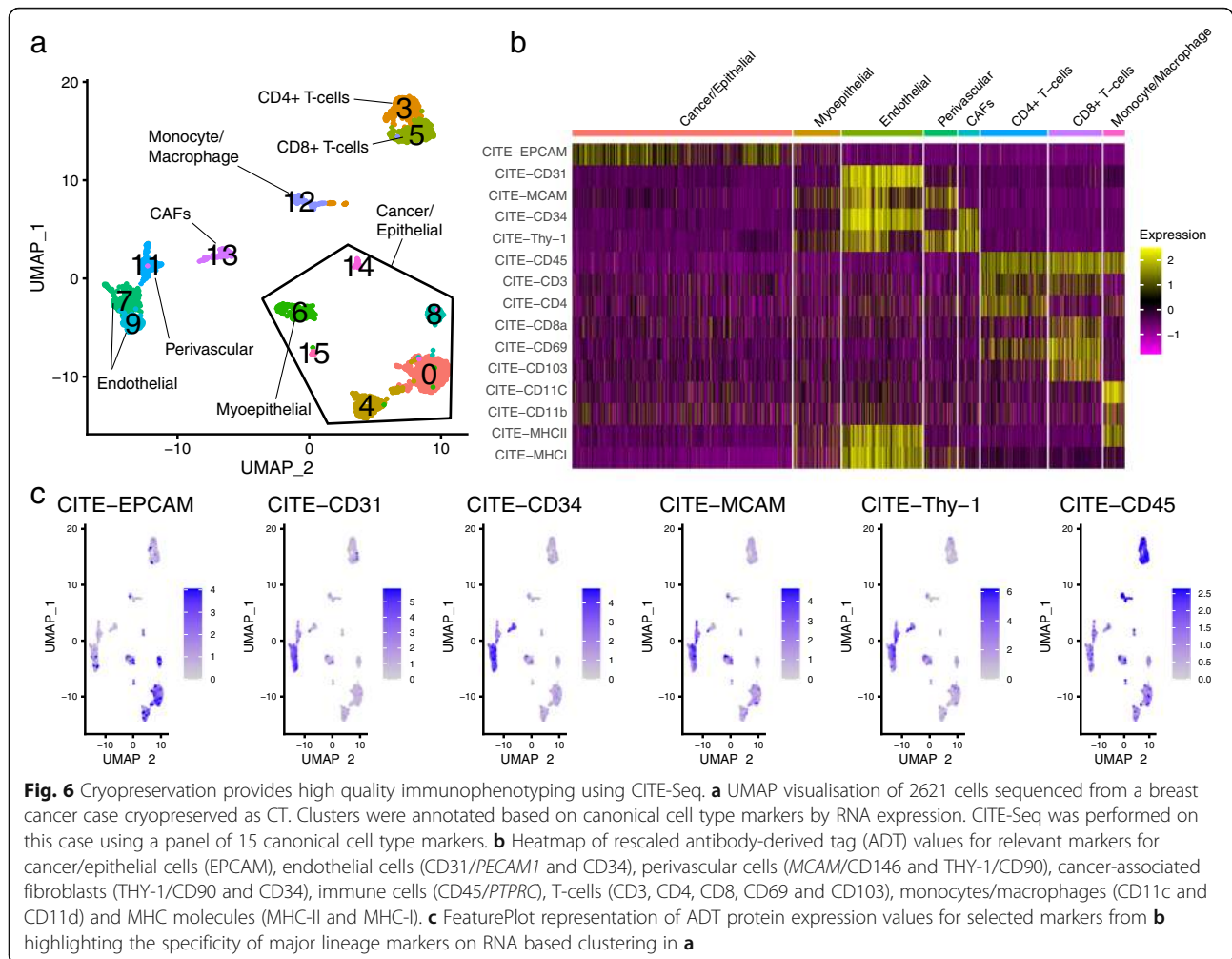
We noticed that cellular stress response pathways such as heat shock protein binding (GO:0031072) and unfolded protein binding (GO:0051082) were commonly enriched across CT and CO conditions, which included several heat-shock related genes including *HSPA1A*, *HSPA1B* and *HSP90AA1* (Additional file 6). Previous studies have reported similar transcriptional artefacts emerging from cryopreservation in the analysis of immune cells from PBMCs [15]. While these genes were upregulated in CT and CO conditions, they were generally robustly expressed across every condition and tumour analysed (Fig. 5d–h). As these heat shock-related pathways (Fig. 5b) and genes (Fig. 5d–g) were modestly upregulated in the CT condition of breast and prostate tumours, it suggests that cells which are first cryopreserved as solid tissue (CT) may undergo an added stress response during cryopreservation or when thawed and dissociated [20, 21] compared to cells processed immediately. It is worth noting that these differences were relatively minor and not detected across every tumour analysed. In contrast, we found a much stronger enrichment of these stress pathways (Fig. 5c) and genes (Fig. 5h) in the CO condition of the M-P1 tumour, suggesting that overnight cold storage of tumour tissue prior to lab processing results in more transcriptional artefacts related to cell stress responses.

When we repeated this analysis on the cluster and cell type level, we found that most of these impacted pathways, including cell stress responses and heat shock,

were well distributed across all cell types, suggesting that these transcriptional artefacts were not unique to a specific cell type (Additional file 2: Figure S6d-f; Additional file 7; Additional file 8). However, we did notice that the total number of pathways unique to cryopreserved cells were skewed towards cancer/epithelial cells, T cells and monocytes/macrophages, which had a total of 44, 14 and 8 pathways shared across the same condition in multiple cases (Additional file 2: Figure S6e-f; Additional file 8). Other cell types that had pathways shared across multiple cases were endothelial cells (5) and B cells (2), while no shared pathways for CAFs and perivascular cells were found (Additional file 2: Figure S6e-f; Additional file 8). Our previous analysis with batch correction (Fig. 1b–g, Fig. 2a–h) and without batch correction (Additional file 2: Figure S1a) found no unique clustering by cryopreservation condition, suggesting that these minor transcriptional changes have no substantial impact on the overall gene expression profile of cryopreserved cells; however, it does warrant caution when interpreting these specific stress response pathways in downstream analyses. Taken overall, cells cryopreserved as CT or CO may undergo an added stress response from cryopreservation or when thawed and dissociated. Consistent with our earlier findings, these conditions result in slightly more variable data compared to cells cryopreserved as suspensions immediately (CCS).

Whole cell cryopreservation allows for highly robust immunophenotyping using CITE-Seq

Immunophenotyping with barcoded-antibody methods such as CITE-Seq can be powerfully applied to simultaneously integrate protein and gene expression in single cells. Although previous studies have applied CITE-Seq to cryopreserved PBMCs, it is yet to be established whether CITE-Seq can be applied to cells cryopreserved as solid tissues [9]. As cell surface markers have been extensively used to characterise immune subpopulations, such additional layers of phenotypic information can be used to profile the tumour immune response in cryopreserved patient samples. Here, we performed CITE-Seq on a single breast cancer case cryopreserved as CT (Fig. 6a) using a panel of 15 canonical cell type markers. We first used a combination of canonical markers from RNA expression to broadly annotate clusters (Fig. 6a; Additional file 2: Figure S7a). From CITE-Seq, we were able to validate our cell type annotations by showing the highly specific antibody-derived tag (ADT) expression levels of canonical markers on corresponding cell types. For example, ADT levels of EPCAM on cancer/epithelial cells (c0, c4, c8, c14 and c15), CD31 (*PECAM1*) and CD34 on endothelial cells (c7 and c9), CD146 (*MCAM*) on perivascular cells (c11), CD90 (*THY1*) and CD34 on CAFs (*c13*) and CD45 (*PTPRC*) on immune cells (c3, c5

Fig. 6 Cryopreservation provides high quality immunophenotyping using CITE-Seq. **a** UMAP visualisation of 2621 cells sequenced from a breast cancer case cryopreserved as CT. Clusters were annotated based on canonical cell type markers by RNA expression. CITE-Seq was performed on this case using a panel of 15 canonical cell type markers. **b** Heatmap of rescaled antibody-derived tag (ADT) values for relevant markers for cancer/epithelial cells (EPCAM), endothelial cells (CD31/*PECAM1* and CD34), perivascular cells (*MCAM*/CD146 and THY-1/CD90), cancer-associated fibroblasts (THY-1/CD90 and CD34), immune cells (CD45/*PTPRC*), T-cells (CD3, CD4, CD8, CD69 and CD103), monocytes/macrophages (CD11c and CD11d) and MHC molecules (MHC-II and MHC-I). **c** FeaturePlot representation of ADT protein expression values for selected markers from **b** highlighting the specificity of major lineage markers on RNA based clustering in **a**

and c12) (Fig. 6b, c; Additional file 2: Figure S7a). Within the immune compartments, CD3 specifically marked T cells, while CD4 and CD8 were more specifically expressed on the respective T cell subpopulations (Fig. 6b; Additional file 2: Figure S7a). ADT levels of the activation marker CD69 and tissue resident marker CD103 were heterogeneously expressed on T cell subpopulations (Fig. 6b). CD11c and CD11d were highly specific to monocyte/macrophage cell clusters (Fig. 6b). Major histocompatibility complexes, MHC-II and MHC-I, were highly expressed by endothelial cells, whereas MHC-II was also detected on monocyte/macrophage clusters (Fig. 6b).

ADT levels, which overcome several technical limitations from gene drop-out, have a greater sensitivity than UMI counts by scRNA-Seq. The average correlation between ADT levels and the corresponding gene expression for this panel of 15 markers was 0.215 (min R^2 = 0.003 and max R^2 = 0.639; Additional file 2: Figure S7b). This ranged significantly for different markers, particularly for lowly expressed immunoregulatory molecules

such as CD4 (*CD4*), CD103 (*ITGAE*), CD11b (*ITGAM*) and CD11c (*ITGAX*), where expression levels of their corresponding genes were lowly detected in comparison to the ADT, with R^2 values of 0.016, 0.005, 0.019 and 0.004, respectively (Additional file 2: Figure S7b). In contrast, highly expressed genes such as the endothelial cell marker CD31 (*PECAM1*) showed much higher correlations (R^2 = 0.639; Additional file 2: Figure S7b). In summary, we show that good quality CITE-Seq data can be generated from cells cryopreserved as solid CT. Such methods can be used to powerfully extract additional phenotypic information from low amounts of cryopreserved clinical tissue, aiding the annotation of single-cell clusters and the detection of clinically relevant molecules such as immune-checkpoints.

Discussion

We show that high-quality scRNA-Seq data can be generated from human cancer samples cryopreserved as dissociated single-cell suspensions (CCS) and solid tissues (CT). For the latter, minimal processing is required

following sample collection and can be conducted routinely in hospital pathology laboratories that have access to −80 °C freezers for short-term storage. These samples can later be transported to research laboratories for long-term storage or further processing. We did however find that CCS samples yielded slightly higher quality data; however, this requires more specialised tissue processing following sample collection before cryopreservation (~ 1–2 h using commercial dissociation kits). The minor decrease in the quality of tumours sequenced as CT is likely due to a reduced efficiency of the cryoprotectant in solid tissue fragments, with the small possibility that these cells are more fragile and prone to transcriptional changes during tissue dissociation. While we used tissues that had been cryopreserved for up to 6 weeks in this study, we have routinely processed samples stored at liquid nitrogen for more than 3 years for scRNA-Seq. Our similar findings for the metastatic melanoma cryopreserved after overnight cold storage warrants some caution for specific heat-shock and stress response pathways in downstream analyses using this method.

Most importantly, we show that the complexity of the TME is conserved following cryopreservation as both CCS and CT. This is an important consideration because an integrated understanding of the neoplastic, stromal and immune states defines tumours and their response to treatment. A limitation of our study is the comparisons of cancers that are mostly of the adenocarcinoma histopathology; however, these cryopreservation methods are likely applicable to other cancer types which may require slightly modified sample preparation and dissociation protocols. Further, we show that multi-omics methods, such as immunophenotyping using CITE-Seq, can be performed using cells cryopreserved as solid tissue pieces, which is impossible when using other preservation methods such as single nuclei sequencing from snap frozen tissues. Our CITE-Seq data is limited to a single breast cancer sample, and future comparisons with data generated from additional fresh tissues can be used to further assess the impact of cryopreservation on CITE-Seq and the integrity of cell-surface proteins. Lastly, our findings show that sample multiplexing methods can be applied to cryopreserved clinical samples to reduce cost and logistics for project scaling, including barcode hashing or genotype based demultiplexing (unpublished data) [22, 23]. Such methods can also be used to further minimise batch effects in larger patient cohorts, allowing for more samples to be simultaneously thawed, processed and sequenced in a single run.

Conclusions

We show that the viable cryopreservation of human cancers provides high-quality single-cells for multi-omics analysis. This can guide experimental designs for tissue biobanking protocols for future clinical scRNA-Seq

studies. Due to the easily adoptable nature of cryopreserving solid tissues in tissue biobanking processes, we envisage our findings to positively impact the sample collection opportunities for future clinical studies, particularly for multi-site collaborative studies, to allow for the centralisation of sample processing and batched analysis.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13073-021-00885-z.

Additional file 1. Clinical information. Clinical details for breast cancer, prostate cancer and metastatic melanoma cases used in this study.

Additional file 2: Supplementary Figures. **Figures S1-S7.**

Additional file 3. Differentially expressed genes for clusters. Differentially expressed genes for integrated clusters. Differential gene expression was performed using the MAST method within Seurat v3 with the RNA assay and default parameters.

Additional file 4. Cluster metrics and annotations. Cluster metric standard deviations, cluster annotations, cluster gene correlations and gene pathways conserved across cryopreserved conditions. a, Standard deviations for silhouette scores, mixing metrics and local structure metrics, computed for the comparisons between the down sampled cells from Fresh Tissue (FT) with cells from FT (positive control), Cryopreserved Cell Suspensions (CCS), Cryopreserved Tissue (CT) and Cryopreserved Overnight (CO). b, Integrated cluster annotations. c, Cluster level correlation values. Adjusted-R^2 values computed using linear regression in R to model log-normalised gene expression values between integrated clustered cells from different cryopreserved replicates. d, Conservation of cluster Gene Ontology (GO) pathways following cryopreservation. Functional enrichment was first performed for all differentially expressed genes in each integrated cluster in FT samples. Enrichment scores for these pathways were then compared to the respective integrated cluster in each cryopreserved replicate (CCS, CT, CO). All DEGs from each cluster were passed on to the ClusterProfiler package for functional enrichment with the CC sub-ontology under the human org.Hs.eg.db database. Specific GO pathways that were unique to FT clusters, CCS vs CT clusters or CT vs CCS clusters are listed with the cluster and number of tumours that they were identified in.

Additional file 5. Differentially expressed genes for all cells across cryopreservation conditions. Differential gene expression was performed using the MAST method within Seurat v3 with the RNA assay and default parameters. Differential gene expression was performed using integrated data. Data from all conditions from each matched tumour datasets were down sampled to the same number of mapped sequencing reads using CellRanger. The number of cells in each condition was randomly down sampled to match the condition with the lowest number of cells.

Additional file 6. Gene pathways for all cells across cryopreservation conditions. Shared Gene Ontology (GO) pathways for each cryopreservation condition across the five tumours analysed. Only pathways detected in the same condition in more than two tumours were analysed. All DEGs from the comparison of all cells across cryopreservation conditions (Additional file 5) were passed on to the ClusterProfiler package for functional enrichment with the CC sub-ontology under the human org.Hs.eg.db database.

Additional file 7. Differentially expressed genes for each cell type across cryopreservation conditions. Differential gene expression was performed using the MAST method within Seurat v3 with the RNA assay and default parameters. Differential gene expression was performed using integrated cluster data. Data from all conditions from each matched tumour dataset were down sampled to the same number of mapped sequencing reads using CellRanger. The number of cells in each condition was randomly down sampled to match the condition with the lowest number of cells.

Additional file 8. Gene pathways for each cell type across cryopreservation conditions. Shared Gene Ontology (GO) pathways for

Wu *et al. Genome Medicine*   (2021) 13:81

Page 16 of 17

each cell type and cryopreservation condition across the five tumours analysed. Only pathways detected in the same cell type and cryopreservation condition in more than two tumours were analysed. All DEGs from the comparison of annotated cell types across cryopreservation conditions (Additional file 7) were passed on to the ClusterProfiler package for functional enrichment with the CC sub-ontology under the human org.Hs.eg.db database.

Acknowledgements

Authors' contributions

A.S. conceived the project and directed the study. S.Z.W. and A.S. wrote the manuscript. All authors reviewed the drafting of the manuscript. E.L., S.O.T., M.N.H., E.A.M., J.B., D.S., C.M. and S.W. organised the access to breast cancer patient tissue. L.H., A.M.J. and P.S. organised the access to patient tissue from prostate cancer patients. G.V.L., R.A.S., J.S. and C.Q. organised the access to melanoma tissue. K.H. collected clinical samples. S.Z.W., K.H. and A.C. performed tumour dissociation for scRNA-Seq. C. C and S.Z.W. performed all Chromium single-cell experiments. C.C. helped perform library preparations and next-generation sequencing of the scRNA-Seq libraries. S.Z.W. performed and interpreted the pre-processing and downstream analysis of the scRNA-Seq data. D.R. supervised the scRNA-Seq analysis. G.A. performed the CITE-Seq staining experiments. B.Z.Y. provided guidance on CITE-Seq experimental parameters and sequencing. N.B. performed bioinformatics analysis of the CITE-Seq experiments. E.L., S.J. and J.P. provided intellectual input. The authors read and approved the final manuscript.

Funding

Availability of data and materials

The scRNA-Seq data from this study has been deposited in the European Genome-Phenome Archive (EGA), which is hosted by the EBI and the CRG, under the accession code EGAS00001005115 [24]. This repository contains demultiplexed paired ended reads (R1 and R2) and Illumina indices processed using the Cellranger software. All code related to the scRNA-Seq analysis can be found on github at https://github.com/sunnyzwu/cryopreservation_scRNAseq [25]. All other relevant data are available from the authors upon request.

Declarations

Ethics approval and consent to participate

Patient tissues used in this work were collected under protocols × 12–0231, × 13–0133, × 16–018, × 17–0312 and × 17–155. The Human Research Ethics Committee approval was obtained through the Sydney Local Health District Ethics Committee, Royal Prince Alfred Hospital zone and the St Vincent's Hospital Ethics Committee. Written consent was obtained from all patients

prior to collection of tissue and clinical data stored in a de-identified manner, following pre-approved protocols. Consent into the study included the agreement to the use of all patient tissue and data for publication. The research described in this study conforms with the principles of the Helsinki Declaration.

Consent for publication

Not applicable.

Competing interests

No conflicts of interests. R.A.S. has received fees for professional services from Merck Sharp & Dohme, GlaxoSmithKline Australia, Bristol-Myers Squibb, Novartis, Myriad, NeraCare and Amgen. G.V.L. is consultant advisor to Aduro, BMS, Mass-Array, Merck MSD, Novartis, Pierre-Fabre, Roche, Sandoz and Qbiotics. The remaining authors declare that they have no competing interests.

Author details

[1]The Kinghorn Cancer Centre and Cancer Research Division, Garvan Institute of Medical Research, Darlinghurst, NSW, Australia. [2]St Vincent's Clinical School, Faculty of Medicine UNSW, Sydney, NSW, Australia. [3]Garvan-Weizmann Centre for Cellular Genomics, Garvan Institute of Medical Research, Sydney, Australia. [4]Chris O'Brien Lifehouse, Camperdown, NSW, Australia. [5]NSW Health Pathology, Department of Anatomical Pathology, St George Hospital, Kogarah, NSW, Australia. [6]School of Medical Sciences, UNSW Sydney, Kensington, Australia. [7]Faculty of Medicine & Health Sciences, Sydney Western University, Campbelltown, NSW, Australia. [8]St George & Sutherland Clinical School, UNSW Sydney, Kensington, Australia. [9]University of Sydney, Camperdown, NSW, Australia. [10]St Vincent's Hospital, Darlinghurst, NSW, Australia. [11]Melanoma Institute Australia, The University of Sydney, Sydney, Australia. [12]The University of Sydney, Sydney, Australia. [13]Royal North Shore Hospital, St Leonards, NSW, Australia. [14]Tissue Pathology and Diagnostic Oncology, Royal Prince Alfred Hospital, Sydney, Australia. [15]BioLegend, San Diego, CA, USA. [16]Department of Breast Surgery, Chris O'Brien Lifehouse, Camperdown, NSW 2050, Australia. [17]Royal Prince Alfred Institute of Academic Surgery, Sydney University, Sydney, Australia. [18]UNSW Cellular Genomics Futures Institute, University of New South Wales, Sydney, Australia.

References

1. Hanahan D, Coussens LM. Accessories to the crime: functions of cells recruited to the tumor microenvironment. Cancer Cell. 2012;21(3):309–22. https://doi.org/10.1016/j.ccr.2012.02.022.
2. Wu SZ, Roden DL, Wang C, Holliday H, Harvey K, Cazet AS, Murphy KJ, Pereira B, Al-Eryani G, Bartonicek N, et al. Stromal cell diversity associated with immune evasion in human triple-negative breast cancer. EMBO J. 2020; 39:e104063.
3. Madissoon E, Wilbrey-Clark A, Miragaia RJ, Saeb-Parsy K, Mahbubani KT, Georgakopoulos N, Harding P, Polanski K, Huang N, Nowicki-Osuch K, et al. scRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. Genome Biol. 2019;21:1.
4. Tran HTN, Ang KS, Chevrier M, Zhang X, Lee NYS, Goh M, Chen J. A benchmark of batch-effect correction methods for single-cell RNA sequencing data. Genome Biol. 2020;21(1):12. https://doi.org/10.1186/s13 059-019-1850-9.
5. Chen J, Cheung F, Shi R, Zhou H, Lu W, Consortium CHI. PBMC fixation and processing for Chromium single-cell RNA sequencing. J Transl Med. 2018; 16(1):198. https://doi.org/10.1186/s12967-018-1578-4.
6. Attar M, Sharma E, Li S, Bryer C, Cubitt L, Broxholme J, Lockstone H, Kinchen J, Simmons A, Piazza P, Buck D, Livak KJ, Bowden R. A practical solution for preserving single cells for RNA sequencing. Sci Rep. 2018;8(1):2151. https://doi.org/10.1038/s41598-018-20372-7.
7. Wohnhaas CT, Leparc GG, Fernandez-Albert F, Kind D, Gantner F, Viollet C, Hildebrandt T, Baum P. DMSO cryopreservation is the method of choice to preserve cells for droplet-based single-cell RNA sequencing. Sci Rep. 2019; 9(1):10699. https://doi.org/10.1038/s41598-019-46932-z.
8. Habib N, Avraham-Davidi I, Basu A, Burks T, Shekhar K, Hofree M, Choudhury SR, Aguet F, Gelfand E, Ardlie K, Weitz DA, Rozenblatt-Rosen O, Zhang F,

Regev A. Massively parallel single-nucleus RNA-seq with DroNc-seq. Nat Methods. 2017;14(10):955–8. https://doi.org/10.1038/nmeth.4407.

9. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, Satija R, Smibert P. Simultaneous epitope and transcriptome measurement in single cells. Nat Methods. 2017;14(9):865–8. https://doi.org/10.1038/nmeth.4380.

10. Plass M, Solana J, Wolf FA, Ayoub S, Misios A, Glazar P, Obermayer B, Theis FJ, Kocks C, Rajewsky N. Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. Science. 2018;360(6391):eaaq1723. https://doi.org/10.1126/science.aaq1723.

11. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun. 2017;8(1):14049. https://doi.org/10.1038/ncomms14049.

12. Guillaumet-Adkins A, Rodriguez-Esteban G, Mereu E, Mendez-Lago M, Jaitin DA, Villanueva A, Vidal A, Martinez-Marti A, Felip E, Vivancos A, et al. Single-cell transcriptome conservation in cryopreserved cells and tissues. Genome Biol. 2017;18(1):45. https://doi.org/10.1186/s13059-017-1171-9.

13. Donlin LT, Rao DA, Wei K, Slowikowski K, McGeachy MJ, Turner JD, Meednu N, Mizoguchi F, Gutierrez-Arcelus M, Lieb DJ, et al. Methods for high-dimensional analysis of cells dissociated from cryopreserved synovial tissue. Arthritis Res Ther. 2018;20(1):139. https://doi.org/10.1186/s13075-018-1631-y.

14. Picelli S, Faridani OR, Bjorklund AK, Winberg G, Sagasser S, Sandberg R. Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc. 2014;9(1): 171–81. https://doi.org/10.1038/nprot.2014.006.

15. Lee JS, Yi K, Ju YS, Shin EC. Effects of cryopreservation and thawing on single-cell transcriptomes of human T cells. Immune Netw. 2020;20(4):e34. https://doi.org/10.4110/in.2020.20.e34.

16. Lun ATL, Riesenfeld S, Andrews T, Dao TP, Gomes T, participants in the 1st Human Cell Atlas J, Marioni JC. EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. Genome Biol. 2019;20:63.

17. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, Hao Y, Stoeckius M, Smibert P, Satija R. Comprehensive integration of single-cell data. Cell. 2019;177(7):1888–902 e1821. https://doi.org/10.1016/j.cell.2019.05.031.

18. Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, Slichter CK, Miller HW, McElrath MJ, Prlic M, et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol. 2015;16(1):278. https://doi.org/10.1186/s13059-015-0844-5.

19. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284–7. https://doi.org/10.1089/omi.2011.0118.

20. O'Flanagan CH, Campbell KR, Zhang AW, Kabeer F, Lim JLP, Biele J, Eirew P, Lai D, McPherson A, Kong E, et al. Dissociation of solid tumor tissues with cold active protease for single-cell RNA-seq minimizes conserved collagenase-associated stress responses. Genome Biol. 2019;20(1):210. https://doi.org/10.1186/s13059-019-1830-0.

21. Denisenko E, Guo BB, Jones M, Hou R, de Kock L, Lassmann T, Poppe D, Clement O, Simmons RK, Lister R, Forrest ARR. Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. Genome Biol. 2020;21(1):130. https://doi.org/10.1186/s13059-020-02048-6.

22. Kang HM, Subramaniam M, Targ S, Nguyen M, Maliskova L, McCarthy E, Wan E, Wong S, Byrnes L, Lanata CM, Gate RE, Mostafavi S, Marson A, Zaitlen N, Criswell LA, Ye CJ. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. Nat Biotechnol. 2018;36(1):89–94. https://doi.org/10.1038/nbt.4042.

23. Stoeckius M, Zheng S, Houck-Loomis B, Hao S, Yeung BZ, Mauck WM 3rd, Smibert P, Satija R. Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. Genome Biol. 2018;19(1):224. https://doi.org/10.1186/s13059-018-1603-1.

24. Swarbrick A. Cryopreservation of human cancers conserves tumour heterogeneity for single-cell multi-omics analysis (EGAS00001005115). European Genome-Phenome Archive. https://ega-archive.org/studies/EGAS00001005115 (2021).

25. Wu SZ. cryopreservation_scRNAseq. Github. https://github.com/sunnyzwu/cryopreservation_scRNAseq (2021).

Publisher's Note