RevGel-seq: instrument-free single-cell RNA sequencing using a reversible hydrogel for cell-specific barcoding

Jun Komatsu[1], Alba Cico[1], Raya Poncin[1], Maël Le Bohec[1], Jörg Morf[1,3], Stanislav Lipin[1], Antoine Graindorge[1], Hélène Eckert[1], Azadeh Saffarian[1], Léa Cathaly[1], Frédéric Guérin[1], Sara Majello[1], Damien Ulveling[1], Anaïs Vayaboury[1], Nicolas Fernandez[1], Dilyana Dimitrova[1], Xavier Bussell[1], Yannick Fourne[1], Pierre Chaumat[1], Barbara André[1], Elodie Baldivia[1], Ulysse Godet[1], Mathieu Guinin[1], Vivien Moretto[1], Joy Ismail[1], Olivier Caille[1], Natacha Roblot[2], Carine Beaupère[2], Alexandrine Liboz[2], Ghislaine Guillemain[2], Bertrand Blondeau[2], Pierre Walrafen[1,4], Stuart Edelstein[1,4]

[1]Scipio bioscience, Paris, France. [2]Sorbonne Université, Inserm, Centre de Recherche Saint-Antoine, CRSA, F-75012 Paris, France. [3]Present address: Skyhawk Therapeutics, Basel, Switzerland. [4]Corresponding authors: pierre.walrafen@scipio.bio, stuart.edelstein@scipio.bio

ABSTRACT

Progress in sample preparation for scRNA-seq is reported based on RevGel-seq, a reversible-hydrogel technology. Barcode bead—cell tandems stabilized by a chemical linker are dispersed in the hydrogel in the liquid state. Upon gelation the tandems are immobilized, cell lysis is triggered by detergent diffusion, and RNA molecules are captured on the adjacent barcode beads. After reverse transcription and preparation for cDNA sequencing, bioinformatic analysis reveals performance quality comparable to microfluidic-based technologies.

Research in many areas of biology focus on single-cell studies, particularly single-cell RNA sequencing (scRNA-seq) mainly using microfluidic-based methods[1-4], with the 10x Genomics Chromium instruments achieving particularly wide usage[5]. Comparative benchmarking has been reported for many of the available methods (including CEL-seq2, MARS-seq, Quartz-seq2, gmcSCRB-seq, Smart-seq2, C1HT-small, C1HT-medium, Chromium, ddSEQ, Drop-seq, ICELL8, and inDrop)[6, 7]. Major efforts using scRNA-seq have been applied to define atlases, notably for human and mouse cells[8, 9], aided by technical optimization of methodologies[6, 10]. Other investigations have focused on specific organs, often in relation to disease states[11], especially cancer immunotherapy[12], as well as on development[13], various forms of multiomics[14], and spatial transcriptomics[15]. While these applications have produced important new findings, further optimization of scRNA-seq could include the following properties:

1. Instrument-free, to avoid an up-front investment for a limited number of trial experiments;
2. Monitoring of initial barcode bead-cell coupling step to assure sufficient yield to proceed;
3. Presence of convenient stopping points to permit interruption of processing for later analysis;
4. Scalable for different numbers of cells in a sample, as well as cells of different sizes;
5. Storage of all ready-to-use reagents and consumables in individual laboratories for immediate usage whenever required;
6. Independent of any requirement to book timeslots on an institutional platform-based instrument;

RevGel-seq is a scRNA-seq preparation method based on a reversible hydrogel technology that provides advances on the above points. The workflow, presented in Fig. 1a, comprises liquid-gel-liquid transitions of the reversible hydrogel that contains cells linked to barcode-beads. Sequencing outputs are evaluated and displayed using Cytonaut, an interactive data analysis platform that includes pre- and post-processing, as well as interactive data visualization (see Fig. 1b). Results establish RevGel-seq performance levels, in comparison to the predominant microfluidics technology, and illustrate a clinically relevant application.

Benchmarking and biological applications. The performance of RevGel-seq was assessed using a barnyard plot for mouse NIH3T3 cells and human HEK293 cells mixed in equal proportions. From 10,000 input cells, an output of 4,500-4,700 cells (for each of three technical replicates) were detected in sequencing data. As shown in Fig. 2a (for one of the technical replicates), the transcripts from the two types of cells were readily separated, with a cell multiplet rate of 6% obtained by doubling the measured hetero-species cell multiplet rate of 3% to account for an equal number of non-detectable homo-species. A hetero-species cell multiplet is defined as a barcode for which more than 1/6 of the detected transcripts correspond to the opposite species.

To evaluate RevGel-seq on a sample with multiple cell types, analyses were performed on peripheral blood mononuclear cells (PBMCs)[7]. PBMC automated classification presented in Fig. 2b reveals that RevGel-seq and a 10x Chromium instrument (using the standard assay) with single-cell 3' reagent kit (v3.1) both identified the same cell types with similar distributions (Fig. 2c, 4,560 cells for 10x sample versus 3,172 cells for RevGel-Seq sample). Finally, to test RevGel-seq on a current research project, pancreatic islet cells were phenotyped in relation to studies on insulin resistance and islet cell adaptation[16]. The automated classification of pancreatic cell types (Fig. 2d) and associated gene counts (Fig. 2e) are in accord with previous studies[17].

General considerations. Recent developments in scRNA-seq methodologies have favored approaches that do not rely on a specific instrument. The RevGel-seq approach progresses significantly on the six properties presented above to improve scRNA-seq:

1. Instrument-free achieved with reversible hydrogel system thereby avoiding an upfront investment;
2. Initial monitoring shown for cardiomyocytes (Supplementary Figures 2 and 3) and beads can be counted after the PCR step (Section I-7-vi in Online Methods);
3. Convenient stopping points as noted in Supplementary Figure 4;
4. Scalability for size illustrated with cardiomyocytes (Supplementary Figures 2 and 3) and for number, up to 30,000 cells, with no loss of performance (Online Section IIB);
5. Ready-to-use reagents (details in Online Methods);
6. Platform-free to be independent of facilities schedules clearly demonstrated;

While other approaches recently proposed include Hive (PerkinElmer) based on nanowells[18]; Parse Biosciences based on split and pool barcoding[19]; and Fluent Biosciences (PIP-seq) based on templated emulsions[20], the RevGel-seq remains the only scRNA-seq technology based on a reversible hydrogel, with the advantages of scalability of cell size and cell number, as well as early stopping points. Other comparisons will be more fully developed in future publications.

Online Methods: appended.

FIGURE LEGENDS:

Fig.1 | Workflow and Analysis Pipeline. a, RevGel-seq workflow steps for sample preparation to characterize scRNA-seq. Barcode beads and cells are attached via a bifunctional chemical linker. These tandems are dispersed in the hydrogel in the liquid state and immobilized upon gelation. Following cell lysis, RNA molecules are captured on the barcodes. After reverse transcription, barcoded cDNAs are PCR-amplified. Further details are presented in the Online Methods. **b,** Workflow of the end-to-end data processing pipeline integrated in the Cytonaut platform. The pre-processing phase inputs the raw sequencing data (FASTQ files) and outputs quality indicators and count matrices, followed by post-processing that inputs the count matrices to perform cell clustering, 2D embedding, and differential gene expression. The Cytonaut-Rover module enables interactive data visualization. Additional details in the Online Methods.

Fig. 2 | Benchmarking and applications. a, Barnyard plot showing for each cell-associated barcode the number of detected mouse NIH3T3 transcripts and the number of detected human HEK293 transcripts, from 10,000 input cells prepared with RevGel-seq and the number of reads/cells down-sampled to a depth of 50,000 raw reads on average. The cell multiplet rate is 6% (including both hetero- and homo-species cell multiplets); for details on criteria see ONLINE METHODS. **b,** Cell classification and 2D embedding from a PBMC sample of 10,000 input cells prepared with RevGel-seq (with doubling of the ratio of beads/cell to increase coupling yield) and sequenced at a depth of 20,000 raw reads in cells on average. Automated cell classification was performed using the SingleR algorithm of Seurat. Unassigned cells had classification uncertainty too high according to the pruneScores method with default parameters. All relevant PBMC sub-types were identified. **c,** Percentage of automatically classified cells for each cell type identified in the same PBMC sample prepared with a reference microfluidic technology (left) and prepared with RevGel-seq (right), with 10,000 input cells and a sequencing depth of 20,000 raw reads in cells on average. The relative proportions of cell types are highly similar between RevGel-seq and the reference microfluidic technology. **d,** Pancreatic islet cells (176,000 raw reads per cell) automatically classified into cell types (left) following the same methodology as in Fig. 2b except for the reference dataset BaronPancreasData[17]. **e,** Percentage of cells belonging to each identified cell type and box plots showing the distribution of gene diversity per cell type for pancreatic islet cells classified in Fig. 2d.

ACKNOWLEDGEMENTS: We gratefully acknowledge helpful discussion with Wilko Duprez, Sahar Sabetnia, Marie-Claude Marchand, Yannick Marie, Laetitia Strehl, Emeline Mundwiller, Justine Guégan, Allison Mallory, Jean-David Bouaziz, and Arturo Londono.

References:

1. Macosko, E.Z. et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214 (2015).
2. Klein, A.M. & Macosko, E. InDrops and Drop-seq technologies for single-cell sequencing. *Lab Chip* **17**, 2540-2541 (2017).
3. Quake, S.R. The cell as a bag of RNA. *Trends Genet* **37**, 1064-1068 (2021).
4. Cable, J. et al. Single cell biology-a Keystone Symposia report. *Ann N Y Acad Sci* **1506**, 74-97 (2021).
5. Zheng, G.X. et al. Massively parallel digital transcriptional profiling of single cells. *Nature communications* **8**, 14049 (2017).
6. Mereu, E. et al. Benchmarking single-cell RNA-sequencing protocols for cell atlas projects. *Nature biotechnology* **38**, 747-755 (2020).
7. Davis-Marcisak, E.F. et al. From bench to bedside: Single-cell analysis for cancer immunotherapy. *Cancer Cell* **39**, 1062-1080 (2021).
8. Han, X. et al. Construction of a human cell landscape at single-cell level. *Nature* **581**, 303-309 (2020).
9. Lahnemann, D. et al. Eleven grand challenges in single-cell data science. *Genome biology* **21**, 31 (2020).
10. Luecken, M.D. et al. Benchmarking atlas-level data integration in single-cell genomics. *Nature methods* **19**, 41-50 (2022).
11. Chalasani, N. et al. Noninvasive stratification of nonalcoholic fatty liver disease by whole transcriptome cell-free mRNA characterization. *Am J Physiol Gastrointest Liver Physiol* **320**, G439-g449 (2021).
12. Samad, T. & Wu, S.M. Single cell RNA sequencing approaches to cardiac development and congenital heart disease. *Semin Cell Dev Biol* **118**, 129-135 (2021).
13. Haniffa, M. et al. A roadmap for the Human Developmental Cell Atlas. *Nature* **597**, 196-205 (2021).
14. Ding, J. et al. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nature biotechnology* **38**, 737-746 (2020).
15. Marx, V. Method of the Year: spatially resolved transcriptomics. *Nature methods* **18**, 9-14 (2021).
16. Beaupere, C., Liboz, A., Fève, B., Blondeau, B. & Guillemain, G. Molecular Mechanisms of Glucocorticoid-Induced Insulin Resistance. *Int J Mol Sci* **22** (2021).
17. Baron, M. et al. A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst* **3**, 346-360.e344 (2016).
18. Gierahn, T.M. et al. Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nature methods* **14**, 395-398 (2017).
19. Rosenberg, A.B. et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176-182 (2018).
20. Clark, I.C., et al. Microfluidics-free single-cell genomics with templated emulsification. *bioRxiv*, doi: 10.1101/2022.06.10.495582 (2022).

Fig. 1

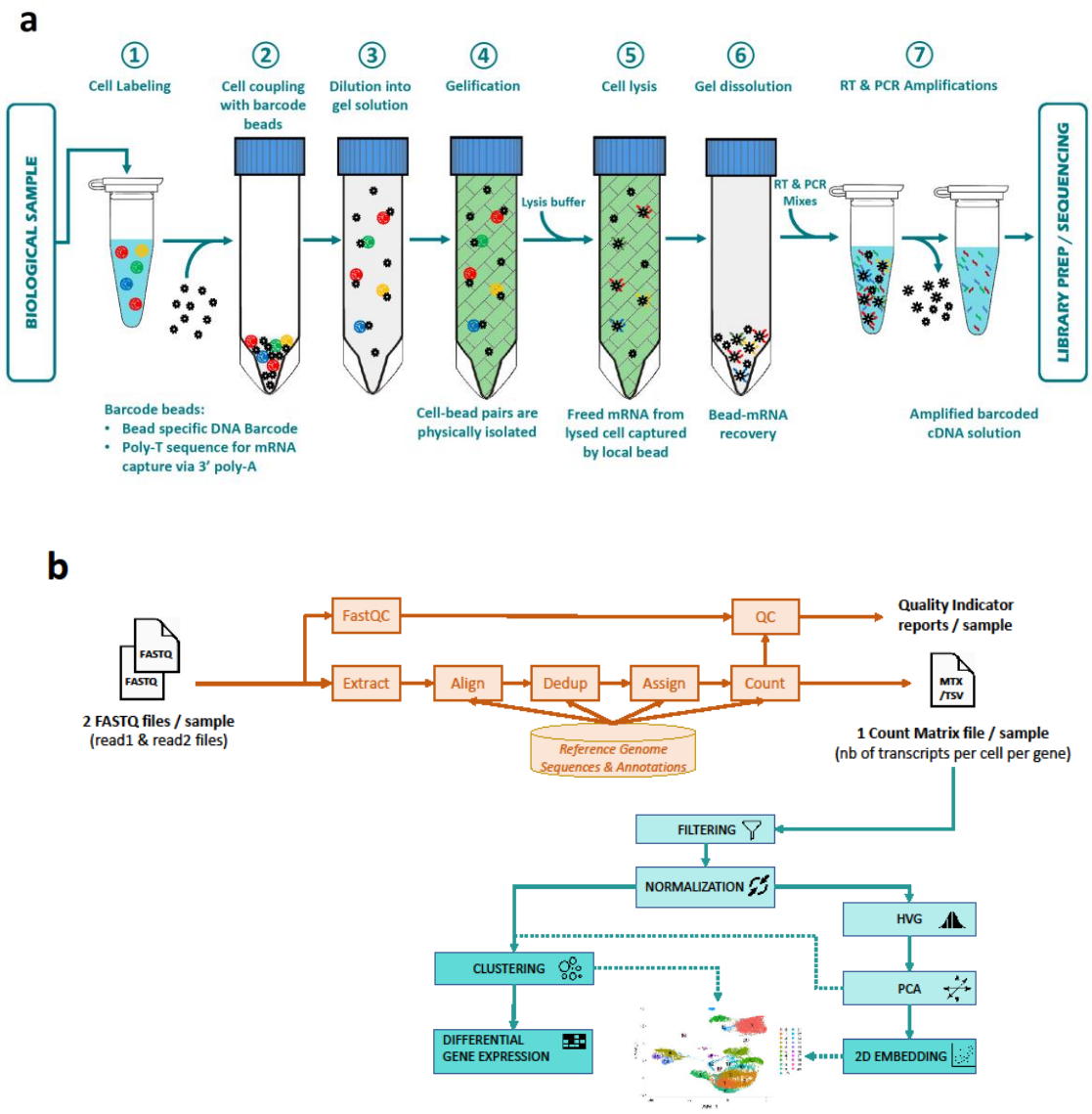
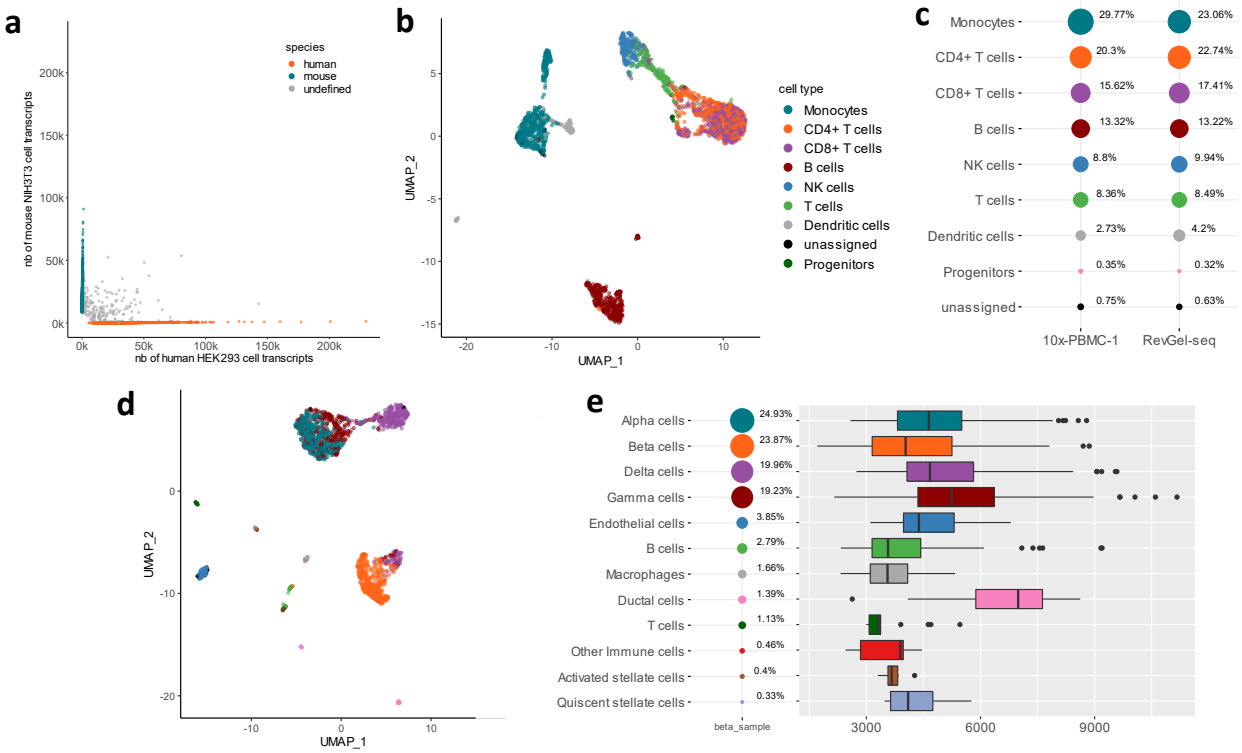

a
1 Cell Labeling
2 Cell coupling with barcode beads
3 Dilution into gel solution
4 Gelification
5 Cell lysis
6 Gel dissolution
7 RT & PCR Amplifications
BIOLOGICAL SAMPLE
Lysis buffer
RT & PCR Mixes
LIBRARY PREP / SEQUENCING
Barcode beads:
• Bead specific DNA Barcode
• Poly-T sequence for mRNA capture via 3' poly-A
Cell-bead pairs are physically isolated
Freed mRNA from lysed cell captured by local bead
Bead-mRNA recovery
Amplified barcoded cDNA solution
b
FASTQ
FASTQ
2 FASTQ files / sample
(read1 & read2 files)
FastQC
QC
Quality Indicator reports / sample
Extract
Align
Dedup
Assign
Count
MTX /TSV
Reference Genome Sequences & Annotations
1 Count Matrix file / sample
(nb of transcripts per cell per gene)
FILTERING
NORMALIZATION
HVG
CLUSTERING
PCA
DIFFERENTIAL GENE EXPRESSION
2D EMBEDDING

Fig. 2


a
species
human
mouse
undefined
nb of mouse NIH3T3 cell transcripts
nb of human HEK293 cell transcripts
b
cell type
Monocytes
CD4+ T cells
CD8+ T cells
B cells
NK cells
T cells
Dendritic cells
unassigned
Progenitors
UMAP_1
UMAP_2
c
Monocytes
CD4+ T cells
CD8+ T cells
B cells
NK cells
T cells
Dendritic cells
Progenitors
unassigned
29.77%
23.06%
20.3%
22.74%
15.62%
17.41%
13.32%
13.22%
8.8%
9.94%
8.36%
8.49%
2.73%
4.2%
0.35%
0.32%
0.75%
0.63%
10x-PBMC-1
RevGel-seq
d
UMAP_1
UMAP_2
e
Alpha cells
Beta cells
Delta cells
Gamma cells
Endothelial cells
B cells
Macrophages
Ductal cells
T cells
Other Immune cells
Activated stellate cells
Quiscent stellate cells
24.93%
23.87%
19.96%
19.23%
3.85%
2.79%
1.66%
1.39%
1.13%
0.46%
0.4%
0.33%
beta_sample
3000
6000
9000

ONLINE METHODS

RevGel-seq: Instrument-free single-cell RNA sequencing using a reversible hydrogel for cell-specific barcoding

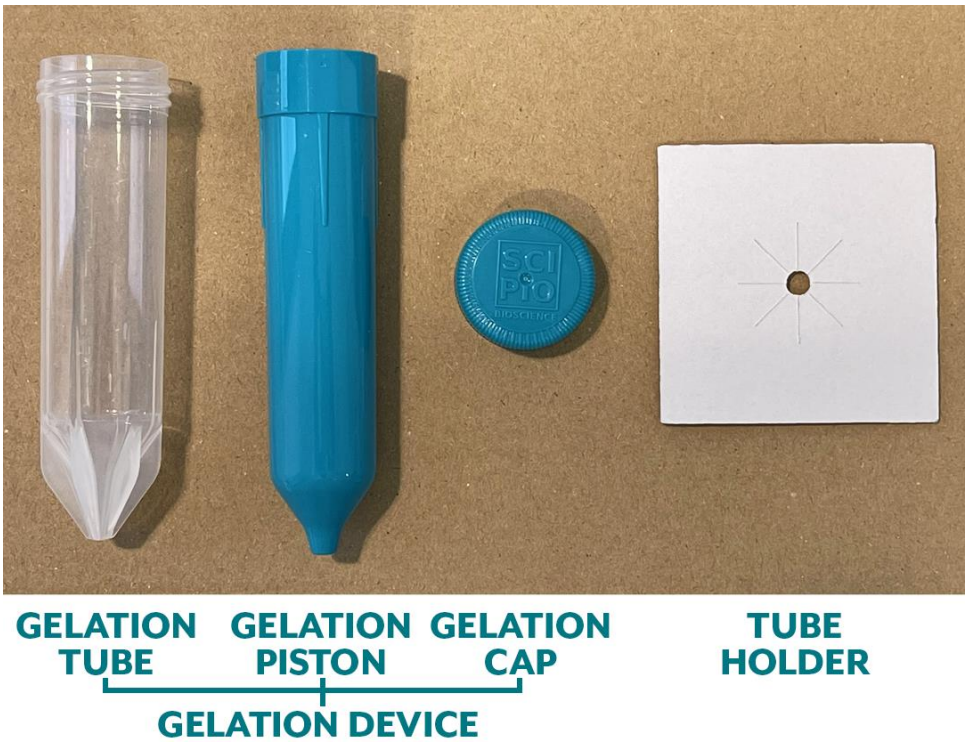
I. scRNA-seq sample preparation workflow.

Experiments were performed with the RevGel-seq protocol, capable of analyzing at least 10,000 cells per run with a specially designed gelation device (see Supplementary Figure 1). The individual steps from coupling cells to barcode beads to library preparation for sequencing shown in Fig. 1a are more fully described below:

1) Cell labeling. Cells are first labelled with a bifunctional chemical linker used to tether individual cells to barcode beads. After an incubation with an excess of chemical linker and washes, the cells are resuspended at a density of 100/µL. Typical experiments analyze 10,000 cells, but scale-up to 30,000 cells is realizable (see section II.B).

2) Cell coupling. For preparation of the barcode-bead tandems, 100 µL of the barcode beads are pipetted into the gelation tube (see Supplementary Figure 1), followed by transfer of 100 µL of the labelled cell suspension (for 10,000 cells) into the suspension of barcode beads. Quantification of cell-bead tandems can be achieved at this point by microscopy, if evaluation of coupling is required.

3) Dilution into gel solution. Cell-bead tandems are diluted into the hydrogel solution and spread along the walls of the gelation device by insertion of a piston to form a thin hydrogel solution layer (Supplementary Figure 1). An additional 2 mL of the hydrogel reagent are added at the bottom of the gelation tube. The gelation tube is then sealed with the gelation cap and homogenization is achieved by rotating the tube. After rotation, the gelation piston is slowly inserted into the gelation tube while keeping both vertical, as the contents rise smoothly along the wall of the gelation tube with a level horizontal meniscus.



Supplementary Figure 1 | Components of the gelation device and the accompanying tube holder.

4) Gelation. The assembled gelation device is then placed vertically in ice and incubated for 20 min for complete gelation to immobilize the cell-bead tandems.

5) Cell lysis. While maintaining verticality, the gelation piston is slowly removed from the gelation tube. The gel remains in the tube, partially collapsed, with no gel remaining on the piston. Then lysis is achieved by incubating the gel for 1 hour in a lysis buffer containing 0.2% sarkosyl, which diffuses freely in the gel, triggering release and hybridization of 3'-polyA$^+$ RNAs on the 3'-polyT extremities of the barcode beads. The procedure can be interrupted at this point by placing the gelation device vertically in a freezer at -80°C. In this case, when ready to continue the gelation device should be thawed at room temperature.

6) Gel dissolution. When lysis is completed, the degelation step is immediately initiated by adding 3.75 mL of the degelation buffer in the gelation tube and sealed with the gelation tube cap. The gelation tube is then held in the vertical position and vortexed for 5 min at maximum speed.

7) RT, PCR, and preparation for DNA sequencing. The RNA-loaded barcode beads are washed, transferred into 1.5mL microtubes, then washed two additional times to prepare for subsequent enzymatic steps.

- **i)** Reverse transcription: The pellet of beads is resuspended in 20 µL of bead wash buffer and supplemented by 75 µL of the RT supermix and 5 µL of reverse transcriptase solution. The sample is then placed in a heat block and incubated 10 min at 25°C followed by 90 min at 42°C. The sample tube is placed on ice for 2 min and then spun briefly on a benchtop centrifuge. (The procedure can be interrupted at this point and the sample stored overnight at 4°C or up to one week at -20°C, then thawed on ice before proceeding to the next steps.)
- **ii)** Exonuclease I treatment: After addition of 5 µL of Exonuclease I, samples are homogenized by pipetting and incubated 50 min at 37°C, after which the enzyme is inactivated by addition of 250 µL of TE-SDS.
- **iii)** Alkaline denaturation: RNAs are released from the beads by adding 100 µL 0.1M NaOH solution. The samples are then spun for 20 sec on a benchtop centrifuge, 100 µL of supernatant is removed, an additional 50 µL 0.1M NaOH solution is added, and samples are rotated for 5 min on a microtube rotator/wheel. Washing is then carried out with 250 µL of TE-TW and adjusting the residual volume to 20 µL using RevGel-seq bead wash buffer.
- **iv)** Second strand synthesis: The S3 supermix is first incubated for 5 min at 70°C, then immediately placed on ice for 1 min, spun, and homogenized. Then 77.5 µL of the S3 super mix is added to the beads, followed by addition of 2.5 µL of the S3 enzyme. The sample is thoroughly homogenized by pipetting and incubated 1 hour at 37°C. Twowashes 250 µL of TE-TW are carried out, followed by a final resuspension in 160 µL with nuclease-free water.
- **v)** Polymerase chain reaction: 240 µL of the PCR supermix is added to the 160 µL of beads and the beads are homogenized by pipetting and split into 8 PCR tubes, with 50 µL in each tube. The PCR program is described below:

□

Step	Cycles	Temperature	Duration

Step 1	Initial denaturation	-	95°C	3 min
Step 2	Denaturation	4	98°C	20 sec
	Annealing		65°C	45 sec
	Extension		72°C	3 min
Step 3	Denaturation	9	98°C	20 sec
	Annealing		67°C	20 sec
	Extension		72°C	3 min
Step 4	Final extension	-	72°C	5 min
Step 5	Store	-	4°C	∞

vi) Samples purification is carried out using SPRIselect following the user guide. PCR strip tubes are spun for 20 sec using a benchtop centrifuge and 45 µL of bead-free supernatant from each tube are pooled in a clean PCR tube. Sample size purification is performed using 27 µL of SPRIselect beads (0.6x of PCR reaction volume) and following the manufacturer's procedure. Elution is performed in 20-40 µL of nuclease-free water. The procedure can be interrupted at this point and the sample stored at -20°C, then thawed on ice before proceeding to the next steps.

vii) Library preparation: Following quantification and size estimation of amplified barcoded cDNA, 600 ng of the purified cDNA amplicons from each PCR tube is sampled and processed according to Illumina Nextera XT library prep kit using a specific library prep primer designed for the RevGel-seq technology.

II. PBMC analysis

A) Sample preparation. Whole blood was provided from *l'Etablissement Français du Sang (EFS)* under a convention for research use. PBMC was isolated from a fresh blood by HistoPaque-1077 (Sigma) according to the instruction. The purified PBMC from the same blood sample was used for both RevGel-seq prep and 10x Genomics Chromium single cell 3' reagent kit (v3.1), and both scRNA sample preparations with 10,000 input cells were started at the same time.

B) Scale-up. Extension of the protocol to greater numbers of cells was verified by performing experiments with 30,000 cells. In this case, the standard protocol presented above was employed, but the volume and reagent quantities were doubled, with an increase in cell density for coupling of 1.5X. Compared to controls with 10,000, very similar results were obtained, as illustrated below with the data for one of the experiments:

Parameter	**10,000 cells**	**30,000 cells**
Capture rate	28%	35%
Median genes/cell*	1130	1310
Median transcripts/cell*	1894	2231
% Mitochondrial transcripts (median)	3.5%	3.3%

*20,000 raw reads per cell

III. Data analysis with Cytonaut, the Scipio bioscience cloud-based bioinformatics platform.

The end-to-end data analysis from raw sequencing data to differential gene expression is performed by Cytonaut cloud software (https://www.cytonaut-scipio.bio), which integrates data pre-processing as well as data post-processing and interactive data visualization. Being cloud-based, Cytonaut can be used via a simple web interface, bypassing the need to install software packages or use computational resources. Workflow of the end-to-end data processing pipeline integrated in the Cytonaut platform is summarized in Fig. 1b, with details on individual steps presented below:

A) Pre-processing pipeline. The data pre-processing pipeline of Cytonaut takes as input the two FASTQ files of the sample (R1 and R2), leverages the specific barcode pattern of RevGel-seq capture beads, and successively performs read quality control, extraction, alignment, deduplication, and assignment. The reference genome sequences and annotations are based on Ensembl release v99 for human species (GRCh38) and mouse species (GRCm38). The pre-processing pipeline provides as output:

- **i)** The count matrix of the input sample, which contains the number of transcripts detected for each gene in each cell;

ii) A set of quality indicators, which include the median number of genes per cell, the median number of transcripts per cell, the median percentage of mitochondrial transcripts per cell, as well as their distribution statistics;

iii) For the NIH 3T3-HEK sample in Fig. 2a, the quality indicators also include statistics on cell purity (where cell purity is the probability that a transcript captured by the cell barcode has been expressed by the main cell coupled to the barcode). Transcripts on barcodes were assigned to each species, mouse or human, if having a species purity of 95% or more; for purity lower than 95% points are represented as gray dots in the barnyard plot. The total cell multiplet rate is extrapolated from the measured hetero-species cell multiplet rate, where a hetero-species cell multiplet is defined as a cell barcode for which cell purity is lower than 2/3 (i.e., for which cell impurity is higher than 1/3).

B) Post-processing pipeline. The data post-processing pipeline of Cytonaut takes as input the sample count matrix provided by the pre-processing pipeline and performs the following steps:

i) Filtering of cells and/or genes is achieved according to application-dependent parameter values which are selected by the user (e.g., minimum number of cells expressing a gene set to 3, minimum number of genes per cell set to 200, maximum percentage of mitochondrial transcripts per cell set to 20%);

ii) Highly Variable Gene (HVG) detection is performed according to application-dependent parameter values which are decided by the user (e.g., number of top genes set to 2000);

iii) Principal Component Analysis (PCA) is run on the top HVG genes, according to application-dependent parameter values which are fixed by the user (e.g., number of PCA dimensions set to 10);

iv) 2D Embedding of cells is carried out according to parameter values, which are decided by the user (e.g., number of neighbors set to 10, UMAP method with minimum distance set to 0.3, or t-SNE method with perplexity set to 30);

v) Cell Clustering is performed according to parameter values, which are set by the user (e.g. Louvain method with resolution set to 0.8, or Leiden method with resolution set to 0.8);

vi) Differential Gene Expression (DGE), according to a statistical test method established by the user (e.g., Wilcoxon-Mann-Whitney test).

C. Post-processing pipeline output. The following items are provided:

i) The DGE matrix, which contains for each cell cluster and each gene the z-score, the log fold change, and the p-values (non-adjusted and adjusted) of the gene in the cells belonging to the cluster compared to the cells belonging to the other clusters.

ii) The attributes of each cell, including the cell quality indicators and the cluster ID of the cell;

iii) The attributes of each gene, including the cell quality indicators and the gene variability indicators;

iv) The AnnData object (h5ad format), which contains all the data allowing to further reproduce the post-processing results.

D) Interactive data visualization. The Cytonaut Rover module provides an interactive visualization of data generated by the post-processing pipeline, allowing exploration of gene content of cell clusters by applying customized filters leading to customized distribution statistics (heatmap, dot plots, violin plots). In particular, the visualization of the expression of discriminant genes in each cell cluster enables annotation of the identified cell types.

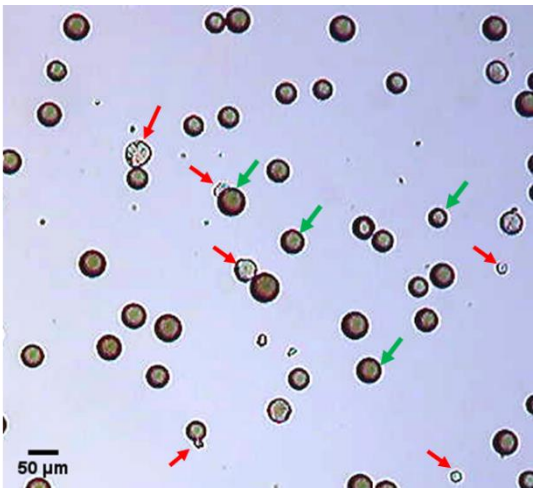
IV. Cardiomyocytes, including an early checkpoint

Cardiomyocyte cells (from Takara Bio, ref. Y10060) were harvested, trypsinized, and labelled (Supplementary Figure 2A), and 10,000 labelled cardiomyocytes were used for RevGel-seq (see Section I above). After cell-bead coupling, 20 µL of the suspension of cell-bead complexes were sampled to count the number of beads coupled to cells by bright-field microscopy (Supplementary Figure 2B), with the rest of the coupled suspension used for scRNA preparation, resulting in eight PCR tubes per sample. Two of the PCR tubes were used for library preparation using Illumina Nextera XT kit and DNA sequencing. Beads were counted to assess bead recovery rate. As shown in Supplementary Figure 2C, the number of cells obtained by sequencing analysis (530, 523) can be closely approximated by counting

cell-bead objects by microscopy (595, 527), hence validating the microscopy coupling counting as an efficient early checkpoint to estimate the number of analyzed cells.

a

b



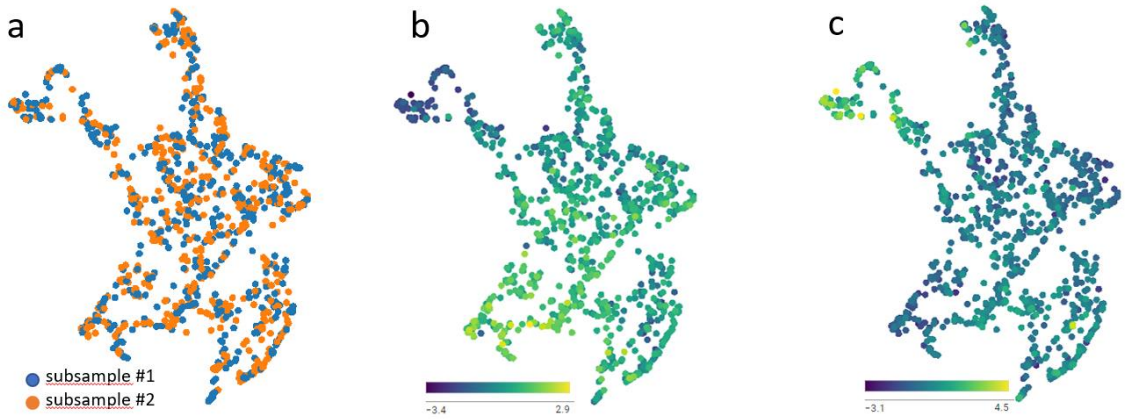
c

	Total cell-coupled beads	Cell-coupled beads/PCR tube	Bead-cell recovery rate (microscopy)	Estimated bead-cell number (microscopy)	Analyzed cell number (sequencing data)
Subsample #1	8,800	1,100	54 %	595	530
Subsample #2	8,800	1,100	48 %	527	523

Supplementary Figure S2 | Assessment of recovered quantity of cells by microscopy. **a,** Microscopic observations of cardiomyocytes following trypsinization (diameter of 20-35 µm). **b,** Microscopic observations of cardiomyocytes following coupling procedure (Red arrows indicate cells, green arrows indicate beads). **c,** Estimated cell quantities obtained from numbers of cell-coupled bead tandems (for all 10,000 input cells) and observed bead recovery rate (for each PCR tube, max 12.5% for each of 8 PCR tubes). Sequencing data was processed according to Section I above. Analyzed cell number is retrieved by post-sequencing analysis (see also Supplementary Figure 3).

The preprocessing results from the sequencing data of the two sub-samples were processed together for dimension reduction and projection on 2D. The Supplementary Figure 3 A-C

present the UMAP graphs for the distribution of cells of two sub-samples in A (two different tubes for cDNA amplification and library preparation) from the same initial cell and capture bead coupling. In the A graph cells from both subsamples are well mixed, indicating the absence of batch effects of split cDNA amplification and subsequent library preparation. The gene expression levels of RYR2 gene and CALD1 are presented in B and C, respectively. The levels of expression of the two genes are well contrasted, implying that most cells are cardiomyocytes (RYR2 high), with a minor fraction of smooth muscle cells (CALD1).



d

sample	median transcript numbers/ cell at 50k RRPC	median gene diversity / cell at 50k RRPC	median % Mitochodrial RNA
Subsample #1	13937	4399.6	7.48
Subsample #2	15892.4	4733.9	7.43

Supplementary Figure S3 | 2D projections of, A) two sub-samples (PCR tubes) from the same cardiomyocyte sample preparation, B) gene expression of RYR2 (cardiomyocyte enriched gene) and C) CALD1 gene expression (smooth muscle cell enriched gene). The Table in panel D shows the statistics of transcripts and gene diversity/analyzed cell at average 50,000 raw reads/cell, as well as mitochondrial RNA fraction.

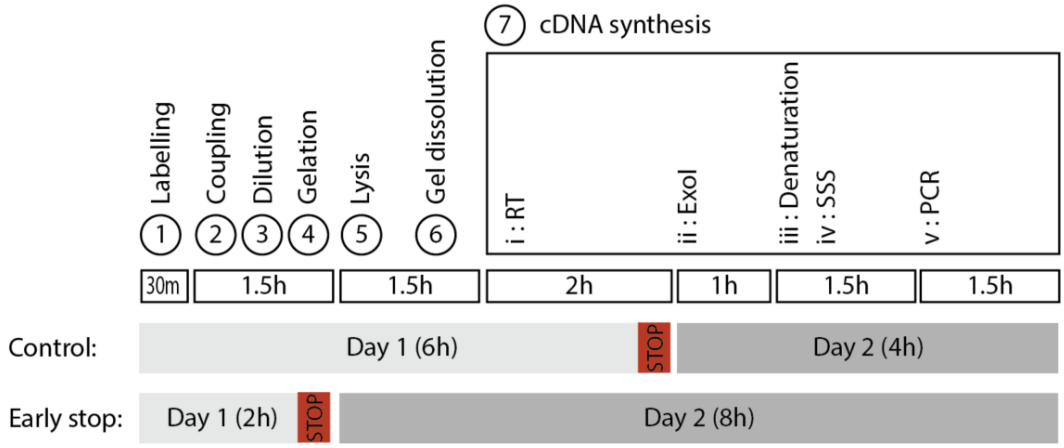
V. Pancreatic islet cells

Pancreas islets were isolated from mouse pancreas by collagenase (1mg/ml, Sigma-Aldrich) injection in the bile duct and handpicked under a binocular microscope (Leica). Islet cells were dissociated with 0,05% trypsin-EDTA (Gibco Thermo Fisher), incubated at 37°C for 5min and resuspended in PBS with 2% fetal calf serum and 0,5 mM EDTA.

VI. Early stopping point

To confirm an early stopping point in the RevGel-seq procedure, experiments were performed using 10,000 input cells of HEK293 and NIH3T3 in a 1:1 ratio, and performances were compared between samples frozen after reverse transcription ("Control") and samples frozen immediately after applying lysis buffer to the hydrogel ("Early stop -80°C" and "Early stop dry ice"), corresponding to ~2h after initial cell labeling; Supplementary Figure 4A). On a subsequent day, the RevGel-seq procedure was continued on all samples, and 3 PCR tubes (~1250 input cells each) were sequenced from each sample. Performance indicators, shown in Supplementary Figure 4A, were very similar for all conditions. Pooling of all 3 conditions for post-processing analysis (see Section III above) showed that cells were not clustering per condition (Supplementary Figure 4C) but rather per cell species (Supplementary Figure 4D), confirming that freezing at this early stopping point did not alter gene expression nor performances.

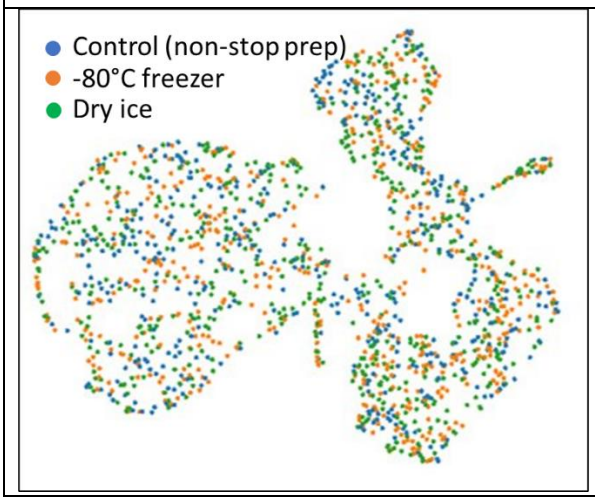
a



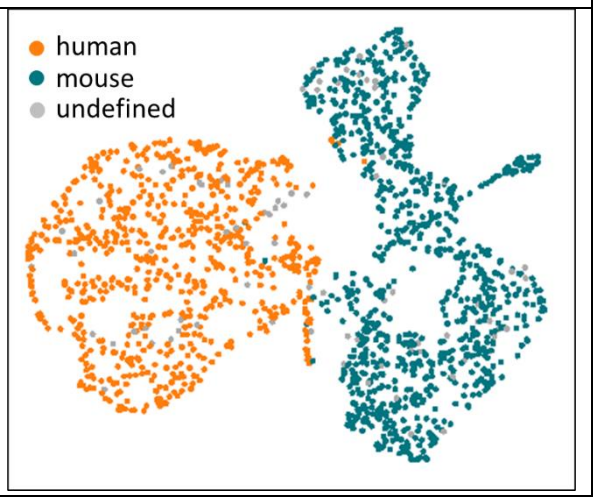
b

	Cell capture rate	Median genes/ human cell	Median transcripts /human cell	Median genes/ mouse cell	Median genes/ mouse cell	Average mito-chondrial transcripts
Control	40 %	5,529	16,875	4,626	12,585	4.7 %
Early stop_-80°C	44 %	5,519	18,439	4,616	12,800	5.4 %
Early stop_Dry ice	45 %	5,375	17,224	4,721	13,525	5.1 %

c



d



Supplementary Figure S4 | Freezing sample at an early stopping point during RevGel-seq protocol does not modify scRNA preparation performances. **a,** Schematic representation of the timing differences between the "Control" and "Early stop" samples (See Section I for specific steps). "Control" samples were obtained by freezing the RT reaction mixture at -20°C overnight, while "Early stop_-80°C" and "Early stop dry ice" samples were obtained by freezing (either at -80°C or in dry ice, respectively) immediately after applying lysis buffer to the hydrogel. RevGel-seq protocol was resumed on a subsequent day for all samples. **b,** Comparison of performance indicators between the tested conditions (average of 3 PCR tubes for each condition). Gene and transcripts/cell at 50k reads/cell with raw read down-sampling. **c, d,** Projection of scRNA seq data on 2D (UMAP) of the pooled samples (Pool of data from 1 PCR tube for all 3 conditions), shown per condition (**c**) and per determined cell species (**d**). ("undefined" corresponds to cells with a transcripts purity < 95%; See Section III A-iii of data processing above).