Construction of multiplexed transcriptome NGS libraries of microdissected tissue samples based on combinational DNA barcode microbeads

Kaitong Dang, [1,†] Yue Zhao,[1,†] Kaiqiang Ye,[1] Yunxia Guo,[1] Wenjia Wang,[1] Qinyu Ge, Xiangwei Zhao

[1]*State Key Laboratory of Bioelectronics, School of Biological Science & Medical Engineering, Southeast University, Nanjing, 210096, China*

[*]Corresponding author: Qinyu Ge, geqinyu@seu.edu.cn(Q.G.); Xiangwei Zhao, xwzhao@seu.edu.cn(X.Z.)

Keywords:
Transcriptome, Combinational DNA barcode, Multiplexing, Fresh frozen tissue, Formalin-fixed paraffin-embedded sample

Abstract

The combination of single-cell RNA sequencing and microdissection techniques that preserves positional information has become a major tool for spatial transcriptome analyses. However, high costs and time requirements, especially for experiments at the single cell scale, make it challenging for this approach to meet the demand for increased throughput. Therefore, we proposed combinational DNA barcode (CDB)-seq as a medium-throughput, multiplexed approach combining Smart-3SEQ and CDB magnetic microbeads for transcriptome analyses of microdissected tissue samples. We conducted a comprehensive comparison of conditions for CDB microbead preparation and related factors and then applied CDB-seq to RNA extracts, fresh frozen (FF) and formalin-fixed paraffin-embedded (FFPE) mouse brain tissue samples. CDB-seq transcriptomic profiles of tens of microdissected samples could be obtained in a simple, cost-effective way, providing a promising method for future spatial transcriptomics.

1 | Introduction

In the last decade, the emergence and rapid development of single-cell RNA sequencing (scRNA-seq) has facilitated studies of cellular heterogeneity and has offered unprecedented insights into the mechanisms underlying diseases in individual organisms.[1-3] Single-cell transcriptome profiles are obtained by combining single-cell isolation techniques, such as microscopic manipulation,[4] fluorescence-activated cell sorting (FACS),[5-6] or microfluidic approaches,[7-8] with compatible library construction protocols. However, these dissociation-based techniques lack spatial context,[9-10] which is essential for cellular bioanalysis.[11] To overcome this limitation, new spatially resolved RNA sequencing techniques have emerged.

Spatially resolved transcriptomics (SRT) provides quantitative expression profiles, while maintaining the original spatial information, allowing a transcriptome atlas to be correlated with cell morphology, physiology, and histology.[12-14] The throughput of single-cell resolution sampling can be divided into three main categories: high (tens of thousands), medium (hundreds to thousands), and low (several to hundreds). Various approaches have been developed to meet the demand for high throughput, such as ST (Spatial transcriptomics),[15] HDST (High-definition spatial transcriptomics),[16] DBiT-seq (Deterministic barcoding in tissue for spatial omics sequencing),[17] MERFISH (multiplexed error-robust fluorescence in situ hybridization),[18] seq-FISH (Next generation sequencing (NGS)-based FISH),[19] Stereo-seq (spatial enhanced resolution omics-sequencing),[20] and Pixel-seq (polony-indexed library-sequencing).[21] Most of these methods use fluorescence in situ hybridization, fluorescence in situ sequencing, in situ capture, or other “in situ” approaches; however, cumbersome experimental protocol, expensive equipment, customized chips, and professional training requirements limit the application of these methods. Nevertheless, for numerous pathology samples available in medical institutions, transcriptome profiling is often required only for certain regions of interest. There are few cost-effective, accessible, sensitive, and reproducible approaches for low to medium throughput spatial transcriptomic studies of pathology samples, especially formalin-fixed paraffin-embedded (FFPE) samples.

Here, we propose combinational DNA barcode (CDB)-seq, a medium-throughput, highly multiplexed transcriptome sequencing library construction approach that couples CDB magnetic microbeads with the optimized Smart-3SEQ RNA sequencing protocol (Figure 1). Compared with the DNA barcode generation concept in Drop-seq with “split and pool” synthesis,[22] we can generate a large number of distinct barcoded beads with known sequences, facilitating downstream data analysis and diminishing bias among samples. Furthermore, CDB beads are easy to collect and wash; more importantly, the supernatant after mRNA capture by beads can be recollected for proteomics or genomics analyses, conferring the potential for integrated multi-omics studies. In addition, Smart-seq2 as the gold standard in RNA library construction methods requires Tn5 transposase to fragment cDNA, making it challenging to achieve multiplexing, as barcode information cannot be retained.[23] Smart-3SEQ is an alternative SMART-based method for 3′-end sequencing library construction from thermally fragmented RNA capable of the accurate quantification of transcript abundances in small samples extracted from FFPE and fresh frozen (FF) specimens.[24] In this study, Smart-3SEQ was optimized and adapted to CDB microbeads to ensure the high sensitivity and accuracy of RNA-seq, while including barcodes. First, we comprehensively studied the conditions for the synthesis of CDB microbeads and applied CDB-seq to total RNA extracts from mouse brains. Then, we collected single-cell resolution samples of FFPE and FF brain tissue sections using laser aided isolation (LAI) [25] to demonstrate the potential of CDB-seq for transcriptome analyses.

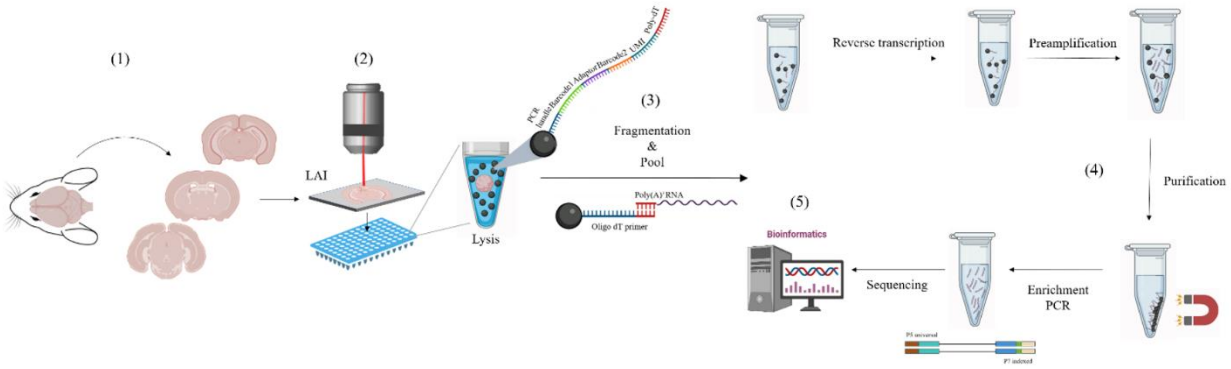

Figure 1. Overview of the CDB-seq protocol. (1) Mouse brain tissue sections are prepared. (2) Target micro-regions on sections are sampled by LAI and collected into 96-well plates loaded with distinct CDB microbeads and lysis buffer. (3) After cell lysis and mRNA capture, CDB microbeads of wells are pooled in a new PCR tube. (4) cDNA sequencing libraries are constructed according to the CDB-seq protocol. (5) The corresponding bioinformatics analysis is performed based on the sequencing data.

2 | MATERIALS AND METHODS

2.1 | Materials

TRIzol™ LS, single cell lysis kit, EDC, $MgCl_2$ (25 mM), Maxima H Minus Reverse Transcriptase (200 U/μL) were purchased from Thermo Fisher Scientific (China). Chloroform, Isopropyl alcohol, 10M NaOH solution, 30% Brij® 35, Betaine (5 M) were purchased from Sigma Aldrich (Shanghai, China). Ethanol, 0.1 M MES (pH 5.0), 1 M Tris-HCl (pH 8.0), 0.5 M EDTA (pH 8.0), 0.05% TWEEN-20, KCl, NaCl, dNTP Mixture (10 mM), Tween-20, 10% (vol/vol), 1× PBS and all nucleic acid primer synthesis were provided by SanGon Biotech (Shanghai, China). RNase Inhibitor (40 U/μL), SMARTScribe Reverse Transcriptase were purchased from Takara (Beijing, China). Carboxylated magnetic beads were provided by Suzhou Knowledge & Benefit Sphere Tech. KAPA HiFi HotStart Read Mix was purchased from Roche (Shanghai, China). Bst 2.0 DNA Polymerase was purchased from NEB. ABclonal One-step DNA Lib Prep Kit for Illumina V2 was purchased from ABclonal (Wuhan, China), VAHTS DNA Clean Beads was provided by Vazyme (Nanjing, China).

2.2 | Animals models and brain dissection

This study was reviewed and approved by the Ethics Committee of Zhongda Hospital Southeast University (20200104005). Mice (male C57Bl/6J, 8 weeks old) were purchased from Shanghai Southern Model Biotechnology Co., Ltd. and were placed under standard conditions with a temperature of 25 °C, relative humidity of 50–60%, noise level of 60 dB or less, and illumination of 10–14 h. Mice were fed ad libitum for 4 days to acclimate to the environment. Mice were executed using the neck-breaking method and the brains were removed and rinsed three times with PBS (1×, pH 7.4) to remove blood. The brains were then snap frozen in 2-methylbutane on dry ice and stored at -80 °C until further processing.

2.3 | RNA extraction

Samples of brain tissue (30–100 μg) were added to 2 mL centrifuge tubes, supplemented with 1 mL of pre-cooled TRIzol at 4 °C, and homogenized in an ice bath using a homogenizer until the liquid became clear and free of obvious particles. Then, the solution was mixed by inverting it 10 times and maintained at room temperature for 5 min. Then, 200 μL of chloroform was added and the mixture was shaken vigorously for 15 s and kept it at room temperature for 5 min. The mixture was centrifuged at 4 °C and 12,000 rpm for 15 min, and 300–400 μL of the supernatant was collected. An equal volume of isopropanol was added and the solution was mixed by inverting it 10 times, then kept at room temperature for 10 min. The sample was then centrifuged at 4 °C and 12,000 rpm for 10 min. The supernatant

was slowly removed, supplemented with 75% ethanol that was pre-cooled at 4 °C, mixed by inversion 10 times, and centrifuged at 4 °C and 12,000 rpm for 5 min; the procedure was then repeated. The supernatant was slowly removed in an ultra-clean bench and the tube was dried using an ultra-clean bench air blower until the white substance in the tube became semi-permeable. Then, 100 μL of ddH_2O was added to dissolve the white substance, followed by mixing by pipetting.

2.4 | Tissue sectioning

The FF or FFPE sections used in this experiment were prepared by Wuhan Saville Biotechnology Co. Unlike conventional tissue sections, which are directly covered and fixed on slides, the commissioned sections were attached to slides/type I ITO slides coated with a PDMS layer to meet the sampling requirements of LAI, and the finished fixed tissue sections were stored in a -80 °C refrigerator to reduce RNA degradation.

2.5 | LAI sampling

Stained sections were placed into the slide holder of a micro-scope and moved to the edge for parameter adjustment. The lens was covered with paper, the laser was turned on, and the current was adjusted to about 1.5–2.0 A. The paper was removed for about 1 s, and the current was then adjusted to 0 A. The shape of the struck area on the tissue section was observed and the size was measured. The diameter of the sampling area (20–40 μm) was adjusted by the current according to the experimental requirements.

The cap was opened and flattened and 4 μL of lysis solution was added (3.6 μL of Single Cell Lysis Solution and 0.4 μL Single Cell DNase I). The section was moved to the target sampling area, the cap was placed vertically under it, the lens was covered with paper, the current was adjusted to the appropriate level, and the paper was removed for 1 s. The cap was placed upside down, followed by reaction at room temperature for about 10 min without mixing. Then, 0.4 μL of Single Cell Stop Solution was added, maintained at room temperature for 2 min, and placed on ice. The reaction was continued for no more than 30 min or stored at -20 °C. LAI methods were adapted from Kim et al. (2018).

2.6 | Combinational Synthesis Approach to Generate Uniquely Barcoded Beads

The barcoded oligonucleotides on the surface of the magnetic beads were synthesized by two rounds of combinations. All sequences used are listed in Table S1 and Table S2.

Briefly, 62.5 μL of carboxylated magnetic beads (50 mg/mL) were added to eight 1.5 mL centrifuge tubes named "1-N," washed twice with 0.1 M MES (pH 5.0), and resuspended in 500 μL of MES (0.1 M). Then, 62.5 μL of 50 mg/mL EDC and 50 μL of 100 μM primer 1-N were added to each centrifuge tube. After incubation in a molecular hybridization oven for 5 h at 28 °C with rotation at 1,500 rpm, maples were washed 3 times with 1 mL of TET buffer (2 μM Tris-HCl, 5 μM EDTA, 0.1% TWEEN-20, 1 mM KCl), resuspended in 312.5 μL of TET buffer, and stored at 4 °C.

In the second round of combination, eight "1-N" tubes were centrifuged to completely remove the supernatant, 32.5 μL of 10× isothermal amplification buffer, 235 μL of enzyme-free water, and 32.5 μL of 10 mM dNTP were added and mixed well by pipetting, followed by dispensing sequentially into wells 1–12 of rows A–H of a 96-well plate (25 μL per well). Primer 2-M (100 μM) was placed in thermal cycler at 72 ℃ for 3 min and then at 4 ℃ for 20 s to remove primer dimer. Then, 6 μL of the corresponding primer 2-M was added to the wells of the 96-well plate by a platoon gun and the 96-well PCR plate was transferred to a thermostatic shaker at 85 °C for 2 min. A volume of 14 μL of Isothermal Amplification Mix (Nuclease-free water 1.2 mL; 10× isothermal amplification buffer 140 μL; Bst 2.0 DNA polymerase (8,000 U/mL） 60 μL) was added to each well of the 96-well PCR plate, which was returned to the shaker at 65 °C for 120 min. A longer incubation time had no deleterious effect.

50μL freshly prepared denatured solution was added to each well of the 96 well PCR plate (152 mM NaOH, 0.5% Brij ® 35) Shake the metal bath at a constant temperature of 85 ℃ for 2 minutes, centrifuge and remove the

supernatant while it is hot, repeat three times, and finally completely remove the supernatant. 50μL neutralization solution was added to each well, vortex oscillation, centrifugation to remove the supernatant, repeat three times. CDB microbeads have been completed.

2.7 | Library Preparation Based on CDB Microbeads

All sequences are listed in Table S3. For CDB-Smart-seq2, 1.5 μL of CDB microbeads and 2.25 μL of Reverse Transcription Mix1 (RNA (10 pg/μL) 1 μL; dNTP (10 mM) 1 μL; RNase Inhibitor (40 U/μL) 25 μL) were added to a 0.2 mL centrifuge tube at 72 °C for 3 min to remove the secondary structure of RNA and oligo-dT. Then, 8.25 μL of Reverse Transcription Mix2 (5× RT Buffer 2.5 μL; $MgCl_2$ (25 mM) 1 μL; TSO (10 μM) 0.25 μL; Maxima H-minus reverse transcriptase (200 U/μL) 1 μL; DTT (100 μM) 0.5 μL; Betaine (5 M) 2 μL; RNase Inhibitor (40 U/μL) 0.25 μL) was added. The PCR program was set to 42 °C, 90 min; 10 cycles (50 °C, 2 min; 42 °C, 2 min); 85 °C, 5 min for reverse transcription. A volume of 25 μL of Pre-Amplified Mix1 (KAPA HiFi HotStart Ready Mix (2×)12.5 μL; IS PCR-oligo (10 μM) 1 μL; Nuclease-free water 1 μL) was added and the PCR program was set to 98 °C, 3 min; 30 cycles (98 °C, 20 s; 67 °C, 15 s; 72 °C, 6 min); 72 °C, 5 min. Then, 25 μL of ddH_2O and 40 μL (0.8 ×) of DNA Clean Beads were added, and the purification process was completed according to the protocol. The cDNA concentration was detected using the Qubit 4.0 fluorimeter. The sequencing library was constructed using the ABclonal One-step DNA Lib Prep Kit for Illumina V2.

For CDB-seq, 4.5 μL of Fragmentation Mix (SMARTScribe first-strand reaction buffer (5×) 2 μL; dNTP (10 mM) 1 μL; RNA (10 pg/μL 1 μL; $MgCl_2$ (80 mM) 1 μL) and 1.5 μL of 10 mg/mL CDB microbeads were added to a 96-well plate and mixed thoroughly. RNA fragmentation was completed at 80 °C for 5 min. Then, 5 μL of TS-RT Mix (DTT (20 μM) 1 μL; Betaine (5 M) 2 μL; RNase Inhibitor (40 U/μL) 0.5 μL; 2S primer (20 μM) 0.5 μL; SMARTScribe reverse transcriptase (100 U/μL) 1 μL) was added to each well of the 96-well plate. The PCR program was set to 42 °C, 30 min; 70 °C, 10 min; 4 °C, ∞. The microbeads were collected from 96 wells into one PCR tube, the supernatant was removed, and Pre-Amplified Mix2 was added (KAPA HiFi HotStart Ready Mix (2×) 12.5 μL; 1S-Primer (PCR) (20 μM) 1.25 μL; 2S-Primer (PCR) (20 μM) 1.25 μL). The PCR program was set to 98 °C, 3 min; 20 cycles (98 °C, 15 s; 60 °C, 30 s; 72 °C, 2 min); 72 °C, 5 min. The product was purified by VAHTS DNA Clean Beads (0.8×). The Amplification Library Mix (KAPA HiFi HotStart Ready Mix (2×) 12.5 μL; P7 indexed (20 μM) 1.25 μL; P5 universal (20 μM) 1.25 μL) was added. The PCR program was set to 98 °C, 3 min; N cycles (98 °C, 15 s; 60 °C, 30 s; 72 °C, 30 s), 30 s; 72 °C, 2 min); 72 °C, 2 min (N is the number of cycles, $N = 19 - \frac{\log(Content\ of\ cDNA\ after\ pre-amplification(ng))}{\log(1.9)}$).

3 | Results

3.1 | A Combinational Approach to Generate CDB Microbeads

To prepare 96 barcoded beads with known sequences, we performed two rounds of "combination" reactions using only 20 oligonucleotide primers (Primer 1-N, Primer 2-M) (Figure 2A). The elaborate primer design enables mRNA capture, the preservation of spatial information, and PCR amplification for downstream library construction (Figure 2B). To characterize CDB microbeads, fluorescent primer 1 (reverse complementary to Adapter) and primer 2 (Poly-dA) modified with Cy3 (orange-red fluorescence) or FAM (green fluorescence) at the 5' end were used to verify whether primers 1-N were covalently immobilized on the microbeads by EDC (1-ethyl-3 (-3-dimethylaminopropyl) carbomiimide hydrochloride) catalysis and whether they were extended to generate capture primers by reverse complementary binding of primers 2-M. The fluorescence intensity of primer 2 was slightly lower than that of primer 1 (Figure 2C) because only a certain percentage of primers 1-N on microbeads could be extended to produce complete DNA barcode capture primers, and this percentage depended on which protocol was used for primer extension. The best extension efficiency obtained in this experiment was 76.92% (Figure 2D).

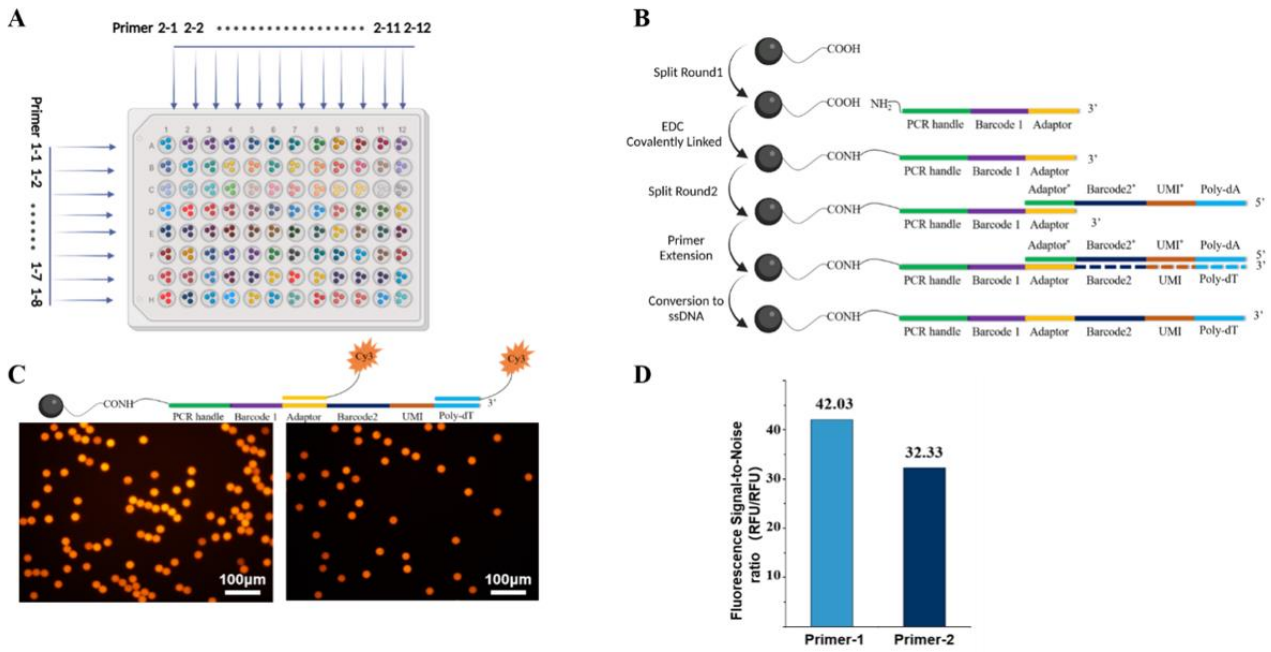

Figure 2. CDB microbead synthesis approach.

(A) Schematic diagram of the CDB microbead synthesis approach.

(B) Procedure and primer design, Primer 1-N sequences were designed (5'→3') as 1) constant sequences (PCR handle) for downstream PCR amplification; 2) DNA Barcodes 1-N (8 in total); 3) ligation sequences (Adaptor) for reverse complementation with the 3' end of primer 2-M. Primer 2-M sequence design (5'→3') was as follows: 1) poly-dA, used to capture mRNA with poly A tails; 2) UMI*, to identify DNA duplicates and eliminate PCR amplification bias; 3) DNA Barcode2-M* (12 in total), combines with Barcode1-N to increase the number of Barcode types; 4) Adaptor*, used for reverse complementation with primer 1-N 3' ends.

(C) Principle of fluorescent primer validation.

(D) Fluorescence signal-to-noise ratio of CDB microbeads.

3.2 | Optimization of Conditions for CDB Microbead Preparation

In the condensation reaction of EDC-catalyzed carboxylated magnetic microbeads with primer 1-N, the one-step coupling cross-linking method using EDC was compared with the two-step coupling method using EDC-bound NHS (*N*-hydroxysuccinimide) for the immobilization efficiency of primer 1-N (Figure S1). The duration of carboxyl activation (Figure S2A), the amount of activation reagent (Figure S2B), and the concentration of primer 1-N were also investigated with respect to the carboxyl activation effect (Figure 3A). The two-step coupling method had no effect on primer 1-N bonding, and a longer EDC activation time did not significantly increase the number of primer 1-N molecules immobilized on microbeads. However, a longer activation duration could reduce the difference in the number of primer molecules bonded to microbeads, and a higher concentration of the EDC solution produced a better carboxyl activation effect at the same time. In addition, the maximum primer carrying capacity of different particle sizes (200 nm, 1 μm, and 20 μm) and the optimal concentration of primer 1-N addition were investigated (Figure 3D). Based on the ssDNA concentration in the supernatant before and after the reaction in each experimental group, molar amounts of 0.25 mg of 200 nm, 1 μm, and 20 μm microbeads immobilized by covalent ligation of primer 1-N were approximately 186.7 pM, 182.7 pM, and 128.3 pM.

After the covalent immobilization of primers 1-N on magnetic beads, it is critical to efficiently complete primer 2-M extension and subsequent primer 2-M removal to ensure a high downstream mRNA capture efficiency and success rate. The effects of the adaptable PCR amplification method in Microwell-seq,[26] isothermal amplification, and conventional PCR on the extension efficiency of primers 1-N were compared (Figures 3B and 3E). In addition, the 95 ℃ thermal removal method, the denaturing solution alkaline wash method, and the 88 ℃ thermal alkaline

wash method for the removal of complementary primers were also compared (Figures 3C and 3F). Only a certain percentage of primer 1-N showed amplification with complete DNA barcodes, and efficiency was highest for the isothermal amplification extension method, with the best extension rate of approximately 76.92%. The three primer removal approaches were relatively similar, and the denaturing solution was chosen owing to its ease of operation.

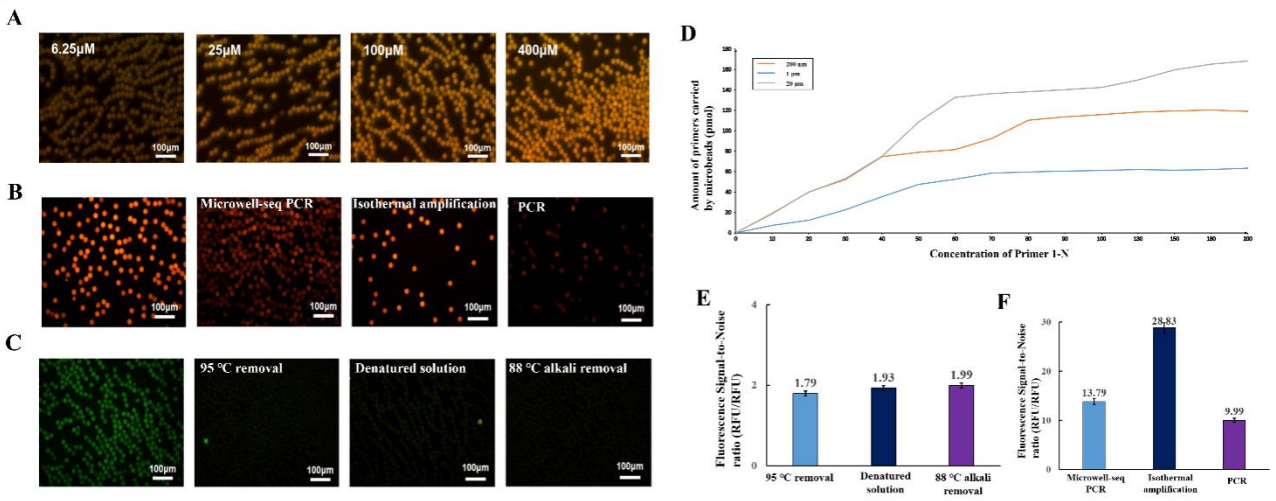

Figure 3. Effects of various conditions on CDB microbead preparation.

(A) Effect of the primer 1-N concentration on the number of primer molecules immobilized on the surface of microbeads.

(B, E) Comparison of the effect of different schemes for primer 1-N extension.

(C, F) Comparison of the effects of different removal schemes for primer 2-M.

(D) Maximum primer carrying capacity of 200 nm, 1 μm, and 20 μm microbeads.

3.3 | CDB-seq: Smart-3SEQ Combined with CDB Microbeads

Smart-3SEQ is an RNA-seq method for gene expression profiling of samples with a low RNA content and little degradation from FFPE or FF tissue sections. It combines the template switching method from Smart-seq2 with the 3′ end targeting approach of 3SEQ to further simplify operational steps and reduce experimental costs.[24] To customize Smart-3SEQ, the capture primer was coupled with an 8 nt barcode and modified with an amino group at the 3′ end for immobilization on carboxylated magnetic beads. Preliminary experiments revealed that the immobilized capture primer decreased the capture rate of mRNA to some degree. To overcome this issue, a pre-amplification step was added at the end of reverse transcription, thereby improving library quality (Figure S3A). Furthermore, CDB microbeads of different sizes showed variation in their primer carrying capacity and sedimentation rates. The size and concentration of cDNA produced by CDB microbeads of 200 nm, 1 μm, and 20 μm were compared using an Agilent 4150 electrophoresis system. The cDNA size ranged from 200 bp to 600 bp, with no significant difference with respect to microbead size. cDNA concentrations of 441 bp (200 nm), 467 bp (1 μm), and 433 bp (20 μm) were used as references for the comparison of mRNA capture and PCR pre-amplification effects (Figure S3B). The calibration concentrations of cDNA produced by the three different sizes of CDB microbeads were 7.18 ng/μL, 1.60 ng/μL, and 4.09 ng/μL, respectively. Based on these findings, 200 nm CDB microbeads met the experimental requirements.

3.4 | Construction of a Sequencing Library Based on CDB Microbeads

To verify the feasibility of CDB-seq, Smart-seq2 was chosen as the gold standard and further adapted for compatibility with CDB microbead-CDB-Smart-seq2. We used mouse brain RNA extracts with 10 pg/μL and RIN values greater than 7.5 to simulate the RNA content of single cells and compared the effectiveness of Smart-seq2, CDB-Smart-seq2, and CDB-seq for transcriptome characterization. According to the sequencing data obtained by CDB-Samrt-seq2 and Smart-seq2, there were almost no differences in AT and CG contents and base sequencing

quality values were both greater than 30, indicating that the probability of incorrect base identification was less than 0.1%. The rate of clean reads for CDB-Smart-seq2 from raw data was 54.6%, which was lower than that of Smart-seq2 at 71.2% (Figures S4A, S4B, and S4D). Both CDB-Smart-seq2 and Smart-seq2 could effectively detect low abundance mRNAs (FPKM < 5); however, the detection rate of CDB-Smart-seq2 for ultra-low abundance mRNAs (FPKM < 1) was slightly lower than that of Smart-seq2 (Figure 4A). The methods achieved full-length transcriptome sequencing, providing good gene coverage (Figure S4C). An average of 41.68% and 43.64% of total reads from both approaches were mapped to the mouse genome, and 66.13% and 70.52% genes overlapped with the gene set from the CellMarker 2.0 website, respectively, without significant differences (Figures 4B and 4D). The use of CDB microbeads instead of free Oligo-dTs did not significantly affect the mRNA capture efficiency and the quality of the resulting cDNA sequencing library.

The average rate of clean reads of CDB-seq sequencing data was 53.26% (Figure S4D), and about 12724 genes were detected, including 1922 genes in CellMarker 2.0. Low abundance mRNA (FPKM < 5) could be effectively detected, similar to the gene detection ability of Smart-seq2 (Figures 4A and 4B). With respect to the gene distribution, among reads mapped to the mouse genome by CDB-seq, an average of 27.15% were in exonic regions, 67.08% were in intronic regions, and 5.77% were in intergenic regions, with no significant differences among batches (Figure 4C). Most genes could be detected by CDB-seq, and only 707, 478, 1167, and 939 genes were detected individually, indicating high correlation between batches (Figure 4E). Moreover, gene detection by the three library preparation methods showed good correlations with each other, demonstrating that CDB-seq can consistently characterize gene expression differences in trace samples with batch-to-batch stability and very high reproducibility (Figure 4F).

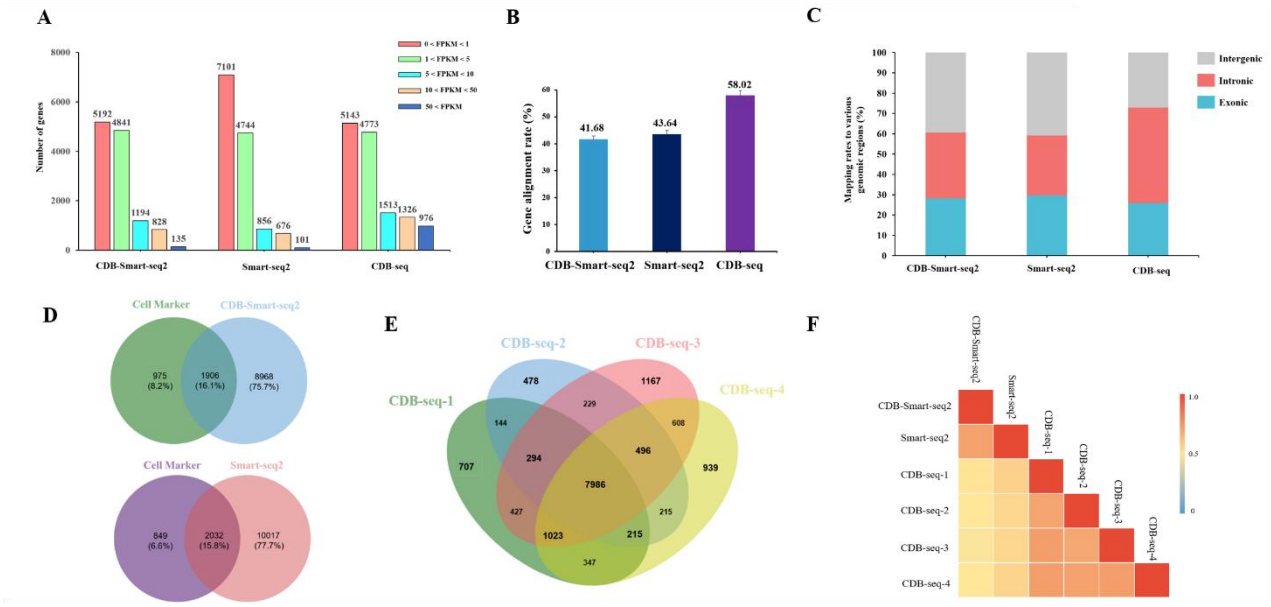

Figure 4. Gene detection ability by different library construction methods.

(A) Gene counts binned by standardized expression FPKM obtained by different methods.

(B) Gene alignment rate for different methods.

(C) Mapping rates to various genomic regions for different methods.

(D) Ratio of genes detected in the CellMarker mouse gene database for different methods.

(E) Repeat rate among CDB-seq batches.

(F) Correlations in gene detection by CDB-seq, Smart-seq2, and CDB-Smart-seq2.

3.5 | CDB-seq transcriptome analysis of mouse brain tissue FF and FFPE sections

FF and FFPE brain sections were attached to slides/type I ITO slides with PDMS layers, fixed by 4% PFA, and then stained with H&E. Based on the staining results, LAI microdissection was performed on the target area, with a single

sampling point of approximately 10 μm in diameter (Figure 5A).

According to the cDNA distribution of CDB-seq-FF and CDB-seq-FFPE by the Agilent 4150 electrophoresis system, the main peak concentrations of fragments within 100–500 bp were 5.00 ng/μL and 9.11 ng/μL (Figures S5A and S5B), respectively. These results indicate that thermal fragmentation of the mRNA by Mg^{2+} and capture by CDB microbeads were successful, followed by reverse transcription and pre-amplification. The libraries size was in the range of 140–400 bp, and concentrations were 7.29 ng/μL and 6.37 ng/μL, meeting the sequencing requirements (Figures S5A and S5B).

The rates of clean reads for CDB-seq-FF and CDB-seq-FFPE were 50.56% and 34.46% (Figure S5C), and the average numbers of genes detected across the 12 samples were 8828 and 3846 (Figure 5B). The sequencing data in the FF group showed greater correlations and uniformity than those of the FFPE group, possibly due to the nucleic acid degradation as a result of FFPE sample treatment and prolonged storage (Figure 5D). An average of 53.96% and 27.04% of detected genes from both groups overlapped with the gene set of CellMarker 2.0 (Figure 5C), which was correlated with the number of genes detected in each sample. Furthermore, for both kinds of samples, CDB-seq detected 5000 and 4800 genes at 0 < FPKM < 5, accounting for 43.54% and 41.42% of the total number of genes (Figures 5E and 5F), respectively, indicating that CDB-seq has excellent detection ability for low-abundance mRNAs.

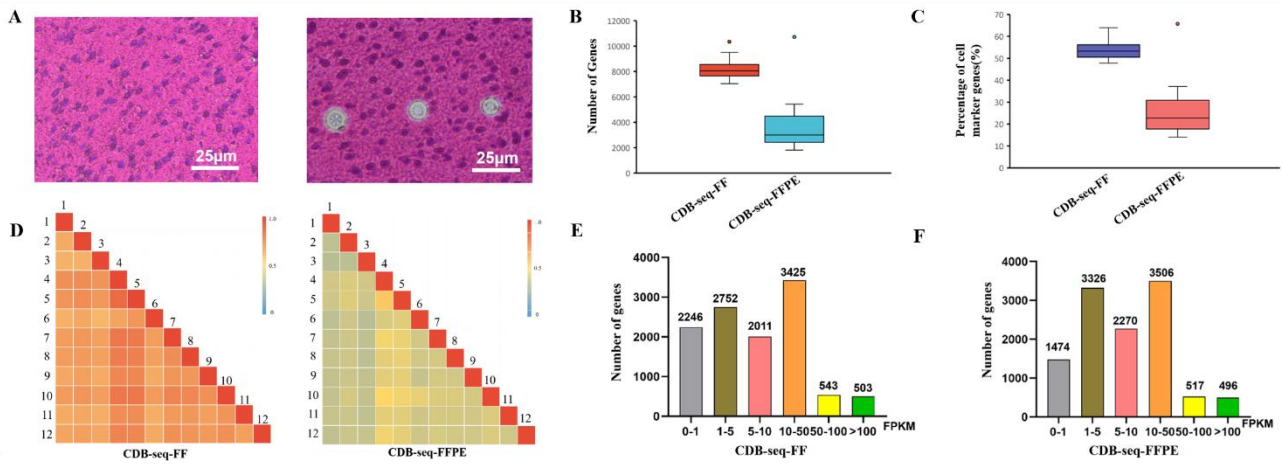

Figure 5. CDB-seq gene detection ability for FF and FFPE samples.

(A) Stained sections before and after LAI sampling. Comparison between CDB-seq-FF and CDB-seq-FFPE.

(B) Gene counts.

(C) Cell marker gene detection ability.

(D) Inter-batch correlations of CDB-seq-FF&FFPE results.

(E, F) Detection ability of CDB-seq for gene abundance after standardization by FPKM.

4 | Discussion

With continuous developments in spatial transcriptome research, many techniques have emerged to enable scRNA-seq while preserving spatial information. Most NGS-based spatially resolved transcriptome technologies apply cDNA library construction methods directly from compatible scRNA-seq protocols or modified versions.[27] However, these methods are often low-throughput RNA sequencing protocols that are not compatible with barcoding and can only be marked specifically by different indexes in PCR library amplification. Consequently, samples cannot be collected for multiplexed library construction and sequencing, resulting in an increase in experimental manipulation and reagent consumption as the number of samples increases.

High-quality RNA can be easily extracted from FF tissue samples for gene expression analyses.[15, 28-30] However, these samples are difficult to preserve for long periods and are not suitable for detailed morphological observations. In contrast, cell morphology and tissue structure are usually well preserved in FFPE samples, providing a good approach for pathological diagnostic specimens.[31] Moreover, FFPE samples can be stored at room temperature for

extended durations, which is beneficial for disease diagnosis, especially since spatial transcriptomic approaches are becoming compatible with FFPE tissues, enabling retrospective analyses of biobank samples spanning decades.[32] However, the cross-linking of proteins and nucleic acids during formalin fixation can lead to RNA breakage,[33] presenting a key challenge in transcriptome analyses using standard protocols. Smart-3SEQ, as an scRNA-seq protocol for FFPE samples, is cost-effective in terms of reagent use and experimental manipulation and is robust and sensitive to the amount of single-cell volume RNA in degraded samples.[24] However, the lack of barcode-based primers limits its application to high-throughput, multiplexed cellular studies, particularly with increasingly advanced single-cell resolution sampling techniques. Therefore, we propose CDB-seq, which combines LAI single-cell resolution sampling technology, CDB microbeads, and Smart-3SEQ and optimizes the original process by adding a pre-amplification step, modulating the magnetic bead purification process, and adjusting reagents and dosages to achieve a stable output (Table S4).

We designed a novel method for the preparation of CDB microbeads based on a combinational concept, enabling the production of barcoded microbeads of 96 known sequences with only 20 primers. Fluorescent primer validation demonstrated that the surface of CDB microbeads had enough primers for mRNA capture in subsequent experiments. This synthesis method required only 64 oligonucleotides (32 × 32) to generate 1024 distinct CDB microbeads or only 144 nucleotides (72 × 72) to generate 5184 distinct CDB microbeads. It can readily accommodate 5000 sample throughput and is compatible with spatial transcriptome profiling of various types of organs. More importantly, this method is very low in cost, with a 97% to 99% cost reduction compared to that of the direct synthesis of equivalent amounts of long DNA barcode oligonucleotides (>100 nts). The number of rounds of combination can be increased to three or more when the sample throughput reaches thousands; however, every round results in a loss of probe capture efficiency (Figure 2). Therefore, CDB-seq can be combined with indexes in sequencing junctions, enabling differentiation by barcoding within a 96-well plate and labeling by indexing between plates for medium-to-high-throughput (thousands of cells) spatial transcriptome studies. Moreover, owing to its barcode-based multiplexing ability, it is more cost-effective and less time-consuming than Smart-3SEQ.

In conclusion, CDB-seq is a promising library construction method for single-cell resolution transcriptome analysis of degraded samples, providing a basis for evaluating gene expression in a wide range of pathological conditions and prognostic processes using archived specimens. This approach can provide new insights into tumor heterogeneity, drug resistance, and prognosis prediction. Furthermore, it is expected to contribute to a wide range of fields, including brain neuroscience, pathology, and embryology, and to advance our knowledge of cellular interactions and disease mechanisms.

Funding: This research was supported by the National Natural Science Foundation of China (81827901). Institutional Review Board Statement: Not applicable.

Acknowledgments: We would like to thank all members of the Zhao laboratories for their invaluable assistance and discussions: Conceptualization, K.D., Q.G. and X.Z.; methodology, K.D. and Y.Z.; software, K.Y. and Y.G.; formal analysis, Y.G. and W.J; investigation, Y.Z. and K.Y.; resources, Q.G.; writing—original draft preparation, K.D.; writing—review and editing, W.W., X.Z.; visualization, K.D. and Y.Z.; supervision, X.Z.; project administration, X.Z.; funding acquisition, X.Z. All authors have read and agreed to the published version of the manuscript. We also thank Shanghai Model Organisms Center for mice, and Shanghai Biotechnology Corporation for RNA sequencing. Moreover, we are grateful to Mogoedit (www.mogoedit.com) for Englishlanguage editing

Conflicts of Interest: The authors declare no conflict of interest

References

1. A. Giustacchini; S. Thongjuea; N. Barkas; P.S. Woll; B.J. Povinelli; C.A.G. Booth; P. Sopp; R. Norfo; A. Rodriguez-Meira; N. Ashley; L. Jamieson; P. Vyas; K. Anderson; A. Segerstolpe; H. Qian; U. Olsson-Stromberg; S. Mustjoki; R. Sandberg; S.E.W. Jacobsen; A.J. Mead *Nat Med* **2017**, *23,* 692.
2. P. van Galen; V. Hovestadt; M.H. Wadsworth; T.K. Hughes; G.K. Griffin; S. Battaglia; J.A. Verga; J. Stephansky; T.J. Pastika; J.L. Story; G.S. Pinkus; O. Pozdnyakova; I. Galinsky; R.M. Stone; T.A. Graubert; A.K. Shalek; J.C. Aster; A.A. Lane; B.E. Bernstein *Cell* **2019**, *176,* 1265.
3. J.S. Jang; Y. Li; A.K. Mitra; L. Bi; A. Abyzov; A.J. van Wijnen; L.B. Baughn; B. Van Ness; V. Rajkumar; S. Kumar; J. Jen *Blood Cancer J* **2019**, *9,* 2.
4. S. Kang; M.S. Nisar; Y. Lu; N. Chang; Y. Huang; H. Ni; S.M. Novikov; Y. Wang; Q. Cui; X. Zhao *Small Methods* **2023**, *7,* e2201379.
5. R.C. Wirka; D. Wagh; D.T. Paik; M. Pjanic; T. Nguyen; C.L. Miller; R. Kundu; M. Nagao; J. Coller; T.K. Koyano; R. Fong; Y.J. Woo; B. Liu; S.B. Montgomery; J.C. Wu; K. Zhu; R. Chang; M. Alamprese; M.D. Tallquist; J.B. Kim; T. Quertermous *Nat Med* **2019**, *25,* 1280.
6. A. Maynard; C.E. McCoach; J.K. Rotow; L. Harris; F. Haderk; D.L. Kerr; E.A. Yu; E.L. Schenk; W. Tan; A. Zee; M. Tan; P. Gui; T. Lea; W. Wu; A. Urisman; K. Jones; R. Sit; P.K. Kolli; E. Seeley; Y. Gesthalter; D.D. Le; K.A. Yamauchi; D.M. Naeger; S. Bandyopadhyay; K. Shah; L. Cech; N.J. Thomas; A. Gupta; M. Gonzalez; H. Do; L. Tan; B. Bacaltos; R. Gomez-Sjoberg; M. Gubens; T. Jahan; J.R. Kratz; D. Jablons; N. Neff; R.C. Doebele; J. Weissman; C.M. Blakely; S. Darmanis; T.G. Bivona *Cell* **2020**, *182,* 1232.
7. Y. Liu; P. Jeraldo; J.S. Jang; B. Eckloff; J. Jen; M. Walther-Antonio *Anal Chem* **2019**, *91,* 8036.
8. J. Bues; M. Biočanin; J. Pezoldt; R. Dainese; A. Chrisnandy; S. Rezakhani; W. Saelens; V. Gardeux; R. Gupta; R. Sarkis; J. Russeil; Y. Saeys; E. Amstad; M. Claassen; M.P. Lutolf; B. Deplancke *Nat Methods* **2022**, *19,* 323.
9. M.D. Young; T.J. Mitchell; F.A. Vieira Braga; M.G.B. Tran; B.J. Stewart; J.R. Ferdinand; G. Collord; R.A. Botting; D.M. Popescu; K.W. Loudon; R. Vento-Tormo; E. Stephenson; A. Cagan; S.J. Farndon; M. Del Castillo Velasco-Herrera; C. Guzzo; N. Richoz; L. Mamanova; T. Aho; J.N. Armitage; A.C.P. Riddick; I. Mushtaq; S. Farrell; D. Rampling; J. Nicholson; A. Filby; J. Burge; S. Lisgo; P.H. Maxwell; S. Lindsay; A.Y. Warren; G.D. Stewart; N. Sebire; N. Coleman; M. Haniffa; S.A. Teichmann; M. Clatworthy; S. Behjati *Science* **2018**, *361,* 594.
10. S.V. Puram; I. Tirosh; A.S. Parikh; A.P. Patel; K. Yizhak; S. Gillespie; C. Rodman; C.L. Luo; E.A. Mroz; K.S. Emerick; D.G. Deschler; M.A. Varvares; R. Mylvaganam; O. Rozenblatt-Rosen; J.W. Rocco; W.C. Faquin; D.T. Lin; A. Regev; B.E. Bernstein *Cell* **2017**, *171,* 1611.
11. A. Rao; D. Barkley; G.S. França; I. Yanai *Nature* **2021**, *596,* 211.
12. M. Asp; J. Bergenstråhle; J. Lundeberg *Bioessays* **2020**, *42,* e1900221.
13. S.K. Longo; M.G. Guo; A.L. Ji; P.A. Khavari *Nat Rev Genet* **2021**, *22,* 627.
14. L. Moses; L. Pachter *Nat Methods* **2022**, *19,* 534.
15. P.L. Ståhl; F. Salmén; S. Vickovic; A. Lundmark; J.F. Navarro; J. Magnusson; S. Giacomello; M. Asp; J.O. Westholm; M. Huss; A. Mollbrink; S. Linnarsson; S. Codeluppi; Å. Borg; F. Pontén; P.I. Costea; P. Sahlén; J. Mulder; O. Bergmann; J. Lundeberg; J. Frisén *Science* **2016**, *353,* 78.
16. S. Vickovic; G. Eraslan; F. Salmén; J. Klughammer; L. Stenbeck; D. Schapiro; T. Äijö; R. Bonneau; L. Bergenstråhle; J.F. Navarro; J. Gould; G.K. Griffin; Å. Borg; M. Ronaghi; J. Frisén; J. Lundeberg; A. Regev; P.L. Ståhl *Nat Methods* **2019**, *16,* 987.
17. Y. Liu; M. Yang; Y. Deng; G. Su; A. Enninful; C.C. Guo; T. Tebaldi; D. Zhang; D. Kim; Z. Bai; E. Norris; A. Pan; J. Li; Y. Xiao; S. Halene; R. Fan *Cell* **2020**, *183,* 1665.
18. G. Wang; J.R. Moffitt; X. Zhuang *Sci Rep* **2018**, *8,* 4847.
19. E. Lubeck; A.F. Coskun; T. Zhiyentayev; M. Ahmad; L. Cai *Nat Methods* **2014**, *11,* 360.
20. A. Chen; S. Liao; M. Cheng; K. Ma; L. Wu; Y. Lai; X. Qiu; J. Yang; J. Xu; S. Hao; X. Wang; H. Lu; X. Chen; X.

Liu; X. Huang; Z. Li; Y. Hong; Y. Jiang; J. Peng; S. Liu; M. Shen; C. Liu; Q. Li; Y. Yuan; X. Wei; H. Zheng; W. Feng; Z. Wang; Y. Liu; Z. Wang; Y. Yang; H. Xiang; L. Han; B. Qin; P. Guo; G. Lai; P. Muñoz-Cánoves; P.H. Maxwell; J.P. Thiery; Q.F. Wu; F. Zhao; B. Chen; M. Li; X. Dai; S. Wang; H. Kuang; J. Hui; L. Wang; J.F. Fei; O. Wang; X. Wei; H. Lu; B. Wang; S. Liu; Y. Gu; M. Ni; W. Zhang; F. Mu; Y. Yin; H. Yang; M. Lisby; R.J. Cornall; J. Mulder; M. Uhlén; M.A. Esteban; Y. Li; L. Liu; X. Xu; J. Wang *Cell* **2022**, *185,* 1777.

21. X. Fu; L. Sun; R. Dong; J.Y. Chen; R. Silakit; L.F. Condon; Y. Lin; S. Lin; R.D. Palmiter; L. Gu *Cell* **2022**, *185,* 4621.

22. E.Z. Macosko; A. Basu; R. Satija; J. Nemesh; K. Shekhar; M. Goldman; I. Tirosh; A.R. Bialas; N. Kamitaki; E.M. Martersteck; J.J. Trombetta; D.A. Weitz; J.R. Sanes; A.K. Shalek; A. Regev; S.A. McCarroll *Cell* **2015**, *161,* 1202.

23. S. Picelli; O.R. Faridani; A.K. Björklund; G. Winberg; S. Sagasser; R. Sandberg *Nat Protoc* **2014**, *9,* 171.

24. J.W. Foley; C. Zhu; P. Jolivet; S.X. Zhu; P. Lu; M.J. Meaney; R.B. West *Genome Res* **2019**, *29,* 1816.

25. S. Kim; A.C. Lee; H.B. Lee; J. Kim; Y. Jung; H.S. Ryu; Y. Lee; S. Bae; M. Lee; K. Lee; R.N. Kim; W.Y. Park; W. Han; S. Kwon *Genome Biol* **2018**, *19,* 158.

26. X. Han; R. Wang; Y. Zhou; L. Fei; H. Sun; S. Lai; A. Saadatpour; Z. Zhou; H. Chen; F. Ye; D. Huang; Y. Xu; W. Huang; M. Jiang; X. Jiang; J. Mao; Y. Chen; C. Lu; J. Xie; Q. Fang; Y. Wang; R. Yue; T. Li; H. Huang; S.H. Orkin; G.C. Yuan; M. Chen; G. Guo *Cell* **2018**, *173,* 1307.

27. C.G. Williams; H.J. Lee; T. Asatsuma; R. Vento-Tormo; A. Haque *Genome Med* **2022**, *14,* 68.

28. K.H. Chen; A.N. Boettiger; J.R. Moffitt; S. Wang; X. Zhuang *Science* **2015**, *348,* aaa6090.

29. M. Kaufmann; A.L. Schaupp; R. Sun; F. Coscia; C.A. Dendrou; A. Cortes; G. Kaur; H.G. Evans; A. Mollbrink; J.F. Navarro; J.K. Sonner; C. Mayer; G.C. DeLuca; J. Lundeberg; P.M. Matthews; K.E. Attfield; M.A. Friese; M. Mann; L. Fugger *Nat Neurosci* **2022**, *25,* 944.

30. Y. Wang; J.L. Fan; J.C. Melms; A.D. Amin; Y. Georgis; I. Barrera; P. Ho; S. Tagore; G. Abril-Rodríguez; S. He; Y. Jin; J. Biermann; M. Hofree; L. Caprio; S. Berhe; S.A. Khan; B.S. Henick; A. Ribas; E.Z. Macosko; F. Chen; A.M. Taylor; G.K. Schwartz; R.D. Carvajal; E. Azizi; B. Izar *Nat Genet* **2023**, *55,* 19.

31. C.R. Merritt; G.T. Ong; S.E. Church; K. Barker; P. Danaher; G. Geiss; M. Hoang; J. Jung; Y. Liang; J. McKay-Fleisch; K. Nguyen; Z. Norgaard; K. Sorg; I. Sprague; C. Warren; S. Warren; P.J. Webster; Z. Zhou; D.R. Zollinger; D.L. Dunaway; G.B. Mills; J.M. Beechem *Nat Biotechnol* **2020**, *38,* 586.

32. S.R. Greytak; K.B. Engel; B.P. Bass; H.M. Moore *Cancer Res* **2015**, *75,* 1541.

33. H. Fraenkel-Conrat; H.S. Olcott *J Am Chem Soc* **1948**, *70,* 2673.