Epitranscriptomic features, such as single-base RNA editing, are sources of transcript diversity in cancer, but little is understood in terms of their spatial context in the tumour microenvironment. Here, we introduce spatial-histopathological examination-linked epitranscriptomics converged to transcriptomics with sequencing (Select-seq), which isolates regions of interest from immunofluorescence-stained tissue and obtains transcriptomic and epitranscriptomic data. With Select-seq, we analyse the cancer stem cell-like microniches in relation to the tumour microenvironment of triple-negative breast cancer patients. We identify alternative splice variants, perform complementarity-determining region analysis of infiltrating T cells and B cells, and assess adenosine-to-inosine base editing in tumour tissue sections. Especially, in triple-negative breast cancer microniches, adenosine-to-inosine editome specific to different microniche groups is identified.
The tumour microenvironment (TME) contains microniches with spatially heterogeneous transcriptomic and epitranscriptomic features, such as alternative splicing or non-synonymous single-base RNA editing. One dynamic epitranscriptomic modification that generates transcript diversity in microniches is adenosine deaminases acting on RNA (ADAR) enzyme-mediated adenosine-to-inosine (A-to-I) editing, including changes in translated and untranslated regions that affect the TME functionally and pathologically. This, single-base-resolution epitranscriptomic analysis of the tumour microniches is the key to understanding how A-to-I editing affects the tumour. In addition, to characterize the microniches, spatial transcriptomic data must be accompanied by spatial epitranscriptomics data. Several technologies that enable gene expression analysis within histopathological and spatial contexts have been applied to depict the spatial landscape of the tumour.

However, spatially barcoded transcripts and in situ barcodes in spatial transcriptomic methodologies use only fractions of the full-length transcriptome due to limitations in reading length in most widely adopted low error next-generation sequencing (NGS), which makes it difficult to analyze epitranscriptomic information, such as alternative splicing or A-to-I editing.

Recent studies incorporated microarray-based spatial technology for long-read NGS to realize the full-length spatial transcriptome after spatial barcoding, but the sequencing accuracy of long-read NGS for distinguishing single-base variants is not yet comparable to that of short-read NGS. Thus, investigation of the epitranscriptomic features in each intratumoural microniche necessitates simultaneous multi-modal analysis of transcriptomic, epitranscriptomic, and spatial information at single-base resolution. Moreover, there were methodological difficulties that laser capture microdissection (LCM) that can isolate ROIs from IF-stained tissues need to overcome for spatial epitranscriptome analysis. Spatial epitranscriptome (as well as immune cell receptor sequences) require a sizable number of samples and high-quality transcriptome data to secure the statistical importance of rare and single-nucleotide level event. Also, RNA is a very fragile material that continues to degrade and the quality can drop by 70% within an hour.

For the in-depth multi-modal analysis of the tumour microniches, we introduce spatial-histopathological examination-linked epitranscriptomics converged to transcriptomics with sequencing (Select-seq), a method that isolates regions of interest (ROIs) as small as single cells from immunofluorescence (IF)-stained tissue and obtains full-length transcriptome data at single-base resolution, connected to the spatial and staining information therein (Fig. 1a). Specifically, the selective isolation of every single ROI enables barcoding of the full portion of the transcriptome, which leads to an in-depth and multi-modal analysis of the tumour microniches. To address previous methodological difficulties, Select-seq was developed to isolate ROIs that contain a very small number of cells in high throughput (Supplementary Fig. 1). Leveraging the advantage of Select-seq, we investigated the hypothesis that cancer stem cells (CSCs) in triple-negative breast cancer (TNBC) have characteristic A-to-I editing-based regulation. We explored the transcriptomic and epitranscriptomic landscape of TNBC tumours, whose microniches contain CSC-like cells. IF staining of two CSC-related proteins, CD44 and ALDH1, was applied to define the ROIs, which were selectively isolated using a pulsed near-infrared (NIR) laser retrieval system that isolates ~100 ROIs in 1 min. The obtained transcriptomic and epitranscriptomic profiles in connection to the spatial and staining information for every ROI (Fig. 1c). We then spatially mapped gene expression, alternative splice variant expression, immune cell receptor sequences, and A-to-I-edited sequences (Fig. 1d).

Transcriptomic features of ALDH1-high-stained regions with tumorigenicty. To investigate the spatial transcriptomic and epitranscriptomic landscape of TNBC, we performed Select-seq on five TNBC patients (Fig. 1 and Supplementary Fig. 2). To search for CSC-like microniches in a primary tumour, we initially investigated 106 target regions (Supplementary Fig. 3) in primary tumour sections from patient A (tissue A, ID: 190603). Fresh-frozen tissue sections were fixed with PFA and stained with IF probes targeting CD44 and ALDH1 to determine the stem cell-like microniche within the tumour. We grouped the targets into four groups: (i) CD44+/ALDH1+, (ii) CD44low−/ALDH1high, (iii) CD44high/ALDH1low−, and (iv) CD44low−/ALDH1low−.
The microniches in the four staining groups with different immunofluorescence patterns were characterized according to the spatial transcriptomic data. We first analyzed the spatial heterogeneity of the tumour, revealing several Lehmann TNBC subtypes within the same tumour (Fig. 3c). The CD44\textsuperscript{low/−}/ALDH1\textsuperscript{high} group and the CD44\textsuperscript{+}/ALDH1\textsuperscript{−} group mostly consisted of immunomodulatory (IM) and mesenchymal stem cell-like (MSL) subtypes, while the CD44\textsuperscript{high}/ALDH1\textsuperscript{low/−} group had a mixed population of basal-like-1 (BL1) TNBC subtypes. Although TNBC is a type of breast cancer that lacks oestrogen receptor (ER), progesterone receptor (PR), and ERBB2 at the protein level, some populations within the tumour section expressed ERBB2 at the gene level, suggesting the intratumoural heterogeneity of the breast cancer type within a given tumour (Fig. 3d, Supplementary data 1).

We also observed four different groups by principal component analysis (PCA) that were aligned to the immunofluorescence staining groups (Fig. 3e) except for the CD44\textsuperscript{+}/ALDH1\textsuperscript{+} group that showed some overlap with The CD44\textsuperscript{low/−}/ALDH1\textsuperscript{high} or CD44\textsuperscript{high}/ALDH1\textsuperscript{low/−} group. To investigate the developmental relationship between the microniches, we examined the RNA expression profiles, transcript isoforms, B cell receptor sequences, and adenosine-to-inosine (A-to-I) editing events. Each transcriptomic and epitranscriptomic data point was mapped to the tissue based on the barcodes.
velocity of the different microniches (Fig. 3f). The RNA velocity plot showed that the CD44low/ALDH1high/ALDH1low/ALDH1high micrioniches tended to develop into CD44high/ALDH1low/ALDH1high micrioniches. Then, we analyzed whether the targets expressed previously reported cancer-related genes. CSC gene expression signature patterns were observed mostly in the CD44low/ALDH1high group and sometimes in the CD44high/ALDH1low/ALDH1high group (Fig. 3d). Furthermore, 22 previously reported upregulated genes in CSCs, such as COL1A2, ENPP2, and PCOLCE, and 15 downregulated genes, such as COL19A1, PLPP2, and HOOK1, were analyzed in the target ROIs. The CD44low/ALDH1high ROIs showed CSC-like gene expression features. When we analyzed the alternative splice variants, the biotype analysis from the ENSEMBL transcript reference showed that protein-coding alternative splice variants constituted 42.3% of all biotypes (Supplementary Fig. 4).

**Tumour infiltrating plasma cell gene signatures and corresponding BCR sequences are mapped onto the tumour microenvironment.** To further assess the tumorigenicity of the CSC-like microniches, we examined the tumour immune microenvironment (Figs. 3g and 4). To effectively present gene expressions of ROIs that have the same staining phenotype and are close to each other, we grouped one to 12 ROIs and defined 22 ROI groups to compare the transcriptomic and epitranscriptomic information (Supplementary Table 2). We confirmed that immunomodulatory subtype-related gene ontology terms (Fig. 3c) and gene signatures related immunosuppression, such as PDL1 (CD274), interleukin 10 (IL10RA, IL10RB), TGF-β (TGFβ), and XBP1, were observed in the CD44low/ALDH1high ROI groups (Fig. 3g, h). We next examined the gene expression signatures for naive B cells or centrocytes (LMO2 and BCL6), memory B cells (PRDM4), and plasmablasts or plasma cells (PRDM1 and XBP1) (Fig. 4a, b). It is interesting to note that the CD44low/ALDH1high micrioniches where PRDM1 and XBP1 were highly expressed had high IGHG1 expression indicating plasma cell infiltration that was suggested to be associated with poor prognosis (Fig. 4c). We were able to extract 7 TCRs and 204 BCRs from the same ROI data we used for gene expression typing analysis. Among them, CDR3 sequences were recovered from 74% of the total BCRs. The extracted BCR Heavy chain comprised of 25.7% G1 isotype and the isotypes for 74.3% were...
not recovered. Although additional data should be acquired, the prevalence of the HCDr3 amino acid sequence during gene rearrangement was confirmed. To further characterize how these microniches are comprised of, we performed single-cell deconvolution using CIBERSORT (Fig. 4d). In the case of CD44-positive IF group, we observed the deconvolution result for each IF group and found out some sites were mainly composed of tumour cells and some other sites were composed of endothelial, fibroblast, and CD44+ T cells. These results were aligned with the tissue type we used and also showed the immune status of our tissue. Also, it is interesting to note that the cancer stem cell microniches with high ALDH1 expression seemed to be located more in the stromal region and through single-cell deconvolution of these microniches we were able to see that the fibroblast population is more observed in these regions. To view this trend at a glance, we brought the gene expression data of the ROIs to each IF staining group and examined the distribution of different cell types. As we displayed B cell population existed in...
which is a local mix of CD44+ distance show common A-to-I editome, but also that some genes had differentially expressed genes (Supplementary Fig. 8). We found out the total 798 A-to-I editing events that were uniquely detected from lysine residue to serine residue. Although the GPX4 203 variant was identified to be a non-stop decaying mRNA, we assessed abnormalities that the GPX4 transcript variant might have caused. We discovered that the CD44low/−/ALDH1high microniches had exceptionally high gene set enrichment for ferroptosis, which is an emerging druggable target and can be further investigated by Select-seq (Supplementary Fig. 9). Also, to assess if the corresponding position is theoretically accessible by the ADAR protein, we predicted the secondary structure with Forna39, and was confirmed to form double-stranded RNA in the site. The amino acid residue was altered from lysine residue to serine residue. The GPX4 gene at position 1106616 reported in the same position36–38, which was neither detected in the germline level nor tumour somatic variant in these cases, but study of the post-transcriptional variant has been limited. In the matched genomic DNA, we confirmed that only adenosine sequence is detected in the germline level nor tumour somatic variant in the corresponding position through Sanger sequencing and next-generation sequencing (Supplementary Fig. 9). Also, to assess if the corresponding position is theoretically accessible by the ADAR protein, we predicted the secondary structure with Forna39, and was confirmed to form double-stranded RNA in the site. The amino acid residue was altered from lysine residue to serine residue. The ADAR gene expression when grouped into four IF staining groups and had no meaningful correlation with the number of ADAR genes detected. Therefore, A-to-I-edited events we found throughout the tissue are meaningful RNA editing events (Supplementary Fig. 6). Among them, 632, 546, and 11,030 events were in repetitive, non-repetitive, and ALC groups, respectively. The exonic proportion of the edited sites was ~1%, and among them, non-synonymous editing consisted of 80% when annotated with the GENCODE reference35. Interestingly, A-to-I-edited sites shared both conserved and characteristic A-to-I-edited variants specific to the spatial groups categorized by IF staining and physical distance between the ROIs. We found that breast tumours maintain a reservoir of subclonal diversity43, we sought to investigate whether the GPX4 A-to-I editing event is related to residual cancer microniche. We applied Select-seq to four other patients who received neoadjuvant chemotherapy and supplied tissues B (ID: 180908T), C (ID: 180807T), D (ID: 190422T), and E (ID: 200710T), from which 161 target regions were selected for Select-seq (Supplementary Fig. 2). After screening the GPX4 gene in the microniche in the four tissues, we separated them into two groups: those with the A-to-I-edited GPX4 variant and those without the variant. We then depicted the volcano plot of the two different microniche groups in tissues B, C, D, and E (Fig. 6d). Specifically, the expression of the iron uptake protein-encoding gene TFRC was downregulated, while the expression of the iron storing ferritin-encoding gene FTH1/FTL and the lipid ROS elimination-related protein-encoding gene GPX4 was upregulated. We suspect that similar to the previously reported A-to-C single-nucleotide

ROIs with high ALDH1 expression. Also, as discussed above, the fibroblast population seems to be larger in ALDH1 high expressing ROIs and a more malignant cell population is observed in CD44 high expressing ROIs.

A-to-I editome specific to CD44low/−/ALDH1high-stained microniches is identified. In addition to the transcriptomic landscape, we analyzed single-base resolution epitranscriptomics profiles to discover characteristic A-to-I-edited transcript variants in CSC-like microniche. This irreversible post-transcriptional deamination of adenosine to inosine in double-stranded RNA is catalyzed by the ADAR family3,32,33 and the converted inosine base is sequenced as a guanine base. The ADAR gene was uniformly expressed, implicating A-to-I events throughout the tissue (Fig. 5a and Supplementary Fig. 5). We observed a total of 12879 A-to-I-edited events using REDTools34 (Fig. 5b) in the first tissue and explored spatial a-to-I edithome in the other four tissues (Supplementary Fig. 7). To validate the A-to-I events didn’t come from technical variation, we checked the correlation between the number of A-to-I events with the number of genes detected and ADAR gene expression. We found out the number of A-to-I events had a strong correlation with ADAR gene expression when grouped into four IF staining groups and had no meaningful correlation with the number of ADAR genes detected. Therefore, A-to-I-edited events we found throughout the tissue are meaningful RNA editing events (Supplementary Fig. 6). Among them, 632, 546, and 11,030 events were in repetitive, non-repetitive, and Alu regions in the genome, respectively. The exonic proportion of the edited sites was ~1%, and among them, non-synonymous editing consisted of 80% when annotated with the GENCODE reference35. Interestingly, A-to-I-edited sites shared both conserved and characteristic A-to-I-edited variants specific to the spatial groups categorized by IF staining and physical distance between the ROIs (Figs. 5c, 6a, b and Supplementary Data 2). We found out the total 798 A-to-I-edited events that were uniquely defined for each spatial group, which had differentially expressed genes (Supplementary Fig. 8). It is interesting to see not only that ROIs with shorter inter-distance show common A-to-I edithome, but also that some genes like BRC2A, SOX9-AS1, and more are specific to spatial group C, which is a local mix of CD44+/ALDH1+ and CD44low/−/ALDH1high cells in the stromal region between the two ducts with CD44high/ALDH1low/− population. In terms of A-to-I edithome specific to IF staining, characteristic non-synonymous A-to-I events were aligned to the GPX4 gene at position 1106616 (rs713041), specifically in GPX4 NM_001039848.4 alternative splice variants from the CD44low/−/ALDH1high ROIs (Fig. 6c), with different frequencies ranging from 0.23 to 0.80, and were validated with capillary electrophoresis sequencing (Supplementary Fig. 5). A single-nucleotide polymorphism (SNP) has been reported in the same position (1106616)36–38, which was neither detected in the germline level nor tumour somatic variant in these cases, but study of the post-transcriptional variant has been limited. In the matched genomic DNA, we confirmed that only adenosine sequence is detected in the germline level nor tumour somatic variant in the corresponding position through Sanger sequencing and next-generation sequencing (Supplementary Fig. 9). Also, to assess if the corresponding position is theoretically accessible by the ADAR protein, we predicted the secondary structure with Forna39, and was confirmed to form double-stranded RNA in the site. The amino acid residue was altered from lysine residue to serine residue. Although the GPX4 203 variant was identified to be a non-stop decaying mRNA, we assessed abnormalities that the GPX4 transcript variant might have caused. We discovered that the CD44low/−/ALDH1high microniches had exceptionally high gene set enrichment for ferroptosis, which is an emerging druggable target and can be further investigated by Select-seq (Supplementary Fig. 5). Other than this specific A-to-I variant, we were able to observe that more portions of ALU sites were edited from adenosine to inosine in ROIs with high ALDH1 expression. More A-to-I-edited Alu sites are closely related to worse patient outcomes40 and we suspect that this further suggests a potential epitranscriptomic signature specific to the ALDH1 high expressing cancer stem cell microniche.

A-to-I editing events in GPX4 gene are related to ferroptosis-related gene expression. Ferroptosis is a non-apoptotic and lipid reactive oxygen species (ROS)-related type of programmed cell death that affects inflammation-associated immunosuppression in tumours41,42. We noted that the CD44low/−/ALDH1high microniche with A-to-I-edited GPX4 variants had high ferroptosis-associated gene set level changes with high expression levels of GPX4. With recent findings that breast tumours maintain a reservoir of subclonal diversity43, we sought to investigate whether the GPX4 A-to-I editing event is related to residual cancer microniche. We applied Select-seq to four other patients who received neoadjuvant chemotherapy and supplied tissues B (ID: 180908T), C (ID: 180807T), D (ID: 190422T), and E (ID: 200710T), from which 161 target regions were selected for Select-seq (Supplementary Fig. 2). After screening the GPX4 gene in the microniche in the four tissues, we separated them into two groups: those with the A-to-I-edited GPX4 variant and those without the variant. We then depicted the volcano plot of the two different microniche groups in tissues B, C, D, and E (Fig. 6d). Specifically, the expression of the iron uptake protein-encoding gene TFRC was downregulated, while the expression of the iron storing ferritin-encoding gene FTH1/FTL and the lipid ROS elimination-related protein-encoding gene GPX4 was upregulated. We suspect that similar to the previously reported A-to-C single-nucleotide
Fig. 4 Spatial analysis of the immune cell repertoire. a Spatial mapping of the signature genes related to B cell maturation (n = 106 biologically independent samples (ROI)). Bar plot indicates mean value and error bar indicates standard deviation. b Six representative BCR sequences recovered using Select-seq and matched for gene expression related to B cell maturation. Overall, both heavy and light chain sequences from 42.9% of samples were recovered using Select-seq. Representative sequences including isotype, CDR3, matched V and J genes for each heavy and light chain. c Matched spatial context of recovered BCR sequences. Most of the BCRs were detected in the CD44high/ALDH1high stromal region, and very few BCRs were detected in the CD44low/-/ALDH1high region, and two were recovered from the CD44/ALDH1– region. d Single-cell deconvolution of each ROI (left) and ROI group (right). Source data for b–d are provided as a Source Data file.

Variations at the genomic level, this alteration causes the over-expression of GPX4 genes. Additionally, pro-ferroptotic VDAC3 gene expression in the A-to-I-edited microniches suggests that these cells will not develop resistance to ferroptosis-inducing drugs such as erastin, which binds to VDAC3. In microniches without A-to-I editing in the GPX4 gene, the TFAP2C, ITGB6, and SPINT2 genes, which are known to be downregulated in CSCs, were overexpressed. Therefore, differentially expressed ferroptosis-related genes and GPX4 A-to-I editing events throughout the tissue were observed and led to our hypothesis that GPX4 A-to-I editing events affect ferroptosis in CSC-like microniches. We further explored whether A-to-I editing in GPX4 genes was related to clinical outcomes in TNBC patients.

We sought to analyze other bulk TNBC transcriptomic datasets that are publicly available in TCGA to see the relation. Bulk transcriptome data from 109 TNBC patients were analyzed, 80 of
which patients had the GPX4 variants that included A-to-I-edited variants, SNPs or somatic mutations at the genome level. We observed the gene set related to ferroptosis was highly enriched when the median frequency of the GPX4 variant in the transcript data was higher (Supplementary Fig. 10). The overall survival rates of the two groups estimated using the Kaplan–Meier method showed that the group with a high frequency of the GPX4 1106616 variant at the transcriptome level (positive ferroptosis gene set enrichment value) showed a trend toward worse clinical outcomes compared to the group with a low frequency of the variant (Supplementary Fig. 10). Furthermore, we compared the correlation of survival with a combination of ADAR genes, total
A-to-I events, and even patients who have SNP within GPX4. We found out none of them showed similar results but only GPX4 11066616 variant has the most strong correlation with survival difference grouped by ferroptosis gene set enrichment value (Supplementary Fig. 11). Although more assessments are required to validate that targeting GPX4 abnormalities would provide therapeutic responses, Select-seq was able to unveil GPX4 A-to-I editing events in relation to CSC-like gene expression signatures in TNBC patients.

Discussion

Using Select-seq, we revealed that the CD44low/-/ALDH1high micrometastases in the stromal regions had transcriptional signatures with high levels of B cell activity and A-to-I editing. Furthermore, we were able to efficiently find the link between the epitranscriptomics signatures of CSC-like micrometastases in TNBC tumours. In particular, we identified GPX4 A-to-I variants related to clinical outcomes. Although the single-base RNA edited variants should further be assessed with additional experiments such as applying recombinant ADAR to the variant in the future or methodologies like cyanoethylation and RT-PCR sequencing, these findings can provide clues to identify useful CSC-related druggable targets as emerging therapeutic targets. With ROI-based, in-depth and multi-modal analysis, Select-seq allows comprehensive spatial epitranscriptomics analysis. Although we could match only a few immune cell receptors of the tumour infiltrating lymphocytes to the existing database, further studies that can pair the antigen-expressing genes of these receptors will lead to deeper studies of how these tissue infiltrating immune cells are interacting in the cancer microenvironment. Also, the studies using alternative splicing variants can be extended to study fusion genes for potential drug targets. By integrating the current state-of-art highly parallel spatial transcriptome analysis and single-cell RNA sequencing, single-cell analysis and other image-based analysis techniques such as in situ sequencing to the spatial transcriptome data in target regions, Select-seq will become a complementary tool that enables the deep analysis of ROIs.

Methods

Region of interest (ROI) isolation. The Spatially-resolved Laser Activated Cell Sorter (SLACS) instrument comprises optical modules and mechanical modules for high-throughput isolation of micrometastases from tissue samples. It already exists for handling the retrieved targets. An X-Y-Z axis motorized stage (ACS Motion Control, Migdal, Israel) was built to control the spatial location of the target. The device can be controlled automatically by communicating with a computer. One stage is for loading sample slides, and the other stage is for loading tubes to receive isolated cells. A charge-coupled device (CCD) camera (Jenoptik, Jena, Germany) was installed to observe where the laser pulse will be applied through the objective lenses. A neodymium-doped yttrium aluminum garnet (Nd:YAG) nanosecond laser was purchased from Continuum (Minilite™ Series ML II; Continuum, San Jose, CA). There is a slit located in the light path between the laser source and the objective lens to control the region to be isolated. The slit is controlled either manually or automatically to adjust the size of the laser pulse. Objective lenses with various magnifications were purchased from Mitutuyo. The long working distance allows more space between the lens and the sample for user convenience.

Communication system for SLACS. We designed two different pieces of software, which were written in Python scripts. The first was built for the user (a pathologist in our case) to select the cells to be isolated. We shared the whole-slide image with the user through a server, and the user ran the software to select the cells of interest while navigating the tissue image through the graphical user interface. After selection, the program produces two files: a text file with locational information about the region of interest and the image file with the selected targets is overlaid with transparent blue on the original image. Both files are required for the automated isolation of the target cells. The second software enables the automatic control of the SLACS instrument. With this software, the users are able to control the slits, change the objective lenses, and move the motorized stages. It also enables automatic target isolation when two files from the first software are loaded. All tissue samples were isolated using an automatic function, while the cell line experiments were performed manually.

Cell culture. Human HEK 293 T (cat # 21573), human IM-9 cells (cat # 18059), and NIH3T3 cells (cat # 21665) were purchased from Korean Cell Line Bank (KCLB) and propagated according to the manufacturer’s instructions. HEK 293 T cells were cultured in DMEM (Thermo Fisher Scientific, Massachusetts, USA) with 1% penicillin-streptomycin (Corning, New York, USA) and 10% foetal bovine serum (HyClone, Massachusetts, USA) at 37 °C under 5% CO2 and 95% atmospheric air; IM-9 cells were cultured in RPMI 1640 (Thermo Fisher Scientific, with 1% penicillin-streptomycin (Corning, New York, USA) and 10% foetal bovine serum (FBS, HyClone) at 37 °C under 5% CO2 and 95% atmospheric air; HuT cells were cultured in DMEM (Thermo Fisher Scientific, Massachusetts, USA) with 1% penicillin-streptomycin (Corning, New York, USA) and 10% FBS (HyClone, Massachusetts, USA) at 37 °C under 5% CO2 and 95% atmospheric air; NIH3T3 cells were cultured in DMEM (Thermo Fisher Scientific, Massachusetts, USA) with 1% penicillin-streptomycin (Corning, New York, USA) and 10% bovine calf serum (HyClone) at 37 °C under 5% CO2 and 95% atmospheric air. Adherent cells such as HEK 293 T cells and NIH3T3 cells were grown to a confluence of 50−80% and treated with TrypLE (Invitrogen, California, USA) for five described above procedures. For confocal imaging, we fixed the tissue sections in 10% formaldehyde for 15 min and washed with DPBS at room temperature; for methanol (MeOH) fixation, smeared cells were serially fixed with 70%, 90%, and 99% MeOH in distilled water for 30 s and rehydrated by dipping for 30 s in reverse order.

Tissue preparation. Tissue sections were acquired from the archives of the biorepository of Lab of Breast Cancer Biology at the Cancer Research Institute, Seoul National University. The preparation of tissue was approved by the Institutional Review Board of Seoul National University Hospital (SNUH, IRB No. 458-009050). The patients provided written consent without compensation for the archiving and use of tissue and blood samples for research purposes. For PFA fixation, we fixed the tissue sections with 4% PFA in PBS on ice for 15 min and then washed them twice with ice-cold PBS containing 1% recombinant RNase inhibitor (Invitrogen, California, USA) and re-spun at 1500 rpm for 5 min. The supernatant was again removed, and cells were resuspended in 10 μl and spread on indium tin oxide (ITO) glass (Fine Chemicals Industry, Republic of Korea).

For paraformaldehyde (PFA) fixation, smeared cells were fixed with 4% PFA solution (Thermo Fisher Scientific, Massachusetts, USA) for 15 min and washed with DPBS at room temperature; for methanol (MeOH) fixation, smeared cells were serially fixed with 70%, 90%, and 99% MeOH in distilled water for 30 s and rehydrated by dipping for 30 s in reverse order.

For haematoxylin and eosin (H&E) staining, we fixed the tissue sections with 4% PFA as described above, except for the air-drying process. For picro-sin fixed the tissue sections in 10% paraformaldehyde (PFA) solution (Thermo Fisher Scientific, Massachusetts, USA) for 1 h. We then blocked the sections in PBS with 1% BSA for 15 min on ice. We washed the sample three times with 100 μl of recombinant RNase inhibitor (Invitrogen, California, USA) and re-spun at 1500 rpm for 3 min. The supernatant was again removed, and cells were resuspended in 10 μl and spread on indium tin oxide (ITO) glass (Fine Chemicals Industry, Republic of Korea).

For haematoxylin and eosin (H&E) staining, we fixed the tissue sections with 4% PFA as described above, except for the air-drying process. For picro-sin fixed the tissue sections in 10% paraformaldehyde (PFA) solution (Thermo Fisher Scientific, Massachusetts, USA) for 1 h. We then blocked the sections in PBS with 1% BSA for 15 min on ice. We washed the sample three times with 100 μl of recombinant RNase inhibitor (Invitrogen, California, USA) and re-spun at 1500 rpm for 3 min. The supernatant was again removed, and cells were resuspended in 10 μl and spread on indium tin oxide (ITO) glass (Fine Chemicals Industry, Republic of Korea).

Modified Smart-seq2. To lyse cells, isolated cells were added to a 0.2 ml thin-wall PCR tube containing 2 μl of a midypotent lysis buffer composed of 0.2% Triton X-100 (Sigma–Aldrich, Germany) and 2μl of recombinant RNase inhibitor (20 U/μl, Takara, Japan), 1 μl of 10 mM oligo- dt primer (Macrogen, Republic of Korea, 5'-AACAGCTGGTATACACCGAGAGACTACGT3’)-1, and 1 μl of 10 mM dNTP mix (Takara, Japan) and were incubated at 72 °C for 15 min and then immediately placed on iced. In the case of PFA-fixed samples, fixed cells were added to 0.2 ml thin-wall PCR tubes containing 2 μl of an enzymatic lysis buffer composed of 2.5 mg/mL Proteinase K (10 mg/mL Sigma–Aldrich, Germany) in nuclease-free water (Invitrogen, California, USA), 1 μl of oligo- dt primer, and 1 μl of dNTP mix (Takara, Japan), and then they were incubated at 50 °C for 1 h, 95 °C for 10 min, and immediately placed on iced afterwards.

DNA isolated from extracted RNA was prepared with SMART-seq2 protocol with the following modifications: we prepared 6 μl of the first-strand reaction mix, containing 0.5 μl SuperScript II (200 U/μl, Invitrogen, California, USA).
USA), 0.25 μl recombinant RNA inhibitor (40 U/μl, Takara, Japan) 2 μl SuperScript II First-Strand Buffer (3x, Invitrogen, California, USA), 0.25 μl dichlorodiphenyltrichloroethane (DDT) (100 mM, Invitrogen, California, USA), 2 μl betaine (5 M, Sigma-Aldrich, Germany), 0.06 μl MgCl2 (1 M, Sigma-Aldrich, Germany), 0.1 μl template switching oligo (TSO) (100 μM, Bioneer, Republic of Korea, 5'-AAGCAGTGATGATACCGAGATGGrC(G)3-5'), and 0.59 μl nuclelease-free water (Qagen, Germany).

After reverse transcription and template switching, cDNA was amplified with an additional PCR mix composed of 12.5 μl of Kapa HotStart HIIF 2x Ready Mix (Kapa Biosystems, Switzerland), 0.25 μl of PCR primers (10 μM, Macrogen, Republic of Korea, 5'-AAGCAGTGATGATACCGAGATGGrC(G)3-5') and 2.25 μl of nuclelease-free water (Qagen, Germany) for 20 or 25 cycles for RNA from single cells21. PCR products were purified using CeleMag beads (Celemics, Republic of Korea). The average fragment length of purified cDNA was determined by electrophoresis within a 1.2% agarose gel, and the concentration of cDNA was determined using the Qubit dsDNA High Sensitivity Assay Kit (Life Technologies, California, USA) according to the manufacturer’s protocol.

Next-generation sequencing. The pTXB1 cloning vector, which introduced hyperactive E54K and L372P mutations into wild-type Trα5, was acquired from Addgene. pTXB1 Trα5 and its mutants were expressed and purified31. Then, 50 bp paired-end sequencing was performed on an Illumina NextSeq sequencing platform, resulting in an average read depth of ~600 M reads per sample.

Read alignment and gene quantification. We demultiplexed and trimmed the raw sequencing reads using Cutadapt32. Then, we filtered out the reads with a sequencing quality less than 15 and a read length shorter than 25 bp. The remaining reads were aligned against the mouse genome (GRCm38) for mouse cell line samples and against the human genome (GRCh38) for human cell line samples and breast cancer samples using STAR aligner33 with the default settings. The number of uniquely mapped reads of each sample was calculated using featureCounts with default parameters for downstream differential expression analysis and RNA-seq by expectation-maximization (RSEM) with default parameters for fragments per kilobase of transcript per million mapped reads (FPKM) count. To quantify and normalize the expression of genes, data from FeatureCounts34 were normalized using the normalization method implemented in DESeq235.

Spatial gene expression visualization. The spatial gene expression was plotted in Python and matplotlib using the gene expression data from normalized Feature-Counts and matched positional information data of isolated samples that were selected and recorded by custom software before the isolation process.

Differential gene expression analysis. For analysis of differential gene expression between samples or selected spatial regions, we conducted differential gene expression analysis using DESeq2. Only genes with a log2-fold change >1 and an adjusted p-value of <0.05 were considered positively differentially expressed genes. Pathway analysis was performed using the R package g:Profiler on the normalized expression counts using DEseq2 and the mapped pathway was visualized using the R package pathview37. Gene expression levels were mapped to corresponding pathways by Kyoto Encyclopedia of Gene and Genomes (KEGG) enrichment or Molecular Signatures Database (MSigDB: http://www.broad.mit.edu/gsea) analysis. The optimal number of stable triple-negative breast cancer (TNBC) subtypes was determined by using the lists of gene sets followed by Lehmann TNBC subtypes.

Fluorescence in situ hybridization (FISH). Stellaris® FISH RNA Probes (Human ERRB2 with Quasar® 570 dye) and the other reagents were purchased from Bio-search Technologies (California, USA). Fresh-frozen dissected tissues at a thickness of 4–10 μm were mounted onto microscope slides. The slide-mounted tissue sections were immersed in fixation buffer (5.7% formaldehyde in RNA-free PBS) for 10 min at room temperature, washed twice with PBS for 5 min, permeabilized with 70% ethanol for 1 h at room temperature and immersed in wash buffer A for 5 min. The tissue sections were incubated with a hybridization buffer containing the probe at 37 °C for 4 h followed by counterstaining the nuclei with wash buffer A consisting of 5 ng/ml DAPI. The slides were rinsed with wash buffer B and mounted with an aqueous mounting medium.

RNA velocity analyses. RNA velocity was calculated on the basis of spliced and unspliced transcript reads. Based on the velocyto pipeline, annotation of spliced and unspliced reads was performed using the Python script velocyto.py and run_smartseq2 options. Annotated reads were filtered based on the number of reads of exonic, intronic, and spanning regions. Principal component analysis (PCA) and t-distributed stochastic neighbour embedding (t-SNE) analyses were performed using filtered reads, and RNA velocity was estimated using a gene-relative model with K-nearest neighbour cell pooling (K = 10). Velocity was projected onto the embedding space of PCA or t-SNE. We used the standard R implementation of velocyto default settings with velocyto.R.

RNA editing analysis by REDtools. RNA editing sites were detected from STAR-aligned reads using REDtools38. Hars parameter settings (-c 10.10 -m 25.25 -3 -q 25.25 -e 0.1 -l -p) were used to assess the backmapped false positives. A list of known adenosine-to-inosine (A-to-I) editing sites was downloaded from the REDPortal database, and only the known editing sites were selected. In total, 20,367 editing sites were detected and applied to further A-to-I editing analysis.

TCR and BCR extraction. The TCR sequences and BCR sequences for each sample were assessed using TraCeR39 and BraCeR39, which allowed the reconstruction of the TCR and BCR sequences from the RNA-seq data. With both software programs, we obtained the CDR3 sequence and the rearranged TCR and BCR genes from the sample that we spatially isolated. TraCeR and BraCeR both first made the reference file that has the possible combination of V and J genes. Then, RNA-seq reads from each cell were aligned against each reference file using the Bowtie2 aligner. The reads that aligned to the appropriate reference were assembled by Trinity RNA-seq assembly software. Finally, contigs assembled by Trinity were used as input to IgBlust, and the resulting output text files provided information about the CDR3 sequence and the rearranged TCR and BCR genes. To display rearranged TCR and BCR genes, we first filtered only those genes whose E value of the V gene was 5×10−3 or less. Then, we reconstructed the genes with the V genes, D genes, and J genes that had the highest E value. The CDR3 region is displayed if the IgBlust output text file gives us information about the CDR3 region. We filtered the sequences with stop codons and poly A tail sequences in the end region. If there were more than three matching genes, we displayed the first three genes from IgBlust.

CIBERSORT. The cell type proportions in the different samples were then estimated with cibersort31, which detects cell-type-specific signature genes using annotated single-cell data. As input, we used the number of uniquely mapped reads of each sample from featureCounts. Cibersort needs a signature matrix, so first, we used the NSCLC PBMCs Single Cell RNA-Seq signature matrix which is online. Calculating cell type proportions was performed with the default parameter.

Public data processing. We obtained an additional 115 TNBC aligned RNA-seq files from TCGA60 via the Genomic Data Commons Data Portal. The clinical data corresponding to each sample were obtained for survival analysis. The number of uniquely mapped reads of each sample was calculated using featureCounts with the same setting we used. Then, pathway analysis related to ferroptosis (hash4216) was performed on each sample compared to the remaining 109 samples except for 6 samples from normal tissues. Samples with a positive number of statistics, in which ferroptosis-related genes were expressed more than other samples, were grouped as upregulated; those with a negative number of statistics were grouped as downregulated and those that were not grouped were divided into a negative group. The upregulated and downregulated groups included 70 and 39 samples, respectively. The density of frequency of A-to-I editing events in each group was expressed through a violin plot, and the black square indicates the median value of frequency. Based on these groups, we conducted a survival analysis using the Limma R python package version 2.35.39. We used Kaplan–Meier analysis comparing the upregulated and downregulated groups, excluding those with <30 days of follow-up. Survival curves for the two groups were compared using the log-rank test (lifelines.statistics.logrank_test).

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

Data availability
The raw sequencing data files generated in this study have been deposited in the Sequence Read Archive as processed bam file under the bioproject code PRJNA779657. Readers can freely access the uploaded datasets through Sequence Read Archive. The TCGA gene expression data and patient’s metadata are available via the NIH genomic data portal (https://portal.gdc.cancer.gov) and the Molecular Signatures Database (MsigDB) (https://www.gsea-msigdb.org/gsea/msigdb/). Kyoto Encyclopedia of Gene and Genomes (KEGG) (https://www.genome.jp/kegg/), and REDPortal database (http://srv00.recas.ba.imfn.it/afaas/). For the remaining analysis, we used the NSCLC PBMCs Single Cell RNA-Seq signature matrix which is online. Calculating cell type proportions was performed using the log-rank test (lifelines.statistics.logrank_test).

Code availability
Source code is available at GitHub (https://github.com/BnNEL-SNU/Select-seq32).
References
1. Hu, Y. et al. Single-cell RNA cap and tail sequencing (scRCAT-seq) reveals subtype-specific isoforms differing in transcript demarcation. Nat. Commun. 11, 5148 (2020).
2. Mella, I. A., Gupte, R., Raj, A. & Rouhanifard, S. H. Visualizing adenosine-to-inosine RNAediting in single mammalian cells. Nat. Methods 14, 801–804 (2017).
3. Nakahama, T. & Kawahara, Y. Adenosine-to-inosine RNA editing in the immune system: friend or foe? Cell. Mol. Life Sci. 77, 2931–2948 (2020).
4. Dominissini, D., Moshitch-Moshkovitz, S., Amariglio, N. & Rechavi, G. Adenosine-to-inosine RNA editing meets cancer. Carcinogenesis 32, 1569–1577 (2011).
5. Lim, C. H. & Chen, S. C. C. The Cancer Editome Atlas: a resource for exploratory analysis of the adenosine-to-inosine RNA editome in cancer. Cancer Res. 79, 3001–3006 (2019).
6. Rodrigues, S. G. et al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. Science 363, 1463–1467 (2019).
7. Chen, W.-T. et al. Spatial transcriptomics and in situ sequencing to study Alzheimer’s disease. Cell 182, 1–16 (2020).
8. Vickovic, S. et al. High-definition spatial transcriptomics for in situ tissue profiling. Nat. Methods https://doi.org/10.1038/s41592-019-0548-y (2019).
9. Moncađa, R. et al. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. Nat. Biotechnol. https://doi.org/10.1038/s41587-019-0392-8 (2020).
10. He, B. et al. Integrating spatial gene expression and breast tumour morphology via deep learning. Nat. Biomed. Eng. https://doi.org/10.1038/s41551-020-0578-x (2020).
11. Lebrigand, K. et al. The spatial landscape of gene expression isoforms in tissue sections. bioRxiv https://doi.org/10.1101/2020.08.24.252296 (2020).
12. Joglekar, A. et al. A spatially resolved brain region- and cell type-specific isoform atlas of the postnatal mouse brain. Nat. Commun. 12, 1–16 (2021).
13. Amarasinghe, S. L. et al. Opportunities and challenges in long-read sequencing data analysis. Genome Biol. 21, 1–6 (2020).
14. Nickerwitz, S. et al. Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling. Nat. Commun. 7, 1–11 (2016).
15. Baccin, C. et al. Combined single-cell and spatial transcriptomics reveal the molecular, cellular and spatial bone marrow niche organization. Nat. Cell Biol. 22, 38–48 (2020).
16. De Cecco, L. et al. Impact of biospecimens handling on biomarker research in breast cancer. BMC Cancer 2009 9, 1–9 (2009).
17. Jiang, Q., Crews, L. A., Holm, F. & Jamieson, C. H. M. RNA editing-dependent epitranscriptome diversity in cancer stem cells. Nat. Rev. Cancer 17, 381–392 (2017).
18. Kim, S. et al. PHLI-seq: constructing and visualizing cancer genomic maps in 3D by phenotype-based high-throughput laser-aided isolation and sequencing. Genome Biol. 19, 158 (2018).
19. Kim, O. et al. Whole genome sequencing of single circulating tumor cells isolated by applying a pulsed laser to cell-capturing microstructures. Small https://doi.org/10.1002/smll.201902607 (2019).
20. Thomsen, E. R. et al. Fixed single-cell transcriptomic characterization of human radial glial diversity. Nat. Methods 13, 87–93 (2015).
21. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. 9, 171–181 (2014).
22. Lehmann, B. D. et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J. Clin. Invest. 121, 2750–2767 (2011).
23. Zhang, K. et al. The collagen receptor discoidin domain receptor 2 stabilizes SNAIL1 to facilitate breast cancer metastasis. Nat. Cell Biol. 15, 677–687 (2013).
24. Wang, C. C. et al. CD164 regulates proliferation, progression, and invasion of human glioblastoma cells. Oncotarget 10, 2041–2054 (2019).
25. Abraham, B. K. et al. Prevalence of CD44+/CD24−/low cells in breast cancer may not be associated with clinical outcome but may favor distant metastasis. Clin. Cancer Res. 11, 1154–1159 (2005).
26. Miao, Q. et al. SOX11 and SOX4 drive the reactivation of an embryonic gene program during murine wounding. Nat. Commun. 10, 1–20 (2019).
27. Gupta, P. B. et al. Identification of selective inhibitors of cancer stem cells by high-throughput screening. Cell 138, 645–659 (2009).
28. Kassambura, A. et al. GenomicScape: an easy-to-use web tool for gene expression data analysis. Application to investigate the molecular events in the differentiaton of B cells into plasma cells. PLoS Comput. Biol. 11, e1004077 (2015).
61. Wulczyn, E. et al. Deep learning-based survival prediction for multiple cancer types using histopathology images. *PLoS One* 15, e0233678 (2020).
62. Lee A. C. et al. Spatial epitranscriptomics reveals A-to-I editome specific to cancer stem cell microniches. Select-seq. Zenodo https://doi.org/10.5281/zenodo.6409223 (2022).

**Acknowledgements**

This research was supported by the Ministry of Science and ICT (MSIT) of the Republic of Korea and the National Research Foundation of Korea (2020M3H1A1073304 to B.K., NRF-2020R1A3B3079653 to S.K.), the Ministry of Education of the Republic of Korea (2021R1I1A1A01045372 to A.L.), and the Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (2022R1A2C2007561 to W.H.).

**Author contributions**

W.H. and S.K. conceived of the idea and supervised the work. A.C.L., Y.L., A.C., and H.L. developed the Select-seq method. A.C.L. developed the automated target isolation program. K.S. developed the B/T-cell receptor extraction pipeline. H.L. developed the A-to-I editing site calling pipeline. H.C. and W.L. developed the RNA-seq preprocessing and analysis pipeline. D.Y. performed the B/T-cell receptor validation experiment. A.C. and S.L. developed the gene ontology analysis pipeline. H.L., T.R., O.K., Y.J., S.S., Y.C., W.H., and S.K. acquired the funds and provide the reagents. C.K. performed RNA-seq related experiments. A.C.L., Y.L., and A.C. performed experiments with assistance from K.S. and W.L. A.C.L., Y.L., A.C., and H.L. wrote the initial manuscript. S.B., T.R., S.S., and Y.C. reviewed and edited the manuscript. J.C., B.K., W.H., and S.K. supervised this work.

**Competing interests**

A.C.L., Y.L., O.K., and S.K. are listed as inventors on patents related to the work applied by the Seoul National University covering the technology (Methods for selectively separating samples from substrate, US 15/770, 765). H.B.L. and W.H. report being a member on the board of directors of and holding stock and ownership interests at DCGen, Co., Ltd., not relevant to this study. The remaining authors declare no competing interests.

**Additional information**

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-30299-3.

**Correspondence** and requests for materials should be addressed to Wonshik Han or Sunghoon Kwon.

**Peer review information** *Nature Communications* thanks Masayuki Sakurai and the other anonymous reviewer(s) for their contribution to the peer review of this work.

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

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

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

© The Author(s) 2022