METHODS AND RESOURCES

Histologically resolved multiomics enables precise molecular profiling of human intratumor heterogeneity

Tao Chen[1,2☯¤a], Chen Cao[1☯¤b], Jianyun Zhang[3,4☯], Aaron Streets[1¤c]*, Tiejun Li[3,4,5]*, Yanyi Huang[1,2,6,7,8,9]*

1 Biomedical Pioneering Innovation Center (BIOPIC), School of Life Sciences, Peking University, Beijing, China, **2** College of Engineering, Peking University, Beijing, China, **3** Department of Oral Pathology, Peking University School and Hospital of Stomatology & National Center of Stomatology & National Clinical Research Center for Oral Diseases & National Engineering Research Center of Oral Biomaterials and Digital Medical Devices, Beijing, China, **4** Beijing Key Laboratory of Digital Stomatology, **5** Research Unit of Precision Pathologic Diagnosis in Tumors of the Oral and Maxillofacial Regions, Chinese Academy of Medical Sciences (2019RU034), Beijing, China, **6** Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China, **7** Beijing Advanced Innovation Center for Genomics (ICG), Peking University, Beijing, China, **8** College of Chemistry and Molecular Engineering, Peking University, Beijing, China, **9** Institute for Cell Analysis, Shenzhen Bay Laboratory, Guangdong, China
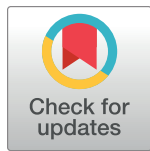
☯ These authors contributed equally to this work.
¤a Current address: Huffington Center on Aging, Baylor College of Medicine, Houston, Texas, United States of America
¤b Current address: Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey, United States of America
¤c Current address: Department of Bioengineering, University of California, Berkeley, Berkeley, California, United States of America
* astreets@berkeley.edu (AS); litiejun22@vip.sina.com (TL); yanyi@pku.edu.cn (YH)

Abstract

Both the composition of cell types and their spatial distribution in a tissue play a critical role in cellular function, organ development, and disease progression. For example, intratumor heterogeneity and the distribution of transcriptional and genetic events in single cells drive the genesis and development of cancer. However, it can be challenging to fully characterize the molecular profile of cells in a tissue with high spatial resolution because microscopy has limited ability to extract comprehensive genomic information, and the spatial resolution of genomic techniques tends to be limited by dissection. There is a growing need for tools that can be used to explore the relationship between histological features, gene expression patterns, and spatially correlated genomic alterations in healthy and diseased tissue samples. Here, we present a technique that combines label-free histology with spatially resolved multiomics in unfixed and unstained tissue sections. This approach leverages stimulated Raman scattering microscopy to provide chemical contrast that reveals histological tissue architecture, allowing for high-resolution in situ laser microdissection of regions of interests. These microtissue samples are then processed for DNA and RNA sequencing to identify unique genetic profiles that correspond to distinct anatomical regions. We demonstrate the capabilities of this technique by mapping gene expression and copy number alterations to histologically defined regions in human oral squamous cell carcinoma (OSCC). Our approach provides complementary insights in tumorigenesis and offers an integrative tool for macroscale cancer tissues with spatial multiomics assessments.

Competing interests: The authors have declared
that no competing interests exist.

Abbreviations: CNA, copy number alteration;
COSMIC, Catalogue of Somatic Mutations in
Cancer; CV, coefficient of variation; DOP-PCR,
degenerate oligonucleotide primed PCR; EMT,
epithelial to mesenchymal transition; EOM, electro-
optical modulator; HE, hematoxylin–eosin; HNSCC,
head and neck squamous cell carcinoma; HOG,
histogram of orientation gradient; LCM, laser
capture microdissection; OPO, optical parametric
oscillator; OSCC, oral squamous cell carcinoma;
PCA, principal component analysis; PEN,
polyethylene naphthalate; PLR, protein-to-lipid
ratio; ROI, regions of interest; SMD-seq, SRS
microdissection and sequencing; SRS, stimulated
Raman scattering; TCGA, the Cancer Genome
Atlas.

Introduction

Recent technological developments for low-sample input genomic analysis have led to high-resolution characterization of cellular heterogeneity in tissues and organs. Specifically whole-genome and whole-transcriptome amplification have enabled comprehensive molecular profiling of single cells [1,2]. Most techniques for single-cell genomic analysis require the dissociation of millimeter-scale biological samples into a single-cell suspension, before the extraction of genomic material, limiting the spatial resolution of these techniques to the size of the dissected tissue sections [3,4]. While high-throughput single-cell RNA sequencing has been a powerful tool for classifying cell types and revealing cell states, cellular state and function are strongly influenced by the cell's microenvironment, including chemical and physical context as well as the composition of neighboring cells [5,6]. Therefore, comprehensive cellular characterization of biological systems would ideally leverage the combination of genome-wide molecular analysis and high spatial resolution. Laser capture microdissection (LCM) [7–9] enables the recovery of submillimeter regions of interest (ROIs) from tissue, allowing for microscopy to guide the selection of cells based on the morphology of the microenvironment. In this way, LCM links histology with molecular measurements, including genome and transcriptome-wide profiling with high-throughput sequencing. However, LCM typically requires fixation and staining to generate histological contrast, and these steps can perturb or even degrade the nucleic acid composition of cells [10,11]. Furthermore, the contrast used to guide dissection in these applications is limited by the chosen staining method [12].

Here, we present an integrative approach for mapping molecular and cellular heterogeneity in tissue samples that leverages stimulated Raman scattering (SRS) microscopy [13] to generate histological images of fresh frozen tissue slices followed by in situ microdissection and multiomic sequencing analyses. This technique, called SRS microdissection and sequencing (SMD-seq), images tissue sections without staining, and instead provides contrast based on the intrinsic chemical composition of the biological specimen. SRS generates signal from molecular vibrations that are specific to chemical bond composition. Imaging a biological sample at multiple Raman frequencies can therefore provide composite images with morphological contrast that corresponds to chemical variation [14–17]. After label-free imaging of fresh or frozen tissue slices, ROIs are dissected directly with the SRS excitation laser, and immediately processed enabling efficient recovery of RNA and DNA from small numbers of cells within an ROI (Fig 1A, Materials and methods). The combination of quantitative chemical imaging and high-resolution genomic analysis facilitates multimodal characterization of the cellular heterogeneity in complex tissues.

The ability to map molecular and cellular heterogeneity is particularly important in cancer. Heterogeneity in cancer is expressed in many facets, including the evolving genetic changes in distinct subpopulations [18,19], and the progressive abnormalities in morphology of the cells, typically identified by histopathology [20]. Molecular and phenotypic aberrant variations are not only common between tumors of different tissue and cell types, but also within a tumor derived from the same patient or even from the same tissue [21]. The clonal diversity during tumorigenesis, in the context of histology and genetic divergence, influences each individual tumor's diagnosis and response to treatment [22]. It is therefore important to characterize the tumor heterogeneity in a comprehensive and precise way.

In this study, we demonstrate the utility of SMD-seq by spatially mapping genomic and transcriptomic profiles in human oral squamous cell carcinoma (OSCC). OSCC is a major subtype of head and neck squamous cell carcinoma (HNSCC). Its aetiological and biological heterogeneity are influenced by distinct risk factors [23,24], which further lead to high morbidity and relatively low overall 5-year survival rate. OSCC is highly invasive, leading to large
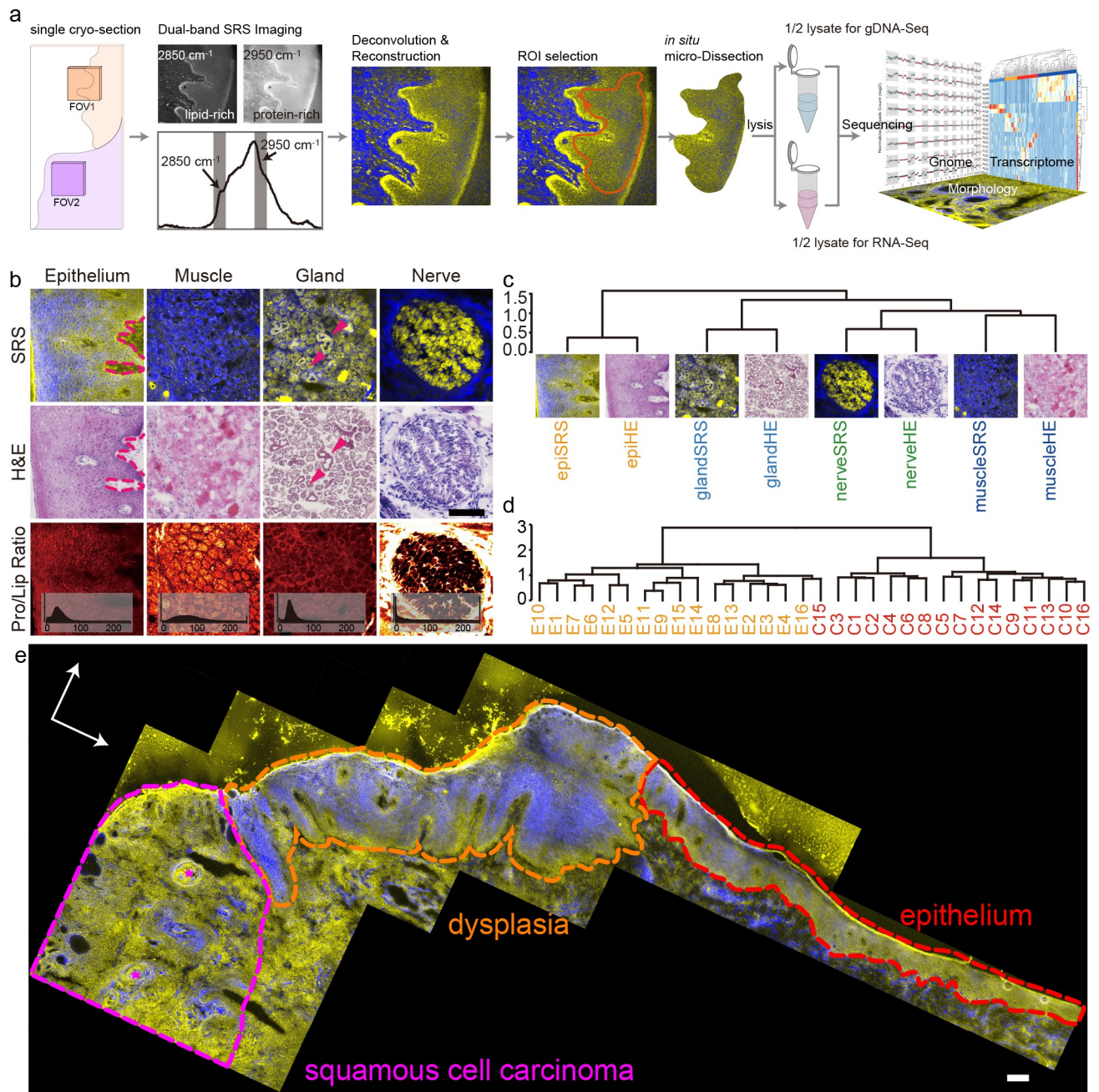
Fig 1. Principle of SMD-seq and evaluation of SRS imaging. (a) The pipeline of SMD-seq, inset of FOV2 represents the complexity around single cancer nest. (b) The comparison between SRS and HE images of different tissues, with the PLR images of SRS. (c) Unsupervised clustering of HOG features of SRS and HE images from different tissue types. (d) Unsupervised clustering of SRS images of epithelium and cancer. (e) Stitched SRS image of a whole sample, showing revealed morphologies including cancer, dysplasia, and epithelium. The magenta stars indicate the keratin pearls. Scale bars are 200 μm. HE, hematoxylin–eosin; HOG, histogram of orientation gradient; PLR, protein-to-lipid ratio; SMD-seq, SRS microdissection and sequencing; SRS, stimulated Raman scattering.

variation in cellular phenotypes over small length scales [25,26]. OSCC cancer nests, for example, measuring as small as 100 μm in diameter, are often interlaced with normal tissues composed of noncancerous cells, and it is challenging to dissect this heterogeneity with existing methods [25]. Additionally, mutation patterns in OSCC have been shown to differ from other

subsites of HNSCC due to its specific anatomical location and tissue characteristics [27]. We used SRS histology to identify progressive abnormalities in the morphology of OSCC cancer cells, and in situ laser microdissection followed by DNA and RNA sequencing identified unique molecular profiles that correspond to distinct anatomical regions of tissue samples from multiple individuals. Transcriptomic analysis was used to identify gene expression profiles that corresponded to specific tissue types and cancer cells, which corroborated histological analysis. Furthermore, transcriptomic and genomic analysis revealed gene fusion events and copy number variations that delineated cancer progression within tissue sections. This combination of imaging and multiomic analysis demonstrates how SMD-seq can provide a powerful tool to accurately investigate complex inter- and intratumor heterogeneity and has the potential to be scalable to any architecturally complex tissue.

Results

Label-free SRS microscopy accurately revealed histological features of OSCC

The recent and rapid development of SRS-based histology [14–17] has proven this technique to be a powerful supplement to traditional histological staining for many tissues. We first examined the ability of SRS imaging to reconstruct histological features in unstained cryo-sections of OSCC that corroborated information obtained with hematoxylin–eosin (HE) staining of OSCC sections. We generated 2-color SRS composite images of 30 μm cryo-sections biopsied from various tissues that provide contrast between lipid-rich and protein-rich structures by exciting CH_2 symmetric vibration and CH_3 vibrations, respectively (Materials and methods, S1 Fig). To validate the efficacy of SRS-based histology for OSCC, we then compared these reconstructed images to adjacent 5 μm cryo-sections that were sampled from the same biopsies and conventionally stained by HE. The lipid–protein contrast in the SRS images recovered the characteristic morphological features of each tissue, including the variation of cell shape and base membrane in the epithelium, the epithelial cells forming the duct wall in the gland tissue [28] and the nerve bundles (Fig 1B). In addition to recapitulating morphological information, SRS images provide information about the chemical composition of the tissue samples. For example, both the muscles and the perineurium, a collagen-rich structure [29,30], around nerve revealed high protein content in SRS images (Fig 1B). Variability in the chemical composition of ROIs can be further evaluated by the histogram of the protein-to-lipid ratio (PLR) from different tissue types, reflecting a distinct molecular signature of the cells within the tissue. In order to confirm that the histopathological utility of SRS was not limited to OSCC, we also identified characteristic features of 2 other oral cavity diseases, the Warthin's tumor and the mucoepidermoid carcinoma (S2 Fig). These results demonstrate the generalizability of SRS imaging for oral cavity histology.

We then quantified the similarity between morphological information retained in SRS and HE images using the histogram of orientation gradient (HOG) [31]. The HOG describes the texture of an image based on intensity variation, which primarily reflects cellular packing in our images. Unsupervised clustering of the HOG from all 8 images presented in Fig 1B, successfully clustered images by tissue type, regardless of imaging modality (Fig 1C). This analysis indicates that SRS images recovered similar cellular organization patterns as HE images. In addition to distinguishing different types of tissue, the ability to distinguish cancerous epithelium from normal epithelium has potential clinical implications. The cellular organization patterns between these tissue states are distinct and can be decoded by HOG features. Unsupervised clustering of the HOG features of 32 SRS images, including 16 cancerous and 16 normal epithelia (S3 Fig), stratified tissue images into 2 groups, accurately reflecting disease

state with only 1 misidentification (Fig 1D). Furthermore, the correlation matrix of all 32 samples' HOG features revealed a weaker correlation among cancers than epithelia (S4 Fig), implying greater variation in cellular organization patterns between individual cancer nests. This heterogeneity could be related to the distinct microenvironment of the cancer nests. For example, the cancer nests in P4 invaded heavily into muscle tissue, while in other patient samples the cancer nests were typically located in connective tissue (S5 Fig). We examined the morphology restoration at different scales on stitched image of a whole cryo-section (9×4 mm^2). Comparing with corresponding HE images, SRS revealed identical features at all scales. Selected ROIs from SRS images can be localized in HE images through correlation of their respective HOG (Figs 1E and S6). The key morphological features were identified in the large-scale SRS image depicted in Fig 1E, including normal epithelium, dysplasia, cancer, and those keratin pearls correlated with OSCC differentiation.

SRS guided microdissection enabled high-quality mRNA and DNA sequencing of cancer nests

In the SMD-seq workflow, after an ROI is identified with SRS histology, that region is immediately excised and recovered from the bulk tissue section for downstream genomic analysis. This is achieved by increasing the power of the pump line of the SRS excitation source and scanning the focal spot around a user-defined ROI boundary (Materials and methods). To validate the accuracy of ROI dissection, we performed HE staining on the remaining tissue section after dissection, as well as adjacent tissue sections (Fig 2A). Because microdissection is performed with the scanning laser, the ROI is defined with high resolution (S7A Fig) and can be recovered immediately after imaging. Depending on the ROI size and tissue type, the dissected sample typically contained about 230 cells on average (S7B Fig). SRS imaging and dissection can be performed in rapid succession, allowing the dissection of multiple ROIs from a single tissue section (Fig 2B), therefore when profiling cancer regions, all the cancer microtissues were paired with a normal epithelial microtissue from the same tissue section to control for intersample variability.

After recovery, each microsample was lysed and divided into 2 equal aliquots for parallel DNA and mRNA extraction. One of the primary advantages of SMD-seq is the efficient recovery of high-quality mRNA, which is made possible because SRS histology avoids fixation, staining, or any other chemical perturbations that may degrade RNA or inhibit recovery. Quantification of housekeeping genes showed that unstained cryo-sections can preserve over 20-fold more mRNA than HE-stained sections (S8 Fig). Another advantage of SMD-seq is that the implementation of whole-genome and whole-transcriptome amplification enables multiomic analysis from a single ROI. Finally, SRS histology and LCM enables molecular profiling on subregions of a tissue section so that cellular heterogeneity can be reduced and ensemble measurement averages over a smaller ROI. Comparing with HNSCC samples collected for the Cancer Genome Atlas, (TCGA), the microsamples measured with SMD-seq presented significantly lower cell-type diversity (*P* value = 0.0098, S7C–S7E Fig, Materials and methods). The roughly 100-μm resolution ROI subselection in SMD-seq allows for profiling of intrasection heterogeneity; however, the lack of single-cell resolution within the microsamples means that some heterogeneity is still obscured.

We applied SMD-seq to 13 cryo-sections from 4 patients (3 males, 1 female) who suffered various stages of OSCC at different ages (Materials and methods). We collected 28 in situ microdissection samples for sequencing, and 27 samples passed the quality control based on mRNA recovery, for further analysis (S1 to S3 Tables, Materials and methods). These microsamples included 12 normal (5 muscle, 2 gland, and 5 epithelium) and 9 cancer samples for
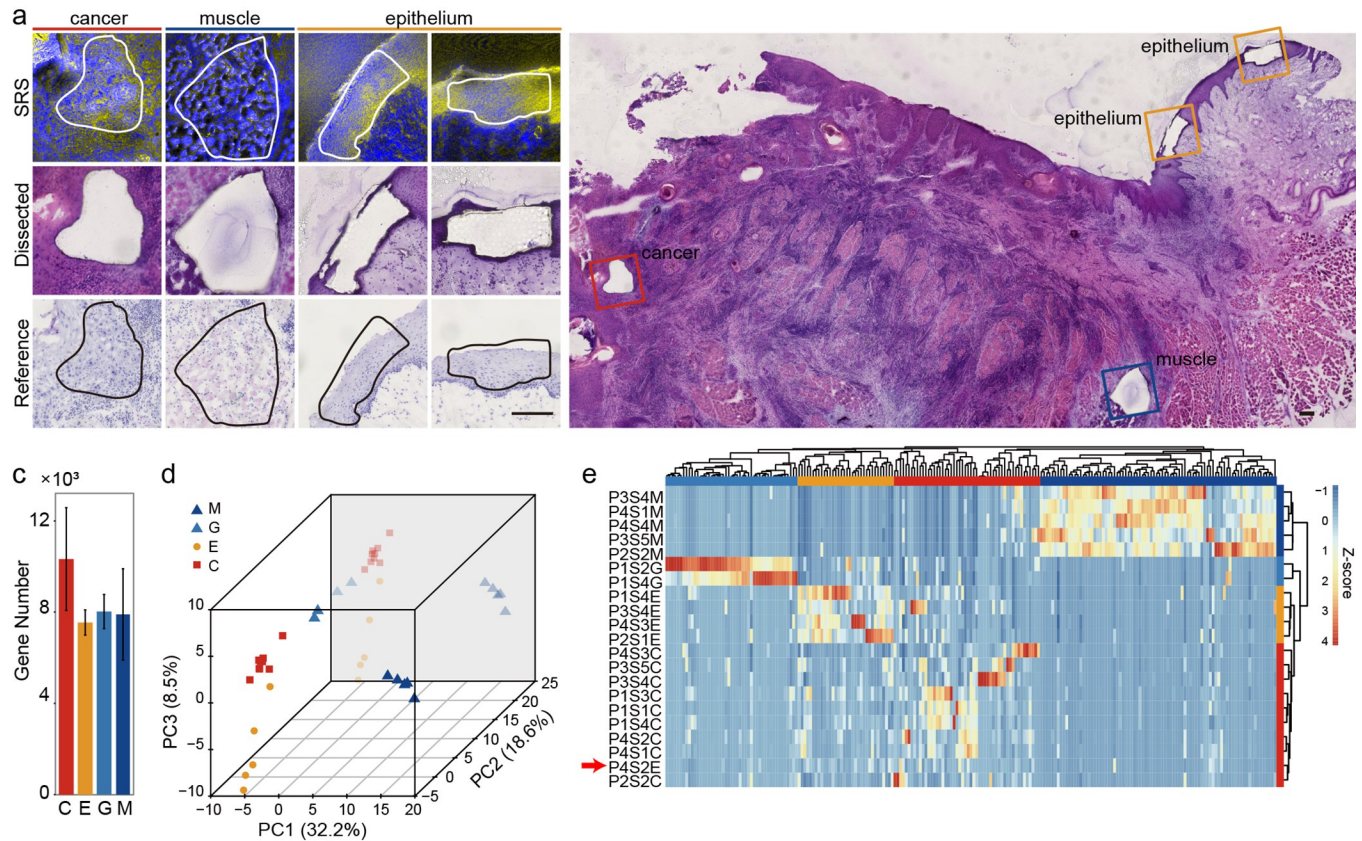
Fig 2. SRS image guided microdissection and evaluation of mRNA recovery. (a) Comparison between SRS images, HE-stained and unstained microtissue section. (b) Epithelium, muscle, and cancer from the same section. (c) Statistics of gene number per sample. (d) PCA plot of all expressed genes. (e) Unsupervised clustering of 217 differently expressed genes from microdissected samples. Scale bars are 200 μm. Source data for panels c–e in S2 Data Sheet 2.1–2.3. HE, hematoxylin–eosin; PCA, principal component analysis; SRS, stimulated Raman scattering.

RNA-seq and 13 normal (muscle, $n = 5$; gland, $n = 2$; and epithelium, $n = 6$) and 8 cancer samples for DNA-seq. RNA-seq recovered around 9,000 genes per microsample on average (FPKM > 0.1, Fig 2C) and highly correlated expression among samples of the same tissue, indicating technical reproducibility (S9 Fig).

Principal component analysis (PCA) separated microsamples into 4 groups of similar gene expression profiles that recapitulated the known tissue types (Fig 2D). Unsupervised hierarchical clustering of microsamples using the top 217 differently expressed genes (Materials and methods) grouped these clusters as distinct tissue types, 3 of which were then annotated using a panel of known marker genes for epithelium, gland, and muscle in agreement with the corresponding tissue types identified by SRS histology (Figs 2D and S10). The fourth group, which was consistently identified as cancer by SRS images, contained highly expressed OSCC marker genes including GSTP1, AKR1B10, FTH1, and FTL (S11 Fig). GSTP1 expression is known to be associated with high malignancy and poor survival rate [32,33]. Expression levels of GSTP1 were further validated with immunofluorescence staining, confirming the confinement of GSTP1 expression to cancer nests (S12 Fig). While the epithelial to mesenchymal transition (EMTs) related gene, such as KRT13 (Fig 3A) [34], was expressed at lower levels, indicating an enhanced migratory capacity and invasiveness [35]. While these OSCC samples originated from the epithelium, they revealed a distinct expression profile compared to the healthy epithelium samples. Unexpectedly, P4S2E, which was identified as epithelium by SRS imaging,
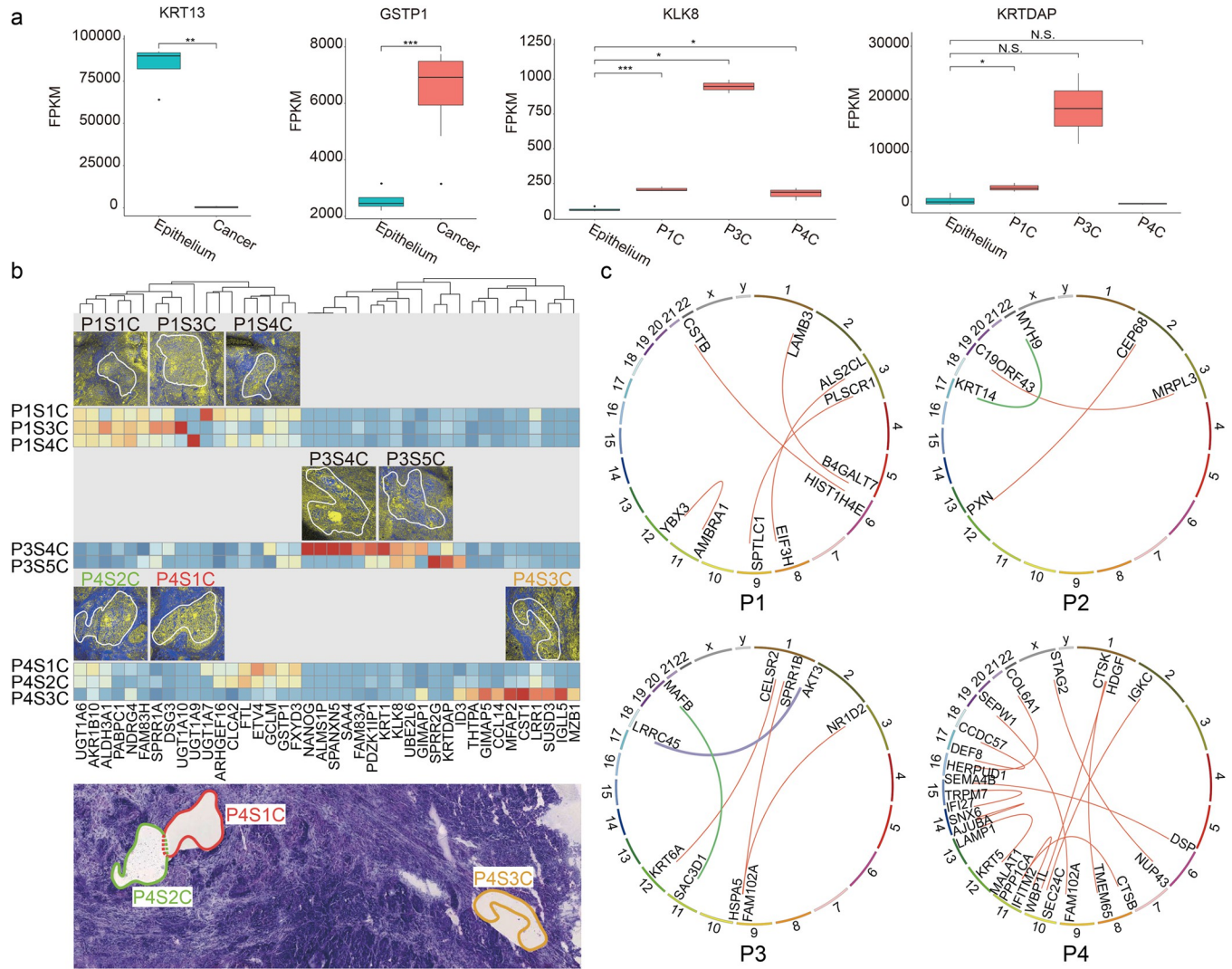
Fig 3. Heterogeneity of gene expression. (a) Expression levels of genes in cancer samples among different patients, in comparison with epithelium. (b) The interpatient and intrapatient cancer heterogeneity. The bottom figure is a projection of 3 slides, showing the relative positions of different cancer nests. (c) Interpatient heterogeneity of gene fusion events. Source data for panel a in S2 Data Sheet 2.4.

https://doi.org/10.1371/journal.pbio.3001699.g003

clustered with the cancer samples based on gene expression profiles (Fig 2E, red arrow). PCA also indicated that P4S2E was more similar to the cancer samples than the epithelium samples (Fig 2D). This observation indicates that microsample P4S2E may have come from a region of the biopsy in which these epithelial cells were developing into carcinoma. This level of molecular characterization is enabled by joint histological profiling with SRS and gene expression analysis.

SMD-seq profiles molecular heterogeneity in OSCC

SMD-seq provides the ability to compare the genome and transcriptome of specific tissue types and disease states within a single tissue section, allowing for a more precise deconvolution of genetic profiles and gene expression profiles than traditional bulk assays. While healthy tissue showed consistent gene expression across samples, OSCC microsamples revealed large patient-to-patient variability in gene expression. KLK8 and KRTDAP only exhibited high

expression in patient P3 (Fig 3A). KLK8 is implicated in malignant progression of OSCC [36], and KRTDAP strongly correlate with the differentiation and maintenance of stratified epithelium. Correspondingly, "keratin pearls," structures indicating high-degree differentiation were present in SRS images of P3 (S13 Fig) [37]. In addition to variability between patients, we found significant variability in gene expression between samples identified as cancer nests from the same patient. Specifically, in P4, the adjacent P4S1C and P4S2C microsamples exhibited similar gene expression; however, P4S3C, which was recovered from a more distant location to previous 2 (S14 Fig), revealed a distinct gene expression profile including expression of CST1, which is related to promoted cell proliferation (Fig 3B). The intratumor molecular variation may indicate a branched evolution of tumorigenesis that could provide insight for tumor diagnosis and treatment.

In order to characterize the cellular heterogeneity within microsamples, we aimed to deconvolve their composition using a published single-cell dataset that profiled transcriptomes of 6,000 single cells from 18 HNSCC patients [38]. We first embedded the RNA-seq data from our tumor microsamples into 20 principal components that capture gene expression variation in the single-cell data from Puram and colleagues (S15A Fig). Most of the cancer microsamples colocalized with the malignant epithelium cluster in the single-cell data, indicating similar gene expression profiles between our sections and individual malignant cells. However, 1 sample colocalized with the fibroblast and endothelial cell clusters. We investigated the cellular composition of these samples, which collectively contribute to their gene expression profile, by identifying the top 1,000 nearest neighbor cells from the Puram data. Then, for each sample, we calculated the percent of these top 1,000 cells that fell into each of the annotated cell clusters (S15B Fig). Five of our microsamples were composed primarily of malignant cells; however, the remaining 3 showed a more diverse composition including tumor resident fibroblasts and T cells. Interestingly, most of the nearest neighbor cells to sample P4S3C were indeed fibroblasts and endothelial cells.

Displacement and recombination of genes, especially oncogenes, has been recognized to drive neoplasia [39] and has become the focus of numerous cancer studies as potential therapeutic targets [40–42]. We then exploited de novo identification of gene fusion events through sequencing these samples. Early fusion events are usually sporadic and hence hard to identify with bulk sequencing. Furthermore, DNA sequencing of small samples with low input heavily relies on whole-genome amplification, which is prone to chimera formation and causes false identification of fusion events. We used RNA-seq of microsamples for de novo identification of gene fusion transcripts to investigate gene fusion events between patients and within tumor samples. Some fusion-related genes were shared between microsamples from separate patients, including KRT6A, which was shared between P1S2C and P3S5C, and FAM102A, which was shared between P1S3C, P3S5C, and P4S2C. Some samples, however, contained unique gene fusion patterns (Fig 3C, S1 Data 1.3 to 1.8). Most of these fusions might be passenger events that came along with cancer development and thus their actual consequences remain unknown. Among those, we found several sets that are consistent with previous observations, including 1 recorded in TCGA (MYH9 and KRT14, from P2) and 2 involving oncogenes AKT3 (fused with LRRC45, from P3) and MAFB (fused with SAC3D1, from P3) (S16 Fig, S2 Table, S3 Data). Sanger sequencing on amplified fragments harboring joint junction of fused genes confirmed these fusions between MYH9 (5′ fusion partner: exon 20) and KRT14 (3′ fusion partner: exon 8), AKT3 (5′ fusion partner: UTR), and LRRC45 (3′ fusion partner: intron), implying the capability of SMD-seq in recurrent fusion event discovery. Fusion events also revealed substantial intrasample heterogeneity (S17 Fig). For example, more fused genes were found in P4S3C, than in P4S1C and P4S2C which were dissected from distant regions of the tissue section, indicating spatially dependent genome instability during tumorigenesis.
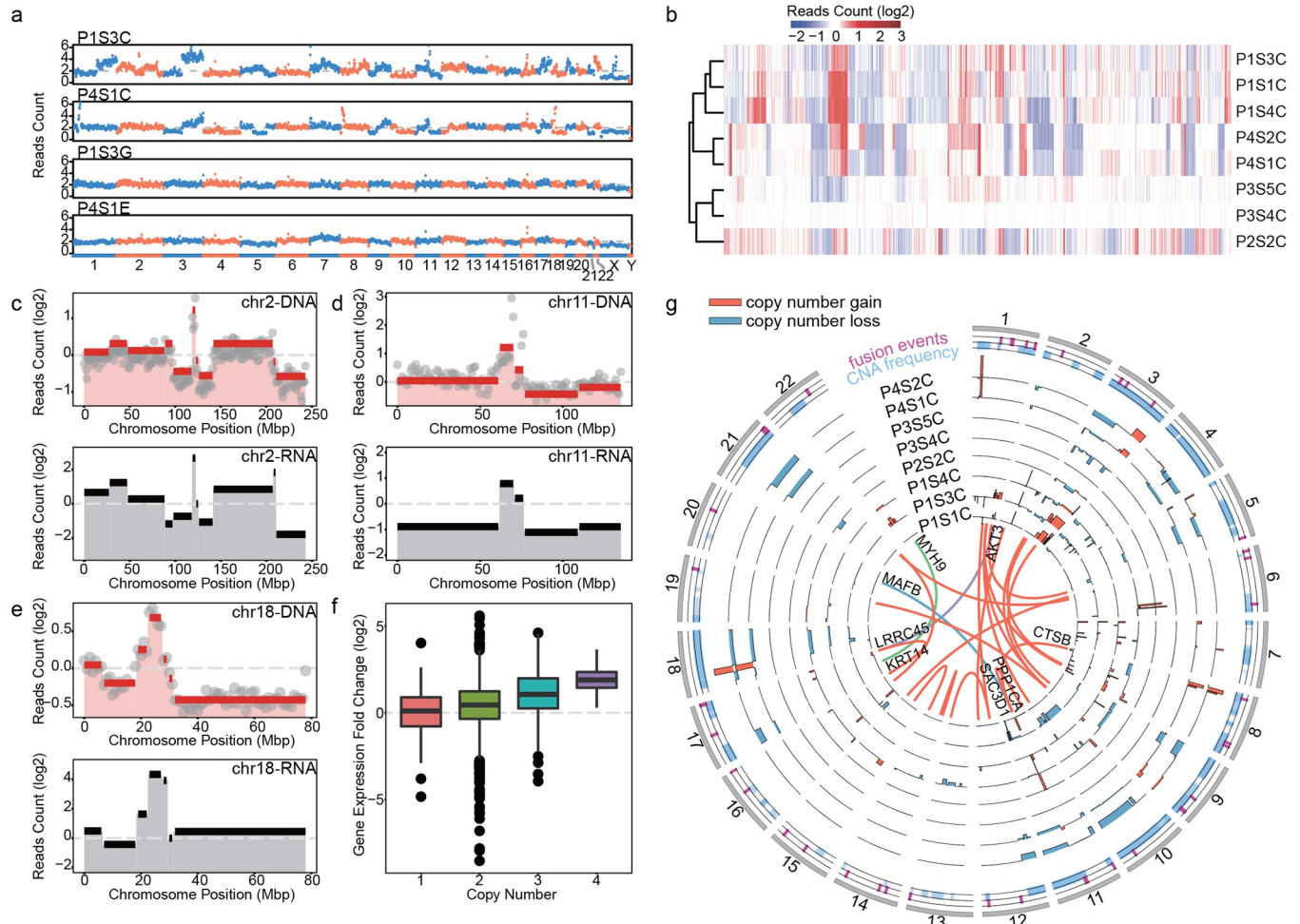
Fig 4. Genomic and transcriptomic analysis of laser dissected tissue samples. (a) Raw reads count across the whole genome of 2 paired cancer-normal tissues dissected from the same slice. Chromosome numbers were labeled at the bottom. (b) Unsupervised clustering of normalized reads counts of all the dissected cancer samples. (c, d) Top panels demonstrated the normalized read count (gray dots) and copy numbers (red lines) identified by CBS algorithm in chromosome 2 and 11 of sample P1S3C and P1S1C, respectively. Mean expression levels within the same segment were shown in black lines in the bottom panel. (e) Averaged read counts of all the cancer samples (top) and their corresponding gene expression values (bottom) of chromosome 18. Genome and transcriptome variations of other chromosomes and samples were shown in S16, S18, and S19 Figs. (f) Averaged gene expression fold changes were computed per 1 M bin across the whole genome and plotted against copy numbers. (g) The distribution of CNA and fusion genes across the genome of all the cancer samples. Orange and blue bars indicated the copy number gains and loss, respectively. Orange lines in the inner circle indicated fusion genes with at least 10 span pair reads, and the green and purple lines represented oncogenes involved gene fusions. The position of fusion events (magenta) and CNAs (light blue) of all the cancer samples were shown in the 2 outmost circles. Source data for panel b–g in S2 Data Sheet 2.5–2.20. CNA, copy number alteration.

https://doi.org/10.1371/journal.pbio.3001699.g004

Moreover, we also observed the existence of an oncogene-involved fusion (RAB3D and MTMR14, verified by Sanger sequencing, S3 Data) in sample P4S2E (S18 Fig), which was identified as healthy epithelium by HE and SRS analyses, again indicating the possibility that this microsample was progressing through the early stages of tumorigenesis.

In parallel to RNA-seq, DNA-seq was performed using half of the lysate from microsamples, enabling the analysis of heterogeneity at the whole-genome level. Copy number alteration (CNA) analysis demonstrated unique patterns of CNAs between cancer and normal samples, and between patients (Figs 4A and S19), indicating the high complexity in OSCC including significant genetic mosaicism and genetic heterogeneity. We subsampled reads to characterize library complexity and read depth, demonstrating the 0.1× was sufficient for accurate and

reproducible CNA (S20 Fig). A few commonly shared large-size ploidy shifts, such as the losses of 3p and 8p and the gains of 3q and 8q, which are shared in other squamous cell carcinomas [43,44], were also observed in our OSCC samples. This analysis also revealed patient-specific CNAs, for example, chromosome 6 showed high instability in P2's cancer sample but this instability was not present in others (S21 and S22 Figs). Unsupervised clustering of the samples based on CNAs further demonstrated the unique CNA patterns between patients (Figs 4B and S21) [45]. Microsamples dissected from different locations within a patient sample displayed subtle discrepancies in copy number profiles (Fig 4B). The CNA pattern of chromosome 1 in P1S1C and P1S3C were different from that of P1S4C, with an obvious gain of 1q in the first 2 ROIs (S22 Fig).

As both genomes and transcriptomes were sequenced from each cancer nest, we were able to perform joint analysis of genomic and transcriptomic variation. The mean expression level of genes within each genomic segment [46] was compared with the copy number in the same regions (Figs 4C–4E, S22, and S23) and summarized across the whole genome (Figs 4F and S24). The average gene expression levels showed positive correlation with the copy number within the same segment (Fig 4F). Among all the OSCC samples, 8 regions of recurrent copy number gain and 5 regions of recurrent copy number loss were identified (q < 0.25, S25 Fig, S1 Data 1.9 to 1.10) [43]. Among these (copy number loss/gain or regions), 11q13.3, 8q24.3, 11p15.4, and 11q24.2 colocalized with differently expressed genes in cancer samples. GSTP1 located within the recurrent focal amplification of 11q13.3 [44,47], which implied that the high expression level of GSTP1 may be the result of increased copy number. FAM83H was also colocalized with a focal amplification region, 8q24.3, and it specifically expressed at higher levels in patient P1. TP53AIP1 and PKP3 both expressed at a lower level in all the patients and located in the regions of recurrent copy number loss 11q24.2 and 11p15.4, respectively. The expression level of GSTP1, FAM83H, TP53AIP1, and PKP3 were all reported to be involved in the development of cancer or affecting patients' survival rate [44,48]. We also analyzed the global correlation between gene fusion events and copy number variations and found a high degree of overlap between them (Fig 4G). Fusion genes CTSB and PPP1CA colocalized with focal amplification regions 8p23.1 and 11q13.3, separately. CTSB was proved to be related to cancer progression and metastasis [49,50], and PPP1CA was reported to contribute to ras/p53-induced senescence [51]. Of the 24 pairs of detected fused genes, 17 pairs (approximately 71%) had at least 1 gene intersected with a focal CNA (S1 Data 1.11). The parallel observation of genomic rearrangement and gene expression fold change may illustrate that the instability of the cancer genome led to gene fusion events that were more likely to occur within amplification and deletion regions (Fig 4G) [52].

Discussion

SMD-seq takes advantage of SRS microscopy, which can capture chemical information without staining, and low-input sequencing to give high-quality genome and gene expression data from well-defined histological regions. This technique enabled dissection of cancer heterogeneity across multiple measurement modalities, including morphology, genome alteration, gene expression, and gene fusion.

Obtaining quantitative molecular information about cell populations in their native environment is challenging with traditional microscopy or genomic analysis alone. The major challenge hindering the progress of these studies is technical: although sequencing analyses, especially single-cell sequencing enables the systematic identification of cell populations in a cancer tissue, most of such studies start with a dissociated cell suspension [53], which retains no spatial information from the sample of origin. On the other hand, current methods in

surgical pathology lack the capability to efficiently isolate specific cell populations in complex tissues/tumors, which can confound molecular results. Spatially resolved transcriptomic profiling or "spatial transcriptomics" refers to a rapidly growing suite of tools that directly address the challenge of high-resolution spatial mapping of molecular heterogeneity [54,55]. These methods fall into 3 broad categories: (1) single-molecule FISH-based methods, which include seqFISH [56] and MERFISH [57]; (2) spatial barcoding and sequencing-based methods that include the 10× Visium platform based on Spatial Transcriptomics [58], DBiT-seq [59], and Slide-seq [60]; and (3) LCM-based approaches that includes our SMD-seq method. Moses and Pachter recently reviewed the spatial transcriptomics literature and created a comprehensive database that catalogues hundreds of methods and applications, annotating their unique advantages and disadvantages [55]. smFISH-based methods tend to have the highest spatial resolution, typically limited by diffraction. Spatial barcoding and sequencing methods are powerful for transcriptome-wide profiling of large tissue sections but with reduced spatial resolution compared to smFISH methods. Both of these types of methods typically rely on a combination of fixation and staining to combine transcriptomic analysis with histological features. SMD-seq provides a complimentary approach to these powerful systems, enabling multiomic molecular analysis of targeted ROIs without fixation or staining. S4 Table compares the unique advantages of a representative subset of spatial transcriptomics methods; reference [55] provides a much more comprehensive overview.

The morphological information recovered by SMD-seq is reconstructed from chemical contrast, unlike gene expression in other spatial transcriptomic approaches [54,55]. Therefore, recovery of histological features does not rely on data reconstruction through gene expression patterns or spatial barcoding. Such independence between morphology and gene expression can lead to previously unobserved characteristics of tumorigenesis. The discordance between the morphology and genomic profile of microsample P4S2E served an example. The sample was identified as normal epithelium by SRS histology, HE staining of an adjacent tissue section. However, the gene expression profile of P4S2E clustered more closely with the cancer samples (Fig 2G), and both the genetic and transcriptomic features of this sample reflected a cancer-like pattern. Furthermore, genes previously reported to be significantly mutated in OSCC, such as FAT1, PPP2R1A, PTEN, HRAS, and CREBBP28, are also found in P4S2E (S26 Fig). This inconsistency between imaging and sequencing could imply that tumorigenesis signature reflected by gene expression profiles may arise before of morphological characteristics, such observation would be difficult to be captured and confirmed by histology reconstructed through expression patterns.

Although SMD-seq can preserve morphology and sequence information with high quality, there are additional limitations. First, SRS imaging in SMD-seq can identify local tissue features; however, there is an intrinsic trade-off between the field of view and resolution of the SRS histology images. In this study, the lower numerical aperture of the objective lens compromised the sensitivity for minor change in subcellular structures, which hindered in depth image analysis. Increasing pixel density of image may reduce the noise level with higher spatial sampling rate (S27 Fig), but prolonged exposure times lead to a higher chance of sample ablation during imaging (S27 Fig) and bring higher risk of the RNA degradation during imaging. Another limitation of SMD-seq is the accessibility of SRS microscopy, which requires specialized light sources and sophisticated optical configurations making the technique difficult to implement in general biology and clinical labs. The integration of SRS microscopy into turn-key system should in principle facilitate adoption of SMD-seq in the clinics [61].

In summary, we have shown that SMD-seq can readily detect spatially dependent transcriptional variation and chromosomal alteration in unfixed tissue sections, and it can discover the subtle and rare genetic alteration events, such as gene fusions, copy number changes, and

differentially expressed transcriptome, with high sensitivity and accuracy, underlying different histological features. In this study, we performed bulk genomic analysis on the dissected microsamples, allowing parallel DNA and RNA sequencing on the same ROI. This small-bulk pool-and-split approach masks the single-cell-level heterogeneity within the sample and blurs the relationship between the genome and transcriptome in single cells. SMD-seq could be extended to single-cell analysis with careful dissociation of microsamples. While they could limit the ability of multiomic characterization, such a development would provide higher resolution molecular profiling, including the analysis of epigenetic information. In combination with deep genome sequencing, the detection of SNVs can also be incorporated into current methods. SMD-seq offers a new direction for cancer research, with the integrated analysis of histology, transcriptome, and genome, it has the potential to enable a more comprehensive understanding of the tumorigenesis process and diagnosis.

Materials and methods

Pipeline of SMD-seq

1. Preparation of tissue section. Tumor samples were collected from 4 patients who suffered various stages of OSCC at different ages and with different genders. All the biopsies were collected by protocols reviewed by and approved by the Ethics Committee of Peking University School and the Hospital of Stomatology (PKUSSIRB-201418116). Tumor samples diagnosed with OSCC were stored in DMEM supplemented with 10% (vol/vol) FBS at 4°C immediately after operation. Samples were stored in ice bucket and transported to lab within 30 min. The samples were then frozen in a cryostat (CM1950, Leica, Germany) at −20°C and were sectioned to 30 μm thin sections. The sections were placed onto polyethylene naphthalate (PEN) membrane slides (RNAase-Free, Leica, Germany). All sections were kept on ice before imaging.

2. SRS imaging and laser microdissection in situ. Each slide was surveyed with SRS microscopy at Raman peaks 2,850 and 2,950 cm^{-1} immediately after sectioning. The 2 Raman peaks represent CH_2 symmetric vibration and CH_3 vibration, respectively. To obtain a fast switch between the 2 peaks, the wavelength of optical parametric oscillator (OPO) was set at 813 nm (corresponding approximately 2,901 cm^{-1}) and switching to 809.8 nm and 816.4 nm by tuning the lyot filter in the cavity. Each image was scanned with a pixel dwelling time of 4 μs and a size of 640 × 640 pixels. To each image, a 2-frame Karlman filtering was applied. After above processing steps, the regions of cancerous cells were able to be identified by the experienced collaborating pathologist. A quick check at different focal planes was done before microdissection to make sure no big morphology change presenting within the 30 μm thickness. The post objective excitation power was 100 mW for pump and 40 mW for Stokes when imaging, and the pump (813 nm) power was increased to 180 to 190 mW to perform microdissection. To dissect an ROI, the laser repeatedly scanned 500 times along the border with a pixel dwelling time of 5 ms, hence effectively incised the PEN membrane from the glass slide, facilitating the cutting of ROIs. The dissected sample was manually transferred from slide to tube loaded with lysis buffer by RNase-free syringe needle. For each tissue section, we selected at least 1 dissected region of tumor and 2 dissected areas of similar size from normal tissues, one of them from the epithelium (the origin of this tumorigenesis) and the other from gland or muscle. To avoid the severe damage of RNA in the tissue sections, the whole procedure needs to be completed promptly.

3. Transcriptome and genome sequencing. The dissected samples were put into lysis buffer [62] separately and immediately centrifuged at 13,000 rpm for 30 s. After lysis, each sample was equally divided into 2 aliquots for RNA-seq and genomic DNA sequencing,

respectively. The protocol of RNA-seq was adapted from the pipeline of single-cell transcriptome analysis [62]. In brief, mRNA was reverse transcribed into first strand cDNA with polyT primer that has an anchor sequence. After other used primers were digested, polyA was added to the 3′ end of cDNA and second strand cDNA was formed and amplified with polyT primer with another anchor sequence by PCR. We employed degenerate oligonucleotide primed PCR (DOP-PCR) for amplifying the whole genome of each lysed tissue sample by the GenomePlex Single Cell Whole Genome Amplification Kit (WGA4-50RXN, Sigma-Aldrich, United States of America). For each sample, 50 ng of amplified genomic DNA and cDNA were used as the start amount of libraries preparation, separately. The pair-end sequencing libraries with approximately 300 bp insert size were constructed following the instructions of NEB Next Ultra DNA Library Prep Kit for Illumina (E7370, New England Biolabs, USA). Illumina HiSeq 2500 systems were used for sequencing.

Stimulated Raman scattering (SRS) microscope

The home-built SRS system used a pump laser integrated OPO (picoEmerald, APE, Germany). It provided 2 spatially and temporally overlapped pulse trains, with the synchronized repetition rate of 80 MHz. One beam, fixed at 1,064 nm, was used as the Stokes light. The other beam, tunable from 780 to 990 nm, served as the pump light. The intensity of the Stokes beam was modulated at 20.2 MHz by a resonant electro-optical modulator (EOM). The overlapped lights were directed into an inverted multiphoton scanning microscope (FV1000, Olympus, Japan). The collinear laser beams were focused into the sample by a 20× objective (UPlan-SAPO, NA 0.75, Olympus, Japan). Transmitted light was collected by a condenser (NA 0.9, Olympus, Japan). After filtering out the Stokes beam, the pump beam was directed onto a large area photo diode (FDS1010, Thorlabs, USA). The voltage from photo diode was sent into lock-in amplifier (HF2LI, Zurich Instruments, Switzerland) to extract the SRS signal. Image was reconstructed through software provided by manufacture (FV10ASW, Olympus, Japan).

Image analysis

For 2-color SRS image, we applied a linear combination approach [63] on each field of view to convert the 2,850 and 2,950 cm^{-1} images into a reconstructed pseudo-color image (S1 Fig), in which we represented protein- and lipid-rich regions with blue and yellow, respectively. For texture analysis, all images were processed with the same procedure for HOG feature extraction. Dual color SRS images were resized to 80×80 pixels and subjected to Matlab (Mathwork, USA) built-in function "extractHOGFeature" to generate feature vectors. Hierarchical clustering was applied to extracted feature vector clustering in R, with built-in function "hclust" ("ward.D" method). Image correlation matrix was generated in R.

Sequencing data analysis

Reproducibility of SMD-seq of these samples were validated by checking the Spearman correlation coefficients (r) of expressed genes (average r = 0.7) and reads count (average r = 0.7) between biological replicates of the same patients. Adaptor contamination and low-quality reads (phred quality <20) were discarded from the raw data. Only samples with coefficient of variation (CV) of reads count per 1 M bin <0.25 (genomic DNA) and gene number more than 6,000 (FPKM > 0.1, RNA) were kept for analysis. For RNA-seq data, TopHat (v2.0.10) were used for sequencing alignment. Reference genome assembly hg19 and gene annotation files were downloaded from UCSC Genome Browser. FPKM values used for analyses were generated by Cufflinks (v2.1.1), and Cuffdiff (v2.2.1) was used for gene expression levels comparison. Significantly different expressed genes between muscle, gland, epithelium, and cancer

were selected under the criteria that *p* value < 0.05 and |log2 (fold change)| >1. Gene functional annotation was performed by The Database for Annotation, Visualization, and Integrated Discovery (DAVID) v6.717 [64]. The purity of tumor samples was estimated by ESTIMATE [65] with gene expression data. They were submitted to ESTIMATE for calculation of each score, then pooled and compared with the database calculated from TCGA HNSCC samples (*n* = 522) by the developer. Gene fusion analysis was carried out by Fusion-Catcher (v0.99.4a) [66] with 4 mapping tools (Bowtie, Bowtie2, BLAT, and STAR). Matched normal samples were used for each patient to exclude the fusion genes that are also found in normal samples. Under following situations, the fusion were discarded: (1) both fusion genes are mutual paralogs; (2) one or both of the fusion genes were pseudogene; (3) reported only by 1 mapping tool or reported by 2 mapping tools only once; (4) no known genes existed in between the fusion genes; and (5) the distance between both genes were less than 100 kbp. Under these criteria, 24 fusion genes were discovered with more than 10 paired reads spanning 2 different genes sequences. The circular diagram of fusion gene was generated by CIRCOS (v0.67–7) [67]. RNA-seq data were used for variant calling by GATK (v3.4–0) according to GATK Best Practices recommendations [68]. We performed duplicate removal, SplitNCigar-Reads, base quality score recalibration before SNP calling and filtered out SNPs by Fisher Strand values (FS > 30.0), Qual by Depth values (QD < 2.0), and sequencing depth passing the quality filter (DP < 10). Annotation of SNPs was performed by SnpEff (v4.0) [69]. Significantly mutated genes in HNSCC were inferred from COSMIC (Catalogue of Somatic Mutations in Cancer) and a comprehensive previous study [70]. Spearman correlation coefficient was computed between tissue samples by function "cor" in R. The unsupervised hierarchical clustering was performed by the function "pheatmap" of package "pheatmap" in R, and the method of measuring the distance in clustering columns was "manhattan". The 3D PCA plot was generated by R package "scatterplot3d."

The sequencing depth of DNA-seq was approximately 0.1×. Genomic DNA sequencing reads were mapped to reference genome by bowtie2 (v2.2.3) [71]. After duplication removal of mappable reads, the counts of aligned reads were calculated in each 1 M bin along the genome (Fig 4A and 4C–4E). For each bin, the read count of each tumor sample was normalized by sequencing depth and the median read count of all normal tissue samples, and the generated copy number went through segmentation by Circular Binary Segmentation [43] (the significance level was set as 0.05). The mean gene expression values of cancer samples within each segment were also calculated and normalized by mean expression values of normal samples that had corresponding qualified (CV < 0.25) gDNA reads (Fig 4C–4E). Function "pheatmap" in R was adopted for CNAs clustering (Fig 4B). GISTIC 2.030 was adopted to analyze the significantly reoccurring focal alterations for the gDNA segmented data.

Supporting information

S1 Materials and methods. Histologically-resolved multiomics enables precise molecular profiling of human intratumor heterogeneity.
(DOCX)

S1 Fig. Reconstruction process of dual-color SRS histology image. SRS, stimulated Raman scattering.
(TIF)

S2 Fig. Dual-color SRS images of Warthin's tumor and mucoepidermoid carcinoma. SRS, stimulated Raman scattering.
(TIF)

S3 Fig. SRS images of epithelium and cancer for unsupervised hierarchical clustering. SRS, stimulated Raman scattering.
(TIF)

S4 Fig. Correlation matrix of samples in 16 cancer samples and 16 epithelium samples.
(TIF)

S5 Fig. Images of cancer infiltrating muscle of P4.
(TIF)

S6 Fig. Localization of SRS subimage in corresponding HE-stained images. HE, hematoxylin–eosin; SRS, stimulated Raman scattering.
(TIF)

S7 Fig. Characterization of microdissection of micro tissues.
(TIF)

S8 Fig. RNA preservation comparison between HE-stained and unstained sections. HE, hematoxylin–eosin.
(TIF)

S9 Fig. Spearman correlation coefficients calculated between SMD-seq samples. SMD-seq, SRS microdissection and sequencing.
(TIF)

S10 Fig. Unsupervised hierarchical clustering and gene annotation of enriched genes in different tissues.
(TIF)

S11 Fig. Gene expression levels of AKR1B10, FTH1, and FTL between cancer and epithelium samples from different patients.
(TIF)

S12 Fig. Immunofluorescence images to show the protein expression level of GSTP1.
(TIF)

S13 Fig. Keratin pearls in cancer nest of P3.
(TIF)

S14 Fig. Three-dimensional location of cancer nests of P4.
(TIF)

S15 Fig. Stromal content evaluation.
(TIF)

S16 Fig. The validation of gene fusion events by Sanger sequencing.
(TIF)

S17 Fig. Fusion events of intratumor ROIs. ROI, regions of interest.
(TIF)

S18 Fig. Gene fusion events in sample P4S2E.
(TIF)

S19 Fig. Sequencing coverage across the whole genome.
(TIF)

S20 Fig. Genome-wide copy number distribution at different sampling rates.
(TIF)

S21 Fig. Comparison of unsupervised clustering of normalized reads count between Ginkgo and our methods.
(TIF)

S22 Fig. Copy number variation of autosomes from all cancer samples.
(TIF)

S23 Fig. Gene expression fold change of autosomes.
(TIF)

S24 Fig. Copy number variation and gene expression fold change of the same sectioned slice.
(TIF)

S25 Fig. Significant focal copy number alterations of all the cancer samples analyzed by GISTIC2.0.
(TIF)

S26 Fig. Significantly mutated genes in OSCC discovered by previous study and COSMIC.
COSMIC, Catalogue of Somatic Mutations in Cancer; OSCC, oral squamous cell carcinoma.
(TIF)

S27 Fig. Effect of image size and sample damage caused by laser.
(TIF)

S1 Table. Patients' information and corresponding dissected tissues.
(TIF)

S2 Table. Summary of RNA-seq datasets.
(TIF)

S3 Table. Summary of genomic DNA sequencing datasets.
(TIF)

S4 Table. Overview of spatial transcriptomic methods.
(TIF)

S1 Data. Summary of genomic sequencing and RNA sequencing data analysis.
(XLSX)

S2 Data. Summary of numeric values presented in figures of main text and supporting information.
(XLSX)

S3 Data. Sanger sequencing results of fusion genes.
(ZIP)

Acknowledgments

Author Contributions

Conceptualization: Tiejun Li, Yanyi Huang.

Data curation: Chen Cao.

Formal analysis: Tao Chen, Chen Cao, Jianyun Zhang, Aaron Streets.

Funding acquisition: Yanyi Huang.

Investigation: Tao Chen, Chen Cao, Jianyun Zhang, Aaron Streets.

Methodology: Tao Chen, Aaron Streets, Tiejun Li, Yanyi Huang.

Project administration: Tiejun Li, Yanyi Huang.

Resources: Tiejun Li, Yanyi Huang.

Supervision: Aaron Streets, Tiejun Li, Yanyi Huang.

Validation: Chen Cao.

Visualization: Tao Chen, Chen Cao, Jianyun Zhang, Yanyi Huang.

Writing – original draft: Tao Chen, Chen Cao, Jianyun Zhang, Aaron Streets, Yanyi Huang.

Writing – review & editing: Tao Chen, Chen Cao, Aaron Streets, Yanyi Huang.

References

1. Kalisky T, Blainey P, Quake SR. Genomic Analysis at the Single-Cell Level. Annu Rev Genet. 2011; 45:431–445. https://doi.org/10.1146/annurev-genet-102209-163607 PMID: 21942365

2. Wu AR, Wang J, Streets AM, Huang Y. Single-Cell Transcriptional Analysis. Annu Rev Anal Chem. 2017; 10:439–462. https://doi.org/10.1146/annurev-anchem-061516-045228 PMID: 28301747

3. Mereu E, Lafzi A, Moutinho C, Ziegenhain C, McCarthy DJ, Álvarez-Varela A, et al. Benchmarking single-cell RNA-sequencing protocols for cell atlas projects. Nat Biotechnol. 2020; 38:747–755. https://doi.org/10.1038/s41587-020-0469-4 PMID: 32518403

4. Hwang B, Lee JH, Bang D. Single-cell RNA sequencing technologies and bioinformatics pipelines. Exp Mol Med. 2018; 50:1–14. https://doi.org/10.1038/s12276-018-0071-8 PMID: 30089861

5. Ren X, Kang B, Zhang Z. Understanding tumor ecosystems by single-cell sequencing: promises and limitations. Genome Biol. 2018; 19:211. https://doi.org/10.1186/s13059-018-1593-z PMID: 30509292

6. Azizi E, Carr AJ, Plitas G, Cornish AE, Konopacki C, Prabhakaran S, et al. Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. Cell. 2018; 174:1293–1308.e36. https://doi.org/10.1016/j.cell.2018.05.060 PMID: 29961579

7. Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, Goldstein SR, et al. Laser Capture Microdissection. Science. 1996; 274:998–1001. https://doi.org/10.1126/science.274.5289.998 PMID: 8875945

8. Espina V, Wulfkuhle JD, Calvert VS, VanMeter A, Zhou W, Coukos G, et al. Laser-capture microdissection. Nat Protoc. 2006; 1:586–603. https://doi.org/10.1038/nprot.2006.85 PMID: 17406286

9. Zhan J, Thakare D, Ma C, Lloyd A, Nixon NM, Arakaki AM, et al. RNA Sequencing of Laser-Capture Microdissected Compartments of the Maize Kernel Identifies Regulatory Modules Associated with Endosperm Cell Differentiation. Plant Cell. 2015; 27:513–531. https://doi.org/10.1105/tpc.114.135657 PMID: 25783031

10. Kube DM, Savci-Heijink CD, Lamblin A-F, Kosari F, Vasmatzis G, Cheville JC, et al. Optimization of laser capture microdissection and RNA amplification for gene expression profiling of prostate cancer. BMC Mol Biol. 2007; 8:25. https://doi.org/10.1186/1471-2199-8-25 PMID: 17376245

11. Pote A, Boghenco O, Marques-Ramos A. Molecular analysis of H&E- and Papanicolau-stained samples—systematic review. Histochem Cell Biol. 2020; 154:7–20. https://doi.org/10.1007/s00418-020-01882-w PMID: 32372108

12. Wang H, Owens JD, Shih JH, Li M-C, Bonner RF, Mushinski JF. Histological staining methods preparatory to laser capture microdissection significantly affect the integrity of the cellular RNA. BMC Genomics. 2006; 7:97. https://doi.org/10.1186/1471-2164-7-97 PMID: 16643667

13. Freudiger CW, Min W, Saar BG, Lu S, Holtom GR, He C, et al. Label-free biomedical imaging with high sensitivity by stimulated Raman scattering microscopy. Science. 2008; 322:1857–1861. https://doi.org/10.1126/science.1165758 PMID: 19095943

14. Ji M, Orringer DA, Freudiger CW, Ramkissoon S, Liu X, Lau D, et al. Rapid, Label-Free Detection of Brain Tumors with Stimulated Raman Scattering Microscopy. Sci Transl Med. 2013; 5:201ra119–201ra119. https://doi.org/10.1126/scitranslmed.3005954 PMID: 24005159

15. Ji M, Lewis S, Camelo-Piragua S, Ramkissoon SH, Snuderl M, Venneti S, et al. Detection of human brain tumor infiltration with quantitative stimulated Raman scattering microscopy. Sci Transl Med. 2015; 7. https://doi.org/10.1126/scitranslmed.aab0195 PMID: 26468325

16. Ozeki Y, Umemura W, Otsuka Y, Satoh S, Hashimoto H, Sumimura K, et al. High-speed molecular spectral imaging of tissue with stimulated Raman scattering. Nat Photonics. 2012; 6:845–851. https://doi.org/10.1038/nphoton.2012.263

17. Shin KS, Laohajaratsang M, Men S, Figueroa B, Dintzis SM, Fu D. Quantitative chemical imaging of breast calcifications in association with neoplastic processes. Theranostics. 2020; 10:5865–5878. https://doi.org/10.7150/thno.43325 PMID: 32483424

18. Reiter JG, Baretti M, Gerold JM, Makohon-Moore AP, Daud A, Iacobuzio-Donahue CA, et al. An analysis of genetic heterogeneity in untreated cancers. Nat Rev Cancer. 2019; 19:639–650. https://doi.org/10.1038/s41568-019-0185-x PMID: 31455892

19. Turajlic S, Sottoriva A, Graham T, Swanton C. Resolving genetic heterogeneity in cancer. Nat Rev Genet. 2019; 20:404–416. https://doi.org/10.1038/s41576-019-0114-6 PMID: 30918367

20. Baba AI, Câtoi C. Comparative Oncology. Bucharest (RO): The Publishing House of the Romanian Academy; 2007.

21. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. N Engl J Med. 2012; 366:883–892. https://doi.org/10.1056/NEJMoa1113205 PMID: 22397650

22. Alizadeh AA, Aranda V, Bardelli A, Blanpain C, Bock C, Borowski C, et al. Toward understanding and exploiting tumor heterogeneity. Nat Med. 2015; 21:846–853. https://doi.org/10.1038/nm.3915 PMID: 26248267

23. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. CA Cancer J Clin. 2017; 67:7–30. https://doi.org/10.3322/caac.21387 PMID: 28055103

24. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. CA Cancer J Clin. 2016; 66:115–132. https://doi.org/10.3322/caac.21338 PMID: 26808342

25. Fan S, Tang Q, Lin Y, Chen W, Li J, Huang Z, et al. A review of clinical and histological parameters associated with contralateral neck metastases in oral squamous cell carcinoma. Int J Oral Sci. 2011; 3:180–191. https://doi.org/10.4248/IJOS11068 PMID: 22010576

26. Edge SB, Byrd DR, Compton CC, Fritz AG, Greene FL, Trotti A. AJCC Cancer Staging Handbook. 7th ed. Edge SB, Byrd DR, Compton CC, Fritz AG, Greene FL, Trotti A, editors. New York: Springer New York; 2010.

27. Vettore AL, Ramnarayanan K, Poore G, Lim K, Ong CK, Huang KK, et al. Mutational landscapes of tongue carcinoma reveal recurrent mutations in genes of therapeutic and prognostic relevance. Genome Med. 2015; 7:98. https://doi.org/10.1186/s13073-015-0219-2 PMID: 26395002

28. Berkovitz BKB, Holland GR, Moxham BJ. Oral anatomy, histology and embryology. 4th ed. Kidlington, Oxford, UK: Mosby Elsevier; 2009.

29. Mayorca-Guiliani AE, Madsen CD, Cox TR, Horton ER, Venning FA, Erler JT. ISDoT: in situ decellularization of tissues for high-resolution imaging and proteomic analysis of native extracellular matrix. Nat Med. 2017; 23:890–898. https://doi.org/10.1038/nm.4352 PMID: 28604702

30. Pavelka M, Roth J. Functional Ultrastructure. Vienna: Springer Vienna. 2015. https://doi.org/10.1007/978-3-7091-1830-6

31. Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection. CVPR '05 Proc 2005 IEEE Comput Soc Conf Comput Vis Pattern Recognit—Vol 1. 2005; 886–893. https://doi.org/10.1109/CVPR.2005.177

32. Troy JD, Weissfeld JL, Diergaarde B, Youk AO, Buch SC, Romkes M, et al. Polymorphisms in NAT2 and GSTP1 are associated with survival in oral and oropharyngeal cancer. Cancer Epidemiol. 2013; 37:505–511. https://doi.org/10.1016/j.canep.2013.02.009 PMID: 23523331

33. Ma H, Yu C, Liu Y, Tan Y, Qiao J, Yang X, et al. Decreased expression of glutathione S-transferase pi correlates with poorly differentiated grade in patients with oral squamous cell carcinoma. J Oral Pathol Med. 2015; 44:193–200. https://doi.org/10.1111/jop.12229 PMID: 25047743

34. De Craene B, Berx G. Regulatory networks defining EMT during cancer initiation and progression. Nat Rev Cancer. 2013; 13:97–110. https://doi.org/10.1038/nrc3447 PMID: 23344542

35. Dongre A, Weinberg RA. New insights into the mechanisms of epithelial–mesenchymal transition and implications for cancer. Nat Rev Mol Cell Biol. 2019; 20:69–84. https://doi.org/10.1038/s41580-018-0080-4 PMID: 30459476

36. Pettus JR, Johnson JJ, Shi Z, Davis JW, Koblinski J, Ghosh S, et al. Multiple kallikrein (KLK 5, 7, 8, and 10) expression in squamous cell carcinoma of the oral cavity. Histol Histopathol. 2009; 24:197–207. https://doi.org/10.14670/HH-24.197 PMID: 19085836

37. Speight PM, Farthing PM. The pathology of oral cancer. Br Dent J. 2018; 225:841–847. https://doi.org/10.1038/sj.bdj.2018.926 PMID: 30412536

38. Puram SV, Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S, et al. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. Cell. 2017; 171:1611–1624.e24. https://doi.org/10.1016/j.cell.2017.10.044 PMID: 29198524

39. Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. Nat Rev Cancer. 2007; 7:233–245. https://doi.org/10.1038/nrc2091 PMID: 17361217

40. Ben-Neriah Y, Daley GQ, Mes-Masson AM, Witte ON, Baltimore D. The chronic myelogenous leukemia-specific P210 protein is the product of the bcr/abl hybrid gene. Science. 1986; 233:212–214. https://doi.org/10.1126/science.3460176 PMID: 3460176

41. Kwak EL, Bang Y-J, Camidge DR, Shaw AT, Solomon B, Maki RG, et al. Anaplastic Lymphoma Kinase Inhibition in Non–Small-Cell Lung Cancer. N Engl J Med. 2010; 363:1693–1703. https://doi.org/10.1056/NEJMoa1006448 PMID: 20979469

42. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. Nature. 2012; 489:519–525. https://doi.org/10.1038/nature11404 PMID: 22960745

43. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics. 2004; 5:557–572. https://doi.org/10.1093/biostatistics/kxh008 PMID: 15475419

44. Kuga T, Kume H, Kawasaki N, Sato M, Adachi J, Shiromizu T, et al. A novel mechanism of keratin cytoskeleton organization through casein kinase Iα and FAM83H in colorectal cancer. J Cell Sci. 2013. https://doi.org/10.1242/jcs.129684 PMID: 23902688

45. The Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature. 2015; 517:576–582. https://doi.org/10.1038/nature14129 PMID: 25631445

46. Garvin T, Aboukhalil R, Kendall J, Baslan T, Atwal GS, Hicks J, et al. Interactive analysis and assessment of single-cell copy-number variations. Nat Methods. 2015; 12:1058–1060. https://doi.org/10.1038/nmeth.3578 PMID: 26344043

47. Yamashita S. p53AIP1 Expression can be a Prognostic Marker in Non-small Cell Lung Cancer. Clin Oncol. 2008; 20:148–151. https://doi.org/10.1016/j.clon.2007.08.006 PMID: 17851056

48. Demirag GG, Sullu Y, Gurgenyatagi D, Okumus NO, Yucel I. Expression of Plakophilins (PKP1, PKP2, and PKP3) in Gastric Cancers. Diagn Pathol. 2011; 6:1. https://doi.org/10.1186/1746-1596-6-1 PMID: 21194493

49. Sevenich L, Schurigt U, Sachse K, Gajda M, Werner F, Müller S, et al. Synergistic antitumor effects of combined cathepsin B and cathepsin Z deficiencies on breast cancer progression and metastasis in mice. Proc Natl Acad Sci U S A. 2010; 107:2497–2502. https://doi.org/10.1073/pnas.0907240107 PMID: 20133781

50. Bengsch F, Buck A, Günther SC, Seiz JR, Tacke M, Pfeifer D, et al. Cell type-dependent pathogenic functions of overexpressed human cathepsin B in murine breast cancer progression. Oncogene. 2014; 33:4474–4484. https://doi.org/10.1038/onc.2013.395 PMID: 24077280

51. Castro ME, Ferrer I, Cascon A, Guijarro MV, Lleonart M, Cajal SR, et al PPP1CA contributes to the senescence program induced by oncogenic Ras. Carcinogenesis. 2007; 29:491–499. https://doi.org/10.1093/carcin/bgm246 PMID: 18204081

52. Mertens F, Johansson B, Fioretos T, Mitelman F. The emerging complexity of gene fusions in cancer. Nat Rev Cancer. 2015; 15:371–381. https://doi.org/10.1038/nrc3947 PMID: 25998716

53. Navin NE. Cancer genomics: one cell at a time. Genome Biol. 2014; 15:452. https://doi.org/10.1186/s13059-014-0452-9 PMID: 25222669

54. Crosetto N, Bienko M, van Oudenaarden A. Spatially resolved transcriptomics and beyond. Nat Rev Genet. 2015; 16:57–66. https://doi.org/10.1038/nrg3832 PMID: 25446315

55. Moses L, Pachter L. Museum of Spatial Transcriptomics. bioRxiv 2021; 2021.05.11.443152. https://doi.org/10.1101/2021.05.11.443152

56. Eng C-HL, Lawson M, Zhu Q, Dries R, Koulena N, Takei Y, et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. Nature. 2019; 568:235–239. https://doi.org/10.1038/s41586-019-1049-y PMID: 30911168

57. Chen KH, Boettiger AN, Moeffitt JR, Wang S, Zhuang X. Spatially resolved, highly multiplexed RNA profiling in single cells. Science. 2015; 348:aaa6090. https://doi.org/10.1126/science.aaa6090 PMID: 25858977

58. Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science. 2016; 353:78–82. https://doi.org/10.1126/science.aaf2403 PMID: 27365449

59. Liu Y, Yang M, Deng Y, Su G, Enninful A, Guo CC, et al. High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. Cell. 2020; 183:1665–1681.e18. https://doi.org/10.1016/j.cell.2020.10.026 PMID: 33188776

60. Rodriques SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, et al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. Science. 2019; 363:1463–1467. https://doi.org/10.1126/science.aaw1219 PMID: 30923225

61. Orringer DA, Pandian B, Niknafs YS, Hollon TC, Boyle J, Lewis S, et al. Rapid intraoperative histology of unprocessed surgical specimens via fibre-laser-based stimulated Raman scattering microscopy. Nat Biomed Eng. 2017; 1:1–13. https://doi.org/10.1038/s41551-016-0027 PMID: 28955599

62. Tang F, Barbacioru C, Nordman E, Li B, Xu N, Bashkirov VI, et al. RNA-Seq analysis to capture the transcriptome landscape of a single cell. Nat Protoc. 2010; 5:516–535. https://doi.org/10.1038/nprot.2009.236 PMID: 20203668

63. Yu Z, Chen T, Zhang X, Fu D, Liao X, Shen J, et al. Label-free chemical imaging in vivo: three-dimensional non-invasive microscopic observation of amphioxus notochord through stimulated Raman scattering (SRS). Chem Sci. 2012; 3:2646. https://doi.org/10.1039/c2sc20316b

64. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. 2003; 4:R60. https://doi.org/10.1186/gb-2003-4-9-r60

65. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013; 4:2612. https://doi.org/10.1038/ncomms3612 PMID: 24113773

66. Nicorici D, Satalan M, Edgren H, Kangaspeska S, Murumagi A, Kallioniemi O, et al. FusionCatcher—a tool for finding somatic fusion genes in paired-end RNA-sequencing data. bioRxiv. 2014. https://doi.org/10.1101/011650

67. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: An information aesthetic for comparative genomics. Genome Res. 2009; 19:1639–1645. https://doi.org/10.1101/gr.092759.109 PMID: 19541911

68. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. Curr Protoc Bioinformatics. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2013. p. 11.10.1–11.10.33. https://doi.org/10.1002/0471250953.bi1110s43

69. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin). 2012; 6:80–92. https://doi.org/10.4161/fly.19695 PMID: 22728672

70. Bamford S, Dawson E, Forbes S, Clements J, Pettett R, Dogan A, et al. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. Br J Cancer. 2004; 91:355–358. https://doi.org/10.1038/sj.bjc.6601894 PMID: 15188009

71. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012; 9:357–359. https://doi.org/10.1038/nmeth.1923 PMID: 22388286