RESEARCH

Novel temporal and spatial patterns of metastatic colonization from breast cancer rapid-autopsy tumor biopsies

Xiaomeng Huang[1,2†], Yi Qiao[1,2†], Samuel W. Brady[3,4], Rachel E. Factor[5], Erinn Downs-Kelly[5], Andrew Farrell[1,2], Jasmine A. McQuerry[3,6], Gajendra Shrestha[3], David Jenkins[7], W. Evan Johnson[7], Adam L. Cohen[8], Andrea H. Bild[9] and Gabor T. Marth[1,2*]

Abstract

Background: Metastatic breast cancer is a deadly disease with a low 5-year survival rate. Tracking metastatic spread in living patients is difficult and thus poorly understood.

Methods: Via rapid autopsy, we have collected 30 tumor samples over 3 timepoints and across 8 organs from a triple-negative metastatic breast cancer patient. The large number of sites sampled, together with deep whole-genome sequencing and advanced computational analysis, allowed us to comprehensively reconstruct the tumor's evolution at subclonal resolution.

Results: The most unique, previously unreported aspect of the tumor's evolution that we observed in this patient was the presence of "subclone incubators," defined as metastatic sites where substantial tumor evolution occurs before colonization of additional sites and organs by subclones that initially evolved at the incubator site. Overall, we identified four discrete waves of metastatic expansions, each of which resulted in a number of new, genetically similar metastasis sites that also enriched for particular organs (e.g., abdominal vs bone and brain). The lung played a critical role in facilitating metastatic spread in this patient: the lung was the first site of metastatic escape from the primary breast lesion, subclones at this site were likely the source of all four subsequent metastatic waves, and multiple sites in the lung acted as subclone incubators. Finally, functional annotation revealed that many known drivers or metastasis-promoting tumor mutations in this patient were shared by some, but not all metastatic sites, highlighting the need for more comprehensive surveys of a patient's metastases for effective clinical intervention.

Conclusions: Our analysis revealed the presence of substantial tumor evolution at metastatic incubator sites in a patient, with potentially important clinical implications. Our study demonstrated that sampling of a large number of metastatic sites affords unprecedented detail for studying metastatic evolution.

Keywords: Tumor evolution, Subclone, Metastatic breast cancer

* Correspondence: gabor.marth@gmail.com
†Xiaomeng Huang and Yi Qiao contributed equally to this work.
[1]Utah Center for Genetic Discovery, University of Utah, Salt Lake City, USA
[2]Department of Human Genetics, School of Medicine, University of Utah, Salt Lake City, USA
Full list of author information is available at the end of the article

Huang *et al. Genome Medicine*   (2021) 13:170

Page 2 of 18

Background

Metastatic breast cancer (MBC) is a deadly disease with a median survival of only 38 months [1]. A previous study estimated that 3 out of 4 patients initially diagnosed with stage I–III disease progressed to MBC [2]. Although the genomic and transcriptomic properties of primary tumors have been described extensively [3–5], metastatic tumors, as well as the processes leading to metastasis, are poorly understood because comprehensive biopsying of metastatic sites is difficult or impossible in living patients. Rapid autopsy programs, in contrast, offer pathologists a comprehensive spatial understanding of the extent of the disease and allow for the collection of fresh tissue samples across all affected organs within hours of the patient's death. This approach has been used to study metastatic tumor evolution in breast cancer with TNBC patients being a smaller subset [6–9] and in other cancer types [10–14]. For example, Savas et al. studied tumor evolution in 3 estrogen-receptor (ER)-positive, human epidermal growth factor receptor 2 (HER2)-negative breast cancer patients, and 1 triple-negative breast cancer patient, using primary tumor and 5–12 matched metastatic samples from the CASCADE program [8]; and Hoadley et al. profiled primary tumors with 4–5 matched metastases genomically and transcriptomically in 2 triple-negative breast cancer patients [7]. More recently, De Mattos-Arruda et al. profiled 7–26 samples per patient from autopsies of 10 patients (5 ER+/HER2-, 3 ER+/HER2+, 1 ER-/HER2+, 1 TNBC) with therapy-resistant breast cancer [9]. These studies found significant heterogeneity in both the primary and metastatic tumors, and complex evolutionary patterns during disease progression. However, critical questions remain unanswered, especially in TNBC patients: for example, whether the ability for the cancer to metastasize fully develops in the primary tumor, as suggested by studies [7], or if early metastatic sites can provide niches where the cancer can further develop metastatic potential not present in the primary tumor, but necessary to invade additional organs. To understand metastatic tumor evolution and disease progression at the subclonal resolution, we studied the primary tumor at diagnosis and at surgery, as well as 28 metastatic samples across seven organs from a metastatic breast cancer patient with aggressive disease, collected via rapid autopsy following the patient's death, with 30 samples in total. Deep whole-genome sequencing allowed us to reconstruct detailed subclone structure and track subclonal expansion across these samples, elucidating the order and timing in which each metastatic site was established, including metastatic colonization events from one organ to another.

Methods

The workflow for this study is shown in Additional file 1: Fig. S1.

Sample collection

The study was reviewed and approved by the human subjects Institutional Review Boards (IRB) of the University of Utah. Informed consent in accordance with the Declaration of Helsinki was obtained from the patient. We collected in total 44 autopsy samples including 2 skin normal tissues, 28 tumor samples, and 14 adjacent/distal normal tissues from a 45-year-old woman with ER-negative, PR-negative, and HER2-negative metaplastic grade III invasive ductal carcinoma of the breast via rapid autopsy program approximately 2 h after death. There were 12 metastatic samples from the lung, 1 from the kidney, 1 from the peri-pancreas, 1 from the skin, 4 from the brain, 3 from the bone, 4 from the liver, and 2 from the peritracheal lymph nodes. All metastatic samples represent individual tumors except sample Bn3 and Bn4, and Ln2 and Ln3, which were from different parts of the same brain tumor, and the same lung tumor, respectively. The tumor sizes vary ranging from 3 to 33 mm (see detailed description in Additional file 2: Table S1). For all autopsy samples, frozen sections were reviewed by pathologists to confirm the tumor type and presence and to quantify necrosis levels. All autopsy samples were stored in RNAlater at −80°C until DNA and/or RNA isolation. FFPE samples for the primary tumor biopsy and the mastectomy biopsy were also available for this study. All tumor samples and 2 normal skin samples were subjected to whole-genome sequencing (WGS). All samples except the FFPE samples and the two normal skin samples were also subjected to bulk RNA sequencing.

Sample process

DNA from FFPE samples was isolated using the Qiagen QIAamp DNA FFPE Tissue Kit. DNA from all 28 autopsy tumor samples and 2 skin normal tissues was isolated using Qiagen's QIAamp DNA Micro Kit. RNA from all 28 autopsy tumors and adjacent/distal normal samples was extracted by using Qiagen RNeasy Micro/Mini Kit.

WGS analysis including somatic SNV and INDEL calling, CNV calling, LOH calling, structural variant calling, and translocation calling

Primary and mastectomy samples were subjected to 45X WGS at the Huntsman Cancer Institute's High Throughput Genomics Core Facility using the Illumina TruSeq. Metastatic tumors and skin biopsy samples were subjected to 60X WGS at the McDonnell Genome Institute at Washington University using NantOmics.

Samples sequenced at Washington University were provided as aligned BAM files. Primary and mastectomy WGS sequencing data were aligned using an identical pipeline to the one used at the McDonnel Genome Institute to the same GRCh37-lite reference genome [15] (ftp://ftp.ncbi.nih.gov/genbank/genomes/Eukaryotes/vertebrates_mammals/Homo_sapiens/GRCh37/special_requests/GRCh37-lite.fa.gz) using BWA-MEM 0.7.15-r1140; Freebayes 0.9.21 was used to identify SNV and INDEL variants called jointly over all samples using the following command line parameters:

- --allele-balance-priors-off
- --report-genotype-likelihood-max
- --genotype-qualities
- --pooled-discrete
- --pooled-continuous

The variants produced by Freebayes were then subjected to quality filtering, including criteria as follows:

- Variant quality > 30
- Per-sample sequencing depth > 15
- Intersecting with 1000G genome accessibility mask
- Inverse-intersecting with low complexity region mask of GRCh37d5
- Filtering out multi-allelic variant sites

Somatic variants were identified when the variant allele frequency (VAF) was below 0.1 or the alternate allele count was less than five in both normal skin samples. To ensure that differences in sequencing provider and depths did not affect variant detection, we showed that the number of somatic variants detected in BrP and BrM (45X, at Huntsman Cancer Institute) were similar to other samples (60X, at Washington University). We detected in total 20,012 somatic variants across all samples. On average, we detected 10,648 ± 493 (standard deviation) variants per metastatic sample, 10,352 variants in BrP, and 10,121 variants in BrM (Additional file 1: Fig. S2). The same trend can be observed for somatic variants on chromosome 14 and chromosomes 2, 7, 9, 11, 14, 15, 16, 21, and 22 (Additional file 1: Fig. S2). SNVs and INDELs were annotated by SnpEFF 4.2.

FACETS [16] was used to identify copy number variants (CNV) and loss of heterozygosity (LOH) events (Additional file 1: Fig. S3A). All copy number calls were then manually curated. We binned the $\log_2$(ratio) value (calculated from the FACETS R package) in 1Mb windows in each of the 28 tumor samples. Then, we clustered the copy number profile by calculating the Euclidean distance between each pair of samples and then used the UPGMA method to cluster the samples.

Except for BrP, BrM, Ln7, Ln9, and Ln1, all other samples which are clustered together had distinct copy number patterns (Fig. 2A, Additional file 1: Fig. S3B).

Structural variants and translocations were identified using the reference-free variant detection algorithm RUFUS [18] and Lumpy [19] followed by visual inspection in IGV [17, 20].

Allele-specific CNV/LOH calling in multiple samples

Heterozygosity and copy number for each sample were derived using FACETS. Allele-specific copy number changes were not generated by FACETS, but were separately inferred using inherited variants falling in somatic CNV regions. By comparing the AF of these variants between samples, we were able to identify the allele-specific copy number changes. For example, the AF of inherited variants on chromosome 3 in pure tumor samples with copy number neutral LOH chromosome 3 is either 1 or 0, whereas the AF of these variants in pure tumor samples with copy number three and both alleles would be 0.33 and 0.67. A scatter plot of AF of these variants between two samples reveals which chromosome is amplified in the copy number amplified sample, as well as whether the amplified chromosome is the same as the ones in the copy number neutral, LOH sample. Additional file 1: Fig. S4A shows the AF of inherited exonic variants on chromosome 3 between Pa1 and Bn2. Variants in red circles represent the homozygous variants in both samples. The AF of variants in yellow circles indicates that the amplified allele in Pa1 became the only allele that remained in Bn2, at copy number 2. The AF of variants in green circles indicates that the unamplified allele in Pa1 was lost in Bn2. This method enabled us to establish whether samples with the same copy number and LOH are the same events. In addition, this method provides higher resolution information such as allele-specific structural variants (including translocation) (Additional file 1: Fig. S4B).

Tumor phylogenetic tree construction

We used somatic short variants on chromosome 14 (which harbored no copy number or LOH events in all samples except Ln9) to construct the phylogenetic tree across all samples and additional somatic variants on copy number-neutral but LOH chromosomes (chromosomes 2, 7, 9, 11, 15, 16, 21, and 22) to refine the phylogenetic tree in each group. We encoded the state of a known somatic variant locus in a sample as a binary value, where 1 indicates the variant is present (AF>0.1) and 0 that the variant is absent (AF<0.1). We used two methods to construct a phylogenetic tree. First, we used the UPGMA clustering method based on the hamming distance matrix calculated between samples (Additional file 1: Fig. S5). Samples in the same group were

confirmed to the cluster. However, this method does not consider the constraint that samples sharing the same variants should share an evolutionary lineage. Therefore, secondly, we developed a method that would incorporate this constraint while simultaneously assuming that (1) all cancer cells are descendants of a single founding clone (i.e., normal cell) and (2) all variants satisfy the infinite sites assumption that the chance the same variant occurs independently in different cells, as well as variants reverting back to the wild type, is extremely low. Therefore, we can describe our problem as a perfect phylogeny problem [21] with complete and cladistic characters which are the states of variants. For each variant, a binary vector v_i^j is calculated where $v_i^j = 1$ if variant i is found in sample j or 0 otherwise. Variants with the same binary vectors are clustered together, which means that they occur in the same clone, albeit the clone can be found in multiple samples. The evolution ordering between any two variants can be established by comparing their binary vectors. If a variant $i1$ occurred in the clone that already contained $i2$, for all samples j, either of the following two conditions must hold true: (1) $v_{i1}^j = 1$ and $v_{i2}^j = 1$ when j contains the descendant clone or (2) $v_{i1}^j = 1$ and $v_{i2}^j = 0$ when j contains the ancestral clone before variant $i1$ occurred. We implemented this method [22] (available via the GitHub repository for all code used in this manuscript). Although Ln9 had acquired an additional chromosome 14, no variants were lost in this process. Thus, this method can still apply to Ln9. The results from the second method showed that Ln9 was the first sample to branch out and the rest of the samples had a common ancestor (Fig. 3A). We then calibrated the length of evolution branches by the number of variants.

Subclonal analysis with SubcloneSeeker
We used AFs of somatic variants on chromosome 14 to reconstruct subclone structure and estimate cell prevalence of each subclone of BrP, BrM, Ln1, Ln7, and Ln9. Except for Ln9, all samples have CN normal chromosome 14; therefore, AF can be used to accurately estimate the cell prevalence. For the subclone structure of samples in each group, in addition to variants on chromosome 14, we also used group-specific variants that are absent in BrP and BrM on the CN neutral chromosomes containing LOH (chromosomes 2, 7, 9, 11, 15, 16, 21, and 22) events shared by all samples in G1–G4 as well as BrP and BrM. Since these variants occurred after the LOH event chronologically, they are most likely to be heterozygous and can be used for subclone analysis within a group. Because all samples in G4 had chr11p15-q25 deletion, we can also accurately estimate cell prevalence from G4-specific variants in this region. Thus, these variants were also used for subclone analysis in G4.

For subclone analysis, we clustered variants with the same level of AF in all samples to a cluster (C1–C28 in Additional file 1: Fig. S6A, S6B, S6C, S7A, S9A, S9B, S10A). We used 0.05 as the allele frequency cutoff for positive somatic variant detection. The ancestral relationship between two subclones satisfy (1) variants in the ancestral clone have larger AF than variants unique to the descendant clone in one sample, (2) variants in the ancestral clone will also be in the descendant clone via clonal lineage, and (3) variants that have ~0.5 AF are in the founding clone of a sample. SubcloneSeeker v2 [23, 24] (https://github.com/yiq/SubcloneSeeker/tree/v2) was used to jointly construct subclone structures, enumerating all possible trees with ±0.1 VAF tolerance to accommodate the VAF measurement error, and estimating cell prevalence for each subclone for individual samples. SubcloneSeeker outputs all subclone structures if multiple solutions exist. The computational outputs were manually reviewed, and the final subclonal evolutionary trajectory was assembled. When multiple solutions are available, we choose the solution that minimizes cellular prevalence violation in a parent-child relationship. For detailed discussions on alternative solutions, see Additional file 3. The cell prevalence for each subclone (shown in Additional file 1: Fig. S6-S10) was corrected for tumor purity.

Identification of monoclonal and polyclonal seeding
A clone presented at a less than 100% cell prevalence in one sample, and then at 100% in another sample, signifies that this subclone emerged in the former sample, and seeded the latter, which can be characterized as a monoclonal seeding event. However, if a clone had a low cellular prevalence in both samples, it is likely to be the result of a polyclonal seeding event, in which two or more subclones in one sample traveled together or separately and seeded the other one.

Mutational signature analysis
We assessed the dynamics of mutational process over time by analyzing somatic mutation patterns attributed to the branches of the phylogenetic tree, including truncal variants (variants occurs in the primary tumor), variants shared by all samples in G1, G2, G3, and G4, respectively, as well as the remaining, group-specific variants for each group. We applied MutationalPatterns [25] to our dataset. Briefly, after de novo extraction of mutational signatures from the mutation count matrix, the contribution of COSMIC mutational signatures (https://cancer.sanger.ac.uk/cosmic/signatures_v2) to the mutational profile was quantified.

Huang *et al. Genome Medicine*   (2021) 13:170

Page 5 of 18

RNA-seq and data processing

RNA-seq was performed using rRNA depletion-based library preparation followed by paired-end Illumina HiSeq sequencing. We obtained RNA-seq data from 42 specimens throughout the patient's body at autopsy, including 28 gross tumors (which also had DNA samples) and 14 surrounding normal samples. RNA-seq data were processed with Rsubread [26, 27] v1.16.1. We aligned the reads to GRCh37 and used only uniquely mapped reads and the Hamming distance to break ties. The maximum indels allowed per alignment was 5. Gene-level expression values were processed to transcript per million mapped reads (TPM). We used the featureCounts function in Rsubread for reads counting. We used the built-in annotation file which includes the exon annotation information from NCBI Build GRCh37.2 and Entrez gene identifier.

RNA-seq-based copy number inference

To infer approximate copy number events from RNA-seq data, we used an approach reported previously for single-cell RNA-seq copy number inference [28], which we had also used previously [29]. This approach relies on the normalization and calculation of 101-gene window expression averages, followed by normalization to samples with little or no tumor purity.

Differential expression (DE) analysis

We used a R workflow package, "RnaSeqGeneEdgeRQL" [30] for normalization, and downstream DE and pathway analysis. Specifically, we normalized data by using calcNormFactors function which applied the trimmed mean of M values (TMM) approach. Next, the DE analysis was performed on samples in four groups (G1–G4) by using EdgeR which implemented empirical Bayes methods that permit the estimation of gene-specific biological variations. We made pairwise comparisons between all four groups and performed a one-way analysis of deviance (ANODEV) for each gene. FDR<0.05 was used for the significance cutoff. Significantly expressed genes in each group were annotated in terms of higher order biological processes or molecular pathways by using the NCI-Nature pathway database in Enrichr [31]. Finally, we performed gene set enrichment analysis (GSEA) using the C2 curated signatures from MSigDB (including 5637 signatures). RnaSeqGeneEdgeRQL package incorporates the Correlation Adjusted MEan RAnk gene set test (CAMERA) [32] method for the enrichment analysis.

Validation of variants in RNA-seq data

Somatic SNVs and INDELs identified by Freebayes from WGS data were validated by RNA-seq data. We randomly picked two samples in each group and Ln7 (total nine samples) with high tumor purity in RNA-seq data for this validation. For any given tumor sample, only somatic short variants that have greater than 0.1 genomic VAF and have a read depth of at least ten in the paired RNA-seq were considered. We use the following workflow to validate the variants:

- If a variant is present in paired RNA-seq data, i.e., having RNA-seq reads containing the variant allele, it is considered "validated."
- If a variant is not found in the paired RNA-seq data but found in the RNA-seq data of other tumor samples genomically determined to also contain the same variant, it is then considered as "validated in other samples."
- If a variant cannot be validated by either of the mentioned steps, we consider the following possibilities:
- Variant dropout in RNA seq data due to sampling: It is reasonable to consider that a variant with low WGS VAF (e.g., 0.1) and low RNA-seq coverage (e.g., 10X) may not be sampled by RNA-seq according to binomial distribution (in the example case, the possibility of sampling 0 alternate allele containing reads, or P_0, is 0.35). We skip such variants with $P_0 > 0.05$. Note that less than five variants were in this category in each sample.
- Variant allele not expressed: this can be the result of unbalanced expression between alleles or false positives in genomic variant calling.

More than 90% of variants can be either "validated" or "validated in other samples" (Additional file 1: Fig. S11).

Results

Clinical presentation showed extremely aggressive metastatic cancer

We studied a 45-year-old woman with ER-negative, PR-negative, and HER2-negative metaplastic grade III invasive ductal carcinoma of the breast (Fig. 1A). At the time of diagnosis, she had a clinical T2N0M0 breast cancer, with staging including an ultrasound (US) and MRI showing a 3.1-cm mass. However, the sentinel node biopsy showed none of the three biopsied lymph nodes had cancer. The patient received neoadjuvant therapy of doxorubicin and cyclophosphamide (AC) followed by 8 weeks of weekly paclitaxel. An ultrasound image after the AC showed enlargement of the breast mass. After paclitaxel, the mass remained stable in size on ultrasound but was more painful. The patient underwent a mastectomy at week 21. At week 39 after diagnosis, MRIs showed brain metastases, and CT scans showed multifocal metastases in the lung, liver, pancreas, bone, skin, and lymph nodes. This led to subsequent courses of chemotherapy and radiation therapy, all without

Huang *et al. Genome Medicine*   (2021) 13:170

Page 6 of 18

Fig. 1 Patient treatment history and sample origins. **A** Treatment history over the course of disease progression as well as imaging history including ultrasound (US) imaging, MRI, and CT scan. **B** Primary breast tumors at diagnosis (BrP) and at mastectomy (BrM) were biopsied. 26 metastatic tumors and 2 normal skin autopsy samples were collected through a rapid-autopsy procedure. Placements of samples are indicative of organ-of-origin, not actual sample locations

response, and 56 weeks after diagnosis, the patient succumbed to the disease. A rapid autopsy was performed approximately 2 h after death: 28 metastatic tumor samples, 14 surrounding normal tissue samples, and 2 normal skin samples were collected (see detailed sample descriptions in Additional file 2: Table S1). We were also able to acquire formalin-fixed paraffin-embedded (FFPE) primary breast tumor samples collected at diagnosis (BrP) and at mastectomy (BrM). All tumor samples and 2 normal skin samples were subjected to whole-genome sequencing (WGS). All samples except the FFPE samples and the two normal skin samples were also subjected to bulk RNA sequencing. We hypothesized that genomic and transcriptomic analysis of this unprecedented

collection of 46 biopsy samples from a single patient would provide the ability to reconstruct the evolutionary course of this aggressive metastatic cancer. According to the wishes of the patient and family, to honor the contribution of the patient, and after IRB approval, we named this study "The Victoria Clark Study."

Genomic characteristics show widespread somatic variants across all tumor sites

We interrogated the WGS data (~60X coverage from the rapid autopsy fresh tissue samples, and ~45X coverage from the FFPE samples, with no discernible quality difference between 60X and 45X samples; see Additional file 1: Fig. S2) with our state-of-the-art multi-sample

tumor data analysis pipeline, identifying inherited and somatically acquired variants including single nucleotide variants (SNVs), short insertions/deletions (INDELs), copy number variations (CNVs), regions of loss-of-heterozygosity (LOH), and chromosomal translocation events. Tumor purity was estimated by FACETS [16]; 28 tumor samples that had >50% tumor content (see Fig. 1B) were selected for subsequent genomic analysis. The patient had no identifiable germline breast cancer predisposition variants, in concordance with earlier clinical testing for *BRCA1* and *BRCA2* germline mutations. However, all tumor samples, including the primary and mastectomy, showed widespread chromosomal aberrations, including CNVs and large regions of LOH (Additional file 1: Fig. S3A). We also found a total of 20,012 somatic variants across all samples. 5149 of these were shared by all samples, including a homozygous *TP53* missense (c.517G>C, p.V173L) SNV, a homozygous *PTEN* frameshift (c.676_697delTCCTCCAATTCAGGACCCACAC, p.S226fs), and a homozygous *RB1* deletion; 11,246 additional variants were shared by at least two samples, and 3617 variants were sample-specific. The average number of variants was 10,648 per sample (range 9975–12,548, see Additional file 2: Table S2). The numbers of variants shared among samples were impacted by the presence of LOH. For example, samples Pa1 and Ln4 retained both copies of chromosome 3 whereas samples Ln3, Ly1, and Bn4 lost one copy. All somatic variants on the lost chromosomal copy are therefore absent in samples Ln3, Ly1, and Bn4 (Additional file 1: Fig. S3A). A high fraction, on average 80% per sample, was already present in the primary tumor BrP, i.e., these were truncal variants. 18% were shared variants with at least one other sample, and the remaining 2% were sample-specific variants.

Chromosomal changes suggest four distinct waves of metastatic colonization

Based on the similarity of their copy number profiles (Additional file 1: Fig. S3B), we were able to cluster 23 of the 28 tumor samples into 4 distinct groups (the shared CNVs in each group are highlighted in red boxes in Fig. 2A). Samples in each group fell into almost perfectly delineated organ groups within the body (G1: abdominal organs, G2: lymph nodes, G3: brain and bones, and G4: lymph nodes). Notably, every group also contains lung sites. Each group signifies a distinct wave of metastatic colonization, as samples in the same group share a common genetic origin. The CNV-based grouping was confirmed by detailed structural variant (SV) analysis in which we identified the exact deletion and amplification breakpoints and the specific deleted or amplified alleles (Fig. 2B).

We then attempted to reconstruct the time order of these metastatic waves. In contrast to longitudinally sampled cancer genomes with an inherent time course, rapid autopsy datasets are collected at a single time point. Therefore, the time order in which these sites were established must be inferred from the data. Here, we were able to use the observed chromosomal changes to infer partial time ordering: samples in G1 all have one chromosome 3 allele amplified, while still retaining the second allele; whereas samples in groups G2–G4 have two copies of the first allele but lost the second allele (Fig. 2C, Additional file 1: Fig. S12). This indicates that tumor sites in G1 were seeded by an earlier tumor subclone than sites in the other groups; therefore, G1 was likely the first metastatic wave. Furthermore, the breast biopsy collected at mastectomy (sample BrM) consists entirely of the G2–G4 genotype, whereas the primary breast tumor (BrP) is a mixture of cells, containing both the earlier and the later genotypes (Fig. 2C). This indicates that a clonal sweep of the G2–G4 lineage occurred in the primary tumor after a G1 precursor escaped. CNV data alone was insufficient to determine the relative time order of the three later waves (G2–G4) and the relationship between the four groups and three lung metastasis samples (Ln1, Ln9, and Ln7) which have distinct copy number profiles and therefore could not be placed in any group (Fig. 2A).

Phylogenetic analysis based on somatic tumor SNVs and short INDELs confirms and refines the four metastatic waves

To further resolve the evolutionary trajectory of the tumor, we constructed a phylogenetic tree among the tumor biopsy samples using somatic SNVs and INDELs (see the "Methods" section). To avoid any confounding effects of CNV and/or LOH on calculating evolutionary distances between samples (e.g., deletions/LOH events cause samples to lose acquired somatic alleles that reside on the deleted chromosome, resulting in a falsely reduced evolutionary distance), we restricted our analysis to somatic variants on chromosome 14 (Additional file 1: Fig. S13A), the only chromosome in our dataset that remained copy number neutral and unaffected by LOH across all tumor samples (Additional file 1: Fig. S3). Figure 3A shows the inferred phylogenetic relationship across all samples.

The sample phylogeny derived from chromosome 14 somatic variants is concordant with the CNV-based grouping, confirming that G1 was seeded by an earlier ancestor, and that the remaining three groups (G2–G4) share a common ancestor. These data also allow the placement of two additional lung samples (Ln7 and Ln9) onto the phylogenetic tree. Based on this phylogeny, most of the tumor evolution (i.e., 85%, as measured by

Fig. 2 (See legend on next page.)
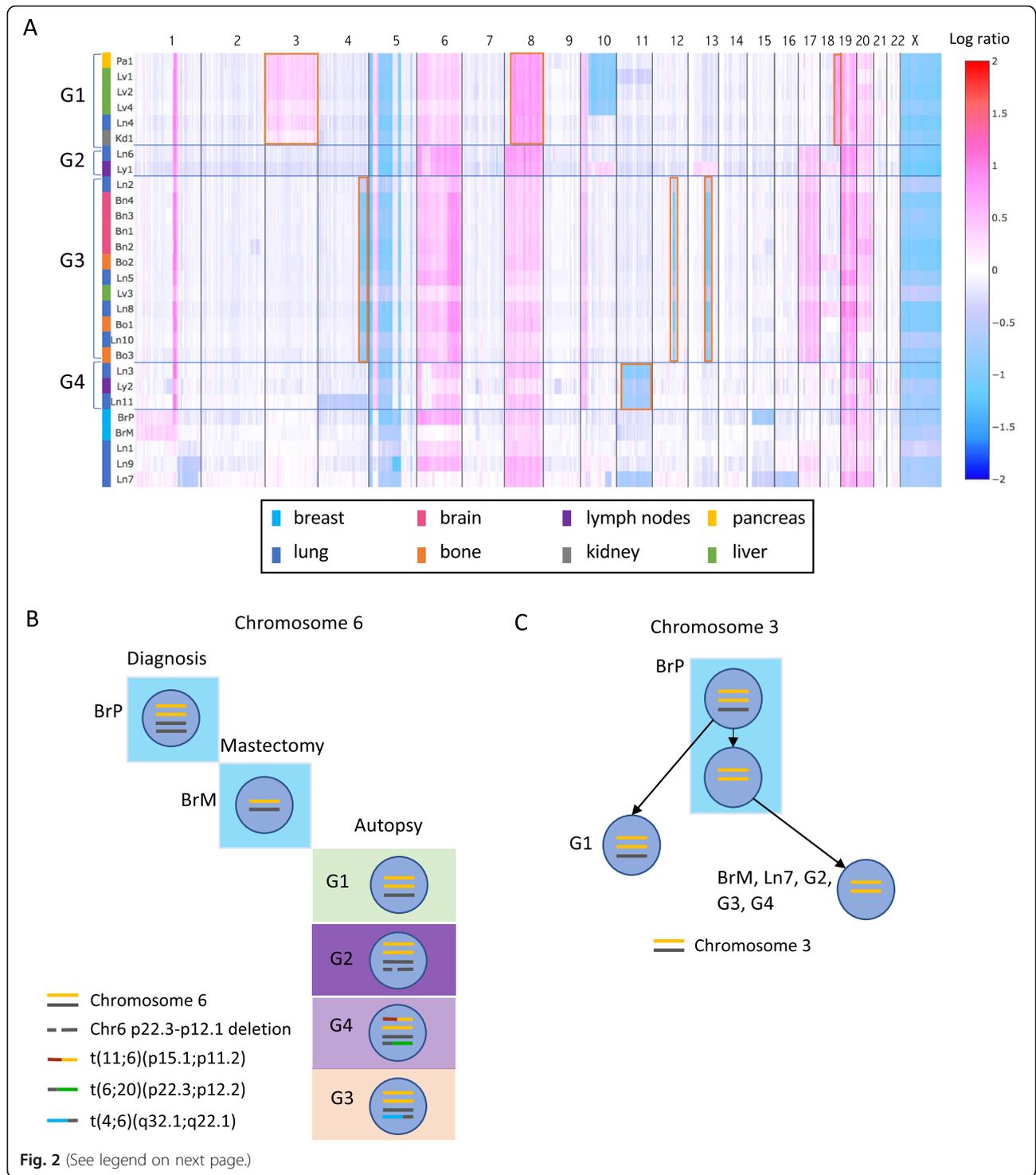
(See figure on previous page.)
Fig. 2 Genomic CNV profiles across samples reveal metastatic waves. **A** Heatmap of copy number profile across all samples. Chromosomal coordinates are on the x-axis, and samples are plotted along the y-axis. Colors in the heatmap represent the $\log_2$ ratio of tumor copy number to normal copy number. Samples with similar CNV profiles are grouped together as G1, G2, G3, and G4. The shared CNVs in each group are highlighted in red boxes. Colored bars beside the sample names represent the host organs: lung(blue), bone(orange), brain(pink), breast(cyan), liver(green), kidney(grey), and pancreas(yellow). **B** An example of samples in the same group sharing complicated structural variants on chromosome 6, further confirming CNV-based grouping. Yellow and black lines are the two alleles of chromosome 6. A deletion occurred between 6p22.3 and 6p12.1 on one black allele in G2. A translocation occurred between chromosome 11 and one yellow allele of chromosome 6, designated as t(11;6)(p15.1;p11.2) (red and yellow line), and a translocation occurred between one black allele of chromosome 6 and chromosome 20, designated as t(6;20)(p22.3;p12.2) (black and green line), in G4. A translocation occurred between chromosome 4 and one black allele of chromosome 6, designated as t(4;6)(q32.1;q22.1) (blue and black line), in G3. **C** Copy number changes on chromosome 3 in different groups and inferred evolution. Yellow and black lines are two alleles of chromosome 3. G1, G2, G3, and G4 represent samples in each group. Each blue circle represents a cell population with certain chromosomal features inside

the accumulation of somatic mutations, Additional file 1: Fig. S13B) had already taken place in the primary tumor, before diagnosis. Once the cancer was able to metastasize; however, it rapidly spread and led to the patient's death. Three samples (BrP, BrM, and Ln1) share mutations with samples across different groups and therefore cannot be placed concordantly as nodes on the phylogenetic tree. Indeed, these samples span multiple branches of the phylogenetic tree (see the colored shade on Fig. 3A) and are likely to be mixtures of cell populations (i.e., tumor subclones) also present in other samples, a conclusion consistent with the mixed CNV profile observed above for these three samples (Fig. 2A). In addition, we refined phylogenetic tree structures in each group using many additional somatic variants on copy number-neutral but LOH chromosomes (chromosomes 2, 7, 9, 11, 15, 16, 21, and 22) (Additional file 1: Fig. S13C).

Subclone-level analysis elucidates the main patterns of metastatic evolution in the patient

To understand the composition of these three heterogeneous sites at the subclonal resolution, as well as the time ordering and trajectory of the major events that were crucial for this patient's metastasis, we carried out subclonal analysis based on the variant allele frequencies of somatic events. In addition to somatic variants on copy number-invariant chromosome 14, we were able to use many additional somatic mutation events on copy number-neutral but LOH chromosomes (chromosomes 2, 7, 9, 11, 15, 16, 21, and 22), in order to refine group-specific subclone structure. Each subclone in this reconstruction is defined, on average, by 33 mutations (see the "Methods" section; Additional file 1: Fig. S6-10; Additional file 2: Table S3-6). This analysis revealed the complex subclonal composition of the primary and mastectomy breast tumors (BrP and BrM) as well as lung metastasis Ln1 and elucidated the critical role that the subclones present in these key, but heterogeneous samples, played in the metastatic process within our patient (Fig. 3B, blue circles labeled with numbers

represent the corresponding subclones). Notably, lung metastasis Ln1 was the site of the initial metastatic escape from the breast. This site contains subclone Sc8 that was derived from primary subclone Sc4 (Additional file 1: Fig. S6A). However, neither Sc4 nor any of its descendant subclones are present in the mastectomy (sample BrM), indicating that lung site Ln1 had likely already been colonized before the mastectomy procedure took place (Additional file 1: Fig. S6A). Furthermore, although primary subclone Sc1 is inferred by our subclonal analysis, this ancestral subclone is no longer present at the primary site (BrP) at the time and location of resection (the observed subclone frequency of Sc1 is zero) (Additional file 1: Fig. S6D). This indicates that metastatic site Ln1 was established even earlier, i.e., before the time of the patient's diagnosis and primary tumor resection. Our analysis also revealed that all four metastatic waves in the patient were seeded by subclones from lung site Ln1: subclone Sc8 gave rise to the first metastatic wave (group G1) (Additional file 1: Fig. S6A); and Sc9 to the three later metastatic waves (G2–G4) (Additional file 1: Fig. S6C). These observations establish lung site Ln1 as a "jumping board" for all subsequent metastatic spread in this patient (Fig. 3C).

Previous studies have reported examples of monoclonal and polyclonal seeding in breast cancer patients [7, 8, 33]. We observed both patterns in our patient (Fig. 4). As the most striking example of *monoclonal seeding*, multiple subclones from lung site Ln10 seeded as many as 9 metastases, primarily in the brain and bones (Fig. 4). Subclone Sc17 colonized 5 distinct sites, including all four metastases in the brain (Fig. 4; Additional file 1: Fig. S9B, C). The lack of additional variants across these sites suggests that they were formed within a very short time period (in contrast to Sc18-21, assuming a similar variant acquisition rate across sites). Conversely, subclones Sc18-Sc21 each seeded a single metastatic site (Fig. 4; Additional file 1: Fig. S9B). These subclones evolved from each other by accumulating mutations gradually and colonizing additional sites in a stepwise manner. This pattern demonstrates that lung site Ln10 acted as a

Huang *et al. Genome Medicine*   (2021) 13:170

Page 10 of 18
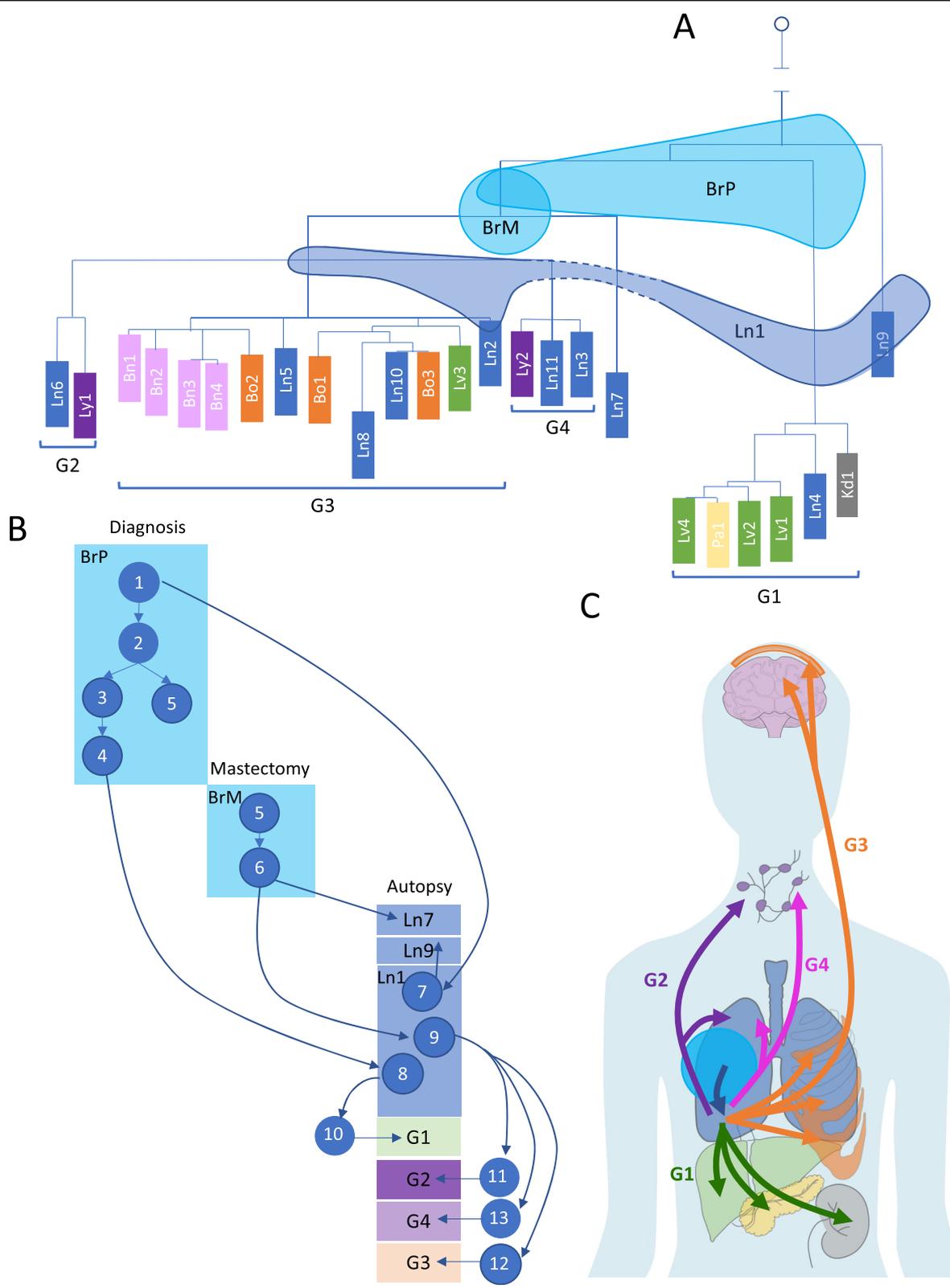
Fig. 3 (See legend on next page.)

(See figure on previous page.)

Fig. 3 Short variants reveal high-resolution phylogenetic relationships among all tumor biopsies, and early subclonal expansion and migration events. **A** Reconstructed phylogenetic tree based on short variants confirms the grouping of the samples and reveals more detailed relationships among samples in groups as well as the placement of Ln7 and Ln9. BrP, BrM, and Ln1 shared variants with different samples, indicating that they are mixtures of cell populations. Three shades spanning multiple branches represent samples BrP, BrM, and Ln1. **B** Reconstructed subclone structures in BrP, BrM, and Ln1 show the heterogeneity of these three samples. The descendant relationship among the subclones suggests the earliest invasion was to the lung and reveal other subclone migrations and expansions from BrP. Each blue circle represents an inferred subclone. Each box represents a sample or a group of samples. **C** Overall migration patterns across all tumor sites, with arrow colors corresponding to metastatic waves. After the first invasion from primary breast tumor to lung (blue arrow), four metastatic waves spread the tumors to other organs

subclonal incubator, in which subclones were able to evolve before colonizing consecutive sites (Fig. 4). We also found examples of *polyclonal seeding* (Fig. 4, green arrows): subclones Sc1 and Sc4 from the primary tumor colonizing lung site Ln1 (Additional file 1: Fig. S6) and subclones Sc24 and Sc26 from pancreas site Pa1 colonizing liver site Lv4 (Additional file 1: Fig. S7). The important distinction between these two polyclonal seeding events is that the former represents *primary tumor to metastasis* seeding, whereas the latter is a *metastasis-to-metastasis* event. It is important to note that in this patient, the vast majority of the metastatic colonization events fall into this latter category, i.e., originated from an already metastatic site, rather than directly from the

primary tumor, consistent with recent findings in other metastatic cancers [8, 13, 34]. Finally, *metastatic recolonization* of an existing tumor site has been noted in a cell line engrafted mouse model [35]. We observed such a recolonization event in this patient (Fig. 4, red arrow): a subclone (Sc9) and its further evolved descendent (Sc16) are observed at a single site (Ln1). However, subclones Sc12 and Sc14, representing intermediary evolutionary steps between Sc9 and Sc16, are found at a different lung site, Ln2 (Fig. 4, G3 Ln2; Additional file 1: Fig. S9A), and at that site only. The most parsimonious explanation for this observation is that, after evolving at site Ln2, subclone Sc16 invaded, i.e., "recolonized" site Ln1, an already established metastatic site.

Fig. 4 Subclone evolution and migration across all samples. Each blue circle represents an inferred subclone. Each box represents a sample or a group of samples. Samples in the same group are shown in a group box-labeled G1–G4. Blue arrows represent subclone evolution or monoclonal seeding. Green arrows represent polyclonal seeding (Sc1, Sc4, and Sc6 seeded Ln1; Sc26 and Sc24 seeded Lv4). Red arrow represents recolonization (Sc16 from Ln2 recolonized Ln1). Ln1 served as the "jumping board" where Sc7 and Sc8, descendants of Sc1 and Sc4 from BrP, and Sc9, descendant of Sc6 from BrM, were all attracted to stay, further evolved, and then colonized other samples. In G3, Ln10 served as an "incubator" where many subclones evolved and went on to seed other sites

Huang *et al. Genome Medicine*   (2021) 13:170

Page 12 of 18

Comprehensive sampling of metastatic tumors is important for understanding subclonal dynamics across sites

Our comprehensive sampling of multiple metastatic sites affords us the ability to evaluate the contribution of each sample to the biological conclusions. 20 of 28 total tumor sites are phylogenetic "leaves" that give rise to no further sites. Were any one of these sites left out of the study, our conclusions would be minimally affected. However, leaving out one of the more complex sites for which our analysis inferred a more critical role would have a larger effect. For example, had we not been able to collect a biopsy at lung site Ln1 (Additional file 1: *Fig. S14A*), we would not have observed a metastatic recolonization event or been able to identify the lung as the critical first site of metastatic escape. Without this site, we would have concluded that the primary tumor site (BrP) gave rise directly to the first metastatic wave (G1) and that the mastectomy site (BrM) gave rise to the 3 later metastatic waves (G2–G4). As a result, substantial insight into the role of the lung in the metastatic spread in this patient would have been lost. Unavailability of sample Ln2 would have resulted in the loss of our ability to infer the recolonization event (Additional file 1: Fig. S14B). Losing sample Ln10 would have obfuscated the identification of the "incubator effect" in the lung. Unavailability of either sample Lv4 or Pa1 would have resulted in the loss of observation of a polyclonal seeding event in G1 (Additional file 1: Fig. S14C). This highlights the importance of comprehensive sampling of metastatic sites for studies aimed at elucidating subclonal dynamics in metastatic expansion.

Functional annotation of somatic variants explains the aggressive metastatic disease observed in the patient and provides a window into organ group-specific metastasis

To assess the driver mechanisms for tumorigenesis and metastasis, we interrogated the overall mutational signatures, variants on well-known oncogenes and tumor suppressors, variants on genes that have been reported to be involved in metastasis, and metastatic sample and group-specific variants.

We first examined the relative contributions of COSMIC mutational signatures at different stages of tumor evolution (Fig. 5A). We observed that signatures 1, 3, 5, and 8 were consistently present throughout the tumor's evolution, suggesting that the mutational processes that caused these signatures were continuously present throughout the disease progression. Signatures 1 and 5 are correlated with age of diagnosis, and signatures 3 and 8 are associated with homologous recombination deficiency (HRD). Consistent with the previous studies [36, 37], TP53-mutated, relatively late diagnosis, TNBC patients had enrichment in signature 3. The dominant

signatures during the early evolution of the tumor (i.e., those associated with truncal variants, present at each tumor site) point overwhelmingly to APOBEC activity (signatures 2 and 13). Strikingly, these signatures were almost completely absent during the later stages of tumor evolution, pointing to cessation/attenuation of APOBEC activity at metastatic sites, in contrast with the general pattern that APOBEC mutational signature increases in prevalence during the course of tumor evolution [12, 38, 39]. Nevertheless, this phenomenon has been observed in the Yates et al. study, wherein one of the studied patients' tumors, APOBEC activity-related mutational signatures, was present early on but was later turned off [40]. Signature 18 was found to be only present in group-shared or group-specific variants, but not in the truncal variants (Fig. 5A). Signature 18 was previously reported to be associated with oxidative DNA damage due to reactive oxygen species (ROS) [41, 42], which can be induced by gamma-radiation [43]. A recent study [44] also showed that gamma-radiation can induce mutations linked to signature 18. The high prevalence of signature 18 that presented in the late tumor evolution may be due to the radiation therapy this patient received after the mastectomy.

Second, we tried to identify driver mutations. A homozygous *TP53* missense (c.517G>C, p.V173L) SNV, a homozygous *PTEN* frameshift (c.676_697delTCCTCCAATTCAGGACCCACAC, p.S226fs), and a homozygous *RB1* deletion occurred in the founding subclone, indicating that these mutations were major primary driver mutations in this patient's tumor. Then, we focused on variants in 127 genes (Additional file 2: Table S7) that have been reported as metastasis-related [45–47]. We found that this patient had somatically acquired deletions and amplifications in 17 of these genes (Fig. 5B); 8 of the genomic lesions were present in the primary tumor, including the *MTDH, ANGPTL4*, and *ID1* genes associated with lung metastasis [46, 48]; and the *IL11* and *TGFB1* genes in which genomic alterations have been associated with bone metastasis [46, 49].

Third, we identified variants in those metastasis-related genes (Additional file 2: Table S7) that may contribute to sample- or group-specific tumor cell survival and expansion (Fig. 5B). Three of the previously reported metastasis-related genes (*CTNNB1, PIK3CA, PIK3CB*) were somatically altered only in group G1 samples, as well as in the G1-progenitor subclones within samples BrP and Ln1; these genes are therefore candidates for promoting abdominal organ-specific metastasis.

We speculate that some of the nonsynonymous coding mutations that were shared among all samples in the same group, albeit without well-established links to tumorigenesis and metastasis, may also contribute to

Fig. 5 Temporal and spatial distribution of mutational signatures, mutated metastasis-related genes, and clinically actionable genes. **A** Genome-wide mutational signatures exhibited at different stages of tumor evolution, annotated to the phylogenetic tree. The number of variants is labeled beside each pie chart. HRD homologous recombination deficiency. ROS reactive oxygen species. **B** Distribution of metastasis-related mutations across genes and samples. Genes that are involved in lung metastasis labeled L; genes that are involved in bone metastasis labeled B. **C** Distribution and type of potentially clinically informative mutations across genes and samples. Gene mutations are shown in yellow, deletions in blue, and amplifications in red

group-specific metastasis (genes *HSPG2* and *ISM1* are mutated in samples in G1, *IQGAP2* and *RUNX1T1* in G2, *SPINK5* and *FGA* in G3, and *PCDH20* in G4) (Additional file 2: Table S8; and Additional file 1: Fig. S15). More functional studies are needed, however, to critically evaluate these findings in the context of tumor development.

Transcriptomic profiling confirms the genomic observation of distinct metastatic waves and identifies group-specific phenotypes

To gain insight into the evolution of the patient's tumor on a molecular/phenotypic level, and thus to complement genomic findings, we collected bulk RNA-seq data from the 28 rapid-autopsy tumor biopsies and 14 surrounding normal tissue samples. First, CNVs were

Huang *et al. Genome Medicine* (2021) 13:170

Page 14 of 18

inferred from RNA-seq data by averaging expressions across 101-gene windows and normalized to the expression of 14 normal samples (see the "Methods" section, Additional file 1: Fig. S16A). 20 tumor samples showed significant copy number changes and recapitulated the presence of large CNVs from the WGS data (major CN events from WGS data were highlighted in dotted boxes in Additional file 1: Fig. S16A). The remaining 8 tumor samples showed no copy number changes, which was likely due to heavy normal tissue contamination. Indeed, PCA analysis and unsupervised k-means clustering concordantly showed that the samples without CNV events clustered together with the normal samples (Additional file 1: Fig. S16B); thus, we excluded these 8 samples from subsequent analysis. Then, we explored what factors played roles in shaping the transcriptomes of the metastases. Unsupervised clustering on 20 metastatic tumor samples showed that although some samples clustered along with their genomic groups (e.g., Ly1 and Ln6, both belonging to G2, were clustered together), and some samples clustered along with their host tissue types (e.g., most of the lung metastases clustered together; all three brain metastases clustered together), neither genomic grouping nor host tissue types were the sole factor that affected the expression (Additional file 1: Fig. S17A). On the one hand, via clustering samples located in the lung and liver respectively because these two organs hosted multiple metastases from different genomic groups, we found that samples in the same genomic group were in fact clustered together (e.g., Ln5, Ln8, and Ln10, which all belong to G3, were more similar with each other than with Ln6 of G2 and Ln3 of G4; Lv1, Lv2, and Lv4, which all belong to G1, were more similar with each other than with Lv3 of G3; Additional file 1: Fig. S17B,C). On the other hand, we clustered the samples that belong to G3, which contained the most diverse host organ types, and found that samples located in the same organ tend to cluster together (Additional file 1: Fig. S17D). This indicates that the environment of the host tissue can shape the transcriptomes of the subclones from the different genomic lineages after these subclones landed and made them adaptive to the respective organs. Next, we compared the gene expression among metastatic samples that belong to different genomic groups. Differential expression (DE) analysis showed that there were 415 genes that were significantly different among the groups (ANODEV test FDR<0.05, logFC>2). Particularly, genes involved in beta1 integrin cell surface interactions (*FGB, FGA,* and *F13A1, VTN, MDK*) were upregulated in G1; genes involved in Wnt signaling network (*FZD8, FZD10, DKK1*) that upregulated in G2; and genes involved in integrin family cell surface interactions (*ITGA10, ITGB7*) and in estrogen receptor alpha network (*ESR2* and *GREB1*) were upregulated in G4. Genes that are upregulated in G3 are involved in multiple mechanisms: *RAB40B* can promote tumor cell invasion by regulating trafficking MM2 and MM9 during invadopodia formation [50, 51]; *RET* activation can drive signaling through MAPK and PI3K pathways [52]; *ALDH1A1* is a marker for cancer cell stemness [53]; and *ST6GALNAC1* encodes the protein in the same family member as *ST6GALNAC5* which has been reported that can mediate infiltration into the brain [54]. Considering the samples in G3 located in multiple tissues, these genes may increase the fitness and invasiveness of samples in G3 to be able to colonize multiple distal organs. This result showed that subclones from different genomic groups probably rely on different survival mechanisms/strategies. Gene set enrichment analysis (GSEA) showed that copy number is the dominant factor in G1 and G3 groups (e.g., gene sets NIKOLSKY_BREAST_CANCER_17Q11_Q21_AMPLICON and 17Q21_Q25_AMPLICON were enriched in samples in G3; gene sets 8Q12_Q22_AMPLICON and 8Q23_Q24 were enriched in samples in G1) and consistent with the underlying genetic alterations.

Delineation of tumor evolution suggests alternative treatment strategies

We identified all somatic variants (CNVs and short variants) in our tumor samples that impacted a gene with known and clinically targetable mutations in the TARGET database [55, 56]. This allowed us to evaluate genetic information that may have impacted the patient's treatment had it been available to the treating oncologist while the patient was still alive. We found that 11 of 20 (Additional file 2: Table S9) potentially targetable alterations were already present in the primary tumor at diagnosis and were subsequently retained in all metastatic samples (Fig. 5C). These alterations include *RB1* loss, *APC* loss, *MYC* amplification, and *AKT2* amplification, as well as alterations involving genes with high cancer patient population frequencies in the *METABRIC* dataset [57], i.e., *FGFR1* amplification (14% patient population frequency) and *AURKA* amplification (6% patient population frequency). These variants would have been the optimal targets for therapy to impact all metastatic sites in the patient. 5 of the 20 were potentially targetable, sample- or sample group-specific alterations: *MEN1* loss is present only at lung site Ln7 and in G4 samples; *CTNNB1, PIK3CA, PIK3CB,* and *RAF1* are amplified only in G1 samples as well as in the samples containing their progenitor subclones (i.e., BrP, Ln1) (Fig. 5C). Because these variants are only present in some but not all metastatic sites, therapies targeting these genes would likely have been ineffective. This observation underlines the necessity of comprehensive
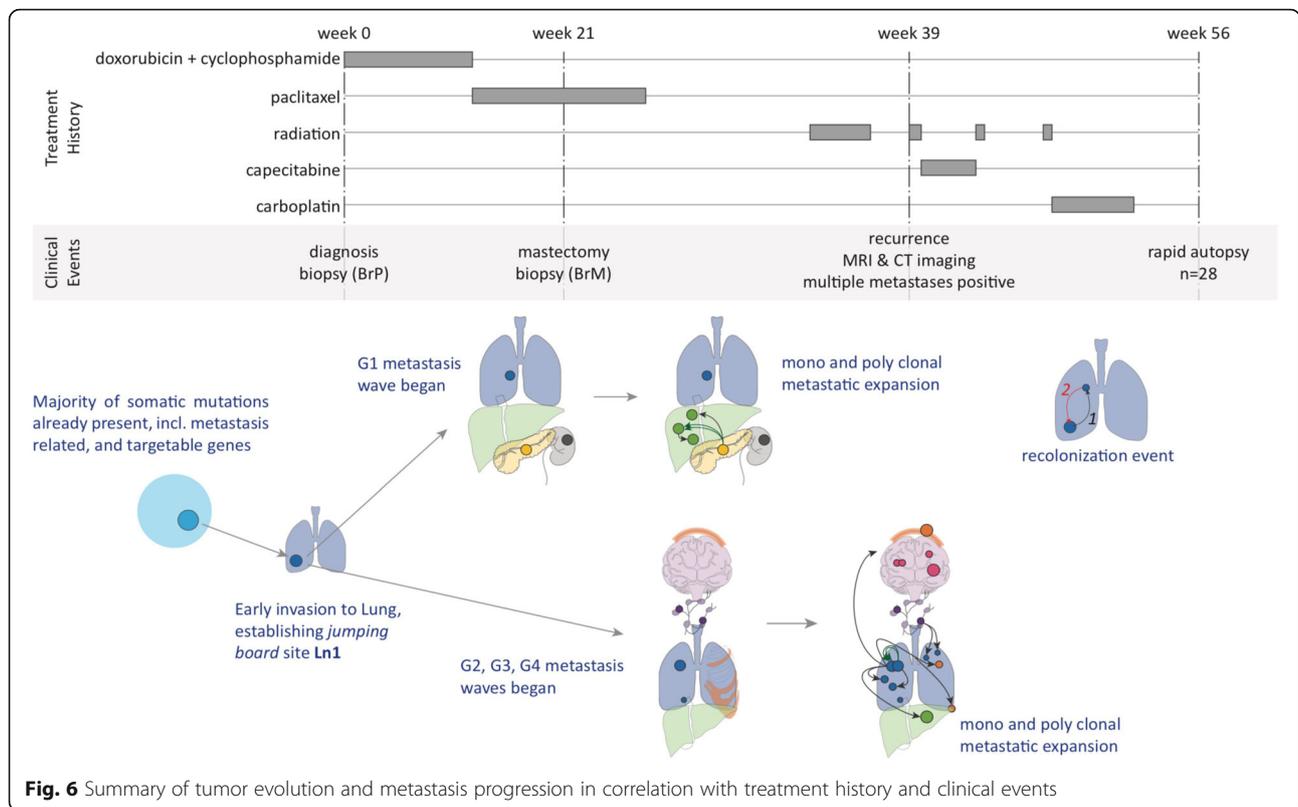
Fig. 6 Summary of tumor evolution and metastasis progression in correlation with treatment history and clinical events

monitoring of metastatic sites for effective therapeutic intervention.

Discussion

The 46 biopsy/autopsy samples in this dataset, the largest number from a single patient to date, allowed us to track the evolution of this metastatic tumor genome at the subclonal resolution, as it spread from the breast to 7 additional organs (Fig. 6). Reconstruction of the evolutionary trajectory of the tumor revealed 4 distinct waves of metastatic colonization, targeting well-delineated groups of organs in the patient.

Our study reveals the existence of "metastatic incubators" where subclones that initially colonized the sites further evolved into founding clones of additional sites. Three independent lines of evidence support our conclusion that these subclones evolved at metastatic sites, rather than that they were already present in the primary. *First*, we observed the evolutionary intermediate clones at lung sites Ln1 (Sc8, Sc9), Ln2 (Sc14-Sc16), and Ln10 (Sc17-Sc21), whereas we did not find any of these intermediate subclones in the breast samples (i.e., in BrP or BrM) or any other sample in our collection. This suggests in situ evolution at, rather than migration from the primary to, those sites. We note that the same argument holds true for other organ sites where evolutionary intermediaries were observed, i.e., liver site Lv4 (Sc24), lymph

node site Ly2 (Sc27), and pancreas site Pa1 (Sc24). *Second*, the evolutionary intermediate subclones are not defined by one or two, but by multiple somatic variants (Additional file 2: Table S3-6). The absence of these variants in BrP and BrM suggests that the intermediate subclones were absent from the breast samples. Previous breast cancer studies [6, 8, 9] observed mutations not found in the primary tumor but shared between multiple metastatic sites and suggest the presence of an already metastatic, common ancestor subclone seeding those sites. Our study provides direct, high-resolution evidence for the presence of such common ancestor clones, and elucidates the extent of subclonal evolution that occurred at these sites. *Third*, COSMIC mutational signature analysis shows the prominent presence of radiation-associated signature 18 in mutations shared across metastatic samples, and its complete absence from the primary tumor (see Additional file 1: Fig. S18). According to the patient's treatment history, radiation therapy was applied after a double mastectomy. This implies that these radiation-associated mutations were acquired after some early metastasis had already occurred, i.e., at "incubator" sites Ln10, Lv4, and Ly2. Subclones evolved at these sites during radiotherapy and then seeded multiple other metastatic sites.

Our study highlights the critical role of the lung in the metastatic process of our patient. The lung was not only

the first site of metastatic escape, but also an organ with multiple metastatic incubator sites. A larger TNBC metastasis cohort with a similar genomic resolution is needed to validate that the lung is a metastatic incubator hotspot or to identify other incubator hotspot organs. Indeed, other studies point to the presence of such metastatic incubators in other cancer types. For example, in a study of metastatic colorectal cancer, Angelova et al. showed that a metastatic site (M10) in the liver was established as a result of multi-step colonization and that metastatic sites in the liver seeded additional sites in the patient's other organs [58]. Identification of these "metastatic incubators" may therefore be critical for effective clinical management of metastatic patients.

The presence of somatic alterations in genes associated with lung metastasis in the primary tumor is consistent with our subclonal migration results (see Sections 4 and 5) indicating that the lung was the first organ of metastasis. The scenarios for distal organs are more complicated. On the one hand, the presence of somatically altered genes associated with bone metastasis in the primary tumor signals that the tumor's ability to metastasize to the bone may, in whole or in part, be already there in the primary tumor. On the other hand, our subclone analysis revealed that all distal organ metastasis sites outside the lung had originated not directly from the primary site, but from lung metastasis sites by new subclones harboring additional mutations. Taken together, these observations suggest that the rampant metastatic invasion was the compound result of early mutations in the primary tumor and later evolution at metastatic sites in the lung.

Once metastasis was established, resection of the primary (breast) site was ineffective against the progression of this aggressive disease. This reinforces the Fisher hypothesis that breast cancer is a systemic disease in which metastasis occurs before diagnosis [59]. Although many potentially targetable mutations were already present at the primary site, and therefore at all tumor sites in the patient, others were only present at a subset of metastatic sites. This highlights the need for more comprehensive sampling of metastatic sites: indeed, targeted treatment strategies identified from assays of the primary site, or only one or two additional metastatic sites, may fail to achieve the desired therapeutic outcome [60]. Therefore, a more comprehensive survey of metastases (or liquid biopsy-based approaches [6, 61] if direct sampling of additional metastatic sites is not practicable) may be necessary for guiding more successful clinical interventions.

This study lays the methodological foundations for tracking a patient's metastases at subclonal resolution. With additional, similarly comprehensive datasets in

hand, it will be possible to assess the generality of these findings and to establish how the specific evolutionary patterns observed in a patient's metastatic tumor evolution can, in the future, inform more effective personalized treatment.

Conclusions

This study reports on a one-of-a-kind triple-negative breast cancer (TNBC) metastatic tumor dataset. This unique dataset allowed us to track the patient's tumor across metastatic colonization in eight distinct organs, at subclonal resolution. The most striking finding of this study is the observation of "metastatic incubators," where substantial subclonal evolution preceded, and may have been necessary for, further metastatic colonization. We found that the lung played a crucial role in metastatic tumor spread in our patient, a finding that must be examined in larger patient cohorts. Our results showcase the level of detail achievable for reconstructing metastatic subclonal evolution when sampling a large number of metastatic sites and combining this with deep genome sequence-based subclonal analysis.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13073-021-00989-6.

Additional file 1: Supplementary Figures S1-S18.

Additional file 2: Tables S1-9.

Additional file 3: Subclone analysis-alternative solutions.

Additional file 4: Somatic SNVs with ssID reference.

Additional file 5: Somatic SNVs vcf file.

Additional file 6: Somatic other variants vcf file.

Authors' contributions
X.H. and Y.Q. jointly contributed to the study design, coordination, genomic and transcriptomic data processing and analysis, variant identification, subclone analysis, sample phylogenetic analysis, subclonal migration pattern determination, and manuscript writing. S.W.B. contributed to the study design, coordination, sample collection, sample processing for DNA and RNA sequencing, RNA-seq analysis, and manuscript writing. R.F. and E.D.-K. contributed to the pathology analysis of cancer specimens and assessment

Huang *et al. Genome Medicine*   (2021) 13:170

Page 17 of 18

writing. A.F. contributed to the structural variant identification. J.A.M. and G.S. contributed to the DNA and RNA isolation. D.J. and W.E.J. helped with RNA-seq analysis. A.L.C. contributed to the imaging reviewing, clinical oversight, and manuscript writing. A.H.B. and G.T.M. conceived the project and contributed to the project coordination. G.T.M. contributed to the manuscript writing. The authors contributed to the manuscript refinement. The authors read and approved the final manuscript.

Availability of data and materials
The gene expression data is available at the Gene Expression Omnibus with accession number GSE184717 [62]. The somatic SNVs have been deposited to dbSNP, whose ssIDs can be found in Additional file 4. The somatic SNV and non-SNV variants can be found in Additional file 5 and Additional file 6, respectively. The custom scripts and processed datasets generated and/or analyzed in the study are available in the Github repository [63] https://github.com/xiaomengh/tumor-evo-rapid-autopsy.

Declarations

Ethics approval and consent to participate
The study was reviewed and approved by the human subjects IRB (#41030) of the University of Utah. Informed consent in accordance with the Declaration of Helsinki was obtained from the patient. According to the wishes of the patient and family, to honor the contribution of the patient, and after IRB approval, we named this study "The Victoria Clark Study."

Consent for publication
Written informed consent was obtained for the publication of the patient's genomic data and clinical details.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Utah Center for Genetic Discovery, University of Utah, Salt Lake City, USA. [2]Department of Human Genetics, School of Medicine, University of Utah, Salt Lake City, USA. [3]Department of Pharmacology and Toxicology, College of Pharmacy, University of Utah, Salt Lake City, USA. [4]Department of Biomedical Informatics, School of Medicine, University of Utah, Salt Lake City, USA. [5]Department of Pathology, School of Medicine, University of Utah, Salt Lake City, USA. [6]Department of Oncological Sciences, School of Medicine, University of Utah, Salt Lake City, USA. [7]Computational Biomedicine, Department of Medicine, Boston University, Boston, USA. [8]Department of Internal Medicine, Huntsman Cancer Institute, University of Utah, Salt Lake City, USA. [9]Department of Medical Oncology & Therapeutics Research, City of Hope, Duarte, USA.

References
1. Caswell-Jin JL, Plevritis SK, Tian L, Cadham CJ, Xu C, Stout NK, et al. Change in survival in metastatic breast cancer with treatment advances: meta-analysis and systematic review. JNCI Cancer Spectr. 2018;2(4):ky062.
2. Mariotto AB, Etzioni R, Hurlbert M, Penberthy L, Mayer M. Estimation of the number of women living with metastatic breast cancer in the United States. Cancer Epidemiol Biomarkers Prev. 2017;26(6):809–15. https://doi.org/10.1158/1055-9965.EPI-16-0889.
3. Network CGA. Comprehensive molecular portraits of human breast tumours. Nature. 2012 Oct 4;490(7418):61–70. https://doi.org/10.1038/nature11412.
4. Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. Nature. 2012;486(7403):405–9. https://doi.org/10.1038/nature11154.
5. Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, et al. Comprehensive molecular portraits of invasive lobular breast cancer. Cell. 2015;163(2):506–19. https://doi.org/10.1016/j.cell.2015.09.033.
6. Murtaza M, Dawson S-J, Pogrebniak K, Rueda OM, Provenzano E, Grant J, et al. Multifocal clonal evolution characterized using circulating tumour

DNA in a case of metastatic breast cancer. Nat Commun. 2015;6(1):8760. https://doi.org/10.1038/ncomms9760.
7. Hoadley KA, Siegel MB, Kanchi KL, Miller CA, Ding L, Zhao W, et al. Tumor evolution in two patients with basal-like breast cancer: a retrospective genomics study of multiple metastases. PLoS Med. 2016;13(12):e1002174. https://doi.org/10.1371/journal.pmed.1002174.
8. Savas P, Teo ZL, Lefevre C, Flensburg C, Caramia F, Alsop K, et al. The subclonal architecture of metastatic breast cancer: results from a prospective community-based rapid autopsy program "CASCADE". PLoS Med. 2016 Dec;13(12):e1002204. https://doi.org/10.1371/journal.pmed.1002204.
9. De Mattos-Arruda L, Sammut S-J, Ross EM, Bashford-Rogers R, Greenstein E, Markus H, et al. The genomic and immune landscapes of lethal metastatic breast cancer. Cell Rep. 2019;27(9):2690–708.e10.
10. Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med. 2012;366(10):883–92. https://doi.org/10.1056/NEJMoa1113205.
11. Zhang J, Fujimoto J, Zhang J, Wedge DC, Song X, Zhang J, et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. Science. 2014;346(6206):256–9. https://doi.org/10.1126/science.1256930.
12. de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, Yates L, et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. Science. 2014;346(6206):251–6. https://doi.org/10.1126/science.1253462.
13. Gundem G, Van Loo P, Kremeyer B, Alexandrov LB, Tubio JMC, Papaemmanuil E, et al. The evolutionary history of lethal metastatic prostate cancer. Nature. 2015;520(7547):353–7. https://doi.org/10.1038/nature14347.
14. Schwarz RF, Ng CKY, Cooke SL, Newman S, Temple J, Piskorz AM, et al. Spatial and temporal heterogeneity in high-grade serous ovarian cancer: a phylogenetic analysis. PLoS Med. 2015 Feb;12(2):e1001789. https://doi.org/10.1371/journal.pmed.1001789.
15. Church DM, Schneider VA, Graves T, Auger K, Cunningham F, Bouk N, et al. Modernizing reference genome assemblies. PLoS Biol. 2011 Jul;9(7):e1001091. https://doi.org/10.1371/journal.pbio.1001091.
16. Shen R, Seshan VE. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. Nucleic Acids Res. 2016;44(16):e131. https://doi.org/10.1093/nar/gkw520.
17. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178–92. https://doi.org/10.1093/bib/bbs017.
18. Farrell JAR. Expanding the horizons of next generation sequencing with RUFUS [Internet]. Boston College; 2014. Available from: https://dlib.bc.edu/islandora/object/bc-ir:104176/datastream/PDF/download/citation.pdf
19. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. Genome Biol. 2014;15(6):R84. https://doi.org/10.1186/gb-2014-15-6-r84.
20. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24–6. https://doi.org/10.1038/nbt.1754.
21. El-Kebir M, Satas G, Oesper L, Raphael BJ. Inferring the mutational history of a tumor using multi-state perfect phylogeny mixtures. Cell Syst. 2016;3(1):43–53. https://doi.org/10.1016/j.cels.2016.07.004.
22. Huang X. SeederSeeker. Github. https://github.com/xiaomengh/tumor-evo-rapid-autopsy/tree/master/Phylogeny. 2021.
23. Qiao Y. SubcloneSeeker v2, https://github.com/yiq/SubcloneSeeker. Github. 2021.
24. Qiao Y, Quinlan AR, Jazaeri AA, Verhaak RG, Wheeler DA, Marth GT. SubcloneSeeker: a computational framework for reconstructing tumor clone structure for cancer variant interpretation and prioritization. Genome Biol. 2014;15(8):443. https://doi.org/10.1186/s13059-014-0443-x.
25. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Med. 2018;10(1):33. https://doi.org/10.1186/s13073-018-0539-0.
26. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. 2013 May 1;41(10):e108. https://doi.org/10.1093/nar/gkt214.
27. Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res. 2019;47(8):e47. https://doi.org/10.1093/nar/gkz114.

28. Tirosh I, Izar B, Prakadan SM, Wadsworth MH 2nd, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science. 2016;352(6282):189–96. https://doi.org/10.1126/science.aad0501.

29. Brady SW, McQuerry JA, Qiao Y, Piccolo SR, Shrestha G, Jenkins DF, et al. Combating subclonal evolution of resistant cancer phenotypes. Nat Commun. 2017;8(1):1231. https://doi.org/10.1038/s41467-017-01174-3.

30. Chen Y, Lun ATL, Smyth GK. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. F1000Res. 2016;5:1438.

31. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016;44(W1):W90–7. https://doi.org/10.1093/nar/gkw377.

32. Wu D, Smyth GK. Camera: a competitive gene set test accounting for inter-gene correlation. Nucleic Acids Res. 2012;40(17):e133. https://doi.org/10.1093/nar/gks461.

33. Brown D, Smeets D, Székely B, Larsimont D, Szász AM, Adnet P-Y, et al. Phylogenetic analysis of metastatic progression in breast cancer using somatic mutations and copy number aberrations. Nat Commun. 2017;8(1):14944. https://doi.org/10.1038/ncomms14944.

34. McPherson A, Roth A, Laks E, Masud T, Bashashati A, Zhang AW, et al. Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. Nat Genet. 2016;48(7):758–67. https://doi.org/10.1038/ng.3573.

35. Kim M-Y, Oskarsson T, Acharyya S, Nguyen DX, Zhang XH-F, Norton L, et al. Tumor self-seeding by circulating cancer cells. Cell. 2009;139(7):1315–26. https://doi.org/10.1016/j.cell.2009.11.025.

36. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415–21. https://doi.org/10.1038/nature12477.

37. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature. 2016;534(7605):47–54. https://doi.org/10.1038/nature17676.

38. McGranahan N, Favero F, de Bruin EC, Birkbak NJ, Szallasi Z, Swanton C. Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. Sci Transl Med. 2015;7(283):283ra54.

39. Gerstung M, Jolly C, Leshchiner I, Dentro SC, Gonzalez S, Rosebrock D, et al. The evolutionary history of 2,658 cancers. Nature. 2020;578(7793):122–8. https://doi.org/10.1038/s41586-019-1907-7.

40. Yates LR, Knappskog S, Wedge D, Farmery JHR, Gonzalez S, Martincorena I, et al. Genomic evolution of breast cancer metastasis and relapse. Cancer Cell. 2017;32(2):169–84.e7.

41. Viel A, Bruselles A, Meccia E, Fornasarig M, Quaia M, Canzonieri V, et al. A specific mutational signature associated with DNA 8-oxoguanine persistence in MUTYH-defective colorectal cancer. EBioMedicine. 2017;20:39–49. https://doi.org/10.1016/j.ebiom.2017.04.022.

42. Pilati C, Shinde J, Alexandrov LB, Assié G, André T, Hélias-Rodzewicz Z, et al. Mutational signature analysis identifies MUTYH deficiency in colorectal cancers and adrenocortical carcinomas. J Pathol. 2017;242(1):10–5. https://doi.org/10.1002/path.4880.

43. Tominaga H, Kodama S, Matsuda N, Suzuki K, Watanabe M. Involvement of reactive oxygen species (ROS) in the induction of genetic instability by radiation. J Radiat Res. 2004;45(2):181–8. https://doi.org/10.1269/jrr.45.181.

44. Rose Li Y, Halliwill KD, Adams CJ, Iyer V, Riva L, Mamunur R, et al. Mutational signatures in tumours induced by high and low energy radiation in Trp53 deficient mice. Nat Commun. 2020;11(1):394. https://doi.org/10.1038/s41467-019-14261-4.

45. Friedl P, Wolf K. Tumour-cell invasion and migration: diversity and escape mechanisms. Nat Rev Cancer. 2003;3(5):362–74. https://doi.org/10.1038/nrc1075.

46. Nguyen DX, Bos PD, Massagué J. Metastasis: from dissemination to organ-specific colonization. Nat Rev Cancer. 2009;9(4):274–84. https://doi.org/10.1038/nrc2622.

47. Nguyen DX, Massagué J. Genetic determinants of cancer metastasis. Nat Rev Genet. 2007;8(5):341–52. https://doi.org/10.1038/nrg2101.

48. Gupta GP, Perk J, Acharyya S, de Candia P, Mittal V, Todorova-Manova K, et al. ID genes mediate tumor reinitiation during breast cancer lung metastasis. Proc Natl Acad Sci U S A. 2007;104(49):19506–11. https://doi.org/10.1073/pnas.0709185104.

49. Kang Y, Siegel PM, Shu W, Drobnjak M, Kakonen SM, Cordón-Cardo C, et al. A multigenic program mediating breast cancer metastasis to bone. Cancer Cell. 2003;3(6):537–49. https://doi.org/10.1016/S1535-6108(03)00132-6.

50. Jacob A, Linklater E, Bayless BA, Lyons T, Prekeris R. The role and regulation of Rab40b-Tks5 complex during invadopodia formation and cancer cell invasion. J Cell Sci. 2016;129(23):4341–53. https://doi.org/10.1242/jcs.193904.

51. Jacob A, Jing J, Lee J, Schedin P, Gilbert SM, Peden AA, et al. Rab40b regulates trafficking of MMP2 and MMP9 during invadopodia formation and invasion of breast cancer cells. J Cell Sci. 2013;126(Pt 20):4647–58. https://doi.org/10.1242/jcs.126573.

52. Paratala BS, Chung JH, Williams CB, Yilmazel B, Petrosky W, Williams K, et al. RET rearrangements are actionable alterations in breast cancer. Nat Commun. 2018;9(1):4821. https://doi.org/10.1038/s41467-018-07341-4.

53. Charafe-Jauffret E, Ginestier C, Iovino F, Tarpin C, Diebel M, Esterni B, et al. Aldehyde dehydrogenase 1-positive cancer stem cells mediate metastasis and poor clinical outcome in inflammatory breast cancer. Clin Cancer Res. 2010;16(1):45–55. https://doi.org/10.1158/1078-0432.CCR-09-1630.

54. Bos PD, Zhang XH-F, Nadal C, Shu W, Gomis RR, Nguyen DX, et al. Genes that mediate breast cancer metastasis to the brain. Nature. 2009;459(7249):1005–9. https://doi.org/10.1038/nature08021.

55. TARGET | www.broadinstitute.org/cancer/CGA [Internet]. [cited 2018 Jun 25]. Available from: http://archive.broadinstitute.org/cancer/cga/target

56. Brastianos PK, Carter SL, Santagata S, Cahill DP, Taylor-Weiner A, Jones RT, et al. Genomic characterization of brain metastases reveals branched evolution and potential therapeutic targets. Cancer Discov. 2015;5(11):1164–77. https://doi.org/10.1158/2159-8290.CD-15-0369.

57. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012;486(7403):346–52. https://doi.org/10.1038/nature10983.

58. Angelova M, Mlecnik B, Vasaturo A, Bindea G, Fredriksen T, Lafontaine L, et al. Evolution of metastases in space and time under immune selection. Cell. 2018;175(3):751–65.e16.

59. Fisher B, Anderson SJ. The breast cancer alternative hypothesis: is there evidence to justify replacing it? J Clin Oncol. 2010;28(3):366–74. https://doi.org/10.1200/JCO.2009.26.8292.

60. André F, Bachelot T, Commo F, Campone M, Arnedos M, Dieras V, et al. Comparative genomic hybridisation array and DNA sequencing to direct treatment of metastatic breast cancer: a multicentre, prospective trial (SAFI R01/UNICANCER). Lancet Oncol. 2014;15(3):267–74. https://doi.org/10.1016/S1470-2045(13)70611-9.

61. De Mattos-Arruda L, Mayor R, Ng CKY, Weigelt B, Martínez-Ricarte F, Torrejon D, et al. Cerebrospinal fluid-derived circulating tumour DNA better represents the genomic alterations of brain tumours than plasma. Nat Commun. 2015;6(1):8839. https://doi.org/10.1038/ncomms9839.

62. Huang X, Qiao Y, Brady S, Cohen A, Bild A, Marth G. Novel temporal and spatial patterns of metastatic colonization from breast cancer rapid-autopsy tumor biopsies. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE184717.

63. Huang X. Tumor-evo-rapid-autopsy. Github. https://github.com/xiaomengh/tumor-evo-rapid-autopsy (2021).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.