Assessment of extended-stage small cell lung cancer tumor heterogeneity from rapid research autopsy

Russell Bonneville

Extended-stage small cell lung cancer (ES-SCLC) is a highly lethal neuroendocrine cancer arising from the bronchial epithelium, with 2-year survival rates less than 10%. SCLC initially exhibits high sensitivity to standard-of-care chemotherapy. However, relapse occurs almost universally, characterized by high resistance to second-line chemotherapy and a median survival of 5 months. Tumor heterogeneity is thought to contribute to the evolution of treatment-resistant clones and delayed relapse in SCLC, but this has been difficult to characterize due to lack of available tumor tissues. Consequently, very little is known of the evolution of SCLC through metastasis and chemotherapy.

Through rapid research autopsy, whole exome sequencing, and bioinformatics analysis, we describe the tumor heterogeneity present within two patients with metastatic, relapsed SCLC. We demonstrate that analysis of tumor heterogeneity permits reconstruction of paths of tumor metastasis, along with the determination of site-specific influences on clonal dynamics.

Our results suggest a model of relapsed ES-SCLC heterogeneity in which genetic diversity develops in the lung early in relapse, followed by co-metastasis of multiple clones and Darwinian selection of specific subclones at metastatic sites. Mutational signature analysis reflected DNA damage from both tobacco smoking and platinum-based chemotherapy, suggesting a potential role of chemotherapy given for initial SCLC onset in promotion of genetic diversity in relapsed SCLC.

Hui-Zi Chen, Michele Wing, Jharna Miya, Eric Samorodnitsky, Melanie Krook, Julie Reeser, Amy Smith, Thuy Dao, Dorrelyn Martin, Aharon G. Freud, Patricia Allenby, Peter Shields, Greg Otterson, and Sameek Roychowdhury have contributed to this research.

INTRODUCTION

Extensive-stage small cell lung cancer (ES-SCLC), also known as oat cell lung cancer, is a highly lethal and aggressive malignancy arising from the bronchial epithelium. It is diagnosed almost exclusively in tobacco smokers[1], and has an overall 2-year survival rate of less than 10%[2]. Notably, ES-SCLC follows a characteristic biphasic disease course. Current first-line treatment with cisplatin or carboplatin-containing chemotherapy regimens[3] produces a profound initial response, often even complete radiographic remission. However, relapse is almost inevitable and is usually highly resistant to chemotherapy and radiation therapy, with a median survival of 5.5 months[4]. This biphasic course suggests an underlying biological cause.

However, despite extensive research, the molecular mechanisms mediating ES-SCLC relapse and metastasis remain poorly understood[5], primarily due to a lack of tumor tissues available for study. Recurrent SCLC progresses rapidly, narrowing the window for SCLC study and rendering patients severely ill, which entirely precludes biopsy of chemoresistant SCLC in many patients. ES-SCLC tends to metastasize extensively throughout the human body, and biopsies are unable to sample the patient's entire tumor burden.

One potential explanation for both the acquisition of treatment resistance and the extreme metastatic potential of ES-SCLC is tumor heterogeneity. Traditionally, oncogenesis has been viewed as a linear process, as normal cells accumulate driver mutations that culminate in oncogenic transformation. However, recent studies have demonstrated that individual tumors can possess remarkable genetic heterogeneity[6]. Moreover, heterogeneity may be present not only within a specific tumor, but among a primary tumor and its metastases as well[7]. For instance, chemotherapy resistance may occur in some instances by Darwinian selection of an already-present resistant subclone[8].

Several previous studies have investigated tumor heterogeneity within a variety of cancers. For instance, Gerlinger et al.[9] utilized whole exome sequencing to construct phylogenetic trees of primary and metastatic renal carcinoma. Avigdor et al.[10] utilized tumor heterogeneity to track the metastatic paths of breast cancer through the body. Brown et al.[11] demonstrated cross-seeding of tumor cells between metastases at different sites. Pisapsia et al.[12] performed research autopsy of fifteen patients (three cancer types), identifying potential metastasis-associated mutations. ES-SCLC has a relatively high median tumor mutational burden of 9.9/Mb[13], making it an ideal cancer for the study of heterogeneity by providing a high probability of mutational events.

In this study, we address sample collection issues through rapid research autopsy, a technique previously used successfully to study several cancer types including breast[10-12, 14] and brain[15] cancer. We have established a rapid research autopsy protocol at the Ohio State University Wexner Medical Center, and we report preliminary findings from analysis of tumor heterogeneity within two ES-SCLC patients. Whole exome sequencing was used to profile germline and somatic mutations within multiple tumors from each patient, and bioinformatics analyses applied to identify and quantify genetically distinct subclones within each tumor sample.

METHODS

Rapid research autopsy: Patients enrolled in the OSU13053 clinical trial were provided the opportunity to consent for rapid research autopsy. Two of these patients are presented in this study. Upon death, the decedents were transported to the OSU Regional Autopsy Center, and rapid research autopsy was performed by the Roychowdhury lab autopsy team and the Autopsy and Tissue Procurement teams of the OSUWMC Department of Pathology. 50 tumors were sampled from patient OSU13053-400 (hereafter patient 400), and 28 from patient OSU13053-431

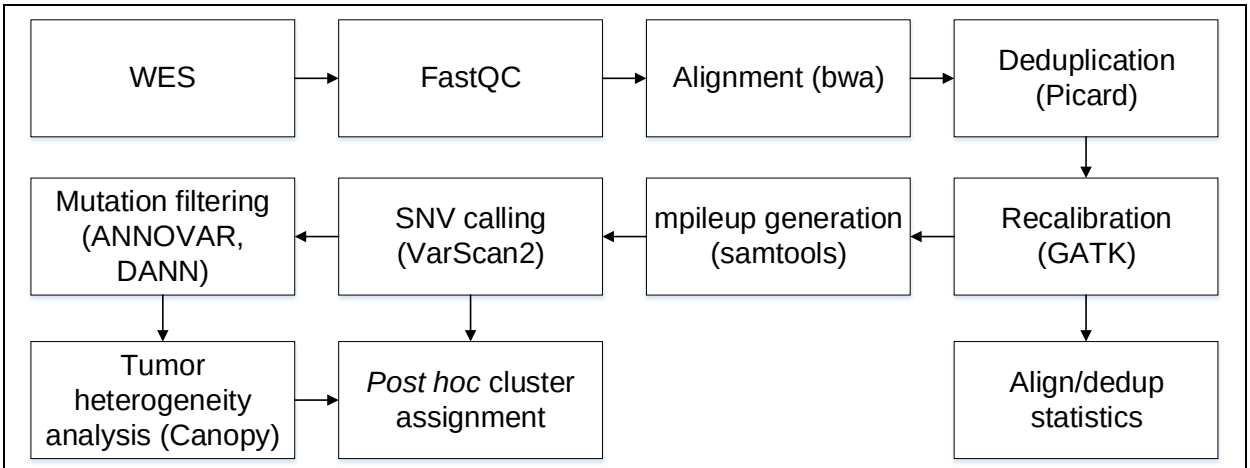

Figure 1: Outline of the bioinformatics workflow. After alignment and pre-processing, Canopy is used to deconvolve subclonal architecture of multiple tumor samples and infer subclonal phylogeny.

(hereafter patient 431). Paraffin embedding, sectioning, and hematoxylin and eosin staining were performed by the CLIA-certified Cancer Genomics Laboratory (CGL, ID#36D2076140, last inspected and certified June 2016) at the Roychowdhury Lab. Samples were evaluated for tumor content by Dr. Aharon Freud.

Whole exome sequencing: 11 samples from patient 400 and 17 from patient 431 were chosen for whole exome sequencing. Cores were extracted and DNA extraction was performed. Whole exome sequencing (WES) was performed using Illumina® HiSeq™ 4000 machines at the Nationwide Children's Hospital Genomics Core.

Sequencing reads were aligned to hg19[16] using bwa[17] version 0.6.2 (*Figure 1*). Optical duplicates were removed using Picard[18] version 2.3.0. Quality recalibration and local realignment around indels was performed with GATK[19] version 3.5. Somatic and germline variants were called using VarScan2[20] version 2.3.9 and annotated with ANNOVAR version 2016-02-01[21]. Variants with average base quality of alternate-supporting reads of at least 22, VarScan2 $P<0.05$, and minimum average distance to 3' read end 0.24 were used for further

analysis. Mutational signatures were called with deconstructSigs[22] version 1.8.0, run on R version 3.3.2. The 30 COSMIC mutational signatures[23] were used as a reference signature set. All aforementioned processing and analyses were performed using the Oakley cluster at the Ohio Supercomputer Center[24].

Analysis of tumor heterogeneity: Canopy[25] was utilized for subclonal inference. First, for each patient, the reference read count, alternate allele count, and variant fractions of all nonsynonymous somatic variants called within each tumor sample were compiled. Variants with at least 100x coverage in all samples, minimum of 20 alt-supporting reads in at least one sample, maximum variant fraction of 50% (to avoid confounding by potential DNA amplification) and a DANN[26] predicted mutational impacted score of 0.96 were used for input to Canopy. Canopy was then run with the following parameters: mutation cluster number from 2 to 9, 10 MCMC clustering runs, τ_{K+1} 0.05, number potential subclones from 3 to 12, 50 chains per subclone number, burnin 100, thinning parameter 5, simulation runs from 20000 to 100000, and writeskip 200. For each patient, Canopy first clusters mutations into several clusters, which are used to learn a phylogenetic tree of subclones and compute the relative population fraction of each subclone in each tumor sample.

Next, minimum angular dissimilarity cluster assignment was performed for all mutations identified in all samples for each patient, to retroactively "graft" them to the phylogenetic tree computed by Canopy. First, for each mutation in any sample, a vector representation of variant allele fractions (VAFs) was constructed as follows:

$$\vec{m}_i = [v_{1i}\ v_{2i} \ldots v_{ni}]^{\mathrm{T}}$$

where $\vec{m}_i$ is the vector for mutation i, n is the number of samples, and v_{ji} is the VAF for sample $j \in 1..n$. A centroid vector was then constructed for all mutations assigned by Canopy to each cluster by computing the element-wise mean of each normalized mutation vector:

$$\vec{c}_K = \frac{1}{|K|}\sum_{i \in K} \frac{\vec{m}_i}{\|\vec{m}_i\|}$$

where *K* is the set of all mutations within a cluster, $\vec{c}_K$ is the centroid vector, and || || represents the L_2 vector norm. Finally, for each mutation *not* clustered by Canopy, the assignment was computed as follows:

$$a_\kappa(\vec{m}) = \arg\min_{K \in \kappa} \left(\cos^{-1} \frac{\vec{m} \cdot \vec{c}_K}{\|\vec{c}_K\| \|\vec{c}_K\|} \right)$$

where κ is the set of all mutational clusters and $a_\kappa(\vec{m})$ is a function assigning a mutation to a cluster.

RESULTS

Patients autopsied: Patient 400 was a 62-year-old male with metastatic SCLC who died 12 months post-diagnosis. First-line therapy included carboplatin/etoposide. After tumor progression, he was enrolled in the OSU-15171 clinical trial (nivolumab, nivolumnab+ipilimumab, or placebo). He subsequently received palliative radiation therapy to the mediastinum and paclitaxel. Patient 431 was a 64-year-old female with metastatic SCLC who died 20 months post-diagnosis. First-line treatment included carboplatin/etoposide. She was also enrolled in OSU-15171 after progression, and subsequently received irinotecan, paclitaxel, and gemcitabine, and was enrolled in the OSU-15303 clinical trial (rovalpituzumab tesirine).

Next-generation sequencing: For both patients, next-generation sequencing of tumor samples collected at autopsy and matched normal blood was performed as previously described (see Methods). Eleven tumors and one normal sample were sequenced from patient 400, with 228x average coverage. Seventeen tumors and one normal were sequenced from patient 431, with 202x average coverage. Average tumor mutational burden was 11.3/Mb in patient 400 and 16.2/Mb in patient 431.

A: Patient 400

Signature	Lung #1	Lung #2	Lung #3	Supraclavicular LN #1	Supraclavicular LN #2	Left hilar LN #1	Left hilar LN #2	Paraaortic LN	Heart	Adrenal	Liver
1	0.110	0.081	0.000	0.000	0.000	0.000	0.000	0.000	0.075	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.061	0.060	0.072	0.000	0.080	0.000
3	0.000	0.000	0.000	0.091	0.000	0.168	0.127	0.101	0.126	0.159	0.092
4	0.220	0.236	0.199	0.000	0.000	0.118	0.144	0.120	0.189	0.112	0.214
5	0.073	0.183	0.334	0.203	0.469	0.438	0.311	0.454	0.066	0.221	0.175
8	0.000	0.000	0.000	0.092	0.087	0.000	0.000	0.000	0.000	0.065	0.000
12	0.101	0.000	0.000	0.000	0.000	0.088	0.152	0.000	0.000	0.000	0.000
13	0.000	0.000	0.062	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
16	0.309	0.322	0.000	0.246	0.180	0.000	0.000	0.123	0.281	0.000	0.000
18	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
19	0.000	0.000	0.136	0.081	0.079	0.070	0.066	0.000	0.000	0.000	0.099
23	0.060	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
24	0.000	0.000	0.000	0.088	0.000	0.000	0.000	0.000	0.000	0.000	0.000
29	0.000	0.000	0.000	0.000	0.061	0.000	0.000	0.000	0.000	0.000	0.000
30	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

B: Patient 431

Signature	Lung #1	Lung #2	Lung #3	Lung #4	Left hilar LN	Paraaortic LN #1	Paraaortic LN #2	Adrenal #1	Adrenal #2	Liver #1	Liver #2	Liver #3	Liver #4	Liver #5	Liver #6	Liver #7	Liver #8
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.062	0.000	0.000	0.000	0.000	0.000
3	0.224	0.242	0.199	0.175	0.331	0.283	0.236	0.285	0.274	0.270	0.156	0.072	0.170	0.276	0.137	0.147	0.159
4	0.256	0.282	0.267	0.207	0.200	0.236	0.285	0.275	0.312	0.333	0.332	0.272	0.313	0.315	0.320	0.348	0.318
5	0.071	0.183	0.234	0.305	0.000	0.000	0.152	0.063	0.000	0.000	0.149	0.257	0.182	0.000	0.254	0.186	0.152
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
12	0.127	0.102	0.151	0.130	0.072	0.106	0.084	0.077	0.112	0.126	0.107	0.000	0.073	0.137	0.000	0.132	0.112
13	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
16	0.000	0.000	0.000	0.000	0.000	0.068	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
18	0.103	0.090	0.000	0.000	0.086	0.141	0.088	0.110	0.101	0.109	0.106	0.000	0.069	0.101	0.088	0.070	0.106
19	0.000	0.000	0.000	0.000	0.077	0.000	0.070	0.061	0.000	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000
23	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
24	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
29	0.000	0.000	0.000	0.000	0.099	0.000	0.000	0.000	0.000	0.000	0.000	0.087	0.000	0.000	0.000	0.000	0.000
30	0.083	0.000	0.000	0.000	0.073	0.000	0.000	0.000	0.000	0.000	0.061	0.000	0.000	0.060	0.000	0.000	0.000

Table 1: Mutational signature analysis within two ES-SCLC patients identifies expected and unanticipated signatures. Each mutational signature reflects the pattern of single-nucleotide changes induced by one or more biological processes. Analysis was performed using deconstructSigs (Rosenthal et al.).

Mutational signature analysis: Nine mutational signatures were detected in patient 400 and seven

in patient 431 (*Table 1*). Of particular note were signature 4, previously associated with smoking[23], and signature 3, associated with double-stranded DNA breaks such as induced by platinum-based chemotherapy. Signature 18, previously reported in neuroblastoma (a neuroendocrine-derived malignancy as is SCLC), was observed in 15 of 17 samples from patient 431.

As both patients 400 and 431 were tobacco smokers, signature 4 was expected. Its presence in 26 of 28 tumor samples (among both patients) is consistent with a smoking-induced malignancy. Signature 3 was detected in 7 of 11 samples from patient 400 and all samples from patient 431.

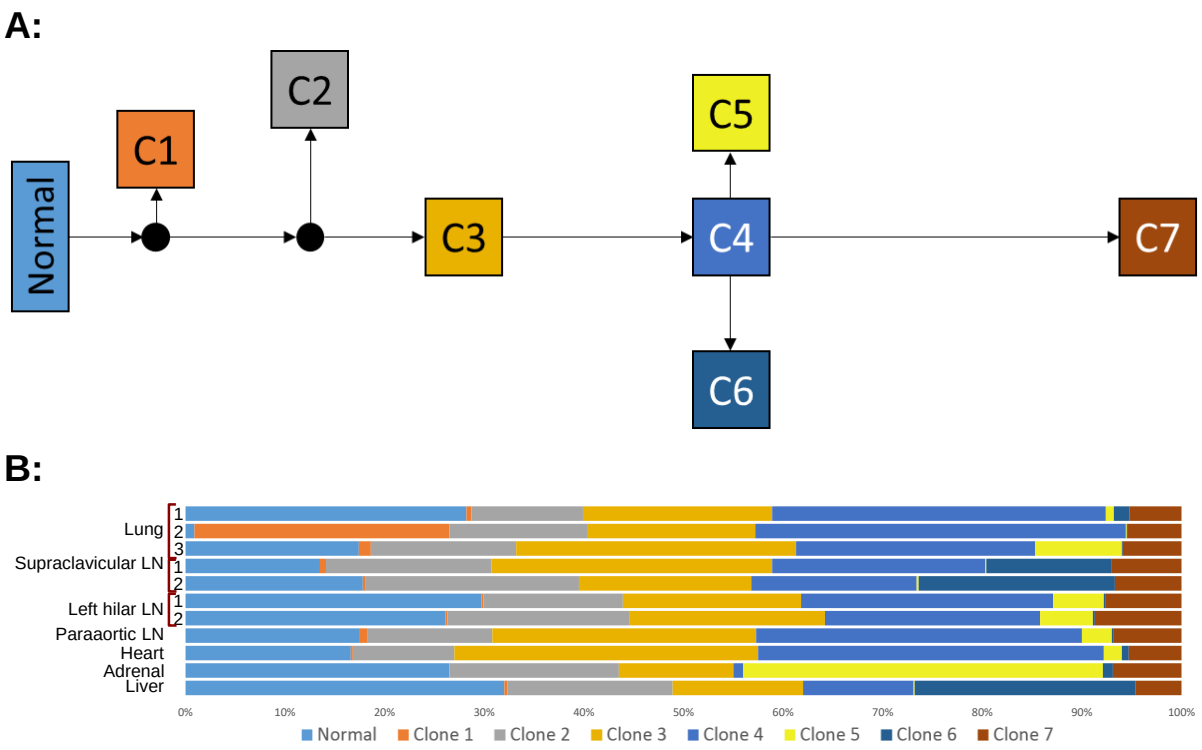

Figure 2: Analysis of tumor heterogeneity within ES-SCLC patient 400. **A:** Phylogenetic tree of seven identified subclones within patient 400. Tumor samples were extracted by rapid research autopsy and compared with normal blood. Length of lines correlates to relative number of SNVs which differentiate identified subclones, and black circles indicate unobserved subclones inferred by Canopy. **B:** Population fraction of each subclone in each tumor sample from patient 400, along with proportion of normal cells inferred. LN: lymph node.

This finding is consistent with homologous repair deficiency[27] and carboplatin therapy, such as received by both patients.

Tumor heterogeneity: Seven distinct genetic subclones were identified in patient 400 (*Figure 2*). Six of seven subclones were identified among the three lung samples at 5% population fraction or greater. Clone 1, a very early clone in the clonal phylogeny, was strongly enriched in lung sample

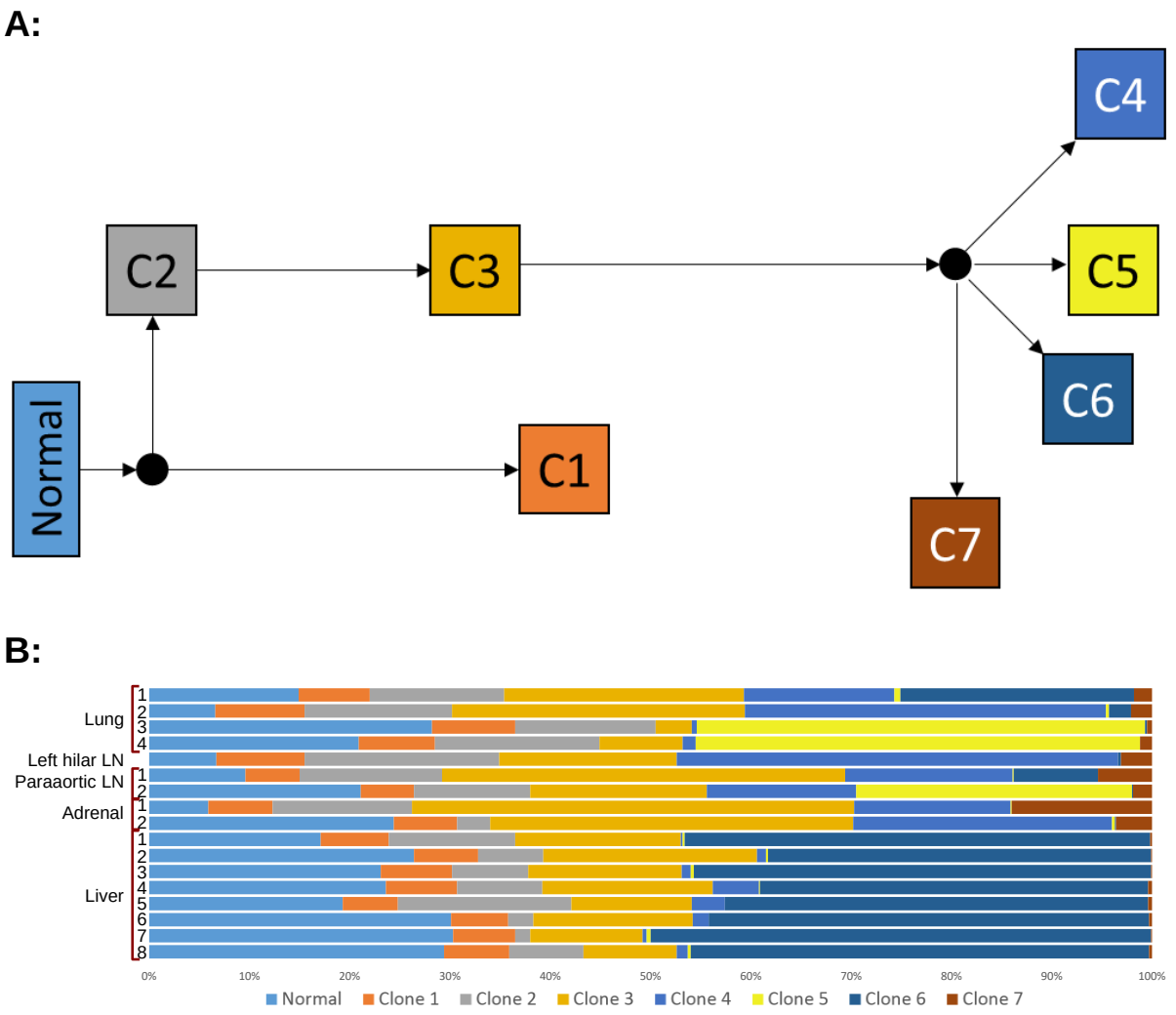

Figure 3: Analysis of tumor heterogeneity within ES-SCLC patient 431. **A:** Phylogenetic tree of seven identified subclones within patient 431. Tumor samples were extracted by rapid research autopsy and compared with normal blood. Length of lines correlates to relative number of SNVs which differentiate identified subclones, and black circles indicate unobserved subclones inferred by Canopy. **B:** Population fraction of each subclone in each tumor sample from patient 431, along with proportion of normal cells inferred. LN: lymph node.

#2 (25.6%). Clones 2 and 3 were detected with at least 10% population fraction in all tumor samples. Clone 5 was identified in lung sample #3, both hilar lymph node samples, the paraaortic lymph node sample and the heart, and was substantially expanded in the adrenal sample (36.1%). Clone 6 was identified at low population fraction (1.6%) in lung sample #1, but at much higher fraction in both supraclavicular lymph node samples (12.6% and 19.7%) and in the liver sample (22.2%). This may suggest a path of SCLC metastasis in this patient from the lung through the supraclavicular lymph nodes to the liver. Clone 7 was detected with low frequency in all samples.

In patient 431, seven distinct genetic subclones were also identified (*Figure 3*). All subclones except clone 7 were identified among the four lung samples at 5% population fraction or greater. Clones 1, 2, and 3 were detected at minimum 5% fraction in all organ sites. Clone 4 was particular to the lung, lymph nodes and adrenal gland. Clone 5 was strongly enriched in lung samples #3 and #4 and the paraaortic lymph node sample. Clone 6 was primarily observed in lung sample #1 (23.3%), paraaortic lymph node sample #1 (8.84%), and in all eight liver samples (38.2%-49.9%). These observations of clone 5 and clone 6 also suggest potential paths of SCLC metastasis in this patient. Clone 7 was detected in the lung, lymph nodes and adrenal glands, with relative enrichment in adrenal sample #1 (14.0%).

DISCUSSION

In this study, we report the first rapid research autopsy-based analysis of tumor heterogeneity within ES-SCLC. We performed whole exome sequencing on multiple tumor samples from two patients, identified DNA mutations, analyzed mutational signatures, and assessed the subclonal architecture present within each patient.

In both patients, almost every identified subclone was detected in at least one lung sample. Furthermore, every tumor profiled exhibited some degree of heterogeneity. This suggests either

co-metastasis of multiple genetic subclones, or potentially cross-seeding of tumor cells among primary and metastatic sites. Mutational signature analysis demonstrated that most tumors from both patients reflect DNA damage inflicted by smoking and potentially carboplatin therapy.

Taken together, all of these findings support a preliminary model of tumor heterogeneity within relapsed ES-SCLC. Under this model, carboplatin-based cytotoxic chemotherapy induces DNA damage in the SCLC cell population, inducing tumor heterogeneity early in the relapse phase of ES-SCLC if not before. Co-metastasis of multiple subclones from the primary lung tumor follows during ES-SCLC relapse. However, site-specific microenvironments of other factors lead to natural selection of some subclones over others, leading to a narrowing of diversity and expansion of subclones with higher site-specific fitness.

Several future directions are available from this study. Our bioinformatics pipeline currently utilizes the variant allele fractions of single-nucleotide variants to determine subclonal architecture and tumor heterogeneity. Utilization of copy-number variations could substantially increase the discriminative power of our approach. Recruitment of additional patients would expand the sample size of our study, allowing observation of potentially different patterns of tumor heterogeneity in other SCLC patients. Although pre-treatment ES-SCLC samples are often difficult to acquire due to SCLC morbidity and rapid disease course, inclusion of pre-treatment samples matched to autopsy samples would allow direct assessment of the impact of carboplatin-based chemotherapy on tumor heterogeneity. An increased sample size would also permit more robust statistical inferences of the subclone-specific mutations themselves, such as enrichment and correlation analyses.

In this study, we demonstrate the power of rapid research autopsy to infer tumor heterogeneity within ES-SCLC patients, with potential clinical significance for first-line SCLC chemotherapy. These methods are in theory applicable to any solid tumor type.

1. Lara, J.D., et al., *Clinical predictors of survival in young patients with small cell lung cancer: Results from the California Cancer Registry.* Lung Cancer, 2017. **112**: p. 165-168.
2. Jett, J.R., et al., *Treatment of small cell lung cancer: diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines.* Chest, 2013. **143**(5, Supplement): p. e400S-e419S.
3. Lara, P.N., et al., *Disease control rate at 8 weeks predicts subsequent survival in platinum-treated extensive stage small-cell lung cancer: results from the southwest oncology group (SWOG) database.* Clinical Lung Cancer, 2016. **17**(2): p. 113-118.e2.
4. Huber, R.M. and A. Tufman, *Update on small cell lung cancer management.* Breathe, 2012. **8**(4): p. 314.
5. Asai, N., et al., *Relapsed small cell lung cancer: treatment options and latest developments.* Therapeutic Advances in Medical Oncology, 2014. **6**(2): p. 69-82.
6. McGranahan, N. and C. Swanton, *Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future.* Cell. **168**(4): p. 613-628.
7. Hong, W.S., M. Shpak, and J.P. Townsend, *Inferring the origin of metastases from cancer phylogenies.* Cancer Research, 2015. **75**(19): p. 4021.
8. Meacham, C.E. and S.J. Morrison, *Tumor heterogeneity and cancer cell plasticity.* Nature, 2013. **501**(7467): p. 328-337.
9. Gerlinger, M., et al., *Intratumor heterogeneity and branched evolution revealed by multiregion sequencing.* New England Journal of Medicine, 2012. **366**(10): p. 883-892.
10. Avigdor, B.E., et al., *Mutational profiles of breast cancer metastases from a rapid autopsy series reveal multiple evolutionary trajectories.* JCI Insight, 2017. **2**(24).
11. Brown, D., et al., *Phylogenetic analysis of metastatic progression in breast cancer using somatic mutations and copy number aberrations.* Nature Communications, 2017. **8**: p. 14944.
12. Pisapia, D.J., et al., *Next-Generation Rapid Autopsies Enable Tumor Evolution Tracking and Generation of Preclinical Models.* JCO Precision Oncology, 2017(1): p. 1-13.
13. Chalmers, Z.R., et al., *Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden.* Genome Medicine, 2017. **9**(1): p. 34.
14. Savas, P., et al., *The Subclonal Architecture of Metastatic Breast Cancer: Results from a Prospective Community-Based Rapid Autopsy Program "CASCADE".* PLOS Medicine, 2016. **13**(12): p. e1002204.
15. Nikbakht, H., et al., *Spatial and temporal homogeneity of driver mutations in diffuse intrinsic pontine glioma.* Nature Communications, 2016. **7**: p. 11185.
16. International Human Genome Sequencing, C., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**: p. 860.

17. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.
18. *Picard Tools - By Broad Institute*. Available from: http://broadinstitute.github.io/picard/.
19. McKenna, A., et al., *The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data.* Genome Research, 2010. **20**: p. 1297-1303.
20. Koboldt, D.C., et al., *VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing.* Genome Research, 2012. **22**(3): p. 568-576.
21. Wang, K., M. Li, and H. Hakonarson, *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.* Nucleic Acids Research, 2010. **38**(16): p. e164-e164.
22. Rosenthal, R., et al., *deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution.* Genome Biology, 2016. **17**: p. 31.
23. Forbes, S.A., et al., *COSMIC: somatic cancer genetics at high-resolution.* Nucleic Acids Research, 2017. **45**(D1): p. D777-D783.
24. *Ohio Supercomputer Center: Oakley*. Available from: https://www.osc.edu/resources/technical_support/supercomputers/oakley.
25. Jiang, Y., et al., *Assessing intratumor heterogeneity and tracking longitudinal and spatial clonal evolutionary history by next-generation sequencing.* Proceedings of the National Academy of Sciences, 2016. **113**(37): p. E5528-E5537.
26. Quang, D., Y. Chen, and X. Xie, *DANN: a deep learning approach for annotating the pathogenicity of genetic variants.* Bioinformatics, 2015. **31**(5): p. 761-763.
27. Polak, P., et al., *A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer.* Nature Genetics, 2017. **49**: p. 1476.