

# Genomic characterization of metastatic ultra-hypermuted interdigitating dendritic cell sarcoma through rapid research autopsy

Hui-Zi Chen<sup>1,2,\*</sup>, Russell Bonneville<sup>1,3,\*</sup>, Lianbo Yu<sup>4</sup>, Michele R. Wing<sup>1</sup>, Julie W. Reeser<sup>1</sup>, Melanie A. Krook<sup>1</sup>, Jharna Miya<sup>1</sup>, Eric Samorodnitsky<sup>1</sup>, Amy Smith<sup>1</sup>, Dorrelyn Martin<sup>1</sup>, Thuy Dao<sup>1</sup>, Qishan Guo<sup>1</sup>, David Liebner<sup>1,4</sup>, Aharon G. Freud<sup>5</sup>, Patricia Allenby<sup>6</sup> and Sameek Roychowdhury<sup>1</sup>

<sup>1</sup>Department of Internal Medicine, Division of Medical Oncology, Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio, USA

<sup>2</sup>Department of Internal Medicine, Hematology and Oncology Fellowship Program, Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio, USA

<sup>3</sup>Biomedical Sciences Graduate Program, The Ohio State University, Columbus, Ohio, USA

<sup>4</sup>Department of Biomedical Informatics, The Ohio State University, Columbus, Ohio, USA

<sup>5</sup>Department of Pathology, Division of Hematopathology, The Ohio State University, Columbus, Ohio, USA

<sup>6</sup>Department of Pathology, Division of Autopsy Services, The Ohio State University, Columbus, Ohio, USA

\*These authors contributed equally to this work

**Correspondence to:** Sameek Roychowdhury, **email:** Sameek.Roychowdhury@osumc.edu

**Keywords:** tumor heterogeneity; research autopsy; hypermutation

**Received:** October 24, 2018

**Accepted:** October 31, 2018

**Published:** January 08, 2019

**Copyright:** Chen et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

**Interdigitating dendritic cell sarcoma (IDCS) is an extremely rare cancer of dendritic cell origin that lacks a standardized treatment approach. Here, we performed genomic characterization of metastatic IDCS through whole exome sequencing (WES) of tumor tissues procured from a patient who underwent research autopsy. WES was also performed on a treatment-naïve tumor biopsy sample obtained from prior surgical resection. Our analyses revealed ultra-hypermutation, defined as >100 mutations per megabase, in this patient's cancer, which was further characterized by the presence of three distinct mutational signatures including UV radiation and APOBEC signatures. To characterize clonal heterogeneity, we used the bioinformatics tool Canopy to leverage single nucleotide and copy number variants to catalog six subclones across various metastatic tumors. Truncal alterations, defined as being present in all clonal tumor cell populations, in this patient's cancer include point mutations in *TP53* and *CDKN2A* and amplifications of *c-KIT* and *APOBEC3A-H*, which are likely driver mutations. In summary, we have performed genomic characterization evaluating tumor mutational burden (TMB) and heterogeneity in a patient with metastatic IDCS. Despite ultra-hypermutation, this patient's cancer was not responsive to treatment with PD-1 inhibition. Our results underscore the importance of characterizing clonal heterogeneity in TMB-high cancers.**

## INTRODUCTION

Interdigitating dendritic cell sarcoma (IDCS) is an extremely rare malignancy of dendritic cell origin with

approximately 100 cases reported to date [1, 2]. Due to its rarity and challenging diagnosis, genomic characterization of this neoplasm has not been previously reported. Furthermore, no standard therapy exists for IDCS, which

tends to affect middle-age adults with median age of diagnosis of 56.5 years [3]. Localized IDCs constitutes 47% of cases and manifests as painless lymphadenopathy, most commonly involving the cervical and axillary nodes. Isolated extra-nodal disease occurs in 25% of cases, involving the liver, lung, spleen, bone marrow and gastrointestinal tract [3]. Distant metastases occur in 39% of cases and most frequently involved lymph nodes, lung, liver and bone marrow [3].

Here we performed whole exome sequencing (WES) of multiple tumors obtained through rapid research autopsy of a patient with metastatic IDCs. To our knowledge, this is the first time IDCs has been extensively sequenced and analyzed. Our findings demonstrate a rare ultra-hypermutated genotype, with >100 somatic mutations per megabase of genome (mut/Mb) [4], characterized by distinct mutational signature profiles and clonal diversity that occurred early in the evolution of this patient's cancer. Phylogenetic analysis revealed inactivation of *TP53* and *CDKN2A* (*p16<sup>INK4a</sup>/p19<sup>ARF</sup>*) as well as amplification of *c-KIT* and *PDGFR $\alpha$*  as truncal alterations. We further detected amplification of *APOBEC3A-H* encoding cytidine deaminases on chromosome 22q that may have attributed to the ultra-hypermutation. In summary, our work describes the genomic landscape and clonal heterogeneity of an ultra-rare cancer through the innovative approach of research autopsy.

## Diagnosis and treatment

Our patient was a 57-year-old Caucasian male who developed an isolated FDG-avid 2.4 × 2.4 cm right-sided neck mass on PET scan in 2016. Biopsy result of this mass showed a “pleomorphic/spindle cell neoplasm”. Immunohistochemical (IHC) analyses demonstrated focal CK7 staining. Clinical evaluation revealed no primary lesion involving the skin or oropharynx. In August 2016, he underwent modified radical right neck dissection and partial submandibular gland excision with one level I lymph node demonstrating complete tumor involvement and suspicious for extranodal extension; remaining lymph nodes (0/29) from levels II, III and IV were negative for tumor infiltration. IHC on the surgical specimen showed positivity for S100, SOX10 and CK7. A diagnosis of IDCs was issued. Given his localized presentation, he received adjuvant radiation to the right level IB-III nodes with a boost to the primary site of disease. In November 2016, he initiated adjuvant nivolumab (3 mg/kg every 2 weeks) given FoundationOne® report showing >100 mut/Mb in his tumor. Following adjuvant immunotherapy, he developed metastatic recurrence that failed to respond to subsequent treatments including combination immunotherapy (CTLA-4/PD-1 inhibitors), chemotherapy and molecularly targeted therapies (Figure 1). His clinical course leading up to the research autopsy is summarized in Figure 1A.

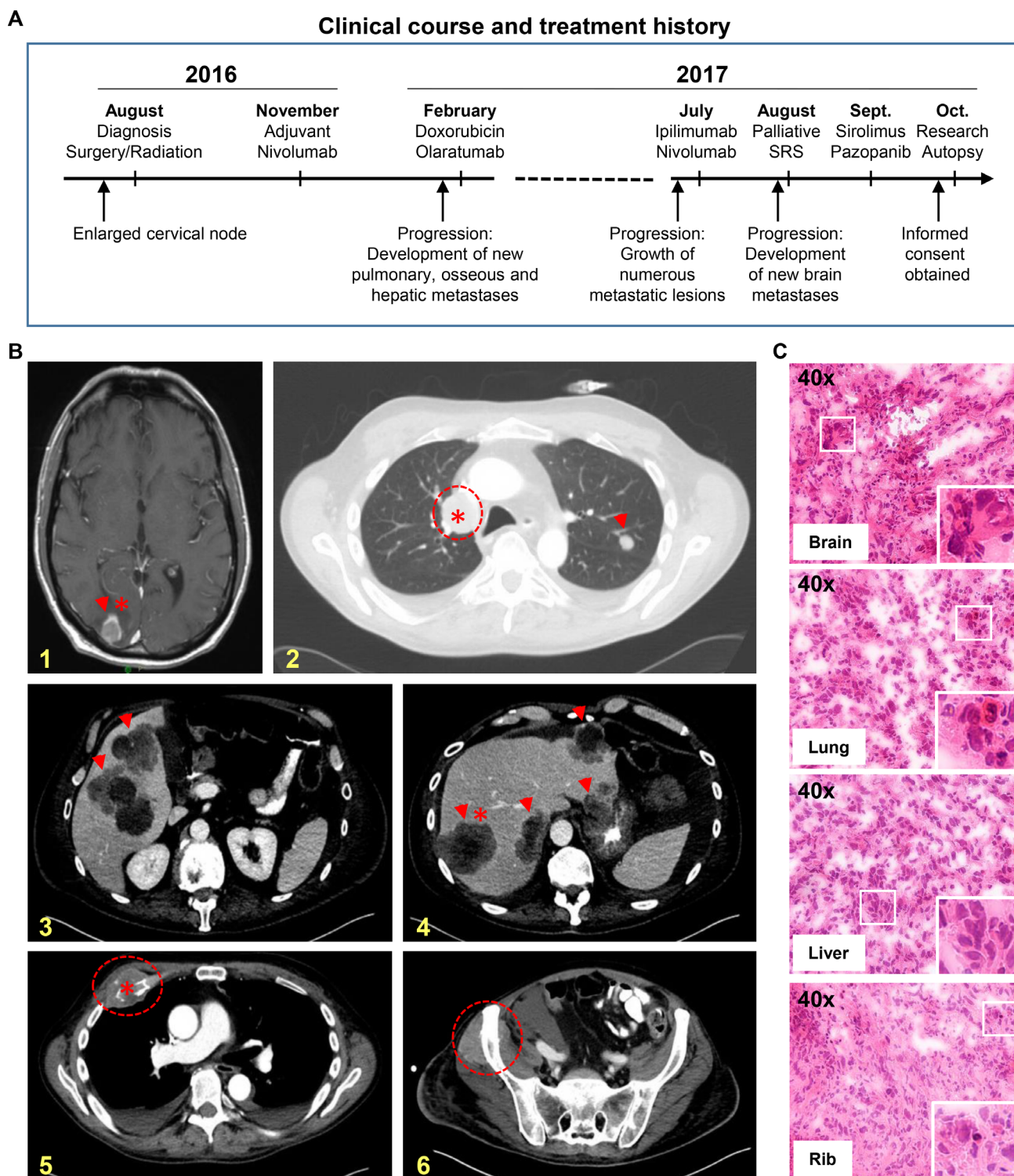
## RESULTS

### Rapid research autopsy

Prior to his death, the patient was consented to an Institutional Review Board (IRB)-approved research autopsy study for patients with advanced cancers (Figure 2). CT scans obtained prior to hospice enrollment showed tumor involvement of multiple organs (Figure 1B) and were used to guide sample procurement at time of research autopsy, which was performed within three hours post-mortem. A total of twenty-four metastatic tumor samples were procured from involved organ sites (Figure 1B, \*tumors). A pathologist assessed the viability and tumor cell (TC) content of autopsy samples (Figure 1C) prior to selection for genomic studies.

### Genomic characterization of IDCs

We performed WES on nine metastatic tumors and the resected pre-treatment tumor (Table 1). Calculation of tumor mutational burden (TMB), including single nucleotide variants (SNVs) and insertions/deletions (indels), demonstrated ultra-hypermutation (130.1–167.0 mut/Mb) in all tumors analyzed (Table 1; Supplementary Table 1). Interrogation of microsatellite status utilizing MANTIS [5] showed that all tumors were microsatellite stable (MSS) despite their ultra-hypermutation (Table 1). To further characterize this patient's cancer, we investigated whether specific substitution mutational signatures, which arise due to different mutagenic processes [6], may be present. This analysis revealed the presence of Signature 2 (APOBEC), Signature 7 (UV light) and Signature 11 (alkylating agent), which are all characterized by C>T substitutions, in all tumor samples (Figure 3A–3B; Supplementary Figure 1 and Supplementary Table 2). Of the three mutational signatures, Signature 7 had the highest prevalence (Figure 3A). We next classified the 7,417 somatic variants (SNVs and indels) in all ten tumor samples of this patient into three categories: Ubiquitous, Shared and Private (Figure 4A–4B). Ubiquitous variants are variants detected in all tumor samples and included driver mutations in *TP53* and *CDKN2A*, while shared variants are those detected in subsets of tumor samples. Finally, private variants are variants unique to each tumor sample. A tumor-centric evolutionary tree was constructed based on the pairwise genetic distance, as measured by number of discordant SNVs, between the wildtype cells (normal) and tumor samples in this patient (Figure 4C). Finally, we interrogated the prevalence of copy number variations (CNVs) using the allele-specific CNV caller FALCON [7]. After manual review and curation, 29 unique CNVs including amplification (including *c-KIT*, *PDGFR $\alpha$* , and *APOBEC3A-H*), deletion and loss-of-heterozygosity events were detected (Supplementary Figure 2 and Supplementary Tables 3–4).



**Figure 1: Research autopsy of a patient with metastatic IDCs.** (A) Summary of clinical course and treatment history of the patient. Research autopsy was performed 3 hours post-mortem. SRS, stereotactic radiosurgery. (B) CT scans depicting organs with metastatic cancer; brain (panel 1, yellow), hilar lymph node (2), liver (3-4), rib (5) and pelvis (6). Arrowheads and dashed circles indicate tumors; asterisks indicate tumors procured at research autopsy. (C) H&E stained frozen sections of metastatic tumors in brain, lung, liver and rib procured from research autopsy. Sections demonstrate high tumor cell content. Enlarged inset at bottom right of each image panel shows dysplastic nuclei with mitotic figures.



## Reconstructing clonal evolution

Integrating the above genomic data (SNVs, indels and CNVs), we used the program Canopy [8] to synthesize a hypothetical model of the clonal evolution of this patient's cancer (Figure 5A). Canopy identified six genetically distinct clonal populations of tumor cells, present in various proportions within each tumor (Figure 5B; Supplementary Table 5). These six clones are differentiated by ten unique groups of genomic alterations (Figure 5A, a–j; Supplementary Tables 6–7), which all contained mutational Signature 7 (Supplementary Figure 3 and Supplementary Table 2). Of note, group *a* contains truncal alterations ancestral to all tumor cells including two well-documented driver mutations, *TP53* P278L and *CDKN2A* R80X, which were classified as ubiquitous (Figure 4C). Interestingly, we also detected a truncal variant *RPA2* V108L. *RPA2* encodes a subunit of the heterotrimeric Replication Protein A complex that is critical for DNA replication, repair, recombination and DNA damage response [9]. Finally, we identified two notable truncal amplifications of regions on the long arms of chromosome 4 (~17.2Mb), containing *c-KIT* and *PDGFR $\alpha$* , and chromosome 22 (~24.6Mb), containing *APOBEC3A-H* (Figure 5A; Supplementary Figure 2).

## DISCUSSION

Extensive multi-regional sequencing of tumors revealed the complex genetic landscape of treatment-refractory cancers by demonstrating intratumor heterogeneity [10, 11], which underscores the limitations of tumor profiling with tissue derived from a single biopsy specimen. From a clinical perspective, tumor heterogeneity contributes to the incomplete therapeutic responses seen in patients receiving different types of anti-cancer therapies including molecularly targeted therapies. Mechanistically, tumor heterogeneity drives acquired therapeutic resistance by facilitating the selection and expansion of therapy-resistant clones. Research autopsy has emerged as a powerful approach to characterize tumor heterogeneity at different metastatic sites in the individual patient. Together with information from treatment-naïve samples, sequencing of metastatic tumor samples from autopsy can provide a comprehensive molecular portrait of advanced cancers that reflects changes in tumor cell populations and genomes over time [12, 13]. Here we performed research autopsy on a patient with widely metastatic IDCS. Genomic profiling of his tumors revealed somatic ultra-hypermutation and tumor heterogeneity in the form clonal diversity. Recently, Campbell *et al.* demonstrated a prevalence of ultra-hypermutation

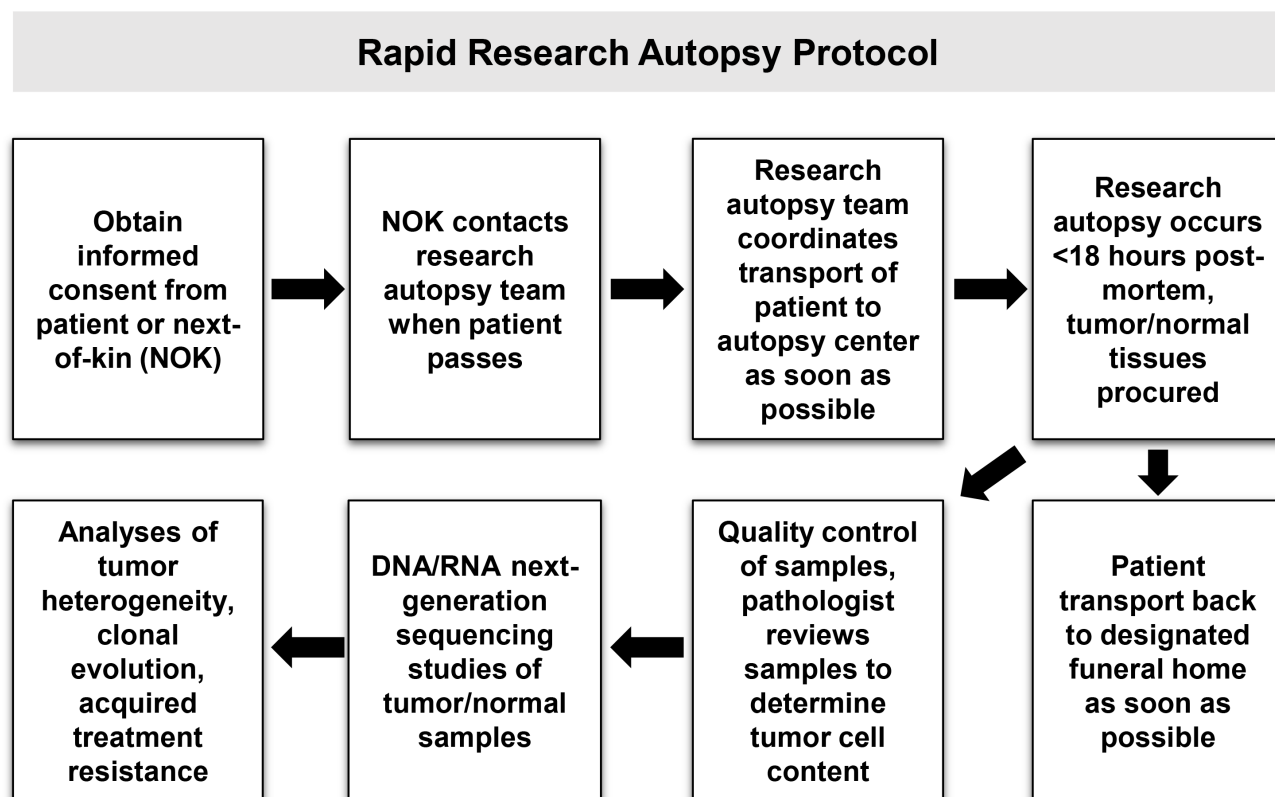


Figure 2: Overview of research autopsy protocol.

**Table 1: Sample characteristics and sequencing metrics**

Sample	Organ involved	% Tumor cells	Median Coverage	Percent 100X	TMB (mut/Mb)	# SNVs	# Indels	Microsatellite
Normal	Blood	N/A	265	88.01	N/A	N/A	N/A	N/A
Tumor 1 (T1)	Biopsy (Bx)	30%	100	50.10	131.5	5,091	21	0.326, MSS
T2	Left lung (L.lung)	80%	213	79.83	158.4	6,139	22	0.330, MSS
T3	Right lung (R.lung)	90%	245	84.19	165.3	6,408	19	0.330, MSS
T4	Liver1	90%	218	82.12	163.3	6,329	20	0.328, MSS
T5	Liver2	80%	233	83.38	158.8	6,160	15	0.327, MSS
T6	Liver3a	80%	270	87.01	150.4	5,826	22	0.318, MSS
T7	Liver3b	90%	286	87.88	156.1	6,050	19	0.324, MSS
T8	Rib	80%	272	86.70	148.9	5,776	15	0.319, MSS
T9	Right iliac (R.ilic)	90%	242	85.20	167.0	6,473	20	0.330, MSS
T10	Brain	50%	257	86.80	130.1	5,043	15	0.320, MSS

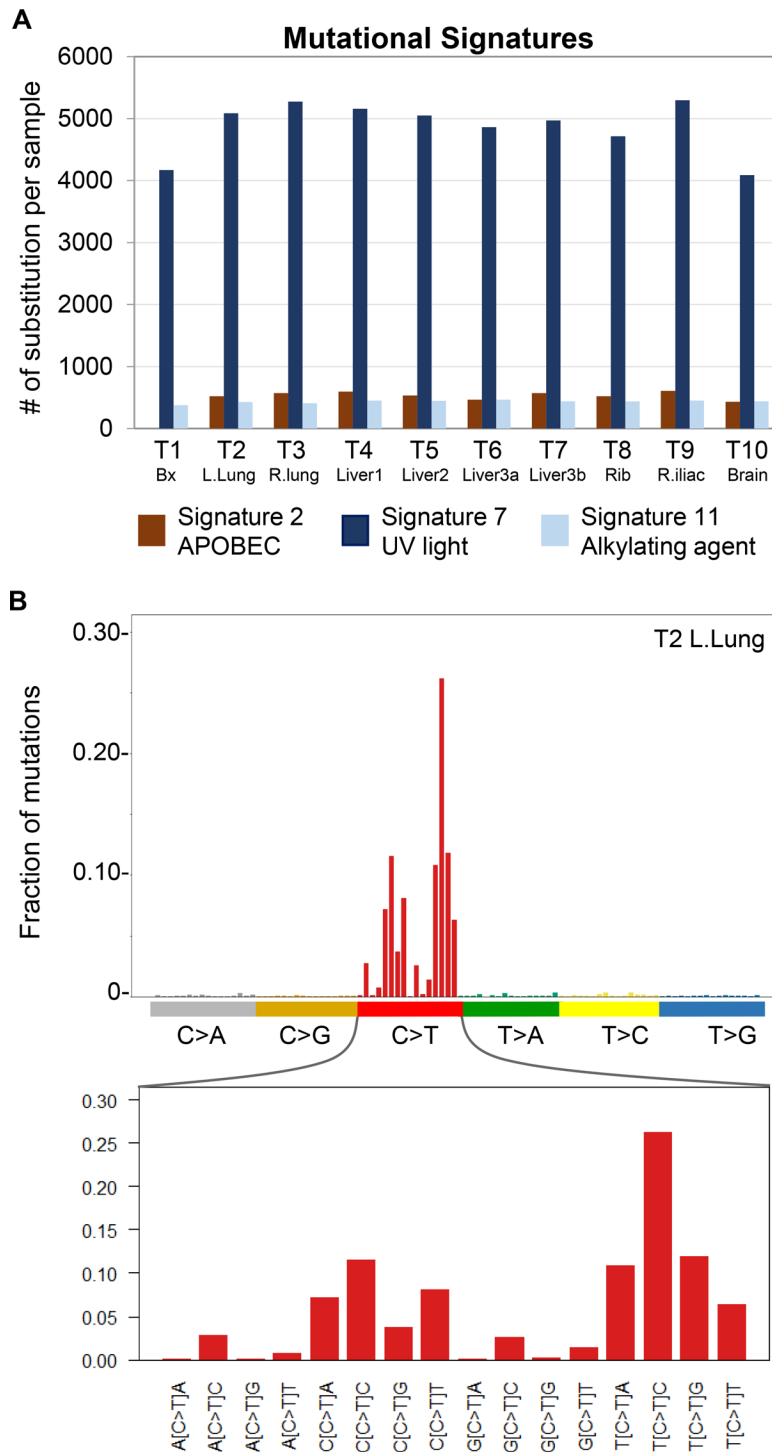
A board-certified pathologist estimated the percentage of tumor cells in each sample. Liver3a and Liver3b samples were procured from two different regions of a large liver tumor. The pre-treatment biopsy sample was obtained from a formalin-fixed paraffin-embedded surgical specimen, while autopsy tumor samples were frozen in OCT. Percentage 100X indicates the percent of variants in each tumor sample with at least 100X coverage. TMB includes synonymous and non-synonymous SNVs as well as insertions/deletions (indels). #SNVs and #Indels columns indicate the total number of somatic SNVs and indels detected in each tumor sample. Microsatellite status was determined using MANTIS, with scores <0.4 indicating microsatellite stability (MSS).

(TMB >100 mut/Mb) of 0.6% in a cohort of 78,452 adult cancers [4]. Therefore, the rare histologic classification of IDCS in this patient is accompanied by a rare ‘genotype’ of ultra-hypermutation.

Hypermutation in cancer may arise from intrinsic defects in DNA damage repair pathways or extrinsic mechanisms such as mutagenic exposure [4]. The majority of ultra-hypermutation detected by Campbell *et al.* was attributed to mutations in mismatch repair (MMR) genes and replication-associated DNA polymerases Polε or Polδ. While we did not detect genomic alterations involving MMR or Polε/Polδ to explain this patient’s ultra-hypermutated cancer, we did detect a truncal amplification of a region on chromosome 22q containing genes encoding the APOBEC3A-H (or A3 subfamily) cytidine deaminases. The deregulated expression and activity of A3A and A3B have been linked to cancer development through enhanced DNA hypermutation and unfaithful RNA editing [14–16]. Furthermore, we identified a truncal *RPA2* variant with a predicted (by DUET [17]) destabilizing missense mutation in the C-terminal winged helix domain important for protein-protein interaction that could have further contributed to defective DNA damage repair and hypermutation. Together with driver mutations in tumor suppressors *TP53* and *CDKN2A*, the above events could have produced ultra-hypermutation seen in this patient’s cancer.

Different models of mutation accumulation have been proposed for hypermutated cancers and could be classified as ‘steady’ versus ‘dynamic’ hypermutation [4]. In the steady model, mutations accumulate gradually

due to continuous mutagenic exposure or germline MMR deficiencies, leading to hypermutation over time. The dynamic model, seen in cancers with Polε or Polδ mutations, is characterized by ‘bursts’ of mutations followed by rise in genome-wide TMB [4]. Through analyses of clonality and mutational signatures, we hypothesize a gradual mode of hypermutation consistent with the former model in our patient’s cancer. Genomic profiling of treatment-naïve tumor revealed high TMB at baseline that may have been the result of UV exposure, producing the driver mutations in *TP53* and *CDKN2A*. TMB analysis of his autopsy tumor samples demonstrated an increase of ~20–37 additional mut/Mb that could be attributed to the truncal amplification of the A3 subfamily of cytidine deaminases, as indicated by the presence of APOBEC-specific Signature 2, and/or potentially therapy-induced mutations. Interestingly, the brain metastasis (T10) was the only autopsy tumor sample that had similar TMB as the treatment-naïve tumor (T1) while still retaining the APOBEC signature. A technical explanation underlying this finding may be due to lower TC contents in both samples affecting variant detection. Alternatively, this may reflect similar biology between the treatment-naïve tumor sample obtained from neck surgery and the brain metastasis; this similarity is demonstrated by the shorter genetic distance between ‘normal’ and T1 or T10 samples relative to other tumor samples in the NJ tree. Finally, the genetic dissimilarity between the brain metastasis and tumor samples from other organs also suggests the presence of unique genetic features of cancer cells with increased propensity for invasion of the central nervous system. Although high TMB has been established

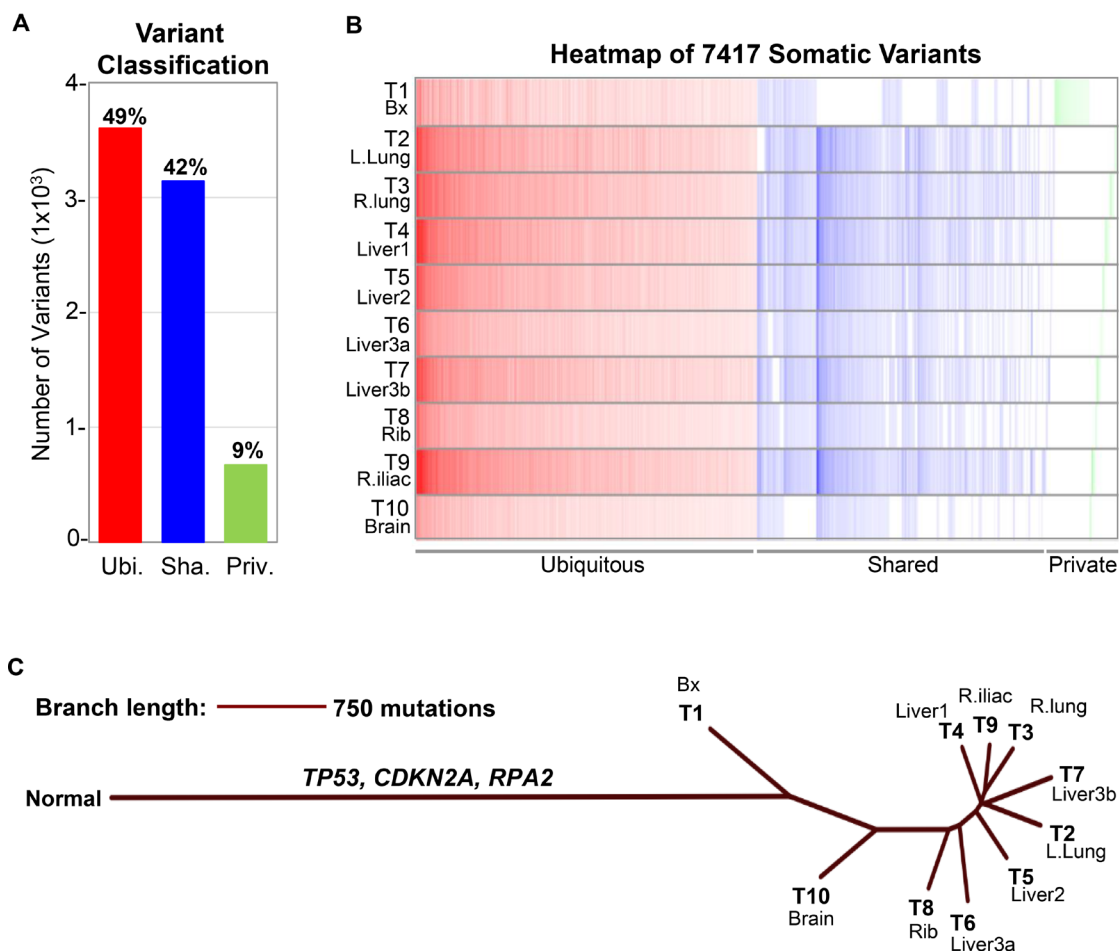


**Figure 3: Detection of mutational signature profiles in metastatic IDCs.** (A) Mutational signatures in tumor samples were called with deconstructSigs. Among the thirty signatures in the COSMIC Mutational Signatures set, only Signatures 2 (APOBEC), 7 (UV) and 11 (alkylating agents) were detected. T1-10 are as labeled and correspond to samples listed in Table 1. (B) Representative lego plot of a sequenced tumor (T2) demonstrating the predominance of C>T transitions in the three signatures detected in this patient's cancer. The X-axis of lego plots contains 96 possible mutation types that result when the six classes of base substitutions (e.g. C>T) are placed in the trinucleotide context of flanking 5' and 3' bases. The Y-axis represents the fraction of base substitutions in sample T2 within each trinucleotide context.

as a clinically useful biomarker of response to checkpoint blockade in multiple human cancers [18, 19], our patient developed progressive cancer while receiving adjuvant therapy with the PD-1 inhibitor nivolumab and had disease progression while being treated with dual CTLA-4/PD-1 blockade. Therefore, his case highlights the need for identification of additional biomarkers to predict clinical benefit for immunotherapy in patients with hypermutated cancers.

In summary, we present the first genomic characterization of metastatic IDCs, an extremely rare neoplasm of dendritic cell origin that lacks any standard therapy. This patient had metastatic IDCs characterized by ultra-hypermutation and clonal heterogeneity, likely through a combination of chronic mutagen exposure (UV),

acquired defects in pathways important for DNA repair (*TP53*, *CDKN2A*, *RPA* mutations), and gain of genes that promote DNA hypermutation (*APOBEC3A-H*). His cancer was aggressive and refractory to multiple anti-cancer therapies including molecularly targeted agents and immunotherapies. In the future, it will be important to study additional patients with this and other rare cancers with hypermutation. Finally, the broader clinical implication of our results is that although patients with hypermutated cancer, originating from either somatic or germline genomic aberrations, are more likely to benefit from checkpoint inhibition, research is still needed to stratify these patients to maximize therapeutic efficacy and identify the different genetic determinants of primary or acquired resistance to immunotherapy.



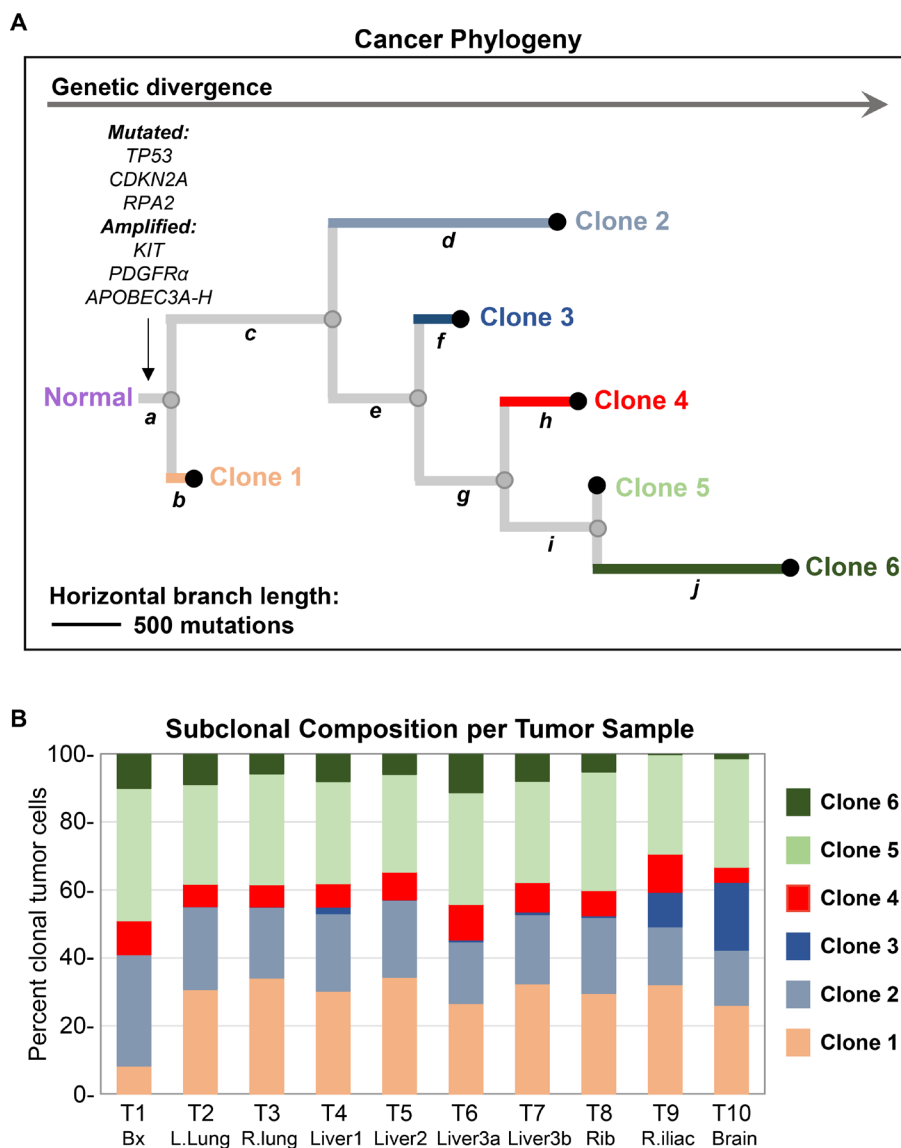
**Figure 4: Classification of somatic variants and construction of neighbor joining (NJ) tree.** (A) The 7,417 somatic variants detected in all ten tumor samples T1-10 from this patient were classified into three categories: Ubiquitous (Ubi.), present in all ten tumor samples; Shared (Sha.), present in some but not all ten tumor samples; Private (Priv.), present in only one tumor sample. Percentage of each variant category are as indicated in the bar graph. Somatic variants analyzed include single nucleotide variants (SNVs) and insertions/deletions (indels). (B) Heatmap demonstrating distribution and variant allele frequency (VAF) of the 7,417 somatic variants in ten tumor samples T1-10. Color intensity corresponds to VAF. (C) NJ tree depicting the evolutionary relationship of the ten tumor samples T1-10 from this patient. Normal represents hypothetical population of wildtype cells without somatic aberrations. The 3,642 ubiquitous variants included driver mutations *TP53* and *CDKN2A* as well as the *RPA2* variant predicted to have loss-of-function. Branch lengths correspond to the genetic distance between samples. T1-10 corresponds to samples listed in Table 1.

## MATERIALS AND METHODS

### Research autopsy

The patient was consented to an IRB-approved clinical study for tumor profiling and body donation. At time of death, the patient's next-of-kin (and/or hospice agency) notified members of the research team.

The deceased was transported to the OSU Medical Center and a limited research autopsy was conducted within 3 hours after patient's passing. Metastatic tumors and adjacent normal tissues were sampled and immediately archived as fresh frozen specimens in OCT medium. After conclusion of the autopsy, the deceased was transported back to the designated funeral home within 24 hours.



**Figure 5: Analysis of cancer phylogeny and clonal evolution.** (A) Phylogram constructed using Canopy integrating SNVs, curated copy number variations (CNVs) and indels detected through WES of tumor samples T1-10. Output from Canopy includes number of clones, clonal fractions and mutational groups (a-j) associated with each clone. In this ultra-hypermuted cancer, downsampling of SNVs was performed in addition to following recommended parameters for Canopy tree building. Somatic SNVs not utilized by Canopy as well as indels were retroactively assigned based on VAF patterns to mutational groups a-j using a maximal likelihood model. Based on this analysis, six different clones of tumor cells were identified in this patient's cancer. Truncal or group a mutations are common or ancestral to all clones of tumor cells. In contrast, mutations in groups b, d, f, h, and j are private to clones 1, 2, 3, 4, and 6, respectively. The length of horizontal branches in the tree is proportional to the number of mutations. (B) Stacked bar graph depicting the percentage of different clonal tumor cell populations as estimated by Canopy in each sample. T1-10 corresponds to samples listed in Table 1.



## Whole exome sequencing

We extracted genomic DNA from frozen tumors and normal (blood) samples and prepared sequencing libraries using an established protocol [20] that included enrichment with the xGEN<sup>®</sup> Exome Research Panel v1.0 from Integrated DNA Technologies. Sequencing was performed on an Illumina HiSeq4000 at The Genomics Services Laboratory at Nationwide Children's Hospital (Columbus, Ohio) and achieved a median depth of 100–286x.

## Alignment, variant calling, and annotation

Sequencing reads were aligned to hg19 using bwa [21] version 0.7.14. Deduplication was performed using Picard [22] version 2.3.0. Quality recalibration and local realignment around indels was performed with Picard and GATK [23] version 3.5. Somatic SNVs (single-nucleotide variations), somatic indels, and germline SNVs were called using VarScan2 [24] version 2.3.9. Somatic SNVs were filtered using bam-readcount as follows: minimum average base quality of variant-supporting reads 22, average variant distance to 3' read end of 0.24, Fisher's exact test  $P \leq 0.05$ , maximum average sum of base quality mismatches 100, maximum average mismatches per base of variant-supporting reads 0.04, and minimum VF (variant fraction) 6%. Germline SNVs were filtered as above, except for Fisher's exact test  $P \leq 0.1$  (as recommended by VarScan2 documentation). Somatic indels were filtered with Fisher's exact test  $P \leq 0.05$  and  $\text{VF} \geq 6\%$ . In addition, a list of repetitive regions in hg19 was generated using the RepeatFinder tool included in MANTIS, modified to output all repeats of 1-mers to 5-mers spanning at least 4 bp. Indels falling within this list were excluded from further analysis.

Somatic SNVs and indels were annotated with ANNOVAR (revision #11f4bb, 2016-02-01) [25]. Mutational signatures from the COSMIC Mutational Signatures set [26] were called with deconstructSigs [27] version 1.8.0 with default settings and exome2genome trinucleotide frequency correction, run on R version 3.3.2. Microsatellite instability testing was performed using MANTIS version 1.0.3, run with three threads and otherwise default settings.

To compute a neighbor joining tree of per-sample phylogeny, we first generate a distance matrix  $D \in \mathbb{Z}^{(N+1) \times (N+1)}$ . For sample  $i \in 1..N$ , define  $S_i$  as the set of somatic SNVs in that sample called as above. Define  $S_{N+1}$  as the empty set, representing germline (with no somatic SNVs by definition). We compute D as follows:

$$d_{ij} = |S_i \Delta S_j|$$

where  $\Delta$  is the set symmetric difference. Neighbor joining was performed over D using RapidNJ [28] version 2.3.2. The resulting tree was plotted with Interactive Tree of Life (iTOL) [29] version 4.2.3.

## Copy number variation

Copy number variations (CNVs) were called using FALCON version 0.2 with threshold 0.3, and the QC procedure provided with Canopy was run with default length and  $\Delta\text{CN}$  settings. The read depth ratio for each sample was computed as the ratio of aligned reads in tumor to aligned reads in normal. CNVs called by FALCON were manually inspected to identify events with major copy number  $>2$  or minor copy number  $<0.5$  in at least one tumor sample, and spanned at least 25% of a chromosome. For each of these CNVs, common breakpoints were estimated across all tumors. FALCON was then re-run in each sample with  $\hat{t}_{chr}$  set to the nearest SNP and threshold 0.2, with manual rescue of CNVs (threshold  $\geq 0.1$  in all cases).

## Subclonal phylogenetic analysis

The reference read count, alternate allele count, and variant fractions of all nonsynonymous somatic variants called within each tumor sample were compiled. Somatic SNVs were further filtered with at least 100x coverage in all samples, minimum of 20 alt-supporting reads in at least one sample, not on a sex chromosome, and a DANN [30] predicted mutational impact score  $\geq 0.96$  ( $n = 893$ ). This list of high-confidence mutations was downsampled to a set of 60 mutations with maximal diversity using a high correlation filter, implemented as follows:

```
while  $|M| > n$ :  
   $i \leftarrow \arg \max_x (\rho_{MxMy}); y > x; x, y \in 1..|M|$   
  delete( $M_i$ )
```

where  $M$  is the set of mutations, with each mutation being a vector of its per-sample VFs,  $n$  is the desired final number of mutations (in this case,  $n = 60$ ), and  $\rho$  is the correlation between VFs in each sample, defined as follows:

$$\rho_{\vec{a}\vec{b}} = \frac{\text{Cov}(\vec{a}, \vec{b})}{\sigma_{\vec{a}}\sigma_{\vec{b}}}$$

where  $\text{Cov}(x, y)$  is the covariance between elements of same-length vectors  $x$  and  $y$ , and  $\sigma_x$  is the population standard deviation of elements of vector  $x$ .

Canopy was then run with an in-house parallelized version of sample.cluster mode with cluster number from 2 to 9, 10 MCMC clustering runs,  $\tau_{k+1}$  0.05, number potential subclones from 3 to 9, 50 chains per subclone number, burnin 100, thinning parameter 5, simulation runs from 20000 to 50000, and writeskip 200. As FALCON does not generate standard deviations for regions with identical major and minor copy number, and Canopy does not support sparse  $\epsilon_M$  or  $\epsilon_m$  matrices, the Frobenius norm of non-NA values of the  $\epsilon_M$  and  $\epsilon_m$  matrices was used for Canopy. The optimal subclone number was selected by highest BIC (Bayesian Information Criterion) score.

Afterwards, the remaining somatic SNVs (not used for Canopy tree building) and indels were retroactively assigned to the resulting tree using a maximal likelihood model. For an arbitrary somatic mutation  $v$ , define  $N$  as the number of tumor samples,  $\vec{r} \in \mathbb{Z}^N$  as the number of alternate-supporting reads for  $v$  in each sample,  $\vec{x} \in \mathbb{Z}^N$  as the coverage of the variant position in each sample, and  $\vec{p} \in \mathbb{R}^N$ , a vector of expected variant fractions in each sample (derived from the tree as shown below). Given known  $\vec{x}$  and  $\vec{p}$ , and modeling alternative read depth as jointly binomial across tumor samples, the probability and log-likelihood can be computed as follows:

$$P(\vec{r} | \vec{x}, \vec{p}) = \prod_{i=1}^N \binom{x_i}{r_i} p_i^{r_i} (1 - p_i)^{(x_i - r_i)}$$

$$\mathcal{L}(v \in z) = \ln(P(\vec{r} | \vec{x}, \vec{p}))$$

$$= \sum_{i=1}^N \left[ \sum_{j=r_i+1}^{x_i} \ln(j) - \sum_{j=1}^{x_i-r_i} \ln(j) + r_i \ln(p_i) + (x_i - r_i) \ln(1 - p_i) \right]$$

To compute  $\vec{p}$ , we utilize the phylogenetic tree and clonal fractions reported by Canopy, along with its per-subclone major and minor copy number estimates. Define  $Z$  as the number of edges of the tree, and  $K$  as the number of tumor subclones + 1 (to account for normal cells). Given the tree, a matrix  $T \in \{0, 1\}^{Z \times K}$  representing the phylogenetic composition of each subclone can be constructed as follows (note that the tree root corresponds to germline):

$$T_{z,k} = \begin{cases} 1, & \text{if the path from root to subclone } k \text{ contains edge } z \\ 0, & \text{otherwise} \end{cases}, z = 1..Z, k = 1..K$$

Next, define  $P \in \mathbb{R}^{K \times N}$  as the clonal fraction matrix provided by Canopy (where entries in the first column represent proportion of normal cells in each sample),  $T$  as the number of CNV regions used for Canopy,  $\tilde{C}^m \in \mathbb{Z}^{(T \times K)}$  as the per-subclone major copy number matrix, and  $\tilde{C}^m \in \mathbb{Z}^{(T \times K)}$  as the per-subclone minor copy number matrix. Any zero values within each column of  $P$  were set to 0.0005, with the remaining entries reduced equally to ensure that each column sums to 1. This is first used to compute  $A \in \mathbb{R}^{(T \times N)}$ , the total copy number of each CNV region in each sample (independent of any specific mutation), as follows:

$$A = (\tilde{C}^m + \tilde{C}^m) P$$

Next, assume without loss of generality that  $v$  is within edge  $z$ . From Canopy, if  $v$  is within a CNV  $t$ ,  $t$  is within edge  $y$ . We can calculate the copy number of  $v$  in each subclone,  $\vec{b} \in \mathbb{Z}^K$ , as follows:

$$\vec{b} = (T_{z,\cdot} \circ C)^T, C = \begin{cases} \tilde{C}^m, & v \text{ is on major allele of CNV } t, \text{ and } y \text{ is further from the root than } z \\ \tilde{C}^m, & v \text{ is on major allele of CNV } t, \text{ and } y \text{ is further from the root than } z \\ \{1\}^K, & v \text{ is not within a CNV, or } z \text{ is further from the root than } y \end{cases}$$

where  $\circ$  is the Hadamard product of matrices. This permits calculation of the copy number of  $v$  in each sample, under

the assumptions that  $v \in z$  and  $v$  is on a major or minor allele, as follows:

$$\vec{s} = P^T \vec{b}$$

Computation of  $\vec{p}$  for mutation  $v$  is now straightforward:

$$p_i = \frac{s_i}{a_i}, i = 1..N, a = \begin{cases} A_{i,t}, & v \text{ is within CNV } t \\ \{2\}^N, & v \text{ is within a CNV, and is on an autosome or chrX in a female} \\ \{1\}^N, & v \text{ is on chrX or chrY in a male} \end{cases}$$

To perform the assignment,  $\mathcal{L}(v \in z)$  is computed as described above for all edges. If  $v$  is within a CNV region and the candidate edge is before the CNV, its possibilities of being on the major or minor allele are both considered. If the candidate edge is after the CNV, only the possibility of it occurring on one copy of the allele is considered, utilizing the simplifying assumptions (also made by Canopy) that each somatic mutation occurs only once and no back-mutations occur. If the candidate edge contains the CNV, all three above possibilities (major/minor/after) must be considered. To avoid numerical errors, if  $p_i = 0$  and  $r_i \neq 0$  for a candidate assignment in any sample, that candidate assignment is discarded. The edge with highest log-likelihood is selected, with ties (such as can occur with deletions or on a different branch than a CNV) broken in favor of the edge furthest from the root node. This approach was implemented and run using Python 2.7.1, using scipy [31] 0.10.1.

## Author contributions

H.C., R.B., and S.R. wrote the manuscript. H.C., R.B., L.Y., M.R.W., J.W.R., M.A.K., J.M., E.S., A.S., D.M., T.D., Q.G. and S.R. participated in the procurement of tumor tissues from research autopsy, prepared samples for whole exome sequencing, and participated in the analyses of sequencing data. D.L., A.G.F. and P.A. critically reviewed the manuscript. A.G.F. assessed tumor samples for tumor cell content and viability prior to sequencing. P.A. performed research autopsy of patients and assisted in the identification of tumor tissues.

## ACKNOWLEDGMENTS

We are grateful for administrative support from Jenny Badillo, the Comprehensive Cancer Center, community support from Pelotonia, and most importantly the patient and his family. We thank the Ohio Supercomputer Center (OSC) for providing us with the computing power and resources for the analyses in this report.

## CONFLICTS OF INTEREST

S.R. participated in Advisory Boards for Incyte Corporation and AbbVie, Inc. (2017). S.R. received honoraria from IDT Integrated DNA Technologies.

(2017). S.R. participated in an Advisory Board for QED Therapeutics. (2018).

## FUNDING

S.R. is supported by an American Cancer Society grant MRS-12-194-01-TBG, the Prostate Cancer Foundation, NHGRI UM1HG006508, NCI UH2CA202971 (Sparkfuse), NCI UH2CA216432 (MSIDx), American Lung Association, and Pelotonia.

H.Z.C. is supported by a Pelotonia Post-Doctoral Research Fellowship and American Society of Clinical Oncology Conquer Cancer Foundation Young Investigator Award.

R.B. is supported by NIGMS T32GM068412.

M.A.K. is supported by a T32 Oncology Training Grant (5T32CA009338).

## Editorial note

This paper has been accepted based in part on peer-review conducted by another journal and the authors' response and revisions as well as expedited peer-review in *Oncotarget*.

## REFERENCES

- Xue T, Jiang XN, Wang WG, Zhou XY, Li XQ. Interdigitating dendritic cell sarcoma: Clinicopathologic study of 8 cases with review of the literature. *Ann Diagn Pathol.* 2018; 34:155–60. <https://doi.org/10.1016/j.anndiagpath.2018.03.008>.
- Pokuri VK, Merzianu M, Gandhi S, Baqai J, Loree TR, Bhat S. Interdigitating dendritic cell sarcoma. *J Natl Compr Canc Netw.* 2015; 13:128–32.
- Saygin C, Uzunaslan D, Ozguroglu M, Senocak M, Tuzuner N. Dendritic cell sarcoma: a pooled analysis including 462 cases with presentation of our case series. *Crit Rev Oncol Hematol.* 2013; 88:253–71. <https://doi.org/10.1016/j.critrevonc.2013.05.006>.
- Campbell BB, Light N, Fabrizio D, Zatzman M, Fuligni F, de Borja R, Davidson S, Edwards M, Elvin JA, Hodel KP, Zahurancik WJ, Suo Z, Lipman T, et al. Comprehensive Analysis of Hypermutation in Human Cancer. *Cell.* 2017; 171:1042–56.e10. <https://doi.org/10.1016/j.cell.2017.09.048>.
- Kautto EA, Bonneville R, Miya J, Yu L, Krook MA, Reeser JW, Roychowdhury S. Performance evaluation for rapid detection of pan-cancer microsatellite instability with MANTIS. *Oncotarget.* 2017; 8:7452–63. <https://doi.org/10.18632/oncotarget.13918>.
- Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL, Boyault S, Burkhardt B, Butler AP, et al. Signatures of mutational processes in human cancer. *Nature.* 2013; 500:415–21. <https://doi.org/10.1038/nature12477>.
- Chen H, Bell JM, Zavala NA, Ji HP, Zhang NR. Allele-specific copy number profiling by next-generation DNA sequencing. *Nucleic Acids Res.* 2015; 43:e23. <https://doi.org/10.1093/nar/gku1252>.
- Jiang Y, Qiu Y, Minn AJ, Zhang NR. Assessing intratumor heterogeneity and tracking longitudinal and spatial clonal evolutionary history by next-generation sequencing. *Proc Natl Acad Sci U S A.* 2016; 113:E5528–37. <https://doi.org/10.1073/pnas.1522203113>.
- Byrne BM, Oakley GG. Replication protein A, the laxative that keeps DNA regular: The importance of RPA phosphorylation in maintaining genome stability. *Semin Cell Dev Biol.* 2018 Apr 20. [Epub ahead of print]. <https://doi.org/10.1016/j.semcdb.2018.04.005>.
- Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A, Tarpey P, Varela I, Phillimore B, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med.* 2012; 366:883–92. <https://doi.org/10.1056/NEJMoa1113205>.
- Gerlinger M, Horswell S, Larkin J, Rowan AJ, Salm MP, Varela I, Fisher R, McGranahan N, Matthews N, Santos CR, Martinez P, Phillimore B, Begum S, et al. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat Genet.* 2014; 46:225–33. <https://doi.org/10.1038/ng.2891>.
- Juric D, Castel P, Griffith M, Griffith OL, Won HH, Ellis H, Ebbesen SH, Ainscough BJ, Ramu A, Iyer G, Shah RH, Huynh T, Mino-Kenudson M, et al. Convergent loss of PTEN leads to clinical resistance to a PI(3)K inhibitor. *Nature.* 2015; 518:240–4. <https://doi.org/10.1038/nature13948>.
- Faltas BM, Prandi D, Tagawa ST, Molina AM, Nanus DM, Sternberg C, Rosenberg J, Mosquera JM, Robinson B, Elemento O, Sboner A, Beltran H, Demichelis F, et al. Clonal evolution of chemotherapy-resistant urothelial carcinoma. *Nat Genet.* 2016; 48:1490–9. <https://doi.org/10.1038/ng.3692>.
- Salter JD, Bennett RP, Smith HC. The APOBEC Protein Family: United by Structure, Divergent in Function. *Trends Biochem Sci.* 2016; 41:578–94. <https://doi.org/10.1016/j.tibs.2016.05.001>.
- Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B, Refsland EW, Kotandeniya D, Tretyakova N, Nikas JB, Yee D, Temiz NA, Donohue DE, et al. APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature.* 2013; 494:366–70. <https://doi.org/10.1038/nature11881>.
- Henderson S, Fenton T. APOBEC3 genes: retroviral restriction factors to cancer drivers. *Trends Mol Med.* 2015; 21:274–84. <https://doi.org/10.1016/j.molmed.2015.02.007>.
- Pires DE, Ascher DB, Blundell TL. DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Res.* 2014; 42:W314–9. <https://doi.org/10.1093/nar/gku411>.

18. Yarchoan M, Hopkins A, Jaffee EM. Tumor Mutational Burden and Response Rate to PD-1 Inhibition. *N Engl J Med*. 2017; 377:2500–1. <https://doi.org/10.1056/NEJMc1713444>.
19. Gibney GT, Weiner LM, Atkins MB. Predictive biomarkers for checkpoint inhibitor-based immunotherapy. *Lancet Oncol*. 2016; 17:e542–e51. [https://doi.org/10.1016/s1470-2045\(16\)30406-5](https://doi.org/10.1016/s1470-2045(16)30406-5).
20. Samorodnitsky E, Jewell BM, Hagopian R, Miya J, Wing MR, Lyon E, Damodaran S, Bhatt D, Reeser JW, Datta J, Roychowdhury S. Evaluation of Hybridization Capture Versus Amplicon-Based Methods for Whole-Exome Sequencing. *Hum Mutat*. 2015; 36:903–14. <https://doi.org/10.1002/humu.22825>.
21. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25:1754–60. <https://doi.org/10.1093/bioinformatics/btp324>.
22. Dietz S, Schirmer U, Merce C, von Bubnoff N, Dahl E, Meister M, Muley T, Thomas M, Sultmann H. Low Input Whole-Exome Sequencing to Determine the Representation of the Tumor Exome in Circulating DNA of Non-Small Cell Lung Cancer Patients. *PLoS One*. 2016; 11:e0161012. <https://doi.org/10.1371/journal.pone.0161012>.
23. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*. 2010; 20:1297–303.
24. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research*. 2012; 22:568–76. <https://doi.org/10.1101/gr.129684.111>.
25. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*. 2010; 38:e164–e. <https://doi.org/10.1093/nar/gkq603>.
26. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, Cole CG, Ward S, Dawson E, Ponting L, Stefancsik R, Harsha B, Kok CY, et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Research*. 2017; 45:D777–D83. <https://doi.org/10.1093/nar/gkw1121>.
27. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biology*. 2016; 17:31. <https://doi.org/10.1186/s13059-016-0893-4>.
28. Simonsen M, Mailund T, Pedersen CNS. Rapid Neighbour-Joining. In: Crandall KA and Lagergren J, eds. *Algorithms in Bioinformatics*. (Berlin, Heidelberg: Springer Berlin Heidelberg). 2008. pgs: 113–22.
29. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*. 2016; 44:W242–W5. <https://doi.org/10.1093/nar/gkw290>.
30. Quang D, Chen Y, Xie X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics*. 2015; 31:761–3. <https://doi.org/10.1093/bioinformatics/btu703>.
31. Jones E, Oliphant T, Peterson P. (2001-present). *SciPy: Open source scientific tools for Python*. <http://www.scipy.org/>.