Published in final edited form as:

*Cell Rep*. 2019 March 05; 26(10): 2651–2666.e6. doi:10.1016/j.celrep.2019.02.028.

# APOBEC Mutagenesis and Copy-Number Alterations Are Drivers of Proteogenomic Tumor Evolution and Heterogeneity in Metastatic Thoracic Tumors

**Nitin Roper**[1], **Shaojian Gao**[1], **Tapan K. Maity**[1], **A. Rouf Banday**[2], **Xu Zhang**[1], **Abhilash Venugopalan**[1], **Constance M. Cultraro**[1], **Rajesh Patidar**[3], **Sivasish Sindiri**[3], **Anna-Leigh Brown**[4], **Alexander Goncearenco**[4], **Anna R. Panchenko**[4], **Romi Biswas**[1], **Anish Thomas**[1], **Arun Rajan**[1], **Corey A. Carter**[5], **David E. Kleiner**[6], **Stephen M. Hewitt**[6], **Javed Khan**[3], **Ludmila Prokunina-Olsson**[2], and **Udayan Guha**[1,7,*]

[1]Thoracic and GI Malignancies Branch, Center for Cancer Research, National Cancer Institute, Bethesda, MD 20814, USA

[2]Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD 20814, USA

[3]Genetics Branch, Center for Cancer Research, National Cancer Institute, Bethesda, MD 20814, USA

[4]National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD 20814, USA

[5]Walter Reed National Military Medical Center, Bethesda, MD 20814, USA

[6]Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, Bethesda, MD 20814, USA

[7]Lead Contact

## SUMMARY

Intratumor mutational heterogeneity has been documented in primary non-small-cell lung cancer. Here, we elucidate mechanisms of tumor evolution and heterogeneity in metastatic thoracic tumors (lung adenocarcinoma and thymic carcinoma) using whole-exome and transcriptome sequencing, SNP array for copy-number alterations (CNAs), and mass-spectrometry-based quantitative



*Correspondence: udayan.guha@nih.gov.

AUTHOR CONTRIBUTIONS

N.R. and U.G. designed the study. N.R., T.K.M., X.Z., A.V., C.M.C., A.R.B., L.P.-O., and U.G. assisted with sample preparation and data analysis. N.R., T.K.M, X.Z., and U.G. performed the proteomic experiments and data analysis. N.R., S.G., R.P., S.S., A.G., A.L.B., A.R.P., and J.K. performed or assisted with computational analysis. N.R. performed the biostatistical analyses. S.H. and D.K. provided guidance and assisted with autopsies. A.V., T.K.M., R.B., and U.G. harvested tissues during autopsies. A.T., A.R., C.A.C., and U.G. were involved with clinical care of patients undergoing autopsies. N.R., L.P.-O., and U.G. wrote the manuscript. All authors reviewed, commented on, and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

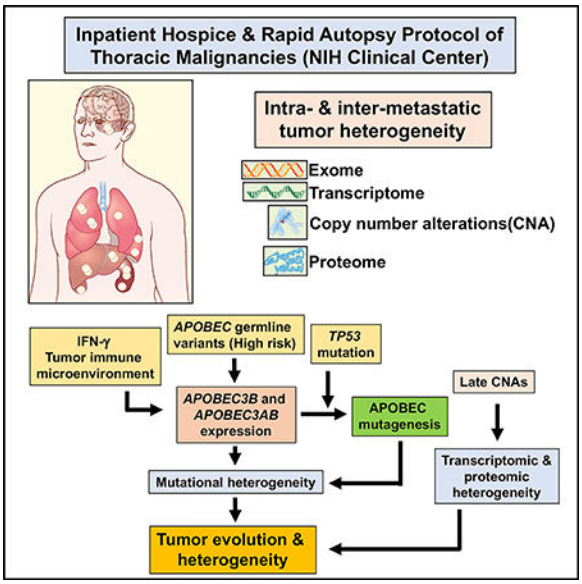
Supplemental Information can be found with this article online at https://doi.org/10.1016/j.celrep.2019.02.028.

proteomics of metastases obtained by rapid autopsy. APOBEC mutagenesis, promoted by increased expression of *APOBEC3* region transcripts and associated with a high-risk *APOBEC3* germline variant, correlated with mutational tumor heterogeneity. *TP53* mutation status was associated with APOBEC hypermutator status. Interferon pathways were enriched in tumors with high APOBEC mutagenesis and IFN-γ-induced expression of *APOBEC3B* in lung adenocarcinoma cells, suggesting that the immune microenvironment may promote mutational heterogeneity. CNAs occurring late in tumor evolution correlated with downstream transcriptomic and proteomic heterogeneity, although global proteomic heterogeneity was significantly greater than transcriptomic and CNA heterogeneity. These results illustrate key mechanisms underlying multi-dimensional heterogeneity in metastatic thoracic tumors.

## In Brief

Roper et al. perform integrated exome, transcriptome, and quantitative proteomics analyses of metastases obtained through rapid autopsy of patients with thoracic malignancies. They identify APOBEC mutagenesis, driven by germline variants, mutant *TP53*, and the tumor microenvironment, and late copy-number alterations as key mechanisms underlying proteogenomic evolution and heterogeneity.

## Graphical Abstract




## INTRODUCTION

Understanding the mechanisms by which metastatic lung adenocarcinoma (LUAD) and thymic carcinoma (TC) evolve may provide greater insight into tumor progression and may guide novel therapeutic avenues. Autopsy programs established to harvest tumor tissue from metastatic sites at the end of life have demonstrated significant heterogeneity depending on tumor type (Campbell et al., 2010; Gundem et al., 2015; Kumar et al., 2016; Liu et al., 2009; Patch et al., 2015; Shah et al., 2009). For example, metastatic pancreatic cancer has high

inter-metastatic heterogeneity of genomic rearrangements (Campbell et al., 2010), but not of driver mutations (Makohon-Moore et al., 2017), whereas metastatic clear-cell renal carcinoma has recurrent driver mutations that occur late within individual metastases (Gerlinger et al., 2014), resulting in high inter-metastatic heterogeneity. Within non-small-cell lung cancer (NSCLC), previous work has characterized the evolution of primary tumors and has demonstrated significant intra-tumor heterogeneity in copy-number alterations and driver mutations (de Bruin et al., 2014; Jamal-Hanjani et al., 2017; Zhang et al., 2014b). In these studies, mutations generated by the cytosine deaminase activity of APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) family of enzymes (substitutions of C with T or G in TCA or TCT motifs (Roberts et al., 2013) were found to occur late in the evolution of primary tumors (de Bruin et al., 2014) and were associated with subclonal mutations (Jamal-Hanjani et al., 2017). However, it is also critically important to understand the evolution of metastatic NSCLC, given metastatic lineages develop early in primary tumor development (Zhao et al., 2016). Thus, the evolution of metastatic disease may be substantially different from primary tumors. Moreover, previous NSCLC tumor heterogeneity studies have largely focused on whole-exome sequencing approaches. The evolution of tumors at the level of transcriptome or proteome and the underlying mechanisms that may generate multidimensional heterogeneity remain largely unknown.

Here, we sought to address the following questions:

1. What is the degree of genomic (mutational and copy number), transcriptomic, and proteomic heterogeneity within and between metastases of a given patient?
2. What is the relationship between these three levels of heterogeneity?
3. What are the potential drivers of such heterogeneity?

To address these questions, we performed rapid (“warm”) autopsies on four patients with LUAD and one patient with TC. The autopsies were initiated within 3 h of death to allow for procurement of sufficient quantity and quality of DNA, RNA, and protein from metastatic tumor tissue for whole-exome and transcriptome sequencing, DNA copy-number analysis, and mass-spectrometry-based proteomics. Our integrated analysis of the genome, transcriptome, and proteome uncovered key mechanisms likely driving proteogenomic heterogeneity within and across metastatic sites.

## RESULTS

### Sampling of Metastatic Tumors through Rapid Autopsy Protocol

We have established a rapid (“warm”) autopsy protocol for thoracic malignancies (including lung cancers and thymic epithelial tumors, among others) at the NIH Clinical Center. Under this protocol, patients with metastatic disease who are near the end of life receive inpatient hospice care. Upon death, an autopsy is performed within 3 h to procure sufficient quantity and high-quality of DNA, RNA, and protein from all possible sites of metastatic disease. For this study, we enrolled four patients with LUAD: patients RA000 and RA004 with oncogenic *KRAS* mutations who were previous smokers and patients RA003 and RA005 with *EGFR* mutations who were both non-smokers. We additionally enrolled patient RA006

with squamous cell TC, a non-smoker, who died within 1.5 years of diagnosis after no response to treatment (Table S1). All our study patients, except for patient RA000, were diagnosed with stage IV disease, had received chemotherapy and/or targeted therapy (range 2–8 lines of therapy), and were previously enrolled in a clinical trial at the NIH Clinical Center (Table S1). For each patient, we harvested 44–183 metastatic tumor lesions from multiple organs, including lung, liver, and kidney (Table S1). We selected a total of 40 tumor samples for further analyses based on tumor content by histology (Figure 1; Data S1). Before analysis, three samples with sequencing data were removed from the study because of low tumor content (Table S1). Additional tumor tissue collected from the likely primary site at the time of diagnosis was available for whole-exome sequencing in three patients (RA000, RA003, and RA006).

## Intra- and Inter-Metastatic Mutational Tumor Heterogeneity Is Highly Variable within Patients and Can Be Extreme

For each patient, we performed whole-exome sequencing (WES) of the metastatic tumors, primary tumors, where available, and matched germline DNA. A range of 182 (RA005) to 1,058 (RA003) non-silent mutations were identified per patient (Table S2). RNA sequencing (RNA-seq) demonstrated a high, independent validation rate of WES (Data S2), similar to previous studies (Network, 2014). We found significant mutational heterogeneity within each patient with non-truncal variants ranging from 67% in RA000 to 99% in RA003 (Figure 2A).

Activating mutations in *EGFR* (RA003) and *KRAS* (RA000, RA004) were present in primary and all metastatic tumors of the respective patients. However, an activating *HRAS* mutation was present in all metastatic sites of patient RA006 at autopsy, but not in the primary tumor at the time of diagnosis. Patients RA003 and RA006 had an average of 12 and 14 non-truncal driver mutations, respectively. In contrast, there was an average of only 3.7 non-truncal driver mutations in the other three patients (RA005, RA004, and RA000), demonstrating variability in metastatic site-specific driver mutation acquisition among patients (Figure 2A). Intra-tumor mutational heterogeneity in driver mutations was found when tested in RA003 tumors L2d/L2e (Figure 2A).

Next, we calculated Jaccard similarity coefficients (defined as the ratio of shared to all mutations between two metastatic tumors) for each patient to quantitatively assess intra- and inter-metastatic tumor genomic heterogeneity (values 0–1 correspond to the range from minimal to maximal heterogeneity (Blokzijl et al., 2016; Makohon-Moore et al., 2017) (Table S3). The means of the Jaccard similarity coefficients for each patient ranged from 0.25 (RA003) to 0.73 (RA000) and were significantly different between patients ($p = 2.2 \times 10^{-16}$, Kruskal-Wallis ranksum test; Figure 2B). Two patients, RA003 and RA006, were clear outliers and exhibited what we termed “extreme” mutational heterogeneity. These two patients had significantly lower combined mean Jaccard similarity coefficients compared to the other patients (mean 0.28 vs. 0.57, $p = 2 \times 10^{-16}$, chi-square test) (Figure 2B). Jaccard similarity coefficients exhibited a similar trend based on expressed mutations identified by RNA-seq analysis (Figure 2C). Additional sequencing resulting in a median exome coverage

of 487× (range 435 to 528) for the tumors from patients RA003 and RA006 did not substantially alter the observed extreme level of heterogeneity (Figure S1; Table S3).

To further evaluate mutational heterogeneity within each patient, we applied PyClone to estimate the proportion of cancer cells identified with a given mutation and the clonal structure of each tumor. As expected, truncal, key driver mutations, such as *EGFR* and *KRAS*, were clonal with high cellular prevalence whereas many shared or private mutations had medium to lower cellular prevalence (subclonal) (Table S2). Patient RA003 had the highest number of subconal mutations (range 243–506), followed by patients RA006 (range 23–103) and RA005 (40–85) and then patients RA004 (range 4–50) and RA000 (2–53) (Table S2). Similarly, patient RA003 had the largest and highest number of subclonal clusters, followed by patients RA006 and RA005 (Figure S2A). In contrast, patients RA004 and RA000 had very few subclonal clusters (Figure S2A), demonstrating differences in clonal structure between patients. However, mutations that may be lost because of copy-number loss were not accounted for in these analyses (McPherson et al., 2016). Collectively, these results indicate that intra- and inter-metastatic mutational heterogeneity can vary considerably among patients, with extreme heterogeneity evident in a subset of patients.

## APOBEC Mutagenesis Strongly Correlates with Mutational Tumor Heterogeneity

Next, we generated phylogenetic trees by two independent methods, and both showed a similar evolutionary pattern (Figures 3 and S2B–S2E). We then analyzed mutational signature profiles for each tumor by layering phylogenetic trees with mutational signatures to elucidate whether specific mutational processes could explain the observed variability in mutational tumor heterogeneity. Smoking signature mutations (C → A) were highly prevalent in patients RA000 and RA004, who were smokers (Figure 3). Mutations generated by the cytosine deaminase activity of APOBEC were most prevalent in patients RA003, RA005, and RA006, all of whom were non-smokers (Figure 3; Data S3). Smoking signature mutations were generally truncal, whereas APOBEC signature mutations were largely in the shared and private branches of the phylogenetic trees (Figures 3A–3E; Data S3).

To further assess the timing of APOBEC-induced mutagenesis, we evaluated mutational signatures in available tumors collected at the time of diagnosis. We found no evidence of APOBEC mutagenesis in these samples, including those from patients RA003 and RA006 (Figures 3A and 3B), indicating that APOBEC signature mutations were acquired later during further progression of metastatic disease. Next, we evaluated the relationship between APOBEC mutagenesis and mutational heterogeneity. APOBEC mutation fold enrichment, a measure of APOBEC mutagenesis (Roberts et al., 2013) (Table S4), strongly correlated with Jaccard similarity coefficients based on WES variants (Pearson rho = −0.66, $p = 2.2 \times 10^{-16}$) and expressed variants by RNAseq (Pearson rho = −0.55, $p = 7.704 \times 10^{-8}$) (Figures 3F and 3G). To validate this finding, we examined the relationship between APOBEC mutagenesis and mutational heterogeneity in the TRACERx cohort of 17 LUAD patients with five or more sites of multi-region exome sequencing (133 total tumors)(Jamal-Hanjani et al., 2017). In line with findings in our cohort, we found APOBEC mutagenesis strongly correlated with Jaccard similarity coefficients (Pearson rho = −0.4689, $p = 1.255 \times 10^{-8}$) in the TRACERx

cohort (Figure 3H). Thus, these results suggest that APOBEC mutagenesis is a significant contributor to the generation of mutational tumor heterogeneity.

## Expression of *APOBEC3* Region Transcripts Correlates with APOBEC Mutagenesis

Previous analyses of early-stage TCGA LUAD patients have shown a weak but significant association between expression of *APOBEC3* transcripts and APOBEC mutagenesis (Chan et al., 2015). We examined RNA-seq data from our metastatic cohort to determine whether *APOBEC3* region transcript expression contributed to the observed variability in APOBEC mutagenesis. Among the LUAD patients, we found *APOBEC3B* to be expressed at higher levels than *APOBEC3A*; this was particularly evident in the tumors of patient RA003 (Figure 4A). APOBEC mutagenesis (measured here by the counts of tCw to tTw and tGw mutations) (Roberts et al., 2013) was highly correlated with *APOBEC3B* expression (Pearson rho = 0.68, $p = 9.4 \times 10^{-5}$), but not with *APOBEC3A* expression (Pearson rho = 0.19, p = 0.34) (Figures S3A and S3b). Analysis of isoform-specific expression of *APOBEC3A* and *APOBEC3B* with custom TaqMan assays confirmed these results (Figures S3C and S3D). Moreover, in multiple tumors of patient RA003 with significant APOBEC mutagenesis, *APOBEC3B* was expressed 20- to 50-fold higher than *APOBEC3A* (Figures 4A and 4B).

To elucidate factors affecting APOBEC mutagenesis in our patients, we genotyped an *APOBEC3* germline variant, rs12628403, associated with increased APOBEC mutagenesis (Nik-Zainal et al., 2014). This germline variant is a proxy for a 30-kb deletion that fuses the coding region of *APOBEC3A* with the 3′ UTR of *APOBEC3B* to generate a chimeric *APOBEC3A-APOBEC3B* (*APOBEC3AB*) transcript. This chimeric transcript is more stable than the *APOBEC3A* transcript and leads to higher APOBEC3A protein levels *in vitro* (Caval et al., 2014). Only patient RA006, the TC patient, was a carrier of the rs12628403 allele and was predicted to generate the *APOBEC3AB* transcript (Figure 4A). Indeed, *APOBEC3AB* expression in tumors of this patient was identified by RNA-seq (Figure 4A) and validated using a TaqMan assay (Figure 4C). However, the *APOBEC3B* transcript was also expressed in tumors from this patient, suggesting that both may have contributed to APOBEC mutagenesis. Expression of *APOBEC3B* (Pearson rho = 0.77, p = 0.08) and *APOBEC3AB* (Pearson rho = 0.62, p = 0.18), but not *APOBEC3A* (Pearson rho = 0.23, p = 0.67), significantly correlated with APOBEC mutagenesis in all tumors from patient RA006 (Figures S4A–S4C).

Although expression of *APOBEC3AB* was lower than *APOBEC3B* (Figure 4C), APOBEC3A encoded by *APOBEC3AB* is considered a more potent inducer of mutagenesis than APOBEC3B (Caval et al., 2014; Landry et al., 2011). Therefore, we quantified the contribution of APOBEC3A and APOBEC3B to APOBEC mutagenesis by calculating YTCA and RTCA enrichment (where Y is a purine and R is a pyrimidine), attributed to differential activity of these enzymes (Chan et al., 2015). In all tumors with high APOBEC mutagenesis from patient RA006, there was significant enrichment of YTCA compared to RTCA (Figure S4D), thereby suggesting APOBEC3A-like mutagenesis as a likely driver of extreme heterogeneity in this metastatic TC patient with a germline *APOBEC3AB* deletion.

Together, our results implicate expression of *APOBEC3* region transcripts as a mediator of APOBEC mutagenesis in metastatic LUAD and TC.

### *TP53* Mutations Are Associated with APOBEC Hypermutator Status in LUAD

Given the role of *TP53* in repressing APOBEC3B expression (Periyasamy et al., 2017), we hypothesized that mutant *TP53*, detected only in the LUAD patient RA003, may be contributing to high *APOBEC3B* expression and increased APOBEC mutagenesis. We found mutant *TP53* was associated with APOBEC hypermutator status (Nik-Zainal et al., 2014) in our cohort and in three independent LUAD datasets (Imielinski et al., 2012; Jamal-Hanjani et al., 2014; Network, 2014) (Table S5). Moreover, mutant *TP53* was also associated with significantly higher counts of APOBEC signature mutations, higher expression of *APOBEC3B*, and an increase in *APOBEC3A*, compared with wild-type *TP53* tumors in the TCGA dataset (Figures 4D–4H). Thus, our results suggest mutant *TP53* contributes to increased *APOBEC3B* expression, APOBEC mutagenesis, and is associated with APOBEC hypermutator status in LUAD.

## Integration of Copy Number, Transcript, and Protein Abundance Highlights Mechanisms of Proteomic Heterogeneity

To evaluate multi-dimensional heterogeneity, we plotted Pearson correlation coefficients (PCCs) for each data type between pairs of tumors for each patient across all genes for which copy number, transcript expression, and protein abundance data were available (Figures 5A–5E). Each patient displayed variable patterns of heterogeneity across each data type. Patient RA003 exhibited the least (Figure 5A), whereas patient RA004 the most (Figure 5D) heterogeneity. Patient RA005 showed the most heterogeneity in gene expression and protein abundance only between tumor L5d and tumors L2a/L2b/L5c (Figure 5C). Patient RA006 showed lower heterogeneity in copy number, gene expression, and protein abundance within three pairs of tumors (L20b/L5a, Li3c/Li1a, L12a/L3a) compared to other pairs (Figure 5B).

Next, we plotted PCCs within each patient for each data type. Protein heterogeneity was significantly greater than gene expression and copy-number alteration (CNA) heterogeneity within all patients (Figure 5F). We then performed pairwise comparison of the PCCs of each data type, including a separate analysis using exome copy-number data that was corrected for tumor purity (Figure S5; Table S6). We found a strong, positive linear correlation between CNA heterogeneity and protein heterogeneity for patients RA004 and RA006, but not for the other patients (Figures S5A, S5D, S5G, S5J, and S5M), providing evidence that CNAs can lead to protein heterogeneity in these patients. Gene expression heterogeneity was also associated with protein heterogeneity but only in patients RA000 and RA006 (Figures 5B and 5N). Together, these results demonstrate high heterogeneity in protein abundance between metastases of these patients that partly stems from heterogeneity in CNAs and gene expression.

## Late-Event CNAs Contribute to Heterogeneity in Gene Expression and Protein Abundance

Next, we performed hierarchical clustering by chromosomal cytoband, gene expression, and protein abundance to further evaluate heterogeneity within each data type. Metastases from

each patient clustered together for CNAs, gene expression, and protein abundance (Figures 6A and S6A–S6C). Metastases from patients RA004 and RA006 showed the lowest correlation in protein abundance (Figures S6D and S6E) and clear differences in CNAs (Figure 6A; Table S7). To explore the downstream effects of CNAs, we plotted gene expression and protein abundance ratios of genes within each chromosomal arm between metastatic lineages of each patient (Figures 6B and 6C; Table S8; Data S4). Arm-level CNAs within tumors of patients RA006 and RA004 corresponded with changes in gene expression and protein abundance of genes at the corresponding arm level. For example, copy-number differences in arm 4p between RA006 tumors corresponded with changes in expression and protein abundance (Figure 6B). Where there were no copy-number differences, such as in arm 7q in patient RA006 and arms 4p/7q of patient RA003, there were no corresponding changes in gene expression and protein abundance (Figure 6C).

Heterogeneity in focal-level CNAs also corresponded with changes in RNA and protein. For example, *CCND1* was highly amplified in patient RA004 tumors L2a, L3a, and L8b (Table S9), which corresponded with high gene and protein expression. On the other hand, tumors L1b and L1c in which there was minimal increase in *CCND1* copy number and gene and protein expression were low (Figures 6D and 6E). Interestingly, among patient RA004 liver and kidney tumors, Li1b and K1, *CCND1* was highly amplified with correspondingly high protein but moderate gene expression, suggesting tissue-specific discordance of gene and protein expression of select genes (Figures 6D and 6E).

Next, we constructed circos plots (Data S5) and phylogenetic trees based on CNAs for each patient. Both arm- and focal-level CNAs largely occurred early in tumor development (i.e., truncal) in patients RA000, RA003, and RA005 (Figures 3 and S7A, S7B, and S7D) but occurred later in tumor development (i.e., shared and private) in patients RA004 and RA006 (Figures 3, S7C, and S7E). These results suggest that late, not early, CNAs likely contributed to the observed changes in gene expression and protein abundance between metastases of these patients. Differential focal and arm-level CNAs may reflect ongoing chromosomal instability as well as selective pressure during evolution of metastatic lineages.

## Enrichment of Interferon Signaling Pathways in Tumors with High *APOBEC3* Expression and Immune Heterogeneity

Next, we sought to decipher common gene sets or pathways within the RNA-seq and mass-spectrometry-based quantitative proteomics data that were heterogeneously enriched within each patient. Using an unbiased approach with single-sample gene set enrichment analysis (ssGSEA) (Barbie et al., 2009; Subramanian et al., 2005), we found interferon (IFN)-signaling pathways (related to possible activity of the IFNs: type I [IFN-α, IFN-β], type II [IFN-γ], and type III [IFN-λ1–4]), were the most significantly and differentially enriched pathways within patient RA003 (tumors L1 and L4a) and RA006 (tumors L20b, L5a, Li1a, and Li3c) (Figures 7A–7D) at both gene expression and protein abundance levels. No common outlier gene sets were identified for the negatively enriched pathways. Within the tumors from patients RA000, RA004, and RA005, no recurrent, common pathways were identified by GSEA of both the transcriptome and proteome (Table S9). The six tumors from patients RA003 and RA006 that were enriched in IFN-signaling pathways also had the

highest expression of *APOBEC3* region transcripts (Figure 4A). To validate these findings in a larger cohort, we interrogated the TCGA LUAD dataset. We found the expression of *STAT1*, a downstream effector of IFN-γ signaling, was higher in LUAD tumors with increased expression of both *APOBEC3A* and *APOBEC3B* (Figures S8A and S8B). Multiple other genes within IFN-α and IFN-γ pathways were also significantly associated with expression of *APOBEC3A* and *APOBEC3B* in the TCGA LUAD dataset (Table S10). To test whether *APOBEC3A* and *APOBEC3B* can be induced by IFNs, we treated three LUAD (A549, HCC4006, and H1975) and two immortalized normal lung epithelial cell lines (HBEC and HPL1D) with IFN-γ. These cell lines were chosen to represent major driver subtypes of LUAD (KRAS and EGFR) and normal lung epithelium. At baseline, expression of *APOBEC3B* was generally higher than expression of *APOBEC3A*. Treatment with IFN-γ led to a significant increase in expression of *APOBEC3B*, whereas expression of *APOBEC3A* increased significantly, but marginally, in only A549 and HCC4006 cells (Figure 7E). Thus, our results suggest IFN-signaling contributed, in part, by the tumor microenvironment, is a potential mechanism of heterogeneity of LUAD tumors with increased *APOBEC3B* transcript expression.

To further interrogate heterogeneity in the immune microenvironment between tumors of each patient in an unbiased manner, we analyzed the gene expression and protein abundance data using ssGSEA based on CIBERSORT immune genes (Newman et al., 2015) and an alternative immune scoring method (Aran et al., 2017) (Table S10). The overall immune signature score for tumors within each patient considerably varied between transcriptome and proteome in patients RA003 and RA005 (Figures 7F and 7G). Patients RA000 and RA004 showed low and high overall immune signature scores, respectively. In contrast, tumors from patient RA006 showed large and consistent differences in immune signature scores across both the transcriptome and proteome (high in L20b and L5a vs. low in L12a, L3a, Li1a, Li3c), demonstrating heterogeneous immune cell infiltration in the tumor microenvironment of this patient (Figures 7F and 7G).

## DISCUSSION

Genomic, transcriptomic and proteomic analyses of tumors from multiple anatomic sites sampled through rapid autopsy offer a unique opportunity to comprehensively explore the biological processes that shape the evolution of metastatic tumors. Here, we have characterized the proteogenomic evolution of metastatic lung and thymic carcinoma through exome and transcriptome sequencing, CNA analysis, and unbiased quantitative mass spectrometry-based proteomics of 37 metastatic tumors acquired through rapid autopsy. Through this integrative analysis, we have uncovered insights into the mechanisms likely driving the mutational, transcriptomic, and proteomic landscape of these metastatic tumors.

At the genomic level, we provide evidence that APOBEC mutagenesis is a driver of mutational heterogeneity in metastatic lung and thymic carcinoma tumors. APOBEC mutagenesis has been described as one of the most common mutational processes second only to “ageing” (Alexandrov et al., 2013). To date, however, within thoracic tumors, this process has been described mostly in primary tumors (Burns et al., 2013a, 2013b; de Bruin et al., 2014). These studies have shown APOBEC mutagenesis to be associated with

subclonal mutations that occur late in the evolution of primary tumors and within spatially distinct regions (de Bruin et al., 2014; Jamal-Hanjani et al., 2017). In our study, we identified a subset of patients with high APOBEC mutagenesis in metastases, but not in the primary tumors, suggesting APOBEC mutagnesis can generate mutations late in the evolution of metastatic disease.

Given that all patients in our cohort received prior treatment, we cannot exclude the possibility that therapy contributed to the observed findings. However, recent studies assessing tumors pre- and post-chemotherapy in multiple tumor types did not find an increase in overall mutational load or a significant increase in APOBEC signature mutations (Liu et al., 2017; Noorani et al., 2017). Moreover, the patients in our cohort with the highest level of APOBEC mutagenesis, RA003 and RA006, received the least therapy. Additionally, we demonstrated increased mutational heterogeneity with increasing APOBEC mutagenesis among the TRACERx cohort, which explored intra-tumor heterogeneity in primary lung cancer among patients without previous chemotherapy, further validating the results in our cohort of metastatic tumors (Figure 3H).

While the existence of mutational heterogeneity in metastases has been previously described (Campbell et al., 2010; Yachida et al., 2010), the mechanisms remained unclear (Vogelstein et al., 2013). Our data suggest APOBEC mutagenesis can generate both putative driver and passenger mutations late in metastases, thereby generating inter-metastatic mutational heterogeneity that in some cases can be extreme. These results stand in contrast to recent genomic studies of the metastases of patients with pancreatic (Makohon-Moore et al., 2017) and prostate (Kumar et al., 2016) cancer, which have shown limited mutational heterogeneity and no significant APOBEC mutagenesis (Roberts et al., 2013). Both of these tumor types have shown no evidence of APOBEC mutagenesis within primary tumors highlighting the likely histologic specificity of this process. Ultimately, the clinical importance of APOBEC mutagenesis will be determined by the response of heterogeneous metastatic tumors—with and without APOBEC signature mutations—to chemotherapy, targeted agents and/or immunotherapy.

Both APOBEC3A and APOBEC3B have been shown to localize to the nucleus (Lackey et al., 2013) leading to potent DNA damage (Taylor et al., 2013), deaminase activity and base substitutions in the genome (Burns et al., 2013a; Shinohara et al., 2012). Upregulation of *APOBEC3B* causes APOBEC signature mutations *in vitro* (Akre et al., 2016). Expression of *APOBEC3A* and *APOBEC3B* also has been associated with APOBEC signature mutations in primary tumors from multiple cancer types including LUAD (Burns et al., 2013a, 2013b; Chan et al., 2015; Leonard et al., 2013). However, it is unclear whether the same transcripts promote APOBEC mutagenesis in metastatic lung and thymic carcinoma tumors. In our set of metastatic tumors, we observed a strong correlation between expression of *APOBEC3B* and *APOBEC3AB* transcripts and APOBEC mutagenesis, suggesting expression of such transcripts may be a more dominant mechanism of APOBEC mutagenesis in metastatic thoracic tumors as opposed to earlier stage disease. These results are in line with recent work in breast cancer that has shown higher *APOBEC3B* expression in metastatic disease compared to early-stage primary tumors (Sieuwerts et al., 2017).

Mutant *TP53* has been previously associated with higher APOBEC signature mutations in breast cancer (Burns et al., 2013a) and there is recent evidence that TP53 can repress *APOBEC3B* expression through direct transcriptional regulation of its promoter (Burns et al., 2013a; Menendez et al., 2017; Periyasamy et al., 2017). Our results suggest mutant *TP53* may be an important contributor of increased *APOBEC3B* and subsequent generation of APOBEC signature mutations in LUAD. In particular, we show mutant *TP53* is associated with APOBEC hypermutator status in LUAD using three large-scale independent datasets. It is important to analyze APOBEC hypermutator and *TP53* mutation status and the resultant mutational heterogeneity in future clinical trials.

Expression of the *APOBEC3AB* transcript was captured in our study by the presence of the *APOBEC3* germline variant, rs1262840, within thymic carcinoma patient RA006. This germline variant has been previously associated with increased APOBEC signature mutations in primary breast cancer tumors (Nik-Zainal et al., 2014). Whether this variant is also associated with high APOBEC mutagenesis and mutational heterogeneity within thymic carcinoma is unknown, as no previous multi-region sequencing study has been performed on this rare tumor type. Additionally, to our knowledge, apart from the current study, this variant has not been examined in relation to APOBEC mutagenesis in metastatic disease. Given that this germline variant can be easily tested utilizing blood DNA, our results warrant further testing of the association between this *APOBEC3* germline variant with APOBEC mutagenesis and mutational tumor heterogeneity in thymic carcinoma as well as other metastatic tumor types.

Our integrated CNA, RNA-seq, and quantitative mass spectrometry study demonstrates that late-event CNAs can be important drivers in the evolution of metastatic cancer through downstream changes in transcript and protein abundance. Early studies in yeast showed CNAs for a given gene lead to proportional increases in protein abundance (Hughes et al., 2000; Pavelka et al., 2010; Rancati et al., 2008; Torres et al., 2007). More recently, studies in primary tumors demonstrated variability in CNA to protein cis-effects (Mertins et al., 2016; Zhang et al., 2014a, 2016). In metastatic disease, multiple studies have reported late-event CNAs (Campbell et al., 2010; Ding et al., 2010; Gerlinger et al., 2014; Hieronymus et al., 2014; Robinson et al., 2015; Yates et al., 2015), but the effect of CNAs on transcript and protein abundance has not been examined.

In one recent metastatic pancreatic cancer study, CNA differences among tumor suppressor genes were not evident at the protein level by IHC suggesting late-event CNAs can be stochastic changes rather than evolutionary selected events (Makohon-Moore et al., 2017). In the current study, all patients had some evidence of late-event CNAs. However, only patients with significant differences in late-event CNAs between tumors exhibited corresponding differences in transcript and protein abundance. In light of the recent association between copy-number heterogeneity and increased recurrence and death in early-stage NSCLC (Jamal-Hanjani et al., 2017), our results raise the question of whether late-event CNAs, through downstream effects on gene expression and protein abundance, can result in worse outcomes for a subset of patients with metastatic cancer. Proteomic heterogeneity induced, in part, by CNAs may also explain why chromosomal instability (CIN) has been associated with poor outcomes in cancer (Carter et al., 2006; Choi et al.,

2009; Jamal-Hanjani et al., 2017; Walther et al., 2008). Importantly, proteomic heterogeneity in our set of metastatic tumors, was much higher than CNA and transcriptomic heterogeneity, suggesting other mechanisms, such as epigenetic and post-translational modifications, also may be important drivers of proteomic heterogeneity.

Through our unbiased analysis of transcriptomic and proteomic data, we found enrichment of IFN signaling within the microenvironment of tumors of patients with the highest *APOBEC3* region transcript expression. Moreover, we show that in LUAD and normal lung epithelial cell lines, *APOBEC3B* expression, and in some LUAD cell lines, *APOBEC3A* expression can be induced by IFNγ treatment. These data suggest IFN signaling within the tumor microenvironment may, in part, influence *APOBEC3* region transcript expression and thereby contribute to heterogeneity in APOBEC signature mutations within the tumors of a given patient.

We also found transcriptomic and proteomic heterogeneity in immune signatures within and between patients. Major advances have been made in the treatment of metastatic tumors, including LUAD and TC, through immunotherapies such as immune checkpoint blockade (Borghaei et al., 2015; Giaccone et al., 2018). Nonetheless, only a subset of patients responds and multiple metastases within a given patient may respond differently because of immune heterogeneity (Reuben et al., 2017). Even without immunotherapy, metastases within a patient may also have differing tumor immune microenvironments, as we demonstrate within TC patient RA006 and as has been recently shown within an ovarian cancer patient (Jiménez-Sánchez et al., 2017). We further demonstrate that, within a given patient, the tumor immune microenvironment may exhibit substantial differences between the transcript and protein expression, adding to the complexity of assessing the immune microenvironment.

One of the strengths of our study is the comprehensive examination of tumor heterogeneity by integrating genomic (exome, CNA), transcriptomic (RNA-seq) and proteomic (global mass spectrometry analysis for protein abundance) data from multiple metastatic tumors procured through rapid autopsy. One of the limitations of our study is that all autopsy patients in this study were diagnosed at the late stage of metastatic disease. Hence, we were unable to conduct a complete temporal analysis of tumor evolution from early- to late-stage disease. The ongoing $TRACER_x$ study (Jamal-Hanjani et al., 2014, 2017) and corecruitment of those patients to the PEACE (Posthumous Evaluation of Advanced Cancer Environment) post-mortem study (Abbosh et al., 2017) will allow for better elucidation of the evolution primary tumor to metastatic advanced disease, including at the end of life.

In conclusion, in this report, we present the heterogeneous genomic, transcriptomic, and proteomic landscape of metastatic lung and thymic carcinoma as well as identify possible mechanisms underlying such multi-level heterogeneity. High activity of the APOBEC3 enzymes, represented by transcript expression, and modulated by germline variation, mutant *TP53*, and the immune microenvironment, can greatly alter the genomic landscape between metastatic tumors of a given patient. Arm-level and focal CNAs occurring later in tumor evolution can generate significant downstream heterogeneity through effects on gene expression and protein abundance. Further studies by comprehensive analyses of multiple

metastatic sites from larger patient populations, including different tumor types, are warranted to validate these mechanisms. Such an endeavor requires development of rapid autopsy programs, meticulous collection and processing of tumors from all possible sites of disease and integrated “omics” analyses. These tumor heterogeneity studies will be integral for evaluating the outcomes of ongoing clinical trials, developing new paradigms in clinical trial design, and, ultimately, to improve survival for patients with metastatic cancer.

## STAR★METHODS

### CONTACT FOR REAGENT AND RESOURCES SHARING

Further information and requests for reagents may be directed to and will be fulfilled by the Lead Contact, Udayan Guha (udayan.guha@nih.gov).

### EXPERIMENTAL MODEL AND SUBJECT DETAILS

**Biospecimen Acquisition—**Samples were obtained from five patients diagnosed with thoracic malignancies who underwent rapid autopsy. Informed consent for rapid autopsy was obtained under an IRB approved protocol 13-C-0131 (NCT01851395) entitled “A Pilot Study of Inpatient Hospice with Procurement of Tissue on Expiration in Thoracic Malignancies.” Patients previously treated at the NCI and with life expectancy less than 3 months were offered inpatient hospice treatment at the Clinical Center of the National Institutes of Health and upon death autopsies were initiated within 3 hours. Clinical information, including the sex, gender of each patient is available in Table S1. One patient, RA003, elected to receive end of life care at home and was subsequently transported to the NIH Clinical Center post-mortem. Prioritization of lesions removed at autopsy was based on CT scan performed within one month before death. All tumors within each patient were removed by an experienced pathologist and macro dissected to remove surrounding non-neoplastic tissue. Punch biopsy needles were used to obtain spatially distinct cores from each tumor. One-third of each tissue core sample was fixed in 10% buffered formalin, one-third in optimal cutting temperature compound (OCT) and the remaining tissue was immediately flash frozen in liquid nitrogen and stored at −80°C. For each tissue sample, a 5-μm section was taken to create a hematoxylin and eosin slide to visualize neoplastic cellularity using a microscope.

For each patient, normal tissue, if available, and/or a blood sample was used as a normal control. DNA and RNA was isolated from approximately 30 mg of snap-frozen tumor tissue using the All Prep DNA/RNA Mini Kit (QIAGEN). RNA was partially degraded with an average RNA integrity number 5.17 but was comparable between organs of different patients and similar in quality to previous postmortem studies. To ensure adequate quality, samples RA003_L2f, RA006_LN2a, RA005_L4a were removed post-sequencing but before analyses due to low tumor content (less than 20 percent) based on Sequenza purity estimates.

### METHOD DETAILS

**Whole-exome sequencing data processing, variants calling, filtering and annotation—**Whole-exome sequencing of tumor and normal samples was performed at a sequencing core at the NCI Frederick National Laboratory at the National Cancer Institute

(NCI). Libraries were constructed and then sequenced as 2 × 126 nt paired-end reads with Illumina HiSeq2500 sequencers. Mean coverage depth was 161x (range 114x to 231x). Raw sequencing data in FASTQ format were aligned against the reference human genome (hg19) with BWA (Li and Durbin, 2009). The alignment BAM files were further processed following GATK's best practices (McKenna et al., 2010) with Picard tools, namely MarkDuplicates, IndelRealigner, and BaseRecalibrator. Somatic variants were then called from the processed BAM files using Strelka (v1.0.10) (Saunders et al., 2012) with the default version of BWA configuration file. The identified somatic variants reported in the "passed" v*cf* files by Strelka were used for further analysis. Variants were functionally annotated using snpEff/snpSift version 3.4 (see URLs) with databases of GRCh37.70 and dbNSFP version 3.4, and the types of variants were filtered using snpSift (Cingolani et al., 2012). If a variant was also reported in one of the three public databases: 1000 Genomes Project, ExAC, and ESP-NHLBI with a MAF greater than 5%, the variant was removed. For each patient, variants identified by Strelka (Saunders et al., 2012) from all tumor regions were combined to get a unique variant list. Using this patient-specific list, if a variant in a particular tumor site was not called by Strelka, the Samtools mpileup was used to retrieve the reference and alternative reads coverage for each SNV site. If the site had > = 2 alternative reads and VAF > = 1%, the SNV was considered present in the tumor site. For short indels, if the variant site was not in the "passed" v*cf* file, the Strelka called "all" v*cf* file is used to retrieve the reference and alternative reads coverage; if missing in the "all" v*cf* file, the indel site was considered absent. Patient RA000 was previously known to have a *KRAS* G12C mutation based on molecular profiling. Although this mutation did not pass variant filtering, it was noted on manual review. Identified missense mutations were manually reviewed using the Integrative Genomics Viewer version 2.4 (Robinson et al., 2011; Thorvaldsdóttir et al., 2013).

**Phylogenetic analysis—**Phylogenetic analysis was conducted using Phangorn (Schliep, 2011) and phytools R packages with all identified variants (silent and non-silent) from all tumor sites in each patient after converting the mutation profile into binary format. The initial phylogenetic relationships between tumor regions for an individual patient was inferred using both the Maximum Parsimony and the Unweighted Pair Group Methods (UPGMA). Phylogenetic trees were then redrawn by hand in Adobe Illustrator with branch length proportional to the number of mutations specific to one tumor (private), two or more tumors (shared) or all tumors (trunk). Driver mutations and focal CNAs were added to the branches. All non-synonymous and synonymous mutations were used for tree construction. Phylogenetic trees were also constructed for each patient using the Treeomics computational tool (Reiter et al., 2017) using bootstrapping values from 1,000 samples except for patient RA004, which failed to run likely due to the large number of tumors. Signature analysis was performed for each individual tumor as well as trunks and each subsequent branch point for each tumor using deconstructSigs (Rosenthal et al., 2016). Mutations that were not signature 1, 2, 4, 5, or 13-type were labeled as "unclassified." Mutational signature analysis was restricted to branches with at least 10 mutations. COSMIC mutational signatures were calculated for each branch using the R "deconstrucSigs" package (Rosenthal et al., 2016). Arm-level amplification and deletions were also added to previously generated phylogenetic

trees for each patient (Figure S7). For these trees, branch lengths were redrawn for visualization purposes only.

**Identification and classification of driver mutations—**All identified nonsynonymous mutations were filtered to include only driver genes based on large-scale non-small cell lung cancer sequencing studies (Campbell et al., 2016; Ding et al., 2008; Govindan et al., 2012; Imielinski et al., 2012; Lawrence et al., 2014; Network, 2014; Weir et al., 2007) and in the COSMIC cancer gene census (downloaded June 2016). We classified all nonsynonymous mutations into the three categories. Category 1 ‘high-confidence driver mutations’ contained all disrupting mutations (nonsense, frameshift, splicing or ‘deleterious’ missense) in tumor suppressor genes or activating amino acid substitutions in non-small cell lung cancer oncogenes as described in lung cancer sequencing studies. Category 2 ‘putative driver mutations’ contained amino acid substitutions located at the same position or up to 5 amino acids away from a substitution present in COSMIC. Category 3 ‘low confidence driver mutations’ contained all other nonsilent mutations in genes that were present in the lists of cancer-related genes described above. Mutations were then analyzed using COSMIC to determine whether the amino acid substitution has been previously identified. Category 2 were further scored as ‘deleterious’ when at least two out of the three predictors classified the mutation as deleterious Functional prediction scores (SIFT, PolyPhen2, and Provean). All category 1 mutations were considered deleterious and category 3 mutations were not included as driver mutations in any analyses.

**APOBEC germline deletion genotyping—**Germline *APOBEC3AB* deletion was genotyped by a proxy SNP rs12628403 using a custom-designed TaqMan genotyping assay, as described previously (Middlebrooks et al., 2016). For patient RA006, the deletion status was also confirmed in all six tumors by Sanger sequencing, and by expression analysis.

**qRT–PCR analysis—**Lung cell lines A549, HCC4006, and H1975 and lung epithelial cell lines HBEC-3KT and HPL1D (Masuda et al., 1997) were grown on 6-well plates in triplicate until confluent, followed by treatment with IFNγ (1 ng/ml, R&D Systems) for 8 and 24 hours. Cells were lysed with RLT buffer supplemented with B-mercaptoethanol. Total RNA for cell lines and tumors was isolated with the QIAGEN All Prep DNA/RNA Mini Kit with on-column DNase I treatment. cDNA was prepared from equal amounts of total RNA for each sample with the RT2 first-strand cDNA kit and random hexamers with an additional DNA removal step (QIAGEN). Expression of *APOBEC3A*, *APOBEC3B*, and *APOBEC3AB* deletion and endogenous controls *GAPDH* and *PPIA* was measured in each cDNA with TaqMan expression assays from Thermo Fisher. Custom assays were used for *APOBEC3B*: (F: TGCTGGGAAAACTTTGTGTACAAT; R: ATG TGTCTGGATCCATCAGGTATCT; Probe: FAM-ATTCATGCCTTGGTACAAA), and *APOBEC3AB* (F: ATCATGACCTACGATGAATTT AAGCA; R: AGCACATTGCTTTGCTGGTG; Probe: FAM-CATTCTCCAGAATCAGGG), and commercial assay Hs00377444_m1 was used for *APOBEC3A*, 4326317E for *GAPDH* and 4326316E for *PPIA* (Thermo Fisher Scientific). Reactions were performed in four technical replicates on QuantStudio 7 (Life Technologies) using TaqMan Gene Expression buffer (Life Technologies); water and genomic DNA were used as negative controls for all assays.

Expression was measured by Ct values (PCR cycle at detection threshold). Expression of *APOBEC3A*, *APOBEC3B* and *APOBEC3AB* was individually normalized by the geometric mean of endogenous controls (*GAPDH* and *PPIA*) based on relative quantification method as ΔCt = Ct (control) – Ct (target).

**Analysis of APOBEC mutagenesis—**APOBEC-signature mutation analysis for all autopsy tumor samples was determined using an R software package kindly provided by Dr. Dmitry A. Gordenin. We used two variables in the file *_sorted_sum_all_fisher_Pcorr.txt: the 'tCw_to_G+tCw_to_T' variable, which represents total counts of APOBEC-signature mutations, and the 'APOBEC_Enrich' variable, which accounts for statistical significance of enrichment and represents the level APOBEC mutagenesis pattern per sample. This second variable is more stringent, as many samples were not enriched at a statistically significant level and were classified as negative for APOBEC-signature mutations. We identified APOBEC 'hypermutators' as those with signature 2 + 13 mutations (TCGA and Broad datasets) or total APOBEC mutations (TRACERx and our dataset) exceeding 1.5 times the length of the interquartile range from the 75th percentile (Nik-Zainal et al., 2014). We also used the same R software package to determine RTCA and YTCA enrichment for patient RA006. A less stringent filtering of whole-exome variants was used to provide sufficient sample size for this analysis. A Benjamin-Hochberg P value of 0.05 was used as a threshold for significance, unless specified otherwise, and all tests were two-sided.

**TCGA analyses—**TCGA data for lung adenocarcinoma were downloaded directly from the Firehose pipeline of the Broad Institute. For gene expression, we used both RNA-SeqV1 (Reads Per Kilobase per Million, RPKM) or RNA-SeqV2 (RNA-Seq by Expectation Maximization, RSEM).

**Copy Number Alteration Analysis—**Copy number alteration (CNA) analysis was performed using MIP array technology (Affymetrix OncoScan FFPE Express 2.0) with 334,183 sequence tag site probes which were used to measure DNA copy number at different loci across the human genome. Copy number data were processed and using the Affymetrix OSCHP-SNP-FASST2 algorithm within the Nexus Copy Number Software, which corrects for tumor ploidy using median centering. We used a log2 ratio cut-off of ± 0.5 to define focal copy number amplifications and deletions and ± 0.25 to define arm-level copy number amplifications and deletions. To minimize overcalling heterogeneity of copy number alterations, we employed the following methods: 1) tumors without ± 0.5 focal amplification/deletion were included if they had at log2 ratio ± 0.2 and tumors without ± 0.25 arm-level amplification/deletion were included if they had a log2 ratio ± 0.10; 2) at least two tumors within a patient were required to have an amplification or deletion above the threshold of ± 0.5 for focal and ± 0.25 for arm; 3) an amplification/deletion was considered truncal if present in > 80% of the tumors within a given patient. Allele-specific focal copy number profiles were determined for primary tumors (available for three patients) using the Sequenza package. Circos plots were generated using segmented GISTIC-output file for all tumors using circos v0.69–4, for every track the min and max are set to −1 and 1 respectively, values between −0.2 and 0.2 are not shown in the figure (Mermel et al., 2011). Arm-level changes as depicted in the copy number phylogenetic trees were determined using

GISTIC. We also determined allele-specific copy number, corrected for tumor ploidy and purity, from whole exome sequencing data using the FACETS package (Shen and Seshan, 2016).

**RNA-sequencing and data processing—**RNA-seq sequencing was performed on 31 out of 37 tumor sites. RNA-seq was done on Illumina HiSeq2500 platform to yield at least 100 million reads/sample using Illumina TruSeq V4 chemistry at 2 × 125 nt paired-end. Sequencing reads were aligned with TopHat version 2.0.13 (Trapnell et al., 2009) against the reference human genome hg19, with UCSC known gene transcripts as the gene model annotation. Expression on gene and isoform level was quantified with Cufflinks version 2.2.1 (Trapnell et al., 2012).

**RNA-seq variant calling and mutation validation—**For RNASeq variants calling, sequencing reads were first aligned to hg19 with STAR version 2.4.2a (Dobin et al., 2013) and then with a second pass alignment to the transcriptome generated by STAR for each patient. For each identified SNV in WES, its expression was confirmed by the presence of sequencing reads of the alternative allele assessed by Samtools mpileup on TopHat generated BAM files from RNASeq data, whereas alternative reads coverage for indels were extracted from v*cf* files generated by the GATK best practices variant calling on RNASeq (see URLs). 69% of whole exome variants had a minimum 1X RNA depth and of these expressed variants 59% were confirmed by RNA-seq (data not shown). 55% of whole exome variants had minimum of 5X RNA depth and of these expressed variants 69% were confirmed by RNA-seq (data not shown). Validation rates for different variant types across all tumor samples were similar (range 42% silent to 56% nonsense) (see Data S1).

**RNA-seq data analysis—**Cufflinks outputted FPKM values for each gene were normalized for all samples within each patient using limma package voom quantile method (Law et al., 2014). This expression data was used to predict enrichment scores among immune genes obtained from CIBERSORT for each sample (Newman et al., 2015) and then using single-sample GSEA (ssGSEA) from GenePattern. Using the R package “fgsea,” GSEA preranked we performed to determine enrichment scores for REACTOME pathways. Pathways with q-value less than 0.05 were considered significantly enriched. Principal component analysis (PCA) was used to combine clustered samples prior to conducting this analysis.

**Protein Extraction—**All but one tumor (RA004 – Li1a) had sufficient tissue for mass-spectrometry (MS)-based proteomic characterization. About 10–15 mg of tumor tissue fresh-frozen in liquid nitrogen was lysed in 400μl of urea lysis buffer (20 mM HEPES pH 8.0, 8 M urea, 1 mM sodium orthovanadate, 2.5 mM sodium pyrophosphate and 1 mM β-glycerophosphate) using a tissue lyser (QIAGEN). Lysates were centrifuged at 14,000 rpm at 4°C for 10 mins and the clear supernatants were transferred to new tubes. Protein concentrations were determined by the Modified Lowry method (BioRad).

**Enzymatic Digestion—**The protein lysate was reduced with 45 mM dithriothreitol (Sigma Aldrich, MO), alkylated with 100 mM iodoacetamide (Sigma Aldrich, MO), and subsequently digested with modified sequencing grade Trypsin (Promega, Madison, WI) at

30°C overnight. The digest was then acidified using 0.1% TFA and the peptides were desalted using solid phase extraction C18 column (Supelco, Bellefonte, PA), and vacuum dried in a centrifugal evaporator.

**TMT-Labeling—**TMT10plex amine reactive reagents (0.8 mg per vial) (Thermo Fisher Scientific) were resuspended in 41 μL of anhydrous acetonitrile (ACN) and all 41 μL of each reagent was added to each sample and mixed briefly on a vortexer. Reactions were incubated at room temperature for 1 h, and then quenched by the addition of 8 μL of 5% hydroxylamine for 15 min and then combined at equal amount. All tumor tissues of LUAD patients RA000, RA003, RA005 and RA006 were pooled together to make a reference channel and labeled with $TMT^{10}$-126. In a separate TMT labeling experiment, tumor tissues from patient RA006 were pooled together to make a reference channel.

**Basic reversed phase liquid chromatography (RPLC) fractionation—**Basic RPLC separation was performed with a XBridge C18, 100 × 2.1 mm analytical column containing 5μm particles and equipped with a 10 × 2.1 mm guard column (Waters, Milford, MA) with a flow rate of 0.25 mL/min. The solvent consisted of 10 mM triethylammonium bicarbonate (TEABC) as mobile phase A, and 10 mM TEABC in ACN as mobile phase B. Sample separation was accomplished using the following linear gradient: from 0 to 1% B in 5min, from 1 to 10% B in 5min, from 10 to 35% B in 30min, and from 35 to 100% B in 5min, and held at 100% B for an additional 3min. A total of 96 fractions were collected during the LC separation in a 96-well plate in the presence of 12.5 μL of 1% formic acid. The collected fractions were concatenated into 12 fractions and dried in a vacuum centrifuge. One tenth of the peptides were injected directly for LC-MS/MS analysis.

**LC-MS/MS analyses—**Peptides separated/fractionated by basic reversed-phase chromatography were analyzed on an LTQ-Orbitrap Elite interfaced with an Ultimate™ 3000 RSLCnano System (Thermo Scientific, San Jose, CA). The dried peptides were loaded onto a nano-trap column (Acclaim PepMap100 Nano Trap Column, C18, 5 μm, 100 Å, 100 μm i.d. × 2 cm) and separated on an Easy-spray™ C18 LC column (Acclaim PepMap100, C18, 2 μm, 100 Å, 75 μm i.d. × 25 cm). Mobile phases A and B consisted of 0.1% formic acid in water and 0.1% formic acid in 90% ACN, respectively. Peptides were eluted from the column at 300 nL/min using the following linear gradient: from 4 to 35% B in 60min, from 35 to 45% B in 5min, from 45 to 90% B in 5min, and held at 90% B for an additional 5min. The heated capillary temperature and spray voltage were 275°C and 2kV, respectively. Full spectra were collected from m/z 350 to 1800 in the Orbitrap analyzer at a resolution of 120,000, followed by data-dependent HCD MS/MS scans of the fifteen most abundant ions at a resolution of 30,000, using 40% collision energy and dynamic exclusion time of 30 s.

**Proteomic Data Analysis—**Peptides and proteins were identified and quantified using the Maxquant software package (version 1.5.3.30) (Tyanova et al., 2016) with the Andromeda search engine (Cox and Mann, 2008). MS/MS spectra were searched against the Uniprot human protein database (May 2013, 38523 entries) and quantification was performed using default parameters for TMT10plex in MaxQuant. Corrected intensities of the reporter ions from TMT labels were obtained from the MaxQuant search. The relative

ratios were calculated for each channel to the reference channel. These ratios were then used to predict enrichment scores of overall immune signatures obtained from CIBERSORT (Newman et al., 2015) and Xcell (Aran et al., 2017) using single-sample GSEA (ssGSEA) from GenePattern and for REACTOME pathways by GSEA preranked through the R package "fgsea." Pathways with p value less than 0.10 were considered significantly enriched. Samples were similarly combined as described for the RNA-seq data analysis. Cufflinks outputted FPKM values for each gene were normalized for all samples within each patient using limma package voom quantile method.

**Construction and Immunohistochemistry of Tissue Microarray—**The physical construction of the TMA followed the guidelines previously used by the NCI Tissue Array Project. Each tumor from each autopsy patient was represented by 1 tumor core of 1mm that was taken from the original paraffin block. Serial 5μm sections were cut from the TMA block and used for immunohistochemical analysis. We used previously reported methods for immunohistochemical staining of TMAs (Hewitt, 2004).

**Integrating copy number, gene expression and protein abundance—**Pearson correlation coefficients (PCCs) were calculated across all common genes in copy number, gene expression and protein abundance data for each patient. Prior to calculating PCCs, gene expression data (RNA-seq FPKM) and protein abundance (protein ratios) were further normalized within each patient using the limma package with its voom quantile method. 3DPlots were created using the R package "scatterplot3d." For arm-level analyses, normalized gene expression and protein abundance data was categorized by chromosomal arm. The mean of clusters of tumors, as determined previously by PCA, were calculated for both sets of data. Ratios were calculated between clusters and then log2 transformed. Probability density plots were generated with 1% outliers removed and x axis of plots were restricted to −1 to 1 (log2 scale) for gene expression and −0.1 and +0.1 (log2 scale) for protein abundance for visualization purposes.

## QUANTIFICATION AND STATISTICAL ANALYSIS

All figures and graphs were generated using the "ggplot2" package available through the R statistical program. Linear regression, correlations and t tests were conducted though the R base packages. All tests were two-tailed and p values less than 0.05 were considered significant.

## DATA AND SOFTWARE AVAILABILITY

The sequencing and genotype data have been deposited at the database of Genotypes and Phenotypes (dbGaP), which is hosted by the National Center for Biotechnology Information (NCBI). The MS proteomics data in this study have been deposited in the ProteomeXchange Consortium (http://proteomecentral.proteomeexchange.org) via the PRIDE partner repository. The accession numbers for the data reported in this paper are dbGaP: phs001432.v1.p1 and PRIDE: PXD012845. Additional analyses, data and code are available at: https://github.com/nitinroper/Rapid-Autopsy-NCI.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

We thank Dr. Ann Berger and Dr. Jennifer Cheng from the Pain and Palliative Care Team; Susan Perry and Emerson Padiernos; clinical and research nurses; and social work and spiritual care teams for their clinical care of the patients in this study who were enrolled in hospice prior to autopsy. We also thank Willie Young and the Pathology residents who assisted with the autopsies. We thank Dr. Dmitry A. Gordenin, Dr. Kin Chan, Dr. Bing Zhang, and Dr. Jing Wing for their helpful discussions. The study was supported by federal funds from the Intramural Research Program, Center for Cancer Research, National Cancer Institute, National Institutes of Health. A.R.B. and L.P-O. were supported by the Intramural Research Program, Division of Cancer Epidemiology and Genetics (DCEG), National Cancer Institute, National Institutes of Health. A.G., A.-L.B., and A.R.P. were supported by the Intramural Research Program, National Library of Medicine, National Institutes of Health.

## REFERENCES

Abbosh C, Birkbak NJ, Wilson GA, Jamal-Hanjani M, Constantin T, Salari R, Le Quesne J, Moore DA, Veeriah S, Rosenthal R, et al.; TRACERx consortium; PEACE consortium (2017). Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. Nature 545, 446–451. [PubMed: 28445469]

Akre MK, Starrett GJ, Quist JS, Temiz NA, Carpenter MA, Tutt AN, Grigoriadis A, and Harris RS (2016). Mutation processes in 293-based clones overexpressing the DNA cytosine deaminase APOBEC3B. PLoS ONE 11, e0155391. [PubMed: 27163364]

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Børresen-Dale AL, et al.; Australian Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MMML-Seq Consortium; ICGC PedBrain (2013). Signatures of mutational processes in human cancer. Nature 500, 415–421. [PubMed: 23945592]

Aran D, Hu Z, and Butte AJ (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. 18, 220. [PubMed: 29141660]

Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, Schinzel AC, Sandy P, Meylan E, Scholl C, et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature 462, 108–112. [PubMed: 19847166]

Blokzijl F, de Ligt J, Jager M, Sasselli V, Roerink S, Sasaki N, Huch M, Boymans S, Kuijk E, Prins P, et al. (2016). Tissue-specific mutation accumulation in human adult stem cells during life. Nature 538, 260–264. [PubMed: 27698416]

Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, Chow LQ, Vokes EE, Felip E, Holgado E, et al. (2015). Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. N. Engl. J. Med 373, 1627–1639. [PubMed: 26412456]

Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B, Refsland EW, Kotandeniya D, Tretyakova N, Nikas JB, et al. (2013a). APOBEC3B is an enzymatic source of mutation in breast cancer. Nature 494, 366–370. [PubMed: 23389445]

Burns MB, Temiz NA, and Harris RS (2013b). Evidence for APOBEC3B mutagenesis in multiple human cancers. Nat. Genet 45, 977–983. [PubMed: 23852168]

Campbell PJ, Yachida S, Mudie LJ, Stephens PJ, Pleasance ED, Stebbings LA, Morsberger LA, Latimer C, McLaren S, Lin ML, et al. (2010). The patterns and dynamics of genomic instability in metastatic pancreatic cancer. Nature 467, 1109–1113. [PubMed: 20981101]

Campbell JD, Alexandrov A, Kim J, Wala J, Berger AH, Pedamallu CS, Shukla SA, Guo G, Brooks AN, Murray BA, et al.; Cancer Genome Atlas Research Network (2016). Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. Nat. Genet 48, 607–616. [PubMed: 27158780]

Carter SL, Eklund AC, Kohane IS, Harris LN, and Szallasi Z (2006). A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. Nat. Genet 38, 1043–1048. [PubMed: 16921376]

Caval V, Suspène R, Shapira M, Vartanian JP, and Wain-Hobson S (2014). A prevalent cancer susceptibility APOBEC3A hybrid allele bearing APOBEC3B 3'UTR enhances chromosomal DNA damage. Nat. Commun 5, 5129. [PubMed: 25298230]

Chan K, Roberts SA, Klimczak LJ, Sterling JF, Saini N, Malc EP, Kim J, Kwiatkowski DJ, Fargo DC, Mieczkowski PA, et al. (2015). An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. Nat. Genet 47, 1067–1072. [PubMed: 26258849]

Choi CM, Seo KW, Jang SJ, Oh YM, Shim TS, Kim WS, Lee DS, and Lee SD (2009). Chromosomal instability is a risk factor for poor prognosis of adenocarcinoma of the lung: Fluorescence in situ hybridization analysis of paraffin-embedded tissue from Korean patients. Lung Cancer 64, 66–70. [PubMed: 18814932]

Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, and Lu X (2012). Using Drosophila melanogaster as a model for genotoxic chemical mutational studies with a new program, SnpSift. Front. Genet 3, 35. [PubMed: 22435069]

Cox J, and Mann M (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol 26, 1367–1372. [PubMed: 19029910]

de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, Yates L, Jamal-Hanjani M, Shafi S, Murugaesu N, Rowan AJ, et al. (2014). Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. Science 346, 251–256. [PubMed: 25301630]

Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, Cibulskis K, Sougnez C, Greulich H, Muzny DM, Morgan MB, et al. (2008). Somatic mutations affect key pathways in lung adenocarcinoma. Nature 455, 1069–1075. [PubMed: 18948947]

Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, Harris CC, McLellan MD, Fulton RS, Fulton LL, et al. (2010). Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature 464, 999–1005. [PubMed: 20393555]

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. [PubMed: 23104886]

Gerlinger M, Horswell S, Larkin J, Rowan AJ, Salm MP, Varela I, Fisher R, McGranahan N, Matthews N, Santos CR, et al. (2014). Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. Nat. Genet 46, 225–233. [PubMed: 24487277]

Giaccone G, Kim C, Thompson J, McGuire C, Kallakury B, Chahine JJ, Manning M, Mogg R, Blumenschein WM, Tan MT, et al. (2018). Pembrolizumab in patients with thymic carcinoma: a single-arm, single-centre, phase 2 study. Lancet Oncol. 19, 347–355. [PubMed: 29395863]

Govindan R, Ding L, Griffith M, Subramanian J, Dees ND, Kanchi KL, Maher CA, Fulton R, Fulton L, Wallis J, et al. (2012). Genomic landscape of non-small cell lung cancer in smokers and never-smokers. Cell 150, 1121–1134. [PubMed: 22980976]

Gundem G, Van Loo P, Kremeyer B, Alexandrov LB, Tubio JMC, Papaemmanuil E, Brewer DS, Kallio HML, Högnäs G, Annala M, et al.; ICGC Prostate Group (2015). The evolutionary history of lethal metastatic prostate cancer. Nature 520, 353–357. [PubMed: 25830880]

Hewitt SM (2004). Design, construction, and use of tissue microarrays. Methods Mol. Biol 264, 61–72. [PubMed: 15020780]

Hieronymus H, Schultz N, Gopalan A, Carver BS, Chang MT, Xiao Y, Heguy A, Huberman K, Bernstein M, Assel M, et al. (2014). Copy number alteration burden predicts prostate cancer relapse. Proc. Natl. Acad. Sci. USA 111, 11139–11144. [PubMed: 25024180]

Hughes TR, Roberts CJ, Dai H, Jones AR, Meyer MR, Slade D, Burchard J, Dow S, Ward TR, Kidd MJ, et al. (2000). Widespread aneuploidy revealed by DNA microarray expression profiling. Nat. Genet. 25, 333–337. [PubMed: 10888885]

Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, Cho J, Suh J, Capelletti M, Sivachenko A, et al. (2012). Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell 150, 1107–1120. [PubMed: 22980975]

Jamal-Hanjani M, Hackshaw A, Ngai Y, Shaw J, Dive C, Quezada S, Middleton G, de Bruin E, Le Quesne J, Shafi S, et al. (2014). Tracking genomic cancer evolution for precision medicine: the lung TRACERx study. PLoS Biol. 12, e1001906. [PubMed: 25003521]

Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, Shafi S, Johnson DH, Mitter R, Rosenthal R, et al.; TRACERx Consortium (2017). Tracking the evolution of non-small-cell lung cancer. N. Engl. J. Med 376, 2109–2121. [PubMed: 28445112]

Jiménez-Sánchez A, Memon D, Pourpe S, Veeraraghavan H, Li Y, Vargas HA, Gill MB, Park KJ, Zivanovic O, Konner J, et al. (2017). Heterogeneous tumor-immune microenvironments among differentially growing metastases in an ovarian cancer patient. Cell 170, 927–938.e20. [PubMed: 28841418]

Kumar A, Coleman I, Morrissey C, Zhang X, True LD, Gulati R, Etzioni R, Bolouri H, Montgomery B, White T, et al. (2016). Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. Nat. Med 22, 369–378. [PubMed: 26928463]

Lackey L, Law EK, Brown WL, and Harris RS (2013). Subcellular localization of the APOBEC3 proteins during mitosis and implications for genomic DNA deamination. Cell Cycle 12, 762–772. [PubMed: 23388464]

Landry S, Narvaiza I, Linfesty DC, and Weitzman MD (2011). APOBEC3A can activate the DNA damage response and cause cell-cycle arrest. EMBO Rep. 12, 444–450. [PubMed: 21460793]

Law CW, Chen Y, Shi W, and Smyth GK (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 15, R29. [PubMed: 24485249]

Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, Meyerson M, Gabriel SB, Lander ES, and Getz G (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. Nature 505, 495–501. [PubMed: 24390350]

Leonard B, Hart SN, Burns MB, Carpenter MA, Temiz NA, Rathore A, Vogel RI, Nikas JB, Law EK, Brown WL, et al. (2013). APOBEC3B upregulation and genomic mutation patterns in serous ovarian carcinoma. Cancer Res. 73, 7222–7231. [PubMed: 24154874]

Li H, and Durbin R (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. [PubMed: 19451168]

Liu W, Laitinen S, Khan S, Vihinen M, Kowalski J, Yu G, Chen L, Ewing CM, Eisenberger MA, Carducci MA, et al. (2009). Copy number analysis indicates monoclonal origin of lethal metastatic prostate cancer. Nat. Med 15, 559–565. [PubMed: 19363497]

Liu D, Abbosh P, Keliher D, Reardon B, Miao D, Mouw K, Weiner-Taylor A, Wankowicz S, Han G, Teo MY, et al. (2017). Mutational patterns in chemotherapy resistant muscle-invasive bladder cancer. Nat. Commun 8, 2193. [PubMed: 29259186]

Makohon-Moore AP, Zhang M, Reiter JG, Bozic I, Allen B, Kundu D, Chatterjee K, Wong F, Jiao Y, Kohutek ZA, et al. (2017). Limited heterogeneity of known driver gene mutations among the metastases of individual patients with pancreatic cancer. Nat. Genet 49, 358–366. [PubMed: 28092682]

Masuda A, Kondo M, Saito T, Yatabe Y, Kobayashi T, Okamoto M, Suyama M, Takahashi T, and Takahashi T (1997). Establishment of human peripheral lung epithelial cell lines (HPL1) retaining differentiated characteristics and responsiveness to epidermal growth factor, hepatocyte growth factor, and transforming growth factor beta1. Cancer Res 57, 4898–4904. [PubMed: 9354455]

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, and DePristo MA (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303. [PubMed: 20644199]

McPherson A, Roth A, Laks E, Masud T, Bashashati A, Zhang AW, Ha G, Biele J, Yap D, Wan A, et al. (2016). Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. Nat. Genet 48, 758–767. [PubMed: 27182968]

Menendez D, Nguyen TA, Snipe J, and Resnick MA (2017). The cytidine deaminase APOBEC3 family is subject to transcriptional regulation by p53. Mol. Cancer Res. 15, 735–743. [PubMed: 28232385]

Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, and Getz G (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. 12, R41. [PubMed: 21527027]

Mertins P, Mani DR, Ruggles KV, Gillette MA, Clauser KR, Wang P, Wang X, Qiao JW, Cao S, Petralia F, et al.; NCI CPTAC (2016). Proteogenomics connects somatic mutations to signalling in breast cancer. Nature 534, 55–62. [PubMed: 27251275]

Middlebrooks CD, Banday AR, Matsuda K, Udquim KI, Onabajo OO, Paquin A, Figueroa JD, Zhu B, Koutros S, Kubo M, et al. (2016). Association of germline variants in the APOBEC3 region with cancer risk and enrichment with APOBEC-signature mutations in tumors. Nat. Genet. 48, 1330–1338. [PubMed: 27643540]

Network, T.; Cancer Genome Atlas Research Network (2014). Comprehensive molecular profiling of lung adenocarcinoma. Nature 511, 543–550. [PubMed: 25079552]

Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, and Alizadeh AA (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods 12, 453–457. [PubMed: 25822800]

Nik-Zainal S, Wedge DC, Alexandrov LB, Petljak M, Butler AP, Bolli N, Davies HR, Knappskog S, Martin S, Papaemmanuil E, et al. (2014). Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. Nat. Genet 46, 487–491. [PubMed: 24728294]

Noorani A, Bornschein J, Lynch AG, Secrier M, Achilleos A, Eldridge M, Bower L, Weaver JMJ, Crawte J, Ong CA, et al.; Oesophageal Cancer Clinical and Molecular Stratification (OCCAMS) Consortium (2017). A comparative analysis of whole genome sequencing of esophageal adenocarcinoma pre- and post-chemotherapy. Genome Res. 27, 902–912. [PubMed: 28465312]

Patch AM, Christie EL, Etemadmoghadam D, Garsed DW, George J, Fereday S, Nones K, Cowin P, Alsop K, Bailey PJ, et al.; Australian Ovarian Cancer Study Group (2015). Whole-genome characterization of chemoresistant ovarian cancer. Nature 521, 489–494. [PubMed: 26017449]

Pavelka N, Rancati G, Zhu J, Bradford WD, Saraf A, Florens L, Sanderson BW, Hattem GL, and Li R (2010). Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. Nature 468, 321–325. [PubMed: 20962780]

Periyasamy M, Singh AK, Gemma C, Kranjec C, Farzan R, Leach DA, Navaratnam N, Pálinkás HL, Vértessy BG, Fenton TR, et al. (2017). p53 controls expression of the DNA deaminase APOBEC3B to limit its potential mutagenic activity in cancer cells. Nucleic Acids Res. 45, 11056–11069. [PubMed: 28977491]

Rancati G, Pavelka N, Fleharty B, Noll A, Trimble R, Walton K, Perera A, Staehling-Hampton K, Seidel CW, and Li R (2008). Aneuploidy underlies rapid adaptive evolution of yeast cells deprived of a conserved cytokinesis motor. Cell 135, 879–893. [PubMed: 19041751]

Reiter JG, Makohon-Moore AP, Gerold JM, Bozic I, Chatterjee K, Iacobuzio-Donahue CA, Vogelstein B, and Nowak MA (2017). Reconstructing metastatic seeding patterns of human cancers. Nat. Commun 8, 14114. [PubMed: 28139641]

Reuben A, Spencer CN, Prieto PA, Gopalakrishnan V, Reddy SM, Miller JP, Mao X, De Macedo MP, Chen J, Song X, et al. (2017). Genomic and immune heterogeneity are associated with differential responses to therapy in melanoma. NPJ Genom. Med 2, 10. [PubMed: 28819565]

Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, Kiezun A, Kryukov GV, Carter SL, Saksena G, et al. (2013). An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. Nat. Genet 45, 970–976. [PubMed: 23852170]

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, and Mesirov JP (2011). Integrative genomics viewer. Nat. Biotechnol 29, 24–26. [PubMed: 21221095]

Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, Montgomery B, Taplin ME, Pritchard CC, Attard G, et al. (2015). Integrative clinical genomics of advanced prostate cancer. Cell 161, 1215–1228. [PubMed: 26000489]

Rosenthal R, McGranahan N, Herrero J, Taylor BS, and Swanton C (2016). DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. 17, 31. [PubMed: 26899170]

Roth A, Khattra J, Yap D, Wan A, Laks E, Biele J, Ha G, Aparicio S, Bouchard-Côté A, and Shah SP (2014). PyClone: statistical inference of colonal population structure in cancer. Nat. Methods 11, 396–398. [PubMed: 24633410]

Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, and Cheetham RK (2012). Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics 28, 1811–1817. [PubMed: 22581179]

Schliep KP (2011). phangorn: phylogenetic analysis in R. Bioinformatics 27, 592–593. [PubMed: 21169378]

Shah SP, Morin RD, Khattra J, Prentice L, Pugh T, Burleigh A, Delaney A, Gelmon K, Guliany R, Senz J, et al. (2009). Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. Nature 461, 809–813. [PubMed: 19812674]

Shen R, and Seshan VE (2016). FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. Nucleic Acids Res. 44, e131. [PubMed: 27270079]

Shinohara M, Io K, Shindo K, Matsui M, Sakamoto T, Tada K, Kobayashi M, Kadowaki N, and Takaori-Kondo A (2012). APOBEC3B can impair genomic stability by inducing base substitutions in genomic DNA in human cells. Sci. Rep 2, 806. [PubMed: 23150777]

Sieuwerts AM, Schrijver WA, Dalm SU, de Weerd V, Moelans CB, Ter Hoeve N, van Diest PJ, Martens JW, and van Deurzen CH (2017). Progressive APOBEC3B mRNA expression in distant breast cancer metastases. PLoS ONE 12, e0171343. [PubMed: 28141868]

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, and Mesirov JP (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 102, 15545–15550. [PubMed: 16199517]

Taylor BJ, Nik-Zainal S, Wu YL, Stebbings LA, Raine K, Campbell PJ, Rada C, Stratton MR, and Neuberger MS (2013). DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. eLife 2, e00534. [PubMed: 23599896]

Thorvaldsdóttir H, Robinson JT, and Mesirov JP (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform 14, 178–192. [PubMed: 22517427]

Torres EM, Sokolsky T, Tucker CM, Chan LY, Boselli M, Dunham MJ, and Amon A (2007). Effects of aneuploidy on cellular physiology and cell division in haploid yeast. Science 317, 916–924. [PubMed: 17702937]

Trapnell C, Pachter L, and Salzberg SL (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–1111. [PubMed: 19289445]

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, and Pachter L (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc 7, 562–578. [PubMed: 22383036]

Tyanova S, Temu T, and Cox J (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat. Protoc 11, 2301–2319. [PubMed: 27809316]

Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr., and Kinzler KW (2013). Cancer genome landscapes. Science 339, 1546–1558. [PubMed: 23539594]

Walther A, Houlston R, and Tomlinson I (2008). Association between chromosomal instability and prognosis in colorectal cancer: a meta-analysis. Gut 57, 941–950. [PubMed: 18364437]

Weir BA, Woo MS, Getz G, Perner S, Ding L, Beroukhim R, Lin WM, Province MA, Kraja A, Johnson LA, et al. (2007). Characterizing the cancer genome in lung adenocarcinoma. Nature 450, 893–898. [PubMed: 17982442]

Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B, Kamiyama M, Hruban RH, Eshleman JR, Nowak MA, et al. (2010). Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature 467, 1114–1117. [PubMed: 20981102]

Yates LR, Gerstung M, Knappskog S, Desmedt C, Gundem G, Van Loo P, Aas T, Alexandrov LB, Larsimont D, Davies H, et al. (2015). Subclonal diversification of primary breast cancer revealed by multiregion sequencing. Nat. Med 21, 751–759. [PubMed: 26099045]

Zhang B, Wang J, Wang X, Zhu J, Liu Q, Shi Z, Chambers MC, Zimmerman LJ, Shaddox KF, Kim S, et al.; NCI CPTAC (2014a). Proteogenomic characterization of human colon and rectal cancer. Nature 513, 382–387. [PubMed: 25043054]

Zhang J, Fujimoto J, Zhang J, Wedge DC, Song X, Zhang J, Seth S, Chow CW, Cao Y, Gumbs C, et al. (2014b). Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. Science 346, 256–259. [PubMed: 25301631]

Zhang H, Liu T, Zhang Z, Payne SH, Zhang B, McDermott JE, Zhou JY, Petyuk VA, Chen L, Ray D, et al.; CPTAC Investigators (2016). Integrated proteogenomic characterization of human high-grade serous ovarian cancer. Cell 166, 755–765. [PubMed: 27372738]

Zhao ZM, Zhao B, Bai Y, Iamarino A, Gaffney SG, Schlessinger J, Lifton RP, Rimm DL, and Townsend JP (2016). Early and multiple origins of metastatic lineages within primary tumors. Proc. Natl. Acad. Sci. USA 113, 2140–2145. [PubMed: 26858460]

## Highlights

- APOBEC mutagenesis correlates with metastatic mutational heterogeneity
- Mutant *TP53* is associated with APOBEC hypermutator status
- IFN-γ signaling and high-risk *APOBEC3* germline variants modulate *APOBEC3* expression
- Late CNAs correlate with downstream transcriptomic and proteomic heterogeneity

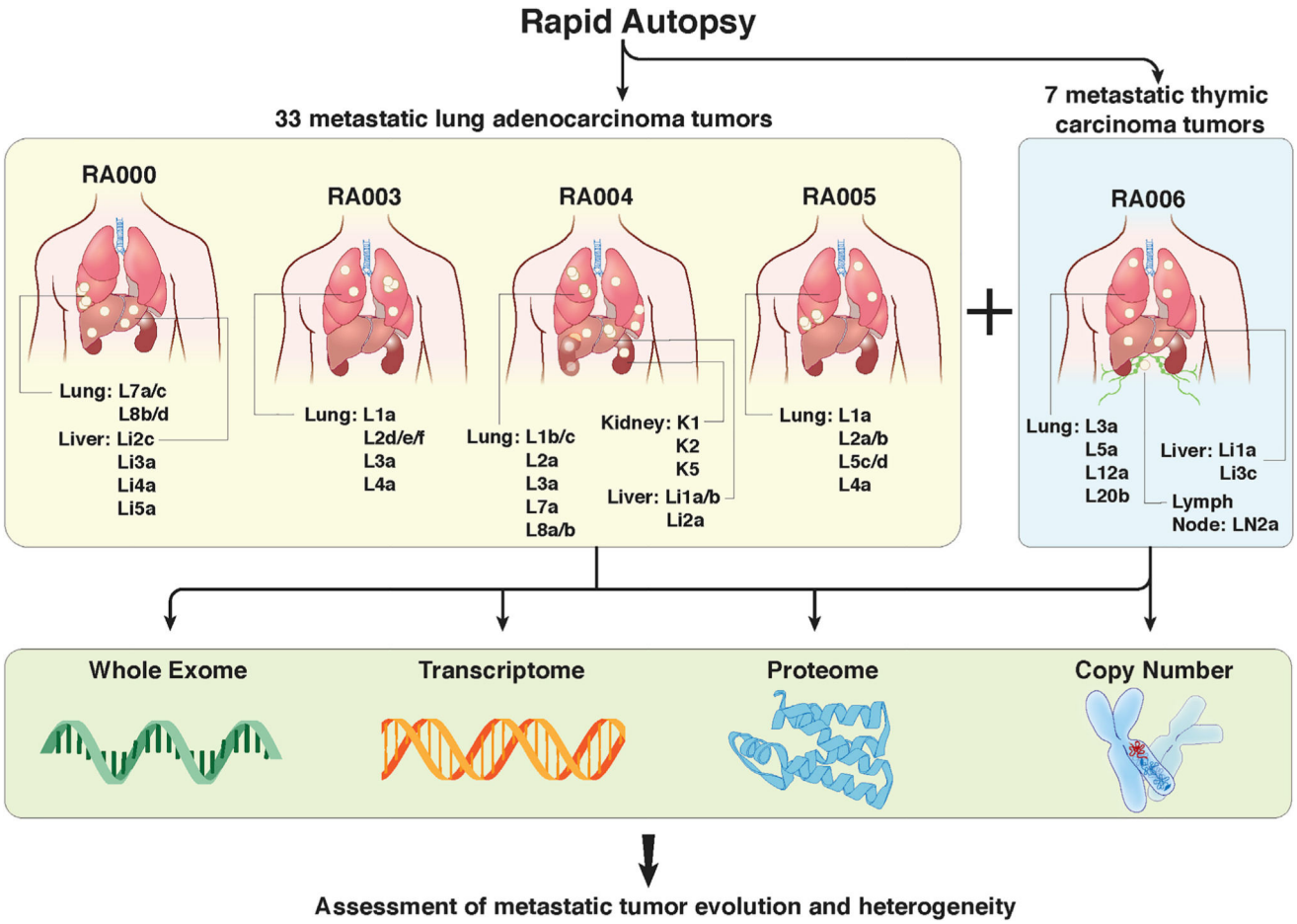



**Figure 1. Flowchart of Study**

Five patients underwent rapid autopsy (defined here as within 3 h of patient death). Thirty-three metastatic lung adenocarcinoma (LUAD) and 7 metastatic thymic carcinoma tumors from lung, liver, and kidney were subjected to whole-exome sequencing (WES), RNA sequencing, copy-number analysis, and mass-spectrometry-based proteomics, followed by assessment of intra- and inter-metastatic tumor heterogeneity. Three samples were removed from the study after sequencing because of low tumor content.

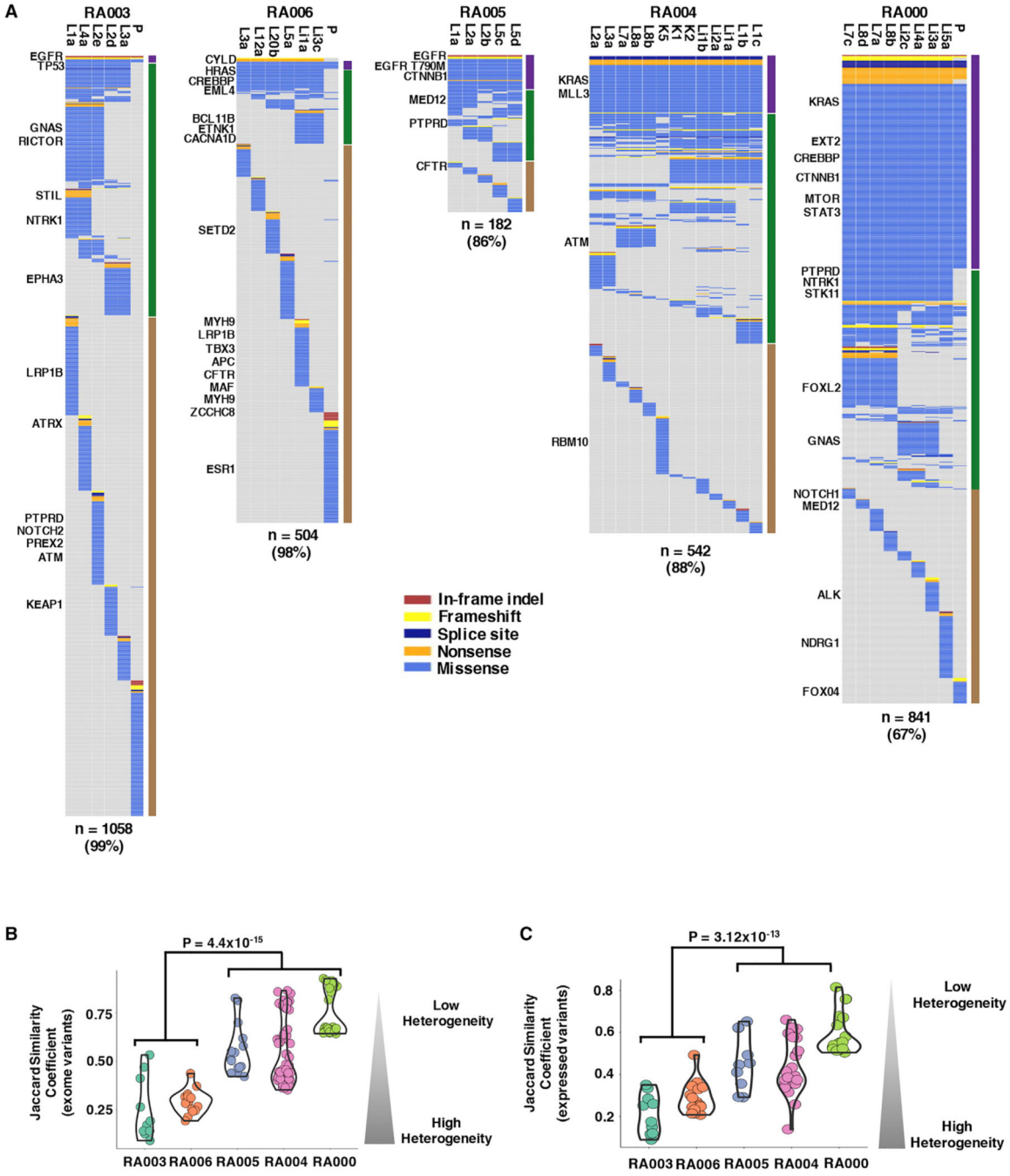



**Figure 2. Intra- and Inter-metastatic Heterogeneity of Somatic Mutations of Tumors from All Autopsy Patients**

(A) Heatmaps depict the distribution of non-silent somatic mutations among metastatic and, where available, primary tumors for each patient. Driver mutations are listed to the left of each heatmap. The total number of non-silent mutations and the percentage of non-truncal mutations are shown below each heatmap. The bars to the right of each heatmap summarizes intra- and inter-metastatic heterogeneity; mutations present in all regions (purple), in more than one, but not all (green), or only in one region (brown).

(B and C) Jaccard similarity coefficients of metastases within each patient based on mutations identified by exome sequencing (B) and expressed variants by RNA-seq (C). Each circle represents the Jaccard similarity coefficient between two metastases. Coefficients range from zero to one representing highest and lowest heterogeneity, respectively. The p value for the difference in mean Jaccard similarity coefficients between two groups of patients is shown. P, tumor sample obtained at diagnosis; L, lung tumor at autopsy; Li, liver tumor at autopsy; K, kidney tumor at autopsy.

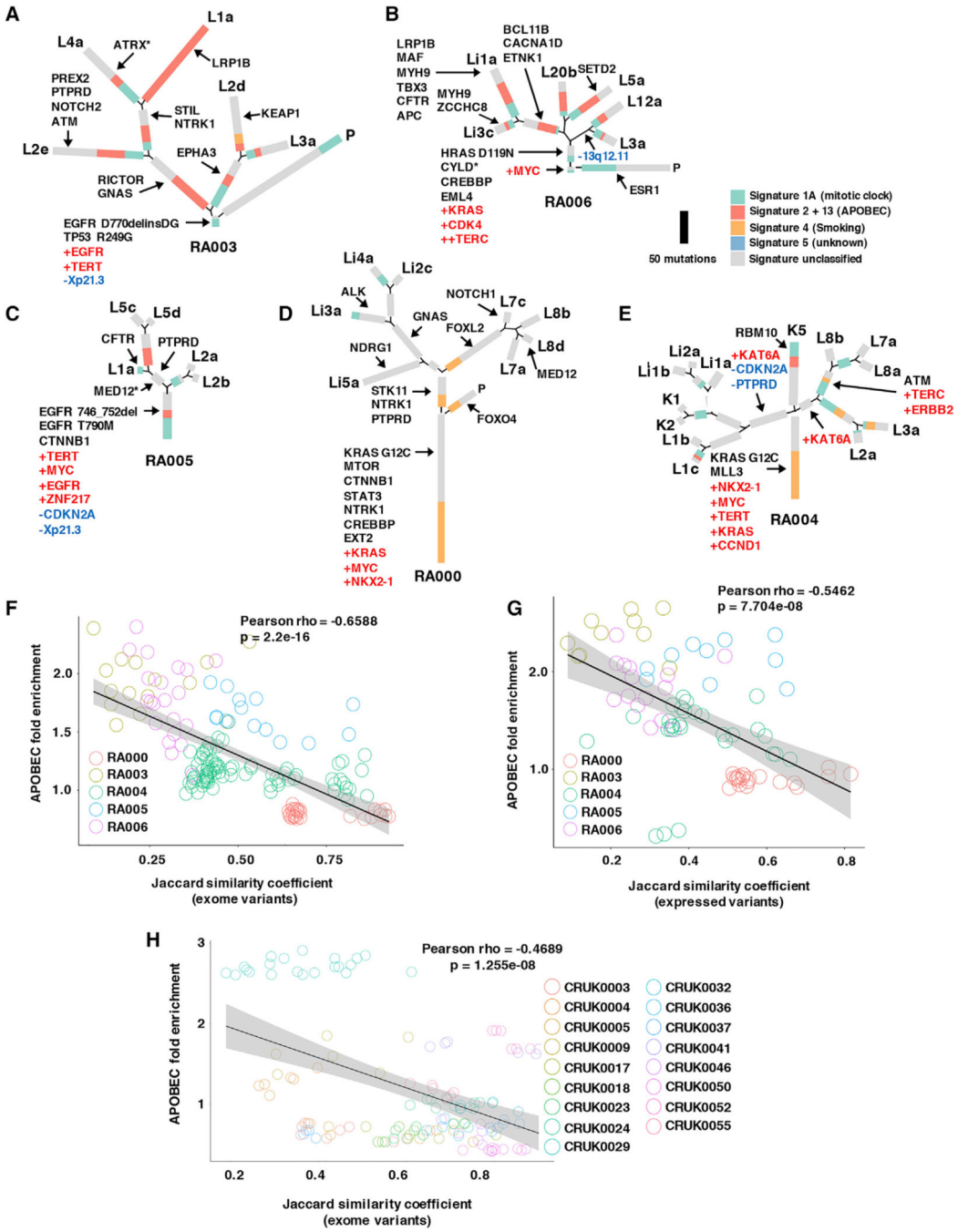



**Figure 3. Inferred Phylogeny, Mutational Signatures, and APOBEC-Associated Heterogeneity**

(A–E) Phylogenetic trees were generated by the Phangorn method from all validated mutations identified by whole-exome sequencing from tumors within patient RA003 (A), RA006 (B), RA005 (C), RA000 (D), and RA004 (E) using the maximum parsimony method. Trees are rooted in mutations common to all tumors within each patient. Trunk and branch lengths are proportional to the numbers of mutations acquired on the corresponding trunk or branch. Each private branch represents mutations unique to each individual tumor. Colors represent COSMIC mutational signatures. Select driver mutations and focal copy-

number amplifications/deletions are mapped to the trunks and branches as indicated. Asterisks denote nonsense mutations.

(F–H) APOBEC fold enrichment correlates with Jaccard similarity coefficients across NCI autopsy patient tumors based on exome variants (F), expressed variants (G), and TRACERx lung adenocarcinoma patient tumors (patients with five or more tumors shown) (H). Each circle represents mean APOBEC fold enrichment and the Jaccard similarity coefficient between two tumors from a given patient. P, tumor sample obtained at diagnosis; L, lung tumor at autopsy; Li, liver tumor at autopsy; K, kidney tumor at autopsy.

P-values for correlation coefficients are shown.

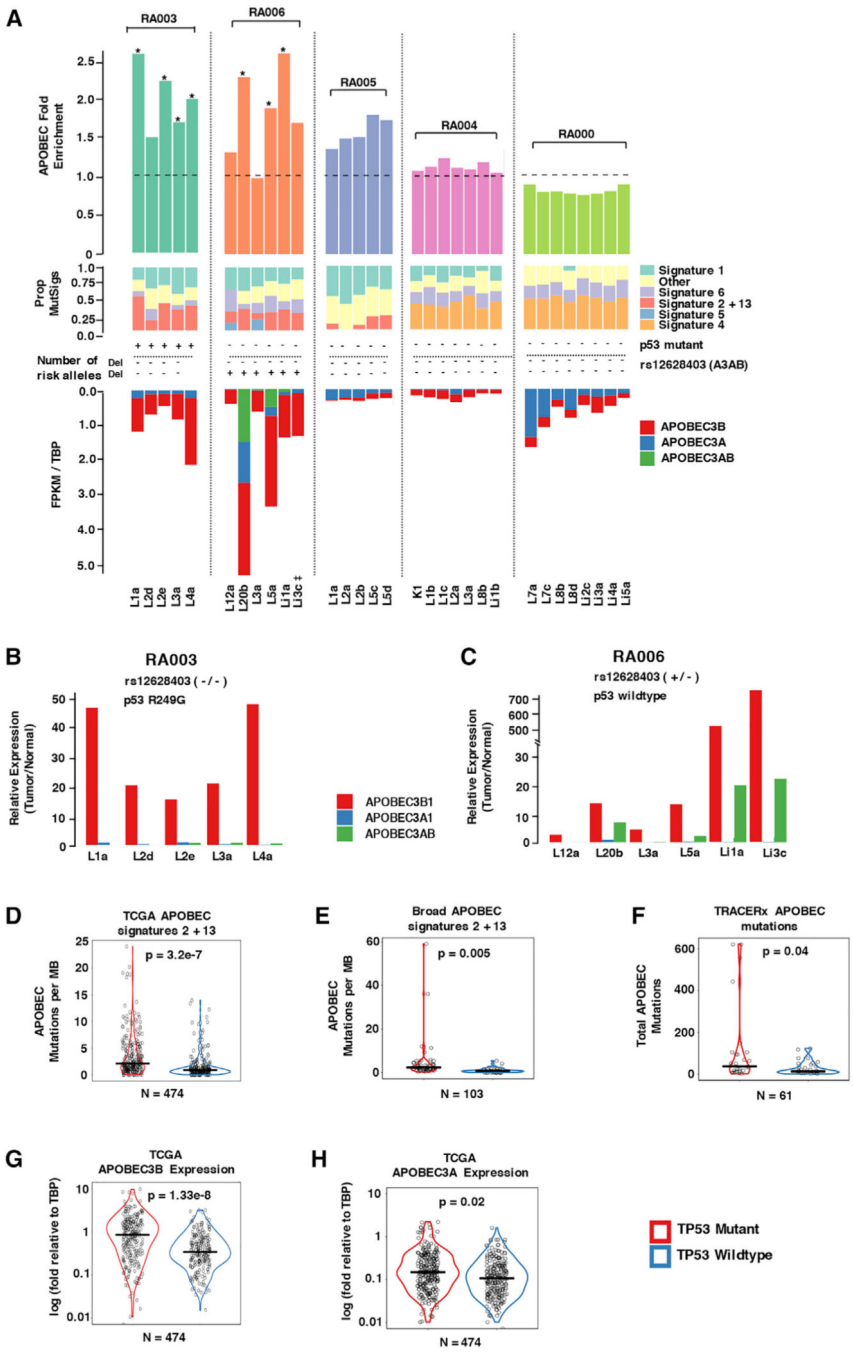



**Figure 4. Relationship between APOBEC Fold Enrichment, Mutational Signatures, *APOBEC3* Region Transcript Expression, *APOBEC3AB* Germline Variant, and Mutant *TP53***

(A) APOBEC fold enrichment of each tumor. Asterisks denote significant enrichment. The dashed line denotes zero enrichment of APOBEC mutations. The proportion of COSMIC signatures for each tumor appears below APOBEC fold enrichment. The number of risk alleles for *APOBEC3* germline variant rs12628403 is shown below the proportion of COSMIC signatures. The expression of *APOBEC3B*, *APOBEC3A*, and *APOBEC3AB* is shown below the proportion of COSMIC signatures.

(B and C) The relative expression of isoforms *APOBEC3B1*, *APOBEC3A1*, and *APOBEC3AB* in tumor relative to normal lung or liver by custom TaqMan assays in patient RA003 (B) and patient RA006 (C). Six of the 37 tumors are not included here because of insufficient RNA for sequencing. (D–F) APOBEC signature mutations per megabase in the TCGA (D) and in the Broad dataset (E) and the total APOBEC mutations in TRACERx dataset (F) in *TP53* mutant compared with *TP53* wild-type tumors.
(G and H) *APOBEC3B* expression (G) and *APOBEC3A* expression (H) in *TP53* wild-type and mutant tumors in the TCGA dataset. p values shown are adjusted for patient age and the number of pack years smoked.
‡visual inspection of RNA-seq data shows expression of *APOBEC3AB*.
See Figure S4E for details.

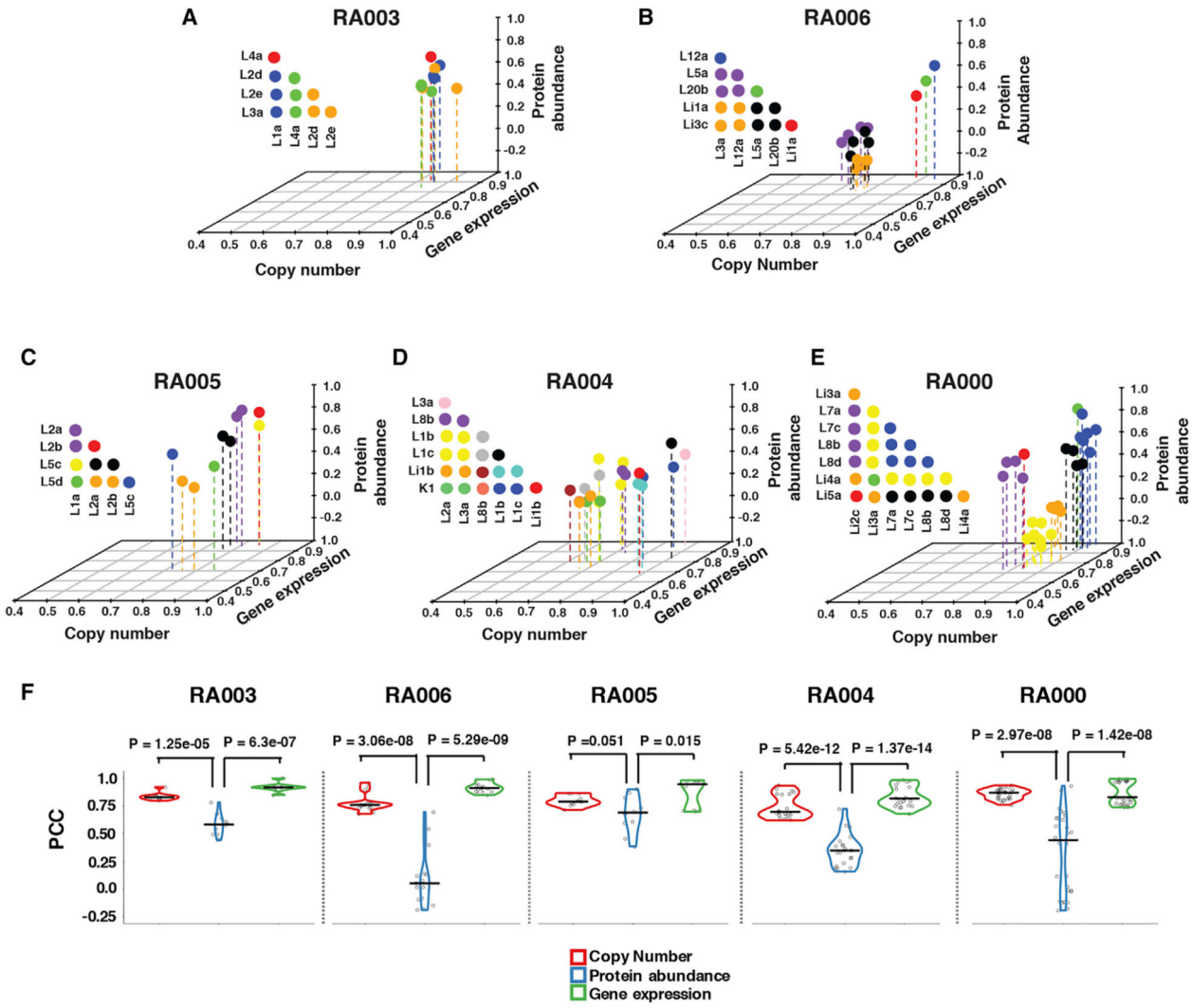



**Figure 5. Intra-Patient Multi-Omic Tumor Relationships**

(A–E) Pearson correlation coefficients (PCCs) are shown on each axis across all common genes identified in copy number, gene expression, and protein abundance datasets between all tumors within patient RA003 (A), RA006 (B), RA005 (C), RA004 (D), and RA000 (E). Each circle represents the relationship between two tumors. Previously designated tumor clusters are assigned the same color. Colors are independently assigned for each patient. (F) PCCs between tumors from each patient grouped by data type: copy number, gene expression, and protein abundance. p values for the difference in mean PCCs between data types are shown for each patient.

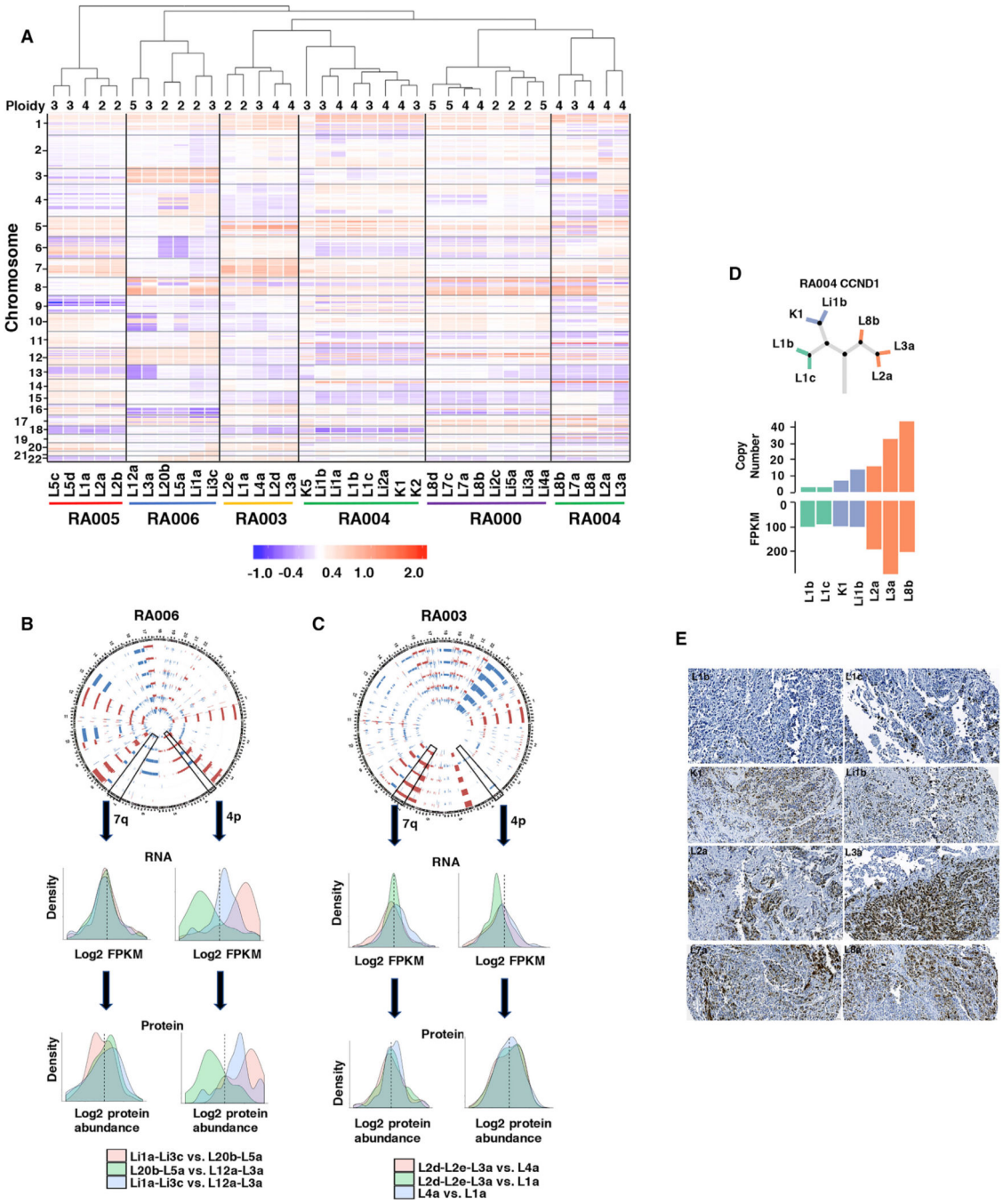



**Figure 6. Copy-Number Heterogeneity Corresponds with Transcriptomic and Proteomic Heterogeneity**

(A) Hierarchical clustering by copy number across the genome across all tumors from all patients (at cytoband resolution). Losses (purple) and gains (red) in log2 scale are depicted relative to mean ploidy. Mean ploidy is shown in the top row (rounded to nearest integer).. (B and C) Copy-number differences in chromosomal arms 4p and 7q between tumors of patient RA006 (B) and patient RA003 (C) and corresponding changes in FPKM and protein abundance. Probability density plots show the log ratios of mean FPKM by chromosomal arm for sets of tumors as displayed. The dashed line on the x axis represents the ratio of 1

(or log ratio 0). For visualization purposes, the x axis was cut at log −1 and 1 for RNA and log −0.1 and 0.1 for protein. The y axis was cut at a probability density of 2..
(D) The phylogenetic tree depicts tumors of patient RA004 with corresponding copy number and RNA-seq FPKM of CCND1 for each tumor..
(E) Protein expression of CCND1 for tumors of patient RA004 as assessed by immunohistochemistry from tissue microarrays.

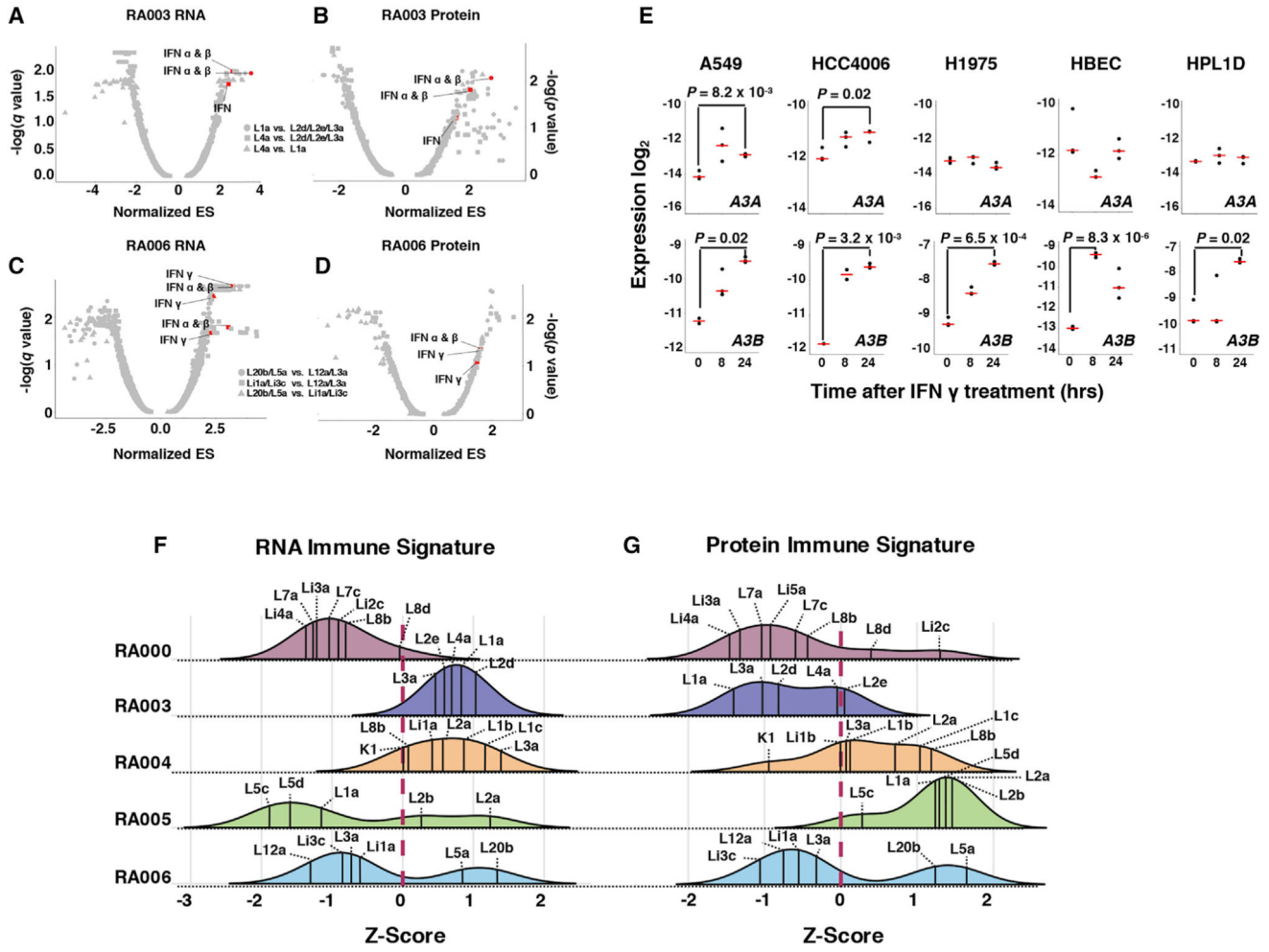



**Figure 7. Site-Specific Enrichment of Interferon Signaling Pathways, Immune Signature Heterogeneity, and the Effect of Interferon on Expression of *APOBEC3* Genes among Lung Cancer and Epithelial Cell Lines**

(A–D) Single-sample gene set enrichment (ssGSEA) analysis of transcriptome and proteome using the REACTOME databases are shown for tumors from patients RA003 (A and B) and RA006 (C and D). Significantly enriched interferon pathways ($q < 0.05$ for transcriptome, $p < 0.10$ for proteome) are colored red..

(E) *APOBEC3A* and *APOBEC3B* expression among lung cancer cell lines A549, HCC4006, and H1975 and immortalized normal lung epithelial cell lines HBEC and HPL1D treated with interferon-gamma. p values were calculated between time points as indicated for the significant changes using two-sided t tests. Data are shown as values for individual biological replicates (n = 3) and mean (red bars), normalized to endogenous controls and presented on the $\log_2$ scale.

(F and G) Immune signature scores within the transcriptome. (F) and proteome (G) are shown between all tumors of patients RA000, RA003, RA004, RA005, and RA006. Scores were normalized across all tumors separately for transcriptome and proteome. ES, enrichment score; IFN, interferon; A3A, APOBEC3A; A3B, APOBEC3B. Symbols denote comparisons between groups of tumors as indicated.

## KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Antibodies | | |
| CCND1 | Agilent | Cat#GA083 |
| Biological Samples | | |
| Metastases collected by rapid autopsy | National Cancer Institute, Bethesda, Maryland | https://clinicaltrials.gov/ct2/show/NCT01851395 |
| Chemicals, Peptides, and Recombinant Proteins | | |
| IFNγ | R&D Systems | Cat#285-IF-100 |
| Critical Commercial Assays | | |
| HiSeq 2500 system | Illumina | N/A |
| Agilent SureSelect XT (All Exon V5 + UTR) | Illumina | N/A |
| TruSeq Stranded Total RNA Library Prep | Illumina | Cat#20020596 |
| TMT10plex | Thermo Scientific | Cat#90110 |
| XBridge C18 | Waters | N/A |
| LTQ-Orbitrap Elite interfaced with an UltimateTM 3000 RSLCnano System | Thermo Scientific | N/A |
| Deposited Data | | |
| Processed data and code used | This paper | https://github.com/nitinroper/Rapid-Autopsy-NCI |
| Raw data | This paper | dbGaP: phs001432.v1.p1 and PRIDE: PXD012845 |
| Experimental Models: Cell Lines | | |
| A549 | ATCC | Cat#CCL-185 |
| HCC4006 | ATCC | Cat#CRL-2871 |
| H1975 | ATCC | Cat#CRL-5908 |
| HBEC3-KT | ATCC | Cat#CRL-4051 |
| HPL1D | Masuda et al., 1997; PMID: 9354455 | N/A |
| Oligonucleotides | | |
| APOBEC3B forward: TGCTGGGAAAAC TTTGTGTACAAT | Middlebrooks et al., 2016; PMID: 27643540 | N/A |
| APOBEC3B reverse: ATGTGTCTGGAT CCATCAGGTATCT | Middlebrooks et al., 2016; PMID: 27643540 | N/A |
| APOBEC3B probe: FAM ATTCATGCC TTGGTACAAA | Middlebrooks et al., 2016; PMID: 27643540 | N/A |
| APOBEC3AB forward: ATCATGACCTA CGATGAATTTAAGCA | Middlebrooks et al., 2016; PMID: 27643540 | N/A |
| APOBEC3AB reverse: AGCACATTGCT TTGCTGGTG | Middlebrooks et al., 2016; PMID: 27643540 | N/A |
| APOBEC3AB probe: FAM-CATTCTCC AGAATCAGGG | Middlebrooks et al., 2016; PMID: 27643540 | N/A |

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| APOBEC3A (Hs00377444_m1) | Thermo Fisher Scientific | N/A |
| GAPDH | Thermo Fisher Scientific | Cat#4326317E |
| PPIA | Thermo Fisher Scientific | Cat#4326316E |
| Software and Algorithms | | |
| PyClone v0.13.0 | Roth et al., 2014; PMID: 24633410 | https://shahlab.ca/projects/pyclone/ |
| deconstructSigs v1.8.0 | Rosenthal et al., 2016; PMID: 26899170 | https://www.rdocumentation.org/packages/deconstructSigs/versions/1.8.0 |
| Treeomics | Reiter et al., 2017: PMID: 28139641 | https://github.com/johannesreiter/treeomics |
| Strelka 1.0.10 | Saunders et al., 2012; PMID: 22581179 | https://sites.google.com/site/strelkasomaticvariantcaller/home |
| STAR | Dobin et al., 2013; PMID: 23104886 | https://github.com/alexdobin/STAR |
| CIBERSORT | Newman et al., 2015; PMID: 25822800 | https://github.com/zomithex/CIBERSORT |
| xCell | Aran et al., 2017; PMID: 29141660 | https://github.com/dviraran/xCell |
| GenePattern | Subramanian et al., 2005; PMID: 16199517 | https://software.broadinstitute.org/cancer/software/genepattern |
| FireBrowse | N/A | http://firebrowse.org/ |
| GISTIC 2.0 | Mermel et al., 2011; PMID: 21527027 | https://software.broadinstitute.org/cancer/cga/gistic |
| Phangorn | Schliep, 2011; PMID: 21169378 | https://www.thermofisher.com/us/en/home/industrial/mass-spectrometry/liquidchromatography-mass-spectrometry-lc-ms/lc-ms-systems/orbitrap-lc-ms.html; https://github.com/KlausVigo/phangorn |
| Maxquant v1.5.3.30 | Tyanova et al., 2016; PMID: 27809316 | https://maxquant.org/ |