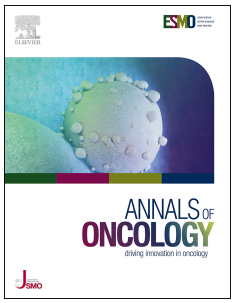


APOBEC mutagenesis, kataegis, chromothripsis in *EGFR*-mutant osimertinib-resistant lung adenocarcinomas

P. Selenica, A. Marra, N.J. Choudhury, A. Gazzo, C.J. Falcon, J. Patel, X. Pei, Y. Zhu, C.K.Y. Ng, M. Curry, G. Heller, Y.-K. Zhang, M.F. Berger, M. Ladanyi, C.M. Rudin, S. Chandarlapaty, C.M. Lovly, J.S. Reis-Filho, H.A. Yu

PII: S0923-7534(22)04142-4

DOI: https://doi.org/10.1016/j.annonc.2022.09.151

Reference: ANNONC 1098

To appear in: *Annals of Oncology*

Received Date: 2 November 2021

Revised Date: 2 August 2022

Accepted Date: 1 September 2022

Please cite this article as: Selenica P, Marra A, Choudhury NJ, Gazzo A, Falcon CJ, Patel J, Pei X, Zhu Y, Ng CKY, Curry M, Heller G, Zhang YK, Berger MF, Ladanyi M, Rudin CM, Chandarlapaty S, Lovly CM, Reis-Filho JS, Yu HA, APOBEC mutagenesis, kataegis, chromothripsis in *EGFR*-mutant osimertinib-resistant lung adenocarcinomas, *Annals of Oncology* (2022), doi: https://doi.org/10.1016/j.annonc.2022.09.151.

Haller 1973R2

JOURNAL: Annals of Oncology

ARTICLE TYPE: Original Article

TITLE:
APOBEC mutagenesis, kataegis, chromothripsis in *EGFR*-mutant osimertinib-resistant lung adenocarcinomas

AUTHORS:
P. Selenica[1#], A. Marra[1#], N.J. Choudhury[2#], A. Gazzo[1], C.J. Falcon[3], J. Patel[1], X. Pei[1], Y. Zhu[1], C.K.Y. Ng[4], M. Curry[5], G. Heller[5], Y-K. Zhang[6], M.F. Berger[1,7,8], M. Ladanyi[8], C.M. Rudin[2,9], S. Chandarlapaty[1,9], C.M. Lovly[6], J.S. Reis-Filho[1*], H.A. Yu[2,9*]

[#]Authors contributed equally
*Co-corresponding authors
[-], Memorial Sloan Kettering Cancer Center, New York City, USA
[2]Department of Medicine, Thoracic Oncology Service, Memorial Sloan Kettering Cancer Center, New York City, USA
[3]Druckenmiller Center for Cancer Research, Memorial Sloan Kettering Cancer Center, New York City, USA
[4]Department for BioMedical Research (DBMR), University of Bern, Bern, CHE
[5]Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York City, USA
[6]Department of Medicine, Division of Hematology and Oncology and Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, USA
[7]Marie-Josée and Henry R. Kravis Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York City, USA
[8]Department of Pathology, Molecular Diagnostics Service, Memorial Sloan Kettering Cancer Center, New York City, USA
[9]Department of Medicine, Weill Cornell Medical College, New York City, USA

CORRESPONDING AUTHORS:
Dr. Helena A. Yu, Department of Medicine, Thoracic Oncology Service, Memorial Sloan Kettering Cancer Center, 545 E 73rd St, New York, NY 10021, USA. Office number 1(646) 608-3912, yuh@mskcc.org

Dr. Jorge S. Reis-Filho, Department of Pathology, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10021, USA. Office number 1(212) 639-8054, reisfilj@mskcc.org

MANUSCRIPT METRICS:

Tables: 0
Figures: 4
Word Count: 4505

References: 43
Supplemental Tables: 5; Figures: 5

ABSTRACT

Background: Studies of targeted therapy resistance in lung cancer have primarily focused on single gene alterations. Based on prior work implicating APOBEC mutagenesis in histological transformation of *EGFR*-mutant lung cancers, we hypothesized that mutational signature analysis may help elucidate acquired resistance to targeted therapies.

Patients and methods: APOBEC mutational signatures derived from an FDA-cleared multigene panel (MSK-IMPACT) using the *SigMA* algorithm were validated against the gold standard of mutational signatures derived from whole exome sequencing. Mutational signatures were decomposed in 3,276 unique lung adenocarcinomas, including 93 paired osimertinib-naïve and resistant *EGFR*-mutant tumors. Associations between APOBEC and mechanisms of resistance to osimertinib were investigated. Whole-genome sequencing (WGS) was performed on available *EGFR*-mutant lung cancer samples (10 paired, 17 unpaired) to investigate large scale genomic alterations potentially contributing to osimertinib resistance.

Results: APOBEC mutational signatures were more frequent in receptor tyrosine kinase (RTK)-driven lung cancers (*EGFR, ALK, RET* and *ROS1;* 25%) compared to lung adenocarcinomas at large (20%, p<0.001); across all subtypes, APOBEC mutational signatures were enriched in subclonal mutations (p<0.001). In *EGFR*-mutant lung cancers, osimertinib-resistant samples more frequently displayed an APOBEC dominant mutational signature compared to osimertinib-naïve samples (28 vs.14% p=0.03). Specifically, mutations detected in osimertinib-resistant tumors but not in pre-treatment samples significantly more frequently displayed an APOBEC dominant mutational signature (44% vs 23%, p<0.001). *EGFR*-mutant samples with APOBEC dominant signatures had enrichment of large scale-genomic rearrangements (p=0.01) and kataegis (p=0.03) in areas of APOBEC mutagenesis.

Conclusions:

APOBEC mutational signatures are frequent in RTK-driven LUADs and increase under the selective pressure of osimertinib in *EGFR*-mutant lung cancer. APOBEC mutational signature enrichment in subclonal mutations, private mutations acquired after osimertinib treatment, and areas of large scale genomic rearrangements highlights a potentially fundamental role for APOBEC mutagenesis in the development of resistance to targeted therapies, which may be potentially exploited to overcome such resistance.

KEYWORDS: EGFR, mutational signatures, APOBEC, tyrosine kinase inhibitor, osimertinib, acquired resistance, structural rearrangements

HIGHLIGHTS:

-Receptor tyrosine kinase-driven lung cancers display higher levels of APOBEC mutagenesis as compared to other lung cancers
-RTK-driven lung adenocarcinoma with evidence of APOBEC mutagenesis are enriched in sub-clonal mutations
-APOBEC is a dominant mutational process in post-osimertinib samples of *EGFR* mutant lung cancers
-APOBEC mutagenesis might have a relevant role in the development of resistance to targeted therapies

INTRODUCTION

Most studies of resistance to targeted therapies have primarily focused on single gene mutations or copy number alterations (CNAs) identified by multigene panel sequencing. In oncogene-driven lung cancers, a single gene alteration that mediates resistance is rarely identified[1], and only a subset of alterations detected are therapeutically actionable, with benefit limited by intratumoral heterogeneity and the subclonal nature of many of these acquired mutations[2]. Spatial and temporal mapping of therapeutic resistance at individual tumor sites suggests that subclonal alterations may be driven by large scale mutational processes occurring during tumor evolution.[3-5] The contribution of genome-wide mutational processes in mediating tumor evolution and therapeutic resistance under the selective pressure of targeted therapies is underexplored. These large-scale patterns of mutagenesis drive genomic diversity and provide a substrate for clonal selection.[4] Merging layers of genomic interrogation, from base pair substitutions to large scale chromosomal rearrangements, provides an integrated assessment of the adaptive processes that propagate resistance to targeted therapies in oncogene-driven lung cancers. Furthermore, as microsatellite instability and homologous recombination deficiency confer sensitivity to immune checkpoint inhibitors [6, 7] and PARP inhibitors respectively, other mutational signatures may provide the basis for the development of new therapeutic strategies in drug-resistance tumors.

Such additional interrogation may be particularly germane to lung cancers driven by *epidermal growth factor receptor* (*EGFR*) alterations. Patients with *EGFR*-mutant lung cancer are effectively treated with EGFR tyrosine kinase inhibitors (TKIs)[11], yet long-term disease control remains elusive. Acquired single-gene mutations and CNAs have been identified in only ~20% of first-line osimertinib-resistance cases.[1] Lineage plasticity[1] and acquired structural rearrangements such as kinase gene fusions[12] are observed in about 25% of osimertinib-resistant tumors; these events are rarely seen in other oncogene-driven lung cancers. The mechanistic basis of the enrichment for these events in post-therapy *EGFR*-mutant lung

cancers has yet to be defined. Prior work elucidating mechanisms of acquired resistance has focused on single gene alterations; Catalogue of Somatic Mutations in Cancer (_) mutational signatures define larger genome-wide trends that may contribute to a better understanding of tumor initiation and evolution. We and others identified enrichment of large scale genomic alterations such as apolipoprotein b mRNA-editing enzyme, catalytic polypeptide-like (APOBEC) mutational signatures in *EGFR*-mutant lung cancers destined for small cell histologic transformation.[13, 14] APOBEC signatures are one of the most prevalent mutational signatures in human cancers, derive from the activity of APOBEC cytosine deaminases, and have been implicated in tumorigenesis and potentially in therapeutic resistance.[15, 16] APOBEC mutational signatures are also the dominant mutational signatures in lung adenocarcinomas (LUADs) with late-evolving subclonal driver alterations.[17]

Given these data, we hypothesized that APOBEC may have a causative role in mediating mechanisms of acquired resistance such as structural rearrangements and lineage plasticity seen in *EGFR*-mutant lung cancer. Validation of such a role would provide rationale for assessing APOBEC enzymatic inhibition in *EGFR*-mutant lung cancers. In the current study, we utilized an algorithm benchmarked for mutational signature detection in multigene sequencing panels[18] to examine the progression of APOBEC mutational signature levels in LUADs derived from clinically obtained multigene panel sequencing and whole-genome sequencing (WGS) to survey tumor evolution under the selective pressure of treatment.

METHODS

Study cohort

We identified all patients with LUAD with targeted hybrid capture, next-generation tumor sequencing (NGS) performed at Memorial Sloan Kettering Cancer Center (MSK) from March 2014 to March 2021 using the FDA-approved MSK Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT) assay NGS platform.[19, 20] Clinical data collection, including

patient demographics and treatments, was approved by the MSK Institutional Review Board/Privacy Board, and all patients provided written consent. Tumor samples were de-identified for further analyses. Of all patients with LUAD sequenced by MSK-IMPACT, 93 patients with *EGFR*-mutant lung cancer treated with osimertinib had available paired pre- and post-treatment tumor samples. For this subgroup, we collected demographic information, clinical characteristics and detailed treatment histories, including line of treatment, time to treatment discontinuation (TTD) on osimertinib, and overall survival (OS) from time of MSK-IMPACT, with left truncation adjustment.

Multi-gene panel sequencing, whole genome sequencing (WGS), and mutational signature analysis

Somatic alterations were detected as described previously[19, 20] utilizing the FDA-cleared New York State Department of Health-approved MSK-IMPACT, with median sequencing coverage depth of 675x (range, 60x-1284x***)***. CNAs and estimated tumor purity were identified using FACETS.[21] The cancer cell fraction (CCF) of each mutation was inferred using ABSOLUTE (v1.0.6) with the ABSOLUTE solutions manually reviewed[22] . Mutations were classified as clonal if the probability of the mutation being clonal was greater than 0.5 or if the lower confidence interval was greater than 0.9 as calculated by ABSOLUTE. To validate findings on MSK-IMPACT and provide further information regarding large-scale genomic rearrangements, WGS was performed on pre- and/or post-osimertinib and matched normal tissue samples for 27 patients by MSK's Integrated Genomics Operations using validated protocols[23, 24], with median sequencing coverage depth of 57x (range 48x-62x). WGS was completed on six pre-treatment, 11 post-treatment, and 10 paired samples derived from formalin-fixed, paraffin-embedded tissue samples or recaptured genomic DNA. Twenty of these patients also had either or both pre-/post-treatment samples analyzed by MSK-IMPACT. To evaluate the number and context type of somatic single-nucleotide variants (SNVs) acquired after osimertinib treatment between paired pre/post samples from the same patient, we

calculated the fraction of substitutions that were private to the pre-osimertinib sample or private to the post-osimertinib sample. For each pair, all somatic SNVs that were detected in one sample were annotated to be either present (shared SNVs) or absent (private SNVs) in the other. To determine if the potential contribution of artifacts stemming from formalin fixation and paraffin embedding (i.e. primarily C>T SNVs), the frequency of SNV nucleotide changes was compared between pre- and post-treatment samples (**Supplementary Table 1**), as well as across all mutations, mutations with less than 10% variant allele fraction and mutations with less than 5% variant allele fraction (**Supplementary Table 2**). Further description of methods is reported in the **Supplementary Methods**. In addition, given that recent studies have demonstrated that formalin fixation can leave predictable and identifiable mutational signatures across the genome (i.e., COSMIC SBS 30[25]), we inspected the mutational profile of the formalin-fixed paraffin-embedded samples that were subjected to sequencing. No detectable levels of these specific signatures were identified (data not shown). Further description of methods is reported in the **Supplementary Methods**.

For MSK-IMPACT samples with at least five somatic SNVs, mutational signatures were computed using *Signature Multivariate Analysis (SigMA)*[18], a tool extensively benchmarked for the analysis of formalin-fixed paraffin-embedded samples subjected to multi-gene panel sequencing, as previously described.[26, 27] A dominant signature for each sample was determined based on the proposed category assigned by *SigMA*, as previously reported.[18] The decomposed exposures (i.e. the fraction of mutations associated with a given mutational process) were converted into percentages to allow a meaningful comparison among samples with different numbers of mutations. Based on the decomposed exposure values, additional categories (i.e. APOBEC < or ≥20%) were computed and utilized in the downstream analyses performed. The decision to adopt the cut-off of 20% was based on the lowest possible mutational signature exposure that can be detected with sufficient confidence in samples with ≥5 SNVs. In addition, *SigMA* was used to classify conventional codon-context mutation types for

each variant. Classifications are based on the six substitution subtypes: C>A, C>G, C>T, T>A, T>C, and T>G, and the nucleotides immediately 5' and 3' to the mutation. Twelve APOBEC-context mutations were identified using the peaks present on the spectra of COSMIC Signature 2 and Signature 13 (4 and 8 codon context mutations, respectively).

Evaluation of SigMA performance *on MSK-IMPACT data*

SigMA performance was evaluated using a simulation strategy to reproduce MSK-IMPACT samples from available whole-exome sequencing (WES) data from The Cancer Genome Atlas (TCGA; version mc3.v0.2.8) LUAD cohort. We considered the APOBEC exposures obtained from TCGA samples as ground truth. MSK-IMPACT sample simulation was performed by filtering TCGA data for only those mutations occurring on regions covered by MSK-IMPACT. COSMIC mutational signatures were computed using *SigMA* and applying "lung cancer" as reference tumor for both the simulated MSK-IMPACT samples and original TCGA WES samples. Performance was assessed using Pearson's coefficient, sensitivity, specificity, and accuracy. Cohen's kappa coefficient was used to assess the agreement between *SigMA*-proposed mutational signature categories, such as APOBEC <20% vs ≥20% and APOBEC dominant vs non-dominant.

Statistical analysis

Statistical analyses were conducted using R (version 3.1.2). Comparisons of categorical and continuous variables were performed by Fisher's exact and Mann-Whitney *U* tests, respectively. Differences in mutational signature exposures across multiple groups were analyzed using the Mann-Whitney *U* test. Pearson's correlation was used to compare mutational signature exposures. Oncoplots for mutational landscape visualization and heatmaps were created using the R packages, *maftools*[28], and *ComplexHeatmap*[29], respectively. Comparisons of frequencies of genes altered by somatic genetic alterations (including non-synonymous

mutations, gene amplifications and/or homozygous deletions) between either pre- and post-osimertinib samples or different APOBEC exposure cut-offs (<20% vs ≥20%) were performed using Fisher's exact test. Comparative analysis in paired samples was performed by the Wilcoxon signed-rank test. Mutation frequencies related to APOBEC mutagenesis in the kataegis and non-kataegis areas were compared using Fisher's exact test. Multiple testing correction using the Benjamini-Hochberg method was applied to control for the false discovery rate whenever appropriate. $P < 0.05$ was considered statistically significant. All tests were two-tailed.

RESULTS

Mutational signatures can be accurately derived from targeted sequencing data

Since only a small fraction of patient samples undergoes WES or WGS, we first sought to validate the performance of the *SigMA* algorithm in calling APOBEC mutational exposures from the MSK-IMPACT targeted gene panel, which is ordered routinely for clinical purposes as an integral element of patient care at MSKCC. WES data from the TCGA LUAD cohort (n=511) served as a benchmark. MSK-IMPACT simulation was performed by the inclusion of SNVs encompassed by the genomic footprint of MSK-IMPACT; only samples with ≥5 SNVs were included, as required by the *SigMA* algorithm (n=390) (**Fig. 1A**). We evaluated two methods for identifying APOBEC-enriched samples in the MSK-simulated cohort: those with POBEC identified as dominant mutational signature by *SigMA* mutational analysis (i.e. the mutational process with the highest level of exposure) and those with high levels of APOBEC exposure, arbitrarily defined as ≥20% exposure. The concordance between each of these designations was tested by signature analysis of the TCGA WES cohort (ground-truth). *SigMA* detected APOBEC as dominant signature in MSK-IMPACT-simulated samples with 74% sensitivity and 97% specificity (**Fig. 1A**). Sensitivity increased to 86% and 85%, respectively, when ≥10 and ≥15 mutations were included with 97% specificity in both groups. Using the Cohen's kappa

correlation test, we observed a substantial agreement (K scores 0.72-0.81) in detecting APOBEC as dominant mutational signature between WES and MSK-IMPACT-simulated panel (**Supplementary Table 3**). Detection of APOBEC exposure ≥20% yielded comparable metrics (**Supplementary Fig. 1A**). When considered as a continuous variable, APOBEC exposure was highly correlated between TCGA WES and MSK-IMPACT-simulated samples (Pearson's r = 0.77, p<0.0001; **Supplementary Fig. 1B**). As additional validation, mutational signatures were computed by *SigMA* in 18 samples for which both WGS and MSK-IMPACT data were available. A significant positive correlation between APOBEC mutational signature exposures detected by MSK-IMPACT and WGS was observed (Pearson's r = 0.91, p<0.0001; **Supplementary Fig. 1C**). These results show the consistency and reliability of *SigMA* in detecting APOBEC mutational signatures in LUAD samples sequenced by the MSK-IMPACT gene panel. Given the higher concordance observed between APOBEC dominant signature calling between TCGA WES and MSK-IMPACT-simulated cohort (K=0.72), this designation was selected for subsequent analysis.

APOBEC mutational signature is enriched in RTK-driven LUADs

We next sought to analyze the frequency of APOBEC mutational signatures in a large cohort of LUADs. We identified 6,034 LUAD samples which had undergone targeted gene sequencing with MSK-IMPACT; 3,276 samples contained the requisite ≥5 SNVs required for *SigMA* signature analysis. Tumors were first sorted according to known oncogenic drivers. We created a RTK driver subgroup encompassing four well-characterized receptor tyrosine kinase (RTK) driver alterations in *EGFR*, *ALK*, *RET*, and *ROS1*, all of which have approved targeted therapies and are enriched in patients with little or no smoking history (**Fig. 1B**). *KRAS*-mutant lung cancers were selected as a molecularly-defined comparator that occur more commonly in patients with smoking histories. The remaining LUADs were classified into the “Other” category.

Mutational signature analysis revealed that APOBEC mutational signatures were more frequently identified as the dominant signature in the RTK Driver group (23%) compared to both the *KRAS* (8%, p<0.001) and Other subgroups (17%, p<0.001) (**Fig. 1C**). Conversely, smoking signature was more frequently the dominant signature in the *KRAS* subgroup (57%) compared to both the RTK Driver (14%, p<0.001) and Other (43%, p<0.001) subgroups. The RTK Driver subgroup also displayed the highest mean APOBEC mutational signature exposure (25% vs. 20% in Other, p < 0.001, and 12% in *KRAS,* p < 0.001, **Fig. 1D**). To take into account potential biases related to the threshold of ≥5 SNVs for *SigMA* signature analysis, a sensitivity analysis using higher cut-offs was performed. A numeric increment in the proportion of cases with APOBEC as dominant mutational signature was found in the RTK Driver group when the ≥10 and ≥15 SNVs thresholds were adopted (43% and 54%, respectively; **Supplementary Fig. 2A-C**). This trend, however, was not observed in the KRAS (11% and 15% using ≥10 and ≥15 SNVs, respectively) and Other (20% and 20% using ≥10 and ≥15 SNVs, respectively) groups where the proportions of APOBEC dominant cases remained constant even if different SNV cut-offs were applied. Conversely, we observed an increment in the proportion of Smoking signature in the KRAS (57%, 71% and 75% using ≥5, ≥10 and ≥15 SNVs, respectively) and Other (43%, 57% and 64% using ≥5, ≥10 and ≥15 SNVs, respectively) groups when higher SNVs thresholds were applied (**Supplementary Fig. 2A-C**). These findings remained consistent when the proportions of the mutational exposures were tested rather than dominant signatures (**Supplementary Fig. 2D-F**). The smoking signature was inversely correlated with APOBEC signature (Pearson's r = -0.36, p <0.001; **Supplementary Fig. 3A**), which was observed consistently across all subgroups. In a univariable model, RTK Driver tumors were significantly more likely to have APOBEC as a dominant signature compared to *KRAS*-mutant tumors (odds ratio [OR] 0.30 for *KRAS* compared to RTK, 95% confidence interval [CI] 0.23-0.39, p<0.001) and Other (OR 0.72 for Other Compared to RTK, 95% CI 0.57-0.91, p<0.001, **Supplementary Table 4**), although results in a multivariate analysis accounting for age, gender, smoking status,

and sample type (primary vs. metastatic) did not reach statistical significance (**Supplementary Table 5**). These findings support the notion that APOBEC mutational signatures are significantly enriched in RTK-driven LUADs.

We further evaluated the proportion of APOBEC-site mutations by clonal status in cases with APOBEC as dominant mutational signature compared to non-dominant. APOBEC-site mutations were detected more frequently at the subclonal rather than clonal level in LUADs displaying dominant APOBEC signature in all considered subgroups ($p < 0.001$; **Supplementary Fig. 3B**), consistent with previous findings that have associated APOBEC mutagenesis with the acquisition of subclonal mutations.[3, 30]

We also assessed spatial and temporal evolution of mutational signatures in LUAD samples from primary (from lung tissue) or metastatic (from tissue other than lung) sites; site of tumor was available for 3,248/3,276 samples (Supp**lemental Fig. 3C**). Metastatic tumors (n=1,364) exhibited a significantly higher mean APOBEC exposure compared to lung primaries (n=1,884, 21% vs. 15%, $p<0.001$, **Supplemental Fig. 3D**, a pattern conserved across all three molecular subgroups (**Supplemental Fig. 3E**). These data support a modest yet statistically significant enrichment in APOBEC signatures in metastatic tumors, suggesting that APOBEC may constitute a relatively late event in the development and progression of LUADs.[31]

Evolution of mutational signatures with osimertinib treatment in EGFR-mutant lung cancers

To evaluate the evolution of APOBEC mutational signature with the selective pressure of targeted therapy, we identified 93 patients with paired osimertinib-naïve and osimertinib-resistant tumor samples that underwent MSK-IMPACT sequencing. Fifty-two received osimertinib as first-line treatment and 41 as later-line treatment. Sixty osimertinib-naïve and 81 osimertinib-resistant tumors had the requisite ≥5 SNVs*;* 54 patients had both samples meet this criterion. APOBEC was the dominant signature in twice as many osimertinib-resistant samples

compared with osimertinib-naïve samples (28% vs 14%, p=0.03, **Fig. 2A, Supplementary Fig. 4A**). Numerically higher APOBEC exposure was also detected in osimertinib-resistant tumors compared to matched osimertinib-naive samples (Wilcoxon signed-rank test, p=0.33, **Supplementary Fig. 4B**). Somatic tumor mutational burden (TMB) was also higher in post-osimertinib compared to pre-osimertinib samples, suggesting a possible correlation between APOBEC mutagenesis and TMB (p<0.001, **Supplementary Fig. 4C**).

To ascertain the contribution of APOBEC mutagenesis to the mutations acquired after treatment with osimertinib, we analyzed 10 cases with paired osimertinib naïve and resistant samples that underwent WGS. Interestingly, we found that the proportion of APOBEC site mutations private to post-treatment samples was significantly higher than those shared with pre-treatment samples (44% vs 23%, p<0.001, **Supplementary Figure 4D**). These data suggest that a high proportion of *EGFR*-mutant lung tumors have evidence of *de novo* APOBEC activity prior to treatment, with further enrichment in APOBEC-driven mutagenesis and acquired APOBEC context mutations following osimertinib.

Concurrent alterations on MSK-IMPACT and mutational signatures

To evaluate whether recurrent mutations in common somatic hotspots might be associated with APOBEC mutagenesis, we focused on cases with both dominant APOBEC mutational signature as well as high exposure. There was no specific pattern of concurrent alterations identified in osimertinib-resistant samples in tumors with APOBEC as dominant signature (**Fig. 2B**). Similarly, no statistically significant enrichment in gene alterations was found when considering tumors with high APOBEC exposure (≥20%) as compared to not. No significant changes in CNAs were found between osimertinib-naïve and resistant samples when stratifying by APOBEC mutational signature <20% vs ≥20% (**Supplementary Fig. 4E-F**). These findings

indicate that APOBEC mutagenesis is unlikely to be involved in the development of canonical resistance SNVs or CNAs, such as EGFR C797S or *MET* amplification.

APOBEC and its relationship with large scale chromosomal changes

Previous reports demonstrate that APOBEC-dependent kataegis and chromothripsis[32] provide a potential etiology for large scale chromosomal rearrangements in the setting of APOBEC mutagenesis, prompting an interest in investigating the potential relationship between structural rearrangements and APOBEC mutagenesis. Kataegis describes a pattern of localized hypermutation within a cancr genome while chromothripsis is a mutational process where thousands of chromosomal rearrangements occur in a single event clustered in a confined genomic region.[33] A subset of patients had sufficient tissue to perform WGS (10 paired pre-/post- samples, 6 pre- only, 11 post- only). In specific cases, kataegis and chromothripsis coincided with areas of the genome enriched with APOBEC mutations, as shown with Case WGS-01 (**Fig. 2C**). Collectively, when assessing all WGS samples, large scale genomic rearrangements including kataegis and chromothripsis were dispersed uniformly across the genome (**Fig. 2D**). Interestingly, tumors displaying dominant APOBEC signature had a statistically significant enrichment of the genome covered by structural variants ($p<0.01$; **Fig. 2E**) and kataegis ($p<0.03$; **Fig. 2F**).

We analyzed matched samples before and after osimertinib treatment to characterize the tumor evolution under therapeutic selection. Case WGS-02 (**Fig. 3A**) was derived from a patient treated with first-line osimertinib for eight months. At progression, MSK-IMPACT identified acquired *MET* amplification and *MET* H1094Y. Mutational signature analysis of the osimertinib-naïve and resistant paired samples for this patient revealed APOBEC mutagenesis restricted to the dominant clone in the osimertinib-resistant sample (Clone 3). Clone 3 harbored a *MET* H1094Y mutation, which is due to a nucleotide level T(C>T)A transition mutation that occurs in the APOBEC-preferred mutational context.

Circos plot of the osimertinib-resistant sample also demonstrates that APOBEC-rich genomic areas co-localize with kataegis and other structural rearrangements (tandem duplications, deletions, inversions and translocations). Similar findings are shown in case WGS-03 (**Fig. 3B**), a patient treated with first-line osimertinib and bevacizumab on a clinical trial for 19 months followed by disease progression; a mechanism of resistance was not identified. Consistent with the findings in WGS-02, clonal decomposition analysis of WGS-03 revealed the emergence of a post-therapy subclone (Clone 3) with a dominant APOBEC mutational signature, which was not observed prior to osimertinib treatment. These two cases suggest that, in a subset of patients, APOBEC mutagenesis may underpin the acquisition of resistance to osimertinib as observed in other tumors treated with targeted therapies[34]. Clones selected for upon osimertinib treatment harbor repertoires of somatic mutations and structural rearrangements distinct from those of the dominant pre-treatment clones, and the development of these new clones may be driven by the acquisition of mutagenic processes (e.g. APOBEC) that were not operative during the early stages of tumorigenesis.

Mechanism of resistance to osimertinib in EGFR-mutant LUADs

We sought to assess the relationship between mutational signatures such as APOBEC and established mechanisms of resistance to osimertinib. Tumors were classified into the following categories of resistance: lineage plasticity (small cell or squamous transformation), off-target (e.g. acquired fusions and *MET* amplification), on-target (*EGFR*-mediated), and unknown. Tumors were considered to be APOBEC-enriched if either the osimertinib-naïve or resistant sample was APOBEC dominant. Lineage plasticity and off-target mechanisms of resistance had the highest proportions of APOBEC-enriched samples (56% lineage plasticity, 41% off-target, 24% unknown, 23% on-target, **Fig. 4A**), corroborating the association between APOBEC and lineage plasticity previously reported by our group and others[13, 14]. Specifically, most cases of acquired fusions (3/5, 60%) and lineage plasticity (9/16, 56%) had APOBEC-as the dominant

mutational signature (**Fig. 4B**). Based on the type of substitution and its trinucleotide context, none of the acquired oncogenic mutations were directly attributable to APOBEC mutagenesis except for the case with an acquired MET H1094Y. There were no associations between pre-treatment or post-treatment mutational signature and overall survival (**Suppl Fig. 5A-D**). Exploration with a larger number of samples is warranted to investigate the relationship between APOBEC mutagenesis and certain mechanisms of resistance, including lineage plasticity and acquired structural rearrangements.

DISCUSSION

In this report, we surveyed mutational signatures in LUADs and identified an association between APOBEC mutagenesis and RTK-driven cancers. We demonstrate that the APOBEC mutational signature can be derived from a routine multi-gene panel test (i.e. MSK-IMPACT) obtained as standard of care in patients with metastatic LUADs. Mutational signatures offer an additional layer of genomic information that may clarify how tumors adapt to selective pressure on a genome-wide level, providing a broader perspective on resistance and nominating interventional strategies that could result in more durable suppression of tumor growth. Although WGS remains the ‘gold standard’ for mutational signature decomposition, by circumventing the need for more extensive WGS or WES data, which may be costly, require specialized core facilities and potentially additional patient material, the methodology described in this study is based upon sequencing data already routinely obtained for care of patients in the clinic. In fact, in a way akin to how TMB and microsatellite instability are now widely reported on commercial multigene panel testing, other mutational signatures, including APOBEC, could become clinically reportable and provide actionable results in the future.

Our data demonstrate that APOBEC mutagenesis, although present in osimertinib-naive samples, appears to occur relatively late in the development and progression of LUADs and increases with osimertinib exposure, with enrichment in metastatic tumors and subclonal as well

as private mutations. Such findings are corrobated by preclinical work that EGFR or ALK inhibitor treatment induces expression and mutagenic activity of APOBEC3B in tumor xenograft models and cell lines.[35] Separately, inhibiting therapy-induced APOBEC3A mutagenesis by gene deletion or RNA inhibitor-meditated suppression in lung cancer cell lines delayed resistance to targeted therapies.[36] Identifying a role for APOBEC mutagenesis in mediating drug resistance presents opportunities for abrogating the development of such resistance by potentially focusing on APOBEC inhibition, an area of active investigation.[37-39] Whilst there are currently no direct inhibitors for APOBEC activity, previous studies[40, 41] have proposed a potential role for ATR and PARP inhibitors. Preclinical studies demonstrate that activation of APOBEC-3A and APOBEC-3B may sensitize tumor cells to DNA damage response pathway inhibitors, such as ATR and PARP inhibitors; *in vitro* studies have shown that ATR inhibition can overcome osimertinib resistance in some cases.[42]

Consistent with previous observations[32], our data demonstrate that APOBEC mutagenesis is associated with kataegis and chromothripsis.[43] These phenomena can increase the probability of translocations leading to activating gene fusions that can drive resistance to EGFR TKIs. The association between APOBEC and lineage plasticity can be considered in analogy with simpler organisms, as lineage plasticity mirrors phenotypic switching seen in bacteria following antibiotic exposure.[44] In bacteria, larger scale genomic processes with functional similarities to APOBEC mutagenesis, including slipped strand repair, homologous recombination, site specific recombination, and transposon insertion and excision can all induce phenotypic switching.[45] APOBEC-derived catastrophic genomic events such as chromothripsis may result in profound reprogramming of cancer cells, creating a milieu that is conducive for phenotypic plasticity and resultant drug resistance. Alternatively, APOBEC mutagenesis may result in greater clonal diversity, providing the required conditions for the selection of de-differentiated subclones that have gained histologic plasticity.

This study has several limitations. Although WGS analysis was performed for a subset of pre- and post-osimertinib treated samples, mutational signature decomposition was largely performed by applying *SigMA* to MSK-IMPACT targeted sequencing data. Despite the agreement between MSK-IMPACT and WES using *SigMA* for the detection of APOBEC mutagenesis, our analysis may have overestimated the extent of APOBEC exposure across samples from all groups analyzed.

Indeed, the five SNVs cut-off applied for the inclusion of samples in our mutational signature analysis was found to display ~75% specificity against the gold-standard WES. The specificity, however, was higher when ≥10 and ≥15 SNVs were used as cut-off values (86% and 85%, respectively). Nonetheless, the inclusion of only samples with more than 10 SNVs would result in the exclusion of a large proportion of samples from the analysis, inevitably creating a selection bias for LUADs with higher mutation burden. Due to limited availability of paired tumor samples obtained before and after a given treatment, we did not investigate whether APOBEC enrichment following targeted therapy would also occur in the context of chemotherapy and immunotherapy treatment. In fact, here we focused on EGFR-mutant lung cancer; the analysis of other RTK driver lung cancers, including *ALK*, *RET* and *ROS1*-positive lung cancers, in sufficiently powered cohorts should be entertained in future studies.

Although the mechanistic basis for the induction of APOBEC mutagenesis in cancer cells and the causes of the selective enrichment for this phenomenon in certain clinical contexts remains to be fully elucidated, the association with RTK-driven lung cancers, targeted therapy exposure, and selected mechanisms of resistance provide clear directions for further study. APOBEC mutagenesis appears to constitute a transient, intermittent process[16], and deciphering the factors that induce and suppress episodic APOBEC mutagenesis during cancer evolution may inform whether this process might be harnessed therapeutically. Mutational

signatures as a biomarker for treatment have been established with homologous repair deficiency and microsatellite instability, which can select patients for treatment with DNA-damaging agents[46] and immunotherapy, respectively. The value of mutational signatures is heighted by the ability to infer signatures from multigene panel sequencing, which is frequently obtained for routine clinical care for patients with LUADs; at present, however, detecting APOBEC mutagenesis clinically is not warranted for lung cancer patients. Ultimately, the clinical relevance of APOBEC mutagenesis centers around therapeutic vulnerability. Future studies focusing on the development of approaches seeking to prevent the development of resistance by blocking the APOBEC mutagenesis process as well synthetic lethal approaches specifically targeting cancer cells with active APOBEC mutagenesis are warranted.

FUNDING: SC supported by NCI-P50CA247749 and NCI-R01CA249666; CML supported in part by EGFR Resisters/LUNGevity Lung Cancer Research Award, GO2 Foundation ALK Positive Research Award, R01CA227833, P30-CA086485, UG1CA233259, and P01CA129243; HAY supported in part by EGFR Resisters/LUNGevity Lung Cancer Research Award, STARR foundation, and R01CA264078. MSK authors supported by National Cancer Institute Cancer Center Core Grant No. P30-CA008748.

FINANCIAL DISCLOSURES:
NJC: Honoraria from MJH Life Sciences; royalties from Wolters Kluwer (Pocket Oncology). **MFB**: consulting or advisory role for Roche, Eli Lilly, PetDx; research funding from Grail. **ML:** consulting or advisory role for Takeda Oncology, Janssen Pharmaceuticals. **CMR** consulting or advisory role for AbbVie, Amgen, Astra Zeneca, Epizyme, Genentech/Roche, Ipsen, Jazz, Lilly, and Syros; serves on the scientific advisory boards of Bridge Medicines, Earli, and Harpoon Therapeutics. **SC:** Consulting or Advisory Role:, Novartis, Lilly, Sanofi, Paige.AI Research Funding: Daiichi Sankyo (Inst), Paige.ai (Inst) Patents, Royalties, Other Intellectual Property: Patents for (1) targeting mutant estrogen receptor (ER) with ER proteolysis targeting chimera (PROTACS) and (2) detecting genomic and histologic alterations in breast cancer using machine learning algorithms (Inst) Travel, Accommodations, Expenses: Bristol Myers. **CML** is a consultant/advisory board member for Pfizer, Novartis, AstraZeneca, Genoptix, Sequenom, Ariad, Takeda, Blueprints Medicine, Cepheid, Foundation Medicine, Roche, Achilles Therapeutics, Genentech, Syros, Amgen, EMD-Serono, Eli Lilly, Daiichi-Sanko, and PUMA and has received research grants (to her University) from Xcovery, AstraZeneca, and Novartis. None of the grants are active at this time. **JSR-F** reports receiving personal/consultancy fees from Goldman Sachs, REPARE Therapeutics, Paige.AI and Eli Lilly, membership of the

scientific advisory boards of VolitionRx, REPARE Therapeutics, Paige.AI and Personalis, membership of the Board of Directors of Grupo Oncoclinicas, and ad hoc membership of the scientific advisory boards of Roche Tissue Diagnostics, Ventana Medical Systems, Novartis, Genentech, MSD, Daiichi Sankyo and InVicro, outside the scope of this study and owns Paige.AI and REPARE Therapeutics stocks. **HY**—consultant or advisory role for AstraZeneca, Janssen, Blueprint Med, C4 Therapeutics, and Daiichi. Research funding: AstraZeneca (Inst), Pfizer (Inst), Daiichi (Inst), Cullinan (Inst), Lilly (Inst), and Novartis (Inst). All remaining authors have declared no conflicts of interest.

REFERENCES:

1. Schoenfeld AJ, Chan JM, Kubota D et al. Tumor Analyses Reveal Squamous Transformation and Off-Target Alterations As Early Resistance Mechanisms to First-line Osimertinib in EGFR-Mutant Lung Cancer. Clin Cancer Res 2020; 26: 2654-2663 PubMed .
2. Roper N, Brown AL, Wei JS et al. Clonal Evolution and Heterogeneity of Osimertinib Acquired Resistance Mechanisms in EGFR Mutant Lung Cancer. Cell Rep Med 2020; 1.
3. de Bruin EC, McGranahan N, Mitter R et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. Science 2014; 346: 251-256 PubMed .
4. Swanton C, McGranahan N, Starrett GJ, Harris RS. APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. Cancer Discov 2015; 5: 704-712 PubMed .
5. Zhang J, Fujimoto J, Zhang J et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. Science 2014; 346: 256-259 PubMed .
6. André T, Shiu K-K, Kim TW et al. Pembrolizumab in Microsatellite-Instability–High Advanced Colorectal Cancer. New England Journal of Medicine 2020; 383: 2207-2218.
7. Marcus L, Lemery SJ, Keegan P, Pazdur R. FDA Approval Summary: Pembrolizumab for the Treatment of Microsatellite Instability-High Solid Tumors. Clin Cancer Res 2019; 25: 3753-3758 PubMed .
8. González-Martín A, Pothuri B, Vergote I et al. Niraparib in Patients with Newly Diagnosed Advanced Ovarian Cancer. New England Journal of Medicine 2019; 381: 2391-2402.
9. Longo DL. Personalized Medicine for Primary Treatment of Serous Ovarian Cancer. New England Journal of Medicine 2019; 381: 2471-2474.
10. Mirza MR, Monk BJ, Herrstedt J et al. Niraparib Maintenance Therapy in Platinum-Sensitive, Recurrent Ovarian Cancer. New England Journal of Medicine 2016; 375: 2154-2164.
11. Soria JC, Ohe Y, Vansteenkiste J et al. Osimertinib in Untreated EGFR-Mutated Advanced Non-Small-Cell Lung Cancer. N Engl J Med 2018; 378: 113-125 PubMed .

12. Leonetti A, Sharma S, Minari R et al. Resistance mechanisms to osimertinib in EGFR-mutated non-small cell lung cancer. British Journal of Cancer 2019; 121: 725-737.
13. Lee J-K, Lee J, Kim S et al. Clonal History and Genetic Predictors of Transformation Into Small-Cell Carcinomas From Lung Adenocarcinomas. Journal of Clinical Oncology 2017; 35: 3065-3074.
14. Offin M, Chan JM, Tenet M et al. Concurrent RB1 and TP53 Alterations Define a Subset of EGFR-Mutant Lung Cancers at risk for Histologic Transformation and Inferior Clinical Outcomes. J Thorac Oncol 2019; 14: 1784-1793 PubMed .
15. Venkatesan S, Rosenthal R, Kanu N et al. Perspective: APOBEC mutagenesis in drug resistance and immune escape in HIV and cancer evolution. Ann Oncol 2018; 29: 563-572 PubMed .
16. Petljak M, Alexandrov LB, Brammeld JS et al. Characterizing Mutational Signatures in Human Cancer Cell Lines Reveals Episodic APOBEC Mutagenesis. Cell 2019; 176: 1282-1294 PubMed .e1220.
17. McGranahan N, Favero F, de Bruin EC et al. Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. Sci Transl Med 2015; 7: 283ra254.
18. Gulhan DC, Lee JJ, Melloni GEM et al. Detecting the mutational signature of homologous recombination deficiency in clinical samples. Nat Genet 2019; 51: 912-919 PubMed .
19. Cheng DT, Mitchell TN, Zehir A et al. Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology. J Mol Diagn 2015; 17: 251-264 PubMed .
20. Zehir A, Benayed R, Shah RH et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat Med 2017; 23: 703-713 PubMed .
21. Shen R, Seshan VE. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. Nucleic Acids Res 2016; 44: e131.
22. Carter SL, Cibulskis K, Helman E et al. Absolute quantification of somatic DNA alterations in human cancer. Nat Biotechnol 2012; 30: 413-421 PubMed .
23. Ng CKY, Piscuoglio S, Geyer FC et al. The Landscape of Somatic Genetic Alterations in Metaplastic Breast Carcinomas. Clin Cancer Res 2017; 23: 3859-3870 PubMed .

24. Weigelt B, Bi R, Kumar R et al. The Landscape of Somatic Genetic Alterations in Breast Cancers From ATM Germline Mutation Carriers. J Natl Cancer Inst 2018; 110: 1030-1034 PubMed .
25. Guo Q, Lakatos E, Al Bakir I et al. The mutational signatures of formalin fixation on the human genome. bioRxiv 2021; 2021.2003.2011.434918.
26. Alexandrov LB, Nik-Zainal S, Wedge DC et al. Signatures of mutational processes in human cancer. Nature 2013; 500: 415-421 PubMed .
27. Alexandrov LB, Kim J, Haradhvala NJ et al. The repertoire of mutational signatures in human cancer. Nature 2020; 578: 94-101 PubMed .
28. Mayakonda A, Lin DC, Assenov Y et al. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res 2018; 28: 1747-1756 PubMed .
29. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 2016; 32: 2847-2849 PubMed .
30. Roper N, Gao S, Maity TK et al. APOBEC Mutagenesis and Copy-Number Alterations Are Drivers of Proteogenomic Tumor Evolution and Heterogeneity in Metastatic Thoracic Tumors. Cell Rep 2019; 26: 2651-2666 PubMed .e2656.
31. Gerstung M, Jolly C, Leshchiner I et al. The evolutionary history of 2,658 cancers. Nature 2020; 578: 122-128 PubMed .
32. Maciejowski J, Chatzipli A, Dananberg A et al. APOBEC3-dependent kataegis and TREX1-driven chromothripsis during telomere crisis. Nat Genet 2020; 52: 884-890 PubMed .
33. Nik-Zainal S, Alexandrov LB, Wedge DC et al. Mutational processes molding the genomes of 21 breast cancers. Cell 2012; 149: 979-993 PubMed .
34. Van Allen EM, Wagle N, Sucker A et al. The genetic landscape of clinical resistance to RAF inhibition in metastatic melanoma. Cancer Discov 2014; 4: 94-109 PubMed .
35. Mayekar MK, Caswell DR, Vokes NI et al. Targeted cancer therapy induces APOBEC fuelling the evolution of drug resistance. bioRxiv 2020; 2020.2012.2018 PubMed .423280.
36. Isozaki H, Abbasi A, Nikpour N et al. APOBEC3A drives acquired resistance to targeted therapies in non-small cell lung cancer. bioRxiv 2021; 2021.2001.2020 PubMed .426852.
37. Olson ME, Harris RS, Harki DA. APOBEC Enzymes as Targets for Virus and Cancer Therapy. Cell Chem Biol 2018; 25: 36-49 PubMed .
38. Matsumoto T, Shirakawa K, Yokoyama M et al. Protein kinase A inhibits tumor mutator APOBEC3B through phosphorylation. Scientific Reports 2019; 9: 8307.

39. Kvach MV, Barzak FM, Harjes S et al. Inhibiting APOBEC3 Activity with Single-Stranded DNA Containing 2'-Deoxyzebularine Analogues. Biochemistry 2019; 58: 391-400 PubMed .

40. Green AM, Budagyan K, Hayer KE et al. Cytosine Deaminase APOBEC3A Sensitizes Leukemia Cells to Inhibition of the DNA Replication Checkpoint. Cancer Research 2017; 77: 4579-4588 PubMed .

41. Buisson R, Lawrence MS, Benes CH, Zou L. APOBEC3A and APOBEC3B Activities Render Cancer Cells Susceptible to ATR Inhibition. Cancer Research 2017; 77: 4567-4578 PubMed .

42. Tanaka K, Yu HA, Yang S et al. Targeting Aurora B kinase prevents and overcomes resistance to EGFR inhibitors in lung cancer by enhancing BIM- and PUMA-mediated apoptosis. Cancer Cell 2021; 39: 1245-1261 PubMed .e1246.

43. Nik-Zainal S, Alexandrov LB, Wedge DC et al. Mutational processes molding the genomes of 21 breast cancers. Cell 2012; 149: 979-993 PubMed .

44. Chantratita N, Wuthiekanun V, Boonbumrung K et al. Biological relevance of colony morphology and phenotypic switching by Burkholderia pseudomallei. J Bacteriol 2007; 189: 807-817 PubMed .

45. A.M. Sousa IMaMOP. Phenotypic switching: an opportunity to bacteria thrive Science against microbial pathogens: communicating current research and technological advances 2011; 252-262.

46. Mukhopadhyay A, Plummer ER, Elattar A et al. Clinicopathological features of homologous recombination-deficient epithelial ovarian cancers: sensitivity to PARP inhibitors, platinum, and survival. Cancer Res 2012; 72: 5675-5682 PubMed .

FIGURE LEGENDS:

Figure 1: Mutational signatures derived from multigene panel sequencing. A) Performance evaluation of *SigMA* comparing whole exome sequencing (WES) data from The Cancer Genome Atlas (TCGA) lung adenocarcinoma (LUAD) cohort to that same data restricted to the MSK-IMPACT genomic footprint. Schematic demonstrating workflow for *SigMA* validation is shown on the left. Sensitivity, specificity, and accuracy are referred to the capacity of *SigMA* to detect APOBEC as dominant mutational signature compared to TCGA cohort. Performance is evaluated at cutoffs for samples with at least 5 single-nucleotide variants (SNVs), 10 SNVs, and 15 SNVs. **B)** Mutational signature heatmap of lung adenocarcinomas (n=3276) from the MSK-IMPACT cohort. Exposure of mutational signatures for each sample according to the group (receptor tyrosine kinase [RTK] driver [n=625], *KRAS* mutant [n=1389), Other [n=1262]). Exposures are reported as continuous variables and colored according to the legend (0 = white; 0.01-0.19 = orange shades; 0.2-1.0 = red shades). **C)** Stacked bar plots of dominant signatures across RTK driver, *KRAS* mutant, and Other groups. Dominant signature was selected according to the category assigned by *SigMA*. **D)** Comparison of APOBEC and smoking-related mutational signature exposures between RTK driver, *KRAS* mutant, and Other groups. The mean exposures attributed to smoking and APOBEC processes are plotted with error bars indicating standard deviations. Mutational signatures are color-coded according to the legend. Statistical comparisons were performed using two-sided Wilcoxon-Mann–Whitney test. RTK driver was considered as reference group. **** = $p < 0.001$

Figure 2: APOBEC mutagenesis in *EGFR*-mutant lung adenocarcinomas treated with osimertinib. **A)** Pie charts showing the proportions of cases with APOBEC dominant, APOBEC mutational signature exposure >20% but non-dominant, Other, and not assessable (<5 single nucleotide variants [SNVs]) signatures between pre- (n=93) and post-osimertinib (n=93) samples. **B)** Repertoire of somatic genetic alterations of lung adenocarcinomas treated with

osimertinib harboring APOBEC mutational signature ≥20% (pre=32; post=49). APOBEC mutational signature exposure, treatment status (pre vs. post osimertinib), dominant mutational signature (APOBEC vs. Other), and acquired resistance mechanism for each sample are represented and color-coded according to the legend. **C)** Region of overlap of chromothripsis and kataegis on chromosome 3 for a single sample with osimertinib resistance (WGS-Post01). APOBEC and non-APOBEC single-nucleotide variants ([SNVs] top) are shown along with regions of kataegis, structural variants ([SVs] upper-middle) with inversions as black dots connected by curved lines and translocations as green dots. A green curved line encompasses the segment of the chromosome that has undergone chromothripsis (middle). Integer copy number [CN] of segments of chromosome 3 are shown in the bottom graph. **D)** Genomic footprint of kataegis and chromothripsis across pre- and post-treatment whole-genome sequenced samples (n=36). **E)** Quantification of SVs and **F)** kataegis across the genome in APOBEC dominant whole-genome sequenced samples as compared to APOBEC non-dominant. Statistical comparisons were performed using two-sided Wilcoxon-Mann–Whitney test.

Figure 3: Large scale molecular patterns detected by WGS in *EGFR*-mutant lung adenocarcinomas. A) Circos plot of pre-treatment (top left) and post-treatment (top right) displaying (from inside to outside) inter-variant distance and apobec classification of single-nucleotide variants (SNVs), regions of kataegis, indels, copy number and structural variants of WGS-02-Pre/Post osimertinib and (B) WGS-03-Pre/Post. Clonality heatmap (bottom left) and fishtail plot (bottom right) showing cancer cell fractions (CCFs), mutations of interest, and mutational signatures in bioinformatically inferred clones.

Figure 4: Correlations between APOBEC and acquired resistance in *EGFR*-mutant lung adenocarcinomas. A) For 93 paired pre- and post-osimertinib samples, mechanism of

resistance subgroup (on target, off target, lineage plasticity, and unknown) and specific alteration acquired were correlated with presence of dominant APOBEC signature in either pre- or post-osimertinib sample. **B)** Table displaying specific acquired molecular alteration and APOBEC status.

SUPPLEMTARY METHODS

MSK-IMPACT targeted sequencing and whole-genome sequencing analyses

Genomic analysis was performed for all patients by the FDA-cleared New York State Department of Health-approved MSK-IMPACT assay with somatic genetic alterations detected, as previously described [1]. WGS sequencing data were processed using our validated bioinformatics pipeline [2, 3]. In brief, sequence reads were aligned to the reference human genome GRCh37 using the Burrows-Wheeler Aligner (BWA v0.7.15) [4]. For both WGS and MSK-IMPACT, somatic single nucleotide variants (SNVs) were detected with MuTect (v1.0) [5]. Insertion and deletions (indels) were detected using Strelka (v2.0.15) [6], VarScan2 (v2.3.7) [7], Platypus (v0.8.1) [8], Lancet (v1.0.0) [9], and Scalpel (v0.5.3) [10]. Copy number alterations (CNAs) and loss of heterozygosity were determined using FACETS [11]. Somatic mutations in tumor suppressor genes that were deleterious/loss-of-function or targeting a mutational hotspot in oncogenes were considered pathogenic. Mutations targeting hotspot loci were annotated using cancerhotspots.org[12].

Formalin-fixed paraffin-embedded (FFPE) artefact evaluation

To evaluate the presence of mutations stemming from the process of formalin fixation and paraffin embedding, the distribution of variant allele fractions (VAFs) across all single nucleotide substitution types was assessed to rule out the presence of C>T low VAF mutations which are a common feature of FFPE artefacts. Mutations were grouped into 3 categories; all mutations, less than 10% VAF and less than 5% VAF to compare if there was any difference in substitution frequencies amongst those 3 groups. Additionally, the distribution of nucleotide change frequencies was compared between pre- and post-treatment samples as well as between shared and private samples in cases where we had matched pre- and post-treatment data

available. Finally, we compared C>T mutations in the APOBEC context versus those that were not in the APOBEC context to ascertain if a significant enrichment of low VAF mutations in the APOBEC or non-APOBEC mutation groups was observed.

Genotyping, cancer cell fraction (CCF) calculation, clonal decomposition and signature analysis for WGS samples

Curated variant calls from each patient were checked in the bam files for each sample from the patient (genotyping) using SUFAM (https://github.com/inodb/sufam). Initial cancer cell fractions (CCFs) for each mutation were computed with ABSOLUTE [13] using variant allele frequencies from SUFAM and copy number alterations from FACETS [11]. CCFs were further refined and clustered using Dirichlet process with PhylogicNDT (PNDT) [14]. Mutation clusters identified by PNDT for each patient were plotted as heatmaps showing CCF levels for each cluster in each sample of the patient. Clusters with less than 5% of total mutations across all samples from a given patient were excluded. PhylogicNDT clusters were depicted using fishplot package [15] (https://github.com/chrisamiller/fishplot). CCFs refined by PhylogicNDT were used to infer the order of clones in pre- and post-treatment, and for other timepoints, fractions were simulated due to lack of data. Mutational signatures (COSMIC v3.1) for each mutation cluster were computed with *MutationalPatterns* [16]. *MutationalPatterns* was run in strict mode with bootstrapping. Median exposures across 100 iterations were picked for each signature. Signatures with the same etiology were combined (e.g., SBS2 and SBS13 for APOBEC).

Structural variant calls and circos plot

Manta [17] and SvABA [18] were used to call structure variants from whole-genome sequencing. The details for the setup and calling can be found at the following code repository, respectively: Manta- https://github.com/ipstone/modules/blob/master/sv_callers/mantaTN.mk,

SvABA: https://github.com/ipstone/modules/blob/master/sv_callers/svabaTN.mk. The VCF outputs are converted to Bedpe format with the vcfToBedpe script (https://github.com/ctsa/svtools) before being intersected with bedtools [19] pairToPair function with '-slop' option set at 25. The intersected structure variant calls are used with other genomics data (SNV, indels, copy number alterations) to generate the circos plots through the signature.tools.lib R package [20](https://github.com/Nik-Zainal-Group/signature.tools.lib).

Assessment of kataegis and chromothripsis in WGS samples

The R package KataegisPortal was used to identify loci of localised hypermutations[21]. The KataegisPortal code and documentation are available at https://github.com/MeichunCai/KataegisPortal. Copy number segment files and structural variant calls were inputted in to ShatterSeek [22] and regions of chromothripsis were filtered to only “high confidence” regions determined by oscillating copy number states and statistically significant breakpoint enrichment/fragment joins as previously described [22]. Kataegis, and chromothripsis regions were visualized using tagore, a python package created to visualize areas of interest on human chromosome ideograms, as previously described [23].

REFERENCES

1. Cheng DT, Mitchell TN, Zehir A et al. Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology. J Mol Diagn 2015; 17: 251-264.

2. Weigelt B, Bi R, Kumar R et al. The Landscape of Somatic Genetic Alterations in Breast Cancers From ATM Germline Mutation Carriers. J Natl Cancer Inst 2018; 110: 1030-1034.

3. Pareja F, Brandes AH, Basili T et al. Loss-of-function mutations in ATP6AP1 and ATP6AP2 in granular cell tumors. Nat Commun 2018; 9: 3533.

4. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 2010; 26: 589-595.

5. Cibulskis K, Lawrence MS, Carter SL et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 2013; 31: 213-219.
6. Saunders CT, Wong WS, Swamy S et al. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics 2012; 28: 1811-1817.
7. Koboldt DC, Zhang Q, Larson DE et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 2012; 22: 568-576.
8. Rimmer A, Phan H, Mathieson I et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat Genet 2014; 46: 912-918.
9. Narzisi G, Corvelo A, Arora K et al. Genome-wide somatic variant calling using localized colored de Bruijn graphs. Commun Biol 2018; 1: 20.
10. Narzisi G, O'Rawe JA, Iossifov I et al. Accurate de novo and transmitted indel detection in exome-capture data using microassembly. Nat Methods 2014; 11: 1033-1036.
11. Shen R, Seshan VE. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. Nucleic Acids Res 2016; 44: e131.
12. Chang MT, Asthana S, Gao SP et al. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. Nat Biotechnol 2016; 34: 155-163.
13. Carter SL, Cibulskis K, Helman E et al. Absolute quantification of somatic DNA alterations in human cancer. Nat Biotechnol 2012; 30: 413-421.
14. Leshchiner I, Livitz D, Gainor JF et al. Comprehensive analysis of tumour initiation, spatial and temporal progression under multiple lines of treatment. bioRxiv 2019; 508127.
15. Miller CA, McMichael J, Dang HX et al. Visualizing tumor evolution with the fishplot package for R. BMC Genomics 2016; 17: 880.
16. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Med 2018; 10: 33.
17. Chen X, Schulz-Trieglaff O, Shaw R et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 2016; 32: 1220-1222.
18. Wala JA, Bandopadhayay P, Greenwald NF et al. SvABA: genome-wide detection of structural variants and indels by local assembly. Genome Res 2018; 28: 581-591.
19. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010; 26: 841-842.

20. Degasperi A, Amarante TD, Czarnecki J et al. A practical framework and online tool for mutational signature analyses show inter-tissue variation and driver dependencies. Nat Cancer 2020; 1: 249-263.

21. Yin X, Bi R, Ma P et al. Multiregion whole-genome sequencing depicts intratumour heterogeneity and punctuated evolution in ovarian clear cell carcinoma. J Med Genet 2020; 57: 605-609.

22. Cortés-Ciriano I, Lee JJ, Xi R et al. Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. Nat Genet 2020; 52: 331-341.

23. Rishishwar L, Conley AB, Wigington CH et al. Ancestry, admixture and fitness in Colombian genomes. Sci Rep 2015; 5: 12376.

SUPPLEMENTARY TABLES

Supplementary Table 1. Frequency of single nucleotide substitutions in pre-treatment and post-treatment samples

	Pre-Treatment Mutations (N=160,067)	**Post-Treatment Mutations (N=180,943)**
C>A	33097 (20.7%)	30682 (16.9%)
C>G	22260 (13.9%)	36275 (20%)
C>T	44299 (27.7%)	57965 (32%)
T>A	16718 (10.4%)	15350 (8.5%)
T>C	28710 (17.9%)	26564 (14.7%)
T>G	14983 (9.4%)	14107 (7.8%)

Supplementary Table 2. Frequency of single nucleotide substitutions in all mutations, mutations at <10% variant allele frequency (VAF) and mutations at <5% VAF

	All mutations (N=300,334)	**Mutations <10% VAF (N=87,635)**	**Mutations <5% VAF (N=4,413)**
C>A	57218 (19%)	13387 (15.3%)	535 (12.1%)
C>G	53702 (17.9%)	15514 (17.7%)	994 (22.5%)
C>T	89833 (29.9%)	28273 (32.3%)	1290 (29.2%)
T>A	28035 (9.3%)	7141 (8.1%)	306 (6.9%)
T>C	46163 (15.4%)	15686 (17.9%)	916 (20.7%)
T>G	25383 (8.4%)	7634 (8.7%)	372 (8.4%)

Supplementary Table 3. Cohen's Kappa correlation coefficients for the agreement in detecting APOBEC as dominant signature in lung adenocarcinoma from the TCGA between whole exome sequencing (WES) and simulated MSK-IMPACT gene panel. SNVs: single nucleotide variants.

	Actual Positive	Actual Negative	K Score
>5 SNVs			0.72
Predicted Positive	37	11	
Predicted Negative	13	329	
>10 SNVs			0.77
Predicted Positive	24	8	
Predicted Negative	4	227	
>15 SNVs			0.81
Predicted Positive	17	4	
Predicted Negative	3	156	

Supplementary Table 4. Univariate analysis for APOBEC as dominant mutational signature

Characteristic	N	Event N	OR[1]	95% CI[1]	p-value	q-value[2]
Group	3,248	470			**<0.001**	**<0.001**
RTK Actionable Driver			—	—		
KRAS Mutant			0.30	0.23, 0.39		
Other			0.72	0.57, 0.91		
Gender	3,005	434			**0.004**	**0.005**
Female			—	—		
Male			1.36	1.10, 1.67		
Smoking status	2,067	319			**<0.001**	**<0.001**
Current			—	—		
Former			1.11	0.79, 1.60		
Never			4.54	3.14, 6.67		
Age category	3,239	469			0.11	0.11
<65 years			—	—		
≥65 years			0.85	0.69, 1.04		
Sample type	3,248	470			**<0.001**	**<0.001**
Metastasis			—	—		
Primary			0.59	0.48, 0.72		

[1]OR = Odds Ratio, CI = Confidence Interval

[2]False discovery rate correction for multiple testing

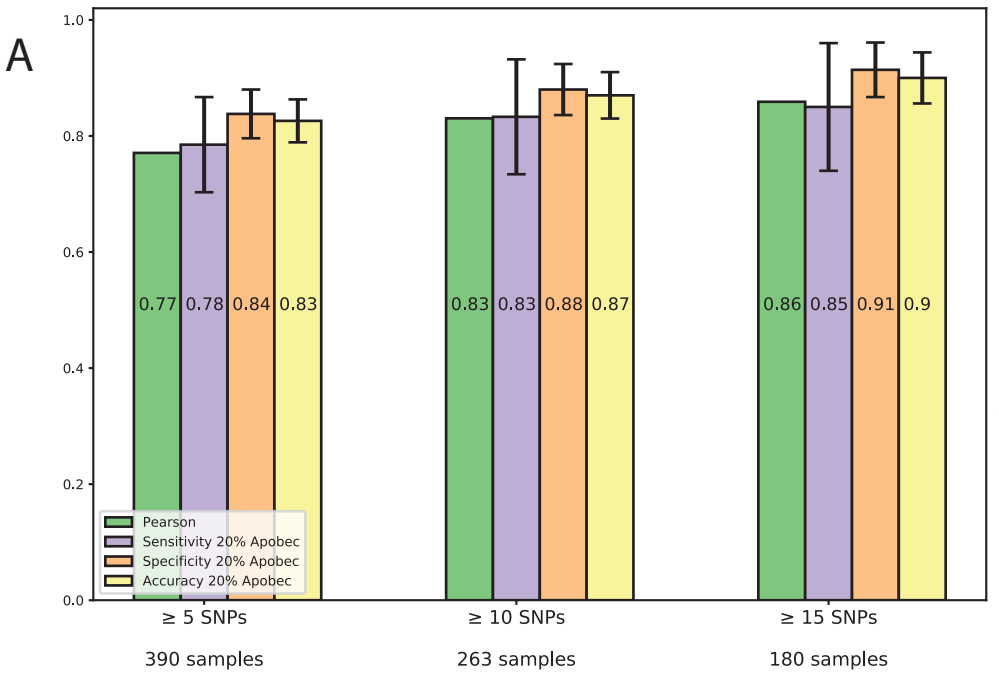
Supplementary Table 5. Multivariate analysis for APOBEC as dominant mutational signature

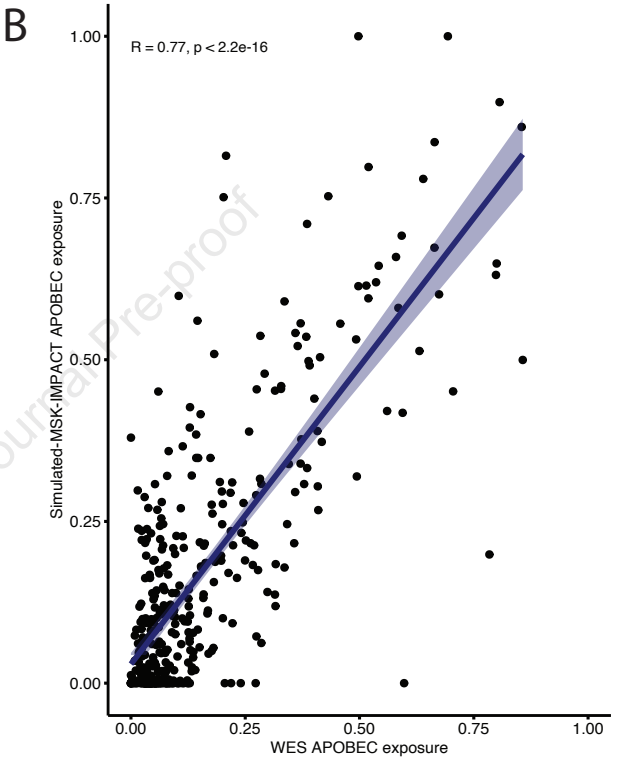
Characteristic	N	Event N	OR[1]	95% CI[1]	p-value	q-value[2]
Group	2,011	309			**<0.001**	**<0.001**
RTK Actionable Driver			—	—		
KRAS Mutant			0.58	0.38, 0.88		
Other			1.24	0.88, 1.77		
Gender	2,011	309			**0.021**	**0.028**
Female			—	—		
Male			1.35	1.05, 1.75		
Smoking status	2,011	309			**<0.001**	**<0.001**
Current			—	—		
Former			1.06	0.74, 1.53		
Never			3.70	2.36, 5.86		
Sample type	2,011	309			**0.037**	**0.037**
Metastasis			—	—		
Primary			0.76	0.59, 0.98		

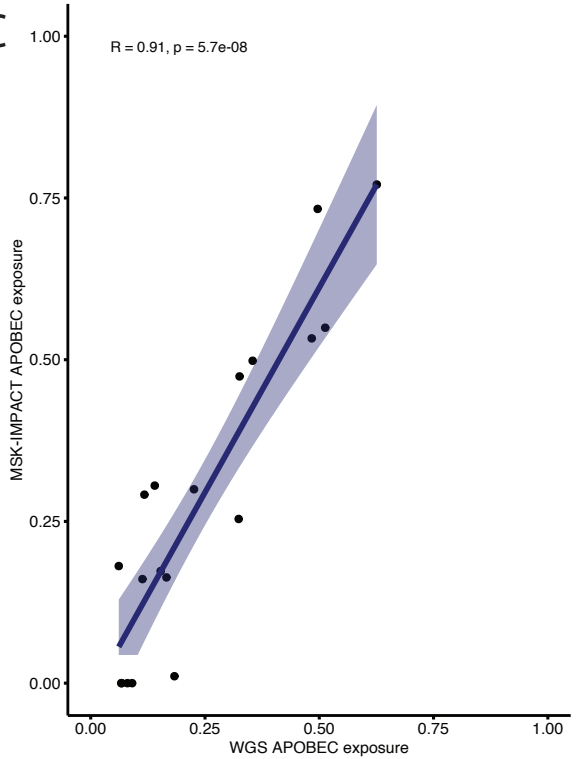
[1]OR = Odds Ratio, CI = Confidence Interval

[2]False discovery rate correction for multiple testing

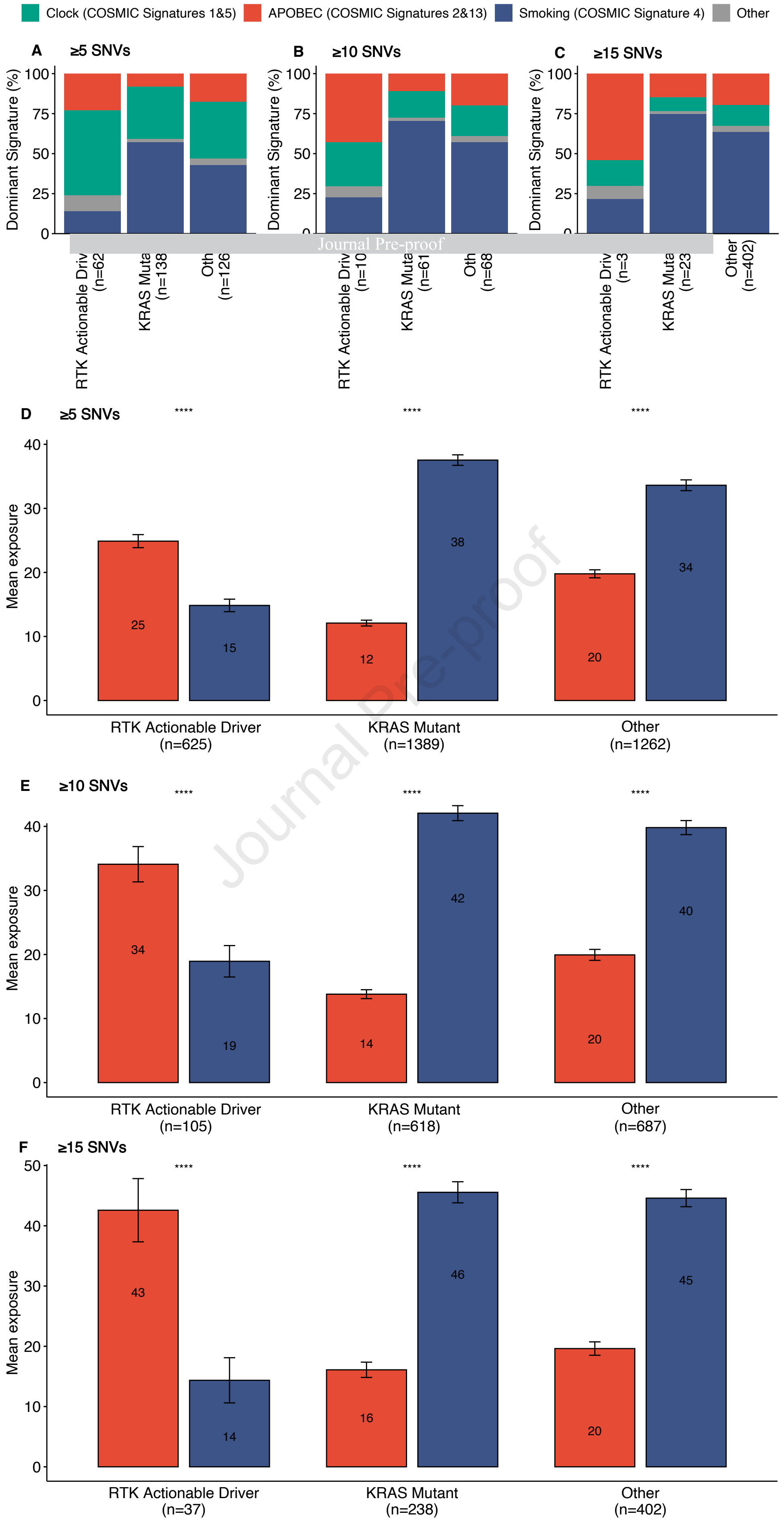
Supplemental Figure 1



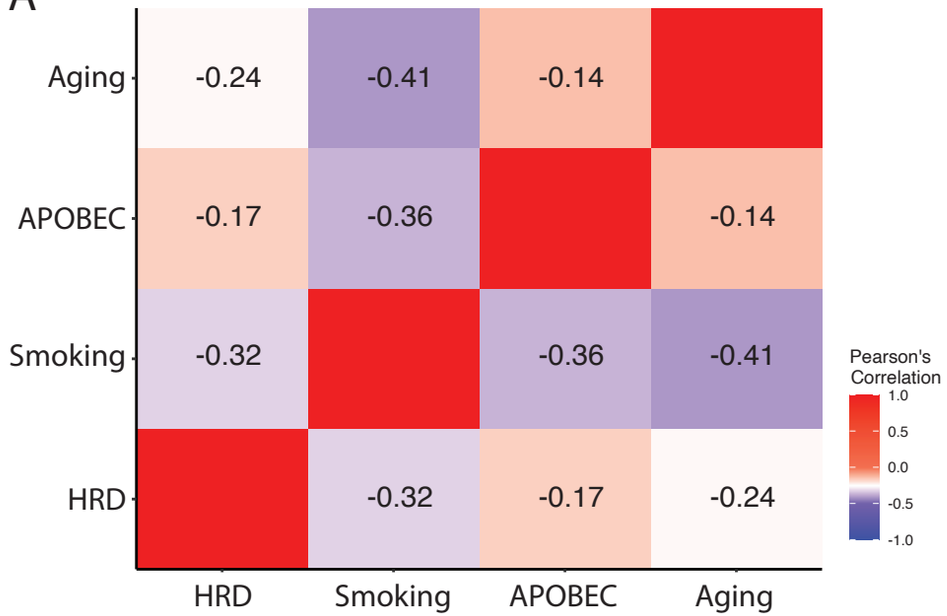


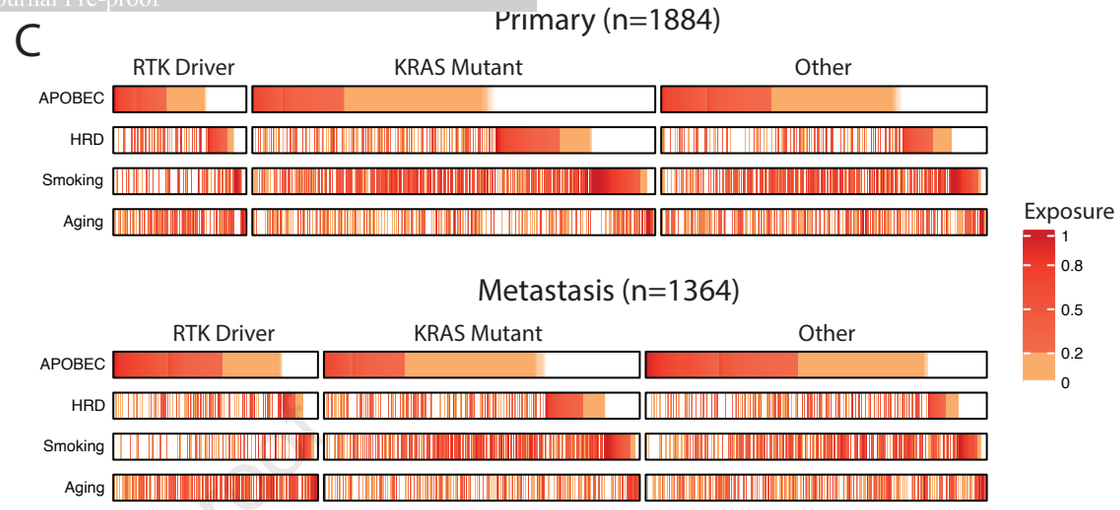


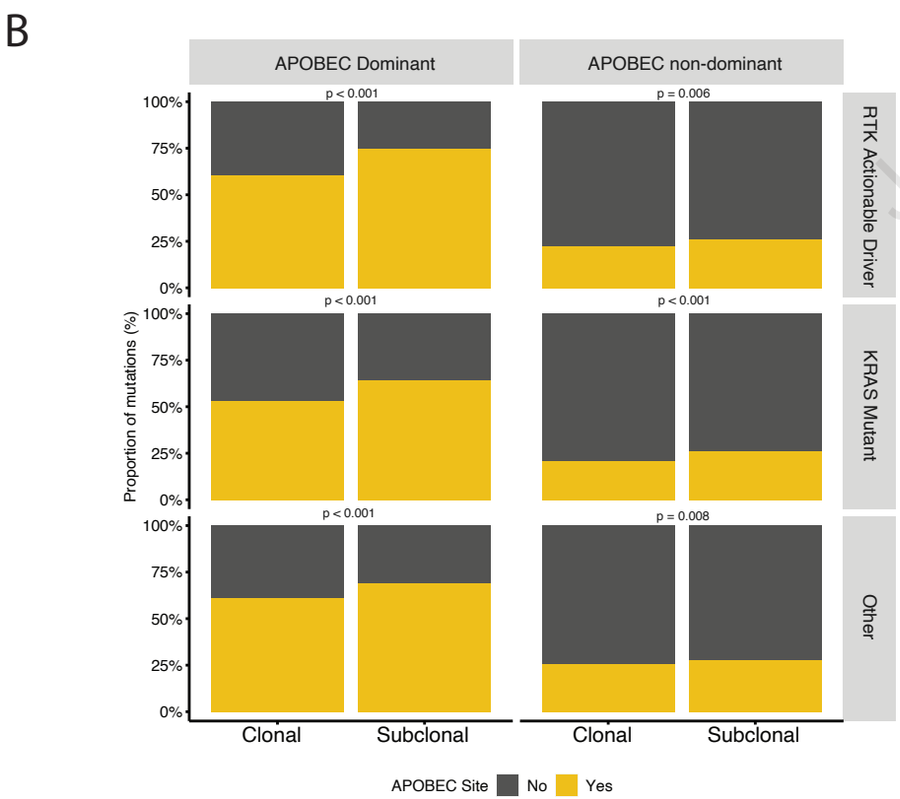
Supplemental Figure 2

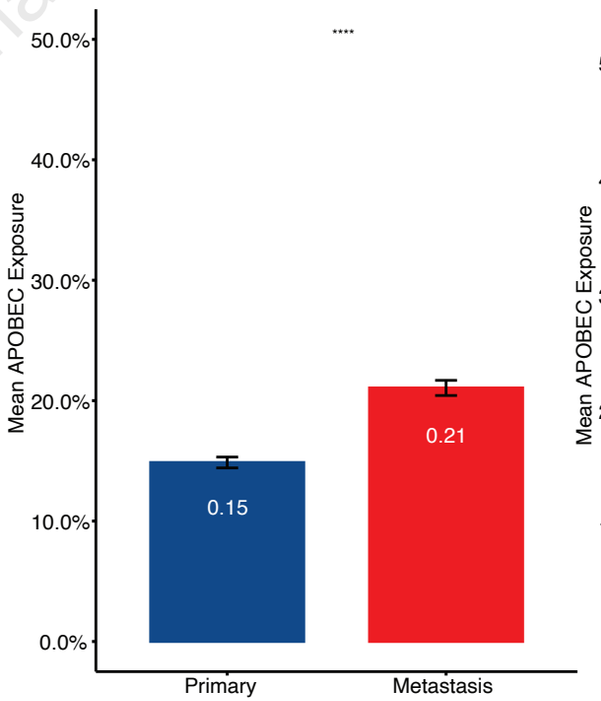


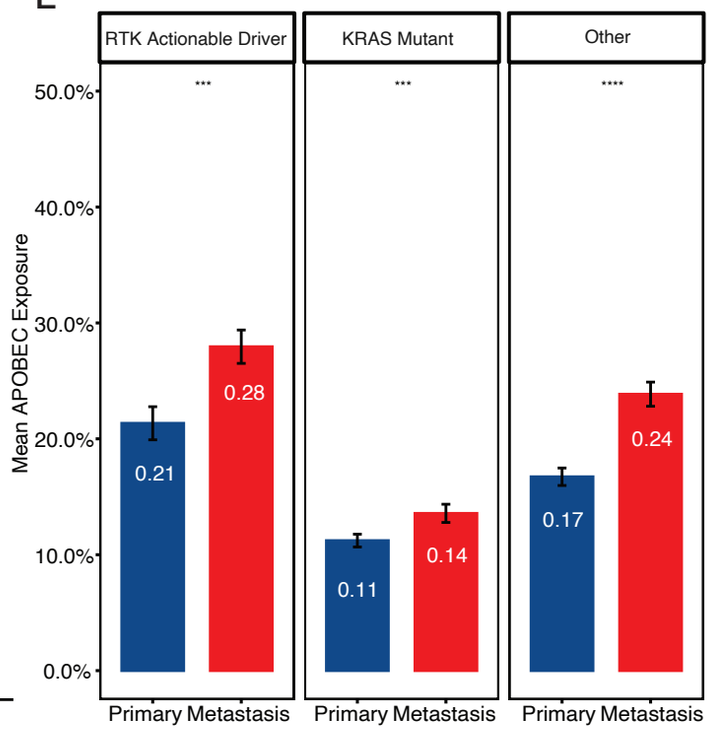
Supplemental Figure 3



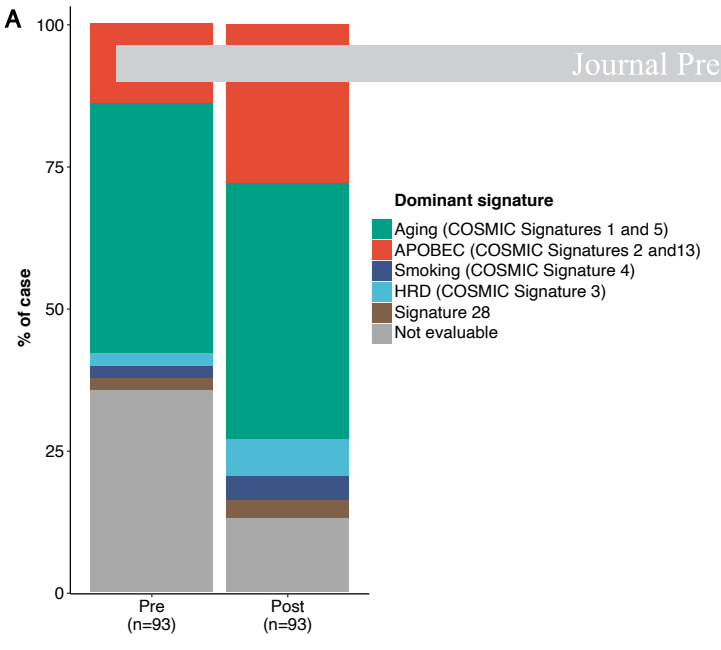


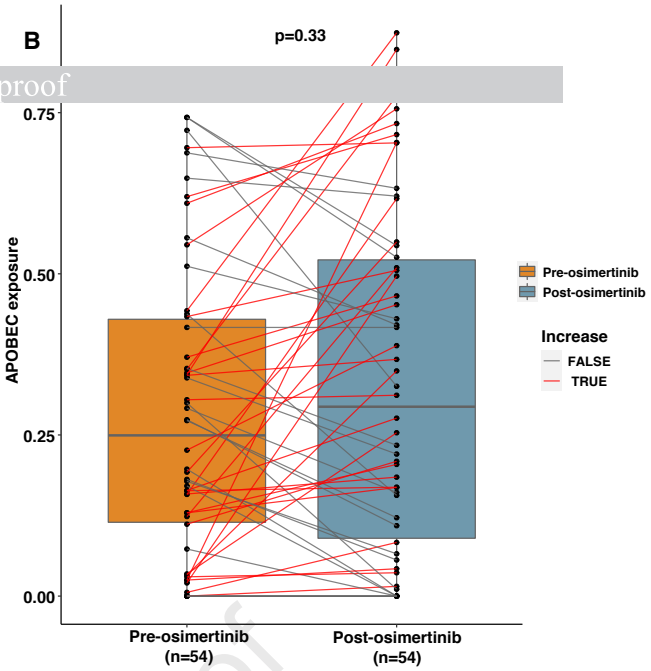


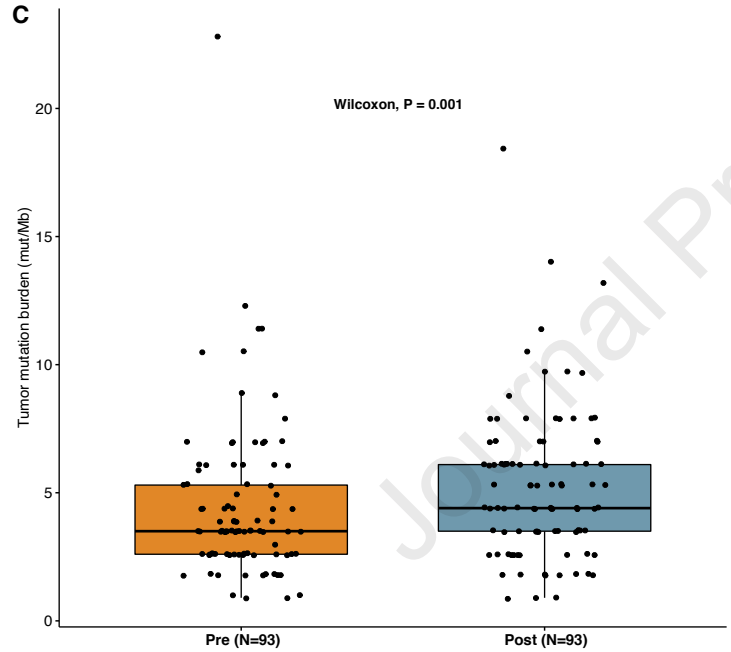


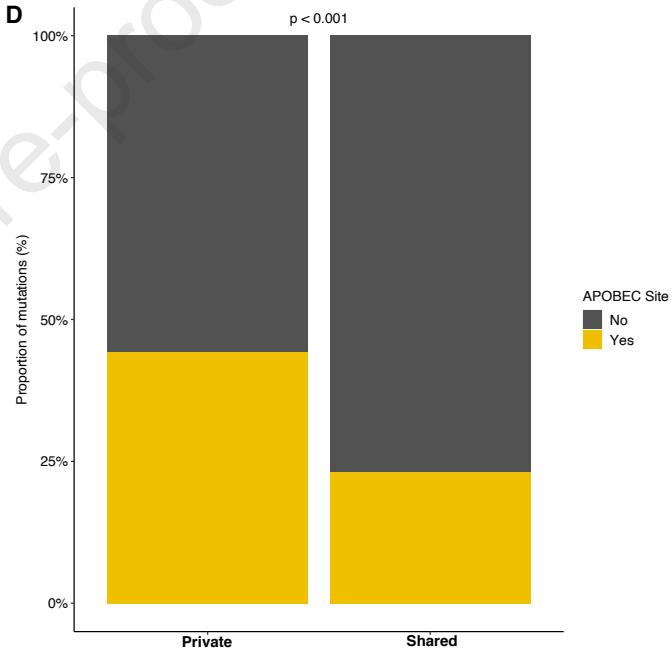


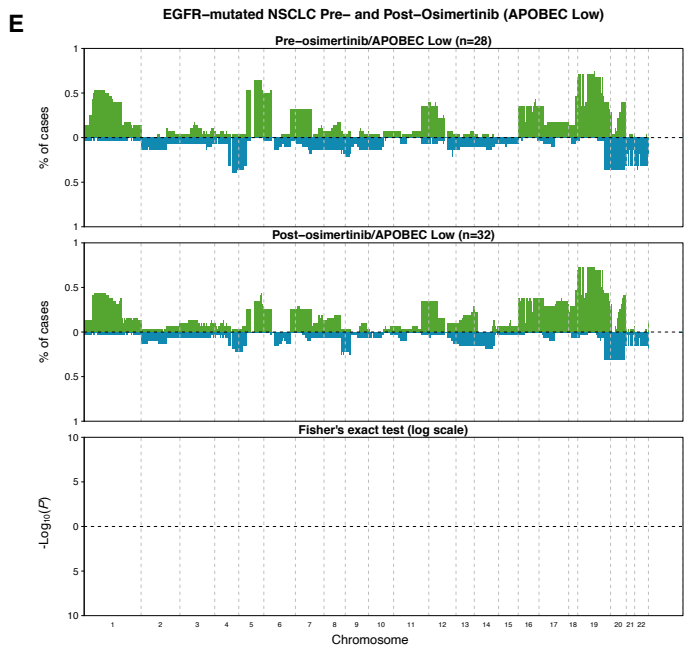
Supplemental Figure 4

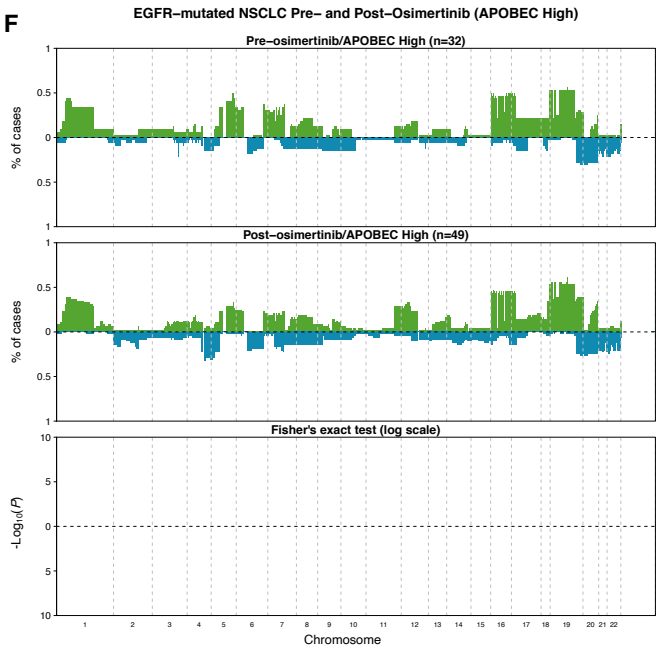




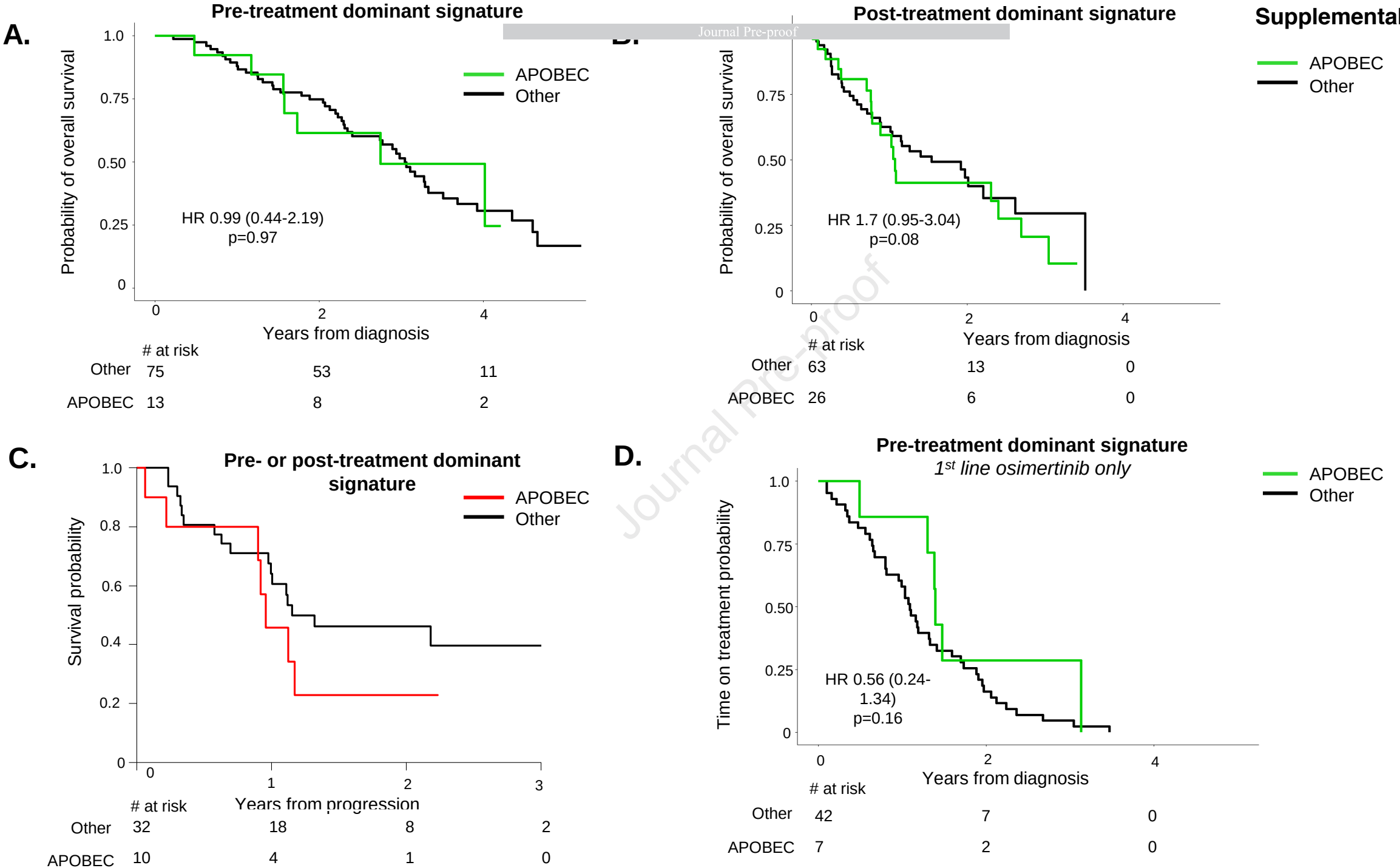








Supplemental Figure 5



SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure 1. Evaluation of APOBEC mutagenesis in lung adenocarcinomas sequenced by MSK-IMPACT targeted multi-gene panel.
A. Performance evaluation of *SigMA* on simulated MSK-IMPACT data from TCGA lung adenocarcinoma (LUAD). Pearson's correlation of APOBEC exposures obtained through whole-exome sequencing (WES) and MSK-IMPACT panel data obtained from simulation through down-sampling of the WES. Sensitivity, specificity, and accuracy are referred to the capacity of *SigMA* to detect APOBEC exposures higher than the 20% threshold (i.e., APOBEC is greater than 20% of the total exposure). The performance was evaluated using three different cutoffs for the minimum number of single nucleotide variants (SNVs) detected from the simulated MSK-IMPACT panel (≥5 SNVs, n=390 samples; ≥10 SNVs, n=263 samples; ≥15 SNVs, n=180=samples).
B. Correlation between APOBEC exposure inferred from WES data (TCGA) and APOBEC exposure from simulated MSK-IMPACT targeted sequencing. Scatter plot depicting the correlation of APOBEC mutational signatures exposure in samples that underwent both WES and simulated MSK-IMPACT (n=390), by down-sampling the TCGA WES data to match the MSK-IMPACT genomic footprint. The regression line and 95% confidence intervals are highlighted in blue. Statistical correlation was performed using the Pearson's correlation test. **C. Correlation between APOBEC exposure calculated from whole-genome sequencing data and APOBEC exposure from MSK-IMPACT targeted sequencing performed in matched samples.** Scatter plot shows the correlation of APOBEC mutational signatures exposure in samples (n=18) that underwent both whole-genome sequencing and MSK-IMPACT sequencing. The regression line and 95% confidence intervals are highlighted in blue. Statistical correlation was performed using the Pearson's correlation test.

Supplementary Figure 2. APOBEC mutagenesis in lung adenocarcinoma according to pre-specified single nucleotide variants (SNVs) cut-offs.
A-C. Stacked bar plots of dominant signatures across RTK driver, KRAS mutant, and Other groups using different single nucleotide variants (SNVs) as input cut-off, ≥ 5 SNVs (A), ≥10 SNVs (B), and ≥15 SNVs (C). Dominant signature was selected according to the category assigned by SigMA. **D-F. Comparison of APOBEC and smoking-related mutational signature exposures between RTK driver, KRAS mutant, and Other groups, using different samples with varying numbers of single nucleotide variants (SNVs) as input cut-off, namely ≥ 5 SNVs (D), ≥10 SNVs (E), and ≥15 SNVs (F).** The mean exposures attributed to smoking and APOBEC processes are plotted with error bars indicating standard deviations. Mutational signatures are color-coded according to the legend. Statistical comparisons were performed using two-sided Wilcoxon-Mann–Whitney test. RTK driver was considered as reference group. **** = $p < 0.001$.

Supplementary Figure 3. APOBEC mutagenesis in primary and metastatic lung adenocarcinomas.
A. Correlations of mutational signatures exposures in lung adenocarcinoma. The heatmap matrix shows the correlation between relevant mutational signatures, such as Aging, APOBEC, Smoking and homologous recombination deficiency (HRD), in lung adenocarcinoma samples from the MSK-IMPACT cohort (n=3276). Legend indicates strength of correlation coefficient

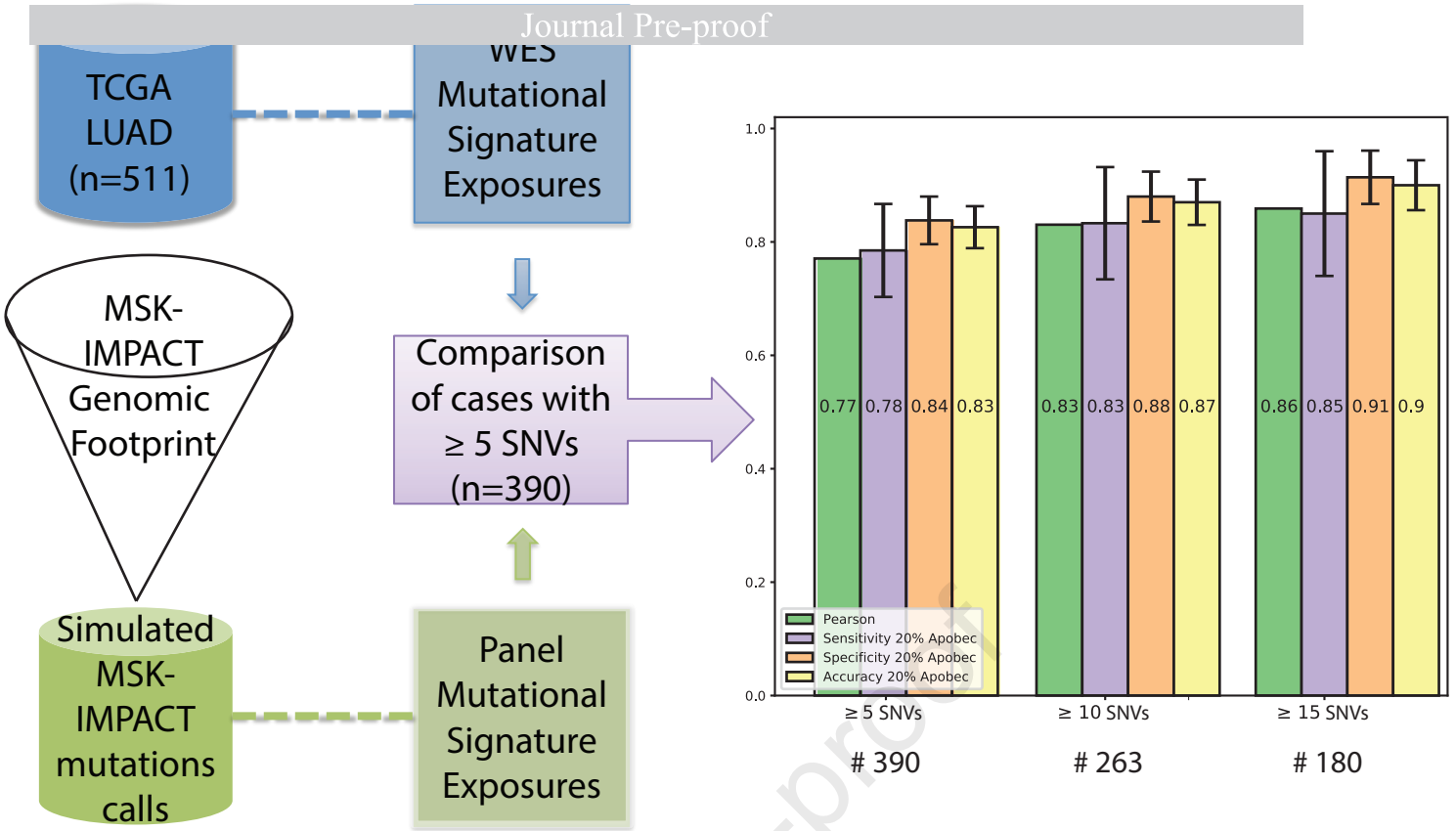
(red: high correlation; blue: weak correlation). Statistical correlation was performed using the Pearson's correlation test. **B. APOBEC site mutations according to the clonal status.** Barplots show the proportion of APOBEC site vs non-APOBEC site mutations by the clonal status assessed for the given mutation. Subgroup analysis by APOBEC dominant vs non-APOBEC dominant (columns) and different molecular subgroups (rows) is displayed (RTK driven [N=625], KRAS mutant [N=1389), Other [N=1262]). **C. Mutational signature heatmaps of primary (n=1884) and metastatic (n=1364) lung adenocarcinomas from MSK-IMPACT cohort.** The heatmap shows the landscape of mutational signatures from primary and metastatic lung adenocarcinomas based on different clinically relevant subtypes (RTK Driver, KRAS mutant and Other). Exposure of mutational signatures for each sample according to the group (RTK driven [N=625], KRAS mutant [N=1389), Other [N=1262]) and dominant mutational signature category. Exposures are reported as continuous variables and colored according to the legend (0 = white; 0.01-0.19 = orange shades; 0.2-1.0 = red shades). **D. APOBEC exposure in primary versus metastatic lung adenocarcinoma.** Plotted are the mean exposure of APOBEC, with error bars indicating standard deviations. Statistical comparisons were performed using two-sided Wilcoxon-Mann–Whitney test. ns, not significant; * = P < 0.05, ** = P < 0.01, *** = P < 0.001. **E. APOBEC exposure in primary versus metastatic lung adenocarcinoma, according to the clinical subtype.** Plotted are the mean exposure of APOBEC, with error bars indicating standard deviations. Statistical comparisons were performed using two-sided Wilcoxon-Mann–Whitney test. ns, not significant; * = p < 0.05, ** = p< 0.01, *** = p< 0.001.

Supplementary Figure 4. APOBEC mutagenesis in pre- and post-osimertinib EGFR-mutant lung adenocarcinomas.

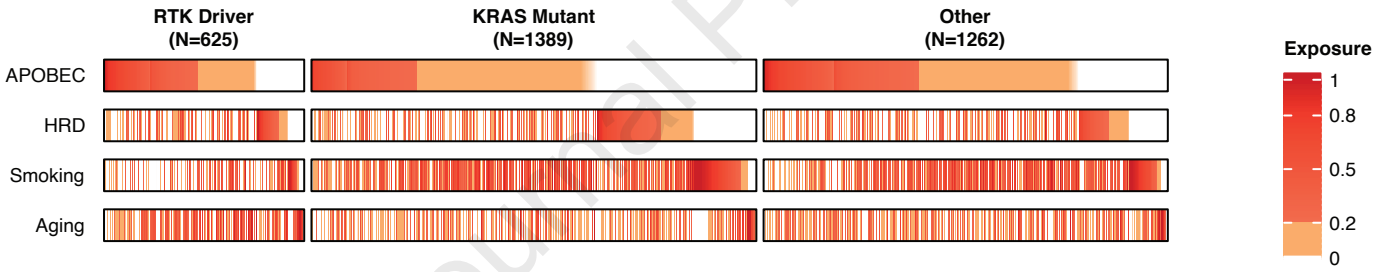
A. Stacked bar plots of dominant signatures across pre- (N=93) and post-osimertinib (N=93) subgroups. Dominant signature was selected according to the category assigned by *SigMA*. **B. APOBEC mutational signature exposure in paired pre- and post-osimertinib samples (n=54).** Plotted are the median, with error bars indicating interquartile ranges. Lines are colored by type of change in exposure (increase vs not-increase) according to the legend. Statistical comparisons were performed using two-sided Wilcoxon signed-rank test. **C. Somatic tumor mutation burden (TMB) APOBEC exposure in in pre-osimertinib and post-osimertinib lung adenocarcinoma.** Plotted are the median, with error bars indicating interquartile ranges. Statistical comparisons were performed using two-sided Wilcoxon-Mann–Whitney test. **D. Proportion of mutations private to the post-treatment or shared with the pre-treatment sample from 9 paired cases analyzed by whole genome sequencing (WGS).** Barplots show the proportion of mutations based on the condition of being an APOBEC site (yellow) or not (grey). P value indicates statistical significance by Fisher's exact test. **E-F. Patterns of copy number alterations in pre-osimertinib (top) and post-osimertinib (middle) lung adenocarcinoma. E)** Patterns of copy number alterations in APOBEC low (<20%) pre-osimertinib (top) and post-osimertinib (middle) lung adenocarcinoma. **F)** Patterns of copy number alterations in APOBEC high (>20%) pre-osimertinib (top) and post-osimertinib (middle) lung adenocarcinoma. At the bottom of each plot, $-\text{Log}_{10}$(P) by two-sided Fisher's exact tests, corrected for multiple testing. The frequencies of amplifications (green bars) and homozygous deletions (blue bars) are plotted above and below the x axis, respectively.

Supplementary Figure 5. **Clinical correlations in osimertinib-treated EGFR-mutant adenocarcinoma samples.**
A. Kaplan-Meier overall survival (OS) curves based on pre-treatment dominant signature (APOBEC vs. other). **B.** Kaplan-Meier OS based on post-treatment dominant signature (APOBEC vs other). **C.** Kaplan-Meier OS based on APOBEC as dominant signature in either the pre- or post-treatment sample. **D.** Kaplan-Meier TTD for patients treated with first line osimertinib stratified according to the pre-treatment dominant signature.

A



B



C

D

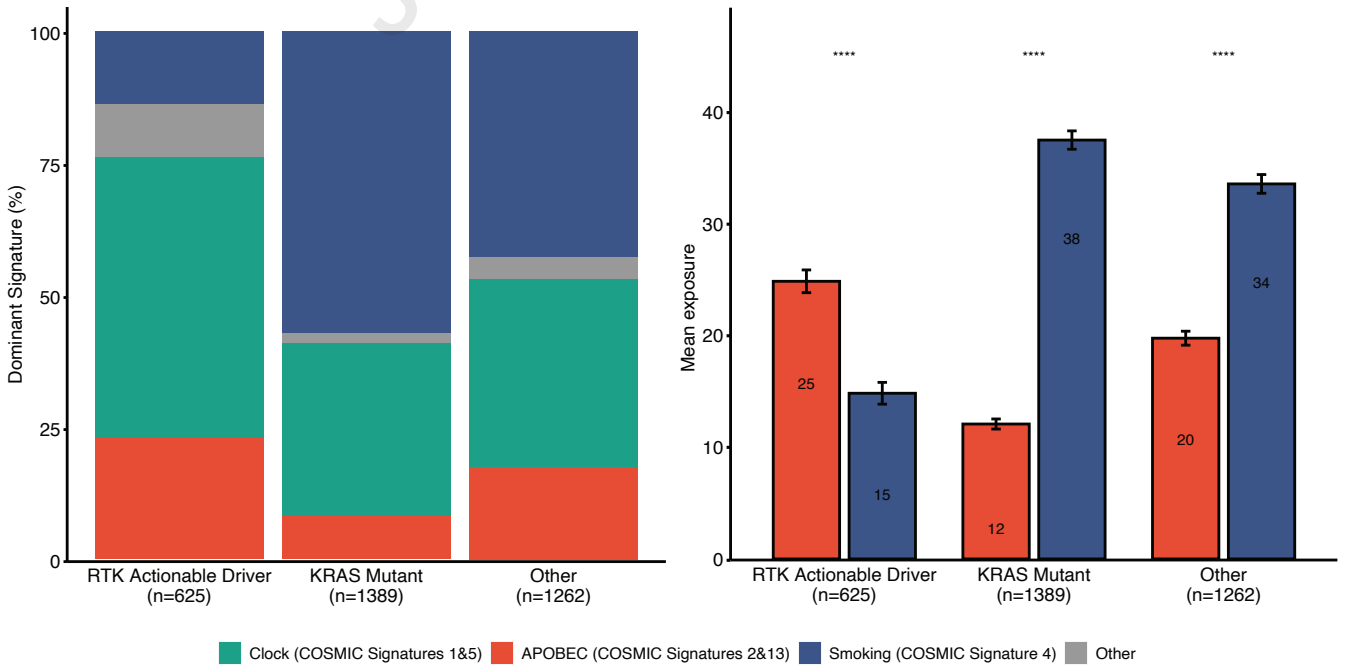


Figure 2

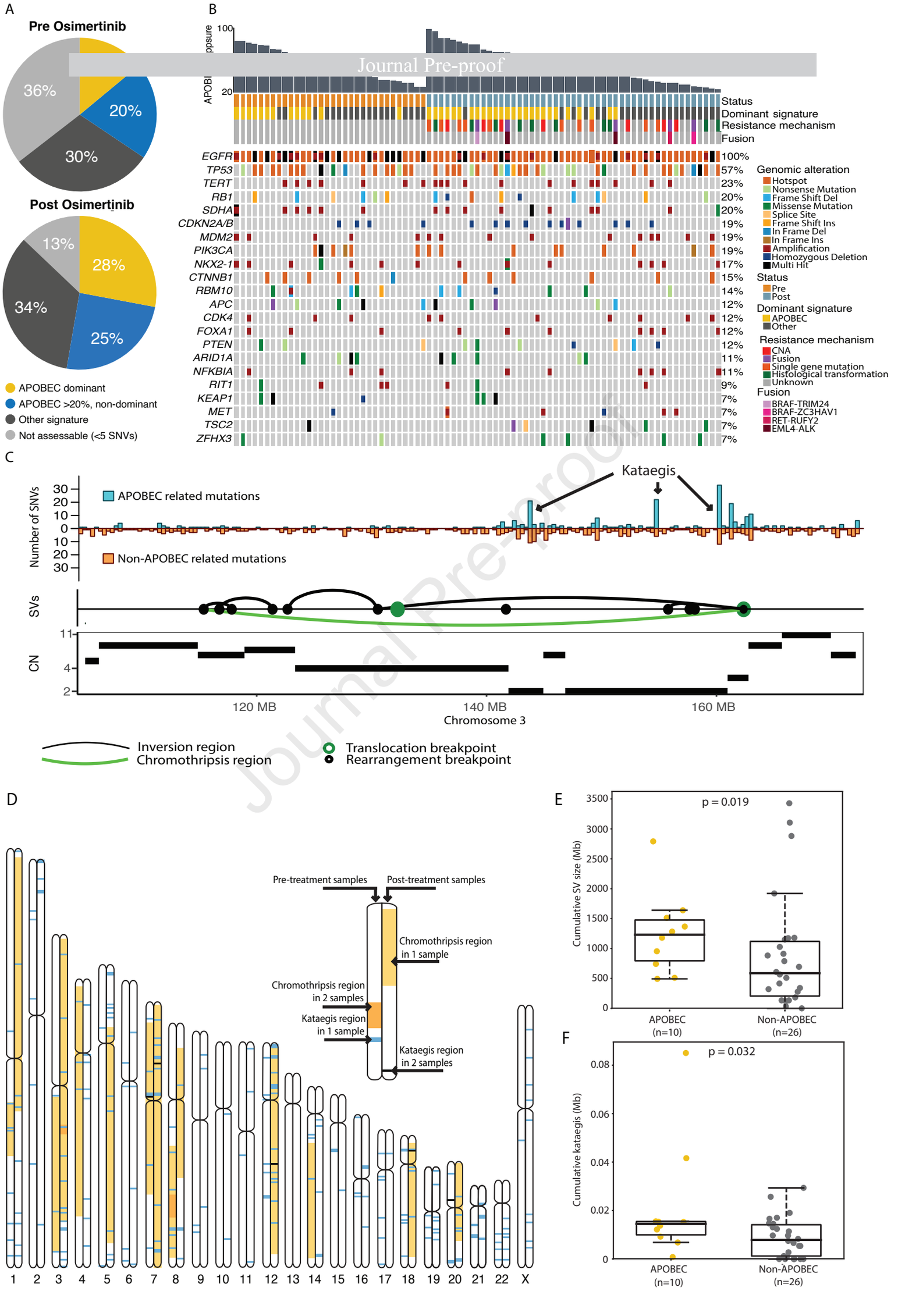
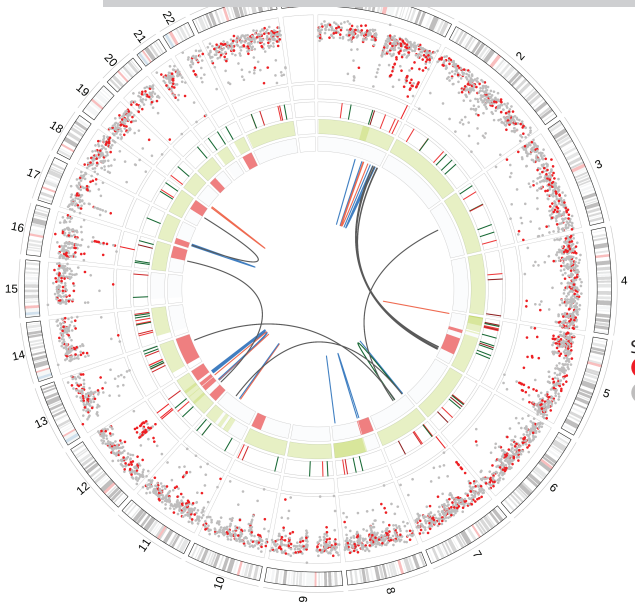

A
Pre Osimertinib
36%
20%
30%
Post Osimertinib
13%
28%
34%
25%
APOBEC dominant
APOBEC >20%, non-dominant
Other signature
Not assessable (<5 SNVs)
B
Journal Pre-proof
ppsure
APOBI
Status
Dominant signature
Resistance mechanism
Fusion
EGFR 100%
TP53 57%
TERT 23%
RB1 20%
SDHA 20%
CDKN2A/B 19%
MDM2 19%
PIK3CA 19%
NKX2-1 17%
CTNNB1 15%
RBM10 14%
APC 12%
CDK4 12%
FOXA1 12%
PTEN 12%
ARID1A 11%
NFKBIA 11%
RIT1 9%
KEAP1 7%
MET 7%
TSC2 7%
ZFHX3 7%
Genomic alteration
Hotspot
Nonsense Mutation
Frame Shift Del
Missense Mutation
Splice Site
Frame Shift Ins
In Frame Del
In Frame Ins
Amplification
Homozygous Deletion
Multi Hit
Status
Pre
Post
Dominant signature
APOBEC
Other
Resistance mechanism
CNA
Fusion
Single gene mutation
Histological transformation
Unknown
Fusion
BRAF-TRIM24
BRAF-ZC3HAV1
RET-RUFY2
EML4-ALK
C
Kataegis
Number of SNVs
APOBEC related mutations
Non-APOBEC related mutations
SVs
CN
120 MB
140 MB
160 MB
Chromosome 3
Inversion region
Chromothripsis region
Translocation breakpoint
Rearrangement breakpoint
D
Pre-treatment samples
Post-treatment samples
Chromothripsis region in 1 sample
Chromothripsis region in 2 samples
Kataegis region in 1 sample
Kataegis region in 2 samples
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X
E
p = 0.019
Cumulative SV size (Mb)
APOBEC (n=10)
Non-APOBEC (n=26)
F
p = 0.032
Cumulative kataegis (Mb)
APOBEC (n=10)
Non-APOBEC (n=26)

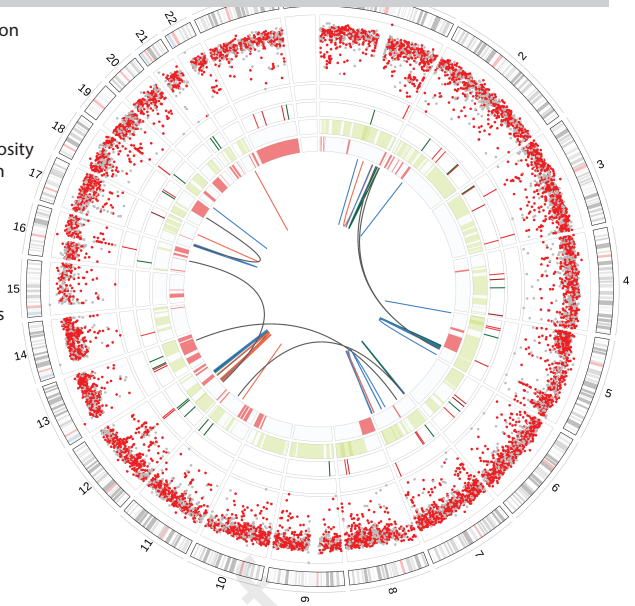
Figure 3

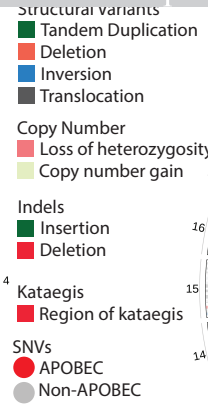
A

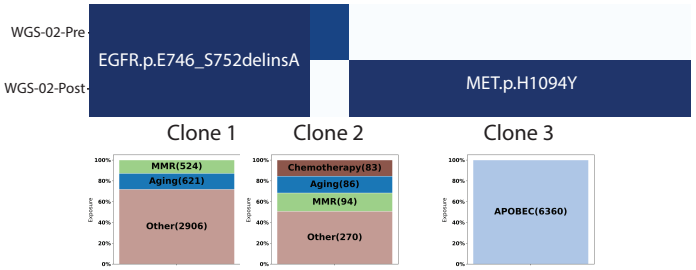
WGS-02-Pre

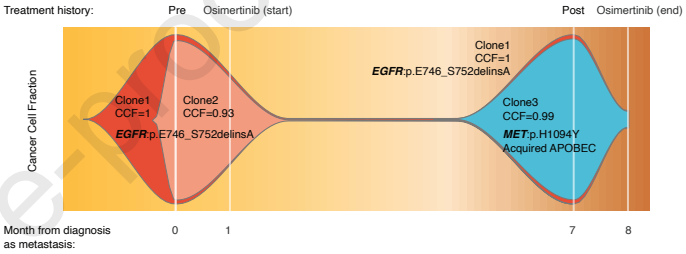
WGS-02-Post







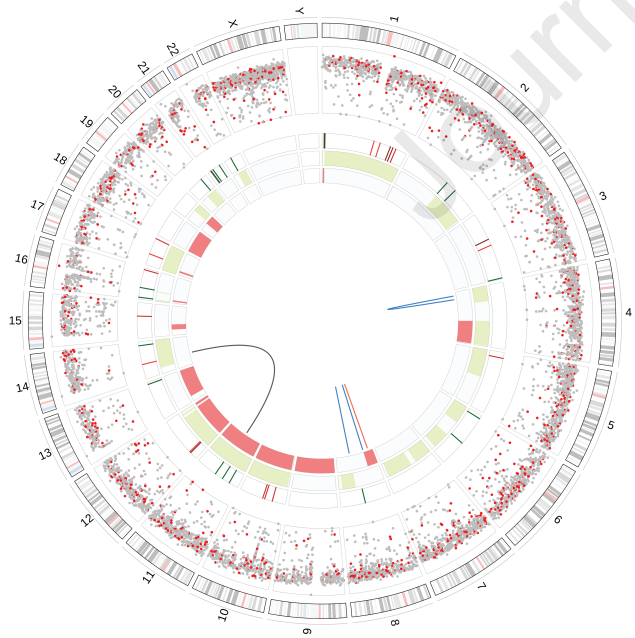


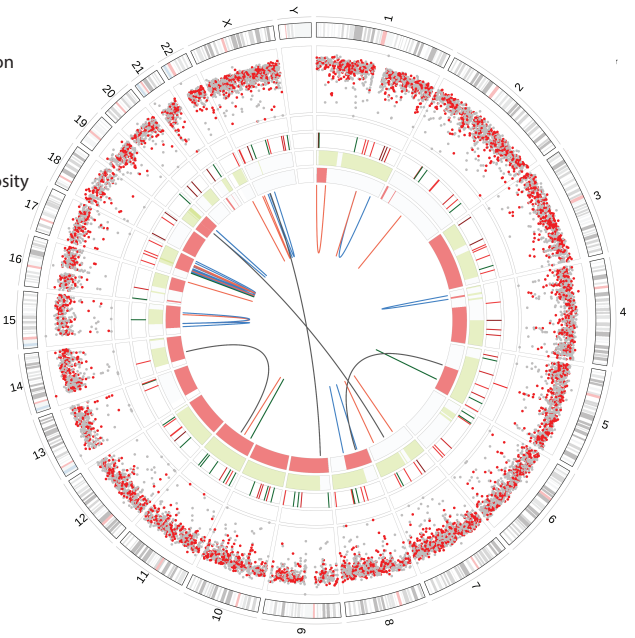


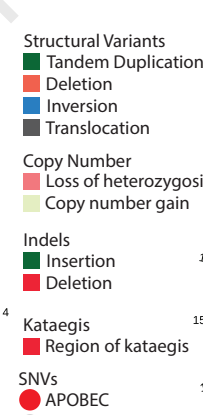
B

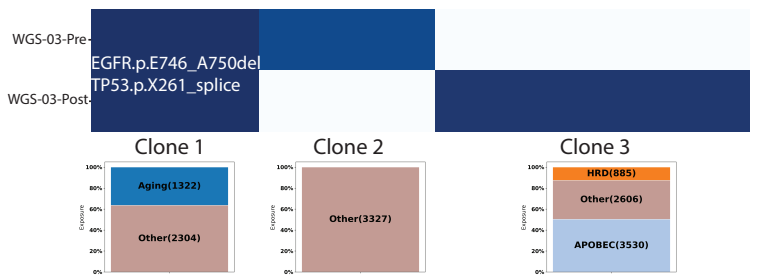
WGS-03-Pre

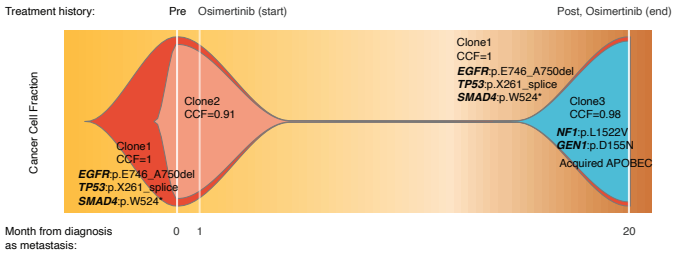
WGS-03-Post



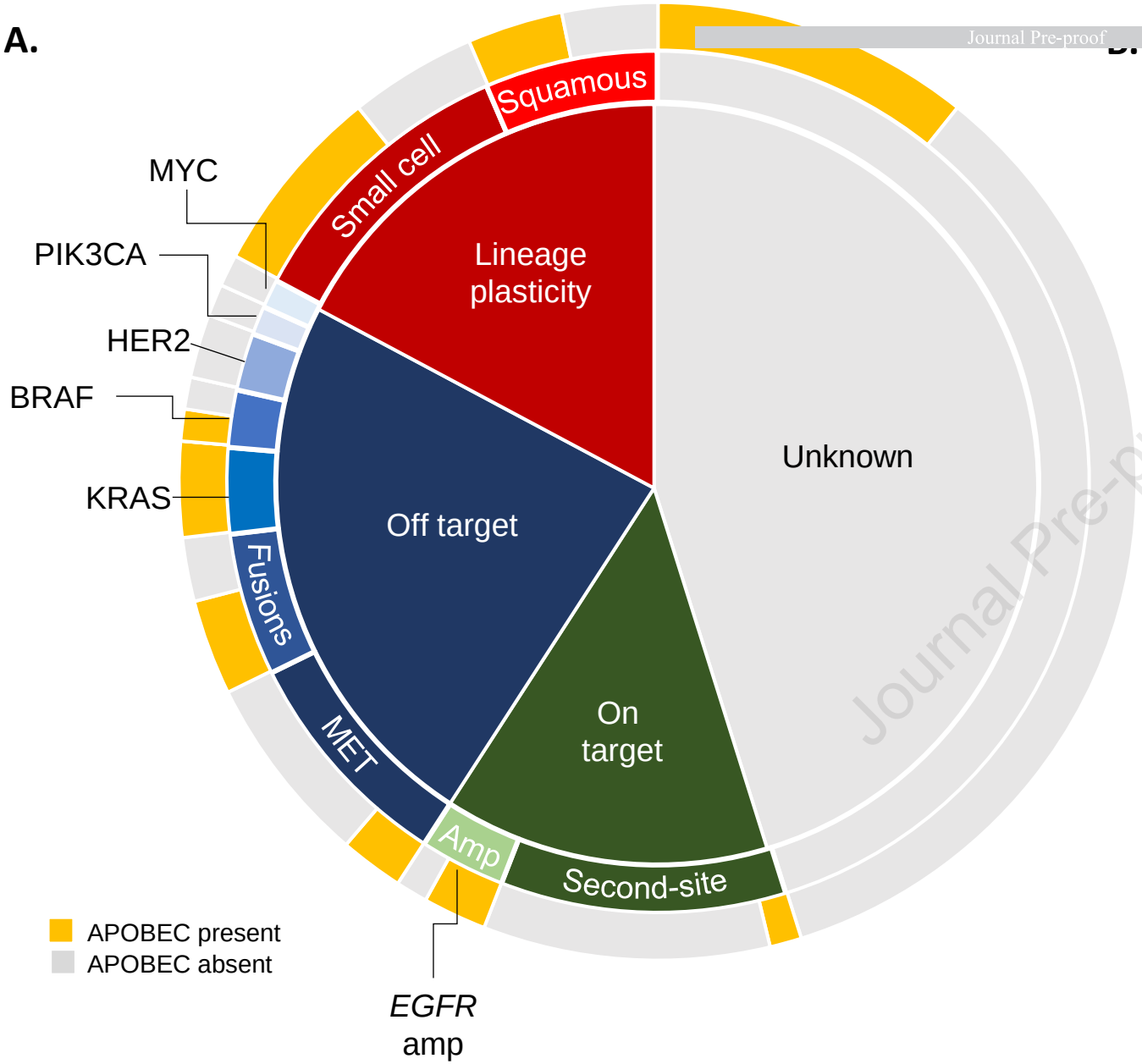








A.



Paired Pre/Post Osimertinib IMPACT Samples			
	APOBEC Signature	Dominant	Non-Dominant
Mechanism of Resistance	**Cohort size**	**31**	**62**
On Target	*EGFR* amp	2	1
	EGFR C797X	1	7
	EGFR G724S	0	1
	EGFR L718V	0	1
	Total	3	10
Off Target	*BRAF V600E*	1	1
	ERBB2 amp	0	2
	KRAS G12X	3	0
	MET amp	2	3
	MET/EGFR amp	0	1
	MET/ERBB2 amp	0	1
	MET exon 14 skipping	0	1
	MYC amp	0	1
	PIK3CA H1047R	0	1
	EML4-ALK fusion	2	0
	TRIM24-BRAF fusion	1	0
	ZC3HAV1-BRAF fusion	0	1
	RUFY2-RET fusion	0	1
	Total	9	13
Lineage Plasticity	Small Cell	6	4
	Squamous Cell	3	3
	Total	9	7
Unknown	Unknown	10	32