Genome Biology

RESEARCH

Open Access

A comprehensive functional atlas of ALK kinase domain variants reveals resistance landscape to ALK inhibitors

Hyeong-Cheol Oh[1,2†], Yeonseung Han[1†], Yoojin Chang[1,3†] and Hyongbum Henry Kim[1,3,4,5,6,7,8,9,10*]

†Hyeong-Cheol Oh, Yeonseung Han and Yoojin Chang contributed equally to this work.

*Correspondence: hkim1@yuhs.ac

[1] Department of Pharmacology, Yonsei University College of Medicine, Seoul, Republic of Korea
Full list of author information is available at the end of the article

Abstract

Background: *ALK* gene fusions are key oncogenic drivers in cancers such as non-small cell lung cancer, where they define a molecular subtype responsive to ALK tyrosine kinase inhibitors (TKIs). However, resistance commonly arises due to single nucleotide variants (SNVs) within the ALK tyrosine kinase domain, many of which remain variants of uncertain significance (VUSs).

Results: To systematically profile resistance, we use prime editing to generate and assess 3,208 *ALK* variants covering 99% of all possible SNVs across exons 20–28, along with intronic variants. We evaluate drug resistance across three generations of ALK TKIs: alectinib, lorlatinib, and zotizalkib. These high-resolution resistance landscapes validate known resistance mutations (e.g., G1202R, L1196M), identify previously uncharacterized resistance-associated VUSs, and reveal distinct patterns of drug-specific and shared resistance across inhibitors. Structural mapping further contextualizes resistance-associated variants relative to the ATP-binding pocket and distal regions associated with resistance.

Conclusions: This study provides a comprehensive functional atlas of ALK tyrosine kinase domain variants under TKI selection, offering a valuable experimental framework for interpreting resistance-associated variants. Although derived from in vitro models and therefore context dependent, this resource complements existing clinical and genomic knowledge and may aid in the functional interpretation of *ALK* variants observed in ALK-driven cancers.

Keywords: ALK fusion variants, Prime editing screening, Variants of uncertain significance

Background

ALK gene fusions play a critical role in the pathogenesis of various cancers [1, 2], including non-small cell lung cancer (NSCLC) [3], anaplastic large cell lymphoma (ALCL) [4, 5], inflammatory myofibroblastic tumor (IMT) [6], breast cancer [7], colorectal cancer [7], and glioma [8, 9]. In particular, *ALK* fusions define distinct molecular

subtypes of NSCLC, accounting for approximately 4–6% of lung adenocarcinomas [3, 10]. These fusions typically result from chromosomal rearrangements involving the region encoding the ALK tyrosine kinase domain (TKD), leading to its constitutive activation and the dysregulation of downstream signaling pathways [3, 11, 12].

ALK-positive lung cancers are generally sensitive to ALK tyrosine kinase inhibitors (TKIs) that target the ALK TKD within the fusion protein [2, 13–18]. However, nearly all patients with advanced ALK-positive NSCLC eventually develop resistance, resulting in disease progression [19]. This resistance often arises from acquired *ALK* mutations [2, 20–24] or activation of bypass signaling pathways [2, 25, 26]. Such mutations, typically caused by single nucleotide variants (SNVs), can alter drug binding affinity and reduce therapeutic efficacy [20–24]. To overcome this challenge, next-generation TKIs have been developed—most notably, the third-generation inhibitor lorlatinib, which addresses resistance to second-generation agents like alectinib [2, 15]. Nevertheless, resistant variants continue to emerge, highlighting the need for a comprehensive understanding of resistance profiles, which would lay foundation for the development of more effective therapies [2, 27–29].

Although several TKI-resistant *ALK* mutations, such as G1202R and L1196M, are well characterized [2, 30], the majority of *ALK* mutations remain variants of uncertain significance regarding their impact on ALK TKI resistance. Currently, only G1202R and L1196M are included in the National Comprehensive Cancer Network (NCCN) guidelines for TKI selection [30]. Therefore, there is a critical need for a comprehensive profile of resistance-conferring *ALK* mutations. This profile should include not only mutations affecting established TKIs such as alectinib and lorlatinib, but also investigational agents like zotizalkib—a fourth-generation TKI that has been evaluated in clinical trials [31]. Establishing such a resource could significantly enhance clinical decision-making, enabling more precise TKI selection and resistance management strategies, and ultimately supporting precision therapy in *ALK* fusion-driven cancers.

In this study, we employed prime editing to systematically assess how *ALK* fusion variants across the entire TKD, a hotspot for acquired mutations following ALK TKI treatment [2], affect drug resistance. We experimentally generated and analyzed a total of 2,953 *ALK* variants, covering 99% (2953 out of 2976) of all possible *ALK* SNVs within the TKD sequence (exons 20–28), and 255 SNVs in intronic regions near exon–intron junctions. We established comprehensive resistance profiles of these variants against alectinib, lorlatinib, and zotizalkib—representing second-, third-, and fourth-generation TKIs, respectively. Our results are consistent with previous clinical findings and results from cell-based assays. Additionally, our findings provide structural insights into drug-resistance mechanisms. We anticipate that these resistance profiles for three generations of NSCLC therapies targeting *ALK* fusions will enhance precision oncology approaches and improve prognostic predictions for individual patients. Moreover, this dataset could be extended to other *ALK* fusion-driven cancers, guiding therapeutic strategies beyond NSCLC.

Results

Optimization of prime editing in an ALK-positive lung cancer cell line

The *EML4-ALK* fusion is the most prevalent *ALK* fusion variant in NSCLC, accounting for approximately 85% of ALK-positive cases [1, 3, 32]. This structural rearrangement

fuses the 5′ end of *EML4* with the 3′ end of *ALK*, which encodes the ALK TKD, thereby activating several oncogenic pathways through the HELP domain of EML4 (Fig. 1A, B) [3, 11, 12, 32]. To comprehensively profile the effects of *ALK* variants on the response to ALK TKIs, we utilized the H3122 cell line, which harbors variant 1—the most common *EML4-ALK* fusion variant (Additional file 1: Fig. S1A).

Given that mismatch repair (MMR) deficiency enhances prime editing efficiency [33] and H3122 cells exhibit functional MMR activity (Additional file 1: Fig. S1B), we generated an *MLH1*-knockout H3122 line (H3122-MLH1KO) using SpCas9 and two single guide RNAs (sgRNAs) targeting *MLH1*—a standard strategy to improve prime editing performance [33, 34]. Deep amplicon sequencing confirmed near-complete disruption of the *MLH1* locus (>99.9% indel frequency; PRJNA1272298). Single-cell clones

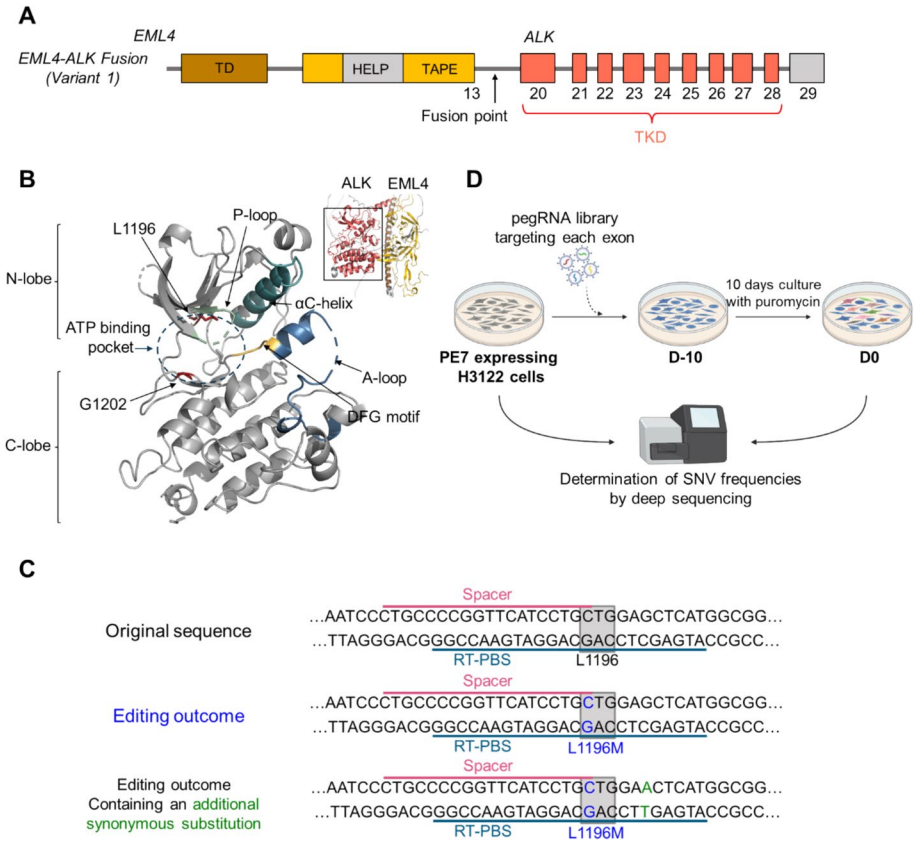

Fig. 1 Cell line generation for functional evaluation of *ALK* variants. **A** Schematic representation of the *EML4–ALK* fusion variant 1 (E13; A20), highlighting key domain-encoding regions: trimerization domain (TD), hydrophobic motif in EML proteins (HELP), tandem atypical propeller domain (TAPE), and tyrosine kinase domain (TKD). The *ALK* TKD spans exons 20–28. **B** Predicted three-dimensional structures of the ALK TKD (left) and full EML4–ALK fusion protein (right) generated using AlphaFold3. The N-lobe (residues 1093–1199) contains a prominent αC-helix (cyan), whereas the C-lobe (residues 1200–1399) includes the glycine-rich P-loop (green), the activation loop (A-loop, blue), and the DFG motif (yellow). The ATP-binding pocket—where representative resistance mutations such as G1202R and L1196M are located—is circled. The ALK domain within the fusion protein is boxed. **C** Schematic of pegRNA design and editing outcomes for the ALK L1196M variant at the endogenous locus. The intended substitution generating L1196M is shown in blue, and an additional synonymous marker substitution is shown in green. Spacer and RT–PBS elements are indicated. **D** Schematic overview of prime editing-based generation of single nucleotide variants (SNVs) in *ALK*

with confirmed *MLH1* knockout were isolated by limiting dilution into 96-well plates and subsequently selected for stable PE expression. Recent studies suggest that PE7 outperforms PEmax in editing efficiency [35]; therefore, we transduced H3122-MLH1KO cells with lentiviral constructs expressing either PEmax [33] or PE7 [35] (Additional file 1: Fig. S1C). To benchmark editing performance, we designed a prime editing guide RNA (pegRNA) library encompassing all possible SNVs within exon 23 and its flanking intronic regions (±5 bp), totaling 1,287 pegRNAs (429 SNVs × 3 pegRNAs per SNV). Deep sequencing of the targeted locus showed that PE7-expressing cells exhibited a 2.1-fold increase in reads per million (RPM) compared to PEmax-expressing cells (77 vs 36; Additional file 1: Fig. S1D). Based on these results, we selected the PE7-expressing H3122-MLH1KO line (hereafter referred to as H3122-PE7) for subsequent experiments.

Prime editing enables saturating SNV generation in the *ALK* TKD sequence

Acquired resistance mutations to ALK TKIs frequently arise within the TKD sequence (exons 20–28), particularly in exons 22 and 23—the core drug-binding domain (Fig. 1B) [2, 36]. Well-known resistance-associated mutations include the gatekeeper mutation L1196M and the solvent-front mutation G1202R, both encoded in exon 23 [2, 30]. To generate all possible SNVs across the *ALK* TKD sequence (exons 20 to 28), we designed a total of 9,531 pegRNAs (= 992 bp × 3 substitutions/bp × 2 ~ 3 pegRNAs (1 ~ 2 regular pegRNAs + 1 engineered pegRNA (epegRNA) [37])/substitution) using DeepPrime-FT [38]—a deep learning-based pegRNA efficiency predictor.

We constructed nine exon-specific pegRNA libraries (E20 to E28) (Methods, Additional file 2: Table S1) and introduced them individually into H3122-PE7 cells. A representative pegRNA design and the corresponding editing outcomes are illustrated in Fig. 1C. Following 10 days of culture to allow sufficient editing (Fig. 1D), we performed deep sequencing to quantify variant allele frequencies, comparing edited samples to unedited controls. After filtering out low-confidence variants (odds ratio (OR) ≤ 3 or adjusted $P \geq 0.05$ in any replicate; Methods) [34], we successfully identified 3,208 significant SNVs, including 99.3% (2,031) of all 2,045 single amino acid variants encoded by the SNVs.

Complete resistance profiling of *ALK* variants against three generations of TKIs

To comprehensively evaluate the resistance profiles of *ALK* variants across three generations of ALK TKIs, prime-edited H3122-PE7 cells were first cultured for 10 days after pegRNA library transduction to establish edited variant populations, followed by an additional 10 days under four experimental conditions: (i) untreated control, and treatment with (ii) alectinib (second-generation), (iii) lorlatinib (third-generation), and (iv) zotizalkib (fourth-generation) (Fig. 2A). To quantify the drug resistance conferred by each SNV, we calculated $\log_2$ fold changes (LFCs) between drug-treated and untreated conditions. These LFCs were standardized using the distribution of LFCs for synonymous SNVs (Methods). Standardized LFCs (sLFCs) for internal replicates of each SNV were frequency-based weighted and averaged to generate a resistance score per replicate. Final resistance scores were obtained by averaging across biological replicates, providing a representative "resistance score" for each variant.

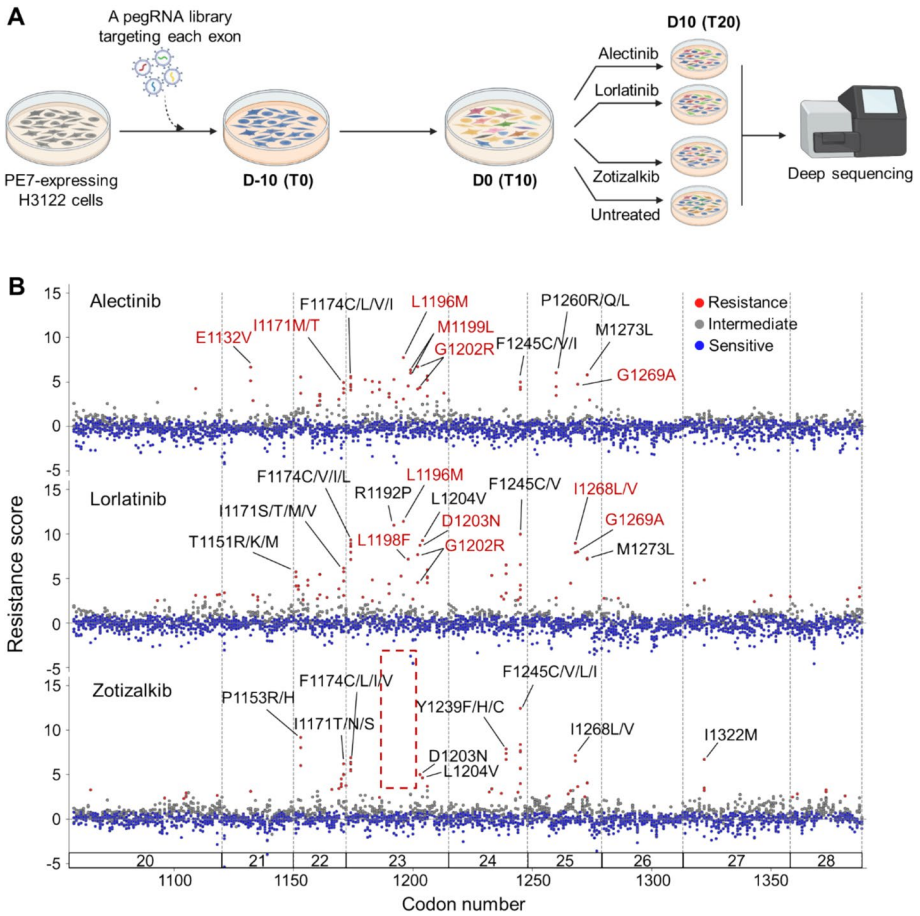

Fig. 2 Landscape of ALK TKI resistance conferred by SNVs in the *ALK* TKD sequence. **A** Schematic overview of the drug-resistance profiling pipeline using pooled pegRNA libraries to introduce SNVs across the *ALK* TKD sequence. Prime editing was initiated by pegRNA library transduction at D-10 (T0), followed by a 10-day editing period to establish variant populations (D0, T10), after which cells were subjected to drug treatment or control conditions for an additional 10 days and harvested at D10 (T20). One pegRNA library was constructed per exon, resulting in nine libraries targeting exons 20–28. **B** Heatmap of resistance scores for SNVs across exons 20–28, stratified by TKI: alectinib (top), lorlatinib (middle), and zotizalkib (bottom). Vertical dashed lines denote exon boundaries. Known and newly identified resistant variants are labeled. Based on three-dimensional structures of ALK bound to alectinib or lorlatinib (PDB: 3AOX and 4CLJ, respectively), mutations located within 6 Å of the drug-binding site are highlighted in red. A boxed region highlights the absence of canonical drug-resistant variants, such as L1196M and G1202R, in cells treated with zotizalkib

Biological replicates showed high concordance across drug conditions (Pearson correlation coefficients: alectinib, r = 0.67; lorlatinib, r = 0.72; zotizalkib, r = 0.68; Additional file 1: Fig. S2A). We then used the mean of SNV resistance scores to calculate the resistance score of each protein variant. These protein-level scores were also highly reproducible across replicates (Pearson r = 0.72, 0.78, and 0.72 for alectinib, lorlatinib, and zotizalkib, respectively; Additional file 1: Fig. S2B). Among the 396 amino acid variants assessed including synonymous mutations, 209, 157, and 30 were encoded by two, three, and four distinct SNVs, respectively. Notably, SNVs encoding the same amino acid substitution yielded highly correlated resistance scores (Pearson r = 0.60, 0.63, and 0.65, respectively; Additional file 1: Fig. S2C), with even stronger concordance observed when the analysis was restricted to missense variants

(Additional file 1: Fig. S2D), confirming the robustness and consistency of our high-throughput resistance assay.

Based on the distribution of resistance scores for synonymous SNVs, we categorized SNVs into three groups: resistant (resistance scores exceeding the 99.7th percentile of the scores of synonymous SNVs in both replicates), sensitive (resistance scores below the 95th percentile in both replicates), and intermediate (SNVs not meeting either threshold; Methods). For alectinib, we identified 2712, 444, and 52 SNVs encoding 1703, 316, and 45 single amino acid variants (SAAVs) as sensitive, intermediate, and resistant, respectively (Fig. 2B and Additional file 2: Table S2). For lorlatinib, 2540, 585, and 83 SNVs encoding 1566, 435, and 75 SAAVs were classified as sensitive, intermediate, and resistant, respectively. For zotizalkib, 2276, 869, and 63 SNVs encoding 1348, 692, and 56 SAAVs were classified as sensitive, intermediate, and resistant, respectively. We provide heatmaps for the resistance scores in Additional file 1: Figs. S3–S8.

Comprehensive profiling provides structural context for *ALK* TKI resistance

To elucidate the structural mechanisms underlying resistance to three generations of ALK TKIs, we examined the 3D spatial distribution of resistant variants in relation to the drug-binding pocket (Fig. 3A, B). The three inhibitors—alectinib, lorlatinib, and zotizalkib—all target the ATP-binding site within the ALK kinase domain, albeit with differences in molecular size and central nervous system penetrance, enabling inhibition of ALK-mediated signaling pathways [2, 31, 36].

We observed that drug resistant variants were distributed in two distinct patterns, with localization either adjacent to, or apart from, the ATP-binding pocket (Fig. 3A, B and Additional file 1: Fig. S9A). Examples of variants in or near the pocket include L1196M, I1171T/M, and G1269A, all residing directly in the ATP-binding pocket, as well as the solvent-front mutations G1202R and D1203N. These variants conferred resistance to alectinib and lorlatinib, consistent with previous reports indicating reduced drug binding affinity associated with these mutations [2, 39–41]. Notably, we found that zotizalkib—a fourth-generation, compact macrocyclic TKI engineered to fit precisely into the ATP-binding pocket [31]—effectively inhibits ALK fusion proteins harboring L1196M or G1202R mutations (Additional file 1: Fig. S9A and Fig. 2B), consistent with prior in vitro data [31].

Strikingly, our high-throughput mutational screening also identified distal sites—particularly residues F1245 and F1174—as recurrent resistance hotspots (Fig. 3A-C and Additional file 1: Fig. S9A). Notably, these residues are spatially separated from the ATP-binding pocket, suggesting that resistance conferred by mutations at these sites is unlikely to arise from direct disruption of drug binding. Previous computational and structural studies have proposed that these residues form an aromatic–aromatic network, and mutations such as F1174C may destabilize this interaction [42, 43]. In this study, structural modeling and prior simulations were used to provide a conceptual framework for interpreting these observations, rather than direct evidence of altered binding kinetics or conformational changes Accordingly, although the functional data demonstrate that these distal variants robustly confer drug resistance, the precise biophysical mechanisms underlying these effects remain to be elucidated.

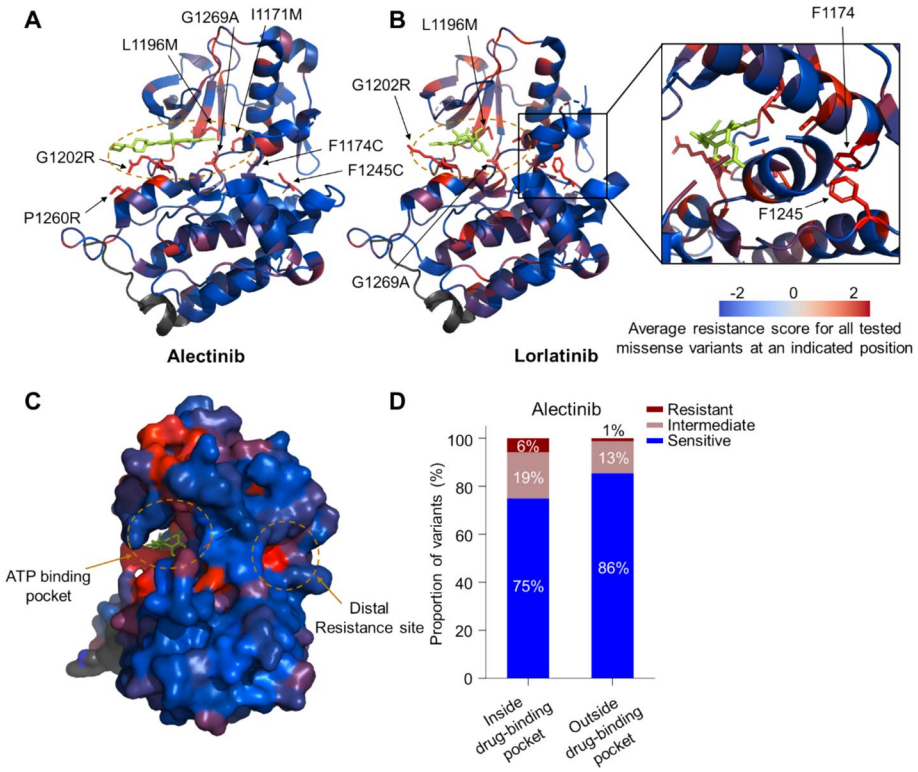

Fig. 3 Spatial distribution of drug-resistant *ALK* variants across the kinase domain. **A** Three-dimensional structure of the ALK TKD in complex with alectinib (PDB: 3AOX). Residue-wise average resistance scores to alectinib are mapped onto the structure using a blue-to-red color gradient (blue: sensitive; red: resistant). The alectinib-binding domain is circled, drug is shown in green and representative resistant variants are labeled. **B** Crystal structure of the ALK TKD in complex with lorlatinib (PDB: 4CLJ). Residue-wise average resistance scores to lorlatinib are overlaid using the same color gradient. The lorlatinib-binding domain is circled. A magnified boxed region highlights drug-resistant variants that form aromatic ring interactions, located distal to the drug-binding site. **C** Three-dimensional surface representation of the ALK TKD bound to lorlatinib. The ATP-binding pocket and distal resistance site are circled. Resistance scores to lorlatinib are shown as a residue-wise color gradient. **D** Stacked bar plots showing the proportion of resistant (red), intermediate (rosy-brown), and sensitive (blue) variants located within or outside the alectinib-binding pocket. Percentages represent the fraction of variants belonging to each functional class

To quantify our observations about mutation distributions, we compared the relative frequency of resistant variants within the drug-binding pocket versus other regions. Among all variants localized to the binding pocket, 6% (15/255) caused resistance to alectinib and 4% (9/229) to lorlatinib; among all variants localized outside the pocket, only 1% (38/2,953) caused resistance to alectinib and 2.5% (75/2,979) to lorlatinib (Fig. 3D and Additional file 1: Fig. S9B). This enrichment analysis underscores that the ATP-binding pocket is a disproportionately frequent site of resistance-conferring variants. Collectively, our structural and functional mapping provides novel insights into the spatial diversity of resistance mechanisms against ALK TKIs, highlighting contributions from both variants located within the ATP-binding pocket and those arising at distal resistance sites.

Comparison of resistance profiles of *ALK* variants with classifications from previous reports

We next compared our high-throughput resistance profiling data with existing clinical and experimental reports. According to current NCCN guidelines for NSCLC management, only two *ALK* mutations—G1202R and L1196M—are recognized as actionable alterations that justify switching from alectinib to lorlatinib [30]. However, our functional analysis classified both G1202R and L1196M as resistant to both alectinib and lorlatinib (Fig. 4A, B), which differs from the clinical interpretation that lorlatinib retains relative activity against these mutations. Notably, this clinical interpretation is primarily based on comparative efficacy across ALK inhibitors rather than absolute equivalence to wild-type sensitivity.

To rule out concentration-dependent effects as a source of this discrepancy, we performed dose–response assays for lorlatinib at three concentrations: 2 nM (below the screening concentration), 5 nM (screening concentration), and 10 nM (above). Resistance scores across these conditions were highly correlated (Fig. 4C, Additional file 2: Table S3), and the resistant phenotype conferred by both G1202R and L1196M remained consistent, confirming that the observed resistance was not due to subtherapeutic exposure.

Moreover, to exclude the possibility that the observed resistance patterns resulted from low-level drug selection or insufficient selective pressure, we performed additional validation experiments using a broad range of drug concentrations. Cell viability was assessed across multiple concentrations for all three ALK inhibitors in 150-mm culture dishes, matching the format and scale of the high-throughput screening platform. In addition to the concentrations used in the primary screen, we included higher drug doses, spanning from approximately half-maximal inhibitory concentrations (IC50) to the 20% inhibitory concentration (IC20) levels (Additional file 1: Fig. S10A and Additional file 2: Table S4).

Using these conditions, we repeated the resistance screening specifically for exon 23—the most clinically relevant exon within the *ALK* TKD sequence—following the same analytical pipeline as in the original high-throughput assay. Resistance scores obtained under these higher and more cytotoxic drug concentrations showed strong correlations with those derived at other tested concentrations across all three inhibitors (Additional file 1: Fig. S10B–D and Additional file 2: Table S5). Importantly, this concordance was maintained even under highly lethal drug exposures, supporting that the resistance phenotypes observed in our study are robust and not attributable to insufficient drug selection pressure. Together, these results further validate the reliability of our high-throughput resistance screening platform.

A literature review [21, 44, 45], including studies cited in the NCCN guidelines [30], further supports our findings; both mutations have been shown to increase the IC50 of lorlatinib by at least 10- to 20-fold relative to that for wild-type *ALK* [29]. Importantly, these observations highlight that our resistance classifications reflect relative resistance under defined in vitro selective pressures, rather than a binary designation of clinical insensitivity.

Notably, zotizalkib, a fourth-generation ALK inhibitor designed to target resistance-associated mutations near the drug-binding pocket via a compact macrocyclic scaffold [31], retained activity against cells harboring G1202R or L1196M (Fig. 4D). This

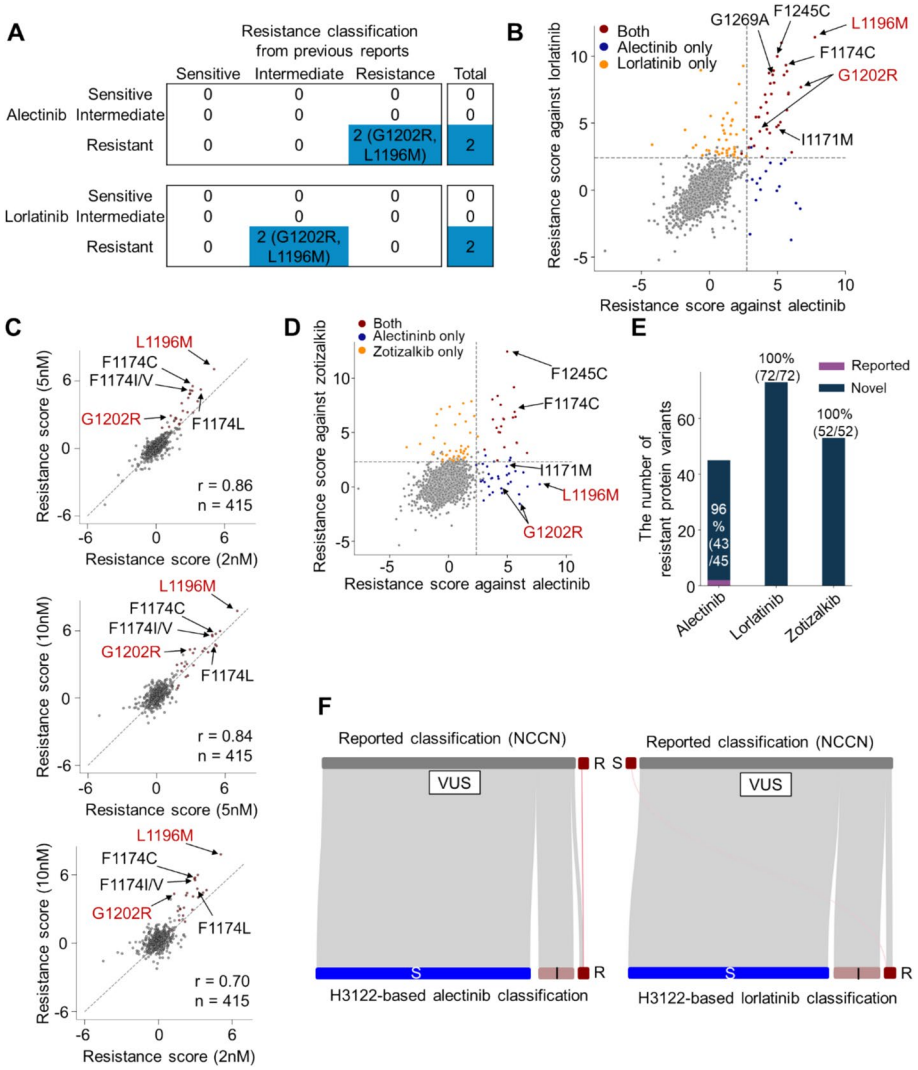

Fig. 4 Comparison of resistance profiles of *ALK* variants with classifications from previous reports. **A** Heatmaps comparing resistance classifications for alectinib (top) and lorlatinib (bottom) between this study (in which mutations are classified as sensitive, intermediate, or resistant) and annotations from prior reports, including the National Comprehensive Cancer Network (NCCN) guidelines (sensitive, intermediate, resistant; intermediate indicates cases where cell-based experimental evidence and clinical guidelines appear to be inconsistent). **B** Scatter plot comparing resistance scores for *ALK* SNVs against alectinib (x-axis) and lorlatinib (y-axis). Variants resistant to both drugs are shown in dark red, alectinib-only resistant variants in blue, and lorlatinib-only resistant variants in orange. Canonical resistance mutations L1196M and G1202R are labeled in red. **C** Scatter plots comparing resistance scores for *ALK* SNVs treated with different concentrations of lorlatinib: 2 nM vs. 5 nM (top), 5 nM vs. 10 nM (middle), and 2 nM vs. 10 nM (bottom). Canonical resistance mutations are labeled in red. **D** Scatter plot comparing resistance scores for *ALK* SNVs against alectinib (x-axis) and zotizalkib (y-axis). Variants resistant to both drugs are shown in dark red, alectinib-only in blue, and zotizalkib-only in orange. Canonical resistance mutations are labeled in red. **E** Bar plot showing the number of ALK protein variants classified as resistant in this study. Variants not annotated as resistant in the NCCN guidelines are highlighted in dark blue and considered novel. **F** Sankey plots showing concordance between resistance classifications in this study and previously reported annotations for alectinib (left) and lorlatinib (right). Variants lacking prior classification were designated as variants of uncertain significance (VUSs). R; resistant, I; intermediate, S; sensitive

result is consistent with recent studies [31] and highlights zotizalkib's potential clinical utility, particularly for patients who develop resistance to lorlatinib. Given the persistent emergence of G1202R as a resistance mutation following lorlatinib therapy [46], our data suggest zotizalkib as a promising therapeutic alternative.

To further validate our resistance classifications, we compared our dataset to the COSMIC (Catalogue Of Somatic Mutations In Cancer) resistance database. Among COSMIC-listed alectinib-resistant variants, 16 of 20 (80%) were classified as resistant in our assay, and the remaining 4 variants (20%) were classified as intermediate; for lorlatinib, the concordance was 100% (16/16), with all COSMIC-listed resistant variants were classified as resistant in our assay (Additional file 1: Fig. S10E and F), underscoring the reliability and clinical relevance of our profiling.

Across our full dataset, we identified 45 alectinib-resistant, 72 lorlatinib-resistant, and 52 zotizalkib-resistant SAAVs, of which 43 (96%), 72 (100%), and 52 (100%), respectively, were previously unreported (Fig. 4E, F and Additional file 1: Fig S10G-H). Notably, 17 (38%), 26 (36%), and 20 (38%) of alectinib-, lorlatinib-, and zotizalkib- resistant variants, respectively, retained sensitivity to other generations of drugs—suggesting that resistance to one TKI may be overcome by switching to another (Additional file 2: Table S2; see Additional file 1: Fig. S10G for direct lorlatinib–zotizalkib comparisons). However, 13 SAAVs conferred resistance to all three inhibitors, underscoring the need for next-generation therapies beyond current ALK TKIs.

Distinct classes of drug-resistant *ALK* variants defined by cell fitness

We investigated whether *ALK* variants could be stratified into distinct functional subtypes by integrating drug resistance and cellular fitness. Prior studies have shown that some oncogenic variants impose intrinsic fitness costs in the absence of inhibitors, consistent with excessive pathway activation and oncogene-induced senescence, giving rise to so-called drug addiction–like resistance states [47–49]. In contrast, other resistance mutations either maintain neutral fitness or confer a proliferative advantage [47].

To quantify fitness scores, *ALK* variants were generated by prime editing, and variant frequencies were compared between two untreated time points: 10 days (T = 10) and 20 days (T = 20) after transduction. LFCs in variant abundance between T = 10 and T = 20 were calculated, standardized using synonymous variants, and weighted by variant frequency to generate a fitness score for each biological replicate, followed by averaging across replicates, using the same analytical framework applied for resistance score calculation (Methods).

Integrating fitness and resistance scores across all three ALK inhibitors revealed three major potential classes of drug-resistant variants: (i) canonical resistance variants, which confer resistance without substantially affecting baseline fitness; (ii) driver-like variants, which confer both drug resistance and a proliferative advantage; and (iii) variants causing a drug addiction–like phenotype, in which cells exhibit resistance but display reduced fitness in the absence of drug treatment (Additional file 1: Fig. S11A–C). This classification framework is consistent with previously described resistance–fitness trade-off models. For example, L1196M, F1174V, and F1245C displayed canonical resistance profiles, whereas G1202R (c.3604G > C, c.3604G > A) showed a drug addiction–like phenotype, characterized by decreased baseline fitness despite robust resistance upon

drug exposure. These findings suggest that fitness trade-offs may shape the evolutionary dynamics of resistant variants, although their therapeutic implications require further investigation.

Individual evaluations of drug resistance in strains expressing *ALK* variants

To validate the resistance phenotypes identified in our high-throughput screen, we performed individual functional assays on eight strains expressing *ALK* variants—P1153R, I1171M, F1174C, F1193L, L1196M, M1199L, G1202R, and F1245C—against three generations of ALK inhibitors. These variants were chosen based on distinct drug response profiles: F1145C, F1174C, and P1153R conferred resistance to all three drugs; G1202R, L1196M, and I1171M were resistant to alectinib and lorlatinib but sensitive to zotizalkib; M1199L showed resistance to alectinib but remained sensitive to lorlatinib and zotizalkib; and F1193L remained sensitive to all drugs tested.

We transduced H3122 cells expressing PE7 with individual lentiviruses encoding each variant-specific pegRNA (Additional file 2: Table S6). Prime-edited cells were mixed with control cells at a 25:75 ratio and cultured for 10 days under either untreated conditions or in the presence of alectinib, lorlatinib, or zotizalkib. Variant frequencies were quantified to assess resistance phenotypes (Fig. 5A).

The frequencies of P1153R-, F1174C-, and F1245C-containing cells increased by 2.6-fold (Day 10 vs. untreated; 65% vs. 25%), 1.6-fold (8.3% vs. 5.4%), and 2.1-fold (51% vs. 24%), respectively, under alectinib treatment; 1.3-fold, 2.7-fold, and 3.0-fold under lorlatinib; and 3.5-fold, 2.7-fold, and 3.7-fold under zotizalkib, confirming their resistance to all three inhibitors. Similarly, frequencies of G1202R-, L1196M-, and I1171M-containing cells increased by 4.7-fold, 3.2-fold, and 2.8-fold under alectinib, and 4.3-fold, 3.2-fold, and 1.9-fold under lorlatinib, but decreased by 0.69-fold, 0.82-fold, and 0.94-fold, respectively, under zotizalkib, supporting its efficacy against these variants (Fig. 5B).

For M1199L, variant-containing cell frequencies increased by 2.2-fold (27% vs. 12%) under alectinib but decreased by 0.39-fold (4.7% vs. 12%) under lorlatinib and 0.47-fold (5.6% vs. 12%) under zotizalkib, indicating retained sensitivity to the latter two agents. By contrast, F1193L-containing cells were consistently depleted across all treatments—0.46-fold (7.6% vs. 17%) with alectinib, 0.75-fold with lorlatinib, and 0.51-fold with zotizalkib—confirming its susceptibility to all three inhibitors (Fig. 5B).

Notably, fold-changes observed in these individual assays showed strong correlation with resistance scores from the high-throughput drug-resistance assay, with Pearson's correlation coefficients of 0.66 for alectinib, 0.82 for lorlatinib, and 0.98 for zotizalkib (Fig. 5C), underscoring the accuracy and reproducibility of our high-throughput platform.

Individual evaluations of drug resistance in single-variant clones

To further validate resistance phenotypes identified in the pooled screen, we evaluated drug sensitivity using single-variant systems. First, H3122 cells harboring newly identified resistant variants (P1153R or F1245C), along with cells harboring well-characterized variants (G1202R or L1196M), were subjected to dose–response assays with three ALK TKIs to determine IC50s (Methods). Both P1153R- and F1245C-containing cells exhibited increased IC50 values relative to cells with wild-type *ALK* across all three inhibitors, whereas G1202R- and L1196M-containing cells

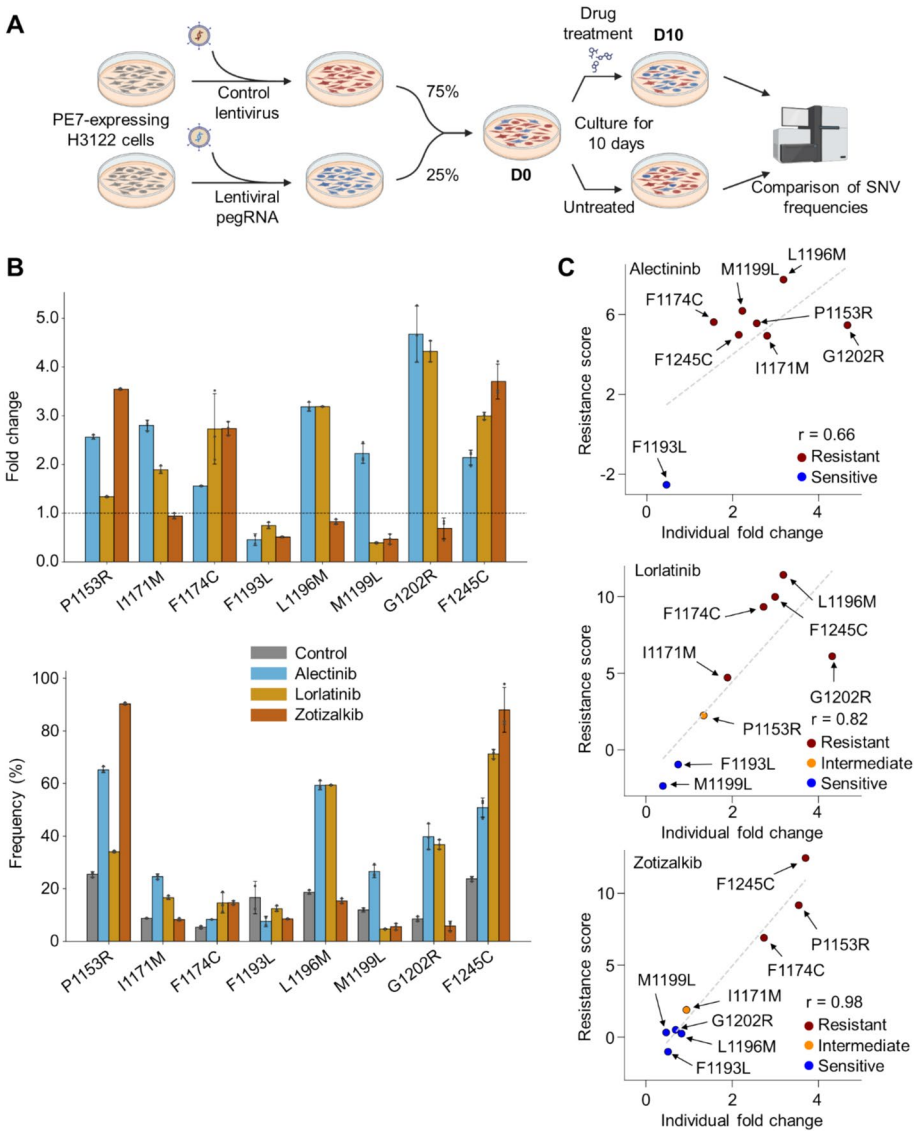

Fig. 5 Individual evaluations of the drug resistance of *ALK* variants. **A** Schematic overview of the strategy for individual validations of drug resistance. Lentiviral vectors encoding pegRNAs designed to introduce eight specific SNVs, along with a control vector, were individually transduced into PE7-expressing H3122 cells. Ten days post-transduction, pegRNA-transduced cells and control (empty vector) cells were mixed and split into untreated and drug-treated groups. SNV frequencies were quantified by targeted deep sequencing 10 days after treatment initiation. **B** Bar plots showing the fold change (top) and percentage (bottom) of sequencing reads harboring each variant after treatment with alectinib (blue), lorlatinib (yellow), zotizalkib (orange), or control solvent (gray) at Day 10. Values represent the mean of three independent biological replicates, with error bars indicating standard deviation. **C** Correlations between resistance scores obtained from high-throughput screening and fold-change values from individual validation assays for alectinib (top), lorlatinib (middle) and zotizalkib (bottom). Pearson correlation coefficients (r) are shown. The SNV classifications from the high-throughput evaluations are indicated

showed lower IC50 values for zotizalkib compared to wild type (Additional file 1: Fig. S12A and Additional file 2: Table S7), consistent with the resistance classification of these variants.

Across variants, IC50 values showed strong concordance with resistance scores derived from the high-throughput screen (Pearson's r = 0.94 for alectinib, 0.82 for lorlatinib, and 0.97 for zotizalkib; Additional file 1: Fig. S12B), confirming that pooled resistance scores accurately recapitulate drug sensitivity measured at the single-variant level.

We next assessed resistance using a conventional Ba/F3 transformation assay to provide an orthogonal validation [50]. Ba/F3 cells expressing *EML4–ALK* harboring a distal resistance variant (F1174C or F1245C) or a canonical resistance variant (L1196M or G1202R) were treated with increasing concentrations of each inhibitor to determine IC50 values. Consistent with results from the pooled screen, F1174C- and F1245C-expressing Ba/F3 cells showed elevated IC50 values across all three TKIs, whereas L1196M and G1202R conferred resistance to alectinib and lorlatinib but increased sensitivity to zotizalkib (Additional file 1: Fig. S12C and Additional file 2: Table S7).

Together, these results demonstrate that resistance profiles derived from prime editing–based high-throughput screening are reproducible across both single-variant cancer cell models and conventional cytokine-dependent transformation assays.

Evaluation of ALK TKIs in NCI-H2228 cells

To assess whether resistance scores defined in H3122 cells are reproducible in an independent ALK fusion–positive lung cancer model, we performed the same resistance profiling in NCI-H2228 cells, which harbor an *EML4–ALK* fusion with a distinct breakpoint (variant 3), in contrast to variant 1 in H3122 cells. We generated PE7-expressing H2228 cells (Methods), introduced all possible SNVs in *ALK* exons 22 and 23 by prime editing, and evaluated variant-specific responses under untreated conditions or treatment with three generations of ALK TKIs.

To exclude insufficient selective pressure as a potential confounder, we first assessed cell viability across a range of drug concentrations for all three ALK TKIs in 150-mm culture dishes and selected doses spanning approximately IC50 to IC20 levels (Additional file 1: Fig. S13A and Additional file 2: Table S8). Using these concentrations and the same experimental and analytical pipeline as in the H3122 screen, we calculated resistance scores for exon 22 and 23 variants in H2228 cells. However, resistance scores derived from H2228 cells showed poor correlation with those obtained in H3122 cells, even under higher drug concentrations (Additional file 1: Fig. S13B and Additional file 2: Table S9).

To investigate the basis for this discrepancy, we examined ALK dependency in H2228 cells using the prime editing–based fitness framework. Fitness scores were calculated by comparing variant frequencies between two untreated time points (T = 10 and T = 20), as described for H3122 cells. We then evaluated whether fitness scores could distinguish loss-of-function variants from intact *ALK* variants using receiver operating characteristic (ROC) analysis, assuming that nonsense SNVs disrupt ALK function and lead to cellular depletion in ALK-dependent contexts, whereas synonymous SNVs preserve ALK function.

In H3122 cells, fitness scores robustly discriminated nonsense from synonymous SNVs, yielding an area under the curve (AUC) of 0.82—comparable to previous functional studies of oncogene dependency in lung cancer, such as EGFR-focused analyses reporting AUC values around 0.85 [34] (Additional file 1: Fig. S13C). In contrast,

fitness scores in H2228 cells failed to distinguish loss-of-function from intact variants (AUC = 0.40), indicating weak or absent dependency on ALK despite the presence of an *EML4–ALK* fusion. Notably, similar observations regarding reduced ALK dependency in H2228 cells have been reported in prior studies [51–53].

Together, these results indicate that resistance classifications defined in an ALK-dependent context (such as H3122 cells) may not be transferable to settings in which tumor cell survival is driven by alternative signaling pathways. Thus, although *ALK* fusion status is necessary, it may not be sufficient to predict functional dependency or drug response, underscoring the importance of cellular context when applying resistance classifications.

Association between *ALK* mutation-defined classifications and clinical outcomes in ALK fusion–positive cancer

Given that our high-throughput screening delineates the impact of *ALK* variants on alectinib response—a drug commonly used in ALK fusion–positive lung cancer—we hypothesized that *ALK* mutation status may hold prognostic value. We analyzed patient survival in relation to somatic *ALK* mutations and their corresponding alectinib resistance classification derived from our screening, using sequencing data from MSK-CHORD (Clinicogenomic Harmonized Oncologic Real-world Dataset) [54]. Multiple variants characterized in our study were also detected in clinical specimens (Additional file 1: Fig. S14A).

Among *EML4–ALK* fusion–positive patients who underwent sequencing after alectinib treatment, those harboring resistant variants exhibited significantly worse overall survival compared to those with wild-type *ALK* ($P = 0.008$) (Fig. 6A). However, the number of patients with resistant variants was limited ($n = 7$), and this analysis should therefore be interpreted with caution. Extending the analysis to all *EML4–ALK* fusion–positive lung cancer cases with *ALK* mutations ($n = 15$), regardless of treatment history or sequencing timepoint, we again observed reduced survival in patients with resistant variants ($n = 13$) compared to those with wild-type *ALK* ($n = 186$; $P = 0.004$) (Fig. 6B). It is important to note that relatively few patients harbored resistant *ALK* variants and that survival analyses for sensitizing variants were not statistically feasible. Therefore, these results should be interpreted as exploratory associations rather than definitive prognostic evidence. Larger, prospectively annotated cohorts will be required to determine the clinical utility of functional resistance classifications in predicting patient outcomes.

Application of *ALK* variant classifications to other *ALK*-driven cancers

Beyond *EML4–ALK* fusion–positive lung cancer, *ALK* rearrangements occur in diverse malignancies [1], including ALCL [4, 5], IMT [6], breast cancer [7], colorectal cancer [7], and glioblastoma [8, 9]. Additionally, several germline *ALK* mutations have been implicated in neuroblastoma [55]. As these cancers typically retain the ALK kinase domain (coding begins in exon 20), we sought to examine whether resistance-associated variants identified in our H3122-based screen overlap with those previously reported in other *ALK*-driven disease contexts.

We used AlphaFold3 to model the protein structures of various *ALK* fusions and wild-type *ALK* (Fig. 6C) [56], confirming that all were structurally aligned with the

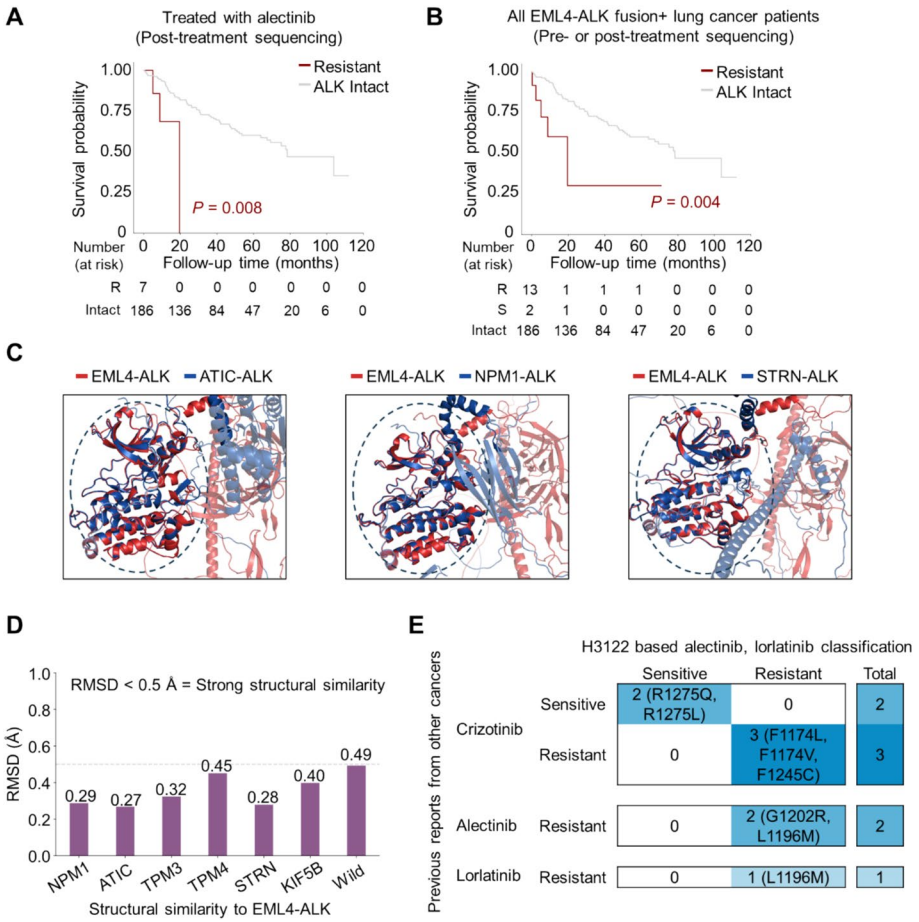

Fig. 6 Clinical application of *ALK* mutation classifications. **A** Kaplan–Meier survival analysis of *EML4–ALK* fusion-positive lung cancer patients who underwent sequencing after alectinib treatment. Patients were grouped based on somatic *ALK* variant status: resistant (n = 7) and wild-type *ALK* (*n* = 186). Groups with fewer than four patients were excluded. P-values comparing survival between the resistant group (dark red) and the wild-type group were calculated using the log-rank test. **B** Kaplan–Meier survival analysis of all *EML4–ALK* fusion-positive adenocarcinoma patients, irrespective of treatment history or sequencing time point. Patients were stratified based on the alectinib resistance classifications determined in this study. P-values comparing the resistant group (dark red; *n* = 13) and wild-type group (n = 186) were calculated using the log-rank test. S; Sensitive. **C** Structural alignment of *EML4–ALK* (red) with other *ALK* fusion proteins (blue), including ATIC–ALK (left), NPM1–ALK (middle), and STRN–ALK (right), with structures predicted using AlphaFold3 and aligned using PyMOL. **D** Root mean square deviation (RMSD) values from pairwise alignments between *EML4-ALK* and other *ALK* fusion variants. Lower RMSD values indicate higher structural similarity to EML4-ALK. **E** Heatmap comparing the drug-resistance profiles of *ALK* mutations reported in other cancers (neuroblastoma, glioma, and inflammatory myofibroblastic tumor; left axis), based on prior literature describing sensitivity or resistance to crizotinib, alectinib, and lorlatinib. These literature-based annotations are compared with resistance classifications from this study, derived using H3122 cells treated with alectinib or lorlatinib (top axis). Mutations are categorized as sensitive or resistant, highlighting concordant results

EML4–ALK fusion present in H3122 cells (Fig. 6D). Based on this conserved structural context, we compared our resistance classifications with previously reported *ALK* mutation profiles across other cancers. Notably, several *ALK* mutations identified as drug-resistant in neuroblastoma [57–61], IMT [62], and glioma [63] matched the resistance profiles from our classification (Fig. 6E).

Together, these observations suggest that the resistance-associated features identified in our screen are not restricted to *EML4–ALK* lung cancer, but may reflect conserved properties of the ALK kinase domain across diverse oncogenic contexts. At the same time, because functional dependency on ALK signaling varies substantially between tumor types, disease-specific validation will be important for translating these classifications beyond ALK-dependent lung cancer.

Discussion

In this study, we systematically profiled the functional consequences of 3,208 variants within the *ALK* TKD sequence using prime editing in a *EML4–ALK*-positive lung cancer model. Our approach provides a high-resolution map of *ALK* variant–specific resistance across three generations of ALK TKIs, uncovering both known and previously uncharacterized resistance-conferring mutations. To our knowledge, this represents the most comprehensive functional interrogation of ALK TKI resistance to date. The strong correlation between our drug-resistance scores and both individual validation results and clinical outcomes in ALK-positive lung cancer underscores the robustness of our approach.

Clinically, our findings have important implications for precision medicine in the treatment of *EML4–ALK*-positive cancers. Our resistance assays identified a diverse spectrum of resistance mutations, with substantial overlap but notable divergence across alectinib, lorlatinib, and zotizalkib. Canonical resistance mutations, such as L1196M and G1202R, predictably conferred resistance [2, 36, 39–41]. Notably, resistance defined in vitro reflects relative drug sensitivity under controlled selective pressures and does not directly equate to absolute clinical insensitivity. In clinical settings, mutations such as G1202R or L1196M may still be considered lorlatinib-responsive because their absolute IC50 values remain within clinically achievable exposure ranges, despite substantial reductions in drug potency relative to that seen with wild-type *ALK* [46]. In addition, we identified distal resistance-associated variants, including F1174C and F1245C, that confer resistance despite their spatial separation from the ATP-binding pocket. Although these observations expand the spectrum of resistance-associated sites within *ALK*, the precise biophysical mechanisms underlying distal resistance remain to be elucidated.

Importantly, our dataset offers potential clinical utility, suggesting that the functional impact of *ALK* mutations may help contextualize resistance mechanisms and potentially inform treatment considerations. Furthermore, the structural conservation of the ALK TKD across ALK-driven malignancies raises the possibility that aspects of this resistance landscape may extend beyond lung cancer. However, applications to other disease contexts should be approached cautiously and may require additional validation.

This study has several limitations. Resistance classifications derived from functional screening are influenced by oncogene dependency in the cellular context tested. Although ALK fusion status defines a major molecular subset of lung cancer, our analysis in NCI-H2228 cells indicates that the presence of an *EML4–ALK* fusion alone does not necessarily confer functional dependency on ALK signaling. In such contexts, resistance patterns defined in ALK-dependent models may not be directly transferable. In addition, although we employed direct endogenous sequencing and multi-pegRNA redundancy to mitigate guide-efficiency and specificity concerns inherent to prime editing–based

screens, we acknowledge that rare unintended editing outcomes or secondary mutations cannot be fully excluded under these experimental conditions. Likewise, although synonymous SNVs were used as empirical neutral references to normalize position- and context-dependent effects, a minority of synonymous changes may exert functional effects. Furthermore, because resistance scores are derived from defined drug concentrations, variants exhibiting highly dose-dependent or context-specific resistance profiles may not be fully captured by a fixed-dose screening framework. Clinical outcome analyses—particularly for zotizalkib—were limited by the availability of patient data. In addition, although higher nominal concentrations were required in vitro to achieve equivalent selective pressure, the relevance of such exposures to patient tolerability or off-target effects will need to be established through pharmacokinetic and clinical studies. Finally, because our analyses were performed in controlled in vitro systems to enable systematic and scalable functional interrogation, in vivo validation will be important to further assess how variant-specific resistance phenotypes manifest under physiological drug exposure and tumor microenvironmental constraints.

Conclusions

Together, the results from our study provide a comprehensive functional map of resistance mutations within the *ALK* TKD, generated at single-variant resolution through prime editing. These findings delineate canonical and distal resistance mechanisms and establish a quantitative framework for interpreting variant-specific drug responses in ALK-positive lung cancer and other ALK-driven malignancies. The resulting resistance atlas represents a valuable resource for guiding drug development and treatment selection, ultimately supporting precision oncology efforts in this clinically important subset of cancers.

Methods

Cell lines and culture

NCI-H3122 (RRID: CVCL_5160) and NCI-H2228 (RRID: CVCL_1543) cells were purchased from Cytion. These H3122 and H2228 cells were maintained in RPMI-1640 medium (Gibco; with 2.05 mM L-glutamine), supplemented with 10 mM HEPES (Welgene), 1 mM sodium pyruvate (Gibco), 10% fetal bovine serum (FBS) (RDT), and 1% penicillin/streptomycin (Gibco). HEK293T cells (ATCC) were cultured in Dulbecco's Modified Eagle Medium (Gibco) with 10% FBS (RDT) and 1% penicillin/streptomycin (Gibco). All cell lines were maintained at 37 °C in a humidified incubator with 5% CO_2. Ba/F3 murine bone marrow-derived pro-B cells were obtained from ATCC. Ba/F3 cells were cultured in RPMI-1640 medium supplemented with 10% FBS, 1% penicillin/streptomycin, and 10 ng/ml of interleukin-3 (IL-3, PeproTech). All cell lines were routinely tested and confirmed to be free of mycoplasma contamination.

Plasmid vector construction

Plasmid constructs were assembled using either Gibson assembly (NEBuilder® HiFi DNA Assembly Master Mix, New England Biolabs) or conventional ligation following restriction enzyme digestion (T4 DNA Ligase, NEB), in accordance with the manufacturer's instructions.

To generate pLenti-PEmax-BSD and pLenti-PE7-BSD, the SpCas9 coding sequence in the LentiCas9-Blast plasmid (Addgene #52962) [64] was replaced with PEmax or PE7 coding sequence, which were PCR-amplified from pCMV-PEmax (Addgene #174820) [33] and pCMV-PE7 (Addgene #214812) [35], respectively.

For pegRNA expression in transfection experiments and evaluation of resistance to ALK TKIs, the Lenti-gRNA-Puro vector (Addgene #84752) was linearized by digestion with BsmBI-v2 (NEB), gel-purified, and ligated with double-stranded DNA inserts encoding pegRNA sequences. These inserts were generated by annealing complementary single-stranded oligonucleotides designed with 4-bp overhangs compatible with the vector backbone.

Plasmid assembly products (2–4 μL of the Gibson or ligation mixture) were transformed into NEB Stable Competent E. coli (NEB). Transformed cells were cultured at 30 °C or 37 °C, and plasmid DNA was extracted using a NICSROprep™ Plasmid DNA Miniprep S&V Kit (Bionics). All constructs were verified by Sanger sequencing (Macrogen).

Lentivirus production

HEK293T cells were seeded at a density of 1×10^7 cells per 150-mm dish approximately 18 h prior to transfection. Lentiviral particles were produced using Lipofectamine 3000 (Thermo Fisher Scientific) according to the manufacturer's instructions. Briefly, 20 μg of transfer plasmid encoding the gene of interest, 15 μg of packaging plasmid psPAX2 (Addgene #12,260), and 5 μg of envelope plasmid pMD2.G (Addgene #12,259) were mixed in 1 mL of Opti-MEM (Gibco) along with 80 μL of P3000 reagent. In parallel, 90 μL of Lipofectamine 3000 was diluted in 1 mL of Opti-MEM and combined with the plasmid mixture. After 15 min of incubation at room temperature, the transfection mix was added to the HEK293T cells. The medium was replaced 6 h post-transfection, and virus-containing supernatant was harvested at 48 h. Supernatants were filtered through a 0.45-μm Millex-HV low-protein-binding membrane (Millipore), aliquoted, and stored at − 80 °C until use.

Generation of cell lines

To generate MMR-deficient H3122 cells, wild-type H3122 cells were transfected with a combination of plasmids: a Cas9-expressing plasmid (pRGEN-Cas9-CMV/T7-Puro-RFP; ToolGen) and two sgRNA-expressing plasmids (pRG2; Addgene #104,174) targeting *MLH1* [34]. The sgRNA sequences used were as follows:

MLH1-sgRNA1, 5′-TGGAGGGGGATACAACAAAG(GGG)−3′
MLH1-sgRNA2, 5′-GCCTGGGGCTGACTGGACAG(GGG)−3′

Transfections were performed using Lipofectamine 3000 following the manufacturer's protocol. A total of 1.5 μg of Cas9-expressing plasmid and 0.5 μg of each sgRNA plasmid were transfected into 3×10^5 cells seeded the previous day. Five days post-transfection, cells were single-cell sorted into 96-well plates. Clonal populations were screened by deep sequencing using primers 21 and 22 (Additional file 2: Table S10). Editing efficiency and indel frequencies were assessed using Cas-Analyzer.

To generate stable H3122 *MLH1*-knockout lines expressing PE2max or PE7, 2.0×10^6 *MLH1*-deficient H3122 cells were seeded into 100-mm dishes and transduced the following day with lentivirus produced from pLenti-PEmax-BSD or pLenti-PE7-BSD. Transductions were carried out at high multiplicity of infection (MOI) in the presence of 10 μg/mL polybrene (Sigma). After 24 h, cells were selected with 8 μg/mL blasticidin (InvivoGen) for 7 days.

To generate *MMR*-deficient H2228 cells, wild-type H2228 cells were transfected using the same approach as described for H3122 cells, with the following modifications. Base editor BE4max (pCMV-BE4max-3xHA; Addgene #112,096) and an sgRNA targeting *MSH6* (pRG2; Addgene #104,174) were used. The sgRNA sequence used was as follows:

MSH6-sgRNA1, 5′-GAATCCcAAGCCCACGTTAG (TGG)—3′

Transfections were performed using an approach similar to that described above. Clonal populations were isolated and screened by targeted deep sequencing using primers 23 and 24 (Additional file 2: Table S10). Editing efficiency and indel frequencies were assessed using BE-Analyzer (95.4% editing efficiency; PRJNA1272298).

To generate stable *MSH6*-deficient H2228 cells expressing PE7, *MSH6*-knockout clones were transduced with lentivirus produced from pLenti-PE7-BSD under the same conditions used for H3122 cells, including a high MOI in the presence of 10 μg/mL polybrene. Stable cell lines were selected with 5 μg/mL blasticidin (InvivoGen) for 7 days.

Design of pegRNA libraries for saturation mutagenesis

To perform saturation mutagenesis across the *ALK* coding region, including the ± 5 nucleotides flanking each exon–intron junctions, we designed comprehensive pegRNA and epegRNA libraries using the human genome reference (hg38) as a template. The final library was divided into nine subsets, each targeting one *ALK* exon (20 through 28).

To enhance editing efficiency and accurately detect genuine edits, we employed the PEER-seq strategy [34]. This method incorporates synonymous substitutions near the intended edit site, facilitating direct sequencing of the genomic locus while reducing background noise and partially bypassing MMR surveillance. Synonymous markers were randomly assigned, resulting in diverse pegRNA designs for each SNV, which were subsequently used to evaluate editing outcomes.

Candidate pegRNAs were ranked by DeepPrime-FT scores [38], calculated using a fine-tuned predictive model trained on PE4max editing efficiency in A549 cells. Guides with right homology arms (RHAs) shorter than 4 nucleotides were excluded. For each SNV, the three guides with the highest scores were selected, with at least two featuring distinct spacer sequences whenever possible. Two of the selected guides were pegRNAs, and the third was an epegRNA, consistent with previous findings that PE7 performs optimally with pegRNAs.

Synonymous substitution sites were chosen based on the following design criteria: (i) preferential use of alternative codons that do not overlap with the primary target base; (ii) avoidance of substitutions within 2 nucleotides of exon boundaries or 5 nucleotides of exon–intron junctions to minimize splicing interference whenever possible; (iii) preference for substitutions that disrupt the protospacer adjacent motif (PAM, NGG); (iv) when PAM disruption was infeasible, priority given to substitutions near the PAM

within the left homology arm; (v) additional preference was given to substitutions in the RHA closest to the primary editing site. To prevent premature transcription termination at poly-T motifs in guide RNAs driven by the U6 promoter [65], additional synonymous mutations were introduced within the reverse transcription template if needed.

Finally, to ensure comprehensive SNV representation across *ALK* exons, the number of guide RNA oligonucleotides per SNV was scaled based on the cumulative DeepPrime-FT score. Specifically, oligos were synthesized at fivefold scale for SNVs with scores below 2, fourfold for scores between 2 and 4, threefold for scores between 4 and 6, and twofold for scores between 6 and 8. If the cumulative score exceeded 70, the number of guides per SNV was limited to two to minimize redundancy while retaining high-scoring candidates.

Construction of pegRNA libraries and cloning

The pegRNA libraries were designed for saturation editing of the *ALK* TKD sequence, targeting exons 20–28 and including ± 5 bp of adjacent splice sites. A pooled library of 256-nucleotide oligonucleotides was synthesized via array-based synthesis (Twist Bioscience) for downstream plasmid construction. Each oligonucleotide contained the following components:

(1) an 18-bp left homology arm terminating at the 3' end of the U6 promoter; (2) a 20-bp guide sequence comprising a 5′ G and a 19-bp spacer; (3) an improved version of the SpCas9 sgRNA scaffold sequence [66]; (4) a reverse transcription (RT) template and primer binding site (PBS); (5) an 8-bp linker (from the pegLIT toolkit) [37] and a 37-bp tevopreQ1 aptamer for the epegRNA format, followed by a 7-bp poly-T terminator; (6) a variable-length buffer sequence for oligo length standardization; and (7) a 20-bp homology arm specific to each exon for subset library amplification.

Unique homology arm sequences for *ALK* were as follows:

ALK exon 20: 5'-TGCGACCGTAATCAAACCAA-3'
ALK exon 21: 5'-GTTCAAATTGCGTGCGACAT-3'
ALK exon 22: 5'-AAATGGATGCCTTGTGCGAA-3'
ALK exon 23: 5'-AAACGGAGCCATGAGTTTGT-3'
ALK exon 24: 5'-AAACTGGAGGCGGCAAATTA-3'
ALK exon 25: 5'-ACCGCGCTCGAAGAATTTAA-3'
ALK exon 26: 5'-TCCTCAGCCGATGAAATTCC-3'
ALK exon 27: 5'-ATGCAATCGGCCTGGTATTT-3'
ALK exon 28: 5'-ACAATTAGTTGGCGCTTCCT-3'

Oligonucleotide pools were PCR-amplified using Q5 High-Fidelity DNA Polymerase (NEB) with the following conditions: 30 s at 98 °C, 30 s at 68 °C, 2 min at 72 °C for 14 cycles. Amplicons were gel-purified. The lentiviral backbone (Lenti-gRNA-Puro, Addgene #84,752) was linearized by BsmBI digestion (Enzynomics, 55 °C for 6 h), followed by gel purification. DNA fragments were assembled using NEBuilder® HiFi DNA Assembly Master Mix. The assembled library was concentrated by isopropanol precipitation with GlycoBlue™ (Invitrogen) and electroporated into Endura electrocompetent cells (Lucigen) using a MicroPulser (Bio-Rad) to generate a plasmid library.

Drug dose–response cell viability assay in H3122 and H2228 cells

For drug dose–response cell viability assays, H3122 and H2228 cells stably expressing PE7 (H3122-PE7, H2228-PE7) were seeded at 2×10^6 cells per 150-mm dish in complete RPMI-1640 medium containing serial dilutions of ALK TKIs [alectinib (Selleckchem, S2762), lorlatinib (Selleckchem, S7536), or zotizalkib (MCE, HY-139279)], in duplicate dishes. TKIs were replenished every 3 days by replacing the medium. Seven days after seeding, cells were harvested by trypsinization and stained with trypan blue (Gibco), after which viable cell numbers were quantified using a Countess Automated Cell Counter (Thermo Fisher Scientific) according to the manufacturer's instructions. Cell viability was calculated as the percentage of viable cells relative to that in the DMSO-treated vehicle control. All experiments were performed in at least two independent biological replicates.

High-throughput drug-resistance screening of *ALK* variants

All screening experiments to measure drug resistance were performed in biological duplicates. To generate exon-specific pegRNA library-expressing H3122 cells, 4×10^6 cells per dish were plated in 150-mm dishes 24 h before transduction. To ensure sufficient cell coverage (> 5,000 cells per pegRNA), a total of 1.8×10^7 cells were seeded for the exon 20, 23, and 27 libraries, and a total of 1×10^7 cells were plated for each of the other exon library transductions. Libraries were transduced at an MOI of 0.5 in the presence of 8 μg/ml polybrene. Cells were selected with 0.5 μg/ml puromycin (Gibco) for 10 days, while being expanded for subsequent experiments. In parallel, H2228 cells were processed using identical criteria but transduced only with exon 22 and exon 23 libraries.

After selection, cells were harvested to maintain ≥ 5,000 × coverage and distributed into four conditions: (1) untreated control, (2) 20 nM alectinib treatment (SelleckChem, S2762), (3) 5 nM lorlatinib treatment (SelleckChem, S7536), and (4) 300 nM zotizalkib treatment (MCE, HY-139279), whereas H2228 cells were treated with 3 μM alectinib, 20 μM lorlatinib, or 300 nM zotizalkib. These concentrations were selected to induce approximately 20–30% reductions in cell viability over a 10-day treatment period, consistent with prior pooled CRISPR-based screening studies [34, 67].

To further assess the robustness of resistance profiles across a broader range of selective pressures, high-dose screening was also performed. In H3122 cells, high-dose conditions included 40 nM and 80 nM alectinib, 10 nM and 20 nM lorlatinib, and 900 nM and 2700 nM zotizalkib. In H2228 cells, high-dose conditions included 6 μM alectinib, 40 μM lorlatinib, and 1000 nM zotizalkib. These higher drug concentrations resulted in more than 50% reductions in cell viability (Additional file 1: Figs. S10 and S13).

Cells were cultured for an additional 10 days under their respective treatment conditions. At the end of the treatment period, cells were harvested for subsequent genomic DNA (gDNA) extraction to enable downstream sequencing and analysis of drug response profiles.

Genomic DNA extraction and deep sequencing

gDNA was isolated using a Wizard® Genomic DNA Purification Kit (Promega) according to the manufacturer's instructions. PCR amplification was carried out in two sequential rounds.

For the first round, 22–66 μg of purified gDNA—corresponding to > 5,000 × coverage per library, based on an estimate of 6 pg gDNA per cell—was amplified using exon-specific

primers and Q5® High-Fidelity DNA Polymerase (NEB) (Additional file 2: Table S10). PCR was performed for 25 cycles with annealing temperatures ranging from 58 °C to 64 °C. The resulting amplicons were pooled, purified using a MEGAquick-Spin Total Fragment DNA Purification Kit (iNtRON Biotechnology), and subjected to size selection via 2% agarose gel electrophoresis followed by gel extraction using the same kit.

In the second (indexing) PCR, 40 ng of purified product from the first round was used as input and amplified using Illumina-compatible indexing primers and Q5® High-Fidelity DNA Polymerase. The reaction was carried out at 60 °C for 8 cycles. Final amplicons were purified using the same kit and sequenced on an Illumina NovaSeq 6000 platform.

Sequencing data processing and variant identification

To identify SNVs from deep sequencing data, we constructed a custom reference database encompassing all intended variants within *ALK* exons 20–28. Reference sequences were derived from the NM_004304.5 transcript and included the full coding sequence with five intronic nucleotides flanking each exon–intron boundary. Each reference entry encoded a specific SNV along with a synonymous substitution; any additional mismatches were excluded. Sequencing reads were preprocessed and aligned to this reference database, and only reads with perfect matches—whether edited or wild-type—were retained for downstream analysis.

To distinguish true prime-edited SNVs from sequencing or background noise, we applied Fisher's exact test to compute ORs and *P*-values by comparing day 0 (D0) samples to unedited controls:

$$\mathrm{OR} = \frac{(SNV\ read\ counts\ (D0) + 1)/(Wild\ type\ read\ counts\ (D0) + 1)}{(SNV\ read\ counts\ (unedited\ cells) + 1)/(Wild\ type\ read\ counts\ (unedited\ cells) + 1)}$$

Variants were classified as true prime-edited events if they met all of the following criteria: (i) $\mathrm{OR} \geq 3$, (ii) false discovery rate < 0.05 after Benjamini–Hochberg correction, and (iii) normalized abundance exceeding 0.5 RPM [34]. These thresholds ensured both statistical significance and sufficient representation in the sequencing data.

Resistance score calculation and classification

LFCs for each SNV were calculated by comparing allele frequencies between drug-treated and untreated cells at day 10 (D10), following exposure to alectinib, lorlatinib, or zotizalkib. To account for variability in editing efficiency due to sequence context and positional bias, we standardized the LFC of each SNV using the LFCs of synonymous SNVs, which were considered neutral as they do not alter amino acid changes. For each exon, the LFCs of synonymous SNVs were modeled using Locally Weighted Scatterplot Smoothing (LOWESS) regression to capture positional trends. Standardization was then performed by subtracting the regressed synonymous LFC from each SNV's LFC and dividing by the standard deviation of synonymous LFCs within the same exon, enabling direct comparison between exons [34].

$$\text{Standardized LFC} = \frac{each\ SNV\ LFC - each\ position\ synonymous\ LFC\ (Lowess\ regressed)}{Standard\ deviation\ of\ synonymous\ SNVs}$$

After standardization, to integrate information across synonymous co-edits (internal replicates), a weighted average of standardized LFCs was computed for each SNV, taking into account the relative allele frequencies of each variant:

$$\text{Weighted LFC} = \frac{\sum \textit{allele frequency of each mutation in SNV} \times \textit{LFC of each mutation (Intended edit)}}{\sum \textit{allele frequency of each SNV (Intended edit)}}$$

Final resistance scores were derived by averaging weighted LFCs across biological replicates. For single amino acid substitutions encoded by multiple SNVs, the resistance score was defined as the mean of all relevant SNV scores.

SNVs generated by the *ALK* exon 20–28 libraries were classified based on thresholds derived from the distribution of synonymous SNV resistance scores. Classifications were defined as follows [34]:

- Resistant: SNV resistance scores that were above the 99.7th percentile relative to scores of the synonymous SNVs in both replicates.

- Sensitive: SNV resistance scores that were under the 95th percentile relative to scores of the synonymous SNV in both replicates.

- Intermediate: SNVs not classified as 'Resistant' or 'Sensitive'.

For protein variant classification, LFCs of protein variants induced by multiple SNVs were calculated as previously described, using the same classification criteria as in SNV-based analyses.

Fitness score calculation and classification of drug-resistant subtypes

Fitness scores were calculated to quantify the effects of individual SNVs on baseline cell growth in the absence of drug treatment. For each SNV, LFCs in variant frequency were calculated by comparing values from untreated samples at day 10 (D10) to those at baseline (D0). To correct for sequence context– and position-dependent effects on editing efficiency, SNV LFCs were standardized using synonymous SNVs. For each exon, the positional trends of synonymous SNV LFCs were estimated using LOWESS regression, and standardized LFCs were obtained by subtracting the regressed synonymous LFC at each position and dividing by the standard deviation of synonymous SNV LFCs [34]. Standardized LFCs were then combined across internal replicates using a weighted average that accounts for read depth and statistical confidence, as described above. The fitness score for each SNV was calculated as the average of weighted LFCs across biological replicates.

To classify resistance subtypes, fitness scores were integrated with drug resistance scores using percentile-based thresholds derived from the distribution of synonymous SNVs within each exon, providing a statistically robust and internally controlled framework [47].

SNVs were classified as follows:

- Canonical: resistance score > 99.7th percentile of those of synonymous SNVs and fitness score between the 5th and 95th percentiles.
- Drug addiction-like: resistance score > 99.7th percentile and fitness score < 5th percentile.
- Driver-like: resistance score > 99.7th percentile and fitness score > 95th percentile.

Drug-resistance screening at variable lorlatinib concentrations

To assess the resistance of *ALK* exon 23 variants in a dose-dependent manner, we performed a drug-resistance screening using the pegRNA library specific to exon 23. All experiments were conducted in biological duplicates. H3122 cells were transduced with lentiviral vectors encoding the pegRNA library at an MOI of 0.5, ensuring a minimum 5,000-fold representation of each pegRNA throughout the experiment.

Twenty-four hours post-transduction, cells were subjected to selection with 0.5 μg/mL puromycin for 10 days. Following selection, cells were seeded into four treatment conditions corresponding to increasing concentrations of lorlatinib (SelleckChem, S7536): 0 nM (untreated), 2 nM, 5 nM, and 10 nM. Cell numbers in each condition were scaled to maintain ≥ 5,000-fold pegRNA coverage.

Cells were cultured under drug treatment for an additional 10 days. At the end of this period, cells were harvested for gDNA extraction. The resulting DNA was used for library preparation and deep sequencing to evaluate variant-specific drug-resistance profiles across the lorlatinib concentration gradient. Resistance scores for each SNV were calculated as described above.

Evaluation of drug resistance using a conventional approach

To introduce specific SNVs associated with drug resistance—P1153R, I1171M, F1174C, F1193L, L1196M, M1199L, G1202R, and F1245C—pegRNAs were designed and cloned into the BsmBI-linearized Lenti-gRNA-Puro vector (Addgene, 87,245). Golden Gate cloning was used to assemble three key fragments into the linearized vector: (1) annealed spacer oligonucleotides; (2) an improved SpCas9 sgRNA scaffold sequence; (3) an annealed pegRNA RT template with a 3′ PBS extension, all with appropriate overhangs for cloning.

For lentiviral transduction, 0.6×10^6 H3122 cells stably expressing PE7 were seeded into 100-mm dishes 24 h prior to infection. Each condition was tested in three biological replicates. Cells were infected with pegRNA-expressing lentiviruses at a high MOI (~ 1). Parallel cultures of 0.6 million cells were transduced with an empty vector as negative controls. At 24 h post-transduction, media were replaced with puromycin-containing media to select for successfully transduced cells.

After 10 days of selection, pegRNA-transduced and empty-vector-transduced cells were mixed at a 25:75 ratio. These mixed populations were divided into four treatment groups: untreated control, or treated with alectinib (20 nM), lorlatinib (5 nM), or zotizalkib (300 nM). Following 10 days of drug exposure, cells were harvested for gDNA extraction. Target regions were amplified using site-specific primers (Additional file 2: Table S10) and subjected to deep sequencing to quantify variant-specific enrichment under each drug condition.

Generation of Ba/F3 cells expressing EML4-ALK variants

cDNAs encoding *EML4–ALK* fusion variant 1 (corresponding to the H3122 fusion), including wild-type and representative mutant forms, were synthesized (gene fragments; Twist Bioscience) and cloned into the pMT_025 expression vector (Addgene #158,579) using Gibson assembly (Additional file 2: Table S11). Recombinant constructs were packaged into lentiviral particles using standard lentiviral production protocols.

For transduction, Ba/F3 cells were seeded at 4×10^6 cells per well in six-well plates and incubated with lentiviral supernatant in RPMI medium supplemented with IL-3 and 10 μg/mL polybrene (Sigma). Plates were centrifuged at 2,000 × g for 90 min at 25 °C (spinoculation) and incubated for an additional 48 h. Successfully transduced cells were selected by withdrawal of IL-3 from the culture medium, allowing the expansion of IL-3–independent Ba/F3 cells expressing each *EML4–ALK* variant.

IC50 determination in H3122 single clones and Ba/F3 variant cell lines

Single-cell H3122 clones carrying P1153R, L1196M, G1202R, or F1245C variants were generated using prime editing. PE7-expressing H3122 cells were transfected with plasmids encoding variant-specific pegRNAs and selected with puromycin for 5 days. Cells were then seeded at limiting dilution (0.3 cells per well) into 96-well plates (four plates per variant) and expanded for 14 days. Clonal genotypes were confirmed by targeted deep sequencing. For genotyping, cells were lysed directly in 96-well plates using 100 μl lysis buffer (50 mM Tris–HCl, 1 mM EDTA, 0.05% SDS, 5% proteinase K; Biofact) at 55 °C for 1 h, followed by two rounds of PCR amplification of the target regions, as described above.

For drug sensitivity assays, H3122 single clones and Ba/F3 variant cell lines were seeded into 96-well plates at a density of 5×10^3 cells per well. H3122 cells were maintained in complete growth medium, whereas Ba/F3 cells were cultured in IL-3-free RPMI medium during drug treatment. After 24 h, cells were treated with serially diluted TKIs and incubated for 72 h. Subsequently, 20 μL of MTT reagent (3-(4,5-dimethylthiazol-2-yl)−2,5-diphenyltetrazolium bromide, Thermo Fisher Scientific) was added to each well and cells were incubated for 1 h. Following removal of the media, 200 μL of DMSO was added to solubilize the formazan crystals. Absorbance was measured at 570 nm using a microplate reader (Infinite M Nano, Tecan), and cell viability was calculated relative to that of untreated controls. IC50 values were determined by nonlinear regression analysis using GraphPad Prism.

Survival analysis

Survival analysis was performed using the "overall survival (in months)" and "survival status" fields from the MSK-CHORD dataset [54], which includes genomic and clinical data. Two patient groups were analyzed: (i) *EML4–ALK* fusion-positive patients treated with alectinib prior to sequencing, and (ii) all *EML4–ALK* fusion-positive NSCLC patients.

Kaplan–Meier survival curves were generated using the lifelines Python package. Log-rank tests were used to compare survival distributions between patients harboring alectinib resistance mutations identified in this study and those with wild-type ALK (Fig. 6A, B).

Data visualization

All data visualizations were performed using Python. Illustrative schematics were created with BioRender.com. Structural visualization and variant mapping were conducted using PyMOL (version 2.5.5). Variants were mapped onto the crystal structures of the ALK TKD bound to alectinib (PDB ID: 3AOX) and lorlatinib (PDB ID: 4CLJ), as well as the full-length structure of EML4–ALK predicted by Alphafold3.

Additionally, AlphaFold3-predicted three-dimensional structures [56] of NPM1–ALK, ATIC–ALK, TPM3–ALK, TPM4–ALK, STRN–ALK, KIF5B–ALK, and EML4–ALK were aligned to compare structural variations across ALK fusion proteins using PyMOL (Fig. 6D).

Statistical analysis

Statistical analyses were performed using Python and Prism, and all tests were conducted as two-tailed unless otherwise specified.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-026-03977-4.

Additional file 1: Fig. S1. Optimization of prime editing in an EML4-ALK fusion positive lung cancer cell line. Fig. S2. Evaluation of the resistance profiles of ALK variants. Fig. S3. Alectinib resistance scores of 2,031 SAAVs. Fig. S4. Lorlatinib resistance scores of 2,031 SAAVs. Fig. S5. Zotizalkib resistance scores of 2,031 SAAVs. Fig. S6. Alectinib resistance scores of 3,208 SNVs. Fig. S7. Lorlatinib resistance scores of 3,208 SNVs. Fig. S8. Zotizalkib resistance scores of 3,208 SNVs. Fig. S9. Structural overview of ALK TKI–resistant variants. Fig. S10. Dose dependence of ALK SNV resistance profiles and comparisons with clinical data. Fig. S11. Integration of cellular fitness and drug resistance across ALK inhibitors. Fig. S12. Individual evaluations of drug resistance in single-variant models. Fig. S13. Evaluation of ALK TKIs in NCI-H2228 cells. Fig. S14. Somatic ALK mutations observed in the real world.

Additional file 2: Table S1. List of all pegRNAs used in this study. Table S2. The results of drug-resistance analysis for SNVs in ALK. Table S3. Comparing resistance scores for ALK SNVs treated with different concentrations of lorlatinib. Table S4. Cell viability of H3122 cells after 7-day treatment with increasing concentrations of TKIs. Table S5. Pairwise comparisons of resistance scores for ALK exon 23 SNVs across increasing concentrations of TKIs. Table S6. List of all pegRNAs used in individual tests and corresponding results. Table S7. MTT and IC50 result. Table S8. Cell viability of H2228 cells after 7-day treatment with increasing concentrations of TKIs. Table S9. Evaluation of ALK TKIs in NCI-H2228 cells. Table S10. Oligonucleotide sequences. Table S11. Genefragment sequences.

Acknowledgements
Not applicable.

Peer review information
Claudia Feng was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. The peer-review history is available in the online version of this article.

Authors' contributions
H.C.O., Y.H., Y.C., and H.H.K. conceived and designed the research. H.C.O., Y.H., Y.C., and H.H.K. analyzed the data. H.C.O., Y.H., and Y.C. performed the experiments. H.C.O., Y.H., Y.C., and H.H.K. wrote the manuscript. All authors read and approved the final manuscript.

Funding
This work was supported, in part, by National Research Foundation of Korea grants funded by the Korean government (MSIT) (RS-2022-NR070713 (H.H.K.) and RS-2025–02214844(H.H.K), RS-2025–18362970 (H.H.K.), The Bio and Medical Technology Development Program of the NRF funded by the Korean government (MSIT) RS-2022-NR067326 (H.H.K.), RS-2022-NR067345 (H.H.K.) and (RS-2023–00260968) (H.H.K.), the Korea Drug Development Fund funded by the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Korea-US Collaborative Research Fund (KUCRF), funded by the Ministry of Science and ICT and Ministry of Health & Welfare, Republic of Korea (grant number: RS-2024–00467177), the Yonsei Signature Research Cluster Program of 2024–22-0165 (H.H.K.), the Brain Korea 21 FOUR Project for Medical Science (Yonsei University College of Medicine), cisionMed (2025-31-0990), the "Regional Innovation System & Education (RISE)" through the Seoul RISE Center, funded by the Ministry of Education (MOE) and the Seoul Metropolitan Government. (2025-RISE-01–022-05) the Yonsei Fellow Program, funded by Lee Youn Jae, the Commercialization Promotion Agency for R&D Outcomes (COMPA) funded by the Ministry of Science and ICT (MSIT) (No.RS-2025-02633786), a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (RS-2025–25467624) and a grant of the MD-PhD/Medical Scientist Training Program (Y.H.) through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea.

Data availability
We have submitted the deep sequencing data from this study to the National Center of Biotechnology Information's Sequence Read Archive under accession number PRJNA1272298 [68]. Support codes for analysis are available at GitHub website [69] and Zenodo [70]. We have provided the datasets used in this study as Additional file 2: Tables S1–S11.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
Based on the findings of this study, Yonsei University has filed a patent listing H.C.O., Y.H., Y. C., and H.H.K. as inventors. H.H.K is the founder and an employee of cisionMed.

Author details
[1]Department of Pharmacology, Yonsei University College of Medicine, Seoul, Republic of Korea. [2]Department of Neurosurgery, Yonsei University College of Medicine, Seoul, Republic of Korea. [3]Graduate School of Medical Science, Brain Korea 21 Plus Project for Medical Sciences, Yonsei University College of Medicine, Seoul 03722, Republic of Korea. [4]Severance Biomedical Science Institute, Yonsei University College of Medicine, Seoul 03722, Republic of Korea. [5]Center for Nanomedicine, Institute for Basic Science (IBS), Seoul 03722, Republic of Korea. [6]Yonsei-IBS Institute, Yonsei University, Seoul 03722, Republic of Korea. [7]Woo Choo Lee Institute for Precision Drug Development, Yonsei University College of Medicine, Seoul 03722, Republic of Korea. [8]Institute for Immunology and Immunological Diseases, Yonsei University College of Medicine, Seoul 03722, Republic of Korea. [9]Won-Sang Lee Institute for Hearing Loss, Yonsei University College of Medicine, Seoul 03722, Republic of Korea. [10]cisionMed, Seoul 07789, Republic of Korea.

Received: 26 August 2025 Accepted: 23 January 2026
Published online: 03 February 2026

References
1. Shreenivas A, Janku F, Gouda MA, Chen HZ, George B, Kato S, et al. ALK fusions in the pan-cancer setting: another tumor-agnostic target? NPJ Precis Oncol. 2023;7(1):101. https://doi.org/10.1038/s41698-023-00449-x.
2. Schneider JL, Lin JJ, Shaw AT. ALK-positive lung cancer: a moving target. Nat Cancer. 2023;4(3):330–43. https://doi.org/10.1038/s43018-023-00515-0.
3. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature. 2007;448(7153):561–6. https://doi.org/10.1038/nature05945.
4. Cao Z, Gao Q, Fu M, Ni N, Pei Y, Ou WB. Anaplastic lymphoma kinase fusions: roles in cancer and therapeutic perspectives. Oncol Lett. 2019;17(2):2020–30. https://doi.org/10.3892/ol.2018.9856.
5. Morris SW, Kirstein MN, Valentine MB, Dittmer KG, Shapiro DN, Saltman DL, et al. Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. Science. 1994;263(5151):1281–4. https://doi.org/10.1126/science.8122112.
6. Debiec-Rychter M, Marynen P, Hagemeijer A, Pauwels P. Alk-atic fusion in urinary bladder inflammatory myofibroblastic tumor. Genes Chromosomes Cancer. 2003;38(2):187–90. https://doi.org/10.1002/gcc.10267.
7. Lin E, Li L, Guan Y, Soriano R, Rivers CS, Mohan S, et al. Exon array profiling detects EML4-ALK fusion in breast, colorectal, and non-small cell lung cancers. Mol Cancer Res. 2009;7(9):1466–76. https://doi.org/10.1158/1541-7786.MCR-08-0522.
8. Bagchi A, Orr BA, Campagne O, Dhanda S, Nair S, Tran Q, et al. Lorlatinib in a child with ALK-fusion-positive high-grade glioma. N Engl J Med. 2021;385(8):761–3. https://doi.org/10.1056/NEJMc2101264.
9. Blandin AF, Giglio R, Graham MS, Garcia G, Malinowski S, Woods JK, et al. ALK amplification and rearrangements are recurrent targetable events in congenital and adult glioblastoma. Clin Cancer Res. 2023;29(14):2651–67. https://doi.org/10.1158/1078-0432.CCR-21-3521.
10. Kris MG, Johnson BE, Berry LD, Kwiatkowski DJ, Iafrate AJ, Wistuba II, et al. Using multiplexed assays of oncogenic drivers in lung cancers to select targeted drugs. JAMA. 2014;311(19):1998–2006. https://doi.org/10.1001/jama.2014.3741.
11. Hirai N, Sasaki T, Okumura S, Minami Y, Chiba S, Ohsaki Y. Monomerization of ALK fusion proteins as a therapeutic strategy in ALK-rearranged non-small cell lung cancers. Front Oncol. 2020;10:419. https://doi.org/10.3389/fonc.2020.00419.
12. Sabir SR, Yeoh S, Jackson G, Bayliss R. Eml4-ALK variants: biological and molecular properties, and the implications for patients. Cancers (Basel). 2017. https://doi.org/10.3390/cancers9090118.
13. Camidge DR, Kim HR, Ahn MJ, Yang JC, Han JY, Lee JS, et al. Brigatinib versus crizotinib in *ALK* -positive non–small-cell lung cancer. N Engl J Med. 2018;379(21):2027–39. https://doi.org/10.1056/NEJMoa1810171.
14. Peters S, Camidge DR, Shaw AT, Gadgeel S, Ahn JS, Kim DW, et al. Alectinib versus crizotinib in untreated ALK-positive non-small-cell lung cancer. N Engl J Med. 2017;377(9):829–38. https://doi.org/10.1056/NEJMoa1704795.
15. Shaw AT, Bauer TM, de Marinis F, Felip E, Goto Y, Liu G, et al. First-line lorlatinib or crizotinib in advanced ALK-positive lung cancer. N Engl J Med. 2020;383(21):2018–29. https://doi.org/10.1056/NEJMoa2027187.
16. Shaw AT, Kim DW, Nakagawa K, Seto T, Crino L, Ahn MJ, et al. Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. N Engl J Med. 2013;368(25):2385–94. https://doi.org/10.1056/NEJMoa1214886.
17. Solomon BJ, Mok T, Kim DW, Wu YL, Nakagawa K, Mekhail T, et al. First-line crizotinib versus chemotherapy in ALK-positive lung cancer. N Engl J Med. 2014;371(23):2167–77. https://doi.org/10.1056/NEJMoa1408440.
18. Soria JC, Tan DSW, Chiari R, Wu YL, Paz-Ares L, Wolf J, et al. First-line ceritinib versus platinum-based chemotherapy in advanced ALK-rearranged non-small-cell lung cancer (ASCEND-4): a randomised, open-label, phase 3 study. Lancet. 2017;389(10072):917–29. https://doi.org/10.1016/S0140-6736(17)30123-X.
19. Lin JJ, Riely GJ, Shaw AT. Targeting ALK: precision medicine takes on drug resistance. Cancer Discov. 2017;7(2):137–55. https://doi.org/10.1158/2159-8290.CD-16-1123.

20. Choi YL, Soda M, Yamashita Y, Ueno T, Takashima J, Nakajima T, et al. EML4-ALK mutations in lung cancer that confer resistance to ALK inhibitors. N Engl J Med. 2010;363(18):1734–9. https://doi.org/10.1056/NEJMoa1007478.
21. Dagogo-Jack I, Rooney M, Lin JJ, Nagy RJ, Yeap BY, Hubbeling H, et al. Treatment with next-generation ALK inhibitors fuels plasma *ALK* mutation diversity. Clin Cancer Res. 2019;25(22):6662–70. https://doi.org/10.1158/1078-0432.CCR-19-1436.
22. Gainor JF, Dardaei L, Yoda S, Friboulet L, Leshchiner I, Katayama R, et al. Molecular mechanisms of resistance to first- and second-generation ALK inhibitors in ALK-rearranged lung cancer. Cancer Discov. 2016;6(10):1118–33. https://doi.org/10.1158/2159-8290.CD-16-0596.
23. Katayama R, Khan TM, Benes C, Lifshits E, Ebi H, Rivera VM, et al. Therapeutic strategies to overcome crizotinib resistance in non-small cell lung cancers harboring the fusion oncogene EML4-ALK. Proc Natl Acad Sci U S A. 2011;108(18):7535–40. https://doi.org/10.1073/pnas.1019559108.
24. Toyokawa G, Hirai F, Inamasu E, Yoshida T, Nosaki K, Takenaka T, et al. Secondary mutations at I1171 in the ALK gene confer resistance to both crizotinib and alectinib. J Thorac Oncol. 2014;9(12):e86. https://doi.org/10.1097/JTO.0000000000000358.
25. Dagogo-Jack I, Yoda S, Lennerz JK, Langenbucher A, Lin JJ, Rooney MM, et al. MET alterations are a recurring and actionable resistance mechanism in ALK-positive lung cancer. Clin Cancer Res. 2020;26(11):2535–45. https://doi.org/10.1158/1078-0432.CCR-19-3906.
26. Katayama R, Shaw AT, Khan TM, Mino-Kenudson M, Solomon BJ, Halmos B, et al. Mechanisms of acquired crizotinib resistance in ALK-rearranged lung cancers. Sci Transl Med. 2012;4(120):120ra17. https://doi.org/10.1126/scitranslmed.3003316.
27. Koopman B, Groen HJM, Schuuring E, Hiltermann TJN, Timens W, den Dunnen WFA, et al. Actionability of on-target ALK resistance mutations in patients with non-small cell lung cancer: local experience and review of the literature. Clin Lung Cancer. 2022;23(2):e104–15. https://doi.org/10.1016/j.cllc.2021.06.011.
28. Recondo G, Mezquita L, Facchinetti F, Planchard D, Gazzah A, Bigot L, et al. Diverse resistance mechanisms to the third-generation ALK inhibitor lorlatinib in ALK-rearranged lung cancer. Clin Cancer Res. 2020;26(1):242–55. https://doi.org/10.1158/1078-0432.CCR-19-1104.
29. Yoda S, Lin JJ, Lawrence MS, Burke BJ, Friboulet L, Langenbucher A, et al. Sequential ALK inhibitors can select for lorlatinib-resistant compound ALK mutations in ALK-positive lung cancer. Cancer Discov. 2018;8(6):714–29. https://doi.org/10.1158/2159-8290.CD-17-1256.
30. Riely GJ, Wood DE, Ettinger DS, Aisner DL, Akerley W, Bauman JR, et al. Non-small cell lung cancer, version 4.2024, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw. 2024;22(4):249–74. https://doi.org/10.6004/jnccn.2204.0023.
31. Murray BW, Zhai D, Deng W, Zhang X, Ung J, Nguyen V, et al. TPX-0131, a potent CNS-penetrant, next-generation inhibitor of wild-type ALK and ALK-resistant mutations. Mol Cancer Ther. 2021;20(9):1499–507. https://doi.org/10.1158/1535-7163.MCT-21-0221.
32. Ou SI, Zhu VW, Nagasaka M. Catalog of 5′fusion partners in ALK-positive NSCLC circa 2020. JTO Clin Res Rep. 2020;1(1):100015. https://doi.org/10.1016/j.jtocrr.2020.100015.
33. Chen PJ, Hussmann JA, Yan J, Knipping F, Ravisankar P, Chen PF, et al. Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. Cell. 2021;184(22):5635-52 e29. https://doi.org/10.1016/j.cell.2021.09.018.
34. Kim Y, Oh HC, Lee S, Kim HH. Saturation profiling of drug-resistant genetic variants using prime editing. Nat Biotechnol. 2024;43(9):1471–84. https://doi.org/10.1038/s41587-024-02465-z.
35. Yan J, Oyler-Castrillo P, Ravisankar P, Ward CC, Levesque S, Jing Y, et al. Improving prime editing with an endogenous small RNA-binding protein. Nature. 2024;628(8008):639–47. https://doi.org/10.1038/s41586-024-07259-6.
36. Voena C, Ambrogio C, Iannelli F, Chiarle R. ALK in cancer: from function to therapeutic targeting. Nat Rev Cancer. 2025;25(5):359–78. https://doi.org/10.1038/s41568-025-00797-9.
37. Nelson JW, Randolph PB, Shen SP, Everette KA, Chen PJ, Anzalone AV, et al. Engineered pegRNAs improve prime editing efficiency. Nat Biotechnol. 2022;40(3):402–10. https://doi.org/10.1038/s41587-021-01039-7.
38. Yu G, Kim HK, Park J, Kwak H, Cheong Y, Kim D, et al. Prediction of efficiencies for diverse prime editing systems in multiple cell types. Cell. 2023;186(10):2256-72 e23. https://doi.org/10.1016/j.cell.2023.03.034.
39. Li J, Huang Y, Wu M, Wu C, Li X, Bao J. Structure and energy based quantitative missense variant effect analysis provides insights into drug resistance mechanisms of anaplastic lymphoma kinase mutations. Sci Rep. 2018;8(1):10664. https://doi.org/10.1038/s41598-018-28752-9.
40. Lu Y, Fan Z, Zhu SJ, Huang X, Zhuang Z, Li Y, et al. A new ALK inhibitor overcomes resistance to first- and second-generation inhibitors in NSCLC. EMBO Mol Med. 2022;14(1):e14296. https://doi.org/10.15252/emmm.202114296.
41. Poei D, Ali S, Ye S, Hsu R. ALK inhibitors in cancer: mechanisms of resistance and therapeutic management strategies. Cancer Drug Resist. 2024;7:20. https://doi.org/10.20517/cdr.2024.25.
42. Ni Z, Wang X, Zhang T, Jin RZ. Molecular dynamics simulations reveal the allosteric effect of F1174C resistance mutation to ceritinib in ALK-associated lung cancer. Comput Biol Chem. 2016;65:54–60. https://doi.org/10.1016/j.compbiolchem.2016.10.005.
43. Burley SK, Petsko GA. Aromatic-aromatic interaction: a mechanism of protein structure stabilization. Science. 1985;229(4708):23–8. https://doi.org/10.1126/science.3892686.
44. Mizuta H, Okada K, Araki M, Adachi J, Takemoto A, Kutkowska J, et al. Gilteritinib overcomes lorlatinib resistance in ALK-rearranged cancer. Nat Commun. 2021;12(1):1261. https://doi.org/10.1038/s41467-021-21396-w.
45. Shaw AT, Solomon BJ, Besse B, Bauer TM, Lin CC, Soo RA, et al. Alk resistance mutations and efficacy of lorlatinib in advanced anaplastic lymphoma kinase-positive non-small-cell lung cancer. J Clin Oncol. 2019;37(16):1370–9. https://doi.org/10.1200/JCO.18.02236.
46. Shiba-Ishii A, Johnson TW, Dagogo-Jack I, Mino-Kenudson M, Johnson TR, Wei P, et al. Analysis of lorlatinib analogs reveals a roadmap for targeting diverse compound resistance mutations in ALK-positive lung cancer. Nat Cancer. 2022;3(6):710–22. https://doi.org/10.1038/s43018-022-00399-6.
47. Coelho MA, Strauss ME, Watterson A, Cooper S, Bhosle S, Illuzzi G, et al. Base editing screens define the genetic landscape of cancer drug resistance mechanisms. Nat Genet. 2024;56(11):2479–92. https://doi.org/10.1038/s41588-024-01948-8.

48. Kong X, Kuilman T, Shahrabi A, Boshuizen J, Kemper K, Song JY, et al. Cancer drug addiction is relayed by an ERK2-dependent phenotype switch. Nature. 2017;550(7675):270–4. https://doi.org/10.1038/nature24037.
49. Suda K, Tomizawa K, Osada H, Maehara Y, Yatabe Y, Sekido Y, et al. Conversion from the "oncogene addiction" to "drug addiction" by intensive inhibition of the EGFR and MET in lung cancer with activating EGFR mutation. Lung Cancer. 2012;76(3):292–9. https://doi.org/10.1016/j.lungcan.2011.11.007.
50. Jung Y, Yu G, Oh HC, Lee JH, Jeon E, Bae J, et al. Comprehensive resistance profiling of chronic myeloid leukaemia associated ABL1 variants against five tyrosine kinase inhibitors using prime editing. Nat Biomed Eng. 2025. https://doi.org/10.1038/s41551-025-01531-4.
51. Kim HR, Kim WS, Choi YJ, Choi CM, Rho JK, Lee JC. Epithelial-mesenchymal transition leads to crizotinib resistance in H2228 lung cancer cells with EML4-ALK translocation. Mol Oncol. 2013;7(6):1093–102. https://doi.org/10.1016/j.molonc.2013.08.001.
52. Choi SH, Kim DH, Choi YJ, Kim SY, Lee JE, Sung KJ, et al. Multiple receptor tyrosine kinase activation related to ALK inhibitor resistance in lung cancer cells with ALK rearrangement. Oncotarget. 2017;8(35):58771–80. https://doi.org/10.18632/oncotarget.17680.
53. DepMap, Broad (2025). DepMap Public 25Q3. Dataset. depmap.org [Internet].
54. Jee J, Fong C, Pichotta K, Tran TN, Luthra A, Waters M, et al. Automated real-world data integration improves cancer outcome prediction. Nature. 2024;636(8043):728–36. https://doi.org/10.1038/s41586-024-08167-5.
55. Bresler SC, Weiser DA, Huwe PJ, Park JH, Krytska K, Ryles H, et al. ALK mutations confer differential oncogenic activation and sensitivity to ALK inhibition therapy in neuroblastoma. Cancer Cell. 2014;26(5):682–94. https://doi.org/10.1016/j.ccell.2014.09.019.
56. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024;630(8016):493–500. https://doi.org/10.1038/s41586-024-07487-w.
57. Bresler SC, Wood AC, Haglund EA, Courtright J, Belcastro LT, Plegaria JS, et al. Differential inhibitor sensitivity of anaplastic lymphoma kinase variants found in neuroblastoma. Sci Transl Med. 2011;3(108):108ra14. https://doi.org/10.1126/scitranslmed.3002950.
58. Debruyne DN, Bhatnagar N, Sharma B, Luther W, Moore NF, Cheung NK, et al. ALK inhibitor resistance in ALK(F1174L)-driven neuroblastoma is associated with AXL activation and induction of EMT. Oncogene. 2016;35(28):3681–91. https://doi.org/10.1038/onc.2015.434.
59. Durand S, Pierre-Eugene C, Mirabeau O, Louis-Brennetot C, Combaret V, Colmet-Daage L, et al. Alk mutation dynamics and clonal evolution in a neuroblastoma model exhibiting two ALK mutations. Oncotarget. 2019;10(48):4937–50. https://doi.org/10.18632/oncotarget.27119.
60. Foster JH, Voss SD, Hall DC, Minard CG, Balis FM, Wilner K, et al. Activity of crizotinib in patients with ALK-aberrant relapsed/refractory neuroblastoma: a children's oncology group study (ADVL0912). Clin Cancer Res. 2021;27(13):3543–8. https://doi.org/10.1158/1078-0432.CCR-20-4224.
61. Mosse YP, Lim MS, Voss SD, Wilner K, Ruffner K, Laliberte J, et al. Safety and activity of crizotinib for paediatric patients with refractory solid tumours or anaplastic large-cell lymphoma: a children's oncology group phase 1 consortium study. Lancet Oncol. 2013;14(6):472–80. https://doi.org/10.1016/S1470-2045(13)70095-0.
62. Sasaki T, Okuda K, Zheng W, Butrynski J, Capelletti M, Wang L, et al. The neuroblastoma-associated F1174L ALK mutation causes resistance to an ALK kinase inhibitor in ALK-translocated cancers. Cancer Res. 2010;70(24):10038–43. https://doi.org/10.1158/0008-5472.CAN-10-2956.
63. Gene-Olaciregui N, Perez-Somarriba M, Santa-Maria V, Cruz O, Gomez-Gonzalez S, Castaneda A, et al. Clinical and Molecular Evolution of an ALK-Driven Infant-Type Hemispheric Glioma Treated Sequentially With Second- and Third-Generation Anaplastic Lymphoma Kinase Inhibitors. JCO Precis Oncol. 2023;7:e2200547. https://doi.org/10.1200/PO.22.00547.
64. Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. Nat Methods. 2014;11(8):783–4. https://doi.org/10.1038/nmeth.3047.
65. Sousa AA, Hemez C, Lei L, Traore S, Kulhankova K, Newby GA, et al. Systematic optimization of prime editing for the efficient functional correction of CFTR F508del in human airway epithelial cells. Nat Biomed Eng. 2025;9(1):7–21. https://doi.org/10.1038/s41551-024-01233-3.
66. Dang Y, Jia G, Choi J, Ma H, Anaya E, Ye C, et al. Optimizing sgrna structure to improve CRISPR-Cas9 knockout efficiency. Genome Biol. 2015;16:280. https://doi.org/10.1186/s13059-015-0846-3.
67. Cuella-Martin R, Hayward SB, Fan X, Chen X, Huang JW, Taglialatela A, et al. Functional interrogation of DNA damage response variants with base editing screens. Cell. 2021;184(4):1081-97 e19. https://doi.org/10.1016/j.cell.2021.01.041.
68. Oh HC, Han Y, Chang Y, Kim H. A comprehensive functional atlas of ALK kinase domain variants reveals resistance landscape to ALK inhibitors. Datasets. Sequence Read Archive. 2026. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1272298.
69. Oh HC, Han Y, Chang Y, Kim H. Functional atlas of ALK kinase domain variants. Github. 2026. https://github.com/po99044/ALK.
70. Oh HC, Han Y, Chang Y, Kim H. Functional atlas of ALK kinase domain variants (v1.0.0). Zenodo. 2026. https://doi.org/10.5281/zenodo.18333259.

Publisher's Note