Supplementary Table S1

Per-variant provenance and codon reachability for the KRAS G12C and ALK composite resistance panels

Companion to: Ronco, M. *Quantifying the preclinical blind spots of mutagen-based TKI-resistance screens from chemistry-free saturation mutagenesis atlases.* 2026.

S1a. KRAS G12C resistance panel (12 variants)

Codon numbering follows ClinVar records anchored to NM_033360.4 (KRAS-4A). The wild-type codon reported here is the codon in the *G12C-mutant background* that is the shared starting point for every preclinical screen in the Awad 2021, Feng 2022, Koga 2021, Georgakopoulos-Soares 2025, and Tang 2022 (Boehringer) studies cited in the main text. Codon 12 is therefore TGT (not GGT); all other panel codons are wild-type at the KRAS reference. Reachability is classified by Hamming distance from the wild-type codon to any codon encoding the target amino acid: *two-step SNV* if the minimum distance is ≥ 2, meaning the variant is suppressed rather than forbidden (main text §2.2); *multi-class SNV* if a single-nucleotide route exists but more than one pyrimidine substitution class produces the amino acid; *class-forced* if a single-nucleotide route exists and only one pyrimidine substitution class produces the amino acid.

Variant	Source panel	WT codon	Target codon	Reachability	Class	Signature concordance	Notes
C12F	Awad 2021 *in vitro* DMS	TGT	TTT (pos 2 G>T)	class-forced	C>A	SBS4 (tobacco)	G12C-mutant background; single-nt via pos 2.
V8L	Awad 2021 *in vitro* DMS	GTA	CTA / TTA	multi-class SNV	—	multi-class SNV	Reachable in one nt via C>G (CTA) or C>A (TTA).
V8M	Awad 2021 *in vitro* DMS	GTA	ATG (requires 2 nt)	two-step SNV	—	two-step, 1 route	Two substitutions required (pos 1 G>A and pos 3 A>G); 51- to 169-fold suppressed.
R68L	Awad 2021 *in vitro* DMS	AGG	all Leu codons need ≥ 2 nt	two-step SNV	—	two-step, 2 routes	Top-ranked variant in Feng chemistry-free atlas under both drugs; 51- to 169-fold suppressed.
R68M	Awad 2021 *in vitro* DMS	AGG	ATG (pos 2 G>T)	class-forced	C>A	SBS4 (tobacco)	Single-nt route via pos 2.
R68S	Awad 2021 clinical cohort	AGG	AGT / AGC	multi-class SNV	—	multi-class SNV	C>A (AGT) or C>G (AGC) at pos 3.
H95D	Awad 2021 clinical cohort	CAT	GAT (pos 1 C>G)	class-forced	C>G	SBS13 at TpC; else rare	H95D under adagrasib: chem-free rank 12 of 1,205, Mutazyme-simulated rank 90 (shift +78 at $\beta = 1$; +0 to +812 across $\beta \in [0.25, 4]$).
H95Q	Awad 2021 clinical cohort	CAT	CAA / CAG	multi-class SNV	—	multi-class SNV	T>A (CAA) or T>G (CAG) at pos 3.
H95R	Awad 2021 clinical cohort	CAT	CGT (pos 2 A>G)	class-forced	T>C	SBS5 (clock-like)	Single-nt route via pos 2.
Y96C	Awad 2021 clinical cohort	TAT	TGT (pos 2 A>G)	class-forced	T>C	SBS5 (clock-like)	Chemistry-free rank 2 under sotorasib, Mutazyme-simulated rank 2 (near-zero shift).
Y96D	Tanaka 2021 case report	TAT	GAT (pos 1 T>G)	class-forced	T>G	low in common NSCLC sigs	First reported clinically in Tanaka 2021; chem-free rank 9 under sotorasib, Mutazyme-simulated rank 223 (shift +214 at $\beta = 1$) — class T>G is 4%–11% of ENU output and 3.6% of Mutazyme II.
Y96H	Awad 2021 *in vitro* DMS	TAT	CAT (pos 1 T>C)	class-forced	T>C	SBS5 (clock-like)	Single-nt route via pos 1.

Summary: 2 two-step SNV (R68L, V8M); 3 multi-class SNV (V8L, R68S, H95Q); 7 class-forced (C12F, R68M, H95D, H95R, Y96C, Y96D, Y96H).

S1b. ALK resistance panel (12 variants)

Codon numbering follows the ALK NM_004304.5 CDS cross-checked against the Oh 2026 Table S11 EML4-ALK v1 WT CDS. Reachability definitions as in S1a.

Variant	Source	WT codon	Target codon	Reachability	Class	Signature concordance	Notes
V1180L	Doi 2022 (alectinib)	GTT	CTT (pos 1 G>C)	class-forced	C>G	SBS13 at TpC; else rare	Chem-free rank 8, Mutazyme-simulated rank 15 under alectinib (class C>G at 5% in Mutazyme).
L1196M	Doi 2022 (alectinib)	CTG	ATG (pos 1 C>A)	class-forced	C>A	SBS4 (tobacco)	Gatekeeper. Rank 1 under both chem-free and Mutazyme-simulated rankings for alectinib.
I1171N	Doi 2022 (alectinib)	ATC	AAC (pos 2 T>A)	class-forced	T>A	low in common NSCLC sigs	Compound-mutation background for the Takei 2025 neladalkib fixed-V1 screen.
I1171T	Doi 2022 (alectinib)	ATC	ACC (pos 2 T>C)	class-forced	T>C	SBS5 (clock-like)	Class T>C — well-supplied by ENU and Mutazyme.
L1198F	Doi 2022 (alectinib)	CTC	TTC (pos 1 C>T)	class-forced	C>T	SBS1 / SBS2 / SBS5 (common)	Class C>T — well-supplied by Mutazyme II (47% of substitutions).
G1202R	Doi 2022 (alectinib)	GGA	CGA (C>G) / AGA (C>T)	multi-class SNV	—	multi-class SNV	Two single-nt routes to Arg via different classes; not class-forced.
G1269A	Takei 2025 (next-gen ALK)	GGA	GCA (pos 2 G>C)	class-forced	C>G	SBS13 at TpC; else rare	Single-nt route via pos 2.
F1174C	Takei 2025 (next-gen ALK)	TTC	TGC (pos 2 T>G)	class-forced	T>G	low in common NSCLC sigs	Pre-existing in a Lin 2024 ALKOVE-1 NVL-655 responder (cis with G1202R).
F1174L	Takei 2025 (next-gen ALK)	TTC	CTC / TTA / TTG	multi-class SNV	—	multi-class SNV	Three single-nt routes to Leu (classes T>C, C>A, C>G); not class-forced.
L1196Q	Takei 2025 (next-gen ALK)	CTG	CAG (pos 2 T>A)	class-forced	T>A	low in common NSCLC sigs	Compound-mutation partner reported with the gatekeeper residue.
C1156Y	Takei 2025 (next-gen ALK)	TGC	TAC (pos 2 G>A)	class-forced	C>T	SBS1 / SBS2 / SBS5 (common)	Class C>T — well-supplied by Mutazyme.
L1152R	Takei 2025 (next-gen ALK)	CTG	CGG (pos 2 T>G)	class-forced	T>G	low in common NSCLC sigs	Largest predicted negative rank shift under Mutazyme × alectinib (+558 at $\beta = 1$; +115 to +817 across $\beta \in [0.25, 4]$).

Summary: 10 class-forced (V1180L, L1196M, I1171N, I1171T, L1198F, G1269A, F1174C, L1196Q, C1156Y, L1152R); 2 multi-class SNV (G1202R, F1174L). No two-step variants in the ALK panel.

Methodological notes

Panel construction. The KRAS G12C panel is a *composite* of patient-observed variants (Awad 2021 clinical cohort; Tanaka 2021 case report) and preclinical *in vitro* DMS variants (Awad 2021 accompanying deep mutational scan). The ALK panel is drawn from Doi 2022 *Targeted Oncology* (six variants recovered in the alectinib Mutazyme II Ba/F3 screen) and Takei 2025 *Pharmaceutical Research* (six single-amino-acid variants from the next-generation ALK TKI screens). Panel membership is treated as fixed input to the analysis; the chemistry-free atlas rankings and class-forced-codon reachability classifications are derived.

Doi 2022 ALK variants. Doi 2022 reports resistance mutations in Figure 3a of the Mutazyme II × alectinib Ba/F3 screen without an explicit numeric ranking. The six Doi variants listed above are the single-amino-acid resistance mutations identified in Fig. 3a after excluding (i) silent passenger mutations (L1152 silent, L1165 silent, R1192 silent, all flagged in the Doi 2022 figure) and (ii) the I1171S sub-variant of the I1171N/T/S compound call (included in Doi but not in our panel because the three sub-variants cannot be distinguished at the position-level NGS readout; we retain I1171N and I1171T as the two independently reported sub-variants).

Takei 2025 ALK variants. The six Takei 2025 single-amino-acid variants listed above are drawn from the Takei *Pharmaceutical Research* manuscript's resistance panel across its alectinib, lorlatinib, zotizalkib, and neladalkib screens. The Takei 2025 fixed-V1 compound-mutation screen reports additional second-step partners (L1256F, E1129V, E1129K, C1156Y, T1151R selected at 500 nM neladalkib on an I1171N background); these are discussed in the main text as compound-mutation predictions and are not included in the single-amino-acid resistance panel for Figure 3.

Reachability classification. Reachability is determined from the Hamming distance between the wild-type codon (from the WT CDS cross-checked against ClinVar nucleotide-level records; see data/reference/codons.yaml and codons_kras.yaml) and the set of codons encoding the target amino acid. The classification is strictly a property of the genetic code and the wild-type codon; it does not depend on which mutagen chemistry a screen uses. The forced class, when a variant is class-forced, is the pyrimidine substitution class of the required single-nucleotide substitution: C>A, C>G, C>T, T>A, T>C, or T>G (all other substitutions are their complements and share a class with their pyrimidine representative).

Chemistry-free rankings. For KRAS G12C the chemistry-free ranking is the Z-score ranking of the 1,205 class-forced variants in the Feng 2022 Twist Biosciences synthetic saturation library under sotorasib or adagrasib. For ALK the chemistry-free ranking is $\sigma_{\text{drug}}(v) = \text{Resistance}(v)$ (the Oh 2026 atlas Resistance score, used without further transformation — see main-text Methods for the rationale, which turns on Oh's Resistance and Fitness scores being orthogonal axes rather than components to be subtracted) over the 1,259 class-forced variants in the Oh 2026 prime-editing saturation library under alectinib, lorlatinib, or zotizalkib. Per-variant scores and ranks are available in the derived data files `feng_awad2021_variant_ranks.csv` and `oh_decomposition_all_drugs.csv` in the paper's accompanying code repository.

Signature concordance column. Each class-forced variant's forced pyrimidine substitution class is annotated with the published COSMIC v3.3 single-base substitution (SBS) signatures (Alexandrov et al. *Nature* 2020;578:94–101, DOI 10.1038/s41586-020-1943-3) that emphasise that class, as a qualitative tumour-context tag. The mapping is: C>A → SBS4 (tobacco smoking); C>T → SBS1 (CpG deamination), SBS2 (APOBEC at TpC), SBS5 (clock-like unknown); C>G → SBS13 (APOBEC at TpC); T>C → SBS5; T>A and T>G → low in all common NSCLC signatures. The two cohort contexts of this paper are not symmetric: KRAS G12C is a predominantly smoking-associated NSCLC population (SBS4 dominates), so C>A variants are maximally tumour-reachable and C>G/T>G variants are relatively rare somatically; ALK-rearranged LUAD is predominantly a never-smoker population (SBS1 and SBS5 dominate), so C>T is the maximally tumour-reachable class and SBS4-emphasised C>A variants (such as L1196M) are rarer somatically than their resistance phenotype would suggest. Variants flagged "low in common NSCLC signatures" are predicted to be acquired at low frequency somatically in both cohorts, so a mutagen-chemistry blind-spot in those classes has lower clinical priority than a blind-spot in a tumour-reachable class — the *tumour-reachable* × *mutagen-unreachable* intersection is the operationally sharp blind-spot. The 96-channel context-aware version of this analysis, with explicit per-signature per-trinucleotide weights and a per-cohort patient-level reweighting, is deferred to a companion paper that requires GENIE patient-level 96-channel catalogues not integrated here.

Supplementary Table S2

Preclinical TKI-resistance screens published 2022–2026 and the mutagen chemistry each uses

The five studies cover both kinases of interest in the main text. The three most recent of them (industry neladalkib ENU, next-generation ALK error-prone PCR, KRAS G12C CBE+ABE) all employ chemistry-biased mutagens.

Study	Gene / drug(s)	Mutagen	Classes supplied
Nuvalent 2025	ALK (neladalkib)	ENU	5 of 6, biased
Takei 2025	ALK (alec/lor/zoti/nela)	error-prone PCR	6 of 6, biased
Diaz 2025	KRAS (ada, RMC-4998, RMC-7977)	CBE + ABE	2 of 6
Oh 2026	ALK (alec/lor/zoti)	prime editing	6 of 6, uniform
Feng 2022	KRAS G12C (sot/ada)	Twist synthetic	6 of 6, uniform

References for this table. Nuvalent 2025: Tangpeerachaikul et al. Takei 2025: Takei et al. Diaz 2025: Diaz et al. Oh 2026: Oh et al. Feng 2022: Feng et al. Full bibliographic details are given in the reference list of the main manuscript.