DISS. ETH NO. 29998

PRECISION GENOME EDITING SCREENS TO STUDY GENETIC VARIANTS OF UNCERTAIN SIGNIFICANCE

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES

(Dr. sc. ETH Zurich)

presented by
OLIVIER DOMINIQUE MARIO BELLI
M.Sc., Université de Strasbourg
D.Ing., Université de Strasbourg

born on 15.10.1993

accepted on the recommendation of
Pr. Randall J. Platt,
Pr. Gerald Schwank
Pr. Andreas Moor

2024

Acknowledgements

First and foremost, I would like to thank Pr. Randall Platt for welcoming me in his team and mentoring throughout my PhD journey. I am particularly grateful for his unwavering support and for allowing me to pursue different ideas, make my own decisions, go to scientific conferences and grow as a scientist. I learned a lot from his vision for science and his optimism.

I would also like to thank my thesis committee members, Pr. Gerald Schwank and Pr. Andreas Moor, for their insightful comments and recommendations during our yearly meetings. These conversations were always very stimulating and I am grateful to have had such inspiring scientists on my committee.

I would also like to extend my gratitude to my previous mentors for encouraging me to pursue my passion for science and for guiding me on the path to becoming a scientist myself. Thank you Michel Ribette, Daniel Lackner and Megan Van Overbeek for being the best role models I could have hoped for and for inspiring me throughout the years.

My gratitude also goes to all current and past Platt lab members. It has been amazing to see this team grow and be surrounded by such intelligent and dedicated scientists on a daily basis. I look forward to witnessing the remarkable accomplishments that each of you will achieve in the future. In particular, I would like to thank Masha, Florian, Raj and Maria for supporting me in difficult times and for the nice moments we shared outside the lab.

Kiki, thank you for your dedication and enthusiasm, it has been so much fun working with you in the past year. I am proud of what we achieved together and I am excited to see the bright future that lies ahead for you in science.

The past six years have been marked by moments of joy and challenging times. Throughout this uphill battle, I am grateful to have had the support of incredible friends who graciously overlooked my evening complaints and busy schedule.

Thank you Leo, Vincent and Arthur for the open-hearted late-night conversations and unwavering solidarity. Your support during difficult times has been invaluable.

Alice, Liore, Thibault, Bruno, we embarked on this journey together 10 years ago and I am proud to see what each of us has achieved. Thank you for all the good laughs and memories.

Selen, thank you for being there for me in my moments of doubts and for encouraging me to push through.

Jaicy and Claire, thank you for the fun moments we shared in Locarno and after, I will forever remember our Basecamp adventures.

Martin, Albert, Laetitia and François, thank you for the great times we shared playing music together. I always looked forward to our weekend sessions after a hard week in the lab and I am proud of what we created together.

Kristina, Coco, Simona and Nico, thank you for becoming my climbing partners but more importantly my trusted friends. Meeting you in the midst of Covid changed so much for me and I will cherish the memories we created together in the gym and outside.

Thank you Mathis and Aurélie for being my best friends throughout the years and for never letting me down, I am lucky to have you by my side.

Kenny, Eugénie, Renaud, Pierre-Yves, Yoann, Hélène, Martin, Clément, Thibault, Paul, Geoffrey, thank you for always reminding me of who I am and that I will always have a team somewhere. Life would not be as fun without you all.

Words cannot adequately convey my gratitude to my parents, Fabienne and Maurice, for their unwavering support, guidance, and for shaping the individual I am today. Their understanding, open-mindedness and encouragement have been invaluable to me throughout this journey. Livia, Serge, Nadine, Roseline and Patrice, thank you for everything you did for me throughout the years, I love you all.

Laura, thank you for being my partner in life and in science for nearly a decade. Thank you for your patience and understanding during this long adventure, it has been a bumpy road but I knew you always had my back. I am proud of what we accomplished together already and you inspire me to become a better person every day.

Finally, thank you Pocky, Mochi, Whisky, Trotsky and Gnocchi for being the coolest pets.

“This universe henceforth without a master seems to him neither sterile nor futile.
Each atom of that stone, each mineral flake of that night-filled mountain,
in itself forms a world.
The struggle itself toward the heights is enough to fill a man’s heart.
One must imagine Sisyphus happy.”

Albert Camus - The Myth of Sisyphus

Table of contents

Summary

Genetic tests play an increasingly pivotal role in predicting disease risks, prognosis and treatment response in the clinics. However, their interpretation is hindered by the prevalence of variants of uncertain significance (VUS) whose phenotypic effects remain elusive. Addressing this issue, multiplexed assays of variant effect (MAVEs) have been developed that enable the interrogation of protein functions and the assessment of variant pathogenicity. However, current workflows are limited in their ability to replicate variant diversity, genomic background, expression levels and cellular context. Here, we leveraged and combined multiple precision genome editing technologies, including cytosine and adenine base editors along with a prime editor, to assess the pathogenicity of a broad spectrum of variants in the genome of cells. We applied our multimodal approach to interrogate EGFR, uncovering pathogenic genetic variants driving oncogenesis and drug resistance. We identified known and novel hits, thus supporting the performance of the screening approach and providing new insights into underappreciated routes to EGFR activation and drug response. In the future, we envision that multimodal precision mutational scanning screens can be widely employed to comprehensively characterize genetic variation down to single nucleotide resolution across diverse genomic backgrounds, thereby contributing to the development of personalized treatment protocols.

Résumé

Les tests génétiques occupent un rôle de premier plan dans la prédiction des prédispositions individuelles aux maladies, de leur évolution clinique et de la réponse aux traitements. Toutefois, leur interprétation est entravée par la prévalence de variants de signification incertaine dont les effets phénotypiques sont actuellement inconnus. Afin de pallier cette problématique, des tests multiplex d’évaluation de l'impact des variants ont été développés, qui permettent d'interroger les fonctions protéiques et d'évaluer leur pathogénicité. Cependant, les protocoles actuels sont limités dans leur capacité à reproduire la diversité, les niveaux d'expression et les contextes génomiques et cellulaires des variants en question. Dans cette thèse, nous avons mis à profit et combiné plusieurs technologies d'édition du génome comme base et prime editing, afin d'évaluer la pathogénicité d'un large spectre de variants dans le génome de cellules en culture. Nous avons appliqué cette approche multimodale à l’étude du récepteur du facteur de croissance épidermique (EGFR), révélant ainsi de nouveaux variants pathogènes impliqués dans l'oncogenèse et la résistance aux traitements. Nos résultats mettent à la fois en lumière des variants connus et inconnus, confirmant la performance de notre approche de criblage et fournissant de nouvelles informations sur différents modes d'activation et de réponse aux traitements de l'EGFR. À l'avenir, nous anticipons que des cribles mutationnels multimodaux de ce type pourront être utilisés à grande échelle pour caractériser des variations génétiques de manière exhaustive dans divers contextes génomiques, contribuant ainsi à l'élaboration de protocoles de traitement personnalisés.

Abbreviations

ABE	Adenine base editor
ACMG	American College of Medical Genetics
AEV	Avian erythroblastosis virus
AKT	Protein kinase B
AREG	Amphiregulin
ATP	Adenosine triphosphate
AUC	Area under the curve
AXL	AXL Receptor Tyrosine Kinase
BRCA1/2	BRCA1/2 DNA Repair Associated
CAMK2	Calcium/Calmodulin Dependent Protein Kinase II
CBE	Cytosine base editor
CBL	Cbl Proto-Oncogene
CCND1	Cyclin D1
CCNE1	Cyclin E1
CDK6	Cyclin Dependent Kinase 6
CDKN2A	Cyclin Dependent Kinase Inhibitor 2A
CFH	Complement Factor H
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
dCas9	Dead Cas9
DDX3X	DEAD-Box Helicase 3 X-Linked
DMEM	Dulbecco's Modified Eagle Medium
DMNT3A	DNA Methyltransferase 3 Alpha
DMS	Deep mutational scanning
DNA	Deoxyribonucleic acid
DSB	Double strand break
dsDNA	Double stranded DNA
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
EMT	Epithelial-to-mesenchymal transition

epegRNA	Engineered prime editing guide RNA
ERK	Mitogen-Activated Protein Kinase
ERRFI1	ERBB Receptor Feedback Inhibitor 1
FACS	Fluorescence-activated cell sorting
FBS	Foetal Bovine Serum
FDR	False discovery rate
GBM	Glioblastoma multiforme
GFP	Green fluorescent protein
GOF	Gain of function
GRB2	Growth Factor Receptor Bound Protein 2
GWAS	Genome-wide association study
HBEGF	Heparin-binding EGF-like growth factor
HDR	Homology directed repair
HER2	Human epidermal growth factor receptor 2
hGH	Human growth hormone
HNSCC	Head and neck squamous cell carcinoma
IC50	Half-maximal inhibitory concentration
IgG	Immunoglobulin G
LFC	Log fold change
LOF	Loss of function
LRIG1	Leucine Rich Repeats And Immunoglobulin Like Domains 1
LTR	Long terminal repeat
LTR	Long terminal repeat
mAb	Monoclonal antibody
MAVE	Multiplexed assays of variant effect
MEK	Mitogen-Activated Protein Kinase Kinase
mESC	Mouse embryonic stem cells
MLH1	MutL Homolog 1
MMEJ	Microhomology-mediated end joining
MMLV	Moloney murine leukemia virus
MOI	Multiplicity of infection
MPRA	Massively parallel reporter assays
nCas9	nickase Cas9

NGS	Next generation sequencing
ngRNA	Nicking guide RNA
NHEJ	Non homologous end joining
NLS	Nuclear localization signal
NPC1	Niemann-Pick disease, type C1
NSCLC	Non-small cell lung cancer
PAM	Protospacer-adjacent motif
PARP	Poly (ADP-ribose) polymerase
PBS	Primer binding site
PCR	Polymerase Chain Reaction
PE	Prime editing
pegRNA	Prime editing guide RNA
PharmGKB	Pharmacogenomics Knowledgebase
PI3K	Phosphoinositide 3-kinase
PIP2	Phosphatidylinositol (4,5)-bisphosphate
PIP3	Phosphatidylinositol (3,4,5)-trisphosphate
PKC	Protein kinase C
PLC	Phospholipase C
PTB	Phosphotyrosine binding
PTEN	Phosphatase And Tensin Homolog
PTPRJ	Protein Tyrosine Phosphatase Receptor Type J
RAF	Proto-Oncogene, Serine/Threonine Kinase
RAS	RAS Proto-Oncogene, GTPase
RNA	Ribonucleic acid
RPMI	Roswell Park Memorial Institute medium
RT	Reverse transcriptase
RTK	Receptor tyrosine kinase
RTT	Reverse transcriptase template
SAMHD1	SAM And HD Domain Containing Deoxynucleoside Triphosphate Triphosphohydrolase 1
SCLC	Small cell lung cancer
scRNA-seq	Single-cell RNA sequencing
SFDA	Chinese National Medical Products Administration
sgRNA	Single guide RNA

SH2	Src homology 2
SHC	SHC Adaptor Protein
SNP	Single-nucleotide polymorphism
SNV	Single nucleotide variants
SOCS4/5	Suppressor Of Cytokine Signaling 4/5
SOS	SOS Ras/Rac Guanine Nucleotide Exchange Factor
SRC	SRC Proto-Oncogene, Non-Receptor Tyrosine Kinase
STAT	Signal Transducer And Activator Of Transcription
TALEN	Transcription Activator-Like Effector Nucleases
TGF-α	Transforming growth factor alpha
TKI	Tyrosine kinase inhibitors
TP53	Tumor Protein P53
tracrRNA	Trans-activating crRNA
USFDA	US food and drug administration
VUS	Variant of uncertain significance
WGS	Whole genome sequencing
WPRE	Woodchuck Hepatitis Virus posttranscriptional regulatory element
WT	Wild type
ZFN	Zinc finger nucleases

Chapter I: Introduction

1.1 Genetic variants in human diseases

1.1.1 Discovery of disease-associated genetic variants

Genetic variation is a fundamental factor shaping disease susceptibility, severity, prognosis and therapeutic outcomes. So far, approximately 800 million single nucleotide variants (SNV) and 1.2 million structural variants have been identified (1) and each human genome is estimated to contain approximately 4 million genetic variants, including 10,000 non-synonymous SNVs (2). This diversity, along with the contribution of numerous variants with minor phenotypic effects to complex traits, presents a significant challenge in comprehending the role of genetic variants in human diseases.

The first efforts to link genotypes and disease phenotypes relied on forward genetics approaches that start from a measurable phenotype and aim to identify genetic features that are specific to affected individuals. In humans, these initially focused on rare monogenic diseases with Mendelian inheritance patterns. In these cases, the study of family trees allows for the identification of genetic markers that specifically co-segregate with the trait of interest within a family due to genetic linkage (3,4). These markers initially consisted in repetitive genomic regions which have the advantages of being found throughout the genome and to be highly polymorphic in lengths, facilitating their probing by Polymerase Chain Reaction (PCR) or restriction fragment length assays (5–7). As a result, genetic markers, while potentially distinct from the genomic loci causing a particular disease, offered a first opportunity for the development of molecular diagnostics tools. Their identification also helped to narrow down the chromosomal localisation of causal genes, eventually leading to their discovery as in the cases of Duchenne muscular dystrophy (8,9) and cystic fibrosis (10–12).

Cosegregation studies using genetic markers are useful for the investigation of highly penetrant monogenic diseases. However, they are laborious and confined to single genomic loci, making them irrelevant in diseases that arise from a non-linear combination of variants with small individual effects (2). These limitations were partly addressed by the publication of the first draft of the human genome (13,14) as well the development of exome sequencing (15,16) and array-based hybridization assays (17–19). Indeed, the rapid advancement of these technologies facilitated the identification of structural and single nucleotide variations in patients on a genome-wide scale, enabling the transition from linkage studies to association studies.

1.1.2 Genome-wide association studies

The first genome-wide association study (GWAS) was published in 2015 by Klein et al. (20). These studies typically compare genetic variant profiles between disease-affected and control patient groups, thus enabling the discovery of mutations associated with complex traits (21). Because GWAS do not require multiple affected individuals from the same family, they also allow for much larger cohorts and enable the exploration of polygenic diseases for which individual mutations have a smaller contribution to the overall phenotype (22). Additionally, the use of whole genome sequencing (WGS) data or single-nucleotide polymorphism (SNP) arrays enables the discovery of genetic variants in coding and non-coding regions, irrespective of previous knowledge about their function. Consequently, approximately 6000 GWAS have been documented since 2005, reporting approximately 400,000 associations with 88% of these involving non-coding loci (23). These data can then serve as a basis to investigate gene functions or to identify genetic risk markers for diseases. For example, the first GWAS identified a SNP in the Complement Factor H (CFH) gene that increases the risk of age-related macular degeneration by 7.4 fold in homozygous individuals (24). In addition, GWAS data can help to rationally design treatment protocols as demonstrated by the identification of coding and non-coding variants of the IL28B gene that predict patient response to treatments against hepatitis C infection (25).

In spite of their popularity, a limitation of GWAS lies in their inability to infer genetic causality for the considered phenotype or, if such causality exists, the underlying mechanism (26). Indeed, common variants might be associated with a trait without contributing to it, simply because they are in linkage imbalance with a biologically relevant locus. For example, the previously mentioned study by Klein et al. initially identified a common intronic SNP associated with macular degeneration before uncovering the causal coding variant by resequencing patient genes (24). Another challenge pertains to the interpretation of the results derived from these studies as GWAS are particularly efficient at capturing common polymorphisms with small contributions to phenotype (27). For example, associations have been found in existing drug target genes with high biological significance where individual SNPs explain less than 1% of the phenotypic variation within the population (28). Similarly, it was estimated that 20 SNPs identified in a GWAS contributed to only 3% of the population variation in height, suggesting that 93,000 individual SNPs would be required to explain 80% of this variation (29). This high sensitivity thus comes at the cost of generating datasets from which it is difficult to extract clinically pertinent information. By contrast, rare variants that are more likely to have a stronger phenotypic impact and provide meaningful biological information can be missed by GWAS due to insufficient prevalence in the studied population or the relyance on SNP arrays, warranting the integration of WGS data on a much broader scale (27).

In conclusion, while GWAS are powerful tools to identify disease-associated genetic loci, including previously unknown regulatory elements, the many hits they generate need to be thoroughly characterized and validated through functional assays to be leveraged as predictive markers or therapeutic targets in the clinics.

1.1.3 Variant databases

Genomic data obtained from large scale sequencing or genotyping studies are typically compared to a reference genome to call variants. This reference is based on a haploid genome assembled from WGS data aggregated across a limited number of donors and thus fails to capture the diversity of actual human genomes (30). The standardization of clinical genetic tests and the emergence of GWAS thus motivated the creation of human genetic variations databases. A prominent example is the genome aggregation database (gnomAD, previously exome aggregation consortium), which currently comprises more than 730,000 human exomes and 76,000 full genomes (1). GnomAD is meant as a reference genome database sourced from individuals around the world that are not affected by pediatric diseases. This extensive repository contributes not only to the understanding of genetic history across human populations but also to the cataloging of genetic variants (31,32), allowing for the identification of genes and non-coding regions under mutational constraint.

Indeed, SNVs are found on average once every 4.9 nucleotides in the gnomAD dataset and each human genome contains approximately 100 loss-of-function (LOF) alleles (33). Genes harboring a lower-than-expected missense and nonsense variant frequency are thus likely subjected to significant natural selection and involved in essential biological functions (1,34,35). Following the same reasoning, gnomAD also contributes to the clinical interpretation of genetic variants as deleterious mutations tend to be depleted at the population scale, indicating that common variants are less likely to be pathogenic. The comprehensive identification of these variations across different human populations can therefore advance more equitable diagnostics.

Adding to the gnomAD effort to provide a representative set of reference genomes, other databases like the UK Biobank focus on combining genotypic and phenotypic data to enable association studies (36). The database thus provides genotyping information, later supplemented with WGS data, along with a broad array of biochemical, clinical, socioeconomic, and lifestyle information about each participant (37). Despite their goal of representing human genetic diversity, these initiatives face challenges related to sampling bias, as illustrated by the presence of age, gender, health and socioeconomic background discrepancies between UK Biobank volunteers and the general population (38).

In addition to population databases, the recent development of genetic testing in the clinics has caused a notable increase in the volume of data obtained from patients affected by various diseases (39). These are aggregated in databases like ClinVar which provide a standardized reporting framework for genetic variants submitted by clinical testing and research laboratories (40). This allows these data to be scrutinized and consolidated by research initiatives like the BRCA Exchange which coordinates the annotations of variants in the BRCA1 and BRCA2 genes by expert panels (41). Similarly, the Catalogue of Somatic Mutations in Cancer (COSMIC) regroups genotyping and WGS data from diverse tumor samples, including information on drug response and zygosity (42).

1.1.4 Challenges in variant classification

Beyond variant cataloging, ClinVar also provides information on their clinical significance following a classification system established by the American College of Medical Genetics (ACMG) (43). Clinical, functional and *in silico* prediction data thus contribute to assessing the likelihood of pathogenicity for a variant, ultimately resulting in its classification as benign, likely benign, uncertain significance, likely pathogenic, or pathogenic. However, determining the relative importance of diverse data types, or the lack thereof, for the same variant can prove challenging and as genetic testing becomes more prevalent in the clinics, the accurate classification of variant clinical impacts is falling behind. Indeed, the fraction of variants of uncertain significance (VUS) in ClinVar has been steadily increasing over time to currently represent 48% of the database while 22.9% of variants with more than one report have conflicting interpretations of pathogenicity (39,44,45) (Figure 1.1).

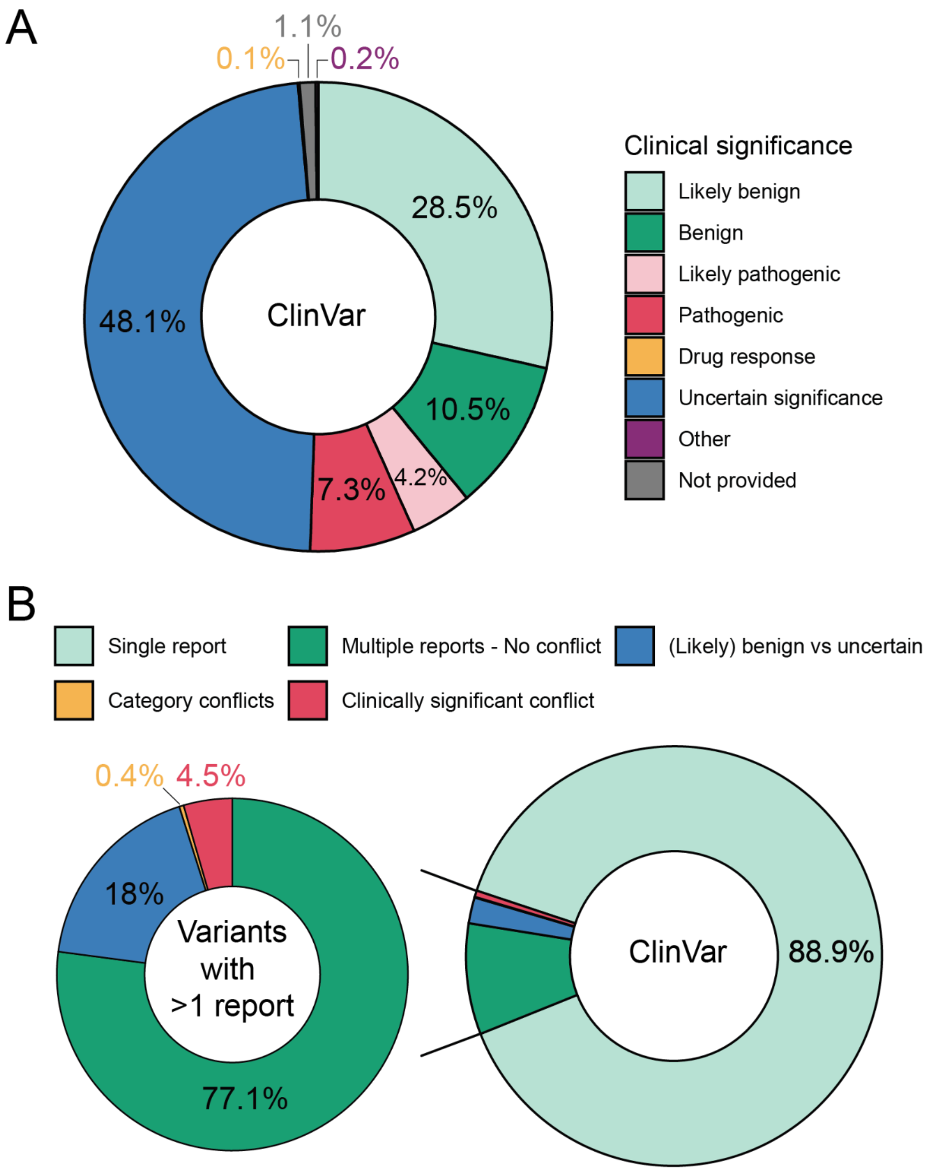


Figure 1.1: **Current classification of variants in ClinVar.** *(A) Reported clinical significance of all variants in the ClinVar database (2 657 180 variants). (B) Proportion of ClinVar variants with conflicting reports of pathogenicity. Category conflicts: e.g. benign vs risk factor, clinically significant conflict: e.g. benign vs pathogenic. Variants with a synonymous conflict (e.g. benign vs non-pathogenic) are ignored. Data obtained from ClinVar Miner* (39) *on 19/12/2023.*

The classification of newly identified variants as VUS may arise from scarce population data, their localisation in non-coding regions with unknown functions, or their co-occurrence with other variants, thereby complicating the establishment of a causal link with the associated disease (46). Additionally, different lines of evidence like residue conservation, variant frequency, family history and functional assays can be difficult to integrate into a unique pathogenicity score (47,48). The accumulation of VUS is particularly prevalent in cancer-associated genes because standardized genotyping panels can include more than a hundred genes at once (49). As a result, the number of VUS identified within a given gene has been shown to correlate with the number of registered clinical tests encompassing it (50).

Adding to this challenge is the observation that pathogenic variants can have incomplete penetrance and that their impact may largely depend on co-occurring mutations in the same or in other genes (51). For example, LOF mutations in BRCA1 and 2 are known to drastically increase the risk of developing breast and ovarian cancers. However, a study using data from the UK Biobank determined that the risk for women harboring a BRCA1/2 pathogenic variant to develop breast cancer ranged from 12.7% to 75.7% depending on their polygenic background (52). Similarly, mutations in the tumor suppressor gene PTEN are associated with a broad range of disease severity stemming from subtle variant effects on protein function, a phenomenon termed genetic quasi-sufficiency (53). This spectrum of variant penetrance and severity can thus complicate their categorization into discrete clinical relevance categories which do not integrate important parameters like zygosity, penetrance and polygenic background.

An additional challenge in variant classification arises from inherent biases in genomics data used in reference databases and GWAS. Indeed, these have been largely skewed towards individuals of European ancestry, resulting in existing polygenic scores being less predictive in patients of other ethnic backgrounds (54). Similarly, commercial SNP arrays were shown to be biased towards variants that are common in Europe, thus missing potentially significant variations in other populations and in turn biasing variant reporting in ClinVar (55). Because common variants are generally considered as less likely to be pathogenic, the absence of population data from specific ethnicities may result in the misclassification of variants that are common within these groups but not represented in reference databases. Encouragingly, this knowledge gap appears to be narrowing down as more diverse genomic data are included in gnomAD, leading to variant reclassification (56).

In conclusion, genetic testing has the potential to improve therapeutic outcomes by allowing for precise patient profiling and targeted treatment courses (57). However, the growing prevalence of VUS in variant databases paradoxically complicates the use of deep sequencing data and leads to heightened uncertainty for patients and clinicians (46). Indeed, a recent study reported that VUS affecting BRCA1/2 were found in 7.7% of breast or ovarian cancer patients (58). This not only represents a diagnostic challenge for physicians but can also lead to loss of therapeutic opportunities as genetic testing results affect access to approved drugs or clinical trials (59). In parallel, VUS can also be a source of psychological distress for patients and lead to unnecessary life-altering treatments like preventive mastectomies in the case of BRCA1/2 germline variants (60,61). As the number of VUS steadily increases, the development of new approaches for their comprehensive and accurate assessment in diverse genetic backgrounds is thus urgently needed to allow for equitable access to treatment.

1.2 Mutagenesis approaches to study genetics variants

1.2.1 Random mutagenesis

Forward or phenotype-driven genetics studies in humans require many individuals to discover genetic variants that are rare or have small effects. Even in model organisms with smaller genomes, naturally occurring mutations are infrequent and do not allow for the comprehensive assessment of gene functions. In this context, random mutagenesis approaches have facilitated the creation of extensive mutant collections, enabling the mapping of gene functions and interactions through mating and linkage analyses.

A first example of this approach was demonstrated by Hermann Muller, for which he received the Nobel prize in 1946 (62). After his initial efforts to increase mutation rates in fruit flies (*Drosophila Melanogaster*) with temperature proved to be inconclusive, he reported in 1924 that exposing germ cells to X-rays helped to rapidly obtain offsprings with diverse phenotypes (63). Another example is the isolation of lambda phage mutants by UV irradiation of lysogenic bacteria carried out by François Jacob's laboratory in the 1950s, which helped to identify regions of the phage genome involved in lysogeny (64,65). Similar approaches using chemical mutagenesis were later used in nematodes (*Caenorhabditis elegans*) (66), zebrafish (*Danio rerio*) (67), and mouse (*Mus Musculus*) (68).

One advantage of inducing mutations through physical or chemical methods is that the mutation rate can generally be finely adjusted to the size of the studied genomes. However, the discovery of recessive variants in diploid genomes is impossible with these methods as the probability of introducing the same mutation in both alleles is negligible. Random mutagenesis screens were thus extensively used in yeasts with haploid life cycles (69) and later in haploid mouse embryonic stem cells (mESC) (70). In this second example, cells were exposed to a DNA-alkylating agent before being treated with the 6-thioguanine anti-cancer drug. This resulted in the emergence of drug-resistant clones whose number was shown to be dependent on the dose of mutagenic agent applied to the cells. Whole-exome sequencing of resistant clones revealed an average of 40 non-synonymous coding mutations per clone, allowing the authors to identify candidate genes mutated in multiple clones and likely involved in drug sensitivity.

While allowing for unbiased genome scanning, the uncontrolled character of mutations introduced in random mutagenesis screens requires the isolation and whole genome sequencing of multiple individuals or cellular clones to identify phenotype-associated mutations. Partially addressing this limitation was the development of insertion screens which rely on the random integration of transposons throughout the genome to disrupt coding regions. These transposons can be engineered to integrate reporter genes, thus facilitating the pre-selection of clones with integration sites within coding sequences (71).

In conclusion, random mutagenesis has been instrumental in the characterisation of many genotype-phenotype relationships and gene networks by producing organisms with diverse phenotypes. It is however limited to phenotype-driven approaches, the identification of LOF mutations and does not allow for the controlled targeting of candidate genes.

1.2.3 Multiplexed assays of variant effects

Recent advances in DNA sequencing (72), high-throughput DNA synthesis (73) and gene editing allowed for the implementation of multiplexed assays of variant effects (MAVEs). These aim to replicate genetic variations of interest in *in vitro* or *in vivo* models to assess their phenotypic impacts in a functional assay. Unlike random mutagenesis, these assays start from a specific coding or non-coding mutation to assess associated phenotypes, thus constituting reverse genetics approaches.

MAVEs generally follow a common structure (74):

1. The construction of a library of genetic variants and its delivery to a cellular system. The variants can be introduced in an endogenous gene or genomically-integrated synthetic reporter system by gene editing or expressed from an exogenous vector.
2. A phenotypic selection step. Examples include drug resistance, cell proliferation or fluorescence-activated cell sorting (FACS).
3. The quantification of the genetic variant library. This usually involves the deep sequencing of the variants themselves or of molecular barcodes associated with a reporter.
4. The assignment of a functional score measuring the phenotypic effect of each variant in relation to the wild-type sequence.

These approaches can thus be evaluated according to 1) their ability to faithfully reflect the genetic context of the studied gene, 2) the biological relevance of the reporter assay which impacts its specificity and sensitivity, 3) their throughput and redundancy which in turn impact statistical power and cost (50,75).

Because MAVEs assess genetic variants in a controlled fashion, they can provide rich information on genetic variant impacts on transcriptional activity (76), mRNA stability (77), protein function (78), protein stability (79) and protein interactions (80), among others. Importantly, clinical interpretations cannot be directly extrapolated from functional data which should thus be reported in terms of functionally “normal” or “abnormal” variants instead of “benign” or “pathogenic” (81). In spite of this, MAVE results from established functional assays are considered by the American College of Medical Genetics as providing strong evidence in the interpretation of genetic variants, alongside co-segregation and population genetics data (43). MAVE results have thus been used in combination with clinical data to reclassify VUS in cancer-associated genes like PTEN, TP53 and BRCA1 (44).

1.2.4 Deep mutational scanning

A prevalent approach to MAVEs consists in the construction of a variant library of the considered genetic element in a plasmid vector. These can be obtained by pooled oligonucleotide assembly (82) or by amplifying the wild-type sequence with an error-prone DNA polymerase (83) or using degenerate (84) or codon-swapping primers (85).

Early studies of protein function used alanine scanning which replaces each residue in a protein of interest with alanines to evaluate their functional relevance (86). Alanine represents an interesting substitute because it is an abundant amino acid with a minimal side chain that does not impact protein main-chain conformation unlike glycine or proline residues. In a seminal study published in 1989, authors used this approach to systematically mutate amino acids in three domains of the human growth hormone (hGH) and identified residues involved in the binding to its receptor.

Alanine scanning was then progressively replaced by deep mutational scans (DMS) which comprehensively assess each possible amino acid substitution in a given protein sequence (87). The resulting library can be used in protein display assays (88) or delivered to reporter cells by plasmid transfection or viral transduction. In this case, a cell should ideally receive a single library element in order to limit experimental noise. In mammalian cells, this could be achieved by viral transduction at a low multiplicity of infection (MOI) (89) or by recombinase-mediated insertion in a synthetic genomic landing pad (90,91).

Many protein-coding genes and applications were explored through the use of DMS. Examples include antibody affinity optimization in mammalian cell display experiments (92) and the investigation of membrane protein variant effect on viral fitness (93). Similar approaches were also used to study non-coding variants in transcriptional enhancers (94) or intronic determinants of alternative splicing (95) in massively parallel reporter assays (MPRA). These use barcoded reporter transcripts or fluorescent proteins whose transcription is affected by the considered genetic element to quantify their transcriptional impact.

The DMS approach presents the significant advantage of allowing for very high-throughput screens, thus enabling the systematic assessment of every possible substitution in a protein sequence. Indeed, the diversity and scale of a DMS experiment are theoretically only limited by the library cloning approach and the number of cells that could be transfected or transduced. However, these screens fail to replicate the genomic context and endogenous expression levels of the considered gene. DMS libraries are indeed derived from protein cDNAs lacking UTRs, splice sites and introns and expressed from synthetic constructs using strong promoters. Furthermore, these studies often require cellular models that do not endogenously express the considered gene and are thus generally conducted in cell lines with low biological relevance like yeasts (96), HEK293 landing pad (97) or engineered knock-out cell lines (89).

1.2.5 CRISPR-Cas9

A major addition to the MAVE toolbox was brought by targeted nuclease technologies like Transcription Activator-Like Effector Nucleases (TALEN) (98) and Zinc Finger Nucleases (ZFN) (99) which enable targeted genome editing in mammalian cells. However, both rely on modular protein components that have to be engineered to define each target site, thereby restricting their multiplexing capabilities. This was addressed by the harnessing of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and associated nucleases in mammalian cells (100,101).

The first of these systems to be characterized was the Cas9 enzyme from *Streptococcus Pyogenes* (SpCas9) associated with a CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA), later fused into a single guide RNA component (sgRNA) (100). Cas9 recognizes this sgRNA, which consists of a 20-nucleotide variable spacer followed by a constant RNA scaffold, to assemble into a ribonucleoprotein. This complex has the ability to scan long stretches of double-stranded DNA (dsDNA) to selectively recognize and bind a sequence complementary to the spacer and followed by a short trinucleotide sequence termed Protospacer Adjacent Motif (PAM) corresponding to NGG for SpCas9. Cas9/sgRNA binding to their target leads to the formation of a DNA R-loop and the activation of the two Cas9 nuclease domains, leading to the introduction of a double strand break (DSB) in the target DNA (102). In mammalian cells, DNA breaks generally activate error-prone DNA repair pathways such as Non Homologous End Joining (NHEJ) (103) or Microhomology-Mediated End Joining (MMEJ) (104), often resulting in the introduction of insertions or deletions (indels) that can lead to frameshifts in coding sequences. In contrast to TALENs and ZFNs, the attractiveness of CRISPR-Cas9 systems thus lies in their versatility as they can be easily retargeted to any genomic site with an available PAM by substituting the sgRNA component.

As a result, the ability of CRISPR-Cas9 to specifically target and cleave endogenous genomic sites has been used extensively to study the effect of gene knock-outs at scale (101). Indeed, the modularity of the CRISPR-Cas9 system allows for sgRNAs to be delivered as a pool to a population of cells expressing Cas9 to introduce thousands of gene knock-outs in parallel. The resulting heterogeneous cell population can then be subjected to phenotypic selection before quantifying sgRNA abundances by deep sequencing and calculating their relative enrichment or depletions between different conditions (105–107). This approach, termed CRISPR screening, allows for diploid knock-outs and deliberate library design unlike random mutagenesis and transposase-mediated random insertion screens (71). It also benefits from lower off-targets compared to previous RNA interference screens (108). As a result, CRISPR screens have been used both *in vitro* and *in vivo* to uncover the genetic bases of many phenotypes at the genome scale like cell viability, drug resistance (109,110), toxin resistance (111,112), enhancer activity (113), cell signaling (114), tumor growth (115,116) and cell differentiation (117).

Although optimized CRISPR screening libraries benefit from high specificity and efficiency at the genome scale (118), concerns have been raised about unexpected large deletions and chromosomal rearrangements at Cas9-induced DSBs (119–121). These could be mitigated by the use of CRISPR interference (CRISPRi) and activation (CRISPRa) screens which use catalytically dead Cas9 (dCas9) fused to transcriptional repressors (122) or activators (123,124) protein domains for transcriptional modulation without the introduction DNA DSB (118,125).

Interestingly, while the predominant focus of CRISPR screens utilizing wild-type Cas9 has been on gene knock-outs, the technology was also leveraged as a random mutagenesis tool. A 2018 study thus reported the identification of drug target genes by using Cas9 to generate in-frame gain-of-function mutations leading to drug resistance in cancer cells (126). However, this approach is largely limited by the availability of PAM sequences and the restricted spectrum of in-frame mutations introduced by NHEJ and MMEJ at a specific DSB. Notably, the study only targets essential genes where out-of-frame mutations would lead to cell death, which constitutes an intrinsic enrichment mechanism for in-frame mutations.

1.2.6 Homology directed repair

In addition to random indels, Cas9-induced DSBs can also be leveraged to introduce precise DNA mutations through homology directed repair (HDR) (101). This can be done by providing cells with a single or double stranded DNA donor homologous to the target site and containing the desired edit. After introduction of a DSB, the donor is used by the cellular DNA repair machinery as a template to precisely repair the break and can thus seamlessly introduce any type of short mutation.

In a screening context, a single stranded donor library is transfected in cells along with a matched sgRNA to introduce mutations in proximity to the target site. This approach has been used for the high-throughput screening of DDX3X and BRCA1 gene variants in HAP1 cells (127,128). This human cell line was derived from partly haploid chronic myeloid leukemia cells which were then further engineered into the fully haploid eHAP line (129). These represent an attractive model for genomic screening as targeting a single allele reduces the heterogeneity of the edited cell population while potentially capturing phenotypes from both dominant and recessive alleles. In the two previously mentioned studies, this allowed for the efficient identification of LOF mutations leading to apoptosis as no allelic compensation was possible. For the same reason, sgRNA-expressing plasmids and repair donors were delivered by transfection instead of viral transduction as a single edit can be installed in the cellular genome regardless of the number of HDR donors present in the cell.

Unlike DMS studies, HDR screens have the advantage of introducing mutations directly in the genome of cells, thus preserving endogenous regulations and expression levels. However, unlike NHEJ which is active throughout the cell cycle, HDR is restricted to the S and G2 phases, thus limiting its use to mitotic cells (130). Furthermore, HDR suffers from a high rate of unwanted indels at the Cas9 cut site because of the engagement of other repair pathways such as NHEJ or MMEJ (131,132). This drastically increases the background noise of a screen and the number of cells that need to be edited to capture enough successful edits. Consequently, various methodologies have been devised to enhance the efficiency and accuracy of HDR through the inhibition of the NHEJ pathway, cell synchronization, chemical modification of the donor, fusion of repair factors or tethering of the donor DNA to Cas9 (133–137). However, none of these have been applied in a screening context so far.

An additional limitation of HDR screens is the requirement to match each donor to an sgRNA delivered separately. This prevents the use of this approach to cover full mammalian genes as an HDR donor library can generally only edit about 200 nucleotides around a single cut site. Similarly, the edits cannot be quantified through a synthetic proxy like genomically integrated sgRNAs in CRISPR knock-out screens. Successful edits thus need to be quantified by deep sequencing of each target site, which has the advantage of discerning precise and unwanted edits but limits the scale of these screens to single exons.

Although few HDR screens have been reported, a study by Findlay et al. (2018) has allowed the reevaluation of BRCA1 variants found in cancer patients (138). Functional scores obtained in the screen were combined with MPRA data from another study, which allowed for the retrospective reclassification of three VUS found in Korean patients into likely pathogenic variants. This holds significance as BRCA1 status is a biological marker predicting the response of cancer patients to poly (ADP-ribose) polymerase (PARP) inhibitor therapies. In spite of their technical limitations, this thus demonstrates the usefulness of Cas9-based HDR screens to help guide therapeutic decisions.

1.2.7 Base editing

The primary limitation of HDR is the need to deliver two distinct components to encode an edit. This was addressed by the development of base editors that rely on a Cas9^{D10A} mutant, termed nickase Cas9 (nCas9), fused to a deaminase domain to introduce base transitions (139,140) or transversions (141–143). The binding of nCas9 to its target leads to the formation of an R-loop, allowing the deaminase to access nucleotides within the non-target strand. In the case of adenine base editors (ABE), this leads to the conversion of adenines to inosines while cytosine base editors (CBE) catalyze the conversion of cytosines to uridines. As a result of its impaired RuvC nuclease domain, nCas9 exclusively generates single-stranded breaks in the guide-complementary strand which biases DNA repair pathways like mismatch repair (MMR) towards the unedited strand. Inosine and uridine are then respectively interpreted as thymine and guanine during DNA repair or replication, allowing for the permanent introduction of C•G to T•A or A•T to G•C edits in both DNA strands.

Base editing has been utilized in a screening context and was shown to accurately identify pathogenic variants in BRCA1, BRCA2 and DNMT3A (144–147). Very recently, a study also demonstrated its use beyond gene-centric applications by screening thousands of cancer-associated genes for cysteins that could constitute drug targets (148). A major improvement of base editors compared to HDR is the absence of DSB which not only limits the introduction of unwanted indels but also mitigates their cellular toxicity compared to that of wild-type Cas9 (149,150). This is particularly relevant in cells with wild-type TP53 such as pluripotent stem cells (151,152) and has recently allowed for the large-scale screening of variants affecting hematopoiesis in primary hematopoietic stem cells (153). Notably, this study also leveraged single-cell RNA sequencing (scRNA-seq) to simultaneously measure cell phenotype and infer introduced edits from the genome-integrated sgRNA.

A significant limitation of these screens pertains to the fact that each base editor only introduces a single type of base conversion, thereby restricting the range of variants assessed in a screen. In addition, each enzyme introduces mutations within an intrinsic editing window of varying sizes, making it challenging to predict the ratios of desired and bystander edits within this window (154,155). This may lead to a disconnect between phenotype and the genotype inferred by sgRNA library sequencing which complicates screen interpretation and validation. Lastly, early versions of ABEs and CBEs have been shown to introduce widespread Cas9-independent base editing in cellular RNAs (156–158). Although these have been mitigated through protein engineering in newer base editors, the biological impact of these off-target effects is currently unknown.

In spite of their limitations, base editors represent powerful tools for variant screening, owing to their high efficiency and scalability as compared to previous gene editing methods. Base editing enzymes that introduce each type of base substitution are now available, constituting a rich toolbox for gene-targeted or genome-wide studies. Interestingly, the ability of base editors to introduce DNA edits without a nucleotide-based donor template recently inspired the development of TALE-based protein-only base editors. These can be targeted to mitochondria to edit mtDNA in cells, thus potentially paving the way to mitochondrial variant screening in the future (159–161).

1.2.8 Prime editing

Prime editing constitutes the latest addition to the CRISPR toolbox. It uses a Cas9^{H840A} nickase fused to the engineered retro-transcriptase of the Moloney murine leukemia virus (M-MLV) paired with a prime editing guide RNA (pegRNA)(162). This pegRNA is composed of a Cas9 sgRNA with a 3’ extension containing a primer binding site (PBS) complementary to the non-target DNA strand, as well as a retrotranscriptase template (RTT) harboring the intended mutation (Figure 1.2). After Cas9 nicking of the non-target strand, the PBS anneals to the nicked non-target strand, functioning as a primer for the retrotranscriptase domain. As a result, the RTT is reverse transcribed into a DNA flap carrying the intended edit. Following flap equilibration, the mismatch between the edited and unedited strands is resolved by the cellular DNA repair machinery, resulting in the permanent integration of the edit in the cell genome. This installment of prime editing, termed PE2, was improved by adding a nicking guide RNA (ngRNA) targeting the non-edited strand to stimulate its replacement after flap equilibration (162). This version, termed PE3, benefits from drastically improved efficiency at the cost of higher levels of unwanted indels attributable to the conversion of the double nicks to DSBs.

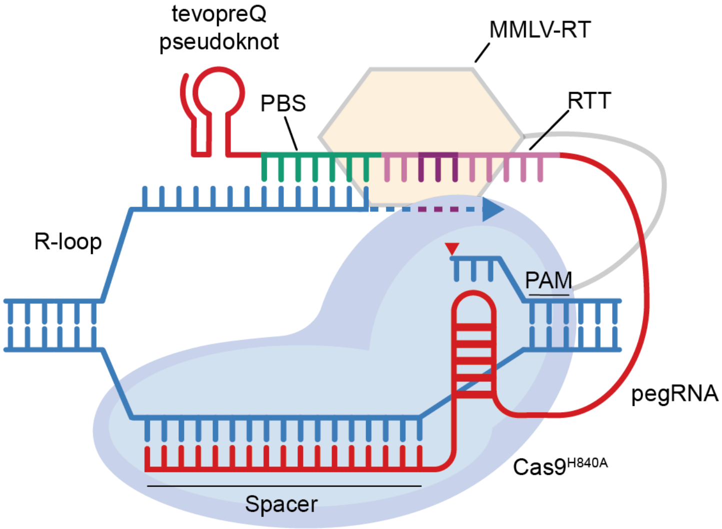


Figure 1.2: Schematic of the prime editing complex at the target site. *RTT: reverse transcriptase template, PBS: primer binding site, pegRNA: prime editing guide RNA, PAM: protospacer adjacent motif, MMLV-RT: moloney murine leukemia virus retrotranscriptase.*

Since the inception of prime editing, the pegRNA has represented an important focus point for optimization. Indeed, prime editing efficiency has been found to be affected by both the lengths and nucleotide compositions of the PBS and RTT (162). Although pooled self-targeting experiments have allowed for the identification of ground rules for pegRNA design, the optimal PBS and RTT compositions appear to be largely target-specific (163). Current pegRNA design tools thus rely on machine learning and *in vitro* assay datasets to predict the best pegRNA architecture (163–167). Independently from PBS and RTT lengths, pegRNA stability has also been shown to be a limiting factor for prime editing. An engineering pegRNA (epegRNA) design was thus developed that adds a pseudoknot RNA structure to its 3' end to protect the extension from exonuclease degradation and improve editing efficiency (168).

Further improvements to the prime editing workflow were achieved by manipulating cellular factors. For instance, the inhibition of the MMR pathway has been shown to prevent excision of the edit from the nicked strand (169). This can be achieved by expressing a dominant negative variant of the MLH1 protein or by directly knocking out the MLH1 gene, thereby converting PE2 to PE4 and PE3 to PE5 (170). Interestingly, this only confers an advantage for substitutions and insertion of sequences up to 13 nucleotides, with larger insertions presumably engaging other DNA repair pathways (165). For the same reason, it was empirically shown that including supplementary synonymous mutations in the RTT of pegRNAs increased the overall prime editing efficiency (171). Other cellular factors were recently shown to inhibit the introduction of long insertions like the cellular 3’ flap nucleases TREX1 and TREX2 (165). Lastly, the prime editing enzyme itself was further engineered to result in newer versions exhibiting improved expression, localization and decreased size (170,172,173).

Although the technology is still in its infancy, it was already used to precisely model and correct patient-derived mutations in organoids (174) and to efficiently generate mouse models (175–177). Beyond the installment of substitutions and short indels, prime editing has also inspired new approaches that use paired pegRNAs or the introduction of integrase recognition sites to introduce large insertions, deletions or chromosomal rearrangements (178–182).

Similarly to base editing and nuclease-based CRISPR screens, prime editing theoretically allows for the inference of edit distribution in a pool of cells from the quantification of genomically-integrated pegRNAs. However, it also presents the important advantage of introducing any type of short mutation without DSB or the need for a separate donor component, thus constituting an attractive *in vitro* screening tool. To date, the only published prime editing screen uses engineered HEK293T cells to study genetic variants in haploidized genomic loci (183). Because of the low efficiency of PE2, the authors opted for the more imprecise PE3 approach and pre-screened high-efficiency pegRNAs targeting 16 loci along the NPC1 gene. They then constructed and delivered 16 distinct pegRNA libraries and sequenced each target site individually to quantify pre- and post-selection variant distributions. Although this represents a successful proof-of-concept, this approach is drastically limited by the need to haploidize the gene of interest, pre-screen functional pegRNAs and conduct one separate screen per gene locus to distinguish intended edits from PE3-induced indels.

In this study, we thus aim to overcome these limitations to study genetic variants in full genes without locus haploidization. As a proof-of-concept, we chose to explore mutations of the epithelial growth factor receptor (EGFR) gene which represent an important therapeutic target in cancer.

1.3 Epithelial Growth Factor Receptor (EGFR)

1.3.1 EGFR activation and signaling

EGFR is a receptor tyrosine kinase (RTK) belonging to the ErbB family and a known proto-oncogene frequently mutated in different cancer types. It is composed of four extracellular domains (I-IV), a transmembrane domain, an intracellular tyrosine kinase domain and a C-terminal tail responsible for signal transduction. Its main ligand, the epithelial growth factor (EGF), was first discovered by Stanley Cohen in 1962 as an unknown factor in murine submaxillary gland extracts that he found to promote early tooth eruption and eyelid opening in newborn mice (184). The unknown protein was isolated in 1965 and shown to promote proliferation and keratinization of epidermal cells in culture (185). It then took until 1982 to finally isolate its receptor, EGFR, from cellular membrane vesicles (186).

In its inactive form, EGFR exists as an autoinhibited monomer (Figure 1.3). Receptor activation is caused by the binding of one of its seven ligands, such as EGF, amphiregulin (AREG) or the transforming growth factor alpha (TGF-α), to extracellular domains I and III (187). Upon binding, the receptor adopts an extended conformation and can form homodimers which interact through a dimerization arm in domain II and the transmembrane domains. This dimerization leads to an asymmetrical interaction between both tyrosine kinase domains, one serving as an allosteric activator for the other (188,189). More specifically, the kinase domain activity is regulated by an αC helix located in its N-terminal lobe (N-lobe). In the inactive kinase domain, this helix is maintained in an outward conformation by the activation loop located in the C-terminal lobe (C-lobe) (190). Upon receptor dimerization, the C-lobe of the "donor" kinase domain interacts with the N-lobe of the "receiver" domain. This interaction disrupts the autoinhibited conformation and allows the αC helix of the receiver to rotate in an active inwards conformation which allows for the formation of the critical Lys745-Glu762 salt bridge within the ATP-binding cleft (190,191). Interestingly, this asymmetric activation mechanism allows for further multimerization of the receptor, each domain thus serving as both an activator and a receiver (192). Furthermore, activated EGFR can also form heterodimers with other ErbB family members which possess different ligand specificities or lack ligands altogether and act as signal amplifiers, such as ErbB2 (or HER2) (192,193).

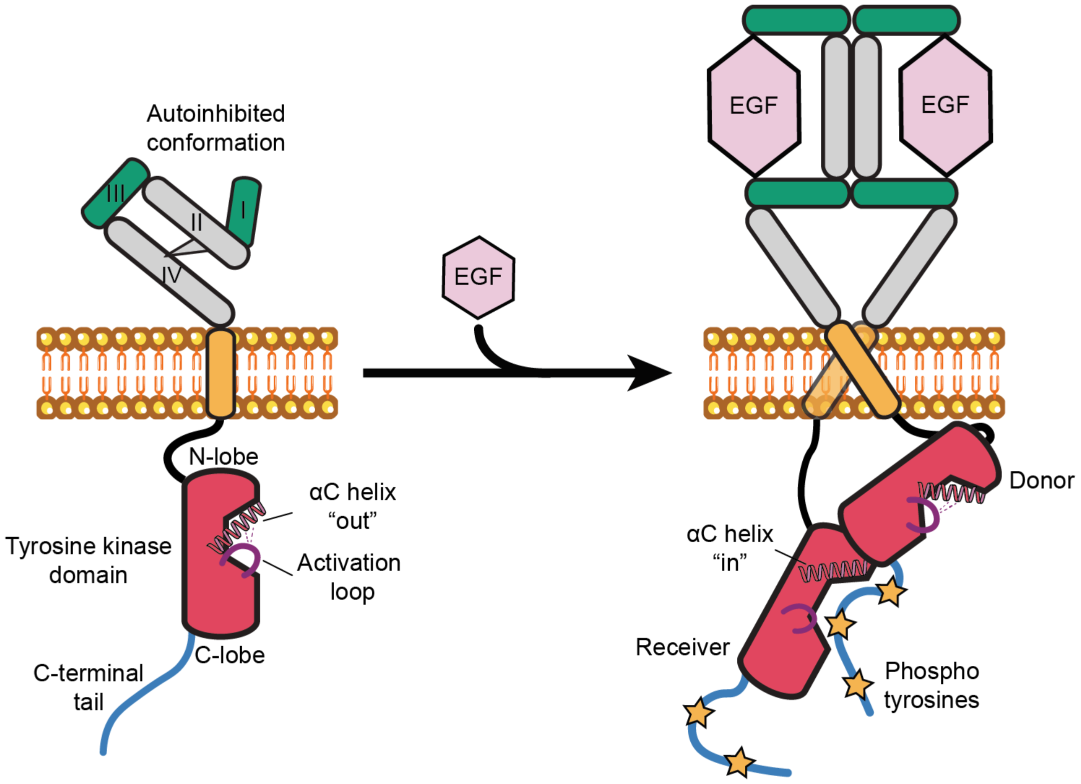


Figure 1.3: Activation of EGFR in response to EGF binding. *The inactive receptor is present at the cell surface as an autoinhibited ("tethered") monomer. EGF binding results in a conformational change allowing the dimerization of the receptor through the interaction of extracellular domains II and transmembrane domains. The C-lobe of the "donor" domain interacts with the N-lobe of the "receiver", resulting in the rotation of the regulatory αC helix in its inward conformation and the allosteric activation of the kinase domain.*

After its activation, the tyrosine kinase domain phosphorylates multiple tyrosine residues of the C-terminal tail. This in turn leads to the phosphorylation of the Tyr869 residue by the Src tyrosine-protein kinase which stabilizes the αC-helix in its active conformation (194,195). Phosphorylated residues of the C-terminal tail are recognised by a wide variety of adaptor proteins containing Src homology 2 (SH2) ot phosphotyrosine binding (PTB) domains. This results in the activation of various signaling cascades, including the RAS-RAF-MEK-MAPK, PI3k-AKT-mTOR and STAT pathways which ultimately promote cell proliferation, migration and apoptosis escape (196–199) (Figure 1.4).

After receptor activation, multiple negative feedback mechanisms can be activated to prevent excessive signaling. First, protein tyrosine phosphatases such as PTPRJ modulate EGFR signaling by dephosphorylating C-terminal phosphotyrosines (200). Additionally, LRIG1, ERRFI1 and SOCS4/5 are EGFR inhibitors which are transcriptionally induced following its activation and directly bind the receptor and other members of the ErbB family, promoting their ubiquitination and degradation (201). Lastly, the phosphorylation of the Tyr1060 residue in the C-terminal tail recruits the Cbl E3 ubiquitin ligase, which leads to receptor ubiquitination, internalization and degradation (202).

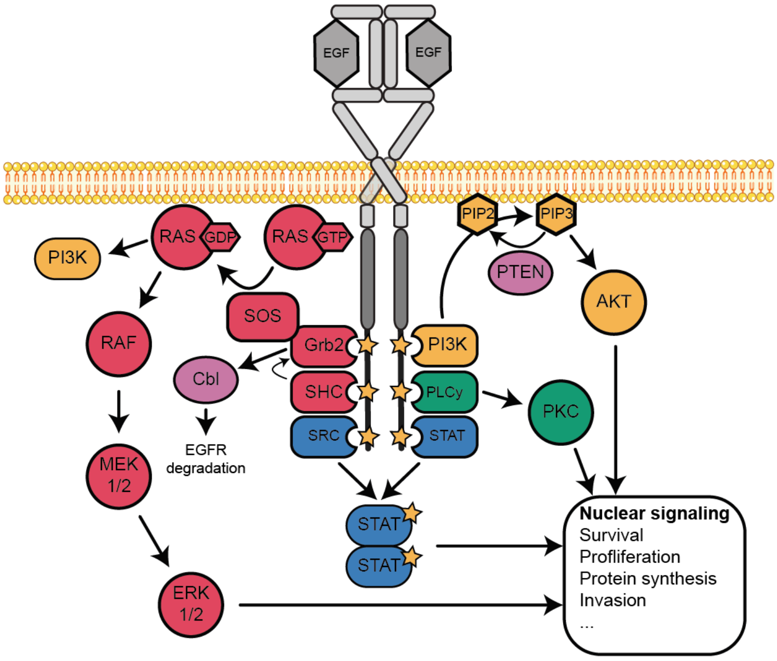
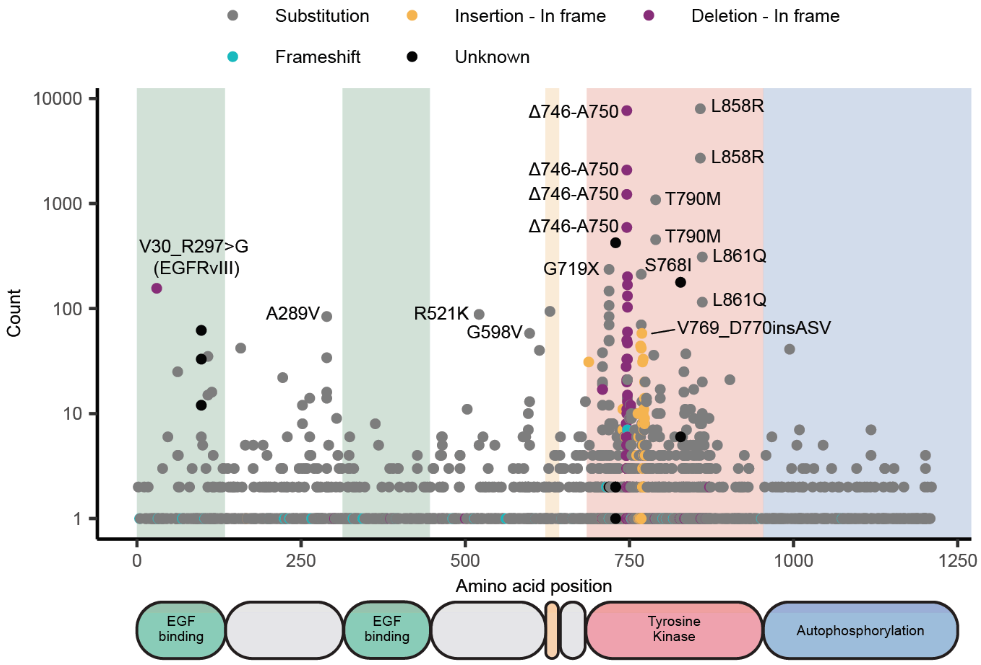


Figure 1.4: **Simplified representation of EGFR downstream signaling pathways.** *Ligand binding activates the tyrosine kinase domain and leads to autophosphorylation of the C-terminal tail. Multiple adapter proteins recognize C-terminal phosphotyrosines and activate downstream signaling pathways such as RAS-RAF-MEK-MAPK (red), PI3k-AKT-mTOR (yellow), STAT (blue) and PLC/PKC (green). Proteins involved in negative feedback loops are shown in pink.*

1.3.2 EGFR in cancer

Because of its role in promoting cell proliferation and survival, EGFR somatic mutations can lead to its constitutive activation and oncogenesis (196). The first link between EGFR and cancer was made with the discovery of the strong homology between EGFR and the v-erb-B transforming protein of the avian erythroblastosis virus (AEV) in 1984 (203). EGFR was later found to be highly expressed in cancer cells (204,205) and its overexpression was shown to result in the EGF-dependent transformation of mouse embryonic fibroblasts (206). Since then, EGFR dysregulation was found to be involved in the emergence of different cancer types, including head and neck squamous cell carcinoma (HNSCC) as well as lung, colorectal, breast, pancreatic and brain cancers.

One frequent mode of EGFR dysregulation is its overexpression or copy number amplification which leads to increased receptor concentration at the plasma membrane and promotes ligand-independent dimerization (207). While this phenomenon is frequently observed in many cancer types, EGFR gain-of-function (GOF) mutations are diverse and appear to be cancer type-specific (Figure 1.5). Notably, these mutations are rare in colorectal and HNSCC but are frequently observed in breast cancer, lung and brain cancers (208–210).



<u>Figure 1.5</u>: Distribution of clinical reports for EGFR variants in the COSMIC database. *Report counts are represented for all cancer types. Protein domains are shown in transparent colors and variant impacts on the EGFR amino sequence are annotated. Total variants: 2128, total reports: 32147.*

In addition to its activity at the plasma membrane, EGFR has also been shown to be internalized and translocated to the nucleus in cancer cells. EGFR nuclear activity is poorly characterized but it is known to bind the promoter of CCND1 and acts as a transcriptional activator, resulting in Cyclin D1 upregulation and cell cycle progression (211). Consequently, nuclear EGFR localization has been associated with poor prognosis in HNSCC and breast cancer (212,213).

1.3.2.1 Non-small cell lung cancer

Lung cancer is the most prevalent cause of cancer-related death, accounting for 20% of cancer mortality. These are categorized into small cell lung cancers (SCLC) and non-small cell lung cancers (NSCLC), the latter representing 85% of all cases (214,215). Among NSCLC, the prevalence of mutated EGFR is approximately 12% worldwide and reaches 49% in Asia (216). These tumors most commonly harbor gain-of-function (GOF) mutations in the EGFR tyrosine kinase domain (217).

Leu858Arg is the most prevalent EGFR mutation in NSCLC (218). The Leu858 residue is located in the activation loop of the C-lobe where it contributes to maintaining the αC helix of the receiver interface in its outward position (190). Its substitution with an arginine was shown to stabilize an intrinsically disordered region surrounding the αC helix, locking it into its active conformation and promoting ligand-independent dimerisation. Because the mutant domain favorably serves as a “receiver” for dimerization, it was termed a “super acceptor” (195,219) and displays a 50-fold enhancement in catalytic activity compared to wild-type (WT) EGFR (190).

Another common mutation type is exon 19 deletions, collectively accounting for 40 to 59% of EGFR mutations in NSCLC (220). More than 100 of these variants have been identified within or adjacent to the β3-αC loop of the kinase domain although Δ746-750 accounts for 75% of cases (221,222). Protein structure modeling has shown that these deletions trapped the regulatory αC helix into its active conformation by shortening the β3-αC loop preceding it, thus promoting dimerization and making Δ746-750 a “super acceptor” (222). However, unlike Leu858Arg, Δ746-750 can serve as both a donor and receiver and has been shown to remain constitutively active upon dimerization disruption by a monoclonal antibody, suggesting differences in their respective modes of activation (223).

1.3.2.2 Glioblastoma multiforme

EGFR mutations or expression alterations are found in 57% of glioblastoma multiforme (GBM) patients (224). These most commonly consist in copy number amplification or large rearrangements like the EGFRvIII variant (EGFRΔExon2-7) which truncates extracellular domains I and II (197,225). This variant shows limited constitutive phosphorylation which leads to decreased recognition by the Cbl E3 ubiquitin ligase and impaired internalization and recycling, thus increasing its localization at the plasma membrane (226,227). Other in-frame deletion variants are found at low frequency in GBM that impact the extracellular domain (EGFRvI, EGFRvII) or the C-terminal tail (EGFRvIV, EGFRvV) (197). Interestingly, vIV and vV are still able to bind ligands and are thought to lack internalization and degradation signals. GBM patients also frequently harbor GOF mutations in the extracellular domains such as Ala265Val/Asp/Thr, Pro545Leu or Arg84Lys which are thought to impair the autoinhibited conformation of the extracellular domain (197).

1.3.3 Treatment options in EGFR-associated cancers

EGFR-driven cancer cells rely on constitutive mitogenic signaling to escape apoptosis and proliferate. This strong imbalance between pro- and anti-apoptotic cellular signals is thought to explain the significant susceptibility of these cells to EGFR inhibition. This process, referred to as oncogene addiction, makes EGFR an ideal therapeutic target in these cancer types (228).

1.3.3.1 Monoclonal antibodies

The first EGFR-targeted molecules developed to counteract oncogenic signaling were monoclonal antibodies (mAb). The potential of murine anti-EGFR mAb to prevent autophosphorylation by competing with EGF binding in the extracellular domains was first demonstrated *in vitro* in 1983 (229). The next year, three mAb clones isolated from mouse hybridomas were shown to inhibit human-derived tumor growth in nude mice (230). The leading clone, mAB-225 (IgG1) was then shown to induce receptor internalization without EGFR autophosphorylation, thus reducing its cell surface concentration (231). Building on these successes, Cetuximab, a chimeric humanized version of mAb-225, was the first to be clinically approved as part of a chemotherapy protocol for the treatment of colorectal cancer in 2004 (232–234). It binds the monomeric inactive receptor and traps it into its autoinhibited conformation, thus preventing both EGF binding and dimerization (235).

Cetuximab is currently approved for the treatment of HNSCC and colorectal cancers, which generally overexpress WT EGFR (236,237). This treatment also recently showed promising results to treat tumors expressing the dimerization-dependent Leu858Arg variant (238). Panitumumab and Necitumumab, other EGFR-targeting mAb, were later approved by the US food and drug administration (USFDA) for the treatment of metastatic colorectal cancer and NSCLC, respectively (239,240).

1.3.3.2 Tyrosine kinase inhibitors

In parallel, tyrosine kinase inhibitors (TKIs) were developed as a class of small molecules that inhibit the kinase activity of RTK and prevent their autophosphorylation. The first EGFR-targeted TKI, named CAQ and developed in 1994, relied on a quinazoline structure to function as an ATP analog that competitively inhibits the EGFR tyrosine kinase activity (241). This inspired the development of multiple similar molecules which are now clinically approved by the USFDA, 11 of which targeting EGFR (242,243).

Among these, Gefitinib (244) and Erlotinib (245) became the first ones to be clinically approved for the treatment of metastatic NSCLC in 2003 and 2004, respectively (246). Interestingly, it was only a year after its approval that Gefitinib was shown to selectively inhibit hyperactive EGFR harboring GOF mutations such as Leu858Arg and Δ746-750 (247,248). This selectivity likely explains the better potency of Gefitinib against this class of mutations compared to Cetuximab (249). Similarly, a phase III clinical trial showed that Gefitinib was superior to a chemotherapy protocol as a first line of treatment in adenocarcinoma patients harboring EGFR mutations (250). Namely, the objective response rate to Gefitinib was 71.2% for patients harboring EGFR mutations against 1.1% for WT EGFR. Comparative clinical studies later showed no difference in response rate between Erlotinib and Gefitinib in EGFR-mutated NSCLC (251). Lapatinib, another first-generation TKI, was granted approval by the USFDA in 2007 and Icotinib was approved by the Chinese National Medical Products Administration (SFDA) in 2011. Although first-generation TKI molecules initially show positive results as cancer therapies, they rapidly give rise to the selection of drug-resistant clones, more than half of which being attributed to the Thr790Met secondary mutation (252,253).

Second-generation TKIs such as Afatinib, Neratinib and Dacomitinib target the same ATP binding pocket as their predecessors but function as irreversible inhibitors forming a covalent bond with the Cys797 residue (254). Afatinib, which also inhibits HER2 and HER4, was the first one to obtain approval from the FDA in 2013. Two phase III clinical trials in NSCLC patients compared survival in patients treated with Afatinib and chemotherapy and showed that the drug significantly improved survival in patients with exon 19 deletions but not Leu858Arg (255). In spite of Afatinib exhibiting a lower IC50 than first-generation TKIs against Thr790Met, it resulted in disappointing clinical outcomes against this variant. This is likely a result of the reduced selectivity of this category of TKIs for mutated EGFR as opposed to wild-type, causing substantial side effects and limiting the dose that could be used in patients (256).

Consequently, irreversible third-generation TKIs like Osimertinib and Olmutinib were specifically developed to counteract Thr790Met (257,258). Osimertinib was granted accelerated approval in 2015 for the treatment of metastatic Thr790Met-positive NSCLC patients. A phase III clinical trial then demonstrated longer progression-free survival with Osimertinib compared to Gefitinib in untreated patients, leading to its approval as a first line treatment in 2018 (259). Osimertinib benefits from a high potency against mutated EGFR, including Thr790Met, and a much better specificity profile compared to Afatinib. However, like previous molecules, its use ultimately results in the emergence of drug resistance (260).

By contrast to Cetuximab, first- and third-generation TKIs are more frequently used in the treatment of NSCLC harboring canonical GOF mutations as they benefit from a high selectivity against these compared to the unmutated receptor (261).

1.4 EGFR and drug resistance

Cellular resistance to anticancer drugs is considered as innate when no response to treatment is observed or when it lasts less than 3 months (262). In the case of TKIs, drug response is generally observed in patients harboring sensitive EGFR mutations but acquired resistances rapidly emerge, invariably leading to disease progression after 1 to 2 years (250,258,259,263).

1.4.1 EGFR-independent resistance

TKI resistance can be acquired through EGFR-independent mechanisms which generally impact the efficacy of all EGFR-targeted drugs. In NSCLC treated with first- or third-generation TKIs, EGFR signaling inhibition is frequently bypassed by the amplification of other RTKs (253,260). Two prominent examples are MET or HER2 which activate the downstream MAPK, PI3K and STAT pathways. The AURA3 phase III clinical trial indeed reported MET and HER2 amplification in 18% and 5% of Thr790Met-positive patients who relapsed after Osimertinib-treatment, respectively (264).

Gene amplification can also affect signaling proteins acting downstream from EGFR like PIK3CA, which encodes the catalytic subunit of the phosphatidylinositol 3-kinase (PI3K) and was found to be amplified in 4% of relapsed patients. Less frequent is the acquisition of GOF mutations in proto-oncogenes like NRAS, KRAS and BRAF which are involved in EGFR signal transduction. Lastly, the amplification of cyclin or cyclin-dependent kinase genes like CCDN1/2, CCNE1, CDK6 and CDKN2A was found in 15% of patients taking part in the AURA3 trial. Consequently, many clinical trials are currently on-going that combine anti-EGFR TKIs with other small molecule inhibitors and mAb to tackle these resistances (260).

Histological changes constitute another poorly understood resistance mechanism. Notably, a histological assessment of 37 NSCLC patients with acquired resistance to first-generation TKIs showed that 14% saw their tumors transform into SCLC (265). This transition is known to lead to EGFR downregulation which makes cells insensitive to TKIs (266). Similarly, epithelial-to-mesenchymal transition (EMT) is associated with increased cell migration and the upregulation of AXL, a RTK whose downstream pathways intersect with those activated by EGFR, thus bypassing its inhibition (267).

1.4.2 Canonical EGFR mutations

Contrary to “off-target” resistance mechanisms affecting other genes, acquired EGFR mutations are generally drug-specific. In the case of first-generation TKIs, the best response rate is achieved against exon 19 deletions like Δ746-750, followed by Leu858Arg (268). Thr790Met is by far the most prominent acquired resistant mutation in this class of drugs and is found in 60% of patients following treatment (253) (Figure 1.6). This variant impacts the “gatekeeper” residue located within the ATP binding pocket and is thought to prevent drug binding by steric hindrance while increasing the affinity to ATP (252,269). While commonly found as a secondary mutation, the ability of Thr790Met to act as an activating mutation on its own is debated although it was shown to confer a growth advantage to epithelial cells *in vitro* and to lead to oncogenesis in inducible mouse models (270,271). Extracellular primary mutations found in GBM like Ala289Asp/Val and Gly598Val are also resistant to Erlotinib and Gefitinib but appear to be more sensitive to Afatinib (272,273). Similarly, EGFRvIII is insensitive to first-generation TKIs and Cetuximab (274–276), possibly due to inefficient blood-brain barrier penetration, while Osimertinib recently showed promising results against this variant (277).

When considering second- and third- generation TKIs, acquired resistance results primarily from mutations affecting Cys797 which is the covalent binding site of the drug. Indeed, Cys797Ser/Gly mutations are reported by different studies in 18 to 25% of patients relapsing after Osimertinib treatment (264,278). Since this molecule is frequently used as a second line of treatment after Thr790Met acquisition, Cys797Ser currently represents a therapeutic dead end for anti-EGFR drugs, thereby motivating the development of a fourth generation of TKIs (279). Interestingly, Cys797Ser remains sensitive to first-generation TKIs when present alone or *in trans* from the Thr790Met allele (260). Other tertiary resistant mutations observed in patients and mouse models after Osimertinib treatment include Leu792Phe/His, Gly796Ser and Leu718Gln (264,280).

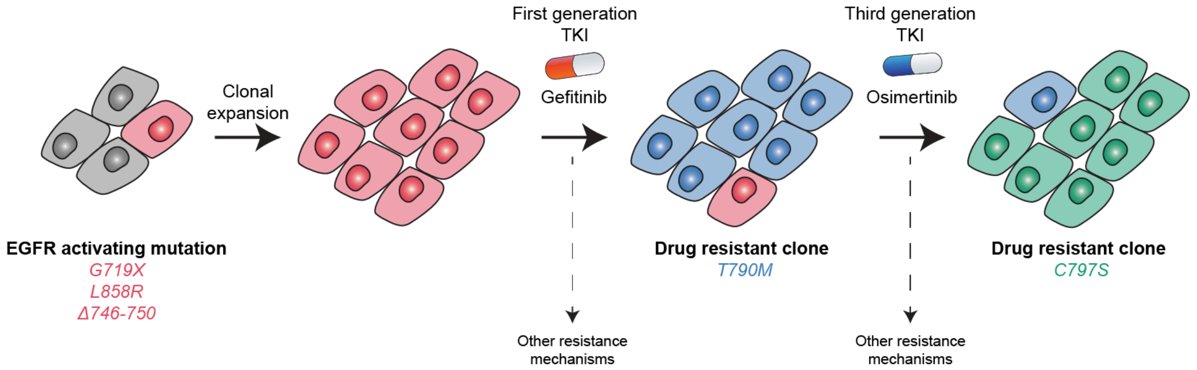


Figure 1.6: Example of drug sensitivity progression in EGFR-mutated NSCLC. *Tumor cells transformed by a canonical EGFR-activating mutation can be treated with a first-generation TKI like Gefitinib. While treatment initially leads to stable disease or tumor regression, resistant clones (e.g. harboring Thr790Met) are rapidly selected which leads to cancer relapse. Third-generation TKIs like Osimertinib selectively inhibit EGFR Thr790Met but eventually also lead to selection of resistant clones harboring Cys797Ser.*

In the case of Cetuximab, impairment of receptor internalization and increased nuclear translocation have been shown to be two potential resistance mechanisms (281,282). Cetuximab is used to treat colorectal cancers in which EGFR mutations are rare. Consequently, a study analyzing colorectal tumors from patients treated with Cetuximab reveals a majority of “off-target” alterations leading to resistance (283). However, two extracellular missense mutations, Gly465Arg and Gly465Glu were identified that were predicted to impact mAb binding. A similar study also identified Ser492Arg in the same domain which was later found in 16% of Cetuximab-treated patients in a phase III clinical study (284,285). In spite of the very low diversity of mutations observed in patients, an *in vitro* study reported that most exon 19 deletions, exon 20 insertions and G719Ala/Cys substitutions found in NSCLC conferred strong resistance to Cetuximab, making this drug unsuitable to treat these cancers (273).

1.4.3 Exon 20 insertions

In-frame exon 20 insertions are found as primary mutations in NSCLC where they represent 4 to 10% of all EGFR variants (286,287). They are a diverse group located within the αC-β4 loop following the αC helix and are thought to displace it to favor its active conformation (288). As a result, the oncogenic character of these mutations appears to be dependent on their position, length and amino acid composition. Indeed, *in vitro* experiments showed that N771_P772insN was constitutively active but not N771_P772insH (288).

Exon 20 insertions generally respond poorly to TKI treatment. In particular, most of these variants have been shown to be resistant to first-generation TKIs (286) with the exception of A763_Y764insFQEA which is located directly within the αC helix and is highly sensitive to Gefitinib (289). Similarly, Osimertinib shows mixed results with the two most frequent of these mutations, D770_N771insSVD and V769_D770insASV, showing sensitivity levels similar to WT EGFR *in vitro* (289). By contrast, another *in vitro* study reported that six exon 20 insertions were resistant to first-, second- and third-generation TKIs, proposing that structural changes prevented drug binding by steric hindrance (290). Additionally, two

Osimertinib phase II clinical trials reported negative outcomes in patients with exon 20 insertion with objective response rates of 0% and 25% (291). These discrepancies demonstrate the risk of considering these different mutations together and the need to stratify patients based on more stringent criteria.

Different treatments have been proposed that selectively target exon 20 insertions. Poziotinib, a second-generation TKI, showed an average potency increase of 100-fold compared to Osimertinib against six resistant insertions *in vitro* (290). In the same study, the authors reported that the drug yielded a 64% remission rate in patients with exon 20 mutations. A phase II clinical trial specifically targeting these patients later reported an objective response rate of 32% although 72% of patients experienced adverse events requiring dose reduction, consistent with the significant side effects observed in other second-generation TKIs (292). Interestingly, patients with an insertion near the αC helix (residues 767-772) responded more positively than those with distal mutations (residues 773+), a feature that was confirmed *in vitro*.

More recently, focus was put on mutation-selective third-generation TKIs with better tolerance profiles. Namely, Mobocertinib, a molecule highly similar to Osimertinib and developed against Thr790Met, yielded objective response rates of 25–28% in phase I/II studies targeting these patients (293). This led the USFDA to grant the drug accelerated approval for the treatment of exon 20 insertion-positive NSCLC in 2021. Lastly, multiple clinical trials are currently evaluating Sunvozertinib which has shown promising potency and selectivity against exon 20 mutations (294). Its measured objective response rate in phase II was 60.8% in patients who previously underwent disease progression after platinum-based chemotherapy, prompting its conditional approval in China in August 2023 (295).

1.4.4 Rare mutations

In addition to canonical EGFR mutations, rare variants are found as primary mutations in 11% of patients and are generally poorly characterized (286,296). This represents a therapeutic challenge as patients harboring these mutations have been shown to generally suffer from worse clinical outcomes.

Among these, Gly719X mutations (most commonly Gly719Ala/Cys/Ser) represent approximately 3% of primary mutations in NSCLC and are associated with mixed response to TKIs. Although clinical studies stratifying these variants separately are sparse, Gly709Cys appears to respond positively to TKI treatment (220). By contrast, Gly709Ala was shown to be 29 and 49 times less sensitive to first- and third-generation TKIs compared to exon 19 deletions, respectively (297). This result was confirmed clinically as patients harboring this variant responded poorly to Gefitinib and Erlotinib but showed lower resistance to Afatinib and Neratinib (220,297).

Ser768Ile and Leu861Gln each represent approximately 1% of EGFR mutations in NSCLC (298). Leu861 is located in the activation loop of the kinase domain where it belongs to the same cluster of hydrophobic residues as Leu858 and contributes to stabilizing the inactive conformation. Leu861Gln was shown to have decreased sensitivity to Gefitinib and Erlotinib *in vitro* but remained sensitive to Osimertinib and Afatinib (273,299). By contrast, Ser768Ile is located in the αC helix and more commonly found associated with other activating variants, making it difficult to establish its drug sensitivity in the clinics (300). However, it has shown poor sensitivity to Gefitinib, Erlotinib and Osimertinib *in vitro* (273,299).

In addition to these, many more very rare EGFR variants have been found as primary or acquired mutations in patients. Indeed, among the more than 2100 EGFR variants in the ClinVar database, 50% are classified as VUS (301). In this context, therapeutic decisions are further complicated by the fact that newly developed TKIs are generally only clinically tested against canonical EGFR mutations (302). Patients harboring uncommon EGFR variants are indeed often excluded from clinical trials (303) although they have been shown to represent 11% of NSCLC patients (296) and to generally suffer from poorer clinical outcomes (286). Given the diversity of these variants and their diverse effects on drug response, there is a strong need for the development of high-throughput assays to elucidate the treatment options for these patients.

1.4.5 Current high-throughput assays for the study of EGFR variant

Over 200 human NSCLC cell lines harboring various driver variants are available (304). While generally well characterized and readily usable in dose-response assays, these often carry mutations in multiple oncogenes and do not represent the full spectrum of rare EGFR variants found in patients. Consequently, a common approach to assess the transforming ability and drug resistance profile of EGFR variants is their overexpression in murine cell lines.

For example, NIH/3T3 are embryonic mouse fibroblasts that can undergo oncogenic transformation when transfected with human EGFR mutants, thereby acquiring a growth advantage and the ability to form colonies on soft agar (305). Similarly, Ba/F3 is a murine pro-B cell line which gains interleukin-3-independent growth upon EGFR expression and activation (306). In both cases, cell proliferation is dependent on EGFR signaling, allowing for the assessment of receptor activation and inhibitor sensitivity in growth assays. These cell lines are thus frequently used to predict the efficacy of different drugs on new EGFR variants discovered in patients (307). For example, Robichaux et al. evaluated the selectivity of different TKIs on 76 Ba/F3 cell lines expressing uncommon patient-derived variants in an arrayed format (286). Their study established precise dose-response curves for each variant and drug combinations before extending their approach to 16 compound mutations.

Other studies report the assessment of variants in a pooled format instead of arrayed assays. For example, Ba/F3 cells were used in random mutagenesis experiments to study secondary drug-resistant variants (308). The authors first established stable cell lines expressing human EGFR activating variants alone or associated with Thr790Met. These lines were then exposed to a mutagenic agent before drug treatment and sequencing of resistant clones. Although it has the advantage of allowing the rapid screening of secondary mutations in different EGFR alleles, this approach identified a limited number of variants including prevalent Osimertinib-resistant mutations like Cys747Ser and Leu718Gln.

In order to maintain a greater control over the investigated mutation, contemporary variant screens generally employ lentiviral delivery of cDNA variants to concurrently evaluate tens of variants. For example, Chakroborty et al. used pooled delivery of 7216 EGFR variants obtained by error-prone PCR in Ba/F3 cells to identify Leu858Arg, Thr790Met and the novel

Ala702Val variant as activating mutations (309). While simultaneously assessing thousands of variants, this approach also identified many false positive hits which were likely passenger mutations present in the same library element as a functionally active variant.

In an approach proposed by Kohsaka et al., the authors used barcoded EGFR variants delivered individually to NIH/3T3 or Ba/F3 cells which were then pooled in competition assays (273). The growth phenotype of 101 EGFR variants was thus assessed under drug treatments or growth factor (FBS or IL-3) deprivation to establish precise activation and drug resistance profiles for each variant. Although establishing EGFR mutant-expressing cell lines in an arrayed format is work-intensive, this approach allowed the assessment of variant combinations in *cis* and *trans*, which is not possible with other methods.
In addition, the authors demonstrated the viability of their approach *in vivo* by injecting their NIH/3T3 libraries as xenografts in nude mice before evaluating the prevalence of each variant in tumor compositions. This *in vivo* approach was later used in another study to evaluate drug combination regimens, thus better replicating treatment conditions implemented in patients (310). Similarly, a recent study used the *in utero* electroporation and transposon-mediated integration of a barcoded EGFR variants library in the mouse genome to infer the transforming potential of 36 variants found in GBM (311).

Lastly, while previously cited studies use mouse cell lines, a recent deep mutational scanning screen was conducted in human NSCLC-derived PC-9 cells (312). Namely, EGFR variant libraries containing thousands of substitutions, insertions or deletions in exons 18–21 of EGFR were delivered in lentiviral vectors containing sgRNAs designed to knock out the endogenous EGFR gene. Targeting a limited number of exons allowed for the direct sequencing of the mutagenised sites after cell treatment with 5 different TKIs and the establishment of comprehensive relative enrichment maps.

In conclusion, high throughput cellular assays have proved their usefulness to rapidly assess the oncogenicity of EGFR variants at scale. However, most of these are performed in mouse cell lines although it is currently unknown if human EGFR can form functional heterodimers with murine receptors. This is significant because EGFR variants can have distinct impacts on the formation of homodimers or heterodimers with other members of the ErbB family. While a recent study applied this approach to human PC-9 cells, it still relies on exogenous variant expression and thus fails to replicate EGFR endogenous expression levels. This too is of particular importance because EGFR overexpression is a drug resistance mechanism in itself and altering receptor stoichiometry at the plasma membrane can lead to constitutive activation. It is thus essential to establish new high-throughput variant screening assays that closely replicate human cellular contexts and expression levels.

1.5 Conclusion and thesis aims

Cancer poses an important health burden and stands as one the primary causes of mortality worldwide. In this context, EGFR is a prominent oncogene commonly impacted by a diverse spectrum of dysregulations and gain-of-function mutations. This motivated the development of TKIs, a class of small molecules that compete with ATP to bind the EGFR kinase domain and inhibit its signaling. Although initially successful in controlling disease progression, these inhibitors invariably lead to the emergence of drug resistance. Although multiple drug-resistant mutations have been identified, 50% of EGFR variants listed in the ClinVar database are still classified as VUS, posing an important therapeutic challenge. It is thus essential to establish efficient methods to rapidly assess EGFR variant pathogenicity and drug sensitivity at scale.

Over the past decade, the assessment of EGFR variants has primarily relied on the overexpression of mutant cDNAs in mouse cell lines. This approach is easily applicable but fails to replicate the endogenous expression levels, genomic context and cellular interaction partners of EGFR. In parallel, base and prime editing have emerged as new high-throughput screening tools capable of precisely introducing thousands of genetic variants in the genome of cells. While base editors have proved their usefulness in gene-centric and genome-wide variant screens, the ground principles of prime editing screens in diploid cells are yet to be established.

The main objective of this thesis is to leverage the emerging base and prime editing technologies and establish complementary precision genome editing screens to evaluate the clinical significance of thousands of EGFR variants in the genome of human cells. Combining both approaches allowed us to explore a broad spectrum of mutations through 1) the unbiased scanning of the EGFR coding sequence with tiling base editing libraries, 2) the targeted introduction of all patient-derived variants listed in the ClinVar or COSMIC databases and amenable to prime editing.

First, I aimed to establish robust and versatile cellular assays to introduce and evaluate EGFR variants *in vitro*. This allowed the screening of thousands of variants in different cell lines with distinct biologically-relevant genetic backgrounds using precision genome editing technologies. These screens provided important information on both primary and secondary EGFR mutations and their role in drug resistance and receptor regulation mechanisms. These results not only illuminated important aspects of EGFR biology but also provided reliable drug sensitivity data, which could assist in the diagnostics and treatment of cancer patients harboring VUS.

Finally, I aimed to establish an experimental framework for the use of prime editing as a screening tool in diploid genomes. This included establishing library and assay design rules that could be easily extended to other genes, cellular models or phenotypes. Notably, the knowledge gained in the process helped us to identify strengths and limitations of the prime editing workflow and to formulate recommendations for the rational design of multimodal precision genome editing screens. With the growing prevalence and specialization of genetic tests, resulting in a growing number of VUS, this study could serve as a reference for future research endeavors seeking to address the current bottleneck of variant classification.

Chapter II: Material and methods

2.1 Cell lines and cell culture

2.1.1 Cell culture and maintenance

H1299 cells were obtained from ATCC (#CRL-5803) and PC-9 cells from Merck (#90071810-1VL). Both cell lines were cultured in RPMI medium (ATCC modification, ThermoFisher #A1049101) supplemented with 10% FBS. Cells were routinely cultured in absence of antibiotics except for screening experiments where the culture medium was supplemented with 1% penicillin-streptomycin (ThermoFisher #15140122).

MCF10A cells were obtained from Cell Lines Service (#305026) and cultured in complete Mammary Epithelial Cell Growth Medium (MEGM Bullet kit, Lonza CC-3150) supplemented with 100ng/mL cholera toxin (Merck #C8052). For EGF deprivation experiments, the same medium was used, omitting the hEGF provided with the Bullet Kit. Culture medium was replaced every 4 days.

HEK293T cells were obtained from Merck (#12022001) and cultured in Dulbecco's Modified Eagle Medium (DMEM) (Merck, #D0822) supplemented with 10% FBS.

All cell lines were cultured at 37 °C in 5.0% CO2 and passaged around 70% confluence using TryplE (ThermoFisher #12604013). In the case of MCF10A cells, detached cells were resuspended in 10mL phosphate-buffered saline, collected in a 15mL falcon tube and spun down at 500g. The supernatant was then aspirated to eliminate residual TryplE and cells were resuspended in fresh medium before seeding.

All screening experiments were conducted in 300 cm^2 cell culture flasks (TPP, #90301).

2.1.2 Cell freezing and storage

For cell freezing, cells were detached using TryplE (ThermoFisher #12604013), resuspended in 10mL phosphate-buffered saline and spun down at 500g. Cell pellets were resuspended at a concentration of 10^7 cells per milliliter in the following freezing solutions:

- For H1299 and PC-9 cells: RPMI medium supplemented with 20% FBS and 10% DMSO.
- For MCF10A cells: complete medium supplemented with 7.5% DMSO.

Resuspended cells were then transferred to cryotubes and placed in an isopropanol chamber at -80°C for 48 hours before transferring the tubes to liquid nitrogen.

2.2 Base editing screens

2.2.1 Lentiviral vectors

For BE3.9Max and ABE8e experiments, individual spacers and libraries were cloned into the pRDA_256 (Addgene, #158581) and pRDA_426 (Addgene, #179097) plasmids, respectively.

2.2.2 Base editing library design

ChopChop (313) was used to design all possible SpCas9 spacers targeting within and 30 nucleotides around EGFR exons and UTRs (1496 unique sgRNAs) as well as 200 unique sgRNAs targeting EGFR introns. sgRNAs containing BsmBI restriction sites or polyT(>4)

were filtered out and an additional G was appended to the 5' end of spacers starting with another nucleotide.
Additionally, 206 spacers (10% of the final library size) with no target site in the human genome were picked at random from the Brunello library (118). Lastly, 103 sgRNAs targeting splice sites of essential genes were obtained from Hanna et al., 2021 (145) and added to the final library. Finally, primer-binding sites and BsmBI restriction sites (in bold) for Golden Gate cloning were added to the 5' and 3' ends of each spacer:
5' - AGGCACTTGCTCGTACGACG(**CGTCTC**)ACACC - 3'
5' - GTTTC(**GAGACG**)TTAAGGTGCCGGGCCCACAT - 3'

Base editing outcomes were predicted for each EGFR-targeting sgRNA using the base editor design tool script from Hanna et al., 2021 (145).

2.2.3 Base editing library cloning

The final oligo library was ordered from Twist Bioscience and amplified in 5 replicates of the following reaction:
25 µL Kapa Taq ReadyMix (Merck #TAQKB), 6 µL Primer mix (2.5 µM each), 1 µL Oligo pool (1 ng/uL) and 17 µL H2O. PCR cycling conditions: 3 min at 95°C, [20 s at 98°C, 15 s at 60°C, 30 s at 72°C] for 20 cycles followed by a final extension 1 min at 72°C.

All PCR replicates were then pooled and desalted with a Qiagen PCR purification kit (Qiagen, #28104). The obtained double stranded library inserts were then cloned into the ABE8e and BE3.9Max lentiviral backbones using Golden Gate Assembly:
50 fmol amplified library, 50 fmol plasmid backbone, 1 µL ThermoFisher Tango buffer (ThermoFisher, #BY5), 1 µL DTT (10 mM), 1 µL ATP (10 mM), 1 µL T7 DNA ligase, 1 µL BsmBI (Esp3I, ThermoFisher #ER0451). Thermocycler conditions: [5 min at 37°C, 5 min at 20°C] for 25 cycles followed by 1h at 37°C and 10 min at 65°C.

Golden Gate assembly products were then desalted and concentrated using isopropanol purification as described in Joung et al., 2017 (105). 100 ng from each purified product was electroporated in each of 3 shots of Endura electrocompetent E. Coli (Lucigen, #60242) following the manufacturer's protocol. Transformed E.Coli cells were then grown overnight in 500mL of Terrific Broth supplemented with 100 mg/L Ampicillin before final plasmid library extraction using the Qiagen Plasmid Maxiprep kit (Qiagen, #12162).

2.2.4 Screening and drug selection

PC-9 or MCF10A cells were infected in duplicates while aiming for a coverage of 500 infected cells per library element and a MOI of 0.3. Infected cells were selected with 3 µg/mL Puromycin from day 2 to day 4. Cells were then split in the different treatment arms on day 11 and final harvest was performed on day 19 while maintaining a minimal cell coverage of 500x. Cell pellets were frozen at -80°C before processing.

Each drug was used at a concentration corresponding to the IC50 described in the literature or measured by cell counting after 4 days of selection. Namely, PC-9 cells were treated with 0.05µM Gefitinib or 0.02µM Osimertinib. MCF10A cells were treated with 0.13µM Gefitinib or 0.3µM Osimertinib.

2.2.5 Validation plasmid cloning

Individual sgRNA spacers were ordered as complementary single stranded oligonucleotide pairs with 5’ sticky ends from Integrated DNA Technologies. Individual oligonucleotides were resuspended in NEB buffer II (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl2, 1 mM DTT, pH 7.9) to a concentration of 100 µM. Complementary oligonucleotides were then mixed in equimolar proportions, heated up to 95°C for 2 minutes and left to cool down at room temperature.

Annealed oligonucleotides were then cloned in base editing lentiviral vectors using Golden Gate cloning by preparing the following reaction mixes: 50 fmol oligonucleotide duplex, 50 fmol plasmid backbone, 1 µL ThermoFisher Tango buffer (ThermoFisher, #BY5), 1 µL DTT (10 mM), 1 µL ATP (10 mM), 1 µL T7 DNA ligase, 1 µL BsmBI (Esp3I, ThermoFisher #ER0451). Thermocycler conditions: [5 min at 37°C, 5 min at 20°C] for 10 cycles followed by 1h at 37°C and 10 min at 65°C.

5 µL of each Golden Gate reaction was transformed into Stbl3 chemically competent E. Coli cells (Invitrogen, #C737303) by heat shock. Transformed cells were then spread on agarose plates supplemented with Ampicillin and incubated overnight at 37°C. Colonies were picked on the next day and sent for Sanger sequencing of plasmid inserts.

Name	Base editor	sgRNA sequence
Thr790Ala,Gln791Arg	ABE8e	caccGATCACGCAGCTCATGCCCTTgttt
Met766Thr	ABE8e	caccGTGGCCATCACGTAGGCTTCCgttt
His773Arg	ABE8e	caccGCCCCACGTGTGCCGCCTGCTgttt
Val845Ala	ABE8e	caccGTGTTTTCACCAGTACGTTCCgttt
Ile853Val,Thr854Ala	ABE8e	caccGCAAGATCACAGATTTTGGGCgttt
Thr790Met,Gln791Ter	BE3.9Max	caccGATCACGCAGCTCATGCCCTTgttt
Gly719Gly/SerS720Phe	BE3.9Max	caccGGGCTCCGGTGCGTTCGGCAgttt
Exon25+1,Leu1038Leu	BE3.9Max	caccGTTCATACCAGAGAGCTCAGGgttt

Table 2.1: sgRNAs used for base editing screen validation. *Cloning overhangs of the oligonucleotide duplexes are shown in lowercase.*

2.2.6 Deep sequencing validation experiments

For validation experiments, all-in-one base editing vectors were individually delivered to cells. Lentivirus production was conducted by co-transfecting the following plasmids into HEK293T cells seeded on the previous day in a 24-well plate:

- 211.4 µg of psPAX2
- 165.9 µg of pMD2
- 422.7 µg of base editing vector

Serum-free OptiMEM medium (Gibco, #31985062) was added to this plasmid mix to a final volume of 250 µL. In a separate tube, 10 µL of Lipofectamine 2000 (Invitrogen, #11668027) was added to 240 µL of OptiMEM. Both solutions were then mixed together, incubated for 10 minutes at room temperature and added dropwise to each well of the 24-well plate.

Viral supernatants were harvested after 48h and cell debris were eliminated using a 0.45 µm filter. Target cells were transduced by adding 250µL of clarified viral supernatant to each well of a 24-well plate and selected with puromycin between from days 2 to 4 after infection. Drug treatments or EGF deprivation started on day 6 after infection and cells were harvested on day 13.

2.2.7 Cell viability assays

For cell viability experiments, cells were seeded in 96-well plates (Greiner #655077) four days after infection, treatments were applied on day 5 and viability measurements were performed on day 10 using the CellTiter-Glo® 2.0 Cell Viability Assay kit (Promega #G9241). Luminescence was measured following the manufacturer’s protocol on a Tecan SPARK reader.

2.3 Prime editing screen

2.3.1 Prime editing vector cloning

For prime editing, a pLentiGuide vector (Addgene, #117986) was modified by replacing the existing BsmBI sites and the sequence in between by the following stuffer, thus reintroducing the BsmBI recognition sites (in bold) and changing the corresponding overhangs (lowercase): caccG**GAGACG**CTATCACCC**CGTCTC**Ttttt.

Additionally, the mCherry CDS was replaced by a dominant negative MLH1$^{\Delta 754\text{-}756}$ variant in order to inhibit MMR in target cells and increase prime editing efficiency. All pegRNAs were cloned in the resulting pLentiGuide_Puro-T2A-MLH1$^{\Delta 754\text{-}756}$ vector.

The prime editor enzyme was delivered to cells via a modified pLenti-PE2-BSD vector (Addgene #161514) in which the blasticidin resistance gene was replaced with a Green Fluorescent Protein (GFP).

2.3.2 Establishment of stable cell lines

For prime editing experiments, clonal cell lines stably integrating the prime editor construct and GFP were derived using lentiviral delivery and FACS isolation of clones with strong fluorescent signal. Before each screening experiment, cells with high GFP expression were resorted to limit construct silencing in the cell population.

2.3.3 Prime editing library design

All EGFR mutations smaller or equal to 3 nucleotides were downloaded from the ClinVar and COSMIC databases (1100 variants in total, database accessed on July 29th, 2021). Each variant was attributed a maximum of 5 SpCas9 spacers designed with ChopChopV2 with a maximum variant to cut site distance of 10 nucleotides.

For each spacer-variant pair, a total of 9 mutagenic pegRNAs were designed with each combination of 10, 12 or 14 nucleotides RTT and 11, 13 or 15 nucleotides PBS. Lastly, for each mutagenic pegRNA, a non-mutagenic equivalent was designed that re-introduces the wild type target sequence to account for potential pegRNA-mediated non-desired edits. pegRNAs containing BsmBI restriction sites or polyT(5) were filtered out, yielding a final library size of 19400 unique pegRNAs.

For each library element, the pegRNA backbone was replaced with a pair of BsmBI sites separated with a random stuffer whose length was designed to normalize the final oligonucleotide length to 215 nucleotides regardless of pegRNA extension design. The tevopreQ sequence was appended to the 3' end of each pegRNA extension followed by a polIII termination signal and a unique 11-nucleotides DNA barcode generated with the DNABarcodes Bioconductor package with a minimal Hamming distance of 2 (314).
Lastly, final oligonucleotides were designed by adding the following 5' and 5' Gibson cloning overhangs:
5'-TATATATCTTGTGGAAAGGACGAAACACCG**[Barcoded_epegRNA]**TTTTCGAGTACTA GGATCCATTAGGCG-3'

2.3.4 Prime editing library cloning

The final oligo library was ordered from Twist Bioscience and amplified in 65 replicates of the following reaction:
Kapa Taq ReadyMix (Merck #TAQKB) 25 µL, 6 µL Primer mix (2.5 µM each), 1 µL Oligo pool (1 ng/uL) and 17 µL H2O. PCR cycling conditions: 3 min at 95°C, [20 s at 98°C, 15 s at 61°C, 30 s at 72°C] for 8 cycles followed by a final extension 1 min at 72°C.

Library cloning was then performed in two steps. First, all PCR replicates were pooled and desalted with a Qiagen PCR purification kit (Qiagen, #28104). The obtained double stranded library inserts were then cloned into the pre-digested pLentiGuide_Puro-T2A-MLH1$^{\Delta754\text{-}756}$ vector by Gibson assembly using the NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs, #E2621). Gibson assembly products were then desalted and concentrated using isopropanol purification as described in Joung et al., 2017 (105). 100 ng from each purified product was electroporated in each of 5 shots of Endura electrocompetent E. Coli cells (Lucigen, #60242) following the manufacturer's protocol. Transformed E.Coli cells were then grown overnight in 500 mL of Terrific Broth supplemented with 100 mg/L Ampicillin. The plasmid library was purified using the Qiagen Plasmid Maxiprep kit (Qiagen, #12162).

The sgRNA scaffold was then inserted in the previously cloned plasmid library using Golden Gate assembly:
50 fmol library plasmid, 50 fmol sgRNA scaffold oligo duplex, 1 µL ThermoFisher Tango buffer (ThermoFisher, #BY5), 1 µL DTT (10 mM), 1 µL ATP (10 mM), 1 µL T7 DNA ligase, 1 µL BsmBI (Esp3I, ThermoFisher #ER0451). Thermocycler conditions: [5 min at 37°C, 5 min at 20°C] for 25 cycles followed by 1h at 37°C and 10 min at 65°C. The final plasmid library was then purified, desalted and amplified following the previously described protocol.

2.3.5 EGFR activation screen

MCF10A cells were infected in duplicates while aiming for a coverage of 500 infected cells per library element and a MOI of 0.3. Infected cells were selected with 3 µg/mL Puromycin from day 2 to day 4. EGF deprivation started on day 14 with a cell coverage of 1000x and the final harvest was performed on day 22. Cell pellets were frozen at -80 degrees before processing.

2.3.6 Validation pegRNA cloning

Individual pegRNAs used for prime editing efficiency assessment were ordered as double-stranded gBlocks (Integrated DNA technologies) with the following overhangs:
5’-TATATATCTTGTGGAAAGGACGAAACACCG**[epegRNA]**TATATATCTTGTGGAAAGGACGAAACACCG-3’
Each pegRNA was cloned in the pLentiGuide_Puro-T2A-MLH1$^{\Delta 754\text{-}756}$ vector by Gibson cloning. Briefly, 25 ng of plasmid backbone pre-digested with BsmBI (ThermoFisher, #ER0451) was mixed with a 3-fold molar excess of gBlock (Integrated DNA Technology) insert and 5µL of Gibson Master Mix (New England Biolabs, #E2611S) in a final reaction volume of 10 µL. The samples were then incubated 15 min at 50°C and transformed into Stbl3 chemically competent E. Coli cells (Invitrogen, #C737303). After heat shock transformation, E. Coli cells were spread on agarose plates supplemented with Ampicillin and incubated overnight at 37°C. Colonies were picked on the next day and sent for Sanger sequencing of plasmid inserts.

Name	Sequence
HEK3 CTT insertion (Addgene #132778)	GGCCCAGACTGAGCACGTGA**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCC**TCTGCCATCAAAGCGTGCTCAGTCTG
EGFR Thr790Met	GTAGTCCAGGAGGCAGCCGAA**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCC**GCTCATCATGCAGCTCATGCCCTTCGGCTGCCTCCTGGCGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAA

Table 2.2: epegRNAs used in prime editing validation experiments. The *sgRNA scaffold is shown in bold.*

2.4 Lentiviral library production

For each sgRNA library, HEK293T cells (Merck, #12022001) seeded in fifteen (base editing) or thirty (prime editing) 15-cm dishes were transfected by preparing the following mixture for each dish: 5548 ng psPAX2 plasmid, 4356 ng pMD2 plasmid, 11087 ng library plasmid, 169 µL PEImax (1g/L), 400 µL NaCl (1.5M), H2O to 2000 µL.
The transfection mixes were vortexed for 10s, incubated at room temperature for 10 minutes and added dropwise to the cells.
Transfected cells were incubated for 48h at 37°C before viral supernatants were harvested, pooled and clarified 5 min at 1000g. Lentiviral solutions were then overlaid on top of a 3mL 20% sucrose cushion and centrifuged in a SW32Ti swing-bucket rotor (90 min at 25,000 rcf, 4°C). Lentiviral pellets were then resuspended in 1mL DMEM + 2% FBS and frozen at -80°C.

2.5 Lentiviral library titration

Lentiviral libraries titers were determined by serial dilution and cell counting of infected MCF10A or PC-9 cells (depending on the experiment). First, 10,000 cells were seeded in each well of a 24-well plate. Right after seeding, one aliquot of concentrated viral solution was thawed from -80°C stocks and serial 10-fold dilutions were performed up to 1:10,000. 50µL of each viral dilution was then used to infect previously seeded cells in 6 replicates. After 48h, 3 of the 6 infected replicates of each viral dilution were selected with 3 µg/mL puromycin while the 3 others had their medium replaced with antibiotics-free medium.

Cell numbers in each well were determined 4 days after infection by aspirating the medium from each well and washing cells with 500µL phosphate-buffered saline. Cells were then detached and resuspended in 200µL of TrypIE (ThermoFisher #12604013) before counting with a Cellometer Auto T4 Bright Field Cell Counter (Nexcelom). The initial viral titer in each well was determined using the following formula:

$$Viral\ titer\ (infectious\ units/\mu L) = (\frac{cell\ number\ (selected)}{cell\ number\ (non\ selected)} \times initial\ cell\ number)\ \div\ viral\ volume$$

The final viral titer was then averaged over the wells with a cell survival fraction between 10 and 90% after accounting for dilution factors.

2.6 Genomic DNA isolation and sequencing

Genomic DNA isolation was performed using the Blood & Cell Culture DNA Midi Kit (Qiagen #13343) according to the manufacturer’s protocol.

Illumina library preparation was performed in two steps. First, 4 µg of genomic DNA were used in each 50 µL PCR1 reaction with primers amplifying the sgRNA library and containing universal Illumina adapters. Forward primers were ordered from Integrated DNA technologies to include three different stagger lengths and were pooled in equal amounts. The final PCR mixes had the following composition:
Kapa Taq ReadyMix (Merck #TAQKB) 25 µL, 6 µL Primer mix (2.5 µM each), 4 µg genomic DNA and H2O to 50 µL. PCR cycling conditions: 3 min at 95°C, [20 s at 98°C, 15 s at 60°C, 30 s at 72°C] for 18 cycles followed by a final extension 30 s at 72°C.
PCR1 products from the same sample were then pooled together. In the case of prime editing screens, PCR1 amplicons were size-selected using a 1.3x Ampure bead ratio and eluted in a volume of H2O equal to the initial PCR product volume.

Name	Sequence	Library element
0477_epeg_library_seq_Fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**(N)**TGGCGGTAATACGGTTATCC	epegRNA barcode
0435_epeg_library_seq_Rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCNGTTACGCGTTAAACCAACTAGAA	epegRNA barcode
0344_Library_Seq1_Fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**(N)**CGGCCGCCTAATGGATC	Full-length epegRNA
0350_Library_Seq_Rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCNNTATATATCTTGTGGAAAGGACGAAACA	Full-length epegRNA
0442_Library_BE_Fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**(N)**TATATATCTTGTGGAAAGGACGAAACA	Base editing sgRNA
0445_Library_BE_Rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCNCCAATTCCCACTCCTTTCAAGAC	Base editing sgRNA

Table 2.3: PCR1 primers used for library deep sequencing. *N nucleotides between parentheses represent staggers of variable lengths (one to three) used to increase sequencing pool diversity on the flow cell.*

For each sample, a single PCR2 reaction was then performed to add P5 and P7 Illumina adapters as well as demultiplexing barcodes. The reaction mix was as follows: Kapa Taq ReadyMix (Merck #TAQKB) 10 µL, 2.5 µL Primer mix (4 µM each), 1 µL PCR1 product and 6.5 µL H2O. PCR cycling conditions: 3 min at 95°C, [20 s at 98°C, 15 s at 61°C, 30 s at 72°C] for 8 cycles followed by a final extension 30 s at 72°C.

Name	Sequence
P5 adapter	AATGATACGGCGACCACCGAGATCTACAC**NNNNNNNN**ACACTCTTTCCCTACACGACGCTCTTCCGATCT
P7 adapter	CAAGCAGAAGACGGCATACGAGAT**NNNNNNNN**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC

Table 2.4: Universal PCR2 primers used for library deep sequencing. *N nucleotides in bold represent demultiplexing barcodes.*

Barcoded PCR2 products were then pooled, desalted and 700 ng of DNA was loaded in each lane of a 2% Agarose EGel (InVitrogen, #A42135). DNA bands were extracted using the Qiagen PCR purification kit (Qiagen, #28104). Samples were then pooled sequenced on an Illumina NextSeq 500 sequencer (SR150 HighOutput) with 5% PhiX.

2.7 Deep sequencing for validation experiments

For screen validation with target site amplicon sequencing, genomic DNA was extracted by resuspending cell pellets in QuickExtract buffer (Lucigen, #QE09050) at a final concentration of 1600 cells/µL and incubating them 5 min at 65°C, 5 min at 68°C and 10 min at 98°C.

The target genomic region was then amplified by preparing the following PCR1 mixes:

Kapa Taq ReadyMix (Merck #TAQKB) 25 µL, 6 µL Primer mix (2.5 µM each), 10 µL genomic DNA (16,000 cells) and H2O to 50 µL. PCR cycling conditions: 3 min at 95°C, [20 s at 98°C, 15 s at 60°C, 30 s at 72°C] for 25 cycles followed by a final extension 30 s at 72°C.
Ampure bead purification of PCR1 products was performed at a 1:1 ratio and size-selected amplicons were eluted in a volume of water equal to the initial PCR product volume. These purified amplicons were then used as templates in a PCR2 reaction to add P5 and P7 Illumina adapters and demultiplexing barcodes:
Kapa Taq ReadyMix (Merck #TAQKB) 10 µL, 2.5 µL Primer mix (4 µM each), 1 µL size-selected PCR1 product and 6.5 µL H2O. PCR cycling conditions: 3 min at 95°C, [20 s at 98°C, 15 s at 61°C, 30 s at 72°C] for 6 cycles followed by a final extension 30 s at 72°C.
All samples were then pooled together and sequenced on a Illumina NextSeq 500 sequencer (SR150 HighOutput) with 5% PhiX.

Name	Sequence	Experiment
0449_EGFR_Ex21_NGS_Fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**(N)**AATGCTGGCTGACCTAAAGC	Val845Ala[ABE8], Ile853Val/Thr854Ala[ABE8e]
0456_EGFR_Ex21_NGS_Rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCNCCAGCCATAAGTCCTCGACG	Val845Ala[ABE8], Ile853Val/Thr854Ala[ABE8e]
0447_EGFR_Ex20_NGS_Fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**(N)**GACATAGTCCAGGAGGCAGC	Met766Thr[ABE8e], His773Arg[ABE8e], Thr790Met,Gln791Ter[BE3.9Max], Thr790Ala,Gln791Arg[ABE8e]
0455_EGFR_Ex20_NGS_Rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCNTTCACCTGGAAGGGGTCCAT	Met766Thr[ABE8e], His773Arg[ABE8e], Thr790Met,Gln791Ter[BE3.9Max], Thr790Ala,Gln791Arg[ABE8e]
0421_EGFR_Ex18_NGS_Fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**(N)**CTCCCCACCAGACCATGAGAG	Gly719Gly,Ser720Phe[BE3.9Max]
0424_EGFR_Ex18_NGS_Rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCNCCAGCTTGTGGAGCCTCTTAC	Gly719Gly,Ser720Phe[BE3.9Max]
0453_EGFR_Ex25_NGS_Fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**(N)**GTGGATGCCGACGAGTACC	Exon25+1,Leu1038Leu[BE3.9Max]
0458_EGFR_Ex25_NGS_Rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCNCCATGTGAGTTTCACTAGATGGT	Exon25+1,Leu1038Leu[BE3.9Max]

Table 2.5: Primers used for target amplicon sequencing in base editing validation experiments. *N nucleotides between parentheses represent staggers of variable lengths (one to three) used to increase sequencing pool diversity on the flow cell.*

Name	Sequence	Experiment
0429_EGFR_T790_NGS_Fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**(N)**CCCGTATCTCCCTTCCCTGA	EGFR Thr790Met
0431_EGFR_T790_NGS_Rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCNGTCTTCACCTGGAAGGGGTC	EGFR Thr790Met
0290_HEK3_NGS_Fwd_1	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**(N)**CTGGCCTGGGTCAATCCTTG	HEK3 CTT insertion
0287_HEK3_NGS_Rev_1	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCNNCCCAGCCAAACTTGTCAACC	HEK3 CTT insertion
0461_HBEGF_ex5_NGS_Fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**(N)**AGCGATTTTCCACTGGGAGG	Diphteria toxin enrichment (HBEGF Glu141His)
0464_HBEGF_ex5_NGS_Rev	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCNNGGGATCGTGGTCAGGGATTG	Diphteria toxin enrichment (HBEGF Glu141His)

Table 2.6: Primers used in prime editing deep sequencing validation and enrichment experiments. *N nucleotides between parentheses represent staggers of variable lengths (one to three) used to increase sequencing pool diversity on the flow cell.*

2.8 Deep sequencing data analysis

2.8.1 Softwares and computational packages

Software or package name	Version
MAGeCK	0.5.9.2
CRISPResso	2.0.42 (Python 2.7)
PyMOL	2.5.5
R	4.3.1 (2023-06-16)
ggplot2	3.4.4
tidyverse	2.0.0
GGally	2.1.2
Python	3.6.13
Numpy	1.19.2
Biopython	1.81
Trimmomatic	0.39
DNABarcodes (Bioconductor)	1.32.0

Table 2.7: Softwares and computational packages used for data analysis and visualization.

2.8.2 Deep sequencing data preprocessing

All sequencing reads were demultiplexed and their quality was assessed using MultiQC (315). All reads were then trimmed and filtered using Trimmomatic (316) with the following parameters: ILLUMINACLIP:TruSeq3-SE:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36.

Experimental samples with less than 2000 attributed reads were excluded from further analyses.

2.8.3 Base and prime editing screen data analysis

sgRNA or epegRNA barcode count tables were obtained using a custom Python script (available at https://github.com/plattlab/). Briefly, sgRNA or barcode sequences were extracted from Illumina sequencing reads by searching for adapter sequences directly preceding and following them:

Adapter sequences used for sgRNA quantification (base editing screens):
5’ GGACGAAACACC**[sgRNA]**GTTTGAGAGCTA 3’
Adapter sequences used for epegRNA quantification (prime editing screens):
5’ CGAAAACGTT**[barcode]**TGACTCGTAT 3’

One mismatch was allowed in each adapter sequence but only perfect sgRNA or barcode matches were considered. Log fold changes and FDR values were then calculated with MAGeCK (317) using default settings after normalizing read counts to non-targeting guides or pegRNAs. sgRNAs or barcodes with an average normalized read count lower than 50 reads in any treatment arm were ignored.

2.8.4 Validation sequencing data analysis

Sequencing data were analyzed with CRISPResso2 (318) using base editing quantification. For prime editing experiments, the HDR quantification mode was used by providing expected edited products. The following parameters were used:

Base editing	
All sgRNAs	--base_editor_output --quantification_window_size 5 --quantification_window_center -16 --trim_sequences --trimmomatic_options_string MINLEN:140
Thr790Ala,Gln791Arg[ABE8e]	--guide_seq ATCACGCAGCTCATGCCCTT --conversion_nuc_from T --conversion_nuc_to C
Met766Thr[ABE8e]	--guide_seq TGGCCATCACGTAGGCTTCC --conversion_nuc_from A --conversion_nuc_to G
His773Arg[ABE8e]	--guide_seq CCCCACGTGTGCCGCCTGCT --conversion_nuc_from T --conversion_nuc_to C
Val845Ala[ABE8e]	--guide_seq TGTTTTCACCAGTACGTTCC --conversion_nuc_from A --conversion_nuc_to G
Ile853Val,Thr854Ala[ABE8e]	--guide_seq CAAGATCACAGATTTTGGGC --conversion_nuc_from T --conversion_nuc_to C
Thr790Met,Gln791Ter[BE3.9Max]	--guide_seq ATCACGCAGCTCATGCCCTT --conversion_nuc_from G --conversion_nuc_to A
Gly719Gly/SerS720Phe[BE3.9Max]	--guide_seq GGGCTCCGGTGCGTTCGGCA --conversion_nuc_from G --conversion_nuc_to A
Exon25+1,Leu1038Leu[BE3.9Max]	--guide_seq TTCATACCAGAGAGCTCAGG --conversion_nuc_from G --conversion_nuc_to A

Table 2.8: CRISPResso2 parameters used for base editing product quantification.

Prime editing	
HBEGF Glu141His	--amplicon_seq AGCGATTTTCCACTGGGAGGCTCAGCCCATGACACCTCTCTCCATGGTAACCCGGGTGGCAGCTAGTTCAAGACAGAACAAGAAGGAGATGGAGTTAGTGCTTTGGCCTCTCTTGGCAATGGCCCACCTGCATAAGCCAAACCCCATTCCTCATACCCTCAGCCTGTCAATCCCTGACCACGATCCC --guide_seq ACCCGGGTTACCATGGAGAG --expected_hdr_amplicon_seq AGCGATTTTCCACTGGGAGGCTCAGCCCATGACACCTGTGTCCATGGTAACCCGGGTGGCAGCTAGTTCAAGACAGAACAAGAAGGAGATGGAGTTAGTGCTTTGGCCTCTCTTGGCAATGGCCCACCTGCATAAGCCAAACCCCATTCCTCATACCCTCAGCCTGTCAATCCCTGACCACGATCCC
EGFR Thr790Met	--amplicon_seq CCCGTATCTCCCTTCCCTGATTACCTTTGCGATCTGCACACACCAGTTGAGCAGGTACTGGGAGCCAATATTGTCTTTGTGTTCCCGGACATAGTCCAGGAGGCAGCCGAAGGGCATGAGCTGCGTGATGAGCTGCACGGTGGAGGTGAG --guide_seq ATAGTCCAGGAGGCAGCCGA --expected_hdr_amplicon_seq CCCGTATCTCCCTTCCCTGATTACCTTTGCGATCTGCACACACCAGTTGAGCAGGTACTGGGAGCCAATATTGTCTTTGTGTTCCCGGACATAGTCCAGGAGGCAGCCGAAGGGCATGAGCTGCATGATGAGCTGCACGGTGGAGGTGAG

Table 2.9: CRISPResso2 parameters used for prime editing product quantification.

2.9 Toxin-based prime editing enrichment experiments

2.9.1 Dual pegRNA vector cloning

To obtain a lentiviral vector expressing both a pegRNA of interest and a pegRNA introducing a toxin-resistant variant in the HBEGF gene, the pLentiGuide_Puro-T2A-MLH1$^{\Delta754-756}$ vector was first digested with the XbaI restriction enzyme (Thermo Scientific, #ER0681). This linearized vector was then used as a backbone for the Gibson cloning of a double-stranded gBlock fragment (Integrated DNA technologies) containing a mouse U6 (mU6) promoter and either of the following pegRNA or epegRNA:

HBEGF Glu141His (pegRNA)	GACCCGGGTTACCATGGAGAG**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCC**ATGACACCTGTGTCCATGGTAACCCG
HBEGF Glu141His (epegRNA)	GACCCGGGTTACCATGGAGAG**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCC**ATGACACCTGTGTCCATGGTAACCCGCGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAA

Table 2.10: epegRNAs used in toxin-based prime editing enrichment experiments. The *sgRNA scaffold is shown in bold.*

The resulting vector contained the HBEGF-targeting pegRNA under the control of a mU6 promoter and a dual BsmBI cloning site following the human U6 promoter, allowing the cloning of pegRNAs of interest.

2.9.2 Toxin-based enrichment of prime edited cells

For toxin-based prime editing enrichment experiments, cells were infected with the dual-pegRNA constructs and selected with 3 µg/mL puromycin between days 2 and 4. On day 9 after transduction, cells were treated with 20 ng/mL diphtheria toxin (Merck, #D0564) until final cell harvest on day 15.

Chapter III: Base editing scanning screens

3.1 Contributions

I designed, conducted and analyzed all the experiments presented in this chapter. I also wrote all computational pipelines used for library design and screen analysis. External packages and code used for data analysis are referenced in the method section.

The **Genomics Facility Basel** (Christian Beisel, Mirjam Feldkamp, Erika Gröflin-Schürch, Ina Nissen-Naidanow, Elodie Vogel Burcklen) quantified, pooled and loaded all deep sequencing libraries.

Dr. Alessio Strano, Dr. Georgios Kalamakis and **Prof. Dr. Randall J. Platt** proofread this chapter.

Prof. Dr. Randall J. Platt proposed the initial project idea, provided fundings and contributed intellectually throughout the progress of this project with supervision and guidance.

3.2 Introduction

Assessing the pathogenicity of EGFR variants in high-throughput requires experimental models enabling the linkage of genotype to phenotype. EGFR variants are generally overexpressed in murine cell lines like Ba/F3 and NIH/3T3 which present the advantage of acquiring measurable growth phenotypes in response to EGFR signaling. However, these cell lines do not recapitulate the cellular context of human cancer, thus potentially missing important interactions and regulation mechanisms. In addition, exogenous delivery of EGFR does not accurately replicate its physiological expression level and genomic context.

In this chapter, we set out to use base editing to assess EGFR variant pathogenicity in the genome of human cells while preserving endogenous expression levels. To this end, a first step is the identification of human cell lines that endogenously express wild-type EGFR and whose proliferation is dependent on its signaling. The cell lines under consideration should not harbor mutations in other oncogenes that confer intrinsic resistance to TKIs, nor should they exhibit significant EGFR copy number amplification, a common occurrence in cancer cell lines (319). Indeed, the ideal cellular model should have as few copies of EGFR as possible to allow for reproducible genomic engineering of EGFR variants and quantitative measurement of their associated phenotype.

Various cell lines derived from cancer, particularly NSCLC, are commercially available. An example is NCI-H1299 which is derived from a lung carcinoma and harbors WT EGFR. EGFR^{Leu858Arg} overexpression in this cell line was shown to confer a growth advantage in colony formation assays (320) and increased EGFR tyrosine phosphorylation (321). However, these exhibit a significant tolerance to TKIs regardless of the expressed EGFR variant, suggesting the existence of an EGFR-independent resistance mechanism. Another example is the NCI-H2073 line which harbors WT EGFR and was shown to be sensitive to different TKIs as measured by relative proliferation rates (289). The introduction of EGFR exon 20 insertions in the genome of these cells was shown to confer them a growth advantage under serum deprivation. However, they show very slow growth, thus complicating their expansion for large-scale experiments (322).

MCF10A is a spontaneously immortalized epithelial cell line isolated from a benign mammary tumor. These cells represent an attractive model because they express wild type EGFR and rely on its signaling for growth. Indeed, MCF10A cells overexpressing hyperactive EGFR variants have been shown to proliferate in the absence of EGF supplementation while cells expressing WT EGFR remain quiescent (323). Additionally, MCF10A are intrinsically sensitive to multiple TKIs and can be further sensitized or acquire resistance through the expression of EGFR mutations like Leu858Arg or Thr790Met, respectively (324). Lastly, unlike many cancer cell lines, MCF10A cells are near-diploid and show no copy number amplification of EGFR, making them an ideal model for gene editing (325).

Many base editors have been developed with different nucleotide specificities and editing windows (154). For our base editing scanning experiment, we decided to evaluate a cytosine base editor (CBE) and an adenine base editor (ABE), which represent the most established categories of base editors to date. ABE8e is the most recent enzyme in the ABE family, displaying higher overall efficiency and uniform editing across its editing window in comparison to earlier iterations (326). Similarly, BE4max has been optimized for increased nuclear localization and activity (327). However, the similar BE3.9Max, which contains a single uracil DNA glycosylase inhibitor domain instead of two, has been shown to outperform BE4max at capturing essential genes in a screening context (145).

In this chapter, we thus set out to perform ABE8e and BE3.9max scanning screens using an sgRNA library tiling the full EGFR coding region. We envisioned that introducing EGFR activating mutations in the genome of MCF10A cells could make edited cells EGF-independent. Consequently, these cells would outcompete unedited cells when deprived of EGF, thereby enabling the connection of genotype to phenotype in a pooled screen. This approach then enabled the assessment of the relative sensitivities of the considered variants to clinically approved TKIs and identified drug-resistant mutations. Lastly, we transposed our screen to a NSCLC cell line harboring an hyperactive EGFR variant and identified cell line-specific drug-resistant variants.

3.3 Pathogenic EGFR variant modeling in MCF10A cells

First, we evaluated base editing efficiency in MCF10A cells by lentiviral delivery of BE3.9max or ABE8e along with an sgRNA predicted to mutate the “gatekeeper” residue Thr790. Illumina sequencing of the target exon revealed efficient editing at the target site with 97% or 95% of reads containing at least one base change 6 days after infection for BE3.9max or ABE8e, respectively (Figure 3.1). Although the total fraction of edited alleles remained stable over time, the relative proportions of editing products varied between day 6 and day 13 with alleles containing 2 or 3 edits being enriched as opposed to single edits. This is likely explained by sequential base editing of less preferred collateral bases through time as the preferred base conversion has already been installed on most alleles. This potentially complicates the accurate prediction of variant effects, but on the other hand also serves to broaden the mutational spectrum that can be assessed in a single experiment. We thus next set out to leverage CBE and ABE base editors to explore the EGFR variant landscape in MCF10A cells.

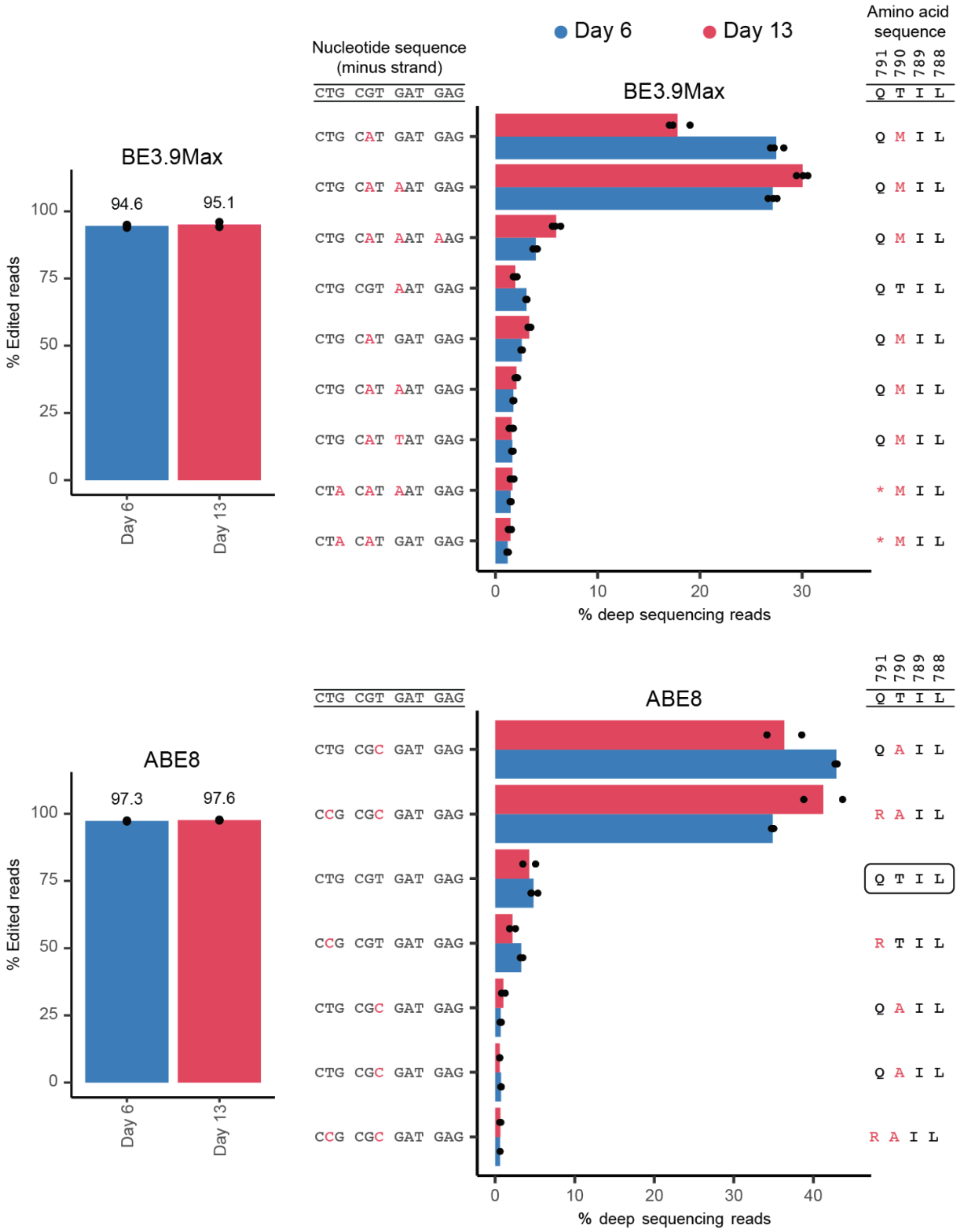


Figure 3.1: **Validation of ABE8e and BE3.9Max editing efficiencies in MCF10A cells.** *Editing efficiencies and read distributions from base editing experiments in MCF10A. An sgRNA targeting Thr790 was delivered with BE3.9Max (n =3) or ABE8e (n =2), target sequencing was performed on day 6 and day 13 after lentiviral infection. Edited nucleotides and amino acids are shown in red, black boxes show unedited sequences.*

After confirming the ability of BE3.9Max to install the Thr790Met amino acid substitution in EGFR, we set out to determine whether this mutation confers TKI resistance to cells. We thus repeated the CBE experiment in MCF10A cells, this time subjecting the cells to Gefitinib treatment for 7 days before deep sequencing. By comparing read proportions between the treated and untreated arms, we observed significant enrichment of reads containing Thr790Met under drug treatment. By contrast, wild-type alleles were significantly depleted, confirming that unedited cells grow slower in presence of Gefitinib.

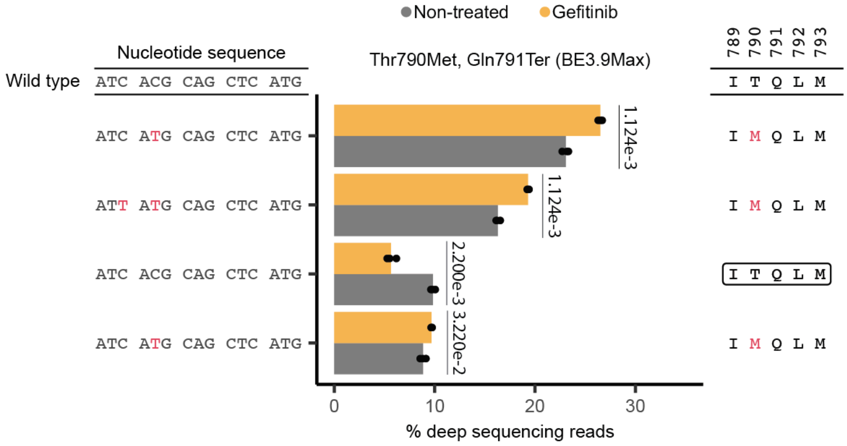


Figure 3.2: The introduction of Thr790Met by base editing confers TKI resistance to MCF10A cells. *MCF10A cells were infected in triplicates with a lentiviral vector expressing BE3.9Max and an sgRNA predicted to introduce the Thr790Met amino acid substitution. Relative proportions of the top four deep sequencing reads at the target site are shown for non-treated controls and cells treated with Gefitinib for 7 days. Edited nucleotides and amino acids are shown in red, black boxes show WT EGFR sequences and p-values are shown for two-sided unpaired t-tests with Holm correction.*

We thus confirmed the ability of ABE8e and BE3.9Max to efficiently introduce mutations in the genome of MCF10A cells as well as to elicit measurable phenotypes in this cell line with regard to TKI drug resistance.

3.4 Identification of EGFR loss of function variants

To assess the pathogenicity of a spectrum of EGFR mutations, we designed a base editing variant scanning library composed of 1496 unique sgRNAs targeting all EGFR exons (Figure 3.3A). We also included 200 sgRNAs targeting an EGFR intronic region and 206 non-targeting sgRNAs as negative controls, as well as 103 sgRNAs targeting splice sites of essential genes as positive editing controls. In order to explore a broad range of genetic variants accessible via base editing, we cloned this library under the control of a U6 promoter in two different lentiviral backbones expressing either ABE8e or BE3.9max. The two resulting all-in-one base editing libraries were quantified by Illumina sequencing and showed no drop-outs as well as a narrow sgRNA distribution with skew ratios of 1.6 (Supplementary Figure S3.1, Supplementary Figure S3.2).

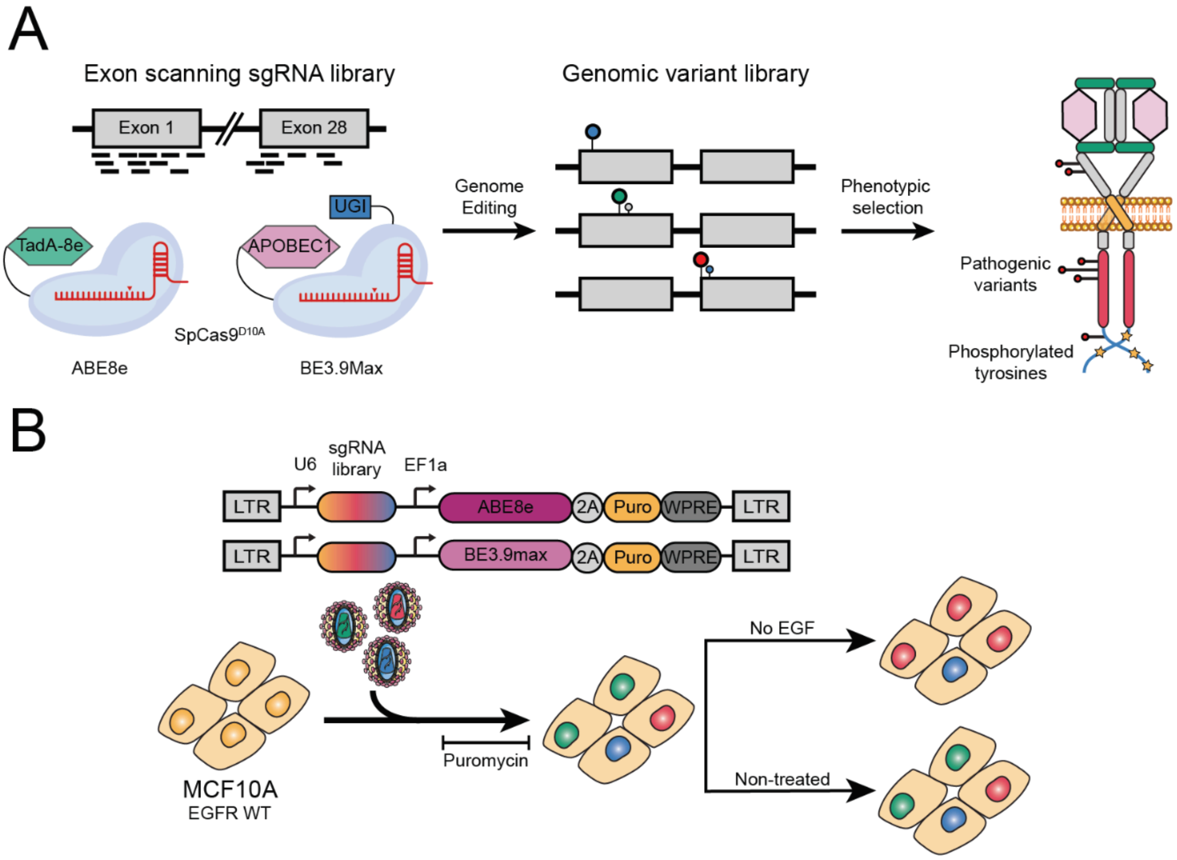


Figure 3.3: **Base editing scanning screens design.** *(A) Schematic of the base editing mutational scanning approach applied to EGFR. (B) Overview of the base editing constructs and workflow to identify activating EGFR mutations. All-in-one base editing libraries were delivered to MCF10A cells in duplicate. Cells were split into non-treated and EGF-deprived treatment arms on day 11 and final harvest was performed on day 19.*

We then delivered both libraries to MCF10A cells via lentiviral infection (Figure 3.3B). Each infection was performed in duplicate at a multiplicity of infection (MOI) of 0.3 with an average coverage of 500 cells per sgRNA. After puromycin selection of transduced cells and EGF deprivation, genomic DNA was extracted and sgRNA libraries were prepared and sequenced while ensuring appropriate coverage and depth. The sgRNA counts were quantified from sequencing data using a custom script and normalized to non-targeting sgRNAs with MAGeCK (317). We then confirmed library coverage and replicate correlation at every time point (Supplementary Figure S3.1, Supplementary Figure S3.2), overall indicating a robust dataset warranting further analysis.

We started our analysis by comparing the relative abundances of sgRNAs between the plasmid library and late (day 19) timepoint in the non-treated arm, which should reveal sgRNAs that impact MCF10A viability. As expected, we observed no change in the non-targeting negative control sgRNAs but a strong depletion of gene editing positive control sgRNAs targeting splice sites of essential genes with both base editors (Figure 3.4). ABE8e and BE3.9Max screens distinguished between essential splice site controls and other sgRNAs with area under the curve (AUC)s of 0.9 and 0.94, respectively (Supplementary Figure S3.3) confirming the ability of our pooled library to efficiently introduce edits and elicit measurable phenotypes.

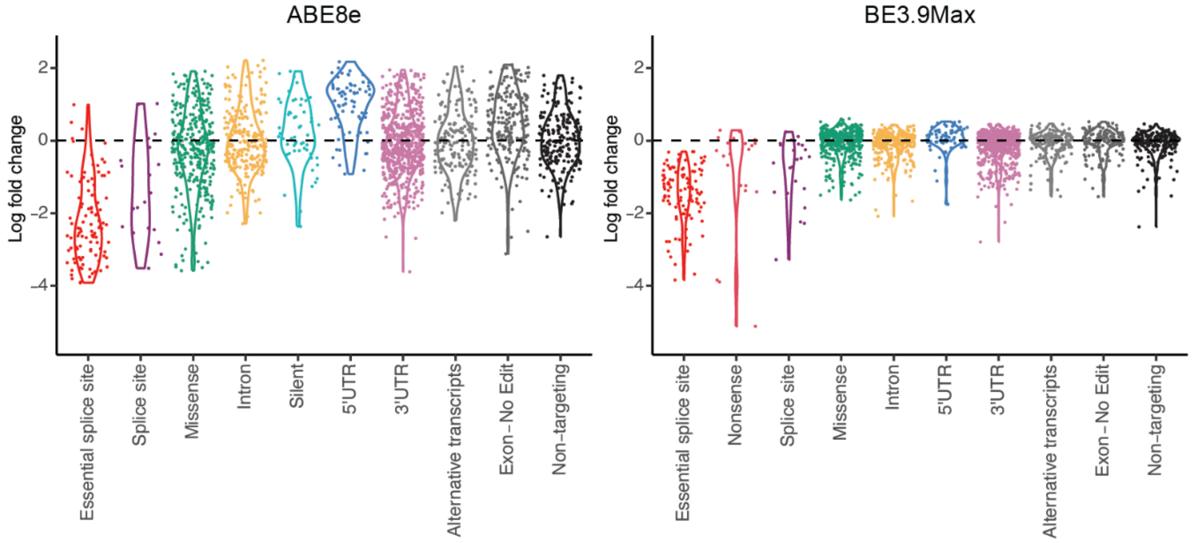
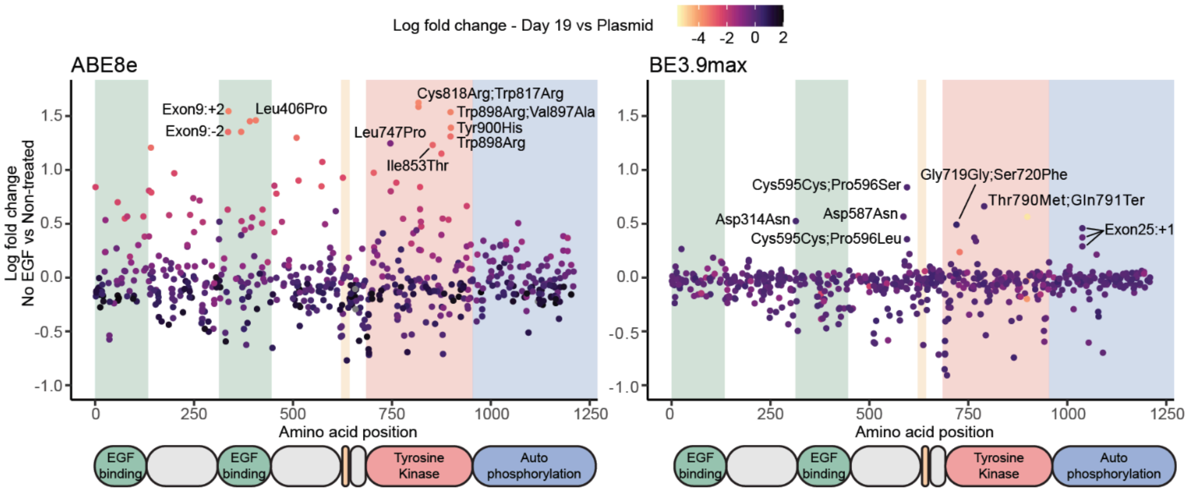


Figure 3.4: Impact of individual sgRNAs on cell viability for the base editing EGFR activation screen in MCF10A cells. *Log fold changes of individual sgRNAs between plasmid and day 19 (non-treated) for each sgRNA target category and predicted mutation type.*

Encouragingly, multiple sgRNAs predicted to introduce splice site, nonsense, and missense mutations in EGFR were significantly depleted at day 19 with both base editors. Loss of function (LOF) genetic variants are indeed expected to impact the viability of MCF10A cells, which rely on EGFR signaling for growth. Together, these data confirm efficient base editing of endogenous loci in MCF10A cells and our capacity to robustly detect LOF variants in EGFR.

3.5 Identification of oncogenic EGFR variants

Towards identifying pathogenic EGFR variants that lead to constitutive EGFR signaling in MCF10A cells, we compared sgRNA library distributions in non-treated and EGF-deprived cells. This analysis revealed a spectrum of enriched and depleted EGFR variants, largely localized to specific functional protein domains (Figure 3.5).



<u>Figure 3.5</u>: Loss-of-function EGFR variants act as confounders in EGF-independent growth screens. *Scatterplot showing the log fold changes (LFC) of individual sgRNAs along the EGFR protein between non-treated and EGF-deprived cells. Log-fold changes between the plasmid library and non-treated cells on day 19 are shown in color gradients. Background colors represent EGFR domains.*

Among the set of sgRNAs enriched in the EGF-deprived condition, we observed some of the sgRNAs previously identified to impact cell viability. Examples include variants predicted to lead to protein truncation through splice site mutations like Exon9(+2)[ABE8e] or Exon14(+2)[ABE8e], as well as Ile853Thr[ABE8e] which has been shown to abolish EGFR kinase activity and autophosphorylation in response to EGF stimulation (328). In addition, we observed that sgRNAs targeting essential splice sites appear enriched under EGF deprivation and that this enrichment seems to correlate with their impact on cell viability (Supplementary Figure S3.4). Taken together, these findings suggest that cells with decreased viability are outgrown more rapidly by healthy cells in non-treated samples, thus making them appear enriched in EGF-deprived samples.

To focus our analysis on activating mutations in EGFR and avoid biasing our results towards confounding LOF variants impacting MCF10A viability, we chose to only look at variants preserving cell viability in the presence of EGF. Namely, all sgRNAs with a log fold change (LFC) inferior to -0.6 when comparing plasmid and late timepoints were excluded in order to leave out >99% of essential splice site controls (Figure 3.6, Supplementary Figure S3.4). This approach identified 19 hits that, while spanning the protein, mostly localized to the tyrosine kinase domain, which is responsible for EGFR autophosphorylation and contains the residues most commonly mutated in NSCLC. Of the 19 hits, 10 were found in ClinVar and labeled as either pathogenic (2) or VUS (8). A further 2 were found in COSMIC and 7 were not listed in either database (Table 3.1).

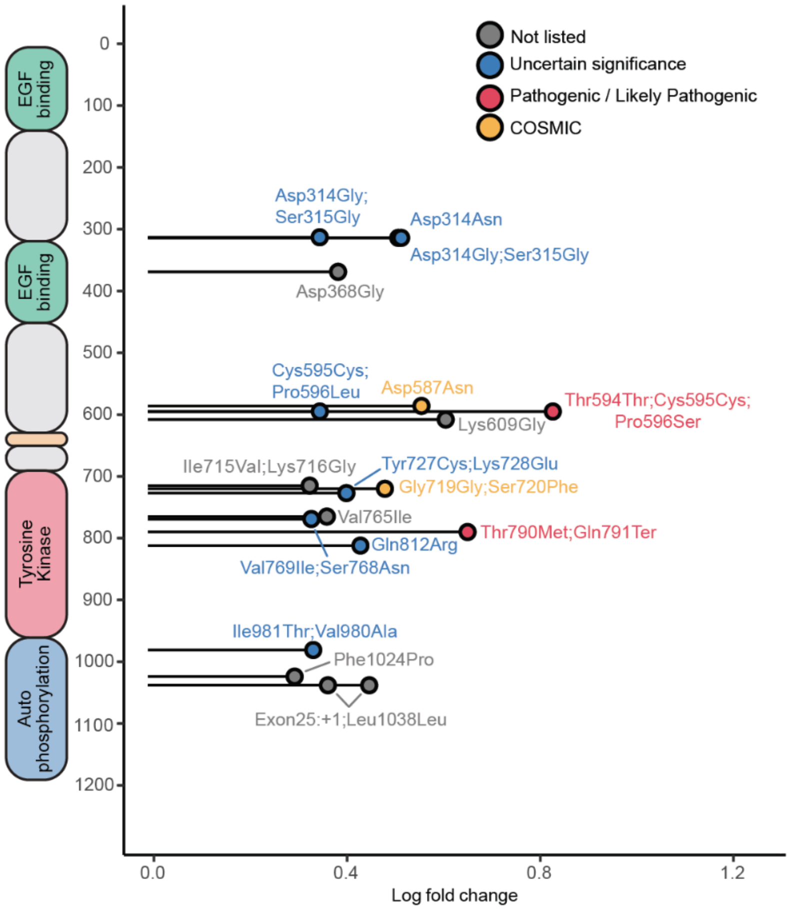


Figure 3.6: Base editing scanning screens identify EGFR activating mutations in MCF10A cells. *Diagram showing combined ABE8e and BE3.9Max screen hits along the EGFR protein. Hits are defined as sgRNAs with LFC > -0.6 between plasmid and day 19 and LFC > 0.3 between non-treated and EGF-deprived cells. Colors represent the classification of the predicted amino acid changes, the COSMIC category is only shown if the mutation is not listed in ClinVar.*

Our screen identified well known pathogenic variants like Thr790Met and Pro596Ser which are both reported as pathogenic in ClinVar. Surprisingly, the majority of other hits introduce mutations that were previously observed in tumor samples but are currently not considered as pathogenic. For example, Ser720Phe is absent from ClinVar but was previously observed in NSCLC patients (329,330) and is adjacent to Gly719 which is a commonly mutated residue in the kinase domain (296). Several additional hits not considered as pathogenic such as Val765Ile[BE3.9Max] and Val769Ile,Ser768Asn[BE3.9Max] affect the αC helix and the αC-β4 loop which are a key regulatory structures for the activation of this domain. Notably, Ser768 and Val769 are conserved residues within the Erbb family which respectively interact with residues of the juxta-membrane region and C-lobe to stabilize the active conformation of the αC helix (331,332). In addition, Ser768 phosphorylation by CAMK2 inhibits EGFR activity and multiple substitutions of this residue like Ser768Ile are known to activate the receptor and confer drug resistance (332,333).

sgRNA sequence	Editor	Log-fold change	FDR	Predicted amino acid change	Database category	Database accession	Comments	References
GTCTGGAAGTACGCAGACGC	ABE8e	0.61674	1.35E-12	Lys609Gly	Not listed		Tyr270, Asp587 and Lys609 form hydrogen bonds contributing to the auto-inhibitory conformation of the inactive receptor.	(334,335)
GGCCGACAGCTATGAGATGG		0.51856	5.33E-13	Asp314Gly, Ser315Gly	Uncertain significance (Ser315Gly)	COSV51810209	Asp314Gly was observed in breast carcinoma.	
GCTCCCAGTACCTGCTCAAC		0.44004	6.79E-07	Gln812Arg	Uncertain significance	VCV002418796.2	Observed in NSCLC patients.	(336)
GGTGTATAAGGTAAGGTCCC		0.41104	3.47E-10	Tyr727Cys, Lys728Glu	Uncertain significance (Tyr727Cys)	VCV001026153.5	Observed in NSCLC patients. Tyr727 is a phosphorylation site of the tyrosine kinase domain.	(337–339)
GCGATCTCCACATCCTGCCGG		0.39385	1.23E-03	Asp368Gly	Not listed			
GTGGGGCCGACAGCTATGAGA		0.35574	2.66E-04	Asp314Gly, Ser315Gly	Uncertain significance (Ser315Gly)	VCV001036295.5	Asp314Gly was observed in breast carcinoma.	
GCTGAATGACAAGGTAGCGCT		0.34238	3.51E-03	Ile981Thr, Val980Ala	Uncertain significance (Ile981Thr)	VCV002105917.1		
GAAGATCAAAGTGCTGGGCTC		0.33496	6.16E-04	Ile715Val, Lys716Gly	Not listed		Lys716 is a ubiquitination site.	(340,341)
GCTGCTGAAGAAGCCCTGCTG		0.30384	4.20E-04	Phe1024Pro	Not listed			
GACCTGCCCGGCAGGAGTCA	BE3.9Max	0.83855	3.13E-127	Thr594Thr, Cys595Cys, Pro596Ser	Pathogenic (Pro596Ser)	VCV002582280.1	Pro596Ser was observed in glioma patients. Pro596 is located in a loop stabilizing the interaction between Asp587 and Lys609 and contributing to receptor auto-inhibition.	(334)
GATCACGCAGCTCATGCCCTT		0.66193	2.27E-80	Thr790Met, Gln791Ter	Drug response (Thr790Met)	VCV000016613.27	Thr790 is the gatekeeper residue of the tyrosine kinase ATP-binding site. Thr790Met is the most prevalent secondary mutation conferring resistance to first-generation TKIs . It has also been shown to increase the EGFR kinase activity and to be oncogenic on its own.	(270,271)
GCGTCAATGTAGTGGGCACAC		0.56653	1.94E-22	Asp587Asn	COSMIC	COSV51849240	Observed in glioma patients. Tyr270, Asp587 and Lys609 form a salt bridge contributing to the auto-inhibitory conformation of the inactive receptor.	(334,335)
GAGCTGTCGGCCCCACAGGCT		0.52444	9.07E-42	Asp314Asn	Uncertain significance	VCV001440221.4	Observed in nasopharyngeal carcinoma patients.	(342)
GGGCTCCGGTGCGTTCGGCA		0.49083	4.44E-45	Gly719Gly, Ser720Phe	COSMIC (Ser720Phe)	COSV51780206	Ser720Phe was observed in NSCLC patients.	(329,330)
GCATACCAGAGAGCTCAGGAG		0.45833	5.69E-33	Exon25:+1, Leu1038Leu	Not listed		Different EGFR C-terminal truncated variants have been observed in glioblastoma. EGFRΔEx26-28 has been shown to promote EGFR signaling and confer anchorage-independent growth in absence of EGF in NIH-3T3 cells.	(343–345)
GTCATACCAGAGAGCTCAGGA		0.3728	8.19E-16	Exon25:+1, Leu1038Leu	Not listed			
GCATCACGTAGGCTTCCTGGA		0.3705	3.44E-20	Val765Ile	Not listed		Located in the regulatory αC-helix. Val765-Met766>X is a similar TKI-resistant indel found in NSCLC.	(346)
GACCTGCCCGGCAGGAGTCAT		0.35593	9.81E-15	Cys595Cys, Pro596Leu	Uncertain significance (Pro596Leu)	VCV001058870.5	Pro596 is located in a loop stabilizing the interaction between Asp587 and Lys609 and contributing to receptor auto-inhibition. It was observed in glioma patients.	(334,347)
GTCCACGCTGGCCATCACGT		0.33849	1.88E-13	Val769Ile, Ser768Asn	Uncertain significance (Ser768Asn)	VCV001015382.5	Ser768 and Val769 are conserved residues located in the loop following the regulatory αC-helix of the kinase domain. They interact with the juxta-membrane region and C-lobe to stabilize the active conformation of the αC helix. Ser768 phosphorylation by CAMK2 inhibits EGFR activity. Multiple Ser768 variants like Ser768Ile are known to activate EGFR and confer drug resistance.	(331–333,346)

Table 3.1: Screen hits of the base editing EGFR activation screen.

Interestingly, the co-occurence of certain screen hits provides insights into the role of autoregulation and phosphorylation in EGFR activation. For example, Asp587 and Lys609 are known to form a salt bridge further stabilized by a loop containing Pro596. This interaction is known to contribute to the auto-inhibitory conformation of the inactive receptor (334), suggesting that its disruption may favor EGFR constitutive activation and constitute a mechanism of transformation. As another example, Ile715Val,Lys716Gly[ABE8e] impacts the Lys716 ubiquitination site (340,341) and Tyr727Cys,Lys728Glu[ABE8] contains a phosphorylation site (337), suggesting a role of these post-translational modifications in EGFR regulation and stability.

Our most surprising result was a cluster of hits found in the C-terminal tail. Among them, we found two Exon 25 splice donor mutations predicted to lead to the C-terminal truncation of the receptor after Exon 25 (EGFRΔEx26-28). Although the autophosphorylation of this domain is essential for signal transduction of WT EGFR stimulated with EGF, it appears not to be required for the downstream signaling of mutant EGFR canonically associated with oncogenesis (348). On the other hand, different truncated EGFR variants have been identified in glioblastoma (343,344), but not NSCLC, and have been reported to be associated with increased receptor activation due to the loss of an auto-inhibitory region of the C-terminal tail (345,349). Another study showed the activating potential of EGFRΔEx26-28 in NIH-3T3 cells which is thought to be due to the role of the C-terminal tail in receptor internalization and degradation (345).

In order to validate that C-terminal truncations lead to EGFR activation, we delivered one of the Exon 25 truncating sgRNAs into MCF10A cells and let selected cells grow in the absence of EGF for 5 days before measuring their viability (Figure 3.7A). We observed that infected cells edited to express EGFRΔEx26-28 displayed significantly higher EGF-independent growth compared to uninfected cells, validating the base editing screen result. To further confirm that the EGFR splice site mutation was driving the observed EGF-independent growth, we then sequenced the exon 25 target site of infected cells and compared read distributions between EGFR-deprived and non-treated samples (Figure 3.7B). In the treated arm, we observed a significant depletion of reads with a wild-type splice site while the two alleles with mutated splice sites appeared enriched.

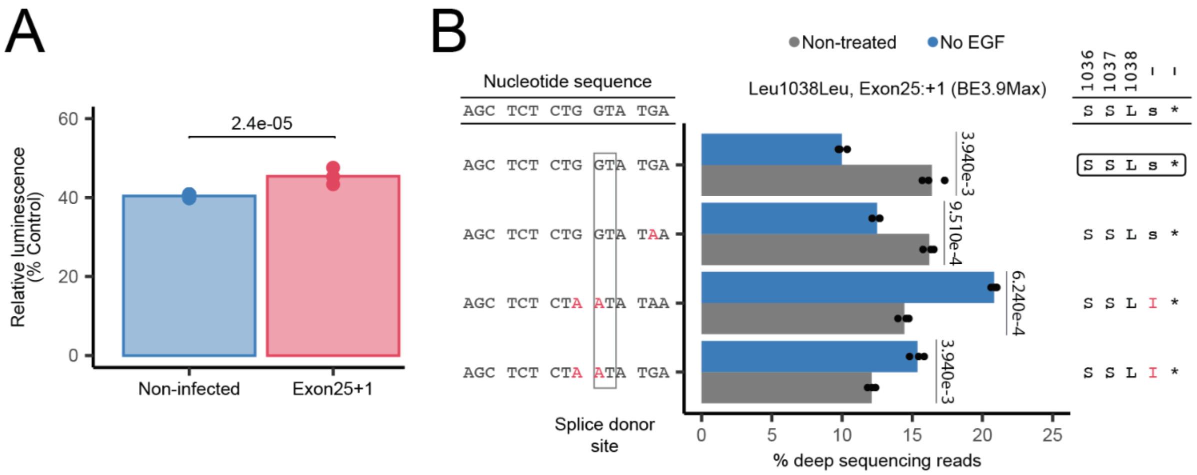


Figure 3.7: **EGFR activation screen hits validation.** *(A) Viability comparison between non-infected cells and cells infected with a sgRNA introducing the Leu1038Leu,Exon25+1[BE3.9Max] mutation as measured by the CellTiter Glo 2.0 assay after 5 days of EGF deprivation. Relative luminescence intensities are compared to non-treated cells seeded at the same time and shown with unpaired t-test p-value (n= 3). (B) Deep sequencing data for Leu1038Leu,Exon25+1[BE3.9Max] hit validation. Relative proportions of the top four deep sequencing reads at the target site are shown for non-treated controls and EGF-deprived cells after 7 days of selection (n= 3). Edited nucleotides and amino acids are highlighted, black boxes show WT EGFR sequences and p-values are shown for two-sided unpaired t-tests with Holm correction.*

Taken together, these results demonstrate the ability of base editing mutational scanning to identify known and unknown EGFR activating variants. Notably, in spite of the limited mutational spectrum introduced by each base editor, our screens identified key domains, residues and post-translational modifications likely involved in receptor regulation and stability. Additionally, using plasmid libraries as a reference, we successfully identified EGFR LOF variants that confound the measured phenotype, thus highlighting the importance to systematically evaluate variant fitness when assessing relative cell proliferation in a pooled manner. In conclusion, base editing scanning screens in the MCF10A cell line offer new opportunities to interrogate multiple aspects of EGFR biology, including structural variants, and broaden the spectrum of likely pathogenic mutations.

3.6 Drug-resistant variant discovery in MCF10A cells

Encouraged by these results, we then set out to expand our screening approach to evaluate the sensitivity of EGFR variants to clinically-approved TKIs in the MCF10A cell line. We thus repeated both base editing screens in the same conditions but replaced the EGF depletion arm with either Gefitinib or Osimertinib treatments, which are first- and third-generation TKIs, respectively (Figure 3.8). Both compounds compete with ATP for binding to the tyrosine kinase active site but employ distinct inhibition mechanisms (244,257). We thus set out to apply our base editing mutational scanning approach to dissect drug-specific resistance profiles.

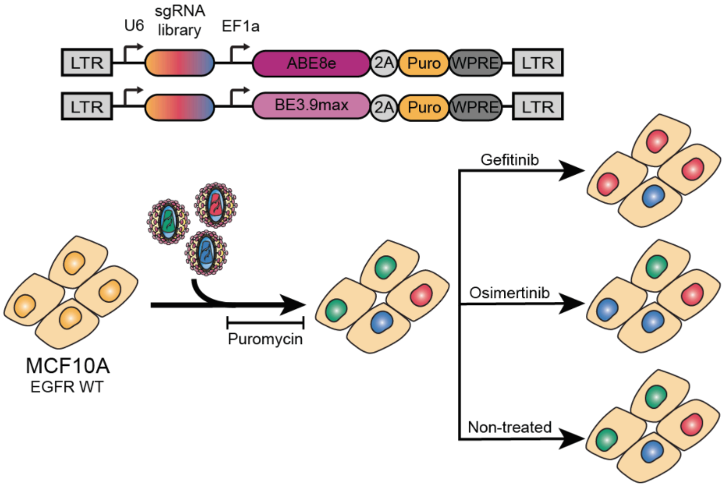


Figure 3.8: **Base editing scanning screen for TKI drug resistance.** *Overview of the base editing constructs and workflow to identify EGFR variants resistant or sensitive to clinically-approved TKIs Gefitinib or Osimertinib. All-in-one base editing libraries were delivered to MCF10A cells in duplicate. Cells were split into three treatment arms on day 11: non-treated, 0.13µM Gefitinib or 0.3µM Osimertinib. Final harvest was performed on day 19.*

We delivered the BE3.9Max and ABE8e EGFR variant scanning libraries to MCF10A cells at MOIs of 0.3 and selected for infected cells with puromycin. We then applied Osimertinib or Gefitinib treatment for 8 days before sgRNA library preparation and sequencing. Library quantification was performed as previously and we confirmed high replicate correlation (Supplementary Figure 3.5, Supplementary Figure 3.6).

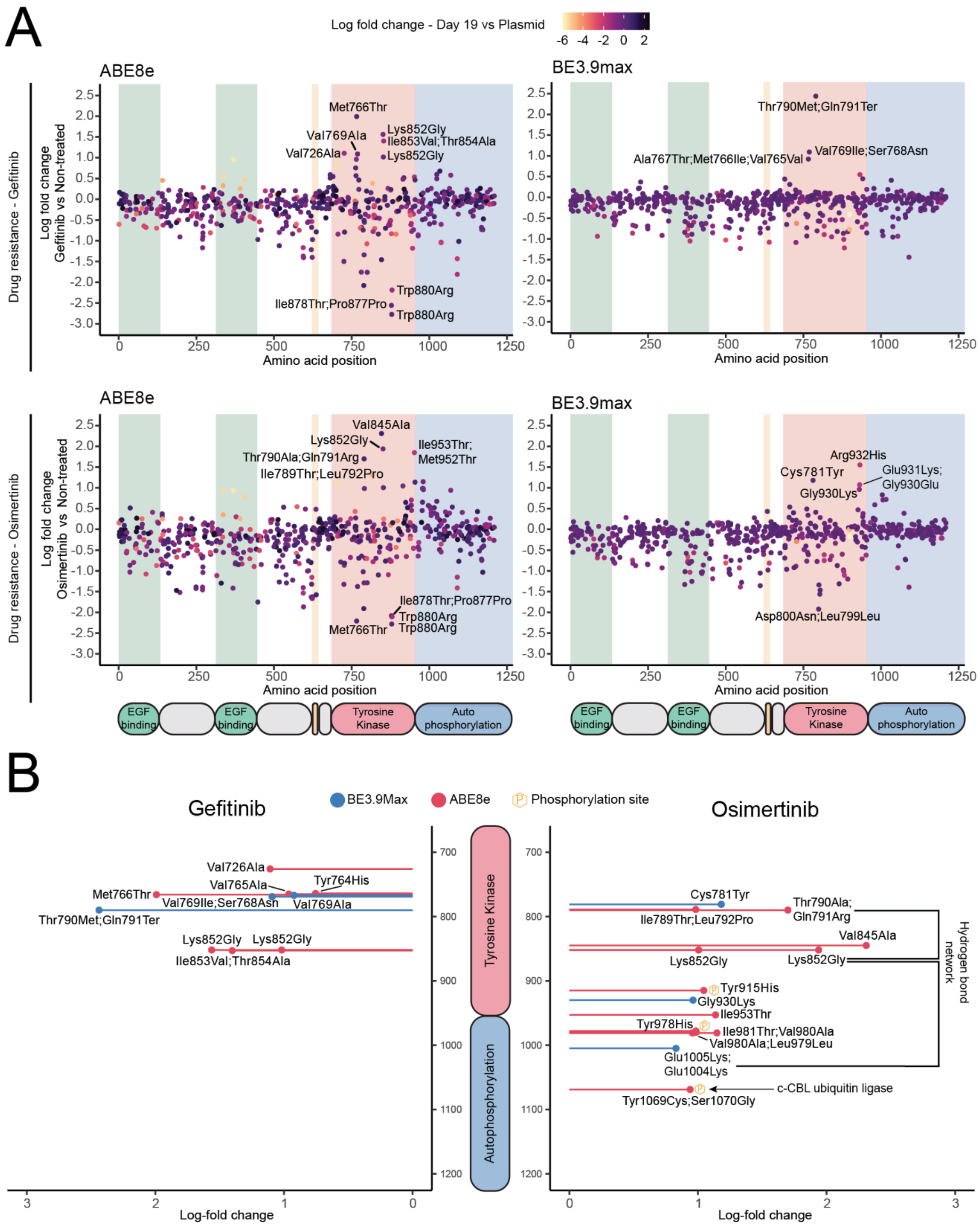


Figure 3.9: Base editing scanning screens identify EGFR variants resistant to first- and third-generation tyrosine kinase inhibitors. *(A) Scatterplot showing the log fold changes (LFC) of individual sgRNAs between non-treated and drug-treated cells. Log-fold changes between the plasmid library and non-treated cells on day 19 are shown in color gradients. Background colors represent EGFR domains. (B) Diagram showing combined ABE8e and BE3.9Max screen hits along the EGFR tyrosine kinase and C-terminal domains for cells treated with Gefitinib or Osimertinib. Hits are defined as sgRNAs with LFC > -1.5 between plasmid and day 19 and LFC > 0.75 between non-treated and TKI-treated cells. Notable post-translational modifications and intramolecular interactions are shown.*

We started our analysis by comparing sgRNA enrichment between treated and non-treated samples while accounting for variant fitness (Figure 3.9A, Supplementary Figure S3.7). Encouragingly, we observed that with both drug treatments a majority of significantly enriched hits were located in the tyrosine kinase domain and included well known drug-resistant variants. For example, Thr790Met is known as the most prevalent Gefitinib-resistant variant and we observed a strong enrichment of the Thr790Met,Gln791Ter[BE3.9Max] sgRNA under Gefitinib treatment. Importantly, this same variant is not enriched under Osimertinib treatment, which is expected as this molecule was specifically developed to counter its emergence. Unexpectedly, Osimertinib selection led to the enrichment of different predicted variants at the same position, namely Thr790Ala,Gln791Arg[ABE8e]. We thus set out to validate this result in a follow-up experiment and measured the viability of MCF10A cells infected with individual sgRNA and base editor pairs followed by TKI treatment for 5 days (Figure 3.10). As predicted by the screen, cells infected with Thr790Ala,Gln791Arg[ABE8e] showed higher resistance to Osimertinib than to Gefitinib while Thr790Met,Gln791Ter[BE3.9Max] showed the opposite resistance profile. This confirms both the specificity and sensitivity of our base editing variant screening approach and highlights the need to understand drug resistance at single base resolution.

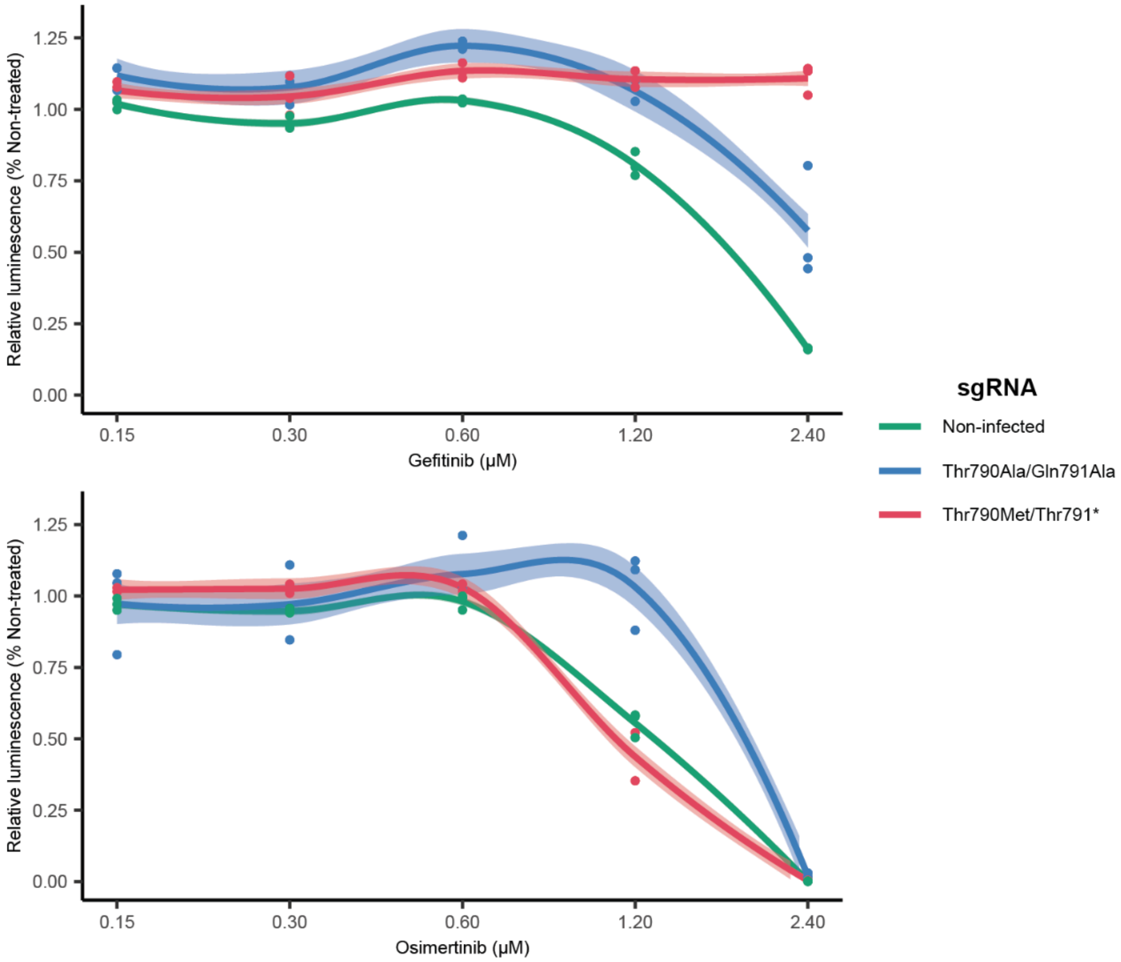


Figure 3.10: TKI response validation of drug resistance screen hits in MCF10A cells. *Relative viabilities of edited and unedited MCF10A cells after (A) Gefitinib or (B) Osimertinib treatment as measured by the CellTiter Glo assay. Cells were infected in triplicates and treated for 5 days with 0, 0.15, 0.3, 0.6, 1.2 or 2.4 μM of the respective drug before viability measurements. Shaded areas represent the standard error.*

Under Gefitinib selection, all enriched screen hits were located in the tyrosine kinase domain and more specifically around the ATP binding pocket which constitutes the binding site of both drugs (Figure 3.9B, Table 3.2). In addition to the Thr790 gatekeeper residue, we identified previously unknown resistant mutations found to affect Val726, Met766 and Thr854, which are all in direct contact with the receptor-bound molecule (Figure 3.11A). We can thus speculate that mutating these residues can directly affect the binding affinity of Gefitinib. Importantly, Thr854Ala is classified as VUS although it has been observed as an acquired resistant mutation in NSCLC patients treated with first-generation TKIs (350). Similarly, Val726Met is not listed in either database while Met766Thr has been shown to be resistant to Gefitinib *in vitro* but is not listed in ClinVar (351). To validate this hit, we delivered a lentiviral construct containing the Met766Thr[ABE8e] sgRNA to MCF10A cells and treated them with Gefitinib for 7 days, similarly to the screening protocol. Illumina sequencing of the target exon revealed a significant enrichment of reads containing the Met766Thr edit under Gefitinib treatment while wild-type alleles were depleted, thus confirming the resistance phenotype (Figure 3.11B).

Similarly to EGFR activation, our Gefitinib resistance screens also identified hits located in the αC-helix like Tyr764His[ABE8e] and Val765Ala[ABE8e] which are not listed in variant databases although both residues are frequently affected by exon 20 insertions. Likewise, Val769Ile,Ser768Asn[BE3.9Max] was previously found to lead to EGFR activation by affecting residues involved in αC-helix stabilization and rotation. These results thus confirm the critical role of the αC helix positioning in both EGFR activation and sensitivity to TKIs.

sgRNA sequence	Editor	Log-fold change	FDR	Predicted amino acid change	Database category	Database accession	Comments	References
GTGGCCATCACGTAGGCTTCC	ABE8e	1.9922	0	Met766Thr	COSMIC	COSV51815137	Observed in NSCLC patients. Met766 is located in the regulatory αC-helix and is part of the Gefitinib binding pocket. Met766Thr was previously shown to confer Gefitinib resistance in vitro.	(351,352)
GCATGTCAAGATCACAGATTT		1.5626	3.77E-107	Lys852Gly	Not listed		In the active conformation Lys852 is involved in an hydrogen-bond network with Gln791, Asp1012 and Asp1014. Destabilizing this network has been shown to reduce Osimertinib binding affinity.	(332,353)
GCAAGATCACAGATTTTGGGC		1.4027	2.26E-39	Ile853Val, Thr854Ala	Uncertain significance (Thr854Ala)	VCV001394709.2	Thr854Ala was observed as an acquired resistant mutation in NSCLC patients treated with 1st generation TKIs. Thr854 is part of the Gefitinib binding pocket.	(350,352,354)
GATACACCGTGCCGAACGCAC		1.1084	3.24E-25	Val726Ala	Not listed		Val726Ala is not listed in databases. However, Val726Met was reported to be insensitive to Gefitinib in NSCLC patients. Val726 has been shown to be directly involved in Gefitinib binding to WT EGFR.	(355,356)
GTCCACGCTGGCCATCACGT		1.0903	1.15E-42	Val769Ala	COSMIC	COSV51782791	Val769 is located in the loop following the regulatory αC helix of the kinase domain and frequently affected by pathogenic and drug-resistant insertions. A NSCLC patient with Val769Ala was reported to respond positively to Erlotinib treatment.	(346,357)
GATGTCAAGATCACAGATTTT		1.018	3.47E-34	Lys852Gly	Not listed		In the active conformation Lys852 is involved in an hydrogen-bond network with Gln791, Asp1012 and Asp1014. Destabilizing this network has been shown to reduce Osimertinib binding affinity.	(332,353)
GCATCACGTAGGCTTCCTGGA		0.96265	7.18E-26	Val765Ala	Not listed		Located in the regulatory αC-helix. Val765-Met766>X is a similar TKI-resistant indel found in NSCLC.	(346)
GACGTAGGCTTCCTGGAGGGA		0.75464	2.95E-17	Tyr764His	Not listed		Located in the regulatory αC-helix. Tyr764 is affected by the A763-Y764>FQEA insertion commonly found in NSCLC.	(346)
GATCACGCAGCTCATGCCCTT	BE3.9Max	2.438	0	Thr790Met, Gln791Ter	Drug response	VCV000016613.27	Thr790Met is the most prevalent TKI-resistant acquired mutation. It also has the strongest resistance to first-generation TKI.	(253)
GTCCACGCTGGCCATCACGT		1.0931	1.14E-81	Val769Ile, Ser768Asn	Uncertain significance (Ser768Asn)	VCV001015382.5	Located the loop following the regulatory αC-helix of the kinase domain. Ser768 phosphorylation by CAMK2 inhibits EGFR activity. Multiple Ser768 variants are known to activate EGFR and confer drug resistance. Ser768Ile in particular has been shown to be resistant to Gefitinib in vitro.	(273,332,333, 346)
GTGGCCATCACGTAGGCTTCC		0.9223	1.82E-53	Ala767Thr, Met766Ile, Val765Val	COSMIC (Met766Ile)	COSV51804547	Met766Ile was observed in melanoma and is part of the Gefitinib binding pocket.	(352,358)

Table 3.2: Screen hits of the base editing Gefitinib resistance screen in MCF10A cells

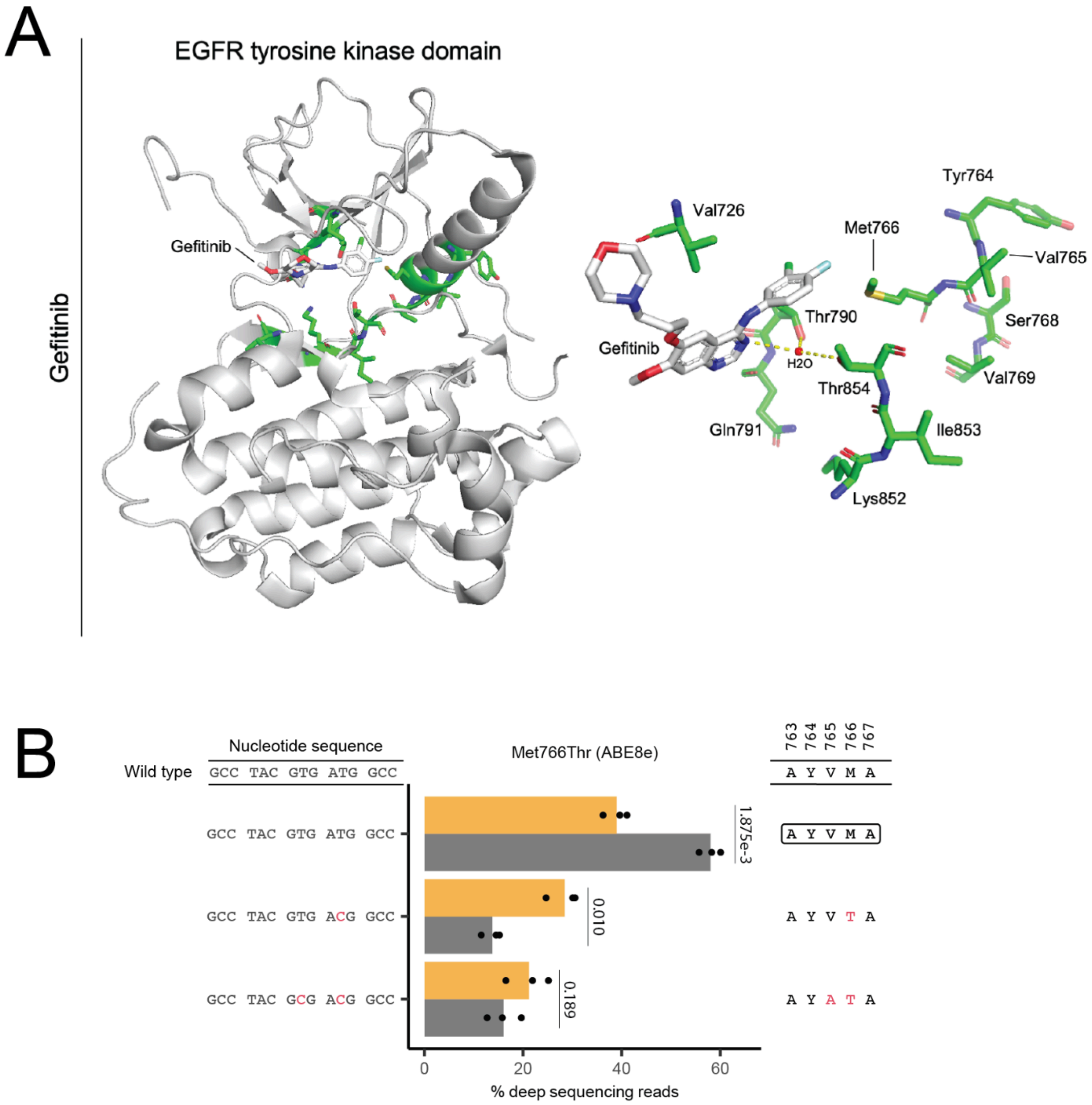


Figure 3.11: MCF10A Gefitinib resistance screens hits and validation. *(A) 3D structures of the EGFR tyrosine kinase domain in complex with Gefitinib (PDB #2ITY) as visualized in PyMol. The carbon backbones of amino acids impacted by screen hits are shown in green and that of the Gefitinib molecule in gray. The relative positions of screen hits around the ATP-binding pocket are shown on the right. (B)* Drug resistance screen validation in MCF10A cells. *Cells were infected with a lentiviral construct expressing the Met766Thr[ABE8e] sgRNA (n= 3). After six days, cells were split into an untreated control and a Gefitinib-treated arm and incubated for 7 days before final harvest. The percentages of reads corresponding to the three most represented alleles are shown for non-treated and drug-treated cells. Edited nucleotides and amino acids are highlighted, black boxes show WT EGFR sequences and p-values are shown for two-sided unpaired t-tests with Holm correction.*

Similarly to Gefitinib, the top three hits for Osimertinib affect residues found in the ATP binding pocket of the tyrosine kinase domain (Figure 3.9B, Figure 3.12A, Table 3.3). For example, Val845Ala[ABE8e] interacts with the Phe795 and Gly796 residues adjacent to Cys797 which is the Osimertinib covalent binding site, hinting at a resistance mechanism involving this residue. Although this mutation is not listed in ClinVar, the similar Val845Leu variant has conflicting reports of pathogenicity in the database. We thus set out to confirm this phenotype by individual sgRNA delivery and target exon sequencing after drug treatment. This revealed the enrichment of the Val845Ala edit and a strong depletion of wild-type alleles under Osimertinib selection, thus validating the resistance phenotype conferred by this variant (Figure 3.12B).

The remaining hits are found throughout the tyrosine kinase and C-terminal domains, highlighting distinct EGFR regulation mechanisms. For example, enriched sgRNAs are found to affect Lys852 and Gln791 which interact with each other in the Osimertinib-bound receptor to form a hydrogen-bond network with residues Asp1012 and Asp1014 of the C-terminal domain. Neither of these variations are listed in ClinVar although perturbing this network by replacing Gln791 with a hydrophobic residue has been predicted to destabilize Osimertinib binding (353). Interestingly, in the inactive conformation, Lys852 is also known to directly interact with another hit, Glu1005, which is part of a C-terminal “electrostatic hook” that inhibits the kinase domain activity. Mutating Glu1005 and Asp1006 has been shown to increase the activity of unstimulated EGFR *in vitro*, thus potentially promoting the observed resistance phenotype (188).

Lastly, our screen identified individual residues involved in important post-translational modification sites. Notably, Cys781Tyr[BE3.9Max] impacts a palmitoylation site involved in EGFR localisation at the plasma membrane (359) while Tyr1069Cys,Ser1070Gly[ABE8e] affects the Tyr1069 residue whose phosphorylation recruits the c-Cbl E3 ubiquitin ligase, leading to EGFR internalization and degradation (202,360). These discoveries are particularly significant as these modifications and their modulation can constitute novel drug targets.

Taken together, the base editing variant scanning results highlight key intra-molecular interactions between EGFR residues involved in enzymatic activity regulation, resulting in new and intricate insights into drug-dependent resistance mechanisms. Interestingly, while top resistant mutations for both drugs are found in the ATP-binding pocket, Osimertinib-resistant hits are also found in the C-terminal domain, hinting at both shared and divergent resistance mechanisms between the drugs. These insights, substantiated by clinical validation, may in the future help clinical decision making.

sgRNA sequence	Editor	Log-fold change	FDR	Predicted amino acid change	Database category	Database accession	Comments	References
GTGTTTTCACCAGTACGTTCC	ABE8e	2.3083	0	Val845Ala	COSMIC	COSV51835641	Val845Ala is not listed in ClinVar but the similar Val845Leu variant has conflicting reports of pathogenicity in the database	
GCATGTCAAGATCACAGATTT		1.938	6.66E-228	Lys852Gly	Not listed		In the active conformation Lys852 is involved in an hydrogen-bond network with Gln791, Asp1012 and Asp1014. Destabilizing this network has been shown to reduce Osimertinib binding affinity.	(332,353)
GATCACGCAGCTCATGCCCTT		1.6989	1.48E-290	Thr790Ala, Gln791Arg	COSMIC (both)	COSV51773932 COSM9583352	Both mutations were observed in cancer patients. In the active receptor, Gln791 interacts with Lys852, Asp1012 and Asp1014. Destabilizing this network has been predicted to reduce Osimertinib binding affinity.	(332,353,361, 362)
GACAATCATCTGGCAGCGAGG		1.4736	1.94E-95	Alternative transcript				
GTGAATGACAAGGTAGCGCTG		1.1473	1.45E-42	Ile981Thr, Val980Ala	Uncertain significance (Ile981Thr)	VCV002105917.1	Observed in lung cancer.	
GTGCGTCTATCATCCAGCCTG		1.1367	7.90E-72	Ile953Thr	Uncertain significance	VCV000965600.6	Observed in lung cancer.	
GTCATATGGCTTGGATCCAA		1.0452	1.49E-30	Tyr915His	COSMIC	COSV51842912	Tyr915 is a phosphorylation site recognised by the c-Src tyrosine kinase.	(363)
GATGTCAAGATCACAGATTTT		1.0052	1.77E-33	Lys852Gly	Not listed		In the active conformation Lys852 is involved in an hydrogen-bond network with Gln791, Asp1012 and Asp1014. Destabilizing this network has been shown to reduce Osimertinib binding affinity.	(332,353)
GAATGACAAGGTAGCGCTGG		0.98683	7.09E-30	Val980Ala, Leu979Leu	Not listed			
GAAGGTAGCGCTGGGGGTCTC		0.98396	3.91E-49	Tyr978His	Uncertain significance	VCV001386901.2	Y978 is phosphorylated in response to EGFR activation and subsequently recruits STAT5.	(364)
GAGCTGCGTGATGAGCTGCA		0.9828	2.98E-25	Ile789Thr, Leu792Pro	COSMIC (Leu792Pro)	COSV51860329	Leu792 is located in the ATP binding pocket of the kinase domain, in contact with bound Osimertinib. Leu792 variants have been found in patients with acquired Osimertinib resistance. Leu792His/Phe/Tyr have been shown to be resistant to Osimertinib in vitro.	(365,366)
GCTGAATGACAAGGTAGCGCT		0.95711	4.66E-21	Ile981Thr, Val980Ala	Uncertain significance (Ile981Thr)	VCV002105917.1		
GCGATACAGCTCAGACCCCAC		0.93918	7.39E-19	Tyr1069Cys, Ser1070Gly	COSMIC (Tyr1069Cys)	COSV51765999	Y1069 phosphorylation recruits the c-Cbl E3 ubiquitin ligase. Tyr1069Cys was shown to increases EGFR signaling and promotes EGF-independent cell growth in vitro	(202,360)
GTAGGAAATTTTAAAAGATGA		0.88042	2.66E-10	3'UTR				
GCAAGAAGATGCACGAAGGC		0.86791	2.42E-12	3'UTR				
GTGAGGCAGATGCCCAGCAGG	BE3.9Max	1.1808	1.92E-54	Cys781Tyr	Not listed		Cys781 is a palmitoylation site involved in EGFR localisation at the plasma membrane.	(359)
GTTCTCCTTTCTCCAGGATGG		0.96163	4.43E-21	Gly930Lys	Not listed			
GCTTCTTCATCCATCAGGGCA		0.82936	2.67E-21	Glu1005Lys, Glu1004Lys	Uncertain significance (Glu1004Lys)	VCV001385349.4	Glu1005 directly interacts with Lys852 in the inactive conformation. It is part of a C-terminal “electrostatic hook” that inhibits the kinase domain activity. Mutating Glu1005 and Asp1006 has been shown to increase the activity of unstimulated EGFR in vitro.	(188)
GCGCTCACACCGTGCGGGGGG		0.80013	7.34E-14	5'UTR				

Table 3.3: Screen hits of the base editing Osimertinib resistance screen in MCF10A cells.

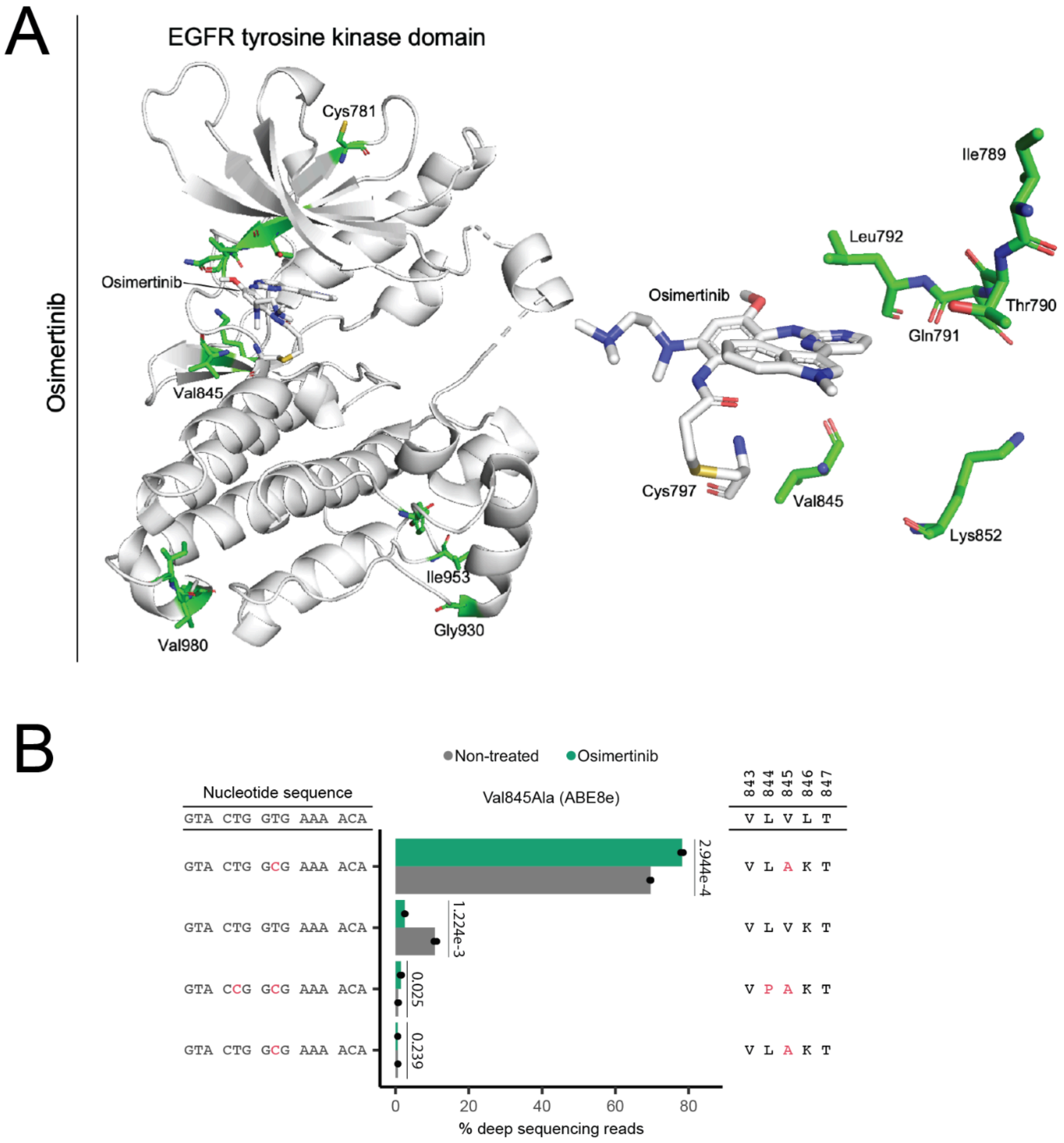


Figure 3.12: MCF10A Osimertinib resistance screens hits and validation. *(A) 3D structures of the EGFR tyrosine kinase domain in complex with Osimertinib (PDB #6JXT) as visualized in PyMol. Hits are shown in green and the Osimertinib molecules in gray. The relative positions of screen hits around the ATP-binding pocket are shown on the right. (B)* Drug resistance screen validation in MCF10A cells. *Cells were infected with the Val845Ala[ABE8e] sgRNA (n= 3) and treated with Osimertinib. The percentages of reads corresponding to the four most represented alleles are shown for non-treated and drug-treated cells. Edited nucleotides and amino acids are highlighted, black boxes show WT EGFR sequences and p-values are shown for two-sided unpaired t-tests with Holm correction.*

3.7 Variant scanning can guide drug prioritization

In addition to drug-resistant variants, our screening data also identified variants that are significantly depleted under drug selection and likely increase drug sensitivity. When comparing the log fold changes of each sgRNA-base editor pair under Gefitinib and Osimertinib selection, we thus observed that variant sensitivities vary between both drugs with some mutations leading to opposite effects (Figure 3.13). For example, Val845Ala[ABE8e] appeared strongly resistant to Osimertinib but sensitive to Gefitinib. On the other hand, other variants like Val726Ala[ABE8e], Val769Ala[ABE8e] and Met766Thr[ABE8e] appeared to confer different levels of resistance to Gefitinib but to be sensitive to Osimertinib. All of these variants are located within the ATP binding pocket, which suggests that their differential drug sensitivities resulted from specific direct interactions with each molecule.

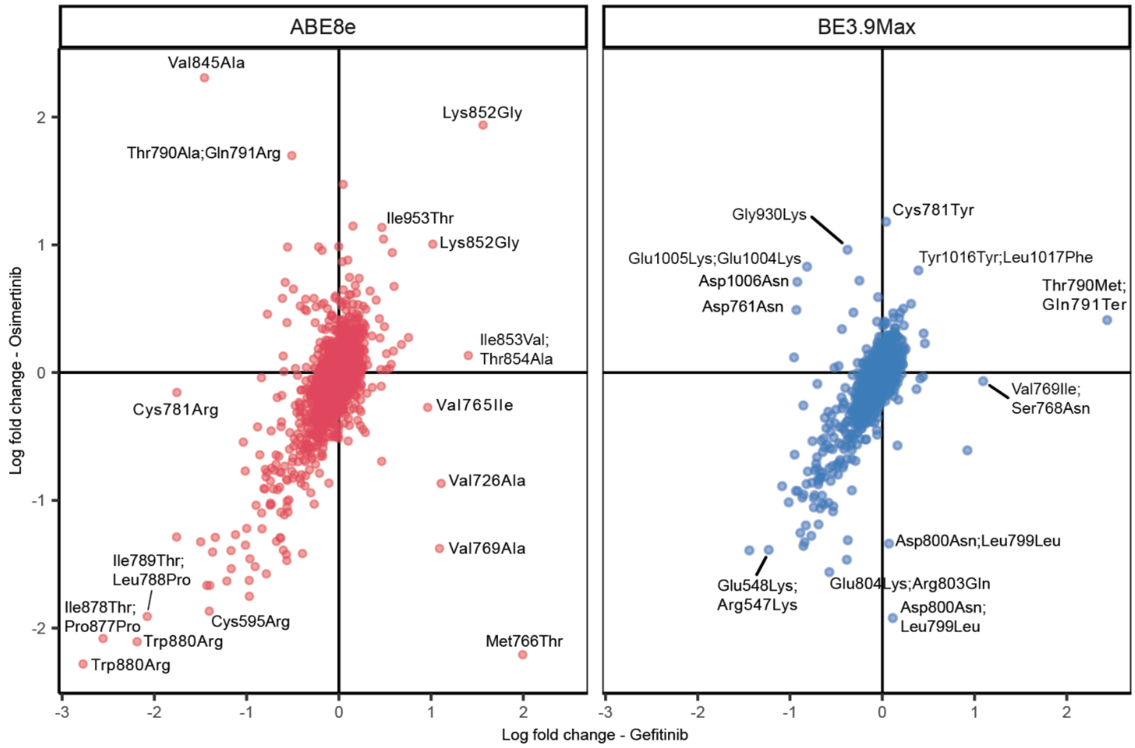


Figure 3.13: Base editing scanning screens identifies EGFR variants with different sensitivities to Gefitinib and Osimertinib. *Scatterplots comparing the log fold changes of sgRNAs in cells treated with either drug for each base editor. Non-targeting and essential splice site controls are not shown as well as sgRNAs with log fold changes < -1.5 between plasmid and day 19.*

We also identified variants that appeared to be resistant to both drugs, such as Lys852Gly[ABE8e], which is currently not listed in ClinVar or COSMIC. Lys852 is known to interact with Gln791 and residues of the C-terminal tail and to contribute to EGFR auto-inhibition. This thus suggests that common resistance mechanisms can emerge from the disruption of auto-inhibition mechanisms, warranting further investigations of this residue. Taken together, these results provide new insights into EGFR variant-dependent drug sensitivities, which may in the future help guide therapeutic decisions for clinicians faced with EGFR variants for which clinical data are currently absent.

3.8 Drug sensitivity profiles vary between primary and secondary EGFR mutations

In cancer patients, TKIs are used to counteract the activity of hyperactive EGFR mutants, such as the common Leu858Arg substitution or Exon 19 deletion before drug resistance emerges (286). We thus set out to apply our base editing variant scanning pipeline to evaluate the impact of secondary EGFR mutations on drug sensitivity in the NSCLC-derived PC-9 cell line. These cells are sensitive to TKIs and harbor EGFR ΔGlu746-Ala750 which is the most prevalent EGFR deletion in lung cancer (330). Additionally, the introduction of the Thr790Met variant in PC-9 cells with base editing has previously been shown to lead to a strong Gefitinib resistance phenotype (367). We thus performed a base editing scanning screen using the same EGFR-targeted sgRNA library and experimental and computational workflows as demonstrated for MCF10A cells.

We started our analysis by comparing sgRNA counts between plasmid and day 19 conditions and confirming high replicate correlation, no shift in the negative control sgRNA population, and the depletion of positive control sgRNAs targeting essential splice sites (Supplementary Figure S3.8, Supplementary Figure S3.9, Supplementary Figure S3.10). As was the case in MCF10A cells, in EGFR-mutant PC-9 cells we also identified a spectrum of EGFR secondary mutations imparting fitness effects. While many of the fitness-altering mutations were shared between MCF10A and PC-9 cells, we identified unique subsets for each cell line (Figure 3.14). Interestingly, tyrosine kinase variants appear to have a stronger impact on viability in EGFR mutant PC-9 compared to EGFR wild-type MCF10A cells. We speculate that this could be due to a diminished tolerance of the mutant receptor to further mutations or to oncogene addiction in PC-9 cells, making them more sensitive to variations in EGFR activity.

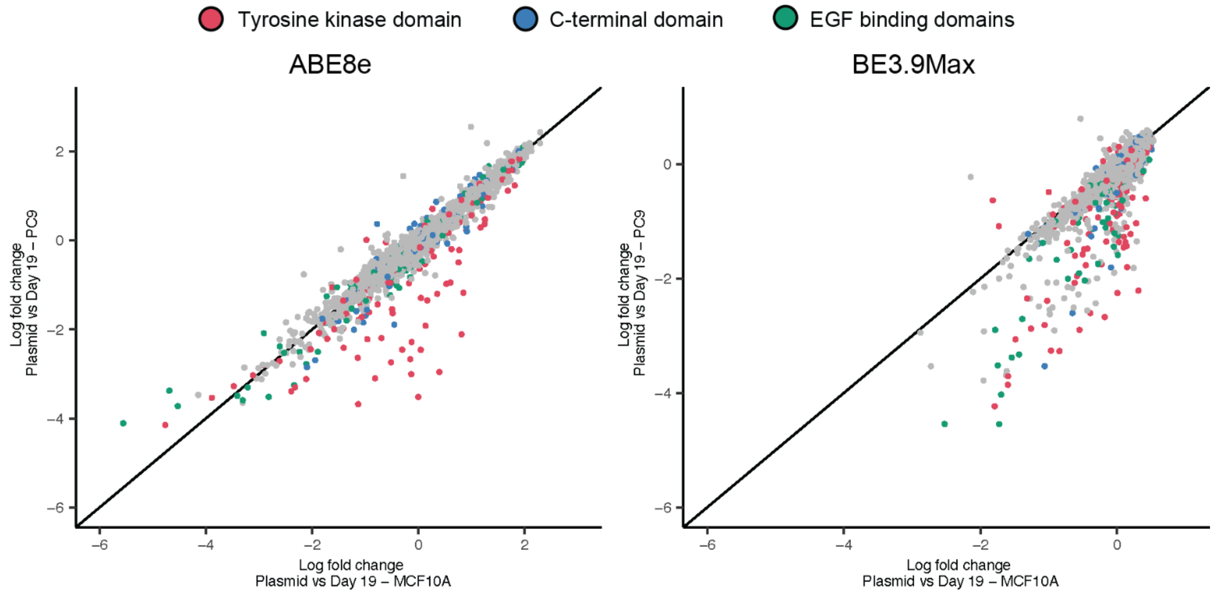


Figure 3.14: EGFR-targeting sgRNAs have different impacts on cell viability in MCF10A and PC-9 cells. *Comparisons of log fold changes between plasmid and day 19 samples in MCF10A and PC-9 cells. sgRNAs targeting essential splice sites and non-targeting controls are not shown.*

Next, we set out to characterize the impact of EGFR secondary mutations on TKI drug resistance (Figure 3.15, Supplementary Figure S3.11). Similarly to MCF10A cells, we observed that the most enriched Gefitinib-resistant hits are located in the tyrosine kinase domain while top Osimertinib hits span both the tyrosine kinase and the C-terminal domains.

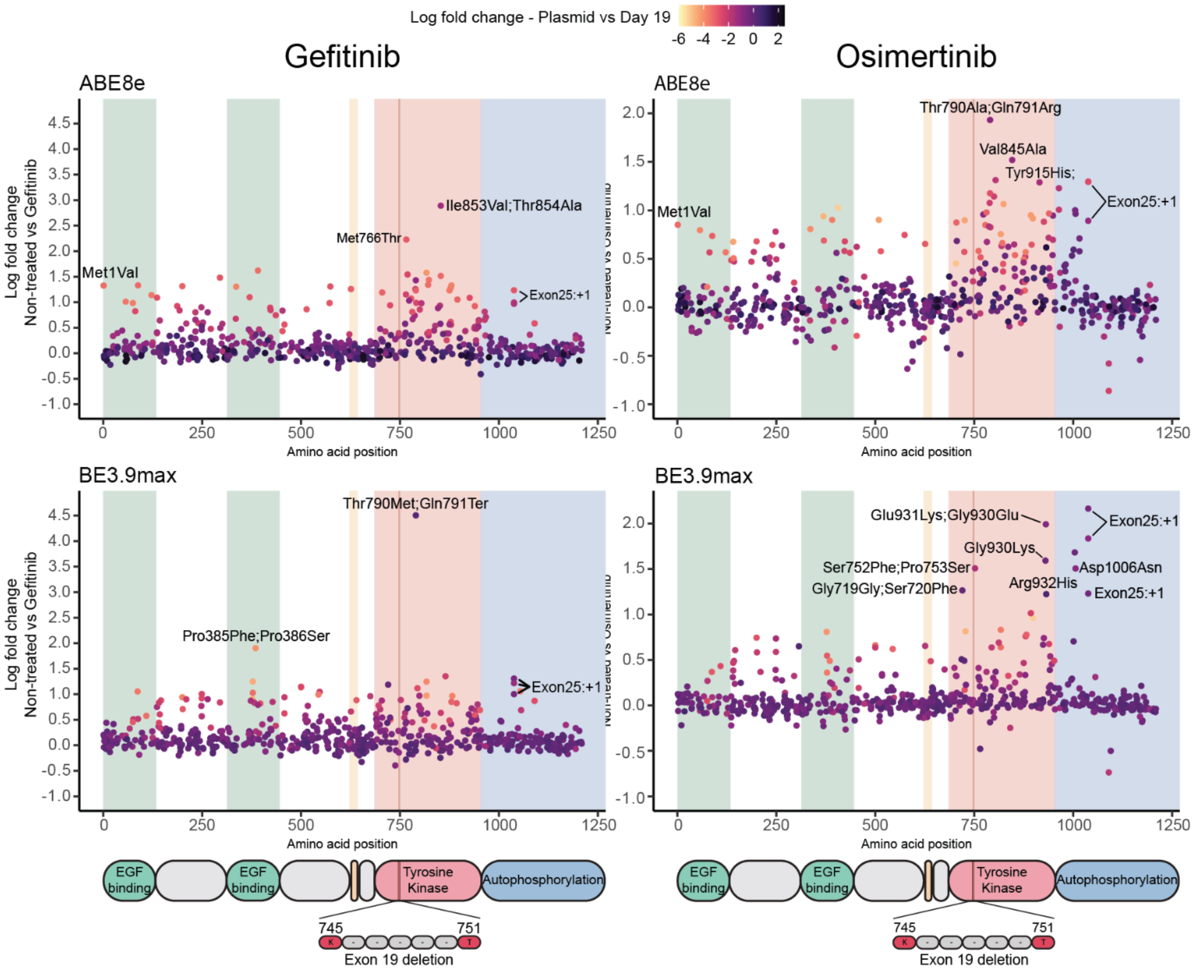


Figure 3.15: Base editing mutational scanning screens identify drug-resistant EGFR variants in PC-9 cells. *Scatterplot showing the log fold changes of individual sgRNAs along the EGFR protein between non-treated and drug-treated cells for each base editor and drug treatment. Log-fold changes between the plasmid library and non-treated cells on day 19 are shown in color gradients. Background colors represent EGFR domains.*

We thus continued our analysis by comparing shared hits between both cell lines. While this revealed a broad-range of shared and distinct variants spanning EGFR, we noticed that the most commonly impacted positions appeared to be the same regardless of the initial EGFR genotype (Figure 3.16). For example, in both cell lines, Thr790Met,Gln791Ter[BE3.9Max], Ile853Val,Thr854Ala[ABE8e], and His773Arg[ABE8e] are enriched under Gefitinib treatment while Thr790Ala,Gln791Arg[ABE8e] and Val845Ala[ABE8e] are enriched with Osimertinib. To confirm a selection of these insights, we delivered the Ile853Val,Thr854Ala[ABE8e] or His773Arg[ABE8e] sgRNAs to PC-9 cells and treated infected cells with Gefitinib or Osimertinib for 7 days. Illumina sequencing of the targeted region revealed that with both sgRNAs, reads corresponding to the unedited alleles were strongly depleted under drug treatment, thus confirming efficient growth inhibition of unedited cells relatively to edited ones (Figure 3.17).

By contrast, we observed an enrichment of reads containing the His773Arg edit with both drugs although this enrichment was not statistically significant in the case of Gefitinib. Furthermore, in Ile853Val,Thr854Ala[ABE8e] samples, reads containing the Thr854Ala edit but not Ile853Val alone appeared enriched under Gefitinib selection, confirming that Thr854Ala is the driver of the observed resistance phenotype. On the contrary, both Ile853Val and Thr854Ala edits appeared to be required for Osimertinib resistance, warranting further mechanistic investigations.

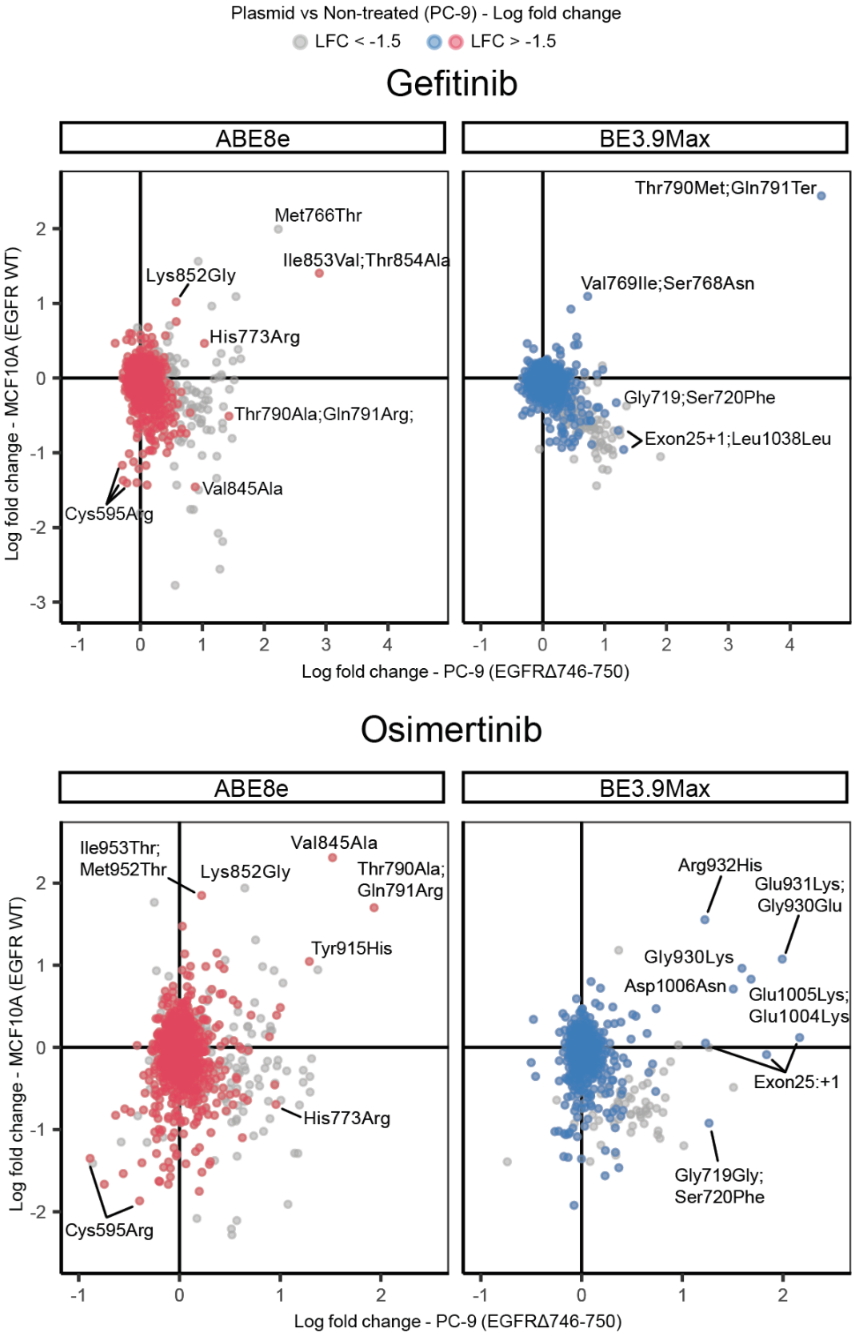


Figure 3.16: Base editing mutational scanning screens identifies EGFR variants with differential drug sensitivities. *Scatterplots comparing the log fold changes of sgRNAs in MCF10A (EGFR WT) and PC-9 (EGFR ΔGlu746-Ala750) cell lines treated with Gefitinib or Osimertinib. sgRNAs with LFC < -1.5 between plasmid and day 19 in PC-9 cells are shown in gray. Non-targeting and essential splice site controls are not shown.*

Finally, we compared sgRNAs that were differentially enriched across the two cell lines. For example, Thr790Ala,Gln791Arg[ABE8e] and Gly719Gly,Ser720Phe[BE3.9max] appeared to only confer Gefitinib resistance in PC-9 cells but not in MCF10A, suggesting the existence of different tyrosine kinase conformations and ligand interactions in the two cell lines. In a follow-up experiment, we delivered the Gly719Gly,Ser720Phe[BE3.9max] sgRNA to PC-9 cells and identified a diverse array of complex products including non-canonical C-to-G edits (Figure 3.17). Among these, only Ser720Phe appeared potentially enriched under drug selection. However, the observed difference between treatment arms was insufficient to conclude on a statistically significant effect. This result thus warrants further investigation, possibly with longer selection times or gene editing approaches generating less complex products, to evaluate Ser720Phe resistance to Gefitinib and Osimertinib in this cellular context.

Surprisingly, with both drugs and both base editors, we noticed the enrichment of sgRNAs predicted to impact the Exon 25 splice donor and leading to the EGFRΔEx26-28 truncation. Interestingly, this truncation was found to drive EGFR activation in our MCF10A screen. However, it appears to confer drug resistance only in PC-9 cells, hinting at a resistance mechanism requiring previous receptor hyperactivation or a specific receptor conformation resulting from the ΔGlu746-Ala750 deletion. Single gRNA follow-up experiments confirmed a significant enrichment of edited alleles with mutated splice donor site under both drug treatments (Figure 3.17). Taken together, these data confirm that the secondary truncation of a constitutively active EGFR variant such as EGFRΔ746-750 is able to maintain its downstream signaling and might constitute a mechanism of resistance to TKIs.

Together, these results highlight the importance of evaluating drug resistance variants in relevant genomic contexts, including pre-existing EGFR mutations, and confirm the relevance of direct sequencing to validate base editor screen hits.

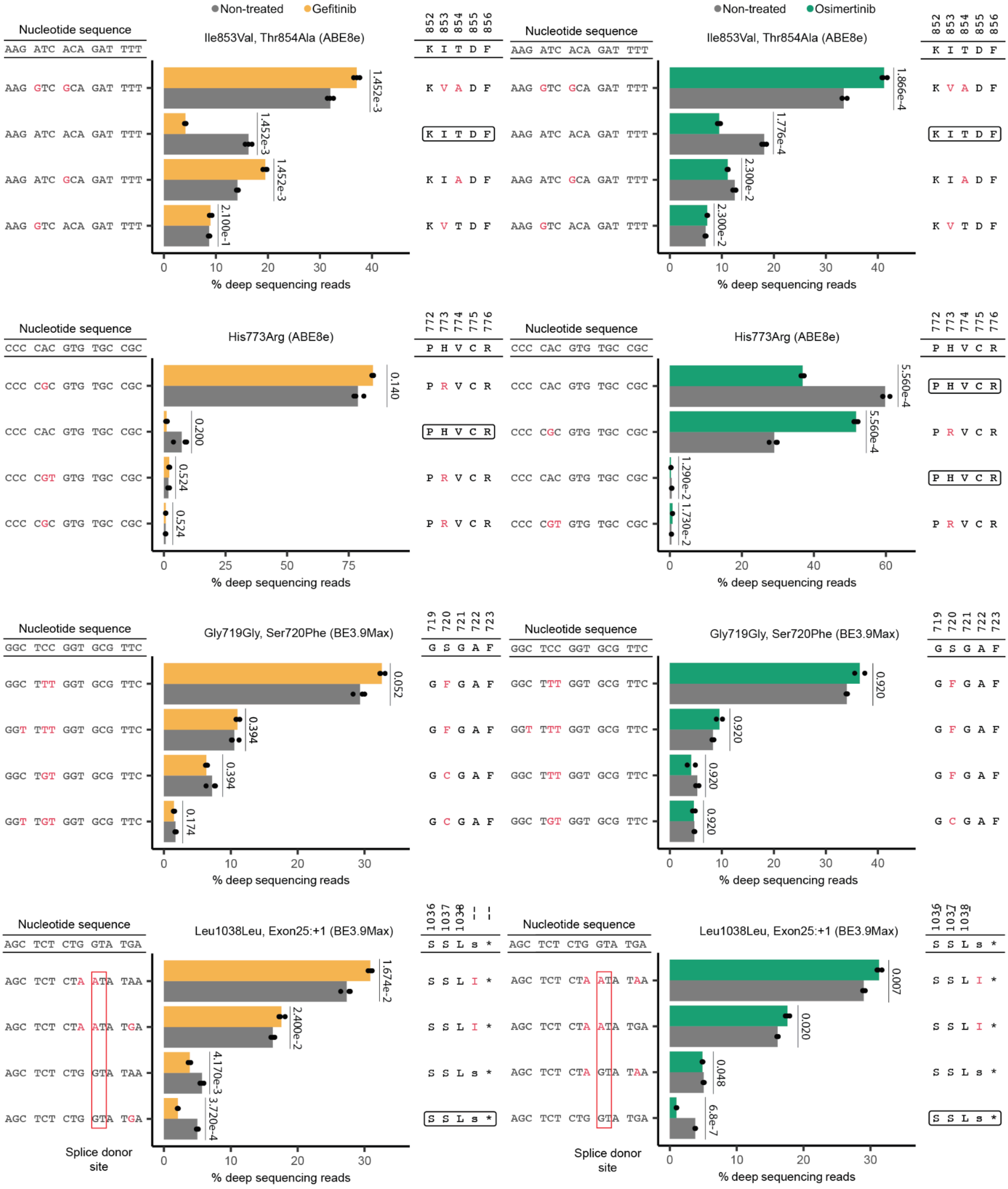


Figure 3.17: Deep sequencing data for individual sgRNA validation in PC-9 cells. *PC-9 cells were infected with the indicated sgRNAs and treated with Gefitinib (n= 3) or Osimertinib (n= 3, for Gly719Gly,Ser720Phe[BE3.9Max] n= 2). The percentages of reads corresponding to the four most represented alleles at each target site are shown for non-treated and drug-treated cells. Edited nucleotides and amino acids are highlighted, black boxes show WT EGFR sequences and p-values are shown for two-sided unpaired t-tests with Holm correction.*

3.9 Conclusion

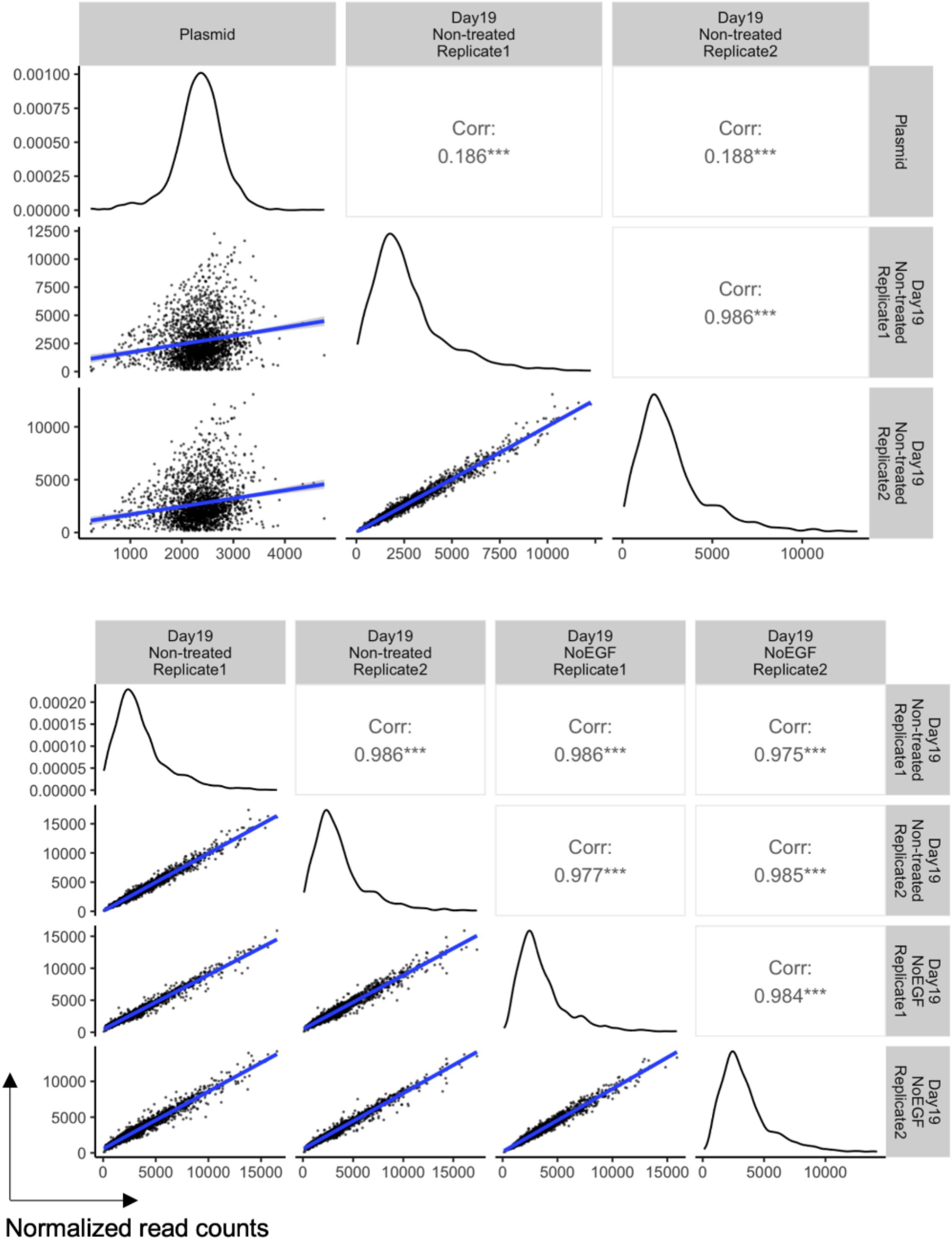
In this chapter, we performed complementary base editing scanning screens in the EGFR gene using cytosine and adenine base editors. We first applied our approach in MCF10A cells harboring WT EGFR and uncovered known and unknown variants leading to EGF-independent growth. Encouragingly, many of our screen hits were located in the tyrosine kinase domain, as frequently observed in NSCLC. However, we also identified activating variants in the extracellular domain which are more typical of GBM, confirming the validity of MCF10A to study EGFR variants in their globality. Although most of our hits are currently not listed in ClinVar or classified as VUS, our data uncovered key protein residues and structures involved in receptor regulation, indicating that these may be bona fide oncogenic mutations. Using plasmid library distribution data as a reference, we also identified EGFR LOF mutations that appeared artificially enriched in activation and drug resistance screens. While we speculate that these variants lead to decreased cell viability by suppressing EGFR signaling, further experiments are needed to determine whether they lead to cell death or quiescence in our different cell lines. Interestingly, some of these mutations appeared to be cell line-specific, highlighting the importance of preserving endogenous expression levels and taking variant fitness into account in screens using cell proliferation as a read-out.

We then expanded our approach to screen for variant responses to clinically-approved TKIs in cell lines harbording WT EGFR or the Δ746-750 activating mutations. This uncovered a spectrum of novel variants leading to distinct drug resistance profiles and hinting at divergent mechanisms of resistance to first- and third-generation TKIs beyond drug-binding pocket interactions. Interestingly, the log fold changes observed with ABE8e and BE3.9Max appeared to consistently follow distinct distributions with ABE8e showing higher variance across sgRNAs. Since the initial distributions of both libraries are very similar, this could be explained by differences in editing efficiency, toxicity and mutational spectrum, thus limiting the direct comparison of screen results between the two enzymes. Furthermore, the prevalence of sgRNAs introducing multiple edits with diverse phenotypic effects in our screen prevents the direct correlation of sgRNA enrichment and drug resistance levels. In their current installment, base editing screens should thus be considered as non-quantitative discovery tools whose hits should be thoroughly validated by direct target sequencing.

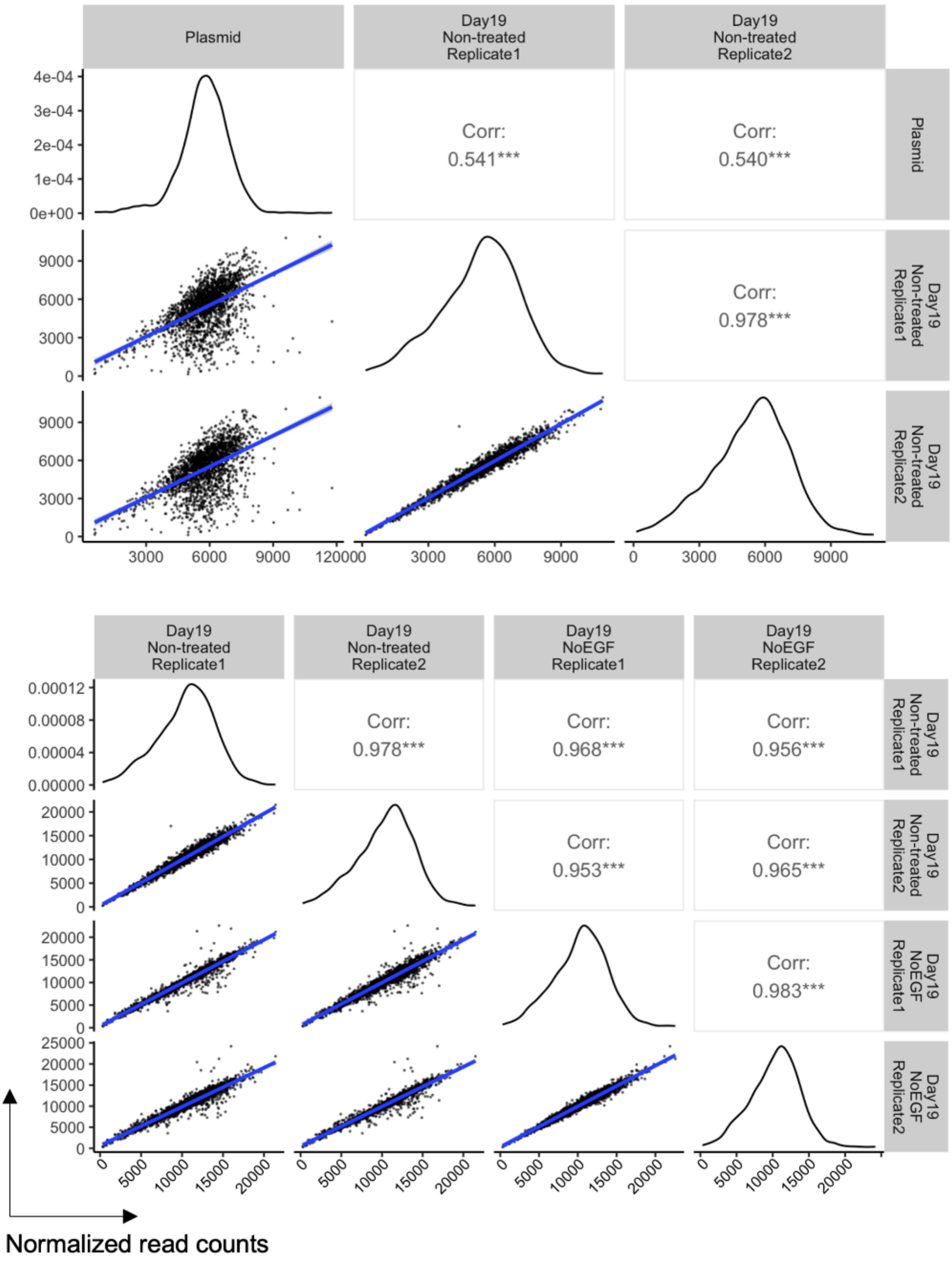
Next, we observed that some variants led to distinct drug sensitivities when introduced in WT EGFR or cells harboring a pre-existing activating mutation. This highlights the importance of assessing EGFR variants in different genetic backgrounds to account for the complexity of primary and secondary mutations found in tumors. Importantly, the direct comparison of screening results across cell lines could be confounded by various levels of gene copy number amplifications in cancer lines like PC-9. Furthermore, our validation data in PC-9 cells show the presence of non-canonical edits which were previously shown to be particularly prevalent in this cell line, thus representing an additional caveat to the comparison of log fold changes across cell lines (368). In this context, diploid MCF10A cells likely represent an ideal model for the pre-engineering of different primary mutations like Leu858Arg in their genome before screening for secondary resistant variants.

In conclusion, base editing scanning screens provided new insights into EGFR activation and resistance to clinically approved TKIs that may help clinical decision making in the future. Our results underscore the need to study EGFR variants in different genomic contexts which is facilitated by the delivery of all-in-one exon-tiling libraries like the ones used in this chapter. Currently, the sensitivity of base editing screens is limited by their restriction to single substitution types and the introduction of bystanders edits. However, the development of base editors with different mutational profiles and of PAM-relaxed Cas9 enzymes will likely allow for increased variant redundancy and the more comprehensive discovery and characterisation of EGFR pathogenic variants in human cells.

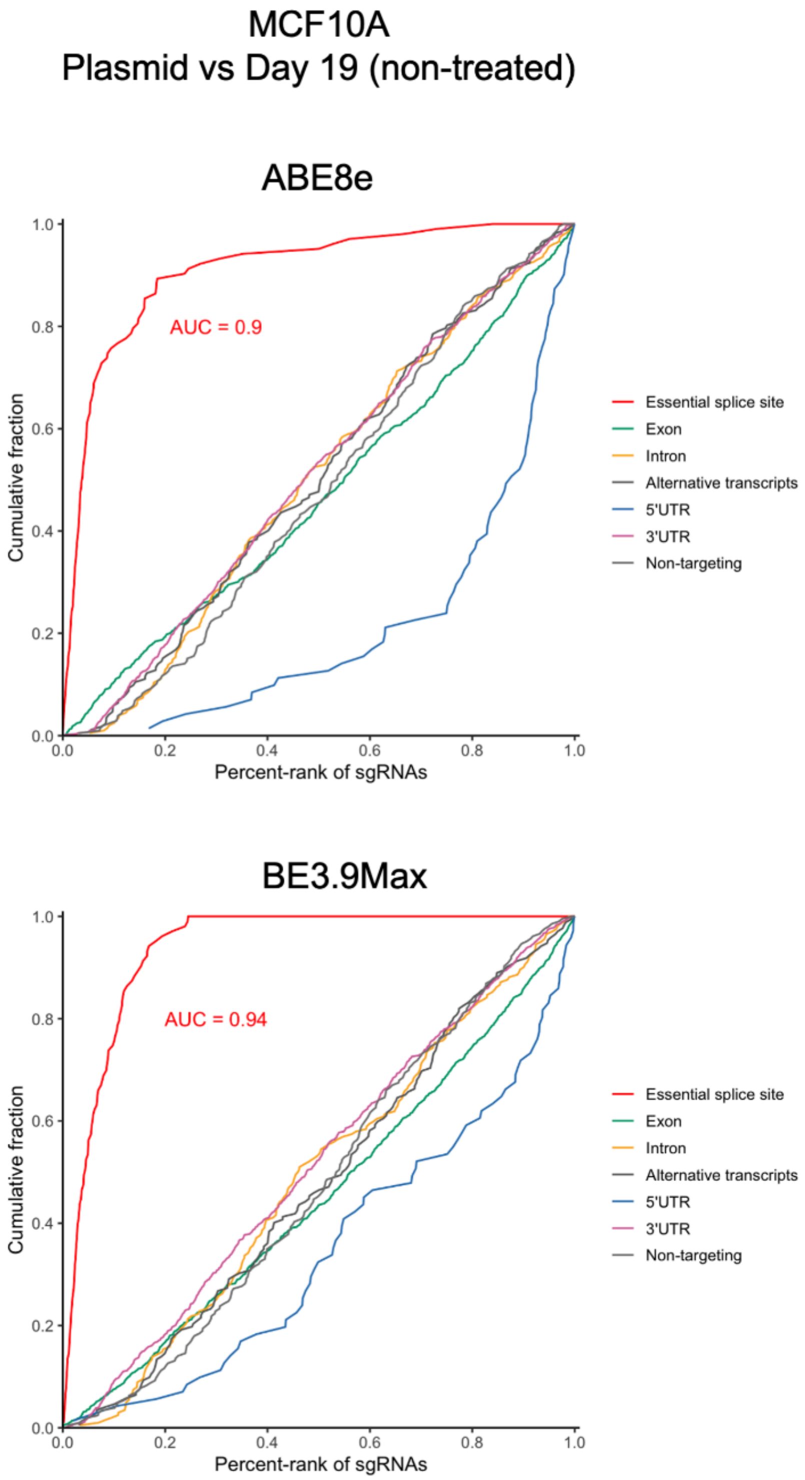
3.10 Supplementary figures



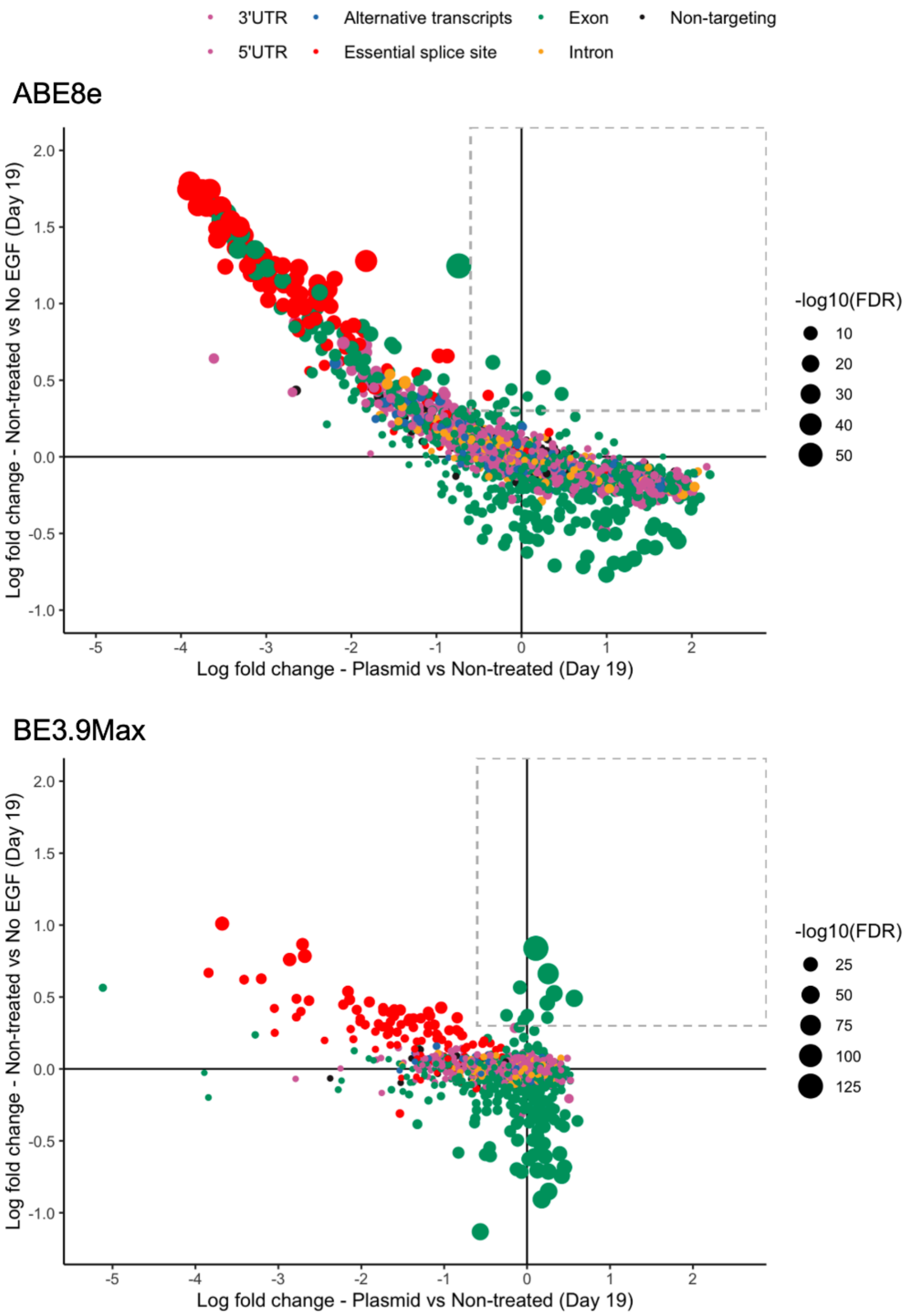
Supplementary Figure S3.1: sgRNA library distributions and sample correlations for the ABE8e EGFR activation screen in MCF10A cells. *Normalized reads counts and Pearson correlation coefficients are shown for all samples*



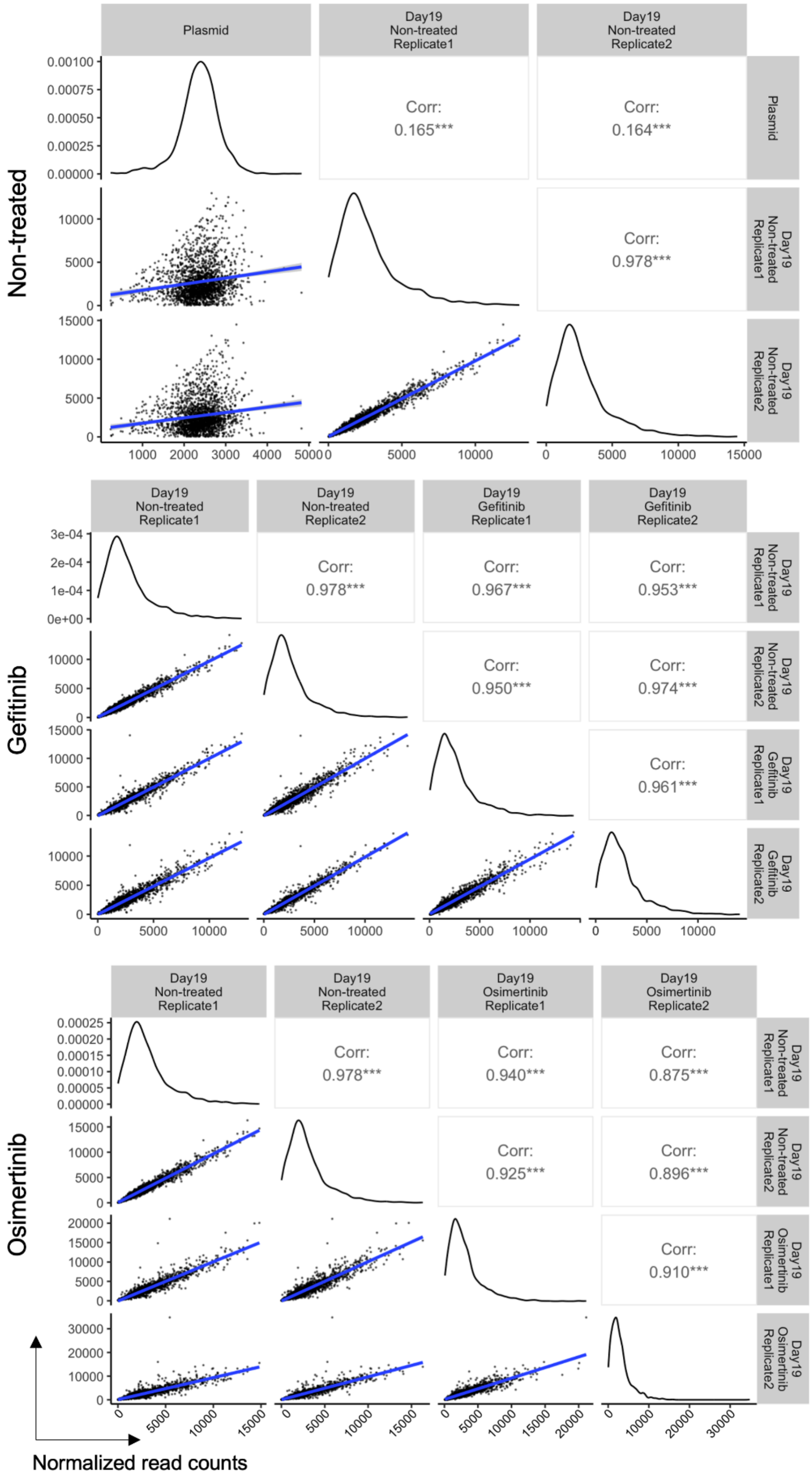
Supplementary Figure S3.2: sgRNA library distributions and sample correlations for the BE3.9Max EGFR activation screen in MCF10A cells. *Normalized reads counts and Pearson correlation coefficients are shown for all samples*



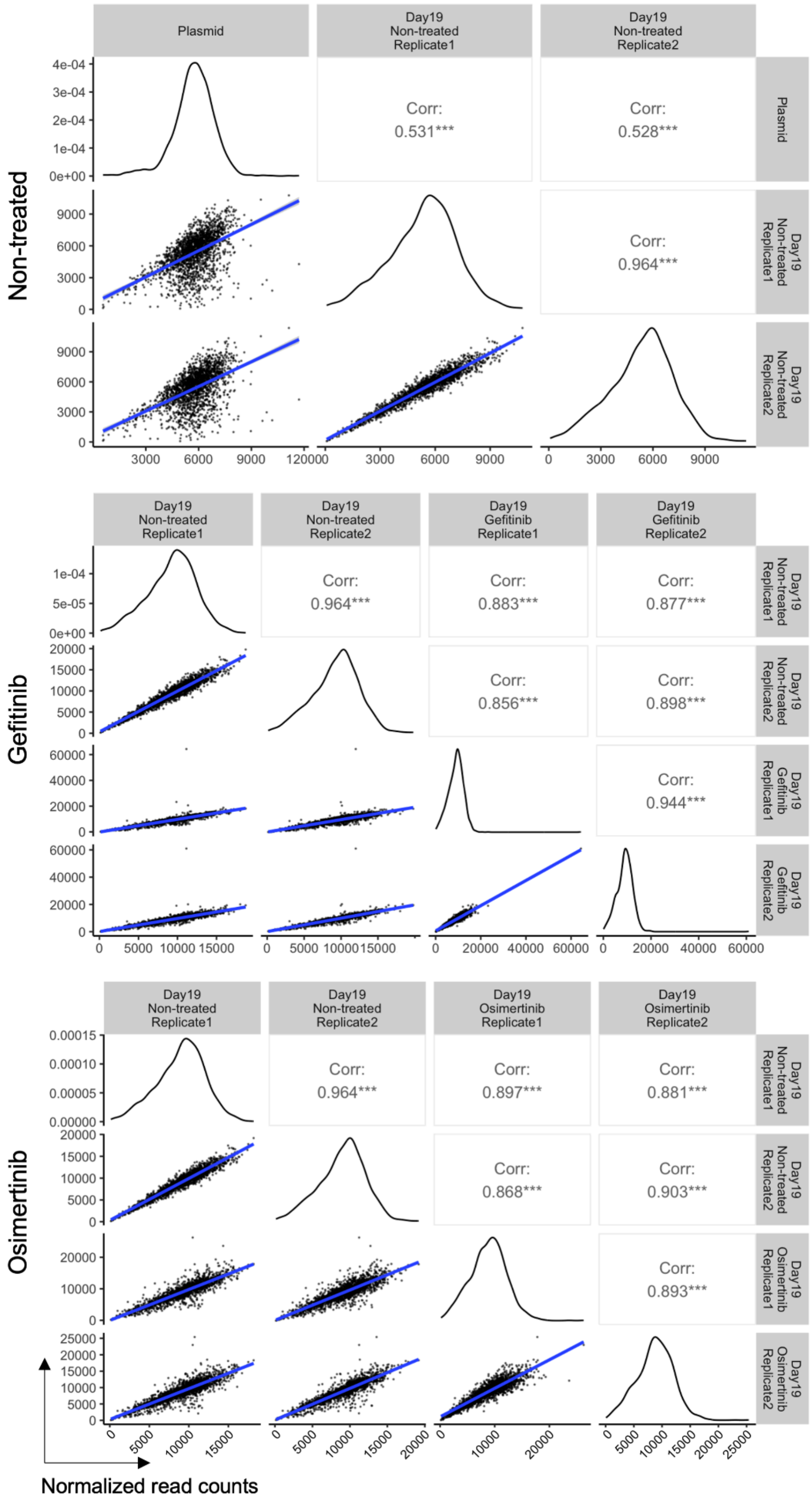
Supplementary Figure S3.3: Impact of individual sgRNAs on cell viability for the base editing EGFR activation screen in MCF10A cells. *Receiver operating characteristic (ROC) curves for each target category or predicted mutation type. Area under the curve (AUC) for essential splice site sgRNAs are shown in red.*



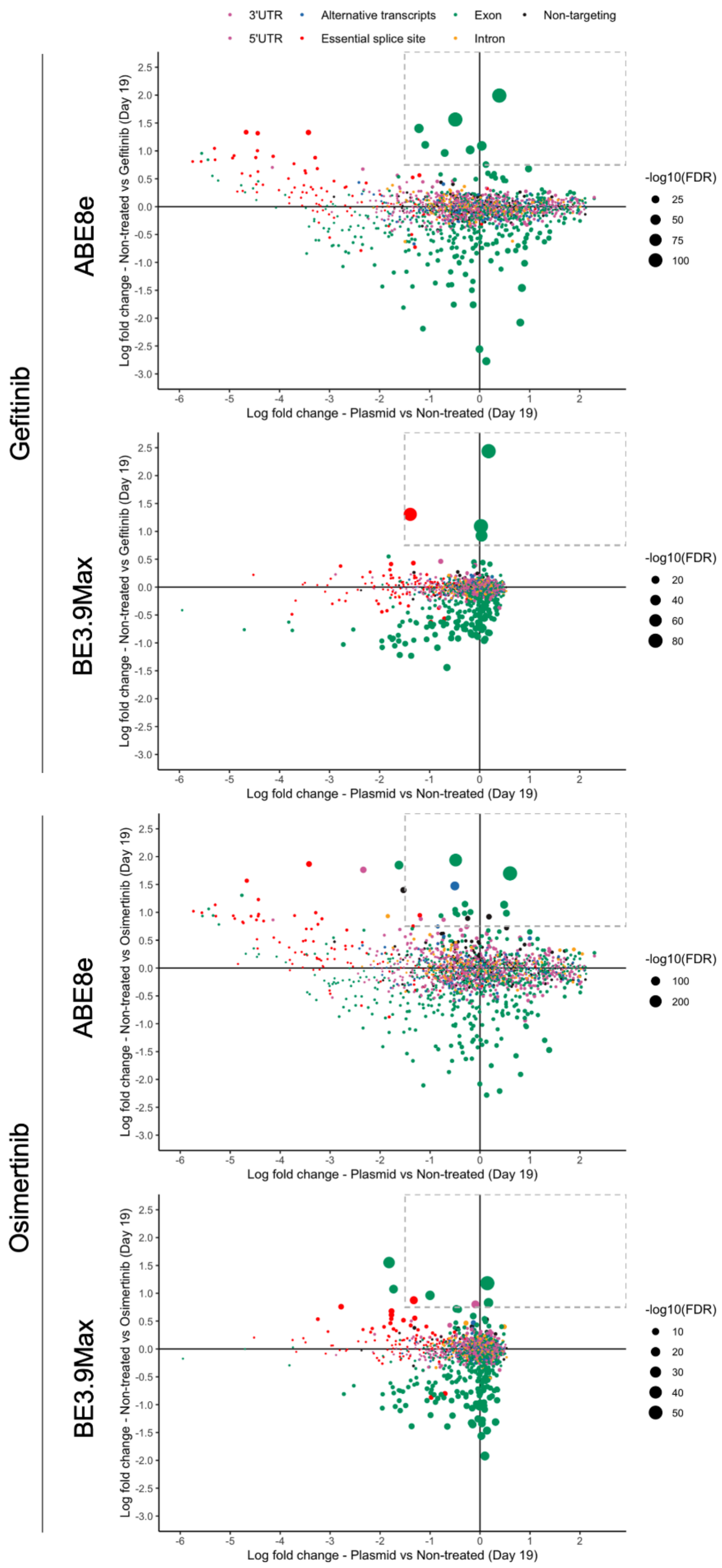
Supplementary Figure S3.4: Comparison of sgRNA impacts on viability and EGF-independent growth in the base editing EGFR activation screen in MCF10A cells. *Scatterplot comparing the individual sgRNA log fold changes (LFC) between plasmid and day 19 and between non-treated and EGF-deprived samples. Negative-log false discovery rates for EGFR activating sgRNAs are shown. Dashed squares represent the hit inclusion criteria: LFC (plasmid vs day 19) > -0.6 and LFC (non-treated vs EGFR deprived) > 0.3.*



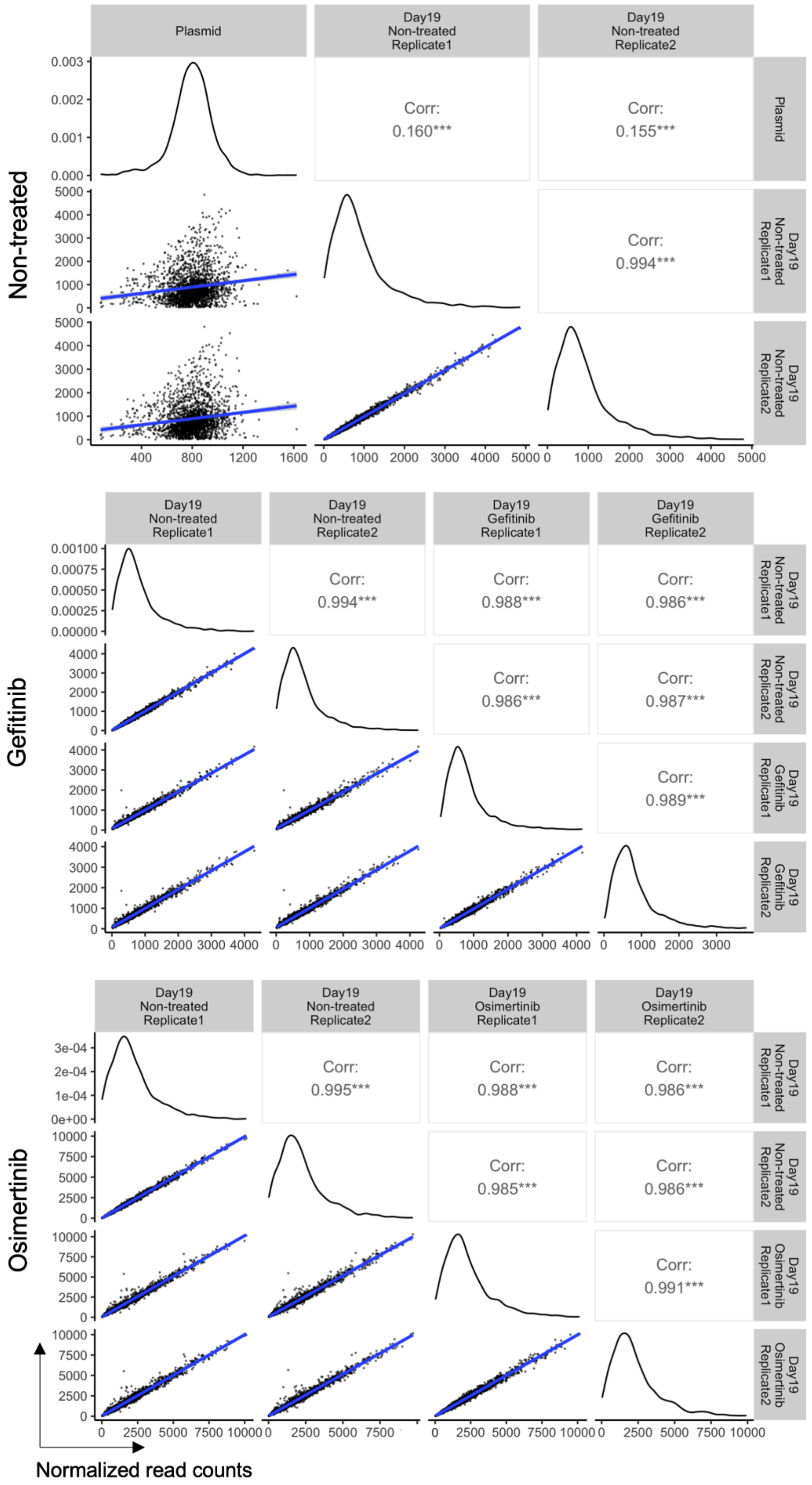
<u>Supplementary Figure S3.5</u>: sgRNA library distributions and sample correlations for the ABE8e drug resistance screens in MCF10A cells. *Normalized reads counts and Pearson correlation coefficients are shown for all samples.*



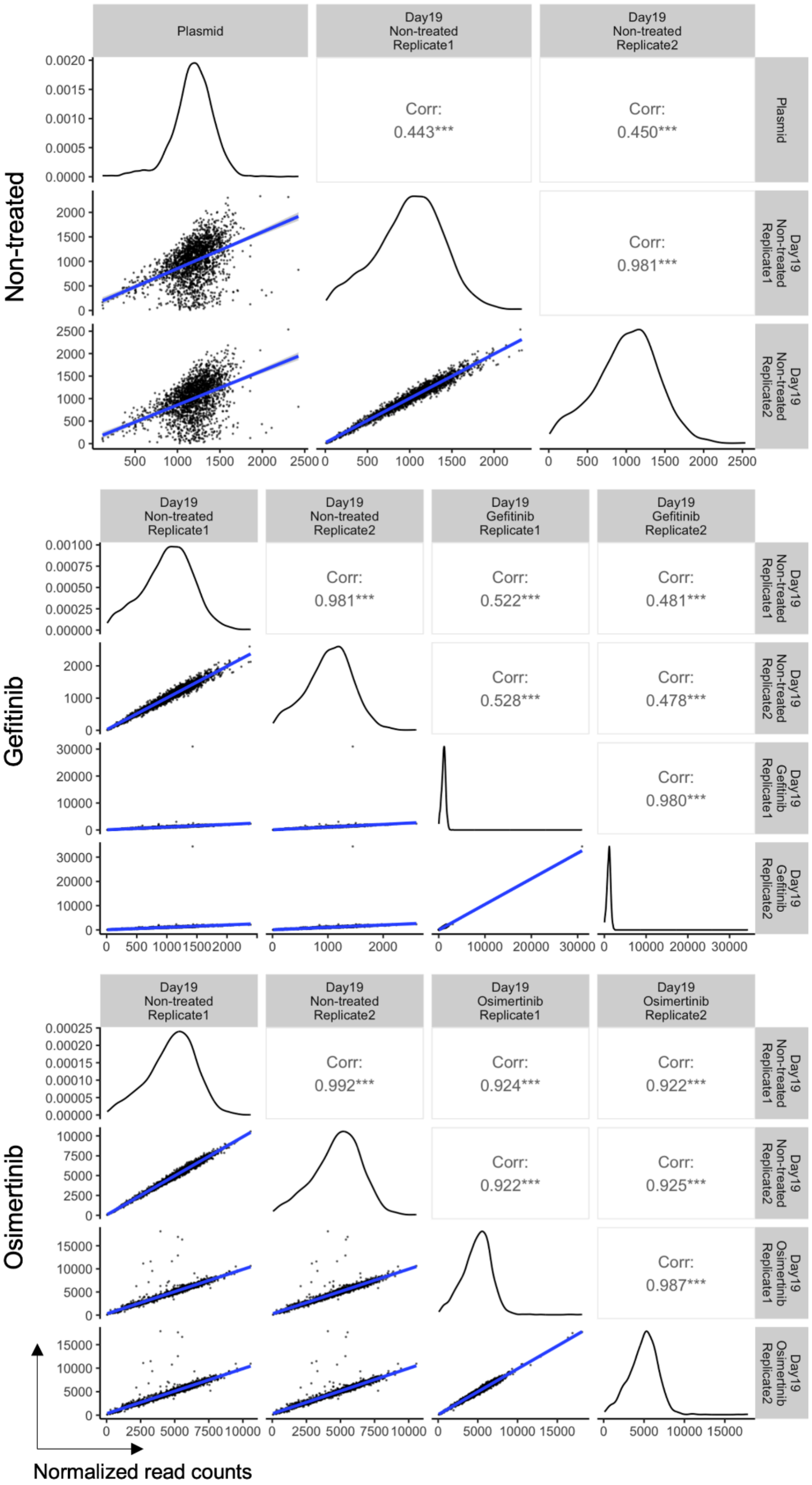
Supplementary Figure S3.6: sgRNA library distributions and sample correlations for the BE3.9Max drug resistance screens in MCF10A cells. *Normalized reads counts and Pearson correlation coefficients are shown for all samples*



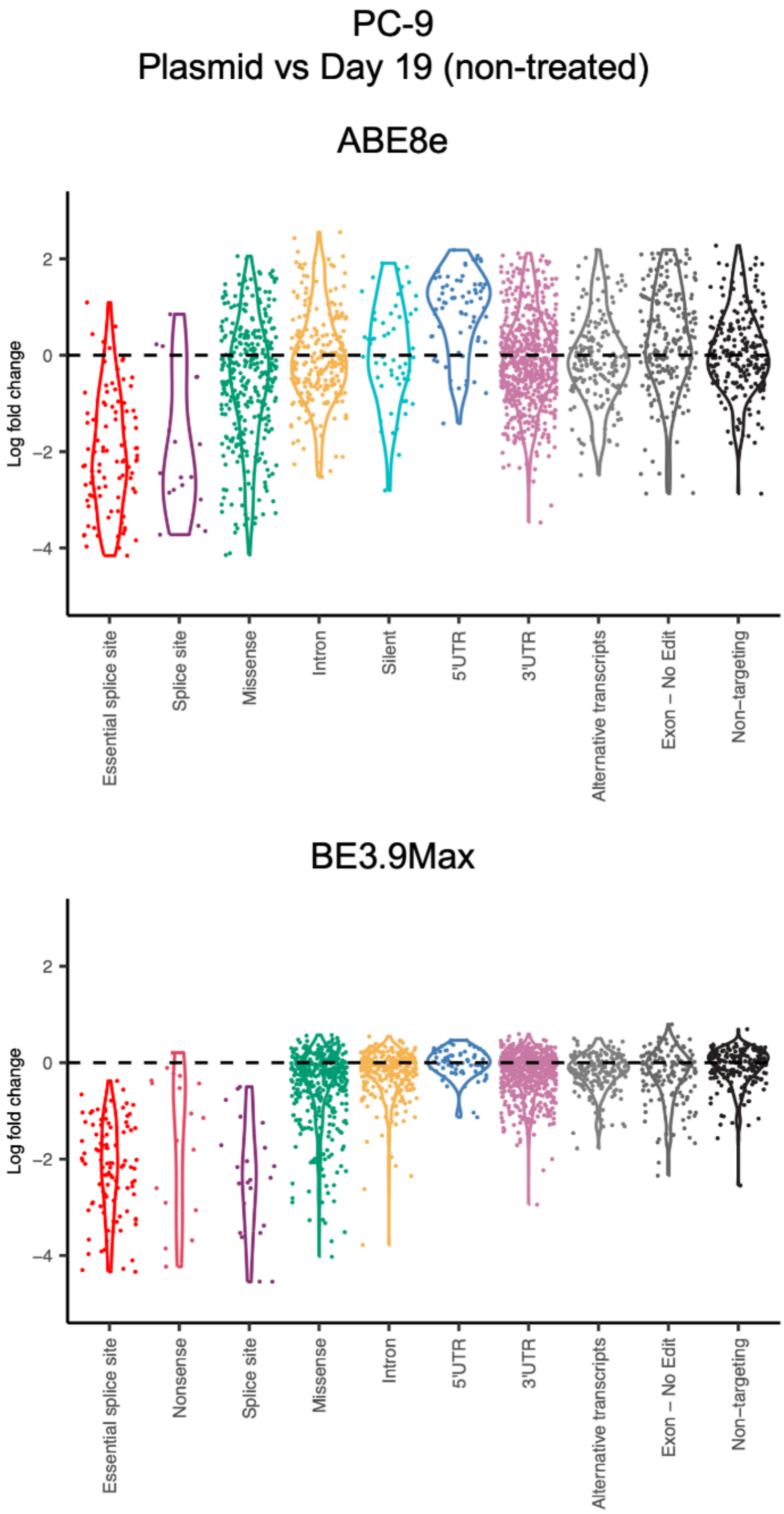
Supplementary Figure S3.7: Comparison of sgRNA impacts on viability and drug resistance in the base editing drug resistance screens in MCF10A cells. *Scatterplot comparing the individual sgRNA log fold changes (LFC) between plasmid and day 19 and between non-treated and TKI-treated samples. Negative-log false discovery rates for drug resistance sgRNAs are shown. Dashed squares represent the hit inclusion criteria: LFC (plasmid vs day 19) > -1.5 and LFC (non-treated vs TKI-treated) > 0.75.*



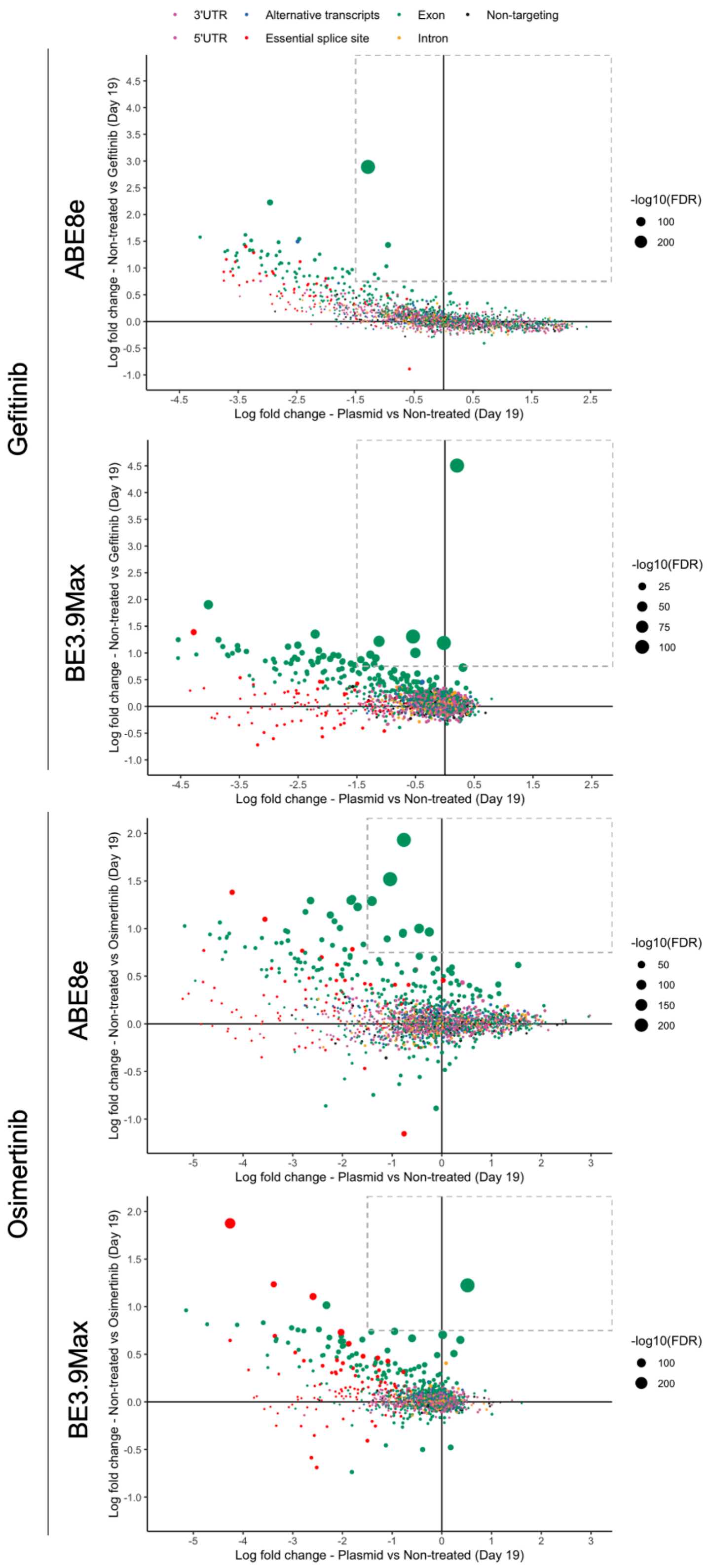
Supplementary Figure S3.8: sgRNA library distributions and sample correlations for the ABE8e drug resistance screens in PC-9 cells. *Normalized reads counts and Pearson correlation coefficients are shown for all samples.*



Supplementary Figure S3.9: sgRNA library distributions and sample correlations for the BE3.9Max drug resistance screens in PC-9 cells. *Normalized reads counts and Pearson correlation coefficients are shown for all samples.*



<u>Supplementary Figure S3.10</u>: Impact of individual sgRNAs on cell viability for the base editing EGFR activation screen in PC-9 cells. *Log fold changes of individual sgRNAs between plasmid and day 19 (non-treated) for each sgRNA target category or predicted mutation type.*



Supplementary Figure S3.11: **Comparison of sgRNA impacts on viability and drug resistance in the base editing drug resistance screens in PC-9 cells.** *Scatterplot comparing the individual sgRNA log fold changes (LFC) between plasmid and day 19 and between non-treated and TKI-treated samples. Negative-log false discovery rates for drug resistance sgRNAs are shown. Dashed squares represent the hit inclusion criteria: LFC (plasmid vs day 19) > -1.5 and LFC (non-treated vs TKI-treated) > 0.75.*

Chapter IV: Patient-derived variant screen with prime editing

4.1 Contributions

I designed, conducted and analyzed all the experiments presented in this chapter. I also wrote all computational pipelines used for library design and screen analysis. External packages and code used for data analysis are referenced in the method section.

The **Genomics Facility Basel** (Christian Beisel, Mirjam Feldkamp, Erika Gröflin-Schürch, Ina Nissen-Naidanow, Elodie Vogel Burcklen) quantified, pooled and loaded all deep sequencing libraries.

Kyriaki Karava provided help with lentivirus concentration and cell harvesting for the prime editing screen presented in Figure 4.5. She also conducted the prime editing experiments in MCF10A cells presented in Figure 4.2 (D-E).

Dr. Rick Farouni provided advice for the statistical analysis and graphical representation of library distribution data (Figure 4.4).

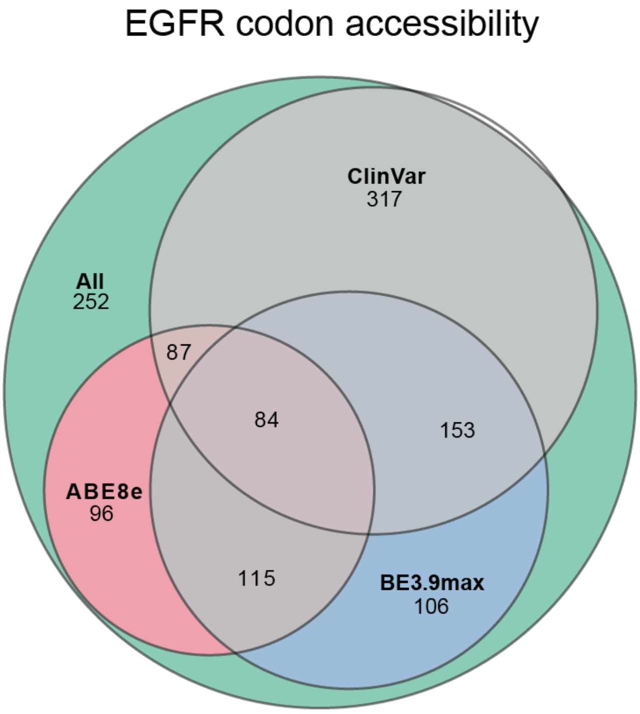
Dr. Alessio Strano, Dr. Georgios Kalamakis and **Prof. Dr. Randall J. Platt** proofread this chapter.

Prof. Dr. Randall J. Platt proposed the initial project idea, provided fundings and contributed intellectually throughout the progress of this project with supervision and guidance.

4.2 Introduction

Base editing screens enable the sensitive identification of phenotype-associated variants at accessible nucleotides but remain limited in the spectrum of mutations they can introduce. Specifically, ABE8e and BE3.9Max coupled with wild type SpCas9 can mutate only about 53% of all EGFR codons and 67% of codons involved in ClinVar entries (Figure 4.1). While this limitation is partly due to PAM availability, which can be circumvented through alternative Cas enzymes (369), most mutations simply cannot be introduced due to the impossibility of base editing chemistry to convert between all codons. For example, the commonly mutated Leu858 residue was not identified in our base editing screens because its codon can only be targeted by BE3.9Max where it is predicted to introduce a synonymous mutation.

In this context, the prime editing technology could facilitate the exploration of a more clinically relevant mutational space. Indeed, prime editing guide RNAs rely on 3’ extensions serving as reverse transcriptase templates to introduce any type of substitution or short indels in proximity to their nicking site (162). This not only drastically increases the diversity of mutations that could be introduced in a single screen but also expands the number of targetable sites as prime editing can introduce edits beyond a strictly defined editing window. In addition, prime editing approaches relying on a single pegRNA generally introduce low levels of unintended edits, thus overcoming the major confounding effect of bystander edits in base editing experiments.



<u>Figure 4.1:</u> Comparison of EGFR codons accessible by ABE8e and BE3.9Max. *Venn diagram showing the number of EGFR codons that can be mutated by ABE8e and BE3.9Max, including synonymous mutations. All possible sgRNAs with an NGG PAM targeting EGFR exons were considered and editing outcomes were predicted using the base editor design tool from Hanna et al., 2021* (145)*.*

In spite of its relatively recent inception, multiple parameters influencing prime editing efficiency have been identified. First, the versatility of the approach is based on the PBS and RTT components of the pegRNA which respectively allow the priming of reverse transcription and encode the intended edit. Consequently, prime editing efficiency is influenced by the nucleotide compositions and melting temperatures of these elements, alongside the targeting efficiency of the spacer. Although general design rules and prediction algorithms have been devised to facilitate pegRNA design, optimal PBS and RTT lengths should be determined for each target to maximize editing efficiency (370). In a screening context, this could be conveniently achieved by including multiple pegRNAs with varying architectures for each variant.

Similarly to previous gene editing technologies, prime editing products are recognized and processed by the cellular DNA repair machinery. Namely, it has been shown that newly introduced edits can be excised by the MMR pathway, leading to MMR-deficient cell lines like HEK293 to be more amenable to prime editing. This inspired PE4, an approach relying on the expression of a dominant-negative MLH1 protein to inhibit the MMR pathway, thus increasing the final edited fraction by an average of 7.7-fold (170). In addition, engineered pegRNAs containing a 3’ pseudoknot motif have been shown to be protected from cellular nucleases, leading to enhanced editing (168).

In spite of these improvements, prime editing efficiencies remain highly variable across targets. For this reason, the only published prime editing screen uses PE3, which leverages an additional nicking guide RNA (ngRNA) to increase editing efficiency at the target but introduces a significant fraction of unwanted indels (183). While the authors circumvented this problem by quantifying edits by direct target site sequencing, this requires the design of separate pegRNA libraries for each protein exon. Furthermore, a high prevalence of unwanted indels can limit the sensitivity of enrichment screens. For example, while introducing indels in the EGFR coding sequence would most likely lead to growth arrest, in-frame insertions or deletions in exons 19 and 20 are known to impact EGFR activation and drug resistance which could bias screening read-outs. Consequently, the less error-prone PE2/PE4 technologies represent more reliable tools to infer variant effects by pegRNA library quantification.

In this chapter we thus set out to introduce hundreds of patient-derived mutations from the ClinVar and COSMIC databases in cells to assess their impact on EGFR activation and drug resistance. To do so, we designed and cloned a barcoded pegRNA library including redundant mutagenic pegRNAs and non-mutagenic controls combining multiple PBS and RTT lengths. Quantifying barcode distributions in treated and non-treated cell populations allowed us to measure variant impact on EGFR activity, as well as to evaluate the effect of different pegRNA architectures on prime editing efficiency. Lastly, we evaluated several optimization approaches to increase prime editing efficiency.

4.3 Pathogenic variant modeling in PC-9 and MCF10A cells

We first set out to evaluate prime editing efficiency in MCF10A and PC-9 cells. To do so, we infected each cell line with a lentiviral construct expressing the PE2 enzyme and GFP driven by an EF1α promoter, and isolated positive clones by FACS (Figure 4.2A). For pegRNA delivery, we designed a lentiviral vector expressing a puromycin resistance gene and a dominant negative MLH1 variant (MLH1$^{\Delta 754\text{-}756}$) which has been shown to drastically improve prime editing in MMR-proficient cells (170). Although endogenous MHL1 knock-out has been used in other studies, we reasoned that inhibiting MMR during pegRNA delivery would be more time-efficient and increase the generalisability of our approach in other cell lines.

We first delivered a published epegRNA introducing a CTT insertion at the HEK3 genomic locus in PE2-expressing PC-9 cells (162). To confirm the stability of our prime editor line, we performed the same experiment in parallel in cells derived from the same clone but frozen right after clonal expansion (passage 2) or kept in culture for 10 passages. Transduced cells were selected with puromycin two days after infection and harvested at day 8 and day 12 for target sequencing and edit quantification (Figure 4.2B). First, we noticed that the initial prime editing efficiency was relatively low with an average of 4.2% of sequencing reads containing the intended edit at day 8. However, this fraction reached 6.7% by day 12, representing a 58% increase in 4 days, suggesting that prime edits accumulate slowly in this cell line. Additionally, we observed at both timepoints a decrease of approximately 20% in prime editing efficiency between cells at passages 2 and 10, suggesting that the genome-integrated construct was progressively silenced during passaging.

Because of the relatively slow kinetics of prime editing in PC-9 cells, we next sought to determine whether the phenotypic selection used in our screens could impact edit accumulation. We thus repeated the same experiment but treated cells with Gefitinib on day 4 (Figure 4.2C). After 4 days of drug treatment, we observed no difference in editing efficiency between treated and non-treated cells. However, after 8 days of selection, we noticed a significant difference in edited fractions, with Gefitinib-treated cells reaching a plateau at 5% of edited alleles. Since the HEK3 edit is not expected to have any biological impact on EGFR signaling, this result suggests that prime editing is likely impacted by the inhibition of cell proliferation under drug selection.

We next set out to leverage prime editing to introduce a variant associated with drug resistance and delivered an epegRNA introducing the EGFR Thr790Met substitution to PE-expressing PC-9 and MCF10A cells (Figure 4.2D-E). Infected cells were treated with Gefitinib for 8 days after 4 (PC-9) or 16 (MCF10A) days of editing, respectively. In both cases, TKI treatment resulted in strong enrichment of the edit, reaching up to 12.9% in PC-9 and 37.6% in MCF10A cells.

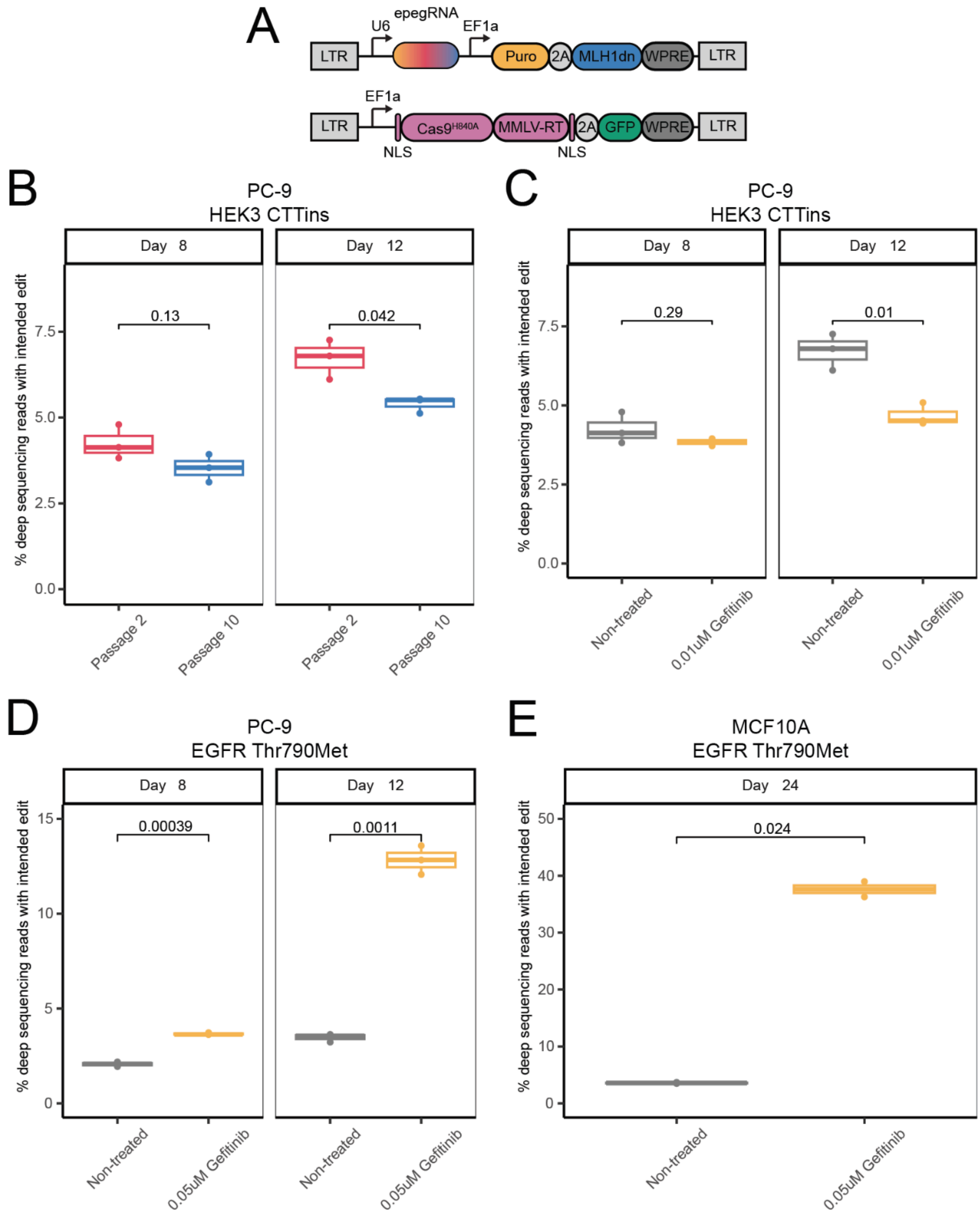


Figure 4.2: Evaluation of PE4 in MCF10A and PC-9 cells. *(A) Lentiviral vectors used for pegRNA and prime editor delivery. MLH1dn: MLH1 dominant negative variant, NLS: nuclear localization signal, LTR: long terminal repeat. (B-C) Deep sequencing quantification of a CTT insertion introduced at the HEK3 genomic locus in PC-9_PE2-2A-GFP cells (n= 3) at (B) passage 2 and passage 10 after stable prime editor cell line isolation or (C) after Gefitinib treatment. (D-E) Deep sequencing quantification of the EGFR Thr790Met edit with and without Gefitinib treatment in (D) PC-9 (n= 3) and (E) MCF10A (n= 2) cells. Days after infection at the time of cell harvest are shown, drug treatment lasted 4 and 8 days in PC-9 cells and 8 days in MCF10A cells. p-values are shown for two-sided unpaired t-tests.*

Although initial prime editing efficiencies are relatively low in PC-9 and MCF10A cells, these results demonstrate the high enrichment of drug-resistant edits in both cell lines. Importantly, edits accumulate slowly in PC-9 cells and this accumulation appears inhibited by TKI treatment. These results suggest that prime editing efficiency could be maximized by leaving as ample time as possible for cells to accumulate edits before initiating drug selection. Additionally, the genomically-integrated PE construct appears to be progressively silenced in cells, leading to decreasing editing efficiencies over time. To mitigate editor silencing, stable cell lines constitutively expressing the prime editor were routinely sorted for high expressors before each of the following experiments.

4.4 pegRNA library design

In order to screen a variety of mutations found in patients, we designed a pegRNA library introducing all possible EGFR variants smaller than 4 nucleotides and listed in ClinVar or that have at least 4 reports in the COSMIC database (Figure 4.3A). Of the 1100 EGFR variants found, 866 (~79%) could be introduced by pegRNAs with a maximum distance of 10 nucleotides between the edit and nicking site (Supplementary Figure S4.1A). Although prime editing can install edits beyond this distance, the optimal RTT length is generally thought to be around 12 nucleotides (163). We thus chose to restrict the size of our editing window in order to maintain the same RTT lengths throughout our library. The remaining 234 variants could not be targeted because no spacer was available in close proximity or because the corresponding pegRNAs contained polyT or BsmBI restriction sites that would preclude their expression or cloning, respectively.

For each accessible variant, a maximum of 5 Cas9 spacers were selected and used to design barcoded epegRNAs with 9 combinations of PBS and RTT lengths corresponding to the optimal range determined in self-targeting assays (Supplementary Figure S4.1B) (163). This redundancy was included to understand how these parameters impact prime editing efficiency and maximize our chances of successful editing. Next, for each mutagenic pegRNA in the library, we designed a matched non-mutagenic pegRNA with the same extension length but introducing the wild-type sequence. These are meant to control for unwanted mutations resulting from DNA nicking or sgRNA scaffold integration at the target site. Lastly, 1000 scrambled pegRNAs were designed with the same extension length distribution as the rest of the library to serve as non-targeting controls for library count normalization.

The resulting barcoded pegRNA library was cloned in two steps (Figure 4.3B, see the Material and Method section). Briefly, an oligonucleotide pool was ordered that replaces the sgRNA scaffold with random stuffers of variable lengths to normalize oligo length across the library. This pool was amplified and cloned into the lentiviral vector expressing a puromycin resistance marker and dominant negative MLH1 mutant via Gibson assembly. The sgRNA scaffold was then added to the final construct through Golden Gate cloning.

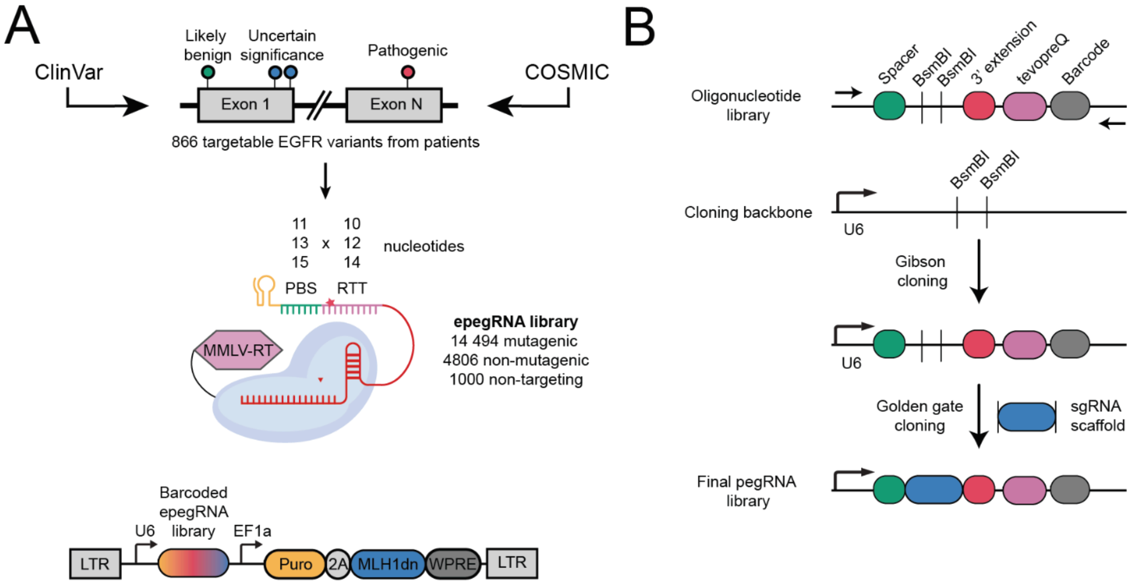


Figure 4.3: Design and cloning of a pooled pegRNA library. *(A) Schematic of the prime editing mutational screening approach. EGFR variants shorter than 4 nucleotides and in ClinVar or reported at least 4 times in COSMIC were used to design a pegRNA library. For each variant, mutagenic pegRNAs were designed with all combinations of three primer-binding sites (PBS) and three retro-transcriptase template (RTT) lengths. For each mutagenic pegRNA, a non-mutagenic control was added to the library which introduces the EGFR WT sequence. Lentiviral pegRNA library and prime editor delivery vectors are shown. (B) pegRNA library cloning steps.*

To quantify our pegRNA plasmid library, we amplified the insert with two different sets of primers either spanning the pegRNA barcode or the whole pegRNA (Figure 4.4A). Surprisingly, we observed that the counts obtained with the two primer sets followed different distributions. Notably, we observed that pegRNAs with longer PBS were under-represented when amplifying the full pegRNA instead of the barcode only (Figure 4.4B). Because the barcode-only amplicons are composed of fixed-length sequences with uniform GC contents, thus limiting PCR and sequencing error biases, we considered them as the “true” library element counts. However, when considering only the barcode counts, we once again observed a negative correlation between PBS length and pegRNA representation, while RTT length appeared to have a minimal impact (Figure 4.4C-D). By contrast, non-targeting pegRNAs with scrambled spacer and extensions sequences appeared over-represented in our library. We thus hypothesize that longer PBS negatively impact pegRNA cloning because of their complementary with the spacer, resulting in the formation of secondary structures during the oligo library amplification or cloning. Similarly, this bias is likely exacerbated when amplifying the full pegRNA for deep sequencing, suggesting that barcode sequencing is preferable for this type of library.

We decided to move forward with our library nonetheless as the global bias was minimal (skew ratio =6.9). However, this should be considered when designing larger pegRNA libraries, especially when including a broad range of PBS lengths.

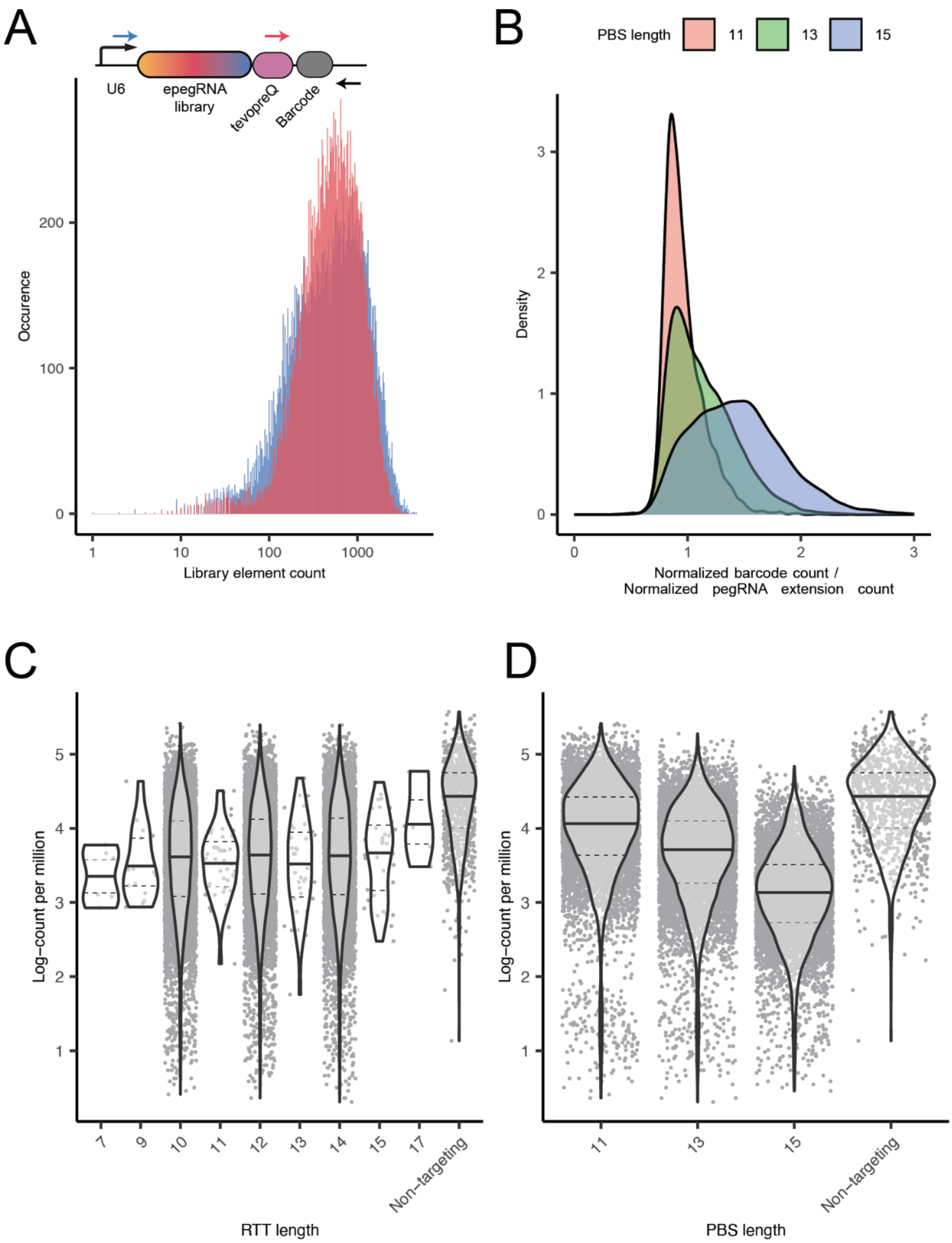


Figure 4.4: Barcoded pegRNA library quantification. *(A) Library barcode count distributions obtained with primer sets amplifying the barcode only (red) or the full pegRNA (blue). (B) Density plot showing the ratio of normalized counts obtained with both primers set and stratified by PBS lengths. (C-D) Scatterplots showing individual barcode counts per million in the final pegRNA library according to the (C) RTT and (D) PBS lengths of the corresponding pegRNAs. Generalized linear model (poisson regression) coefficients: PBS length = -0.2155608, (p-value < 2e-16), RTT length = 0.0089518, (p-value < 2e-16), intercept = 9.0445693 (p-value < 2e-16) on 20561 degrees of freedom.*

4.5 Screening of patient-derived mutations in MCF10A cells

We next set out to introduce patient-derived mutations in MCF10A cells and assess their EGFR activation potential. We thus delivered the pegRNA library to PE-expressing MCF10A cells and let them accumulate edits for 14 days before EGF depletion. Cells harvesting was performed after 8 days of selection and followed by pegRNA barcode sequencing. A cellular coverage of at least 1000x was maintained throughout the workflow. After confirming high replicate correlations (Supplementary Figure S4.2), we compared pegRNA counts between the non-treated and EGF-deprived arms of the screen (Figure 4.5A, Supplementary Figure S4.3). As expected, we observed no change in the distributions of non-targeting pegRNAs or non-mutagenic pegRNAs, suggesting that no measurable growth effect could be attributed to prime editing byproducts or indels.

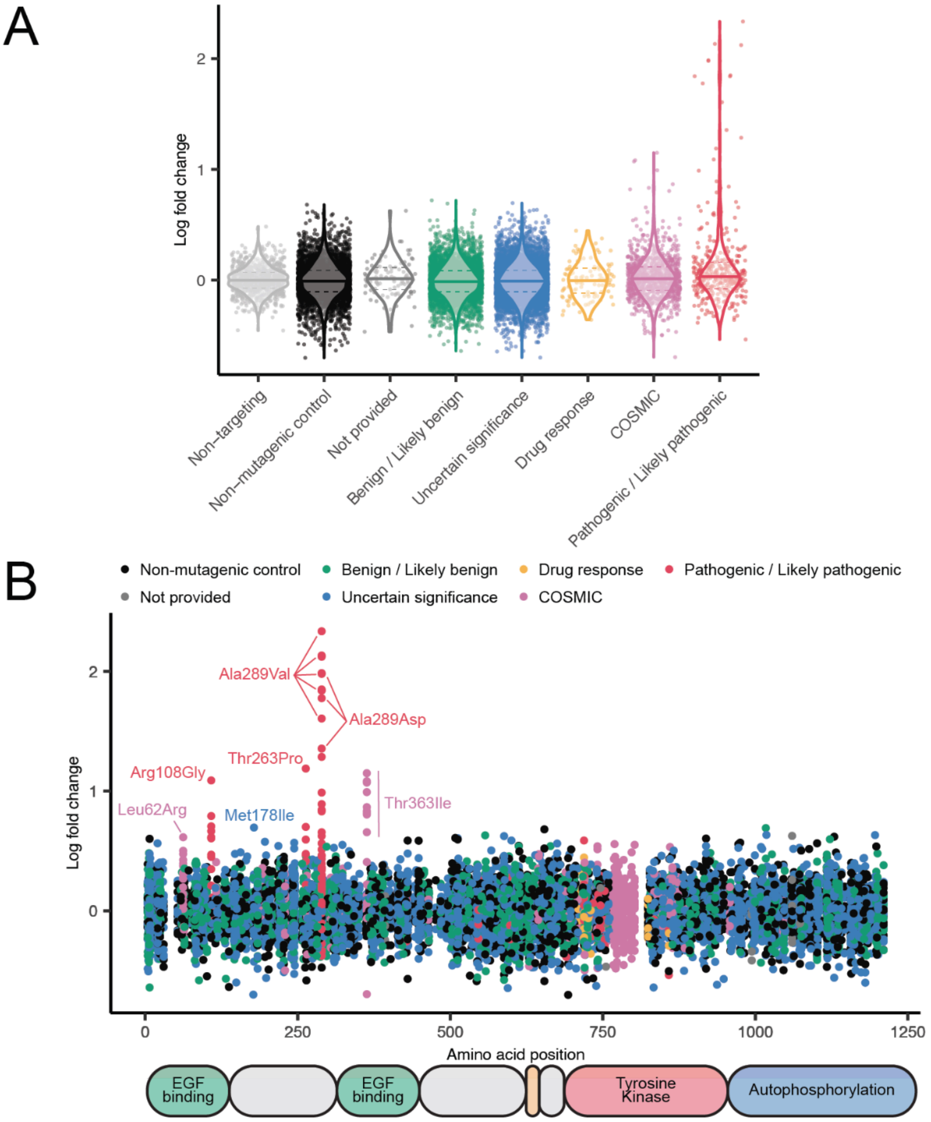


Figure 4.5: Prime editing screen of patient-derived variants identify EGFR activating mutations in MCF10A cells. *(A) Violin plots showing the log-fold changes of individual pegRNAs between non-treated and EGF-deprived cells stratified by clinical significance category. Solid horizontal lines represent populations medians and dashed lines represent 0.25 and 0.75 quantiles. (B) Log-fold changes of individual pegRNAs between non-treated and EGF-deprived cells along the EGFR protein.*

When considering pegRNAs individually, our screen revealed an enrichment of pegRNAs introducing the known pathogenic variants Ala289Val, Ala289Asp, Thr263Pro and Arg108Gly which are commonly found in glioblastoma (Figure 4.5B) (347,371). Additionally, we identified Thr363Ile, a variant present in COSMIC and found in glioma patients but absent from ClinVar. For all of these hits, individual pegRNAs of different extension lengths introducing the same edit were significantly enriched together, suggesting that these did not arise through technical noise.

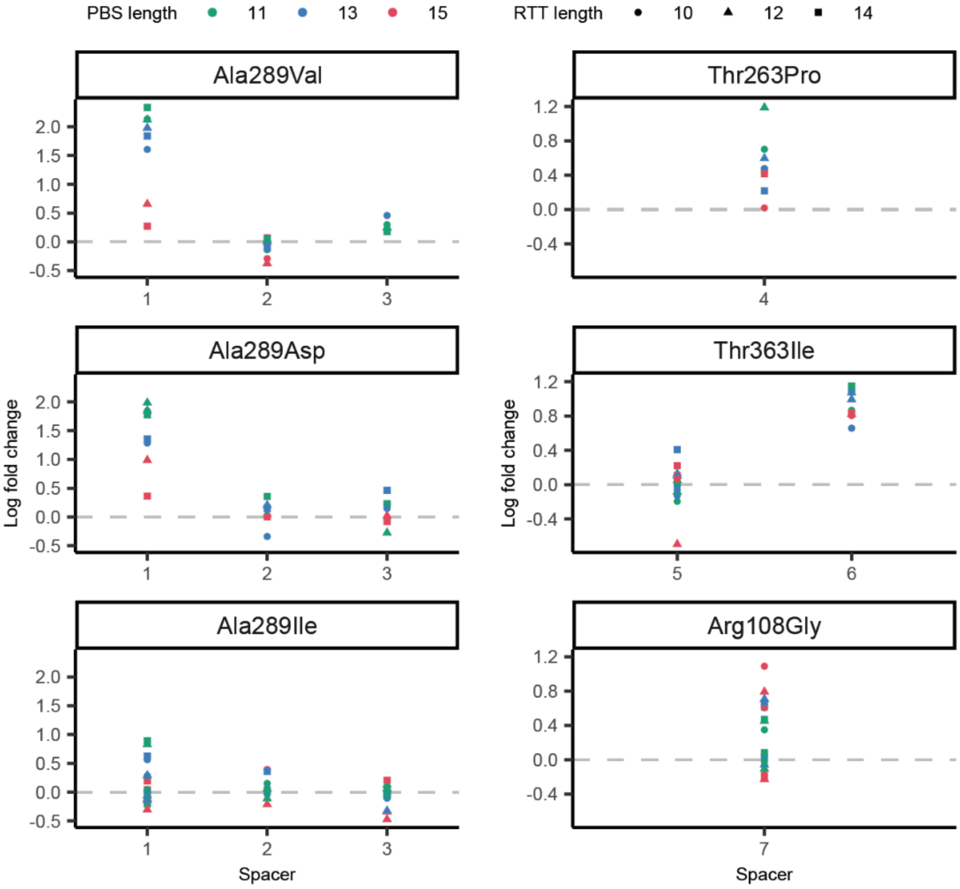


Figure 4.6: Relative enrichments of pegRNAs introducing the top 6 screen hits. *All pegRNAs introducing each of the top 6 screening hits are shown stratified by spacer sequence. PBS and RTT lengths are represented by color and shape, respectively*

While identifying pathogenic mutations, our screen misses the majority of known EGFR activating variants, in particular those in the tyrosine kinase domain. This is likely due to the low prime editing efficiency in MCF10A cells. Indeed, when considering the top 6 hits in our screen, we observed an important variability in enrichment between the corresponding pegRNAs (Figure 4.6). Notably, for a given edit, significant enrichment is generally observed for pegRNAs using the same Cas9 spacer while others produce no measurable phenotype. Furthermore, 3 of the 6 enriched variants are introduced by the same spacer, indicating that the corresponding pegRNAs enable prime editing regardless of the introduced mutation. Although PBS and RTT lengths also appear to have an impact on enrichment in some cases, the number of hits in our screen is too small to establish generalized pegRNA design rules. Taken together, this suggests that spacer sequence is a critical parameter for prime editing efficiency and that optimized PBS and RTT lengths, while impacting the final editing rate, are not sufficient to rescue an inactive pegRNA.

Taken together, these results demonstrate that prime editing screens represent a promising avenue to expand genetic diversity in variant screens and reveal pathogenic mutations undiscoverable by base editing. However, in its current installment prime editing suffers from low efficiency, thus limiting the sensitivity of such a screen. While PBS and RTT lengths appear to impact editing, our data suggest that spacer sequence is the most critical parameter in pegRNA design. Consequently, increasing spacer redundancy should be a priority when designing a pegRNA library. This could be done by allowing for a wider editing window between the edit and the nick or by using prime editors with relaxed PAM sequences to allow for more spacers to be targeted. If library size is a limiting factor, this approach can be combined with scoring algorithms to prioritize high-performance spacers.

4.6 Diphtheria toxin-based enrichment of prime edited cells

Multiple approaches have been devised to increase prime editing efficiency, including optimized enzymes and manipulation of cellular factors. Another approach is the enrichment of cells with successfully edited alleles. This is generally done by targeting a synthetic reporter alongside the intended target that is then used for the enrichment of editing-competent cells. In the case of prime editing, this was done by restoring a splice site in a GFP synthetic gene (372) or by removing a frameshift in a synthetic linker between two fluorescent proteins (373). These approaches have the advantage of yielding a gain-of-function phenotype that allows for easy enrichment of edited cells by FACS. However, in a screening context this would require the sorting of millions of cells in order to preserve library coverage. We thus took inspiration from a toxin-based enrichment system initially developed for base editing (374). Instead of restoring the activity of a reporter, this system disrupts the expression of the Heparin-binding EGF-like growth factor encoded by the endogenous HEBGF gene. This broadly-expressed membrane-anchored protein serves as a receptor to the diphtheria toxin, meaning that its disruption renders edited cells resistant to the toxin and allows for their simple enrichment by supplementing cell culture media with the toxin.

First, we designed a pegRNA introducing the HBEGF^{Glu141His} substitution, which has been shown to prevent toxin binding (375). In order to test toxin-based enrichment of independent edits, we inserted the HBEGF^{Glu141His} pegRNA or epegRNA under the control of a mouse U6 promoter in our lentiviral vector expressing the EGFR^{Thr790Met} epegRNA (Figure 4.7A). Both the pegRNA and epegRNA resulted in substantial HBEGF editing in PC-9 cells with 24.7% and 34.2% of alleles containing the intended edit, respectively (Figure 4.7B). Encouragingly, after 6 days of diphtheria toxin selection, the edited fraction was about 98% in both cases, confirming the selection of resistant cells carrying homozygous edits. We then quantified the EGFR^{Thr790Met} edit in selected and non-selected cells and saw an average increase of 1.38-fold when using the HBEGF-targeting pegRNA and 1.86-fold with the epegRNA (Figure 4.7C). Interestingly, the initial EGFR^{Thr790Met} fraction was slightly lower with the HBEGF^{Glu141His} epegRNA vector, possibly due to its increased stability leading to a greater competition between both epegRNAs to bind the prime editor.

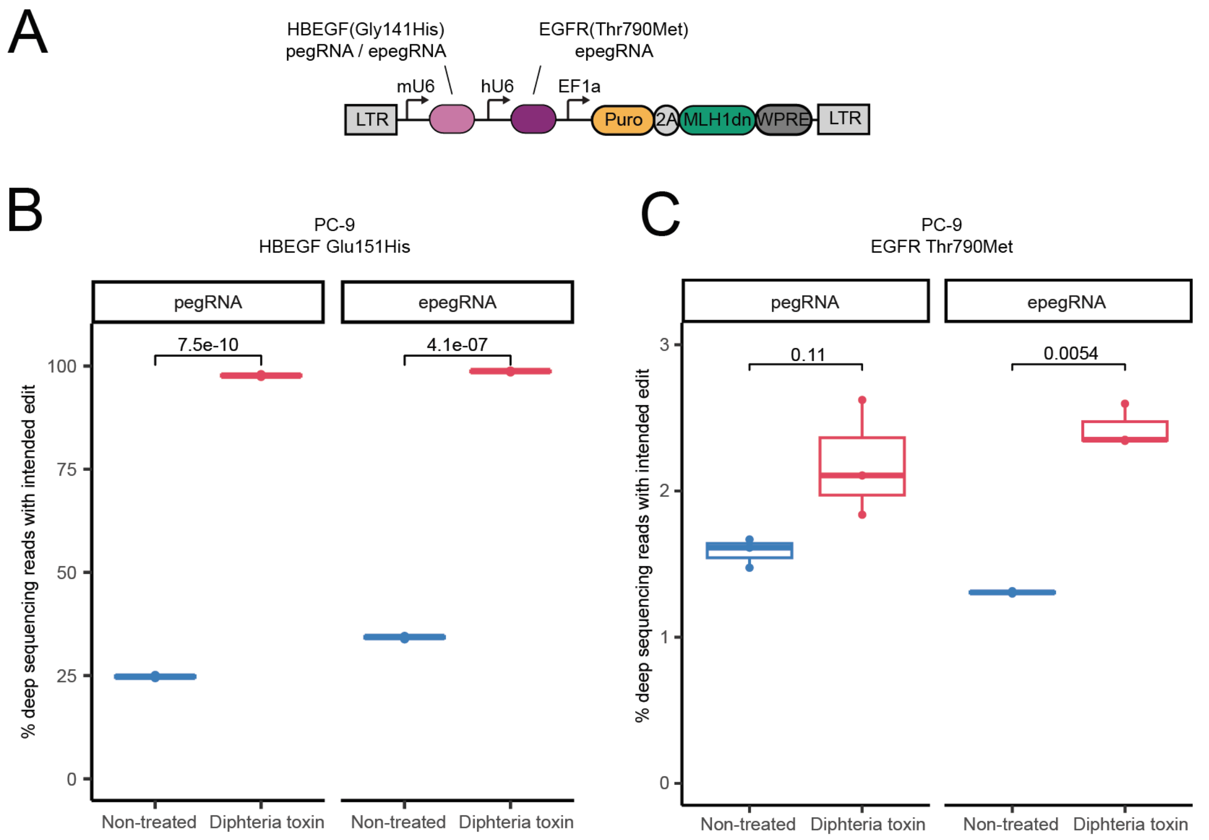


Figure 4.7: **Toxin-based enrichment of prime editing products in PC-9 cells.** *(A) Dual pegRNA lentiviral vector for the PE4-mediated introduction of HBEGF^{Glu141His} and EGFR^{Thr790Met} edits. The HBEGF mutation is installed by a pegRNA or an epegRNA while EGFR^{Thr790Met} is introduced by an epegRNA. Deep sequencing quantification of (B) HBEGF^{Glu141His} and (C) EGFR^{Thr790Met} edits in PC-9 cells infected with the toxin-based enrichment construct (n= 3) and selected with 20 ng/mL diphtheria toxin between day 9 and day 15 after transduction. p-values are shown for unpaired two-sided t-tests.*

In this part, we thus confirmed the feasibility of toxin-based enrichment of prime editing products. While more practical than FACS, this approach yielded lower enrichments than the 2- to 8-fold reported with fluorophore-based methods (372). This is likely due to the fact that cells must harbor a homozygous HBEGF mutation to acquire toxin resistance while a single edited reporter copy is enough to yield a measurable fluorescent signal. A fraction of cells with the intended edit but only a single copy of HBEGF^{Glu141His} is thus probably lost during toxin selection. Additionally, in our experiment, the HBEGF editing efficiency was much higher than that measured at the EGFR locus, suggesting that a significant number of toxin-resistant cells did not acquire the EGFR^{Thr790Met} edit. We thus speculate that the enrichment ratio could be improved by modulating the efficiency of the HBEGF^{Glu141His} pegRNA to make it more similar to that of the intended edit. This could be done by changing the pegRNA itself or performing toxin selection earlier after infection. However, targeting HBEGF using a pegRNA instead of an epegRNA did not improve the final edited EGFR^{Thr790Met} fraction in our experiment. In spite of these promising results, more experiments are thus required to optimize the toxin-based enrichment conditions.

4.7 Other optimization avenues

Various efforts have been made to enhance prime editor expression and nuclear localization (170,173). In order to determine if prime editor expression is a limiting factor *in vitro*, we infected PC-9 cells with increasing concentrations of PE2-expressing lentivirus. After antibiotics selection, transduced cells were infected with another lentiviral vector expressing the HEK3^{CTTins} epegRNA and cultured for 9 days before harvesting. Target sequencing and allele quantification revealed that the final edited fraction correlates with the PE lentiviral concentration (Figure 4.8A). Notably, cells infected at an MOI of 2 exhibited an average 45% increase in editing efficiency in comparison to cells infected at an MOI of 0.5, suggesting that prime editor expression is a limiting factor in our experimental setup.

Interestingly, we repeated the same experiment in the H1299 lung cancer cell line and measured a significantly higher overall editing efficiency (Figure 4.8B). Indeed, while the effect of lentiviral MOI appeared to have a smaller impact, the highest measured edited fraction was 56.7% in this cell line against 14.6% in PC-9 cells. H1299 cells are proficient in MMR, suggesting that other endogenous cellular factors determine prime editing efficiency in the two cell lines (376).

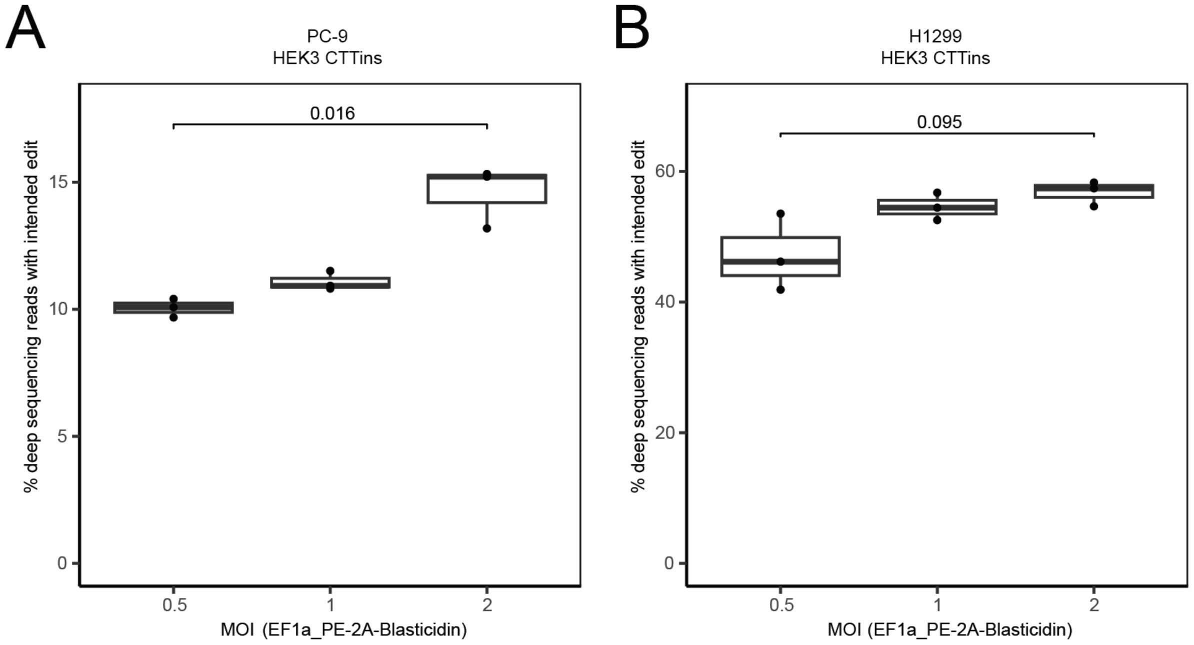


Figure 4.8: Prime editor expression limits PE4 efficiency in PC-9 and H1299 cells. *Deep sequencing quantification of HEK CTT insertion in (A) PC-9 and (B) H1299 cells infected with different concentrations of a lentiviral vector expressing the PE2-2A-Blasticidin S deaminase construct.*

In conclusion, prime editor expression and cell line selection are two potential avenues to increase screening sensitivity. While identifying cell lines that are both proficient in prime editing and produce biologically-relevant phenotypes is challenging, the identification of new endogenous factors impacting editing efficiency may allow for the engineering of optimized cell lines in the future. Additionally, prime editor overexpression through lentiviral delivery is limited by the cellular toxicity of high viral MOIs. In this context, recombinase-based delivery of multiple prime editor copies expressed from inducible promoters might increase their final expression while limiting gene silencing.

4.8 Conclusion

In this chapter, we designed and cloned a lentiviral pegRNA library to introduce hundreds of EGFR mutations observed in patients with PE4. We then delivered this library to MCF10A cells to identify mutations conferring EGF-independent growth and identified known extracellular mutations found in GBM as well as Thr363Ile, which was observed in patients but is currently not listed in ClinVar. This demonstrates for the first time the utility of prime editing screens without the need for locus haploidization with a library spanning a whole gene instead of hand-picked exons.

While quantifying our library, we uncovered a bias in pegRNA cloning and amplification likely resulting from the complementarity of the PBS and the spacer components. Such bias has not been reported in the literature and demonstrates the importance of quantifying pegRNA elements from a barcode amplified independently instead of amplifying the whole pegRNA sequence. Similarly, the impact of PBS lengths on library skewness should be considered during library design and could potentially be mitigated by reducing the number of PCR cycles or optimizing amplification conditions to minimize intramolecular interactions.

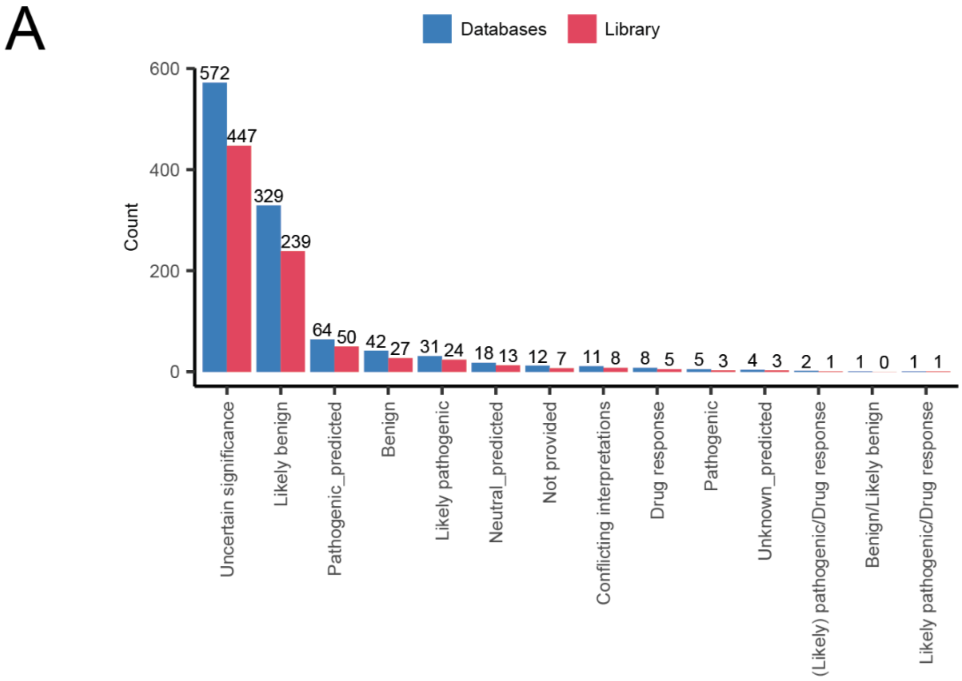
Despite identifying positive hits, our screen missed the majority of known EGFR pathogenic variants. This is likely due to the low prime editing efficiency in our cell line which should be confirmed by individual pegRNA validation in the kinase domain. While we observed significantly higher editing levels in other cell lines like H1299, finding cells with minimal chromosomal aberrations that produce measurable phenotypes in response to EGFR signaling remains challenging. Increasing the sensitivity of our screen thus requires improving prime editing efficiency in MCF10A cells which could be done by introducing a complete MLH1 gene knockout in these cells instead of the dominant-negative protein and by providing more time for cells to accumulate edits before EGF deprivation. Our data also demonstrate that prime editor expression is a limiting factor in our cell lines. However, viral delivery at high MOI is associated with important cell toxicity. Preferable approaches thus include improving prime editor delivery via plasmid electroporation or transposition and using enzymes with optimized nuclear localization like PEmax (170).

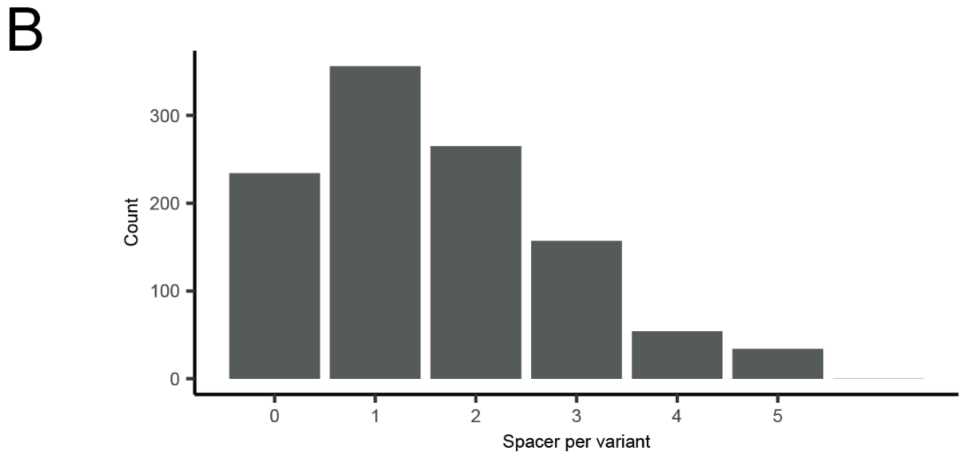
Contrary to base editing screens, prime editing libraries allow for built-in redundancy by introducing the same variant with pegRNAs using different spacer sequences and different RTT and PBS lengths. This not only increases the credibility of screening hits as multiple pegRNAs are enriched together, but also the likelihood of efficiently introducing a given variant. Furthermore, we observed in our screen that enriched pegRNAs generally rely on the same spacer, suggesting that spacer redundancy is a key parameter in PE4-based screens. Broadening the editing window beyond 10 nucleotides could thus allow for more spacers to be included in our library, thereby decreasing our false negative rate.

Lastly, we demonstrated toxin-based enrichment as a potential avenue to improve prime editing efficiency and screening sensitivity in the future. Although it only yielded a modest increase in the final edited fraction, this approach can be enhanced by finely tuning the editing efficiency at the HBEGF locus to maximize co-enrichment. Contrary to FACS-based enrichment, this method presents the advantage of being highly scalable and readily applicable in many cell types including primary cells (374). However, the biological impact of toxin-resistant HBEGF variants should be comprehensively characterized before it could be implemented in a screening workflow.

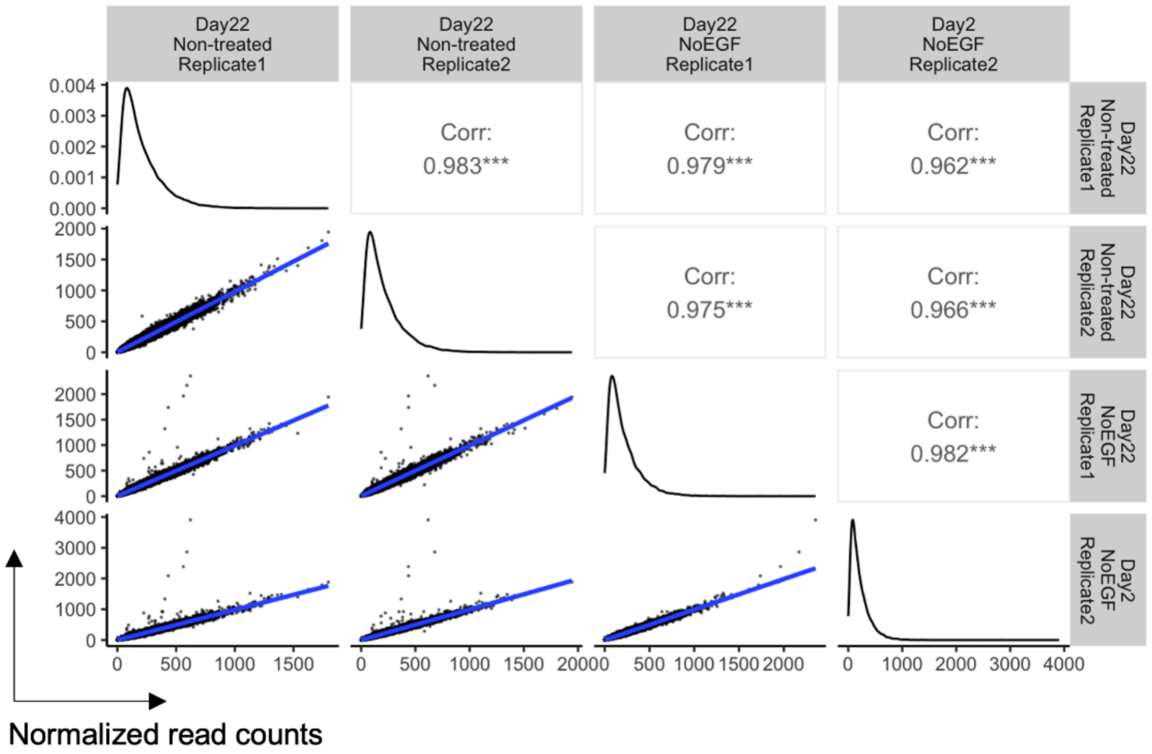
In conclusion, prime editing screens represent a promising tool to characterize pathogenic mutations undiscoverable by base editing, including insertions and deletions, in their genomic context. Additionally, prime editing allows for the design of pegRNA libraries from patient-derived variants, thus representing a complementary approach to unbiased variant scanning and paving the way to personalized genetic screens. While currently limited by low editing efficiency, we envision that future iterations of the prime editing technology will enable its broad use in *in vitro* models including patient-derived primary cells.

4.9 Supplementary figures

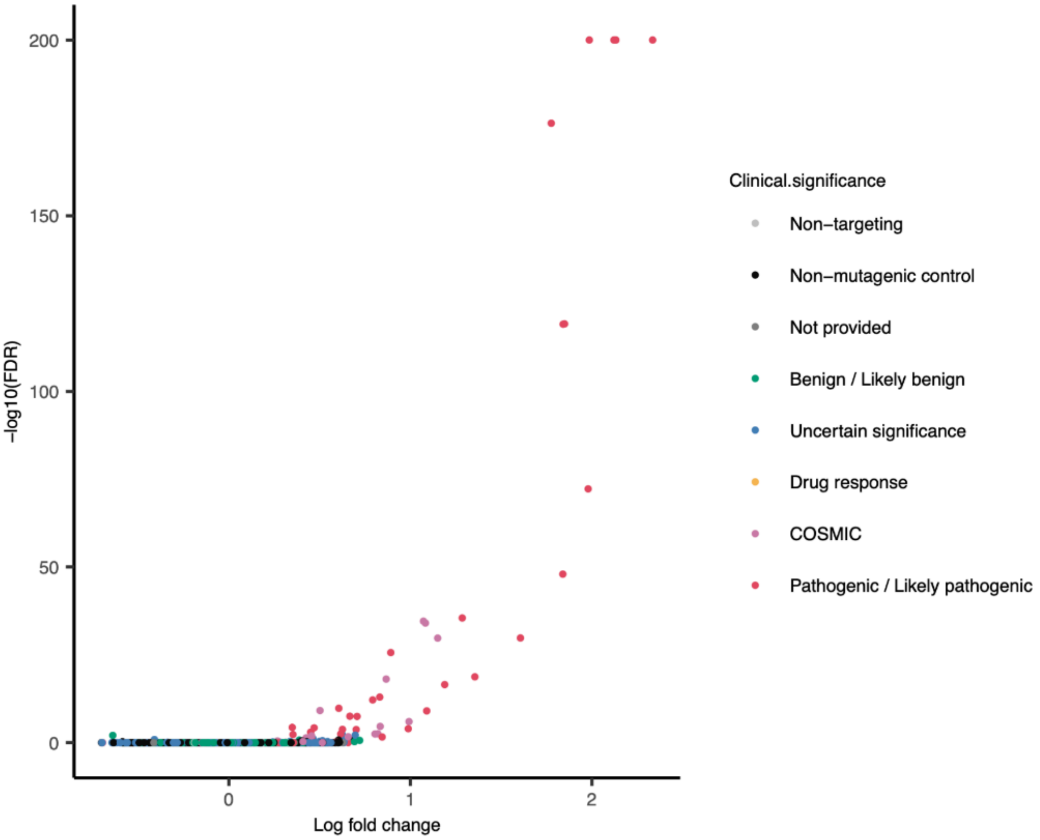




Supplementary Figure S4.1: pegRNA library characteristics. *(A) Number of EGFR variants from COSMIC and Clinvar considered for library design (blue, n= 1100) and present in the final library (red, n= 828) according to their clinical significance. (B) Distribution of the number of unique spacers introducing a considered variant in the final pegRNA library.*



Supplementary Figure S4.2: epegRNA library distributions and sample correlations for the prime editing EGFR activation screen in MCF10A cells. *Normalized barcode counts and Pearson correlation coefficients are shown for all samples.*



Supplementary Figure S4.3: Prime editing activation screen leads to the enrichment of EGFR variants conferring EGF-independent growth. *Volcano plot showing the log fold change enrichment and negative log false discovery rate of individual pegRNAs between non-treated and EGF-deprived cells. Colors represent the clinical significance of the corresponding variants.*

Chapter V: Discussion

5.1 Biological insights into EGFR variants

As genetic testing becomes increasingly prevalent in disease diagnosis and characterization, the prevalence of variants of uncertain significance hinders their interpretation and the adoption of personalized treatment protocols. In the case of lung and brain cancers, multiple tyrosine kinase inhibitors are approved for the treatment of tumors with mutated EGFR. However, their efficacy against uncommon EGFR variants remains largely unknown, leading to loss of therapeutic opportunities in patients. Due to the diverse nature of these variants, recruiting an adequate number of patients in clinical trials is challenging. Consequently, functional assays have emerged as a cost-effective solution for their characterization. However, existing multiplexed assays evaluating EGFR variant pathogenicity have relied on exogenous gene delivery in mouse cell lines, thereby failing to accurately replicate biologically-relevant cellular contexts, including EGFR expression levels and interaction partners. In this thesis, we developed complementary multiplexed assays of variant effects using cytosine base editors, adenine base editors, and prime editing technology. We tested our approach by introducing thousands of mutations in EGFR in multiple cell lines in response to EGF deprivation or treatment with the clinically-approved TKIs Gefitinib or Osimertinib. Unlike conventional CRISPR screens, these precision editing screens increase our ability to link genotype to phenotype at single nucleotide resolution while preserving endogenous genomic contexts.

One advantage of our screening approach is its ability to preserve endogenous expression levels which allows the precise assessment variant fitness. Indeed, mutant library overexpression poses the risk of promoting EGFR dimerization at the plasma membrane or saturating receptor internalization mechanisms, thus confounding activity measurements. In base editing screens, relative variant fitness can be easily extrapolated by taking sgRNA distributions at an early timepoint or in the plasmid library as a reference. By doing so, we showed that LOF EGFR variants can appear as false positives in phenotypic selection assays and bias screen results. Furthermore, our data show varying sensitivities of PC-9 cells to mutations compared to MCF10A, underscoring the importance of systematically considering variant fitness in this type of screen. Importantly, we chose to exclude variants leading to loss of cell viability from our screen hits. However, it should be noted that EGFR downregulation or LOF can be a transient mechanism of drug tolerance enabling the cells to regain viability through the accumulation of other oncogenic mutations or activation of alternative pathways (377,378).

When applied to EGFR activating mutations, our screens identified GOF variants in the tyrosine kinase domain, which are typical of NSCLC, as well as in the relatively understudied extracellular and C-terminal domains, more prevalently mutated in glioblastoma (GBM). This not only validates the relevance of MCF10A cells for variant screening across different cancer types, but also unveils a range of EGFR hyperactivation mechanisms. Namely, in addition to known and unknown mutations affecting the αC helix of the kinase domain, base editing scanning screens identified phosphorylation and ubiquitination sites involved in receptor downregulation. In the extracellular domains, multiple hits affected residues that interact in the tethered conformation of the inactive receptor, suggesting that these mutations could promote extracellular domain dimerisation. More surprisingly, we identified and validated splice site variants truncating the C-terminal tail after exon 25. Although this truncation was previously shown to lead to oncogenic transformation *in vitro* (345), we then demonstrated its resistance to both Osimertinib and Gefitinib when associated with an exon 19 deletion, warranting further characterization of its activation mechanism. Interestingly, gene fusions linking exon 25 to SEPT14 were reported in glioblastoma and NSCLC patients and shown to activate EGFR signaling, possibly through the same mechanism (379,380). Moreover, another study showed that this fusion combined with an EGFR exon 19 deletion was resistant to Gefitinib and Osimertinib (380).

Applied to drug resistance, base editing scanning screens revealed diverse mechanisms of resistance to Gefitinib and Osimertinib. Indeed, Gefitinib-resistant mutations were almost exclusively found in close proximity to the drug-binding site within the tyrosine kinase domain while Osimertinib-resistant variants occurred both in the tyrosine kinase and the C-terminal domains. Interestingly, multiple of the Osimertinib screen hits were found to be involved in autoinhibitory protein structures or recognition sites for other cellular factors, which could constitute targets for combined therapies. The existence of these divergent resistance mechanisms between first- and third-generation TKIs is currently unexplained. Indeed, mutations in the ATP binding pocket can be easily attributed to drug-specific interactions. However, it is unclear why general regulation mechanisms such as the mutations of Tyr1069, a ubiquitin-ligase recognition site, are found to confer resistance to Osimertinib but not Gefitinib. One possible explanation is the difference in inhibition mechanism of both drugs. Gefitinib is a reversible inhibitor, suggesting that mutations that increase affinity towards ATP could reverse drug binding. By contrast, a variant increasing the EGFR concentration at the cell surface might provide a greater advantage to cells in presence of an irreversible inhibitor like Osimertinib. An alternative explanation is that the distinct hits identified with both TKIs stem from a difference in screening sensitivity. Indeed, the Gefitinib resistance conferred by Thr790Met is, to our knowledge, the strongest acquired resistance to any TKI. Since this variant is by far the most enriched in each of our Gefitinib screens, we can speculate that its prevalence in our final library counts might obscure less enriched hits with more subtle resistance phenotypes.

Another important finding of our screens is the differences in drug-resistant variants between cell lines harboring wild-type EGFR and the Δ746-750 deletion. This difference likely results from different conformational constraints of the αC helix and ATP binding cleft in PC-9 cells, leading to altered intramolecular interactions within the kinase domain. A similar mechanism was previously demonstrated with the Gly724Ser variant which confers Osimertinib resistance when associated with exon 19 deletion but not Leu858Arg (381). However, in our case we compare a wild-type receptor with one that is not only structurally different but also constitutively active in cells showing oncogene addiction, which could play a role in the observed drug resistance differences. Additionally, PC-9 cells show EGFR gene copy number amplification, thus likely modifying the fraction of edited present in individual cells compared to diploid MCF10A (382). It would thus be interesting to investigate these variants in combination with pre-engineered Leu858Arg in MCF10A cells to determine whether they lead to resistance in any hyperactive receptor or solely in conjunction with a specific mutation. In summary, these results illustrate the need to consistently document acquired drug-resistant mutations alongside the associated primary mutation in patients and for EGFR variant screens to systematically evaluate compound mutations rather than isolated variants

5.2 Limitations of our screens

While the main limitation of our prime editing screen lies in low editing efficiencies, base editing scanning screens are mostly limited by the introduction of bystander edits and the lack of control on the introduced mutations. This drastically complicates the prediction of editing products and their proportions at the target site and the inference of genotype-phenotype relationships. For example, the Thr790Met,Gln791Ter[BE3.9Max] sgRNA used in our screens is predicted to introduce two variants with opposite phenotypic effects, namely drug resistance and receptor LOF. For this reason, the log fold changes measured in base editing screens cannot be directly extrapolated as relative resistance levels and thorough validation is required to classify variants on a resistance continuum. Similarly, it is currently impossible to determine the false negative rate in our screens as the exact spectrum of mutations effectively introduced in cells is unknown.

Target sequencing at different timepoints demonstrated that bystander edits abundance and proportions can be altered by precisely timing base editing experiments. For example, using inducible base editors could help to limit the retargeting of mutated sites and favor single edits. However, an ideal editing timeframe likely cannot be extrapolated across different targets, thus limiting the use of this approach in pooled experiments. To solve this issue, sensor libraries have been developed that leverage lentiviral constructs containing an sgRNA and its cognate target site (368). These allow for the pooled quantification of editing products introduced by each sgRNA and the prediction of their genomic impact in relevant cellular conditions, including non-canonical edits. Conveniently, such libraries can be directly used for screening as they can edit both the endogenous and the reporter targets in the same cell. The resulting datasets can then be used to normalize screen results or to design optimized follow-up libraries with known mutational profiles (383).

In addition, the implementation of sensor libraries could allow for the normalization and integration of screening results obtained with different gene editing technologies. Indeed, current differences in editing efficiencies and phenotype amplitudes between our screens prevent the direct comparison of log fold changes across base editors. In future installments of base editing screens, we also anticipate that all-in-one lentiviral constructs expressing different base editors can be pooled together by conveniently adding an editor-specific molecular barcode adjacent to the sgRNA to allow for read deconvolution. This will enable a single screening experiment to leverage multiple base editors, thereby increasing variant diversity while reducing reagent costs and lab work.

Another limitation of base editors pertains to their biased editing spectrum which limits the number of robust controls that can be used for variant classification. Indeed, the recommendations formulated by the American College of Medical Genetics for the interpretation of genetic variants through functional assays require that these include a set of multiple known benign and pathogenic variants to serve as references for the quantification of functional effects (81). While synonymous variants could reasonably be used as references for benign variants, our screens lack a number of pathogenic references. For example, while our Osimertinib screens identified multiple enriched variants, very few resistant mutations are known for this drug and none were introduced by our libraries. This limitation could be partly solved by expanding the spectrum of introduced variants with base editors showing a broader targeting capabilities. Indeed, multiple engineered Cas9 variants like SpRY-Cas9 (369) or Cas9-NG (384) show relaxed PAM requirements and a base editor using the latter was recently shown to identify Cys797, which could not be targeted in our screen, as an essential residue in EGFR (148).

Beyond gene editing constraints, a major limitation of our cellular models is that EGF is only one of the seven known ligands of EGFR. One theory to explain the difference in EGFR mutations found in GBM and NSCLC is that oncogenic extracellular mutations found in GBM favor low affinity ligands like epiregulin or amphiregulin (385). Consequently, our screens might miss ligand-specific variants as low affinity ligands are not present in our cell culture media. Likewise, some of our hits might behave differently in different tissues and tumors with various ligand availability. Expanding our screening approach to different tissue-relevant conditions or xenograft tumor models might thus provide unprecedented insights in the cancer-specific EGFR variant landscapes.

Lastly, drug resistance screen hits are contingent on the applied drug treatment, meaning that variants conferring insufficient drug resistance to yield measurable enrichment at the considered TKI concentrations cannot be captured. In this study, all drug treatments were conducted using the drugs IC50, which are likely significantly different from the *in vivo* bioavailability of these molecules. It is thus possible that variants showing resistance in our screen would be completely eliminated at TKI concentration used in patients. However, variants conferring partial drug tolerance remain clinically relevant, as they can enable cells to acquire other drug resistant phenotypes. Tumor xenograft represents an interesting tool to tackle this question as oncogenically transformed MCF10A cells can form tumors in mice, allowing the direct *in vivo* transposition of our activation and drug resistance screens.

5.3 Future directions for EGFR variant screens

An important open question in EGFR variant biology is the contribution of other ErbB receptors to oncogenic signaling. Indeed, while it is known that canonical oncogenic variants like Leu858Arg behave as “super acceptors” for homodimerization, their impact on heterodimerization with other receptors remains elusive. This contribution can be easily assessed by transposing our screens into MCF10A lines with inducible or knocked out ErbB receptors and comparing screening outcomes between induced and uninduced cells (386). In a complementary approach, interesting hits can be introduced in cells that are then subjected to genome-wide knock-out screens to identify genes involved in variant-associated phenotypes.

Similarly, different EGFR activating mutations result in distinct conformational constraints on the kinase domain, favoring specific homodimerization conformations. As observed in our PC-9 drug resistance screens, it is important to replicate EGFR variant screens in cell lines with different mutational backgrounds to evaluate the impact of compound mutations on drug resistance. MCF10A cells likely represent an ideal model for this application as different primary mutations can be introduced in their genome and easily enriched through EGF depletion to produce homozygous lines. In this context, base editing scanning screens present the advantage of relying on compact all-in-one libraries that can be readily deployed into multiple of these pre-engineered cell lines to screen for secondary mutations.

In addition to compound mutations *in cis*, previous studies have shown differences in drug resistance profiles for variant combinations *in trans* (273). For example, Thr790Met and Cys797Ser alleles respectively respond to second- and first-generation TKIs when found *in trans* but their combination *in cis* renders the cell insensitive to both drugs (387). Additionally, oncogenic EGFR variants have different multimerization requirements (388) with some acting almost exclusively as dimerization acceptors, thus requiring the co-expression of WT EGFR to produce a phenotype (219,222). In spite of this, mutant zigosity is often overlooked in functional assays and is not systematically reported in variant databases. Besides the co-delivery of synthetic genes, methods to genomically introduce heterozygous mutations by modulating HDR efficiency have been developed (389). However, like other HDR approaches, these are not applicable at the scale of a gene. In the case of a base editing scanning screen, this could be achieved by designing a library of mismatched sgRNA to reduce their editing efficiency, thus increasing the likelihood of obtaining heterozygous edits. However, this strategy would probably necessitate thorough optimizations and compromise editing efficiency and diversity. Another potential approach is the genomic introduction of a recoded copy of the gene with alternative codons. In this scenario, only the endogenous alleles could be edited during the screen, allowing the assessment of variants in a pseudo-heterozygous context.

Lastly, while precision genome editing screens expand the spectrum of genetic variations that could be studied at scale, our screens rely on a simple proliferation read-out that misses on mechanistic nuances of the measured phenotype. For example, *in vitro* experiments have shown that Thr790Met can either be rapidly selected from preexisting variants or slowly emerge from drug-tolerant cell populations (378). These different paths to resistance have implications for other facets of cellular adaptation which subsequently influence the ultimate drug tolerance profile of tumor cells. Future EGFR variant screens will thus benefit from assessing variant signaling via phosphoproteomics and transcriptomics at the single cell level, enabling a more comprehensive understanding of cell phenotypes.

5.4 Potential optimization avenues for prime editing screens

Low editing efficiency represents the primary bottleneck for prime editing screens. Consequently, a first step towards increasing screening sensitivity is the adoption of the PEmax enzyme which was optimized for high nuclease activity and nuclear localization (170). Additionally, our results illustrate that increasing prime editor expression represents a significant advantage. This could potentially be achieved by replacing lentiviral transduction with plasmid transfection or adenoviral delivery depending on the considered cell line (390). Interestingly, a recent study demonstrated that the nuclease and reverse transcriptase domains did not need to be fused to enable prime editing, suggesting that the expression of both components could be optimized independently (391). Additionally, since the beginning of this project, different preprints have been released that use prime editing in a pooled screening context and proposed different methods to enhance its efficiency. These include the direct knock-out of MLH1 (392), the introduction of additional silent mutations in the RTT to bypass MMR (393) and the use of an optimized sgRNA scaffold (394).

Interestingly, one of these reports uses the PE3 approach by co-delivering pegRNA and nicking sgRNA pairs (394). In this study, the authors controlled for unwanted edits by designing two separate libraries introducing alternative and reference alleles, respectively, and calculating log fold change ratios between the two. These reference pegRNAs, similar to our non-mutagenic controls, could thus be incorporated in a single PE3 library and allow the identification of phenotypes induced by unwanted edits. However, implementing this approach would require prior confirmation that mutagenic and non-mutagenic pegRNAs targeting the same locus indeed introduce the same spectrum and frequency of unwanted indels. Additionally, imprecise editing in PE3 screens targeting coding regions would likely lead to significant loss of information. Indeed, in the case of EGFR, we can speculate that a pegRNA/nicking sgRNA pair introducing an in-frame indel in exons 19 or 20 would lead to a strong phenotype and render all edits introduced by the same pair effectively undiscoverable.

The prime editing screens reported to this day tend to favor long screen durations to increase both edit accumulation and strong phenotypic selection. In a recent preprint, Gould et al. thus conducted a screen targeting TP53 for a total of 34 days (395). In another manuscript by Chardon et al., the authors screened for TKI-resistant EGFR variants using PE4 but applied drug selection for 19 days to increase hit enrichment (392). While it is unlikely that spontaneous drug resistance would emerge in this timeframe, we have shown that drug treatment inhibits the accumulation of edits. In this context, it might thus be preferable to increase the editing time before drug treatment than increasing selection time.

Additionally, recent studies suggest that the slow accumulation of prime editing products is a consequence of dNTP availability, constituting a bottleneck to reverse transcription (396,397). Notably, cellular dNTP concentration is dependent on cell cycle progression, potentially explaining decreased prime editing efficiencies reported in serum-deprived cells that remain in G1 phase (390). This impact of the cell cycle could also partially account for the significant efficiency differences observed between different cell lines and suggests that rapidly proliferating cells could constitute better models for prime editing screens. Interestingly, the inhibition of the SAMHD1 deoxynucleotide triphosphohydrolase has been shown to increase intracellular dNTP concentrations and to increase prime editing efficiency in cells (396,397). However, the perturbation of nucleotide metabolism might constitute an important confounding factor in cellular assays and would require thorough cell line characterisation before being employed in a screen using cell growth as a read-out.

Beyond the optimization of prime editing efficiency, our study also illustrates the critical role of pegRNA redundancy, through the use of multiple spacers, in ensuring the successful introduction of variants. This could be maximized by increasing the size of the considered editing window or by using a PAM-flexible prime editor accessing more spacers at the cost of lower editing efficiencies (398). Another elegant approach was recently proposed by Gould and colleagues that uses sensor libraries similar to those previously used in base editing screens (395). These self-targeting constructs allow for the assessment of individual pegRNAs during the screen and the exclusion of weak editors from the final screen analysis. Alternatively, they can be used in pre-screening experiments to design compact and optimized pegRNA libraries, thus allowing for higher cell coverage and sensitivity in the final screen.

Lastly, we demonstrated the feasibility of toxin-based prime edit enrichment by using a dual-pegRNA vector targeting the endogenous HBEGF gene. This method, unlike FACS-based methods, presents the advantage of allowing for the selection of edited cells regardless of the experiment scale. Furthermore, diphtheria toxin-based enrichment does not require any adjustment in library or experiment design compared to a classical screen and yielded a modest but significant increase in global editing efficiency in PC-9 cells. However, it is important to note that secreted HB-EGF is a known ligand of EGFR and other ErbB receptors (399). Although toxin-selected cells should be homogeneously homozygous for HBEGF^{Glu141His}, the impact of this variant on HB-EGF-mediated signaling should be evaluated before implementing this enrichment scheme in a screening context.

In conclusion, very little is currently known regarding the exact mechanism of prime editing. While the modulation of cellular factors has shown significant benefits in increasing editing efficiency, approaches that preserve cellular pathways integrity are preferable to allow for more relevant cellular models. It is thus likely that future prime editing screening approaches will rely on enhanced pegRNA designs as well as optimized nuclease and reverse transcriptase domains that will drastically improve the prime editing pipeline.

5.5 Future applications of precision genome editing screens

MAVEs have been predominantly used to study variants constituting disease risk factors or that predict drug response as these results can directly be leveraged in the clinics. To this end, the Pharmacogenomics Knowledge base (PharmGKB) aggregates curated information and guidelines regarding the use of genetic tests to inform drug treatments (400). The database currently contains annotations on 471 drugs and lists 35 “genes with substantial evidence to support their importance in pharmacogenomics” and 9 “genes which are important in tumor pharmacogenomics”, thus representing an important resource for screen target prioritization. Similarly, genes with approved genetic tests and an important number of VUS or conflicting reports can be prioritized as they could benefit the most from precise variant interpretations (50). A predominant example aside from EGFR is BRCA1/2 status, which informs treatment with PARP inhibitors in breast and ovarian cancers (401).

In addition to increasing the scale and accuracy of MAVEs, precision genome editing screens have the potential to expand their scope to applications that are incompatible with gene overexpression. Namely, a recent study reported a potentially deleterious impact of synonymous variants introduced in yeast via homologous recombination. While later criticized by other groups, this resurgence of interest in non-synonymous variations underscores the need for screening tools capable of studying these variants within their genomic context (402–404). Moreover, the controlled introduction of variants at genomic loci is particularly relevant for genes like PTEN that produce diverse phenotypes resulting from complex interactions between protein abundance and functions (405). Future iterations of these screens including single cell read-outs and controlled zygosity could thus help to elucidate the elusive functional differences between PTEN variants leading to autism spectrum disorder and tumor syndromes (406).

In this context, scRNA-seq tools have been developed to simultaneously profile cell transcriptomes and infer genetic perturbations at the single cell level. This has been done indirectly by capturing sgRNAs in experiments combining CROP-seq and base editors (407) or by delivering variant cDNA libraries harboring 3’ barcodes that could be sequenced at the mRNA level (408). However, recent approaches also allow for the direct targeted identification of genomic variants from cell transcriptomes. For example, a modified scRNA-seq workflow enabling the amplification of a site of interest from the cDNA pool after RNA capture and cell barcoding was used to determine the mutational profiles of individual patient-derived cancer cells (409). While this approach is directly applicable to single exons proximal to the transcript 3’ end, it does not allow for the screening of variants in a full-length transcript. This was solved by TISCC-seq, a method combining base editing, long read sequencing of target gene transcripts and short read transcriptome profiling (410). In the case of base editing, this has the advantage of stratifying cells based on actual variants instead of predicted edits profiles. Transposed to prime editing, these approaches could also help to identify and filter out unedited cells from downstreams analyses. Importantly, these technologies are currently limited to variants in transcribed regions and could be confounded by mutations leading to decreased transcriptional activity or RNA degradation. However, they hold the potential to distinguish between homozygous and heterozygous variants, thereby helping towards the identification of allelic dominance, compensation and buffering mechanisms (411).

To this day, MAVEs have generally ignored sequence insertions, deletions and inversions because of the difficulty to precisely introduce chromosomal rearrangements at the genomic level or to generate diverse libraries of these variants. Recent prime editing technologies like twinPE leverage matched pegRNA pairs to generate accurate large chromosomal rearrangements, thus unlocking a broad new spectrum of genetic variations that could be introduced *in vitro* (180). Similarly, the recently discovered CRISPR-associated transposases can introduce large payloads at specific genomic target sites (412,413). While still extremely inefficient in mammalian cells, these hold the potential to study largely unexplored copy number variations or replace full exons at once in a multiplexed way (414).

Aside from the large-scale exploration of genetic variants, precision genome editing screens also have direct applications in the clinics. For example, they could allow for the *in vitro* replication of an array of variants found in a patient by non-invasive tumor profiling to evaluate their relative competition under drug selection and better model heterogeneous tumors (415). In the context of EGFR, a patient-specific primary mutation can be introduced in MCF10A cells before screening for secondary mutations with an unbiased base editing scanning library or a targeted prime editing library introducing common variants. The sensitivity of the resulting compound mutations to a panel of drugs could then be assessed to establish personalized treatment courses anticipating the emergence of resistance.

Similarly, the relatively low toxicity of precision genome editing tools enables their direct application in patient-derived primary cells or induced pluripotent stem cells, facilitating the evaluation of risk alleles influenced by the polygenic background of the patient. Lastly, the interaction between specific mutations and cellular polygenic backgrounds could be assessed in a reversed approach using an isogenic cell line harboring a pathogenic variant of interest. Prime editing screens could then be used to screen for a broad range of polymorphisms present in population databases like gnomAD to identify common variations that modify pathogenic mutation phenotypes.

5.6 Concluding remarks and outlook

In this thesis, my goal was to establish an experimental framework for multimodal precision genetic screens and demonstrate how prime and base editing could be combined to assess genetic variants at scale while preserving their genomic context and expression levels. We introduced thousands of mutations in the EGFR gene of human cancer and non-cancer cells using two base editors and one prime editor and quantified the effect of the introduced variants on EGFR activation and drug resistance by library sequencing. By doing so, we uncovered diverse and cell-specific mechanisms of EGFR activation and regulation. These insights contribute to the existing knowledge on EGFR biology and call for future studies to characterize variant-specific signaling pathways and protein interaction partners.

These results highlight the distinct advantages and limitations of each technology. Namely, although base editors introduce a limited spectrum of amino acid substitutions, their high editing efficiency makes them ideal tools to identify important residues in applications where precision is not a limiting factor such as the mapping of drug binding sites. In turn, prime editing is more adapted to study patient-derived variants which are diverse in nature. Prime and base editing thus currently represent complementary tools for the exploration of protein function and disease mechanisms. In the future, we anticipate that the rapid improvements to prime editing efficiency will allow for saturating mutagenesis of full genes, making it the gold standard for comprehensive genetic variant screens. In the meantime, targeted exon mutagenesis coupled with direct sequencing and reporter libraries could serve as powerful tools for targeted mutagenesis studies.

Because of their distinct strengths and weaknesses, base and prime editing screen results currently cannot be directly integrated and compared. However, base editing screen results can potentially be used to narrow-down critical protein residues and design targeted pegRNA screens to further explore interesting hits using direct target sequencing. Furthermore, the recent publication of the AlphaMissense model trained on datasets including amino acid sequence, predicted protein structures and population genetics data suggests that multimodal functional assays can in the future be used by machine learning algorithms to better predict genetic variant significance (416).

Importantly, although both technologies are known to introduce DNA DSB at low frequency in cells, their relatively low toxicity compared to nuclease-based gene editing makes them ideal tools to study genetic variants in human primary cells. We thus anticipate that precision gene editing screens can be directly deployed to diverse cell types and genetic backgrounds to provide unprecedented insights into genotype-phenotype relationships. Lastly, as gene editing tools and cellular models become more accurate, we anticipate that precision variant screens data could become an essential component of personalized medicine to allow for accurate diagnostics and drug response prediction in the clinics.

References

1. Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, et al. A genome-wide mutational constraint map quantified from variation in 76,156 human genomes [Internet]. bioRxiv; 2022 [cited 2024 Jan 17]. p. 2022.03.20.485034. Available from: https://www.biorxiv.org/content/10.1101/2022.03.20.485034v2
2. Marian AJ. Clinical Interpretation and Management of Genetic Variants. JACC Basic Transl Sci. 2020 Oct 26;5(10):1029–42.
3. White RL, Lalouel JM, Lathrop GM, Leppert MF, Nakamura Y, O'Connell P. Mapping Approaches to Gene Identification in Humans. West J Med. 1987 Oct;147(4):423–7.
4. Menon AG, Klanke CA, Su YR. Identification of disease genes by positional cloning. Trends Cardiovasc Med. 1994;4(3):97–102.
5. Weber JL, May PE. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet. 1989 Mar;44(3):388–96.
6. Hudson TJ, Engelstein M, Lee MK, Ho EC, Rubenfield MJ, Adams CP, et al. Isolation and chromosomal assignment of 100 highly informative human simple sequence repeat polymorphisms. Genomics. 1992 Jul;13(3):622–9.
7. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet. 1980 May;32(3):314–31.
8. Kunkel LM. Analysis of deletions in DNA from patients with Becker and Duchenne muscular dystrophy. Nature. 1986 Jul;322(6074):73–7.
9. Hoffman EP, Brown RH, Kunkel LM. Dystrophin: the protein product of the Duchenne muscular dystrophy locus. Cell. 1987 Dec 24;51(6):919–28.
10. Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, et al. Identification of the cystic fibrosis gene: genetic analysis. Science. 1989 Sep 8;245(4922):1073–80.
11. Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science. 1989 Sep 8;245(4922):1066–73.
12. Rommens JM, Iannuzzi MC, Kerem BS, Drumm ML, Melmer G, Dean M, et al. Identification of the Cystic Fibrosis Gene: Chromosome Walking and Jumping. Science. 1989 Sep 8;245(4922):1059–65.
13. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001 Feb;409(6822):860–921.
14. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature. 2004 Oct;431(7011):931–45.
15. Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, et al. Genome-wide in situ exon capture for selective resequencing. Nat Genet. 2007 Dec;39(12):1522–7.
16. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature. 2009 Sep;461(7261):272–6.
17. Vissers LELM, de Vries BBA, Osoegawa K, Janssen IM, Feuth T, Choy CO, et al. Array-Based Comparative Genomic Hybridization for the Genomewide Detection of Submicroscopic Chromosomal Abnormalities. Am J Hum Genet. 2003 Dec 1;73(6):1261–70.
18. Kennedy GC, Matsuzaki H, Dong S, Liu W min, Huang J, Liu G, et al. Large-scale genotyping of complex DNA. Nat Biotechnol. 2003 Oct;21(10):1233–7.
19. Shen R, Fan JB, Campbell D, Chang W, Chen J, Doucet D, et al. High-throughput SNP genotyping on universal bead arrays. Mutat Res Mol Mech Mutagen. 2005 Jun 3;573(1):70–82.
20. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, et al. Complement Factor H Polymorphism in Age-Related Macular Degeneration. Science. 2005 Apr 15;308(5720):385–9.

21. Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A, et al. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007 Jun;447(7145):661–78.
22. Risch N, Merikangas K. The Future of Genetic Studies of Complex Human Diseases. Science. 1996 Sep 13;273(5281):1516–7.
23. Fabo T, Khavari P. Functional characterization of human genomic variation linked to polygenic diseases. Trends Genet. 2023 Jun 1;39(6):462–90.
24. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, et al. Complement Factor H Polymorphism in Age-Related Macular Degeneration. Science. 2005 Apr 15;308(5720):385–9.
25. Tanaka Y, Nishida N, Sugiyama M, Kurosaki M, Matsuura K, Sakamoto N, et al. Genome-wide association of IL28B with response to pegylated interferon-α and ribavirin therapy for chronic hepatitis C. Nat Genet. 2009 Oct;41(10):1105–9.
26. Chang J, Tian J, Yang Y, Zhong R, Li J, Zhai K, et al. A Rare Missense Variant in TCF7L2 Associates with Colorectal Cancer Risk by Interacting with a GWAS-Identified Regulatory Variant in the MYC Enhancer. Cancer Res. 2018 Sep 4;78(17):5164–72.
27. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. Nat Rev Genet. 2019 Aug;20(8):467–84.
28. Hirschhorn JN. Genomewide Association Studies — Illuminating Biologic Pathways. N Engl J Med. 2009 Apr 23;360(17):1699–701.
29. Goldstein DB. Common Genetic Variation and Human Traits. N Engl J Med. 2009 Apr 23;360(17):1696–8.
30. Church DM, Schneider VA, Graves T, Auger K, Cunningham F, Bouk N, et al. Modernizing Reference Genome Assemblies. PLOS Biol. 2011 Jul 5;9(7):e1001091.
31. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. Nature. 2015 Oct;526(7571):68–74.
32. Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, et al. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature. 2013 Jan;493(7431):216–20.
33. MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, et al. A Systematic Survey of Loss-of-Function Variants in Human Protein-Coding Genes. Science. 2012 Feb 17;335(6070):823–8.
34. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020 May;581(7809):434–43.
35. Minikel EV, Karczewski KJ, Martin HC, Cummings BB, Whiffin N, Rhodes D, et al. Evaluating drug targets through human loss-of-function genetic variation. Nature. 2020 May;581(7809):459–64.
36. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018 Oct;562(7726):203–9.
37. Li S, Carss KJ, Halldorsson BV, Cortes A, Consortium UBWGS. Whole-genome sequencing of half-a-million UK Biobank participants [Internet]. medRxiv; 2023 [cited 2024 Jan 17]. p. 2023.12.06.23299426. Available from: https://www.medrxiv.org/content/10.1101/2023.12.06.23299426v1
38. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. Am J Epidemiol. 2017 Nov 1;186(9):1026–34.
39. Henrie A, Hemphill SE, Ruiz-Schultz N, Cushman B, DiStefano MT, Azzariti D, et al. ClinVar Miner: Demonstrating utility of a Web-based tool for viewing and filtering ClinVar data. Hum Mutat. 2018;39(8):1051–60.
40. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, et al. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res. 2016 Jan 4;44(D1):D862–8.
41. Cline MS, Liao RG, Parsons MT, Paten B, Alquaddoomi F, Antoniou A, et al. BRCA Challenge: BRCA Exchange as a global resource for variants in BRCA1 and BRCA2.

PLOS Genet. 2018 Dec 26;14(12):e1007752.

42. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, et al. COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res. 2017 Jan 4;45(D1):D777–83.

43. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015 May 1;17(5):405–24.

44. Fayer S, Horton C, Dines JN, Rubin AF, Richardson ME, McGoldrick K, et al. Closing the gap: Systematic integration of multiplexed functional data resolves variants of uncertain significance in BRCA1, TP53, and PTEN. Am J Hum Genet. 2021 Dec 2;108(12):2248–58.

45. Harrison SM, Dolinsky JS, Johnson AEK, Pesaran T, Azzariti DR, Bale S, et al. Clinical laboratories collaborate to resolve differences in variant interpretations submitted to ClinVar. Genet Med. 2017 Oct 1;19(10):1096–104.

46. Federici G, Soddu S. Variants of uncertain significance in the era of high-throughput genome sequencing: a lesson from breast and ovary cancers. J Exp Clin Cancer Res. 2020 Mar 4;39(1):46.

47. MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, et al. Guidelines for investigating causality of sequence variants in human disease. Nature. 2014 Apr;508(7497):469–76.

48. Goldgar DE, Easton DF, Byrnes GB, Spurdle AB, Iversen ES, Greenblatt MS, et al. Genetic evidence and integration of various data sources for classifying uncertain variants into a single model. Hum Mutat. 2008;29(11):1265–72.

49. Maxwell KN, Hart SN, Vijai J, Schrader KA, Slavin TP, Thomas T, et al. Evaluation of ACMG-Guideline-Based Variant Classification of Cancer Susceptibility and Non-Cancer-Associated Genes in Families Affected by Breast Cancer. Am J Hum Genet. 2016 May 5;98(5):801–17.

50. Starita LM, Ahituv N, Dunham MJ, Kitzman JO, Roth FP, Seelig G, et al. Variant Interpretation: Functional Assays to the Rescue. Am J Hum Genet. 2017 Sep 7;101(3):315–25.

51. Forrest IS, Chaudhary K, Vy HMT, Petrazzini BO, Bafna S, Jordan DM, et al. Population-Based Penetrance of Deleterious Clinical Variants. JAMA. 2022 Jan 25;327(4):350–9.

52. Fahed AC, Wang M, Homburger JR, Patel AP, Bick AG, Neben CL, et al. Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. Nat Commun. 2020 Aug 20;11(1):3635.

53. Berger AH, Knudson AG, Pandolfi PP. A continuum model for tumour suppression. Nature. 2011 Aug;476(7359):163–9.

54. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet. 2019 Apr;51(4):584–91.

55. Osman A, Jonasson J. Cross-ethnic analysis of common gene variants in hemostasis show lopsided representation of global populations in genetic databases. BMC Med Genomics. 2022 Mar 25;15(1):69.

56. Sharo AG, Zou Y, Adhikari AN, Brenner SE. ClinVar and HGMD genomic variant classification accuracy has improved over time, as measured by implied disease burden. Genome Med. 2023 Jul 13;15(1):51.

57. Morash M, Mitchell H, Beltran H, Elemento O, Pathak J. The Role of Next-Generation Sequencing in Precision Medicine: A Review of Outcomes in Oncology. J Pers Med. 2018 Sep;8(3):30.

58. Fanale D, Fiorino A, Incorvaia L, Dimino A, Filorizzo C, Bono M, et al. Prevalence and Spectrum of Germline BRCA1 and BRCA2 Variants of Uncertain Significance in Breast/Ovarian Cancer: Mysterious Signals From the Genome. Front Oncol [Internet]. 2021 [cited 2024 Jan 17];11. Available from:

https://www.frontiersin.org/articles/10.3389/fonc.2021.682445

59. Li MM, Datto M, Duncavage EJ, Kulkarni S, Lindeman NI, Roy S, et al. Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. J Mol Diagn. 2017 Jan 1;19(1):4–23.
60. Makhnoon S, Shirts BH, Bowen DJ. Patients' perspectives of variants of uncertain significance and strategies for uncertainty management. J Genet Couns. 2019;28(2):313–25.
61. Vos J, Otten W, van Asperen C, Jansen A, Menko F, Tibben A. The counsellees' view of an unclassified variant in BRCA1/2: recall, interpretation, and impact on life. Psychooncology. 2008;17(8):822–30.
62. Gelling C. Hermann Muller on Measuring Mutation Rates. Genetics. 2016 Feb;202(2):369–70.
63. Muller HJ. Artificial Transmutation of the Gene. Science. 1927 Jul 22;66(1699):84–7.
64. Jacob F, Fuerst CR, Wollman EL. Study of Defective Lysogenic Bacteria. II.-Physiological Types resulting from Prophage Mutations. Ann Inst Pasteur. 1957;93(6):724–53.
65. Ichikawa-Ryo H, Kondo S. Indirect mutagenesis in phage lambda by ultraviolet preirradiation of host bacteria. J Mol Biol. 1975 Sep 5;97(1):77–92.
66. Hirsh D, Vanderslice R. Temperature-sensitive developmental mutants of Caenorhabditis elegans. Dev Biol. 1976 Mar 1;49(1):220–35.
67. Mullins MC, Hammerschmidt M, Haffter P, Nüsslein-Volhard C. Large-scale mutagenesis in the zebrafish: in search of genes controlling development in a vertebrate. Curr Biol. 1994 Mar 1;4(3):189–202.
68. Nolan PM, Peters J, Strivens M, Rogers D, Hagan J, Spurr N, et al. A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. Nat Genet. 2000 Aug;25(4):440–3.
69. Forsburg SL. The art and design of genetic screens: yeast. Nat Rev Genet. 2001 Sep;2(9):659–68.
70. Forment JV, Herzog M, Coates J, Konopka T, Gapp BV, Nijman SM, et al. Genome-wide genetic screening with chemically mutagenized haploid embryonic stem cells. Nat Chem Biol. 2017 Jan;13(1):12–4.
71. Noorani I, Bradley A, de la Rosa J. CRISPR and transposon in vivo screens for cancer drivers and therapeutic targets. Genome Biol. 2020 Aug 19;21(1):204.
72. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008 Nov;456(7218):53–9.
73. Cleary MA, Kilian K, Wang Y, Bradshaw J, Cavet G, Ge W, et al. Production of complex nucleic acid libraries using highly parallel in situ oligonucleotide synthesis. Nat Methods. 2004 Dec;1(3):241–8.
74. Gasperini M, Starita L, Shendure J. The power of multiplexed functional analysis of genetic variants. Nat Protoc. 2016 Oct;11(10):1782–7.
75. Gelman H, Dines JN, Berg J, Berger AH, Brnich S, Hisama FM, et al. Recommendations for the collection and use of multiplexed functional data for clinical variant interpretation. Genome Med. 2019 Dec 20;11(1):85.
76. Patwardhan RP, Hiatt JB, Witten DM, Kim MJ, Smith RP, May D, et al. Massively parallel functional dissection of mammalian enhancers in vivo. Nat Biotechnol. 2012 Mar;30(3):265–70.
77. Zhao W, Pollack JL, Blagev DP, Zaitlen N, McManus MT, Erle DJ. Massively parallel functional annotation of 3′ untranslated regions. Nat Biotechnol. 2014 Apr;32(4):387–91.
78. Majithia AR, Tsuda B, Agostini M, Gnanapradeepan K, Rice R, Peloso G, et al. Prospective functional classification of all possible missense variants in PPARG. Nat Genet. 2016 Dec;48(12):1570–5.
79. Kim I, Miller CR, Young DL, Fields S. High-throughput analysis of in vivo protein

stability. Mol Cell Proteomics MCP. 2013 Nov;12(11):3370–8.

80. Sahni N, Yi S, Taipale M, Fuxman Bass JI, Coulombe-Huntington J, Yang F, et al. Widespread Macromolecular Interaction Perturbations in Human Genetic Disorders. Cell. 2015 Apr 23;161(3):647–60.

81. Brnich SE, Abou Tayoun AN, Couch FJ, Cutting GR, Greenblatt MS, Heinen CD, et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. Genome Med. 2019 Dec 31;12(1):3.

82. Melnikov A, Rogov P, Wang L, Gnirke A, Mikkelsen TS. Comprehensive mutational scanning of a kinase in vivo reveals substrate-dependent fitness landscapes. Nucleic Acids Res. 2014 Aug 18;42(14):e112.

83. Wilson DS, Keefe AD. Random Mutagenesis by PCR. Curr Protoc Mol Biol. 2000;51(1):8.3.1-8.3.9.

84. Jain PC, Varadarajan R. A rapid, efficient, and economical inverse polymerase chain reaction-based method for generating a site saturation mutant library. Anal Biochem. 2014 Mar 15;449:90–8.

85. Kitzman JO, Starita LM, Lo RS, Fields S, Shendure J. Massively parallel single-amino-acid mutagenesis. Nat Methods. 2015 Mar;12(3):203–6.

86. Cunningham BC, Wells JA. High-Resolution Epitope Mapping of hGH-Receptor Interactions by Alanine-Scanning Mutagenesis. Science. 1989 Jun 2;244(4908):1081–5.

87. Fowler DM, Fields S. Deep mutational scanning: a new style of protein science. Nat Methods. 2014 Aug;11(8):801–7.

88. Starita LM, Young DL, Islam M, Kitzman JO, Gullingsrud J, Hause RJ, et al. Massively Parallel Functional Analysis of BRCA1 RING Domain Variants. Genetics. 2015 Jun 1;200(2):413–22.

89. Giacomelli AO, Yang X, Lintner RE, McFarland JM, Duby M, Kim J, et al. Mutational processes shape the landscape of TP53 mutations in human cancer. Nat Genet. 2018 Oct;50(10):1381–7.

90. Boonen RACM, Rodrigue A, Stoepker C, Wiegant WW, Vroling B, Sharma M, et al. Functional analysis of genetic variants in the high-risk breast cancer susceptibility gene PALB2. Nat Commun. 2019 Nov 22;10(1):5296.

91. Matreyek KA, Starita LM, Stephany JJ, Martin B, Chiasson MA, Gray VE, et al. Multiplex assessment of protein variant abundance by massively parallel sequencing. Nat Genet. 2018 Jun;50(6):874–82.

92. Forsyth CM, Juan V, Akamatsu Y, DuBridge RB, Doan M, Ivanov AV, et al. Deep mutational scanning of an antibody against epidermal growth factor receptor using mammalian cell display and massively parallel pyrosequencing. mAbs. 2013 Jul 1;5(4):523–32.

93. Wu NC, Young AP, Dandekar S, Wijersuriya H, Al-Mawsawi LQ, Wu TT, et al. Systematic Identification of H274Y Compensatory Mutations in Influenza A Virus Neuraminidase by High-Throughput Screening. J Virol. 2013 Jan 15;87(2):1193–9.

94. Inoue F, Ahituv N. Decoding enhancers using massively parallel reporter assays. Genomics. 2015 Sep 1;106(3):159–64.

95. Rosenberg AB, Patwardhan RP, Shendure J, Seelig G. Learning the Sequence Determinants of Alternative Splicing from Millions of Random Sequences. Cell. 2015 Oct 22;163(3):698–711.

96. Mighell TL, Evans-Dutson S, O'Roak BJ. A Saturation Mutagenesis Approach to Understanding PTEN Lipid Phosphatase Activity and Genotype-Phenotype Relationships. Am J Hum Genet. 2018 May 3;102(5):943–55.

97. Suiter CC, Moriyama T, Matreyek KA, Yang W, Scaletti ER, Nishii R, et al. Massively parallel variant characterization identifies NUDT15 alleles associated with thiopurine toxicity. Proc Natl Acad Sci. 2020 Mar 10;117(10):5394–401.

98. Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, et al. A TALE nuclease architecture for efficient genome editing. Nat Biotechnol. 2011 Feb;29(2):143–8.

99. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. Genome editing with

engineered zinc finger nucleases. Nat Rev Genet. 2010 Sep;11(9):636–46.

100. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science. 2012 Aug 17;337(6096):816–21.

101. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science. 2013 Feb 15;339(6121):819–23.

102. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-Guided Human Genome Engineering via Cas9. Science. 2013 Feb 15;339(6121):823–6.

103. Chang HHY, Pannunzio NR, Adachi N, Lieber MR. Non-homologous DNA end joining and alternative pathways to double-strand break repair. Nat Rev Mol Cell Biol. 2017 Aug;18(8):495–506.

104. Truong LN, Li Y, Shi LZ, Hwang PYH, He J, Wang H, et al. Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. Proc Natl Acad Sci. 2013 May 7;110(19):7720–5.

105. Joung J, Konermann S, Gootenberg JS, Abudayyeh OO, Platt RJ, Brigham MD, et al. Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. Nat Protoc. 2017 Apr;12(4):828–63.

106. Shalem O, Sanjana NE, Zhang F. High-throughput functional genomics using CRISPR–Cas9. Nat Rev Genet. 2015 May;16(5):299–311.

107. Bock C, Datlinger P, Chardon F, Coelho MA, Dong MB, Lawson KA, et al. High-content CRISPR screening. Nat Rev Methods Primer. 2022 Feb 10;2(1):1–23.

108. Mohr SE, Smith JA, Shamu CE, Neumüller RA, Perrimon N. RNAi screening comes of age: improved techniques and complementary approaches. Nat Rev Mol Cell Biol. 2014 Sep;15(9):591–600.

109. Wang T, Wei JJ, Sabatini DM, Lander ES. Genetic screens in human cells using the CRISPR/Cas9 system. Science. 2014 Jan 3;343(6166):80–4.

110. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelson T, et al. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Science. 2014 Jan 3;343(6166):84–7.

111. Koike-Yusa H, Li Y, Tan EP, Velasco-Herrera MDC, Yusa K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nat Biotechnol. 2014 Mar;32(3):267–73.

112. Zhou Y, Zhu S, Cai C, Yuan P, Li C, Huang Y, et al. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. Nature. 2014 May 22;509(7501):487–91.

113. Canver MC, Smith EC, Sher F, Pinello L, Sanjana NE, Shalem O, et al. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. Nature. 2015 Nov;527(7577):192–7.

114. Breslow DK, Hoogendoorn S, Kopp AR, Morgens DW, Vu BK, Kennedy MC, et al. A CRISPR-based screen for Hedgehog signaling provides insights into ciliary function and ciliopathies. Nat Genet. 2018 Mar;50(3):460–71.

115. Chow RD, Guzman CD, Wang G, Schmidt F, Youngblood MW, Ye L, et al. AAV-mediated direct in vivo CRISPR screen identifies functional suppressors in glioblastoma. Nat Neurosci. 2017 Oct;20(10):1329–41.

116. Chen S, Sanjana NE, Zheng K, Shalem O, Lee K, Shi X, et al. Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis. Cell. 2015 Mar 12;160(6):1246–60.

117. Li QV, Dixon G, Verma N, Rosen BP, Gordillo M, Luo R, et al. Genome-scale screens identify JNK–JUN signaling as a barrier for pluripotency exit and endoderm differentiation. Nat Genet. 2019 Jun;51(6):999–1010.

118. Sanson KR, Hanna RE, Hegde M, Donovan KF, Strand C, Sullender ME, et al. Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. Nat Commun. 2018 Dec 21;9(1):5416.

119. Kosicki M, Tomberg K, Bradley A. Repair of double-strand breaks induced by

CRISPR–Cas9 leads to large deletions and complex rearrangements. Nat Biotechnol. 2018 Sep;36(8):765–71.

120. Cullot G, Boutin J, Toutain J, Prat F, Pennamen P, Rooryck C, et al. CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations. Nat Commun. 2019 Mar 8;10(1):1136.

121. Lazar NH, Celik S, Chen L, Fay M, Irish JC, Jensen J, et al. High-resolution genome-wide mapping of chromosome-arm-scale truncations induced by CRISPR-Cas9 editing [Internet]. bioRxiv; 2023 [cited 2024 Jan 17]. p. 2023.04.15.537038. Available from: https://www.biorxiv.org/content/10.1101/2023.04.15.537038v1

122. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, et al. CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. Cell. 2013 Jul 18;154(2):442–51.

123. Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat Methods. 2013 Oct;10(10):973–6.

124. Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK. CRISPR RNA-guided activation of endogenous human genes. Nat Methods. 2013 Oct;10(10):977–9.

125. Kampmann M. CRISPRi and CRISPRa screens in mammalian cells for precision biology and medicine. ACS Chem Biol. 2018 Feb 16;13(2):406–16.

126. Neggers JE, Kwanten B, Dierckx T, Noguchi H, Voet A, Bral L, et al. Target identification of small molecules using large-scale CRISPR-Cas mutagenesis scanning of essential genes. Nat Commun. 2018 Feb 5;9(1):502.

127. Findlay GM, Daza RM, Martin B, Zhang MD, Leith AP, Gasperini M, et al. Accurate classification of BRCA1 variants with saturation genome editing. Nature. 2018 Oct;562(7726):217–22.

128. Radford EJ, Tan HK, Andersson MHL, Stephenson JD, Gardner EJ, Ironfield H, et al. Saturation genome editing of DDX3X clarifies pathogenicity of germline and somatic variation. Nat Commun. 2023 Dec 6;14(1):7702.

129. Essletzbichler P, Konopka T, Santoro F, Chen D, Gapp BV, Kralovics R, et al. Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. Genome Res. 2014 Dec;24(12):2059–65.

130. Hustedt N, Durocher D. The control of DNA repair by the cell cycle. Nat Cell Biol. 2017 Jan;19(1):1–9.

131. Yang H, Ren S, Yu S, Pan H, Li T, Ge S, et al. Methods Favoring Homology-Directed Repair Choice in Response to CRISPR/Cas9 Induced-Double Strand Breaks. Int J Mol Sci. 2020 Jan;21(18):6461.

132. Chu VT, Weber T, Wefers B, Wurst W, Sander S, Rajewsky K, et al. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat Biotechnol. 2015 May;33(5):543–8.

133. Lin S, Staahl BT, Alla RK, Doudna JA. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. eLife. 3:e04766.

134. Shams F, Bayat H, Mohammadian O, Mahboudi S, Vahidnezhad H, Soosanabadi M, et al. Advance trends in targeting homology-directed repair for accurate gene editing: An inclusive review of small molecules and modified CRISPR-Cas9 systems. BioImpacts BI. 2022;12(4):371–91.

135. Richardson CD, Ray GJ, DeWitt MA, Curie GL, Corn JE. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. Nat Biotechnol. 2016 Mar;34(3):339–44.

136. Renaud JB, Boix C, Charpentier M, De Cian A, Cochennec J, Duvernois-Berthet E, et al. Improved Genome Editing Efficiency and Flexibility Using Modified Oligonucleotides with TALEN and CRISPR-Cas9 Nucleases. Cell Rep. 2016 Mar 8;14(9):2263–72.

137. Aird EJ, Lovendahl KN, St. Martin A, Harris RS, Gordon WR. Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. Commun Biol. 2018 May 31;1(1):1–6.

138. Kim HK, Lee EJ, Lee YJ, Kim J, Kim Y, Kim K, et al. Impact of proactive

high-throughput functional assay data on BRCA1 variant interpretation in 3684 patients with breast or ovarian cancer. J Hum Genet. 2020 Mar;65(3):209–20.

139. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature. 2016 May;533(7603):420–4.

140. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature. 2017 Nov;551(7681):464–71.

141. Chen L, Hong M, Luan C, Gao H, Ru G, Guo X, et al. Adenine transversion editors enable precise, efficient A•T-to-C•G base editing in mammalian cells and embryos. Nat Biotechnol. 2023 Jun 15;1–13.

142. Kurt IC, Zhou R, Iyer S, Garcia SP, Miller BR, Langner LM, et al. CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. Nat Biotechnol. 2021 Jan;39(1):41–6.

143. Tong H, Wang X, Liu Y, Liu N, Li Y, Luo J, et al. Programmable A-to-Y base editing by fusing an adenine base editor with an N-methylpurine DNA glycosylase. Nat Biotechnol. 2023 Aug;41(8):1080–4.

144. Kweon J, Jang AH, Shin HR, See JE, Lee W, Lee JW, et al. A CRISPR-based base-editing screen for the functional assessment of BRCA1 variants. Oncogene. 2020 Jan;39(1):30–5.

145. Hanna RE, Hegde M, Fagre CR, DeWeirdt PC, Sangree AK, Szegletes Z, et al. Massively parallel assessment of human variants with base editor screens. Cell. 2021 Feb 18;184(4):1064-1080.e20.

146. Sangree AK, Griffith AL, Szegletes ZM, Roy P, DeWeirdt PC, Hegde M, et al. Benchmarking of SpCas9 variants enables deeper base editor screens of BRCA1 and BCL2. Nat Commun. 2022 Mar 14;13(1):1318.

147. Lue NZ, Garcia EM, Ngan KC, Lee C, Doench JG, Liau BB. Base editor scanning charts the DNMT3A activity landscape. Nat Chem Biol. 2023 Feb;19(2):176–86.

148. Li H, Ma T, Remsberg JR, Won SJ, DeMeester KE, Njomen E, et al. Assigning functionality to cysteines by base editing of cancer dependency genes. Nat Chem Biol. 2023 Nov;19(11):1320–30.

149. Enache OM, Rendo V, Abdusamad M, Lam D, Davison D, Pal S, et al. Cas9 activates the p53 pathway and selects for p53-inactivating mutations. Nat Genet. 2020 Jul;52(7):662–8.

150. Haapaniemi E, Botla S, Persson J, Schmierer B, Taipale J. CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. Nat Med. 2018 Jul;24(7):927–30.

151. Fiumara M, Ferrari S, Omer-Javed A, Beretta S, Albano L, Canarutto D, et al. Genotoxic effects of base and prime editing in human hematopoietic stem cells. Nat Biotechnol. 2023 Sep 7;1–15.

152. Ihry RJ, Worringer KA, Salick MR, Frias E, Ho D, Theriault K, et al. p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. Nat Med. 2018 Jul;24(7):939–46.

153. Martin-Rufino JD, Castano N, Pang M, Grody EI, Joubran S, Caulier A, et al. Massively parallel base editing to map variant effects in human hematopoiesis. Cell. 2023 May 25;186(11):2456-2474.e24.

154. Anzalone AV, Koblan LW, Liu DR. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. Nat Biotechnol. 2020 Jul;38(7):824–44.

155. Arbab M, Shen MW, Mok B, Wilson C, Matuszek Ż, Cassa CA, et al. Determinants of Base Editing Outcomes from Target Library Analysis and Machine Learning. Cell. 2020 Jul 23;182(2):463-480.e30.

156. Grünewald J, Zhou R, Garcia SP, Iyer S, Lareau CA, Aryee MJ, et al. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. Nature. 2019 May;569(7756):433–7.

157. Rees HA, Wilson C, Doman JL, Liu DR. Analysis and minimization of cellular RNA editing by DNA adenine base editors. Sci Adv. 2019 May 8;5(5):eaax5717.

158. Grünewald J, Zhou R, Iyer S, Lareau CA, Garcia SP, Aryee MJ, et al. CRISPR DNA base editors with reduced RNA off-target and self-editing activities. Nat Biotechnol. 2019 Sep;37(9):1041–8.

159. Mok BY, de Moraes MH, Zeng J, Bosch DE, Kotrys AV, Raguram A, et al. A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. Nature. 2020 Jul;583(7817):631–7.

160. Yi Z, Zhang X, Tang W, Yu Y, Wei X, Zhang X, et al. Strand-selective base editing of human mitochondrial DNA using mitoBEs. Nat Biotechnol. 2023 May 22;1–12.

161. Cho SI, Lee S, Mok YG, Lim K, Lee J, Lee JM, et al. Targeted A-to-G base editing in human mitochondrial DNA with programmable deaminases. Cell. 2022 May 12;185(10):1764-1776.e12.

162. Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature. 2019 Dec;576(7785):149–57.

163. Kim HK, Yu G, Park J, Min S, Lee S, Yoon S, et al. Predicting the efficiency of prime editing guide RNAs in human cells. Nat Biotechnol. 2021 Feb;39(2):198–206.

164. Mathis N, Allam A, Kissling L, Marquart KF, Schmidheini L, Solari C, et al. Predicting prime editing efficiency and product purity by deep learning. Nat Biotechnol. 2023 Aug;41(8):1151–9.

165. Koeppel J, Weller J, Peets EM, Pallaseni A, Kuzmin I, Raudvere U, et al. Prediction of prime editing insertion efficiencies using sequence features and DNA repair determinants. Nat Biotechnol. 2023 Oct;41(10):1446–56.

166. 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 May 11;186(10):2256-2272.e23.

167. Mathis N, Allam A, Tálas A, Benvenuto E, Schep R, Damodharan T, et al. Predicting prime editing efficiency across diverse edit types and chromatin contexts with machine learning [Internet]. bioRxiv; 2023 [cited 2024 Jan 17]. p. 2023.10.09.561414. Available from: https://www.biorxiv.org/content/10.1101/2023.10.09.561414v1

168. Nelson JW, Randolph PB, Shen SP, Everette KA, Chen PJ, Anzalone AV, et al. Engineered pegRNAs improve prime editing efficiency. Nat Biotechnol. 2022 Mar;40(3):402–10.

169. Ferreira da Silva J, Oliveira GP, Arasa-Verge EA, Kagiou C, Moretton A, Timelthaler G, et al. Prime editing efficiency and fidelity are enhanced in the absence of mismatch repair. Nat Commun. 2022 Feb 9;13(1):760.

170. 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 Oct 28;184(22):5635-5652.e29.

171. Li X, Zhou L, Gao BQ, Li G, Wang X, Wang Y, et al. Highly efficient prime editing by introducing same-sense mutations in pegRNA or stabilizing its structure. Nat Commun. 2022 Mar 29;13(1):1669.

172. Doman JL, Pandey S, Neugebauer ME, An M, Davis JR, Randolph PB, et al. Phage-assisted evolution and protein engineering yield compact, efficient prime editors. Cell. 2023 Aug 31;186(18):3983-4002.e26.

173. Velimirovic M, Zanetti LC, Shen MW, Fife JD, Lin L, Cha M, et al. Peptide fusion improves prime editing efficiency. Nat Commun. 2022 Jun 18;13(1):3512.

174. Schene IF, Joore IP, Oka R, Mokry M, van Vugt AHM, van Boxtel R, et al. Prime editing for functional repair in patient-derived disease models. Nat Commun. 2020 Oct 23;11(1):5352.

175. Liu Y, Li X, He S, Huang S, Li C, Chen Y, et al. Efficient generation of mouse models with the prime editing system. Cell Discov. 2020 Apr 28;6(1):1–4.

176. Davis JR, Banskota S, Levy JM, Newby GA, Wang X, Anzalone AV, et al. Efficient prime editing in mouse brain, liver and heart with dual AAVs. Nat Biotechnol. 2023 May 4;1–12.

177. Ely ZA, Mathey-Andrews N, Naranjo S, Gould SI, Mercer KL, Newby GA, et al. A

prime editor mouse to model a broad spectrum of somatic mutations in vivo. Nat Biotechnol. 2023 May 11;1–13.
178. Choi J, Chen W, Suiter CC, Lee C, Chardon FM, Yang W, et al. Precise genomic deletions using paired prime editing. Nat Biotechnol. 2022 Feb;40(2):218–26.
179. Wang J, He Z, Wang G, Zhang R, Duan J, Gao P, et al. Efficient targeted insertion of large DNA fragments without DNA donors. Nat Methods. 2022 Mar;19(3):331–40.
180. Anzalone AV, Gao XD, Podracky CJ, Nelson AT, Koblan LW, Raguram A, et al. Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. Nat Biotechnol. 2022 May;40(5):731–40.
181. Yarnall MTN, Ioannidi EI, Schmitt-Ulms C, Krajeski RN, Lim J, Villiger L, et al. Drag-and-drop genome insertion of large sequences without double-strand DNA cleavage using CRISPR-directed integrases. Nat Biotechnol. 2023 Apr;41(4):500–12.
182. Zheng C, Liu B, Dong X, Gaston N, Sontheimer EJ, Xue W. Template-jumping prime editing enables large insertion and exon rewriting in vivo. Nat Commun. 2023 Jun 8;14(1):3369.
183. Erwood S, Bily TMI, Lequyer J, Yan J, Gulati N, Brewer RA, et al. Saturation variant interpretation using CRISPR prime editing. Nat Biotechnol. 2022 Jun;40(6):885–95.
184. Cohen S. Isolation of a Mouse Submaxillary Gland Protein Accelerating Incisor Eruption and Eyelid Opening in the New-born Animal. J Biol Chem. 1962 May 1;237(5):1555–62.
185. Cohen S. The stimulation of epidermal proliferation by a specific protein (EGF). Dev Biol. 1965 Dec 1;12(3):394–407.
186. Cohen S, Ushiro H, Stoscheck C, Chinkers M. A native 170,000 epidermal growth factor receptor-kinase complex from shed plasma membrane vesicles. J Biol Chem. 1982 Feb 10;257(3):1523–31.
187. Schneider MR, Wolf E. The epidermal growth factor receptor ligands at a glance. J Cell Physiol. 2009;218(3):460–6.
188. Jura N, Endres NF, Engel K, Deindl S, Das R, Lamers MH, et al. Mechanism for Activation of the EGF Receptor Catalytic Domain by the Juxtamembrane Segment. Cell. 2009 Jun 26;137(7):1293–307.
189. Zhang X, Gureasko J, Shen K, Cole PA, Kuriyan J. An Allosteric Mechanism for Activation of the Kinase Domain of Epidermal Growth Factor Receptor. Cell. 2006 Jun 13;125(6):1137–49.
190. Yun CH, Boggon TJ, Li Y, Woo MS, Greulich H, Meyerson M, et al. Structures of Lung Cancer-Derived EGFR Mutants and Inhibitor Complexes: Mechanism of Activation and Insights into Differential Inhibitor Sensitivity. Cancer Cell. 2007 Mar 13;11(3):217–27.
191. Ferguson KM. A structure-based view of Epidermal Growth Factor Receptor regulation. Annu Rev Biophys. 2008;37:353–73.
192. Huang Y, Bharill S, Karandur D, Peterson SM, Marita M, Shi X, et al. Molecular basis for multimerization in the activation of the epidermal growth factor receptor. Dötsch V, editor. eLife. 2016 Mar 28;5:e14107.
193. Yarden Y, Sliwkowski MX. Untangling the ErbB signalling network. Nat Rev Mol Cell Biol. 2001 Feb;2(2):127–37.
194. Sato KI. Cellular functions regulated by phosphorylation of EGFR on Tyr845. Int J Mol Sci. 2013 May 23;14(6):10761–90.
195. Shan Y, Eastwood MP, Zhang X, Kim ET, Arkhipov A, Dror RO, et al. Oncogenic Mutations Counteract Intrinsic Disorder in the EGFR Kinase and Promote Receptor Dimerization. Cell. 2012 May 11;149(4):860–70.
196. Shostak K, Chariot A. EGFR and NF-κB: partners in cancer. Trends Mol Med. 2015 Jun 1;21(6):385–93.
197. Huang PH, Xu AM, White FM. Oncogenic EGFR Signaling Networks in Glioma. Sci Signal. 2009 Sep 8;2(87):re6–re6.
198. Wee P, Wang Z. Epidermal Growth Factor Receptor Cell Proliferation Signaling Pathways. Cancers. 2017 May;9(5):52.
199. Jones RB, Gordus A, Krall JA, MacBeath G. A quantitative protein interaction

network for the ErbB receptors using protein microarrays. Nature. 2006 Jan;439(7073):168–74.

200. Tarcic G, Boguslavsky SK, Wakim J, Kiuchi T, Liu A, Reinitz F, et al. An Unbiased Screen Identifies DEP-1 Tumor Suppressor as a Phosphatase Controlling EGFR Endocytosis. Curr Biol. 2009 Nov 17;19(21):1788–98.

201. Segatto O, Anastasi S, Alemà S. Regulation of epidermal growth factor receptor signalling by inducible feedback inhibitors. J Cell Sci. 2011 Jun 1;124(11):1785–93.

202. Grøvdal LM, Stang E, Sorkin A, Madshus IH. Direct interaction of Cbl with pTyr 1045 of the EGF receptor (EGFR) is required to sort the EGFR to lysosomes for degradation. Exp Cell Res. 2004 Nov 1;300(2):388–95.

203. Downward J, Yarden Y, Mayes E, Scrace G, Totty N, Stockwell P, et al. Close similarity of epidermal growth factor receptor and v-erb-B oncogene protein sequences. Nature. 1984 Feb;307(5951):521–7.

204. Gusterson B, Cowley G, Smith JA, Ozanne B. Cellular localisation of human epidermal growth factor receptor. Cell Biol Int Rep. 1984 Aug 1;8(8):649–58.

205. Cowley GP, Smith JA, Gusterson BA. Increased EGF receptors on human squamous carcinoma cell lines. Br J Cancer. 1986 Feb;53(2):223–9.

206. Velu TJ, Beguinot L, Vass WC, Willingham MC, Merlino GT, Pastan I, et al. Epidermal-growth-factor-dependent transformation by a human EGF receptor proto-oncogene. Science. 1987 Dec 4;238(4832):1408–10.

207. Chakraborty S, Li L, Puliyappadamba VT, Guo G, Hatanpaa KJ, Mickey B, et al. Constitutive and ligand-induced EGFR signalling triggers distinct and mutually exclusive downstream signalling networks. Nat Commun. 2014 Dec 15;5(1):5811.

208. Ogino S, Meyerhardt JA, Cantor M, Brahmandam M, Clark JW, Namgyal C, et al. Molecular Alterations in Tumors and Response to Combination Chemotherapy with Gefitinib for Advanced Colorectal Cancer. Clin Cancer Res. 2005 Sep 15;11(18):6650–6.

209. Byeon HK, Ku M, Yang J. Beyond EGFR inhibition: multilateral combat strategies to stop the progression of head and neck cancer. Exp Mol Med. 2019 Jan;51(1):1–14.

210. Weber F, Fukino K, Sawada T, Williams N, Sweet K, Brena RM, et al. Variability in organ-specific EGFR mutational spectra in tumour epithelium and stroma may be the biological basis for differential responses to tyrosine kinase inhibitors. Br J Cancer. 2005 May 23;92(10):1922–6.

211. Bitler BG, Goverdhan A, Schroeder JA. MUC1 regulates nuclear localization and function of the epidermal growth factor receptor. J Cell Sci. 2010 May 15;123(10):1716–23.

212. Lo HW, Xia W, Wei Y, Ali-Seyed M, Huang SF, Hung MC. Novel prognostic value of nuclear epidermal growth factor receptor in breast cancer. Cancer Res. 2005 Jan 1;65(1):338–48.

213. Psyrri A, Yu Z, Weinberger PM, Sasaki C, Haffty B, Camp R, et al. Quantitative determination of nuclear and cytoplasmic epidermal growth factor receptor expression in oropharyngeal squamous cell cancer by using automated quantitative analysis. Clin Cancer Res Off J Am Assoc Cancer Res. 2005 Aug 15;11(16):5856–62.

214. Ferlay J, Autier P, Boniol M, Heanue M, Colombet M, Boyle P. Estimates of the cancer incidence and mortality in Europe in 2006. Ann Oncol. 2007 Mar 1;18(3):581–92.

215. Ganti AK, Klein AB, Cotarla I, Seal B, Chou E. Update of Incidence, Prevalence, Survival, and Initial Treatment in Patients With Non–Small Cell Lung Cancer in the US. JAMA Oncol. 2021 Dec 1;7(12):1824–32.

216. Melosky B, Kambartel K, Häntschel M, Bennetts M, Nickens DJ, Brinkmann J, et al. Worldwide Prevalence of Epidermal Growth Factor Receptor Mutations in Non-Small Cell Lung Cancer: A Meta-Analysis. Mol Diagn Ther. 2022 Jan 1;26(1):7–18.

217. Jordan EJ, Kim HR, Arcila ME, Barron D, Chakravarty D, Gao J, et al. Prospective comprehensive molecular characterization of lung adenocarcinomas for efficient patient matching to approved and emerging therapies. Cancer Discov. 2017 Jun;7(6):596–609.

218. Leduc C, Merlio JP, Besse B, Blons H, Debieuvre D, Bringuier PP, et al. Clinical and molecular characteristics of non-small-cell lung cancer (NSCLC) harboring EGFR

mutation: results of the nationwide French Cooperative Thoracic Intergroup (IFCT) program. Ann Oncol. 2017 Nov 1;28(11):2715–24.

219. Red Brewer M, Yun CH, Lai D, Lemmon MA, Eck MJ, Pao W. Mechanism for activation of mutated epidermal growth factor receptors in lung cancer. Proc Natl Acad Sci. 2013 Sep 17;110(38):E3595–604.

220. Li K, Yang M, Liang N, Li S. Determining EGFR-TKI sensitivity of G719X and other uncommon EGFR mutations in non-small cell lung cancer: Perplexity and solution (Review). Oncol Rep. 2017 Mar 1;37(3):1347–58.

221. van Alderwerelt van Rosenburgh IK, Lu DM, Grant MJ, Stayrook SE, Phadke M, Walther Z, et al. Biochemical and structural basis for differential inhibitor sensitivity of EGFR with distinct exon 19 mutations. Nat Commun. 2022 Nov 10;13(1):6791.

222. Brown BP, Zhang YK, Kim S, Finneran P, Yan Y, Du Z, et al. Allele-specific activation, enzyme kinetics, and inhibitor sensitivities of EGFR exon 19 deletion mutations in lung cancer. Proc Natl Acad Sci. 2022 Jul 26;119(30):e2206588119.

223. Cho J, Chen L, Sangji N, Okabe T, Yonesaka K, Francis JM, et al. Cetuximab response of lung cancer-derived EGF receptor mutants is associated with asymmetric dimerization. Cancer Res. 2013 Nov 15;73(22):6770–9.

224. Brennan CW, Verhaak RGW, McKenna A, Campos B, Noushmehr H, Salama SR, et al. The Somatic Genomic Landscape of Glioblastoma. Cell. 2013 Oct 10;155(2):462–77.

225. Li X, Zhao L, Chen C, Nie J, Jiao B. Can EGFR be a therapeutic target in breast cancer? Biochim Biophys Acta BBA - Rev Cancer. 2022 Sep 1;1877(5):188789.

226. Huang HJS, Nagane M, Klingbeil CK, Lin H, Nishikawa R, Ji XD, et al. The Enhanced Tumorigenic Activity of a Mutant Epidermal Growth Factor Receptor Common in Human Cancers Is Mediated by Threshold Levels of Constitutive Tyrosine Phosphorylation and Unattenuated Signaling*. J Biol Chem. 1997 Jan 31;272(5):2927–35.

227. Gan HK, Cvrljevic AN, Johns TG. The epidermal growth factor receptor variant III (EGFRvIII): where wild things are altered. FEBS J. 2013;280(21):5350–70.

228. Sharma SV, Gajowniczek P, Way IP, Lee DY, Jiang J, Yuza Y, et al. A common signaling cascade may underlie “addiction” to the Src, BCR-ABL, and EGF receptor oncogenes. Cancer Cell. 2006 Nov 1;10(5):425–35.

229. Sato JD, Kawamoto T, Le AD, Mendelsohn J, Polikoff J, Sato GH. Biological effects in vitro of monoclonal antibodies to human epidermal growth factor receptors. Mol Biol Med. 1983 Dec;1(5):511–29.

230. Masui H, Kawamoto T, Sato JD, Wolf B, Sato G, Mendelsohn J. Growth Inhibition of Human Tumor Cells in Athymic Mice by Anti-Epidermal Growth Factor Receptor Monoclonal Antibodies1. Cancer Res. 1984 Mar 1;44(3):1002–7.

231. Sunada H, Magun BE, Mendelsohn J, MacLeod CL. Monoclonal antibody against epidermal growth factor receptor is internalized without stimulating receptor phosphorylation. Proc Natl Acad Sci U S A. 1986 Jun;83(11):3825–9.

232. Peng D, Fan Z, Lu Y, DeBlasio T, Scher H, Mendelsohn J. Anti-epidermal growth factor receptor monoclonal antibody 225 up-regulates p27KIP1 and induces G1 arrest in prostatic cancer cell line DU145. Cancer Res. 1996 Aug 15;56(16):3666–9.

233. Baselga J, Pfister D, Cooper MR, Cohen R, Burtness B, Bos M, et al. Phase I studies of anti-epidermal growth factor receptor chimeric antibody C225 alone and in combination with cisplatin. J Clin Oncol Off J Am Soc Clin Oncol. 2000 Feb;18(4):904–14.

234. Cunningham D, Humblet Y, Siena S, Khayat D, Bleiberg H, Santoro A, et al. Cetuximab Monotherapy and Cetuximab plus Irinotecan in Irinotecan-Refractory Metastatic Colorectal Cancer. N Engl J Med. 2004 Jul 22;351(4):337–45.

235. Li S, Schmitz KR, Jeffrey PD, Wiltzius JJW, Kussie P, Ferguson KM. Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. Cancer Cell. 2005 Apr 1;7(4):301–11.

236. Barber TD, Vogelstein B, Kinzler KW, Velculescu VE. Somatic Mutations of EGFR in Colorectal Cancers and Glioblastomas. N Engl J Med. 2004 Dec 30;351(27):2883–2883.

237. Grandis JR, Melhem MF, Gooding WE, Day R, Holst VA, Wagener MM, et al. Levels of TGF-α and EGFR Protein in Head and Neck Squamous Cell Carcinoma and Patient

Survival. JNCI J Natl Cancer Inst. 1998 Jun 3;90(11):824–32.

238. Marrocco I, Giri S, Simoni-Nieves A, Gupta N, Rudnitsky A, Haga Y, et al. L858R emerges as a potential biomarker predicting response of lung cancer models to anti-EGFR antibodies: Comparison of osimertinib vs. cetuximab. Cell Rep Med. 2023 Aug 15;4(8):101142.

239. Yarden Y, Pines G. The ERBB network: at last, cancer therapy meets systems biology. Nat Rev Cancer. 2012 Aug;12(8):553–63.

240. You KS, Yi YW, Cho J, Park JS, Seong YS. Potentiating Therapeutic Effects of Epidermal Growth Factor Receptor Inhibition in Triple-Negative Breast Cancer. Pharmaceuticals. 2021 Jun;14(6):589.

241. Ward WHJ, Cook PN, Slater AM, Davies DH, Holdgate GA, Green LR. Epidermal growth factor receptor tyrosine kinase: Investigation of catalytic mechanism, structure-based searching and discovery of a potent inhibitor. Biochem Pharmacol. 1994 Aug 17;48(4):659–66.

242. Carles F, Bourg S, Meyer C, Bonnet P. PKIDB: A Curated, Annotated and Updated Database of Protein Kinase Inhibitors in Clinical Trials. Molecules. 2018 Apr;23(4):908.

243. Abourehab MAS, Alqahtani AM, Youssif BGM, Gouda AM. Globally Approved EGFR Inhibitors: Insights into Their Syntheses, Target Kinases, Biological Activities, Receptor Interactions, and Metabolism. Molecules. 2021 Nov 4;26(21):6677.

244. Wakeling AE, Guy SP, Woodburn JR, Ashton SE, Curry BJ, Barker AJ, et al. ZD1839 (Iressa): An Orally Active Inhibitor of Epidermal Growth Factor Signaling with Potential for Cancer Therapy. Cancer Res. 2002 Oct 15;62(20):5749–54.

245. Moyer JD, Barbacci EG, Iwata KK, Arnold L, Boman B, Cunningham A, et al. Induction of apoptosis and cell cycle arrest by CP-358,774, an inhibitor of epidermal growth factor receptor tyrosine kinase. Cancer Res. 1997 Nov 1;57(21):4838–48.

246. Karachaliou N, Fernandez-Bruno M, Bracht JWP, Rosell R. EGFR first- and second-generation TKIs—there is still place for them in EGFR-mutant NSCLC patients. Transl Cancer Res. 2019 Jan;8(Suppl 1):S23–47.

247. Sordella R, Bell DW, Haber DA, Settleman J. Gefitinib-Sensitizing EGFR Mutations in Lung Cancer Activate Anti-Apoptotic Pathways. Science. 2004 Aug 20;305(5687):1163–7.

248. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, et al. Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non–Small-Cell Lung Cancer to Gefitinib. N Engl J Med. 2004 May 20;350(21):2129–39.

249. Mukohara T, Engelman JA, Hanna NH, Yeap BY, Kobayashi S, Lindeman N, et al. Differential Effects of Gefitinib and Cetuximab on Non–small-cell Lung Cancers Bearing Epidermal Growth Factor Receptor Mutations. JNCI J Natl Cancer Inst. 2005 Aug 17;97(16):1185–94.

250. Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, et al. Gefitinib or Carboplatin–Paclitaxel in Pulmonary Adenocarcinoma. N Engl J Med. 2009 Sep 3;361(10):947–57.

251. Lim SH, Lee JY, Sun JM, Ahn JS, Park K, Ahn MJ. Comparison of Clinical Outcomes Following Gefitinib and Erlotinib Treatment in Non–Small-Cell Lung Cancer Patients Harboring an Epidermal Growth Factor Receptor Mutation in Either Exon 19 or 21. J Thorac Oncol. 2014 Apr 1;9(4):506–11.

252. Kobayashi S, Boggon TJ, Dayaram T, Jänne PA, Kocher O, Meyerson M, et al. EGFR Mutation and Resistance of Non–Small-Cell Lung Cancer to Gefitinib. N Engl J Med. 2005 Feb 24;352(8):786–92.

253. Yu HA, Arcila ME, Rekhtman N, Sima CS, Zakowski MF, Pao W, et al. Analysis of Tumor Specimens at the Time of Acquired Resistance to EGFR-TKI Therapy in 155 Patients with EGFR-Mutant Lung Cancers. Clin Cancer Res. 2013 Apr 15;19(8):2240–7.

254. Li D, Ambrogio L, Shimamura T, Kubo S, Takahashi M, Chirieac LR, et al. BIBW2992, an irreversible EGFR/HER2 inhibitor highly effective in preclinical lung cancer models. Oncogene. 2008 Aug;27(34):4702–11.

255. Yang JCH, Wu YL, Schuler M, Sebastian M, Popat S, Yamamoto N, et al. Afatinib

versus cisplatin-based chemotherapy for EGFR mutation-positive lung adenocarcinoma (LUX-Lung 3 and LUX-Lung 6): analysis of overall survival data from two randomised, phase 3 trials. Lancet Oncol. 2015 Feb;16(2):141–51.

256. Camidge DR, Pao W, Sequist LV. Acquired resistance to TKIs in solid tumours: learning from lung cancer. Nat Rev Clin Oncol. 2014 Aug;11(8):473–81.

257. Cross DAE, Ashton SE, Ghiorghiu S, Eberlein C, Nebhan CA, Spitzler PJ, et al. AZD9291, an Irreversible EGFR TKI, Overcomes T790M-Mediated Resistance to EGFR Inhibitors in Lung Cancer. Cancer Discov. 2014 Sep 1;4(9):1046–61.

258. Jänne PA, Yang JCH, Kim DW, Planchard D, Ohe Y, Ramalingam SS, et al. AZD9291 in EGFR Inhibitor–Resistant Non–Small-Cell Lung Cancer. N Engl J Med. 2015 Apr 30;372(18):1689–99.

259. Soria JC, Ohe Y, Vansteenkiste J, Reungwetwattana T, Chewaskulyong B, Lee KH, et al. Osimertinib in Untreated EGFR-Mutated Advanced Non–Small-Cell Lung Cancer. N Engl J Med. 2018 Jan 11;378(2):113–25.

260. Passaro A, Jänne PA, Mok T, Peters S. Overcoming therapy resistance in EGFR-mutant lung cancer. Nat Cancer. 2021 Apr;2(4):377–91.

261. Zhai X, Ward RA, Doig P, Argyrou A. Insight into the Therapeutic Selectivity of the Irreversible EGFR Tyrosine Kinase Inhibitor Osimertinib through Enzyme Kinetic Studies. Biochemistry. 2020 Apr 14;59(14):1428–41.

262. Santoni-Rugiu E, Melchior LC, Urbanska EM, Jakobsen JN, de Stricker K, Grauslund M, et al. Intrinsic Resistance to EGFR-Tyrosine Kinase Inhibitors in EGFR-Mutant Non-Small Cell Lung Cancer: Differences and Similarities with Acquired Resistance. Cancers. 2019 Jul;11(7):923.

263. Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. Lancet Oncol. 2012 Mar 1;13(3):239–46.

264. Chmielecki J, Mok T, Wu YL, Han JY, Ahn MJ, Ramalingam SS, et al. Analysis of acquired resistance mechanisms to osimertinib in patients with EGFR-mutated advanced non-small cell lung cancer from the AURA3 trial. Nat Commun. 2023 Feb 27;14(1):1071.

265. Sequist LV, Waltman BA, Dias-Santagata D, Digumarthy S, Turke AB, Fidias P, et al. Genotypic and Histological Evolution of Lung Cancers Acquiring Resistance to EGFR Inhibitors. Sci Transl Med. 2011 Mar 23;3(75):75ra26-75ra26.

266. Niederst MJ, Sequist LV, Poirier JT, Mermel CH, Lockerman EL, Garcia AR, et al. RB loss in resistant EGFR mutant lung adenocarcinomas that transform to small-cell lung cancer. Nat Commun. 2015 Mar 11;6(1):6377.

267. Zhang Z, Lee JC, Lin L, Olivas V, Au V, LaFramboise T, et al. Activation of the AXL kinase causes resistance to EGFR-targeted therapy in lung cancer. Nat Genet. 2012 Aug;44(8):852–60.

268. Lee CK, Wu YL, Ding PN, Lord SJ, Inoue A, Zhou C, et al. Impact of Specific Epidermal Growth Factor Receptor (EGFR) Mutations and Clinical Characteristics on Outcomes After Treatment With EGFR Tyrosine Kinase Inhibitors Versus Chemotherapy in EGFR-Mutant Lung Cancer: A Meta-Analysis. J Clin Oncol [Internet]. 2015 Apr 20 [cited 2024 Jan 17]; Available from: https://ascopubs.org/doi/10.1200/JCO.2014.58.1736

269. Yun CH, Mengwasser KE, Toms AV, Woo MS, Greulich H, Wong KK, et al. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. Proc Natl Acad Sci. 2008 Feb 12;105(6):2070–5.

270. Vikis H, Sato M, James M, Wang D, Wang Y, Wang M, et al. EGFR-T790M Is a Rare Lung Cancer Susceptibility Allele with Enhanced Kinase Activity. Cancer Res. 2007 May 15;67(10):4665–70.

271. Regales L, Balak MN, Gong Y, Politi K, Sawai A, Le C, et al. Development of new mouse lung tumor models expressing EGFR T790M mutants associated with clinical resistance to kinase inhibitors. PloS One. 2007 Aug 29;2(8):e810.

272. Vivanco I, Robins HI, Rohle D, Campos C, Grommes C, Nghiemphu PL, et al. Differential Sensitivity of Glioma- versus Lung Cancer–Specific EGFR Mutations to EGFR

Kinase Inhibitors. Cancer Discov. 2012 May 9;2(5):458–71.

273. Kohsaka S, Nagano M, Ueno T, Suehara Y, Hayashi T, Shimada N, et al. A method of high-throughput functional evaluation of EGFR gene variants of unknown significance in cancer. Sci Transl Med. 2017 Nov 15;9(416):eaan6566.

274. Learn CA, Hartzell TL, Wikstrand CJ, Archer GE, Rich JN, Friedman AH, et al. Resistance to Tyrosine Kinase Inhibition by Mutant Epidermal Growth Factor Receptor Variant III Contributes to the Neoplastic Phenotype of Glioblastoma Multiforme. Clin Cancer Res. 2004 May 6;10(9):3216–24.

275. Haas-Kogan DA, Prados MD, Tihan T, Eberhard DA, Jelluma N, Arvold ND, et al. Epidermal Growth Factor Receptor, Protein Kinase B/Akt, and Glioma Response to Erlotinib. JNCI J Natl Cancer Inst. 2005 Jun 15;97(12):880–7.

276. An Z, Aksoy O, Zheng T, Fan QW, Weiss WA. Epidermal growth factor receptor and EGFRvIII in glioblastoma: signaling pathways and targeted therapies. Oncogene. 2018 Mar;37(12):1561–75.

277. Chagoya G, Kwatra SG, Nanni CW, Roberts CM, Phillips SM, Nullmeyergh S, et al. Efficacy of osimertinib against EGFRvIII+ glioblastoma. Oncotarget. 2020 Jun 2;11(22):2074–82.

278. Lee J, Kim HS, Lee B, Kim HK, Sun JM, Ahn JS, et al. Genomic landscape of acquired resistance to third-generation EGFR tyrosine kinase inhibitors in EGFR T790M-mutant non–small cell lung cancer. Cancer. 2020;126(11):2704–12.

279. Patel H, Pawara R, Ansari A, Surana S. Recent updates on third generation EGFR inhibitors and emergence of fourth generation EGFR inhibitors to combat C797S resistance. Eur J Med Chem. 2017 Dec 15;142:32–47.

280. Starrett JH, Guernet AA, Cuomo ME, Poels KE, van Alderwerelt van Rosenburgh IK, Nagelberg A, et al. Drug Sensitivity and Allele Specificity of First-Line Osimertinib Resistance EGFR Mutations. Cancer Res. 2020 May 15;80(10):2017–30.

281. Brand TM, Iida M, Wheeler DL. Molecular mechanisms of resistance to the EGFR monoclonal antibody cetuximab. Cancer Biol Ther. 2011 May 1;11(9):777–92.

282. Li C, Iida M, Dunn EF, Ghia AJ, Wheeler DL. Nuclear EGFR Contributes to Acquired Resistance to Cetuximab. Oncogene. 2009 Oct 29;28(43):3801–13.

283. Bertotti A, Papp E, Jones S, Adleff V, Anagnostou V, Lupo B, et al. The genomic landscape of response to EGFR blockade in colorectal cancer. Nature. 2015 Oct;526(7572):263–7.

284. Montagut C, Dalmases A, Bellosillo B, Crespo M, Pairet S, Iglesias M, et al. Identification of a mutation in the extracellular domain of the Epidermal Growth Factor Receptor conferring cetuximab resistance in colorectal cancer. Nat Med. 2012 Feb;18(2):221–3.

285. Price TJ, Newhall K, Peeters M, Kim TW, Li J, Cascinu S, et al. Prevalence and outcomes of patients (pts) with EGFR S492R ectodomain mutations in ASPECCT: Panitumumab (pmab) vs cetuximab (cmab) in pts with chemorefractory wild-type KRAS exon 2 metastatic colorectal cancer (mCRC). J Clin Oncol. 2015 May 20;33(15_suppl):e14623–e14623.

286. Robichaux JP, Le X, Vijayan RSK, Hicks JK, Heeke S, Elamin YY, et al. Structure-based classification predicts drug response in EGFR-mutant NSCLC. Nature. 2021 Sep;597(7878):732–7.

287. Harrison PT, Vyse S, Huang PH. Rare epidermal growth factor receptor (EGFR) mutations in non-small cell lung cancer. Semin Cancer Biol. 2020 Apr 1;61:167–79.

288. Ruan Z, Kannan N. Altered conformational landscape and dimerization dependency underpins the activation of EGFR by αC–β4 loop insertion mutations. Proc Natl Acad Sci. 2018 Aug 28;115(35):E8162–71.

289. Floc'h N, Martin MJ, Riess JW, Orme JP, Staniszewska AD, Ménard L, et al. Antitumor Activity of Osimertinib, an Irreversible Mutant-Selective EGFR Tyrosine Kinase Inhibitor, in NSCLC Harboring EGFR Exon 20 Insertions. Mol Cancer Ther. 2018 Apr 30;17(5):885–96.

290. Robichaux JP, Elamin YY, Tan Z, Carter BW, Zhang S, Liu S, et al. Mechanisms and

clinical activity of an EGFR and HER2 exon 20–selective kinase inhibitor in non–small cell lung cancer. Nat Med. 2018 May;24(5):638–46.

291. Friedlaender A, Subbiah V, Russo A, Banna GL, Malapelle U, Rolfo C, et al. EGFR and HER2 exon 20 insertions in solid tumours: from biology to treatment. Nat Rev Clin Oncol. 2022 Jan;19(1):51–69.

292. Elamin YY, Robichaux JP, Carter BW, Altan M, Tran H, Gibbons DL, et al. Poziotinib for EGFR exon 20 mutant NSCLC: clinical efficacy, resistance mechanisms and impact of insertion location on drug sensitivity. Cancer Cell. 2022 Jul 11;40(7):754-767.e6.

293. Wang J, Lam D, Yang J, Hu L. Discovery of mobocertinib, a new irreversible tyrosine kinase inhibitor indicated for the treatment of non-small-cell lung cancer harboring EGFR exon 20 insertion mutations. Med Chem Res. 2022;31(10):1647–62.

294. Wang M, Yang JCH, Mitchell PL, Fang J, Camidge DR, Nian W, et al. Sunvozertinib, a Selective EGFR Inhibitor for Previously Treated Non–Small Cell Lung Cancer with EGFR Exon 20 Insertion Mutations. Cancer Discov. 2022 Jul 6;12(7):1676–89.

295. Dhillon S. Sunvozertinib: First Approval. Drugs. 2023 Nov 1;83(17):1629–34.

296. Brindel A, Althakfi W, Barritault M, Watkin E, Maury JM, Bringuier PP, et al. Uncommon EGFR mutations in lung adenocarcinoma: features and response to tyrosine kinase inhibitors. J Thorac Dis [Internet]. 2020 Sep [cited 2024 Jan 17];12(9). Available from: https://jtd.amegroups.org/article/view/43784

297. Kobayashi Y, Togashi Y, Yatabe Y, Mizuuchi H, Jangchul P, Kondo C, et al. EGFR Exon 18 Mutations in Lung Cancer: Molecular Predictors of Augmented Sensitivity to Afatinib or Neratinib as Compared with First- or Third-Generation TKIs. Clin Cancer Res. 2015 Nov 30;21(23):5305–13.

298. Kobayashi Y, Mitsudomi T. Not all epidermal growth factor receptor mutations in lung cancer are created equal: Perspectives for individualized treatment strategy. Cancer Sci. 2016;107(9):1179–86.

299. Banno E, Togashi Y, Nakamura Y, Chiba M, Kobayashi Y, Hayashi H, et al. Sensitivities to various epidermal growth factor receptor-tyrosine kinase inhibitors of uncommon epidermal growth factor receptor mutations L861Q and S768I: What is the optimal epidermal growth factor receptor-tyrosine kinase inhibitor? Cancer Sci. 2016;107(8):1134–40.

300. Leventakos K, Kipp BR, Rumilla KM, Winters JL, Yi ES, Mansfield AS. S768I Mutation in EGFR in Patients with Lung Cancer. J Thorac Oncol Off Publ Int Assoc Study Lung Cancer. 2016 Oct;11(10):1798–801.

301. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018 Jan 4;46(D1):D1062–7.

302. Ramalingam SS, Vansteenkiste J, Planchard D, Cho BC, Gray JE, Ohe Y, et al. Overall Survival with Osimertinib in Untreated, EGFR-Mutated Advanced NSCLC. N Engl J Med. 2020 Jan 2;382(1):41–50.

303. Meyers DE, Pasternak M, Dolter S, Grosjean HAI, Lim CA, Stukalin I, et al. Impact of Performance Status on Survival Outcomes and Health Care Utilization in Patients With Advanced NSCLC Treated With Immune Checkpoint Inhibitors. JTO Clin Res Rep. 2023 Apr 1;4(4):100482.

304. Hynds RE, Frese KK, Pearce DR, Grönroos E, Dive C, Swanton C. Progress towards non-small-cell lung cancer models that represent clinical evolutionary trajectories. Open Biol. 2021 Jan 13;11(1):200247.

305. Di Fiore PP, Pierce JH, Fleming TP, Hazan R, Ullrich A, King CR, et al. Overexpression of the human EGF receptor confers an EGF-dependent transformed phenotype to NIH 3T3 cells. Cell. 1987 Dec 24;51(6):1063–70.

306. Jiang J, Greulich H, Jänne PA, Sellers WR, Meyerson M, Griffin JD. Epidermal Growth Factor–Independent Transformation of Ba/F3 Cells with Cancer-Derived Epidermal Growth Factor Receptor Mutants Induces Gefitinib-Sensitive Cell Cycle Progression. Cancer Res. 2005 Oct 3;65(19):8968–74.

307. Koga T, Suda K, Mitsudomi T. Utility of the Ba/F3 cell system for exploring on-target

mechanisms of resistance to targeted therapies for lung cancer. Cancer Sci. 2022 Mar;113(3):815–27.

308. Ercan D, Choi HG, Yun CH, Capelletti M, Xie T, Eck MJ, et al. EGFR mutations and resistance to Irreversible pyrimidine based EGFR inhibitors. Clin Cancer Res Off J Am Assoc Cancer Res. 2015 Sep 1;21(17):3913–23.

309. Chakroborty D, Kurppa KJ, Paatero I, Ojala VK, Koivu M, Tamirat MZ, et al. An unbiased in vitro screen for activating epidermal growth factor receptor mutations. J Biol Chem. 2019 Jun 14;294(24):9377–89.

310. Ikeuchi H, Hirose T, Ikegami M, Takamochi K, Suzuki K, Mano H, et al. Preclinical assessment of combination therapy of EGFR tyrosine kinase inhibitors in a highly heterogeneous tumor model. Oncogene. 2022 Apr;41(17):2470–9.

311. Yu K, Kong K, Lozzi B, Luna-Figueroa E, Cervantes A, Curry R, et al. In vivo functional characterization of EGFR variants identifies novel drivers of glioblastoma. Neuro-Oncol. 2023 Mar 1;25(3):471–81.

312. An L, Wang Y, Wu G, Wang Z, Shi Z, Liu C, et al. Defining the sensitivity landscape of EGFR variants to tyrosine kinase inhibitors. Transl Res. 2023 May 1;255:14–25.

313. Labun K, Montague TG, Krause M, Torres Cleuren YN, Tjeldnes H, Valen E. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. Nucleic Acids Res. 2019 Jul 2;47(W1):W171–4.

314. Buschmann T, Bystrykh LV. Levenshtein error-correcting barcodes for multiplexed DNA sequencing. BMC Bioinformatics. 2013 Sep 11;14(1):272.

315. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016 Oct 1;32(19):3047–8.

316. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114–20.

317. Li W, Xu H, Xiao T, Cong L, Love MI, Zhang F, et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biol. 2014 Dec 5;15(12):554.

318. Clement K, Rees H, Canver MC, Gehrke JM, Farouni R, Hsu JY, et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nat Biotechnol. 2019 Mar;37(3):224–6.

319. Helfrich BA, Raben D, Varella-Garcia M, Gustafson D, Chan DC, Bemis L, et al. Antitumor Activity of the Epidermal Growth Factor Receptor (EGFR) Tyrosine Kinase Inhibitor Gefitinib (ZD1839, Iressa) in Non–Small Cell Lung Cancer Cell Lines Correlates with Gene Copy Number and EGFR Mutations but not EGFR Protein Levels. Clin Cancer Res. 2006 Dec 4;12(23):7117–25.

320. Zheng C, Li X, Ren Y, Yin Z, Zhou B. Coexisting EGFR and TP53 Mutations in Lung Adenocarcinoma Patients Are Associated With COMP and ITGB8 Upregulation and Poor Prognosis. Front Mol Biosci [Internet]. 2020 [cited 2024 Jan 17];7. Available from: https://www.frontiersin.org/articles/10.3389/fmolb.2020.00030

321. Chen YR, Fu YN, Lin CH, Yang ST, Hu SF, Chen YT, et al. Distinctive activation patterns in constitutively active and gefitinib-sensitive EGFR mutants. Oncogene. 2006 Feb;25(8):1205–15.

322. Tagal V, Wei S, Zhang W, Brekken RA, Posner BA, Peyton M, et al. SMARCA4-inactivating mutations increase sensitivity to Aurora kinase A inhibitor VX-680 in non-small cell lung cancers. Nat Commun. 2017 Jan 19;8:14098.

323. Bessette DC, Tilch E, Seidens T, Quinn MCJ, Wiegmans AP, Shi W, et al. Using the MCF10A/MCF10CA1a Breast Cancer Progression Cell Line Model to Investigate the Effect of Active, Mutant Forms of EGFR in Breast Cancer Development and Treatment Using Gefitinib. PLoS ONE. 2015 May 13;10(5):e0125232.

324. Hoshi H, Hiyama G, Ishikawa K, Inageda K, Fujimoto J, Wakamatsu A, et al. Construction of a novel cell-based assay for the evaluation of anti-EGFR drug efficacy against EGFR mutation. Oncol Rep. 2017 Jan;37(1):66–76.

325. Soule HD, Maloney TM, Wolman SR, Peterson WD, Brenz R, McGrath CM, et al. Isolation and characterization of a spontaneously immortalized human breast epithelial

cell line, MCF-10. Cancer Res. 1990 Sep 15;50(18):6075–86.
326. Richter MF, Zhao KT, Eton E, Lapinaite A, Newby GA, Thuronyi BW, et al. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. Nat Biotechnol. 2020 Jul;38(7):883–91.
327. Koblan LW, Doman JL, Wilson C, Levy JM, Tay T, Newby GA, et al. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. Nat Biotechnol. 2018 Oct;36(9):843–6.
328. Kancha RK, von Bubnoff N, Peschel C, Duyster J. Functional Analysis of Epidermal Growth Factor Receptor (EGFR) Mutations and Potential Implications for EGFR Targeted Therapy. Clin Cancer Res. 2009 Jan 15;15(2):460–7.
329. Sonobe M, Manabe T, Wada H, Tanaka F. Mutations in the epidermal growth factor receptor gene are linked to smoking-independent, lung adenocarcinoma. Br J Cancer. 2005 Aug 8;93(3):355–63.
330. Shigematsu H, Lin L, Takahashi T, Nomura M, Suzuki M, Wistuba II, et al. Clinical and Biological Features Associated With Epidermal Growth Factor Receptor Gene Mutations in Lung Cancers. JNCI J Natl Cancer Inst. 2005 Mar 2;97(5):339–46.
331. Aertgeerts K, Skene R, Yano J, Sang BC, Zou H, Snell G, et al. Structural Analysis of the Mechanism of Inhibition and Allosteric Activation of the Kinase Domain of HER2 Protein. J Biol Chem. 2011 May 27;286(21):18756–65.
332. Mirza A, Mustafa M, Talevich E, Kannan N. Co-Conserved Features Associated with cis Regulation of ErbB Tyrosine Kinases. PLOS ONE. 2010 Dec 13;5(12):e14310.
333. Huang LC, Ross KE, Baffi TR, Drabkin H, Kochut KJ, Ruan Z, et al. Integrative annotation and knowledge discovery of kinase post-translational modifications and cancer-associated mutations through federated protein ontologies and resources. Sci Rep. 2018 Apr 25;8(1):6518.
334. Uzilov AV, Ding W, Fink MY, Antipin Y, Brohl AS, Davis C, et al. Development and clinical application of an integrative genomic approach to personalized cancer therapy. Genome Med. 2016 Jun 1;8(1):62.
335. Ferguson KM, Berger MB, Mendrola JM, Cho HS, Leahy DJ, Lemmon MA. EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization. Mol Cell. 2003 Feb;11(2):507–17.
336. Koyama N, Jinn Y, Takabe K, Yoshizawa M, Usui Y, Inase N, et al. The characterization of gefitinib sensitivity and adverse events in patients with non-small cell lung cancer. Anticancer Res. 2006;26(6B):4519–25.
337. Guha U, Chaerkady R, Marimuthu A, Patterson AS, Kashyap MK, Harsha HC, et al. Comparisons of tyrosine phosphorylated proteins in cells expressing lung cancer-specific alleles of EGFR and KRAS. Proc Natl Acad Sci U S A. 2008 Sep 16;105(37):14112–7.
338. Gow CH, Chang YL, Hsu YC, Tsai MF, Wu CT, Yu CJ, et al. Comparison of epidermal growth factor receptor mutations between primary and corresponding metastatic tumors in tyrosine kinase inhibitor-naive non-small-cell lung cancer. Ann Oncol Off J Eur Soc Med Oncol. 2009 Apr;20(4):696–702.
339. Stabile LP, Lyker JS, Gubish CT, Zhang W, Grandis JR, Siegfried JM. Combined targeting of the estrogen receptor and the epidermal growth factor receptor in non-small cell lung cancer shows enhanced antiproliferative effects. Cancer Res. 2005 Feb 15;65(4):1459–70.
340. Akimov V, Fehling-Kaschek M, Barrio-Hernandez I, Puglia M, Bunkenborg J, Nielsen MM, et al. Magnitude of Ubiquitination Determines the Fate of Epidermal Growth Factor Receptor Upon Ligand Stimulation. J Mol Biol. 2021 Oct 15;433(21):167240.
341. Ding K, Jiang X, Wang Z, Zou L, Cui J, Li X, et al. JAC4 Inhibits EGFR-Driven Lung Adenocarcinoma Growth and Metastasis through CTBP1-Mediated JWA/AMPK/NEDD4L/EGFR Axis. Int J Mol Sci. 2023 May 15;24(10):8794.
342. Li YY, Chung GTY, Lui VWY, To KF, Ma BBY, Chow C, et al. Exome and genome sequencing of nasopharynx cancer identifies NF-κB pathway activating mutations. Nat Commun. 2017 Jan 18;8(1):14121.
343. Frederick L, Wang XY, Eley G, James CD. Diversity and Frequency of Epidermal

Growth Factor Receptor Mutations in Human Glioblastomas. Cancer Res. 2000 Mar 1;60(5):1383–7.

344. Cho J, Pastorino S, Zeng Q, Xu X, Johnson W, Vandenberg S, et al. Glioblastoma-derived epidermal growth factor receptor carboxyl-terminal deletion mutants are transforming and are sensitive to EGFR-directed therapies. Cancer Res. 2011 Dec 15;71(24):7587–96.

345. Park AKJ, Francis JM, Park WY, Park JO, Cho J. Constitutive asymmetric dimerization drives oncogenic activation of epidermal growth factor receptor carboxyl-terminal deletion mutants. Oncotarget. 2015 Mar 12;6(11):8839–50.

346. Vyse S, Huang PH. Targeting EGFR exon 20 insertion mutations in non-small cell lung cancer. Signal Transduct Target Ther. 2019 Mar 8;4(1):1–10.

347. Lee JC, Vivanco I, Beroukhim R, Huang JHY, Feng WL, DeBiasi RM, et al. Epidermal Growth Factor Receptor Activation in Glioblastoma through Novel Missense Mutations in the Extracellular Domain. PLOS Med. 2006 Dec 19;3(12):e485.

348. Cho J, Kim S, Du J, Meyerson M. Autophosphorylation of the carboxyl-terminal domain is not required for oncogenic transformation by lung-cancer derived EGFR mutants. Int J Cancer. 2018 Aug 1;143(3):679–85.

349. Kovacs E, Das R, Wang Q, Collier TS, Cantor A, Huang Y, et al. Analysis of the Role of the C-Terminal Tail in the Regulation of the Epidermal Growth Factor Receptor. Mol Cell Biol. 2015 Sep 1;35(17):3083–102.

350. Bean J, Riely GJ, Balak M, Marks JL, Ladanyi M, Miller VA, et al. Acquired Resistance to EGFR Kinase Inhibitors Associated with a Novel T854A Mutation in a Patient with EGFR-Mutant Lung Adenocarcinoma. Clin Cancer Res Off J Am Assoc Cancer Res. 2008 Nov 15;14(22):7519–25.

351. Ruan Z, Katiyar S, Kannan N. Computational and Experimental Characterization of Patient Derived Mutations Reveal an Unusual Mode of Regulatory Spine Assembly and Drug Sensitivity in EGFR Kinase. Biochemistry. 2017 Jan 10;56(1):22–32.

352. Verma N, Rai AK, Kaushik V, Brünnert D, Chahar KR, Pandey J, et al. Identification of gefitinib off-targets using a structure-based systems biology approach; their validation with reverse docking and retrospective data mining. Sci Rep. 2016 Sep 22;6(1):33949.

353. Lin L, Lu Q, Cao R, Ou Q, Ma Y, Bao H, et al. Acquired rare recurrent EGFR mutations as mechanisms of resistance to Osimertinib in lung cancer and in silico structural modelling. Am J Cancer Res. 2020;10(11):4005–15.

354. Zhang L, Yang X, Ming Z, Shi J, Lv X, Li W, et al. Molecular Characteristics of the Uncommon EGFR Exon 21 T854A Mutation and Response to Osimertinib in Patients With Non-Small Cell Lung Cancer. Clin Lung Cancer. 2022 Jun;23(4):311–9.

355. Pallis AG, Voutsina A, Kalikaki A, Souglakos J, Briasoulis E, Murray S, et al. “Classical” but not “other” mutations of EGFR kinase domain are associated with clinical outcome in gefitinib-treated patients with non-small cell lung cancer. Br J Cancer. 2007 Dec 3;97(11):1560–6.

356. Todsaporn D, Mahalapbutr P, Poo-Arporn RP, Choowongkomon K, Rungrotmongkol T. Structural dynamics and kinase inhibitory activity of three generations of tyrosine kinase inhibitors against wild-type, L858R/T790M, and L858R/T790M/C797S forms of EGFR. Comput Biol Med. 2022 Aug;147:105787.

357. Xing K, Zhou X, Zhao X, Sun S, Luo Z, Wang H, et al. A novel point mutation in exon 20 of EGFR showed sensitivity to erlotinib. Med Oncol Northwood Lond Engl. 2014 Jul;31(7):36.

358. Siroy AE, Boland GM, Milton DR, Roszik J, Frankian S, Malke J, et al. Beyond BRAF(V600): clinical mutation panel testing by next-generation sequencing in advanced melanoma. J Invest Dermatol. 2015 Feb;135(2):508–15.

359. Guo H, Wang J, Ren S, Zheng LF, Zhuang YX, Li DL, et al. Targeting EGFR-dependent tumors by disrupting an ARF6-mediated sorting system. Nat Commun. 2022 Oct 12;13(1):6004.

360. Koivu MKA, Chakroborty D, Tamirat MZ, Johnson MS, Kurppa KJ, Elenius K. Identification of Predictive ERBB Mutations by Leveraging Publicly Available Cell Line

Databases. Mol Cancer Ther. 2021 Mar;20(3):564–76.

361. Sano A, Sakurai S, Kato H, Suzuki S, Yokobori T, Sakai M, et al. Expression of receptor tyrosine kinases in esophageal carcinosarcoma. Oncol Rep. 2013 Jun;29(6):2119–26.

362. Zhuo M, Guan Y, Yang X, Hong L, Wang Y, Li Z, et al. The Prognostic and Therapeutic Role of Genomic Subtyping by Sequencing Tumor or Cell-Free DNA in Pulmonary Large-Cell Neuroendocrine Carcinoma. Clin Cancer Res Off J Am Assoc Cancer Res. 2020 Feb 15;26(4):892–901.

363. Stover DR, Becker M, Liebetanz J, Lydon NB. Src phosphorylation of the epidermal growth factor receptor at novel sites mediates receptor interaction with Src and P85 alpha. J Biol Chem. 1995 Jun 30;270(26):15591–7.

364. Schulze WX, Deng L, Mann M. Phosphotyrosine interactome of the ErbB-receptor kinase family. Mol Syst Biol. 2005;1:2005.0008.

365. Chen K, Zhou F, Shen W, Jiang T, Wu X, Tong X, et al. Novel Mutations on EGFR Leu792 Potentially Correlate to Acquired Resistance to Osimertinib in Advanced NSCLC. J Thorac Oncol Off Publ Int Assoc Study Lung Cancer. 2017 Jun;12(6):e65–8.

366. Yang Z, Yang N, Ou Q, Xiang Y, Jiang T, Wu X, et al. Investigating Novel Resistance Mechanisms to Third-Generation EGFR Tyrosine Kinase Inhibitor Osimertinib in Non-Small Cell Lung Cancer Patients. Clin Cancer Res Off J Am Assoc Cancer Res. 2018 Jul 1;24(13):3097–107.

367. Coelho MA, Li S, Pane LS, Firth M, Ciotta G, Wrigley JD, et al. BE-FLARE: a fluorescent reporter of base editing activity reveals editing characteristics of APOBEC3A and APOBEC3B. BMC Biol. 2018 Dec 28;16(1):150.

368. Sánchez-Rivera FJ, Diaz BJ, Kastenhuber ER, Schmidt H, Katti A, Kennedy M, et al. Base editing sensor libraries for high-throughput engineering and functional analysis of cancer-associated single nucleotide variants. Nat Biotechnol. 2022 Jun;40(6):862–73.

369. Walton RT, Christie KA, Whittaker MN, Kleinstiver BP. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. Science. 2020 Apr 17;368(6488):290–6.

370. Kim HK, Yu G, Park J, Min S, Lee S, Yoon S, et al. Predicting the efficiency of prime editing guide RNAs in human cells. Nat Biotechnol. 2020;9(7):1–9.

371. Binder ZA, Thorne AH, Bakas S, Wileyto EP, Bilello M, Akbari H, et al. Epidermal Growth Factor Receptor Extracellular Domain Mutations in Glioblastoma Present Opportunities for Clinical Imaging and Therapeutic Development. Cancer Cell. 2018 Jul 9;34(1):163-177.e7.

372. Simon DA, Tálas A, Kulcsár PI, Biczók Z, Krausz SL, Várady G, et al. PEAR, a flexible fluorescent reporter for the identification and enrichment of successfully prime edited cells. Lapinaite A, Stainier DY, Hamilton JR, editors. eLife. 2022 Feb 23;11:e69504.

373. Schene IF, Joore IP, Baijens JHL, Stevelink R, Kok G, Shehata S, et al. Mutation-specific reporter for optimization and enrichment of prime editing. Nat Commun. 2022 Mar 1;13(1):1028.

374. Li S, Akrap N, Cerboni S, Porritt MJ, Wimberger S, Lundin A, et al. Universal toxin-based selection for precise genome engineering in human cells. Nat Commun. 2021 Jan 21;12:497.

375. Louie GV, Yang W, Bowman ME, Choe S. Crystal Structure of the Complex of Diphtheria Toxin with an Extracellular Fragment of Its Receptor. Mol Cell. 1997 Dec 1;1(1):67–78.

376. Hou Y, Gao F, Wang Q, Zhao J, Flagg T, Zhang Y, et al. Bcl2 Impedes DNA Mismatch Repair by Directly Regulating the hMSH2-hMSH6 Heterodimeric Complex. J Biol Chem. 2007 Mar;282(12):9279–87.

377. Ramirez M, Rajaram S, Steininger RJ, Osipchuk D, Roth MA, Morinishi LS, et al. Diverse drug-resistance mechanisms can emerge from drug-tolerant cancer persister cells. Nat Commun. 2016 Feb 19;7(1):10690.

378. Hata AN, Niederst MJ, Archibald HL, Gomez-Caraballo M, Siddiqui FM, Mulvey HE, et al. Tumor cells can follow distinct evolutionary paths to become resistant to epidermal

growth factor receptor inhibition. Nat Med. 2016 Mar;22(3):262–9.

379. Frattini V, Trifonov V, Chan JM, Castano A, Lia M, Abate F, et al. The integrated landscape of driver genomic alterations in glioblastoma. Nat Genet. 2013 Oct;45(10):1141–9.

380. Wang W, Lv W, Wang H, Xu Y, Yan J, Shen HM, et al. A novel acquired EGFR-SEPT14 fusion confers differential drug resistance to EGFR inhibitors in lung adenocarcinoma. Genes Dis. 2023 Nov 1;10(6):2241–4.

381. Brown BP, Zhang YK, Westover D, Yan Y, Qiao H, Huang V, et al. On-target Resistance to the Mutant-Selective EGFR Inhibitor Osimertinib Can Develop in an Allele-Specific Manner Dependent on the Original EGFR-Activating Mutation. Clin Cancer Res. 2019 Jun 3;25(11):3341–51.

382. Okabe T, Okamoto I, Tamura K, Terashima M, Yoshida T, Satoh T, et al. Differential Constitutive Activation of the Epidermal Growth Factor Receptor in Non–Small Cell Lung Cancer Cells Bearing EGFR Gene Mutation and Amplification. Cancer Res. 2007 Mar 1;67(5):2046–53.

383. Ryu J, Barkal S, Yu T, Jankowiak M, Zhou Y, Francoeur M, et al. Joint genotypic and phenotypic outcome modeling improves base editing variant effect quantification [Internet]. Genetic and Genomic Medicine; 2023 Sep [cited 2024 Jan 17]. Available from: http://medrxiv.org/lookup/doi/10.1101/2023.09.08.23295253

384. Nishimasu H, Shi X, Ishiguro S, Gao L, Hirano S, Okazaki S, et al. Engineered CRISPR-Cas9 nuclease with expanded targeting space. Science. 2018 Sep 21;361(6408):1259–62.

385. Hu C, Leche CA, Kiyatkin A, Yu Z, Stayrook SE, Ferguson KM, et al. Glioblastoma mutations alter EGFR dimer structure to prevent ligand bias. Nature. 2022 Feb;602(7897):518–22.

386. Hayat A, Carter EP, King HW, Ors A, Doe A, Teijeiro SA, et al. Low HER2 expression in normal breast epithelium enables dedifferentiation and malignant transformation via chromatin opening. Dis Model Mech. 2023 Feb 1;16(2):dmm049894.

387. Niederst MJ, Hu H, Mulvey HE, Lockerman EL, Garcia AR, Piotrowska Z, et al. The Allelic Context of the C797S Mutation Acquired upon Treatment with Third-Generation EGFR Inhibitors Impacts Sensitivity to Subsequent Treatment Strategies. Clin Cancer Res. 2015 Aug 31;21(17):3924–33.

388. Cho J. Mechanistic insights into differential requirement of receptor dimerization for oncogenic activation of mutant EGFR and its clinical perspective. BMB Rep. 2020 Mar 3;53(3):133.

389. Paquet D, Kwart D, Chen A, Sproul A, Jacob S, Teo S, et al. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. Nature. 2016 May;533(7601):125–9.

390. Wang Q, Liu J, Janssen JM, Tasca F, Mei H, Gonçalves MAFV. Broadening the reach and investigating the potential of prime editors through fully viral gene-deleted adenoviral vector delivery. Nucleic Acids Res. 2021 Nov 18;49(20):11986–2001.

391. Liu B, Dong X, Cheng H, Zheng C, Chen Z, Rodríguez TC, et al. A split prime editor with untethered reverse transcriptase and circular RNA template. Nat Biotechnol. 2022 Sep;40(9):1388–93.

392. Chardon FM, Suiter CC, Daza RM, Smith NT, Parrish P, McDiarmid T, et al. A multiplex, prime editing framework for identifying drug resistance variants at scale [Internet]. bioRxiv; 2023 [cited 2024 Jan 17]. p. 2023.07.27.550902. Available from: https://www.biorxiv.org/content/10.1101/2023.07.27.550902v1

393. Kim Y, Oh HC, Lee S, Kim HH. Saturation resistance profiling of EGFR variants against tyrosine kinase inhibitors using prime editing [Internet]. bioRxiv; 2023 [cited 2024 Jan 17]. p. 2023.12.03.569825. Available from: https://www.biorxiv.org/content/10.1101/2023.12.03.569825v1

394. Ren X, Yang H, Nierenberg JL, Sun Y, Chen J, Beaman C, et al. High-throughput PRIME-editing screens identify functional DNA variants in the human genome. Mol Cell. 2023 Dec 21;83(24):4633-4645.e9.

395. Gould SI, Wuest AN, Dong K, Johnson GA, Hsu A, Narendra VK, et al. High throughput evaluation of genetic variants with prime editing sensor libraries [Internet]. bioRxiv; 2023 [cited 2024 Jan 17]. p. 2022.10.26.513842. Available from: https://www.biorxiv.org/content/10.1101/2022.10.26.513842v4

396. Ponnienselvan K, Liu P, Nyalile T, Oikemus S, Joynt AT, Kelly K, et al. Addressing the dNTP bottleneck restricting prime editing activity [Internet]. bioRxiv; 2023 [cited 2024 Jan 17]. p. 2023.10.21.563443. Available from: https://www.biorxiv.org/content/10.1101/2023.10.21.563443v2

397. Levesque S, Verma A, Bauer DE. Nucleotide metabolism constrains prime editing in hematopoietic stem and progenitor cells [Internet]. bioRxiv; 2023 [cited 2024 Jan 17]. p. 2023.10.22.563434. Available from: https://www.biorxiv.org/content/10.1101/2023.10.22.563434v1

398. Kweon J, Yoon JK, Jang AH, Shin HR, See JE, Jang G, et al. Engineered prime editors with PAM flexibility. Mol Ther. 2021 Jun 2;29(6):2001–7.

399. Miyamoto S, Yagi H, Yotsumoto F, Kawarabayashi T, Mekada E. Heparin-binding epidermal growth factor-like growth factor as a novel targeting molecule for cancer therapy. Cancer Sci. 2006;97(5):341–7.

400. Gong L, Whirl-Carrillo M, Klein TE. PharmGKB, an Integrated Resource of Pharmacogenomic Knowledge. Curr Protoc. 2021 Aug;1(8):e226.

401. George A, Kaye S, Banerjee S. Delivering widespread BRCA testing and PARP inhibition to patients with ovarian cancer. Nat Rev Clin Oncol. 2017 May;14(5):284–96.

402. Shen X, Song S, Li C, Zhang J. Synonymous mutations in representative yeast genes are mostly strongly non-neutral. Nature. 2022 Jun;606(7915):725–31.

403. Kruglyak L, Beyer A, Bloom JS, Grossbach J, Lieberman TD, Mancuso CP, et al. No evidence that synonymous mutations in yeast genes are mostly deleterious [Internet]. bioRxiv; 2022 [cited 2024 Jan 17]. p. 2022.07.14.500130. Available from: https://www.biorxiv.org/content/10.1101/2022.07.14.500130v1

404. Dhindsa RS, Wang Q, Vitsios D, Burren OS, Hu F, DiCarlo JE, et al. A minimal role for synonymous variation in human disease. Am J Hum Genet. 2022 Dec 1;109(12):2105–9.

405. Post KL, Belmadani M, Ganguly P, Meili F, Dingwall R, McDiarmid TA, et al. Multi-model functionalization of disease-associated PTEN missense mutations identifies multiple molecular mechanisms underlying protein dysfunction. Nat Commun. 2020 Apr 29;11(1):2073.

406. Spinelli L, Black FM, Berg JN, Eickholt BJ, Leslie NR. Functionally distinct groups of inherited PTEN mutations in autism and tumour syndromes. J Med Genet. 2015 Feb;52(2):128–34.

407. Jun S, Lim H, Chun H, Lee JH, Bang D. Single-cell analysis of a mutant library generated using CRISPR-guided deaminase in human melanoma cells. Commun Biol. 2020 Apr 2;3:154.

408. Ursu O, Neal JT, Shea E, Thakore PI, Jerby-Arnon L, Nguyen L, et al. Massively parallel phenotyping of coding variants in cancer with Perturb-seq. Nat Biotechnol. 2022 Jun;40(6):896–905.

409. Nam AS, Kim KT, Chaligne R, Izzo F, Ang C, Taylor J, et al. Somatic mutations and cell identity linked by Genotyping of Transcriptomes. Nature. 2019 Jul;571(7765):355–60.

410. Kim HS, Grimes SM, Chen T, Sathe A, Lau BT, Hwang GH, et al. Direct measurement of engineered cancer mutations and their transcriptional phenotypes in single cells. Nat Biotechnol. 2023 Sep 11;1–9.

411. Battle A, Khan Z, Wang SH, Mitrano A, Ford MJ, Pritchard JK, et al. Impact of regulatory variation from RNA to protein. Science. 2015 Feb 6;347(6222):664–7.

412. Strecker J, Ladha A, Gardner Z, Schmid-Burgk JL, Makarova KS, Koonin EV, et al. RNA-guided DNA insertion with CRISPR-associated transposases. Science. 2019 Jul 5;365(6448):48–53.

413. Klompe SE, Vo PLH, Halpin-Healy TS, Sternberg SH. Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. Nature. 2019

Jul;571(7764):219–25.

414. Lampe GD, King RT, Halpin-Healy TS, Klompe SE, Hogan MI, Vo PLH, et al. Targeted DNA integration in human cells without double-strand breaks using CRISPR-associated transposases. Nat Biotechnol. 2024 Jan;42(1):87–98.

415. Clark TA, Chung JH, Kennedy M, Hughes JD, Chennagiri N, Lieber DS, et al. Analytical Validation of a Hybrid Capture-Based Next-Generation Sequencing Clinical Assay for Genomic Profiling of Cell-Free Circulating Tumor DNA. J Mol Diagn JMD. 2018 Sep;20(5):686–702.

416. Cheng J, Novati G, Pan J, Bycroft C, Žemgulytė A, Applebaum T, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science. 2023 Sep 19;381(6664):eadg7492.