AWARD NUMBER: W81XWH-12-1-0269

TITLE: Functional Analysis of Somatic Mutations in Lung Cancer

PRINCIPAL INVESTIGATOR: Matthew Meyerson, MD, PhD

CONTRACTING ORGANIZATION: Dana-Farber Cancer Institute
Boston, MA 02215-5450

REPORT DATE: October 2015

TYPE OF REPORT: Final Report

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
October, 2015	Final Report	01 August, 2012 – 31 July, 2015

4. TITLE AND SUBTITLE

Functional Analysis of Somatic Mutations in Lung Cancer

5a. CONTRACT NUMBER
W81XWH-12-1-0269

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Matthew Meyerson, Aruna Ramachandran, Peter Choi, Alice Berger, Hugh Gannon, Heidi Greulich, Xiaoyang Zhang

E-Mail: matthew_meyerson@dfci.harvard.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Dana-Farber Cancer Institute
Boston, MA 02215-5450

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

We have conducted functional investigations of somatic mutations in lung cancer, focusing on genes/allelic variants individually and also using high throughput combinatorial approaches. We have gained insights into the impact of somatic mutations in *RBM10*, *MAP2K1*, *ERBB2*, *EGFR* and a host of other oncogenes or tumor suppressor proteins, on the initiation and maintenance of lung cancer. Notably, we have revealed for the first time the mechanism by which amplifications in an enhancer region (a novel super-enhancer), upstream of the *MYC* promoter, drives carcinogenesis. During the course of this project, we have also developed novel *in vitro* and *in vivo* experimental and analytical approaches, including genome editing methods, to augment our own research capabilities and that of the broader scientific community. The significance of our findings lies in the fact that we have now gained valuable insights into the molecular mechanisms by which somatic mutations in lung cancer—that were previously detected by large-scale, genomics approaches and were of unknown significance—incite tumorigenesis. The next step would be to translate this knowledge to devising more effective and tumor-specific targeted therapies, to benefit lung cancer patients.

15. SUBJECT TERMS
somatic mutation, lung cancer, therapeutic target, RBM10, MYC, MAP2K1, EGFR, ERBB2, L1000, PC9, signaling pathway, therapeutic resistance, oncogene

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	36	19b. TELEPHONE NUMBER *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

1. INTRODUCTION

The subject of this research proposal is to understand the functional basis of how somatic mutations in genes observed in human lung cancers contribute to carcinogenesis. Thanks to the genomics revolution—spurred by the rapid advances in and decreasing costs of next-generation DNA sequencing, advances in analytical methods—studies from laboratories around the world, such as ours, have been able to shed light on genetic aberrations that cause cancer. Our understanding of the mechanisms by which these genome alterations cause cancer, however, has not advanced at a similar pace. Thus, we still have little knowledge of the molecular mechanisms by which these individual genome alterations cause cancer. The purpose of this project thus is to illuminate the function of somatic gene mutations in lung cancer. The scope of the research includes analysis of hundreds of mutated alleles of lung cancer-relevant genes.

2. KEYWORDS

Lung cancer
Somatic mutation
Oncogene
Acquired resistance
Splicing
Signaling
Kinase
Gene expression
RBM10
MYC
MGA
ERBB2
EGFR
NRF2
MAP2K1
L1000

3. ACCOMPLISHMENTS

3A: The hypothesis, aims and major goals of the project were:

Hypothesis or Objective

We hypothesize that lung cancer somatic mutation data include novel oncogenic alleles that may also be therapeutic targets. Our objective is to identify these oncogenic alleles through systematic functional assays and determine which are suitable for therapeutic intervention.

Specific Aims

Specific Aim 1. Determine if prioritized somatic alleles reported in lung adenocarcinoma and lung squamous cell carcinoma whole-exome sequencing experiments are oncogenic.

Specific Aim 2. Measure signaling pathway perturbation in response to ectopic expression of candidate oncogenes.

Specific Aim 3. Test sensitivity of engineered cells and lung cancer cell lines expressing novel oncogenic alleles to candidate small molecules and shRNAs.

Statement of Work

Task 1: Create cell line models of ectopic expression with selected lung cancer somatic mutations

Task 2: Measure oncogenic transformation of engineered cells

Task 3: Perform RNA sequencing on tumor cell lines harboring wild-type or mutant *RBM10*

Task 4: Assess mRNA levels of 1000 landmark genes in engineered BEAS-2B cells

Task 5: Assess inhibitor efficacy in Ba/F3 cells

Task 6: Assess inhibitor efficacy in lung cancer cell lines

3B: Accomplishments under these goals:

We have met or made substantial progress toward the goals outlined in this proposal. We have conducted large-scale functional investigations of mutated alleles observed in lung cancers, to reveal novel activities or validate predicted biological outcomes associated with these mutations. In addition we have also deeply interrogated the function of many critical genes with somatic mutations in lung cancer, with the potential for long-term clinical impact. Each line of investigation incorporates elements spanning multiple tasks/goals as described in our SOW; thus for clarity the structure of this report will focus on the independent projects themselves while making reference to the appropriate task(s) they fall under.

3B (i): *RBM10* (relevant to tasks 1, 2, 3)

We recently reported recurrent somatic mutations in *RBM10 (*RNA binding motif 10), an X-chromosomal gene linked to TARP syndrome in males [1, 2], in about 7% of lung adenocarcinoma cases [3, 4] (**Figure 1**). Interestingly, mutations in *RBM10* are more prevalent in tumors from males, in contrast to *EGFR* mutations that occur more frequently in females (**Figure 1**, right panel), [3]. RBM10 is a regulator of RNA splicing [5-7]; its role in cancer, however, is a mystery. Our group is the first to report its being genetically altered in any tumor type [3, 4] at a frequency comparable to ALK translocations (~4%) in lung adenocarcinoma [8]. Together with our discovery of somatic alterations in another splicing factor (*U2AF1*) in 3-4% of lung adenocarcinomas, this reveals a new and hitherto unsuspected role for splicing factors in this cancer, with potential therapeutic significance.

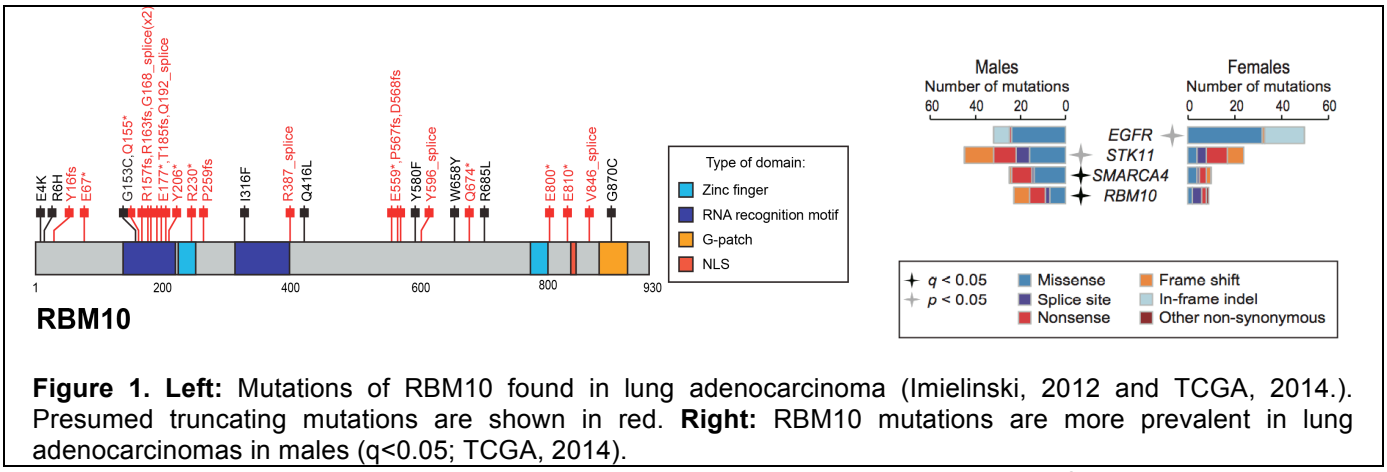
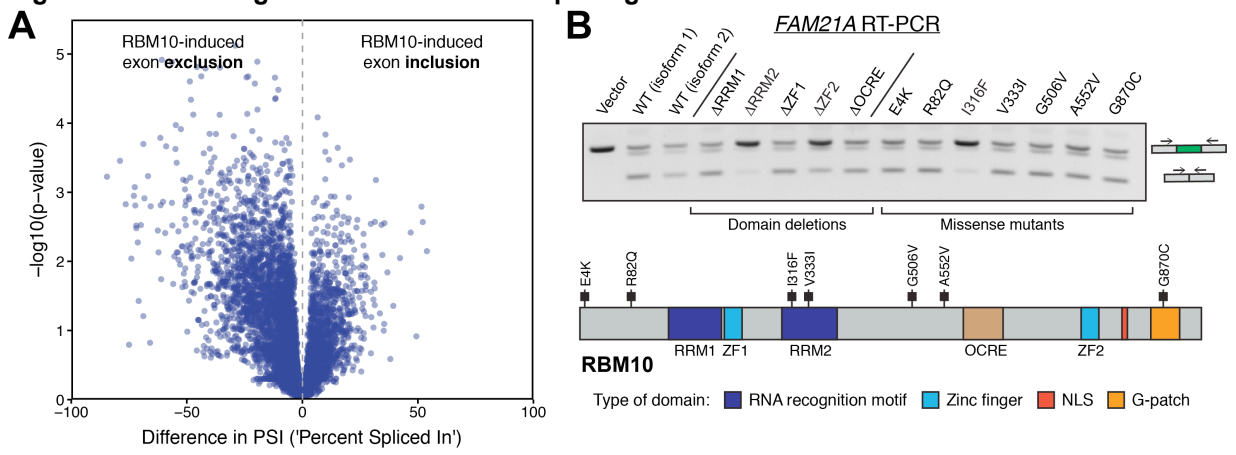


Figure 1. Left: Mutations of RBM10 found in lung adenocarcinoma (Imielinski, 2012 and TCGA, 2014.). Presumed truncating mutations are shown in red. **Right:** RBM10 mutations are more prevalent in lung adenocarcinomas in males (q<0.05; TCGA, 2014).

Figure 2: RBM10-regulated cassette exon splicing

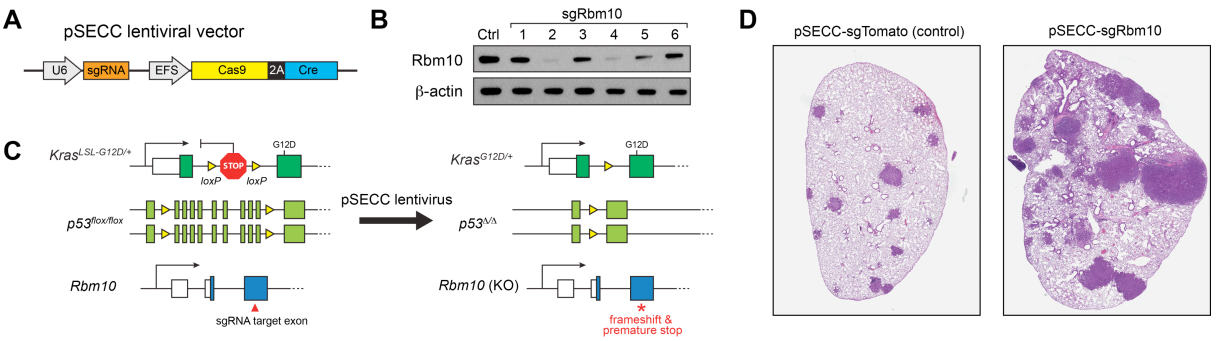


(A) Splicing differences in cassette exons between *RBM10*-mutant cells expressing wild-type RBM10 versus a control protein (BFP). Negative PSI values represent cassette exons that are more excluded in the presence of RBM10 and positive PSI values represent cassette exons that are more included in the presence of RBM10. **(B)** Splicing assay to identify domains and amino acid residues critical for RBM10-mediated regulation of RNA splicing. HEK293T cells deficient in *RBM10* were transfected with wild-type RBM10 or various RBM10 mutants and the effects on FAM21A splicing were assessed by RT-PCR.

Using lung adenocarcinoma cell lines that harbor loss-of-function mutations in *RBM10*, we have been investigating changes in RNA expression and splicing regulated by RBM10. Cells are first stably infected with tetracycline-regulated RBM10 or control BFP expression constructs. Upon treatment with the tetracycline-analog doxycycline and induction of wild-type RBM10 expression, we have observed a significant inhibition of proliferation in multiple *RBM10*-mutant cell lines (described in Annual Report, 2014). To characterize the mechanism responsible for RBM10-mediated growth suppression, we prepared RNA-sequencing libraries and sequenced biological triplicates of both control BFP and RBM10-expressing samples on the Hiseq 2000 platform. From the RNA-sequencing results, we identified differentially expressed genes using the Cufflinks and DESeq2 software packages, and differentially spliced exons using JuncBASE software. In agreement with previous reports, we found that RBM10 most frequently induces cassette exon exclusion (**Figure 2**). We are currently validating the most significantly differentially spliced exons by RT-PCR and downstream functional assays. In addition, we developed a splicing assay to better characterize how RBM10 functions as a regulator of RNA splicing. We generated several RBM10 expression constructs with deletions in critical functional domains, such as the RNA-recognition motifs (RRM1 and 2), and substitutions in amino acids found to be mutated in human lung adenocarcinoma. These were transfected into *RBM10*-knockout HEK293T cells and the effects on splicing of the RBM10 target transcript *FAM21A* were assessed by RT-PCR (**Figure 2B**). Our analysis revealed that the second RNA recognition motif region (RRM2) is particularly important, as deletion or mutation of this motif significantly impaired the ability of RBM10 to induce *FAM21A* exon exclusion. We are continuing to apply this assay to characterize additional *RBM10* mutations identified in both lung cancer and other cancer types.

In human lung adenocarcinoma, *RBM10* mutations consist mainly of loss-of-function alterations, such as frameshift and nonsense mutations, strongly suggesting that RBM10 functions as a tumor suppressor. To

Figure 3: Rbm10 loss-of-function increases lung tumor burden



(A) All-in-one lentiviral vector for CRISPR-targeting and Cre-mediated recombination (Jacks lab). **(B)** NIH3T3 cells were infected with lentiviruses expressing Cas9 and either a control sgRNA (Ctrl) or Rbm10-specific sgRNAs (sgRbm10 1-6). **(C)** Depiction of *Kras*$^{LSL\text{-}G12D/+}$, *p53*$^{flox/flox}$ and *Rbm10* alleles before and after delivery of pSECC lentivirus. Cre-mediated recombination results in expression of mutant *Kras*^{G12D} and Cas9-mediated targeting of *Rbm10* results in a frameshift mutation leading to premature termination. **(D)** H&E staining of lung sections from *Kras*$^{LSL\text{-}G12D/+}$/*p53*$^{flox/flox}$ mice, approximately 20 weeks following treatment with either control sgTomato or sgRbm10 lentivirus.

investigate the consequences of *RBM10* mutations in vivo, we have taken advantage of a technique developed by the lab of Dr. Tyler Jacks (Koch Institute) [9]. By delivering a single vector (pSECC) that expresses all components of the CRISPR/Cas9 system to the adult mouse lung, we are able to target any gene of interest and determine the functional effects without necessitating the creation of new germline-targeted mouse models (**Figure 3A**). By co-expressing Cre recombinase, Cas9-targeting can also be studied in the context of well-established cancer models, such as the *Kras*$^{G12D/+}$ and *p53*$^{flox/flox}$ models (**Figure 3C**). We designed several single-guide RNAs (sgRNAs) specific for mouse *Rbm10* and identified the sgRNAs that were most effective at

reducing Rbm10 expression (**Figure 3B**). These were cloned into the pSECC vector and lentivirus was delivered into the lungs of $Kras^{G12D/+}/p53^{flox/flox}$ mice. In this genetic background, delivery of Cre via pSECC lentivirus results in expression of mutant $Kras^{G12D}$ and loss of *p53* expression. Mice were sacrificed approximately 20 weeks following treatment with pSECC lentivirus and lungs were harvested for histopathological analysis. Expression of an *Rbm10*-specific sgRNA (sgRbm10) resulted in greater tumor burden compared to a control sgRNA (sgTomato) (**Figure 3D**). We are currently carrying out RNA-sequencing analysis to characterize differences in gene expression and splicing patterns between sgTomato and sgRbm10 targeted lung tumors. To identify functionally relevant targets downstream of Rbm10, we will then compare splicing events regulated by Rbm10 across both our human and mouse lung tumor data.

3B (ii): *EGFR* (relevant to tasks 1, 2, 5, 6)

We have made expression constructs expressing multiple EGFR-mutant protein isoforms discovered in human lung cancers, including EGFR exon20 insertion mutants. We verified that expression of these variants induced oncogenic transformation in the Ba/F3 model cell line, and evaluated the effect of 2nd and 3rd generation small molecule EGFR inhibitors on the growth of Ba/F3 cells expressing these mutants. We also conducted a comprehensive gain-of-function screen in the *EGFR*-mutant PC9 lung cancer cell line, to define signaling mechanisms that mediate resistance to EGFR-targeted therapies [10]. Additionally, we identified an EGFR G724S kinase domain mutation from a patient colorectal adenocarcinoma and revealed that this mutation confers susceptibility to the anti-EGFR antibody cetuximab [11]. Finally, we have developed novel single cell sequencing approaches to uncover EGFR mutational variants in glioblastoma and their role in development of tumor heterogeneity [12].

Effect of small molecules on growth of Ba/F3 cells transformed by expression of mutant EGFR variants.

We constructed expression vectors expressing *EGFR* mutants found in human lung cancer. Constructs included EGFR WT (wild type), L858R, T790M, L858R+T790M and various exon20 insertion mutants. We confirmed that expression of these mutants promoted oncogenic transformation of the IL3-dependent Ba/F3 cell line, allowing for IL3-independent growth. We next evaluated the effectiveness of various EGFR inhibitors in reducing growth of these transformed cells. As shown in **Figure 4**, the 2nd generation inhibitors afatinib and neratinib, originally developed against the EGFR T790M mutant, show

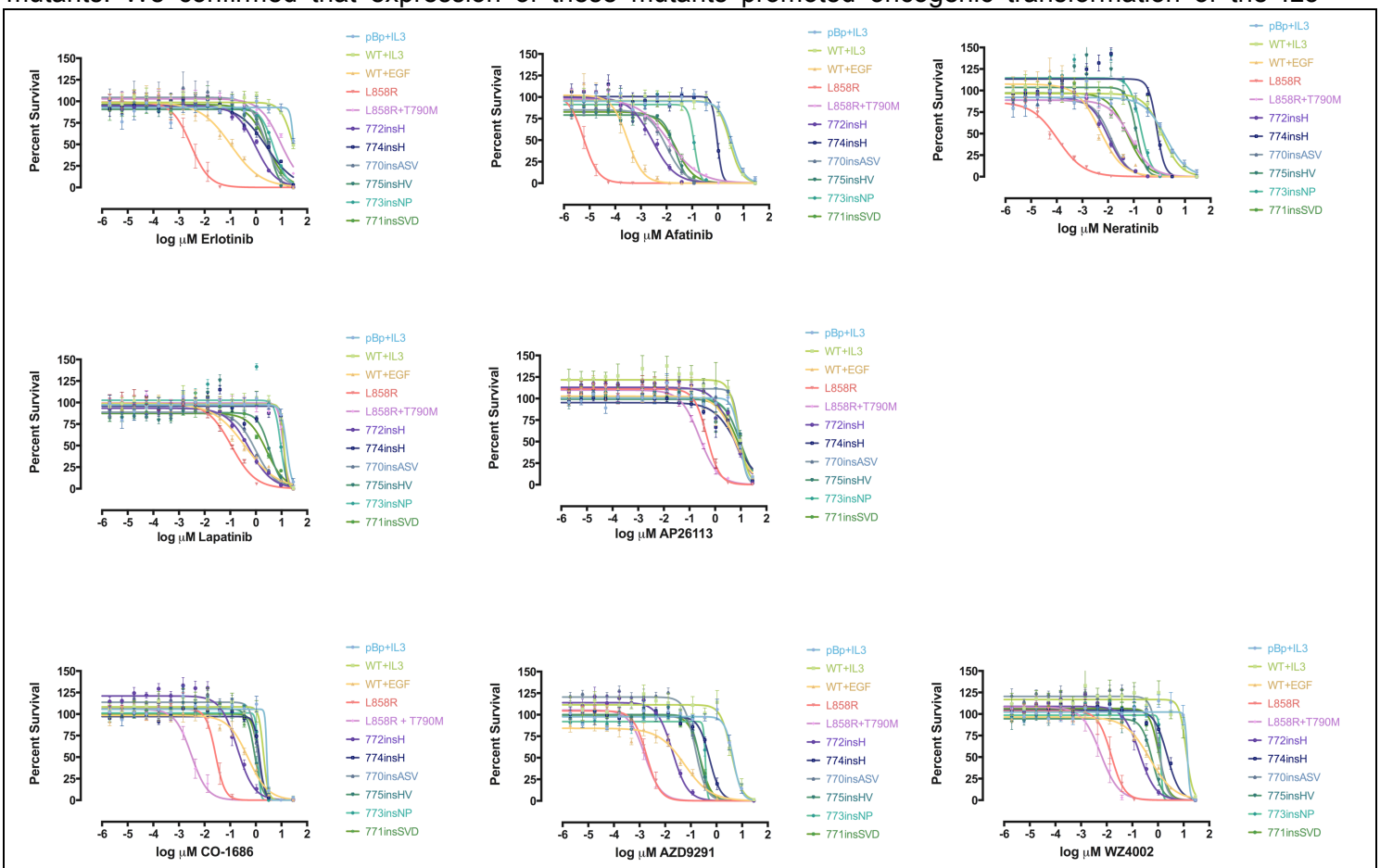


Figure. 4. Effect of EGFR inhibitors on growth of Ba/F3 cells transformed with EGFR exon 20 insertion mutants. Afatinib and neratinib (top panel) are effective against some EGFR exon20 insertion mutants, compared to erlotinib, but also show significant toxicity aginst wild type EGFR (yellow lines). Lapatinib and the AP compound are ineffective. The 3rd generation molecules (CO, AZ and WZ) are somewhat effective against some exon 20 mutants but also target wild type EGFR better, rendering them unsuitable for therapeutic use.

efficacy against these mutants. Unfortunately, they are more potent against wild type EGFR-expressing cells, therefore rendering them inappropriate for patient treatment due to on-target toxicity. Lapatinib and AP26113 were ineffective against the EGFR exon20 insertion mutants. The 3rd generation inhibitor CO-1686 and AZD9291 reduced growth of T790M- and EGFR L858R-expressing cells but, as with afatinib and neratinib, also inhibited wild type EGFR-expressing cells more effectively than cells expressing EGFR exon 20 insertion mutations. Our results underscore the recalcitrance of EGFR exon20 mutants to devise specific and effective treatments for this class of EGFR mutants.

Screening for kinases conferring resistance to EGFR inhibitors:

We conducted a gain-of-function screen of all human kinases to identify candidates that would confer resistance to erlotinib-sensitive cells. The non-small cell lung cancer line PC9 was used here as it is mutated for *EGFR* and consequently sensitive to the EGFR inhibitor erlotinib. PC9 cells were infected with a lentivirus library expressing the entire collection of human kinase ORFs. 18 ORFs scored as positive hits i.e. induced erlotinib resistance, and were experimentally validated. The Src signaling pathway emerged as a novel pathway that induced resistance to erlotinib. Inhibition of this pathway by dual treatment with dasatinib, a Src kinase family inhibitor, along with erlotinib, reversed resistance i.e. strongly inhibited growth of PC9 cells (**Figure 5**). From our analyses, the MEK/ERK and PI3K/AKT axes emerge as dominant signaling mechanisms that promote resistance to EGFR-targeted therapy [10]. Thus, we are now in a position to test drugs targeting these and other signaling pathways that converge on oncogenic EGFR-signaling, for potential therapeutic application.

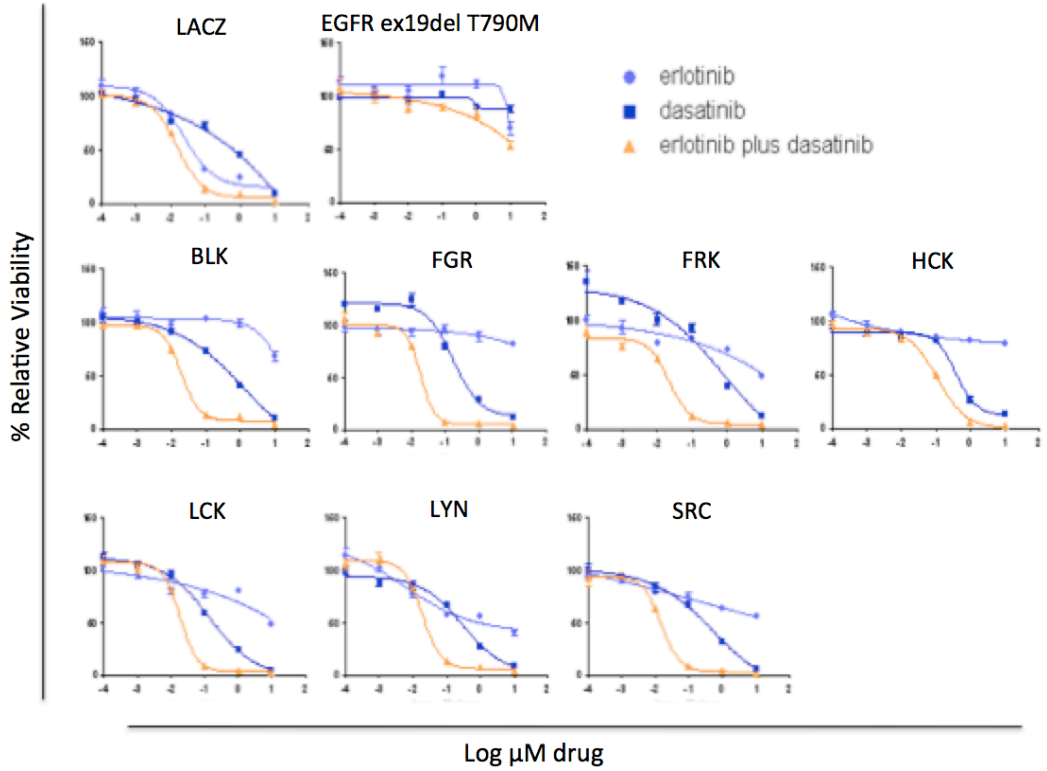


Figure 5: PC9 cells infected with resistance-inducing Src family ORFs show cell death only upon treatment with both erlotinib and dasatinib, and not with the individual drugs.

3B (iii): *ERBB2* (relevant to tasks 1, 2, 5, 6)

We recently identified novel oncogenic extracellular mutations of *ERBB2* in lung adenocarcinoma that are drug-sensitive in isogenic engineered cells [13]; however, mutation of lung cancer gene *ERBB2* appears not to be as tightly correlated with inhibitor response in actual tumor cells as is the case with *EGFR* and erlotinib [13-15]. In an effort to better understand the relationship between *ERBB2* mutation and inhibitor response, we are studying the set of *ERBB2* mutations found in cancer cell lines, defining which mutations are oncogenic and which cell lines respond to ERBB2 inhibitors. All of these cancer cell lines have been genomically characterized by the Broad-Novartis Cancer Cell Line Encyclopedia [16], facilitating identification of genomic features that are associated with inhibitor response or resistance. Because in most cases it is not possible to determine somatic vs. germline status for these mutations from the cell line sequencing alone, we will rely more heavily on the set of *ERBB2* mutations found independently to be somatic in one or more patient samples, as reported in the COSMIC database.

The majority of ERBB2 mutations found in cancer cell lines have proven to be oncogenic in NIH-3T3 cells (sample data: **Figure 6**, left panel). We have completed testing of these mutants in batches, and will re-test all 18 cancer cell line *ERBB2* mutations in a single experiment in triplicate with known germline SNPs as negative controls to accurately distinguish weak transformation from a lack of phenotype.

Testing of sensitivity of cancer cell lines harboring mutations of *ERBB2* to ERBB2 inhibitors is also ongoing (sample data: **Figure 6**, right panel). In the experiment shown in Figure 12, we used HCC827 (*EGFR* exon 19 del) to define maximum sensitivity to afatinib, NCI-H1781 (*ERBB2* G776>VC) to define typical response of an *ERBB2* mutant cancer cell line known to respond to afatinib, and AN3CA (*FGFR2* N549K) as our negative control, resistant to afatinib. Test cell lines NCI-H2110 (*ERBB2* N319Y) and VMCUB1 (*ERBB2* S653C) respond to afatinib similarly to sensitive cell line NCI-H1781. Test cell lines HCC461 (*ERBB2* A386D) and NCI-H1563 (*ERBB2* S335C; not shown) are resistant to afatinib, similar to negative control cell line AN3CA. As we expand the number of cell lines tested, we will also test the FDA-approved inhibitors, trastuzumab and lapatinib.

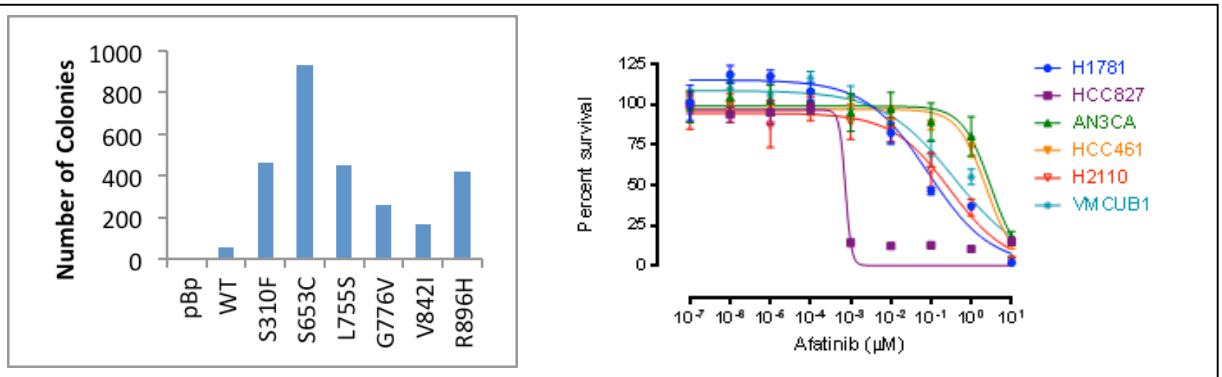


Figure 6. *ERBB2* mutant alleles are oncogenic, but cell lines harboring these mutations display differential sensitivity to afatinib. Left panel, example of soft agar assay with NIH-3T3 cells ectopically expressing the indicated *ERBB2* mutations found in cancer cell lines. All alleles tested in this experiment are oncogenic. Right panel, example of CellTiter-Glo survival assay following 72 hr treatment with the indicated concentrations of the dual EGFR-ERBB2 inhibitor, afatinib.

We are studying the relationship between *ERBB2* mutation and response to targeted therapies. Unlike the case with *EGFR* and erlotinib [17, 18], *ERBB2* mutation does not consistently sensitize cancer cells to ERBB2 inhibitors, although these inhibitors are biochemically effective, as evidenced by activity in engineered cell line models [13]. Furthermore, lung adenocarcinoma patients whose tumors harbor mutations of *ERBB2* do not dramatically respond to ERBB2 inhibition [14, 15, 19]. In at least one ERBB2 mutant cancer cell line, a combination of ERBB2 and Mek inhibition shows synergistic effects on viability [13]. We therefore hypothesized that additional pathways are activated in the *ERBB2* mutant cells exhibiting primary resistance to ERBB2 inhibition.

With the goal of identifying genomic features that segregate with resistance to the ERBB2 inhibitor, afatinib, we selected 20 ERBB2 mutant cancer cell lines genomically characterized by the Broad-Novartis Cancer Cell Line Encyclopedia [16]. Some of these ERBB2 variants are likely to be somatic and oncogenic, some may be somatic passenger mutations, and, because no matched normal samples are available for these cell lines, we must assume that some ERBB2 variants will be SNPs.

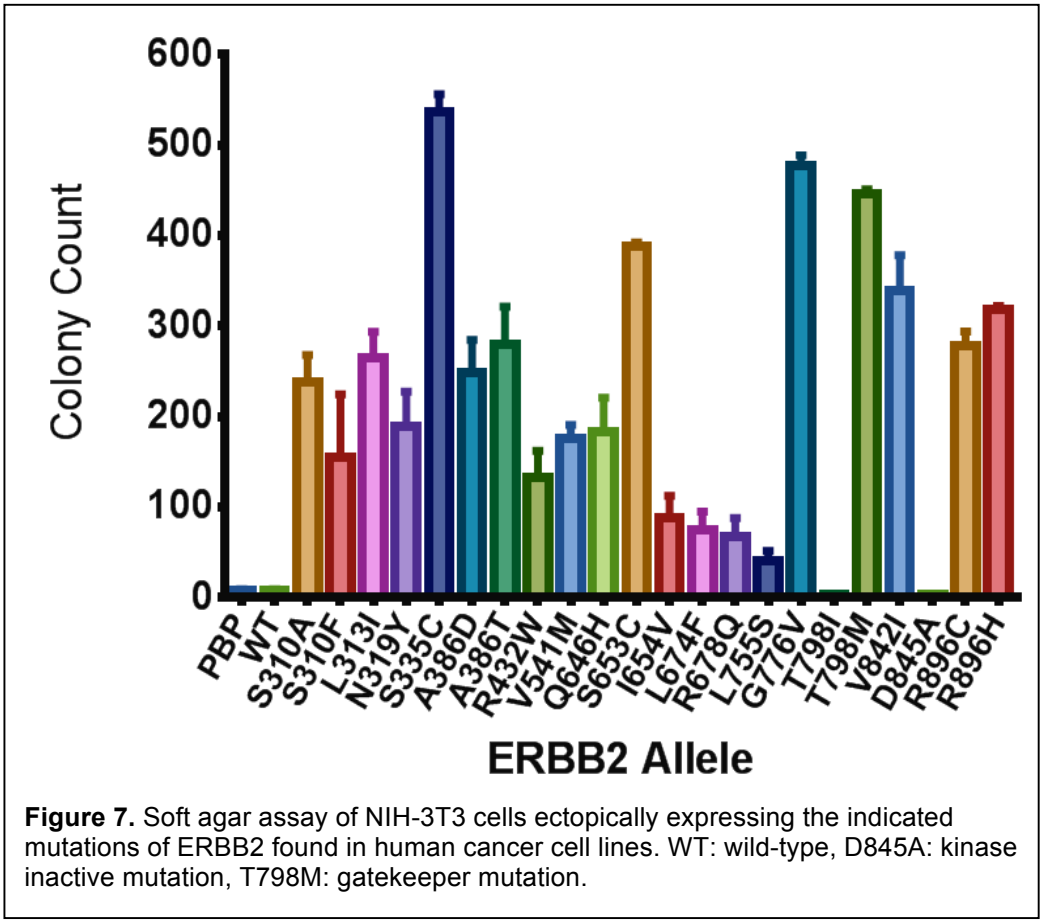


Figure 7. Soft agar assay of NIH-3T3 cells ectopically expressing the indicated mutations of ERBB2 found in human cancer cell lines. WT: wild-type, D845A: kinase inactive mutation, T798M: gatekeeper mutation.

In order to focus our analysis on cell lines harboring only oncogenic variants of ERBB2, we first assessed the observed alleles of ERBB2 for oncogenic potential. We introduced the mutations and appropriate controls into the wild-type ERBB2 cDNA, ectopically expressed the mutant cDNAs in NIH-3T3 cells, and measured anchorage independent proliferation with soft agar assays. With the surprising exception of T798I [20], located at the Gatekeeper residue analogous to Bcr-Abl T315I [21] and EGFR T790M [22], all mutations tested supported soft agar colony formation (**Figure 7**). The high success rate was likely due at least in part to prioritization of alleles also found to be somatic in independent patient samples.

Although we initially measured the response of the *ERBB2* mutant cancer cell lines to ERBB2 inhibition ourselves, we were recently able to leverage data generated at the Broad Institute by an NIH-funded Cancer Target Discovery and Development (CTDD) project, in which cancer cell lines, including those harboring the mutations assessed in **Figure 7**, were tested for viability upon treatment with a series of small molecules, including several ERBB2 inhibitors. We focused our analyses on survival data generated following treatment with afatinib, one of the most potent small molecule inhibitors of mutant ERBB2 that is also FDA approved. Using primary data generated by CTDD, we calculated IC50s of afatinib treatment, shown in **Table 1**, for the 20 *ERBB2* mutant cancer cell lines plus one control cell line harboring wild-type amplified *ERBB2*.

In order to identify genomic features in these cancer cell lines that segregate with increased resistance to afatinib, we used an algorithm called REVEALER developed by our colleague at the Broad Institute, Dr. Pablo Tamayo, which uses mutual information to establish nonlinear correlations between an input vector and known genomic features (manuscript submitted). We examined gene expression, pathway expression, reverse-phase protein array, shRNA sensitivity, and compound sensitivity data for features that correlate with increased resistance of ERBB2 mutant cancer cell lines to afatinib (data not shown). Interestingly, although the input IC50s comprise a continuous variable, the REVEALER output indicate a sharp demarcation from sensitivity to resistance between the eighth (SW403) and ninth (TE441T) *ERBB2* mutant cell lines, as ranked by afatinib response, suggesting a fundamental biological difference between the slightly sensitive and very

resistant cells. Combining results from analyses of the five data types listed above (other data types, such as gene copy number and metabolite expression were deemed uninformative for our analysis), we found five pathways to be upregulated (or causing dependencies) in the more resistant ERBB2 mutant cell lines: Mek, PI3K/MTOR, Notch, Rho, and anti-apoptotic proteins.

Lineage	ERBB2 mutation	Afatinib IC50 (mM)
BT474_BREAST	wt ERBB2	<0.1
VMCUB1_URINARY_TRACT	S653C	0.1402
DV90_LUNG	I654V	0.7446
NCIH1781_LUNG	776G>VC	<1
CW2_LARGE_INTESTINE	L755S	1.336
SNUC2A_LARGE_INTESTINE	R678Q	1.343
NCIH2110_LUNG	N319Y	2.786
SW403_LARGE_INTESTINE	I654V	3.139
TE441T_SOFT_TISSUE	R432W	3.525
LN229_CENTRAL_NERVOUS_SYSTEM	L755S	3.764
G361_SKIN	I654V	4.518
SNU1040_LARGE_INTESTINE	V777M	4.649
NCIH1563_LUNG	S335C	5.206
OC314_OVARY	A386T	5.381
GCT_SOFT_TISSUE	I654V	5.479
MFE319_ENDOMETRIUM	R896H	7.051
OVCAR8_OVARY	G776V	7.504
KPL1_BREAST	I654V	8.761
NCIH1793_LUNG	V541M	9.429
J82_URINARY_TRACT	R678Q	9.962
EFO27_OVARY	V842I	15.52

Table 1: IC50s of afatinib treatment of cancer cell lines harboring mutations in *ERBB2*

3B (iv): *Assess mRNA levels of L1000 genes in response to expression of somatically mutated alleles* (relevant to task 4)

We have developed molecular and genetic assays to systematically phenotype 194 somatic mutations identified in primary lung adenocarcinomas, by assessing the impact of expression of these alleles on gene expression of L1000 genes, a group of a thousand carefully validated "cancer landmark genes" that captures the overall transcriptional output of the tumor cell [23], http://www.lincscloud.org/l1000/. The L1000 signature is thus a surrogate for broader effects on the transcriptome, allowing us to infer biological outcomes associated with global gene expression changes caused by specific perturbations. We developed a new method, expression-based variant impact phenotyping (eVIP), which analyzes mutant-induced expression changes to distinguish impactful from neutral somatic mutations. As a complementary approach, we assessed which mutations are epistatic to EGFR or capable of initiating xenograft tumor formation *in vivo*. Using eVIP, we identified 69% of mutations analyzed as impactful whereas 31% appear functionally neutral. A subset of the impactful mutations induce xenograft tumor formation in mice and/or confer resistance to cellular EGFR inhibition. Among these are 24 rare or non-canonical somatic variants in clinically-actionable or –relevant oncogenes including *EGFR* S645C, *ARAF* S214C and S214F, *ERBB2* S418T, and *BRAF* P367R and L485S. We further discovered that 92% of tested missense mutations in the *KEAP1* and *STK11* tumor suppressor genes inactivate protein function.

We have generated an expression library of 354 constructs, including 194 mutant alleles identified in primary lung adenocarcinomas and have profiled each allele in gene expression, EGFR epistasis and in vivo tumor formation assays (**Figure 8**), using three complementary experimental approaches:
1) eVIP, a novel gene-agnostic method to determine which mutations impact gene function,
2) a genetic epistasis screen to determine which mutations confer resistance to cellular EGFR inhibition, and
3) an *in vivo* pooled xenograft assay to determine which mutations promote tumor formation.

1) Expression-based variant impact phenotyping (eVIP)

We introduced the mutated ORF library via lentiviral transduction and measured transcriptional changes 96 hours later using L1000 gene expression profiling [23]. This method allows for the assessment of 978 "landmark" transcripts in a high-throughput manner. After normalization to a set of invariant genes, we generated differential gene expression profiles ("signatures") for further analysis. To verify that ORF constructs were adequately expressed, the experiment included 80 cDNAs known to be detected by L1000 probes. 79 of these 80 expression controls were indeed the top differentially expressed gene in the respective gene signature, confirming that most ORFs are robustly expressed as expected.

By comparatively analyzing mutant and wild-type-induced gene expression signatures, we discovered that both the signal strength of the signature and the signature identity (which transcripts are altered) contain information which can differentiate two signatures. Weak induction of gene expression changes results in a signature more similar to noise; therefore, we used replicate correlation as a quantitative proxy for signal strength. In some cases, mutated ORFs such as *CTNNB1* S33N showed an increase in signal strength relative to the cognate wild-type ORF (**Figure 9A**), and we hypothesized that this directionality (mutant>WT) was indicative of a dominant mutation (DOM). By contrast, other mutated ORFs, such as *STK11* D194Y, had decreased signal strength compared to wild-type (**Figure 9B**), suggesting likely loss-of-function impact on gene function (LOF).

In other instances, the signal strength of the wild-type and mutant signatures were not significantly different. For example, both *ARAF* S214F and V145L induced transcriptional changes with similar signal strengths as wild-type *ARAF* (**Figure 9C-D**). However, when we compared the identity of the signatures (correlation between the top up- and down-regulated genes in each signature), *ARAF* S214F induced substantially different gene expression changes than wild-type *ARAF* (**Figure 9E**), whereas gene expression induced by *ARAF* V145L was indistinguishable from that induced by wild-type *ARAF* (**Figure 9F**). Thus, S214F represents an impactful mutation of *ARAF* whereas V145L represents a neutral mutation in this context. We classified such mutations with a similar signal strength to wild-type but significantly different

signature identities as impactful: not otherwise specified (NOS).

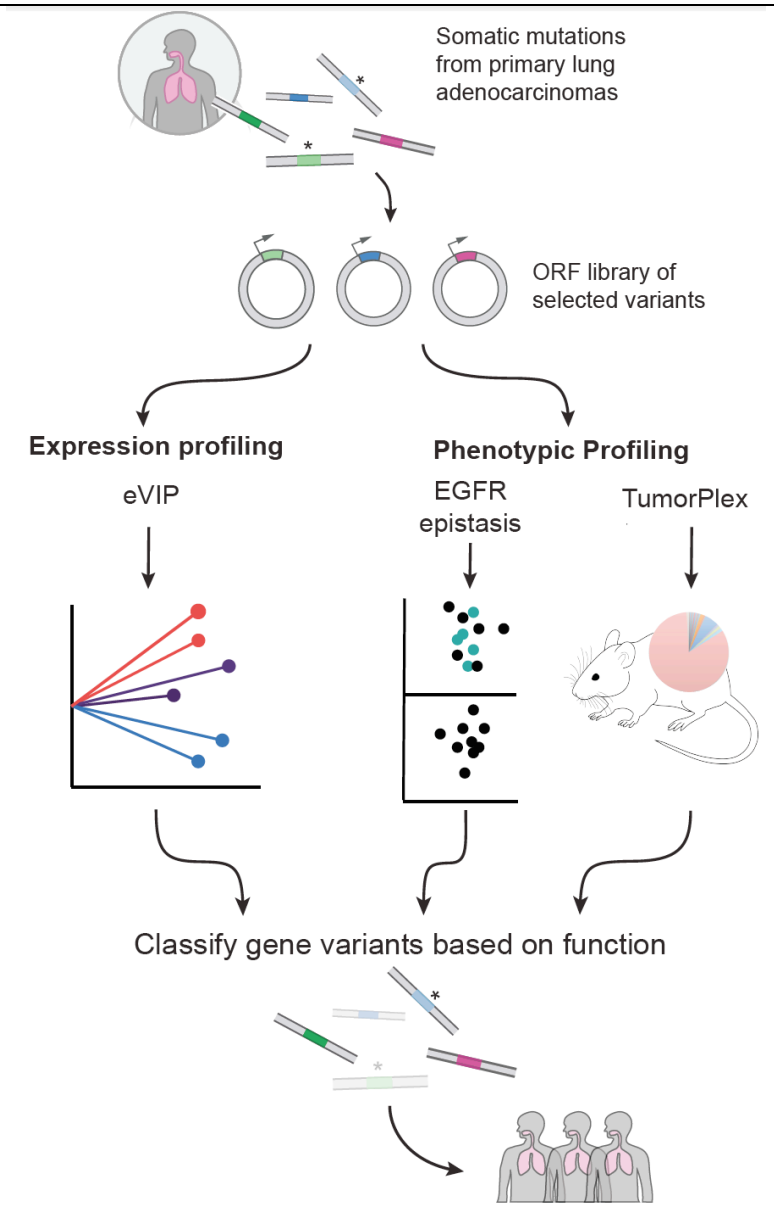


Figure 8. A Pipeline to Translate Genotype Information into Clinical Hypotheses.
Somatic variants are identified from primary lung adenocarcinomas using next-generation sequencing. Selected variants are incorporated into cDNA ORF expression constructs by site-directed mutagenesis and subcloning into an expression vector backbone. The mutant allele library is then profiled using gene-agnostic (eVIP) and pathway-directed assays (EGFR epistasis, TumorPlex). Phenotypic information from the screens is integrated to classify mutations as impactful/neutral and by their effect on cancer-related phenotypes. Together this information leads to clinical hypotheses as to whether a specific variant is likely to be causally involved in tumorigenesis and the impact of the mutant genotype on response to therapy.

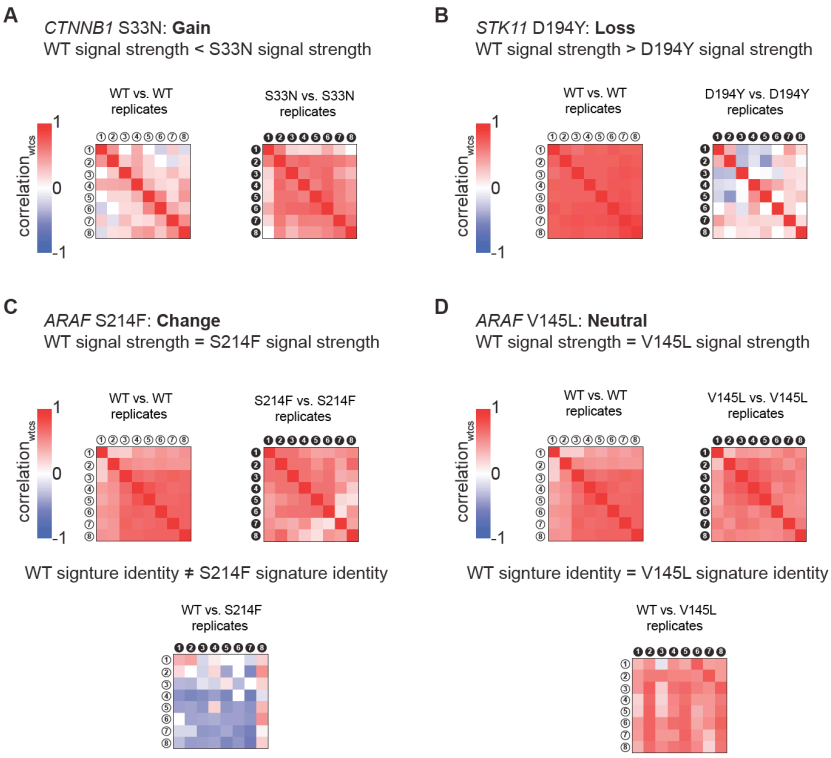


Figure 9. Expression-based variant impact phenotyping.
Comparing signal strength and signature identity of gene expression changes induced upon replicate introduction of a wild-type (WT) ORF or a mutant ORF is used to predict the functional impact of the mutation. Signal strength is determined by comparing the correlation between replicate introductions of the same ORF. (**A**) *CTNNB1* S33N shows a dominant/gain-of-function phenotype where the mutant signal strength is greater than the wild-type. (**B**) *STK11* D194Y shows a loss-of-function phenotype where the mutant signal strength is less than the wild-type. In cases where there is similar signal strength between wild-type and mutant ORFs, the signature identity can determine a functional change. The signature identity is determined by comparing the correlation between replicate introductions of the wild-type versus the mutant ORF. (**C**) *ARAF* S214F has a change-of-function where the signature identity is different. (**D**) *ARAF* V145L is determined to be a neutral mutation because there is no difference in signal strength or signature identity.

2) A genetic epistasis screen to identify genes that confer resistance to EGFR inhibition.

Here, we leveraged a recently developed erlotinib-rescue assay in PC9 lung adenocarcinoma cells [10] to determine which mutations represent gain-of-function mutations that are epistatic to EGFR. PC9 cells harbor an activating *EGFR* exon 19 in-frame deletion and these cells are naturally sensitive to the EGFR inhibitor erlotinib. Expressing alleles such as mutant *KRAS* that re-activate downstream signaling pathways can rescue the erlotinib-induced lethality in this cell type [10]. To identify other mutations epistatic to EGFR, we assessed each ORF for their potential to rescue erlotinib sensitivity at two erlotinib doses after 72 hours of treatment. Replicate conditions and both doses were highly correlated. 62 of 351 ORFs (17.7%) corresponding to 14 genes rescued cell viability in 3 μM erlotinib, including numerous mutant alleles of *KRAS*, *EGFR*, *RIT1*, and *BRAF* (Robust Z score > 2; **Figure 10A**). In addition to canonical recurrent hotspot mutations, many rare mutations, e.g. *PIK3CA* p.E600K, *BRAF* p.A762E, *BRAF* p.L613F, *BRAF* p.P367R, *RIT1* p.A192T, and *KRAS* p.D33E) were able to rescue cell viability of PC9 cells in erlotinib (**Figure 10B**).

3) An in vivo assay to determine tumorigenic potential of putative oncogenes.

In parallel, we developed a multiplexed *in vivo* tumorigenesis assay (TumorPlex) sensitized to read out variants in the EGFR/RAS pathway. We introduced each of the barcoded alleles into SALE-Y cells in arrayed format, pooled in groups of ~70, and injected cells subcutaneously into nude mice (**Figure 10C**). We excluded known activating alleles of *EGFR/KRAS* pathway genes from experimental pools to suppress potential

"jackpot" effects. To deconvolute which ORFs were able to confer tumor-forming capacity to SALE-Y cells, we harvested tumor DNA, amplified, and sequenced ORF barcodes from each tumor. By identifying barcodes that were significantly increased in abundance in tumors versus the pre-injection pools, the top hits were non-canonical alleles in known EGFR/RAS pathway genes such as *BRAF* p.H574Q, *BRAF* p.P367R, *KRAS* p.D33E, *EGFR* p.S645C, *ERBB2* p.S418T, and *RIT1* p.R122L (**Figure 10D-E**).

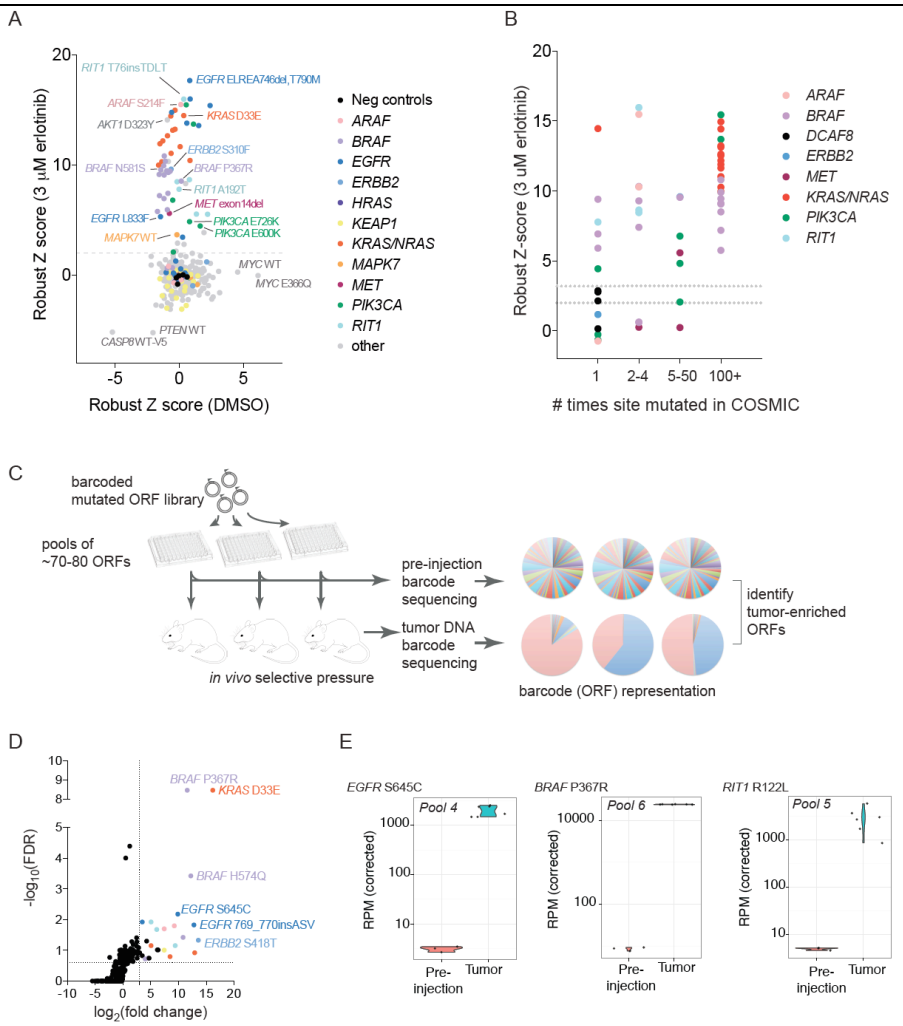


Figure 10. Functional Identification of Rare Oncogenic Mutations.
(**A**) Cell viability of PC9 cells after mutant allele library infection and 72 hours of treatment with 3 μM erlotinib or DMSO. Data shown is the average robust Z score of two replicates per ORF condition. A dashed line indicates the threshold used to select ORFs for validation (Z = 2). (**B**) Relationship between mutation frequency in COSMIC (x-axis) and ability to rescue cell viability in erlotinib (y-axis). Two dashed lines indicate the Z score thresholds used to select ORFs for validation (Z > 2) and the threshold at which all ORFs retested in validation (Z > 3). (**C**) Experimental overview of the TumorPlex pooled xenograft formation assay. Expression clones were introduced into SALE-Y cells in arrayed 96-well format. Infected cells were selected with puromycin, expanded, and then pooled into pools of ~70 ORFs. During in vitro passaging, three pre-injection timepoints were collected for DNA extraction before injection of cells subcutaneously into nude mice in six replicate injection sites per pool. Resulting tumors were dissected, DNA extracted, and sequencing libraries prepared which incorporated sample-specific barcodes. After PCR amplification, sample-specific and ORF-specific barcodes were deconvoluted to quantify representation of reads of each ORF in each sample. ORFs that were enriched in tumors compared to pre-injection samples were identified by calculation of the T statistic after correction of read distributions across samples by linear modeling. (**D**) Volcano plot showing TumorPlex results showing distribution of ORFs from all pools. Each datapoint represents data generated from all pre-injection and tumor replicates for a given ORF. The log2 fold-change (x-axis) was calculated by comparing the median corrected reads-per-million (RPM) of each ORF from the pre-injection samples to the median corrected RPM in tumor samples. (**E**) Violin plots of representative mutants identified as enriched in tumor samples. Data points are corrected RPM values of individual replicates.

3B (v): *MAP2K1* (relevant to tasks 1, 2, 5, 6)

The advent of next generation sequencing has facilitated the genomic characterization of patients' tumors and enabled large-scale functional genomic analyses in vitro. The Cancer Cell Line Encyclopedia (CCLE, http://www.broadinstitute.org/ccle) and Sanger Cell Lines Project (cancer.sanger.ac.uk) have created publicly accessible databases documenting genomic alterations in hundreds of cancer cell lines. These data can potentially be paired with genome-scale functional genomics screens of essentiality to match genetic dependencies with mutations or gene expression patterns. One such genome-scale shRNA screen performed in 216 cancer cell lines across a variety of lineages is publicly available (www.broadinstitute.org/achilles). The uniquely large number and variety of cell lines screened allows for comparison of survival differences in different genetic contexts after knockdown of individual genes, leading to hypotheses generation and subsequent preclinical validation.

To mirror "exceptional responder" studies in vitro, we mined cell line RNAi and genomic data sets to identify genes upon which a small number of cell lines are uniquely dependent. We first explored known relationships between the dependencies and genomic features seen in current targeted therapies. One example discussed above is the inhibition of cells harboring the BCR-ABL1 translocation by the tyrosine kinase inhibitor, imatinib. To evaluate genetic dependency in such cells, we assessed the impact of ABL1 knockdown in the 30 hematopoietic and lymphoid cell lines included in the data set. Of these lines, the three known to harbor the BCR-ABL1 translocation are the most sensitive to ABL1 knockdown (**Figure 11A**).

We ordered all genes in the 21 lung cancer cell lines included in the data set by dependency score. We then calculated an outlier score: the difference in gene-level scores between the cell line with the lowest score and the cell line with the gene-level score at the 90th percentile. Once we generated a ranked list of outlier scores across the lung cancer cell lines, we searched for genes with the highest outlier scores that are also significantly mutated in lung cancers to identify genes whose exceptional dependency may be due to known oncogenic drivers in lung cancers. Of the known oncogenic drivers in lung cancers, *MAP2K1* showed the highest outlier score (1.7, top 1% of outlier scores) (**Figure 11C**).

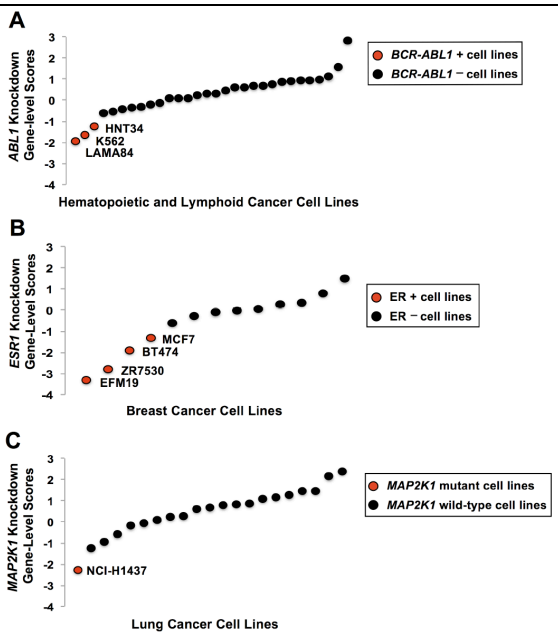


Figure 1. (**A**) Each data point represents the normalized gene-level dependency scores for *ABL1* calculated using ATARiS (gene solution 1) in 30 hematopoietic and lymphoid cell lines. The three most sensitive lines (indicated by red circles) are the only lines within this data set known to harbor the BCR-ABL1 translocation. (**B**) Each data point represents the normalized gene-level dependency scores for ESR1 in the 13 breast cancer cell lines tested. The 4 most sensitive lines (indicated by red circles) are the only breast lines within this data set known to be positive for the estrogen receptor. (**C**) Each data point represents the normalized gene-level dependency scores for *MAP2K1* in 21 lung cancer cell lines. The most sensitive line (indicated by the red circle) is the only lung line in this set known to harbor a *MAP2K1* mutation.

To confirm MAP2K1 dependence in NCI-H1437 cells, we used shRNAs to knockdown MAP2K1 transcript levels. MAP2K1 shRNAs effectively decreased MEK1 protein levels (**Figure 12A**), resulting in decreased viability in the NCI-H1437 cells but not in A549 cells, which harbor an activating *KRAS* mutation (G12S) (**Figure 12B**). Consistent with our results, the MAP2K1 dependency score for A549 cells indicates no effect (gene-level score = -0.2).

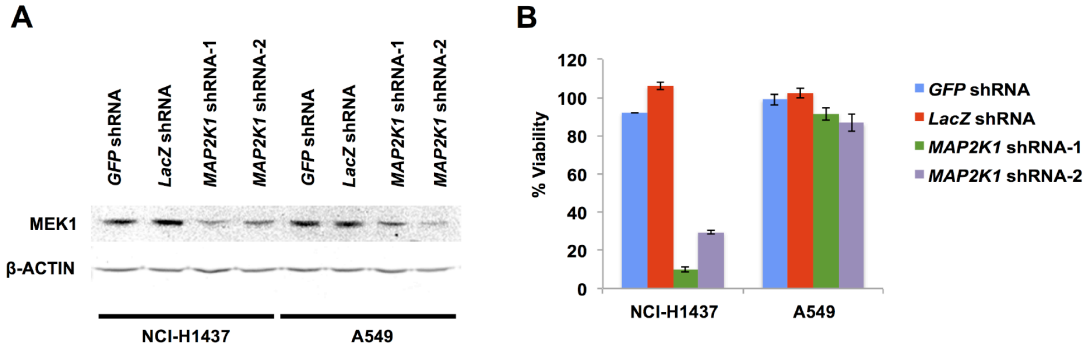


Figure 12. (**A**) Immunoblot of MEK1 and β-ACTIN (loading control) protein levels 5 days after lentiviral infection of control shRNAs (GFP and LacZ) and two independent MAP2K1 shRNAs in NCI-H1437 and A549 cells. (**B**) Cell viability beginning 7 days after lentiviral infections. Viability is normalized to mock infected controls.

Finally, in mouse xenograft experiments, we have shown that treatment with the MEK inhibitor, trametinib, significantly decreases tumor growth (**Figure 13**). Taken together, our results demonstrate a novel *MAP2K1* dependency in tumor cells harboring these mutations. We therefore believe that cancers that bear driver mutations in *MAP2K1* may particularly benefit from treatment with MEK1 inhibitors.

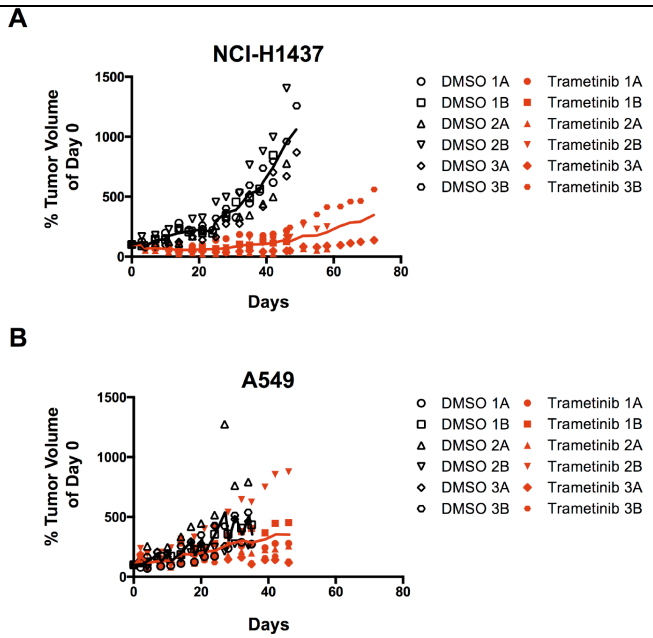


Figure 13. (**A**) Xenograft tumor growth after NCI-H1437 subcutaneous injection in nude mice. Mice were treated daily with either vehicle (5% DMSO) or 0.3mg/kg trametinib. Each number represents a single mouse and the letters A and B represent each tumor. Tumor volume was normalized to its size at the start of drug treatment. Solid lines indicate the average of the six tumors for each treatment. The differences between trametinib and control treatments after day 8 were statistically significant ($p<0.007$). Control mice were all sacrificed during the experiment due to excessive tumor burden. All trametinib treated mice were sacrificed at the conclusion of the experiment. (**B**) Xenograft tumor growth after A549 subcutaneous injection in nude mice. Mice were treated daily with either vehicle (5% DMSO) or 0.3mg/kg trametinib. For each tumor, the volume was normalized to the size at the start of drug treatment. Control and trametinib treated mice were all sacrificed during the experiment due to excessive tumor burden.

3B (v): *MYC* (relevant to tasks 1, 2)

Whole genome analysis approaches are revealing recurrent cancer-associated somatic alterations in non-coding DNA regions. The significance of these findings, however, is often unknown. We combined somatic copy number analysis of 12 tumor types with tissue-specific epigenetic profiling data to identify significant regions of focal amplification harboring super-enhancers. We identified a distinct focal amplification peak at ~450 kb 3' to the *MYC* oncogene on chromosome 8 in lung adenocarcinoma, termed MYC-Lung Adenocarcinoma Super-Enhancer (MYC-LASE). We localized the activity of the super-enhancer element to five distinct regions, named e1 to e5 and used DNAse1 hypersensitivity and site-directed mutagenesis/luciferase reporter assays to further narrow down the active region to the e3 element (**Figure 14**).

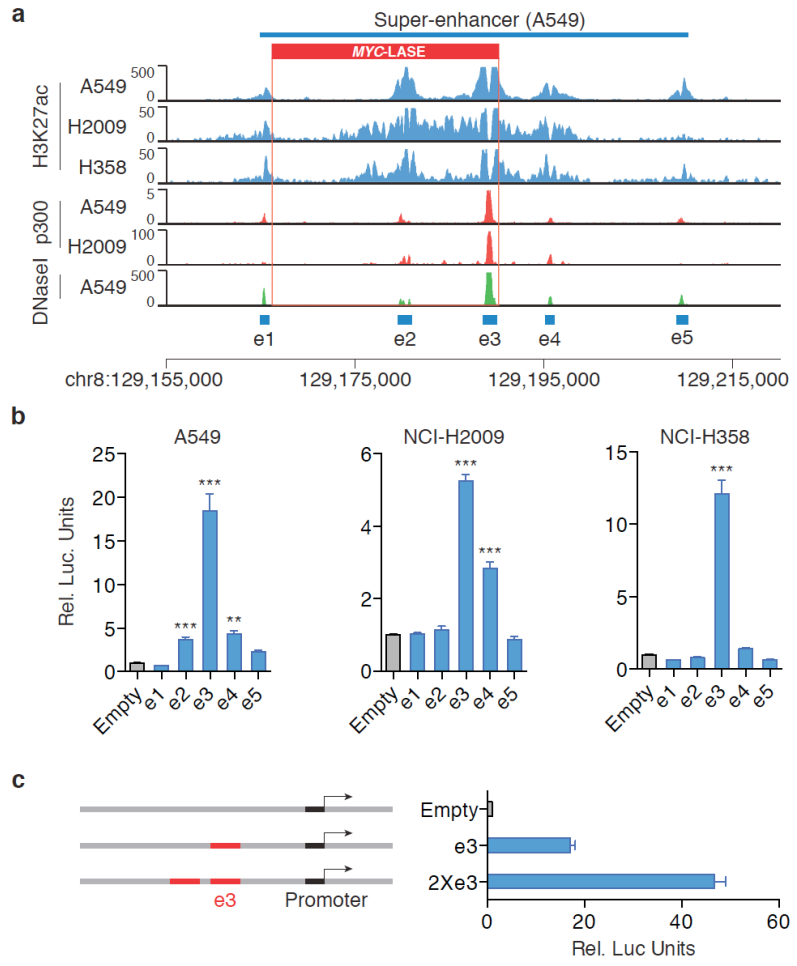


Figure. 14: The activity of the MYC-LASE is predominantly driven by the e3 constituent enhancer. (**a**) H3K27ac, p300 binding and DNase I hypersensitivity profiles in A549, NCI-H2009 and NCI-H358 cells reveal the constituent enhancers e1-e5 within the super-enhancer region. (**b**) Luciferase reporter assay measuring enhancer activity of e1-e5 in A549, NCI-H358 and NCIH2009 lung adenocarcinoma cells. The pGL3 plasmid without the enhancer region (Empty) is used as a negative control. (Y-axis) Relative Luciferase units are normalized to Renilla signal ± SEM. The P-value is derived from a t-test; (**) p≤ p0.01; (***) p≤0.001. (**c**) Duplication of the e3 enhancer (2Å~e3) amplified the enhancer activity that is measured by luciferase reporter assay. The p-value is derived from a t-test; (***) p≤0.001.

In addition to e3 repression, we used the CRISPR/Cas9 system to specifically delete the e3 enhancer in NCI-H2009 cells. Two separate pairs of guide RNAs (sg-e3del #1) and (sg-e3del #2) were used to recruit active Cas9 to the boundaries of the e3 enhancer region and to delete e3 by non-homologous end joining (NHEJ). Deletion of the e3 enhancer region resulted in a ~30% reduction in MYC expression and a significant impairment of both anchorage independent and clonogenic growth. These results suggest that copy number gain of the e3 enhancer region drives MYC over-expression, which further contributes to the tumorigenic phenotype.

We have shown that two distinct focal amplifications of super-enhancers, approximately 450 kb 3' to MYC in lung adenocarcinoma (MYC-LASE) and 800 kb 3' to MYC in endometrial carcinoma (MYCECSE), are physically associated with the *MYC* promoter and correlate with MYC overexpression. A constituent enhancer of MYC-LASE, e3, is the most active component of the super-enhancer and is driven by the transcription factors NFE2L2 and CEBPB. CRISPR/Cas9-mediated repression or deletion of the e3 enhancer in a MYC-LASEamplified lung adenocarcinoma cell line resulted in significant reductions in the expression of MYC and its target genes. Furthermore, repression or deletion of e3 impairs anchorage-independent and clonogenic growth, consistent with an oncogenic function of the amplified super-enhancer.

These results are now part of a manuscript that is provisionally accepted in *Nature Genetics* and that we expect will be e-published soon.

3C: Opportunities for training and professional development provided by the project.

This research project has not only offered numerous and significant opportunities for the training and professional development of scientists directly involved in this research project, but has also contributed more broadly to the productivity and scientific goals of the laboratory. In particular, work funded through this project has allowed fellows to present their research at national and international symposia (described in section 6) and, more importantly, to successfully compete for prestigious research fellowships and faculty positions. Some examples are noted below:

1. Marcin Imielinski, MD, PhD
Dr. Imielinski, a recent clinical fellow in Pathology in the Meyerson laboratory, led two critical studies on somatic gene mutations in lung adenocarcinoma: one, a systematic analysis that revealed mutations in novel genes, including—for the first time—splicing factors and epigenetic regulators (Imielinski, et al., *Cell* 2012), and the other on novel mutations in *ARAF* and *RAF1* that formed the basis of a patient's exceptional response to the drug sorafenib (Imielinski, et al., *J Clin Investig.,* 2014). These accomplishments and the training that Dr. Imielinski received in the Meyerson laboratory were instrumental in his securing a faculty position at Weill Cornell Medical College, NY, in 2014.

2. Peter Choi, PhD
Dr. Choi is a research fellow in the Meyerson laboratory, pursuing studies on the role of the tumor suppressor, *RBM10*, in lung cancer. Peter's work in this regard has led to his obtaining two highly competitive postdoctoral fellowships to date: the F32 Postdoctoral Individual National Research Service Award (NRSA) from the National Institutes of Health and a grant from the International Association for the Study of Lung Cancer (IASLC). In addition to his important work outlining the function of *RBM10* mutations in lung cancer, Peter has also developed novel genome editing methods, using the CRISPR-Cas9 system, to accurately model chromosomal translocations *in vitro* (Choi and Meyerson, *Nat Commun.* 2014), and has jointly led a groundbreaking study (Zhang, Choi, Francis et al., *Nature Genetics*, accepted for publication), on the role of *MYC* enhancer amplification in lung adenocarcinoma. Peter will be applying for academic faculty positions in the near future and I expect he will be a strong candidate given his impressive track record, enabled in great part by support from this grant.

3. Hugh S. Gannon, PhD
Dr. Gannon is a research fellow in the Meyerson laboratory and is interested in identifying cellular dependencies associated with mutations in signaling pathway-related genes in lung cancer. Among other discoveries, he has recently reported that mutations in the *MAP2K1* gene, which he identified first in a lung cancer cell line and later in other tumor lines, conferred extreme dependency to reduction in MEK1 activity, either via genomic editing approaches, transcript knockdown or treatment with the MEK inhibitor, trametinib (Gannon et al., manuscript under revision). Support from this grant has enabled Dr. Gannon's research and also helped him to secure independent research funding from two sources: a postdoctoral fellowship from the American Association for Cancer Research, and an F32 NRSA grant from NIH.

4. Xiaoyang Zhang, PhD
Dr. Zhang is a post-doctoral research fellow in the Meyerson laboratory and is investigating mechanisms by which mutations and amplifications of MYC and its protein partner, MGA, drive lung carcinogenesis. Like his colleagues, Drs. Gannon and Choi, Dr. Zhang is also the recipient of two competitive postdoctoral research fellowships, from the Lung Cancer Research Foundation and the American Association for Cancer Research, thanks to work funded by this grant. Dr. Zhang's work has shed light on the role of genomic amplification of *MYC* enhancers in lung cancer, the significance of which had been unknown to date. This study has recently been accepted for publication in *Nature Genetics*.

5. Alice Berger, PhD
Alice Berger, a post-doctoral fellow in the Meyerson laboratory, is working to understand the impact of somatic mutations in lung adenocarcinoma genes using high throughput *in vitro* and *in vivo* approaches, as described in this report. This is a completely novel approach to address the functional relevance of thousands of allelic

variants in lung cancer simultaneously *in vitro* and in rodent models. A manuscript describing this work on which she is co-first author, along with Dr. Angela Brooks, a former laboratory fellow, will be communicated in the next few weeks for consideration for publication. Dr. Berger is also the lead author on a paper demonstrating the function of the Ras-effector protein, Rit1, on lung tumorigenesis (Berger et al., *Oncogene*, 2014). Additionally, she is joint first author with Dr. Imielinski on the 2012 Cell paper, described above, on discovery of novel recurrently mutated genes in lung adenocarcinoma. Dr. Berger has had an exceptionally productive postdoctoral training period so far and has received fellowships both from the American Cancer Society and a K99/R00 Pathway to Independence (PI) award from the National Cancer Institute. Of note, her K99 application was awarded a score of '10', the highest possible score, testament to her outstanding research productivity and potential. Dr. Berger is presently interviewing for a faculty position at several leading academic institutions, for which she is undoubtedly a frontline contender.

6. Angela Brooks, PhD

Dr. Brooks is a former research fellow in computational biology in the Meyerson laboratory and since 2014 is Assistant Professor of Biomolecular Engineering at the University of California, Santa Cruz. Dr. Brooks's work focuses on understanding how mutations in RNA splicing regulators cause cancer. She has published on splicing alterations that occur in lung and other cancers (Brooks et al., *PLoS One*, 2014) and has collaborated with Dr. Berger, above, on analysis of mutated alleles in lung cancer, lending her computational expertise to make predictions about functional impact. She received a post-doctoral fellowship award from the Damon Runyon Cancer Research Foundation and is currently the recipient of a prestigious Damon Runyon-Dale F. Frey Award for Breakthrough Scientists.

7. Hideo Watanabe, MD, PhD

Dr. Watanabe was a research fellow in the laboratory and joined the faculty at the Mount Sinai School of Medicine, NY, in 2014. Dr. Watanabe was a long-term and important member of the Meyerson laboratory who conducted pioneering work on the role of lineage oncogenes in lung cancer and who mentored and trained numerous junior colleagues during his tenure in the laboratory, including Drs. Choi, Zhang and others. His research efforts overlapped with and fueled multiple projects within the group, including those described in this report, to shed light on lung adenocarcinoma biology.

8. Tanaz Sharifnia, PhD

Dr. Sharifnia obtained her PhD in 2014 from the Harvard Biological and Biomedical Sciences program, under the mentorship of Dr. Matthew Meyerson. Her thesis work centered on genetic and chemical modifiers of EGFR dependence in non-small cell lung cancer, made possible in part by support from this grant, as described in this report and last year's annual report, and was published in the Proceedings of the National Academy of Sciences, USA. Dr. Sharifnia is now a scientist at the Broad Institute of MIT and Harvard, working in the laboratory of Dr. Stuart Schreiber.

3D: Dissemination of results to communities of interest.

The results of our work have been distributed to the scientific and medical communities through our publications. A comprehensive list of the laboratory's publications can be found at:

http://www.ncbi.nlm.nih.gov/pubmed/?term=meyerson+m

3E: Plans for accomplishing goals during the next reporting period.

Not applicable as this is the final project report.

4. IMPACT

The research outcomes of this project have been impactful in the following ways:

a) *Scientific impact:* We have made significant inroads into understanding how multiple individual somatic mutations in dozens of lung cancer-associated genes contribute to the initiation and maintenance of lung cancer. In addition, many of the techniques developed here, e.g. CRISPR-mediated chromosomal translocations, will be of use to researchers within the cancer research community and other disciplines.

b) *Impact on training the next generation of cancer researchers:* The fight against cancer is powered by the caliber and drive of our scientists and clinicians. In this regard, the Meyerson laboratory has been a leader in training researchers in lung cancer genome discovery and the development of experimental and analytical approaches in cancer genomics. Dr. Meyerson has mentored dozens of trainees, the majority of whom have gone on to pursue successful research or clinical careers within the Boston area, nationally, or internationally (some examples noted in Section 3C). Several former trainees are now tenured professors e.g. Roman Thomas, Professor of Medicine at the University of Koln in Germany; Neil Hayes, tenured Associate Professor of Oncology at the University of North Carolina and a leader of The Cancer Genome Atlas' RNA sequencing efforts; and Tom Laframboise at Case Western Reserve University. Many other post-doctoral fellows have pursued industrial research careers while a number of clinician-scientist trainees have accepted academic positions focusing on clinical research and care. Many former lab members also hold prominent academic staff positions at the Broad Institute, Dana-Farber Cancer Institute and Harvard Medical School. Thus, support from grants such as this one has helped to mentor and train the next generation of leaders in cancer research.

c) *Impact on Public Health:* Our studies have revealed new facets of lung cancer biology, including mechanisms of cancer causation, spread and resistance to treatment. These questions are of direct relevance to our goal—and indeed, that of cancer researchers worldwide—namely, to identify new gene candidates and molecular pathways that can be targeted therapeutically. Lung cancer is responsible for the greatest mortality and healthcare burden of all cancers globally; translationally-oriented studies such as this are therefore crucial to expanding our knowledge of this lethal malignancy, to devise better ways to combat it.

5. CHANGES

We have made minor changes to our systems and methods, in response to the evolution of our research outcomes and findings during the tenure of the grant, while adhering to the overall goals of the proposal. This includes, in many cases, conducting more in-depth, systematic or mechanistic studies, especially in areas of potentially high clinical and therapeutic impact, e.g. development of the L1000 eVIP studies [section 3B (iv)].

Changes include changes in cell lines used for functional studies (typically using different or more lines than originally proposed, for rigor); analysis of different genes and alleles, based on emerging evidence of their importance from our own work and that of other groups; and the use of orthogonal approaches to verify our results.

6. PRODUCTS

6A. Publications, conference papers and presentations:

6A (i). Publications:

A list of relevant publications from this last reporting period (Sep 2014 – Oct 2014) is shown below.

For a complete list, please refer to: http://www.ncbi.nlm.nih.gov/pubmed/?term=meyerson+m

- Zhang X, Choi PS, Francis JM, Imielinski M, Watanabe H, Cherniack AD, Meyerson M. Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. 2015 Nature Genetics (under revision, provisionally accepted for publication).

- Xu C, Buczkowski KA, Zhang Y, Asahina H, Beauchamp EM, Terai H, Li YY, Meyerson M, Wong KK, Hammerman PS. NSCLC Driven by DDR2 Mutation Is Sensitive to Dasatinib and JQ1 Combination Therapy. Mol Cancer Ther. 2015 Oct;14(10):2382-9. doi: 10.1158/1535-7163.MCT-15-0077.

- Terai H, Tan L, Beauchamp EM, Hatcher JM, Liu Q, Meyerson M, Gray NS, Hammerman PS. Characterization of DDR2 Inhibitors for the Treatment of DDR2 Mutated Nonsmall Cell Lung Cancer. ACS Chem Biol. 2015 Sep 25. [Epub ahead of print]

- Park E, Kim N, Ficarro SB, Zhang Y, Lee BI, Cho A, Kim K, Park AK, Park WY, Murray B, Meyerson M, Beroukhim R, Marto JA, Cho J, Eck MJ. Structure and mechanism of activity-based inhibition of the EGF receptor by Mig6. Nat Struct Mol Biol. 2015 Sep;22(9):703-11.

- Wilson FH, Johannessen CM, Piccioni F, Tamayo P, Kim JW, Van Allen EM, Corsello SM, Capelletti M, Calles A, Butaney M, Sharifnia T, Gabriel SB, Mesirov JP, Hahn WC, Engelman JA, Meyerson M, Root DE, Jänne PA, Garraway LA. A functional landscape of resistance to ALK inhibition in lung cancer. Cancer Cell. 2015 Mar 9;27(3):397-408.

- Sharifnia T, Rusu V, Piccioni F, Bagul M, Imielinski M, Cherniack AD, Pedamallu CS, Wong B, Wilson FH, Garraway LA, Altshuler D, Golub TR, Root DE, Subramanian A, Meyerson M. Genetic modifiers of EGFR dependence in non-small cell lung cancer. Proc Natl Acad Sci U S A. 2014 Dec 30;111(52):18661-6.

- Tchaicha JH, Akbay EA, Altabef A, Mikse OR, Kikuchi E, Rhee K, Liao RG, Bronson RT, Sholl LM, Meyerson M, Hammerman PS, Wong KK. Kinase domain activation of FGFR2 yields high-grade lung adenocarcinoma sensitive to a Pan-FGFR inhibitor in a mouse model of NSCLC. Cancer Res. 2014 Sep 1;74(17):4676-84.

6A (ii). Conference abstracts and talks:

- Zhang X, Meyerson M and colleagues. Focal amplification of super-enhancers in lung cancer. NCI-Lung-SPORE, June 2015, National Cancer Institute, Bethesda, MD.

- Choi PS, Brooks AN, Meyerson M. Characterization of RNA Splicing Factor Mutations in Lung Adenocarcinoma. IASLC 16th World Conference on Lung Cancer, Denver, CO.

- Alice Berger, Angela N. Brooks, Xiaoyun Wu, Larson Hogstrom, Itay Tirosh, Federica Piccioni, Mukta Bagul, Cong Zhu, Yashaswi Shretha, Candace Chouinard, Nathan Kaplan, David Root, Pablo Tamayo, Ryo Sakai, Bang Wong, Ted Natoli, David Lahr, Atanas Kamburov, Aravind Subramanian, Gad Getz, Todd Golub, Matthew Meyerson, and Jesse Boehm. Expression-based variant impact phenotyping of

lung adenocarcinoma somatic mutations. NCI-Lung-SPORE, June 2015, National Cancer Institute, Bethesda, MD.

- Alice Berger, Angela N. Brooks, Xiaoyun Wu, Larson Hogstrom, Itay Tirosh, Federica Piccioni, Mukta Bagul, Cong Zhu, Yashaswi Shretha, Candace Chouinard, Nathan Kaplan, David Root, Pablo Tamayo, Ryo Sakai, Bang Wong, Ted Natoli, David Lahr, Atanas Kamburov, Aravind Subramanian, Gad Getz, Todd Golub, Matthew Meyerson, and Jesse Boehm. High-throughput gene expression profiling as a generalizable assay for determination of mutation impact on gene function. Translation of the Cancer Genome , American Association for Cancer Research, February 2015, San Francisco, CA.

- M Meyerson. AACR Special Conference on Translation of the Cancer Genome *[Invited speaker]* 2015.

- M Meyerson. Mount Sinai School of Medicine *[Invited speaker]* 2015.

- M Meyerson. United States and Canadian Academy of Pathology *[Invited speaker]* 2015.

- M Meyerson. ASIP Annual Meeting at Experimental Biology *[Invited speaker]* 2015.

- M Meyerson. 39th Annual UNC Linberger Cancer Center Symposium *[Invited speaker]* 2015.

- M Meyerson. AACR Advances in Brain Cancer Research *[Keynote speaker]* 2015.

- M Meyerson. Caring for Carcinoid Foundation Annual Symposium *[Invited speaker]* 2015.

- M Meyerson. Mexican Society of Oncology, SMEO *[Invited speaker]* 2014.

- M Meyerson. Mexican National Cancer Institute *[Invited speaker]* 2014.

- M Meyerson. Korean Global R&D Center Symposium *[Invited speaker]* 2014.

- M Meyerson. 5th International Symposium on Cancer Spiral: Infection, Immunity and Cancer, Kobe, Japan *[Invited speaker]* 2015.

- M Meyerson. University of Tokyo *[Invited speaker]* 2015

- M Meyerson. Japanese Society of Medical Oncology 2015 Annual Scientific Meeting *[Invited speaker]* 2015.

6B. Websites or other internet sites:

Meyerson laboratory website, Harvard Medical School: https://dms.hms.harvard.edu/bbs/fac/Meyerson.php

Meyerson laboratory publications, PubMed: http://www.ncbi.nlm.nih.gov/pubmed/?term=meyerson+m

6C. Technologies or techniques:

All technologies and techniques are described in the relevant publications and will be made available to the scientific community according to federal guidelines.

6D. Inventions, patent applications, licenses:

None

6E. Other products:

As stated in section 6C, above, all cell lines and other biological reagents generated are already or will shortly be described in relevant publications and will be made freely available to the scientific community as per federal requirements.

7. PARTICIPANTS AND OTHER COLLABORATING ORGANIZATIONS

7A. Individuals who have worked on this project:

Name:	Matthew Meyerson
Project Role:	Principal Investigator
Researcher Identifier:	MMEYERSON01 (NIH eRA commons ID)
Nearest person work months:	1.92
Contribution to Project:	Responsible for overall research strategy, direction and communication.
Funding Support:	Detailed in section 7B

Name:	Alice Berger
Project Role:	Post-doctoral fellow
Researcher Identifier:	
Nearest person work months:	8
Contribution to Project:	Worked on developing and optimizing L1000 and eVIP/tumorplex/EGFR epistasis assays and analyzing RIT1 function
Funding Support:	American Cancer Society, K99 Pathway to Independence award

Name:	Marcin Imielinski
Project Role:	Post-doctoral fellow, now faculty at Weill Cornell Medical College, NY
Researcher Identifier:	
Nearest person work months:	2
Contribution to Project:	Computational analysis of impact of somatic variants in lung cancer and implications for treatment efficacy and emergence of resistance
Funding Support:	Mass General Hospital fellowship training program in Pathology

Name:	Tanaz Sharifnia
Project Role:	Graduate Student
Researcher Identifier:	None available
Nearest person work months:	2
Contribution to Project:	Worked on cellular mechanisms involved in resistance to EGFR-targeting drugs in lung cancer
Funding Support:	

Name:	Angela Brooks
Project Role:	Post-doctoral fellow, now faculty at University of California, Santa Cruz
Researcher Identifier:	ABROOKS2011 (NIH eRA commons ID)
Nearest person work months:	4
Contribution to Project:	Computational analysis of functional impact of somatically mutated alleles in lung cancer
Funding Support:	Damon Runyon Fellowship, Damon Runyon Frey Award

Name:	Hugh Gannon
Project Role:	Post-doctoral fellow
Researcher Identifier:	HGANNON (NIH eRA commons ID)
Nearest person work months:	7
Contribution to Project:	Worked on MAP2K1 alterations in lung adenocarcinoma and dependencies in multiple model systems
Funding Support:	American Association for Cancer Research, F32 NRSA grant from NIH

Name:	Peter Choi
Project Role:	Post-doctoral fellow
Researcher Identifier:	PSCHOI (NIH eRA commons ID)
Nearest person work months:	8
Contribution to Project:	Worked on analysis of mutations in *RBM10* and their functional consequences, in lung adenocarcinoma.
Funding Support:	NIH F32 Postdoctoral National Research Service Award (NRSA) and International Association for the Study of Lung Cancer (IASLC)

Name:	Xiaoyang Zhang
Project Role:	Post-doctoral fellow
Researcher Identifier:	
Nearest person work months:	8
Contribution to Project:	Analysis of mechanism of MYC enhancer amplification in lung cancer
Funding Support:	Lung Cancer Research Foundation and American Association for Cancer Research

7B. Active other support of the PI.

Attached in Appendix

7C. Other organizations involved as partners:

None

8. SPECIAL REPORTING REQUIREMENTS

None

9. APPENDICES

9A. Research Support of PI

Active:

1. R35CA197568 (Meyerson) 08/01/2015 – 06/30/2022
National Institutes of Health/National Cancer Institute
How Do Genome Alterations Cause Human Lung Cancer?
This project is aimed at understanding the mechanism of how significant alterations in the DNA of lung tumors such as loss or gain of whole chromosomes or chromosome arms; mutations within genes that regulate transcript splicing; and genomic amplification of "lineage" oncogenes, cause lung cancer. We believe the insights so gained will uncover new therapeutic approaches to combat this disease.
Role: Principal Investigator

2. American Cancer Society Research Professor Award (Meyerson) 01/01/2015 – 12/31/2019
American Cancer Society
Genome-Inspired Approaches to Cancer Discovery and Therapeutics
This project focuses on a number of research areas, including application of whole genome sequence analysis to find the major missing genome alterations that cause human lung cancer; evaluation of how somatic genome alterations in lung cancers predict their response to immunomodulatory therapies; and germline variants associated with increased susceptibility to lung cancers with somatic *EGFR* mutation in patients with origins in eastern Asia
Role: Principal Investigator

3. P01CA154303 (Meyerson) 05/11/2012 – 04/30/2017
National Institutes of Health/National Cancer Institute
Protein Kinase Therapeutic Targets for Non-Small Cell Lung Carcinoma
This Program Project will advance the development of targeted therapies for non-small cell lung cancer, by developing three protein kinases, EGFR, TBK1, and DDR2, as targets for inhibition.
Role: Program Director, Project Leader, Core Director

4. R01CA109038 (Meyerson) 03/01/2010 – 12/31/2015
National Institutes of Health/National Cancer Institute
NKX2-1, A Candidate Lineage Survival Oncogene in Lung Adenocarcinoma
This study aims to identify the transcriptional and functional targets of NKX2-1 in lung adenocarcinoma, and perform shRNA screens to identify genes that are specifically required for the survival of lung adenocarcinoma cells with amplified *NKX2-1*.
Role: Principal Investigator

5. R01CA116020 (Eck/Meyerson) 04/01/2012 – 03/31/2017
National Institutes of Health/National Cancer Institute
Inhibitor-sensitive and -resistant EGFR Mutants from Lung Cancer and Glioblastoma
The aims of this project are to characterize the response of novel EGFR mutants to enzymatic inhibitors and to antibody therapies; and to perform structural and functional studies of the exon 20 insertion mutants of EGFR, which are resistant to gefitinib, erlotininb and cetuximab.
Role: Principal Investigator (with Michael Eck)

6. Claudia Adams Barr Award (Meyerson/Pellman) 10/01/2014 – 09/30/2016
Dana-Farber Cancer Institute
Elucidating the Mechanisms of Chromosomal Rearrangements in Cancer at the Single Cell Level
This study will investigate the mechanistic basis of chromothripsis, a catastrophic chromosome rearrangement event seen in cancer cells. We will use live-cell imaging coupled with single-cell sequencing to study how micronuclei that form during aberrant cell division can contribute to the development of chromothripsis.

Role: Principal Investigator (with David Pellman)

7. Bayer Pharmaceuticals (Broad Institute) 09/01/2013 – 08/31/2016
Discovery of Therapeutic Agents Selective for Cancer Genome Alterations
The overall goal of this collaboration is to develop candidate therapies to better treat human cancers.
Role: Project Leader, member of Joint Steering Committee and Joint Research Committee

Pending:

8. The V Foundation (Meyerson) 09/01/2015 – 08/31/2018
The contribution of germline genomic variation to EGFR mutation frequency in non-small cell lung cancers in Latin American populations: implications for cancer risk prediction, screening and intervention.
The goals of this project are:
(i). To determine the population ancestry admixture in lung adenocarcinomas from Columbia and Mexico by genome-wide molecular genotyping
(ii). To elucidate the relationship between fraction of Native American ancestry and EGFR mutation, and identify specific ancestry markers associated with increased risk of somatic EGFR mutation in lung adenocarcinoma

9B. Statement of Work

Task 1: Create cell line models of ectopic expression with selected lung cancer somatic mutations (months 1-12)

1a. Select approximately 200 mutant alleles from lung cancer somatic mutation data (months 1-4)
1b. Introduce these mutations into wild-type cDNAs by site-directed mutagenesis (months 2-12)
1c. Transduce NIH-3T3 cells with wild-type and mutant cDNAs (months 4-24)
1d. Transduce BEAS-2B cells with wild-type and mutant cDNAs (months 4-24)

Task 2: Measure oncogenic transformation of engineered cells (months 4-24)

2a. First 100 NIH-3T3 soft agar assays (months 4-14)
2b. Second 100 NIH-3T3 soft agar assays (months 15-24)
2c. AALE soft agar assays in parallel where appropriate (months 13-24)

Task 3: Perform RNA sequencing on tumor cell lines harboring wild-type or mutant *RBM10* (months 13-30)

3a. Cell line sequencing (months 13-24)
3b. Data analysis (months 24-30)

Task 4: Assess mRNA levels of 1000 landmark genes in engineered BEAS-2B cells (months 12-36)

4a. First 100 L1000 assays (months 12-24)
4b. Second 100 L1000 assays (months 25-32)
4c. Use L1000 results to mine the Connectivity Map database (months 24-36)

Task 5: Assess inhibitor efficacy in Ba/F3 cells (months 25-36)

5a. Ectopically express validated oncogenic alleles and corresponding wild-type alleles in Ba/F3 cells (months 25-30)
5b. Withdraw IL-3 and test IL-3 independence (months 26-31)
5c. Use signaling pathway data with assistance from the Connectivity Map to help identify inhibitors and combinations of inhibitors that impinge on ectopically expressed mutant protein (months 27-36)
5d. Measure engineered Ba/F3 cell survival upon inhibitor treatment using WST assays (months 28-36)
5e. Measure engineered Ba/F3 cell survival upon transduction with specific shRNAs targeting the mutant proteins of interest (months 30-36)

Task 6: Assess inhibitor efficacy in lung cancer cell lines (months 28-36)

6a. Use data from the Cancer Cell Line Encyclopedia and Sanger Institute to select lung cancer cell lines with appropriate genomic lesions (months 28-30)
6b. Measure survival of selected lung cancer cell lines with specific shRNAs and inhibitors proven effective in Ba/F3 cells (months 30-36)
6c. Review Project Achilles (Broad) and MGH/Sanger Institute chemical sensitivity data to identify additional cell lines that may harbor lesions of interest (months 32-34)
6d. Sequence relevant genes and confirm inhibitor sensitivity in these additional cell lines (months 32-36)

9C. References

1. Gripp, K.W., et al., *Long-term survival in TARP syndrome and confirmation of RBM10 as the disease-causing gene.* Am J Med Genet A, 2011. **155A**(10): p. 2516-20.
2. Johnston, J.J., et al., *Massively parallel sequencing of exons on the X chromosome identifies RBM10 as the gene that causes a syndromic form of cleft palate.* Am J Hum Genet, 2010. **86**(5): p. 743-8.
3. *Comprehensive molecular profiling of lung adenocarcinoma.* Nature, 2014. **511**(7511): p. 543-50.
4. Imielinski, M., et al., *Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing.* Cell, 2012. **150**(6): p. 1107-20.
5. Wang, Y., et al., *Integrative analysis revealed the molecular mechanism underlying RBM10-mediated splicing regulation.* EMBO Mol Med, 2013. **5**(9): p. 1431-42.
6. Inoue, A., et al., *RBM10 regulates alternative splicing.* FEBS Lett, 2014. **588**(6): p. 942-7.
7. Bechara, E.G., et al., *RBM5, 6, and 10 differentially regulate NUMB alternative splicing to control cancer cell proliferation.* Mol Cell, 2013. **52**(5): p. 720-33.
8. Kwak, E.L., et al., *Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer.* N Engl J Med, 2010. **363**(18): p. 1693-703.
9. Sanchez-Rivera, F.J., et al., *Rapid modelling of cooperating genetic events in cancer through somatic genome editing.* Nature, 2014. **516**(7531): p. 428-31.
10. Sharifnia, T., et al., *Genetic modifiers of EGFR dependence in non-small cell lung cancer.* Proc Natl Acad Sci U S A, 2014. **111**(52): p. 18661-6.
11. Cho, J., et al., *Colon cancer-derived oncogenic EGFR G724S mutant identified by whole genome sequence analysis is dependent on asymmetric dimerization and sensitive to cetuximab.* Mol Cancer, 2014. **13**: p. 141.
12. Francis, J.M., et al., *EGFR variant heterogeneity in glioblastoma resolved through single-nucleus sequencing.* Cancer Discov, 2014. **4**(8): p. 956-71.
13. Greulich, H., et al., *Functional analysis of receptor tyrosine kinase mutations in lung cancer identifies oncogenic extracellular domain mutations of ERBB2.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(36): p. 14476-81.
14. De Greve, J., et al., *Clinical activity of afatinib (BIBW 2992) in patients with lung adenocarcinoma with mutations in the kinase domain of HER2/neu.* Lung Cancer, 2012. **76**(1): p. 123-7.
15. Mazieres, J., et al., *Lung cancer that harbors an HER2 mutation: epidemiologic characteristics and therapeutic perspectives.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2013. **31**(16): p. 1997-2003.
16. Barretina, J., et al., *The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity.* Nature, 2012. **483**(7391): p. 603-7.
17. Paez, J.G., et al., *EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy.* Science, 2004. **304**(5676): p. 1497-500.
18. Mok, T.S., et al., *Gefitinib or Carboplatin-Paclitaxel in Pulmonary Adenocarcinoma.* N Engl J Med, 2009. **361**(10): p. 947-957.
19. Gandhi, L., et al., *Phase I study of neratinib in combination with temsirolimus in patients with human epidermal growth factor receptor 2-dependent and other solid tumors.* J Clin Oncol, 2014. **32**(2): p. 68-75.
20. Trowe, T., et al., *EXEL-7647 inhibits mutant forms of ErbB2 associated with lapatinib resistance and neoplastic transformation.* Clin Cancer Res, 2008. **14**(8): p. 2465-75.
21. Branford, S., et al., *High frequency of point mutations clustered within the adenosine triphosphate-binding region of BCR/ABL in patients with chronic myeloid leukemia or Ph-positive acute lymphoblastic leukemia who develop imatinib (STI571) resistance.* Blood, 2002. **99**(9): p. 3472-5.
22. Pao, W., et al., *Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib Is associated with a second mutation in the EGFR kinase domain.* PLoS Med, 2005. **2**(3): p. e73.
23. Peck, D., et al., *A method for high-throughput gene expression signature analysis.* Genome Biol, 2006. **7**(7): p. R61.