HHS Public Access

Author manuscript

*Mol Diagn Ther*. Author manuscript; available in PMC 2018 August 01.

Published in final edited form as:

*Mol Diagn Ther*. 2017 August ; 21(4): 431–436. doi:10.1007/s40291-017-0275-y.

# Identification of a novel somatic mutation leading to allele dropout for *EGFR* L858R genotyping in non-small cell lung cancer

**Helio A. Costa, PhD**[1], **Joel W. Neal, MD, PhD**[2], **Carlos D. Bustamante, PhD**[1,3], and **James L. Zehnder, MD**[4,*]

[1]Stanford University School of Medicine, Department of Genetics, Stanford, CA, 94305, USA

[2]Stanford University School of Medicine, Department of Medicine, Division of Oncology, Stanford, CA, 94305, USA

[3]Stanford University School of Medicine, Department of Biomedical Data Science, Stanford, CA, 94305, USA

[4]Stanford University School of Medicine, Department of Pathology, Stanford, CA, 94305, USA

## Abstract

**Objective—**While PCR-based genotyping methods abound in molecular testing for lung cancer therapy, these approaches may not provide the robust sensitivity to detect accurate genotypes in a variable cancer genomic background.

**Methods—**Here, we describe a study of a clinical tumor specimen containing a novel somatic single nucleotide variant that caused allele drop-out in *EGFR* L858R genotyping, resulting in a false negative interpretation and impacting patient clinical management.

**Results—**We demonstrate that a subsequent unbiased next generation sequencing approach correctly identified the driver mutation, and therefore may be more reliable for somatic variant detection.

**Conclusions—**These findings magnify the potential pitfalls of PCR amplification based approaches and stress the importance of unbiased and sensitive molecular testing strategies for therapeutic marker detection as molecular testing becomes the standard for determining clinical management of cancer patients.

## 1. Introduction

Lung cancer is the leading cause of cancer mortality and accounts for approximately 27% of all cancer deaths in adults in the United States [1]. Further, patients presenting with

*Correspondence to: zehnder@stanford.edu, +1-650-723-9532, (J.L.Z.).

**5. Compliance with ethical standards**

5.1 Conflict of interest

J.W.N. is a consultant for Clovis Oncology, CARET, Nektar, Boehringer Ingelheim, ARMO BioScience and ARIAD. C.D.B. is on the scientific advisory boards (SAB) of AncestryDNA, BigDataBio, Etalon DX, Liberty Biosecurity, and Personalis. He is also a founder and SAB chair of Identify Genomics. None of these entities played a role in the design, execution, interpretation, or presentation of this study.

advanced non-small cell lung cancer (NSCLC) who are left untreated generally have a poor prognosis with a four-to-five month median survival time [2]. However, the use of therapeutic agents targeting the epidermal growth factor receptor (EGFR) have dramatically improved the clinical management of the approximately 15% of patients in the United States and 50% of patients in Asia with lung adenocarcinomas harboring *EGFR* mutations. The most common of these mutations are exon 19 deletions and the *EGFR* L858R mutation. When present, the L858R mutation has been shown to increase patient sensitivity to first-, second-, and third-generation EGFR tyrosine kinase inhibitors (TKIs) leading to longer progression-free survival on TKI therapy as compared to those receiving chemotherapy [3, 4].

As a result, accurate and rapid detection of underlying *EGFR* mutations in lung cancer specimens is crucial for effective management of patients. Common molecular diagnostic approaches for genotyping the therapeutically actionable mutations in the *EGFR* locus include PCR-based methods that utilize specifically designed primers to interrogate the region of interest. However, amplification-based genotyping assays may have artifacts such as allele dropout that represent a risk of false negative results and misdiagnosis. A common mechanism for allele dropout is the inability of primers to stably bind to their specific complementary sequence binding site as a result of non-complementation from single nucleotide variants [5]. Here we report a specimen that was falsely determined to be negative for the *EGFR* L858R mutation as a result of allelic dropout caused by a nearby novel silent coding mutation (c.2571G>A, p.G857G), and was subsequently correctly genotyped as an L858R carrier using an unbiased target-capture based sequencing approach. These findings stress the importance of unbiased molecular diagnostic methodologies for sensitive and specific molecular markers for precision cancer care.

## 2. Materials and methods

### 2.1 Specimens and DNA extraction

The patient specimen described in this study was obtained from a fine needle aspirate of a right supraclavicular lymph node from the Stanford Thoracic Oncology Clinic under institutional review broad-approved protocols. Pathological review diagnosed the specimen as metastatic adenocarcinoma consistent with a lung primary and with 20% tumor purity. A NCI-H1975 non-small cell cancer cell line (ATCC CRL-5908, American Type Culture Collection, Manassas, VA, USA) was used as the *EGFR* L858R 20% sensitivity control for the ‘SNaPshot’ genotyping assay (SNaPshot Multiplex Kit, Applied Biosystems, Foster City, CA, USA). Genomic DNA from both patient and control samples was extracted using a Qiagen DNeasy Blood and Tissue DNA Purification kit (QIAamp, Qiagen, Inc., Valencia, CA, USA).

### 2.2 SNaPshot genotyping

The specimens were genotyped by first PCR amplifying exon 21 of the EGFR gene. The exon 21 amplicons were treated to remove primers and single nucleotides, and subsequently amplified to identify the L858R mutation in a ‘SNaPshot’ reaction that consists of a dideoxy single-base extension of an unlabeled oligonucleotide extension primer. The sequence for

the c.2573 extension primer is as follows: '(GACT)x9 AGATCACAGATTTTGGGC'. The primer is designed to bind to a complimentary template in the presence of fluorescently labeled ddNTPs and DNA polymerase. The polymerase extends the primer by one nucleotide, adding a single labeled ddNTP to its 3′ end. SNaPshot products were then run on an ABI3500xl genetic analyzer for capillary electrophoresis fragment analysis and the data was analyzed with Genemapper software. The sensitivity of this assay is approximately 20% EGFR mutation detection in a background of EGFR wild type DNA.

### 2.3 Sequencing, mapping, variant calling and annotation

The patient specimen was sequenced on an Illumina MiSeq instrument producing 100bp paired end reads. We assessed the sequencing run for multiple quality metrics including number of reads, read depth, and selector on-target rate (Table 1). We performed bioinformatic analysis on the specimen as previously described [6]. In brief, sequencing reads were mapped to the human reference genome (hg19) using the BWA mapping algorithm (v 0.6.2)[7]. Mapped reads were filtered to include only properly paired reads with Phred quality scores greater than 30. Variant calling was performed using VarScan 2 [8] with tailored post processing filters to improve variant call confidence. All tools were run using the best practices outlined in their respective manuals unless otherwise indicated. Called variants were annotated for a series of functional predictions, conservation scores, in addition to publicly available database annotations using a combination of perl scripts and ANNOVAR [9].

## 3. Results

A supraclavicular lymph node fine needle aspirate from a 49-year-old Chinese male was reviewed by a surgical pathologist and diagnosed as a metastatic adenocarcinoma originating from a lung primary with approximately 20% tumor purity. The specimen was concurrently submitted for testing on a targeted EGFR mutation panel that includes the L858R allele, and a high-throughput 198 gene pan-cancer somatic mutation sequencing panel which includes the entire coding region of the *EGFR* gene.

### 3.1 False negative EGFR L858R genotype

Targeted *EGFR* L858R genotyping results from a dideoxy single-base extension 'SNaPshot' assay is shown in Figure 1. The test consists of PCR amplification of the L858R containing exon followed by a single base pair extension reaction with a 54-nucleotide long extension primer. The extension primer anneals 5′-adjacent to the L858R causing nucleotide (c.2573) and incorporates a single fluorescently labelled ddNTP. Slight PCR product size differences are due to the influence of different fluorescent dyes associated with the specific incorporated nucleotide. The positive control for this clinical genotyping assay (Figure 1C) demonstrates a 61 nucleotide wild type fragment for the L858 codon in red, and 59 nucleotide fragment corresponding to the spike-in L858R coding (c.2573T>G) mutation at 20% allele frequency in blue.

The findings from the positive control (Figure 1C) are in contrast to the single prominent wild type peaks produced by both the normal control (Figure 1B) and the patient specimen

(Figure 1A). Clear absence of any other fragment peaks suggested that the patient specimen was negative for the L858R (c.2573T>G) mutation.

### 3.2 NGS sequencing of the EGFR locus detects L858R mutation

In parallel to the targeted *EGFR* genotyping, the patient specimen underwent a large capture-based next generation sequencing (NGS) pan-cancer panel that included the *EGFR* locus. Table 2 shows the variants identified in the *EGFR* exon 21 region that passed our quality control filtering criteria. The criteria include filtering nucleotide changes that do not produce an amino acid alteration. Following this criteria, we identified the causal L858R c.2573T>G mutation at 23.9% allele frequency within the specimen. This high confidence variant was sequenced at a 2648X depth of coverage with 613 reads supporting the c.2573T>G variant.

### 3.3 Identification of a silent EGFR mutation

To assess the conflicting genotyping and sequencing results we analyzed the raw mapping read alignment (.bam) file from the NGS sequencing for any potential discrepancies within the data. Additional bioinformatic analysis identified an adjacent point mutation to the L858 allele causing a silent amino acid change (c.2571G>A, p.G857G) (Figure 2). Figure 2 shows an IGV bam pileup visualization of the locus [10]. The G857G (c.2571G>A) allele is shown centered with black vertical lines bounding it. The G857G is in *cis* with the L858R allele in virtually all reads and is present at a nearly identical depth (2586X, 619 variant supporting reads) and allele frequency (23.2%).

The presence of the G857G variant disrupted the 3′ extension primer binding for the ‘SNaPshot’ genotyping assay for this patient. As the G857G variant was uniquely in *cis* with the L858R variant on the same somatic allele, the wild type allele was preferentially amplified resulting in a false negative result (Figure 3C). This is in contrast to normal specimens harboring two wild type *EGFR* L858 (c.2573T) alleles producing a single PCR fragment (red) (Figure 3A) and specimens with only the L858R and wild type *EGFR* alleles producing a mutant PCR fragment (blue) and a wild type PCR fragment (red) (Figure 3B).

The G857G variant appears to represent a novel somatic variant that has not previously been reported in common cancer somatic databases such as COSMIC, cBioPortal, and the ICGC Data Portal [11–13]. However, estimating the occurrence of silent synonymous mutations is challenging as variant submission is biased towards changes that produce amino acid substitutions. Thus is it difficult to determine the potential rate of allele dropout in *EGFR* L858R PCR-based genotyping cases due to synonymous variant reporting bias. Bioinformatic queries into the 1000 Genomes Phase 3 germline dataset and ExAC germline variant browser demonstrate that this variant is not present in any of the available germline data across a variety of diverse populations [14, 15].

## 4. Discussion

These findings demonstrate a novel somatic variant (p.G857G, c.2571G>A) resulting in the incorrect genotyping of a lung cancer patient for the *EGFR* L858R allele using a commonly implemented single base pair extension method. The presence of these PCR-based

genotyping pitfalls has the potential to be exacerbated as continuing molecular testing guidelines suggest EGFR mutation status genotyping as standard of lung cancer care [16]. These shortcomings may be further exacerbated by reporting biases in commonly used mutation databases to guide appropriate primer design and test development. Immediate approaches to alleviate some of these issues include using multiple alternative primer sets or preferentially using methods insusceptible to allelic dropout events. NGS approaches, such as the capture-based methodology we present here, also present addressable limitations such as potential difficulties in successfully capturing the region of interest, or false positive findings due to low sequencing depth. These issues highlight the significance of appropriate assay design and thorough assay clinical validation. This study underscores the clinical principle that a multitude of factors can lead to false negative results in molecular testing including poor DNA yield, low tumor purity or limitations of specific methodologies, and that repeat testing with a new methodology can be clinically important. Here we have demonstrated the utility of NGS sequencing technology for specimen mutation status detection and the importance of the use and development of robust methodologies in precision molecular patient care.

## Acknowledgments

**Funding:** This research was supported by grants from the National Institutes of Health.

5.2 Funding

This research was supported by grants from the National Institutes of Health (ClinGen 4U01HG007436-04 to H.A.C. and C.D.B.; NCI 2PO1CA49605 to J.L.Z.).

## References

1. Hankey BF, Ries LA, Edwards BK. The surveillance, epidemiology, and end results program: a national resource. Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 1999; 8(12):1117.
2. Rapp E, Pater J, Willan A, Cormier Y, Murray N, Evans W, et al. Chemotherapy can prolong survival in patients with advanced non-small-cell lung cancer--report of a Canadian multicenter randomized trial. Journal of Clinical Oncology. 1988; 6(4):633–641. [PubMed: 2833577]
3. Sequist L, Yang J, Yamamoto N, O'Byrne K, Hirsh V, Mok T, et al. Phase III Study of Afatinib or Cisplatin Plus Pemetrexed in Patients With Metastatic Lung Adenocarcinoma With EGFR Mutations. Journal of Clinical Oncology. 2013; 31(27):3327–3334. [PubMed: 23816960]
4. Jänne P, Yang J, Kim D, Planchard D, Ohe Y, Ramalingam S, et al. AZD9291 in EGFR Inhibitor–Resistant Non–Small-Cell Lung Cancer. New England Journal of Medicine. 2015; 372(18):1689–1699. [PubMed: 25923549]
5. Blais J, Lavoie S, Giroux S, Bussières J, Lindsay C, Dionne J, et al. Risk of Misdiagnosis Due to Allele Dropout and False-Positive PCR Artifacts in Molecular Diagnostics. The Journal of Molecular Diagnostics. 2015; 17(5):505–514. [PubMed: 26146130]
6. Newman A, Bratman S, To J, Wynne J, Eclov N, Modlin L, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nature Medicine. 2014; 20(5): 548–554.
7. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25(14):1754–1760. [PubMed: 19451168]
8. Koboldt D, Zhang Q, Larson D, Shen D, McLellan M, Lin L, et al. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Research. 2012; 22(3): 568–576. [PubMed: 22300766]

9. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Research. 2010; 38(16):e164–e164. [PubMed: 20601685]

10. Robinson J, Thorvaldsdóttir H, Winckler W, Guttman M, Lander E, Getz G, et al. Integrative genomics viewer. Nature Biotechnology. 2011; 29(1):24–26.

11. Forbes S, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Research. 2014; 43(D1):D805–D811. [PubMed: 25355519]

12. Cerami E, Gao J, Dogrusoz U, Gross B, Sumer S, Aksoy B, et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data: Figure 1. Cancer Discovery. 2012; 2(5):401–404. [PubMed: 22588877]

13. Zhang J, Baran J, Cros A, Guberman J, Haider S, Hsu J, et al. International Cancer Genome Consortium Data Portal--a one-stop shop for cancer genomics data. Database. 2011; 2011(0):bar026-bar026. [PubMed: 21930502]

14. Auton A, Abecasis G, Altshuler D, Durbin R, Abecasis G, Bentley D, et al. A global reference for human genetic variation. Nature. 2015; 526(7571):68–74. [PubMed: 26432245]

15. Lek M, Karczewski K, Minikel E, Samocha K, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016; 536(7616):285–291. [PubMed: 27535533]

16. Lindeman N, Cagle P, Beasley M, Chitale D, Dacic S, Giaccone G, et al. Molecular Testing Guideline for Selection of Lung Cancer Patients for EGFR and ALK Tyrosine Kinase Inhibitors: Guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. Journal of Thoracic Oncology. 2013; 8(7): 823–859. [PubMed: 23552377]

**Key points**

- Correct EGFR mutation genotyping is critical for the proper therapeutic management of lung cancer patients.
- This study highlights a major pitfall of a common clinical genotyping methodology that results in a false negative result for a clinically actionable EGFR mutation.
- We demonstrate an alternative next-generation sequencing method that is more reliable and robust.

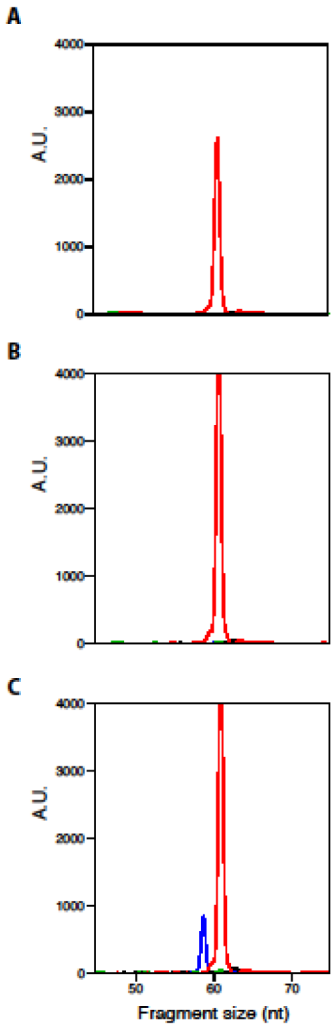


**Fig. 1. SNaPshot genotyping fragment analysis**

(**A,B**) Patient specimen (**A**) and normal control specimen (**B**) with a wild type *EGFR* L858 (c.2573T) 61 nucleotide (nt) PCR fragment (**C**) Positive control specimen with a wild type *EGFR* L858 (c.2573T) 61 nucleotide PCR fragment (red) and a mutant *EGFR* L858R (c. 2573T>G) 59 nt PCR fragment at approximately 20% allele frequency (blue). Y-axis represents fluorescence arbitrary units (A.U.).

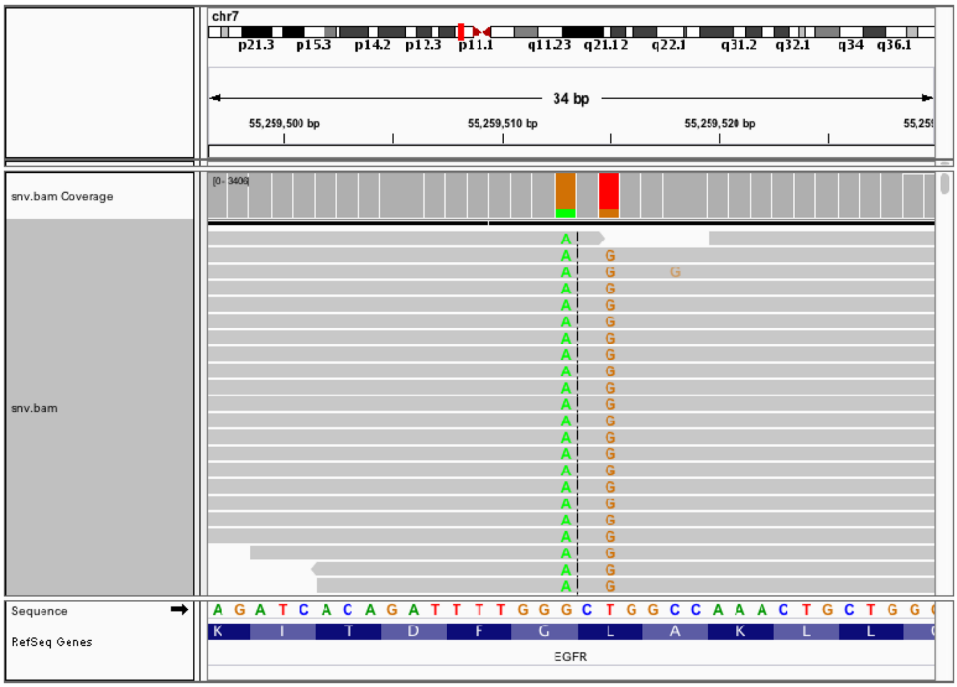


**Fig. 2. Integrative Genomics Viewer alignment mapping for L858R locus**

Read alignment (.bam) file view of patient's sequenced EGFR locus. The silent G857G (c. 2571G>A) is shown in green with a black vertical dotted line bounding the sequence. The L858R (c.2573T>G) variant is shown two nucleotides away in brown. Both variants are found at approximately 23% variant allele frequency and are in *cis* in virtually all >2500 supporting variant reads.

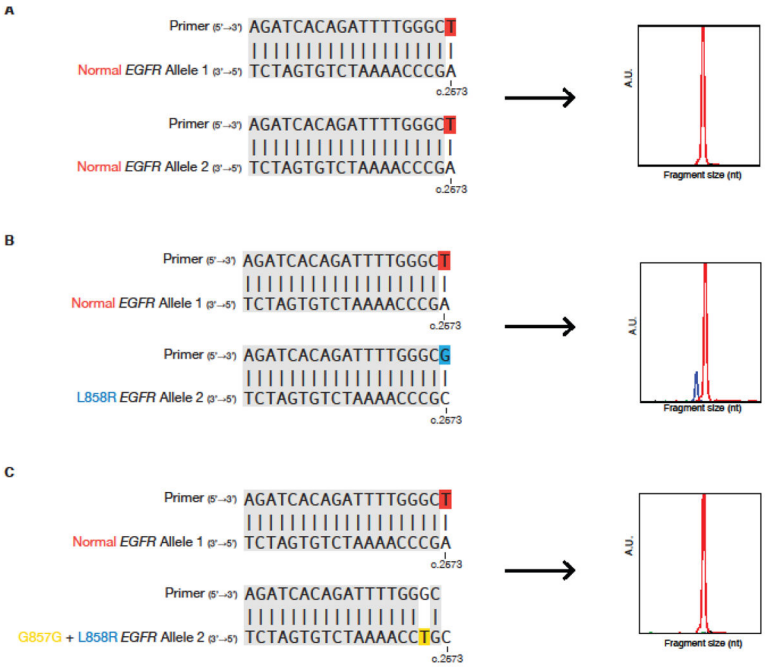


**Fig. 3. *EGFR* L858R allelic dropout diagram**

Primer-DNA hybridization of the last 19 3′-end nucleotides of the 'SNaPshot' extension primer and the resulting PCR fragment analysis plots are shown across several allelic contexts. Gray boxes represent primer-DNA complementary regions. (**A**) A normal control specimen with two wild type *EGFR* L858 (c.2573T) alleles will produce a single wild type PCR fragment (red). (**B**) A positive control specimen with a wild type *EGFR* L858 (c. 2573T) allele and a mutant 20% allele frequency *EGFR* L858R (c.2573T>G) allele will produce a wild type PCR fragment (red) and a mutant PCR fragment (blue). (**C**) A specimen with a wild type *EGFR* L858 (c.2573T) allele and an allele with G857G (c.2571G>A) (yellow) and L858R (c.2573T>G) in *cis* will disrupt the primer binding and single base pair extension reaction resulting in a single wild type PCR fragment (red).

**Table 1**

Specimen sequencing coverage metrics

| On-target capture rate | Uniquely mapping pairs | Median depth of coverage |
|---|---|---|
| 83.4% | 6,008,824 | 2401 |

**Table 2**

Summary of identified mutations in EGFR exon 21 capture region

| Gene | Coding mutation | Amino acid change | Genomic location | Mutation type | Depth of coverage | Variant supporting reads | Variant Allelic Frequency | Filtering criteria status |
|---|---|---|---|---|---|---|---|---|
| EGFR | c.2571G>A | p.G857G | Chr7:g.55259513 | Synonymous | 2586 | 619 | 23.9% | Rejected |
| EGFR | c.2573T>G | p.L858R | Chr7:g.55259515 | Missense | 2648 | 613 | 23.2% | Accepted |