PNAS
www.pnas.org

Supplementary Information for

YES1 amplification is a mechanism of acquired resistance to EGFR inhibitors identified by transposon mutagenesis and clinical genomics

Pang-Dian Fan, Giuseppe Narzisi, Anitha D. Jayaprakash, Elisa Venturini, Nicolas Robine, Peter Smibert, Soren Germer, Helena A. Yu, Emmet J. Jordan, Paul K. Paik, Yelena Y. Janjigian, Jamie E. Chaft, Lu Wang, Achim A. Jungbluth, Sumit Middha, Lee Spraggon, Huan Qiao, Christine M. Lovly, Mark G. Kris, Gregory J. Riely, Katerina Politi, Harold Varmus and Marc Ladanyi

Corresponding authors: Pang-Dian Fan, Harold Varmus, and Marc Ladanyi
Email: fanp@mskcc.org, varmus@med.cornell.edu, and ladanyim@mskcc.org

This PDF file includes:

Supplementary text
Figs. S1 to S4
References for SI reference citations

www.pnas.org/cgi/doi/10.1073/pnas.1717782115

Supporting Information

SI Materials and Methods

Library preparation and sequencing. *piggyBac* insertion sites were identified by generating Illumina-compatible libraries from DNA fragments that span the *piggyBac* sequence and the surrounding genomic DNA, using a modified TraDIS-type method(29). Genomic DNA (200 ng) from 188 isolated clones was sheared in a Covaris 96 microTUBE Plate to give fragments around 350 base pairs in length. End repair, A-tailing and ligation were performed using the Bioo NextFlex rapid DNA sequencing kit. Following ligation and removal of excess adapters, a two-step nested PCR protocol was performed to create sequencing libraries. For the first PCR (12 cycles), primers F-nest_1 (designed to bind the adapter sequence) and Pb-51 (designed to bind the *piggyBac* sequence) were used to specifically amplify fragments that contain the *piggyBac* end sequence. Products were purified, and re-amplified with F-X and N2-X primers (16 cycles) to append indices and library ends. Using a dual indexing strategy, the combination of 8 F-X primers and 12 N2-X primers allows 96 index combinations. The F-X primer binds to the adapter sequence, and adds the i5 index and the P5 end of the library. The N2-X primer binds to the *piggyBac* sequence at a position immediately adjacent and proximal to the binding site of the Pb-51 primer used in the first amplification. It also contains the i7 index sequence and the P7 sequence for binding to the Illumina flowcell. PCRs were performed with Phusion polymerase using high GC buffer. Sequencing was performed on a Miseq using a 2x75bp run with custom sequencing primers for index 1 and read 2. The PCR strategy was designed so that for read 2, the first 17 nucleotides sequenced (tatctttctagggttaa) correspond to the end of *piggyBac*, to enable unambiguous identification of *piggyBac* end sequences and precise delineation of insertion sites.

Primer sequences:

F-nest_1: TCTTTCCCTACACGACGCTCTTCCGATCT
Pb-51: CGCTATTTAGAAAGAGAGAGCAATATTTCA

F-X: AATGATACGGCGACCACCGAGATCTACACxxxxxxxxACACTCTTTCCCTACACGACGCTCTTCCGATCT
N2-X: CAAGCAGAAGACGGCATACGAGATxxxxxxAGAATGCATGCGTCAATTTTACGCAGAC

Custom index 1 sequencing primer: GTCTGCGTAAAATTGACGCATGCATTCT
Custom read 2 sequencing primer: AGAATGCATGCGTCAATTTTACGCAGAC

F-1: AATGATACGGCGACCACCGAGATCTACACTAGATCGCACACTCTTTCCCTACACGACGCTCTTCCGATCT
F-2: AATGATACGGCGACCACCGAGATCTACACCTCTCTATACACTCTTTCCCTACACGACGCTCTTCCGATCT
F-3: AATGATACGGCGACCACCGAGATCTACACTATCCTCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
F-4: AATGATACGGCGACCACCGAGATCTACACAGAGTAGAACACTCTTTCCCTACACGACGCTCTTCCGATCT
F-5: AATGATACGGCGACCACCGAGATCTACACGTAAGGAGACACTCTTTCCCTACACGACGCTCTTCCGATCT
F-6: AATGATACGGCGACCACCGAGATCTACACACTGCATAACACTCTTTCCCTACACGACGCTCTTCCGATCT
F-7: AATGATACGGCGACCACCGAGATCTACACAAGGAGTAACACTCTTTCCCTACACGACGCTCTTCCGATCT
F-8: AATGATACGGCGACCACCGAGATCTACACCTAAGCCTACACTCTTTCCCTACACGACGCTCTTCCGATCT

N2-1: CAAGCAGAAGACGGCATACGAGATTGGTCAAGAATGCATGCGTCAATTTTACGCAGAC

N2-2: CAAGCAGAAGACGGCATACGAGATCACTGTAGAATGCATGCGTCAATTTTACGCAGAC
N2-3: CAAGCAGAAGACGGCATACGAGATCTGATCAGAATGCATGCGTCAATTTTACGCAGAC
N2-4: CAAGCAGAAGACGGCATACGAGATTACAAGAGAATGCATGCGTCAATTTTACGCAGAC
N2-5: CAAGCAGAAGACGGCATACGAGATCGTACGAGAATGCATGCGTCAATTTTACGCAGAC
N2-6: CAAGCAGAAGACGGCATACGAGATCCACTCAGAATGCATGCGTCAATTTTACGCAGAC
N2-7: CAAGCAGAAGACGGCATACGAGATATCAGTAGAATGCATGCGTCAATTTTACGCAGAC
N2-8: CAAGCAGAAGACGGCATACGAGATGCCTAAAGAATGCATGCGTCAATTTTACGCAGAC
N2-9: CAAGCAGAAGACGGCATACGAGATCGTGATAGAATGCATGCGTCAATTTTACGCAGAC
N2-10: CAAGCAGAAGACGGCATACGAGATACATCGAGAATGCATGCGTCAATTTTACGCAGAC
N2-11: CAAGCAGAAGACGGCATACGAGATATTGGCAGAATGCATGCGTCAATTTTACGCAGAC
N2-12: CAAGCAGAAGACGGCATACGAGATAAGCTAAGAATGCATGCGTCAATTTTACGCAGAC

index sequences (replacing "X" in generic primer sequences above)
N2-1: TGGTCA
N2-2: CACTGT
N2-3: CTGATC
N2-4: TACAAG
N2-5: CGTACG
N2-6: CCACTC
N2-7: ATCAGT
N2-8: GCCTAA
N2-9: CGTGAT
N2-10: ACATCG
N2-11: ATTGGC
N2-12: AAGCTA

F-1: TAGATCGC
F-2: CTCTCTAT
F-3: TATCCTCT
F-4: AGAGTAGA
F-5: GTAAGGAG
F-6: ACTGCATA
F-7: AAGGAGTA
F-8: CTAAGCCT

Bioinformatics analysis. Illumina paired-end reads were aligned to the reference genome (hg19) using bwa mem v0.7.8 [1] with default parameters. PCR duplicates were marked using picard tools[1] v1.83 and were not used for downstream analysis. In order to generate a list of candidate *piggyBac* insertion sites we developed custom software (using the BamTools API[2]) to analyze the pattern of soft-clipped sequences in the aligned reads. Reads were extracted if the soft-clipped portion of the sequence (either forward or reverse complemented) contained a perfect match (≥ 13 bp) to the *piggyBac* sequence. A candidate insertion site was considered if all soft-clipped sequences (matching *piggyBac*) for a predefined minimum number of reads start at the same coordinate in the genome. Read orientation and alignment position of the mate are collected for all reads supporting the insertion site and used to infer the orientation of the insertion and compute statistics to remove false-positive candidates. For actual *piggyBac* insertion sites one would expect the insert size distribution, and the standard deviation (std) of insert sizes, to follow the expected distribution, measured by mapping the read pairs to the reference genome. However, due to ambiguity in read mapping or library artifacts, this may not hold. **Fig. S4A** and **Fig. S4B** show that most of the candidate insertions sites follow the expected insert size distribution inferred after alignment (mean=100,

[1] http://broadinstitute.github.io/picard/
[2] https://github.com/pezmaster31/bamtools

std=55) and **Fig. S4C** shows good correlation between mean and std of the fragment insert size. The exception is a suspiciously high number of sites characterized by a small std. These sites are likely to be false-positives and in fact the correlation analysis (**Fig. S4C**) shows that most of them have significantly higher mean insert size. Based on these results we compiled a list of high confidence insertion sites using a conservative set of filters: >10 supporting reads (matching the *piggyBac* sequence) and insert size std > 10. After filtering, our list contained 1927 distinct candidate insertion sites in 188 clones. We then focused our attention to genes with multiple independent insertion events.

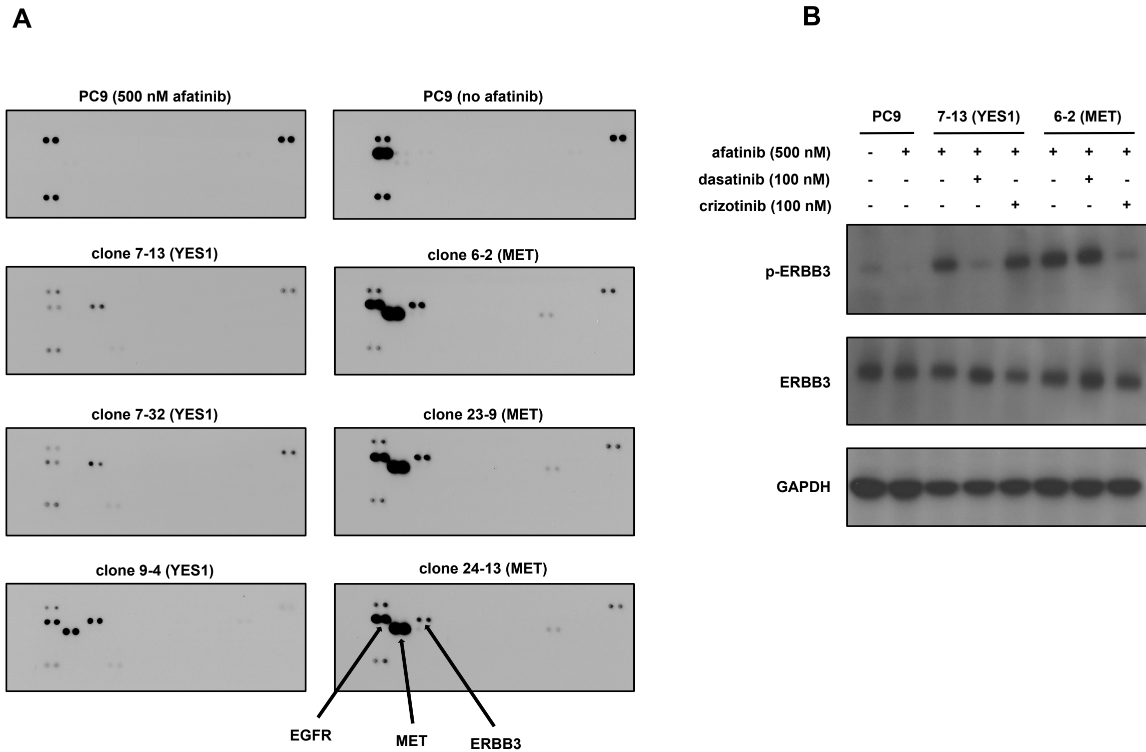

Fig. S1. Tyrosine phosphorylation of ERBB3 in *YES1* and *MET* clones. (A) Lysates from PC9 cells with and without 500 nM afatinib and from *YES1* and *MET* clones maintained in 500 nM afatinib were hybridized to human phospho-RTK arrays (R&D Systems, ARY001B). The pan phospho-tyrosine antibody for the array kit detects phosphorylation of MET in clone 9-4, presumably at a site different from the tyrosine recognized by the phospho-MET antibody used in Figure 1B. (B) Lysates from PC9 cells, clone 7-13 (*YES1*), and clone 6-2 (*MET*) treated with the indicated inhibitors for 60 minutes were subjected to immunoblot analysis with antibodies against the indicated proteins

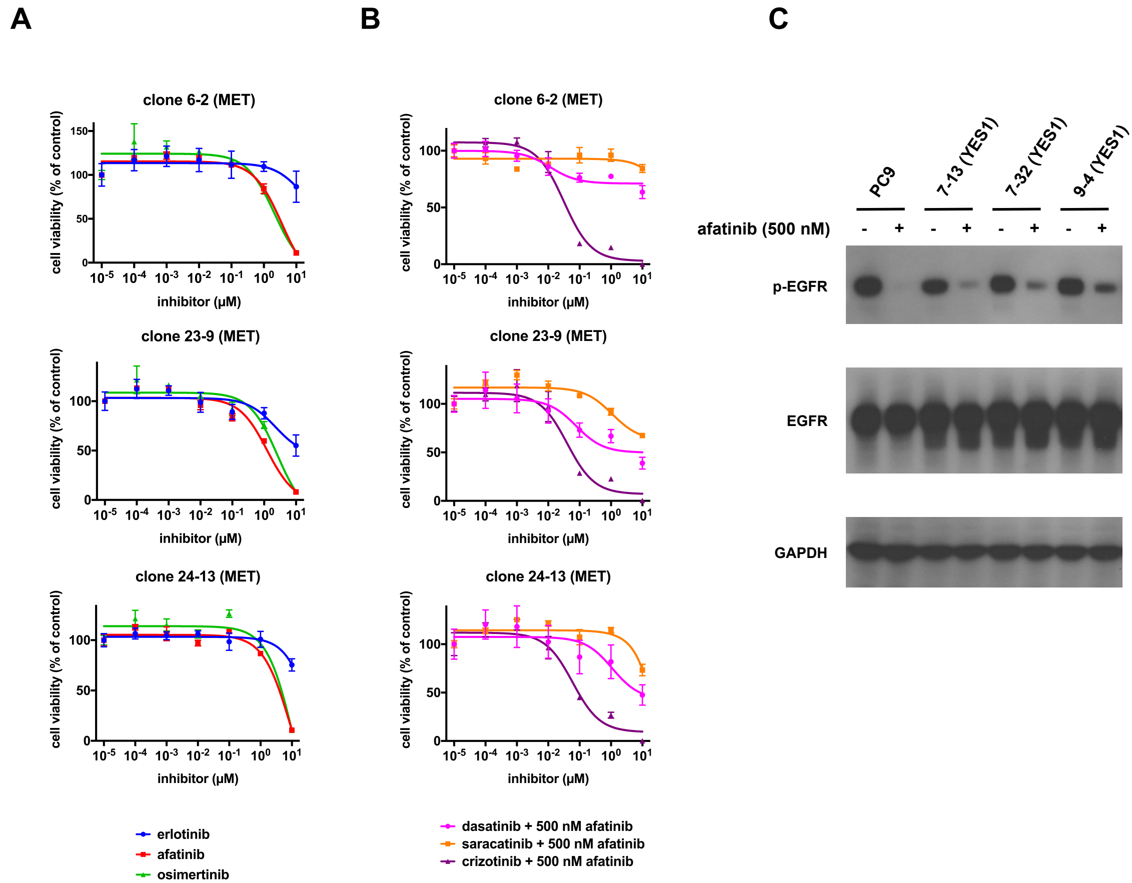

Fig. S2. *MET* clones are resistant to EGFR inhibitors from all three generations but sensitive to additional blockade of MET kinase activity. *YES1* clones retain EGFR kinase activity. (A) and (B) *MET* clones were seeded in 96-well plates and treated with EGFR inhibitors or the indicated inhibitors in combination with 500 nM afatinib for 96 hours. Cell viability was assayed as described in Methods. Data are expressed as a percentage of the value for cells treated with vehicle control and are means of triplicates. The experiments were performed 3 times with similar results. (C) PC9 cells were treated with and without 500 nM afatinib for 60 minutes. *YES1* clones were maintained in 500 nM afatinib or grown without afatinib for 72 hours. Cell lysates were subjected to immunoblot analysis with antibodies against the indicated proteins.

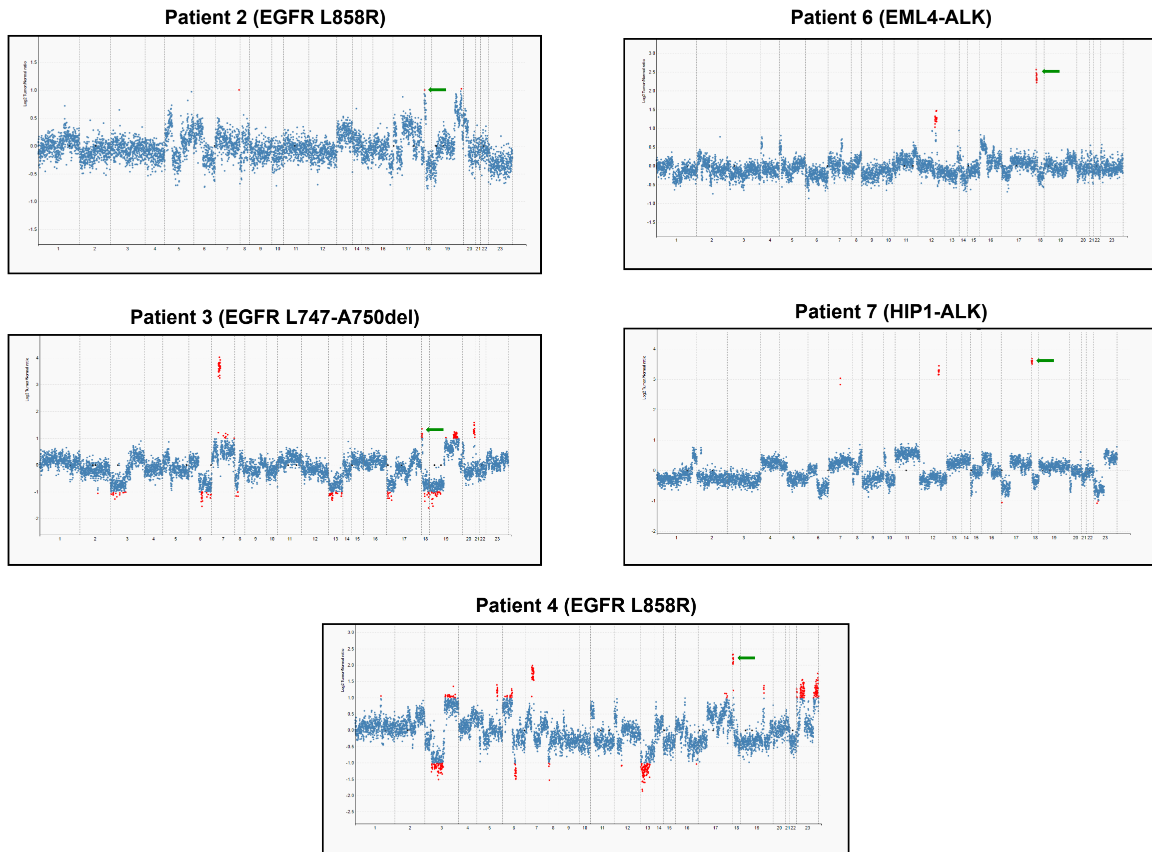

Fig. S3. Amplification of *YES1* in post-TKI tumor samples from patients with acquired resistance to EGFR and ALK inhibitors. Copy number plots for post-TKI samples from patients 2, 3, 4, 6 and 7. Each dot represents a target region in the MSK-IMPACT targeted capture assay. Red dots are target regions exceeding a fold change cutoff of 2-fold. The log-ratios (y-axis) comparing tumor versus normal coverage values are calculated across all targeted regions (x-axis). Green arrows indicate focal amplification of *YES1* (11 coding exons targeted).

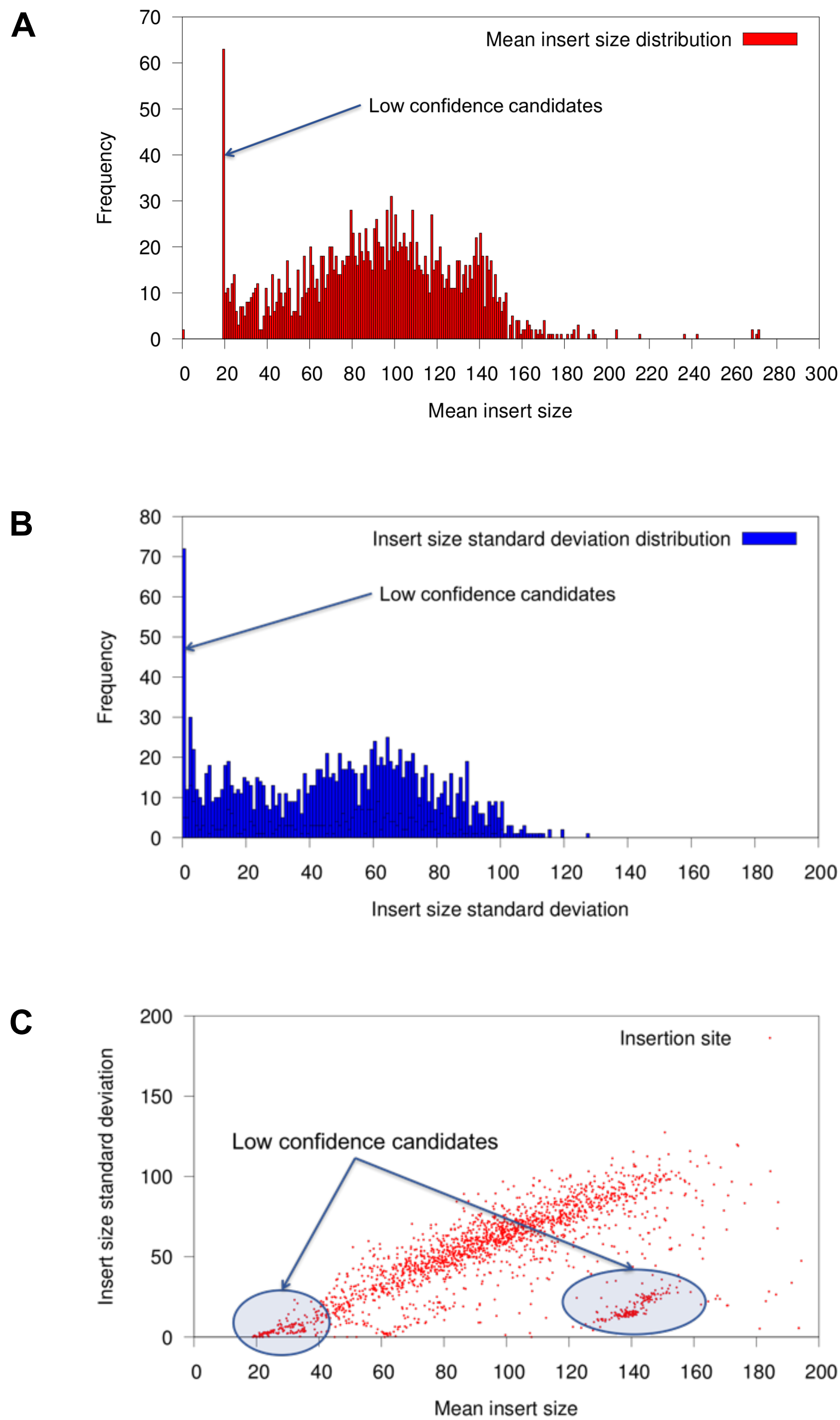

Fig. S4. Mean insert size and standard deviation of insert size for candidate insertion sites. (A) Mean insert size distribution of candidate insertion sites. (B) Standard deviation of insert sizes for candidate insertion sites. (C) Correlation between mean insert size and standard deviation of insert size for the candidate insertion sites.

References

1. Li H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v1 [q-bio.GN]