EGFR Variant Heterogeneity in Glioblastoma Resolved through Single-Nucleus Sequencing

Joshua M. Francis¹,², Cheng-Zhong Zhang¹, Cecile L. Maire², Joonil Jung¹, Veronica E. Manzo², Viktor A. Adalsteinsson¹,³,⁴, Heather Homer⁵, Sam Haidar², Brendan Blumenstiel¹, Chandra Sekhar Pedamallu¹,² Azra H. Ligon⁵,⁶,⁷,⁸, J. Christopher Love¹,³,⁴, Matthew Meyerson¹,²,⁸,⁹ and Keith L. Ligon²,⁵,⁶,⁷,⁸
ABSTRACT

Glioblastomas (GBM) with EGFR amplification represent approximately 50% of newly diagnosed cases, and recent studies have revealed frequent coexistence of multiple EGFR aberrations within the same tumor, which has implications for mutation cooperation and treatment resistance. However, bulk tumor sequencing studies cannot resolve the patterns of how the multiple EGFR aberrations coexist with other mutations within single tumor cells. Here, we applied a population-based single-cell whole-genome sequencing methodology to characterize genomic heterogeneity in EGFR-amplified glioblastomas. Our analysis effectively identified clonal events, including a novel translocation of a super enhancer to the TERT promoter, as well as subclonal LOH and multiple EGFR mutational variants within tumors. Correlating the EGFR mutations onto the cellular hierarchy revealed that EGFR truncation variants (EGFRvIII and EGFR carboxyl-terminal deletions) identified in the bulk tumor segregate into nonoverlapping subclonal populations. In vitro and in vivo functional studies show that EGFRvIII is oncogenic and sensitive to EGFR inhibitors currently in clinical trials. Thus, the association between diverse activating mutations in EGFR and other subclonal mutations within a single tumor supports an intrinsic mechanism for proliferative and clonal diversification with broad implications in resistance to treatment.

SIGNIFICANCE: We developed a novel single-cell sequencing methodology capable of identifying unique, nonoverlapping subclonal alterations from archived frozen clinical specimens. Using GBM as an example, we validated our method to successfully define tumor cell subpopulations containing distinct genetic and treatment resistance profiles and potentially mutually cooperative combinations of alterations in EGFR and other genes. Cancer Discov; 4(8); 956–71. ©2014 AACR.

See related commentary by Gini and Mischel, p. 876.

INTRODUCTION

Glioblastoma (GBM), also known as diffuse astrocytoma WHO grade IV, is the most common primary brain tumor in adults and a leading example of a tumor containing a high degree of cellular heterogeneity (1). Recurrent genetic alterations in GBM have been identified (2–4), with frequent amplification of several receptor tyrosine kinases (RTK) including EGFR, PDGFRA, and MET. Genomic heterogeneity in tumors of the central nervous system, especially GBM, is an emerging problem with substantial ramifications for treatment strategies (5, 6). One of the most prominent examples is the mosaic pattern of different RTK amplifications present in different subpopulations of tumor cells (7, 8).

As the most frequent RTK amplification in glioblastoma, EGFR is focally amplified in about half of all primary glioblastomas and in 95% of the classic subtype (9). Recent reports by The Cancer Genome Atlas (TCGA) revealed unprecedented complexity at the focally amplified EGFR locus within individual tumors, including clusters of chromosomal breakpoints (10, 11), and the expression of multiple EGFR variants and extracellular domain missense mutations (4, 12). However, outside of the EGFRvIII variant and carboxyl-terminal truncations, it is currently unknown whether many such variants are oncogenic and how such variants coexist at a single-cell level (13–15). Such patterns are important with regard to whether aberrations may functionally cooperate within individual cells or between cells with differing mutations. The functional importance of such distinctions is underscored by recent work showing the differential effects and interactions of variant receptors on EGFR activation within individual cells (16). Furthermore, extrachromosomal amplicons containing EGFRvIII (a constitutively active form of EGFR resulting from the deletion of exons 2–7) were recently shown to be dynamically lost and/or silenced in response to EGFR inhibitor treatment and then to re-emerge following removal of this selective pressure (5). The expression of EGFRvIII is also known to be regionally heterogeneous by IHC and RTPCR, and it has been proposed that there may be functional synergy between tumor cells expressing the oncogenic variant (EGFRvIII) and those with the slower-growing counterpart (EGFRwt; refs. 5, 14, 17). However, whether this heterogeneity...
is due to differences in the genomes of single cells has not been fully assessed.

Traditional statistical analysis of bulk average sequencing data cannot resolve the complexity at the EGFR locus because both the high copy number and the nonmitotic segregation of EGFR variants present in extrachromosomal amplicons prevent accurate clonality inference (18, 19). Although FISH can assess the amplification of different RTK genes (EGFR, PDGFR, and MET), it is difficult to distinguish between multiple variants of these oncogenes by FISH, as it is generally limited to the characterization of known alterations spanning large genomic regions, and cannot probe single-nucleotide variants (SNV), short deletions, or discover de novo alterations.

Single-cell sequencing has great potential to dissect coexisting genomic alterations. Here, we describe a population-based framework for single-cell genomic analysis and demonstrate its potential by dissecting the subclonal populations in two primary glioblastomas with focal EGFR amplifications. Our methodology uses a novel approach for confident inference of LOH and enables unambiguous characterization of the subclonal tumor cell populations with implications for EGFR-targeted therapies in brain cancer and other cancers.

RESULTS

Multiple Somatic Events Can Target EGFR in a Single Glioblastoma

To characterize the frequency of compound alterations of EGFR in glioblastoma, we examined the RNA-sequencing data of 76 cases of glioblastoma from TCGA with focally amplified EGFR (4, 12), and observed that 71% (54 of 76) had expression of wild-type (WT) EGFR along with at least one aberrant EGFR variant, with 30% (23 of 76) expressing two or more variants. In 34% (28 of 76) of tumors, missense mutations involving the EGFR extracellular domain coexisted with structural variants. We then performed an exhaustive search of structural rearrangements of EGFR from whole-genome sequencing data in the 25 publicly available cases. Joint analysis of the DNA and RNA-sequencing data not only confirmed the presence of multiple variant EGFR transcripts, but also further revealed that the transcript encoding identical oncogenic EGFR variants could have been generated by distinct alterations at the DNA level.

As an example, our analysis suggested that TCGA-19-2624 had expression of wild-type EGFR together with both EGFRVIII (deletion of exons 2–7) and EGFRVII (deletion of exons 14–15). At the DNA level, two distinct genomic rearrangements were found to give rise to EGFRVIII, and two other events were found to generate EGFRVII (Fig. 1A). Coexpression of EGFRVII and EGFRVIII was previously described (20), but the frequency of variant reads suggested that EGFRVIII is more prevalent in this sample. In addition, an extrachromosomal domain missense mutation, G63K, was also observed at the DNA and RNA level. On the basis of the five EGFR aberrations alone, TCGA-19-2624 could exist as 32 possible combinations of cellular clones (21). Similar scenarios were discovered in additional samples (Supplementary Fig. S1A–S1D and Supplementary Table S1), with an overall prevalence of 44% (7 out of 16 with focal EGFR amplifications) harboring multiple EGFR mutations, including both rearrangements and single-nucleotide substitutions. Bulk analyses cannot resolve whether such heterogeneity commonly occurs in a single tumor clone, mutually exclusive clones, or other complex combinations of multiple, yet functionally cooperative variants (Fig. 1B). To answer this question, we sought to assess the clonal structure of EGFR aberrations by single-nucleus sequencing (SNS).

SNS Approach

We selected for study two primary glioblastomas (BT325 and BT340) with focal EGFR amplification present as extrachromosomal amplicons based on prior clinical analysis. Bulk tumor and SNS were performed to examine tumor heterogeneity within a single sampled region taken from each tumor (~1 cm³ in size).

The experimental and analytical framework is summarized in Fig. 2. Single nuclei (SN) from tumor resections were isolated by flow cytometry and a limited multiple displacement amplification (MDA)–based whole-genome amplification was performed using a protocol that had been optimized for even genome coverage (see Supplementary Methods). Approximately 100 multiplexed SNS libraries were constructed and sequenced to low depth (~0.01×) for quality evaluation (Supplementary Methods and Supplementary Fig. S2A and S2B). The quality assessment of single-cell DNA libraries requires an entirely new system of metrics than those used for bulk genomic DNA sequencing due to inherent challenges associated with starting with only two alleles of template DNA (21–23). Here, we described two metrics, both of which can be estimated from low-pass sequencing (median depth, ~0.01×): one directly measures the amplitude of overamplification (whole-genome amplification pileup metric) and the other measures the unevenness of amplification (modified from Zong and colleagues; ref. 21). Both metrics showed consistent results (Supplementary Fig. S2C).

The 50 to 60 best SN libraries were selected and sequenced to a combined depth of 30×. In parallel, whole-genome sequencing was performed on dissociated nuclear DNA from the tumor to serve as the bulk tumor reference and on matched blood DNA as germline reference. We performed joint analysis of SN and bulk DNA sequencing data to infer and guide the clonality estimates of somatic mutations and the subclonal tumor populations. In comparison with prior studies, which examined only the single-cell genomes, performing joint analysis allowed us to more rigorously establish and confirm the presence of subclonal alterations using methods that have been previously established for bulk genome sequencing and validate findings from SNS.

EGFR Copy Number Is Highly Variable between Single Cells

Sequencing of the bulk BT325 tumor revealed the presence of a 935-kb amplicon containing four genes, including EGFR, with an additional deletion corresponding to EGFRVIII (Fig. 3A and Supplementary Fig. S3A). Bulk tumor analysis suggested an approximate 50% reduction in read density within this region, consistent with a mixture of amplified wild-type EGFR and EGFRVIII deletion. A total of 60 SN libraries from BT325 were sequenced and all the tumor-derived nuclei harbored the EGFR amplification as...
Figure 1. Genomic complexity of the EGFR locus in GBM cannot be resolved using bulk tumor sequencing. A, RNA sequencing (RNA-Seq) data from GBM patient TCGA-19-2624 revealed coexistence of multiple EGFR aberrations, including EGFRvIII (16% frequency), EGFRvII (2% frequency), and a G63K mutation (5% frequency); analysis of whole-genome sequencing further revealed the presence of four distinct intragenic rearrangements producing two variants of EGFRvIII and two variants of EGFRvII, each with different allelic frequencies as indicated by read counts. B, the five distinct EGFR aberrations could result in \(2^5 = 32\) possible clone patterns at the cellular level. Two extreme cases are that all five variants are present in all cells (possibility 1) or that the five variants each reside in different cells (possibility 3); alternatively, the five variants can exist in up to 26 unique combinations at the cellular level (possibility 2).

Figure 2. Experimental and analytic workflow of SN and bulk tumor sequencing for the characterization of tumor heterogeneity. WGA, whole genome amplification; WGS, whole-genome sequencing; WES, whole-exome sequencing; QC, quality control; BL, blood; TP, tumor primary.
Figure 3. Heterogeneities in the copy number and mutation status of focally amplified EGFR in two glioblastomas. **A**, read coverage plots of BT325 showing concurrent amplification of EGFR wild-type and EGFRvIII (bulk average). SNS plots show the same features as bulk but each cell varies in the ratio of EGFR wild-type and EGFRvIII amplification. **B**, quantified levels of total EGFR (exons 1, 8–28) and wild-type EGFR (exons 2–7) across all tumor nuclei using read depths from these regions. EGFRvIII (total minus wild-type) is coamplified with wild-type in all tumor nuclei, but to different degrees. The copy numbers of representative SN libraries shown in **A** are denoted by † and *. **C**, FISH for total EGFR and the chromosome 7 centromere (CEP7) confirming varying total EGFR copy-number levels in BT325. SureFISH assay (far right nuclei) designed to specifically detect retained and deleted exons shows that individual cells range from low (top) to high levels of vIII deletion (bottom). **D**, read coverage plots of BT340 showing two overlapping deletions encompassing exons 14 and 15 (bulk average). Three representative SN libraries (lower tracks) show different mutation patterns. Cell A contained only EGFRvIII truncation, cell B contained only the longer EGFRvIII (vIII-extended), and cell C contained only wild-type EGFR. **E**, quantified levels of total EGFR (exons 1–13 and 17–28) and wild-type EGFR (exons 14–15) read depth within single tumor nuclei demonstrate three distinct populations containing amplified EGFRvIII, EGFRvIII-extended, or wild-type EGFR. Discordant read pair signatures confirmed the presence of vIII or vIII-extended variants in population A and B; no discordant pairs were found in population C. Asterisks denote the representative nuclei displayed in **D**.
Single-Nucleus Sequencing of GBM Mutational Heterogeneity

Large differences in total EGFR copy number showed an extremely broad range, from five copies to more than 200 copies in different nuclei, and was assessed by discordant read pairs spanning C-terminal breakpoints that occur in the presence of wild-type EGFR.

Multiple EGFRvIII Variants Exist in Nonoverlapping Cell Populations

We next performed sequencing of the BT340 tumor. This tumor harbored a 535-kb amplicon containing the EGFR locus (Supplementary Fig. S3B). Bulk sequencing showed a marked decrease in read density in the region of exons 14 and 15 indicative of EGFRvII (Fig. 3D, top). An additional region of decreased read density indicative of a 7.6-kb deletion that included alternative exon 16 was also found. Discordant read pairs in the bulk tumor sequencing data corroborated the presence of these deletions and defined their precise boundaries as being distinct (Supplementary Fig. S3C). We denoted this novel deletion variant as EGFRvII-extended.

Examination of SNS libraries showed three distinct cell populations with respect to exons 14 to 16 of the EGFR locus (Fig. 3D). EGFRvII and EGFRvII-extended segregated into two populations based on read depth and discordant read pairs corresponding to the rearrangements (Fig. 3D, cell A and cell B). Cells that did not harbor any discordant read pairs supporting either EGFRvIII or EGFRvII variants were inferred to retain these exons (Fig. 3D, cell C). The initial amplicon was derived from a single copy of the wild-type allele (allele A) as assessed by germline SNVs; however, the exact number of copies per amplicon cannot be determined.

We further characterized the copy number of different EGFR variants within each individual cell. The total EGFR copy number was estimated from the read depths in exons 1 to 13 and in exons 17 to 28, and the wild-type EGFR copy number was estimated from the read depth in exons 14 and 15. In 35 of the 48 SN libraries, we were able to confidently assess the absolute number of EGFR copies. The total EGFR copy number showed an extremely broad range, from five copies to more than 200 copies in different nuclei, and was consistent with our FISH analysis (Fig. 3E, gray bars and Supplementary Fig. S3D). From 35 nuclei, we identified 26 containing EGFRvII and three EGFRvII-extended with no overlap. Further examination of the other six nuclei within the non-vIII cluster revealed amplified EGFR, with three nuclei showing discordant read pairs spanning C-terminal breakpoints that occur in the presence of wild-type EGFR.

Determination of EGFRvII Evolutionary Hierarchy Based on Tumor Subclones

The common genomic boundaries of the overall EGFR amplicon in all three populations in BT340 suggested that the clonal diversification occurred after the common ancestral EGFR amplification. Although the vII-deleted and the vII deletion existed in distinct cell populations, their clonal association could not be unambiguously determined (e.g., the larger vII-deleted deletion may theoretically be derived from the smaller vII). We therefore sought to characterize the clonal hierarchy based on coexisting alterations unrelated to EGFR to infer the clonal origin of the two EGFR variants.

SNS allows one to reconstruct the lineage evolution history in tumors using copy-number profiling and single-nucleotide variant clustering (24–27). However, the generic clustering approach previously used in these studies cannot identify the exact clonal or subclonal alterations that differentiate subclonal populations as in bulk analysis (19, 28), and can be confounded by artifacts that are introduced during whole-genome amplification (22). To circumvent this limitation, we developed and applied an integrative approach to identify distinct tumor populations based on joint detection of clonal and subclonal events from the bulk tumor and SN genomes. Of particular importance, by requiring true variants to be present in at least two SN libraries, or with evidence in the bulk tumor DNA, we were able to exclude the majority of random MDA artifacts generated within individual libraries, and, at the same time, increase detection sensitivity for subclonal events that are present in the bulk library at low allelic frequencies.

Profiling Clonal Aberrations Increases Sensitivity for Detection of Subclonal Alterations

We first focused on clonal events inferred from bulk tumor analysis that could be used as markers for SNS analysis. In both BT340 and BT325, we identified characteristic glioblastoma somatic copy-number alterations (SCNA), including homozygous loss of CDKN2A/CDKN2B, hemizygous loss of chromosome 10, as well as gains of chromosome 7 (Fig. 4A and Supplementary Fig. S4A and S4B). BT325 was found to have a common activating mutation within the promoter region of TERT (chr 5:1295228, G→A) that is observed in >80% of glioblastomas (4, 29–31). In BT340, we observed a nonreciprocal translocation of a putative super enhancer within chromosome 10 (32) to the TERT promoter that could be an alternative mechanism for activating TERT in glioblastoma (Fig. 4C).

We then characterized the SNS libraries of BT340 according to the clonal deletion events. In regions of hemizygous deletion or LOH, SN should harbor only one consensus haplotype. We used this feature to cluster individual nuclei based on the differences in their observed genotypes in LOH regions and infer the consensus haplotype within these regions (Supplementary Fig. SSA–SSH and Supplementary Methods). Using the germline heterozygous sites to infer the haplotype information from chromosome 10, we clustered the 48 nuclei derived from BT340 into two groups (Fig. 4D). The 44 tumor-derived nuclei (black circles) showed minimal
presence of the deleted haplotype (whole-genome amplification and sequencing errors), and the four normal diploid nuclei (open circles) exhibited a roughly 1:1 ratio between the deleted and the retained haplotypes. The clustering of nuclei was further confirmed by haplotype analysis in hemizygously deleted regions in 9p and from the average copy ratio of chromosome 7, which is polysomic in tumor but disomic in normal nuclei (Supplementary Fig. S6A–S6C).

Haplotypeing of 60 SNS libraries derived from BT325 based on deleted regions in chromosome 10 and chromosome 9p identified 56 libraries as tumor derived (Supplementary Fig. S7A–S7C, middle). Of the other four libraries, two showed an approximate 1:1 ratio between the deleted and the retained haplotypes, consistent with having been derived from normal diploid cells (open circles); the other two exhibited a 1:2 ratio between the deleted and the retained haplotype, implying a 1:1 mixture of tumor and normal nuclei (Supplementary Fig. S7D and S7E).

The haplotype-based method for clustering SNS libraries allowed for unprecedented sensitivity and specificity by using a common event in cancer (loss of heterogeneity) rather than sequencing read depths that can be grossly distorted by whole-genome amplification. Moreover, haplotype-based clustering uses the digital nature of single-cell genotypes (in contrast with allele frequencies in bulk sequencing data) and the clonal structure of tumor populations to achieve high accuracy at low sequencing depths (~1x) even for libraries with high amplification bias.

**Complex Chromosomal Rearrangements Drive Diversification**

Chromothripsis has been observed to be a frequent event in glioblastoma and often results in the disruption of regions that undergo recurrent somatic copy-number alterations (10, 33). In both BT325 and BT340, we observed that homozygous inactivation of CDKN2A/CDKN2B resulted from two events, one a simple deletion and the other a complex genomic rearrangement event resembling chromothripsis. In BT325, the chromothripsis event involved chromosomes 1, 3, 8, and 9 and resulted in three derivative chromosomes, one of which was subsequently amplified (Supplementary Fig. S7F). In a 1.25-Mb region encompassing the CDKN2A locus, we observed alternating copy-number events between three states (0, 1, and 2) from the bulk read depth characteristic of chromothripsis. The SNS coverage captured the retained regions after chromothripsis, and the signal was not influenced by the background contribution of contaminating normal cells. In BT340, a complex double-minute amplification involving multiple segments of chromosome 6 and chromosome 9 resulted in homozygous inactivation of CDKN2A/2B (Supplementary Fig. S8A–S8E).

**Subpopulations of BT325 Are Distinguished by a Chromothricptic Event**

Clonality analysis of SCNA events from bulk tumor sequencing of BT325 suggested three subclonal deletions within chromosomes 1, 4, and 6 at a copy ratio of approximately 0.61 (Supplementary Fig. S4A and S4B). Zooming into these regions, we observed alternating copy-number alterations involving chromosomes 1, 4, and 6 that are characteristic of chromothripsis (Supplementary Fig. S9A–S9E). Of the 56 tumor-derived nuclei, 46 nuclei (red circles) harbored hemizygous deletions in all the chromothriptic regions,
whereas six nuclei (blue circles) were heterozygous at these loci (Supplementary Fig. S9B). Four additional SNS libraries had a haplotype ratio suggesting a 1:1 mixture of the tumor clones (magenta circles). This “all-or-none” scenario of all deletions clearly established that these events had been acquired within a single event as implied in chromothripsis.

On the basis of the presence of shared clonal features such as EGFRvIII, TERT promoter mutations, and loss of chromosome 10, but the absence of the chromosome 1, 4, and 6 chromothriptic event in six nuclei, we infer that the minor subclone split from the major subclone at this point (Supplementary Fig. S9F).
Subclonal Deletions Define Two Subpopulations in BT340

Analysis of SCNAs from the BT340 bulk tumor also showed copy ratios of approximately 0.20, 0.64, and 0.86, corresponding to noninteger copy numbers, indicating that such deletions were likely subclonal (Fig. 4A and B).

Regions of subclonal LOH were observed in chromosome 6p11.2 from 53.6 to 57 Mb, in 6q12 from 63.3 to 66.1 Mb, and in 9p21.1 from 27.1 to 30.9 Mb in the bulk tumor sequencing (Fig. 5A and Supplementary Fig. S6D). Clustering analysis of the 44 tumor-derived nuclei at these regions of chromosomes 6 and 9 resulted in two unique populations (Fig. 5B and Supplementary Fig. S6D). Cluster 1 (pink circles), consisting of 35 nuclei, was characterized by hemizygous deletions in 6p11.2 and 6q12 and homozygous deletions in 9p21.1 (27.1–30.9 Mb, 31.8–32 Mb). The second cluster of nine nuclei (green circles) was characterized by the retention of segments from 6p11.2 and 6q12 and hemizygous deletion of segments from 9p21.1. The haplotype ratio between the deleted (D) and the retained (R) in hemizygously deleted regions in 6p11.2 confirmed the hemizygous loss of these regions in cluster 1.

We then compared the subclonal copy numbers statistically inferred from the bulk analysis to the clonal fractions from single-cell analysis. The bulk average allelic copy number of the subclonal deletion in 9p21.1 is approximately 0.2 after correcting for tumor purity (90%) and ploidy (2.1). Clustering of single tumor nuclei found that 20% (9 of 44) retained these regions and 80% (35 of 44) harbored the deletion (Fig. 5C). The same agreement was observed for regions in chromosome 6 in which the bulk average copy number is approximately 0.70. Finally, the copy ratio of 0.86 from the bulk analysis, corresponding to a subclonal deletion of chromosome 16 at approximately 19% clonal fraction, was associated to 7 out of 35 nuclei within cluster 1 with hemizygous loss (Supplementary Fig. S6E). The excellent concordance between the bulk clonality analysis and the single-cell analysis highlights the power of our population-based approach to robustly and efficiently resolve heterogeneous subclonal populations even at low clonal frequencies. The consistency between the clonality analysis of bulk sequencing and population analysis of SN on not only validates the statistical rationale behind bulk tumor analysis (34), but further sets a standard paradigm for single-cell–based tumor analysis. Furthermore, single-cell analysis clearly resolved that the subclonal chromosome 16 loss occurred in a different population than the subclone that retained segments from chromosomes 6 and 9, although their clonal fractions are similar (~19% and ~20%). Thus single-cell sequencing provides an unambiguous solution to associate independent subclonal alterations with distinct subclonal populations.

EGFRvII Variants Reside within Distinct Clones

Another application of single-cell sequencing in dissecting heterogeneous tumors is demonstrated by correlating alterations in amplified loci with other subclonal events, which is impossible to infer from bulk tumor sequencing data. Nucleus-by-nucleus comparison of EGFR mutation status to the clonal hierarchy constructed from subclonal deletions above revealed that EGFRvII and EGFRvII-extended were exclusively associated with only one of the two subclonal populations: EGFRvII was present only in cluster 1 with 6p11.2 and 6q12 deletions, whereas EGFRvII-extended was present only in cluster 2 that retained these chromosomal segments (Fig. 5D). In addition, cluster 2 also contained cells with a known oncogenic C-terminal truncation (ex. 25–28; ref. 15) and a second C-terminal deletion (ex. 26–28), all of which occur in the context of wild-type EGFR amplification. Three cells from cluster 1 and two cells from cluster 2 were found not to contain focally amplified EGFR and, thus, could not be assigned as being a variant. In summary, 89% (39 of 44) of tumor cells were found to contain EGFR variants, with EGFRvII being the most common, followed by EGFRvII-ext., EGFR (del25–28), and EGFR (del25–26). Remarkably, all EGFR variants were mutually exclusive, suggesting that the EGFR mutations were most likely independently derived and expanded. In particular, the convergent evolution of EGFRvII variants suggested its oncogenic potential.

EGFRvII Is Oncogenic in the Absence of EGFRvIII

EGFRvII was initially described in GBM two decades ago, but was reported to be nononcogenic (35, 36). Subsequent studies identified the presence of EGFRvII coexisting with EGFRvIII and, thus, EGFRvIII was inferred to be the driving oncogenic variant (20). However, in TCGA-19-2624, the EGFRvII variant was found to be more prevalent than EGFRvIII at both the DNA and RNA levels, arguing in favor of its functional relevance. Furthermore, the convergence of independent clonal diversifications in BT340 resulting in EGFRvII and EGFRvII-extended further underscored its oncogenic potential. We therefore assessed the transformation potential of EGFRvII and the previously characterized EGFRvIII using two model systems (Fig. 6A).

Ba/F3 cells are dependent upon supplemental IL3 for survival but can be transformed to ligand independence by expression of oncogenic EGFR mutants (37, 38). Ectopic expression of EGFRvII and EGFRvIII permitted Ba/F3 cells to grow in the absence of IL3, whereas cells containing the empty parental vector or wild-type EGFR did not survive (Fig. 6B). Constitutive phosphorylation of EGFR (Y1068) and Akt (S473) was observed in serum-starved EGFRvII- and EGFRvIII-expressing cells, demonstrating ligand-independent activation (Fig. 6C).

To evaluate the oncogenic potential of the EGFRvII variant in a genetically relevant model of GBM, we expressed EGFR variants in E14.5 neural stem cells derived from Ink4a/Arf−/− mice (39). All variants demonstrated constitutive phosphorylation of EGFR as compared with wild-type EGFR-expressing cells (Fig. 6D). Consistent with published reports, the expression of EGFRvIII induced tumor growth when subcutaneously grafted in nude mice (Fig. 6E; ref. 13). Expression of EGFRvII was also tumorigenic, albeit tumors developed at a slower rate (11 weeks for EGFRvII compared with 6 weeks for EGFRvIII; Fig. 6E and F). Histologically, EGFRvIII and EGFRvII both generated high-grade gliomas, but we observed that the EGFRvII tumors exhibited oligodendrogial features.
Single-Nucleus Sequencing of GBM Mutational Heterogeneity

RESEARCH ARTICLE

A
Extracellular domain (ECD)

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>II</td>
<td>III</td>
<td>IV</td>
</tr>
<tr>
<td>VIII (J6-273)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VII (J521-603)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Kinase domain (KD)


B
IL3-independent proliferation


C


D


E


F

Figure 6. EGFRvII transforms Ba/F3 and Ink4a/Arf−/− neural stem cells. A, schematic representation of the EGFR vii and vili mutants. B, proliferation of Ba/F3 cells stably containing empty vector or expressing EGFR wild-type, EGFRvII, or EGFRvIII were plated in the presence or absence of IL3. Proliferation was measured over 5 days. C, immunoblot of Ba/F3 cells stably expressing empty vector, EGFR wild-type, EGFRvII, or EGFRvIII. Cells were serum starved for 3 hours and then stimulated with 25 ng/mL of EGF for 10 minutes. D, immunoblot of Ink4a/Arf-null neurospheres stably expressing EGFR wild-type, EGFRvII, or EGFRvIII. E, xenografted Ink4a/Arf-null neurospheres stably expressing EGFRvII or EGFRvIII form subcutaneous tumors after 6 or 11 weeks, respectively. F, measurement of the tumor volumes of subcutaneous xenografts stably expressing EGFRvII or EGFRvIII after 6 or 11 weeks, respectively. Cells expressing wild-type EGFR failed to form tumors.

and a corresponding lower rate of proliferation. These findings suggest that these similar mutations produce different functional effects.

To examine the drug sensitivities of EGFRvII and EGFRvIII, transformed Ba/F3 cells were tested for sensitivity to a panel of EGFR inhibitors. Ba/F3 cells dependent on EGFRvII were killed by erlotinib, lapatinib, afatinib, and neratinib (Supplementary Fig. S10A–S10D), with the irreversible inhibitors afatinib and neratinib exhibiting the lowest IC50 values for both EGFRvII- and EGFRvIII-expressing cells (Supplementary Fig. S10C and S10D). These results are similar to previous reports for EGFRvIII-expressing Ba/F3 cells (38), and suggested that EGFRvIII may be a therapeutic target for kinase inhibition.

DISCUSSION

Our identification of subclone-specific rearrangements within the EGFR gene expands upon the recent discoveries of heterogeneous genomic alterations in oncogenic kinase drivers in glioblastoma (7, 8, 14, 40). These results further suggest that a substantial fraction of the multiple EGFR variants reported in the TCGA glioblastoma collection may represent distinct subclonal populations within the tumor.

High-resolution genomic approaches have begun to reveal genomic complexity and heterogeneity to a degree not previously understood. These methods include clonality analysis of bulk tumor sequencing (18, 28, 41), regional sampling of tumors (6, 42, 43), and SNS (24–26). Among these techniques, SNS has the unique power to resolve cellular heterogeneity, but can be technically challenging and difficult to interpret. Our methods for LOH detection by allele-specific delineation of haplotypes address limitations in genome coverage due to MDA and expand upon the applications of single-cell DNA analysis. As single-cell sequencing studies expand in size and scope, our methods provide the basis for maximizing sequencing information, given the numbers...
Figure 7. Summary of cellular genomic heterogeneity in glioblastoma. Left, different RTK amplifications can lead to clonal diversification after common clonal mutations. Right, mutations in a single RTK can also lead to clonal diversification through multiple variants, SNVs, and amplification of these different alleles. The presence of multiple concurrent RTK amplifications/mutations implies potential need for multiple drugs targeting the same RTK via different mechanisms.
of cells sequenced. Clonality inference by combining bulk tumor analysis with single-cell population analysis further provides a unique strategy for resolving clonal evolution and coassociation of independent genomic alterations at the single-cell level. The combination of SNS and joint calling with bulk sequencing allowed us to detect minor subclonal populations that were present in as low as 10% of the tumor at relatively low sequencing depth.

Genomic diversification in cancer is known to contribute to targeted and cytotoxic therapy failure and emergence of resistance (5, 44, 45). Large-scale genomic studies of cancer have outlined sets of alterations that affect growth-promoting and tumor-suppressing pathways that are necessary to drive tumorigenesis. In glioblastoma, the RB, TP53, and PTEN tumor suppressors are frequently inactivated directly or indirectly through aberrations in pathway factors such CDKN2A/CDKN2B or MDM2 (4). These genomic alterations serve as the foundation upon which RTKs like EGFR, MET, and PDGFRα can become amplified and evolve as the tumor begins to compete with each other for growth advantage. As FISH studies have demonstrated, RTKs within a tumor can be amplified alone or in combination (Fig. 7, left; refs. 7, 8). Presumably, clones with diverse growth signaling pathways will gain an advantage and expand as competition ensues.

Here, our study reveals an additional paradigm for the diversification of growth-signaling pathways through the acquisition of multiple variants of a single RTK gene (Fig. 7, right). The convergence of subclone-specific EGFRvIle deletions reflects the dynamic evolution of amplified EGFR copies and pinpoints its oncogenic potential. We further demonstrated that overexpression of EGFRvIle does cause transformation in both Ba/F3 and Ink4a/Arf−/− neural stem cell models, a variant previously thought not to be oncogenic because it co-occurred with EGFRvIII (20, 35). TCGA RNA sequencing even revealed the expression of EGFRvII in 9% (7 of 76) of cases with focally amplified EGFR (4). Our functional characterizations found that constitutive expression of EGFRvIle results in downstream activation of AKT signaling, consistent with that of EGFRvIII, but not with enhanced ERK activation. It is possible that EGFRvII activates an alternative pathway such as STAT3/5 and poses a more direct means of inducing transcriptional changes (16). Therefore, the presence of multiple EGFR variants in a single tumor highlights the intratumoral heterogeneity of glioblastoma conferred by the plasticity of EGFR amphi
cons.

Heterogeneous expression of oncogenic EGFR mutants and cooperativity of EGFR with itself or other RTKs have each been hypothesized to contribute to tumor growth and resistance to therapy (14, 16, 40). It is conceivable that heterogeneity and cooperativity mechanisms play a role in RTK diversification and will affect the efficacy of targeted therapies in cancer, and, in particular, may explain the lack of response in glioblastoma. Further complicating this scenario is recent work demonstrating how EGFRvIII-containing cells can evade EGFR inhibitors through ampiclon loss and down-regulation of the oncogenic receptor (5). It is striking that small-molecule inhibitors can be so successful in non–small-cell lung cancers, but remain ineffective in glioblastomas, despite EGFR being functionally significant (2, 3, 46–49).

This work suggests that combining multiple EGFR inhibitors that act through different mechanisms on the receptor could be required in patients with multiple variants in nonoverlapping subclones and also potentially prevent the activation of a resistance pathway. Such concepts deserve further evaluation in preclinical models.

As cancer genomic research approaches the discovery of the plurality of somatic mutations, it becomes crucial to understand the distribution pattern of somatic mutations among individual tumor cells and clones and how these mutations work together to drive tumor fitness (50). Single-cell analysis has the potential to greatly contribute to our understanding of co-dependency of mutations and tumor heterogeneity. The input requirements of this approach also make it feasible to deliver great insights into biopsies, necrotic tissues, or tumors with great immune infiltration to delineate the full spectrum of cell-specific oncogenic drivers. Once a thorough understanding of the tumor state can be integrated, it is then possible to select combinations of drugs to prevent resistance due to emergence of preexisting subclonal populations of tumor cells not detected in bulk tumor sequencing.

**METHODS**

**Isolation of SN and Preparation of Bulk Tumor DNA**

Anonymized tumor and matching blood specimens were collected following informed consent and processed under protocols approved by the Institutional Review Boards of the Dana-Farber Cancer Institute and the Broad Institute (protocol 11-433). Two cases of de novo GBM, designated as brain tumor BT325 and BT340, were molecularly dissociated into pools of nuclei as previously described (24). A portion of the dissociated material was selected to represent bulk tumor. Genomic DNA was isolated from the pooled nuclei and matched blood using DNaseasy columns (Qiagen).

**Preparation of SN DNA Libraries**

SN were sorted directly into 96-well PCR plates containing 5 µL of phosphate-buffered saline using the BD FACSAria IIu flow cytometer calibrated to deposit SN into each well. A 4′,6-diamidino-2-phenylindole (DAPI)-based DNA stain was used to select intact nuclei. Empty wells were included on every plate to control for background MDA amplification. See Supplementary Methods for additional details.

**Quality Assessment of SN DNA Libraries**

We generated SN DNA libraries from approximately 100 SN with the aim of selecting the best libraries from each tumor for deeper sequencing and analysis. Quality assessment was done by multiplexed sequencing of SN libraries to an average depth of 0.01× per library. Two criteria were used to select the best SNS libraries for deeper sequencing:

- First, we removed libraries with more than 20% read sequences not aligned to the human reference. We then evaluated the evenness of the remaining MDA libraries by two metrics. The first metric was adapted from the method recently described (21) to measure the evenness of genome-wide coverage by evaluating the distribution of reads in a DNA library across different genomic loci stratified by the sequencing coverage at each locus. The second metric, termed the pileup metric, estimates the percentage of sequenced reads that were accumulated in local regions at densities 10 times higher than the chromosomal average. The spacing between each read and the adjacent reads (left and right) was used to infer if the read was in a pileup region. Reads in pileup regions were then downsampled to the...
average density estimated from the flanking nonpileup regions. The downsampling was iterated until no pileup region exceeding a certain size (e.g., 100 reads) was remaining. The pileup metric was calculated from the percentage of reads removed during the downsampling procedure.

The two metrics showed a strong correlation between each other (Supplementary Fig. S2C); both can be used to estimate of the evenness of coverage for SNS libraries to guide the optimization of SNS workflow or select better libraries for deeper sequencing. In addition, removing the pileup reads can smooth the sequence coverage and reveal true copy-number profiles for each SNS library, which can be used to identify (tumor) samples of interest.

Two-Step Strategy for Somatic Variant Detection by SNS

Whole-genome amplification by MDA is known to generate artifacts, including single-base errors and random chimera, which can be indistinguishable from true somatic mutations. Given these considerations, we adopted a two-step strategy to analyze somatic mutations by SNS of many tumor nuclei. First, we detected all clonal and subclonal somatic variants, including single-nucleotide variants and chromosomal rearrangements, from the pool of SNS complemented with bulk whole-genome sequencing. Second, we searched for the detected variants within each SNS and in the bulk tumor to infer the clonality of these variants and construct the clonal hierarchy of these variants. See expanded methods for additional details.

Detection of Somatic Copy-Number Alterations

Whole-genome amplification by MDA from single-nuclear DNA can generate multiple distortions to the original DNA copy-number profile, which pose a considerable challenge to the detection of true somatic copy-number events in SN libraries. We, therefore, did not use the read depth signal from SN libraries to detect copy-number breakpoints but only inferred copy-number breakpoints from read depths in the bulk tumor against the germline reference using SegSeq (51). At 10× sequencing depth, the read depth signal from the bulk tumor should be sufficient to detect somatic copy-number alterations of 10 kb or longer at or above 15% clonal fraction, which is comparable with the smallest subclone we can identify from 50 to 60 single cells.

As the read depth signal often resulted in oversegmentation, we merged the copy-number breakpoints inferred from the bulk tumor read depth with the list of nonreciprocal chromosomal translocation breakpoints detected from the pool of bulk tumor and SN libraries. We then identified true chromosomal breaks as supported both by read depth difference across the breakpoint and by the presence of somatic rearrangement signatures on the higher copy-number side. This final set of chromosomal breakpoints was then used to segment the copy-number profile for the bulk tumor as well as for each SN.

Haplotype-Based Detection of LOH

A unique feature of many tumor genomes is the LOH of large chromosomal regions as well as smaller homoeologous deletions, leading to the elimination of one or both germline haplotypes (refs. 52, 53; Supplementary Fig. S4A). For cells with LOH in a region of the genome, the retained haplotype, or allele type (54), can be directly determined from the sequencing data (Supplementary Fig. SSB). If whole-genome amplification from single cells were complete and represented all alleles, then the haploid genotype of cells with LOH could be readily distinguished from the heterozygous genotype of diploid cells (Supplementary Fig. SSB, left). However, single-cell sequencing is inevitably incomplete and not all alleles are represented in the resulting data (Supplementary Fig. SSB, right). Single-base errors introduced during MDA can further confound the inference of true genotype.

We devised a clustering-based approach to dissect the cell populations harboring heterozygous and/or homozygous genotypes in each genomic region. First, the pairwise distance between single-cell genotypes was calculated from the fraction of germline heterozygous sites covered in both libraries with different genotypes (Supplementary Fig. SSC); this distance is large if cells are discordant for heterozygosity/homozygosity, and small if they are both homozygous (concordant). We then performed clustering analysis of cellular populations to identify cells with LOH in a given region, which formed a very tight cluster (Supplementary Fig. SS, right). Supplementary Figure SSD shows the two clusters formed by 60 SNS libraries from BT325 identified by LOH in chromosome 10, one cluster displaying a unique haploid genotype (LOH), and a second cluster that was heterozygous.

By pooling the genotype data of LOH cells in cluster 1, we can determine both the retained haplotype “R” and the deleted haplotype “D” and assign individual or combined “RD” haplotypes to each nucleus at all covered germline heterozygous sites (Supplementary Fig. SSE). This analysis showed that cluster 1 contains BT325 nuclei with only allele “R,” whereas cluster 2 contains an admixture of alleles “R” and “D” on chromosome 10 (Supplementary Fig. SSF). The ratio of the deleted to retained haplotypes is approximately 0 in LOH nuclei, and close to 1 in diploid nuclei. Interestingly, the D:R ratio is near 0.5 in another population of nuclei (Supplementary Fig. SSG).

We inferred that the population with a D:R ratio of 0.5 represents libraries derived from a 1:1 mixture of tumor and normal nuclei. The normal/tumor mixture was later confirmed by the presence of clonal drivers at subclonal quantities in these samples (Supplementary Fig. S7). Therefore, the haplotype analysis also provides an accurate method to identify and characterize cellular mixtures in presumably single-nuclear samples.

FISH on FFPE Tissue

FISH was performed on 5-μm formalin-fixed paraffin-embedded (FFPE) tissues, which were baked at 60°C for at least 2 hours, then deparaffinized and digested using methods described previously (55). See Supplementary Methods for probes used in this study. Tissue sections and probes were codenatured, hybridized at least 16 hours at 37°C in a darkened humid chamber, washed in 2× saline-sodium citrate (SSC) at 70°C for 10 minutes, rinsed in room-temperature 2× SSC, and counterstained with DAPI (Abbott Molecular/Vysis, Inc.). Slides were imaged using an Olympus BX51 fluorescence microscope. Individual images were captured using an Applied Imaging system running CytoVision Genus version 3.9, and all aberrations detected by FISH were reviewed and confirmed by a cytogeneticist (A.H. Ligon).

BAC Probes in FISH

Bacterial artificial chromosome (BAC) probes were all obtained from the CHORI BACPAC Resources Center (Oakland, CA) and included: clone RP11-815K24 (corresponding to EGFR, at 7p11.2), clone RP11-121K16 (corresponding to 6q13), and clone RP11-482110 (mapping to 9p21.3). Commercially available probes (all obtained from Abbott Molecular) included CEP6 (6p11.1-q11), CEP7 (7p11.1-q11.1), and CEP9 (9p11-q11). Final homebrew probe concentrations were 100 ng/μL. Final concentrations for commercial reference probes followed the manufacturer’s directions.

Agilent SureFISH

Specific detection of the EGFRVIII and EGFR wild-type alleles was performed using custom-designed pooled oligonucleotide probe sets (SureFISH), which tiled exons 2 to 7 (red probe, retained in wild-type
EGFR and exons 9 to 28 (green probe, used to enumerate copies of both wild-type EGFR and EGFRvIII). Hybridization was performed according to the manufacturer’s protocol (Agilent Technologies) on 4-μm FFPE sections from BT325.

**Expression Constructs**

Wild-type EGFR and EGFRvIII p-BABE-puro vectors were prepared as previously described (38). QuickChange site-directed mutagenesis ( Stratagene ) was used for generating EGFR vIII using wild-type EGFR as a template. All plasmids were confirmed by sequencing.

**Cell Culture and Generation of Cell Lines by Viral Transduction**

Ba/F3 cells were obtained, thawed, and maintained as previously described (38). Proliferation assays were performed as previously described (37). The Ba/F3 cells were confirmed to be the parental line through a cell viability test following removal of IL3.

Mouse neural stem cells were derived from E14.5 ganglionic eminences of p16Ink4a:p19Arf-null embryos, genotyped by qPCR on genomic DNA as described (13, 39). Cells were transduced by EGFR wild-type, EGFRvII, or EGFRvIII p-BABE retrovirus and selected with puromycin after 3 days for two passages. Transduction of the EGFR variant was confirmed by immunoblot. For subcutaneous xenograft studies, EGFRvII- or EGFRvIII-transduced mouse neural stem cells (106 cells) were injected in the left flank of nude mice with EGFR wild-type injected on the right as control. Five mice were injected for each EGFR variant. All animal experiments were performed in accordance with Dana-Farber Cancer Institute animal facility regulations and policy under the protocol #09-016.

**Drug Sensitivity Assay**

For growth inhibition assays, Ba/F3 cells (10,000 cells) were plated in 180 μL of media in 96-well flat-bottom plates. Twenty-four hours after plating, drugs prepared in 20 μL of cell culture medium were added to cells. All drugs were purchased from LC Laboratories. The concentrations of erlotinib and lapatinib used for the assay ranged from 0.001 to 100 nmol/L. The concentrations of afatinib and neratinib ranged from 0.001 to 100 nmol/L. The cells were incubated for another 72 hours and the viable cell numbers were measured using CellTiter Glo (Promega).

**Immunoblotting and Antibodies**

Cultures were serum-starved for 3 hours before EGF stimulation and harvesting. EGF (PeproTech) stimulations were performed using 25 ng/mL for 10 minutes. Anti-EGFR, phospho-EGFR (pT1068), AKT, phospho-AKT (pS473), ERK1/2, and phospho-ERK1/2 (pT202/Y204) antibodies were purchased from Cell Signaling Technology. Anti-vinculin was from Sigma-Aldrich.

**Disclosure of Potential Conflicts of Interest**

M. Meyerson reports receiving a commercial research grant from Bayer and has ownership interest (including patents) in and is a consultant/advisory board member for Foundation Medicine. No potential conflicts of interest were disclosed by the other authors.

**Authors’ Contributions**

Conception and design: J.M. Francis, C.-Z. Zhang, C.L. Maire, J. Jung, V.A. Adalsteinsson, M. Meyerson, K.L. Ligon
Development of methodology: J.M. Francis, C.-Z. Zhang, C.L. Maire, J. Jung, B. Blumentzid, K.L. Ligon
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): J.M. Francis, C.-Z. Zhang, C.L. Maire, V.E. Manzo, S. Haidar, J.C. Love, K.L. Ligon

**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** J.M. Francis, C.-Z. Zhang, C.L. Maire, S. Haidar, C.S. Pedamallu, A.H. Ligon, K.L. Ligon

**Writing, review, and/or revision of the manuscript:** J.M. Francis, C.-Z. Zhang, C.L. Maire, J. Jung, A.H. Ligon, M. Meyerson, K.L. Ligon
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): J.M. Francis, C.-Z. Zhang, H. Homer, S. Haidar, K.L. Ligon

**Study supervision:** J.M. Francis, C.-Z. Zhang, J.C. Love, M. Meyerson, K.L. Ligon

**Acknowledgments**

The authors thank Dr. Juliana Chmielecki for generating the EGFRvII expression construct and Drs. Scott Carter and Marcin Iemielski for helpful discussions. The authors also thank the Brigham and Women’s Hospital Tissue Biorepository Core (Dr. William Richards, Director) and the Department of Pathology faculty and staff of the Dana-Farber Cancer Institute Tissue Bank for their help and dedication. The authors also thank Leslie Gaffney for help in preparing figures.

**Grant Support**

This work was supported by the Bridge Project, a collaboration between The David H. Koch Institute for Integrative Cancer Research at MIT and the Dana-Farber/ Harvard Cancer Center (DF/HCC), and the National Brain Tumor Society (to J.C. Love, K.L. Ligon, and M. Meyerson). This work was supported in part by grants to K.L. Ligon from NCI R01CA170592, P01CA142536, and the Sontag Foundation, and to M. Meyerson from NCI R01CA16020.

Received November 15, 2013; revised May 18, 2014; accepted May 22, 2014; published OnlineFirst June 3, 2014.

**REFERENCES**


EGFR Variant Heterogeneity in Glioblastoma Resolved through Single-Nucleus Sequencing


Cancer Discovery 2014;4:956-971. Published OnlineFirst June 3, 2014.

Updated version
Access the most recent version of this article at: doi:10.1158/2159-8290.CD-13-0879

Supplementary Material
Access the most recent supplemental material at:
http://cancerdiscovery.aacrjournals.org/content/suppl/2014/06/05/2159-8290.CD-13-0879.DC1

Cited articles
This article cites 55 articles, 27 of which you can access for free at:
http://cancerdiscovery.aacrjournals.org/content/4/8/956.full#ref-list-1

Citing articles
This article has been cited by 20 HighWire-hosted articles. Access the articles at:
http://cancerdiscovery.aacrjournals.org/content/4/8/956.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.