bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Integrated time-course omics analysis distinguishes immediate therapeutic response from acquired resistance

Genevieve Stein-O'Brien[1,2*], Luciane T Kagohara[3*], Sijia Li[3*], Manjusha Thakar[3], Ruchira Ranaweera[3,4], Hiroyuki Ozawa[5], Haixia Cheng[6], Michael Considine[3], Alexander V Favorov[3,7], Ludmila V Danilova[3,7], Joseph A Califano[8], Evgeny Izumchenko[9], Daria A Gaykalova[9], Christine H Chung[3,4,+], Elana J Fertig[3,+].

Affiliations:

1. Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD, USA
2. Lieber Institute for Brain Development, Baltimore, MD, USA
3. Department of Oncology, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD, USA
4. Department of Head and Neck-Endocrine Oncology, Moffitt Cancer Center, Tampa, FL, USA
5. Department of Otorhinolaryngology-Head and Neck Surgery, Keio University School of Medicine, Tokyo, Japan
6. Department of Surgery - Otolaryngology–Head and Neck Surgery, University of Utah, Salt Lake City, UT
7. Laboratory of Systems Biology and Computational Genetics, Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia
8. Department of Surgery, UC San Diego Moores Cancer Center, La Jolla, CA, USA
9. Department of Otolaryngology-Head and Neck Surgery, Johns Hopkins University, Baltimore, MD, USA

* These authors contributed equally for this manuscript.
+ Co-corresponding authors

Email addresses:
Genevieve Stein-O'Brien: gsteinobrien@jhmi.edu
Luciane T Kagohara: ltsukam1@jhmi.edu
Sijia Li: sli61@jhu.edu
Manjusha Thakar: mthakar3@jhmi.edu
Ruchira Ranaweera: Ruchira.Ranaweera@moffitt.org
Hiroyuki Ozawa: ozakky@cb.mbn.or.jp
Haixia Cheng: haixia.cheng@hci.utah.edu
Michael Considine: mconsid3@jhmi.edu
Alexander Favorov: favorov@sensi.org
Ludmila Danilova: ldanilo1@jhmi.edu
Joseph A Califano: jcalifano@ucsd.edu
Evgeny Izumchenko: izumchen@jhmi.edu
Daria A Gaykalova: dgaykal1@jhmi.edu
Christine H Chung: Christine.Chung@moffitt.org
Elana J Fertig: ejfertig@jhmi.edu

Key words: Acquired resistance, data integration, time-course analysis, genomics, epigenetics

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Background: Targeted therapies specifically act by blocking the activity of genes that are critical for tumorigenesis. However, most cancers acquire resistance and long-term disease control is rarely observed. Knowing the timing of molecular changes responsible for the development of acquired resistance can enable optimization of alterations to patients' treatments. Clinically, acquired therapeutic resistance can only be studied at a single time point in resistant tumors. To determine the dynamics of these molecular changes, we obtained high throughput omics data weekly during the development of cetuximab resistance in a head and neck cancer model.

Results: An unsupervised algorithm, CoGAPS, quantified the evolving transcriptional and epigenetic changes. Further applying a PatternMarker statistic to the results from CoGAPS enabled novel heatmap-based visualization of the dynamics in these time-course omics data. We demonstrate that transcriptional changes resulted from immediate therapeutic response and resistance whereas epigenetic alterations only occurred with resistance. Integrated analysis demonstrated delayed onset of changes in DNA methylation relative to transcription, suggesting that resistance was stabilized epigenetically.

Conclusions: Genes with epigenetic alterations associated with resistance that had concordant expression changes were hypothesized to stabilize resistance. These genes include *FGFR1*, which was associated with EGFR inhibitor resistance previously. Thus, integrated omics analysis distinguishes the timing of molecular drivers of resistance. Our findings are a relevant step into the better understanding of the time course progression of changes resulting in acquired resistance to targeted therapies. This is an important contribution to the development of alternative treatment strategies that would introduce new drugs before the resistant phenotype develops.

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

BACKGROUND

Cancer targeting therapeutic agents inhibit specific key role players in the regulation of molecular pathways essential for tumor development and maintance [1]. These therapies prolong survival but are not curative. Most patients will develop acquired resistance within the first few years of treatment [2]. Although a wide variety of molecular alterations that confer resistance to treatment have been described, the mechanisms and timing of their evolution are still poorly characterized [3,4]. Serial biopsies during the prolonged treatment period are invasive, expensive, and impractical for patients. Thus, molecular alterations associated with acquired resistance from previous studies are only known when the tumor has already developed resistance. As a result, the understanding of how these changes develop and are incorporated by the tumor cells is restricted to few time points along the treatment. The lack of adequate time course datasets makes it challenging to delineate the two predominant hypotheses for how therapeutic resistance develops: the presence of small populations of resistant cells that will survive the treatment and repopulate the tumor; or the development of *de novo* resistance by the tumor cells [3,5]. Characterization of the dynamics of genomic alterations induced during acquired resistance can identify targetable oncogenic drivers and determine the best time point to introduce alternative therapeutic strategies to avoid resistance establishment [6].

Inhibitors against Epidermal Growth Factor Receptor (EGFR) represent a common class of targeted therapeutics. Cetuximab, a monoclonal antibody against EGFR, is FDA approved for the treatment of metastatic colorectal cancer and head and neck squamous cell carcinoma (HNSCC) [7]. As with other targeted therapies, stable response is not observed for a long period and virtually all patients invariably develop acquired resistance [8]. Recent advances in *in vitro* models of acquired cetuximab resistance [9] provide a unique opportunity to study the time-course of genetic events resulting in acquired resistance. Cell lines chronically exposed to the targeted agent develop resistance and can be sequentially collected during the course of treatment to evaluate the progressive molecular changes. Previous studies to assess the mechanisms of acquired cetuximab resistance have been limited to comparing the genomic profile of the parental sensitive cell line to stable clones with acquired resistance [9–11]. Therefore, these studies fail to capture the dynamics of acquired molecular alterations during the evolution of therapeutic resistance. Approaches using combined experimental and bioinformatics tools that would adjust to different tumor

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

models and therapeutic agents are fundamental tools to overcome issues related to sample availability and serial time point data analysis.

Even with advances in experimental sampling approaches, time-course high throughput data alone is insufficient to determine molecular drivers of therapeutic resistance. A novel serial, multi-platform genomics analysis is essential to untangle specific and targetable signaling changes that drive cetuximab resistance in HNSCC. Current supervised bioinformatics algorithms that find time-course patterns in genomic data adjust linear models to correlate molecular profiles with known temporal patterns [12–15]. However, these algorithms cannot quantify the rate of genomics alterations relative to that of the input phenotype. Other algorithms [16–21] enhance such inference by using prior knowledge of gene relationships to find coherent, dynamic regulatory relationships that are linked to pathways. Many of these algorithms trace individual phenotypes or individual genomics platforms. Their ability to determine drivers of gene expression associated with acquired resistance from time-course data in multiple experimental conditions and multiple genomics data modalities is emerging [22]. On the other hand, unsupervised compressed sensing algorithms can simultaneously quantify the dynamics and infer the gene regulatory networks directly from the input time course data [23]. Nonetheless, visualization tools for inference of dynamics from unsupervised algorithm are limited.

In this study, we used an *in vitro* HNSCC cell line model to induce resistance and measure the molecular changes using multiple high throughput assays while the resistant phenotype developed. We measured DNA methylation and gene expression to determine the functional impact of the inferred epigenetic alterations. The association of DNA methylation with acquired cetuximab resistance *in vitro* and *in vivo* have been previously described [24]. An intrinsic sensitive HNSCC cell line (SCC25) was treated for 11 weeks with cetuximab or phosphate buffered saline (PBS) and cells were collected weekly. In total, high throughput analysis for gene expression and methylation was performed in 11 treated and 11 untreated samples. Using the 22 samples generated, we used the Bayesian non-negative matrix factorization algorithm CoGAPS [25] to infer specific patterns of gene expression and DNA methylation that developed according to the gradual establishment of the acquired cetuximab resistance. Gene expression also had a pattern associated with immediate

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

therapeutic response. We selected genes uniquely associated with these changes using a PatternMarker statistic [25]. Plotting expression or methylation of genes with the PatternMarker statistic enabled novel visualization of the dynamics of these alterations from high throughput data. Analysis of the CoGAPS patterns demonstrated that onset of methylation changes associated with resistance were temporally delayed relative to expression changes and involved different genes, suggesting that epigenetic alterations stabilize the gene signatures relevant to the resistant phenotype. Specifically, the DNA methylation PatternMarker genes were selected as putative epigenetic drivers. Therefore, we next performed correlation analysis of PatternMarkers of the DNA methylation patterns with gene expression to identify the subset of genes with tight temporal concordance implying direct epigenetic regulation. Among these genes *FGFR1* had the strongest correlation between gene expression and DNA methylation in a cetuximab resistant sub-clone generated from the same parental cell line. Previous studies associate *FGFR1* increased expression with acquired cetuximab resistance in HNSCC patients [26–28]. In this study, we also demonstrated that epigenetic changes of *FGFR1* are observed in HNSCC tumors in TCGA. This work represents the first integrated time-course analyses to determine the drivers of acquired resistance, suggesting a direct link between epigenetic regulation of *FGFR1* gene expression and the development of acquired resistance. Both the experimental and bioinformatics methods developed could be applicable to other molecular platforms, therapeutics, and cancer types.

METHODS

Cell lines and materials

SCC25 cells were purchased from American Type Culture Collection (ATCC). Cells were cultured in Dulbecco's Modified Eagle's medium and Ham's F12 medium supplemented with 400ng/mL hydrocortisone and 10% fetal bovine serum and incubated at 37°C and 5% carbon dioxide. SCC25 was authenticated using short tandem repeat (STR) analysis kit PowerPlex16HS (Promega, Madison, WI) through the Johns Hopkins University Genetic Resources Core Facility. Cetuximab (Lilly, Indianapolis, IN) was purchased from the Johns Hopkins Pharmacy.

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Experimental protocol to establish time-course during acquisition of cetuximab resistance in SCC25

The HNSCC cell line SCC25 (intrinsically sensitive to cetuximab) was treated with 100nM cetuximab every three days for 11 weeks (generations G1 to G11). On the eighth day, cells were harvested. Sixty thousand cells were replated for another week of treatment with cetuximab and the remaining cells were separately collected for: (1) RNA isolation (gene expression analysis); (2) DNA isolation (DNA methylation analysis); (3) proliferation assay and (4) storage for future use. All steps were repeated for a total of 11 weeks. In parallel with the cetuximab treated cells, we generated controls that received the same correspondent volume of PBS (phosphate buffered saline). Cells were plated in several replicates each time at the same initial density. The replicates were then harvested and pooled to provide enough cells for genetic, epigenetic and proliferation assays. To achieve adequate final cell confluence and number of cells for the experimental analysis of each generation, cetuximab and PBS treated cells were plated in different flask sizes. Cells treated with cetuximab were plated in multiple T75 ($75cm^2$) flasks (60,000 cells/flask) that were combined on the eighth day. PBS treated cells were plated in a single T175 ($175cm^2$) flask (60,000 cells). This design was selected considering the growth inhibition of the earliest cetuximab generations and to control confluence of the PBS controls at the collection time (**Supplemental Fig. 1**).

Cell proliferation and colony formation assays

Cell proliferation events were measured using the Click-iT Plus EdU Flow Cytometry Assay Kit Alexa Fluor 488 Picolyl Azide (Life Technologies, Carlsbad, CA) according to manufacturer's instructions. The cetuximab generations were considered resistant when the frequency of proliferating cells was higher than in the PBS control generations.

Anchorage-independent growth assay was used to further confirm the development of resistance. The parental SCC25 and the late G10 resistant cells were treated with different concentrations of cetuximab 10nM, 100nM and 1000nM. Number of colonies was compared to the same cells treated with PBS. Colony formation assay in Matrigel (BD Biosciences, Franklin Lakes, NJ) was performed as described previously [29].

Stable SCC25 cetuximab resistant single clones (CTXR clones)

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Resistance to cetuximab was induced in an independent passage of SCC25 cells. After resistance was confirmed, single cells were isolated and grown separately to generate the isogenic resistant single cell clones (CTXR). In total, 11 CTXR clones were maintained in culture without addition of cetuximab. With the exception of one clone (CTXR6), all CTXR clones presented substantial survival advantage compared to the parental SCC25, as reported by Cheng et al. [30].

Proliferation assay was performed to confirm cetuximab resistance in the CTXR clones compared to the parental SCC25. A total of 1000 cells were seeded in 96-well plates in quadruplicate for each condition. PBS or cetuximab (10nM, 100nM or 1000nM) was added after 24 and 72 hours and cells were maintained in culture for 7 days. AlamarBlue reagent (Invitrogen, Carlsbad, CA) at a 10% final concentration was incubated for 2 hours and fluorescence was measured according to the manufacturer's recommendations (545nm excitation, 590nm emission). Resistance in the CTXR clones was confirmed when the proliferation rates were higher than in the PBS treated SCC25 cells.

RNA-sequencing (RNA-seq) and data normalization

RNA isolation and sequencing were performed for the parental SCC25 cells (G0) and each of the cetuximab and PBS generations (G1 to G11) and the CTXR clones at the Johns Hopkins Medical Institutions (JHMI) Deep Sequencing & Microarray Core Facility. RNA-seq was also performed for two additional technical replicates of parental SCC25 cell line to distinguish technical variability in the cell line from acquired resistance mechanisms. Total RNA was isolated from a total of $1x10^6$ cells using the AllPrep DNA/RNA Mini Kit (Qiagen, Hilden, Germany) following manufacturer's instructions. The RNA concentration was determined by the spectrophotometer Nanodrop (Thermo Fisher Scientific, Waltham, MA) and quality was assessed using the 2100 Bioanalyzer (Agilent, Santa Clara, CA) system. An RNA Integrity Number (RIN) of 7.0 was considered as the minimum to be used in the subsequent steps for RNA-seq. Library preparation was performed using the TrueSeq Stranded Total RNAseq Poly A1 Gold Kit (Illumina, San Diego, CA), according to manufacturer's recommendations, followed by mRNA enrichment using poly(A) enrichment for ribosomal RNA (rRNA) removal. Sequencing was performed using the HiSeq platform (Illumina) for 2X100bp sequencing. Reads were aligned to hg19 with MapSplice [31] and gene expression counts were quantified

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

with RSEM [32]. Gene counts were upper-quartile normalized and log transformed for analysis following the RSEM v2 pipeline used to normalize TCGA RNA-seq data [33]. All RNA-seq data from this study is available from GEO (GSE98812) as part of SuperSeries GSE98815.

DNA methylation hybridization array and normalization

Genome-wide DNA methylation analysis was performed on the same samples as RNA-seq using the Infinium HumanMethylation450 BeadChip platform (Illumina) at the JHMI Sidney Kimmel Cancer Center Microarray Core Facility. Briefly, DNA quality was assessed using the PicoGreen DNA Kit (Life Technologies) and 400ng of genomic DNA was bisulfite converted using the EZ DNA Methylation Kit (Zymo Research, Irvine, CA) following manufacturer's recommendations. A total volume of 4μL of bisulfite-converted DNA was denatured, neutralized, amplified and fragmented according to the manufacturer's instructions. Finally, 12μL of each sample were hybridized to the array chip followed by primer-extension and staining steps. Chips were image-processed in the Illumina iScan system. Data from the resulting iDat files were normalized with funnorm implemented in the R/Bioconductor package minfi (version 1.16.1) [34]. Methylation status of each CpG site was computed from the signal intensity in the methylated probe (*M*) and unmethylated probe (*U*) as a *β* value as follows:

$$\beta = \frac{M}{M+U}.$$

Annotations of the 450K probes to the human genome (hg19) were obtained from the R/Bioconductor package FDb.InfiniumMethylation.hg19 (version 2.2.0). Probes on sex chromosomes or annotated to SNPs were filtered from analysis. The CpG island probe located closest to the transcription start site was selected for each gene. Genes with CpG island probes less than 200bp from the transcription start site were retained to limit analysis to CpG island promoter probes for each gene. Probes are said to be unmethylated for $\beta < 0.1$ and methylated for $\beta > 0.3$ based upon thresholds defined in TCGA analyses [33]. All DNA methylation data from this study is available from GEO (GSE98813) as part of SuperSeries GSE98815.

Hierarchical clustering and CoGAPS analysis

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Unless otherwise specified, all genomics analyses were performed in R and code for these analyses is available from https://sourceforge.net/projects/scc25timecourse.

The following filtering criterion for genes from the profiling of the time course data from generations of cetuximab treated cells was used. Genes from RNA-seq data were selected if they had log fold change greater than 1 between any two time points of the same condition and less than 2 between the replicate control samples at time zero (5,940 genes). CpG island promoter probes for each gene were retained if the gene switched from unmethylated ($\beta < 0.1$) to methylated ($\beta > 0.3$) in any two samples of the time course (1,087 genes). We used the union of the sets of genes retained from these filtering criteria on either data platform for analysis, leaving a total of 6,445 genes in RNA-seq and 4,703 in DNA methylation.

Hierarchical clustering analysis was performed with Pearson correlation dissimilarities between genes and samples on all retained genes. CoGAPS analysis was performed on both log transformed RNA-seq data and DNA methylation β values, independently using the R/Bioconductor package CoGAPS [35] (version 2.9.2). CoGAPS decomposed the data according to the model

$$D_{i,j} \sim \mathcal{N}\left(\sum_{k=1}^{p} A_{i,k} P_{k,j}, \Sigma_{i,j}\right),$$

where $\mathcal{N}$ represents a univariate normal distribution, matrices $\mathbf{A}$ and $\mathbf{P}$ are learned from the data for a specified number of dimensions p, $\Sigma_{i,j}$ is an estimate of the standard deviation of each row and column of the data matrix $\mathbf{D}$, and i represents each gene and j each sample. In this decomposition, each row of the pattern matrix $\mathbf{P}$ quantifies the relative association of each sample with a continuous vector of relative gene expression changes in the corresponding column of $\mathbf{A}$. These relative gene weights are called meta-pathways, and the reduction of the high-dimensional data into fewer reductions through this matrix factorization represents an established form of compressed sensing in genomics. The standard deviation of the expression data was 10% of the signal with a minimum of 0.5. The standard deviation of DNA methylation data under the assumption that β values follow a beta distribution is

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$\Sigma_{i,j}^{\beta} = \sqrt{\frac{\beta_{i,j}\left(1-\beta_{i,j}\right)}{M_{i,j}+U_{i,j}+1}}.$$

CoGAPS was run for a range of 2 to 10 dimensions p for expression and 2 to 5 for DNA methylation. Robustness analysis with ClutrFree [36] determined that the optimal number of dimensions p for expression was 5. DNA methylation was run in 4 parallel sets using GWCoGAPS [25]. In DNA methylation, the maximum number of patterns that modeled resistance mechanisms over and above technical variation in replicate samples of SCC25 was three. Gene sets representative of the meta-pathway were derived for each pattern using the PatternMarkers statistics [25]. Gene set activity was estimated with the gene set statistic implemented in calcCoGAPSStat of the CoGAPS R/Bioconductor package [35]. Comparisons between DNA methylation and gene expression values for PatternMarkerGenes or from CoGAPS patterns and amplitudes were computed with Pearson correlation.

Cetuximab resistance signatures and EGFR network

In a previous study, CoGAPS learned a meta-pathway from gene expression data corresponding to overexpression of the $HRAS^{Val12D}$ in the HaCaT model of HPV- HNSCC premalignancy. That study associated the CoGAPS HaCaT-HRAS meta-pathway with gene expression changes in acquired cetuximab resistance in the HNSCC cell line UMSCC1 [37]. In the current study, we applied the PatternMarkers statistics [25] to the previously published CoGAPS analysis of these data to derive a gene set from this meta-pathway called HACAT_HRAS_CETUXIMAB_RESISTANCE or HACAT_RESISTANCE. In addition, we searched MSigDB [38] (version 5.2) for all gene sets associated with resistance to EGFR inhibition. In this search, we found the gene sets COLDREN_GEFITINIB_RESISTANCE_DN and COLDREN_GEFITINIB_RESISTANCE_UP representing resistance to the EGFR inhibitor gefitinib in non-small-cell lung cancer cell lines [39]. Gene sets of transcription factor targets were obtained from experimentally validated targets annotated in the TRANSFAC [40] professional database (version 2014.1).

Sources and analysis of human tumor genomics data

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Genomics analyses of TCGA was performed on level 3 RNA-seq and DNA methylation data from the 243 HPV-negative HNSCC samples from the freeze set for publication [33]. DNA methylation data was analyzed for the same CpG island promoter probes obtained in the cell line studies. Pearson correlation coefficients were computed in R to associate different molecular profiles.

Analysis was also performed on gene expression data measured with Illumina HumanHT-12 WG-DASL V4.0 R2 expression beadchip arrays on samples from patients treated with cetuximab from Bossi et al [41], using expression normalization and progression-free survival groups as described in the study. Data was obtained from the GEO GSE65021 series matrix file. We performed t-tests in R on the probe that had the highest standard deviation of expression values for each gene.

RESULTS

Prolonged exposure to cetuximab induces resistance

SCC25 is among the most sensitive HNSCC cell lines to cetuximab but can acquire resistance during a long-term exposure to cetuximab [30]. This cell line is one of two HNSCC cell lines shown in the literature to develop isogenic clones that are resistance to cetuximab from long term exposure [10], with a third model developing resistance to erlotinib from contamination by HeLa [9]. In this study, cetuximab resistance was induced by exposing the SCC25 cells to the targeted therapeutic agent for a period of eleven weeks (CTX-G1 to –G11) (**Supplemental Fig. 1**). The SCC25 cells treated with PBS were used as time-matched controls (PBS-G1 to –G11). Response to cetuximab was determined by comparing the proliferation rates between CTX and PBS generations. Proliferation of the PBS generations was stable throughout the eleven weeks (G1 to G11). Conversely, proliferation of the CTX generations progressively increased over each week (**Fig. 1**). Relative to the untreated controls, the growth of the treated cells was initially (CTX-G1) inhibited until CTX-G3. Starting at CTX-G4, the cells became resistant to the anti-proliferative effects of cetuximab and gained stable growth advantages compared to the untreated controls.

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Comparison of proliferation rates between generations of CTX treated cells relative to generations of cells treated PBS enabled us to conclude that cell growth advantages arise from chronic cetuximab treatment and were associated with resistance rather than prolonged cell culturing. We mirrored the changes in proliferation rates with clinical responses seen in HNSCC tumors treated with cetuximab (**Fig. 1**, top panel). Specifically, we inferred that the decreased growth rates in CTX-G1 to –G3 represented initial stages of treatment with a decrease in tumor size. Then, the switch from decreased to increased growth rates during CTX-G3 to -G4 represented stable disease without tumor outgrowth. Finally, the higher proliferation in cetuximab-treated cells starting at CTX-G4 represented rapid outgrowth after acquired resistance.

Because higher proliferation in treated than untreated cells started at CTX-G4, this was the timepoint at which acquired cetuximab resistance began and all subsequent timepoints continue to acquire stable cetuximab resistance. The resistant CTX generation 10 (CTX-G10) also presented enhanced anchorage-independent growth when compared to the parental SCC25 (G0) at different concentrations of cetuximab (two-way anova with multiple comparisons p-value < 0.01 for each concentration, **Supplemental Fig. 2**), representing the stabilization of cetuximab resistance in later generations.

Treatment vs. control gene expression changes dominated clustering and immediate therapeutic response was confounded with changes from acquired resistance

To characterize the gene expression changes occurring as cells acquire cetuximab resistance, we collected RNA-seq data for the parental SCC25 cell line (G0) and from each generation of CTX- and PBS-treated cells. The RNA-seq data hierarchical unsupervised clustering separated genes with expression changes in treated and untreated generations (**Fig. 2A**). Clustering analysis of samples (**Supplemental Fig. 3**) further distinguished clusters with gene expression changes in stages with cetuximab sensitivity (CTX-G1 to CTX-G3), early stages of resistance (CTX-G4 to CTX-G8), and late stages of resistance (CTX G9-G11). Expression changes at these distinct stages were shared between numerous genes. Confounding by changes resulting from immediate therapeutic response made identification of resistance-specific gene expression changes impossible with clustering.

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Similar separation of stages of cetuximab response were observed in clustering analysis of gene signatures previously described in HNSCC and non-small cell lung cancer cell line models resistant to cetuximab or gefitinib (anti-EGFR small molecule), respectively [37,39] (**Supplemental Fig. 4**). For these genes, changes during early stages of resistance clustered for CTX-G4 to CTX-G6 as distinct from later stages for CTX G7-11. Nevertheless, these gene signatures also clustered samples with gene expression changes at early stages (CTX-G1 to G3) as distinct from samples from PBS treated generations. However, these analyses were insufficient to quantify the relative dynamics of genes associated with immediate response to therapy or subsequent acquired resistance.

CoGAPS analysis of gene expression distinguished patterns of acquired resistance from immediate therapeutic response

To define gene expression signatures for treatment effect and cetuximab resistance, we applied CoGAPS [25] Bayesian matrix factorization algorithm to the time-course gene expression data. Bayesian non-negative matrix factorization with algorithms such as CoGAPS have already proven highly effective in relating gene expression changes to patterns related to EGFR inhibition [42], perturbation of nodes in the EGFR network [43], and time course dynamics of targeted therapeutics. CoGAPS is an unsupervised matrix factorization algorithm that simultaneously infers the relative magnitude of genes in concordantly transcribed gene sets in each sample. These sets provide a low-dimensional representation that reconstructs the signal of the input genomics data. Therefore, CoGAPS is a robust, compressed sensing algorithm for genomics. These relative magnitudes across samples are called patterns and quantify the separation of distinct experimental conditions. The gene sets are inferred simultaneously, and are continuous to quantify the relative magnitude of gene weights in each set. A single gene may have non-zero magnitude in several distinct gene sets, representing the fact that a single gene can have distinct roles in different biological processes (such as immediate therapeutic response and acquired resistance). A recently developed PatternMarker statistic [25] selects the genes that are unique to each of the inferred patterns, and therefore represent biomarkers unique to the corresponding biological process.

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We identified five CoGAPS patterns in the time course gene expression dataset: three patterns that distinguished the experimental conditions (cetuximab vs. PBS) (**Fig. 2B** and **Fig. 2C** and **Supplemental Fig. 5**); one pattern that represented changes in gene expression from the parental cell lines and subsequent generations; and one pattern that was constant and corresponded to signature of highly expressed genes (**Supplemental Fig. 2**). We applied the PatternMarker statistic to define genes that were uniquely associated with each of these patterns. We excluded the technical, flat pattern to focus with genes with expression changes. By design, genes selected with the PatternMarker statistic are selected to not be multiply regulated regulated. Therefore, limiting the heatmap to these genes enabled visualization of the dynamics of gene expression changes in our time-course dataset (**Fig. 2B**). The relative magnitude of CoGAPS pattern weights for each sample quantified the dynamics of gene expression changes (**Fig. 2C**).

Similar to the separation seen with clustering (**Supplemental Fig. 5**), the first CoGAPS expression pattern distinguished cetuximab from PBS at every generation (expression pattern 1, **Fig. 2B** and **Fig. 2C**, top). These genes had an immediate transcriptional induction in response to cetuximab treatment. Gene set analysis to determine the function of CoGAPS patterns was performed with an enrichment analysis on all gene weights obtained from the CoGAPS analysis. By performing the analysis on gene weights and not only the PatternMarker genes in **Fig. 2B**, we accounted for multiple regulation of genes in pathways. Gene set enrichment analysis on confirmed that published resistance signatures [37,39] were significantly enriched in this pattern (**Supplemental Fig. 6**; one-sided p-values of 0.002 and 0.003 for resistance gene sets COLDREN_GEFITINIB_RESISTANCE_DN and HACAT_HRAS_CETUXIMAB_RESISTANCE, respectively). However, the transcriptional changes in this pattern were not associated with acquired resistance to cetuximab, and even decreased modestly as resistance developed. Further, enrichment by transcription factor AP-2alpha targets (*TFAP2A*; one-sided p-value of 0.05) confirmed previous work indicating that transcription by AP-2alpha is induced as an early feedback response to EGFR inhibition [44]. Based upon these findings, we concluded that pattern 1 was associated with immediate response to cetuximab although it includes genes that were also associated with cetuximab resistance in previous studies.

The second CoGAPS expression pattern quantified divergence of the cetuximab treated cells from controls at generation CTX-G4 (expression pattern 2, **Fig. 2B** and **Fig. 2C**, middle) which was the time point that

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cetuximab treated cells presented significant and stable growth advantage over PBS controls (**Fig. 1**). Therefore, expression pattern 2 obtained gene expression signatures associated consistently with the development of cetuximab resistance. Gene set statistics of transcription factor targets of EGFR on CoGAPS gene weights were significantly down-regulated in this acquired resistance pattern (**Supplemental Fig. 6**). One striking exception was c-Myc, which trended with acquired resistance (p-value of 0.06), consistent with the role of this transcription factor in cellular growth. Resistance signature COLDREN_GEFITINIB_RESISTANCE_DN gene signature was significantly down-regulated in expression pattern 2 (p-value of 0.04).

CoGAPS expression pattern 3 represented a gradual repression of gene expression with cetuximab treatment (**Fig. 2B** and **Fig. 2C**, bottom). This expression pattern trended to significant enrichment in the COLDREN_GEFITINIB_RESISTANCE_DN resistance signature (**Supplemental Fig. 6**, one-sided p-value 0.12) and down-regulated in the HACAT_HRAS_CETUXIMAB_RESISTANCE resistance signature (**Supplemental Fig. 6**, one-sided p-value 0.09). This confirmed that expression pattern 3 was associated with repression of gene expression during acquired cetuximab resistance.

Significant enrichment of the acquired resistance signature in CoGAPS expression patterns 1-3 (**Supplemental Fig. 6**) suggested that genes defined from case-control experimental designs of acquired resistance provide a mixture of genes associated with early response to cetuximab and genes associated with acquired resistance. In addition, the published resistance signatures [37,39] included genes that the CoGAPS and PatternMarker analysis associated with immediate response to treatment (**Supplemental Fig 4**). Inclusion of immediate response genes from expression pattern 1 in the published resistance signatures arose from the design of the experiments in the original publications [37,39]. Specifically, the resistance signatures derived from data that was collected at a single time point when the cell models have already developed resistance. At the same time point in our time-course data, gene expression changes included both immediate response genes and longer-term expression changes due to acquired resistance. Therefore, both sets of genes have significant expression changes in resistant cells when compared only to their parental cell line. These immediate response genes cannot be eliminated without including any additional samples at intermediate time points. This observation is consistent with recent studies demonstrating that

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

time-course proliferation data increases the accuracy in drug-response metrics by removing the confounding effects of variability in cell growth/division rates and treatment effects [45,46]. Thus, the gene expression signature in CoGAPS patterns from the time course were able to parse apart transcriptional changes specific to immediate therapeutic response from those specific to acquired resistance.

Changes in DNA methylation inferred with CoGAPS were associated with resistance to cetuximab, but not the immediate response to treatment observed in gene expression

To determine the timing of the methylation changes associated with acquired resistance, we also measured DNA methylation in each cetuximab generation of SCC25 cells and PBS controls (**Fig. 3A**). Application of the CoGAPS matrix factorization algorithm to the methylation data revealed a total of 3 patterns (**Fig. 3B and Fig. 3C**): gradual increase of DNA methylation in controls (DNA methylation pattern 1, middle); rapid demethylation in CTX generations starting at CTX-G4 (DNA methylation pattern 2, bottom); and rapid increase in DNA methylation in CTX generations starting at CTX-G4 (DNA methylation pattern 3, top). In contrast to the gene expression data, there was no immediate shift in DNA methylation resulting from cetuximab treatment.

Comparing the CoGAPS patterns from gene expression and DNA methylation revealed strong anti-correlation between gene expression and DNA methylation in resistant patterns **(Supplemental Fig. 7A)**. We observed that the gene expression changes associated with acquired resistance occurred gradually and were evident in early generations (**Fig. 4**, top). The DNA methylation was consistent in cetuximab treatment and control PBS in DNA methylation patterns 2 and 3 during early generations. Then, rapid accumulation in DNA methylation changes started after generations CTX-G4 and CTX-G5 in both patterns 2 and 3 (**Fig. 4**, bottom), concurrent with the onset of the observed growth advantage over the PBS control (**Fig. 1**). Changes in DNA methylation were delayed relative to those of gene expression in acquired cetuximab resistance (**Fig. 4**, dashed vertical lines). These dynamics suggests that DNA methylation stabilized the gene expression signatures crucial to the maintenance of acquired cetuximab resistance.

Epigenetic regulation of *FGFR1* expression was associated with acquired cetuximab resistance in the time course and in stable cetuximab resistant clones

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The gene signatures from the CoGAPS resistance patterns for expression and methylation had low correlation (**Supplemental Fig. 7B**) and there was little overlap between their respective PatternMarker genes. The low overlap between genes and their timing differences indicated that alterations to transcription were independent of DNA methylation. However, we hypothesized that the DNA methylation changes stabilized the resistant phenotype. Therefore, the CoGAPS gene signatures from each data modality were insufficient to define the functional DNA methylation regulation of acquired resistance. Nonetheless, we hypothesized that epigenetic regulation contributed to stabilizing the resistant phenotype. Characterizing the role of such epigenetic regulations is critical to understand the stable resistant phenotype. Moreover, identifying these epigenetic drivers can provide targets to overcome such stable resistance.

To ascertain potential drivers of the stable cetuximab resistant phenotype induced by DNA methylation, we defined genes that are PatternMarkers [25] of the DNA methylation patterns associated with stable acquired cetuximab resistance (methylation patterns 2 and 3). We then applied correlation analysis to determine genes that were epigenetically silenced. Specifically, we performed correlation analysis between DNA methylation and gene expression for each of the DNA methylation PatternMarker genes (**Fig. 5**). *FGFR1* was among these genes. This finding was consistent with previous studies that associate differential expression of *FGFR1* with resistance to EGFR inhibitors, including cetuximab, in different tumor types *in vitro* and *in vivo* [26–28]. Given the tight temporal regulation of these genes and the previous work on *FGFR1*, we hypothesized that this set of genes represented epigenetic drivers of acquired resistance.

To delineate whether our presumptive drivers resulted from clonal expansion of resistant cells or from the development of new epigenetic alterations to drive resistance, we measured DNA methylation and gene expression on a panel of eleven isogenic stable cetuximab resistant clones derived from SCC25 cells previously [30]. Briefly, SCC25 was continuously treated with cetuximab until resistance developed, and then single cell clones were isolated and profiled in the absence of cetuximab treatment. Despite being derived from parental SCC25 cells, the single cell clones and time course generations displayed widespread differences. Significantly greater heterogeneity was observed among the cetuximab resistant single-cell clones in both expression and methylation profiles (**Supplemental Fig. 8** and **9**, respectively) and cellular morphology (**Supplemental Fig. 10**). **Fig. 5A and 5B** demonstrate that higher heterogeneity among single

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cell clones was also observed in the epigenetically regulated PatternMarker genes from the CoGAPS analysis that are shown in **Figure 4D**. These results suggest that different mechanisms of resistance may arise in the same HNSCC cell line.

We hypothesized that epigenetically regulated genes shared along the time course patterns and resistant single-cell clones may implicate common mechanisms acquired during evolution of the stable resistance phenotype. To test this hypothesis, we also performed correlation analysis for each of the epigenetically regulated genes in our resistant set (**Fig. 5**) in the resistant clones and parental cell lines. Nine of the epigenetically regulated PatternMarker genes also had significantly anti-correlated gene expression and DNA methylation in the stable cetuximab resistant clones (**Supplemental Fig. 11**). Of these, only *FGFR1* was demethylated and reexpressed in a cetuximab resistant clone relative to the parental SCC25 cell line **(Fig. 6)**. In this analysis, epigenetic regulation of gene expression for *FGFR1* occurred in only one of the resistant clones (CTXR10). This clone was among the fastest growing under cetuximab treatment (**Supplemental Fig. 12**). This observation suggested that the pooled data from the time course captured clonal outgrowth of a cetuximab resistant clone with similar molecular features (*FGFR1* demethylation) to CTXR10, and that therefore clonal outgrowth was the dominant mechanism of resistance in our resistance model.

FGFR1 **observed dynamics** ***in vitro*** **recapitulates relationships from** ***in vivo*** **tumor genomics and acquired cetuximab resistance**

In order to validate our *in vitro* findings, we further investigated the pattern of expression and methylation of *FGFR1* and *EGFR* in other publicly available datasets. Using gene expression and DNA methylation data from The Cancer Genome Atlas (TCGA) for 243 HPV-negative HNSCC pretreatment samples [33]**,** we verified that the up-regulation of *EGFR* and *FGFR1* is not concomitant **(**Pearson correlation coefficient = -0.06, p value = 0.33, **Fig. 7A)**. We found that *FGFR1* gene expression and DNA methylation status were significantly negatively correlated (Pearson correlation r of -0.32, p value < 0.0001, **Fig. 7B)**, in TCGA samples, suggesting that *FGFR1* transcription was epigenetically regulated in a significant proportion of HPV-negative HNSCC tumors.

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Bossi et al. [41] collected gene expression data from cetuximab-treated HNSCC patients with recurrent metastasis with either short- (SPFS, median 3 months surival) or long-progression-free survival (LPFS, median 19 months survival). Using this dataset, we verified that *EGFR* expression in SPFS is significantly lower than the LPSF group (**Fig. 7C**) (log fold change -1.0, t-test p-value 0.0003). The opposite was observed for *FGFR1*, with overexpression in SPFS vs. LPSF (**Fig. 7D**, log fold change 0.9, t-test p-value 0.003**)**. However, Bossi et al.[41] lacked DNA methylation data to assess whether *FGFR1* was epigenetically regulated in these samples. Nonetheless, this finding in combination with the data from TCGA, supports our findings that the non-responder phenotype was accompanied by loss of *EGFR* expression and gain in *FGFR1* expression as a result of *FGFR1* promoter demethylation acquired during the development of cetuximab resistance.

DISCUSSION

Numerous short time course genomics studies of therapeutic response have been performed [45,47,48], but this is the first time that genetic and epigenetic changes were measured for a prolonged exposure (11 weeks) to a targeted therapeutic agent. Using our novel robust time course experimental approach, we characterized the molecular alterations during the development of acquired cetuximab resistance in HNSCC *in vitro*. By collecting cells over experimentally equivalent cultures (cetuximab and PBS control generations), we could measure changes in proliferation and multiple genomics data platforms as resistance developed. We applied this approach to the intrinsic cetuximab sensitive cell line SCC25 to track the molecular progression in acquired cetuximab resistance. This cell line was selected as one of only two HNSCC cell lines known to develop an isogenic resistance cell line with acquired cetuximab resistance [10]. Thus, this was the first study to our knowledge to enable characterization of the dynamics at the early stages of therapeutic resistance, which cannot be measured in patients due to the complexity of early detection of resistance and obtaining repeat biopsy samples.

Determining the dynamics of the molecular alterations responsible for resistance requires integrated, time-course bioinformatics analysis to quantify the dynamics of these alterations. Based upon previous

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

performance of Bayesian, non-negative matrix factorization algorithms in inferring dynamic regulatory networks for targeted therapeutics [48,49], we selected CoGAPS [35] for analysis of gene expression data from our time course experiment. In this dataset, CoGAPS analysis of gene expression data from cetuximab resistant clones distinguished the patterns for immediate gene expression changes and patterns for long-term changes associated with acquired resistance. Gene expression signatures for resistance to EGFR inhibitors in two additional cell lines (one HNSCC and one non-small cell lung cancer) from previous studies [37,39] were significantly enriched in both types of CoGAPS patterns. These previous resistance signatures were learned from case-control studies that compared gene expression for sensitive cells to that of resistant cells, without multiple time point measurements. Therefore, we concluded that time course data was instrumental in parsing signatures of immediate therapeutic response from signatures of acquired resistance.

Pooling cells to obtain paired measure of methylation and gene expression enabled us to evaluate whether changes in DNA methylation impact gene expression. CoGAPS analysis of DNA methylation data observed only changes associated with acquired resistance, in contrast to the immediate expression changes observed with cetuximab treatment. Thus, while therapeutic response can drive massive changes in gene expression, only the subset of expression changes associated with the development of resistance have corresponding epigenetic signatures suggesting that epigenetic landscape was important for the creation of acquired resistance. The CoGAPS patterns in gene expression that were associated with acquired cetuximab resistance gradually changed over the time course. On the other hand, the CoGAPS patterns for DNA methylation changes had a sharp transition at the generation at which resistance was acquired (CTX-G4). These patterns reflect a delayed, but more rapid change in DNA methylation. Our data is consistent with previous observations that gene expression changes precede DNA methylation alterations in genes critical for cancer progression. *P16*^{INK4A} and *GSTP1* are tumor suppressor genes for which transcription silencing was found to occur prior to DNA hypermethylation and chromatin changes. The temporal delay observed between expression and methylation patterns in our time course provides transcriptome wide evidence of this phenomena. Specifically, that epigenetic changes are necessary to stabilize gene expression aberrant profile and will be followed by modification into a silenced methylation state that will result in tumor progression [50,51]. Our integrated RNA-seq and DNA methylation analysis corroborated the fact that gene expression

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

changes occur earlier to epigenetic alterations and suggest that in acquired cetuximab resistance to cetuximab DNA methylation is essential to maintain the changes in gene expression. Future investigation into the chromatin remodeling mechanisms will test whether chromatin alterations follow the changes in expression and occur in combination with altered methylation patterns to drive epigenetic regulation of resistance.

In a recent study, gene expression changes are associated with a transient resistant phenotype present in melanoma cell lines prior to vemurafenib administration [52]. Once the melanoma cells are exposed to the drug, additional changes in gene expression are detected and are later accompanied by changes in chromatin structure [52]. These findings, together with our time course observations, suggest that in the heterogeneous tumor environment the existence of some cells expressing specific marker genes can trigger cellular reprogramming as soon as the targeted therapy is initiated. Upon drug administration, the number of genes with aberrant expression increases, and is followed by other epigenetic and genetic changes that will shift the transient resistant state into a stable phenotype. This finding on acquired resistance development could dramatically change the course of treatment with targeted therapeutic agents. The precise characterization of resistant gene signatures and their timing could be used to determine the correct point during the patients' clinical evolution to introduce alternative therapeutic strategies. This way, secondary interventions would start before the stable resistant phenotype is spread among the tumor cells resulting in prolonged disease control and substantial increased in overall survival.

The timing delays between alterations in DNA methylation and gene expression pose a further computational challenge for integrated, time course genomics analyses. The vast majority of integrated analysis algorithms assume one-to-one mapping of genes in different data platforms or seek common patterns or latent variables across them [53]. Such approaches would fail to capture the early changes from cetuximab treatment that impact only gene expression, time delays between DNA methylation and gene expression patterns, and different gene usage in each pattern. It is essential to develop new integrated algorithms to simultaneously distinguish both patterns that are shared across data types and that are unique to each platform. For time course data, these algorithms must also model regulatory relationships that may give rise to timing delays, such as epigenetic silencing of gene expression. However, as we observed with the unanticipated changes in

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

DNA methylation following and not preceding gene expression, they must also consider delays resulting from larger phenotypic changes such as the stability of the therapeutic resistant phenotype.

In spite of the complexities of the data integration, the weight of each sample in patterns inferred by CoGAPS reflected the dynamics of the process in each data modality. These patterns were learned completely unsupervised from the data, and did not require any gene selection or comparison between time points relative to any reference control. This analysis demonstrates that applying matrix factorization algorithms for compresses sensing in genomics can reconstruct signals associated with phenotypes from time course, omics data. The genes associated with CoGAPS patterns had weights that were non-zero in multiple patterns. The PatternMarker [25] statistic enabled further selection of the genes that were uniquely associated with each pattern. Creating a heatmap of the genomics profiles for these genes enabled novel, heatmap-based visualization of the temporal dynamics in the omics data. In the case of DNA methylation, these pattern marker genes also included genes representing driver alterations in resistance. However, transcriptional regulation by epigenetic alterations or in pathways involves simultaneous co-regulation of multiple genes. This co-regulation was reflected in the reuse of genes in CoGAPS gene weights associated in each pattern. Therefore, estimates of pathway dynamics from transcriptional data required accounting for all genes with gene set enrichment statistics instead of the PatternMarker statistic. Thus, we hypothesize that the PatternMarker statistic is robust for visualization, biomarker identification, and functional alterations in DNA over time, whereas it is robust only for visualization of and biomarker selection from time-course transcriptional data.

Among the genes we observed with the canonical relationship between expression and methylation, *FGFR1* presented with loss of CpG methylation accompanied by increased gene expression. *FGFR1* is a receptor tyrosine kinase that regulates downstream pathways, such as PI3K/AKT, and RAS/MAPK, that are also regulated by *EGFR* [54]. Its overexpression has been previously associated with EGFR inhibitors resistance [26–28]. To our knowledge this is the first study showing epigenetic regulation of *FGFR1* in HNSCC and the association of that epigenetic regulation with acquired cetuximab resistance. In this case, *FGFR1* induction through promoter demethylation in concordance with down regulation of *EGFR* appears to be the dominant mechanism. The novel cell culture protocol and time course analysis we developed here is what enabled us

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

to see the clonal outgrowth of this particular mechanism. These results are also relevant for further translational studies into the role of *FGFR1* as a potential biomarker of acquired cetuximab resistance and potential target to overcome that resistance. *FGFR1* is a potential target for combined targeted therapy with *EGFR*, and inhibitors against this target are already the focus of clinical trials [54].

The main limitation of the current study was the use of a single cell line model to induce resistance and collect the time course data for gene expression and epigenetics analysis. SCC25 is intrinsically sensitive to cetuximab and one of the only two available models in HNSCC that are known to acquire resistance to cetuximab from long term exposure [9]. From a single cell line model, we generated two groups of samples (CTX and PBS generations) over the course of 11 weeks. High throughput measurements and analysis were performed for a total of 22 samples. The collection of multiple data points in the analysis had to be accounted for when determining the number of cell models to be included. We nonetheless compare our data to gene signatures from the other isogenic HNSCC resistance model 1CC8 [10] and an independent resistance model to an EGFR inhibitor in non-small cell lung cancer [39]. Besides the number of samples, we also had to take into consideration the potential batch and technical effects of broad cross-platform profiling. Nevertheless, the analysis of HNSCC patient samples from TCGA [33] and another study [41] validated our finding that *FGFR1* is up-regulated and demethylated in HNSCC and associated with resistance to cetuximab.

The *in vitro* protocol for time course sampling developed in this study has the additional advantage of aggregating potentially heterogeneous mechanisms of resistance increasing the signal of changes in any cetuximab resistant subclone. For example, we observed epigenetic regulation of *FGFR1* in the pooled cells, but only a single stable clone generated from the same SCC25 cell line in a previous study (CTXR10) had upregulation of FGFR1 [30]. This finding suggests that tumor heterogeneity also plays a role in acquired resistance to target therapies and enables different pathways to be used to bypass the silenced target within the same tumor. The heterogeneity in methylation profiles reflected the complexity of the resistance mechanisms that can arise from combination therapies in heterogeneous tumors. Future work extending these protocols to *in vivo* models is essential to determine the role of the microenvironment in inducing therapeutic resistance. Developing *in vivo* models with acquired therapeutic resistance presents numerous

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

technical challenges that must first be addressed before such time course sampling is possible [9]. Pinpointing precise molecular predictors of therapeutic resistance will facilitate the identification of unprecedented biomarkers and reveal the mechanisms by which to overcome acquired therapeutic resistance to most therapies used to treat cancer.

CONCLUSIONS

By developing a new experimental protocol and using robust bioinformatics tools, we measured the changes in gene expression and DNA methylation during the progression from an intrinsic responsive state to cetuximab to the acquired resistant phenotype. Using the CoGAPS matrix factorization algorithm, we observed massive changes in gene expression and identified and discriminated the different patterns associated with resistance or cell culturing conditions. This analysis demonstrates that compressed sensing matrix factorization algorithms can identify gene signatures associated with the dynamics of phenotypic changes from time-course, genomics data. In this case, the gene expression patterns relevant to resistance were later followed by epigenetic alterations. Our main conclusion is that the resistant phenotype is driven by gene expression changes that would confer the cancer cells adaptive advantages to the treatment with cetuximab. Finally, the stability of the resistant state is dependent on epigenetic changes that will make these new gene signatures heritable to expand the phenotype to the daughter cells. This finding is significant to clinical practice, since it suggests the resistant phenotype could be reversed if alternative interventions are used in combination from the beginning or introduced before epigenetic alterations to the genes driving acquired resistance.

List of abbreviations

AKT – AKT serine/threonine kinase
ATCC – American Type Culture Collection
CoGAPS – Coordianted Gene Activity in Pattern Sets
CTX - cetuximab
CTXR – single cell cetuximab resistant clone
DNA – deoxyribonucleic acid

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

EGFR – Endogenous Growth Factor Receptor
FDA – Food and Drug Administration
FGFR1 – Fibroblast Growth Factor Receptor 1
GEO – Gene Expression Omnibus
GSTP1 – Glutathione S-Transferase pi 1
GWCoGAPS – Genome Wide Coordinated Gene Activity in Pattern Sets
HNSCC – Head and neck squamous cell carcinoma
HRAS – HRAS proto-oncogene
JHMI – Johns Hopkins Medical Institutions
LPFS – Long-progression-free survival
MAPK – Mitogen activated kinase-like protein
PBS – Phosphate buffered saline
PI3K – Phosphoinositide-3-kinase regulatory subunit 1
RIN – RNA integrity number
RNA – Ribonucleic acid
RNA-seq – Ribonucleic acid sequencing
rRNA – Ribosomal ribonucleic acid
SNP – Single nucleotide polymorphism
SPFS – Short-progression-free survival
STR – Short tandem repeat
TCGA – The Cancer Genome Atlas
TFAP2A – Transcription factor AP-2 alpha

Acknowledgements

We thank JHMI Deep Sequencing & Microarray Core and SKCCC Microarray Core Facility on performing and providing advice on RNA-Seq and DNA methylation hybridization arrays, respectively; S. Boca, B. Kerr, S. Floor, C. Mak, T. Ou, D. Sidransky, L. M. Weiner, F. Zamuner, K. Zambo, and members of NewPISlack for critical comments and feedback during the preparation of the manuscript.

DECLARATIONS

Funding

This work was supported by NIH Grants R01CA177669, R21DE025398, P30 CA006973. R01 DE017982, and SPORE P50DE019032.

Competing interests

None of the authors have competing interests to declare.

Author contributions

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

G.S., L.T.K, S.L., C.H.C. and E.J.F. planned, designed and wrote the manuscript with input from all authors. G.S., L.T.K., S.L., M.T., C.H.C. and E.J.F. contributed to the development of methodology. G.S., S.L. and E.J.F. performed analysis and interpretation of data (e.g., computational analysis). R.R., H.O., H.C., M.C., A.F., L.V.D., J.A., D.A.G. participated in development of methodology and provided technical and material support. R.R., L.V.D., E.I. and D.A.G. participated in review, and/or revision of the manuscript. All authors discussed the data and contributed to the manuscript preparation. C.H.C. and E.J.F. instigated and supervised the project.

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Sawyers C. Targeted cancer therapy. Nature. 2004;432:294–7.

2. Hyman DM, Taylor BS, Baselga J. Implementing Genome-Driven Oncology. Cell. 2017;168:584–99.

3. Hata AN, Niederst MJ, Archibald HL, Gomez-Caraballo M, Siddiqui FM, Mulvey HE, et al. Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition. Nat. Med. 2016;22:262–9.

4. Engelman JA, Zejnullahu K, Mitsudomi T, Song Y, Hyland C, Park JO, et al. MET Amplification Leads to Gefitinib Resistance in Lung Cancer by Activating ERBB3 Signaling. Science. 2007;316:1039–43.

5. Misale S, Yaeger R, Hobor S, Scala E, Janakiraman M, Liska D, et al. Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. Nature. 2012;486:532–6.

6. Pietrantonio F, Vernieri C, Siravegna G, Mennitto A, Berenato R, Perrone F, et al. Heterogeneity of Acquired Resistance to Anti-EGFR Monoclonal Antibodies in Patients with Metastatic Colorectal Cancer. Clin. Cancer Res. 2017;23:2414–22.

7. Vincenzi B, Zoccoli A, Pantano F, Venditti O, Galluzzo S. Cetuximab: from bench to bedside. Curr. Cancer Drug Targets. 2010;10:80–95.

8. Boeckx C, Weyn C, Vanden Bempt I, Deschoolmeester V, Wouters A, Specenier P, et al. Mutation analysis of genes in the EGFR pathway in Head and Neck cancer patients: implications for anti-EGFR treatment response. BMC Res. Notes. 2014;7:337.

9. Quesnelle KM, Wheeler SE, Ratay MK, Grandis JR. Preclinical modeling of EGFR inhibitor resistance in head and neck cancer. Cancer Biol. Ther. 2012;13:935–45.

10. Wheeler DL, Huang S, Kruser TJ, Nechrebecki MM, Armstrong EA, Benavente S, et al. Mechanisms of acquired resistance to cetuximab: role of HER (ErbB) family members. Oncogene. 2008;27:3944–56.

11. Narayan M, Wilken JA, Harris LN, Baron AT, Kimbler KD, Maihle NJ. Trastuzumab-Induced HER Reprogramming in "Resistant" Breast Carcinoma Cells. Cancer Res. 2009;69:2191–4.

12. Ernst J, Nau GJ, Bar-Joseph Z. Clustering short time series gene expression data. Bioinforma. Oxf. Engl. 2005;21 Suppl 1:i159-168.

13. Aryee MJ, Gutiérrez-Pabello JA, Kramnik I, Maiti T, Quackenbush J. An improved empirical bayes approach to estimating differential gene expression in microarray time-course data: BETR (Bayesian Estimation of Temporal Regulation). BMC Bioinformatics. 2009;10:409.

14. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. 2004;3:Article3.

15. Lin D, Shkedy Z, Yekutieli D, Burzykowski T, Göhlmann HWH, De Bondt A, et al. Testing for trends in dose-response microarray experiments: a comparison of several testing procedures, multiplicity and resampling-based inference. Stat. Appl. Genet. Mol. Biol. 2007;6:Article26.

16. Fernández MA, Rueda C, Peddada SD. Identification of a core set of signature cell cycle genes whose relative order of time to peak expression is conserved across species. Nucleic Acids Res. 2012;40:2823–32.

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

17. Ernst J, Vainas O, Harbison CT, Simon I, Bar-Joseph Z. Reconstructing dynamic regulatory maps. Mol. Syst. Biol. 2007;3:74.

18. Bonneau R, Reiss DJ, Shannon P, Facciotti M, Hood L, Baliga NS, et al. The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. Genome Biol. 2006;7:R36.

19. Liao JC, Boscolo R, Yang Y-L, Tran LM, Sabatti C, Roychowdhury VP. Network component analysis: reconstruction of regulatory signals in biological systems. Proc. Natl. Acad. Sci. U. S. A. 2003;100:15522–7.

20. Seok J, Xiao W, Moldawer LL, Davis RW, Covert MW. A dynamic network of transcription in LPS-treated human subjects. BMC Syst. Biol. 2009;3:78.

21. Naegle KM, Welsch RE, Yaffe MB, White FM, Lauffenburger DA. MCAM: multiple clustering analysis methodology for deriving hypotheses and insights from high-throughput proteomic datasets. PLoS Comput. Biol. 2011;7:e1002119.

22. Wise A, Bar-Joseph Z. SMARTS: reconstructing disease response networks from multiple individuals using time series gene expression data. Bioinformatics. 2015;31:1250–7.

23. Ochs MF, Fertig EJ. Matrix factorization for transcriptional regulatory network inference. IEEE Symp Comput Intell Bioinforma Comput Biol Proc [Internet]. 2012. p. 387–96. Available from: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6217256

24. Ogawa T, Liggett TE, Melnikov AA, Monitto CL, Kusuke D, Shiga K, et al. Methylation of death-associated protein kinase is associated with cetuximab and erlotinib resistance. Cell Cycle Georget. Tex. 2012;11:1656–63.

25. Stein-O'Brien G, Carey J, Lee W, Considine M, Favorov A, Flam E, et al. PatternMarkers and Genome-Wide CoGAPS Analysis in Parallel Sets (GWCoGAPS) for data-driven detection of novel biomarkers via whole transcriptome Non-negative matrix factorization (NMF). bioRxiv. 2016;83717.

26. Azuma K, Kawahara A, Sonoda K, Nakashima K, Tashiro K, Watari K, et al. FGFR1 activation is an escape mechanism in human lung cancer cells resistant to afatinib, a pan-EGFR family kinase inhibitor. Oncotarget. 2014;5:5908–19.

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

28. Koole K, Brunen D, van Kempen PMW, Noorlag R, de Bree R, Lieftink C, et al. FGFR1 Is a Potential Prognostic Biomarker and Therapeutic Target in Head and Neck Squamous Cell Carcinoma. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 2016;22:3884–93.

29. Hatakeyama H, Cheng H, Wirth P, Counsell A, Marcrom SR, Wood CB, et al. Regulation of heparin-binding EGF-like growth factor by miR-212 and acquired cetuximab-resistance in head and neck squamous cell carcinoma. PloS One. 2010;5:e12702.

30. Cheng H, Fertig EJ, Ozawa H, Hatakeyama H, Howard JD, Perez J, et al. Decreased SMAD4 expression is associated with induction of epithelial-to-mesenchymal transition and cetuximab resistance in head and neck squamous cell carcinoma. Cancer Biol. Ther. 2015;16:1252–8.

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

31. Wang K, Singh D, Zeng Z, Coleman SJ, Huang Y, Savich GL, et al. MapSplice: Accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Res. 2010;38:e178–e178.

32. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.

33. Lawrence MS, Sougnez C, Lichtenstein L, Cibulskis K, Lander E, Gabriel SB, et al. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature. 2015;517:576–82.

34. Fortin J-P, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol. [Internet]. 2014 [cited 2017 Feb 27];15. Available from: http://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0503-2

35. Fertig EJ, Ding J, Favorov AV, Parmigiani G, Ochs MF. CoGAPS: an R/C++ package to identify patterns and biological process activity in transcriptomic data. 2010; Available from: http://bioinformatics.oxfordjournals.org/content/26/21/2792.short

36. Bidaut G. Interpreting and Comparing Clustering Experiments Through Graph Visualization and Ontology Statistical Enrichment with the ClutrFree Package. link.springer.com [Internet]. Boston, MA: Springer US; 2010. p. 315–33. Available from: http://www.springerlink.com/index/10.1007/978-1-4419-5714-6_19

37. Fertig EJ, Ren Q, Cheng H, Hatakeyama H, Dicker AP, Rodeck U, et al. Gene expression signatures modulated by epidermal growth factor receptor activation and their relationship to cetuximab resistance in head and neck squamous cell carcinoma. BMC Genomics. 2012;13:160.

38. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. 2005;102:15545–50.

39. Coldren CD. Baseline Gene Expression Predicts Sensitivity to Gefitinib in Non-Small Cell Lung Cancer Cell Lines. Mol. Cancer Res. 2006;4:521–8.

40. Matys V, Fricke E, Geffers R, Gössling E, Haubrock M, Hehl R, et al. TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic Acids Res. 2003;31:374–8.

41. Bossi P, Bergamini C, Siano M, Cossu Rocca M, Sponghini AP, Favales F, et al. Functional Genomics Uncover the Biology behind the Responsiveness of Head and Neck Squamous Cell Cancer Patients to Cetuximab. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 2016;22:3961–70.

42. Fertig EJ, Ozawa H, Thakar M, Howard JD, Kagohara LT, Krigsfeld G, et al. CoGAPS matrix factorization algorithm identifies transcriptional changes in AP-2alpha target genes in feedback from therapeutic inhibition of the EGFR network. Oncotarget [Internet]. 2016;5. Available from: http://www.impactjournals.com/oncotarget/index.php?journal=oncotarget&page=article&op=view&path%5B%5D=12075&path%5B%5D=38216

43. Fertig EJ, Ren Q, Cheng H, Hatakeyama H, Dicker AP, Rodeck U, et al. Gene expression signatures modulated by epidermal growth factor receptor activation and their relationship to cetuximab resistance in head and neck squamous cell carcinoma. BMC Genomics. 2012;13:160.

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

44. Fertig EJ, Ozawa H, Thakar M, Howard JD, Kagohara LT, Krigsfeld G, et al. CoGAPS matrix factorization algorithm identifies transcriptional changes in AP-2alpha target genes in feedback from therapeutic inhibition of the EGFR network. Oncotarget. 2016;7:73845–64.

45. Hafner M, Niepel M, Chung M, Sorger PK. Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. Nat. Methods. 2016;13:521–7.

46. Harris LA, Frick PL, Garbett SP, Hardeman KN, Paudel BB, Lopez CF, et al. An unbiased metric of antiproliferative drug effect in vitro. Nat. Methods. 2016;13:497–500.

47. Michna A, Schötz U, Selmansberger M, Zitzelsberger H, Lauber K, Unger K, et al. Transcriptomic analyses of the radiation response in head and neck squamous cell carcinoma subclones with different radiation sensitivity: time-course gene expression profiles and gene association networks. Radiat. Oncol. [Internet]. 2016 [cited 2017 Mar 26];11. Available from: http://ro-journal.biomedcentral.com/articles/10.1186/s13014-016-0672-0

48. Ochs MF, Rink L, Tarn C, Mburu S, Taguchi T, Eisenberg B, et al. Detection of Treatment-Induced Changes in Signaling Pathways in Gastrointestinal Stromal Tumors Using Transcriptomic Data. Cancer Res. 2009;69:9125–32.

49. Hill SM, Heiser LM, Cokelaer T, Unger M, Nesser NK, Carlin DE, et al. Inferring causal molecular networks: empirical assessment through a community-based effort. Nat. Methods. 2016;13:310–8.

50. Bachman KE, Park BH, Rhee I, Rajagopalan H, Herman JG, Baylin SB, et al. Histone modifications and silencing prior to DNA methylation of a tumor suppressor gene. Cancer Cell. 2003;3:89–95.

51. Stirzaker C, Song JZ, Davidson B, Clark SJ. Transcriptional gene silencing promotes DNA hypermethylation through a sequential change in chromatin modifications in cancer cells. Cancer Res. 2004;64:3871–7.

52. Shaffer SM, Dunagin MC, Torborg SR, Torre EA, Emert B, Krepler C, et al. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. Nature. 2017;546:431–5.

53. Tyekucheva S, Marchionni L, Karchin R, Parmigiani G. Integrating diverse genomic data using gene sets. Genome Biol. 2011;12:R105.

54. Babina IS, Turner NC. Advances and challenges in targeting FGFR signalling in cancer. Nat. Rev. Cancer. 2017;

FIGURES LEGENDS

Figure 1 – ***In vitro*** **time course reflects clinical evolution of cetuximab response and evolution of acquired resistance.** Intrinsic cetuximab sensitive HNSCC cell line SCC25 was treated with cetuximab (red) or PBS (black) for 11 generations to develop acquired resistance. Proliferation assay (flow) of cetuximab treatment (red line) and PBS treated cells (black line) measured cetuximab response for all SCC25 generations (bottom). Treatment response was

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

divided into three stages based upon the measured proliferation rates and clinical stages (top). While proliferation of the PBS generations was stable throughout the eleven weeks, proliferation of the CTX generations progressively increased over each week. Relative to the untreated controls, the growth of the treated cells was initially (CTX-G1) inhibited until CTX-G3. Starting at CTX-G4, the cells became resistant to the anti-proliferative effects of cetuximab and gained stable growth advantages compared to the untreated controls.

Figure 2 – CoGAPS analysis identifies signatures of resistance to EGFR inhibitors and separate resistant and control generations. (A) Heatmap of gene expression values in 11 generations of SCC25 cells treated with 100nM of cetuximab (red columns) to acquire resistance and with PBS as control (black columns). (**B)** Heatmap of gene expression values for PatternMarker genes identified with CoGAPS analysis of gene expression data from 11 generations of SCC25 cells treated with PBS as control (black columns) and with 100nM of cetuximab (red columns) to acquire resistance. Rows are colored according to which CoGAPS pattern the PatternMarker statistic assigned each gene, and sorted by the PatternMarker statistic. (**C)** CoGAPS patterns inferred from gene expression data over generations of PBS control (black lines) or treatment with 100nM of cetuximab (red lines).

Figure 3 – Dynamics of DNA methylation alterations and association with gene expression patterns in acquired cetuximab resistance. (A) Heatmap of DNA methylation values in 11 generations of SCC25 cells treated with PBS as control (black columns) and with 100nM of cetuximab (red columns) to acquire resistance. **(B)** Heatmap of DNA methylation values for genes selected by CoGAPS DNA methylation patterns analysis in the same SCC25 cetuximab and PBS generations. **(C)** CoGAPS patterns inferred from DNA methylation data over generations of PBS control (black lines) or treatment with 100nM of cetuximab (red lines).

Figure 4 – DNA methylation and expression CoGAPS patterns demonstrate delayed onset of epigenetic changes in acquired resistance. CoGAPS patterns for gene expression (top) and DNA methylation (bottom) of patterns associated with acquired cetuximab resistance in SCC25 cetuximab generations (red) relative to PBS generations (black). Vertical dashed line represents time at which patterns for SCC25 generation separated from pattern for PBS generations. The timing of methylation changes distinguishing cetuximab resistant generations was delayed in DNA methylation relative to that of gene expression.

Figure 5 – Clonal heterogeneity does not reflect signature of epigenetically regulated genes observed in bulk time-course analysis. (A) Heatmap of gene expression values for DNA

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

methylation PatternMarker genes for acquired resistance that were anti-correlated between expression and DNA methylation (**Fig. 4D**). Data includes 11 generations of SCC25 cells treated with PBS as control (black columns labeled PBS) and with 100nM of cetuximab (red columns labeled cetuximab) to acquire resistance and gene expression data from independent, stable cetuximab resistant clones in absence of cetuximab treatment (CTX resistant clones). Gene expression heatmap on a red-yellow scale indicated in the color key. **(B)** Heatmap of DNA methylation data in conditions described in (a), on a blue-yellow scale indicated in the color key.

Figure 6 – Epigenetic regulation of *FGFR1* in acquired cetuximab resistance is validated in stable SCC25 resistant clones. (A) Expression of *FGFR1* gene expression relative to DNA methylation in stable cetuximab resistant clones. **(B)** QRT-PCR of *FGFR1* gene expression in CTXR10 relative to the parental cell line (greater than 30 fold change). **(C)** Western blot comparing FGFR1, phosphor-FGFR1, EGFR, and phospho-EGFR in CTXR10 relative to the parental SCC25 cell line. In the resistant cell clone, increased levels of FGFR1 were associated with increased levels of phospho-FGFR1 and decrease in EGFR and phospho-EGFR.

Figure 7 – FGFR1 gene expression and DNA methylation patterns were confirmed in independent HNSCC tumor samples datasets. (A) Scatter plot of gene expression for *EGFR* and *FGFR1* in HPV-negative HNSCC samples from TCGA demonstrated that only a few HNSCC cases present increased levels of both genes and that there is no significant correlation between the expression of both genes concomitantly. **(B)** DNA methylation of *FGFR1* was anti-correlated with *FGFR1* expression in HPV-negative HNSCC TCGA samples, suggesting that up-regulation of FGFR1 might be a result of promoter hypomethylation in primary tumors. **(C)** *EGFR* expression was significantly overexpressed in a group of HNSCC patients with long progression free survival relative to patients with short progression free survival in gene expression data from Bossi et al. **(D)** *FGFR1* was significantly overexpressed in patients with short progression free survival relative to patients with long progression free survival in this same dataset.

Supplementary Figures Captions

Supplemental Figure 1 – Time course approach to induce resistance to cetuximab and measure gene expression and DNA methylation changes. Intrinsic cetuximab sensitive HNSCC cell line SCC25 were treated with cetuximab (red) or PBS (black) for 7 days. In the eighth day, cells

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

were collected and pooled from multiple replicate cultures to provide adequate amounts for total RNA isolation for RNA-seq, genomic DNA isolation for DNA methylation array, proliferation assay (flow), for storage (frozen) and to be plated again to continue treatment until resistance to cetuximab developed. Each collection point was called a generation (from CTX-G0 to CTX-G11).

Supplemental Figure 2 – Colony formation assay in matrigel for anchorage-independent growth confirmed acquired cetuximab resistance of CTX-G10 (red) relative to the parental cell line (CTX-G0, black) at different concentrations of cetuximab (0nM, 10nM, 100nM and 1000nM).

Supplemental Figure 3 – Heatmap and hierarchical clustering of gene expression values in 11 generations of SCC25 cells treated with PBS as control (black columns) and with 100nM of cetuximab (red columns) to acquire resistance.

Supplemental Figure 4 – **A.** Heatmap of gene expression values in 11 generations of SCC25 cells treated with 100nM of cetuximab (red columns) to acquire resistance and with PBS as control (black columns). Genes selected for visualization were associated with cetuximab resistance from previous gene expression studies comparing sensitive and resistant cells without regard for timing. These studies provided three gene sets, colored along rows of the heatmap. **B.** Average of z-score gene expression values for genes in each of the resistance signatures over generations of PBS control (black lines) or treatment with 100nM of cetuximab (red lines).

Supplemental Figure 5 – Expected gene expression values for genes in each CoGAPS pattern inferred from gene expression data over generations of PBS control (black lines) or treatment with 100nM of cetuximab (red lines). Patterns included a pattern reflecting technical artifacts between untreated controls at time 0 and subsequent generations (pattern 4) and a flat pattern for highly expressed genes (pattern 5), excluded from analysis in main figures. Heatmap of gene expression values for PatternMarker genes identified for all of these patterns. Rows were colored according to which CoGAPS pattern the PatternMarker statistic assigned each gene, and sorted by the PatternMarker statistic.

Supplemental Figure 6 – Heatmap of gene set analysis scores for targets of transcription factors in the EGFR network, targets of the AP-2alpha transcription factors associated with cetuximab response, and cetuximab resistance signatures in CoGAPS patterns. A score of 100 indicated upregulation of the targets with a p-value of 0 and -100 downregulation with p-values of 0. Matrix elements with a star indicated p-values below 0.05 for either up or down-regulation of the gene set.

bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Gene expression heatmap was colored on a red-green scale where as the gene set statistics heatmap was colored on a blue-red scale, with values indicated in the respective color keys.

Supplemental Figure 7 – **A.** Heatmap of Pearson correlation coefficients between CoGAPS gene expression and DNA methylation patterns. Row colors for expression patterns match the colors for patterns in Figure 2,3. The column colors for methylation patterns are selected to match the color of the corresponding expression pattern with maximum anti-correlation. **B.** As in **A** for CoGAPS gene weights (meta-pathway values) corresponding to patterns in DNA methylation (columns) and gene expression (rows).

Supplemental Figure 8 – Heatmap of gene expression values for 11 generations of SCC25 cells treated with PBS as control (black columns labeled PBS) and with 100nM of cetuximab (red columns labeled cetuximab) to acquire resistance and gene expression data from independent, stable cetuximab resistant clones in absence of cetuximab treatment (CTX resistant clones).

Supplemental Figure 9 – Heatmap of DNA methylation values for 11 generations of SCC25 cells treated with PBS as control (black columns labeled PBS) and with 100nM of cetuximab (red columns labeled cetuximab) to acquire resistance and gene expression data from independent, stable cetuximab resistant clones in absence of cetuximab treatment (CTX resistant clones).

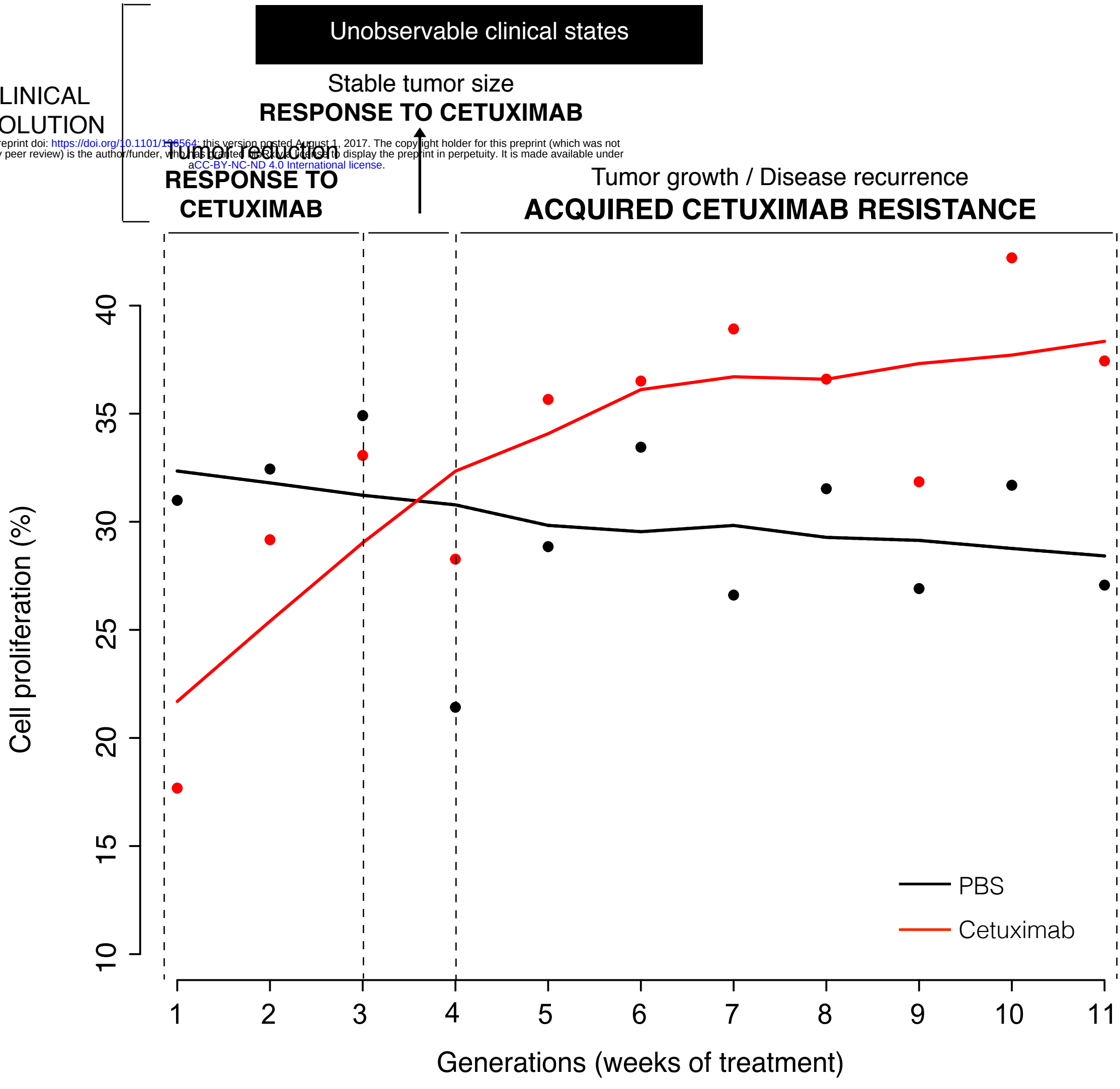
Supplemental Figure 10 – Brightfield microscopy images of the representative clones' morphology above with corresponding clone specific CoGAPS patterns of DNA methylation above.

Supplemental Figure 11 – Epigenetically regulated pattern marker genes associated with resistance having significant anti-correlation between gene expression and DNA methylation in the cetuximab single cell resistant clones.

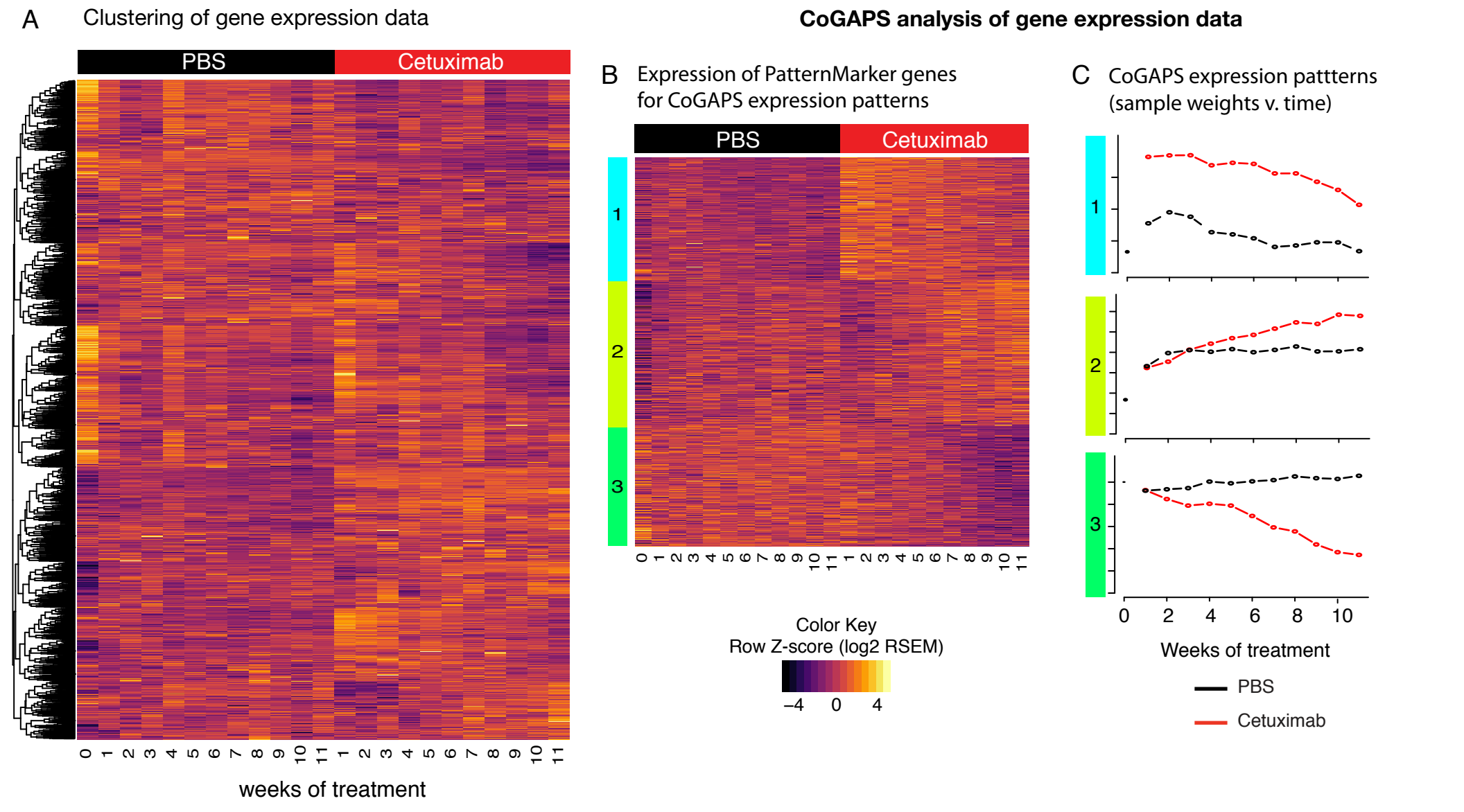
Supplemental Figure 12 – Cell proliferation assay using AlamarBlue (Invitrogen, Carlsbad, CA) to compare proliferation rates under different concentrations of cetuximab in the resistant single cell clones (CTXR4, 7, 10 and 11) and the parental SCC25 cell line to confirm resistance when treated with different concentrations of cetuximab.

Proliferation assay

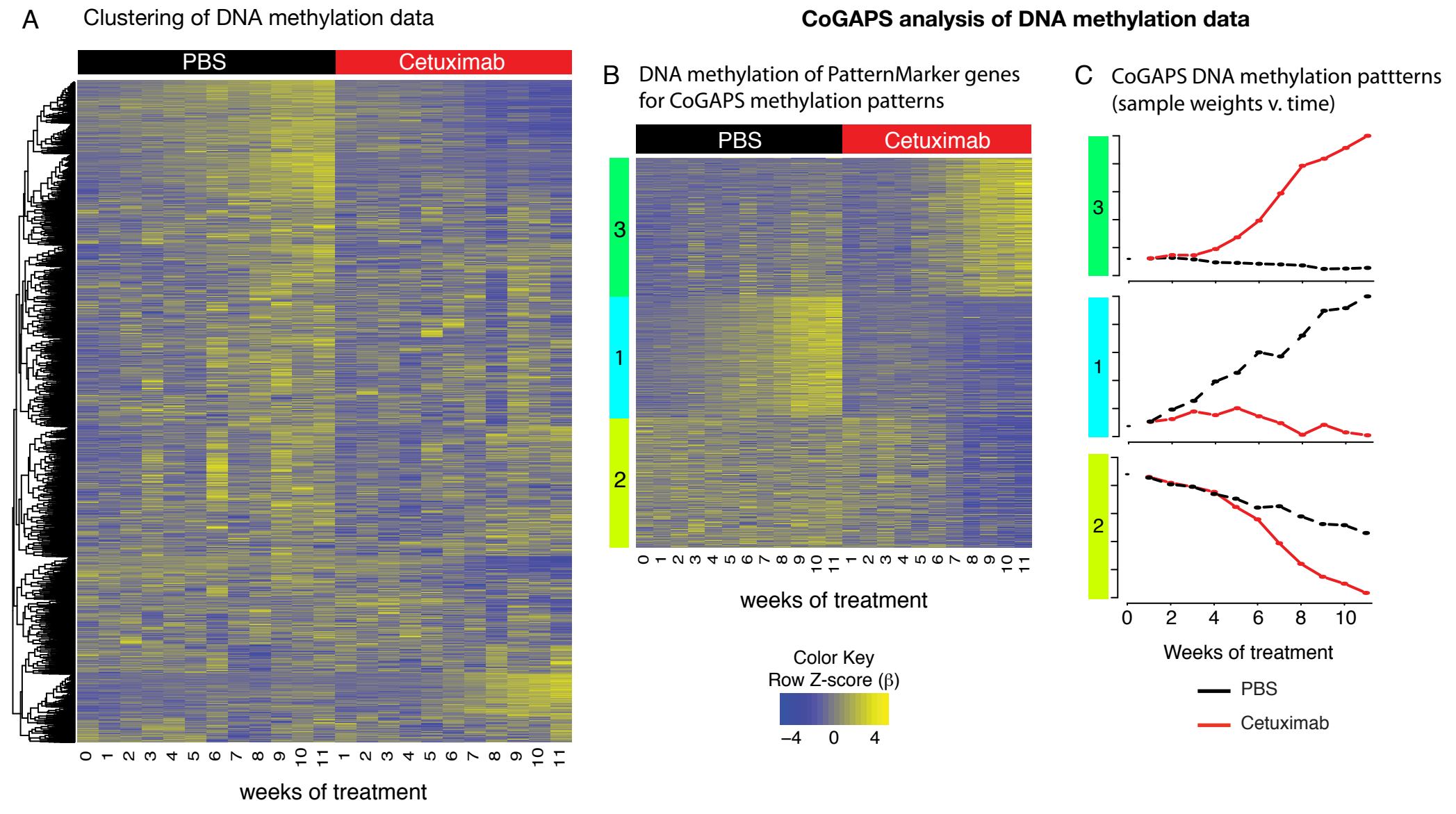
bioRxiv preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

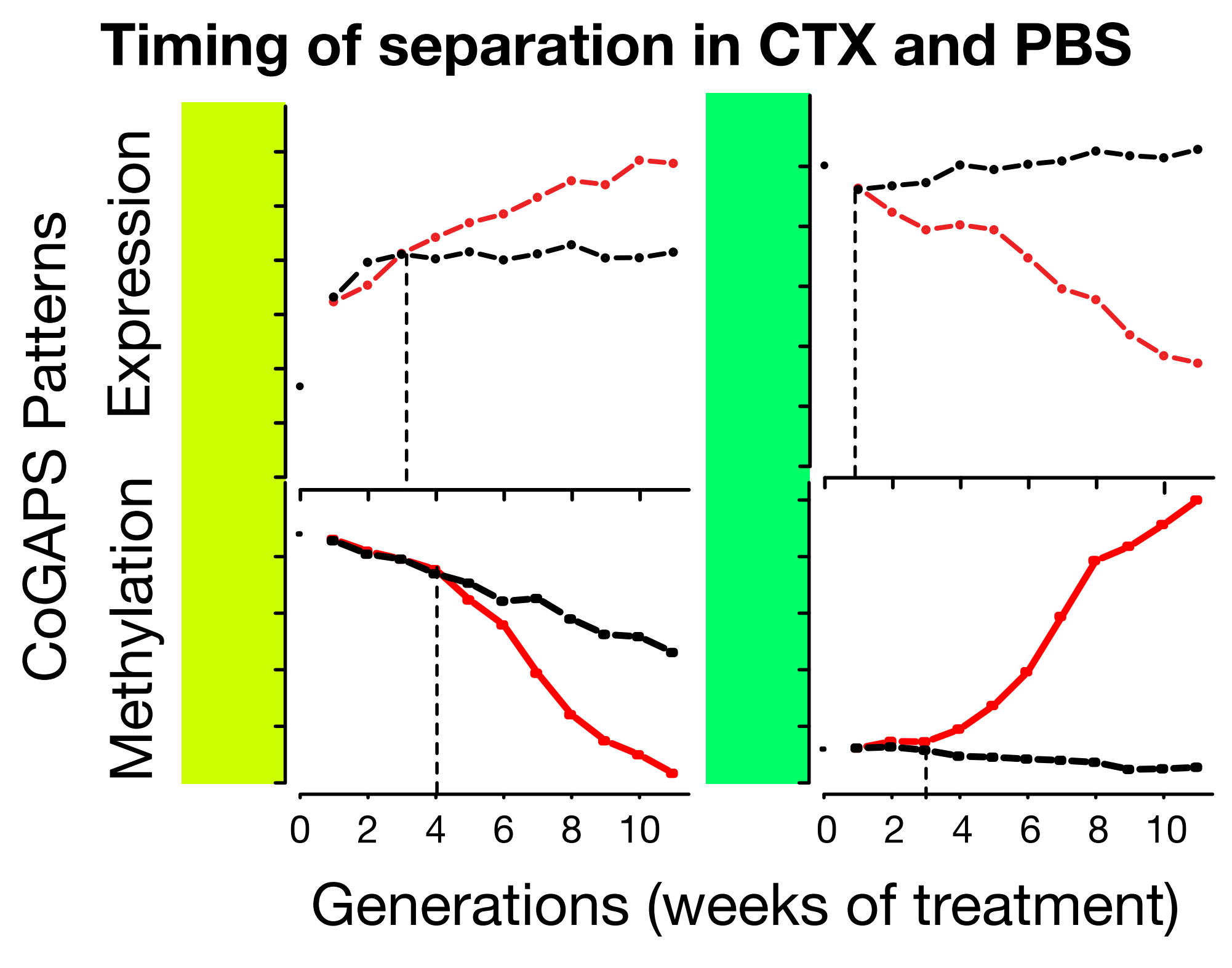


CoGAPS analysis of gene expression data

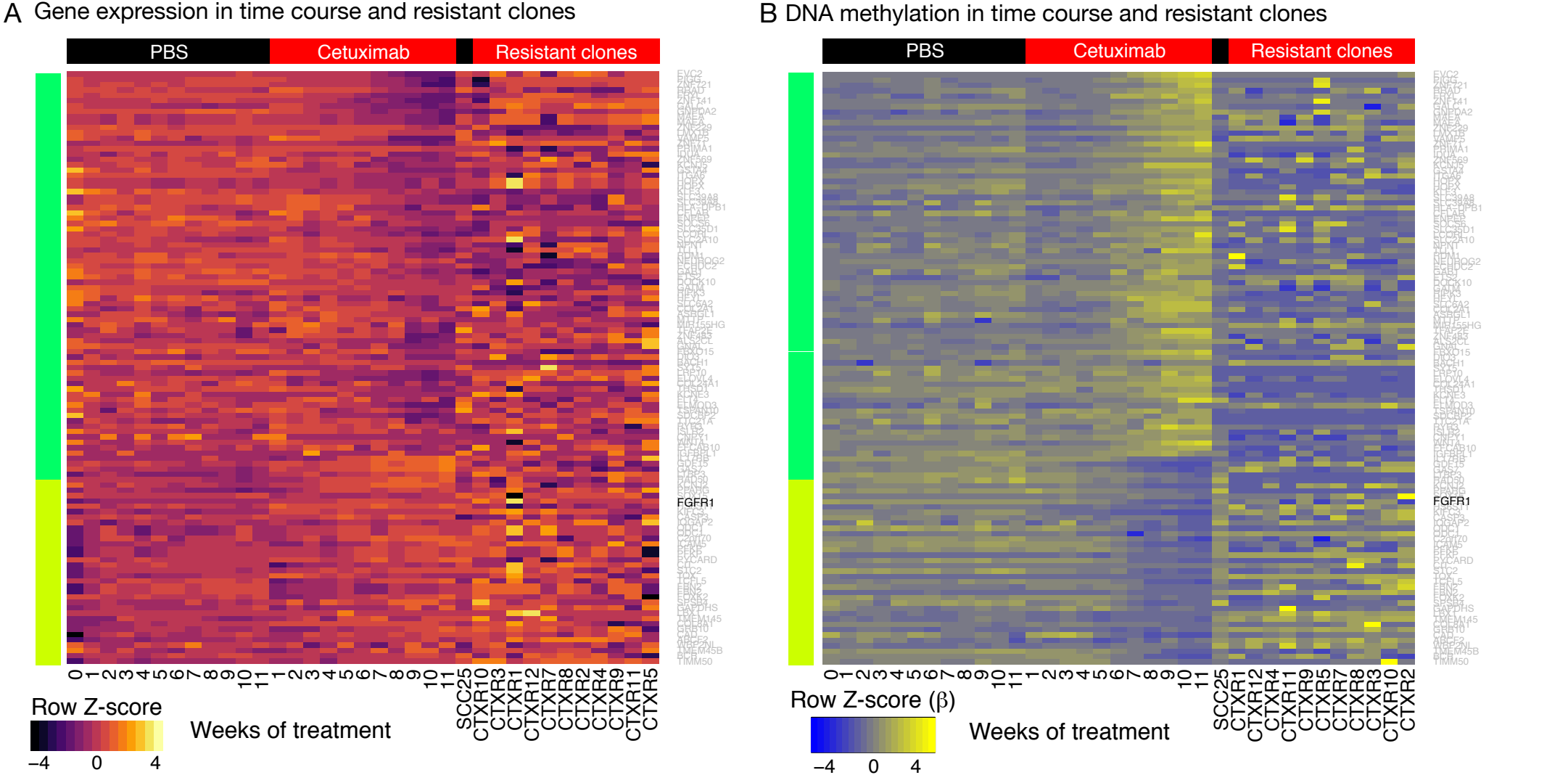


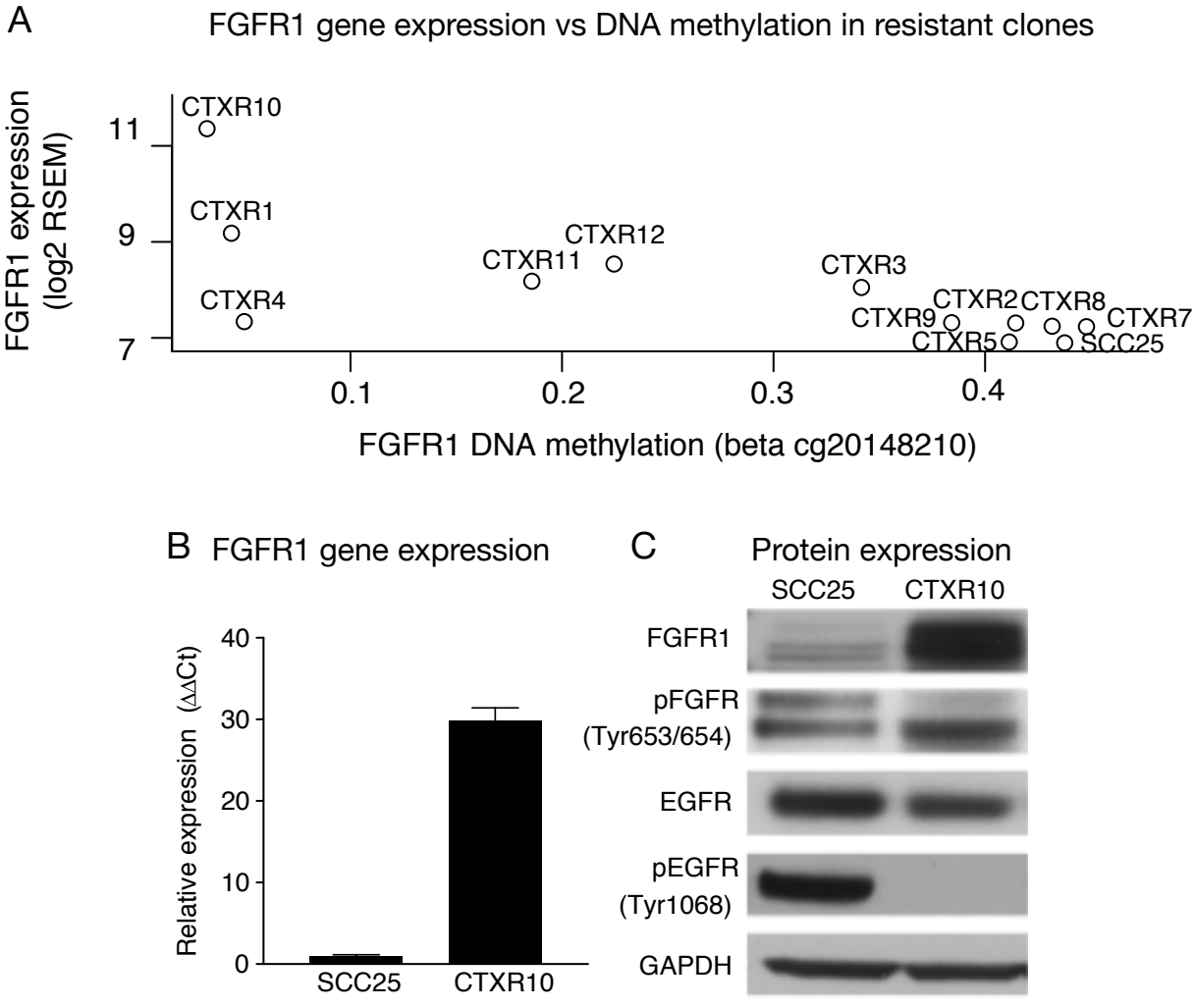
CoGAPS analysis of DNA methylation data




Timing of separation in CTX and PBS
CoGAPS Patterns
Expression
Methylation
0
2
4
6
8
10
Generations (weeks of treatment)

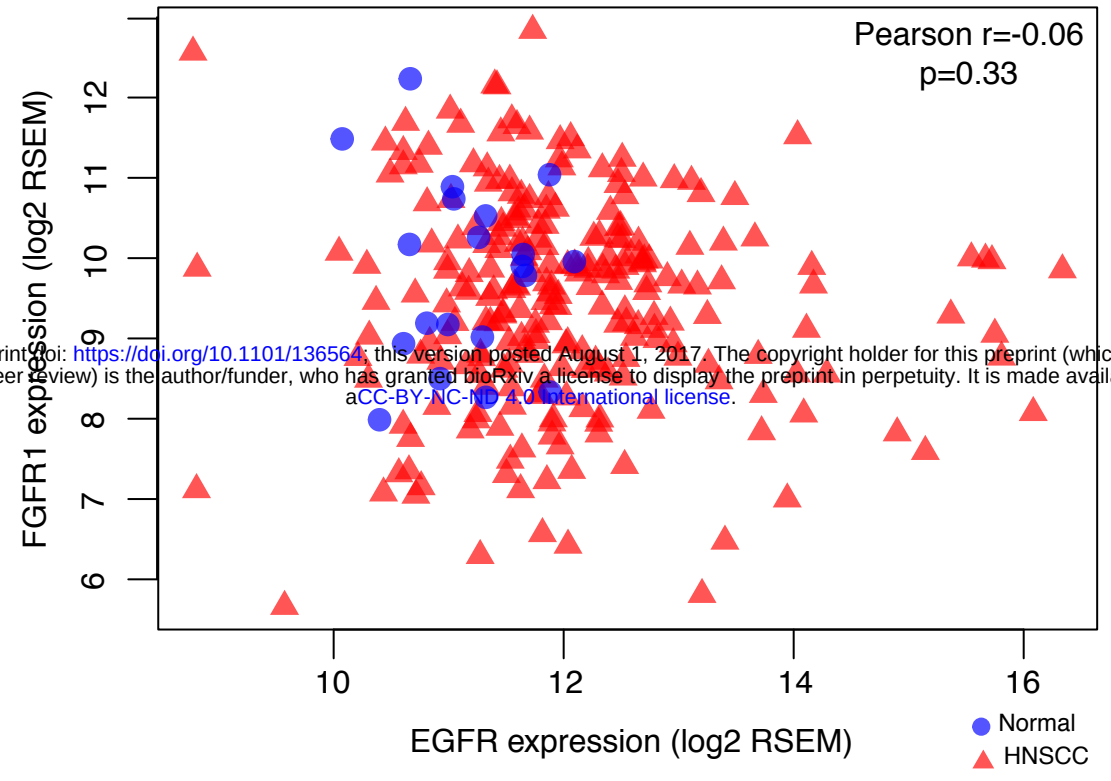
PatternMarker genes for CoGAPS methylation patterns that are anti-correlated with expression in time course data



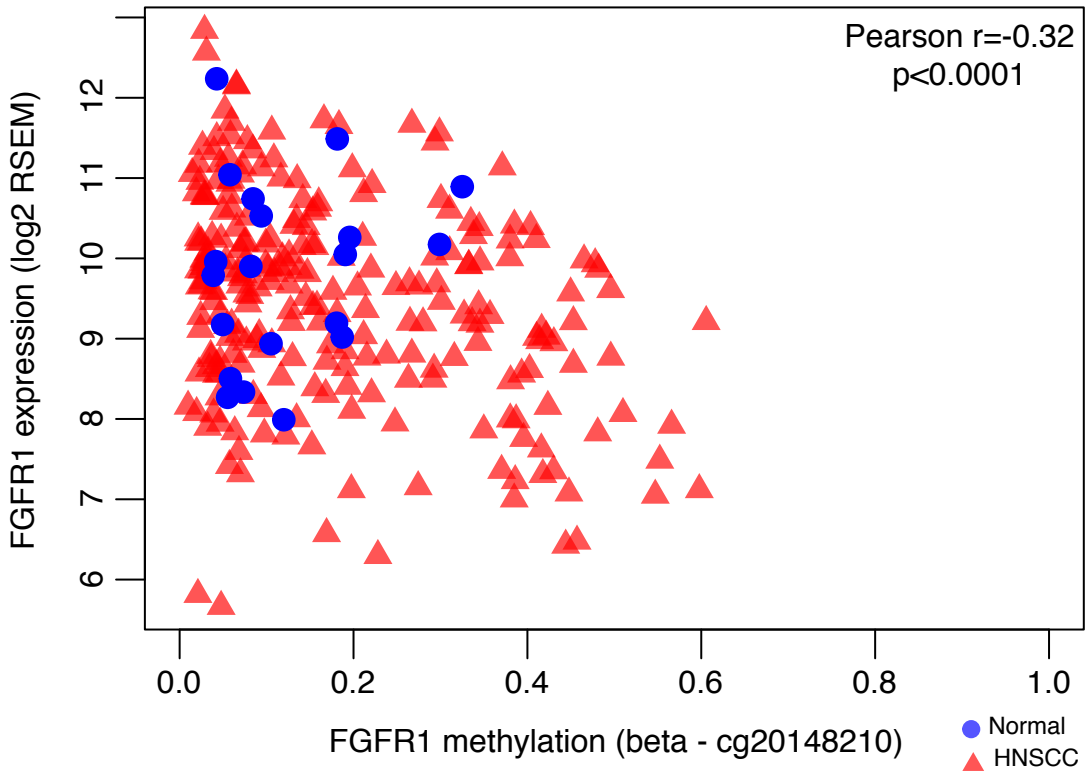

A
FGFR1 gene expression vs DNA methylation in resistant clones
FGFR1 expression (log2 RSEM)
11
9
7
CTXR10
CTXR1
CTXR4
CTXR12
CTXR11
CTXR3
CTXR2
CTXR8
CTXR9
CTXR7
CTXR5
SCC25
0.1
0.2
0.3
0.4
FGFR1 DNA methylation (beta cg20148210)
B
FGFR1 gene expression
Relative expression (ΔΔCt)
40
30
20
10
0
SCC25
CTXR10
C
Protein expression
SCC25
CTXR10
FGFR1
pFGFR (Tyr653/654)
EGFR
pEGFR (Tyr1068)
GAPDH

preprint doi: https://doi.org/10.1101/136564; this version posted August 1, 2017. The copyright holder for this preprint (which was not by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

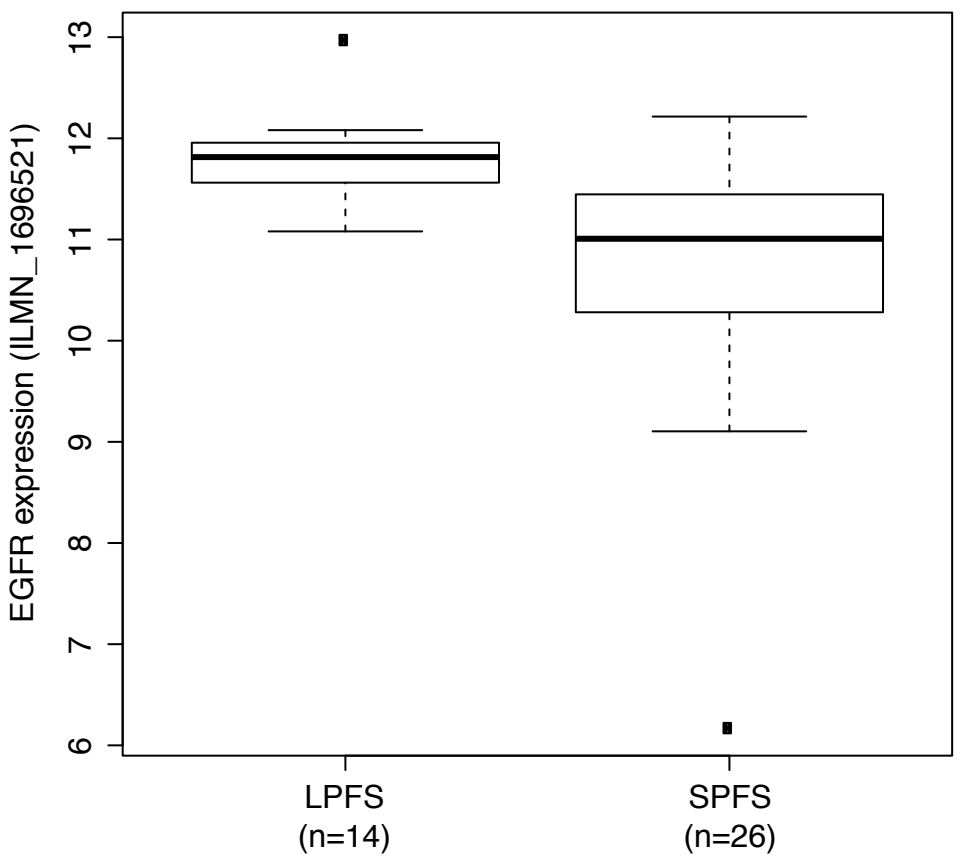
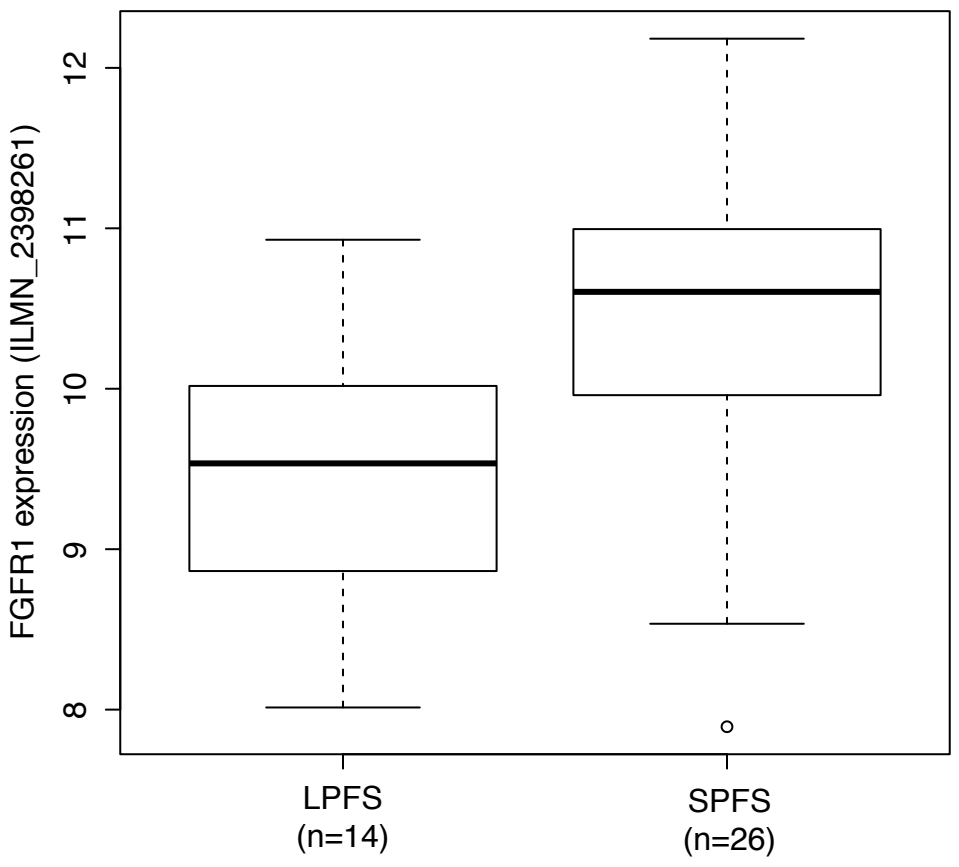
A



B



C



D