Untangling the gene-epigenome networks: Timing of epigenetic regulation of gene expression in acquired cetuximab 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
Michael F Ochs: ochsm@tcnj.edu
Christine H Chung: Christine.Chung@moffitt.org
Elana J Fertig: ejfertig@jhmi.edu

Key words:

ABSTRACT

Targeted cancer therapeutics induce transcription changes in hundreds of genes simultaneously due to the mechanism of action of the treatment, changes in cellular proliferation, and development of resistance. While almost all patients treated with targeted therapeutics develop resistance, the timing and interplay among regulatory pathways responsible for acquired resistance remain unknown. Here we developed a robust combination of experimental and bioinformatics tools to measure genomic changes during the development of resistance. Using the targeted therapeutic cetuximab, an EGFR blocking agent, on an *in vitro* model of head and neck squamous cell carcinoma (HNSCC), we characterized the genetic and epigenetic alterations that occur while cells acquired cetuximab resistance. We confirmed that gene expression signatures from previous gold-standard studies comparing only pre and post resistance samples conflate genes associated with early therapeutic response and genes associated with acquired therapeutic resistance. In contrast, analysis of the time course data with CoGAPS non-negative matrix factorization found substantial gene expression changes uniquely associated with acquired therapeutic resistance. Specifically, analysis of DNA methylation data in our time course identified patterns of gene expression anti-correlated with DNA methylation that changed over time with the acquisition of resistance. Further, this novel experimental and bioinformatics approach identified a robust temporal delay between gene expression changes and DNA methylation, suggestive of an epigenetic mechanism to stabilize the critical alterations for the resistant phenotype. Together this method is generalizable to cross-platform time course analysis of high-throughput genomics data in other dynamic systems where development of resistance leads to recurrent disease.

INTRODUCTION

Targeted therapies inhibit specific molecular pathways responsible for tumor development[1]. Intrinsically responsive patients present prolonged survival, however the treatment is not curative. Most patients will acquire resistance to the therapy and relapse usually within two years of initial treatment[2]. Prevailing evidence suggests that resistance is a result from the recruitment of alternative pathways[3]. Mechanisms leading to acquired resistance vary within the same tumor type and are currently not well characterized. Inhibitors against Epidermal Growth Factor Receptor (EGFR) represent a common class of targeted therapeutics. Cetuximab monoclonal antibody against EGFR is FDA approved for the treatment of metastatic CRC and HNSCC[4]. As with other targeted therapies, stable response is not observed for a long period and virtually all patients invariably develop acquired resistance[5].

The majority of studies to characterize the molecular landscape of acquired resistance are case control studies. These studies compare the genomic profiles of samples prior to treatment and samples after the resistance develops. While these studies can determine genomic profiles that have changed due from prolonged treatment, they cannot pinpoint their precise association with the acquisition of resistance. Characterization of the dynamics of genomic alterations induced during acquired cetuximab resistance could identify targetable oncogenic drivers. However, repeated biopsies at regular time intervals throughout treatment are invasive, expensive, and impractical for patients. Therefore, development of informative model systems of the acquisition of resistance is extremely important.

Recent advances in *in vitro* models of acquired cetuximab resistance[6] 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[6–8]. Therefore, these studies are blind to the evolution of acquired molecular alterations in therapeutic resistance.

Even with advances to experimental sampling of acquired resistance, 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 bioinformatics algorithms that find time-course patterns in genomic data use unsupervised approaches that at best infer known functional forms[9–13] or adjust linear models to correlate molecular profiles with known temporal patterns[14–17]. Other algorithms[18–23] 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. Therefore, 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 has not been established.

In this study, we present a sampling protocol to follow molecular progression using multiple high throughput assays while resistant phenotype develops. We selected gene expression and DNA methylation based upon previous association of DNA methylation with acquired cetuximab resistance *in vitro* and *in vivo*[24]. We demonstrated that the Bayesian non-negative matrix factorization algorithm CoGAPS[10] infers specific patterns of gene expression and DNA methylation that develop according to the gradual establishment of the acquired cetuximab resistance. The DNA methylation pattern inferred with the time course CoGAPS analysis identified *FGFR1* epigenetic expression regulation as the alteration that is correlated with gene expression changes in acquired cetuximab resistance. Such epigenetic regulation of *FGFR1* is observed in HNSCC tumors and changes in *FGFR1* gene expression are associated with acquired cetuximab resistance in HNSCC patients, suggesting integrated analyses are essential to determine the drivers of acquired resistance. Both the experimental and bioinformatics methods developed here are applicable to other molecular platforms, therapeutics, and cancer subtypes.

MATERIAL AND METHODS

Time course establishment of cetuximab resistant SCC25 clones

HNSCC sensitive cell line, SCC25 (purchased from ATCC), was treated with 100nM cetuximab every three days for 11 weeks (generations C1 to C11). On the eighth day, cells were collected: 60,000 cells were replated for another week of treatment with cetuximab; the remaining cells were separately collected for: (1) RNA and DNA isolation, (2) proliferation assay and (3) storage for future use. This cycle was 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 at an initial density selected to reach ~70% confluence at the end of each generation. The replicates were then collected and pooled to provide enough cells for genetic, epigenetic and proliferation assays (**Figure 1A**). To achieve the same final cell confluence and adequate 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 (75cm2) flasks (60,000 cells/flask) that were combined by 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 confluency of the PBS controls at the collection time.

Response to cetuximab was measured by cell proliferation using the Click-iT Plus EdU Flow Cytometry Assay Kit with Alexa Fluor 488 picolyl azide (Life Technologies, Carlsbad, CA) according to manufacturer's instructions. The induced resistance to cetuximab was verified by the ability of a late resistant generation (C10) to form colonies in Matrigel (BD Biosciences, Franklin Lakes, NJ) in comparison to the parental SCC25 (C0), both were treated with different concentrations of cetuximab: 0nM, 10nM, 100nM and 1000nM. The colony formation assay was performed as described previously[11].

SCC25 was authenticated using short tandem repeat (STR) analysis kit PowerPlex16HS (Promega, Madison, WI).

Stable SCC25 resistant single clones (CTXR clones)

Induced cetuximab resistance, single cell isolation from SCC25, gene expression and DNA methylation measurements were performed as previously described[12].

The current analysis included the 11 clones with substantial survival advantage compared to the parental SCC25 as reported in Cheng et al.[12], excluding CTXR6. This design enabled genomics profiling of the resistant clones and the parental SCC25 cell line in a single DNA methylation microarray chip, thereby reducing the interference of technical artifacts between DNA methylation array batches.

Proliferation assay was performed to confirm cetuximab resistance in the single cell clones. SCC25 and resistant cell lines (CTXR4, 7, 10 and 11) were counted with trypan blue staining using a TC20 Automated Cell Counter (Bio-Rad, Hercules, CA). A total of 1000 cells in 100ul final volume were seeded in 96-well plates in quadruplicate. PBS or cetuximab (10nM, 100nM or 1000nM) was added after 24 hours. Fresh media with PBS or cetuximab was added after 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). Proliferation rate was determined by comparing PBS treatment to Cetuximab treatment.

RNA-sequencing sample preparation and data normalization

RNA isolation and sequencing were performed for the parental SCC25 (C0) and each of the cetuximab and PBS generations (C1 to C11) and the cetuximab resistant clones at the Johns Hopkins Medical Institutions (JHMI) Deep Sequencing & Microarray Core Facility. RNA-sequencing was also performed for two additional technical replicates of parental SCC25 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 RNAseq. 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[13] and gene expression counts were quantified with RSEM[14]. Gene counts were upper-quartile normalized and log transformed for analysis following the RSEM v2 pipeline used to normalize TCGA RNA-seq data[15]. All RNA-seq data from this study is available from GEO (ID pending) as part of SuperSeries.

DNA methylation hybridization array and normalization

Genome wide DNA methylation analysis was performed on the same samples as RNA-sequencing 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 4uL of bisulfite-converted DNA was denatured, neutralized, amplified and fragmented according to the manufacturer's instructions. Finally, 12uL of each sample were hybridized to the array chip followed by primer-extension and staining steps. Chips were image-processed in the Illumina's iScan system. Data from the resulting iDat files were normalized with funnorm implemented in the R/Bioconductor package minfi (version 1.16.1)[16]. 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 of 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[15]. All DNA methylation data from this study is available from GEO (ID pending) as part of SuperSeries.

Hierarchical clustering and CoGAPS analysis

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 b values, independently using the R/Bioconductor package CoGAPS[17] (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. 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

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

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[18] determined that the optimal number of dimensions p for expression was 5. DNA methylation is run in 4 parallel sets using GWCoGAPS[10]. 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[10]. Gene set activity was estimated with the gene set statistic implemented in calcCoGAPSStat of the CoGAPS R/Bioconductor package[17]. Comparisons between DNA methylation and gene expression values in the data or from CoGAPS were computed with Pearson correlation.

We compared the DNA methylation data from the time course to data from stable cetuximab resistant SCC25 clones generated previously[12]. To compare epigenetic profiles, CoGAPS was also run on their combined DNA methylation data. In this case, CoGAPS was run only for only the 1,087 genes that switch methylation status. This analysis was run for 2 to 10 dimensions which w assessed as described above.

Cetuximab resistance signatures and EGFR network

Previously, CoGAPS learned a meta-pathway from gene expression data corresponding to overexpression of the $HRAS^{Val12D}$ in the HaCaT model of HPV- HNSCC premalignancy. The gene expression values from this meta-pathway associated with acquired cetuximab resistance in the HNSCC cell line UMSCC1 in that previous study[19]. In the current study, we applied the PatternMarkers statistics[10] to the previously published CoGAPS analysis of these data to derive a gene set from this metapathway called HACAT_HRAS_CETUXIMAB_RESISTANCE or HACAT_RESISTANCE in the manuscript. In addition, we searched MSigDB[20] (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 (NSCLC) cell lines[21]. Gene sets of transcription factor targets were obtained from experimentally validated targets annotated in the TRANSFAC[22] professional database (version 2014.1).

Sources and analysis of human tumor genomics data

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[15]. 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[23], 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 with enhanced proliferation and growth advantage

In order to evaluate the time course progression of genetic and epigenetic changes during the development of acquired cetuximab resistance, we propose a new experimental approach to collect transcriptional, epigenetic, and proliferation data as cells acquire resistance to cetuximab (**Figure 1A**). We apply this approach to the *de novo* cetuximab sensitive HNSCC cell line SCC25 and refer to each point of profiling as a generation (C1 to C11). Cell proliferation analysis (EdU incorporation and flow cytometry) reveals marked differences in proliferative rates across generations in the controls vs. the cetuximab treated cells (**Figure 1B**). Proliferation of the SCC25 controls (PBS generations) is stable through the generations (C1 to C11). Conversely, proliferation of the cetuximab generations progressively increases over each generation. Relative to controls, the growth dynamics of the treated cells is initially inhibited (C1) until generation C3. Starting at C4, the cells become stably resistant to the anti-proliferative effects of cetuximab and gain stable growth advantages absent in the controls. The absence of changes in the proliferation frequency of the PBS generations is an indication that proliferation advantages arise from chronic cetuximab treatment rather than a long period of cell culturing. Anchorage-independent growth of C0 (parental) and C10 cetuximab treated cells further confirms that generation C10 presents significantly ($p<0.05$) higher ability to grow in a semi-solid medium than C0, demonstrating that prolonged cetuximab treatment *in vitro* results in resistance to EGFR blockade enhancing anchorage-independent growth.

Treatment vs. control dominates gene expression clusters

To determine the functional genomics changes occurring as cells acquire cetuximab resistance we performed RNA-seq analysis of SCC25 cells from each generation. Hierarchical clustering of these RNA-seq data for genes in known resistance signatures for EGFR inhibitors (cetuximab and gefitinib)[19,21] in different cell models (HNSCC and NSCLC) successfully separates treatment and control samples **(Figures 2A** and **2B**, shown for all genes in **Supplemental Figure 1)**. Nevertheless, none of these signatures are of sufficient resolution to distinguish between treatment effects as the resistance develops at cetuximab generation C4. This inability of known signatures to distinguish the timing of acquired resistance in generations of cells is not surprising given that all of these signatures are defined in single time point case-control paradigms. Recent work to increase accuracy in drug-response metrics illustrate the confounding effects of variability in cell growth/division rates and/or delayed treatment effects[24,25]. Therefore, robust time-course bioinformatics algorithms to accurately determine the timing of the molecular changes from these assays can remediate these known limitations.

CoGAPS analysis of gene expression defines patterns of acquired resistance

To define gene expression signatures for treatment effect and cetuximab resistance, respectively, we applied CoGAPS algorithm to the time course gene expression data. CoGAPS simultaneously infers gene expression signatures and the relative magnitude of these signatures in each sample. These sample magnitudes separate distinct experimental conditions and each is referred to as a pattern. This dataset has a total of five patterns: three patterns that distinguish the experimental conditions (cetuximab vs. PBS) (**Figure 2C** and **Supplemental Figure 2**), one pattern that represents changes in gene expression from the parental cell lines and subsequent generations, and one pattern that is constant and corresponds to signature of highly expressed genes (**Supplemental Figure 2**).

Similar to the separation seen with clustering (**Supplemental Figure 1**), the first CoGAPS pattern (pattern 1) distinguishes cetuximab from PBS at every generation (**Figure 2C**). Pattern 1 gene expression signature illustrates an immediate transcriptional induction in response to cetuximab treatment. CoGAPS gene set statistics confirms that known gene sets associated with resistance to EGFR inhibitors[19,21] are significantly enriched in this pattern (**Figure 2D**; one-sided p-values of 0.002 and 0.003 for COLDREN_GEFITINIB_RESISTANCE_DN and HACAT_HRAS_CETUXIMAB_RESISTANCE, respectively). However, the transcriptional changes in this pattern are not associated with acquired resistance to cetuximab, and even decrease modestly as resistance develops. Further, enrichment of the transcription factor AP-2alpha targets (*TFAP2A*; one-sided p-value of 0.05) confirms previous work indicating that transcription by AP-2alpha is induced as an early feedback response to EGFR inhibition[26]. These enrichment statistics are consistent with these sets being defined in case and control experimental designs, rather than a time course of acquired resistance.

The second CoGAPS pattern (pattern 2) shows the cetuximab treated cells diverging from controls at generation C4 (**Figure 2C**) — the time point that cetuximab treated cells present significant and stable growth advantage over PBS controls (**Figure 1B**). Therefore, pattern 2 obtains gene expression signatures associated consistently with the development of cetuximab resistance. CoGAPS gene set statistics show that the vast majority of transcription factors (TFs) downstream of EGFR remain down-regulated during acquired resistance (**Figure 2D**). One striking exception is c-Myc, which trends with acquired resistance (p-value of 0.06). This association may reflect the growth advantage of the resistant cells. Only the COLDREN_GEFITINIB_RESISTANCE_DN gene signature is significantly down-regulated in pattern 2 (p-value of 0.04).

CoGAPS expression pattern 3 represents a gradual repression of gene expression with cetuximab treatment (**Figure 2C**). This pattern is trending to significant enrichment in the COLDREN_GEFITINIB_RESISTANCE_DN signature (one-sided p-value 0.12) and down-regulation in HACAT_HRAS_CETUXIMAB_RESISTANCE (one-sided 0.09). Both the dynamics in the pattern and gene set enrichment confirms that pattern 3 is associated with repression of gene expression during acquired cetuximab resistance. Significant enrichment of the acquired resistance signature in CoGAPS patterns 1-3 suggests 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. Thus, the gene expression signature in CoGAPS patterns from the time course are specific to the transcriptional changes associated with acquired therapeutic resistance.

Changes in DNA methylation occur concomitantly with gene expression changes associated with resistance to cetuximab, but not gene expression changes that occur as an immediate response to treatment

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 (**Figure 3A**). Application of the CoGAPS matrix factorization algorithm to the methylation data reveals a total of 3 patterns (**Figures 3B and C**): (1) gradual increase of DNA methylation in controls, (2) rapid demethylation in cetuximab treated generations starting at C4, and (3) rapid increase in DNA methylation in cetuximab treated generations starting at C4. In contrast to the gene expression data, there is no immediate shift in DNA methylation resulting from cetuximab treatment. All together, CoGAPS methylation patterns demonstrate that demethylation of genes occurs gradually across the generations, evident at low levels in early generations and is suggestive of clonal outgrowth **(Figure 3B and C, Pattern 2)**.

Comparing the CoGAPS patterns from gene expression and DNA methylation reveals strong anti-correlation between gene expression and DNA methylation in resistant patterns **(Figures 4A** and **4B)**. The temporal resolution of this relationship is remarkably precise and recapitulates the phenotypic changes captured in the growth curves **(Figures 1B and 4C)**. In spite of this correlation, we observe that the gene expression changes associated with acquired resistance occur more gradually over all generations of cetuximab resistance. In contrast, the DNA methylation is consistent with cetuximab treatment and control PBS in patterns 2 and 3 during early generations. Additionally, rapid accumulation in DNA methylation changes starting after generations C4 and C5 (**Figures 3 and 4, Patterns 2 and 3**), concurrent with the start of the observed growth advantage over the PBS control. These dynamics suggests that DNA methylation changes have an important role in stabilizing the gene expression signatures crucial to acquired cetuximab resistance.

The gene signatures from the anti-correlated DNA methylation and gene expression CoGAPS patterns have low correlation (**Supplemental Figure 3**). We hypothesize that the timing differences between DNA methylation and gene expression render the CoGAPS gene signatures from each data modality insufficient to indicate regulation of expression by DNA methylation. To ascertain potential drivers of the stable cetuximab resistant phenotype induced by DNA methylation, we defined genes that are PatternMarkers[10] of the DNA methylation patterns associated with stable acquired cetuximab resistance (methylation patterns 2 and 3). We then correlated the gene expression profiles of each of these PatternMarkers genes to the DNA methylation values (**Figure 4D**).

Increased *FGFR1* expression is associated with cetuximab related down-regulation of *EGFR*

To evaluate heterogeneity within the pooled cell lines in the time course experiment, we measured DNA methylation and gene expression on a panel of eleven isogenic stable cetuximab resistant clones derived from SCC25 previously[12]. 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 SCC25, the single cell clones and time course generations display widespread differences. Significantly greater heterogeneity among the cetuximab resistant single cell clones in both expression and methylation profiles (**Supplemental Figure 4 and 5**, respectively) and cellular morphology (**Supplemental Figure 9**). **Figure 5A and 5B** demonstrate that higher heterogeneity among single cell clones is also observed in the epigenetically regulated PatternMarker genes from the CoGAPS analysis that are shown in Figure 4D. Further CoGAPS analysis combining DNA methylation data from the time course with DNA methylation data from the single cell clones also reflect the increased heterogeneity of the stable cetuximab resistant clones relative to the time course (**Supplemental Figure 7**). These results suggest that different mechanisms of resistance may arise in the same HNSCC cell line. Therefore, we hypothesize 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.

Nine of the epigenetically regulated pattern marker genes associated with resistance from Figure 4D also have significantly anti-correlated gene expression and DNA methylation in the stable cetuximab resistant clones (**Supplemental Figure 8**). Of these, only *FGFR1* was demethylated and reexpressed in a cetuximab resistant clone relative to the parental SCC25 cell line **(Figure 4A-C)**. This finding is 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*[27–29]. In this analysis, epigenetic regulation of gene expression for *FGFR1* occurs in only one of the resistant clones (CTXR10). This clone is among the fastest growing under cetuximab treatment (**Supplemental Figure 6**), suggesting that the pooled data from the time course captures clonal outgrowth of a cetuximab resistant clone with similar molecular features (*FGFR1* demethylation) to CTXR10.

***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 investigate 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[15]**,** we verified that the up-regulation of *EGFR* and *FGFR1* is not concomitant **(**Pearson correlation coefficient = -0.0633, p value = 0.3258, **Figure 6A)**. We found that *FGFR1* gene expression and DNA methylation status are negatively correlated (Pearson correlation r=-0.3219, p value<0.0001) **(Figure 6B)**, in TCGA samples, suggesting that *FGFR1* transcription is epigenetically regulated *in vivo* in HPV-negative HNSCC tumors.

Bossi et al.[23] collected gene expression data from HNSCC patients with recurrent metastasis with either short- (SPFS, median 3 months surival) or long-progression-free survival (LPFS, median 19 months survival) to cetuximab. Using this dataset, we verified that *EGFR* expression in SPFS is significantly lower than the LPSF group (**Figure 6C**) (log fold change -1.0, t-test p-value 0.0003). The opposite is observed for *FGFR1*, with overexpression in SPFS vs. LPSF (**Figure 6D**), log fold change 0.9, t-test p-value 0.003**)**. However, this study lacks DNA methylation data to assess whether *FGFR1* is epigenetically regulated in these samples. Nonetheless, this finding in combination with the data from TCGA support our findings that in resistance to cetuximab, the non-responder phenotype is accompanied by loss of *EGFR* expression and *FGFR1* gain as a result of promoter demethylation.

DISCUSSION

Here we present a novel time course experimental and bioinformatics approach for the study of 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. Numerous time course genomics studies of short-term therapeutic response have been performed in the literature[24,30,31]. To our knowledge, this study is the first to collect time course multi-platform genomics data during the acquisition of acquired targeted therapeutic resistance.

In addition to acquiring time course genomics data, establishing the molecular changes associated with acquired cetuximab resistant requires robust time-course bioinformatics analysis that can account for multiple experimental conditions. Based upon previous performance in inferring dynamic regulatory networks for targeted therapeutics[31], we selected a Bayesian non-negative Matrix Factorization algorithm called CoGAPS[17] 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. Unexpectedly, gene expression signatures for resistance to EGFR inhibitors from previous studies were significantly enriched in both types of CoGAPS patterns. These previous resistance signatures were learned from case-control studies that compare gene expression for sensitive cells to that of the resistant cells, unable to distinguish the timing of therapeutic response. Therefore, we concluded that including the time course data in this study was essential to determine gene signatures that are unique to the resistant phenotype.

Combining gene expression and DNA methylation data from the time course 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 is important for the creation of acquired resistance. The CoGAPS patterns in gene expression that are associated with acquired cetuximab resistance gradually change over the time course. On the other hand, the CoGAPS patterns for DNA methylation changes have a sharp transition at the generation at which resistance is acquired. These patterns reflect a later, 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 examples of tumor suppressor genes that transcription silencing was found to occur prior to DNA hypermethylation and chromatin changes, suggesting that epigenetic changes are necessary to stabilize gene expression aberrant profile by locking a silenced chromatin state that will result in tumor progression[32,33]. Although the lack of investigation into chromatin modifications, our integrated RNA-Seq and DNA methylation analysis corroborate the fact that gene expression 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. Further investigation into the chromatin remodeling mechanisms might elucidate the hypothesis that they follow the changes in expression and occur in combination with altered methylation patterns.

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[34]. These 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 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 resistance phenotype.

Our time course approach allowed us to follow the progression of DNA methylation changes at the different points of cetuximab treatment. We found that for a significant proportion of genes, promoter methylation and mRNA levels are negatively correlated. Among these genes, *FGFR1* presented with loss of CpG methylation accompanied by increase in 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*[35]. Its over-expression has previously been associated with EGFR inhibitors resistance[27–29]. 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 and the time course analysis enables us to see the clonal outgrowth of this one 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[35]. DNA methylation of *FGFR1* must also be considered when evaluating its utility as a biomarker in HNSCC in future studies.

We recognize that a limitation of the current study is the use of only one cell line model to induce resistance and collect the time course data for gene expression and epigenetics analysis. However, we had to take into consideration the broad cross-platform profiling to identify the patterns associated with acquired resistance. Since here we do not perform only initial and end time point analysis, multiple data points in the analysis had to be accounted for when determining the number of cell models to be included. Even using only one in vitro model, we demonstrated that our approach and findings could be generalized to HNSCC patients sample since TCGA[15] and another study[23] data validated our main 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 observe 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[12]. 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 reflects 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 technical challenges that must first be addressed before such time course sampling is possible[6]. Pinpointing precise molecular predictors of therapeutic resistance will facilitate unprecedented biomarkers and reveal the mechanisms by which to overcome acquired therapeutic resistance to all therapies used to treat cancer.

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; D. Sidransky for critical comments during the preparation of the manuscript. This work was supported by NIH Grants R01CA177669, R21DE025398 and P30 CA006973 and SPORE P50DE019032.

Author contributions

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.

References

1. Sawyers, C. Targeted cancer therapy. *Nature* **432,** 294–297 (2004).
2. Hyman, D. M., Taylor, B. S. & Baselga, J. Implementing Genome-Driven Oncology. *Cell* **168,** 584–599 (2017).
3. Engelman, J. A. *et al.* MET Amplification Leads to Gefitinib Resistance in Lung Cancer by Activating ERBB3 Signaling. *Science* **316,** 1039–1043 (2007).
4. Vincenzi, B., Zoccoli, A., Pantano, F., Venditti, O. & Galluzzo, S. Cetuximab: from bench to bedside. *Curr. Cancer Drug Targets* **10,** 80–95 (2010).
5. Boeckx, C. *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* **7,** 337 (2014).
6. Quesnelle, K. M., Wheeler, S. E., Ratay, M. K. & Grandis, J. R. Preclinical modeling of EGFR inhibitor resistance in head and neck cancer. *Cancer Biol. Ther.* **13,** 935–945 (2012).
7. Wheeler, D. L. *et al.* Mechanisms of acquired resistance to cetuximab: role of HER (ErbB) family members. *Oncogene* **27,** 3944–3956 (2008).
8. Narayan, M. *et al.* Trastuzumab-Induced HER Reprogramming in 'Resistant' Breast Carcinoma Cells. *Cancer Res.* **69,** 2191–2194 (2009).
9. Ogawa, T. *et al.* Methylation of death-associated protein kinase is associated with cetuximab and erlotinib resistance. *Cell Cycle Georget. Tex* **11,** 1656–1663 (2012).
10. Stein-O'Brien, G. *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* 83717 (2016). doi:10.1101/083717
11. Hatakeyama, H. *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* **5,** e12702 (2010).
12. Cheng, H. *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.* **16,** 1252–1258 (2015).
13. Wang, K. *et al.* MapSplice: Accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res.* **38,** e178–e178 (2010).
14. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12,** 323 (2011).
15. Lawrence, M. S. *et al.* Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **517,** 576–582 (2015).
16. Fortin, J.-P. *et al.* Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol.* **15,** (2014).
17. Fertig, E. J., Ding, J., Favorov, A. V., Parmigiani, G. & Ochs, M. F. CoGAPS: an R/C++ package to identify patterns and biological process activity in transcriptomic data. (2010).
18. Bidaut, G. in *link.springer.com* 315–333 (Springer US, 2010). doi:10.1007/978-1-4419-5714-6_19
19. Fertig, E. J. *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* **13,** 160 (2012).
20. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 15545–15550 (2005).
21. Coldren, C. D. Baseline Gene Expression Predicts Sensitivity to Gefitinib in Non-Small Cell Lung Cancer Cell Lines. *Mol. Cancer Res.* **4,** 521–528 (2006).
22. Matys, V. *et al.* TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.* **31,** 374–378 (2003).
23. Bossi, P. *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.* **22,** 3961–3970 (2016).

24. Hafner, M., Niepel, M., Chung, M. & Sorger, P. K. Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. *Nat. Methods* **13,** 521–527 (2016).

25. Harris, L. A. *et al.* An unbiased metric of antiproliferative drug effect in vitro. *Nat. Methods* **13,** 497–500 (2016).

26. Fertig, E. J. *et al.* CoGAPS matrix factorization algorithm identifies transcriptional changes in AP-2alpha target genes in feedback from therapeutic inhibition of the EGFR network. *Oncotarget* **7,** 73845–73864 (2016).

27. Azuma, K. *et al.* FGFR1 activation is an escape mechanism in human lung cancer cells resistant to afatinib, a pan-EGFR family kinase inhibitor. *Oncotarget* **5,** 5908–5919 (2014).

28. Bertotti, A. *et al.* The genomic landscape of response to EGFR blockade in colorectal cancer. *Nature* **526,** 263–267 (2015).

29. Koole, K. *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.* **22,** 3884–3893 (2016).

30. Michna, A. *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.* **11,** (2016).

31. Ochs, M. F. *et al.* Detection of Treatment-Induced Changes in Signaling Pathways in Gastrointestinal Stromal Tumors Using Transcriptomic Data. *Cancer Res.* **69,** 9125–9132 (2009).

32. Bachman, K. E. *et al.* Histone modifications and silencing prior to DNA methylation of a tumor suppressor gene. *Cancer Cell* **3,** 89–95 (2003).

33. Stirzaker, C., Song, J. Z., Davidson, B. & Clark, S. J. Transcriptional gene silencing promotes DNA hypermethylation through a sequential change in chromatin modifications in cancer cells. *Cancer Res.* **64,** 3871–3877 (2004).

34. Tyekucheva, S., Marchionni, L., Karchin, R. & Parmigiani, G. Integrating diverse genomic data using gene sets. *Genome Biol.* **12,** R105 (2011).

35. Babina, I. S. & Turner, N. C. Advances and challenges in targeting FGFR signalling in cancer. *Nat. Rev. Cancer* (2017). doi:10.1038/nrc.2017.8

FIGURE 1

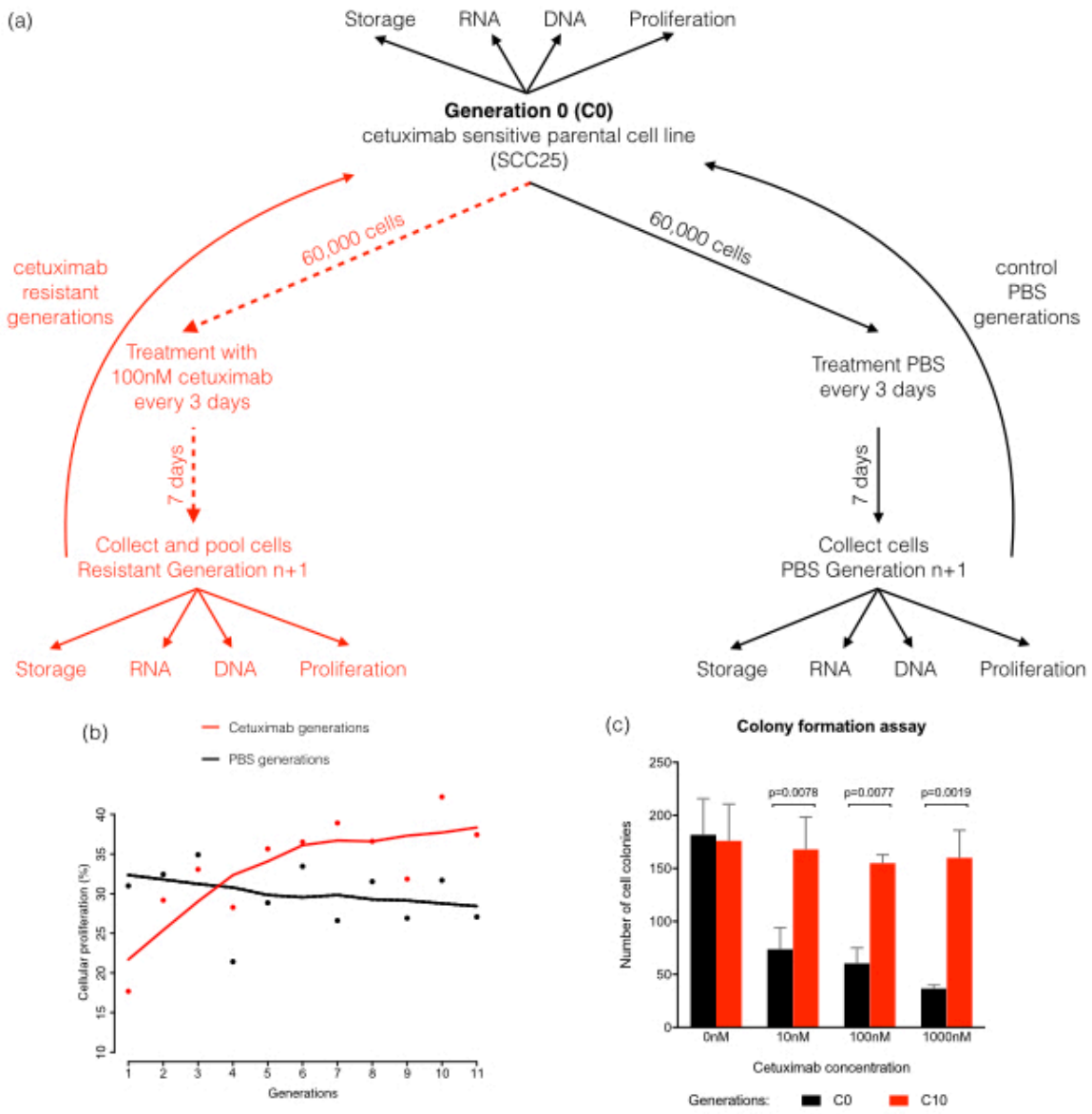


Figure 1 - Time course approach to induce resistance to cetuximab and measure gene expression and DNA methylation changes. (a) Intrinsic cetuximab sensitive HNSCC cell line SCC25 were treated with cetuximab (red) or PBS (black) for 7 days. In the eighth day, cells were collected and pooled from multiple 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 C1 to C10). (b) Proliferation assay from cetuximab treatment (red line) and PBS treated cells (black line) for all SCC25 generations. (c) Colony formation assay in matrigel for anchorage-independent growth confirms acquired cetuximab resistance of C10 (red) relative to the parental cell line (C0, black) at different concentrations of cetuximab (0nM, 10nM, 100nM and 1000nM).

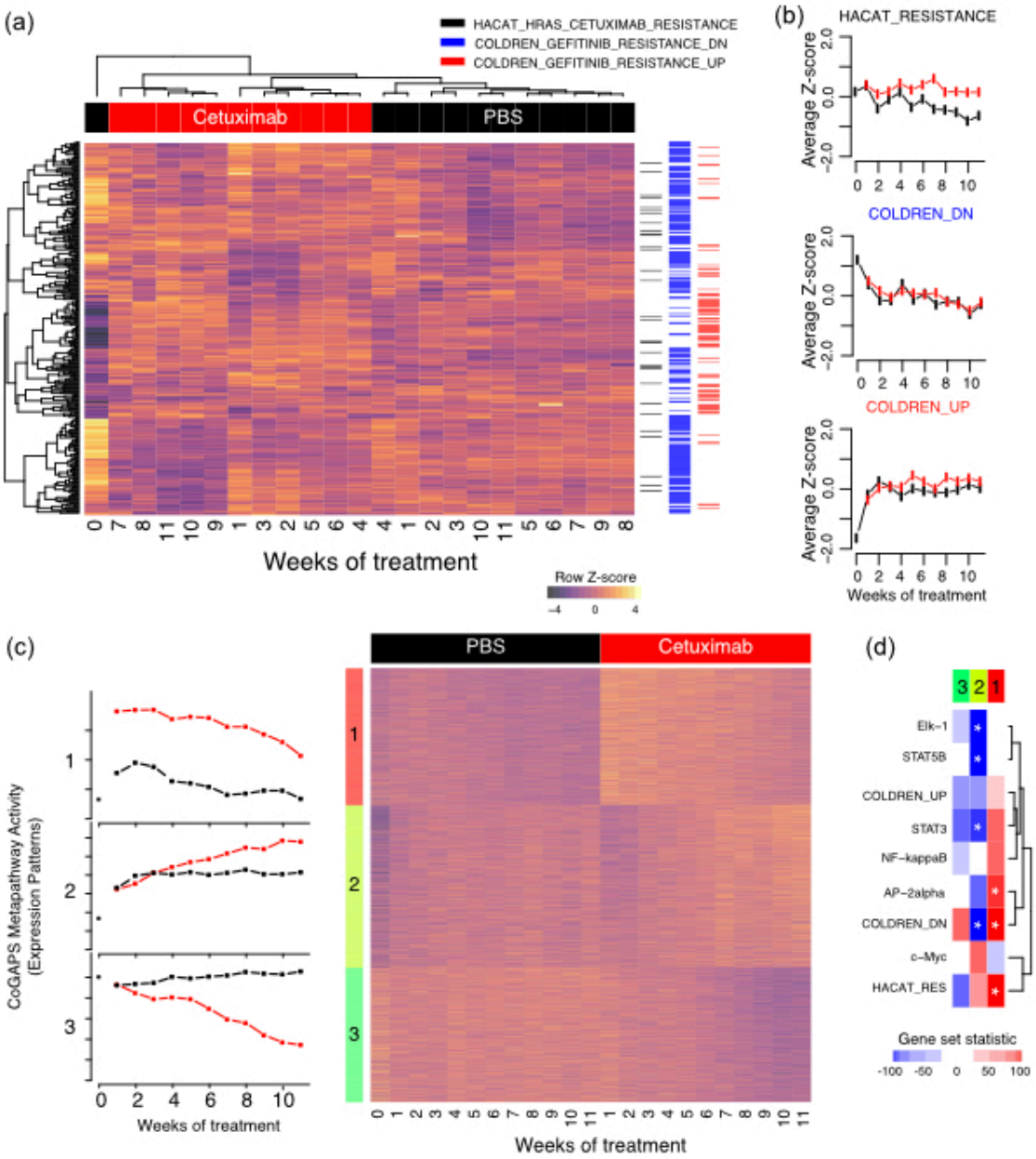


Figure 2 – Gene expression of resistance signatures to EGFR inhibitors separate resistant and control generations, and CoGAPS analysis reflects the dynamics of acquired cetuximab resistance. (a) Heatmap of gene expression values in 11 generations of SCC25 cells treated with 100nM of cetuximab (red columns) to acquire resistance and with PBS as control (black columns). Genes selected for visualization are associated with cetuximab resistance from previous gene expression studies comparing sensitive and resistant cells without regard for timing. These studies provide 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). (c) CoGAPS pattern inferred from gene expression data over generations of PBS control (black lines) or treatment with 100nM of cetuximab (red lines) and 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. (d) 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. A score of 100 indicates upregulation of the targets with a p-value of 0 and -100 downregulation with p-values of 0. Matrix elements with a star indicate p-values below 0.05 for either up or down-regulation of the gene set. Gene expression heatmap is colored on a red-green scale where as the gene set statistics heatmap is colored on a blue-red scale, with values indicated in the respective color keys.

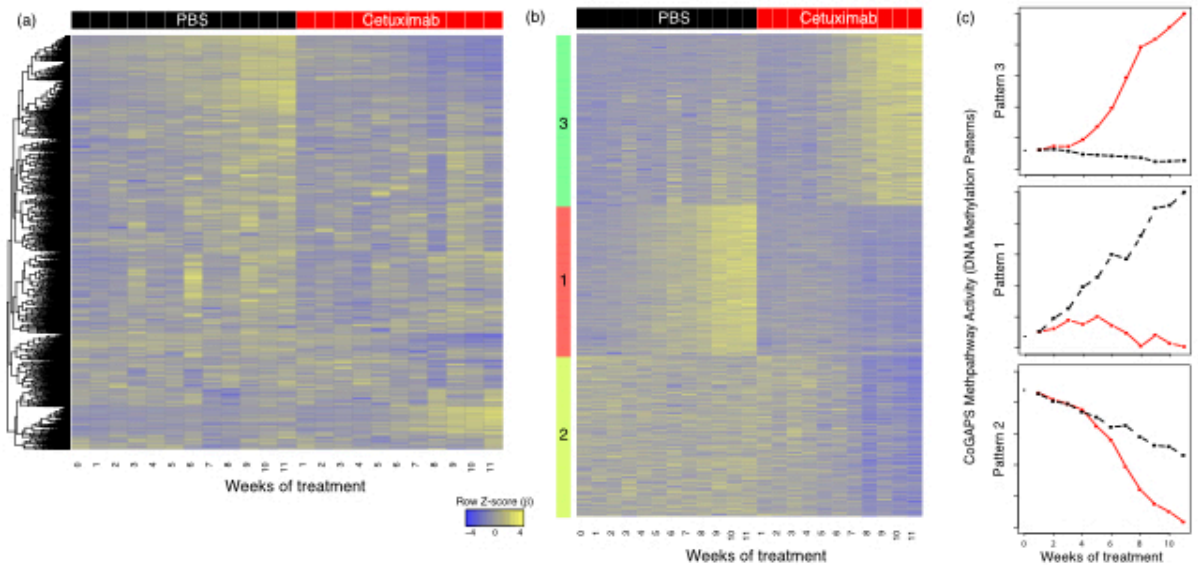


Figure 3 – Dynamics of DNA methylation alterations and association with gene expression CoGAPS patterns in acquired cetuximab resistance. (a) Heatmap of DNA methylation values in 11 generations of SCC25 cells treated with PBS as control (black columns) and with 100nM of cetuximab (red columns) to acquire 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).

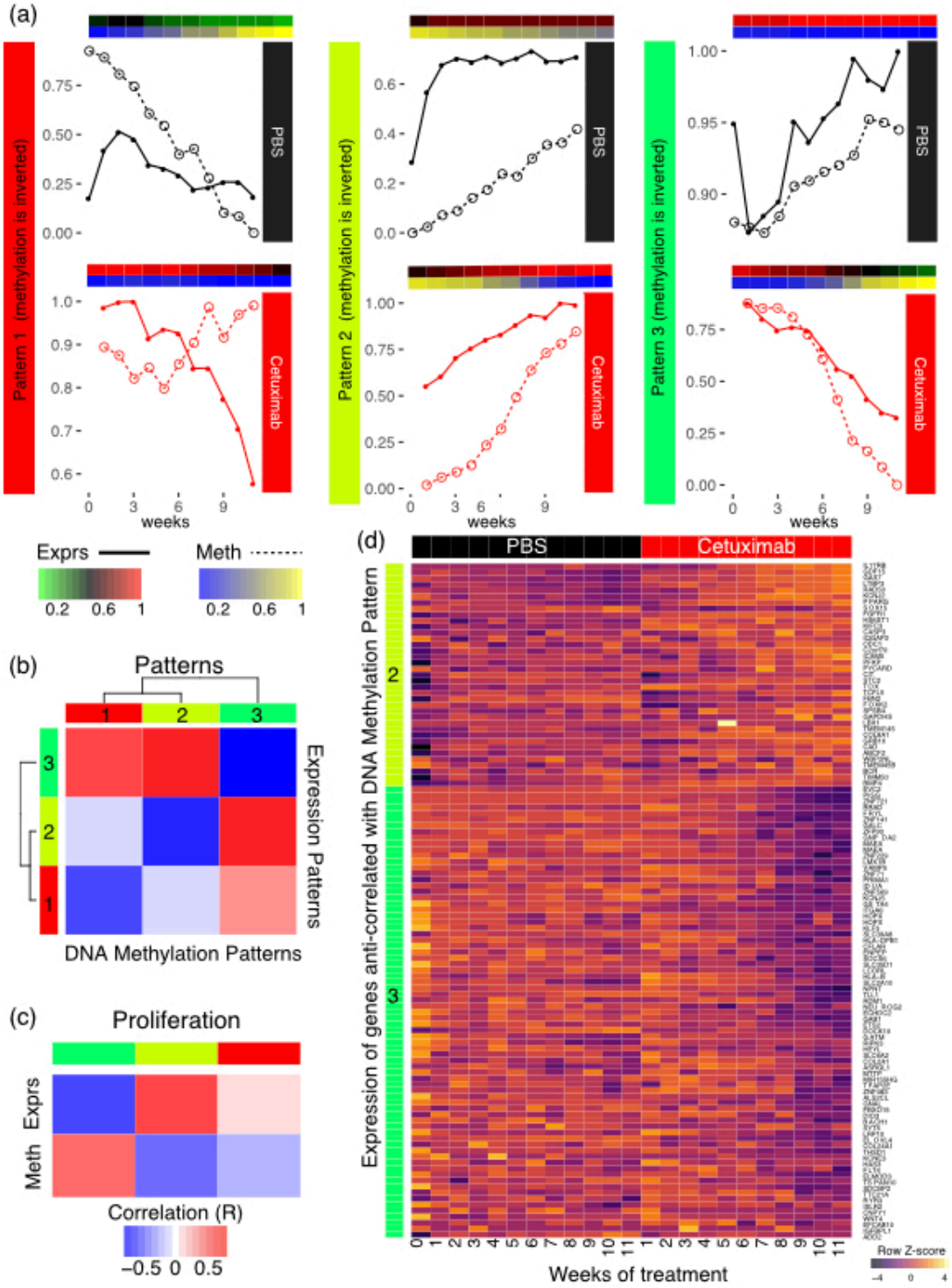


Figure 4 – Dynamics of DNA methylation alterations in CoGAPS are associated with epigenetic regulation of gene expression and the dynamics of acquired cetuximab resistance. (a) CoGAPS patterns for gene expression (red-green heatmap, solid lines) and DNA methylation (blue-yellow heatmap, dashed lines) of most anti-correlated patterns. Y-axis for gene expression increases with increasing values of the CoGAPS pattern for gene expression (labeled on left) and decreases with the CoGAPS patterns for DNA methylation (labeled on right) to reflect anti-correlation. (b) 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 3. The column colors for methylation patterns are selected to match the color of the corresponding expression pattern with maximum anti-correlation. (c) Heatmap of Pearson correlation coefficients for measured proliferation rates and CoGAPS patterns for gene expression (top) and DNA methylation (bottom). Columns are colored according to matched CoGAPS patterns between the data types defined in (b). (d) Heatmap of gene expression data for PatternMarker genes for CoGAPS analysis of DNA methylation (Supplemental Figure 3) with expression significantly anti-correlated with the CoGAPS DNA methylation pattern for acquired cetuximab resistance (pattern 2, yellow; pattern 3, green) to reflect discrepancies in timing between

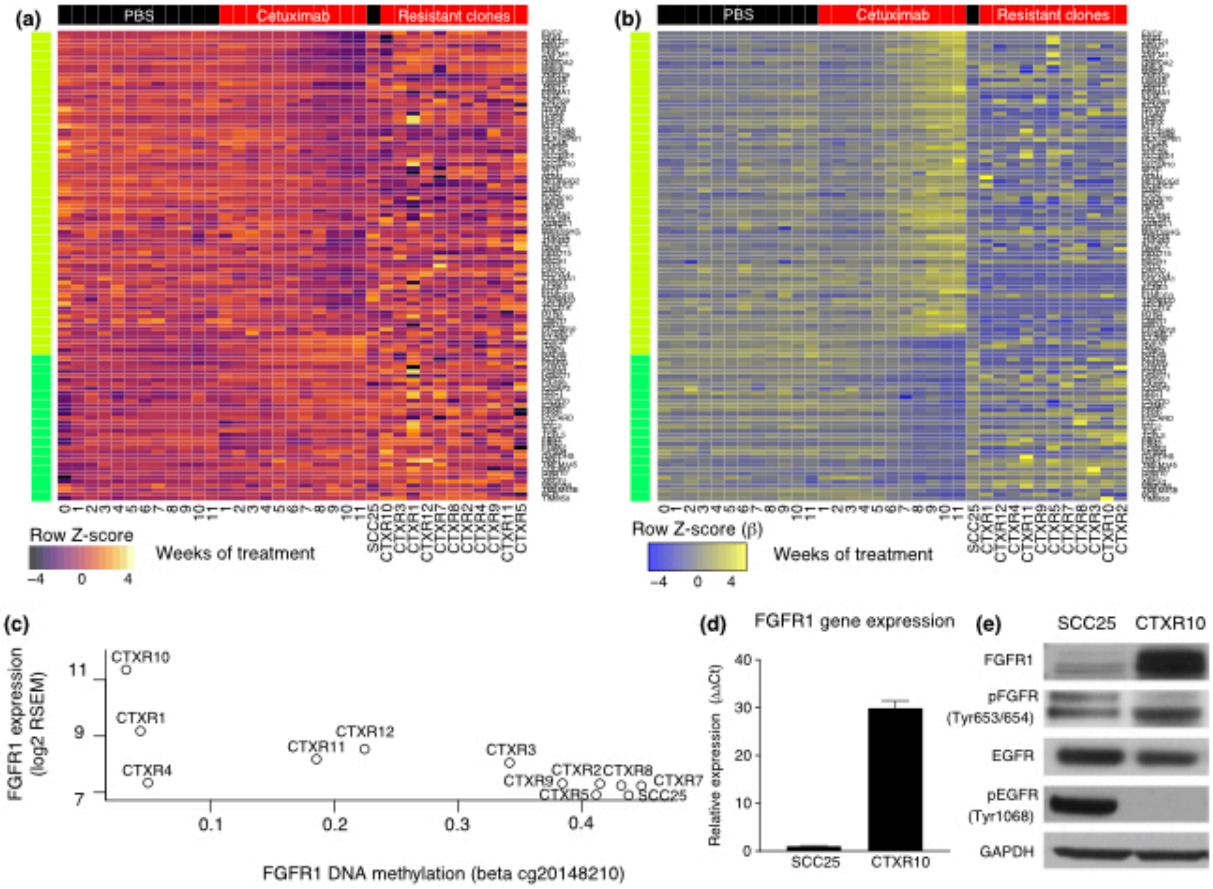
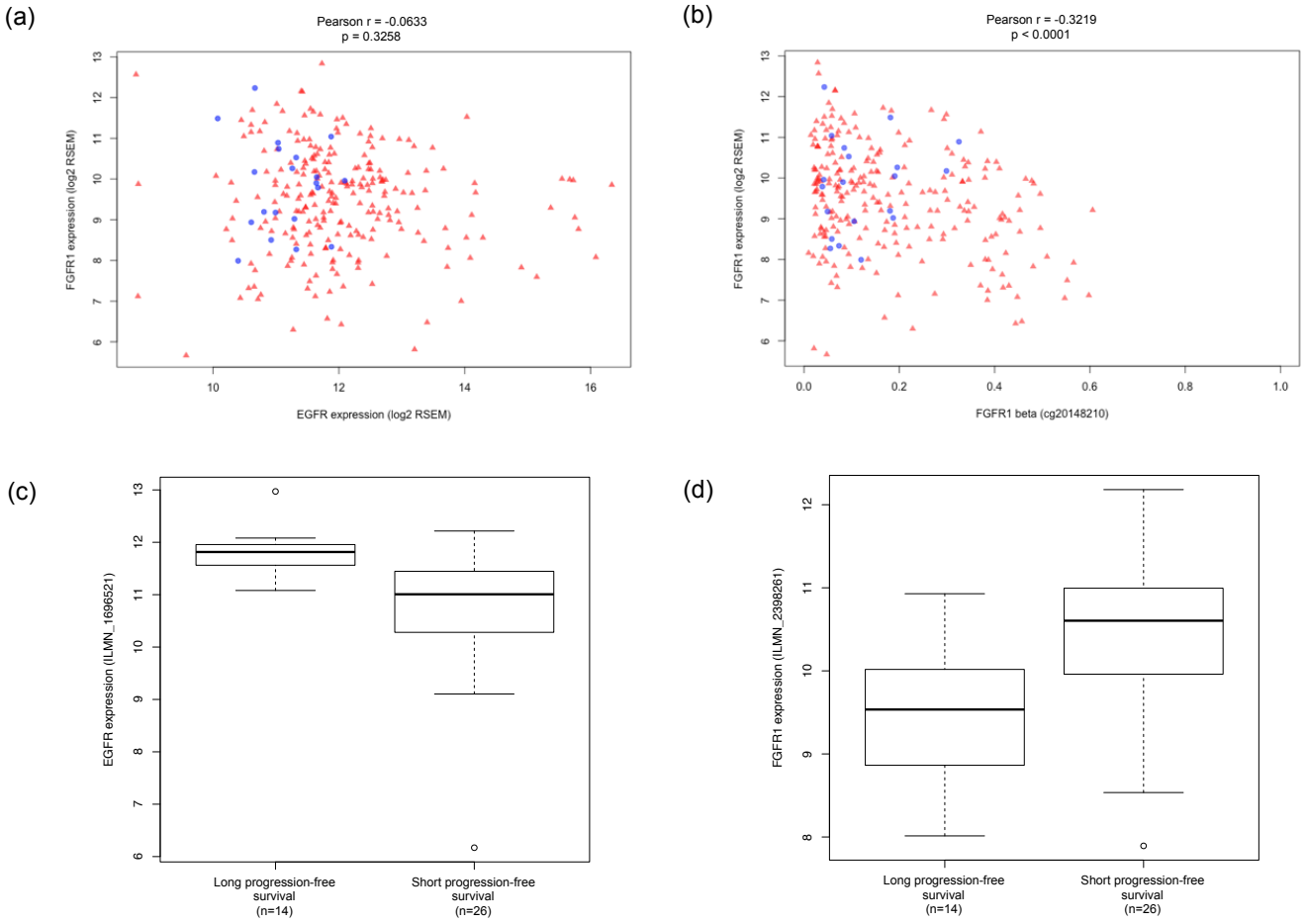
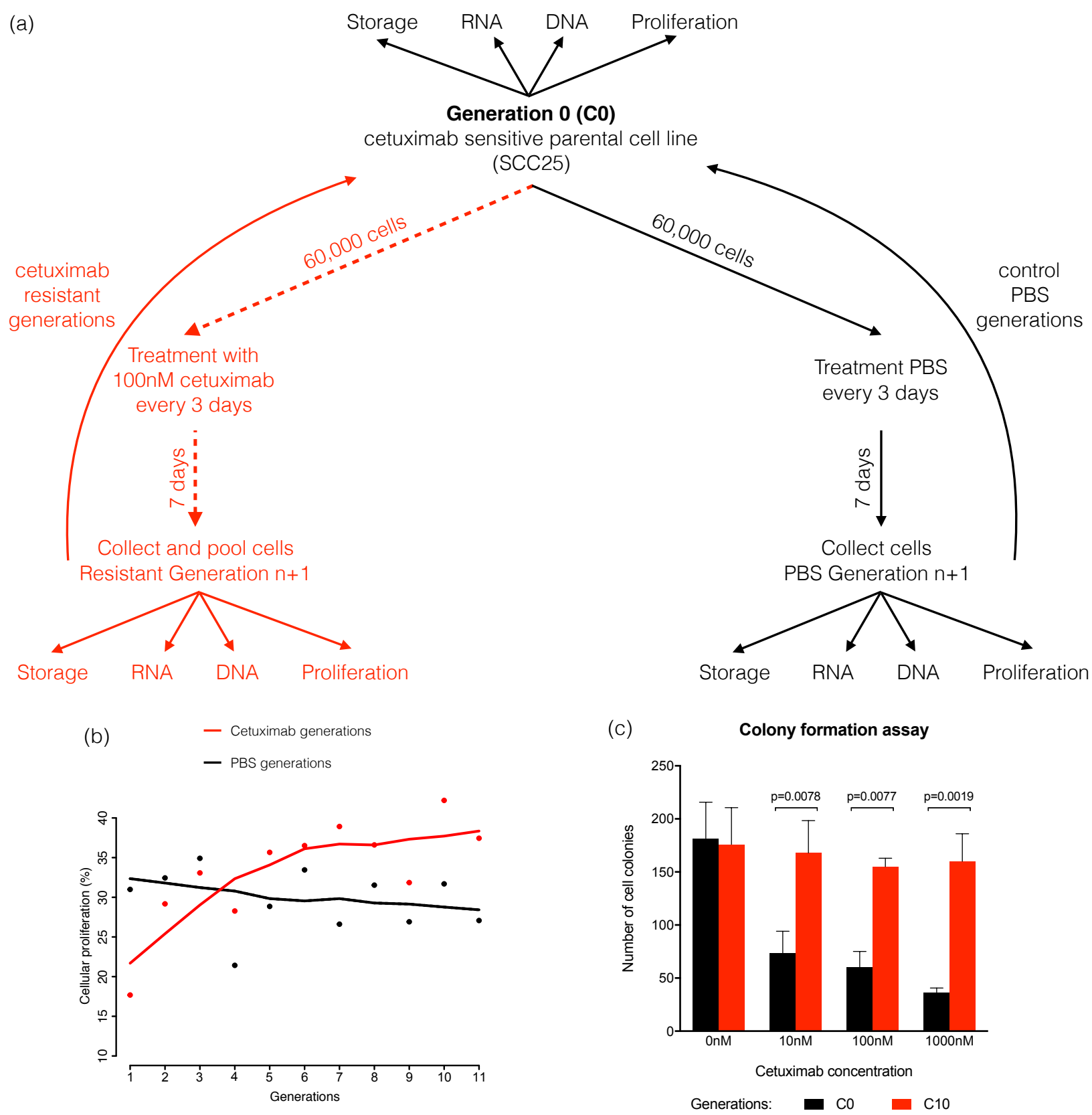


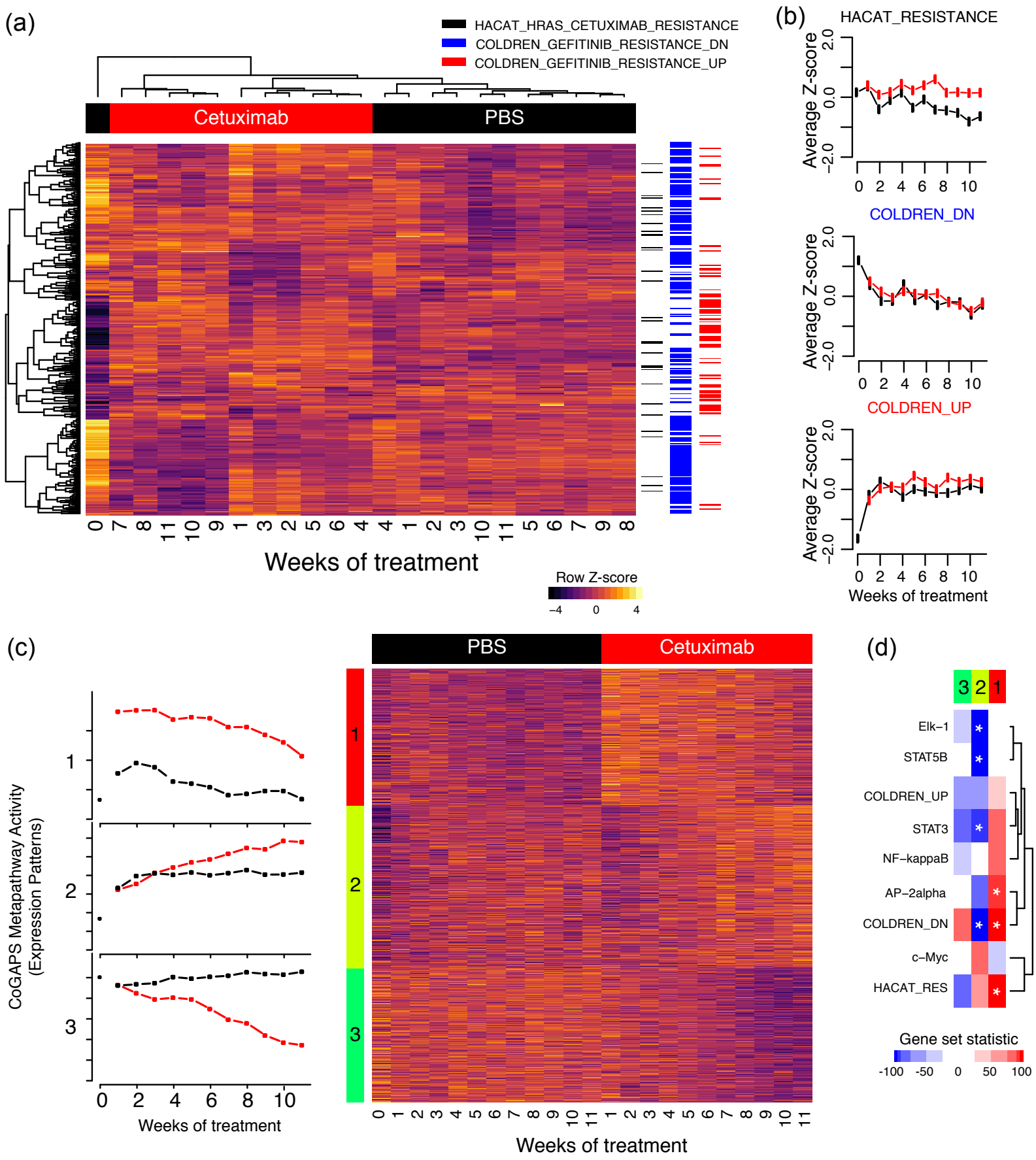
Figure 5 – DNA methylation and gene expression patterns identified by CoGAPS in independent, stable SCC25 cetuximab resistant clones associate epigenetic regulation of *FGFR1* with acquired resistance to cetuximab *in vitro*. (a) Heatmap of gene expression values for PatternMarker DNA methylation genes that are anti-correlated with corresponding CoGAPS patterns for acquired resistance (Figure 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-green 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. (c) Expression of *FGFR1* gene expression relative to DNA methylation in stable cetuximab resistant clones. (d) QRT-PCR of *FGFR1* gene expression in CTXR10 relative to the parental cell line (greater than 30 fold change). (e). 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 is associated with slightly increased levels of phospho-FGFR1 and decrease in EGFR and phospho-EGFR.



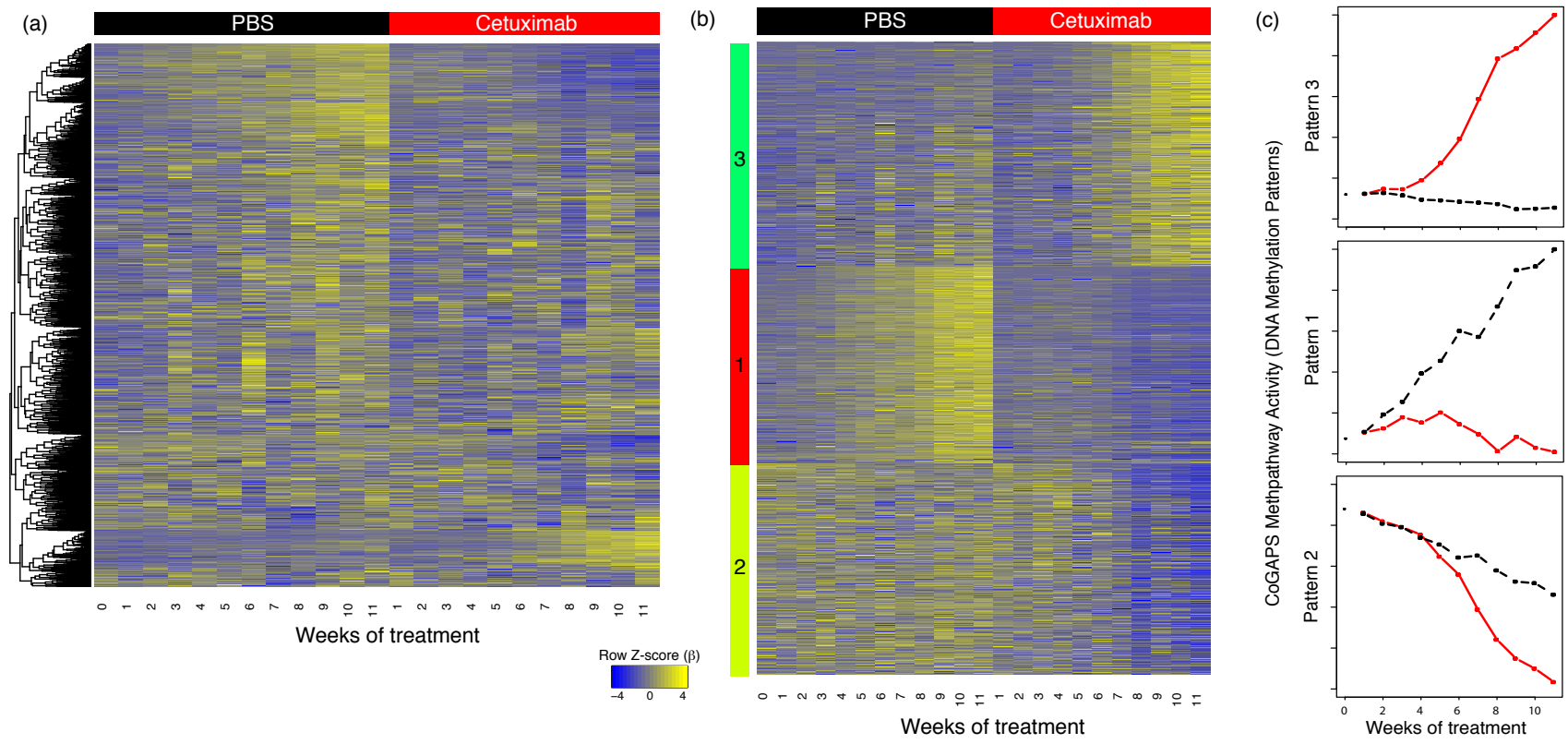
<u>Figure 6</u> - Additional datasets confirms FGFR1 gene and protein overexpression *in vitro*, epigenetic regulation of *FGFR1 in vivo*, and inverse relationship between *EGFR* and *FGFR1* expression in *in vivo* cetuximab resistance. (a) Scatter plot of gene expression for *EGFR* and *FGFR1* in HPV-negative HNSCC samples from TCGA demonstrates 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* is anti-correlated with *FGFR1* expression in HPV-negative HNSCC, suggesting that up-regulation of FGFR1 might be a result of promoter hypomethylation in primary tumors. (c) *EGFR* expression is 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* is significantly overexpressed in patients with short progression free survival relative to patients with long progression free survival in this same dataset.



<u>Figure 1</u> - Time course approach to induce resistance to cetuximab and measure gene expression and DNA methylation changes. (a) Intrinsic cetuximab sensitive HNSCC cell line SCC25 were treated with cetuximab (red) or PBS (black) for 7 days. In the eighth day, cells 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 C1 to C10). (b) Proliferation assay from cetuximab treatment (red line) and PBS treated cells (black line) for all SCC25 generations. (c) Colony formation assay in matrigel for anchorage-independent growth confirms acquired cetuximab resistance of C10 (red) relative to the parental cell line (C0, black) at different concentrations of cetuximab (0nM, 10nM, 100nM and 1000nM).



<u>Figure 2</u> – Gene expression of resistance signatures to EGFR inhibitors separate resistant and control generations, and CoGAPS analysis reflects the dynamics of acquired cetuximab resistance. (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 are associated with cetuximab resistance from previous gene expression studies comparing sensitive and resistant cells without regard for timing. These studies provide 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). (c) CoGAPS pattern inferred from gene expression data over generations of PBS control (black lines) or treatment with 100nM of cetuximab (red lines) and 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. (d) 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. A score of 100 indicates upregulation of the targets with a p-value of 0 and -100 downregulation with p-values of 0. Matrix elements with a star indicate p-values below 0.05 for either up or down-regulation of the gene set. Gene expression heatmap is colored on a red-green scale where as the gene set statistics heatmap is colored on a blue-red scale, with values indicated in the respective color keys.



<u>Figure 3</u> – Dynamics of DNA methylation alterations and association with gene expression CoGAPS 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).

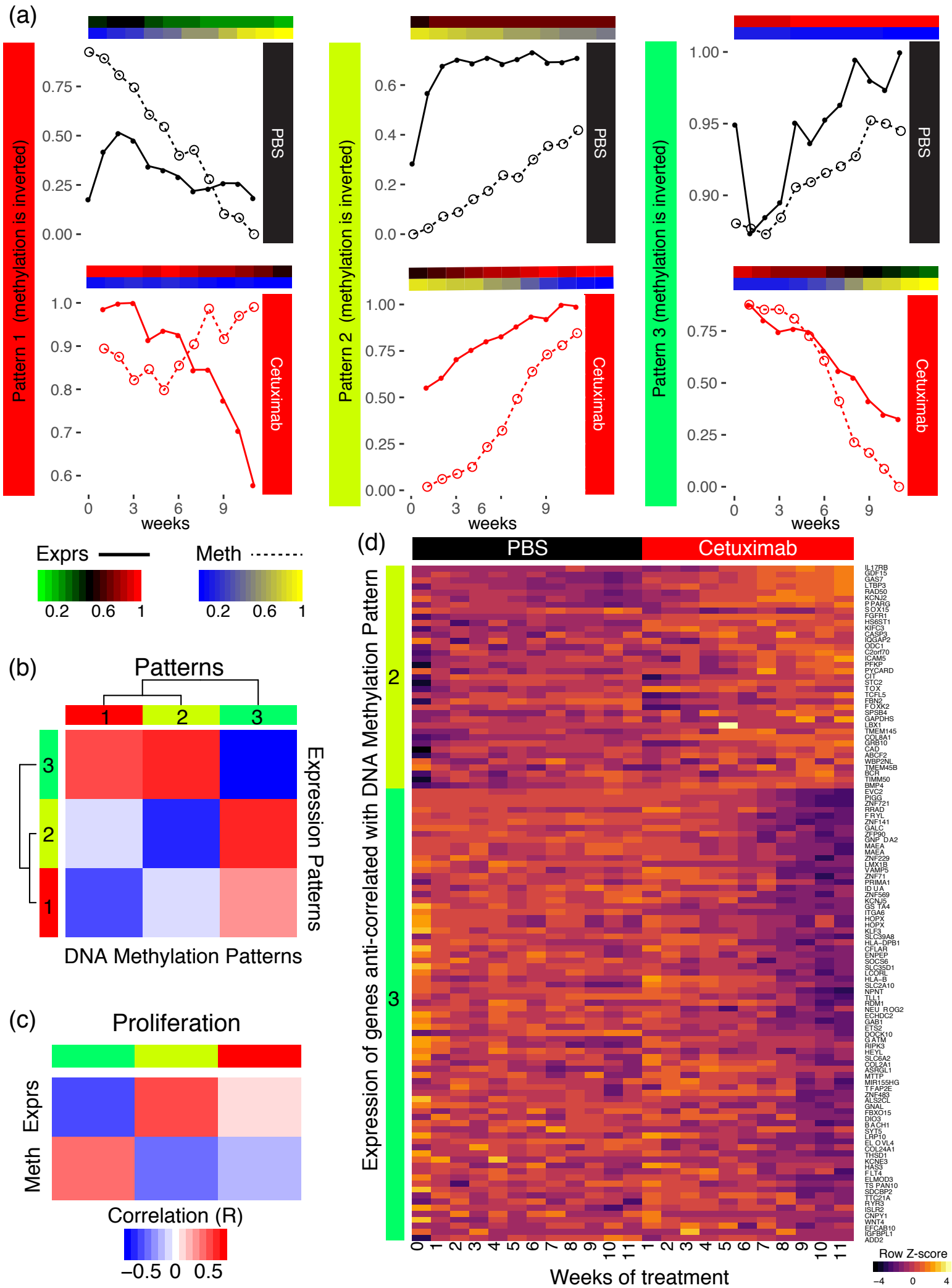
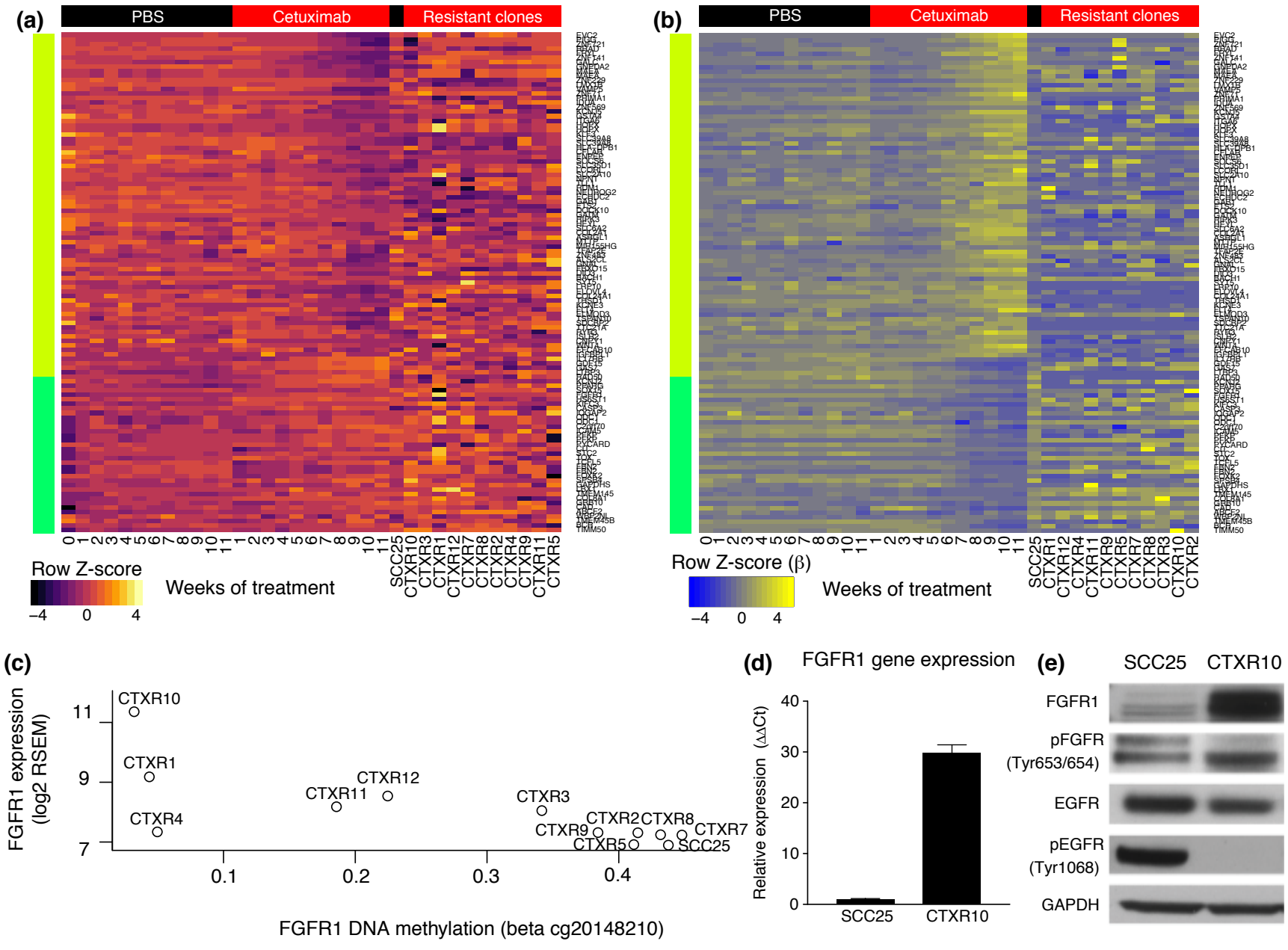
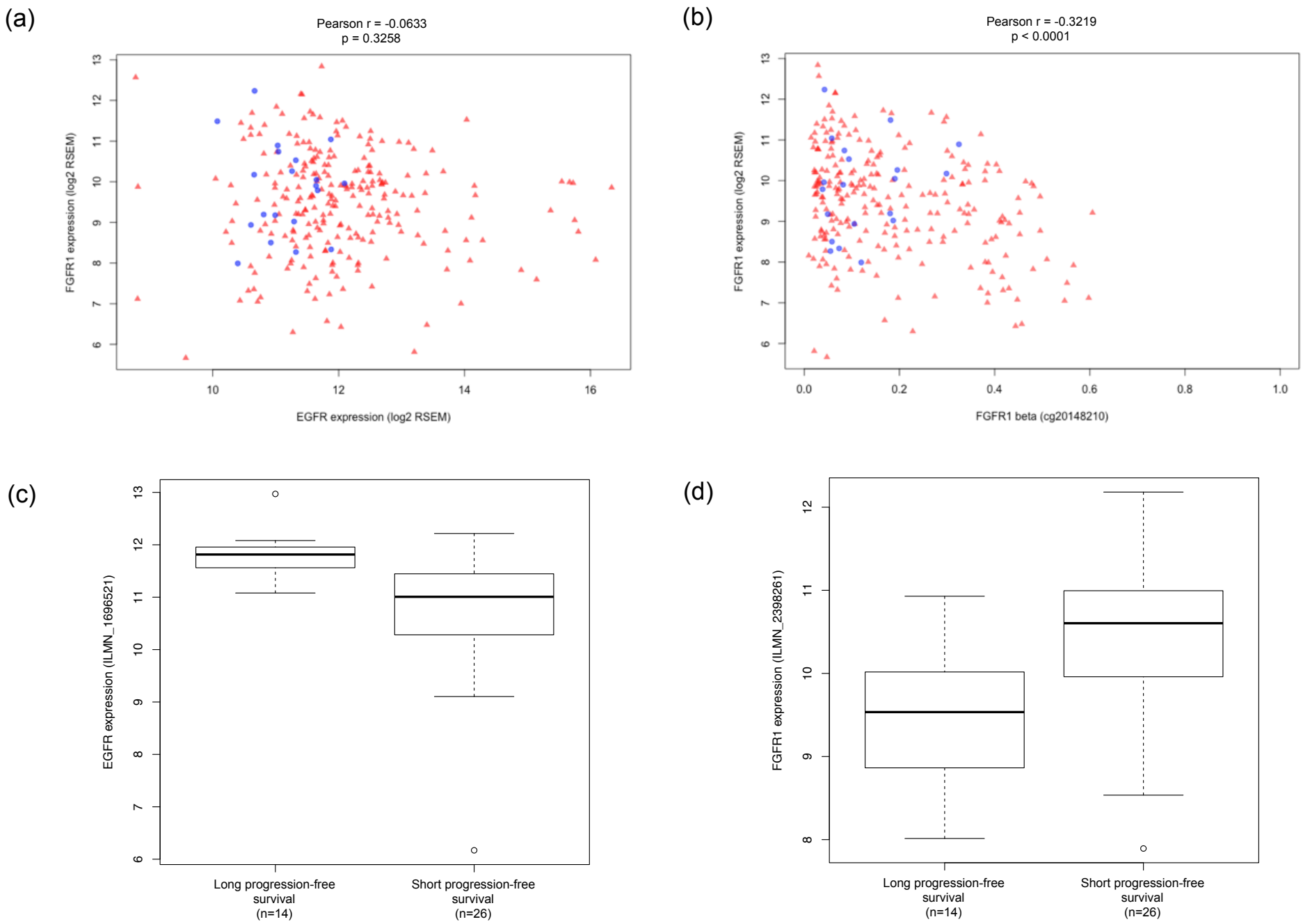


Figure 4 – Dynamics of DNA methylation alterations in CoGAPS are associated with epigenetic regulation of gene expression and the dynamics of acquired cetuximab resistance. (a) CoGAPS patterns for gene expression (red-green heatmap, solid lines) and DNA methylation (blue-yellow heatmap, dashed lines) of most anti-correlated patterns. Y-axis for gene expression increases with increasing values of the CoGAPS pattern for gene expression (labeled on left) and decreases with the CoGAPS patterns for DNA methylation (labeled on right) to reflect anti-correlation. (b) 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 3. The column colors for methylation patterns are selected to match the color of the corresponding expression pattern with maximum anti-correlation. (c) Heatmap of Pearson correlation coefficients for measured proliferation rates and CoGAPS patterns for gene expression (top) and DNA methylation (bottom). Columns are colored according to matched CoGAPS patterns between the data types defined in (b). (d) Heatmap of gene expression data for PatternMarker genes for CoGAPS analysis of DNA methylation (Supplemental Figure 3) with expression significantly anti-correlated with the CoGAPS DNA methylation pattern for acquired cetuximab resistance (pattern 2, yellow; pattern 3, green) to reflect discrepancies in timing between changes in DNA methylation and gene expression (a).



<u>Figure 5</u> – DNA methylation and gene expression patterns identified by CoGAPS in independent, stable SCC25 cetuximab resistant clones associate epigenetic regulation of *FGFR1* with acquired resistance to cetuximab *in vitro*. (a) Heatmap of gene expression values for PatternMarker DNA methylation genes that are anti-correlated with corresponding CoGAPS patterns for acquired resistance (Figure 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-green 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. (c) Expression of *FGFR1* gene expression relative to DNA methylation in stable cetuximab resistant clones. (d) QRT-PCR of *FGFR1* gene expression in CTXR10 relative to the parental cell line (greater than 30 fold change). (e). 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 is associated with slightly increased levels of phospho-FGFR1 and decrease in EGFR and phospho-EGFR.



<u>Figure 6</u> - Additional datasets confirms FGFR1 gene and protein overexpression *in vitro*, epigenetic regulation of *FGFR1 in vivo*, and inverse relationship between *EGFR* and *FGFR1* expression in *in vivo* cetuximab resistance. (a) Scatter plot of gene expression for *EGFR* and *FGFR1* in HPV-negative HNSCC samples from TCGA demonstrates 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* is anti-correlated with *FGFR1* expression in HPV-negative HNSCC, suggesting that up-regulation of FGFR1 might be a result of promoter hypomethylation in primary tumors. (c) *EGFR* expression is 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* is significantly overexpressed in patients with short progression free survival relative to patients with long progression free survival in this same dataset.