Deciphering the phenotypic heterogeneity and drug response in cancer cells using activity and interactions in chromatin domains

Neetesh Pandey*[1], Madhu Sharma*[1], Arpit Mathur[1], Chukwuemeka George Anene-Nzel[2,6,5], Muhammad Hakimullah[2], Priyanka Patel[3], Indra Prakash Jha[1], Omkar Chandra[1], Shreya Mishra[1], Jui Bhattacharya[7], Ankur Sharma[2,4,5], Roger Foo[2], Kuljeet Sandhu[7], Amit Mandoli[3], Ramanuj DasGupta[2], Vibhor Kumar[1,2#]

1. Department of computational Biology, Indraprastha Institute of Information Technology, New Delhi.
2. Genome Institute of Singapore, A*STAR, Singapore
3. National Institute of Pharmaceutical Education and Research-Ahmedabad (NIPER-A), Palaj, Gandhinagar 382355, Gujarat, India.
4. Harry Perkins Institute of Medical Research, QEII Medical Centre, Nedlands, Western Australia, Australia
5. Curtin Medical School, Curtin University, Bentley, Western Australia, Australia
6. Cardiovascular Disease Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore.
7. Department of Biological Sciences, Indian Institute of Science Education and Research (IISER) - Mohali, SAS Nagar, 140306, Punjab, India.

*equal contribution
co-corresponding authors, email: vibhor@iiitd.ac.in

Abstract

The effect of colocalization of genes in the topologically associated domains (TADs) and their activity orchestration in cancer samples and cells, hand-in-hand with drug-response have not been comprehensively studied. Here, we analyzed patterns in the activity of TADs in cancer-cells along with chromatin interaction profiles to understand their modes of drug-response. Analysis of transcriptomes of 819 cancer cell lines revealed that their response to multiple drugs were more correlated with the activity of TADs than individual genes. Applying our approach to 9014 cancer patients' data of 20 different cancer-types further strengthened the association between drug response and the activity of TADs, instead of individual genes, highlighting similarity despite

cancer heterogeneity. CRISPR-mediated knock-out of regulatory sites inside a TAD associated with cisplatin-response of oral cancer cells and discovery of primate-specific gain of synteny of genes within a TAD containing EGFR gene and its consequence, also demonstrate greater utility of TAD-activity based analysis.

Introduction

Chromatin organisation and orchestration of activity of different segments of genomes are important influencing factors in the transition of cellular states and response of cells. Multiple groups have investigated the link between chromatin organization and different properties of cancer-cells [1] and variable mutation rates in their genomes [2]. Jones *et al.* have mentioned that more than 50% of human cancers develop mutations in enzymes involved in chromatin organisation [3]. Even though genetic mutation and escape from apoptosis are involved in the development of cancerous state, regulatory changes in genes through perturbation in the activity of different chromatin segments always become instrumental in such transition. It is evident from previous reports that cancer therapy tends to spare some cancer-cells, even when they lack resistance mutations. [4]. In the past, several studies have tried to find reasons for drug-resistance of cancer-cells. Using mutation and expression profiles, a group involved in Cancer Cell Line Encyclopedia (CCLE) attempted to estimate response for many different drugs [5]. Many other groups have also tried to provide non-genetic explanations using reprogramming and stemness which is linked with increased xenobiotic resistance, in part, due to the efflux pumps expression [6–11], greater DNA repair effectiveness, [6] resilience and stress response [12–14]. While others have attempted to explain it in terms of the heterogeneity of tumors and the capacity of a few cancer-cells to transition states. In every scenario, chromatin organization plays a crucial role in the growth of cancer and its drug-resistant versions [15,16].

Chromatin is known to be organized in the form of different domains, often called topologically associated domains (TAD). TADs are naturally defined genomic regions that are segregated [17]. TADs have been found to be largely stable across cell-types [18,19] and during differentiation [20]. Besides mammalian genomes, other non-mammalian genomes are also organized in such domains. In fact, a few studies have demonstrated the conservation of a few TAD boundaries

across evolution such that approximately 54% of TAD boundaries in the human genome are also conserved in homologous regions in the genome of mice [18]. The gene co-localization also exhibit a pattern driven by cellular and evolutionary mechanisms, such as, a gene-dense locus 3p21.31 on chromosome 3 is known to harbor the largest cluster of tumor suppressor genes. In the past, reports of numerous deletions in locus 3p21.31 on chromosome 3 in a number of tumors have been made. The colocalisation and synteny of genes involved in similar or related functions could be due to multiple reasons. One of the reasons is selection pressure to avoid the recombination of beneficial alleles in a locus. Salem *et al.* have shown that paralogs are more often in the same TAD than random pairs of genes [21]. TADs define colocalised gene sets that are activated by eukaryotic cells in an orchestrated manner in different developmental stages and cellular states. The mutation rates in TADs and their boundaries have been studied in the context of the development of cancer and survival analysis [1]. A recent analysis by Akedmir *et al.* [22] revealed that changes in somatic mutational load in cancer genomes were co-localized with TAD boundaries. Several studies have discussed TAD-boundary disruption during the development of cancer. TAD boundaries insulate genes and enhancers and their activity from the effect of neighboring domains [23]. Hence, the abnormality in activity of genes inside a TAD could be a more direct indication of effective disruption of the TAD boundary. Therefore, we used approach of estimation the activity of gene-set consisting of genes inside a TAD to study their role in cancer and drug response.

First we analysed chromatin interaction landscape for head and neck squamous cell carcinoma (HNSC) cell lines to comprehend the genomic domain activation pattern in drug resistance. Later, we expanded our study by analyzing the transcriptome of 819 cell lines reported in the CCLE database and response for 544 drugs reported in the CTRPV2 portal. Further, we used our approach to draw inferences from transcriptome profiles of 9014 patients across 20 cancer types made available by the TCGA consortium.

Results

Our integrative analysis of chromatin interaction profiles and single-cell transcriptomes of patient derived HNSC cell lines, laid the foundation for analysis of larger data-sets of transcriptomes in CCLE and TCGA databases using our unique approach.

Oral cancer Hi-C profiling, TAD patterns and their association with drug resistance

We profiled chromatin-interaction landscape of HNSC cell lines using Hi-C protocol. Specifically, we used patient-derived primary cultures (PDCs) of HNSC established by Chia *et al.* [24], which include : primary HNSC cells of two patients' primary tumors (HN148P, HN137P) and metastatic cancer-cells of patient HN120 (HN120M) (Supplementary Figure S1). We also performed ChIP-seq using antibodies for H3K27ac and H3K4me3 histone modifications in HN137P, HN137M cell lines. For single-cell transcriptome, we used UniPath [25] to calculate TAD gene-set enrichment in every single-cell. To calculate TAD activity using the bulk expression profile, we used GSVA [26]. Here, we call TAD gene-set enrichment score as TAD-activity. A brief outline of the work has also been shown in Figure 1A.

TAD activity alterations with cisplatin resistance development

We calculated TAD-activity scores for all TADs in the union list of TADs detected using Hi-C profile of HNSC cells. We contrasted the TAD-activity scores of oral cancer cell lines and their cisplatin-resistant variants. For this purpose, we used the single-cell expression profiles of HNSC cell lines HN137P (paired-end), HN137M, HN120M and their cisplatin resistant versions HN137PCR, HN137MCR and HN120MCR published by Sharma *et al.* [27] and further confirmed for correctness using mutational profiles [28]. We found that 14 TADs had upregulation (P-value < 0.05, fold change > 1.5) in activity in cisplatin-resistant versions of all three HNSC cell lines (HN137PCR, HN120MCR, HN137MCR) (Supplementary Table 1). Among 14 TADs, three TADs are from 13q13.1 band and harbor genes associated with actin filament reorganisation, mitotic sister chromatid cohesion and nucleotide excision repair like BRCA2, FRY and PDS5B. Some commonly upregulated TADs in other regions also include genes associated with mitotic sister chromatid cohesion, response to DNA damage, TGF-beta activation, drug transport and mitochondrial translational elongation known to involved in resistance and response to cisplatin (for references, see, Supplemental File-1).

Overall, 20 TADs had consistently downregulated activity in cisplatin resistant versions of multiple cell lines (HN137PCR, HN120MCR, HN137MCR) (Figure 1B-C, Supplementary Table 5). Since we allowed overlap among TAD regions, eight out of 20 commonly downregulated

TADs lied with in region chr11:70150000-71470000 (hg19) in 11q13.3 and 11q13.4 bands and showed some hierarchy. For brevity we call them as 11q13.3-4 TADs. All the 11q13.3-4 TADs have 5 known keratin associated protein genes (KRTAP5-7, KRTAP5-10, KRTAP5-9, KRTAP5-8, and KRTAP5-11), DHCR7 and NADSYN1. Out of eight 11q13.3-4 TADs the longer TADs also contain SHANK2, CTNN and a few more genes with both known and unknown functions. Previous studies, have reported copy number amplification of genes in 11q13.3 and 11q13.4 regions in HNSC and other types of cancer. However, our analysis provides an important therapeutical insight that HNSC cells with higher activity of 11q13.3-4 TADs are sensitive towards cisplatin.

Visualization of the chromatin interaction landscape of HNSC cells in chr11 revealed that the genomic region in 11q13.3-11q13.4 bands is connected to other regions in same chromosome (Figure 1D, regions R1, R2 R3, R4, R5, R6, R7, R8). Whereas, such intensity of interaction of 11q13.3-4 bands and R1-R8 regions was not found in other cell types, like cardiomyocytes and astrocytes (Supplementary Figure S2). The 11q13.3-4 TADs could be part of a larger regulatory condensate in HNSC cells. In addition, to 11q13.4 TADs, (with genes KRTAP5-7, KRTAP5-8, KRTAP5-9, KRTAP5-9, KRTAP5-10, KRTAP5-11) and the region R1 connected to them also has many genes of family KRATAP5 (KRTAP5-2, KRTAP5-1, KRTAP5-4, KRTAP5-3, and KRTAP5-5). Whereas regions R2, R3 and R4 have significant enrichment (Pval < 0.05) for gene-set of ‘positive regulation of cell growth’, central carbon metabolism in cancer and ETS family genes respectively, which are known to promote tumor progression. The region R7 has enrichment for folate receptor activity which is known to be a upregulated in oral cancer (Figure 1E)[29]. Further we chose H3K27ac ChIP-seq peaks (in HN137 cell lines) present in 11q13.3-4 TADs and regions connected to them (R1-R8). The motif enrichment analysis for selected peaks showed top motifs for TFs EWS: ERG, FLI1, and ETV2, which belong to the ETS family (Figure 1F). Involvement of ERG and FLI1 in multiple cancers has been reported previously [30]. The FLI1 is itself known to be involved in cell-cycle regulation [31]. Overall, our approach of analysis of HNSC cell profiles revealed important TADs in 11q13.3-11q13.4 bands lying at the hub of chromatin interaction of regions which could be involved in response to stress response and could be important targets for influencing viability of HNSC cells.

Generalization of relationship between drug sensitivity and TAD activity through analysis of CCLE data

The insights from changes in the activity of a few TADs in HNSC cell lines in the presence of cisplatin highlighted the need for comprehensive analysis of association of TADs with drug-response in wider range of cancer cell lines. Hence, we first compiled a cumulative TAD list by making a union list of TADs detected in our HNSC cell lines, and TADs reported in the TADKB database [32]. Our approach is based on the reports by multiple researchers on the conservation of TAD boundaries across cell-types [18,20], which we further verified in cancer cells also. We found that majority of TAD boundaries from our common TAD list, are conserved across different cell-types including cancer cells also (Supplementary Figure S3A-B). Such observation encouraged us to use our common TAD list to calculate their TAD activity across 819 CCLE cell lines[5]. We calculated the correlation between the activity of TADs in the union list and the IC50 value of drugs reported for CCLE cell lines[33] (see Methods). Compared to natural TADs, the null-model TADs with random boundaries had lower correlation between their activity level and drug PIC50 values (Supplementary Figure S4). For HNSC cell lines there was high correlation between the activity of a few TADs and PIC50 value of many drugs. Such as, the activity of TAD in location hg19:chr13_58180000_59940000 had a high correlation (R=0.79) with IC50 value of the drug Importazole. Similarly the activity of 11q13.3-4 TAD (chr11:70150000-71300000) had substantial negative correlation (less than -0.3) with PIC50 values of more than 10 drugs (supplementary Figure S5).

Further we performed clustering of values of correlation between TAD-activity and drug PIC50 values (see Figure 2A). We divided the TADs into 8 major clusters and observed a class of TADs whose activity had both a positive correlation with PIC50 values for a group of drugs and a negative correlation with PIC50 values for another group of drugs. As shown in Figure 2A, cluster 3 TAD activity had a negative correlation with PIC50 values of drugs in the D1 group but a positive correlation with PIC50 values of the D2 group of drugs. On the other hand, TADs in cluster-1 and cluster-5 had a negative correlation with PIC50 value of most drugs. In addition, drugs belonging to group D3 showed quite a distinct pattern in each cluster. Drugs belonging to group D4 had a PIC50 value that negatively correlated with the activity of TADs in cluster-5 and positively

correlated with cluster-2 TAD (Figure 2A). Such results can potentially guide toward better drug combinations according to TAD-activity.

When we repeated the same analysis using cell lines for all cancer types in the CCLE database, the correlation between TAD activity and drug PIC50 value decreased (Figure 2B). Such a decrease in correlation values indicates that the association of TAD activity with response of cancer cell towards drugs is influenced by the context of cancer type. We found varying correlations for the same set of TAD and drugs across different cancer-types (Supplementary Figure S6). Nevertheless, there were still a few TADs whose activity had a notable correlation with drug response even when cell-lines of all types of cancers in the CCLE set were used. We again divided the TADs into 8 clusters based on correlation with IC50 values (Figure 2B). Such as, TADs activity in clusters 4,5 and 8 (Figure 2B) exhibited a strong negative correlation with the PIC50 value of multiple drugs. Activity of TADs in Cluster 1 is positively correlated with PIC50 value of multiple drugs (drug-sensitive). The biological process term that is most enriched for TADs in cluster1 is 'negative regulation of IL-6' which corroborates with previous findings [34]. The biological process of keratinisation, keratinocyte differentiation, cornification and epidermis development appeared among top enriched terms for TADs in clusters 3 and 8. The drugs for which TADs in clusters 3 and 8 have negative correlation with PIC50 value of drugs mostly belong to diverse groups. TADs in Cluster 3 and 6 also have enriched terms with glucoronidation known to be involved in drug resistance [35]. There were many drugs which had high correlation between their IC50 values and only one or two TADs. Such as drug 1,9-Pyrazolonthrone had negative correlation (below -0.30) only with activity of TAD at chr20:43940000-44410000 (see supplementary file -2). Overall, the correlation values of TAD activity with drug PIC50 for multiple types of cancer cells, highlighted a few common mechanisms of response towards multiple drugs in spite of heterogeneity among different types of cancers.

While using all cancer cell lines of the CCLE database for associating TAD activity with drug-response, we found an interesting pattern: for many drugs, the correlation between the activity of a few TADs and the IC50 values was substantially higher than the best correlation for any single-gene (Figure 2C). (Supplementary Figure S8 A-B). Such results highlight the fact that even if

genes involved in response to a drug could vary in different types of cancer, the mechanism would remain the same. Such as many paralogous genes with similar functions often co-localise. We found that for many drugs, even if we used many genes in machine learning model to predict PIC50 values, the accuracy achieved is much lower than using TAD activity based prediction (Figure 2D). We also tried to compare the correlations (with IC50 values) of TADs with their own respective genes to confirm the hypothesis of co-functionality. We found that for many TADs none of their own gene had comparable correlation between expression and IC50 value (Figure 2E-F). The observations were similar for both the cases, i.e, drug sensitivity and drug resistance (Supplementary Figure S9). A few examples, include the drug Ciclopirox, LY-2183240, Lenvatinib, Axitinib, and Foretinib, where association between IC50 value was significantly greater with TAD activity than any of their genes. Such results indicate that joint activity of genes in a TAD could be more effective in deciding response toward drug than using single genes.

TAD activity pattern and its influence in TCGA PAN-cancer

There could be multiple reasons for variation in TAD activity among patients, such as i) change in the expression level of genes ii) mutation at TAD boundary or genomic structural variation which could cause shuffling of genomic regions causing smaller TAD formation. ii) CNV of the genomic region consisting of the TAD. Nevertheless, each of these events has an effect on the gene-set activity of TADs. Hence we studied variations in TAD-activity using the cancer sample transcriptomes provided by the TCGA consortium. We used the same union list comprising of TAD detected in Hi-C profiles in HNSC cell lines and TADs reported in the TADKB database [32]. The calculation of differential activity of TADs in tumor and normal samples and their clustering reveal multiple TADS with cancer-specific activation. However, there were multiple TADs which had similar differential activity in multiple cancer types. As shown in Figure 3A, TADs belonging to cluster 12 and cluster 11 showed differential activity in multiple types of cancer. However clusters of TADs made using their differential activity, could not highlight any particular biological process with significant enrichment (FDR < 0.1), except for cluster-11. Such results indicate that TAD activity perturbed in cancers of different types and stages, affected multiple biological processes.

The analysis to estimate differential activity of TADs in TCGA cancer samples indicated that there is a need to further classify or re-group the samples based on different criteria to highlight targetable pathways or biological processes. Therefore, we focused more on TADs associated with survival in 20 different types of cancer (see methods, supplementary Figure S10). The number of TADs associated with survival in 2 or 3 cancer types was substantially more than TADs with survival-association in only one cancer type (Figure 3B) (supplemental File 3). Analysis of Hazard ratios revealed that many TADs could have significant associations with survival in multiple cancers and their effects could vary based on the context of cancer type (Figure 3D). The results of LUSC and BRCA cancer are provided in supplementary figures (Supplementary Figure S13 A-B). To confirm top survival-associated TADs without the effect of bias due to conditional dependencies, we also used Bayesian network modeling where every TAD was represented by a node (see Methods) (Supplementary Figure S11).

Further, we tested the hypothesis that TAD association with survival could be artifact of the effect of a few genes, which could be independently more influential for survival than the activity of its host TAD. We found that for multiple TADs, their activity had a higher influence on survival than any of the genes inside them. The fraction of such TADs with more influence on survival than any of their genes were 57%, 56% and 44% in HNSC, BRCA and CESC cancer, respectively (Figure 3C). Such as a TAD containing EGFR gene and located at chr7:51500000-57100000 has more significant association between its activity and survival than any of its genes in HNSC, BRCA and CESC. For HNSC, other examples are TADs located at chr3:97760000-98290000 and chr1:156110000-156900000 which have more association with survival than any of the genes lying with in their boundaries (Figure 3E-G). While the TAD at chr3:97760000-98290000 hold multiple olfactory receptor genes, the TAD at location chr1_156110000_156900000 harbors genes for which enriched pathway terms are "signalling to P38 via RIT and RIN" and "NGF-independent TRKA Activation". Overall, our analysis using TAD activity, revealed its potential to highlight the possible effect of colocalisation of genes in influencing cancer and possibly a few unknown mechanisms or conserved coordination of genes towards a major biological process that influence survival in cancer.

Further we hypothesized that TADs associated with high malignancy and low survival could also be involved in multi-drug resistance of cancer cells. We sorted the TADs associated with survival (P-val < 0.05) in TCGA-HNSC patients according to the number of drugs with which they had an absolute correlation values higher than 0.3 for HNSC cell lines in the CCLE database. The top 50 TADs (survival-associated) providing resistance (negative correlation with PIC50; r <-0.3) towards the highest number of drugs had a significantly higher hazard ratio than the top 50 TADs correlated with sensitivity towards the maximum number of drugs (see Figure 4A). The same pattern was observed when top 10 or top 100 TADs were used. Out of top 10 TADs (with survival P-value < 0.05) correlated towards resistance with highest number of drugs (r < -0.3 with PIC50) , 50% had hazard ratio above 1.25. Whereas none of the top 10 TADs (with survival P-value < 0.05) correlated for sensitivity towards highest number of drugs (r > 0.3 with PIC50), had a hazard ratio above 1.25 (Figure 4B). A similar trend was observed when top-50 or top-100 TAD categories were used or enrichment of TADs with a hazard ratio less than 0.75 were counted (see Figure 4C). Thus, our analysis reveals coherence of survival outcomes in patients with influence of TAD-activity on drug-response in cancer cell lines. It also highlights a few TADs as potential markers of malignancy and multi-drug-resistance to be considered for careful therapeutic decisions.

Case study of TADs in 11q13.3-11q13.4 region

We focused on 11q13.3-4 TADs as they appeared to be at center of hub of chromatin interaction in chromosome 11 in HNSC cells and located on genomic site with high copy number amplification in HNSC tumor of multiple patients in TCGA cohort (Figure 5A). We visualised and investigated Hi-C profile of HNSC cells around other TADs with top association with survival, however we could not find such a strong hub like property (supplementary Figures S2 and S14). Intrigued by peculiarity of 11q13.3-4 TADs, we first tested the hypothesis that the presence of KRTAP5 family genes could be influencing its importance (Figure 5B). The association between enrichment of KRTAP5 gene-set and TCGA-HNSC patient survival was insignificant (P-val=0.054). Moreover, the promoters of KRTAP5 family genes in 11q13.3-4 TADs did not have H3K27ac or H3K4me4 histone modifications at their promoters in 120M, 137P and HN137M HSC cell lines (supplementary Figure S15). The CNV of gene-set in 11q13.3-4 TADs is significantly (P-val < 0.05) associated with survival of HNSC patients (Figure 5C). Therefore, in

addition to KRTAP5 family genes, there could be other genes and regulatory elements in 11q13.3-4 TADs which could be essential for HNSC cells. The synteny of genes in 11q13.3-4 TADs appear to be mostly preserved across most of the mammals (Figure 5D) (see Methods) and point towards a possible conservative force for the colocalisation of genes for supporting each other towards a common biological role. We looked for cis-regulatory sites in 11q13.3-4 TADs which could be essential for HNSC cells. We found two locations with high chromatin-accessibility in tumor samples of TCGA-HNSC patients and overlap with eRNA in HNSC samples as reported by the TCGA consortium (Figure 5E) (Supplementary Figure S16). CRISPR based knockout using gRNA for those two locations decreased the viability of HNSC cell line Cal27. However, CRISPR based knockout using the same guide RNA in control cells (HEK293) did not have any effect (Figure 5F). RT-qPCR results showed increased expression of housekeeping FADD, DHCR, RNF, CTTN, PPFIA in Cal27 cells, which indicated that CRISPR-KO of selected regulatory sites induce cell death . In control HEK293 cells, the genes FADD DHCR, RNF, CTNN and PPFIA did not show any change in expression upon CRISPR-KO of selected regulatory sites (Figure 5 G). Thus, our results indicate that identifying important TADs can lead to finding relevant enhancers to be targeted to affect the viability of cancer cells.

Case Study of TAD with EGFR gene

As demonstrated above the activity of TAD at chr7:51500000-57100000 containing EGFR gene seems to have more significant association with survival than EGFR gene in multiple cancer types. In our analysis for HNSC, EGFR gene expression had a lower significance of association with survival than neighboring genes like PSPH, LANCL2, and SEC61G in the same TAD. For brevity, we call the TAD located at chr7:51500000-57100000 as EGFR-TAD. Our results raise important question regarding the importance of colocalisation and function of neighboring genes of EGFR. Therefore we used a multipronged approach to understand EGFR-TAD. The EGFR-TAD tends to have high copy number in tumors of a few TCGA-HNSC patients (Figure 6A). Moreover, the genes in EGFR-TAD, gained synteny in apes/old world monkeys even if it remained non-syntenic throughout other mammals and new world monkeys (Figure 6B, supplementary Figure s17B). The gain of synteny of genes in apes and old world monkeys in EGFR-TAD hints about their collaborative effect in a common cellular mechanism involved in development of brain and severity of cancer which we further investigated. The effects of CNVs of individual genes inside

the EGFR-TAD on survival also tend to be similar like EGFR gene in multiple types of cancer (Figure 6C), including HNSC and Glioblastoma. A gene PSPH in the same TAD is involved in endogenous serine production through aerobic glycolysis (Figure 6D) needed by cancer cells. Surprisingly, the EGFR-TAD has a very high enrichment of genomic variations associated with Serine and Glycine measurements in blood and brain volume and glioblastoma (Figure 6E). The high enrichment of GWAS mutations associated with brain volume and gain of synteny in primates with larger brain implies primate-specific adaptive selection on the locus, though at the cost of increased cancer susceptibility. Further investigation revealed a systematic cooperation of genes in EFGR-TAD for a crucial process needed for brain development as well as cancer. It is to be noticed that aerobic glycolysis, known as Wardburg effect is used by cancer cells and brain cells for production of molecules like amino acids and nucleotides as well as fatty acid synthesis from glucose [36] for cellular proliferation and maintenance. Analysis based on CCLE cell-line transcriptome also shows that many drugs with high positive correlation between their PIC50 value and activity of the EFGR-TAD affect glycolysis and related processes. Such as in breast cancer cell lines top 3 drugs with positive correlation of PIC50 value with EGFR-TAD are Piperlongumine, cerulenin and AZD6482 which have been shown to affect glycolysis and related process directly or indirectly (see supplementary Figure S19 and supplementary table 3). Similar trends were observed for HNSC and LUSC cell lines (Supplementary Figure S19 and supplementary table 3). We also found that multiple genes in the EGFR-TAD have been shown to be associated with aerobic glycolysis and related processes (Figure 6F, supplementary Table-4). Such results indicate that genes in EGFR-TAD are promoting cell proliferation in two modes. First mode is the activation of RTK pathways through EGFR. The second mode is, by supporting aerobic glycolysis for synthesis of biomolecules.

Discussion

Here, we used a novel approach of using TAD-activity to study cancer heterogeneity and drug-response. In addition to gene-set, TAD-activity could also be partially reflective of enrichment of activity of other set of regulatory elements in a TAD such as enhancer-set. One of the reasons for using our approach is that there are multiple factors involved in formation and strengthening of TAD boundaries such as CTCF, G4 quadruplexes [37] and repeats [38]. Barutcu *et al.* have shown that deleting CTCF-rich region near TAD boundaries does not necessarily disrupt the TADs [39].

The high number of conserved interactions within TADs also indicate that TADs are regulatory units. Hence our approach of using TAD-activity could be truly a direct and effective way to estimate effect of disruption in TAD-based regulatory units.

The analysis of HNSC cell lines transcriptomes highlighted TADs in 11q13.3–11q13.4 band which could cause cisplatin-sensitivity but resistance towards multiple drugs like azacitidine and importazole in oral cancer. Most notably, a region interacting with TADs at 11q13.3 and 11q13.4 band also harbor many genes for keratin associated proteins. Keratin-associated proteins have not been commonly reported for association with properties of cancer-cells except by a few anecdotal studies [40]. It could be possible that clusters of genes for keratin-associated proteins and other response genes colocalise for co-expression for a major biological process. Nevertheless, the case study of 11q13.3-4 TADs highlighted benefit of our approach in reducing search space for the identification of cis-regulatory elements for the purpose of cancer therapeutics. It is possible that sgRNA designed by us for CRISPR based knock-out of an eRNA-location in a 11q13.4-4 TAD could also be effective for other cancer type also especially for tumor with amplification of the same region. Thus it could help to reduce dosage of cisplatin and open new avenue for cancer therapeutics with minimal side-effect.

Our analysis clearly indicates that TAD activity could be used as a biomarker for predicting drug response or survival estimation to cope with heterogeneity among tumors. In many cases TAD-activity could be better biomarker than any gene for drug-response and survival analysis. The current paradigm of therapy for cancer is rapidly evolving with new combinations of drugs. Developing a framework for selecting effective drug combinations to prevent redundancies can be aided by the knowledge of common mechanism highlighted by TAD activity based analysis.

Our multipronged analysis highlighted role of multiple genes in EGFR-TAD in aerobic glycolysis and related functions, which reveal a link between an evolutionary force associated with development larger brain and severity in cancer. In addition to cell-cycle and drug-resistance, aerobic glycolysis has also also been associated with immune evasion of tumor cells by modulation of T cell activity [41]. Hence targeting glycolysis or related genes in EGFR-TAD can possibly help in improving effectiveness of cancer immunotherapy [41][42]. Thus the example of EGFR-TAD shows that study of collaborative role of genes in TADs can lead to identification of

functionally effective targets and biomarkers to support current regime of cancer diagnostics and therapeutics.

Methods

Experimental method

The culture of HNSC patient-derived cell lines (HN137P, HN120P, HN120M, and HN148P) as well as their ChiP-seq and Hi-C based profiling and sequencing, was done at Genome Institute of Singapore.

Culture of HNSC PDC and their cisplatin-resistant model

The creation of the PDC model of HNSC is described by Chia *et al.* [24]. Cells were kept in a humidified atmosphere of 5% CO_2 at 37 °C. The cell line identity was authenticated and confirmed by Chia *et al.* through a comparison of STR profiles cell lines to their original tumor. The cisplatin version of the same HNSC cell lines was made by Sharma *et al.* by treating mice xenograft model with cisplatin for 3 weeks. The cells dissociated from the residual xenograft model were isolated into single-cell suspension followed by flow-cytometric analysis of cell-type and stem-state markers. Flow cytometry data were analyzed by FlowJo.

Hi-C profiling

In situ Hi-C was performed as previously described [43]. Five million HN137 and HN120 cells were cross-linked with 1.5% formaldehyde for 10 minutes at room temperature, after which glycine was added to stop the reaction. Cross-linked cells were lysed in 500 µL cold lysis buffer (10 mmol/L Tris-HCl, pH 8.0, 10 mmol/L NaCl, 0.2% (vol/vol) Igepal CA630, mixed with protease inhibitors) on ice for 30 minutes. The pellet was washed twice with cold 1× NEBuffer 2, and chromatin was solubilized in 50 µL 0.5% SDS and incubated at 62°C for 10 minutes. SDS was quenched using Triton X-100 (Sigma 93443), and chromatin was digested overnight by adding 25 µL 10X NEBuffer 2 and 100 U Mbo I (New England Biolabs, catalog No. R0147). Digestion efficiency was confirmed by isolating DNA from a 25-µL aliquot of the suspension and performing agarose gel electrophoresis. Digested ends were filled and labeled with biotin by adding 37.5 µL 0.4 mmol/L Biotin-14-dATP (Thermo Fisher Scientific, catalog No. 19524016), 1.5 µL 10 mmol/L

dCTP, 1.5 µL of 10 mmol/L dGTP, 1.5 µL of 10 mmol/L dTTP, 8 µL of 5 U/µL DNA polymerase I, and large Klenow fragment (New England Biolabs, catalog No. M0210); and incubated at 37°C for 90 minutes. Blunt-end ligation was performed by adding 463 µL water, 120 µL 10X NEB T4 DNA ligase buffer, 200 µL 5% Triton X-100, 12 µL 10 mg/mL BSA, and 5 µL T4 DNA Ligase (New England Biolabs, catalog No. M0202). The solution was incubated at 16°C for 4 hours. After ligation, cross-link was reversed overnight, and DNA was purified with the phenol-chloroform method. Next, Hi-C libraries were treated with T4 DNA Polymerase (New England Biolabs, catalog No. M0203) to remove biotin from unligated ends. To do this, 5 µg Hi-C library was combined with 2.5 µL of 1 mmol/L dATP, 2.5 µL of 1 mmol/L dGTP, 10 µL of 10X NEBuffer 2.1, and 5 µL of T4 DNA Polymerase (New England Biolabs, catalog No. M0203), and water was added to a final volume of 100 µL. The reactions were incubated at 20°C for 4 hours and stopped by incubating at 75°C for 20 minutes. Samples were sheared to a mean fragment length of 400 base pairs (bp) using Covaris M220. Pull-down of biotinylated DNA was performed with Dynabeads My One T1 Streptavidin beads (Thermo Fisher Scientific, catalog No.65601), and library preparation was performed using NEBNext Ultra II DNA Library Prep Kit for Illumina. Samples were quantified with KAPA, and sequencing was performed on a Hiseq 4000.

ChIP-seq profiling

Cells were cross-linked for 30 minutes (10% FBS/DMEM containing 1.5% PFA) followed by the addition of glycine (1/20 volume of 2.5M) and incubation on ice (5 min.). It was re-dissolved in a buffer followed by sonication. The buffer used for re-dissolving contained 10 mM Tris-HCl, pH8, 1 mM EDTA, 0.5 mM EGTA, and 1X protease inhibitor cocktail (Roche). For pulling the chromatin H3K4me3 or H3K27ac antibody-coupled Dynabeads and 150 µl of ChIP buffer (433 mM NaCl, 0.43% sodium deoxycholate, 4.3% Triton X-100) were used. For the input control library, 1/10 of the chromatin mixture was kept separately. RIPA buffer was used to wash the beads which were later incubated in 100 µl of elution buffer (50 mM Tris-HCl, pH8, 1% SDS, 10 mM EDTA) at 65 °C for 20 min with brief vortexing. De-crosslinking of immunoprecipitated chromatin and input samples was done by incubating at 65 °C overnight and subsequently incubating with 0.2 µg/µl of RNase A at 37 °C for 1 h and with 0.2 µg/µl of proteinase K at 55 °C for 2 h. DNA was isolated from the chromatin using phenol/chloroform/isoamyl alcohol (49:49:2) and ethanol and precipitated in presence of 40 µg of glycogen. Pellet was then suspended in 10mM

Tris-JCL, pH8 after washing with 70% ethanol. Library preparation and multiplexing for ChIP-seq libraries were performed using the NEBNext ChIP-seq library Prep kit (NEB). The sequencing of ChIP-seq and input control libraries was performed using Illumina HiSeq-2500.

CRISPR knockout of enhancers in 11q13.4 region

sgRNAs were designed using online CRISPR designing tool "CHOPCHOP" (https://chopchop.rc.fas.harvard.edu/). The sequence for coordinates chr11:71127012-71133012 and chr11:71124803-71130803 were taken from NCBI and given as input sequences for sgRNA designing. Two sgRNA for each coordinate were chosen. The oligos obtained were 20-bp long. Further, complementary pairs for the selected sgRNAs were made using reverse complement tool (https://www.bioinformatics.org/sms/rev_comp.html). Necessary oligos and primers specified by online tool were ordered from Sigma Aldrich.

The details about the preparation of sgRNA construct and transfection into cell lines are described in supplementary methods.

Computational analysis

Analysis of Hi-C profile

To construct genome-wide maps with millions of interactions between genomic locus pairs, Hi-C creates hundreds of millions of read-pairs [44]. In our analysis HiCUP tool was used to process the Hi-C data. The same pipeline was also used to remove artifacts for down-stream analysis [45,46].

Analysis using HiCUP pipeline

The HiCUP pipeline involves numerous programs like hicup_truncater, hicup_mapper, hicup_filter, hicup_deduplicator, and hicup_digester- which prepares the genome digest file used by hicup_filter. The paired FASTQ files were used to generate Hi-C di-tag paired reads, which were then mapped to the hg19 reference genome. This HiCUP script controlled the pipeline carrying output from one section of the program to the next and executed each script sequentially. The configuration file for the HiCUP program also contains the parameters for the complete pipeline [45]. The majority of reads produced by the HiCUP Mapper script are undoubtedly

authentic Hi-C products, a substantial minority are likely not and should be eliminated. For the purpose of finding legitimate Hi-C pairs, the HiCUP filter script is used for analysis of paired reads along with the file produced by the HiCUP Digester. This might not necessarily be the case, though, as the sequenced region may contain the Hi-C ligation junction. As a result, potentially useful data will be lost when such reads are eliminated from the Hi-C pipeline during the mapping process. By locating ligation junctions among reads and removing sequences downstream of the restriction enzyme recognition site, the HiCUP truncater script helps to fix this [45].

Pipeline implementation

We used the DpnII digest profile of the hg19 reference genome created using hicup_digester for computational analysis. This file represented all possible DpnII fragments in the genome and was used to identify Hi-C artifacts. The alignment process was performed using HiCUP v0.5.10 with Bowtie2 (hg19), and the filtering step's minimum and maximum di-tag range were set to 150 and 1000, respectively. The default settings for the other parameters were used. The final BAM output was produced with a satisfying unique alignment score of more than 65%.

Generation of Hi-C contact matrix

We first converted the binary alignment mapped file into arrowhead input format using the samtools view command to create the Hi-C contact matrix [47]. The normalized Hi-C contact matrix (.hic file) was created using the Juicer tool's precommand[48] with the help of the arrowhead input file. The dump command of Juicer Tools was used to export the observed contact count for each region of the Hi-C map with balanced Knight-Ruiz (KR) normalization. With this, we obtained the text Hi-C matrix (.txt) at 25 kb resolution for all the chromosomes.

Detection of topologically associated domains (TADs)

We used the DomainCaller approach from Dixon *et al.* to identify topologically associated domains at 25 kb on all chromosomes [18,49]. We started with a square contact matrix of the 25 kb Hi-C contact map, and subsequently, directionality indices were created using DI_from_matrix.pl with a bin size of 25,000 and a window size of 125,000. This vector of directionality indices was provided as input for the HMM_calls.m script. Further, we obtained TAD calls by processing HMM via file_ends_cleaner.pl, converter_7col.pl,

hmm_probablity_correcter.pl, hmm-state_caller.pl, and, hmm-state_domains.pl. The HMM probability correction script was run with the following parameters as input: min = 2, prob = 0.99, and bin size = 25,000 [49]. Secondly, we also got TADs from the TADKB database, which gives TADs at resolutions of 50 kb and 10 kb from a variety of cell types, namely human HMEC, NHEK, GM12878, IMR90, K562, HUVEC, and KBM7 [19,32]. TAD locations for each cell type were obtained with the help of the Directionality Index (DI) [32].

Analysis of ChIP-seq profiles

Analysis of ChIP-seq profiles was done using DFilter [50] to call peaks and make custom tracks. The parameters used for calling peaks were according to the instructions provided on the DFilter help page (https://reggenlab.github.io/DFilter/).

TAD gene-set procurement and calculation of TAD activity score

The TADs gene-set was constructed using genes found in certain genomic locations on the hg19 genome version after intersecting with the corresponding promoter regions. The widespread availability of big data sets from multiple assays with numerous complex phenotypes inspired us to study the functional mechanism of TADs in the context of transcriptomics. To facilitate the interpretation and organization of TAD gene-sets in large microarray experiments, gene set enrichment was done. This approach has the advantages of noise & dimension reduction and better biological interpretability [51]. The Gene Set Variation Analysis (GSVA) R-package was used to calculate gene set enrichment for genes in TADs for every sample [26]. GSVA is an unsupervised and non-parametric technique that eliminates the traditional strategy of directly modeling phenotypes within the enrichment score.

We defined TADs activity scores as enrichment scores of its gene-set so that we could better understand biological events. For TCGA (https://portal.gdc.cancer.gov/) [52] patient data, RNAseq gene expression profiles of 20 cancer types out of 24 cancer types (PANCANCER) were transformed into TAD activity scores. The data from the Cancer Cell Line Encyclopedia (CCLE)[5] with 819 samples from different cancer cell lines were used to calculate TAD activity scores. 819 CCLE-cancer cell lines were chosen based on their coherence with 544 drugs in the Cancer Therapeutics Response Portal (CTRPV2)[33]. In order to estimate change in TAD activity on application of drug we also used the single-cell expression profiles of HNSC cancer cell lines

published by Sharma et al. [27] to calculate TADs activity scores using UniPath R-package [25]. The gene enrichment scores thus obtained were analysed for differential activity among resistant and non-resistant cell lines using the Wilcoxon rank-sum test (Supplementary Table 1-4).

Calculation of survival P-value

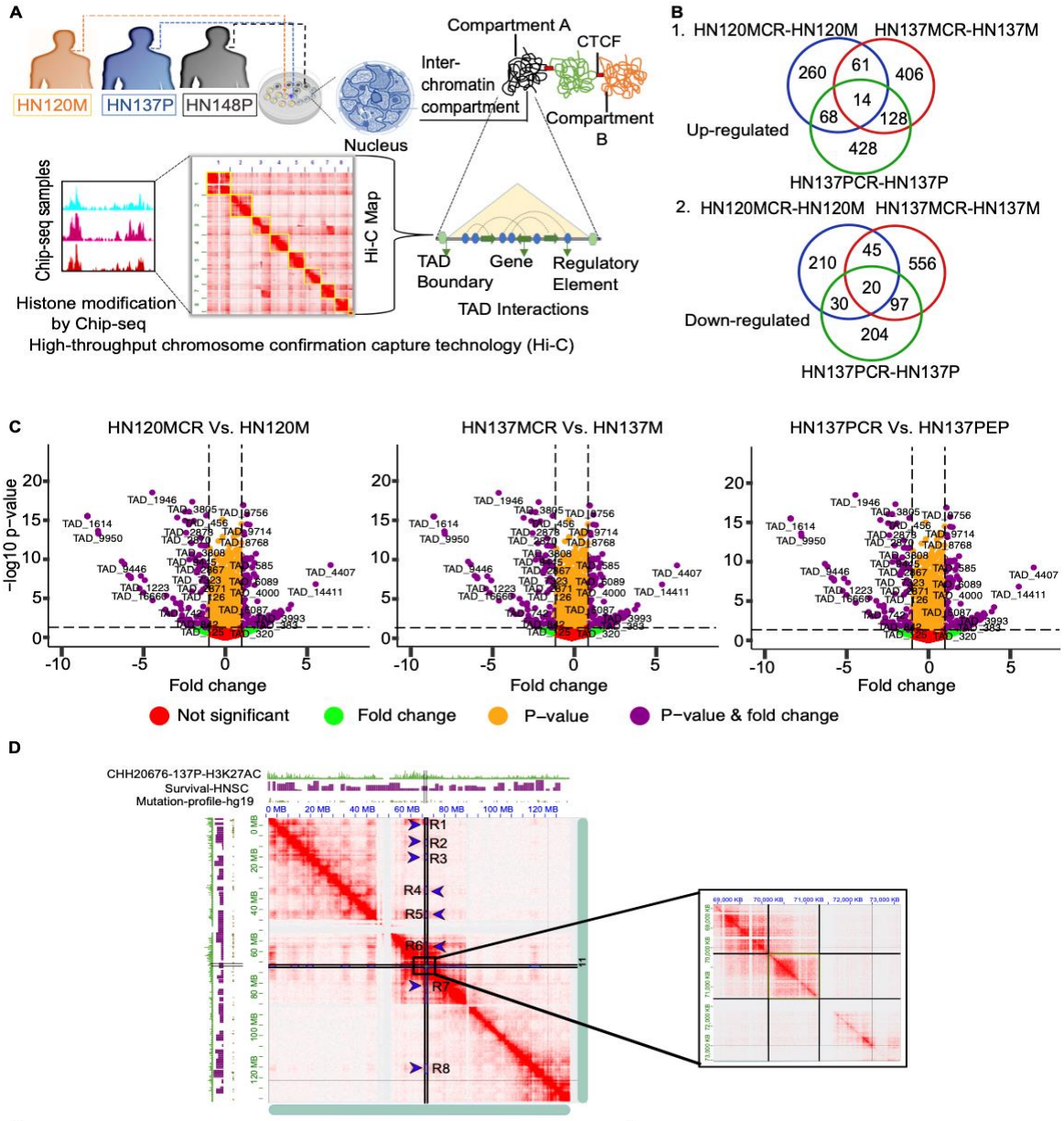
To calculate the significance of association of TAD activity and gene expression with survival for TCGA patients, R packages survival and survminer were used [53]. Input files were GSVA scores and clinical data containing days to death, gender, etc, as fetched from the Xena browser [54]. 'Status' column 0 or 1 in clinical data stands for alive and dead, respectively. For calculating the survival P-value for each TAD, the median of each row of GSVA was calculated, and another "median-group" column was created, signifying 1 for GSVA greater than the median and 0 for GSVA less than the median. The clinical information (days-to-death, status (0/1), and patient ID) was merged with GSVA scores and median-group column. Hence for each TAD, we created a file from which data was fit in the 'survfit' function from R package survival and survminer. Survfit output includes survival P-value for each TAD. In addition, we also calculated the hazard ratio for each TAD using the coxph function. Different cut-off combinations were implemented to identify TADs significance in survival (high-low), such as 51-49, 60-40, and 70-30 percentiles, to define the group of patients based on the activity of a TAD, in addition to the MaxStat package [55]. Subsequently, TADs which were common among such combinations, were selected for further analysis (Supplementary Figure S16).

To study the association of patient survival with drug resistance, we created two sets of TADs, i.e., Positive_set and Negative_set, with correlation above 0.3 and below -0.3, respectively. We examined their survival hazard ratio in multiple categories. Furthermore, we also explored the distribution of TADs across different cut-offs of hazard ratio.

Synteny analysis

We obtained the orthologous genes and the corresponding chromosomal coordinates of 69 mammals from Ensembl (https://asia.ensembl.org/index.html). We also downloaded the genome fasta files for the species that were not represented in Ensembl from DNA Zoo database (https://www.dnazoo.org/) and align the genome sequences against that of human genome using

last (https://gitlab.com/mcfrith/last). The alignment chain files were then used to map the orthologous positions of genes. We implemented ‘plotRect’ and ‘plotSegments’ function of 'plotgardener' R-package[56] to plot the synteny maps from the TSS coordinates of orthologous genes across mammals.



E

Region	Rank	P-value
R1	High sulphur keratin-associated protein	3.10e-06
R2	Positive regulation of cell growth	8.20e-05
R3	Central carbon metabolism in cancer	2.10e-08
R4	ETS	4.10e-02
R5	Protein binding	5.90e-02
R6	General substrate transporter	8.20e-02
R7	Folate receptor	2.70e-02
R8	Immunoglobulin domain	1.30e-04

F

Rank	Motif	Name	P-value
1	ACAGGAAGTG	ERG(ETS)	1e-3
2	TTAACCCTTTCATTC	ZNF652	1e-2
3	CGGTGACGTCAC	CRE(bZIP)	1e-2
4	CCGGGTCACGTGA	E-box(bHLH)	1e-2
5	ATTTCCCAGAATGCC	ZNF143|STAF(Zf)	1e-2

Figure 1: Analysis of chromatin interaction profiles and domain activity of head and neck cancer patient-derived cell lines. A) A description of the profiling and analysis of the chromatin

interaction profiles of three HNSC cell lines; metastatic cell line HN120M from patient HN120, primery tumor cells HN137P and HN148P from patients HN137 and HN148 respectively. The union list of TAD boundaries derived from Hi-C profiles of HNSC cell lines were used for downstream analysis. **B)** Venn diagrams showing the number of TADs commonly up-regulated or down-regulated in cisplatin-resistant versions of HNSC cell lines. **C)** The fold-change and significance (P-value) of the TAD activity between HNSC cell lines and their cisplatin-resistant versions are depicted on the volcano plots. **D)** The chromatin interaction visulation for chromosome 11 using Hi-C profile of 137P cells. The TADs (in the 11q13.3 band) downregulated in all three HNSC cell lines is the present region inside the rectangle indicated in the middle. Other regions with distal interaction with the TAD in 11q13.3-11q13.4 bands are shown with arrows as R1, R2, R3 .. R8). **E)** The enriched gene-ontology terms of genes present in different regions interacting with 11q13.3-11q13.4 bands. **F)** The enriched motifs in H3K27ac peaks in down-regulated TAD in 11q13.3-11q13.4 band and R1, R2 …. R8 region.

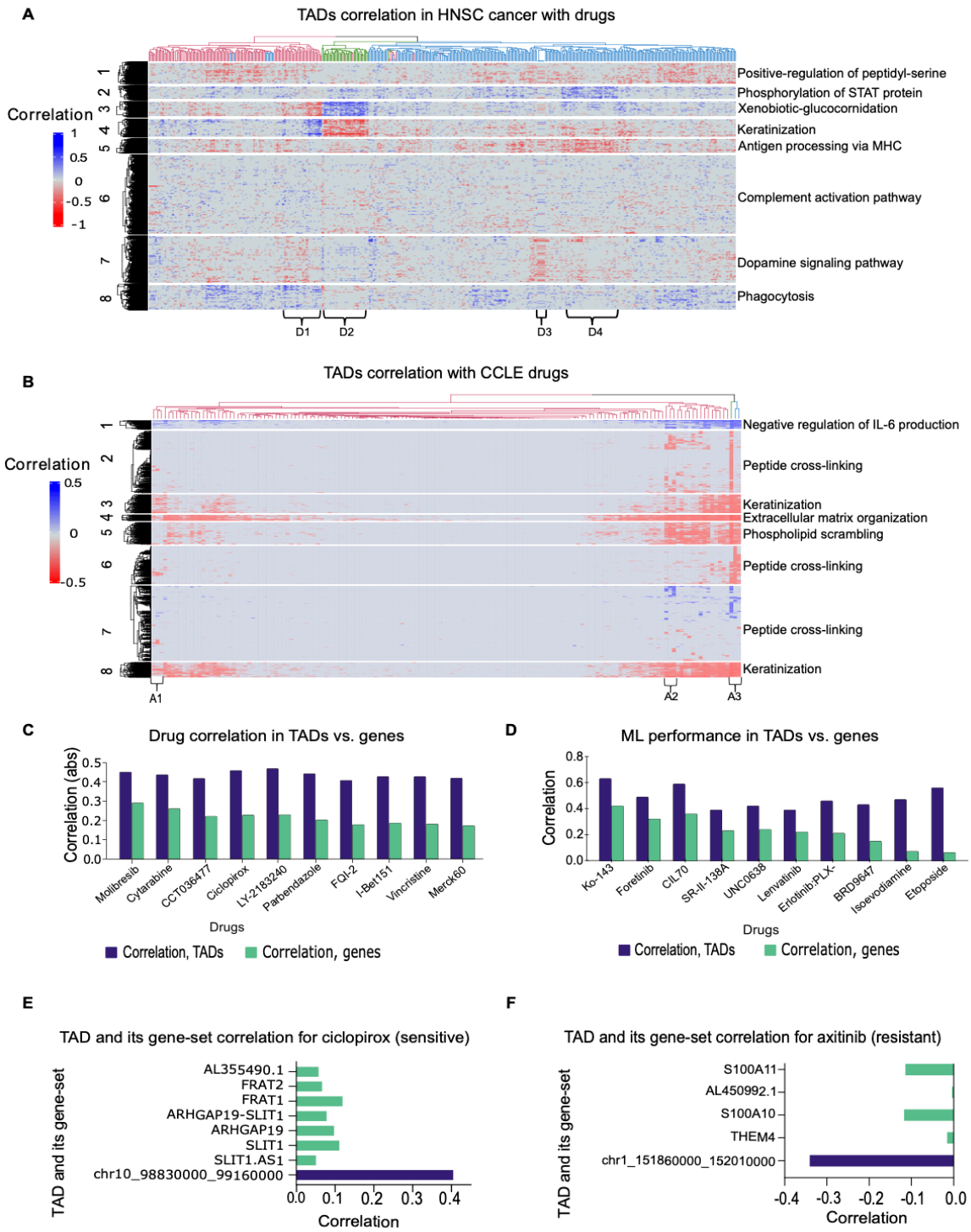


Figure 2: Pattern of correlation between TAD activity and drug response. A) Heatmap based illustration for correlation among TAD activity and PIC50 value for drugs in HNSC cell lines from CCLE database. Only those TADs (in rows) are shown whose activity has an absolute correlation

value of more than 0.3 with PIC50 value of at least one drug. Similar to this, only drugs that have an absolute correlation value of more than 0.3 with the activity of at least one TAD are displayed. Four sub-clusters of drugs are also highlighted here. **B)** Heatmap of correlation between TAD activity and PIC50 value of drugs for cell lines from all cancer types present in CCLE database. Only those TADs (rows) or drugs (columns) are shown which have atleast one absolute correlation value of more than 0.3. A few sub-clusters of drugs are also shown. **C)** Barplot of best positive correlation achieved for few drugs between their PIC50 values with TAD activity (blue) or single gene expression (green). **D)** The actual and predicted PIC50 values using a machine learning model are shown in a bar chart where the features were either TAD activity or gene expression. For many drugs, even when numerous genes are used, the prediction of PIC50 values (the effectiveness or accuracy) was noticeably lower than the value attained by TAD activity score. **E)** Correlation with PIC value for drug (Ciclopirox) with activity of a TAD (red) (chr10_98830000_99160000) across and its respective genes (yellow). The correlation values were calculated using PIC50 values of available cell lines with PIC50 value and transcriptome profile in CCLE database irrespective of cancer type. As shown, none of the genes individually have as high a correlation with PIC50 value of Ciclopirox as their host TAD activity. **F)** Comparison of correlation values of activity of a TAD (red) (chr1_151880000_152010000) and expression of genes in its respective gene-set (yellow) with PIC50 value of axitinib across available cell-lines of all cancer-types.

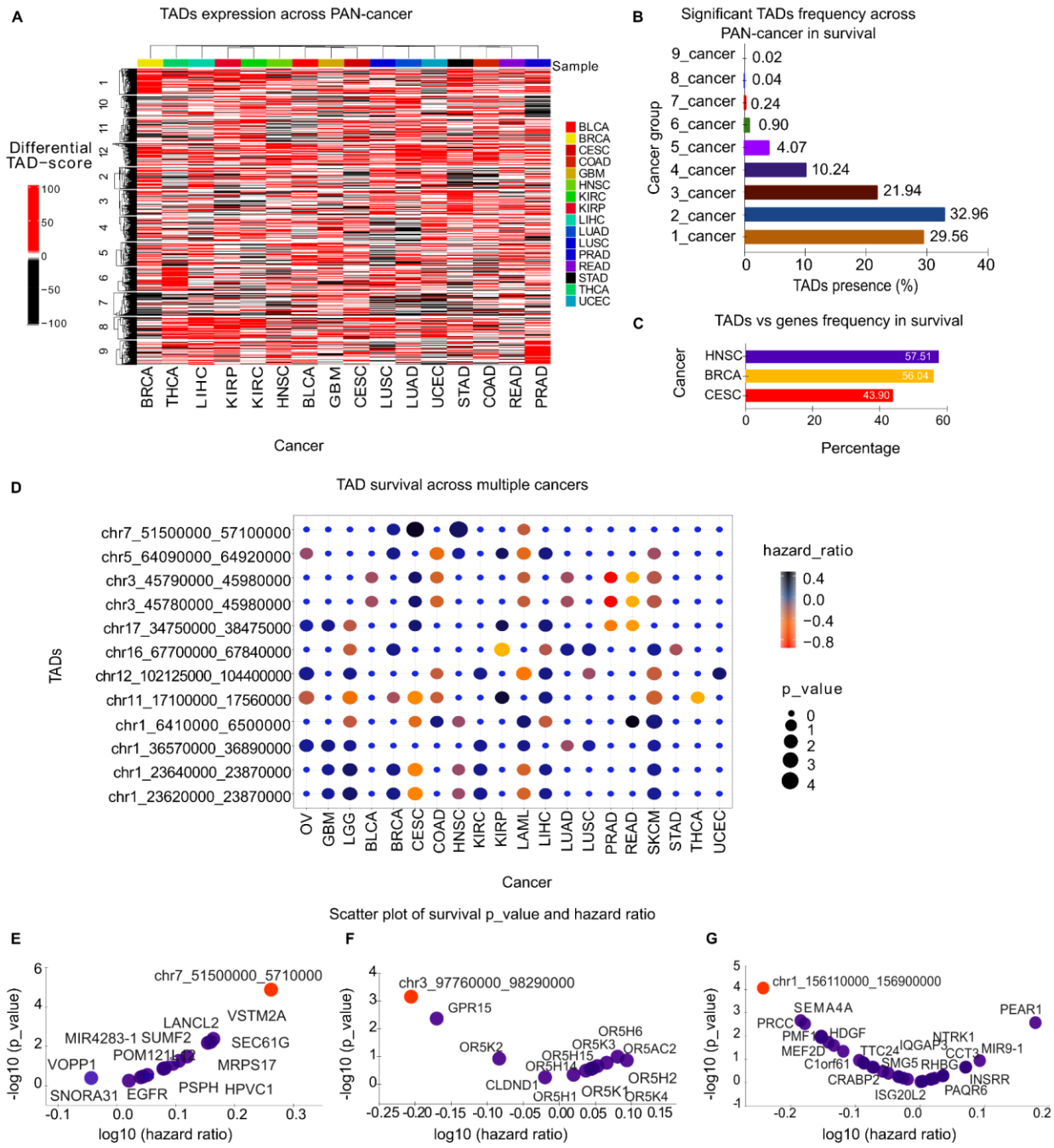


Figure 3: Activity of TADs across multiple cancer TCGA samples and their association with survival. A) Heatmap of fold changes of median TAD-activity (tumor Vs normal sample) across 16 cancer types, calculated using TCGA data. **B)** Here the proportion of numbers of TADs associated with survival in only one or 2 or 3 cancer type are shown. For association the P-value threshold was 0.05 and proportion was calculated by division by number of TADs which had association with survival in at least one cancer type. **C)** Percentage of TADs having more

significant association with survival than any of their genes. Percentage was calculated with respect to all TADs which had association with survival in the given cancer type. **D)** Based on their P-values and hazard ratios for survival, some TADs patterns across PAN-cancer are graphically shown by dot plots. **E)** TAD (chr7_51500000_57100000)-gene survival P-value comparisons are illustrated using scatter plots. **F)** TAD (chr3_97760000_98290000)-gene survival P-value comparisons are illustrated using scatter plots. **G)** TAD (chr1_156110000_156900000)-gene survival P-value comparisons are illustrated using scatter plots.

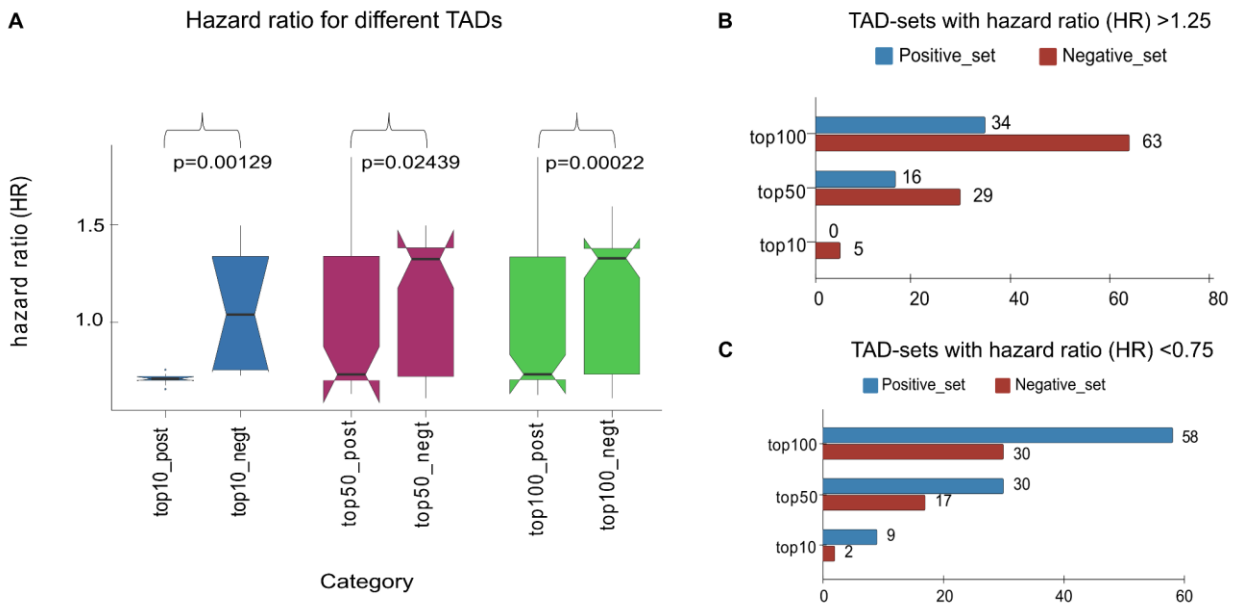


Figure 4: **Coherence between association of TAD activity with survival among TCGA patients and relationship between TAD and drug-response in CCLE cell lines. A)** hazard-ratio distribution of TADs correlated with drug-response, across the top 10, 50, and 100 categories. Here top10_post contain those TADs which are 10 most frequently positively correlated (R > 0.3) with PIC50 values of drugs. Similarly, top10_neg contain those TADs which have top 10 highest frequency of negative correlation with drug PIC50 value. Top50_post, top50_negt, top100_post and top100_negt groups are on the same trend. The hazard ratio for activity of TADs negatively correlated with PIC50 value (positive correlation with IC50) was above one (low survival). Similarly, the inverse is true for positively associated TADs. **B)** Frequency for TADs in different groups (top50_pos or top50_neg) for hazard ratio cutoffs greater than 1.25 (in TCGA cohort). **C)** Frequency of TADs in different groups, for hazard ratio less than 0.75.

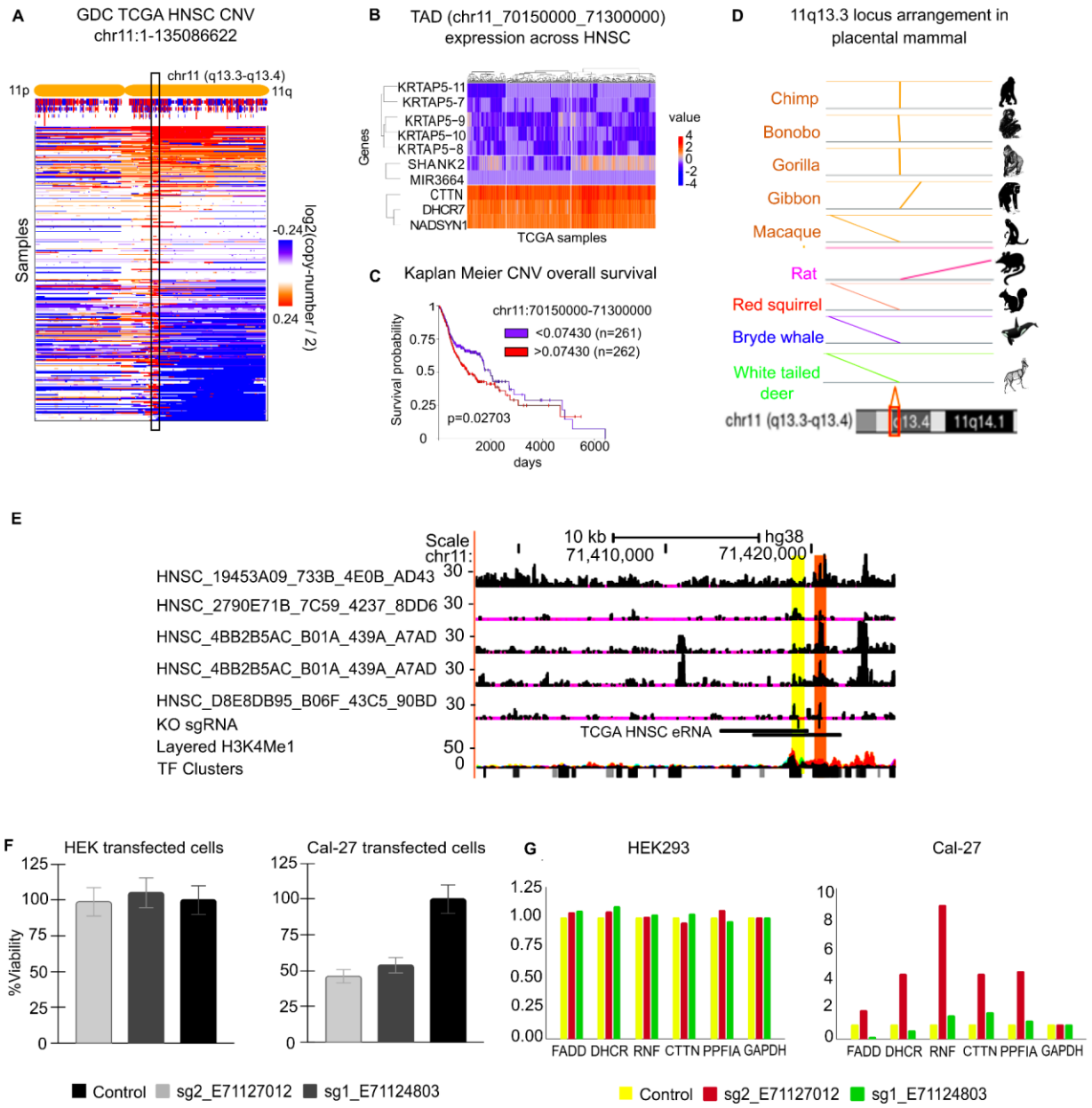


Figure 5: Case study of 11q13.3-4 TAD for identifying and targeting relevant regulatory sites in cancer. A) A view of copy number variations of genomic region in chromosome 11 in Tumor samples of head and Neck Cancer (HNSC) patients from TCGA. The box shows the regions in 11q13.3 - 11q13.4 bands, which also contains high number of CNV. **B)** Heatmap of expression of genes in TAD (hg19:chr11-70150000_71300000) in 11q13.3-11q13.4 band in TCGA-HNSC samples. C) Kaplan Meier (KM) plot showing association of copy number variation (CNV) in genomic location chr11-70150000_71300000, made using Xena browser D) A view of conserved

synteny of genes in TAD (hg19:chr11-70150000_71300000) of 11q13.4-11q13.4 band across 9 mammalian species. For synteny of across 69 mammals, see supplementary Figure S17. **E)** The UCSC-browser based view of open-chromatin profile (ATAC-seq) of 5 tumor samples of HNSC patients from TCGA consortium. **F)** Bar-plots showing viability of two cell lines (Cal-27, HEK293) after CRISPR-KO of inter-genic regions overlapping TCGA-HNSC-eRNA in a TAD in 11q13.3-11q13.4 band. **G)** rt-PCR based quantification of expression of genes after CRISPR based knock out of two sites overlapping TCGA-eRNA in TAD (hg19:chr11-70150000_71300000) of 11q13.3-11q13.4 band. The names of gRNAs shown here (sg1_E71124803 , sg2_E71127012) also contain their respective locations (hg19, chr11:71124803 and chr1171127012) corresponding to their 5-prime ends. The CRISPR-KO based results is shown for HNSC cell line (cal-27) and a control cell line (HEK293).

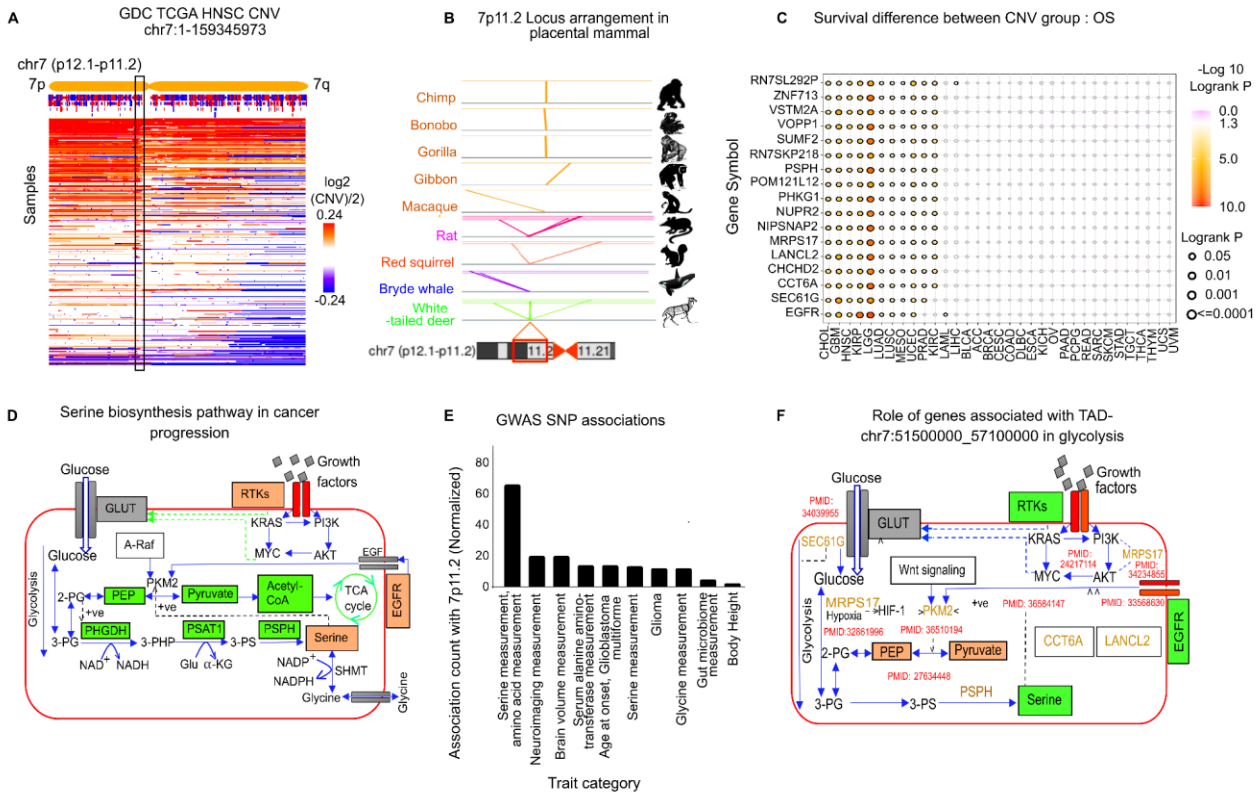


Figure 6: **Detail case study for topologically associated domain containing EFGR gene. A)** The visualisation of copy number variation in chromosome 7 in HNSC tumor samples in TCGA cohort. The TAD region chr7:51500000-57100000) containing EGFR gene (EFGR-TAD) is shown by a box. **B)** The brief view of conservation of synteny of genes in EFGR-TAD. Here the

state of synteny is shown only for nine mammals, taking human as reference. For visualising synteny in 69 mammal species, see (Supplementary Figure S17) **C)** The association of survival among TCGA-HNSC patients with copy number variation (CNV) of genes in the EGFR-TAD. **D)** An overview of major processes involved with intrinsic Serine biosynthesis as side product of glycolysis. **E)** Top enriched traits based on presence of GWAS (Genome wide association study) single nucleotide variants present EGFR-TAD. **F)** An overview of involvement of genes in EFGR-TAD in glycolysis and associated processes.

References

1. Li L, Barth NKH, Pilarsky C, Taher L. Cancer Is Associated with Alterations in the Three-Dimensional Organization of the Genome. Cancers . 2019;11. doi:10.3390/cancers11121886

2. Polak P, Karlić R, Koren A, Thurman R, Sandstrom R, Lawrence M, et al. Cell-of-origin chromatin organization shapes the mutational landscape of cancer. Nature. 2015;518: 360–364.

3. Jones PA, Issa J-PJ, Baylin S. Targeting the cancer epigenome for therapy. Nat Rev Genet. 2016;17: 630–641.

4. Pisco AO, Huang S. Non-genetic cancer cell plasticity and therapy-induced stemness in tumour relapse: “What does not kill me strengthens me.” Br J Cancer. 2015;112: 1725–1732.

5. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modeling of anticancer drug sensitivity. Nature. 2012;483: 603.

6. Kenyon J, Gerson SL. The role of DNA damage repair in aging of adult stem cells. Nucleic Acids Res. 2007;35: 7557–7565.

7. DeNicola GM, Karreth FA, Humpton TJ, Gopinathan A, Wei C, Frese K, et al. Oncogene-induced Nrf2 transcription promotes ROS detoxification and tumorigenesis. Nature. 2011;475: 106–109.

8. Nakasone ES, Askautrud HA, Kees T, Park J-H, Plaks V, Ewald AJ, et al. Imaging tumor-stroma interactions during chemotherapy reveals contributions of the microenvironment to resistance. Cancer Cell. 2012;21: 488–503.

9. Rosenzweig SA. Acquired resistance to drugs targeting receptor tyrosine kinases. Biochem

Pharmacol. 2012;83: 1041–1048.

10. Challen GA, Little MH. A side order of stem cells: the SP phenotype. Stem Cells. 2006;24: 3–12.

11. Zhou S, Schuetz JD, Bunting KD, Colapietro AM, Sampath J, Morris JJ, et al. The ABC transporter Bcrp1/ABCG2 is expressed in a wide variety of stem cells and is a molecular determinant of the side-population phenotype. Nat Med. 2001;7: 1028–1034.

12. Dean M, Fojo T, Bates S. Tumour stem cells and drug resistance. Nat Rev Cancer. 2005;5: 275–284.

13. Donnenberg VS, Donnenberg AD. Multiple drug resistance in cancer revisited: the cancer stem cell hypothesis. J Clin Pharmacol. 2005;45: 872–877.

14. Medema JP. Cancer stem cells: the challenges ahead. Nat Cell Biol. 2013;15: 338–344.

15. Morgan MA, Shilatifard A. Chromatin signatures of cancer. Genes Dev. 2015;29: 238–249.

16. Kleppe A, Albregtsen F, Vlatkovic L, Pradhan M, Nielsen B, Hveem TS, et al. Chromatin organisation and cancer prognosis: a pan-cancer study. Lancet Oncol. 2018;19: 356–369.

17. Tena JJ, Santos-Pereira JM. Topologically Associating Domains and Regulatory Landscapes in Development, Evolution and Disease. Front Cell Dev Biol. 2021;9: 702787.

18. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012;485: 376–380.

19. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014;159: 1665–1680.

20. Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, et al. Chromatin architecture reorganization during stem cell differentiation. Nature. 2015;518: 331–336.

21. Ibn-Salem J, Muro EM, Andrade-Navarro MA. Co-regulation of paralog genes in the three-dimensional chromatin architecture. Nucleic Acids Res. 2017;45: 81–91.

22. Akdemir KC, Le VT, Kim JM, Killcoyne S, King DA, Lin Y-P, et al. Somatic mutation distributions in cancer genomes vary with three-dimensional chromatin structure. Nat Genet. 2020;52: 1178–1188.

23. Gong Y, Lazaris C, Sakellaropoulos T, Lozano A, Kambadur P, Ntziachristos P, et al. Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. Nat Commun. 2018;9: 542.

24. Chia S, Low JL, Zhang X, Kwang XL, Chong FT, Sharma A, et al. Phenotype-driven

precision oncology as a guide for clinical decisions one patient at a time. Nat Commun. 2017;8. doi:10.1038/s41467-017-00451-5

25. Chawla S, Samydurai S, Kong SL, Wu Z, Wang Z, Tam WL, et al. UniPath: a uniform approach for pathway and gene-set based analysis of heterogeneity in single-cell epigenome and transcriptome profiles. Nucleic Acids Res. 2021;49: 1801.

26. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14: 7.

27. Sharma A, Cao EY, Kumar V, Zhang X, Leong HS, Wong AML, et al. Longitudinal single-cell RNA sequencing of patient-derived primary cells reveals drug-induced infidelity in stem cell hierarchy. Nat Commun. 2018;9: 4931.

28. Ramazzotti D, Angaroni F, Maspero D, Ascolani G, Castiglioni I, Piazza R, et al. Variant calling from scRNA-seq data allows the assessment of cellular identity in patient-derived cell lines. Nature communications. 2022. p. 2718.

29. Saba NF, Wang X, Müller S, Tighiouart M, Cho K, Nie S, et al. Examining expression of folate receptor in squamous cell carcinoma of the head and neck as a target for a novel nanotherapeutic drug. Head Neck. 2009;31: 475.

30. Martens JHA, Mandoli A, Simmer F, Wierenga B-J, Saeed S, Singh AA, et al. ERG and FLI1 binding sites demarcate targets for aberrant epigenetic regulation by AML1-ETO in acute myeloid leukemia. Blood. 2012;120: 4038–4048.

31. Fuchs O. Importance and Presentation of Transcription Factor Fli-1 in Hematopoiesis and in Hematological and Other Malignancies. ARC Journal of Hematology. 2017;2: 23–37.

32. Liu T, Porter J, Zhao C, Zhu H, Wang N, Sun Z, et al. TADKB: Family classification and a knowledge base of topologically associating domains. BMC Genomics. 2019;20: 1–17.

33. Rees MG, Seashore-Ludlow B, Cheah JH, Adams DJ, Price EV, Gill S, et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. Nat Chem Biol. 2016;12. doi:10.1038/nchembio.1986

34. Ghandadi M, Sahebkar A. Interleukin-6: A Critical Cytokine in Cancer Multidrug Resistance. Curr Pharm Des. 2016;22. doi:10.2174/1381612822666151124234417

35. Allain EP, Rouleau M, Lévesque E, Guillemette C. Emerging roles for UDP-glucuronosyltransferases in drug resistance and cancer progression. Br J Cancer. 2020;122: 1277–1287.

36. DeBerardinis RJ, Chandel NS. Fundamentals of cancer metabolism. Science Advances. 2016 [cited 29 Oct 2023]. doi:10.1126/sciadv.1600200

37. Zhang R, Shu H, Wang Y, Tao T, Tu J, Wang C, et al. G-Quadruplex Structures Are Key Modulators of Somatic Structural Variants in Cancers. Cancer Res. 2023;83: 1234.

38. Ni L, Liu Y, Ma X, Liu T, Yang X, Wang Z, et al. Pan-3D genome analysis reveals structural and functional differentiation of soybean genomes. Genome Biol. 2023;24. doi:10.1186/s13059-023-02854-8

39. Barutcu AR, Maass PG, Lewandowski JP, Weiner CL, Rinn JL. A TAD boundary is preserved upon deletion of the CTCF-rich Firre locus. Nat Commun. 2018;9: 1444.

40. Berens EB, Sharif GM, Schmidt MO, Yan G, Shuptrine CW, Weiner LM, et al. Keratin-associated protein 5-5 controls cytoskeletal function and cancer cell vascular invasion. Oncogene. 2017;36: 593–605.

41. Wu L, Jin Y, Zhao X, Tang K, Zhao Y, Tong L, et al. Tumor aerobic glycolysis confers immune evasion through modulating sensitivity to T cell-mediated bystander killing via TNF-α. Cell Metab. 2023;35. doi:10.1016/j.cmet.2023.07.001

42. Chang H, Xu Q, Li J, Li M, Zhang Z, Ma H, et al. Lactate secreted by PKM2 upregulation promotes Galectin-9-mediated immunosuppression via inhibiting NF-κB pathway in HNSCC. Cell Death Dis. 2021;12: 1–13.

43. Lee DP, Tan WLW, Anene-Nzelu CG, Lee CJM, Li PY, Luu TDA, et al. Robust CTCF-Based Chromatin Architecture Underpins Epigenetic Changes in the Heart Failure Stress-Gene Response. Circulation. 2019;139: 1937–1956.

44. Forcato M, Nicoletti C, Pal K, Livi CM, Ferrari F, Bicciato S. Comparison of computational methods for Hi-C data analysis. Nat Methods. 2017;14: 679–685.

45. Wingett S, Ewels P, Furlan-Magaril M, Nagano T, Schoenfelder S, Fraser P, et al. HiCUP: pipeline for mapping and processing Hi-C data. F1000Res. 2015;4: 1310.

46. Aljogol D, Thompson IR, Osborne CS, Mifsud B. Comparison of Capture Hi-C Analytical Pipelines. Front Genet. 2022;0. doi:10.3389/fgene.2022.786501

47. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10. doi:10.1093/gigascience/giab008

48. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell systems. 2016;3: 95.

49. Dali R, Blanchette M. A critical assessment of topologically associating domain prediction tools. Nucleic Acids Res. 2017;45: 2994–3005.

50. Kumar V, Muratani M, Rayan NA, Kraus P, Lufkin T, Ng HH, et al. Uniform, optimal signal processing of mapped deep-sequencing data. Nat Biotechnol. 2013;31: 615–622.

51. Hung J-H, Yang T-H, Hu Z, Weng Z, DeLisi C. Gene set enrichment analysis: performance evaluation and usage guidelines. Brief Bioinform. 2012;13: 281–291.

52. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a Shared Vision for Cancer Genomic Data. N Engl J Med. 2016;375: 1109–1112.

53. Tm T, Grambsch PM. Modeling survival data: extending the Cox model. New York Springer; 2000.

54. Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. Nat Biotechnol. 2020;38: 675–678.

55. Lausen B, Schumacher M. Maximally Selected Rank Statistics. Biometrics. 1992. p. 73. doi:10.2307/2532740

56. Kramer NE, Davis ES, Wenger CD, Deoudes EM, Parker SM, Love MI, et al. Plotgardener: cultivating precise multi-panel figures in R. Bioinformatics. 2022;38: 2042–2045.