NEOPLASIA
www.neoplasia.com

Original Research

TCF7L2 silencing results in altered gene expression patterns accompanied by local genomic reorganization

Markus A. Brown [a,b,1]**; Gabrielle A. Dotson** [c,1]**;**
Scott Ronquist [c]**; Georg Emons** [a]**; Indika Rajapakse** [c,*]**;**
Thomas Ried [a,*]

[a] Genetics Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA
[b] Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD, USA
[c] Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI, USA

Abstract

Canonical Wnt signaling is crucial for intestinal homeostasis as TCF4, the major Wnt signaling effector in the intestines, is required for stem cell maintenance. The capability of TCF4 to maintain the stem cell phenotype is contingent upon β-catenin, a potent transcriptional activator, which interacts with histone acetyltransferases and chromatin remodeling complexes. We used RNAi to explore the influence of TCF4 on chromatin structure (Hi-C) and gene expression (RNA sequencing) across a 72-hour time series in colon cancer. We found that TCF4 reduction results in a disproportionate up-regulation of gene expression, including a powerful induction of *SOX2*. Integration of RNA sequencing and Hi-C data revealed a TAD boundary loss, which occurred concomitantly with the over-expression of a cluster of *CEACAM* genes on chromosome 19. We identified EMT and E2F as the 2 most deregulated pathways upon TCF4 depletion and *LUM, TMPO*, and *AURKA* as highly influential genes in these networks using measures of centrality. Results from gene expression, chromatin structure, and centrality analyses were integrated to generate a list of candidate transcription factors crucial for colon cancer cell homeostasis. The top ranked factor was c-JUN, an oncoprotein known to interact with TCF4 and β-catenin, confirming the usefulness of this approach.

Neoplasia (2021) 23, 257–269

Introduction

The Wnt signaling transcription factor, TCF4, is crucial for homeostasis of the mammalian intestine [1]. Loss of TCF4, in either embryonic or adult mice, results in ablation of the proliferative compartment of the intestine [2]. The capability of TCF4 to drive a crypt progenitor phenotype and maintain intestinal homeostasis is contingent upon its binding partner. When Wnt signaling is active, cytoplasmic β-catenin migrates to the nucleus where it binds TCF4, resulting in target gene expression. β-catenin is a potent transcriptional activator which influences the surrounding chromatin by recruiting histone acetyltransferases, chromatin remodeling factors, and RNA polymerase associated factors [3–6]. Therefore, the binding of a βcat/TCF4 complex significantly enhances the recruitment of the cellular machinery necessary for transcription, thereby driving a crypt progenitor phenotype [9].

Conversely, when TCF4 is bound by the TLE family of transcriptional repressors, target gene expression is repressed. Despite the presence of TCF4 throughout the intestine, active Wnt signaling is sequestered to the base of the colonic crypts, thereby limiting βcat/TCF4 complex formation, and the resulting crypt progenitor phenotype, to cells of the crypt base [7–9].

In colorectal cancer, mutations in the Wnt signaling pathway, primarily in *APC*, result in constitutive Wnt signaling activity [10,11]. High levels of nuclear β-catenin result in the constitutive expression of Wnt target genes, such as *MYC* and *CCND1* [12–15]. Given that epigenetic modifications and chromatin remodeling have been reported to occur prior to the expression

of Wnt target genes, the high levels of nuclear β-catenin in colorectal cancer likely influence the surrounding chromatin structure [5,16].

It has become evident that understanding chromatin structure lends insight into the regulation of gene expression [17,18]. Aspects of chromatin structure include the accessibility of genomic loci to the transcriptional machinery as well as the 3-dimensional configuration of the chromatin, which may facilitate cooperativity between genomic regions sharing a particular 3-dimensional space. Division of the genome into euchromatin and heterochromatin domains, referred to as A and B compartments, respectively, reflects the interaction between chromatin structure and function at the chromosomal level [17]. Chromosome folding brings distant sites along the linear genome in close spatial proximity, forming topologically associating domains (TADs), which are insulated regions of the genome sharing epigenetic modifications, gene expression, and replication timing [18–20]. Given the influence of chromatin structure on gene expression, the dissection of a cellular response to a stimulus requires an understanding of both structural and functional dynamics [21,22].

Herein, we explored the dynamical influence of silencing *TCF7L2*, the gene encoding TCF4, on chromatin structure and gene expression across a 72-hour time series in the colon cancer cell line, SW480. Silencing of *TCF7L2* not only allows us to investigate the changes which occur as a result of silencing a major transcription factor, it also represents a clinically relevant model as *TCF7L2* expression correlates with resistance to chemoradiotherapy (CRT), a treatment modality in rectal cancer [23–25]. For the time series, an inverse RNAi transfection protocol was designed and optimized, which significantly reduced confounding factors typically found in time series data. A 4DN approach to data analysis was used, which integrated structural (Hi-C) and functional (RNA sequencing) data and allowed us to identify influential genes in the most dynamically changing networks (Fig. 1A). We then identified candidate reprogramming factors, to be perturbed alongside, or *in lieu of, TCF7L2* to debilitate the colorectal cancer cell.

Materials and methods

Cell culture

The SW480, DLD1, COLO201, HCT116, and LS174T colon cancer cell lines were obtained from the American Type Culture Collection. SW480, DLD1, and COLO201 cells were grown in RPMI-1640 growth medium (Thermo Fisher; 11875093), HCT116 cells were grown in McCoy's 5A (Modified) medium (Thermo Fisher; 16600082), and LS174T cells were grown in Minimum Essential medium (Thermo Fisher; 11095080). All growth media were supplemented to 10% FBS (Thermo Fisher; 10082147) and 2mM L-glutamine (Thermo Fisher; 25030149). Cells were incubated at 37°C in 5% CO_2. No antibiotics were used. Proper cell line identity was confirmed using STR profiling and the absence of mycoplasma contamination was confirmed routinely using PCR (Genlantis; MY01050).

siRNA inverted transfections

Cells were seeded in a 6-well tissue culture plate (Corning; 353046) at a density of 150,000 cells per well. The wells contained 2mL of RPMI-1640 growth medium. Silencer Select siRNAs (Thermo Fisher), Negative Control No.1 (4390843) and siTCF7L2-s13880 (4392420), were delivered to the cells using the lipofectamine RNAiMAX reagent (Thermo Fisher; 13778150) according to the manufacturer's instructions. This resulted in the addition of 7.5μL of RNAiMAX and 25pmol of siRNA per well of the 6-well plate. The siRNA:lipofectamine complexes were not removed during the time course. The time series was constructed thus: 0- and 72-hour time points were transfected with siNEG and siTCF7L2, respectively, 24 hours after seeding. The 48-hour time point was transfected, with siTCF7L2, 24 hours after the 0- and 72-hour time points. The 24-hour time point was transfected, with

siTCF7L2, 24 hours after the 48-hour time point. The cells were harvested 24 hours later, resulting in exposure to siTCF7L2 for 72, 48, and 24 hours. The 0-hour time point spent 0 hours in siTCF7L2 and serves as the transfection control. This procedure is illustrated in Supplementary Fig. 1.

Quantitative PCR

RNA was extracted from at least 3 independent inverted transfections using the RNeasy Mini Kit and on-column DNase treatment (Qiagen; 74104, 79254), according to the manufacturer's instructions. RNA concentration was determined using a NanoDrop 1000 spectrophotometer (Thermo Fisher). Synthesis of cDNA was performed using the Verso cDNA Synthesis kit (Thermo Fisher; AB1453B) based on the manufacturer's instructions. The optional RT Enhancer was used and equal amounts of both the Anchored oligo dT and Random Hexamers were added. Quantitative PCR was performed for *TCF7L2, SOX2, CEACAM1*, and *YWHAZ* on an ABI PRISM 7000 sequence detection system (Applied Biosystems) using *Power* SYBR Green PCR Master Mix (Thermo Fisher; 4367659), according to the manufacturer's instructions. Primer sequences can be found in Supplementary Table 7. The thermocycling protocol consisted of a 10-minute hold at 95°C followed by 40 two-step cycles consisting of 15 seconds at 95°C and 1 minute at 60°C. Fold change was determined using the *YWHAZ* reference gene and the $2^{-\Delta\Delta CT}$ method.

Western blot

Protein was extracted using a 1% NP-40 Lysis Buffer solution containing a protease inhibitor cocktail (Millipore; 11836153001). To extract protein, 350μL of the 1% NP-40 solution was added to each well of a 6-well plate and cells were scraped into the buffer using a cell scraper and transferred to a 1.5 mL Eppendorf tube. The mixture was incubated on ice for 30 minutes and spun at 14,000 rpm for 30 minutes at 4°C. The supernatant was transferred to a new tube. Protein was prepared for quantification using the Pierce BCA Protein Assay Kit (Thermo Fisher; 23225), according to the manufacturer's instructions, and quantified on a SpectraMAX M2e microplate reader (Molecular Devices). Nuclear protein samples (20 μg) were electrophoresed using a 4% to 12% Bis-Tris gel (Thermo Fisher; NP0322BOX) in a XCell SureLock Mini-Cell (Thermo Fisher; EI0002) with MOPS Running Buffer (Thermo Fisher; NP0001). Proteins were transferred into a PVDF membrane (Millipore; IPVH00010) using Transfer Buffer (Thermo Fisher; NP00061) and an XCell II Blot Module. Membranes were incubated with antibodies targeting TCF4 (Origene; TA333645), active β-catenin (Millipore; 05-665), or β-actin (CST; 4970S). An HRP-linked secondary antibody (CST; 7074S, 7076S) was applied and detected using Pierce ECL Western Blotting Substrate (Thermo Scientific; 32209). Blots were imaged in an Azure c600 Gel Imaging Station (Azure Biosystems).

RNA sequencing and data processing

RNA was extracted from 3 independent inverted transfections using the RNeasy Mini Kit and on-column DNAse treatment (Qiagen; 74104, 79254), according to the manufacturer's instructions. RNA concentration was determined using a NanoDrop 1000 spectrophotometer (Thermo Fisher). RNA integrity was determined using the 4200 TapeStation (Agilent). RNA samples with an RNA integrity number (RIN) of 9 or higher were used for library preparation. Libraries were prepared by the CCR Sequencing Facility using the TruSeq Stranded Total RNA Library Prep Gold kit (Illumina; 20020598). Ribosomal RNA (rRNA) was removed using biotinylated, target-specific oligos combined with Ribo-Zero rRNA removal beads. The RNA was then fragmented and the cleaved RNA fragments were copied into first strand cDNA using reverse transcriptase and random primers, followed by second strand cDNA synthesis using DNA Polymerase I and RNase H. The resulting

double-strand cDNA was used as input for Illumina library preparation with end-repair, adapter ligation, and high-fidelity PCR. The final purified product was quantified by qPCR. Samples were sequenced on a HiSeq2500 using Illumina TruSeq v4 chemistry generating 125 base pair, paired-end reads. RNA sequencing data was processed using the CCBR Pipeliner Version 3.0. Briefly, FASTQC was used to assess sequencing quality and Cutadapt was used to remove adapter sequences and perform quality trimming, respectively [26,27]. Kraken, KronaTools, and FastQ Screen were used to assess microbial contamination [28,29]. Our samples were confirmed to be free of microbial contamination. STAR (2-pass) was used to align reads to the hg19 reference genome [30]. Picard, Preseq, SAMtools, and RSeQC were used to assess alignment quality [31–33]. Gene expression was quantified using RSEM [34]. Transcripts per million were computed to normalize gene expression values. Differential gene expression was performed using limma [35,36]. Volcano plots were generated using the EnhancedVolcano package [37].

Gene set enrichment analysis

Gene Set Enrichment Analysis was performed using R and the fgsea package, available through Bioconductor [38,39]. Differentially expressed genes were pre-ranked according to their $\log_2$FC values. Gene set enrichment was calculated using the fgsea command and the Hallmark gene sets available from MSigDB (https://www.gsea-msigdb.org/gsea/msigdb/collections.jsp). The minSize and maxSize parameters were set to 15 and 500, respectively. The heatmap was generated using the ComplexHeatmap package [40].

Hi-C

The *in situ* Hi-C protocols from Rao et al [19] were adapted with slight modifications. For each Hi-C library, approximately 3×10^6 cells were incubated in 250µl of ice-cold Hi-C lysis buffer (10mM Tris–HCl pH 8.0, 10mM NaCl, 0.2% Igepal CA630) with 50µl of a protease inhibitor cocktail on ice for 30 minutes and washed with 250 µL lysis buffer. The nuclei were pelleted by centrifugation at 2500 g for 5 minutes at 4°C, re-suspended in 50 µL of 0.5% sodium dodecyl sulfate and incubated at 62°C for 10 minutes. Afterwards, 145 µL of water and 25 µL of 10% Triton X-100 were added and incubated at 37°C for 15 minutes.

Chromatin was digested with 200 units of MboI (NEB) overnight at 37°C with rotation. Chromatin end overhangs were filled in and marked with biotin-14-dATP (Thermo Fisher Scientific) by adding the following components to the reaction: 37.5 µL of 0.4 mM biotin-14-dATP (Life Technologies), 1.5 µL of 10 mM dCTP, 1.5 µL of 10 mM dGTP, 1.5 µL of 10 mM dTTP, and 8 µL of 5 U/µL DNA Polymerase I, Large (Klenow) Fragment (NEB). The marked chromatin ends were ligated by adding 900µl of ligation master mix consisting of 663 µL of water, 120 µL of 10X NEB T4 DNA ligase buffer (NEB), 100 µL of 10% Triton X-100, 12 µL of 10 mg/mL BSA, 5 µL of 400U/µL T4 DNA Ligase (NEB), and incubated at room temperature for 4 hours.

Chromatin reverse crosslinking was performed by adding 50 µL of 20 mg/mL proteinase K (NEB) and 120 µL of 10% sodium dodecyl sulfate and incubated at 55°C for 30 minutes, adding 130 µL of 5M sodium chloride and incubated at 68°C overnight. DNA was precipitated with ethanol, washed with 70% ethanol, and dissolved in 105 µL of 10 mM Tris–HCl, pH 8. DNA was sheared on a Covaris S2 sonicator. Biotinylated DNA fragments were pulled with the MyOne Streptavidin C1 beads (Life Technologies). To repair the ends of sheared DNA and remove biotin from unligated ends, DNA-bound beads were resuspended in 100 µL of mix containing 82 µL of 1X NEB T4 DNA ligase buffer with 10 mM ATP (NEB), 10 µL of 10 (2.5 mM each) 25 mM dNTP mix, 5 µL of 10U/µL NEB T4 PNK (NEB), 4 µL of 3U/µl NEB T4 DNA polymerase (NEB), and 1 µL of 5U/µL NEB DNA polymerase I, Large (Klenow) Fragment (NEB).

After end-repair, dATP attachment was carried out in 100 µL of mix consisting of 90µl of 1X NEBuffer 2, 5 µL of 10 mM dATP, 5 µL of 5U/µL NEB Klenow exo minus (NEB), and incubated at 37°C for 30 minutes. The beads were then cleaned for Illumina sequencing adaptor ligation which was done in a mix containing 50 µL of 1X T4 ligase buffer, 3 µL T4 DNA ligases (NEB), and 2 µL of a 15 µM Illumina indexed adapter at room temperature for 1 hour. DNA was dissociated from the bead by heating at 98°C for 10 minutes, separated on a magnet, and transferred to a clean tube.

Final amplification of the library was carried out in mul- tiple PCR reactions using Illumina PCR primers. The reactions were performed on a 25 µL scale consisting of 25 ng of DNA, 2 µL of 2.5 mM dNTPs, 0.35 µL of 10 µM each primer, 2.5 µL of 10X PfuUltra buffer, and 0.5 µL of PfuUltra II Fusion DNA polymerase (Agilent). The PCR cycle conditions were set to 98°C for 30 seconds as the denaturing step, followed by 16 cycles of 98°C 10 seconds, 65°C for 30 seconds, 72°C for 30 seconds, then with an extension step at 72°C for 7 minutes.

After PCR amplification, the products from the same library were pooled and fragments ranging in sized of 300–500 bp were selected with AMPure XP beads (Beckman Coulter). The sized selected libraries were sequenced to produce paired-end reads on the Illumina HiSeq 2500 platform with the V4 for 125 cycles. Quality control metrics of the Hi-C data is given in Supplementary Table 1.

Hi-C matrix generation

Paired end reads were processed using the juicer pipeline with default parameters [41]. Reads were mapped to reference genome hg19, with "-s" (site parameter) MboI. Reads with MAPQ >30 were kept for further analysis. Data was extracted and input to MATLAB using Juicebox tools command "dump". Knight-Ruiz (KR) normalization was applied to all matrices, observed over expected (O/E) matrices were used for A/B compartmentalization and identification of topologically associating domains (TADs). Rows and columns for which more than 10% of entries had zeros were removed from the matrix.

A/B compartmentalization

Hi-C data was partitioned into A/B compartments according to the sign of the chromatin accessibility metric termed the Fiedler vector (the eigenvector corresponding to the second smallest eigenvalue of the normalized Laplacian matrix). Positive vector values were assigned to the "A" compartment representing euchromatic loci and negative values were assigned to the "B" compartment representing heterochromatic loci. The magnitude of the Fiedler values indicates the level of openness or closedness of the chromatin at a given loci. Hi-C data was at 100 kb resolution.

TAD calling

TADs were designated using spectral identification as described in [42]. Briefly, the Fiedler vector is calculated for a normalized Hi-C adjacency matrix and initial TADs are organized according to neighboring regions whose Fiedler values have the same sign. The initial TAD structure is repartitioned if, for a given domain, the Fiedler number falls below a user-specified threshold (λthr). This ensures that TADs are not overly large and repartitions the domains until the Fiedler number is larger than the threshold or until the TAD reaches the smallest allowable TAD size (default is 3). This process is performed iteratively for a set of chromosome-specific Hi-C contact maps. For our analysis, λthr was chosen to ensure a median TAD size of 900kb, as the expected median TAD size in mammalian genomes is 880kb (rounded to 900 kb since bins are in intervals of 100kb) [43]. λthr was chosen individually for each chromosome to ensure each TAD clustering set would have the same approximate median TAD size. Hi-C data was at 100kb resolution.

Transcription factor enrichment analysis

Transcription factor enrichment analysis was performed using oPOSSUM-3 software [44]. oPOSSUM-3 tests for over- representation of transcription factor binding sites (TFBS) in the DNA sequences of a set of genes, using Z-score and Fisher exact probability to determine significance. For our analysis, 64 genes from a differentially conformed region (*CEACAM* region) of chromosome 19 were submitted as query and compared against a background set of genes in the oPOSSUM-3 database. Both sets of genes were compared to the JASPAR CORE database of preferential TFBS, representing 719 vertebrate TFBS, and the rate of TFBS occurrence for each set of genes was measured [45]. The following analysis parameters were used: 8-bit minimum profile specificity, 0.40 conservation cutoff, 85% matrix score threshold, and a 5kb flanking length.

ChIP sequencing analysis

To explore key TF binding patterns throughout the genome and specifically at TAD boundaries, we obtained publicly available SP1, CTCF, and TCF4 ChIP-seq data for the HCT116 colorectal cancer cell line from the Gene Expression Omnibus using the accession numbers GSM1010902 (SP1), GSM1010903 (CTCF), and GSM782123 (TCF4). Two replicates were available for SP1 and CTCF, hence replicate profiles were averaged together. Profiles were then binned at 100kb resolution to be compatible with our RNA-seq and Hi-C data. The percentage of nonzero CTCF and TCF4 peak intensities that coincided with TAD boundaries called on the Hi-C data are reported in Supplementary Tables 5 and 6. Additionally, because TFEA analysis revealed that SP1 is the most highly enriched TF near the *CEACAM* loci on chromosome 19, we evaluated SP1 peak intensities at these loci. To evaluate the influence of TCF4 binding on the *CEACAM* loci, we aligned the TCF4 ChIP-seq profile to the region as well. Because ChIP-seq data for SP1 and TCF4 were acquired from different experiments, direct com- parison of their binding behavior at the *CEACAM* loci was facilitated by performing min-max normalization on each profile.

Synthetic 5C map

We constructed a synthetic 5C contact map derived from a genome-wide 1Mb Hi-C contact map for genomic regions containing genes in the EMT and E2F gene networks. This was done by locating the genomic bins corresponding to the network genes, extracting the interchromosomal and network-specific intrachromosomal interaction frequencies for those bins, and stitching them together in genomic order. Our working set of network genes was sourced from the KEGG database and included 189 EMT genes and 189 E2F genes distributed across all chromosomes in the genome. Some of the 378 total genes occupy the same 1Mb genomic bins resulting in a 333 by 333 1Mb-resolution synthetic 5C adjacency matrix.

Permutation test

To determine whether the EMT and E2F gene networks are more closely interacting than other regions of the genome, we performed a permutation test with 1000 randomly sampled sets of 333 1 Mb genomic bins across the genome. The Fiedler number for each set was computed and the distribution of numbers was plotted with the observed Fiedler number corresponding to the combined EMT and E2F gene networks. Our null hypothesis is that the EMT/E2F gene networks and random gene sets do not differ and our significance level is 0.01.

Von Neumann entropy

Entropy is a measure of "disorder" in a given system – the higher the entropy, the greater the disorder. In the context of genome structure, the higher the entropy, the more conformations available to the system [46]. If the distant ends of a genomic region, e.g., a gene, interact to form a loop, there are fewer conformations available to the gene and thus the entropy of that genomic region is reduced. In this way, entropy can be considered a signature for chromatin conformation. We compute the Von Neumann Entropy (VNE) - multivariate entropy - using the following equation:

$$H = -\sum_{i=1}^{n} \lambda_i ln\lambda_i$$

Here, *H* takes on positive values with larger values indicating higher chromatin accessibility.

Data availability

The RNA sequencing and Hi-C data sets generated in this study have been deposited to the Gene Expression Omnibus and can be accessed via their accession numbers, GSE151934 and GSE151970, respectively.

Results

Gene expression is disproportionately up-regulated across the time series

The time series data was generated by sequential addition of a small, interfering RNA (siRNA) targeting the 3' UTR of *TCF7L2* at equally spaced time points (0, 24, 48, and 72 hours) according to a modified siRNA transfection protocol (Supplementary Fig. 1A, see Methods). The modified protocol was designed to mitigate varying cell cycle distributions, which occur naturally due to varying growth times and are a major confounding factor in extended (>48 hours) time series data (Supplementary Fig. 1C and D). The modified protocol resulted in ~70% of the cells in the G_1 phase of the cell cycle at each time point (Supplementary Fig. 1B).

To assess the efficacy of siRNA-mediated *TCF7L2* silencing, the abundance of *TCF7L2* transcripts was determined using quantitative PCR (qPCR). Transcript levels decreased progressively across the time series resulting in an ~2-fold decrease by 72 hours (Fig. 1B). The amount of TCF4, the protein product of *TCF7L2*, was determined using a Western blot. TCF4 protein levels decreased marginally after 24 hours, ~35% after 48 hours and ~70% after 72 hours (Fig. 1C). Principal component analysis (PCA) of the RNA sequencing data demonstrated a directed change in the global gene expression program, with biological replicates clustering together (Fig. 1D). The number of differentially regulated genes which exhibited both a statistically significant ($P < 0.05$) and a 4-fold change in expression increased dramatically across the time series (g). However, the balance was positively skewed with nearly twice the number of genes up-regulated as down-regulated. Considering that SW480 cells have constitutive Wnt signaling activity, and therefore persistent transcription through βcat/TCF4 complexes, the general increase in gene expression following the reduction in TCF4 is unexpected.

To determine the biological relevance of the disproportionate up-regulation of gene expression, the expression profiles of individual genes with known biological roles were assessed. The expression of prominent Wnt signaling target genes, such as *MYC* and *CCND1*, decreased consistently across the time series, similar to that of *TCF7L2* [14,47,48]. It has been previously reported that loss of βcat/TCF4- mediated transcription results in the up-regulation of genes associated with intestinal differentiation [9]. We confirm this observation as genes associated with differentiation, such as *MUC2* and *LGALS4*, were significantly up-regulated following *TCF7L2* silencing ($P < 0.05$)(Supplementary Fig. 2A).
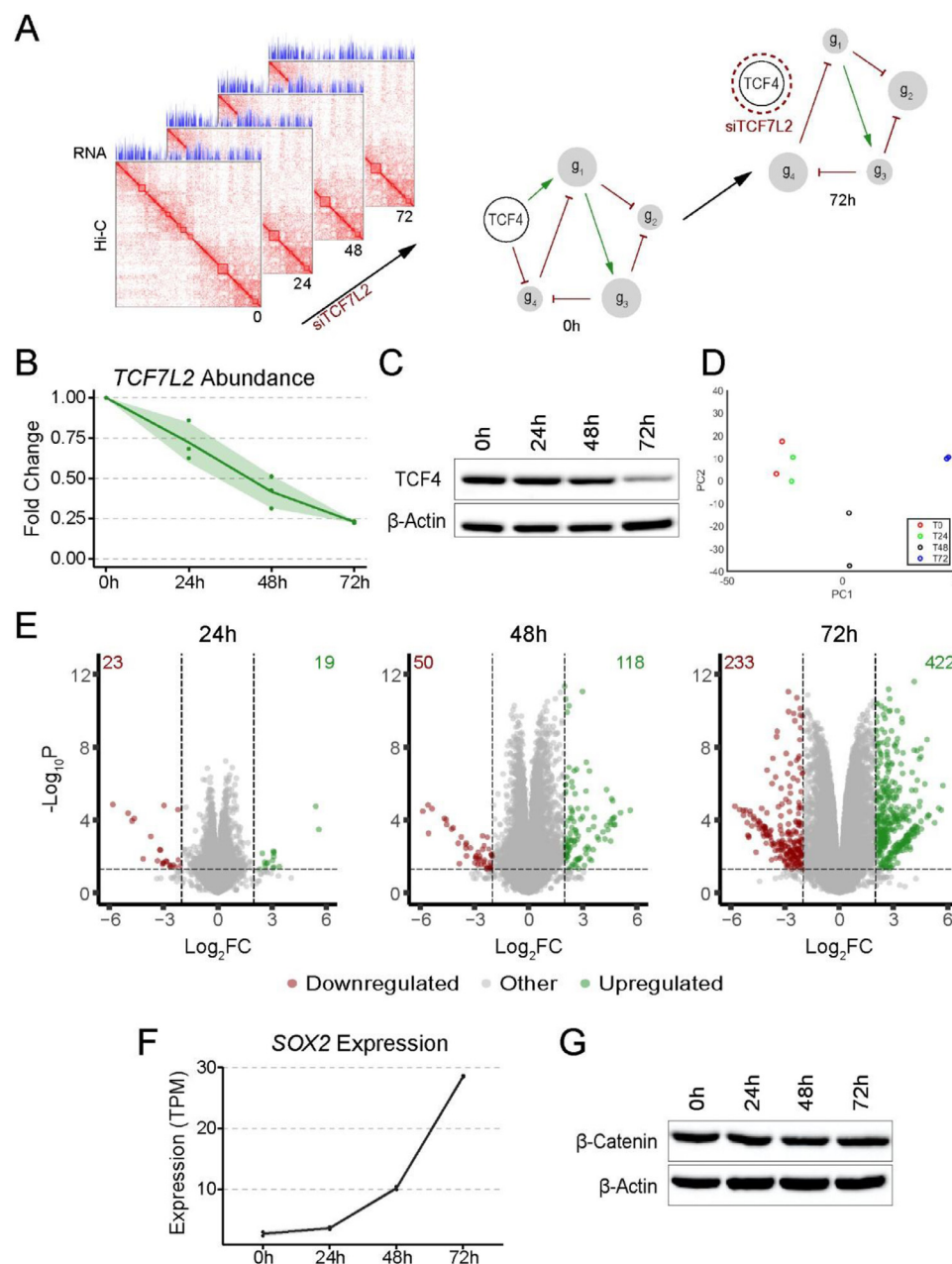
Fig. 1. Conceptual approach and gene expression dynamics following *TCF7L2* silencing. (A) Schematic summary of the experimental approach. We sought to explore how silencing *TCF7L2* impacted the colon cancer gene network in terms of chromatin structure and gene expression. (B) qPCR demonstrated progressive silencing of *TCF7L2* over time with a 75% reduction in *TCF7L2* transcript levels after 72 hours. Each dot represents a biological replicate, with 3 replicates plotted per time point ($n = 3$). The green line represents the mean, while the green-shaded ribbon represents the standard deviation. (C) Western blot analysis demonstrated a 70% decrease in TCF4 protein abundance by the last time point with β-actin as loading control. (D) PCA of gene expression (TPM) over the time course, with time points differentiated by color ($n = 2$). (E) Volcano plots show differential gene expression genome-wide with thresholds set at $\log_2 FC = 2$, and $P = 0.05$ (equivalent to 1.3 on the $-\log 10P$ scale). Genes in red are significantly down-regulated and undergo a >4-fold decrease in expression, while those in green are significantly up-regulated and experience a >4-fold increase in expression. The number of genes which fall within these regions are shown. (F) Expression profile of *SOX2*, which increases dramatically across the time series. The dots correspond to each biological replicate ($n = 2$), the shaded ribbon represents the standard deviation. (G) Western blot for active β-catenin over time with β-actin as loading control. No noticeable change in active β-catenin protein levels occurs during the time series. (Color version of figure is available online.)

SOX2 up-regulation is a consequence of TCF7L2 *silencing*

We identified a powerful ($\log_2 FC \sim 3.55$) up-regulation of *SOX2*, which began after 24 hours and continued to increase until 72 hours post-transfection, which we found was dose-dependent in relation to the levels of *TCF7L2* (Fig. 1F, Supplementary Fig. 2B). This may represent direct

repression of *SOX2* by TCF4, as TCF4 has been previously reported to bind the *SOX2* promoter in colorectal cancer [49]. To determine the influence of nuclear β-catenin on this interaction, we performed a Western Blot (Fig. 1G). Levels of active β-catenin did not fluctuate over the time series, indicating that the interaction between TCF4 and *SOX2* is independent of β-catenin,

in this system, supporting our direct repression hypothesis. Given the potency of the *SOX2* up-regulation, we also investigated whether a concomitant shift in nuclear structure occurred at the *SOX2* locus. *SOX2* resides at a TAD boundary in SW480 cells, however the boundary remained unchanged across the time series, indicating that nuclear structure was not a contributing factor in the up-regulation of *SOX2* (Supplementary Fig. 3A). The observation that *TCF7L2* silencing results in the over-expression of *SOX2* suggests that *TCF7L2* may enforce a transcriptional module which potently represses *SOX2* under physiological conditions.

Genome-wide A/B compartments are maintained over time

Hi-C was performed to capture genome-wide changes in chromosome structure across the time series to determine the influence of TCF4 on chromatin organization. *TCF7L2* silencing resulted in a significant, genome-wide increase in both the number and strength of pairwise chromatin interactions, as evidenced by matrix subtraction between 0 and 72 hour Hi-C contact maps (Fig. 2A). The contact maps demonstrate a higher interaction frequency at 72 hours, while the algebraic difference between the 2 contact maps reveals that the overwhelming direction of change is an increase in chromatin interaction from 0 to 72 hours. Out of all possible 1Mb-scale chromatin interactions, SW480 cells at 0 hours harbored 74% of those interactions, whereas by 72 hours the proportion of interactions had risen to 87%, demonstrating that the chromatin became more compact over time.

To further assess changes in genome-wide chromatin organization, the genome was assigned to A/B compartments using a spectral graph theory approach, which calculates the Fiedler number for each 100kb bin (Supplementary Fig. 3B)[42]. The sign of the Fiedler number, positive or negative, indicates within which compartment the bin resides. An A/B compartment switch occurred in ∼4% of the genome at any given point during the time series, which encompassed 1,305 genes. K-means clustering identified 8 distinct switching patterns in which a specific genomic locus switched compartments unidirectionally (A-to-B or B-to-A) or bidirectionally (a permutation of A-to-B-to-A-to-B) over time (Fig. 2B). Approximately 52% of the regions demonstrating an A/B compartment switch underwent a single switch, while the remaining 48% underwent multiple switching events. Despite the switching events, the overall proportion of A/B compartments genome-wide remained relatively unchanged with most loci residing in the same chromatin state across the time series (52% A and 48% B)(Supplementary Table 2). Furthermore, ∼1.3% of the genes within the A/B switching regions were significantly differentially expressed following the compartment switch.

TAD boundary loss in the CEACAM gene cluster

Local partitioning of chromatin into TADs provided insight into the coupling of structure and function. Genes occupying the same TADs have been shown to be co-expressed, likely by accessing the same cluster of transcriptional machinery due to their spatial proximity [18,50]. The number of TADs genome-wide fluctuated by ∼5% over the time series, while the number of common TADs shifted by ∼7% (Supplementary Tables 3 and 4). To gain insight regarding which factors may be involved in the TAD domain switching, we used publicly available CTCF and TCF4 ChIP-seq data sets from the HCT116 colorectal cancer cell line (see Methods, Supplementary Tables 5 and 6). We observed that the binding locations of CTCF in HCT116 matched ∼75% of the TAD boundaries of SW480 while the binding locations of TCF4 matched ∼41% of the TAD boundaries. However, while TAD boundary binding by CTCF was relatively consistent across chromosomes, the binding of TCF4 fluctuated, with chromosome 19 showing the highest levels of TCF4 binding.

On chromosome 19, an ∼10Mb span of the genome houses a group of *CEACAM* family genes, which are involved in colorectal cancer progression

and metastasis [51]. At the initial time point, *CEACAM5, CEACAM6*, and *CEACAM7* were located within the same TAD, while *CEACAM1* and *CEACAM8*, were located in an adjacent, smaller TAD (Fig. 2C). Changes in expression of the 2 *CEACAM* gene clusters after 24 hours were negligible. However, at the 48-hour time point, the boundary separating the 2 TADs was lost, joining *CEACAM5, CEACAM6*, and *CEACAM7* into a larger TAD with *CEACAM1* and *CEACAM8*. Combination of the TAD domains as well as a significant increase in expression for both *CEACAM* groups occurred seemingly simultaneously. Gene expression for both groups continued to increase following the TAD boundary loss (Fig. 2C). The increase in gene expression likely represents increased accessibility to the transcriptional machinery for both gene clusters. Silencing of *TCF7L2* in multiple colon cancer cell lines demonstrates that the up-regulation of *CEACAM1* is present in both APC mutated (SW480, DLD1) as well as APC wild-type (HCT116, LS174T) colon cancer cell lines (Fig. 2D). The up-regulation of *CEACAM1* is however weaker in COLO201, which is derived from a metastatic site.

TAD boundaries of the CEACAM gene cluster show enrichment for the SP1, KLF4, ZFX, and MZF1 transcriptions factors

To identify which transcription factors may be mediating this TAD partitioning loss, a transcription factor enrichment analysis was performed using oPOSSUM-3 (Fig. 2E)[44]. Over-representation of TFBS in the DNA sequences of the 64 genes occupying the differentially conformed region of chromosome 19 were investigated. The top 4 identified transcription factors were SP1, KLF4, ZFX, and MZF1 (Fig. 2F). The top hit, SP1, is found ubiquitously and is involved in chromatin remodeling. KLF4 is involved in embryonic development and binds a GC-rich motif highly similar to SP1. ZFX plays a role in self-renewal of hematopoietic stem cells and MZF1 has been associated in inflammatory bowel disease. To further explore the role of SP1, and TCF4, in mediating this TAD boundary change, we used publicly available SP1 and TCF4 ChIP-seq data sets (see Methods) from the HCT116 colorectal cancer cell line. We found that both SP1 and TCF4 bind the site of the TAD boundary observed in SW480 and it is therefore feasible that they are responsible for the repartitioning of this domain (Fig. 2C).

Coordinated pathway responses suggest a loss in cell cycle progression and DNA synthesis capabilities

To assess the dynamics of TCF4 loss on pathway behavior, Fast Gene Set Enrichment Analysis (fgsea) was performed using the 50 Hallmark gene sets and the log2FC expression values from each time point to generate a normalized enrichment score (NES) [38,39]. The NESs from fgsea were compared and the 20 gene sets (10 up-regulated and 10 down-regulated) with the greatest change by the last time point were plotted (Fig. 3A). Strong decreases in expression for both MYC Target gene sets as well as the E2F Targets and G2M Checkpoint gene sets were observed. All 4 of these gene sets are involved in growth and the regulation of cell cycle progression. The E2F pathway is also involved in DNA synthesis, and links DNA synthesis to cell cycle progression [52,53].

Interestingly, the Notch, Cholesterol homeostasis, Estrogen Response, and MTORC1 signaling pathways all demonstrated a similar NES profile over time, suggesting coordinated function. Several articles identify interplay between these pathways, supporting this hypothesis [54,55]. The EMT, Apoptosis, and Interferon Gamma pathways also demonstrated a coordinated network of up-regulated pathways. The most dynamically up-regulated and down-regulated gene sets upon *TCF7L2* silencing were EMT and E2F, respectively.
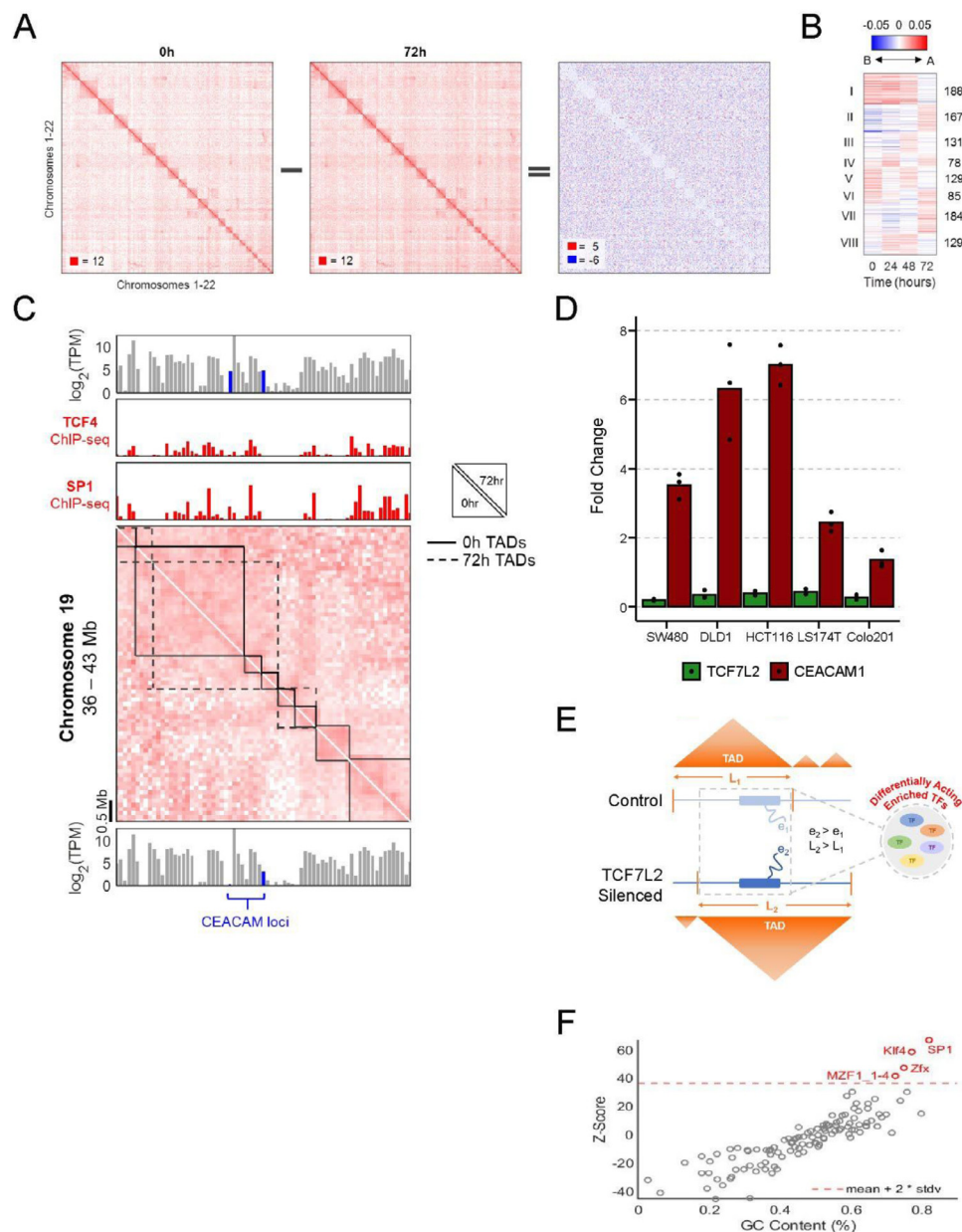
Fig. 2. Changes in global and local Hi-C partitioning over time. (A) Algebraic difference between genome-wide Hi-C contact maps at 0 and 72 hours demonstrates an overall increase in contact frequency over time. (B) Global partitioning (A/B compartmentalization) dynamics revealed that 1091 100-kb genomic bins switched compartments at one or more time points genome-wide. The Fiedler number (see Methods) of each loci are k-means clustered into 8 groups (denoted to the left) exhibiting different switching dynamics. The number of genomic bins in each cluster is specified on the right. A 95% change in Fiedler number from one time point to another was set as a threshold to remove inherent noise. (C) Local partitioning (TAD organization) of the region on Chromosome 19 (34.6 - 45.2 Mb) containing two *CEACAM* gene groups. Hi-C contact maps are shown at 100kb resolution with TAD domains at 0h denoted by solid, black lines and at 72h denoted by dashed, black lines. The *CEACAM* gene groups are denoted by blue lines in the expression array. ChIPseq data for TCF4 and SP1 binding demonstrated that TCF4 and SP1 may bind within the TAD boundary region. (D) qPCR was performed on *TCF7L2* silenced cells at the 72-hour time point to determine the expression of *CEACAM1* in various colon cancer cell lines. Expression of *TCF7L2* and *CEACAM1* was normalized to the negative control (Time 0 - not plotted), which was uniformly set to 1 for each cell line. All cell lines tested demonstrated an up-regulation of *CEACAM1*, however the weakest response was observed in COLO201. (E) Diagram illustrating coupled chromosome structure and gene expression between two conditions as well as a possible explanation for this coupling, i.e., condition specific transcription factor activity. (F) Transcription factor enrichment analysis for the 64 genes found in the differentially conformed region of chromosome 19 (see 2C) showed enrichment for SP1, KLF4, ZFX, and MZF1 when compared against a background of 24,752 genes.
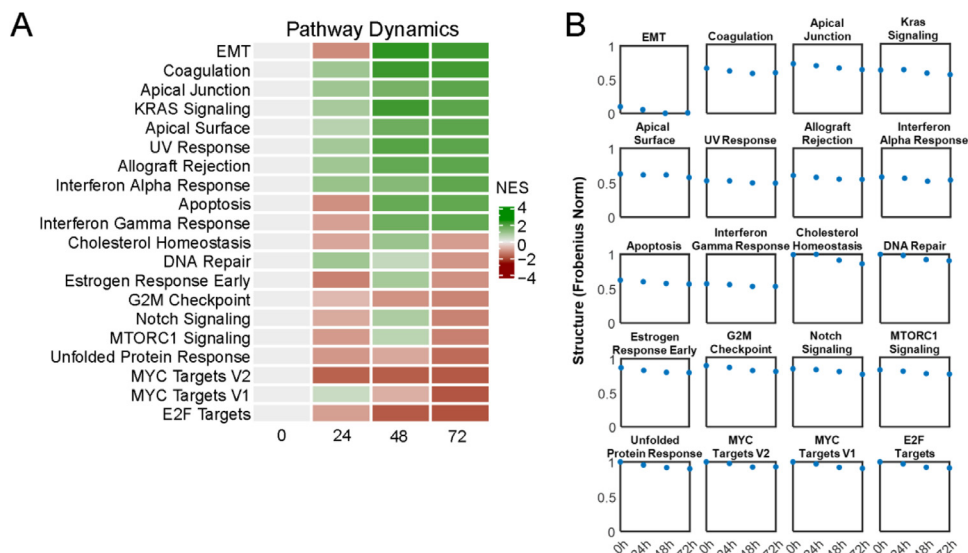
Fig. 3. Pathway Level Gene Expression and Structural Dynamics. (A) Fast Gene Set Enrichment Analysis (fgsea) was performed using the Hallmark Gene Sets and $\log_2$FC for each time point to generate a normalized enrichment score (NES). The NESs of the 10 most up- and down-regulated pathways were then plotted as a heatmap to show pathway level gene expression over time. The NES at the initial time point is set to 0. (B) The montage of plots reflect the change in Frobenius norm of pathway-specific contact frequency data for each time point and pathway of interest. Norm values are min-max normalized across all pathways represented. Each pathway is represented as 4 points, with each point corresponding to a time point, with higher values indicating a more open conformation. The E2F Targets, MYC Targets V1 and V2, Coagulation, and KRAS Signaling pathways all demonstrate high degrees of openness.

Pathway level structure-function relationships

We then sought to compare structure-function relationships in these pathways, utilizing the Frobenius norm, which describes the space of a vector (gene expression) or matrix (chromatin interaction frequencies) [56] (Fig. 3B). We find that, generally, the pathways which demonstrate more drastic changes in expression also occupy a more open conformation. For instance, strongly down-regulated E2F Targets, MYC Targets V1, V2, and Unfolded Protein Response pathway genes reside in a more open conformation than do the moderately down-regulated genes in the MTORC1, Notch, G2M Checkpoint, and Estrogen Response Early pathways. This remains largely true for the up-regulated pathways with the Coagulation, Apical Junction, and KRAS Signaling pathway genes residing in a more open conformation than the Allograft Rejection, Interferon Alpha Response, Apoptosis, and Interferon Gamma Response pathway genes.

However, the exceptions are the EMT, Cholesterol Homeostasis, and DNA Repair pathways. The EMT pathway undergoes the most dramatic shift in expression, yet is in the most constricted conformation, while the Cholesterol Homeostasis and DNA Repair pathways have the least change in expression yet still reside in one of the most open conformations.

EMT and E2F signaling genes are highly connected in SW480 cells

The 2 most dynamic gene sets, EMT and E2F, were then probed to gauge the connection between these networks in 3-dimensional space. A Hi-C-derived, synthetic 5C contact map was generated to highlight the inter-loci interaction frequencies among EMT and E2F signaling genes. The contact map is comprised of 1Mb bins containing genes in either of the pathways (see Methods). Compared to a synthetic 5C contact map composed of randomly sampled, gene-containing genomic bins, we observe that the EMT and E2F signaling genes contact map is denser, suggesting that these genes are more closely interacting than other genes in the genome (Fig. 4A). Permutation testing reveals that these genes are indeed more inter-connected than genes outside either network ($P < 0.001$), with a Fiedler number that surpasses

those of 1000 sets of randomly sampled genes (Supplementary Fig. 4A and B).

To understand the evolving behavior of individual genes in the EMT, E2F, and Wnt signaling networks, a gene-level, structure-function analysis was performed. Each gene was examined at 5kb resolution and was assigned a 35kb-length window – 5kb overlapping the transcription start site and 15kb flanking either side of that 5kb region – creating a 7 by 7 sub-matrix for each gene. Several network-based approaches have previously been used to characterize behaviors in dynamically changing genomes [57,58]. We apply one such approach - a derivative of VNE - to measure local chromatin organization of individual gene regions [59]. Higher VNE values indicate that the number of conformations available to the gene and its immediate neighborhood are higher, indicating that chromatin is more accessible. We computed VNE on the EMT, E2F, and Wnt signaling gene sub-matrices and plotted them as a function of gene expression, creating a phase portrait (Fig. 4B).

The phase portraits show the gene-level trajectory of chromatin accessibility and expression for the 10 most dynamic genes, as determined by the area of their ellipse. Overlap between the gene ellipses demonstrates a homogeneous concerted pathway response in the EMT and E2F gene sets, indicating that the genes, in their respective pathways, are in a similar chromatin environment and demonstrate co-ordinated changes in expression. The WNT gene network undergoes a greater diversity in its response, indicative of varying chromatin environments and diverse transcriptional regulation at the individual gene loci. Genes with different expression patterns, yet whose phase portraits still overlap, are indicative of regulation at the transcriptional level (such as transcription factor binding), rather than at the chromatin level. For example, *CCNY* and *LEF1* are present in similar chromatin environments, as indicated by their overlapping phase portraits, yet show opposite functional patterns, indicating that their regulation occurs at the transcriptional level. Overall, we observe that the fitted ellipses change dynamically in terms of structure and function, confirming that *TCF7L2* silencing influences chromatin structure.
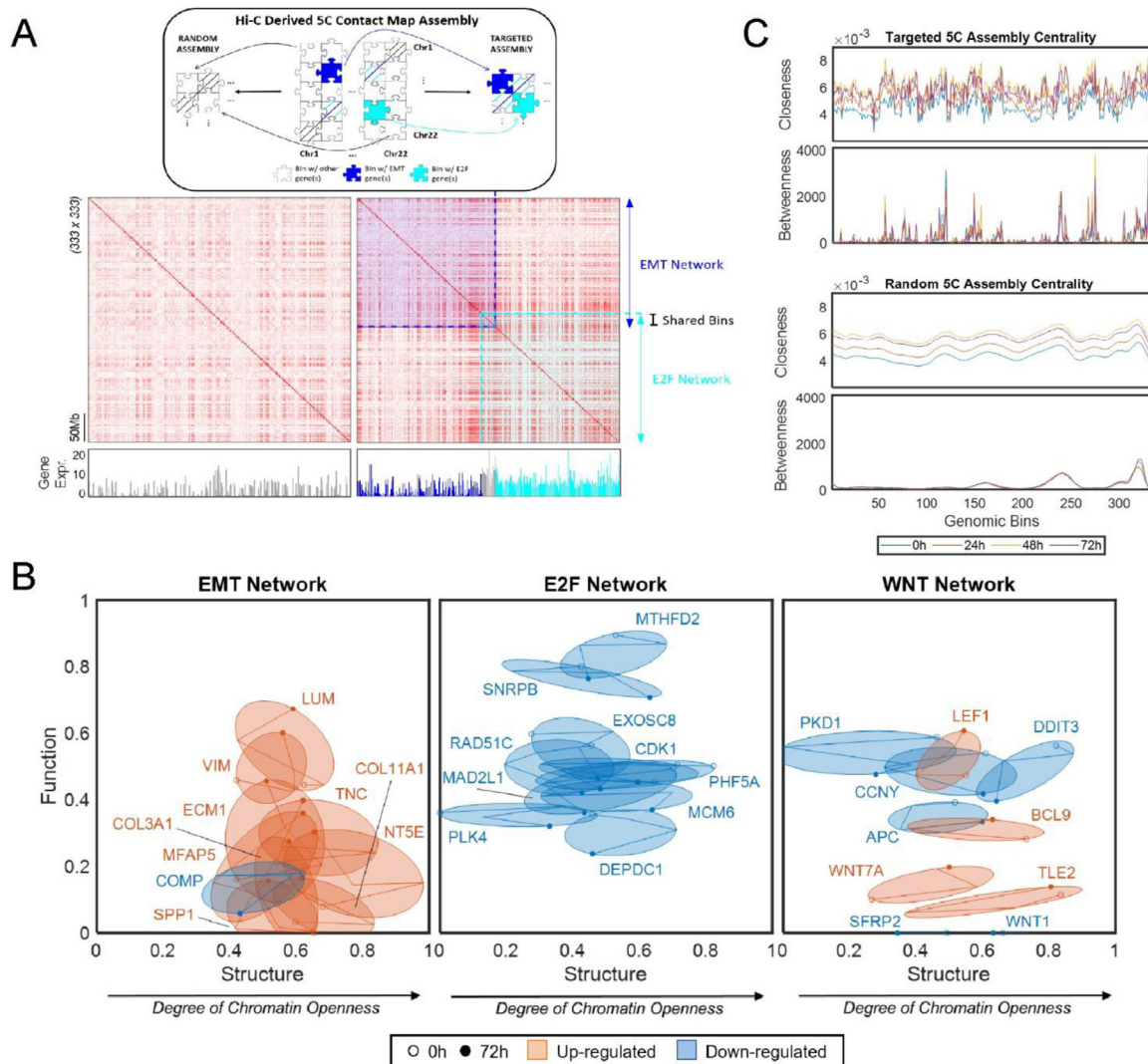
Fig. 4. Centrality Analyses and Gene Level Structure-Function Relationships. (A) Gene network-level synthetic 5C contact maps were generated by extracting bins corresponding to genes in the EMT and E2F signaling networks from a 1Mb adjacency matrix and stitched together (top) to create a synthetic contact map (right). Synthetic 5C contact map for randomly sampled gene bins (left) shown for comparison. (B) Structure versus function phase portrait for the EMT, E2F, and WNT signaling pathway genes. Min-Max normalized Von Neumann Entropy (see Methods) and gene expression represent structure and function, respectively. The 10 most dynamic genes (those whose fitted ellipse have the largest areas) in each network are shown. Genes that decrease in expression over time are labelled in blue and genes that increase in expression are labelled in orange. Phase portraits reveal the extent to which chromatin structure and gene expression are coupled. (C) Closeness centrality increases over time, indicating that the network becomes increasingly compact. An increase in betweenness centrality is also observed over time which reaches a maximum at 48 hours. Peaks in betweenness centrality plot identify bins (genes) which significantly regulate the network. The Random 5C Assembly demonstrates a progressive increase in closeness centrality, reflective of the general increased contact frequencies over time.

Network interactions are preserved as genes move spatially closer over time

To further explore chromatin topology, we performed network centrality analysis on inter-loci interaction data from our synthetic 5C contact maps. Network centrality describes the influence or importance of nodes in a network, which yields insight into how the network relays information [60,61,58]. We calculated 4 measures of centrality: betweenness, closeness, degree, and eigenvector. Betweenness centrality identifies nodes which lie along the shortest path between other nodes, closeness centrality is determined by the average farness to all other nodes, degree centrality

captures the number of edges connected to a node, and eigenvector centrality determines the influence of a node based on the influence of the neighboring nodes.

Roughly 57% of the EMT and E2F network genes underwent a significant change in betweenness centrality, the majority of which experienced a decrease in their betweenness centrality score (Fig. 4C). *LUM, TMPO,* and *AURKA* had the highest betweenness centrality scores at any given time point, suggesting that they are highly influential in their respective networks. Since silencing of *TCF7L2* resulted in decreased betweenness centrality scores for these genes, they represent the most likely nodes by which TCF4 may influence the EMT and E2F networks.

Table 1

Ranked candidate reprogramming factors.

Gene	Network	Gene Expression (Fold Change)	Structure (Frobenius Norm)	Betweenness Centrality (Fold Change)	Biological Function
JUN	EMT	0.56	87.11	4.29	Stabilizes βcat/TCF4 complexes, decreases MYC degradation
SNAI2	EMT	−1.35	81.57	2.81	Promotes invasion and metastasis in EMT
TP53	E2F	−0.74	66.56	2.45	Crucial tumor suppressor involved in DNA repair/apoptosis
MXD3	E2F	−0.10	78.36	−4.21	Suppresses MYC dependent cell transformation
MSX1	EMT	0.12	85.91	−3.35	Transcriptional repressor in limb-pattern formation
MYBL2	E2F	−0.68	149.56	−2.27	Regulates cell survival, proliferation, and differentiation
PRRX1	EMT	−1.36	59.32	3.00	Regulator of muscle creatine kinase
E2F8	E2F	−0.21	151.82	−1.29	Regulates progression from G1 to S phase
DNMT1	E2F	-0.12	79.52	1.02	May silence tumor suppressor genes by methylation

A total of 215 genes from the EMT and E2F signaling networks were ranked according to the magnitude of their change in gene expression, Frobenius norm, and betweenness centrality, between the 0- and 72-hour time points, irrespective of direction. The list of candidate factors was then filtered by removing genes which are not well characterized as transcription factors as well as those whose motif binding information was unavailable.

Closeness centrality scores provided insight into the evolving proximity of the network genes. We observe that closeness centrality increased with time for all loci with either of the latter 2 time points (48 and 72 hours) having the highest scores (Fig. 4C). This accompanies the previous observation that pairwise chromatin interactions increase over time, that is, chromatin organization is becoming denser. Since a higher closeness centrality score indicates that a node is relatively closer to all other nodes in the network, this trend suggests that the genes become more organized with time. We observe that the eigenvector and degree centralities of our network genes fluctuate very little with time-meaning the same gene interactions with a similar magnitude of connectivity occur at each time point (Supplementary Fig. 4C).

4DN analysis provides insight into TF-driven controllability of CRC cells

Our findings demonstrate that silencing *TCF7L2* is sufficient to mediate structural and functional changes that influence the behavior of colon cancer cells. We suspect that there are a number of reprogramming factors that, when silenced alongside *TCF7L2*, could further destabilize the colorectal cancer cell. Genes that have significant nodal influence represent ideal candidate reprogramming factors, as a result of their ability to destabilize and promote wide-spread changes in the network. Betweenness centrality is a measure of nodal influence and is particularly characteristic of cross-talk control. We find that in the SW480 colon cancer network, 187 1Mb bins containing 215 genes from the EMT and E2F signaling gene networks harbored a significant change in betweenness centrality from 0 to 72 hours. These genes were ranked according to their magnitude of change in betweenness centrality, with higher magnitudes corresponding to a higher ranking. The genes were then ranked separately by their percent change in gene expression and Frobenius norm between 0 and 72 hours to incorporate aspects of gene function and structure. The average of the 3 rankings for each gene was computed and the genes ordered accordingly, with lower rankings indicating a gene with the most controllability potential. Considering that transcription factors regulate a network of genes, we summarize the rankings and dynamics for the most likely transcription factor-encoding genes from our candidate gene list in Table 1. The highest ranked candidate was c-JUN, an oncogenic subunit of the AP-1 transcription factor, whose activity is augmented in many cancer types, and has been shown to interact with TCF4 and β-catenin to form a ternary, transcriptionally-competent complex [62].

Discussion

TCF4 as a transcriptional repressor

Colorectal cancers exhibit constitutively active Wnt signaling, most commonly due to mutations in *APC* [10]. The over-expression of Wnt target genes is driven by βcat/TCF4 complexes, as TCF4 is the major Wnt signaling transcription factor in colorectal cancer [13]. Due to the central role of Wnt signaling in colorectal tumorigenesis as well as the capability of β-catenin to modulate the chromatin environment, we explored the dynamical influence of TCF4 on chromatin structure and gene expression in the SW480 colon cancer cell line.

TCF7L2 silencing resulted in a progressive deregulation of the transcriptome with the most dynamic changes occurring at the last time point. Surprisingly, nearly twice the number of genes were up-regulated upon *TCF7L2* silencing. While difficult to distinguish between direct and indirect effects, these results hint at a repressive role for TCF4. In support of this, naturally occurring dominant-negative variants of TCF4, have been described [63,64]. The gene *SOX2* demonstrated a powerful increase in expression, which occurred progressively over the time series. We reason that this induction is a result of decreased repression by dominant negative TCF4 isoforms, as TCF4 has been shown to bind the *SOX2* promoter [49]. Additionally, this interaction occurs independently of fluctuating β-catenin levels, in line with repression by a dominant negative isoform. Furthermore, interactions between canonical Wnt and *SOX2* have been described in various cell types. In lung epithelia, an antagonistic role between canonical Wnt signaling and *SOX2* has been described in which activation of Wnt signaling led to the loss of *SOX2* and interfered with the development of the bronchiolar epithelium [65]. In gastric tumorigenesis, *SOX2* functions as a tumor suppressor which antagonizes the growth of Wnt-driven adenomas [66]. These 2 examples demonstrate a mutual, antagonizing relationship between *SOX2* and canonical Wnt signaling, which influences the development of normal tissue (lung) as well as adenomas (stomach). Here, we observe an antagonizing role between TCF4 and *SOX2* in colon cancer. Taken together, these results suggest that TCF4 maintains a transcriptional module for repressing *SOX2*, which may be necessary for generating and/or maintaining the intestinal lineage.

TCF4 influences local genome organization

Time series Hi-C experiments demonstrated that TCF4 influences chromatin organization both globally and locally. Silencing of *TCF7L2* led

to an increase in pairwise chromatin interactions, demonstrating an overall compaction of the chromatin across the genome. This may be due to a decrease in βcat/TCF4 binding, which would lead to decreased β-catenin-mediated chromatin acetylation. The high levels of nuclear β-catenin present in colorectal cancer may be sufficient for mediating a genome-scale chromatin effect. However, the compaction was not sufficient to cause a decrease in gene expression.

Approximately 4% of the genome underwent an A/B compartment switch at any given time point. The silencing of a single transcription factor leading to compartment switching in ~4% of the genome is comparable to over-expressing 4 transcription factors which influence ~20% of the genome [67]. Despite changes in A/B compartment switching, the total number of bins which resided in the A/B compartments remained stable over time (52% A, 48% B). This supports the existence of a genome-wide 'balance' of open and closed chromatin domains. However, the A/B compartment balance in normal cells of the mouse is different, at 40% A and 60% B [67]. This may be a reflection of the difference in species or could be an indication that, since cancer cells have enlarged nuclei, they may be able to support a larger percentage of the genome in the A compartment, and therefore support a more diverse transcriptome. However, an increased nuclear volume could merely reflect the increased DNA content of cancer cells. Another explanation for the relatively stable A/B compartment balance is a result of noise within the data. Since a single time series was performed for Hi-C, we are unable to exclude this possibility.

TAD analysis demonstrated that the number of TAD domains fluctuated by ~5% over time. Incorporating CTCF ChIP-seq data from HCT116 demonstrated that the majority (80%) of the TAD boundaries are likely bound by CTCF, whereas 41% are likely to be bound by TCF4. On chromosome 19, a TAD domain change was coordinated with an increase in gene expression in the *CEACAM* family genes. The most dramatic change in the expression of these genes occurred concomitantly as the TAD boundary loss, hinting at a causal link between the two. Following the TAD domain coalescence, the expression of both *CEACAM* gene groups continued to increase, likely by accessing the same transcriptional machinery, as they are part of the same gene family. Incorporation of ChIP-seq data for TCF4 and SP1, the transcription factor with the highest degree of enrichment in the switching region, suggested that TCF4 and SP1 are responsible for the TAD domain change as peaks for both transcription factors were present near the initial TAD boundary. Given the current understanding of structure-function relationships, we interpret the dynamics of the event thus: *TCF7L2* silencing influences the activity of SP1, which results in loss of the TAD boundary. Boundary loss increases the probability that the 2 *CEACAM* groups will come in close contact, thereby allowing them to coordinate recruitment of the transcriptional machinery.

Structure-function relationships in SW480

From fgsea, we observed down-regulation of the MYC, E2F, and the G2/M gene sets, all which mediate growth and cell cycle progression. In addition to regulating growth, genes in the E2F pathway are also responsible for DNA synthesis [52,53]. Additionally, despite showing the greatest decrease in pathway gene expression, E2F genes are still expressed to a higher degree than genes in EMT, the most up-regulated gene set (Fig. 4A - Gene Expression). Previous evidence has demonstrated that TCF4 expression corresponds with resistance to CRT and that loss of TCF4 increases sensitivity [23,24,68]. Furthermore, it was been suggested that the mutation rate in intestinal stem cells is relatively equal to the mutation rate in liver stem cells, despite the higher proliferative rate of the former [69,70]. We hypothesize that high levels of DNA synthesis/DNA proofreading activity, mediated by E2F, could maintain DNA sequence integrity in the face of continued stem cell proliferation and, in colorectal cancers, this same DNA program may repair the damage caused by CRT. Loss of TCF4 results in cell cycle arrest,

ensuring that cells with hampered DNA repair capabilities do not propagate in the stem cell crypt.

Investigation of structure and function relationships in the top 20 most dynamic pathways demonstrated that pathways whose genes resided in a more open conformation also demonstrated more dramatic changes in expression. Pathways which demonstrated less dramatic changes in expression also resided in chromatin environments which were more constricted. This trend was observed in 85% of the most dynamically acting pathways. According to these results, we find that the conformation of the chromatin is not a determinant of gene expression, but rather, functions as a filter. Genes which reside in open conformations are easily accessed by the transcriptional machinery and thus experience a full response, while genes in a more constricted environment have less exposure and therefore undergo fewer dynamic changes in expression. The general decrease in chromatin accessibility across the time points is a result of the genome-wide chromatin compaction which occurred as a result of *TCF7L2* silencing. In the remaining 15% of the pathways, specifically EMT, Cholesterol Homeostasis, and DNA Repair, the connection between expression and structure was not pronounced. Despite showing the most dynamic change in expression, the EMT pathway genes resided in the most constricted conformation out of the top 20 pathways. Conversely, the Cholesterol and DNA Repair pathway genes resided in the most accessible conformation yet underwent fewer dynamic changes in expression then other pathways. Therefore, in some instances, chromatin conformation is not a determinant of gene expression.

Candidate reprogramming factors for colorectal cancer

Centrality analysis allowed us to determine the most influential genes in their respective networks. The genes with the highest betweenness centrality were *LUM, TMPO,* and *AURKA*, identifying these genes as the most useful targets for propagating a stimulus through their respective pathways (EMT or E2F). We subsequently identified a list of factors which can be used in concert with loss of TCF4 to profoundly perturb the colon cancer cell. The top ranked candidate reprogramming factor was c-JUN, an oncoprotein which forms part of the AP-1 transcription factor. Previous work has shown that c-JUN forms a complex with TCF4 and β-catenin, which stabilizes β-catenin, and drives expression of the *JUN* promoter, initiating a self-stimulating feedback loop [71,72]. Abrogation of the c-JUN, TCF4, β- catenin complex led to reduced tumor number and size, as well as prolonged lifespan [62]. High levels of c-JUN and Wnt signaling activity resulted in increased resistance to cisplatin whereas abrogation of c-JUN or Wnt signaling activity increased the sensitivity of ovarian cancer cells to treatment.

Conclusion

In conclusion, we find that *TCF7L2* influences both structure and function in the SW480 colon cancer cell line. Loss of TCF4 results in a potent up-regulation of *SOX2*, which in a saturated WNT environment, is not dependent upon β-catenin, hinting at a repressive role for TCF4. When TCF4 is present, chromatin resides in a more open conformation genome-wide, which is likely mediated by βcat/TCF4 complexes. At the local chromatin level, the expression of a group of *CEACAM* genes situated at a TAD boundary was greatly increased upon TAD boundary loss, a change which is likely mediated by TCF4 or SP1. The expression of pathway specific genes corresponded to the accessibility of those genes, suggesting that chromatin conformation acts as a filter to attenuate the potency of gene expression regulation. The 2 most dynamic pathways following *TCF7L2* silencing – EMT and E2F – interact in 3-dimensional space and the most dynamic pathway genes inhabit overlapping structure-function space. Incorporation of chromatin structure, gene expression, and centrality data allowed us to generate a list of transcription factors which would be most effective at perturbing the colorectal cancer cell alongside TCF4. Our top

ranked transcription factor, c-JUN, is known to bind TCF4 and β-catenin, thereby confirming the validity of network connectivity analysis in identifying relevant biological interactions.

Acknowledgments

Competing interests

All authors declare that they have no competing financial or nonfinancial interests that might have influenced the performance or presentation of the work described in this manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neo.2020.12.010.

References

[1] Korinek V, Barker N, Moerer P, Donselaar EV, Huls G, Peters PJ, Clevers H. Depletion of epithelial stem-cell compartments in the small intestine of mice lacking tcf-4. *Nat Genetics* 1998;**19**(4):379–83.

[2] van Es JH, Haegebarth A, Kujala P, Itzkovitz S, Koo B-K, Boj S F, Korving J, Born MVD, Naarden AVO-, Robine S, et al. A critical role for the wnt effector tcf4 in adult intestinal homeostatic self-renewal. *Mol Cell Biol* 2012;**32**(10):1918–27.

[3] Agalioti T, Lomvardas S, Parekh B, Yie J, Maniatis T, Thanos D. Ordered recruitment of chromatin modifying and general transcription factors to the ifn-β promoter. *Cell* 2000;**103**(4):667–78.

[4] Hecht A, Vleminckx K, Stemmler MP, Roy FV, Kemler R. The p300/cbp acetyl-transferases function as transcriptional coactivators of β-catenin in vertebrates. *EMBO J* 2000;**19**(8):1839–50.

[5] Barker N, Hurlstone A, Musisi H, Miles A, Bienz M, Clevers H. The chromatin remodelling factor brg-1 interacts with β-catenin to promote target gene activation. *EMBO J* 2001;**20**(17):4935–43.

[6] Mosimann C, Hausmann G, Basler K - Rad. Parafibromin/hyrax activates wnt/wg target gene transcription by direct association with β-catenin/armadillo. *Cell* 2006;**125**(2):327–41.

[7] Barker N, Es JHV, Kuipers J, Kujala P, Born MVD, Cozijnsen M, Haegebarth A, Korving J, Begthel H, Peters PJ, et al. Identification of stem cells in small intestine and colon by marker gene lgr5. *Nature* 2007;**449**(7165):1003–7.

[8] Sato T, Es JHV, Snippert HJ, Stange DE, Vries RG, Born MVD, Barker N, Shroyer NF, de Wetering MV, Clevers H. Paneth cells constitute the niche for lgr5 stem cells in intestinal crypts. *Nature* 2011;**469**(7330):415–18.

[9] de Wetering MV, Sancho E, Verweij C, Lau WD, Oving I, Hurlstone A, Horn KVD, Batlle E, Coudreuse D, Haramis A - P, et al. The β-catenin/tcf-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. *Cell* 2002;**111**(2):241–50.

[10] Kinzler K W, Vogelstein B. Lessons from hereditary colorectal cancer. *Cell* 1996;**87**(2):159–70.

[11] Rowan AJ, Lamlum H, Ilyas M, Wheeler J, Straub J, Papadopoulou A, Bicknell D, Bodmer WF, Tomlinson IPM. Apc mutations in sporadic colorectal tumors: a mutational "hotspot" and interdependence of the "two hits. *Proc Natl Acad Sci* 2000;**97**(7):3352–7.

[12] Morin P J, Sparks A B, Korinek V, Barker N, Clevers H, Vogelstein B, Kinzler KW. Activation of β-catenin-tcf signaling in colon cancer by mutations in β-catenin or apc. *Science* 1997;**275**(5307):1787–90.

[13] Korinek V, Barker N, Morin PJ, Van Wichen D, De Weger R, Kinzler KW, Vogelstein B, Clevers H. Constitutive transcriptional activation by a β-catenin-tcf complex in apc-/- colon carcinoma. *Science* 1997;**275**(5307):1784–7.

[14] He T-C, Sparks AB, Rago C, Hermeking H, Zawel L, da Costa LT, Morin PJ, Vogelstein B, Kinzler KW. Identification of c-myc as a target of the apc pathway. *Science* 1998;**281**(5382):1509–12.

[15] Tetsu O, McCormick F. β-catenin regulates expression of cyclin d1 in colon carcinoma cells. *Nature* 1999;**398**(6726):422–6.

[16] Parker DS, Ni YY, Chang JL, Li J, Cadigan KM. Wingless signaling induces widespread chromatin remodeling of target loci. *Mol Cell Biol* 2008;**28**(5):1815–28.

[17] Lieberman-Aiden E, Van Berkum NYL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 2009;**326**(5950):289–93.

[18] Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 2012;**485**(7398):376–80.

[19] Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 2014;**159**(7):1665–80.

[20] Palstra RJ, Tolhuis B, Splinter E, Nijmeijer R, Grosveld F, de Laat W. The β-globin nuclear compartment in development and erythroid differentiation. *Nat Genet* 2003;**35**(2):190–4.

[21] Rajapakse I, Groudine M. On emerging nuclear order. *J Cell Biol* 2011;**192**(5):711–21.

[22] Ried T, Rajapakse I. The 4d nucleome. *Methods* 2017;**123**:1.

[23] Michael Ghadimi B, Grade M, Difilippantonio MJ, Varma S, Simon R, Montagna C, Füzesi L, Langer C, Becker H, Liersch T, et al. Effectiveness of gene expression profiling for response prediction of rectal adenocarcinomas to preoperative chemoradiotherapy. *J Clin Oncol* 2005;**23**(9):1826.

[24] Kendziorra E, Ahlborn K, Spitzner M, Rave-Fränk M, Emons G, Gaedcke J, Kramer F, Wolff HA, Becker H, Beissbarth T, et al. Silencing of the wnt transcription factor tcf4 sensitizes colorectal cancer cells to (chemo-) radiotherapy. *Carcinogenesis* 2011;**32**(12):1824–31.

[25] Cunningham D, Atkin W, Lenz H, Lynch HT, Minsky B, Nordlinger B, et al. Colorectal cancer. *Lancet North Am Ed* 2010;**375**(9719):1030–47.

[26] Andrews S, et al. *Fastqc: A Quality Control Tool for High Throughput Sequence Data*; 2010.

[27] Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 2011;**17**(1):10–12.

[28] Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* 2014;**15**(3):1–12.

[29] Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in a web browser. *BMC Bioinformatics* 2011;**12**(1):385.

[30] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. Star: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;**29**(1):15–21.

[31] Daley T, Smith AD. Predicting the molecular complexity of sequencing libraries. *Nat Methods* 2013;**10**(4):325–7.

[32] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and samtools. *Bioinformatics* 2009;**25**(16):2078–9.

[33] Wang L, Wang S, Li W. Rseqc: quality control of rna-seq experiments. *Bioinformatics* 2012;**28**(16):2184–5.

[34] Li B, Dewey CN. Rsem: accurate transcript quantification from rna-seq data with or without a reference genome. *BMC Bioinformatics* 2011;**12**(1):323.

[35] Law CW, Chen Y, Shi W, Smyth GK. Voom: Precision weights unlock linear model analysis tools for rna-seq read counts. *Genome Biol* 2014;**15**(2):R29.

[36] Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Gene Mol Biol* 2004;**3**(1):1544–6115.

[37] Blighe K, Rana S, Lewis M. Enhancedvolcano: Publication-ready volcano plots with enhanced colouring and labeling. *R package version* 2019;**1**(0). https://github.com/kevinblighe/EnhancedVolcano.

[38] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci* 2005;**102**(43):15545–50.

[39] Sergushichev A. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *BioRxiv* 2016:060012.

[40] Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 2016;**32**(18):2847–9.

[41] Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. Juicer provides a one-click system for analyzing loop-resolution hi-c experiments. *Cell Syst* 2016;**3**(1):95–8.

[42] Chen J, Hero AO III, Rajapakse I. Spectral identification of topological domains. *Bioinformatics* 2016;**32**(14):2151–8.

[43] Bonev B, Cavalli G. Organization and function of the 3d genome. *Nat Rev Genet* 2016;**17**(11):661.

[44] Kwon AT, Arenillas DJ, Hunt RW, Wasserman WW. Opossum-3: advanced analysis of regulatory motif over-representation across genes or chip-seq datasets. *G3: Genes, Genomes. Genetics* 2012;**2**(9):987–1002.

[45] Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, Lee RVD, Bessy A, Cheneby J, Kulkarni SR, Tan G, et al. Jaspar 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res* 2018;**46**(D1):D260–6.

[46] Phillips R, Kondev J, Theriot J, Garcia H. *Physical Biology of the Cell*. Garland Science; 2012.

[47] Weidinger G, Thorpe CJ, Wuennenberg-Stapleton K, Ngai J, Moon RT. The sp1-related transcription factors sp5 and sp5-like act downstream of wnt/β-catenin signaling in mesoderm and neuroectoderm patterning. *Curr Biol* 2005;**15**(6):489–500.

[48] Takahashi M, Nakamura Y, Obama K, Furukawa Y. Identification of sp5 as a downstream gene of the β-catenin/tcf pathway and its enhanced expression in human colon cancer. *Int J Oncol* 2005;**27**(6):1483–7.

[49] Hatzis P, Flier LGVD, Driel MAV, Guryev V, Nielsen F, Denissov S, Nijman IJ, Koster J, Santo EE, Welboren W, et al. Genome-wide pattern of tcf7l2/tcf4 chromatin occupancy in colorectal cancer cells. *Mol Cell Biol* 2008;**28**(8):2732–44.

[50] Cook PR, Marenduzzo D. Transcription-driven genome organization: a model for chromosome structure and the regulation of gene expression tested through simulations. *Nucleic Acids Res* 2018;**46**(19):9895–906.

[51] Beauchemin N, Arabzadeh A. Carcinoembryonic antigen-related cell adhesion molecules (ceacams) in cancer progression and metastasis. *Cancer Metastasis Rev* 2013;**32**(3-4):643–71.

[52] Ishida S, Huang E, Zuzan H, Spang R, Leone G, West M, Nevins JR. Role for e2f in control of both dna replication and mitotic functions as revealed from DNA microarray analysis. *Mol Cell Biol* 2001;**21**(14):4684–99.

[53] Ren B, Cam H, Takahashi Y, Volkert T, Terragni J, Young RA, Dynlacht BD. E2f integrates cell cycle progression with DNA repair, replication, and g2/m checkpoints. *Genes Dev* 2002;**16**(2):245–56.

[54] Chan SM, Weng AP, Tibshirani R, Aster JC, Utz PJ. Notch signals positively regulate activity of the mTOR pathway in t-cell acute lymphoblastic leukemia. *Blood J Am Soc Hematol* 2007;**110**(1):278–86.

[55] Boulay A, Rudloff J, Ye J, Zumstein-Mecker S, O'Reilly T, Evans DB, Chen S, Lane HA. Dual inhibition of mTOR and estrogen receptor signaling in vitro induces cell death in models of breast cancer. *Clin Cancer Res* 2005;**11**(14):5319–28.

[56] Strang G, Strang G, Strang G, Strang G. *Introduction to linear algebra*, 3. MA: Wellesley-Cambridge Press Wellesley; 1993.

[57] Seaman L, Chen H, Brown M, Wangsa D, Patterson G, Camps J, Omenn GS, Ried T, Rajapakse I. Nucleome analysis reveals structure–function relationships for colon cancer. *Mol Cancer Res* 2017;**15**(7):821–30.

[58] Liu S, Chen H, Ronquist S, Seaman L, Ceglia N, Meixner W, Chen P-Y, Higgins G, Baldi P, Smale S, et al. Genome architecture mediates transcriptional control of human myogenic reprogramming. *iScience* 2018;**6**:232–46.

[59] P-Y Chen, L Wu, S Liu, and I Rajapakse. Fast incremental von neumann graph entropy computation: theory, algorithm, and applications. *arXiv preprint arXiv:1805.11769*, 2018.

[60] Borgatti SP. Centrality and network flow. *Soc Netw* 2005;**27**(1):55–71.

[61] Newman MEJ. *Networks: An Introduction*. Oxford; 2010.

[62] Nateri AS, Spencer-Dene B, Behrens A. Interaction of phosphorylated c-JUN with tcf4 regulates intestinal cancer development. *Nature* 2005;**437**(7056):281–5.

[63] Vacik T, Stubbs JL, Lemke G. A novel mechanism for the transcriptional regulation of WNT signaling in development. *Genes Dev* 2011;**25**(17):1783–95.

[64] Cadigan KM, Waterman ML. Tcf/lefs and WNT signaling in the nucleus. *Cold Spring Harb Perspect Biol* 2012;**4**(11):a007906.

[65] Hashimoto S, Chen H, Que J, Brockway BL, Drake JA, Snyder JC, Randell SH, Stripp BR. β-catenin–sox2 signaling regulates the fate of developing airway epithelium. *J Cell Sci* 2012;**125**(4):932–42.

[66] Sarkar A, Huebner AJ, Sulahian R, Anselmo A, Xu X, Flattery K, Desai N, Sebastian C, Yram MA, Arnold K, et al. Sox2 suppresses gastric tumorigenesis in mice. *Cell Rep* 2016;**16**(7):1929–41.

[67] Stadhouders R, Vidal E, Serra F, Di Stefano B, Dily FL, Quilez J, Gomez A, Collombet S, Berenguer C, Cuartero Y, et al. Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. *Nat Genet* 2018;**50**(2):238–49.

[68] Braun R, Anthuber L, Hirsch D, Wangsa D, Lack J, McNeil NE, Heselmeyer-Haddad K, Torres I, Wangsa D, Brown MA, et al. Single cell-derived primary rectal carcinoma cell lines reflect intratumor heterogeneity associated with treatment response. *Clin Cancer Res* 2020:3468–80.

[69] Blokzijl F, De Ligt J, Jager M, Sasselli V, Roerink S, Sasaki N, Huch M, Boymans S, Kuijk E, Prins P, et al. Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* 2016;**538**(7624):260–4.

[70] Gehart H, Clevers H. Tales from the crypt: new insights into intestinal stem cells. *Nat Rev Gastroenterol Hepatol* 2019;**16**(1):19–34.

[71] Mann B, Gelos M, Siedow A, Hanski ML, Gratchev A, Ilyas M, Bodmer WF, Moyer MP, Riecken EO, Buhr HJ, et al. Target genes of β-catenin–t cell-factor/lymphoid- enhancer-factor signaling in human colorectal carcinomas. *Proc Natl Acad Sci* 1999;**96**(4):1603–8.

[72] Gan XQ, Wang JY, Xi Y, Wu ZL, Li YP, Li. Nuclear dvl, c-JUN, β-catenin, and tcf form a complex leading to stabilization of β-catenin– TCF interaction. *J Cell Biol* 2008;**180**(6):1087–100.