Scarless genome editing reveals risk SNP rs6983267-G and lncRNA CCAT2 dictate targetable PI3K signaling dependence in WNT-dysregulated CRC

Nicole B. Coggins
UC Davis: University of California Davis https://orcid.org/0000-0002-0790-6548
Henriette O'Geen
UC Davis: University of California Davis
Paul C. Lott
UC Davis: University of California Davis
David J. Segal
UC Davis: University of California Davis
Luis Carvajal Carmona (lgcarvajal@ucdavis.edu)
University of California Davis https://orcid.org/0000-0001-7129-2918

Research Article

Keywords: scarless genome editing, colorectal cancer, SNP rs6983267, CCAT2 expression, PI3K dysregulation

Posted Date: July 1st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-667797/v1

Scarless genome editing reveals risk SNP rs6983267-G and lncRNA CCAT2 dictate targetable PI3K signaling dependence in WNT-dysregulated CRC

Nicole B. Coggins[1], **Henriette O'Geen**[1], **Paul C. Lott**[1], **David J. Segal**[1,2,3*] **and Luis G. Carvajal-Carmona**[1,2,4*]

[1] Genome Center, University of California, Davis. Davis, CA 95616, USA

[2] Department of Biochemistry and Molecular Medicine, School of Medicine, University of California, Davis. Sacramento, CA 95817, USA

[3] Innovative Genomics Institute, University of California, Berkeley, Berkeley, CA 94720, USA

[4] University of California Davis Comprehensive Cancer Center. Sacramento, CA 95817, USA

* Senior authors

Correspondence: lgcarvajal@ucdavis.edu

ABSTRACT

Genome-wide association studies have identified numerous loci associated with increased risk for colorectal cancer (CRC), including 8q24.21 which contains a known enhancer of proto-oncogene *MYC*. However, the role of candidate functional SNP rs6983267 within this locus remains unclear. Here, we generate isogenic cellular models of risk SNP rs6983267 in human CRC line, HCT-116. Comprehensive molecular characterization reveals risk allele-G drives enhancer DNA contacts with downstream regions that include *MYC*. Absence of risk allele leads to activation of lncRNA *CCAT2*. Rather than changes in *MYC* expression, we observe activation of alternative growth factor signaling pathways with loss of both risk allele and *CCAT2* expression. Analysis of TCGA CRC cases demonstrates low *CCAT2* expression combined with non-risk rs6983267 genotype correlate with higher frequency of PI3K mutations in CRC patients displaying WNT dysregulation. Together, these provide a potential biomarker for therapeutically targetable PI3K dysregulation in CRC and application in cancer precision medicine.

INTRODUCTION

Colorectal cancer (CRC) is the second leading cause of cancer-related deaths worldwide[1–3]. The majority of CRC cases are adenocarcinomas, in which tumor cells originate from the epithelial lining of either the colon or rectum. Molecular progression of CRC adenocarcinomas is a well-studied stepwise accumulation of genetic and epigenetic alterations known as the adenoma-carcinoma sequence[4]. Specifically, cancer initiation in the form of benign adenomas, or polyps, begins with aberrant activation of the WNT signaling pathway[4–6]. This is followed by second driver mutations within the Ras/Raf/MAPK pathway, boosting cellular proliferation, and finally dysregulation of the TGF-β and PI3K/Akt signaling pathways, ultimately resulting in malignant carcinomas[4–6]. The discovery of these key pathways has been a crucial element to improving our understanding of CRC pathogenesis for treatment and intervention. Efforts to understand genetic mechanisms of CRC predisposition have also highlighted rare, highly penetrant germline variants and their critical role in CRC development, such as mutations in *APC* and in DNA mismatch repair genes *MLH1*, *MSH2*, *MSH6*, *PMS1* and *EPCAM* observed in Familial Adenomatous Polyposis (FAP) and Hereditary Nonpolyposis CRC (HNPCC), respectively[6]. Extension of these analyses to common variants like Single Nucleotide Polymorphisms (SNPs) via Genome-Wide Association Studies (GWAS) have further identified genomic loci associated with increased risk for CRC. Cumulatively, it has been predicted that GWAS-identified risk loci can explain up to 17% of sporadic CRC heritability[7,8]. For many of these risk loci, however, the determination of a causal SNP that drives the risk association, and more importantly, the mechanism by which it functions to progress CRC remains lacking. In order to apply our knowledge of these risk loci to the advancement of cancer precision medicine, it is essential to understand the role each driver variant plays in CRC initiation and progression.

The multicancer risk locus 8q24.21[9–14] represents one such locus for which the definitive role of the candidate risk SNP rs6983267 remains elusive. The locus is home to proto-oncogene

MYC which encodes the c-Myc protein, a key transcriptional regulator that functions downstream of a number of signaling pathways, including WNT signaling, to promote tumor cell proliferation, migration, invasion and metastasis across virtually all cancer types[15–18]. Risk SNP rs6983267 has been shown to reside within an enhancer that interacts with the *MYC* gene promoter via CTCF-mediated chromatin looping[19–22] and is located within a binding site for WNT signaling transcription factor TCF7L2, which binds preferentially to the risk allele[19–21]. At the 8q24.21 locus, β-catenin has been shown to complex with TCF7L2 and bind the enhancer element at rs6983267 for WNT gene activation[21]. The enhancer has been extensively characterized, residing within a super enhancer that spans hundreds of kilobases[23–25]. While *MYC* has been independently validated as a target of this enhancer in a number of studies, the specific effect of rs6983267 genotype status on *MYC* expression, and by extension, its function in CRC pathogenesis, remains inconsistent both *in vitro* and in patient samples[19–21,26–30].

In this study, we present a comprehensive molecular characterization of rs6983267 and its function in CRC. Utilizing scarless CRISPR/Cas9 genome editing[31], we engineered 12 isogenic lines modeling all three rs6983267 genotypes in the CRC cell line HCT-116. Through our functional studies, we observed significant loss of DNA contacts downstream of the 8q24.21 enhancer in the homozygous non-risk genotype, which included the *MYC* promoter. However, *MYC* expression remained stable in the three genotypes. Rather, activation of alternative signaling pathways, such as PI3K and Ras/Raf/MAPK, also known to utilize c-Myc transcriptional activity, was observed with the loss of risk allele-driven WNT enhancer activity at the *MYC* promoter. We also detected activation of lncRNA *CCAT2* in our homozygous non-risk lines, which demonstrated a robust ability to compensate for the absence of DNA WNT enhancer element transcriptional regulatory activity. Extension of this relationship to The Cancer Genome Atlas (TCGA) CRC patient samples revealed a significant correlation between rs6983267 genotype and *CCAT2* expression with genetic alterations in the PI3K pathway, suggesting a requirement of CRC cells to dysregulate growth-promoting pathways beyond WNT

when activity of the 8q24.21 enhancer is low. These results demonstrate that rs6983267 genotype and *CCAT2* expression, together, influence WNT dependence in CRC.

RESULTS

Modeling risk SNP rs6983267 in HCT-116 human colon cancer cells

We demonstrated efficacy of a scarless and selection-free CRISPR/Cas9 editing pipeline previously developed by our group[31] by generating isogenic clones containing a single engineered allele of rs6983267 in the human CRC cell line, HCT-116, which is naturally homozygous risk (G/G) for this SNP (Figure 1a, Supplementary Figure 1a). However, frequency of editing by Homology-Directed Repair (HDR) remained low is these cells with use of a single-stranded donor template. Likely, this was due to the fact that HCT-116 cells cannot efficiently perform single-strand annealing (SSA)-mediated CRISPR/Cas9 editing due to loss of FA pathway function, which is required for single-stranded donor templates[32]. To address this issue, we tested alternative double-stranded donor templates, including a Homologous Recombination (HR) template and a more recently established Homology-Mediated End Joining (HMEJ) template. It has been suggested that the HMEJ donor can increase HDR efficiency of CRISPR/Cas9 editing due to its ability to be utilized within either the HMEJ or HR DNA repair pathways, thus increasing the probability of the donor to be used as a template for DSB repair, even in post-mitotic cells[33]. Indeed, when we assessed HDR and NHEJ editing frequencies via targeted deep Next-Generation Sequencing (NGS), we observed a ten-fold increase in precisely edited alleles when transiently transfecting Cas9/sgRNA with the HMEJ donor versus a single-stranded donor (Supplementary Figure 1b).

Following donor optimization, we performed two rounds of HCT-116 editing and single cell cloning o to produce isogenic clones containing either a single edited non-risk allele (heterozygous) or both edited non-risk alleles (homozygous non-risk) (Supplementary Figure 1c,d). Precisely edited alleles were confirmed by Sanger Sequencing (Supplementary Figure

1e). Positive clones were further validated to be devoid of off-target events by Sanger sequencing of the top five predicted off-target sites based on sgRNA on-target site sequence homology (Supplementary Table 1). After validation, we selected three heterozygous and three homozygous non-risk isogenic clones with no known off-target CRISPR/Cas9 events for downstream functional characterization. For consistency, we also validated three isogenic clones that did not receive any CRISPR/Cas9-mediated genome editing from both rounds of selection for a total of six homozygous risk isogenic clones for comparison. All downstream analyses were performed on these 12 isogenic cell lines.

rs6983267 demonstrates risk allele-specific recruitment of TCF7L2.

We first sought to validate the risk allele binding preference of WNT transcription factor, TCF7L2, in our models. We performed allele-specific chromatin immunoprecipitation utilizing our ChIPnQASO pipeline[31] with our engineered heterozygous lines on TCF7L2 at the rs6983267 binding site (Figure 1b). We also included the nearest CTCF binding site, ~200bp upstream, and active enhancer histone modification H3K27ac[34,35] surrounding the SNP in our analysis (Figure 1b). We observed near complete allelic preference of TCF7L2 binding to the risk allele G over the non-risk allele T (Figure 1c), demonstrating that our engineered endogenous genotype models replicate the effect observed in naturally heterozygous cells and exogenous reporter systems[19,20]. Interestingly, we also observed a much weaker but still significant preference in CTCF binding and H3K27ac modification enrichment for the risk allele as well (Figure 1c). This latter novel finding was intriguing and suggested allele-specific differences in overall enhancer activity, although H3K27ac abundance is not known to affect enhancer activity[36].

To confirm our allelic imbalance results within the heterozygous lines at the genotypic level in our homozygous lines, we performed ChIP-qPCR for the TCF7L2, CTCF and H3K27ac sites in all 12 isogenic lines and measured relative enrichment. We did not observe changes in

overall CTCF binding or H3K27ac enrichment between the different genotypes but observed complete ablation of TCF7L2 binding within the homozygous non-risk lines, recapitulating at the genotypic level, our observations of strong allelic TCF7L2 binding preferences (Figure 1d,e,f). TCF7L2 binding enrichment in the heterozygous lines, however, was not significantly different from that in the homozygous risk lines (Figure 1d). This result suggested that either TCF7L2 binding becomes limited with availability of two risk alleles or that a single risk allele can maximize TCF7L2 recruitment and maintain its overall binding at a level to which HCT-116 cells have adapted. Since the homozygous risk genotype is the natural status of HCT-116 cells, the latter explanation seems more plausible. Our allelic and genotypic findings indicate a mechanism in which rs6983267 risk allele status robustly effects TCF7L2 recruitment to the 8q24.21 enhancer. While TCF7L2 recruitment can be fulfilled by just a single risk allele, it does not significantly affect CTCF recruitment or H3K27ac enrichment to a degree that translates to genotype-level differences (Figure 1b).

rs6983267 drives enhancer interactions with several 8q24.21 genes.

Previously assumed to be a “gene-desert”, the 8q24.21 region has only recently been discovered to be rich in CRC-related noncoding transcripts[37–45]. To better understand the consequences of differential TCF7L2 recruitment to the 8q24.21 enhancer, we investigated changes in DNA interactions with the enhancer region surrounding rs6983267. Previous work has shown that enhancer elements can exert their transcriptional regulatory activity on gene targets via direct contact with the genes’ promoters mediated by CTCF/cohesin chromatin looping[46–48]. Most enhancer-promoter function occurs in *cis* to regulate the transcription of genes located within the same Topologically Associated Domain (TAD) [49–53]. To identify interactions, we performed circularized chromosome conformation capture and high throughput sequencing (4C-seq) on all 12 isogenic lines, using a ~600bp bait sequence centered at rs6983267. Sequencing reads were mapped to human reference genome Hg38 and normalized

using pipe4C [54]. Significant interactions were called using the PeakC, which uses a monotonic regression to calculate a background model from biological replicates for a given experimental condition and calls peaks whose mean enrichment significantly exceed background coverage at a given genomic location followed by Benjamini-Hochberg adjusted FDR[54,55].

Consistent with previous 3C[20,21] and captureHi-C[56] studies, we observed significant interactions in the un-edited homozygous risk lines with the *MYC* promoter (Figure 2a). In addition, we observed downstream interactions with nearby long noncoding RNA (lncRNA) gene *CASC11* and *PVT1* isoform promoter regions, both of which have been implicated in cancer progression and therapeutic resistance[57–67]. We also observed significant upstream interactions at two other 8q24 regions involving promoters of lncRNA genes *CASC19, CCAT1* and *PCAT1* (Figure 2a), which have also been linked to CRC pathogenesis[68–74]. Importantly, both *PVT1* and *PCAT1* have demonstrated the ability to regulate c-Myc translation and protein stability[40,42,66,75]. The cumulative interactions of the 8q24.21 enhancer depict a regulatory element that directly interacts with several CRC-related genes at the region harboring rs6983267, including the proposed *MYC* target.

In our heterozygous lines, we detected a decrease in significant DNA interactions both upstream and downstream of the enhancer region, including loss of contacts with the *MYC/CASC11* shared promoter region (Figure 2a). In homozygous non-risk lines, the interaction with the *MYC/CASC11* region was no longer significant. Instead, we identified a significant gain in upstream interactions with the *PCAT1* promoter region (Figure 2a). Interaction with *CASC19* and *CCAT1* remained stable across all genotypes (Figure 2a). Utilizing an allele-specific variation of 4C-seq[76], we performed an allele-specific assessment of DNA interactions within our heterozygous lines. In agreement with the genotype-level data, we observe a significant interaction for both risk and non-risk alleles at the *CASC19/CCAT1* promoter region. We also observe that only the risk allele contained significant enrichment of contacts downstream of the rs6983267 enhancer, near the *MYC/CASC11* promoter region

(Supplementary Figure 2). This data confirmed the downstream interactions of the enhancer with *MYC* and *CASC11* are specific to the risk allele G.

For further confirmation of our DNA interaction findings, we overlaid our 4C-seq interaction data with ENCODE HCT-116 CTCF, TCF7L2 and H3K27ac ChIP-seq data[77] and with published HCT-116 Hi-C chromatin interaction data[78]. We observed that significant interactions clustered with CTCF ChIP-seq peaks, supporting the involvement of CTCF-mediated chromatin looping to bring these promoters within close proximity of the rs6983267 enhancer (Figure 2b). Furthermore, virtually all significant interactions called across all isogenic lines were located within the rs6983267 TAD (Figure 2c). While TCF7L2 and H3K27ac ChIP-seq peaks were enriched at significant contacts in the *MYC/CASC11* promoter region observed in both the homozygous risk and heterozygous lines (Figure 2, orange highlight), the *PCAT1* promoter region, which was enriched as a DNA contact in the homozygous non-risk lines, was devoid of both (Figure 2, blue highlight). These correlations highlight the importance of the TCF7L2 transcription factor for risk allele-driven enhancer activity while also suggesting an alternative function of rs6983267 at the *PCAT1* promoter that is both risk allele- and TCF7L2-independent.

CCAT2 expression functions as a redundant regulatory mechanism of _MYC_ activation in the absence of the rs6983267 risk allele.

To more thoroughly explore the effect of these genotypic differences in DNA contacts of the rs6983267 enhancer region, we measured the expression of 11 genes (including six of their isoforms) within the rs6983267 TAD using RT-qPCR [79] (Supplementary Table 2). RT-qPCR analyses revealed no significant changes in *MYC* expression across rs6983267 genotypes (Figure 2d). Moreover, most genes maintained consistently low expression levels (Supplementary Figure 3). *CCAT2* was the only rs6983267 TAD gene with dramatic genotype-associated changes across our isogenic lines. While *CCAT2* expression was extremely low in

the homozygous risk and heterozygous lines, we detected an 8-fold upregulation in the homozygous non-risk lines (Figure 2e). As a lncRNA gene that encompasses rs6983267, *CCAT2* has been previously shown to be involved in CRC pathogenesis. Exogenous *CCAT2* overexpression in HCT-116 cells and mouse xenografts increased colony formation, cellular migration, tumor growth and metastatic formation. *CCAT2* overexpression also led to c-Myc upregulation by enhancing TCF7L2 binding and activity at the *MYC* promoter which increased overall WNT signaling[38]. The mechanism of TCF7L2 binding regulation by *CCAT2* is remarkably similar to what we observed at the DNA level for the rs6983267 enhancer. Specifically, both *CCAT2* expression and risk allele presence increase TCF7L2 binding for *MYC* transcriptional activation. The DNA enhancer element achieves this by binding TCF7L2 itself and bringing the *MYC* promoter into close proximity for transcription factor activity while *CCAT2* RNA molecules can achieve this by stabilizing TCF7L2 bound to the *MYC* promoter. Thus, these two mechanisms represent redundant regulatory pathways within CRC cells to maintain c-Myc levels above a required threshold for constitutive cellular growth and proliferative signaling.

To test these hypotheses, we assessed TCF7L2 enrichment across our 12 isogenic lines at two sites upstream of the *MYC* transcriptional start site (TSS). If risk-allele enhancer activity and *CCAT2* expression represented functionally redundant mechanisms for *MYC* transcriptional activation, we would expect similar TCF7L2 enrichment at the *MYC* promoter for both the homozygous risk and homozygous non-risk cell lines. In line with this hypothesis, we did not observe significant differences in TCF7L2 enrichment at either region across our isogenic lines (Figure 2f). Lastly, we assessed proliferation and invasion ability of our isogenic clones. Again, there was no significant difference detected for either assay between genotypes (Supplementary Figure 4). Our results demonstrate that TCF7L2-mediated *MYC* transcriptional activation and downstream cellular effects continue to the same degree despite the complete lack of risk allele-mediated TF recruitment to the enhancer and that this is likely due to increased *CCAT2* expression and function.

PI3K and Ras/Raf/MAPK are activated by loss of rs6983267 risk allele in the absence of *CCAT2* expression.

We were able to demonstrate activation of *CCAT2* transcription in the absence of the rs6983267 risk allele, but presence of at least one risk allele prevented this effect (Figure 2e). Despite this, *MYC* expression as well as cellular proliferation and invasion rates were also stable between the un-edited homozygous risk and heterozygous lines (Supplementary Figure 3). These data correspond with our genotype-level ChIP data demonstrating that just a single risk allele could recruit TCF7L2 to the same degree as two risk alleles in our models, allowing for consistent transcriptional activation of the proposed target gene *MYC*. This, however, does not rule out the possibility of other genes being affected by rs6983267 risk allele status. In fact, transcriptome-wide analysis of an isogenic HCT-116 line containing a homozygous deletion of the 2Kb enhancer region surrounding risk SNP rs6983267 discovered that more than 1000 genes were differentially expressed when compared to wildtype HCT-116 cells[36]. Despite this finding, no other target gene besides *MYC* has been proposed for this regulatory element in CRC. Furthermore, no study, to our knowledge, has dissected the specific targets of this functional regulatory risk SNP from gene targets of the greater enhancer element that encompasses it.

To pursue these questions, we conducted RNA-seq on all 12 isogenic genotype lines and performed differential expression analysis to identify genes affected by rs6983267 genotype status. When compared to un-edited homozygous risk lines, the heterozygous lines showed over 400 genes that were significantly differentially expressed (DE); 268 genes significantly upregulated and 153 genes significantly downregulated (Figure 3a). In stark contrast, differential expression analysis revealed just 13 DE genes between the homozygous non-risk and un-edited homozygous risk genotype lines, 7 upregulated and 6 downregulated (Figure 3b). The 7 upregulated genes included *APOL6*, *CRPPA*, *DBP*, *H19*, *APLF*, *AC009065.3*, and *FHAD1*. The

6 downregulated genes were *LINC00668*, *TENM3*, *GCNT3*, *ZNF615*, *MMP7* and *RNF128*. When hierarchical clustering and PCA were performed, isogenic lines containing the two homozygous genotypes were virtually indistinguishable (Figure 3c, Supplementary Figure 5). Given the heterozygous lines demonstrated 100s of DE genes in our analysis compared to un-edited homozygous risk, this nearly opposite result for the homozygous non-risk lines was striking. This was yet another supporting piece of evidence that *CCAT2* contributes to the compensatory effect we observed in the absence of rs6983267 risk allele-mediated DNA enhancer activity in HCT-116 cells.

We further interrogated the DE gene set from the heterozygous comparison. Pathway enrichment analysis of the 268 significantly upregulated DE genes produced three major functional enrichment clusters: circadian rhythm dysregulation, cell motility and extracellular matrix (ECM) reorganization, and growth factor signaling pathway activation (Figure 3d). Closer investigation revealed activation of both the Mitogen-Activated Protein Kinase Ras-Raf-MEK-ERK (Ras/Raf/MAPK) signaling pathway and Phosphatidylinositol 3-Kinase (PI3K) signaling pathway. Like the WNT signaling pathway, both Ras/Raf/MAPK and PI3K pathways are commonly dysregulated across virtually all cancer types, including CRC[80–85]. When aberrantly activated, both pathways have demonstrated the ability to promote cancer cell survival, proliferation and invasion/migration[86–93], such as through activation of integrin and plasminogen-mediated ECM remodeling[94–100] which were also identified in the pathway enrichment analysis. Importantly, Ras/Raf/MAPK and PI3K have also displayed the ability to activate and stabilize c-Myc expression, respectively, for perpetuation of their pro-proliferation and anti-apoptotic signaling cascades within the cell[87,92,101–103]. It is plausible that upregulation of PI3K and Ras/Raf/MAPK within our heterozygous models contributed to the stabilization of *MYC* expression to compensate for the loss of transcriptional activation of the 8q24.21 WNT enhancer caused by reduced DNA-DNA contacts with the *MYC* promoter for maintenance of overall pro-growth signaling within the cell.

PI3K and Ras/Raf/MAPK activation is a specific effect of rs6983267 risk allele loss within the greater 8q24.21 enhancer.

To validate these transcriptomic effects, we processed RNA-seq data from the previously published HCT-116 enhancer deletion line[36] (Figure 3e) using the same pipeline and cutoff thresholds from our risk SNP analysis. DE analysis found 439 significantly upregulated and 483 significantly downregulated genes. We assessed DE gene overlap between our genotype lines and the homozygous enhancer deletion line. One hundred and forty-eight genes were significantly differentially expressed in both the enhancer deletion and our heterozygous lines compared to un-edited homozygous risk. For our homozygous non-risk lines, there was just a single DE gene that overlapped with the enhancer deletion, *ZNF615*. When we assessed directionality of expression change for the SNP heterozygotes and enhancer deletion common DE genes, either up- or down-regulated compared to homozygous risk, we observed that 128 out of the 148 genes were differentially expressed in the same direction. Of these, 91 were upregulated in both datasets and 37 were downregulated (Figure 3f). Pathway enrichment analysis of these commonly downregulated DE genes showed enrichment for target genes of transcription factors *CEBPA* and *LMO2* (Supplemental Figure 6), both of which have been shown to oppose WNT signaling and exert tumor suppressor activity in CRC[104,105]. The shared downregulated genes also showed enrichment for transcription factor activity and negative regulation of RNA polymerase II (Supplemental Figure 6). On the other hand, the commonly upregulated gene set demonstrated enrichment for the same pathways observed in the independent analysis of our heterozygous genotype lines, namely, plasminogen and integrin activity, extracellular matrix interaction and organization and activation of the Ras/Raf/MAPK and PI3K pathways (Figure 3g). This correlation further indicated that the transcriptional effects we identified, described above to be involved in the inhibition of WNT signaling and activation of PI3K and Ras/Raf/MAPK, were specific effects of the loss of regulatory function exerted by the risk allele of rs6983267 to maintain WNT signaling and WNT-mediated activation of c-Myc.

Furthermore, in the enhancer deletion, this effect could not be rescued by endogenous *CCAT2* upregulation as the deletion fully encompassed the gene (Figure 3e).

While *MYC* expression remained unchanged between our engineered heterozygous lines and the un-edited homozygous risk lines, it is possible that the downstream gene targets of c-Myc transcription factor activity could have changed with *MYC* expression shifting from control by WNT to that of PI3K and Ras/Raf/MAPK, as indicated by the pathway enrichment analyses. To address this, we identified a comprehensive list of validated direct c-Myc targets from the Ingenuity Pathway Analysis software (Qiagen) and identified common c-Myc targets between our heterozygous and homozygous non-risk genotype lines and the enhancer deletion line. Out of 287 possible direct targets, the enhancer deletion showed 37 significantly DE genes (Figure 3h). *HNF4A*, *PRDM1*, *GATA3*, *EBI3*, *ITGA1*, *GSR*, *TERT* and *MYC* itself were the most downregulated, while *AXL*, *AOPEP*, *GFI1*, *EGFR* and *FOSL1* were the most upregulated (Supplementary Table 3). The heterozygotes showed 31 DE direct c-Myc target genes (Figure 3h). Of these, *FBX032, HNF4A, NBN* and *DNMT3B* were most significantly downregulated, while *JUN*, *FOSL1*, *EZR, SLC2A1* and *AXL* demonstrated the greatest upregulation. Overall, 12 direct c-Myc target genes were significantly differentially expressed in both the enhancer deletion and risk SNP heterozygotes when compared to un-edited homozygous risk, with directionality consistent for all 12 common genes (Figure 3h, Supplementary Table 3). Importantly, two of the most upregulated gene targets in both datasets, *FOSL1*, also known as Fra-1, and *AXL*, are both known downstream effectors of the PI3K signaling cascade[106–109]. Likewise, *JUN*, which was significantly upregulated in the heterozygotes, encodes a major downstream transcriptional activator of MAPK signaling, c-Jun[87,110]. Altogether, these results support an activation of PI3K and Ras/Raf/MAPK with the loss of risk allele-mediated enhancer activity in the absence of CCAT2 compensatory activation and function, which we identify in both the heterozygous SNP lines and homozygous enhancer deletion line. Furthermore, we identify 128 genes significantly differentially expressed with loss of either the entire enhancer or

risk allele, of which 12 are verified direct c-Myc transcriptional targets. These 128 common DE genes represent a subset of 8q24.21 WNT enhancer-affected genes that are under the specific control of rs6983267 genotype status in HCT-116 cells and thus warrant further investigation.

rs6983267 genotype status and *CCAT2* expression dictate CRC WNT and PI3K dependence in TCGA

The activation of alternative signaling pathways that target downstream *MYC* expression observed in the heterozygous cell models and the activation of *CCAT2* to stabilize WNT signaling in the homozygous non-risk cells prompted our interest to delve further into the interplay of regulatory elements at the 8q24.21 locus and their relationship to WNT signaling. For this investigation, we utilized the CRC TCGA dataset[111]. Initially, we observed no significant difference in *CCAT2* expression between rs6983267 genotypes (Supplementary Figure 7a). This is consistent with previous findings[38,112,113]. Using the Consensus Molecular Subtype (CMS) classifications for CRC[114], we stratified CRC patient samples by subtype and investigated the correlation with both rs6983267 genotype and *CCAT2* expression. Again, we did not observe any significant correlations with rs6983267 genotypes (Supplementary Figure 7b). However, we did observe significantly higher *CCAT2* expression in the CMS2 subtype (Figure 4a). Since CMS2 represents the canonical WNT-dependent CRC subtype, this is in-line with the hypothesis that *CCAT2* functions as a WNT signaling-specific lncRNA[38]. Interestingly, when we further stratify by risk SNP genotype within each CMS and compare patients of the same genotype across the four CMS groups, we observed that CMS2-homozygous non-risk (TT) patients demonstrated higher *CCAT2* expression levels compared to those of CMS1 (Figure 4a). For both heterozygous (GT) and homozygous risk (GG) groups, *CCAT2* expression within the CSM2 subtype was significantly higher than both CMS1 and CSM3 (Figure 4c,d). These results suggested the rs6983267 risk allele contributes to the increased *CCAT2* expression observed in WNT-dependent CRC tumors.

Based on our findings highlighting the PI3K pathway in rs6983267 function, we utilized TCGA data to examine the correlation between mutations within the WNT and PI3K pathways, *CCAT2* expression and rs6983267 genotype. In TCGA, *APC* and *PIK3CA* are the two most commonly mutated genes in the WNT and PI3K/Akt pathways, respectively (Supplementary Figure 8). When we interrogated the frequency of cases with *APC* mutations and *APC-PIK3CA* co-mutations, chi-squared testing revealed a statistically significant correlation with rs6983267 genotype; CRC patients with the homozygous non-risk genotype displayed greater frequency of harboring somatic mutations in both *APC* and *PIK3CA* ($P = 0.0159$, Table 1). This discovery corroborated our findings *in vitro*, suggesting that rs6983267 genotype status and its effect on the 8q24.21 DNA enhancer's function to perpetuate WNT signaling plays a role in dictating whether a tumor cell requires activation of other pathways, such as PI3K, to achieve a robust proliferative advantage for CRC initiation and progression. We performed the same mutational analysis with *CCAT2* expression, stratifying patients into no, low and high expression groups. Unlike rs6982367 genotype, expression of *CCAT2* alone did not produce a significant correlation with *APC-PIK3CA* co-mutation frequency ($P = 0.2167$, Supplementary Table 4). However, upon assessment of a synergistic effect of both risk allele status and lncRNA expression via further stratification patient samples by risk SNP genotype within each *CCAT2* expression group, we observed a highly significant correlation with APC-PIK3CA co-mutation frequency ($P < 0.0001$, Table 2). While both the no/low a *CCAT2*-expressing CRCs followed the pattern observed at the cohort-wide level, in which the homozygous non-risk patients had the highest frequency of *APC-PIK3CA* co-mutations compared to patients harboring at least one risk allele, the high *CCAT2*-expressing group diverged from this pattern. Within the high-*CCAT2* expressing group, homozygous non-risk patients actually had the lowest frequency of *APC* and *PIK3CA* co-mutations compared to the heterozygous, which instead demonstrated the highest co-mutation frequency, and homozygous risk high-*CCAT2* expressors (Table 2). In fact, the high-*CCAT2* expressing, homozygous non-risk patients showed the lowest co-mutation

frequency across all *CCAT2* expression-rs6983267 genotype combinatorial groups (Table 2). Interestingly, overall, rs6983267 heterozygous patients demonstrated the greatest consistency of *APC* and *PIK3CA* mutation co-occurrence across *CCAT2* expression groups, suggesting that *CCAT2* expression and function is least critical on a rs6983267 heterozygous background (Table 2). Importantly, the no *CCAT2*-expressing, homozygous non-risk patients, in other words, patients with the lowest 8q24.21 WNT enhancer activity, demonstrated the greatest co-mutational burden of *APC* and *PIK3CA* of the entire TCGA CRC cohort (Table 2). Together, these results support a critical role of both rs6983267 genotype status and *CCAT2* expression to promote WNT signaling for CRC initiation and progression. Loss of either enhancing element, particularly rs6983267 risk allele-driven enhancement, results in increased frequency to dysregulate other growth-promoting pathways, such as PI3K/Akt. Our analysis of CRC patients from TCGA identify *CCAT2* expression and rs6983267 genotype as a potential combinatorial biomarker for PI3K mutational burden in canonical WNT-dependent CRC, a pathway for which a number of targeted therapeutics have either been approved for cancer treatment or for which clinical trials are currently ongoing.

DISCUSSION

In this study, we present a comprehensive molecular characterization of risk SNP rs6983267 utilizing scarless CRISPR/Cas9 genome-edited isogenic lines derived from the HCT-116 CRC cell line that precisely model all three genotypes for the SNP. Through interrogation of our models, we demonstrated significant changes in the activity of the 8q24.21 WNT enhancer as a result of genotype status that effect both local gene expression and global transcriptional regulation within CRC cells. We confirmed the mechanism of preferential binding of WNT transcription factor TCF7L2 to the risk (G) allele, which resulted in the absence of TCF7L2 binding to the enhancer region within the homozygous non-risk genotype lines. We discovered that the risk (G) allele drives 3D interactions with the *MYC*, *CASC11* and *PVT1* promoter

regions located downstream of the enhancer element through TCF7L2 recruitment and CTCF-mediated looping. The non-risk allele was observed to drive upstream enhancer DNA interactions, such as that with *PCAT1*, and absence of the risk allele resulted in complete loss of interactions downstream of the enhancer that include the *MYC* promoter. Despite these significant changes, there was no measurable difference in *MYC* expression across rs6983267 genotypes. Rather, we observed activation of two compensatory mechanisms for c-Myc regulation: increased expression of lncRNA CCAT2 in the homozygous non-risk genotype and activation of the Ras/Raf/MAPK and PI3K pathways. We confirmed this effect in an HCT-116 line harboring a 2Kb homozygous deletion of the 8q24.21 enhancer element. Our data depicts a model in which two elements, rs6983267 risk allele and *CCAT2* expression, function to maintain activity of a WNT enhancer. Loss of both DNA- and RNA-driven enhancer element activity requires the tumor cells to activate alternative growth-signaling pathways, such as PI3K and Ras/Raf/MAPK. In TCGA data, both rs6983267 risk allele and *CCAT2* expression correlated with *PIK3CA* mutational burden in *APC* mutant cases. Being that the PI3K pathway is a highly druggable pathway in cancer, this discovery merits further investigation into the application of a *CCAT2* expression-rs6983267 genotype combinatorial biomarker for the presence of PI3K-activating mutations in canonical WNT-driven CRC.

The transcription factor c-Myc is a master regulator of cellular proliferation, anti-apoptosis and angiogenesis that nearly all cancer types exploit to some degree in order to survive and gain advantage over neighboring cells[15–18]. Because of its central role in regulation of multiple signaling pathways, it is no surprise that c-Myc would have a number of regulatory mechanisms controlling its mRNA- and protein-level expression. In fact, it has been reported that in addition to this risk SNP enhancer, the *MYC* gene has 9 other enhancer elements within the 8q24 region alone[115]. In light of this, activity loss of just a single enhancer may not affect *MYC* expression to a degree that cannot be accounted for by other regulatory mechanisms, as we have observed in our CRC cell models. Additionally, crosstalk between critical signaling

pathways in response to environmental or intracellular change is not a novel observation, especially in cancer cells, where adaptation and pathway redundancy have become defining characteristics of drug resistance and disease relapse[93]. In CRC specifically, one mechanism of resistance to PI3K targeted therapies is the overactivation of WNT signaling[92,116,117]. The reciprocal may be viable as well, in which inhibition of WNT signaling via reduced efficacy of a *MYC*-targeting WNT enhancer, would result in the overactivation of PI3K, for which *MYC* is also a target. In fact, it has been shown that transcription factor activity of nuclear β-Catenin requires PI3K signaling to exert maximal activation of WNT target genes and inhibition of PI3K signaling reduced expression of WNT genes *MYC*, *LEF* and *CCND1* without affecting nuclear β-Catenin levels. Because of this, high nuclear β-Catenin levels combined with high PI3K signaling levels proved to be a reliable biomarker for CRC distal metastasis formation[118]. Here, our results also support the synergistic relationship of WNT signaling with PI3K and MAPK in the maintenance and fine-tuning of c-Myc activity for CRC progression.

We and others have demonstrated the value of precise disease-associated variant modeling in established cancer cell lines. Indeed, these studies have helped to uncover specific molecular functions for many genomic variants. However, many of these functions remain validated solely on cancerous backgrounds. Because of this, direct translation of these findings to mechanisms of patient risk represent a critical limitation of established cancer cell line models. To take our study as an example, we observed activation of the lncRNA CCAT2 in response to loss of function of the 8q24.21 DNA enhancer element with the absence of rs6983267 risk alleles. A direct correlation of these findings to patient data would support the hypothesis that homozygous non-risk CRC patients demonstrate the highest expression of *CCAT2*. However, as we have reported, this is not the case. Rather, *CCAT2* expression and rs6983267 genotype status appear to function synergistically to promote WNT signaling in tumor cells, with the greatest effect observed in CRC cases that develop due to aberrant canonical WNT signaling as defined by the CMS2 subtype. These disconnects can perhaps be

attributed to the modeling of risk-associated variants, which seemingly function at the initiation of tumorigenesis, in already transformed cellular models. The emergence of more translational *in vitro* models may provide a resolution to this issue, such as patient-derived organoids (PDOs), which can be derived from both normal and cancerous tissue and have proven to more reliably recapitulate the characteristics of the original tumor. Despite these challenges, the field has gained immense knowledge from established cancer cell lines. In this study specifically, while we could not directly translate all *in vitro* risk SNP-mediated alterations to CRC patient data, our experiments highlighted the significance of PI3K and Ras/Raf/MAPK pathways in rs6983267's WNT signaling function, which we used to direct our analyses in TCGA CRC patient datasets. In this way, established cancer cell lines can, and have, provided critical information for variant function in cancer pathogenesis. Importantly, established cell lines also remain significantly more tolerant to current methods of genomic manipulation. The development of more efficient methods of genomic editing in PDOs and other normal *in vitro* models will provide an important component for characterization and translational application of these cancer risk-associated variants to patient care.

Ultimately, our study illustrates a complex and tightly regulated network of DNA elements, lncRNA and signaling cascades that control c-Myc expression and activity. The balance between WNT, Ras/Raf/MAPK and PI3K signaling pathways provide cancer cells with the ability to adapt to changes, both environmental and internal, in order to maintain selective advantages throughout tumorigenesis. The interplay of WNT, Ras/Raf/MAPK and PI3K activation in response to targeted therapies as a mechanism of resistance, for example, has proven to be a difficult hurdle in the treatment of cancer. This study helps to further illuminate the role that the 8q24.21 *MYC* enhancer, and more precisely, the role that functional CRC risk SNP rs6983267 plays in this intricate network. It is our goal that these findings can contribute to the advancement of precision cancer medicine and improve our ability to predict therapeutic efficacy at the level of the individual.

CONCLUSIONS

Most cancer risk associated SNPs identified by GWAS remain clinically unactionable despite mechanisms of function having been characterized for many of these variants. One limitation has been elucidating the role of SNP function within the development of its associated cancer. Here, we demonstrate application of a known CRC functional SNP, rs6983267, and expression of a related long non-coding RNA, CCAT2, to predict dysregulation of the PI3K pathway in CRC through control of a WNT-specific enhancer. The PI3K pathway is highly targetable in cancer, with a number of targeted therapeutics either approved or in clinical trials. Our study reveals how risk SNP rs6983267 may be used to inform CRC patient treatment and enhance our efforts towards precision medicine.

METHODS

CRISPR/Cas9 HDR design and cloning. Risk SNP rs6983267-targeting guide RNA was designed, cloned into the gRNA-Cloning Vector plasmid from the Church Lab (Addgene Plasmid #41824) via Gibson Assembly (NEB) and tested as previously described[31]. ***Symmetrical single stranded donor template*** (ssSYM) was designed centered around risk SNP rs6983267 and flanked by 35 bases of homology to the non-gRNA-complementary strand of HCT-116 genomic DNA and synthesized by IDT. ***Asymmetrical single stranded donor template*** (ssASYM) was designed as previously described and synthesized by IDT[119]. ***Homologous Recombination double stranded donor template*** (dsHR) was designed centered around risk SNP rs6983267 flanked by 800 bases of homology to the HCT-116 genomic sequence and synthesized by IDT as a gBlock. The gBlock was cloned into the pCR4-TOPO vector backbone (Thermo Fisher Scientific) using the TopoTA Clonign Kit for Sequencing (Thermo Fisher Scientific). ***Homology-Mediated End Joining double stranded donor template*** (dsHMEJ) was designed using the dsHR donor. Primers were designed to include the risk SNP-targeting gRNA sequence with

PAM sites oriented towards the 3' end followed by 20 bases of homology with the 5' and 3' ends of the dsHR homology arms (dsHMEJ_Cloning_Fwd, dsHMEJ_Cloning_Rev). 10ng of purified dsHR plasmid donor was amplified with dsHMEJ_Cloning primers and Phusion High-Fidelity PCR Master Mix (NEB) in a 25uL reaction. PCR products were purified with QIAquick PCR Purification kit (Qiagen). 50ng of purified PCR product was incubated with 10uL GoTaq Green Master Mix (Promega) in a 20uL reaction at 70C for 30min for 3'-A tailing. PCR products were again purified with QIAquick PCR Purification kit (Qiagen). 4uL of 3'-A-tailed PCR product (15-25ng total) was cloned into the pCR4-TOPO vector backbone (Thermo Fisher Scientific) using the TopoTA Clonign Kit for Sequencing (Thermo Fisher Scientific) following manufacturer's protocol with one adjustment: RT incubation of vector and PCR insert was increased to 10min. gRNA plasmid, hCas9 vector (Church Lab, Addgene Plasmid #41815) and double stranded donor templates (dsHR, dsHMEJ) were transformed into NEB 5-alpha competent E. coli cells (high efficiency). Plasmids were purified with the Plasmid *Plus* Midi Kit (Qiagen).

KASP primer design. KASP primers can be designed using the Primerpicker software (KBiosciences) or custom ordered directly from the LGC Group (https://www.lgcgroup.com/products/kasp-genotyping-chemistry/). The resulting primers sequences, two allele-specific primers of 40-50bp length named A1 and A2 and two universal primers of 20-30bp length named C1 and C2 are reconstituted at a 100uM concentration for use.

Cell culture and transfections. All human cell lines were obtained from ATCC and cultured according to ATCC guidelines. HCT-116 cells were maintained in McCoy's 5A Medium supplemented with 10% Fetal Bovine Serum and 1X Penicillin/Streptomycin. Transfections were carried out in 6-well plates. HCT-116 cells were seeded at a concentration of 500,000 cells per well. When the cells reached 70% confluency approximately 24 hours later, the cells were transfected with 7.5µL of Lipofectamine 3000 (Thermo Fisher Scientific) and either ***Cas9-NHEJ***

for T7 Assay*:** 1ug of hCas9 plasmid, 1ug of guide RNA vector plasmid and 1ug of eGFP plasmid (Lonza) per well or ***Cas9-HDR for cloning: 1ug of hCas9 plasmid, 1ug of guide RNA vector plasmid and 800ng of ssODN donor template or 1ug of plasmid donor template per well. Cells were incubated at 37C with 5% CO_2 for 72 hours, with media replaced every 24 hours for HCT-116 cells.

Amplicon sequencing and allele modification analysis. HCT-116 cells were transfected with Lipofectamine 3000, plasmid Cas9, plasmid gRNA and ssODN or plasmid donor template. Non-transfected HCT-116 cells were used as a control. 72 hours later, genomic DNA was extracted using Quick-gDNA MiniPrep Kit (Genesee Scientific) and 100ng genomic DNA was used for PCR amplification with GoTaq Green Master Mix (Promega) resulting in a 217-bp amplicon. Each forward primer contained a unique 5-bp bar code sequence at the 5' end for multiplexing. All amplicons were purified using QIAQuick PCR Purification Kit (Qiagen) and pooled at equal concentrations for Illumina sequencing. Amplicon sequencing was performed by the CCIB DNA Core Facility at Massachusetts General Hospital (Cambridge, MA) following their instructions for CRISPR Sequencing. Sequencing data was processed using FLASH2 to overlap forward and reverse reads into a single long read. Using FASTX barcode splitter, single long reads were demultiplexed by identifying barcodes at the beginning or end of the sequence read, allowing for one mismatch. Processed fastq files were analyzed for NHEJ frequency with the CRISPResso online tool (http://crispresso.rocks) using default settings. For increased stringency, HDR-positive reads were identified from fastq files using the grep function to select reads with 100% homology to the expected HDR-positive sequence after 20-bp trimming on either side of the read.

Single-cell limiting dilutions for 96-well plate seeding and colony expansion. Transfected cells were resuspended to a single-cell suspension and counted using a hemocytometer.

Seeding 2 to 4 96-well plates per gRNA targeting a specific mutation was sufficient to produce enough clones for screening and identifying HDR-positive clones. For each 96-well plate, 100 cells were added to 10mL of fresh medium and distributed at a 100µL volume per well using a multichannel pipette. 1-2 days later, wells were checked for the presence of a single cell. 5-7 days later, 100µL of fresh media was added to each well. Another 5-7 days later, media was changed for all wells. Cell colonies were transferred to 24-well plates at 50% confluency.

Genomic DNA extraction and KASP genotyping of clones. Cell clones were harvested for genomic DNA when they reached 80% confluency in 24-well plates. For this, each clone was dissociated with 100µL of TypLE Express (Thermo Fisher Scientific) and 30µL of cells were transferred to a new 24-well plate containing fresh 500uL pre-warmed media per well, while 70µL were pelleted for genomic DNA extraction. Extraction was performed using the Quick-gDNA MiniPrep Kit (Genesee Scientific) eluted into 40µL Elution Buffer. Genomic DNA was quantified using Qubit 2.0 Fluorometer (Thermo Fisher Scientific) producing concentrations between 1-20ng/µL. KASP genotyping reactions (LGC Group) were performed in a 384-well plate as follows: 10ng genomic DNA, 5µL 2x KASP Master Mix, 0.14µL Assay (A1, A2, C1 primers) Mix, H_2O up to 10µL total reaction volume. Genotyping plates were run on CFX384 Touch Real-Time PCR Detection System (Biorad) using KASP manufacturer's settings and cluster analysis visualized with the Biorad CFX Manager 3.1 Allelic Discrimination Viewer.

Off-target analysis of isogenic clones produced by CRISPR-HDR. Putative off-target (OT) sites were identified by consensus of output of three online software designed to identify OTs based on sequence homology: CHOPCHOP (https://chopchop.cbu.uib.no), Cas-OFFinder (http://www.rgenome.net/cas-offinder/) and CC Top (https://crispr.cos.uni-heidelberg.de). Primers were designed using Primer3 software to produce a 500-600bp region around OT site. PCR amplification of 50ng of genomic DNA from each clone with GoTaq Green Mastermix

(Promega) was performed following manufacturer's instructions and PCR products visualized on a 1% agarose gel for correct size and specificity. PCR products were then purified with QIAQuick PCR Purification kit (Qiagen) and submitted to the UC Davis DNA Sequencing Core for Sanger sequencing. Sanger sequencing results were analyzed with MacVector (MacVector Inc.) against Sanger sequencing results of genomic DNA from the non-treated parental cell line for presence of sequence mismatches within 100bp of the OT site, indicative of NHEJ events.

Chromatin immunoprecipitation real-time qPCR and quantitative allele-specific occupation (ChIPnQASO) assays. ChIP was performed as previously described with modifications[120]. In summary, 10 million cells in a 10cm dish were cross-linked with a 10min incubation in 1% Formaldehyde. Cross-linking was quenched by 0.125M Glycine incubated for 5min. Cells were scraped, pelleted and lysed via 15min incubation in Cell Lysis Buffer (5mM PIPES, 85mM KCl, 1% Igepal, 1x Protease Inhibitors (Pierce Protease Inhibitor Mini Tablets, Thermo Fisher Scientific); pH 8.0) on ice. Nuclei were pelleted and lysed via 30min incubation in Nuclear Lysis Buffer (50mM Tris-HCl, 10mM EDTA, 1% SDS; pH 8.0) on ice. DNA was sheared to 100-500bp length using the Covaris E220 Focused-ultrasonicator using the following settings: 1mL volume, PIP 140, Duty 5%, CPB 200, Time 240sec. For each ChIP, 3 volumes of IP Dilution Buffer (16.7mM Tris-HCl, 167mM NaCl, 1.2mM EDTA, 1.1% Triton X 100, 0.01% SDS, 1x Protease Inhibitors; pH 8.0) was added to 15ug of chromatin and 2ug of antibody (TCF7L2 (D31H2 Cell Signaling Technology); CTCF (C48H11 Cell Signaling Technology); H3K27ac (39133 Active Motif)). Chromatin was incubated with antibody at 4C overnight on a rotating platform. Staph A cells were added to chromatin and rotated at room temp for 15min. Chromatin-antibody complexes were pelleted and washed twice with Wash Buffer 1 (50mM Tris-HCl, 2mM EDTA, 0.2% Sarkosyl; pH 8.0) and four times with Wash Buffer 2 (100mM Tri-HCl, 500mM LiCl, 1% Igepal, 1% Deoxycholic Acid; pH 8.0). Immunoprecipitated chromatin was eluted in ChIP Elution Buffer (50mM $NaHCO_3$, 1% SDS) via 15min shaking at RT. Crosslinks

were reversed at 67C overnight with 10% 5M NaCl. Finally, immunoprecipitated DNA was treated with 10ug RNase, incubated at 37C for 20min, and purified using QIAQuick PCR Purification Kit (Qiagen). ***Real-time qPCRs*** were performed in 384-well plate as follows: 1.5uL Immunoprecipitated DNA or 1:50 diluted Input DNA, 6.5uL 2x IQ Syber Green Supermix (Bio-Rad), 1.0uL each of 5uM forward and reverse primers, H^2O up to 13µL total reaction volume. Each condition was run in technical triplicates. ***KASP ChIPnQASO reactions*** were performed in 384-well plate as follows: 1µL Immunoprecipitated or Input DNA, 5µL 2x KASP Master Mix, 0.14µL Assay (A1, A2, C1 primers) Mix, H^2O up to 10µL total reaction volume. Genotyping plates were run on CFX384 Touch Real-Time PCR Detection System (Biorad) using KASP manufacturer's settings and cluster analysis visualized with the Biorad CFX Manager 3.1 Allelic Discrimination Viewer. Allelic ratios were calculated by setting the Input DNA RFU values to 50% allelic percentages and using the conversion factor to convert the Immunoprecipitated DNA RFU values to allelic percentages relative to input DNA. All ChIPnQASO reactions were performed in biological triplicate and standard deviation was calculated for each run.

Real-time Reverse Transcription-qPCR (RT-qPCR) and analysis. Cells for each genotype clone were seeded in 6-well plates and allowed to grow until 60% confluent. RNA was extracted and isolated using RNeasy Plus Mini kit (Qiagen) and Qiashredder (Qiagen) following manufacturer's instructions. RNA was quantified with Nanodrop and 2ug RNA reverse transcribed using Superscript IV VILO Master Mix (Thermo Fisher Scientific) following manufacturer's instructions. RT-qPCR reactions were performed in 384-well plate as follows: 4.5 uL cDNA sample, 6.5 uL of IQ Syber Supermix (BioRad) and 1.0 uL 5uM primers each were added to each well of 384-well plate. All cDNA samples were diluted 1:300 prior to loading. All samples were run in technical triplicate. RT-qPCR reactions were run on CFX384 Touch Real-Time PCR Detection System (Biorad) using IQ Syber manufacturer's settings. Relative expression for each sample was calculated using the delta-delta-Ct method. Genotype clones

were averaged, and standard deviation calculated for plotting. Statistical significance was calculated via one-way ANOVA followed by post hoc Tukey HSD analysis.

4C-Sequencing and analysis. 4C libraries were generated following previously published protocol[121] using restriction enzymes DpnII and CviQI with the following modifications: *(Cross-Linking)* 10 million cells in a 10cm dish were cross-linked with a 10min incubation in 1% Formaldehyde. Cross-linking was quenched by 0.125M Glycine incubated for 5min. Cells were scraped and pelleted for lysis. *(Sequencing)* In order to conduct allele-specific 4C-seq analyses in the heterozygous clones, libraries were prepared to be sequenced paired-end with a genotyping primer and a sequencing primer as previously described[76]. Libraries were QCed with Bioanalyzer High Sensitivity DNA kit (Agilent) and quantified with Qubit dsDNA High Sensitivity assay kit (Thermo Fisher) then pooled at equimolar ratios. Due to low complexity, 4C-Seq library pool was further pooled 1:1 with unrelated RNA-Seq libraries and sequenced on two lanes of PE150 sequencing on a HiSeq4000. Sequencing data for each genotype clone was combined into a single FASTQ file for analysis. Libraries were analyzed following previously published protocol[54] using the pipe4C[54] pipeline in R developed by the de Laat lab (https://github.com/deLaatLab/pipe4C) and peakC[55] significant cis-interaction peak calling software developed by the de Wit Lab (https://github.com/deWitLab/peakC). Raw FASTQ files are deposited on NCBI GEO.

RNA-Sequencing and analysis. Cells for each genotype clone were seeded in 6-well plates and allowed to grow until 60% confluent. RNA was extracted and isolated using RNeasy Plus Mini kit (Qiagen) and Qiashredder (Qiagen) following manufacturer's instructions. RNA libraries were generated using the NEBNext Ultra Directional RNA Library Prep kit (NEB) following manufacturer's instructions. 1ug of RNA was used for each library. Libraries were QCed with Bioanalyzer High Sensitivity DNA kit (Agilent) and quantified with Qubit dsDNA High Sensitivity

assay kit (Thermo Fisher) then pooled at equimolar ratios for a single lane of PE150 sequencing on a HiSeq4000. Sequencing reads were demultiplexed and index/adapter trimmed and aligned to the Hg38 reference genome with STAR Universal Aligner version 2.5.3a using the following settings: ***Indexed reference genome:*** Ensembl reference genome and annotation files for Hg38 release 77 were downloaded and complied, Genome was indexed using the following code:

> "STAR --runMode genomeGenerate --runThreadN 12 –genomeDir /STAR_INDEX_HG38 --genomeFastaFiles GRCh38_r77.all.fa --sjdbGTFfile Homo_sapiens.GRCh38.77.gtf --sjdbOverhang 149"

Sample read alignment: alignment of sample reads was performed with the following code:

> "STAR --runThreadN 24 --genomeDir /STAR_INDEX_HG38 --outFileNamePrefix /STAR/SampleName_ --outSAMtype BAM SortedByCoordinate --outWigType bedGraph --quantMode TranscriptomeSAM GeneCounts --readFilesCommand zcat --readFilesIn Sample-Read1.fastq.gz Sample-R2.fastq.gz"

Differential Expression (DE) analysis was performed with Limma-Voom (EdgeR) software in R Studio. Gene count files were combined into a single file and non-expressed and lowly-expressed genes were removed from the analysis by removing genes with gene counts in less than 3 of the 24 samples using the "filterByExpr" function. Normalization and DE analysis were performed using default Limma-Voom settings followed by empirical Bayes moderation of the standard errors towards a common value using the "eBayes" function. DE gene lists from pairwise comparisons were exported into .csv files and utilized for GO term analysis using DAVID (https://david.ncifcrf.gov). ***Tak et al. (2015) dataset analysis:*** FASTQ files containing sequencing data of homozygous enhancer deletion line (SRR2242932.fq, SRR2242933.fq, SRR2242934.fq) and wildtype control HCT116 lines (SRR2242920.fq, SRR2242921.fq, SRR2242922.fq, SRR2242923.fq, SRR2242924.fq, SRR2242925.fq, SRR2242926.fq, SRR2242927.fq, SRR2242928.fq, SRR2242929.fq, SRR2242930.fq, SRR2242931.fq) were

downloaded from NCBI Sequence Read Archive using sratoolkit. Indexed reference genome was generated using "--sjdbOverhang 74" since the FASTQ files were produced from a SE75 sequencing run. For all analyses, significantly differentially expressed genes were identified using log2 fold change less than -1.0 or greater than 1.0 and FDR adjusted p-value less than 0.05 cutoffs. Gene count files and raw FASTQ files are deposited on NCBI GEO.

Cellular proliferation (MTS) assay. Cells from each genotype clone were seeded in 96-well plates in technical triplicate at a confluency of 50,000 cells in 200uL of fresh medium per well. At each timepoint (24hr, 48hr, 72hr post-seeding) 20uL of Cell Titer 96 Aqueous One Solution (Promega) was added to each well and cells were incubated for 2hr at 37C. After incubation, 96-well plates were measured for absorbance at 490nm on Tecan M1000 Pro plate reader. All absorbances were normalized against the background absorbance of wells without cells. Absorbances for technical replicates for each clone were averaged. Clones for each genotype were further averaged and standard deviation calculated for plotting. Statistical significance was calculated via one-way ANOVA followed by post hoc Tukey HSD analysis.

Cellular Invasion assay. Cells from each genotype clone were seeded in QCM ECMatrix Cell Invasion Assay 96-well plates (8um, fluorometric) (Milipore-Sigma) in technical triplicate at a confluency of 50,000 cell in 100uL of serum-free medium (supplemented with 5% BSA) per well. Receiving wells were filled with 150uL of full medium (+ 10% FBS). As a control, all clone cells were identically seeded with receiving wells that contained serum-free medium. Cells were incubated for 72hr and invasion measured following manufacturer's instructions using a Tecan M1000 Pro plate reader. All absorbances were normalized against the background absorbance of wells without cells. Absorbances for technical replicates for each clone were averaged. Clones for each genotype were further averaged and standard deviation calculated for plotting.

Statistical significance was calculated via one-way ANOVA followed by post hoc Tukey HSD analysis.

TCGA CRC dataset analysis. ***SNP rs6983267 genotyping:*** Simple nucleotide variation (*.birdseed.data.txt) files from Affymetrix Genomewide SNP6 genotyping arrays and related metadata was downloaded from GDC Legacy Archive[122] for all COAD individuals (n=461). Genotypes were extracted from the birdseed files and compiled into PLINK PED/MAP files utilizing annotation information provided in GenomeWideSNP_6.na35.annot.db sqlite database and internal scripts. For each individual, normal genotyping was extracted and only one normal sample was selected for each individual using PLINK (v.1.9)[123]. For individuals where no Blood Derived Normal sample was available, Solid Tissue Normal sample was used. SNP_A-8283428 probeset for rs6983267 was extracted for all COAD individuals with normal genotyping data (n=417) using PLINK. ***CCAT2 expression:*** RNA-Seq HT-Seq gene count files were downloaded from GDC Legacy Archive[122] for all COAD individuals with rs6983267 genotyping data (n=352) and *CCAT2* expression extracted and converted to log counts per million (logCPM) expression. Cases were binned into *CCAT2* expression quartiles based on logCPM expression values (no expression n=111, low expression n=65, moderate expression n=88, high expression n=88). ***Consensus Molecular Subtyping:*** CRC CMS assignments were obtained from the CRC Subtyping Consortium (https://www.synapse.org/#!Synapse:syn4978511) based on previously published CMS for TCGA COAD samples (CMS1 n=71, CMS2 n=135, CMS3 n=46, CMS4 n=100)[114]. ***APC and PIK3CA somatic mutation analysis:*** Somatic mutation data for *APC* and *PIK3CA* for all COAD individuals with normal genotyping (n=413) and *CCAT2* expression (n=352) were obtained from GDC Data Portal (https://portal.gdc.cancer.gov) using UUID. Statistical significance of *CCAT2* expression across genotypes was calculated via one-way ANOVA followed by post hoc Tukey HSD analysis. Statistical significance of somatic mutation frequencies was calculated via Chi-squared test.

AVAILABILITY OF DATA AND MATERIALS

HCT-116 SNP genotype clones 4C-Seq FASTQ files are available in NCBI GEO (Accession numbers TBD). HCT-116 SNP genotype clones 4C-Seq differential interaction peak calls are available in the supplementary file “Coggins2021_4CseqPeakCSignificantInteractions-SuppTable”.

All primer and oligo sequences utilized in this study are available in the supplementary file “Coggins2021_PrimersOligosList-SuppTable2”.

HCT-116 SNP genotype clones RNA-Seq FASTQ files and raw gene count files are available in NCBI GEO (Accession numbers TBD). HCT-116 SNP genotype clones RNA-Seq differentially expression analysis log2 fold change and statistical values are available in supplementary file “Coggins2021_RNAseqExpressionMYCTargets-SuppTable3”. Qiagen IPA curated c-Myc direct target gene list is available in supplementary file “Coggins2021_ RNAseqExpressionMYCTargets-SuppTable3”.

HCT-116 homozygous enhancer deletion RNA-Seq FASTQ files from Tak et al. were obtained from NCBI Sequence Read Archive: homozygous enhancer deletion HCT-116 lines (SRR2242932.fq, SRR2242933.fq, SRR2242934.fq); wildtype control HCT-116 lines (SRR2242920.fq, SRR2242921.fq, SRR2242922.fq, SRR2242923.fq, SRR2242924.fq, SRR2242925.fq, SRR2242926.fq, SRR2242927.fq, SRR2242928.fq, SRR2242929.fq, SRR2242930.fq, SRR2242931.fq).

MAIN FIGURE LEGENDS

Figure 1| Preferential binding of transcription factor TCF7L2 to the rs6983267 risk allele and its effects on local enhancer function. (a) Scarless genome editing and clonal selection pipeline producing 3 homozygous reference, 3 heterozygous and 6 un-edited homozygous risk

clonal lines derived from HCT-116 modeling rs6983267 genotypes. **(b)** Graphical representation of enhancer region encompassing risk SNP rs6983267. TCF7L2 transcription factor (red) binding site, adjacent CTCF insulator (purple) binding site, histone mark H3K27ac (green) associated with active enhancers surrounding risk SNP. **(c)** Chromatin Immunoprecipitation and Quantitative Allele-Specific Occupancy (ChIPnQASO) in heterozygous lines (n=3) shows strong preference of TCF7L2 binding to the risk allele (orange bars) over the non-risk allele (blue bars) and slight preference for adjacent CTCF binding and surrounding H3K27ac marks. Asterisks (*) indicate significantly different groups by Student's t-test (*) p-value < 0.05, (***) p-value < 0.001. Error bars show standard deviation. (d,e,f) Chromatin Immunoprecipitation comparing overall binding of TCF7L2 (red), CTCF (purple), H3K27ac (green) in across all genotypes (n=6 for GG, n=3 for GT and TT) shows ablation of TCF7L2 binding in the absence of the risk (G) allele. Grey bars indicate negative control IgG. Asterisks (*) indicate significantly different groups by One-Way ANOVA followed by Post-Hoc Tukey HSD. (*) p-value < 0.05. Error bars show standard deviation.

Figure 2| Local regulatory interactions for risk SNP rs6983267 in HCT-116 genotype clones. (Top) Genes annotated within +/- 1Mb of SNP. (a) 4C-Seq DNA interactions (significant interactions in red called by PeakC software) generated for all genotypes (n=6 for GG, n=3 for GT and TT) of rs6983267 show differential interactions. Loss of the risk (G) allele shows loss of interactions with the *MYC* promoter (highlighted in orange) and gain of interactions near the *PCAT1* promoter (highlighted in blue). Bait indicates region encompassing SNP used for interaction detection. **(b)** ENCODE ChIP-Seq signal tracks for CTCF, TCF7L2 and H3K27ac histone marks in HCT-116 cell line. **(c)** Hi-C chromatin contacts at 5Kb resolution from Rao et al. in HCT-116 cell line. Visualized using Juicebox.js software. TAD containing rs6983267 and its interactors indicated by dotted line outlining enrichment of contacts within the region. **(d,e)** Relative RNA expression of *MYC* and lncRNA *CCAT2* across genotypes (n=6 for

GG, n=3 for GT and TT) measured by RT-qPCR. Asterisks (*) indicate significantly different groups by One-Way ANOVA followed by Post-Hoc Tukey HSD. (**) p-value < 0.01, (***) p-value <0.001. Error bars show standard deviation. Error bars show standard deviation. **(f)** Chromatin Immunoprecipitation comparing overall binding of TCF7L2 across all genotypes (n=6 for GG, n=3 for GT and TT) shows no statistically significant difference in TCF7L2 enrichment within the *MYC* promoter at multiple binding sites (+1900bp, +1200bp) identified by ENCODE ChIP-seq upstream of the *MYC* TSS across genotypes. Grey bars indicate negative control IgG. Groups were assessed with One-Way ANOVA followed by Post-Hoc Tukey HSD. Error bars show standard deviation.

Figure 3| SNP status of rs6983267 affects a subset of enhancer gene that function in the PI3K and Ras/Raf/MAPK pathways in HCT-116 cells. (a,b) Volcano plots showing significantly differentially expressed genes (DEGs) across rs6983267 genotype clones. Significantly downregulated (blue dots) and upregulated (red dots) genes were selected using log2 fold change <-1.0 or >1.0 and FDR adjusted p-value < 0.05 cut-off values for heterozygous **(a)** and homozygous non-risk (T/T) **(b)** compared to parental homozygous risk (G/G). **(c)** Unsupervised hierarchical clustering heatmap of top 500 DEGs shows clustering of heterozygous clones (n=3) away from both homozygous risk (n=3) and homozygous non-risk clones (n=3). **(d)** DAVID Gene Ontology (GO) analysis of significantly upregulated DEGs in heterozygous compared to parental homozygous risk (G/G) shows enrichment for pathways involved in extracellular matrix remodeling and cell motility and migration (orange bar terms) as well as enrichment for aberrantly activated cell growth and proliferation pathways PI3K and Ras/Raf/MAPK (red bar terms). **(e)** UCSC Genome Browser shot of Enhancer Deletion in HCT-116 cells by Tak et al. (black bar, middle), SNP rs6983267 (light blue vertical line) and ENCODE ChIP-Seq signal tracks for H3K27ac, TCF7L2 and CTCF. **(f)** Heatmap of 148 common DEGs between heterozygotes and enhancer deletion (upregulated in blue, downregulated in red)

indicate putative target genes of enhancer SNP regulation. **(g)** DAVID GO analysis of commonly upregulated genes from 5d shows enrichment for same pathways involved in extracellular matrix remodeling and cell motility and migration (orange bar terms) as well as enrichment for aberrantly activated cell growth and proliferation pathways such as PI3K, Ras and MAPK (red bar terms) seen for Heterozygous enrichment analysis alone. **(h)** Direct c-Myc targets identified by IPA (Qiagen) significantly DE in enhancer deletion and heterozygotes show putative target genes of enhancer SNP regulation within the c-Myc interactome.

Figure 4| rs6983267 risk allele correlates with higher CCAT2 in CRC tumors with aberrant canonical WNT signaling and affect *APC* and *PIK3CA* co-mutations in TCGA colorectal adenocarcinoma (COAD) patient cohort. (a) Stratification of COAD patients by consensus molecular subtype (CMS) shows significantly higher CCAT2 expression overall within the CMS2 subtype. F (3,348) = 17.82. **(b, c, d)** Further stratification by rs6983267 genotype (TT, F (3,60) = 4.155; GT, F (3,151) = 7.907, GG, F (3,126) = 6.056) shows this effect is strengthened with the presence of the risk allele (G) as both heterozygous (GT) and homozygous risk (GG) groups show significant increase in CCAT2 expression within the CMS2 subtype compared to CMS1 and CMS3. Asterisks (*) indicate significantly different groups by One-Way ANOVA followed by Post-Hoc Tukey HSD. (*) p-value < 0.05, (**) p-value < 0.01, (****) p-value <0.0001. Error bars show standard deviation.

Acknowledgments.

We are grateful to the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786-01, for performing all HiSeq4000 runs in this study. The results published here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. NBC received

funding from NIH Initiative for Maximizing Student Diversity (IMSD) T32 Training Fellowship (5R25GM056765-16), NIH Molecular and Cellular Biology T32 Training Fellowship (2T32GM007377-36), NCI IMAT Diversity Supplement (parent grant R21CA204563), NCI CRCHD Diversity Supplement (parent grant U54CA233306). DJS received funding from NIH/NCI (R21CA204563). LGC-C receives funding from NIH/NCI (R01CA223978, U54CA233306, P30CA093373 and R21CA199631) and The Auburn Community Cancer Endowed Chair in Basic Science.

Author Contributions.

LGC-C, DJS and NBC conceived the project. NBC and HO'G designed and performed the molecular experiments. NBC and PCL designed and performed the TCGA CRC dataset analyses. LGC-C, DJS and NBC wrote the manuscript with input from all authors.

Declaration of Interests.

The authors declare no competing interests.

REFERENCES

1. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **68**, 394–424 (2018).
2. Peery, A. F. *et al.* Burden and Cost of Gastrointestinal, Liver, and Pancreatic Diseases in the United States: Update 2018. *Gastroenterology* **156**, 254-272.e11 (2019).
3. Keum, N. & Giovannucci, E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 713–732 (2019).
4. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**,

759–767 (1990).

5. Vogelstein, B. & Kinzler, K. W. The multistep nature of cancer. *Trends Genet.* **9**, 138–141 (1993).

6. Kinzler, K. W. & Vogelstein, B. Lessons from hereditary colorectal cancer. *Cell* **87**, 159–170 (1996).

7. Frampton, M. J. E. *et al.* Implications of polygenic risk for personalised colorectal cancer screening. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **27**, 429–434 (2016).

8. Sud, A., Kinnersley, B. & Houlston, R. S. Genome-wide association studies of cancer: current insights and future perspectives. *Nat. Rev. Cancer* **17**, 692–704 (2017).

9. Tomlinson, I. *et al.* A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet* **39**, 984–988 (2007).

10. Haiman, C. A. *et al.* Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat. Genet.* **39**, 638–644 (2007).

11. Schumacher, F. R. *et al.* A common 8q24 variant in prostate and breast cancer from a large nested case-control study. *Cancer Res.* **67**, 2951–2956 (2007).

12. Ahmadiyeh, N. *et al.* 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. *Proc. Natl. Acad. Sci.* **107**, 9742 LP – 9746 (2010).

13. Sahasrabudhe, R. *et al.* The 8q24 rs6983267G variant is associated with increased thyroid cancer risk. *Endocr. Relat. Cancer* **22**, 841–849 (2015).

14. Zhu, M. *et al.* Association between 8q24 rs6983267 polymorphism and cancer susceptibility: a meta-analysis involving 170,737 subjects. *Oncotarget* **8**, 57421–57439 (2017).

15. Dang, C. V, Le, A. & Gao, P. MYC-induced cancer cell energy metabolism and therapeutic opportunities. *Clin. cancer Res. an Off. J. Am. Assoc. Cancer Res.* **15**, 6479–6483 (2009).

16. Dang, C. V. MYC on the Path to Cancer. *Cell* **149**, 22–35 (2012).

17. Stine, Z. E., Walton, Z. E., Altman, B. J., Hsieh, A. L. & Dang, C. V. MYC, Metabolism, and Cancer. *Cancer Discov.* **5**, 1024–1039 (2015).

18. Dejure, F. R. & Eilers, M. MYC and tumor metabolism: chicken and egg. *EMBO J.* **36**, 3409–3420 (2017).

19. Tuupanen, S. *et al.* The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat. Genet.* **41**, 885 (2009).

20. Pomerantz, M. M. *et al.* The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat. Genet.* **41**, 882–884 (2009).

21. Wright, J. B., Brown, S. J. & Cole, M. D. Upregulation of c-MYC in cis through a Large Chromatin Loop Linked to a Cancer Risk-Associated Single-Nucleotide Polymorphism in Colorectal Cancer Cells. *Mol. Cell. Biol.* **30**, 1411–1420 (2010).

22. Sur, I. K. *et al.* Mice lacking a Myc enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. *Science* **338**, 1360–1363 (2012).

23. Lovén, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).

24. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).

25. Dave, K. *et al.* Mice deficient of Myc super-enhancer region reveal differential control mechanism between normal and pathological growth. *Elife* **6**, (2017).

26. Prokunina-Olsson, L. & Hall, J. L. No effect of cancer-associated SNP rs6983267 in the 8q24 region on co-expression of MYC and TCF7L2 in normal colon tissue. *Mol. Cancer* **8**, 96 (2009).

27. Takatsuno, Y. *et al.* The rs6983267 SNP Is Associated with MYC Transcription Efficiency, Which Promotes Progression and Worsens Prognosis of Colorectal Cancer. *Ann. Surg.*

Oncol. **20**, 1395–1402 (2013).

28. Yang, B. *et al.* No evidence that associations of incident, sporadic colorectal adenoma with its major modifiable risk factors differ by chromosome 8q24 region rs6983267 genotype. *Mol. Carcinog.* **53 Suppl 1**, E193-200 (2014).
29. Nan, H. *et al.* Aspirin use, 8q24 single nucleotide polymorphism rs6983267, and colorectal cancer according to CTNNB1 alterations. *J. Natl. Cancer Inst.* **105**, 1852–1861 (2013).
30. Gong, J. *et al.* A polymorphic MYC response element in KBTBD11 influences colorectal cancer risk, especially in interaction with an MYC-regulated SNP rs6983267. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **29**, 632–639 (2018).
31. Coggins, N. B., Stultz, J., O'Geen, H., Carvajal-Carmona, L. G. & Segal, D. J. Methods for Scarless, Selection-Free Generation of Human Cells and Allele-Specific Functional Analysis of Disease-Associated SNPs and Variants of Uncertain Significance. *Sci. Rep.* **7**, 15044 (2017).
32. Richardson, C. D. *et al.* CRISPR-Cas9 Genome Editing In Human Cells Works Via The Fanconi Anemia Pathway. *bioRxiv* (2017).
33. Yao, X. *et al.* Homology-mediated end joining-based targeted integration using CRISPR/Cas9. *Cell Res.* **27**, 801–814 (2017).
34. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311–318 (2007).
35. Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
36. Tak, Y. G. *et al.* Effects on the transcriptome upon deletion of a distal element cannot be predicted by the size of the H3K27Ac peak in human cells. *Nucleic Acids Res.* **44**, 4123–4133 (2016).
37. Nissan, A. *et al.* Colon cancer associated transcript-1: a novel RNA expressed in

malignant and pre-malignant human tissues. *Int. J. cancer* **130**, 1598–1606 (2012).

38. Ling, H. *et al.* CCAT2, a novel noncoding RNA mapping to 8q24, underlies metastatic progression and chromosomal instability in colon cancer. *Genome Res.* **23**, 1446–1461 (2013).

39. Ge, X. *et al.* Overexpression of long noncoding RNA PCAT-1 is a novel biomarker of poor prognosis in patients with colorectal cancer. *Med. Oncol.* **30**, 588 (2013).

40. Tseng, Y.-Y. *et al.* PVT1 dependence in cancer with MYC copy-number increase. *Nature* **512**, 82–86 (2014).

41. Kim, T. *et al.* Long-range interaction and correlation between MYC enhancer and oncogenic long noncoding RNA CARLo-5. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 4173–4178 (2014).

42. Prensner, J. R. *et al.* The long non-coding RNA PCAT-1 promotes prostate cancer cell proliferation through cMyc. *Neoplasia* **16**, 900–908 (2014).

43. Zhang, Z. *et al.* Long non-coding RNA CASC11 interacts with hnRNP-K and activates the WNT/beta-catenin pathway to promote growth and metastasis in colorectal cancer. *Cancer Lett.* **376**, 62–73 (2016).

44. Wang, J. J. *et al.* Expression and Function of Long Non-coding RNA CASC19 in Colorectal Cancer. *Zhongguo Yi Xue Ke Xue Yuan Xue Bao.* **39**, 756–761 (2017).

45. Shen, P., Pichler, M., Chen, M., Calin, G. A. & Ling, H. To Wnt or Lose: The Missing Non-Coding Linc in Colorectal Cancer. *Int. J. Mol. Sci.* **18**, (2017).

46. Ong, C.-T. & Corces, V. G. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat. Rev. Genet.* **12**, 283–293 (2011).

47. Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84–98 (2012).

48. Zhang, Y. *et al.* Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature* **504**, 306–310 (2013).

49. Hughes, J. R. *et al.* Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat. Genet.* **46**, 205–212 (2014).

50. Symmons, O. *et al.* Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* **24**, 390–400 (2014).

51. Mifsud, B. *et al.* Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **47**, 598–606 (2015).

52. Schoenfelder, S. *et al.* The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res.* **25**, 582–597 (2015).

53. Tang, Z. *et al.* CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell* **163**, 1611–1627 (2015).

54. Krijger, P. H. L., Geeven, G., Bianchi, V., Hilvering, C. R. E. & de Laat, W. 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. *Methods* **170**, 17–32 (2020).

55. Geeven, G., Teunissen, H., de Laat, W. & de Wit, E. peakC: a flexible, non-parametric peak calling package for 4C and Capture-C data. *Nucleic Acids Res.* **46**, e91 (2018).

56. Jäger, R. *et al.* Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. *Nat. Commun.* **6**, 6178 (2015).

57. Barsotti, A. M. *et al.* p53-Dependent induction of PVT1 and miR-1204. *J. Biol. Chem.* **287**, 2509–2519 (2012).

58. Guo, K. *et al.* The expression pattern of long non-coding RNA PVT1 in tumor tissues and in extracellular vesicles of colorectal cancer correlates with cancer progression. *Tumour Biol. J. Int. Soc. Oncodevelopmental Biol. Med.* **39**, 1010428317699122 (2017).

59. Wu, H. *et al.* lncRNA PVT1 Promotes Tumorigenesis of Colorectal Cancer by Stabilizing miR-16-5p and Interacting with the VEGFA/VEGFR1/AKT Axis. *Mol. Ther. Nucleic Acids* **20**, 438–450 (2020).

60. Chai, J. *et al.* A feedback loop consisting of RUNX2/LncRNA-PVT1/miR-455 is involved

in the progression of colorectal cancer. *Am. J. Cancer Res.* **8**, 538–550 (2018).

61. Ping, G. *et al.* Silencing long noncoding RNA PVT1 inhibits tumorigenesis and cisplatin resistance of colorectal cancer. *Am. J. Transl. Res.* **10**, 138–149 (2018).

62. Wang, C., Zhu, X., Pu, C. & Song, X. Upregulated plasmacytoma variant translocation 1 promotes cell proliferation, invasion and metastasis in colorectal cancer. *Mol. Med. Rep.* **17**, 6598–6604 (2018).

63. Fan, H., Zhu, J.-H. & Yao, X.-Q. Long non-coding RNA PVT1 as a novel potential biomarker for predicting the prognosis of colorectal cancer. *Int. J. Biol. Markers* **33**, 415–422 (2018).

64. Fan, H., Zhu, J.-H. & Yao, X.-Q. Knockdown of long non-coding RNA PVT1 reverses multidrug resistance in colorectal cancer cells. *Mol. Med. Rep.* **17**, 8309–8315 (2018).

65. Shang, A.-Q. *et al.* Knockdown of long noncoding RNA PVT1 suppresses cell proliferation and invasion of colorectal cancer via upregulation of microRNA-214-3p. *Am. J. Physiol. Gastrointest. Liver Physiol.* **317**, G222–G232 (2019).

66. Jin, K. *et al.* Long non-coding RNA PVT1 interacts with MYC and its downstream molecules to synergistically promote tumorigenesis. *Cell. Mol. Life Sci.* **76**, 4275–4289 (2019).

67. He, F. *et al.* Long noncoding RNA PVT1-214 promotes proliferation and invasion of colorectal cancer by stabilizing Lin28 and interacting with miR-128. *Oncogene* **38**, 164–179 (2019).

68. Qiao, L. *et al.* Down regulation of the long non-coding RNA PCAT-1 induced growth arrest and apoptosis of colorectal cancer cells. *Life Sci.* **188**, 37–44 (2017).

69. Qiao, L. *et al.* Knockdown of long non-coding RNA prostate cancer-associated ncRNA transcript 1 inhibits multidrug resistance and c-Myc-dependent aggressiveness in colorectal cancer Caco-2 and HT-29 cells. *Mol. Cell. Biochem.* **441**, 99–108 (2018).

70. Wang, A.-H., Fan, W.-J., Fu, L. & Wang, X.-T. LncRNA PCAT-1 regulated cell

proliferation, invasion, migration and apoptosis in colorectal cancer through targeting miR-149-5p. *Eur. Rev. Med. Pharmacol. Sci.* **23**, 8310–8320 (2019).

71. Wang, X.-D., Lu, J., Lin, Y.-S., Gao, C. & Qi, F. Functional role of long non-coding RNA CASC19/miR-140-5p/CEMIP axis in colorectal cancer progression in vitro. *World J. Gastroenterol.* **25**, 1697–1714 (2019).
72. Chen, S., Liu, Y., Wang, Y. & Xue, Z. LncRNA CCAT1 Promotes Colorectal Cancer Tumorigenesis Via A miR-181b-5p/TUSC3 Axis. *Onco. Targets. Ther.* **12**, 9215–9225 (2019).
73. Shang, A. *et al.* Long non-coding RNA CCAT1 promotes colorectal cancer progression by regulating miR-181a-5p expression. *Aging (Albany. NY).* **12**, 8301–8320 (2020).
74. Gu, C. *et al.* Long non-coding RNA CCAT1 promotes colorectal cancer cell migration, invasiveness and viability by upregulating VEGF via negative modulation of microRNA-218. *Exp. Ther. Med.* **19**, 2543–2550 (2020).
75. Yeh, E. *et al.* A signalling pathway controlling c-Myc degradation that impacts oncogenic transformation of human cells. *Nat. Cell Biol.* **6**, 308–318 (2004).
76. Holwerda, S. J. B. *et al.* Allelic exclusion of the immunoglobulin heavy chain locus is independent of its nuclear localization in mature B cells. *Nucleic Acids Res.* **41**, 6905–6916 (2013).
77. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* **9**, e1001046 (2011).
78. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320.e24 (2017).
79. Ozawa, T. *et al.* CCAT1 and CCAT2 long noncoding RNAs, located within the 8q.24.21 'gene desert', serve as important prognostic biomarkers in colorectal cancer. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **28**, 1882–1888 (2017).
80. Bos, J. L. *et al.* Prevalence of ras gene mutations in human colorectal cancers. *Nature*

327, 293–297 (1987).

81. Watanabe, M., Ishiwata, T., Nishigai, K., Moriyama, Y. & Asano, G. Overexpression of keratinocyte growth factor in cancer cells and enterochromaffin cells in human colorectal cancer. *Pathol. Int.* **50**, 363–372 (2000).

82. Davies, H. *et al.* Mutations of the BRAF gene in human cancer. *Nature* **417**, 949–954 (2002).

83. Rajagopalan, H. *et al.* Tumorigenesis: RAF/RAS oncogenes and mismatch-repair status. *Nature* **418**, 934 (2002).

84. Bardelli, A. *et al.* Mutational analysis of the tyrosine kinome in colorectal cancers. *Science* **300**, 949 (2003).

85. Samuels, Y. *et al.* High frequency of mutations of the PIK3CA gene in human cancers. *Science* **304**, 554 (2004).

86. Sato, S., Fujita, N. & Tsuruo, T. Involvement of 3-phosphoinositide-dependent protein kinase-1 in the MEK/MAPK signal transduction pathway. *J. Biol. Chem.* **279**, 33759–33767 (2004).

87. Fang, J. Y. & Richardson, B. C. The MAPK signalling pathways and colorectal cancer. *Lancet. Oncol.* **6**, 322–327 (2005).

88. Cheung, L. W. T. *et al.* High frequency of PIK3R1 and PIK3R2 mutations in endometrial cancer elucidates a novel mechanism for regulation of PTEN protein stability. *Cancer Discov.* **1**, 170–185 (2011).

89. Chen, D. *et al.* BRAFV600E mutation and its association with clinicopathological features of colorectal cancer: a systematic review and meta-analysis. *PLoS One* **9**, e90607 (2014).

90. Das Thakur, M. & Stuart, D. D. Molecular pathways: response and resistance to BRAF and MEK inhibitors in BRAF(V600E) tumors. *Clin. cancer Res. an Off. J. Am. Assoc. Cancer Res.* **20**, 1074–1080 (2014).

91. Kim, E. K. & Choi, E.-J. Compromised MAPK signaling in human diseases: an update. *Arch. Toxicol.* **89**, 867–882 (2015).

92. Thorpe, L. M., Yuzugullu, H. & Zhao, J. J. PI3K in cancer: divergent roles of isoforms, modes of activation and therapeutic targeting. *Nat. Rev. Cancer* **15**, 7–24 (2015).

93. Janku, F., Yap, T. A. & Meric-Bernstam, F. Targeting the PI3K pathway in cancer: are we making headway? *Nat. Rev. Clin. Oncol.* **15**, 273–291 (2018).

94. Ganesh, S. *et al.* Urokinase receptor and colorectal cancer survival. *Lancet (London, England)* **344**, 401–402 (1994).

95. Wei, Y. *et al.* Regulation of integrin function by the urokinase receptor. *Science* **273**, 1551–1555 (1996).

96. Danø, K. *et al.* Cancer invasion and tissue remodeling--cooperation of protease systems and cell types. *APMIS* **107**, 120–127 (1999).

97. Ossowski, L. & Aguirre-Ghiso, J. A. Urokinase receptor and integrin partnership: coordination of signaling for cell adhesion, migration and growth. *Curr. Opin. Cell Biol.* **12**, 613–620 (2000).

98. Kaneko, T., Konno, H., Baba, M., Tanaka, T. & Nakamura, S. Urokinase-type plasminogen activator expression correlates with tumor angiogenesis and poor outcome in gastric cancer. *Cancer Sci.* **94**, 43–49 (2003).

99. Ahmed, N., Oliva, K., Wang, Y., Quinn, M. & Rice, G. Downregulation of urokinase plasminogen activator receptor expression inhibits Erk signalling with concomitant suppression of invasiveness due to loss of uPAR-beta1 integrin complex in colon cancer cells. *Br. J. Cancer* **89**, 374–384 (2003).

100. Brungs, D. *et al.* The urokinase plasminogen activation system in gastroesophageal cancer: A systematic review and meta-analysis. *Oncotarget* **8**, 23099–23109 (2017).

101. Schild, C. *et al.* PI3K signaling maintains c-myc expression to regulate transcription of E2F1 in pancreatic cancer cells. *Mol. Carcinog.* **48**, 1149–1158 (2009).

102. Takeda, K., Naguro, I., Nishitoh, H., Matsuzawa, A. & Ichijo, H. Apoptosis signaling kinases: from stress response to health outcomes. *Antioxid. Redox Signal.* **15**, 719–761 (2011).

103. Tsai, W.-B. *et al.* Activation of Ras/PI3K/ERK pathway induces c-Myc stabilization to upregulate argininosuccinate synthetase, leading to arginine deiminase resistance in melanoma cells. *Cancer Res.* **72**, 2622–2633 (2012).

104. Liu, Y. *et al.* LMO2 attenuates tumor growth by targeting the Wnt signaling pathway in breast and colorectal cancer. *Sci. Rep.* **6**, 36050 (2016).

105. Heuberger, J. *et al.* A C/EBPα-Wnt connection in gut homeostasis and carcinogenesis. *Life Sci. alliance* **2**, e201800173 (2019).

106. Young, M. R. & Colburn, N. H. Fra-1 a target for cancer prevention or intervention. *Gene* **379**, 1–11 (2006).

107. He, J. *et al.* Fra-1 is upregulated in gastric cancer tissues and affects the PI3K/Akt and p53 signaling pathway in gastric cancer. *Int. J. Oncol.* **47**, 1725–1734 (2015).

108. Scaltriti, M., Elkabets, M. & Baselga, J. Molecular Pathways: AXL, a Membrane Receptor Mediator of Resistance to Therapy. *Clin. cancer Res. an Off. J. Am. Assoc. Cancer Res.* **22**, 1313–1317 (2016).

109. Antony, J. & Huang, R. Y.-J. AXL-Driven EMT State as a Targetable Conduit in Cancer. *Cancer Res.* **77**, 3725–3732 (2017).

110. Kappelmann, M., Bosserhoff, A. & Kuphal, S. AP-1/c-Jun transcription factors: regulation and function in malignant melanoma. *Eur. J. Cell Biol.* **93**, 76–81 (2014).

111. Muzny, D. M. *et al.* Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).

112. Redis, R. S. *et al.* CCAT2, a novel long non-coding RNA in breast cancer: expression study and clinical correlations. *Oncotarget* **4**, 1748–1762 (2013).

113. Kasagi, Y. *et al.* The Expression of CCAT2, a Novel Long Noncoding RNA Transcript,

and rs6983267 Single-Nucleotide Polymorphism Genotypes in Colorectal Cancers. *Oncology* **92**, 48–54 (2017).

114. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350–1356 (2015).

115. Rennoll, S. & Yochum, G. Regulation of MYC gene expression by aberrant Wnt/β-catenin signaling in colorectal cancer. *World J. Biol. Chem.* **6**, 290–300 (2015).

116. Muellner, M. K. *et al.* A chemical-genetic screen reveals a mechanism of resistance to PI3K inhibitors in cancer. *Nat. Chem. Biol.* **7**, 787–793 (2011).

117. Tenbaum, S. P. *et al.* β-catenin confers resistance to PI3K and AKT inhibitors and subverts FOXO3a to promote metastasis in colon cancer. *Nat. Med.* **18**, 892–901 (2012).

118. Ormanns, S., Neumann, J., Horst, D., Kirchner, T. & Jung, A. WNT signaling and distant metastasis in colon cancer through transcriptional activity of nuclear β-Catenin depend on active PI3K signaling. *Oncotarget* **5**, 2999–3011 (2014).

119. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotech* **34**, 339–344 (2016).

120. O'Geen, H., Frietze, S. & Farnham, P. J. Using ChIP-seq Technology to Identify Targets of Zinc Finger Transcription Factors. in *Engineered Zinc Finger Proteins: Methods and Protocols* (eds. Mackay, J. P. & Segal, D. J.) 437–455 (Humana Press, 2010). doi:10.1007/978-1-60761-753-2_27

121. van de Werken, H. J. G. *et al.* 4C technology: protocols and data analysis. *Methods Enzymol.* **513**, 89–112 (2012).

122. Grossman, R. L. *et al.* Toward a Shared Vision for Cancer Genomic Data. *N. Engl. J. Med.* **375**, 1109–1112 (2016).

123. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, (2015).

Figures

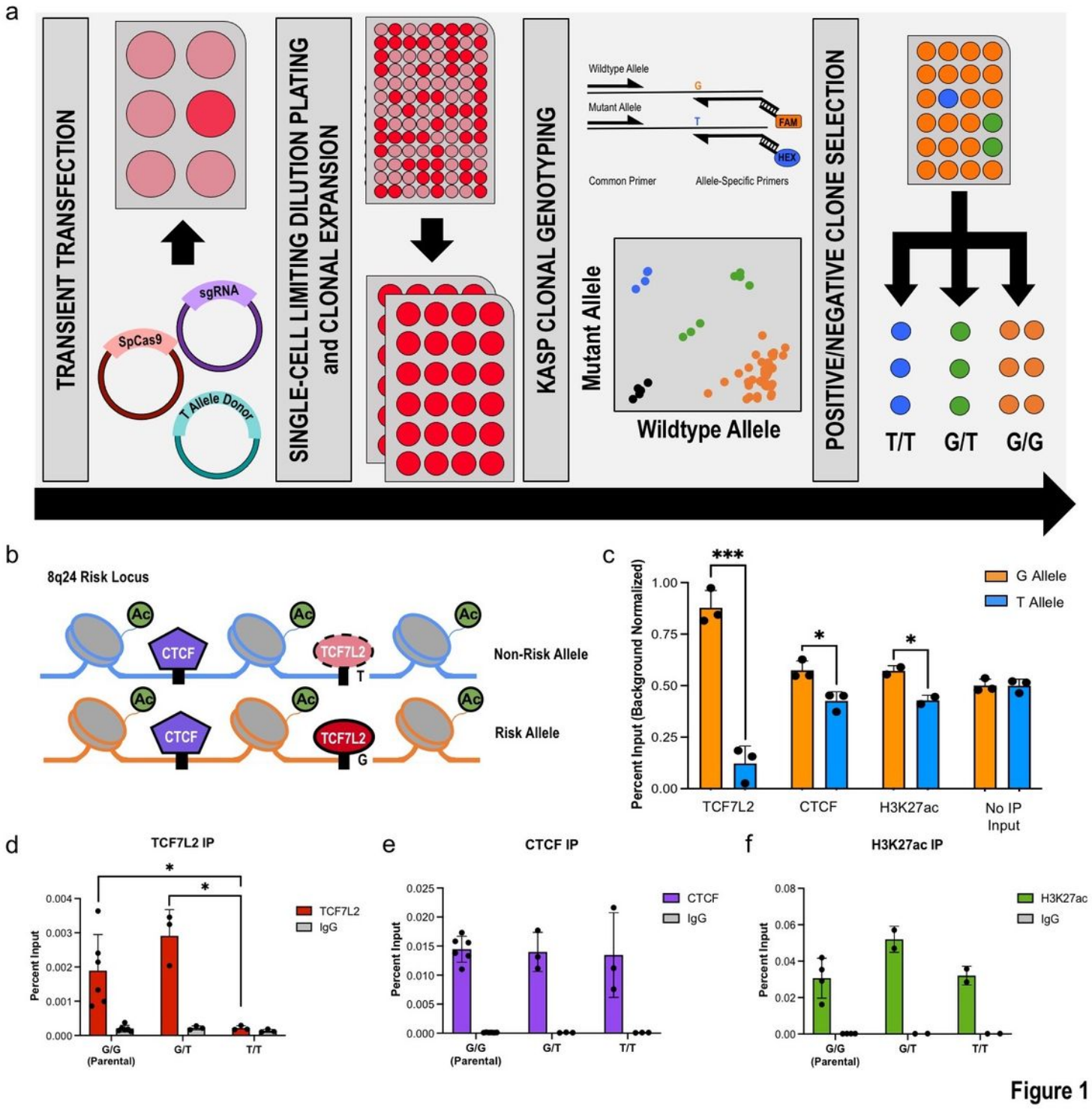

Figure 1

Figure 1

Preferential binding of transcription factor TCF7L2 to the rs6983267 risk allele and its effects on local enhancer function. (a) Scarless genome editing and clonal selection pipeline producing 3 homozygous reference, 3 heterozygous and 6 un-edited homozygous risk clonal lines derived from HCT-116 modeling

rs6983267 genotypes. (b) Graphical representation of enhancer region encompassing risk SNP rs6983267. TCF7L2 transcription factor (red) binding site, adjacent CTCF insulator (purple) binding site, histone mark H3K27ac (green) associated with active enhancers surrounding risk SNP. (c) Chromatin Immunoprecipitation and Quantitative Allele-Specific Occupancy (ChIPnQASO) in heterozygous lines (n=3) shows strong preference of TCF7L2 binding to the risk allele (orange bars) over the non-risk allele (blue bars) and slight preference for adjacent CTCF binding and surrounding H3K27ac marks. Asterisks (*) indicate significantly different groups by Student's t-test (*) p-value < 0.05, (***) p-value < 0.001. Error bars show standard deviation. (d,e,f) Chromatin Immunoprecipitation comparing overall binding of TCF7L2 (red), CTCF (purple), H3K27ac (green) in across all genotypes (n=6 for GG, n=3 for GT and TT) shows ablation of TCF7L2 binding in the absence of the risk (G) allele. Grey bars indicate negative control IgG. Asterisks (*) indicate significantly different groups by One-Way ANOVA followed by Post-Hoc Tukey HSD. (*) p-value < 0.05. Error bars show standard deviation.

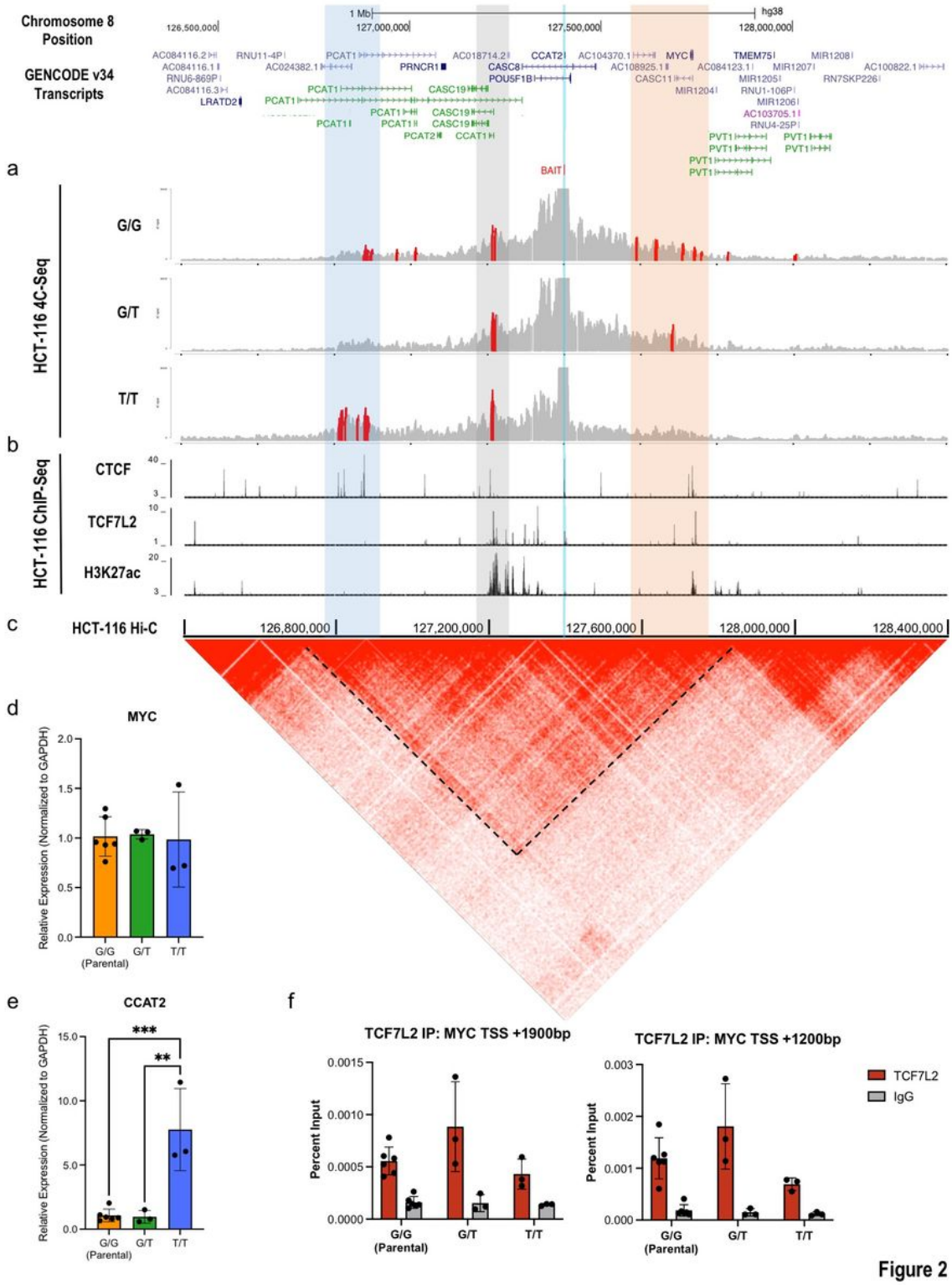

Figure 2

Figure 2

Local regulatory interactions for risk SNP rs6983267 in HCT-116 genotype clones. (Top) Genes annotated within +/- 1Mb of SNP. (a) 4C-Seq DNA interactions (significant interactions in red called by PeakC software) generated for all genotypes (n=6 for GG, n=3 for GT and TT) of rs6983267 show differential interactions. Loss of the risk (G) allele shows loss of interactions with the MYC promoter (highlighted in orange) and gain of interactions near the PCAT1 promoter (highlighted in blue). Bait indicates region

encompassing SNP used for interaction detection. (b) ENCODE ChIP-Seq signal tracks for CTCF, TCF7L2 and H3K27ac histone marks in HCT-116 cell line. (c) Hi-C chromatin contacts at 5Kb resolution from Rao et al. in HCT-116 cell line. Visualized using Juicebox.js software. TAD containing rs6983267 and its interactors indicated by dotted line outlining enrichment of contacts within the region. (d,e) Relative RNA expression of MYC and lncRNA CCAT2 across genotypes (n=6 for GG, n=3 for GT and TT) measured by RT-qPCR. Asterisks (*) indicate significantly different groups by One-Way ANOVA followed by Post-Hoc Tukey HSD. (**) p-value < 0.01, (***) p-value <0.001. Error bars show standard deviation. Error bars show standard deviation. (f) Chromatin Immunoprecipitation comparing overall binding of TCF7L2 across all genotypes (n=6 for GG, n=3 for GT and TT) shows no statistically significant difference in TCF7L2 enrichment within the MYC promoter at multiple binding sites (+1900bp, +1200bp) identified by ENCODE ChIP-seq upstream of the MYC TSS across genotypes. Grey bars indicate negative control IgG. Groups were assessed with One-Way ANOVA followed by Post-Hoc Tukey HSD. Error bars show standard deviation.

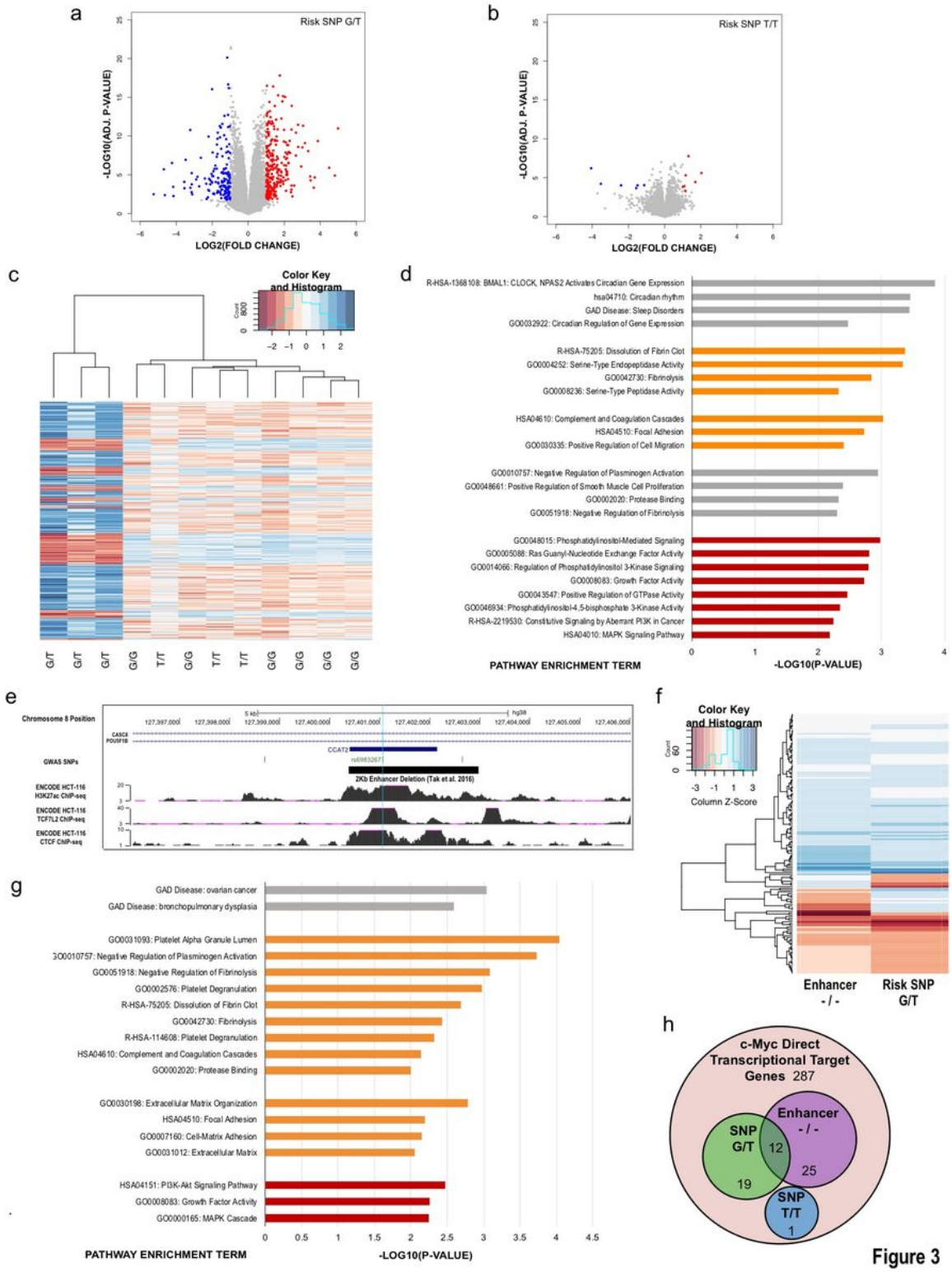

Figure 3

SNP status of rs6983267 affects a subset of enhancer gene that function in the PI3K and Ras/Raf/MAPK pathways in HCT-116 cells. (a,b) Volcano plots showing significantly differentially expressed genes (DEGs) across rs6983267 genotype clones. Significantly downregulated (blue dots) and upregulated (red dots) genes were selected using log2 fold change <-1.0 or >1.0 and FDR adjusted p-value < 0.05 cut-off values for heterozygous (a) and homozygous non-risk (T/T) (b) compared to parental homozygous risk

(G/G). (c) Unsupervised hierarchical clustering heatmap of top 500 DEGs shows clustering of heterozygous clones (n=3) away from both homozygous risk (n=3) and homozygous non-risk clones (n=3). (d) DAVID Gene Ontology (GO) analysis of significantly upregulated DEGs in heterozygous compared to parental homozygous risk (G/G) shows enrichment for pathways involved in extracellular matrix remodeling and cell motility and migration (orange bar terms) as well as enrichment for aberrantly activated cell growth and proliferation pathways PI3K and Ras/Raf/MAPK (red bar terms). (e) UCSC Genome Browser shot of Enhancer Deletion in HCT- 116 cells by Tak et al. (black bar, middle), SNP rs6983267 (light blue vertical line) and ENCODE ChIP-Seq signal tracks for H3K27ac, TCF7L2 and CTCF. (f) Heatmap of 148 common DEGs between heterozygotes and enhancer deletion (upregulated in blue, downregulated in red) indicate putative target genes of enhancer SNP regulation. (g) DAVID GO analysis of commonly upregulated genes from 5d shows enrichment for same pathways involved in extracellular matrix remodeling and cell motility and migration (orange bar terms) as well as enrichment for aberrantly activated cell growth and proliferation pathways such as PI3K, Ras and MAPK (red bar terms) seen for Heterozygous enrichment analysis alone. (h) Direct c-Myc targets identified by IPA (Qiagen) significantly DE in enhancer deletion and heterozygotes show putative target genes of enhancer SNP regulation within the c-Myc interactome.

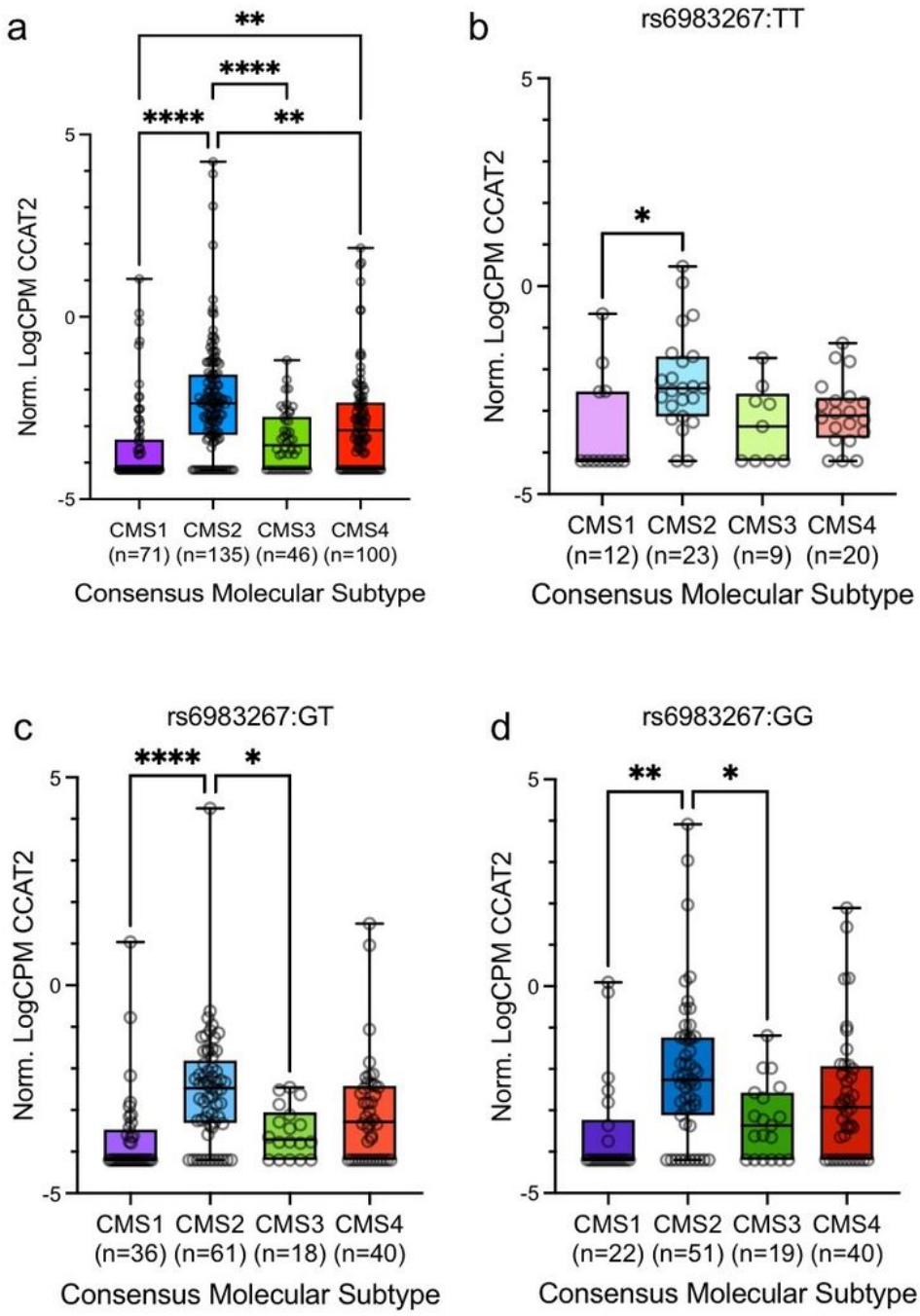

Figure 4

Figure 4

rs6983267 risk allele correlates with higher CCAT2 in CRC tumors with aberrant canonical WNT signaling and affect APC and PIK3CA co-mutations in TCGA colorectal adenocarcinoma (COAD) patient cohort. (a) Stratification of COAD patients by consensus molecular subtype (CMS) shows significantly higher CCAT2 expression overall within the CMS2 subtype. F (3,348) = 17.82. (b, c, d) Further stratification by rs6983267 genotype (TT, F (3,60) = 4.155; GT, F (3,151) = 7.907, GG, F (3,126) = 6.056) shows this effect is

strengthened with the presence of the risk allele (G) as both heterozygous (GT) and homozygous risk (GG) groups show significant increase in CCAT2 expression within the CMS2 subtype compared to CMS1 and CMS3. Asterisks (*) indicate significantly different groups by One-Way ANOVA followed by Post-Hoc Tukey HSD. (*) p-value < 0.05, (**) p-value < 0.01, (****) p-value <0.0001. Error bars show standard deviation.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Coggins20214CseqPeakCSignificantInteractionsSuppTable.xlsx
- Coggins2021PrimersOligosListSuppTable2.xlsx
- Coggins2021RNAseqExpressionMYCTargetsSuppTable3.xlsx
- Coggins2021SupplementaryMaterials.pdf
- Coggins2021MainTable1.jpeg
- Coggins2021MainTable2.jpeg