

Published in final edited form as:

Science. 2012 May 11; 336(6082): 736–739. doi:10.1126/science.1217277.

Epigenomic enhancer profiling defines a signature of colon cancer

Batool Akhtar-Zaidi¹, Richard Cowper-Sal-lari², Olivia Corradin¹, Alina Saiakhova¹, Cynthia F. Bartels¹, Dheepa Balasubramanian¹, Lois Myeroff³, James Lutterbaugh³, Awad Jarrar⁴, Matthew F. Kalady^{3,4,5}, Joseph Willis^{3,6}, Jason H. Moore², Paul J. Tesar^{1,3}, Thomas Laframboise^{1,3}, Sanford Markowitz^{1,3,7}, Mathieu Lupien², and Peter C. Scacheri^{1,3}

[Assistant Professor]

¹Department of Genetics, Case Western Reserve University, Cleveland, OH 44106

²Norris Cotton Cancer Center, Institute for Quantitative Biomedical Sciences, Department of Genetics, Dartmouth Medical School, Lebanon, NH 03756, USA

³Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH 44106, USA.

⁴Department of Colorectal Surgery, Cleveland Clinic Foundation, Cleveland, Ohio 44195, USA

⁵Cancer Biology Department, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, Ohio 44195, USA

⁶Department of Pathology, Case Western Reserve University, Cleveland, OH 44106

⁷Department of Medicine, Case Western Reserve University, Cleveland, OH 44106

Abstract

Cancer is characterized by gene expression aberrations. Studies have largely focused on coding sequences and promoters, despite the fact that distal regulatory elements play a central role in controlling transcription patterns. Here we utilize the histone mark H3K4me1 to analyze gain and loss of enhancer activity genome wide in primary colon cancer lines relative to normal colon crypts. We identified thousands of variant enhancer loci (VELs) that comprise a signature that is robustly predictive of the in vivo colon cancer transcriptome. Furthermore, VELs are enriched in haplotype blocks containing colon cancer genetic risk variants, implicating these genomic regions in colon cancer pathogenesis. We propose that reproducible changes in the epigenome at enhancer elements drive a unique transcriptional program to promote colon carcinogenesis.

REPORT

Although non-coding functional elements play a central role in establishing gene expression patterns that drive normal development, cell-type identity, and evolutionary processes, their potential involvement in the context of common cancers remains unknown. The mono- and di-methylated forms of H3K4 (H3K4me1/2) broadly mark multiple classes of gene enhancer elements (1-3). Here, we present an epigenomic comparison of H3K4me1-marked gene enhancer elements in a cohort of colorectal cancer (CRC) cell lines and normal colon epithelial crypt cells, from which colon cancer is derived.

We performed H3K4me1 ChIP-seq analysis on 3 preparations of normal epithelial crypts as well as primary CRC cell lines derived from 2 early stage tumors (V432 and V703), 2 late stage tumors (V8 and V9P), and 5 liver metastases (V400, V457, V481, V503, V9M). On average, we detected ~ 71,000 peaks significantly enriched for H3K4me1 at a False Discovery Rate of less than 5% (Table S1). The distribution of H3K4me1 relative to annotated genes is similar between colon cancer samples and crypt controls, with the majority of H3K4me1 sites mapping to intergenic and intronic regions located distal to transcription start sites (fig. S1). We compared H3K4me1 patterns between all 12 colon samples and 9 unrelated human cell types (4). H3K4me1 patterns in tumors are more similar to colon crypt than non-colon cells, consistent with the notion that colon tumors are derived from colon crypts (fig. S2). Moreover, there is less variation between the colon samples than between unrelated cell types.

We identified thousands of H3K4me1 sites, or Variant Enhancer Loci (VELs), that are differentially enriched (lost or gained) in each of the CRC samples compared to normal colon crypts (Fig. 1A). On average, less than 0.05% of VELs map to regions altered in DNA copy number, and thus, the vast majority of VELs are unlikely to be the result of copy number variations related to malignant transformation. VELs comprise 28-61% of all putative enhancers present in a given CRC sample (Fig. 1B). ChIP-seq analysis of H3K27ac, an epigenetic mark of active enhancer elements, revealed that ~40% of gained VELs acquire H3K27ac in CRC. 70% of lost VELs are enriched for H3K27ac in normal crypts, and show virtually no detectable levels of H3K27ac in CRC (fig. S3). We also performed global mapping of DNase I hypersensitive sites in two CRC lines using DNase-seq (5). Consistent with acquisition and loss of enhancer marks, virtually all gained VELs map to open chromatin sensitive to DNase I digestion, and lost VELs map to DNase I-insensitive regions (fig S3D and E). Collectively, the data indicate that multiple changes in chromatin state and function accompany the changes in H3K4me1 at VELs. Lastly, we verified that H3K4me1 sites are functionally active using luciferase reporter assays (fig. S4).

A higher number of VELs than expected by random chance are common to multiple CRC samples. Specifically, we detected 2604 gained VELs common to five or more lines, and 2047 lost VELs common to six or more CRC lines ($P < 0.001$). Both unique and common VELs are distributed relatively evenly among the CRC samples (Fig. 1D). 197 VELs are shared between all 9 CRC samples. The universally common VELs are dispersed throughout the genome on multiple chromosomes, and do not appear to cluster in any meaningful way (fig. S5).

We ranked VELs by their level of specificity in crypts and nine unrelated cell types. Compared to a control set of H3K4me1 sites invariant between CRC samples and crypts, lost VELs are highly crypt specific, while gained VELs are relatively non-crypt specific (fig. S6A). These relationships also held true for common VELs (fig. S6B). We also determined that 67-92% of gained VELs map to H3K4me1-marked loci in any one of the nine non-colon cell types, compared to 9-11% for lost VELs and 24-31% for control enhancers (Fig. 1E). Collectively, these data indicate that in colon cancer, the chromatin configuration is altered by acquisition of putative enhancer marks that are normally found in non-colon cell types, and loss of putative enhancer marks that typify normal crypt differentiation status; the net effect leading to a less colon specific phenotype.

Multiple approaches were used to assess the relationship between VELs and gene expression. Compared to control genes not linked to gained VELs, genes linked to gained VELs are generally expressed higher in CRC samples than crypts, and genes linked to lost VELs are expressed lower in CRC samples than crypts (Fig. 2A and fig. S7). For all CRC samples, the effect of lost VELs on gene repression is more pronounced than the effect of

gained VELs on gene overexpression, indicating that lost VELs are more likely than gained VELs to confer a functional effect. Overexpressed genes are 1.6 - 6.2 times more likely than randomly selected control genes to have gained VELs (Fig. 2B). Repressed genes are 2.8 – 8.7 times more likely than controls to have lost VELs (Fig. 2C). One VEL is generally sufficient to confer an effect on gene expression, and additional VELs confer more marked changes in a relatively quantitative fashion (Fig. 2, D and E, and fig. S8). Genes associated with gained VELs are generally expressed at high levels in crypt controls and become further elevated in CRC (Fig. 2F, and fig. S9). Genes associated with lost VELs are expressed at mid-high levels in crypt controls and generally become either attenuated or silenced in CRC (Fig. 2F, and fig. S9). These results are consistent with the above findings indicating that the majority of lost VELs lose the active H3K27ac enhancer mark, while the minority of gained VELs acquire H3K27ac. We also found that correlations of global gene expression between CRC samples and crypts improved when VEL genes were not considered (fig. S10A). Common VELs are also enriched for genes frequently dysregulated in the CRC cell lines (fig. S10B). Collectively, the data indicate that gained and lost VELs are highly predictive of local cancer-specific overexpressed and repressed genes, respectively. Consistent with these positive correlations, lost VEL gene promoters often show decreases of H3K4me3 and/or H3K27ac in CRC relative to crypts, and gained VEL gene promoters show increases of H3K4me3 and/or H3K27ac in CRC relative to crypts (fig. S11). However, there is also a class of VEL genes that do not show measurable differences in promoter-associated H3K4me3/H3K27ac between normal crypts and CRC, but clearly show expression changes (fig. S11, B and C).

If VELs are indeed cancer-related events, then aberrantly expressed genes associated with common VELs ought to validate as aberrantly expressed in primary tumors. We determined that overexpressed genes associated with gained VELs common to 5-9 lines, and repressed genes associated with lost VELs common to 6-8 lines validated as aberrantly expressed in primary tumors at a rate 2-8 times higher than that determined when the VEL was not considered (Fig. 3, A and B). These results suggest that VELs are a signature predictive of the *in vivo* colon cancer transcriptome more robustly than the aberrant gene expression patterns associated with the colon cancer cell lines from which the VELs themselves were identified. PDGH, a known colon tumor suppressor gene associated with the VEL signature and repressed in CRC is shown in Figure 3C (6).

Twenty SNPs have been identified through GWAS studies to confer risk to CRC (7-18). We utilized Variant Set Enrichment analyses (VSE) to test whether enhancers and VELs were significantly enriched among the 20 CRC-risk SNPs (or variants in LD with the CRC risk SNPs (clusters), designated as the Annotated Variant Set, or AVS). Among the 20 clusters of SNPs comprising the AVS, 16 (80%) overlapped at least one H3K4me1 site in colon crypt (Fig. 4A). Similar analyses in nine other cell types indicated that the CRC AVS association was specific to H3K4me1 enhancers in colon crypt and HepG2 cells (Fig. 4B). Furthermore, significant associations were detected between the AVS and low frequency lost VELs (L1 and L2, Fig. 4B), and not common gained or lost VELs. An example is shown in Figure 4C. Of the 8 SNPs associated with unique lost VELs, five (rs719725, rs6983267, rs10505477, rs7014346, rs3802842) were associated with enhancers in crypt and HepG2 cells, and not in any other cell types, indicating that SNP/enhancer associations exclusive to the disease-relevant tissue are particularly important. Although VSE tests for enrichment of enhancers in linkage disequilibrium with the CRC AVS as a whole, we did detect multiple instances in which individual risk SNPs (or variants in strong LD with the risk SNP) overlapped VELs, despite the lack of significance with the entire AVS. For example, rs4444235 was significantly associated with gains common to 7 CRC lines ($P=0.004$). Rs4444235 maps to the enhancer of BMP4 and increases its expression (19). Accordingly, gained VELs at this locus correlate with increased BMP4 expression in CRC

cell lines. Furthermore, lost VELs associated with risk SNPs rs719725 and rs9929218 were associated with reduced expression of potential target genes, JMJD2C and TMED6, in CRC samples containing the lost VELs. Collectively, these findings provide further evidence that enhancers and VELs are relevant to CRC pathogenesis.

Our epigenomic comparison of H3K4me1-marked gene enhancer elements in colon cancer cells suggest that central changes at enhancers drive a unique transcriptional program to promote colon carcinogenesis. Lost VELs appear to be more of a contributor to this signature than gained VELs, as lost VELs confer a greater functional effect on expression than gained VELs, are better predictors of gene expression in primary tumors than gained VELs, typify colon crypt identity, are far more concordant across tumors than gained VELs, and are more robustly associated with CRC-risk SNPs than gained VELs. Most, but not all, VELs are linked to changes in promoter-associated H3K4me3 and H3K27ac. Thus, VELs capture novel and global information about the chromatin state that is related to gene expression. Moreover, these findings suggest that some of the VEL genes identified in this study would likely remain undiscovered through analysis of these promoter marks alone. Lastly, the majority of VELs are common to at least two of nine (>20%) CRC samples. The commonality of the epigenetic colon cancer signature captured by VELs contrasts with the marked heterogeneity in mutations in colon cancer candidate driver genes revealed by genome sequencing and suggests either that VELs capture pathway outputs that are downstream of sets of gene mutations or that they capture epigenetic alterations that are independent of and more common than gene mutations (20-22). Clearly, the number of enhancers consistently altered across multiple CRC tumors is likely far greater than genes commonly mutated in colon cancer. These findings, even when adjusted for the notion that enhancers are 2-5 times more prevalent than genes, suggest that the epigenetic terrain at gene enhancer elements in colon cancer is less heterogenous than the genetic landscape of protein coding genes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

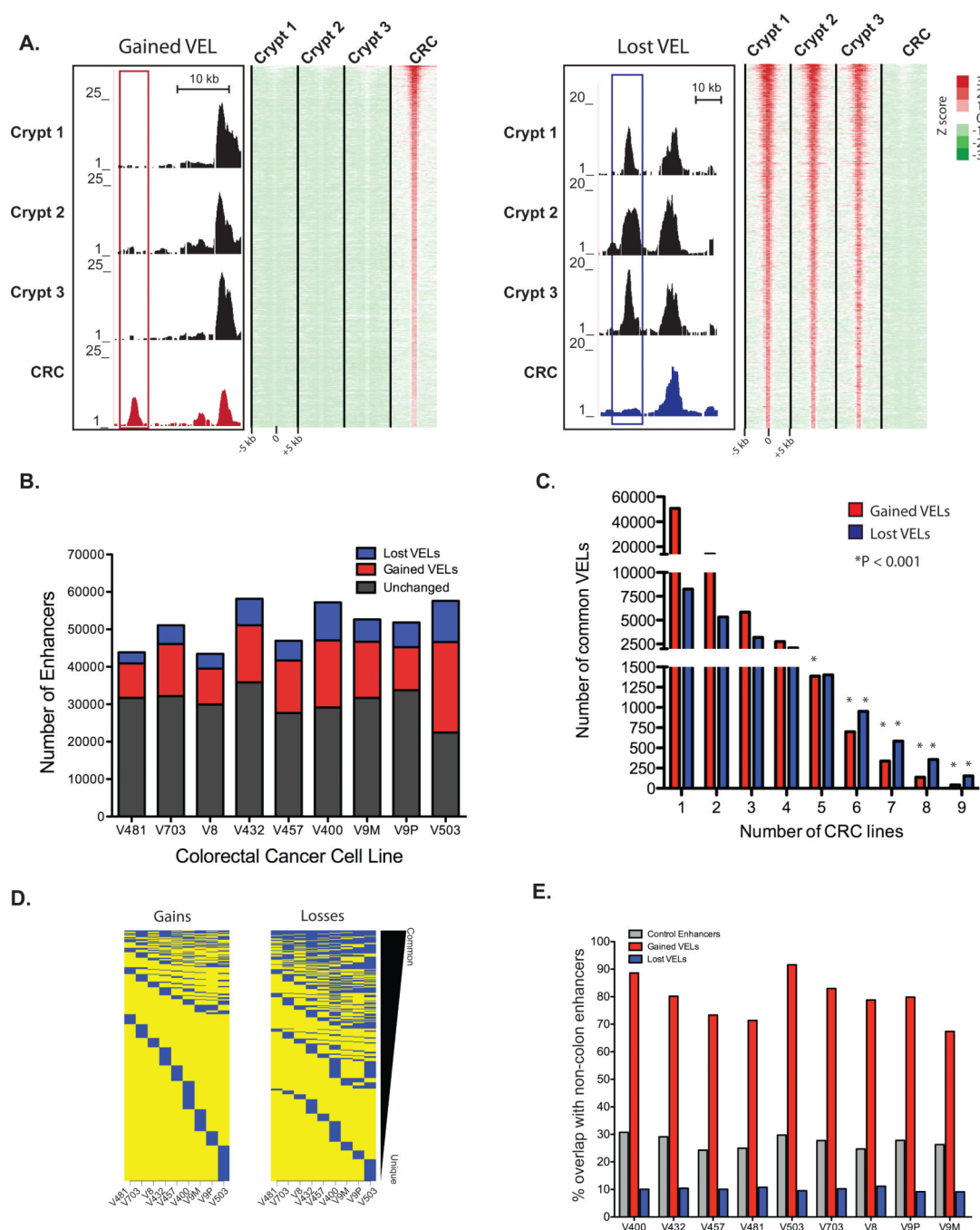
Acknowledgments

We thank Angela Ting and Kishore Guda for helpful comments and discussion, Zhancheng Zhang for providing Perl scripts for data analysis, Pavel Manaenkov for assistance with data visualization, and Simone Edelheit, Nick Beckloff, and Neil Molyneaux from the CWRU Genomics Core for sequencing and informatics assistance. This work was supported in part by the following NIH grants: R01HD056369 to PCS, R01CA1555004 to ML, R01-LM009012 and R01-LM010098 to JHM and RC, IP50CA150964 and NIH U01 CA152756 to SM, and 5T32GM008056-29 to OC. All data is currently being deposited in Genbank and will be made publically available upon publication.

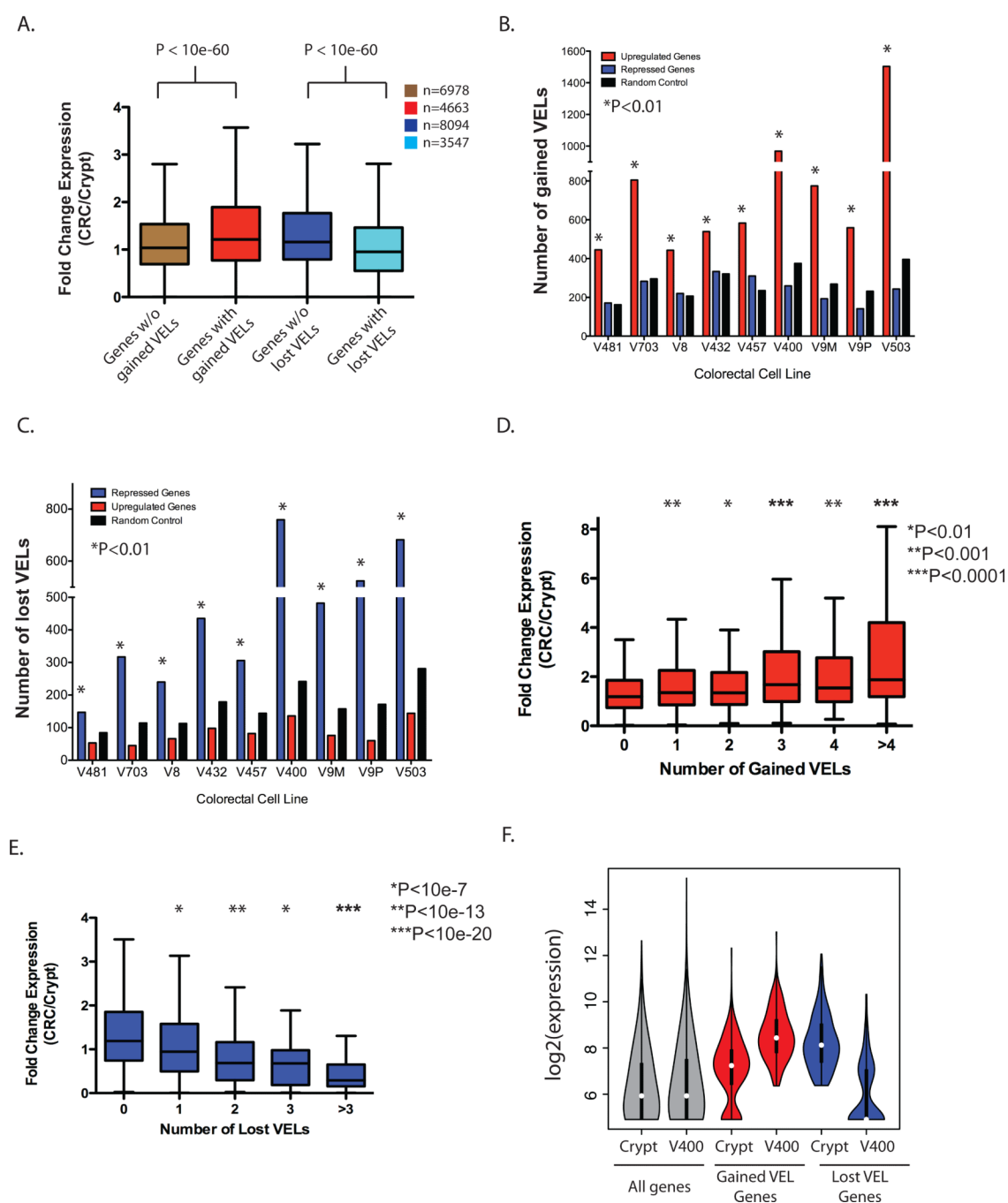
REFERENCES

1. Heintzman ND, et al. *Nature*. 2009; 459
2. Rada-Iglesias A, et al. *Nature*. 2011; 470
3. Zentner GE, Tesar PJ, Scacheri PC. *Genome Res*. 2011; 21
4. Ernst J, et al. *Nature*. 2011; 473
5. Boyle AP, et al. *Cell*. 2008; 132
6. Myung SJ, et al. *Proc Natl Acad Sci U S A*. 2006; 103
7. Peters U, et al. *Hum Genet*. 2011
8. Figueiredo JC, et al. *Cancer Epidemiol Biomarkers Prev*. 2011; 20
9. Cui R, et al. *Gut*. 2011; 60
10. Houlston RS, et al. *Nat Genet*. 2010; 42

11. Houlston RS, et al. Nat Genet. 2008; 40
12. Tomlinson IP, et al. Nat Genet. 2008; 40
13. Tenesa A, et al. Nat Genet. 2008; 40
14. Broderick P, et al. Nat Genet. 2007; 39
15. Lascorz J, et al. Carcinogenesis. 2010; 31
16. Zanke BW, et al. Nat Genet. 2007; 39
17. Poynter JN, et al. Cancer Res. 2007; 67
18. Hindorff, LA., et al. 2012. <http://www.genome.gov/gwastudies>
19. Lubbe SJ, et al. Oncogene. 2011
20. Bass AJ, et al. Nat Genet. 2011; 43
21. Sjoblom T, et al. Science. 2006; 314
22. Wood LD, et al. Science. 2007; 318

**Figure 1.**

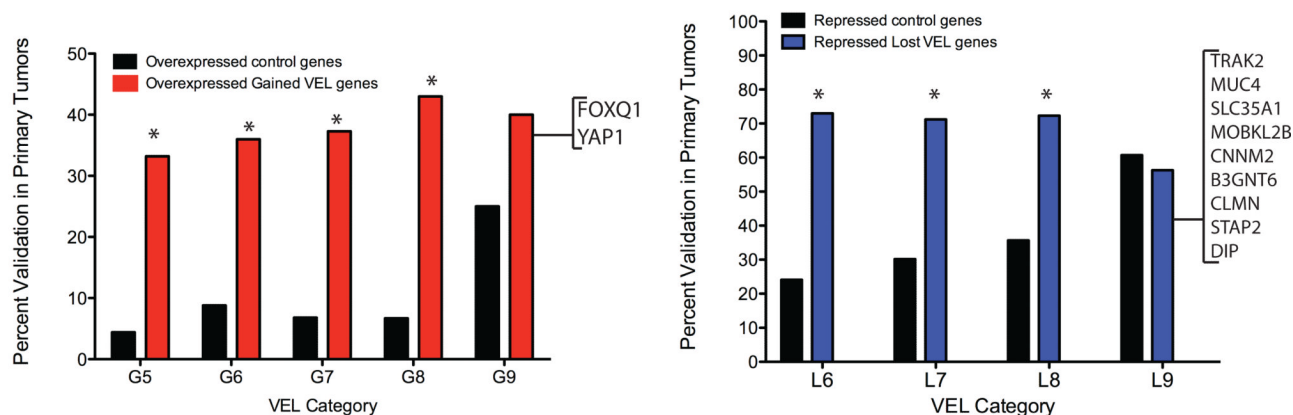
H3K4me1 ChIP-seq identifies Variant Enhancer Loci (VELs). (A) UCSC browser views of H3K4me1 profiles from 3 normal crypts and a CRC cell line (V400), illustrating an example of a gained (left) and a lost (right) VEL. Heatmaps show the corresponding H3K4me1 ChIP-seq signals \pm 5kb of VEL midpoints. (B) Number of VELs and unchanged H3K4me1-sites in CRC samples relative to normal controls. (D) Number of unique and common VELs. (D) Distribution of VELs among CRC lines. Blue = VEL. (E) Percentage of control enhancers and VELs that overlap with H3K4me1 sites in any of nine non-colon cell types. All comparisons are significant by Chi Squared test ($P < 10e-10$).

**Figure 2.**

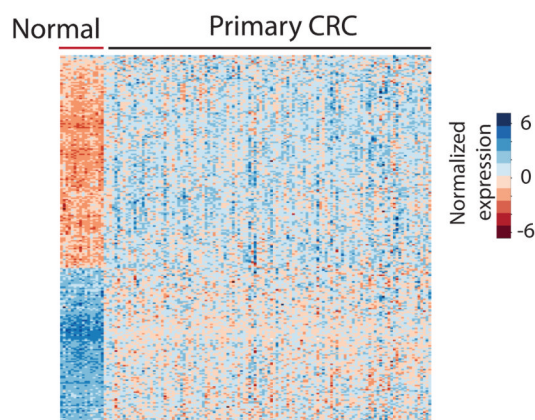
VELs correlate with aberrant gene expression. (A) Fold change in expression of VEL and control genes for a representative CRC line (V400). Number of (B) gained and (C) lost VELs associated with over-expressed and repressed genes, respectively. Fold change in expression of genes associated with variable numbers of (D) gained VELs and (E) lost VELs in CRC sample V400. (F) Levels of all genes (grey) and aberrantly expressed genes (> 1.5 -fold relative to crypts) associated with VELs in CRC sample V400.

*P<10e-7

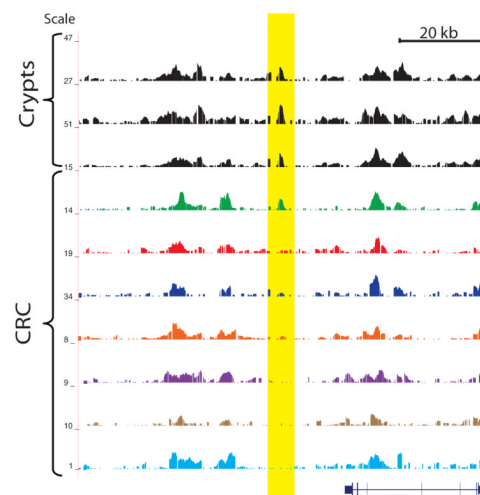
A.



B.



C.

**Figure 3.**

Common VELs predict aberrant gene expression in primary tumors. (A) (left) Red bars represent the percentage of overexpressed genes associated with gained VELs common to five or more lines (G5-G9) that validate as overexpressed in primary tumors. Black bars represent the baseline predictive power when the VEL is not considered, i.e., the percentage of overexpressed genes in 5 or more cell lines that validate as overexpressed in primary tumors. G9 genes that validated as overexpressed in primary tumors are listed in brackets. (right) same as left, but for lost VELs common to six or more lines (L6-L9). (B) Heatmap of expression of VEL-associated genes in panel A (red and blue bars) in normal colon tissue (n=16) and primary CRC tumors (n=120). (C) UCSC Browser view of H4K4me1 ChIP-seq signals across the PDGH locus, associated with a lost VEL common to 6 CRC samples (highlighted in yellow).

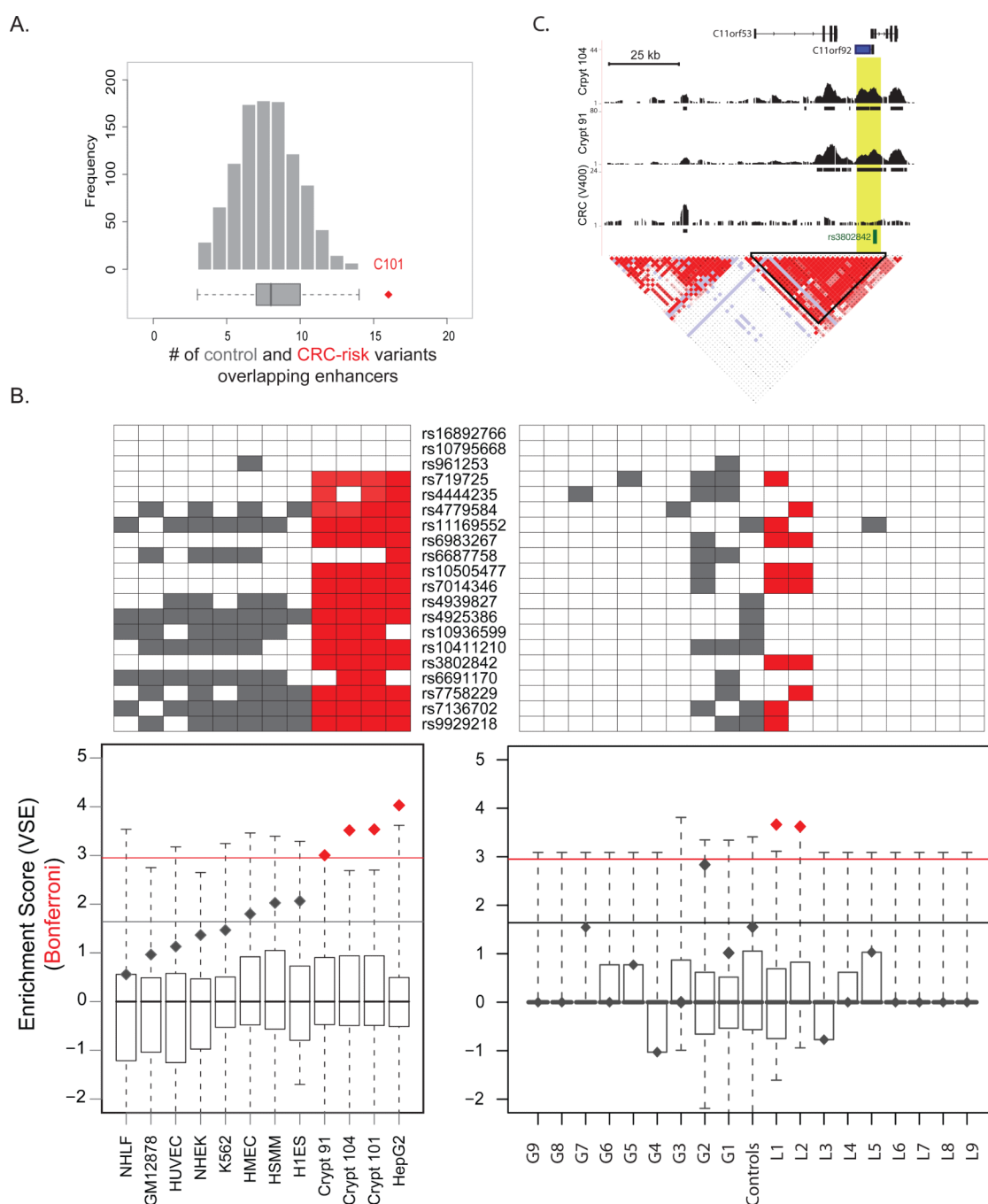


Figure 4.

Colon enhancers and VELs are associated with genetic risk variants for CRC. (A) Results of VSE analysis showing that 16 of 20 CRC-risk SNP clusters map to H3K4me1-marked enhancers in a colon crypt sample (C101, red diamond), compared to a null distribution (grey). (B) (left) The results of VSE analyses testing the association between CRC-risk SNPs and H3K4me1-sites in 10 cell types. The red line represents the significance threshold ($P < 0.001$). The lower horizontal line represents the unadjusted significance threshold. The individual CRC SNPs found to be associated with H3K4me1-enhancers in each cell type are indicated in boxes, above each boxplot. (right) VSE analysis of the CRC-risk SNPs and VELs. Control enhancers are H3K4me1 sites that are unchanged between CRC samples and

crypts. L1 corresponds to unique lost VELs; L2-L9 to losses common to 2-9 lines. G1 corresponds to unique gained VELs; G2-G9 to gains shared between 2-9 lines. (C) Example of a lost VEL directly overlapping a CRC risk SNP shown within the relevant haplotype block structure (red).