Supplementary Information

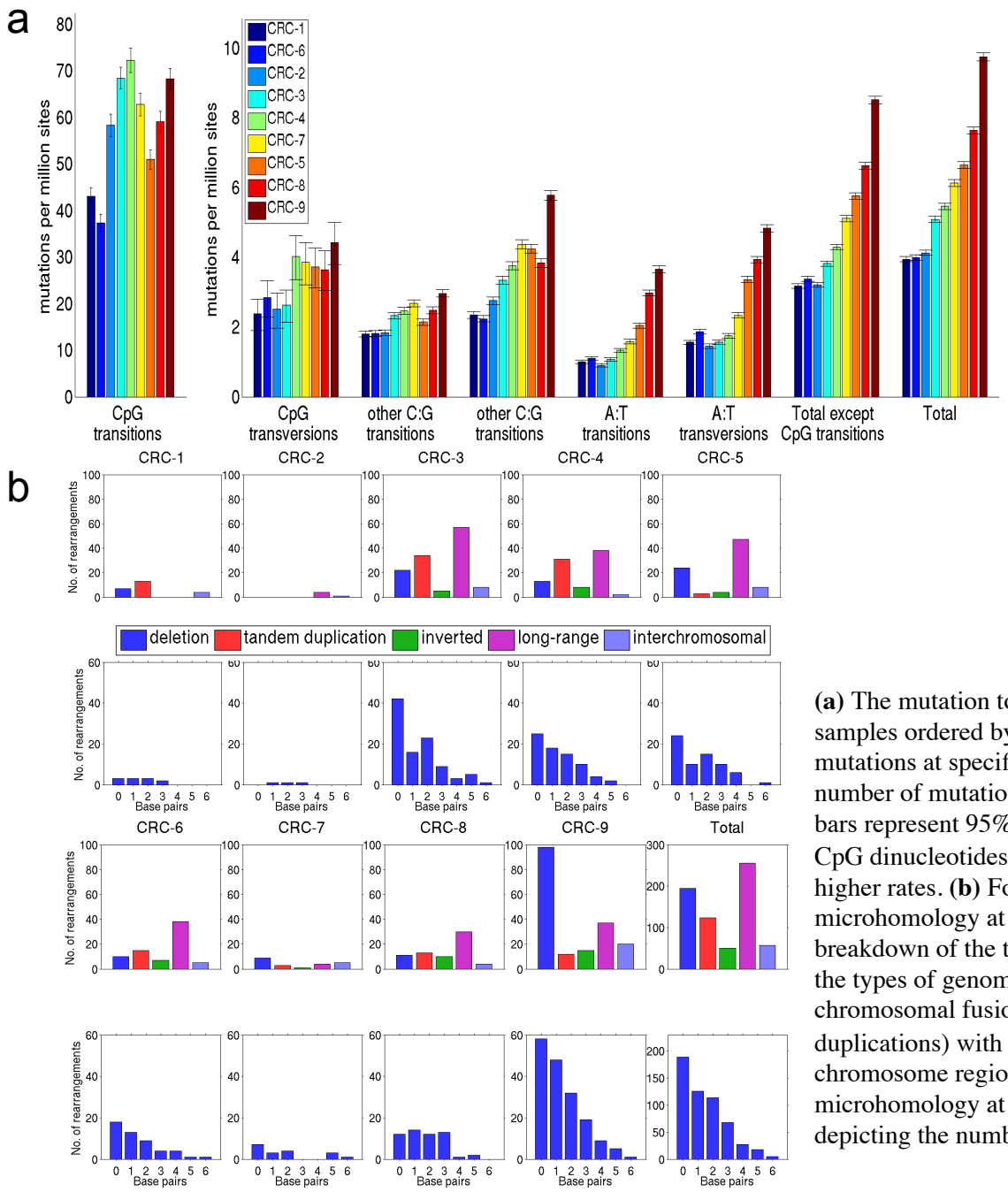
for

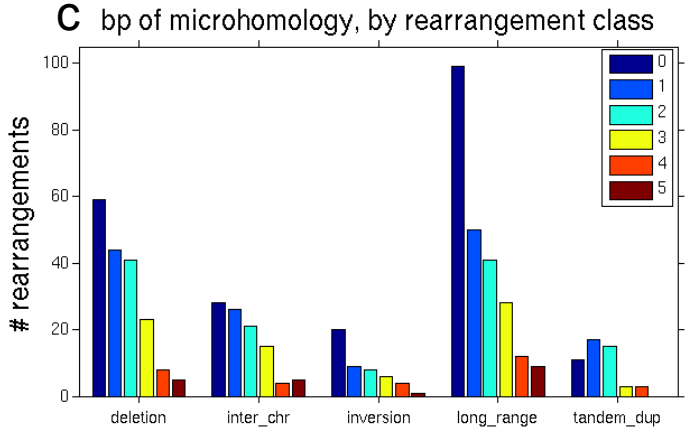
Genomic Sequencing of Colorectal Adenocarcinomas Identifies a Recurrent *VTI1A-TCF7L2* Fusion

Contents

Nature Genetics: doi:10.1038/ng.936

Supplementary Figure 1:
Degree of microhomology and types of intrachromosomal rearrangements

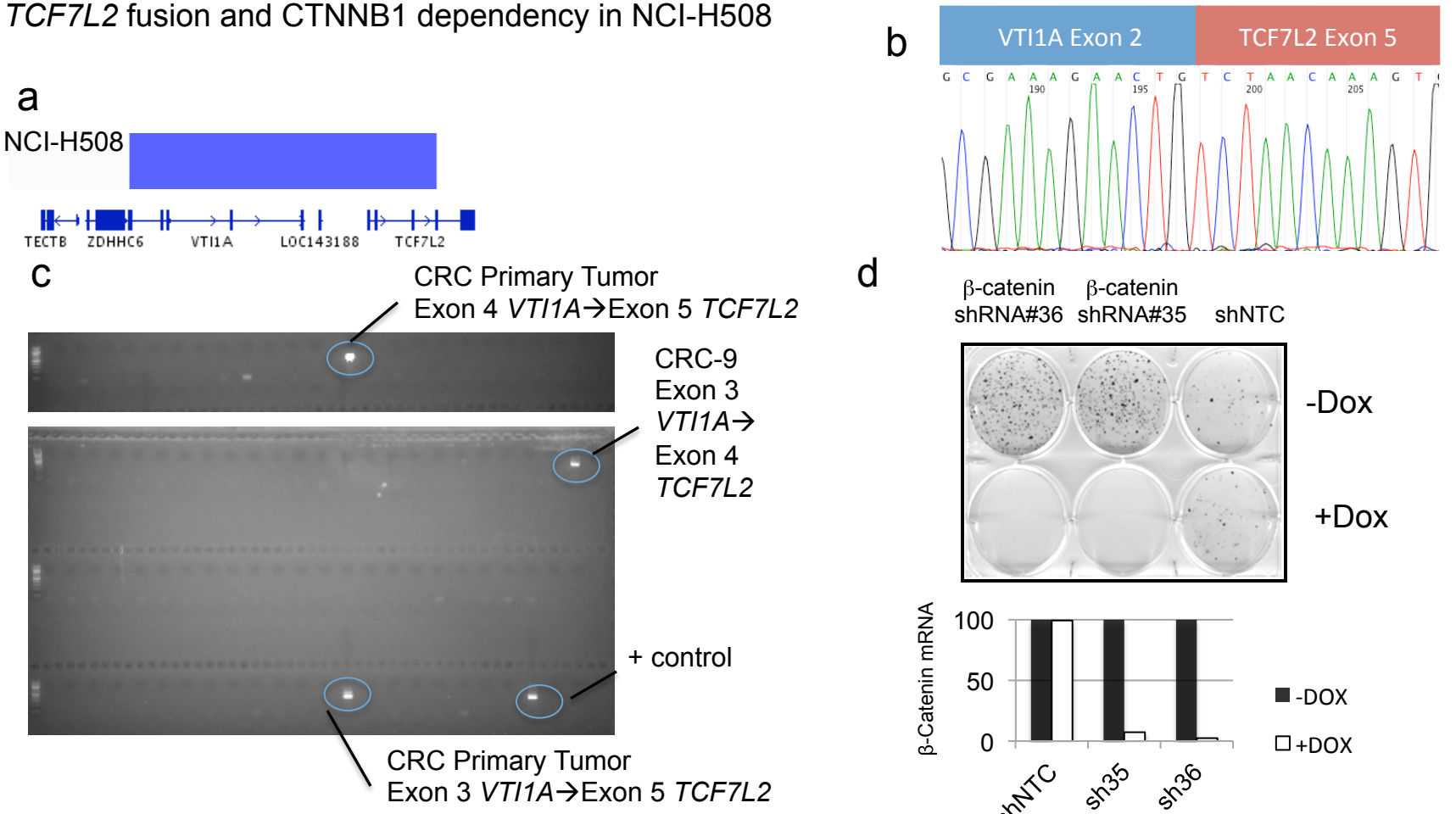




(a) The mutation total rates (right) for each sample are shown, with samples ordered by total mutation rate. Category-specific rates for mutations at specific base contexts are also shown, with rates calculated as number of mutations per million covered sites at risk for mutation. Error bars represent 95% confidence intervals. The rate of C>T transitions at CpG dinucleotides are shown on a separate axis (left) due to their much higher rates. **(b)** For each of the tumor genomes are shown the spectrum of microhomology at sites of breakpoints of rearrangements (bottom) and the breakdown of the types of rearrangements (top). The color scheme shows the types of genomic rearrangements identified (deletions, inter-chromosomal fusions, inversions, long-range deletions, tandem duplications) with long-range event representing those bringing together chromosome regions spanning >1Mb separation. **(c)** The amount of microhomology at breakpoints from each type of rearrangement with colors depicting the number of basses with homology.

Nature Genetics: doi:10.1038/ng.936

Supplementary Figure 2: Validation of the *VTI1A-TCF7L2* fusion and CTNNB1 dependency in NCI-H508



(a) Segmented SNP array-derived derived deletion in chromosome 10 in cell line NCI-H508 showing location of deletion interrupting the adjacent genes *VTI1A* and *TCF7L2*. **(b)** Sequence traces from cloned cDNA from NCI-H508 shows the transcript fusing Exon 2 to *VTI1A* to Exon 5 of *TCF7L2*. **(c)** The agarose gel containing PCR products from a panel of primary colorectal adenocarcinomas screened for the presence of the *VTI1A-TCF7L2* fusion constructs. The marked bands obtained from 3 of 97 samples were subsequently excised, cloned and sequenced with the resulting exon structure of the fusion marked. Of note, these identified fusions all encode an in-frame fusion protein. **(d)** anchorage-independent growth of NCI-H508 cells infected with doxycycline inducible shRNA vectors targeting CTNNB1 (#35 and #36) or a non-targeting control when cultured without (top) or with doxycycline (bottom) to induce shRNA expression. The bottom plot shows the reduction in *CTNNB1* transcript as measured by quantitative real-time PCR.

Nature Genetics: doi:10.1038/ng.936

Supplementary Table 2: *Chart of Key Mutations by Tumor Sequenced*

	TP53	*APC*	*KRAS*	*NRAS*	*PIK3CA*	*SMAD4*
CRC-1	frameshift (heterozygous)	nonsense + frameshift	G13D			
CRC-2		nonsense + frameshift	G13D			
CRC-3						L540R
CRC-4	nonsense (homozygous)	nonsense (homozygous)				
CRC-5	splice-site (homozygous)	nonsense (heterozygous)		G12V	F83Y	
CRC-6	frameshift (heterozygous)*		G13D			G510E
CRC-7	frameshift (heterozygous)	nonsense + nonsense		G13C		
CRC-8		nonsense (homozygous)*	G13D			
CRC-9		nonsense + frameshift	G13D			

* indicates mutation identified upon manual review of sequencing data
(these events not included in overall significance analysis)

Nature Genetics: doi:10.1038/ng.936

Supplementary Table 3: *List of Genes Recurrently Mutated or Rearranged*

Gene	Tumors with Alterations	Gene_length (bp)	Number of rearrangements	Tumors with Rerrangements	coding_length (bp)	Number of Mutations	Tumors with Mutations
APC	7	138,719	0	0	8,532	11	7
IMMP2L	6	899,238	7	5	528	1	1
MACROD2	6	2,057,696	15	6	1,278	0	0
A2BP1	5	1,694,209	17	5	1,257	0	0
FHIT	5	1,502,098	8	5	444	0	0
KRAS	5	45,675	0	0	570	5	5
TP53	5	19,144	0	0	1,182	5	5
TTN	5	281,432	0	0	100,270	7	5
CADM2	3	1,115,445	3	3	1,314	0	0
COL4A5	3	257,622	2	1	5,067	2	2
COL6A3	3	90,196	0	0	9,534	3	3
CSMD3	3	1,214,086	1	1	11,124	3	2
FAM190A	3	1,474,686	5	3	2,703	0	0
PRMT8	3	102,714	2	2	1,185	1	1
AGBL4	2	1,491,100	2	2	1,512	0	0
AKAP9	2	169,798	0	0	11,724	2	2
ANK1	2	243,537	2	2	5,694	0	0
ATG2A	2	22,718	1	1	5,817	1	1
ATRNL1	2	855,373	1	1	4,140	1	1
B4GALT5	2	80,939	1	1	1,167	1	1
C2orf78	2	32,959	0	0	2,769	2	2
CASS4	2	47,229	1	1	2,361	1	1
CNTNAP5	2	890,000	1	1	3,921	1	1
DLC1	2	431,558	0	0	4,587	2	2
DPP6	2	1,101,577	3	2	2,598	0	0
DYM	2	416,908	2	2	2,010	0	0
EMR1	2	52,883	0	0	2,661	2	2
FAT3	2	544,374	0	0	13,674	2	2
FER1L5	2	62,051	0	0	6,282	2	2
FSTL4	2	416,072	3	2	2,529	0	0
GABBR2	2	420,810	0	0	2,826	2	2

Nature Genetics: doi:10.1038/ng.936

HECW2	2	393,359	1	1	4,719	1	1
LPXN	2	51,294	2	2	1,176	0	0
MED12L	2	347,135	2	1	6,438	1	1
MGAM	2	110,868	0	0	5,574	2	2
NBEA	2	730,418	1	1	8,841	1	1
NLRC5	2	94,027	1	1	5,601	1	1
NLRP4	2	45,277	1	1	2,985	1	1
NRAS	2	12,438	0	0	570	2	2
NRG1	2	1,125,291	4	1	1,938	1	1
PAK7	2	301,651	5	2	2,160	0	0
PARK2	2	1,380,245	5	2	1,398	0	0
PDZRN4	2	386,143	0	0	3,111	2	2
PLXNA4	2	525,357	0	0	5,685	2	2
PRKG1	2	1,304,330	2	2	2,061	0	0
PTPRD	2	2,298,264	2	2	5,739	0	0
PUS7L	2	30,185	1	1	2,106	1	1
RALGAPA2	2	319,856	1	1	5,622	1	1
RGMB	2	27,198	1	1	1,437	1	1
RYR2	2	791,587	5	2	14,904	1	1
RYR3	2	555,127	0	0	14,613	2	2
SCN1A	2	84,480	3	1	6,030	1	1
SLC24A3	2	510,252	8	2	1,935	0	0
SMAD4	2	54,827	0	0	1,659	2	2
SPAG16	2	1,126,110	2	2	1,896	0	0
STXBP5L	2	516,559	0	0	3,561	2	2
SYNGAP1	2	33,620	1	1	4,032	1	1
TACC1	2	124,843	5	2	2,418	0	0
TBC1D9B	2	45,786	0	0	3,753	2	2
TCF7L2	2	217,426	2	1	1,809	1	1
TMPRSS11F	2	76,672	1	1	1,317	1	1
TOPBP1	2	61,289	0	0	4,569	2	2
TTC28	2	701,850	4	2	7,446	0	0
VKORC1L1	2	81,544	2	2	531	0	0
WNK1	2	158,394	0	0	7,149	2	2
ZFHX3	2	275,749	2	2	11,112	0	0
ZHX2	2	192,855	2	2	2,514	0	0
ZMAT4	2	367,228	6	2	690	0	0

Nature Genetics: doi:10.1038/ng.936

Supplementary Table 5: *Primers Used in Validation Studies*

Screening for VTI1A-TCF7L2 Fusion Across cDNA set

First Round PCR	
5' UTR VTI1A	TTTCCCTGACCTAGGCTTTG
Exon 6 TCF7L2	GGATGGGGGATTTGTCCTAC
Nested PCR	
Exon 1 VTI1A	CCGACTTCGAAGGTTACGAG
Exon 5 TCF7L2	TACGTCGGCTGGTAAGTGTG

Real-Time PCR primers for Quantification of VTI1A-TCF7L2 knockdown

qRT-PCR set1A	TGGTTGCAAATGTGGAGAAA
qRT-PCR set1B	GCACCACTGGCACTTTGTTA
qRT-PCR set2A	CAGCTTGAAGAAGCGAAAGAA
qRT-PCR set2B	ACGTGATAAGAGGCGTGAGG

Supplementary Note:

Human Research Subjects

The patients whose cancer and germline DNA were sequenced in this report were recruited at the time of their already scheduled surgery as part of the treatment for their colorectal cancer diagnosis. Patients were subjected to informed consent for the collection of their tumor for use in research studies, including those involving genomic analysis. Among patients whose tumors were collected and provided to our team as part of collaborative research, cases were selected based upon availability of DNA and pathologic and computational estimates of tumor purity and ploidy.

Nature Genetics: doi:10.1038/ng.936