SUPPLEMENTARY INFORMATION

The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling

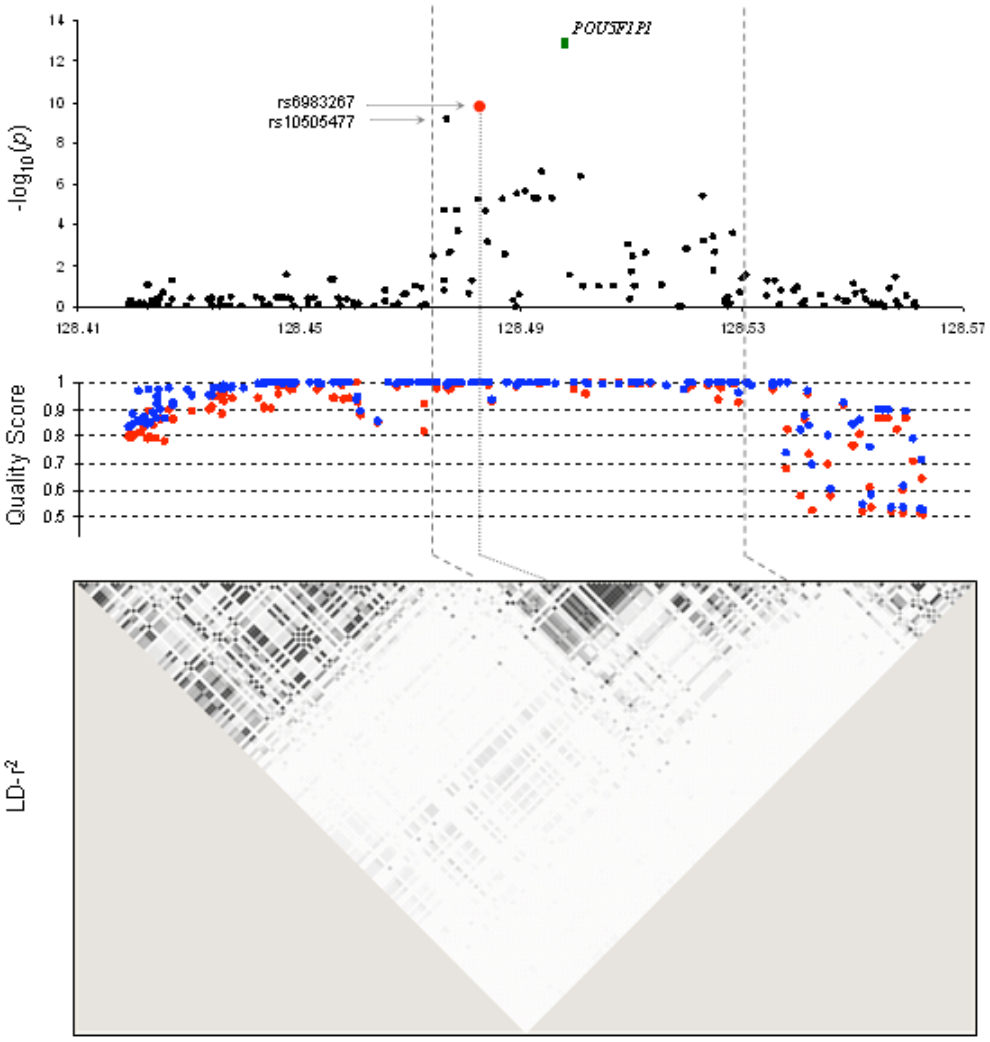
Sari Tuupanen[1], Mikko Turunen[2], Rainer Lehtonen[1], Outi Hallikas[2], Sakari Vanharanta[1,7], Teemu Kivioja[2,6], Mikael Björklund[2], Gonghong Wei[2], Jian Yan[2], Iina Niittymäki[1], Jukka-Pekka Mecklin[3], Heikki Järvinen[4], Ari Ristimäki[5], Mariachiara Di-Bernardo[8], Phil East[9], Luis Carvajal-Carmona[9], Richard S Houlston[8], Ian Tomlinson[9], Kimmo Palin[6, 10], Esko Ukkonen[6], Auli Karhu[1], Jussi Taipale[2*], Lauri A. Aaltonen[1]*

[1] Department of Medical Genetics, Genome-Scale Biology Research Program, Biomedicum Helsinki, P.O. Box 63, 00014 University of Helsinki, Helsinki, Finland
[2] Institute of Biomedicine, Genome-Scale Biology Research Program, Biomedicum Helsinki, P.O. Box 63, 00014 University of Helsinki, Helsinki, Finland and Department of Molecular Medicine, National Public Health Institute
[3] Department of Surgery, Jyväskylä Central Hospital, 40620 Jyväskylä, Finland
[4] The Second Department of Surgery, Helsinki University Hospital, 00029 Helsinki, Finland
[5] Genome-Scale Biology Research Program, P.O. Box 63, 00014 University of Helsinki, Helsinki, Finland; Department of Pathology, HUSLAB and Haartman Institute, Helsinki University Central Hospital, Helsinki, Finland; Department of Pathology, University of Oulu and Oulu University Hospital, Oulu, Finland
[6] Department of Computer Science, P.O. Box 68, FIN-00014 University of Helsinki, Helsinki, Finland
[7]Present address: Cancer Biology and Genetics Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA
[8] Section of Cancer Genetics, Institute of Cancer Research, Sutton. Surrey. SM2 5NG. UK
[9] Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Headington, Oxford OX3 7BN
[10] Present address: Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1HH, UK

CONTENTS

Nature Genetics: doi:10.1038/ng.406

Supplementary Figure 1



Supplementary Figure 1. Association and LD structure at the 8q24 locus. Shown in the upper panel is the joint analysis (single-locus test of association) of CORGI and NSCCG1 imputed HapMap SNPs. Plotted in the middle panel is the quality score of each of the imputed SNPs; red for SNPs imputed in CORGI and blue for NSCCG1. In the lower box are estimated statistics of the square of the correlation coefficient (r2), derived from HapMap project data using Haploview software (v4.0). The values indicate the LD relationship between each pair of SNPs: the darker the shading, the greater extent of LD. The predicted transcript *POUF51P1* in the local area is shown. Positions are those of UCSC March 2006 assembly; NCBI build 36.1.

Nature Genetics: doi:10.1038/ng.406

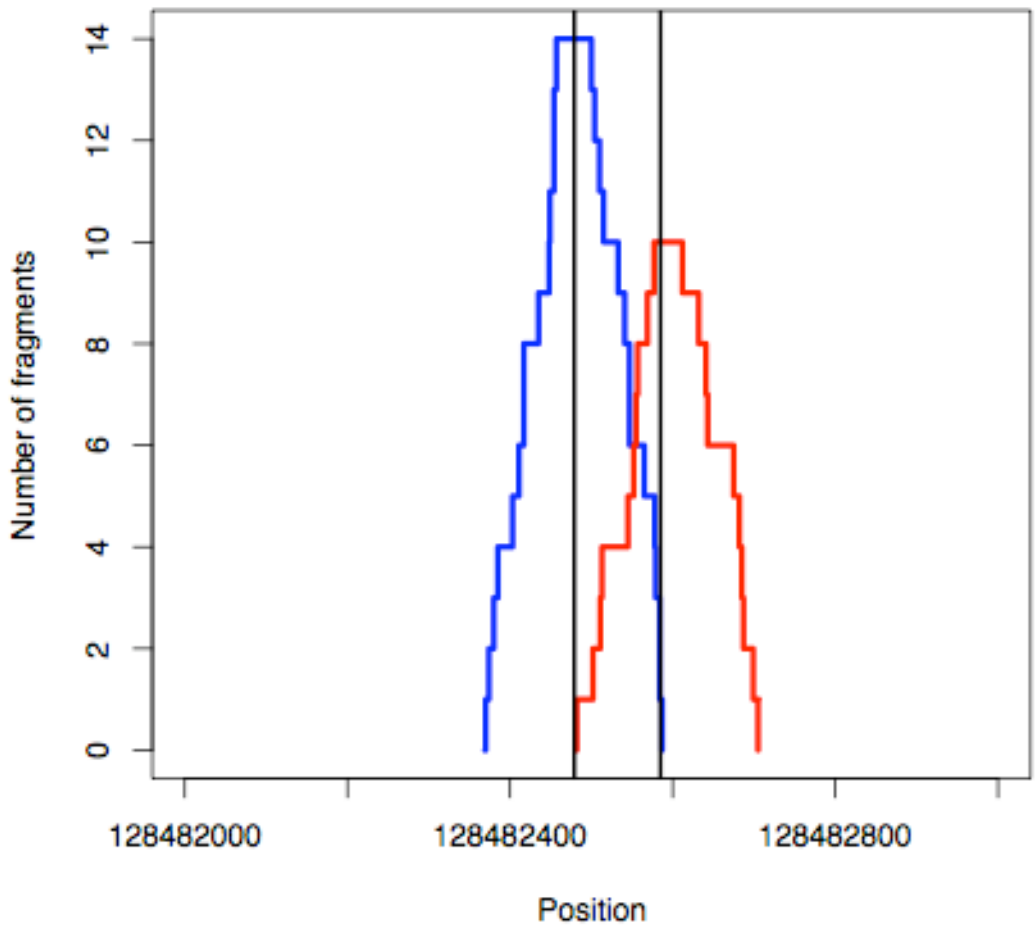
Supplementary Figure 2

Binding site # Match	Score	Binding site ID	Binding site position match Mus Musculus	Human	
D[10201][14041]	=39.81	Hunchback.pfm	(133546,133555)	<=> (164894,164903)	+
D[10204][14043]	=45.66	jtcf4.pfm	(133634,133642)	<=> (164982,164990)	-
D[10205][14044]	=75.59	Hunchback.pfm	(133663,133672)	<=> (165011,165020)	-
D[10208][14047]	=113.92	HMG-IY.pfm	(133682,133697)	<=> (165030,165045)	-
D[10209][14050]	=136.71	Dof2.pfm	(133729,133734)	<=> (165077,165082)	-
D[10212][14053]	=173.55	jtcf4.pfm	(133740,133748)	<=> (165088,165096)	-
D[10213][14054]	=213.45	Su(H).pfm	(133759,133774)	<=> (165107,165122)	-
D[10214][14055]	=246.69	SP1.pfm	(133789,133798)	<=> (165137,165146)	+
D[10217][14057]	=248.96	AP2alpha.pfm	(133846,133854)	<=> (165196,165204)	-
D[10218][14058]	=285.77	SRY.pfm	(133876,133884)	<=> (165226,165234)	+
D[10223][14062]	=317.95	c-MYB1.pfm	(133900,133907)	<=> (165250,165257)	-
D[10228][14065]	=334.29	MYB-ph3.pfm	(133957,133965)	<=> (165307,165315)	-
D[10229][14066]	=368.87	SPI-1.pfm	(133970,133975)	<=> (165320,165325)	+

Supplementary Figure 2. List of the transcription factor binding sites that are shown in the Figure 2a. The EEL predicted enhancer element in the Figure 2a contains 13 transcription factor binding sites which are indicated by the boxes in the Figure 2a. Shown here are the transcription factors, locations of the binding sites relative to the input sequences and EEL scores.

Nature Genetics: doi:10.1038/ng.406

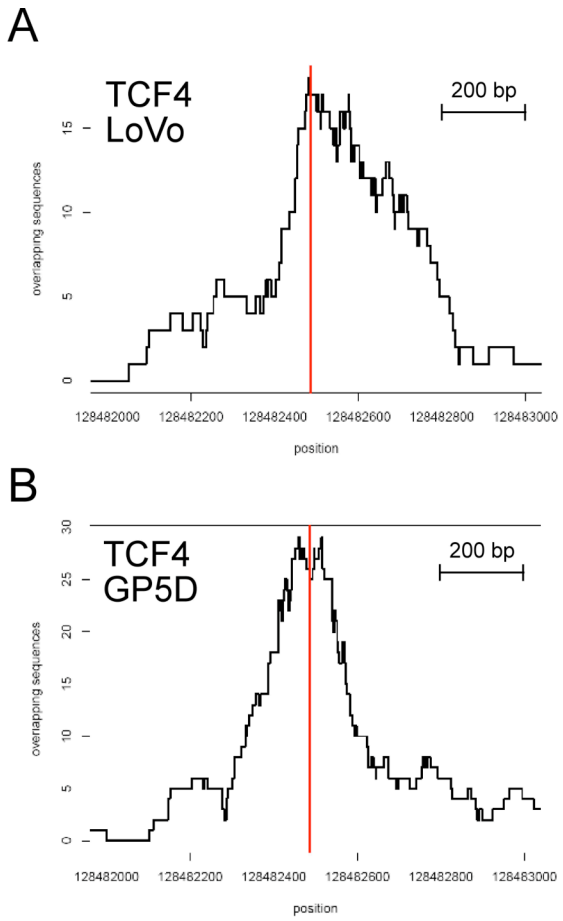
Supplementary Figure 3



Supplementary Figure 3. Analysis of TCF4 ChIP-sequenced fragments from LoVo cells indicates that TCF4 site at rs6983267 is occupied. Sequenced fragments containing only the TCF4 site at rs6983267 (blue) or only the weaker TCF4 site 106 bp downstream (red). Note that occupancy of the weaker TCF4 site cannot explain the observed peak at rs6983267. The TCF4 sites are indicated by black vertical lines.

Nature Genetics: doi:10.1038/ng.406

Supplementary Figure 4



Supplementary Figure 4. Chip-by-sequencing identifies the SNP region as a strong binding site for TCF4. Sequence histogram of the TCF-4 ChIP in LoVo cells (**A**) and GP5D (**B**) around the SNP rs6983267 site in the chromosome 8. The histogram in (**A**) is identical with the data shown in Fig. 3e and is shown to facilitate comparison of the transcription factor binding in the SNP region. The red vertical bar indicates the SNP site. Note that the sequencing reads are highly concentrated within a <200 bp region around the SNP site.

Nature Genetics: doi:10.1038/ng.406

Supplementary Figure 5

Supplementary Figure 5. *LacZ* embryos with mutated TCF4 sites. Embryos injected with a construct were the two conserved TCF sites in MYC-335 have been mutated show only ectopic staining. Of the 33 embryos analyzed in total, five were genotyped positive for the transgene. Three of those showed *LacZ* staining patterns, but the patterns are inconsistent.

Nature Genetics: doi:10.1038/ng.406

Supplementary Table 1

Supplementary Table 1. Highest-scoring predicted enhancer-elements near the *MYC* gene (from genome-wide EEL alignment)

rank	start position (relative to MYC TSS)	length (bp)	EEL score
1	-334506*	1255	323
2	-89662	448	305
3	-255343	3441	277
4	-405297	765	265
5	-345319	1711	248
6	487824	285	247
7	-92160	478	238
8	-441366	924	222
9	-56284	558	203
10	-36280	581	193

EEL score is based on 180 different TF binding sites. Approximately 0.8 and 4 elements / aligned gene are found with EEL scores of more than 323 and 193, respectively.
*MYC-335

Nature Genetics: doi:10.1038/ng.406

Supplementary Table 2

Supplementary Table 2. Oligonucleotides used in this study.

Method	Primer name/Description	Sequence
TF binding assay	consensus TCF4 binding site	cagcaCATCAAAGGcactg
	G allele	cagcaGATGAAAGGcactg
	T allele	cagcaGATGAAAGTcactg
	scrambled	cagcaAGAAGTGGAcactg
EMSA	TCFconsensus	ACGCTACcctttgatgTTACCGT
	TCFconsensusR	ACGGTAAcatcaaaggGTAGCGT
	bio-TCFconsensus	ACGCTACcctttgatgTTACCGT
	bio-TCFconsensusR	ACGGTAAcatcaaaggGTAGCGT
	TCFscrambled	ACGCTACtggacttctTTACCGT
	TCFscrambledR	ACGGTAAagaagtccaGTAGCGT
	TCFsnpG	ACGCTACcctttcatcTTACCGT
	TCFsnpGR	ACGGTAAgatgaaaggGTAGCGT
	TCFsnpT	ACGCTACactttcatcTTACCGT
	TCFsnpTR	ACGGTAAgatgaaagtGTAGCGT
Luciferase assay	Cloning	GGCGCAGGTACCGAAACCACCTTGGACTGGAA
		CAGGAAGATATCCAAGCACTTGGAGCACACAT
ChIP	Weak positive control (cyclinD)	AGCGGGGCGATTTGCATTTCTA
		AGACCCAAAAGCCATCCCTGA
	strong positive control CHR7 5890255	TTACTCTCTCCCTCTTTCATCTGTTGC
		GGCAGGCTCACTTTCCCACA
	Negative control 1	CTTGCCCCTCTATTCCCCACCAAC
		CCCCTTCCCATCACTCACTGTCC
	Negative control 2	CTCCTCCTCCCCTCTGGTCTTTCC
		CCAATCTGTTCTGCCCACTCCATC
	rs6983267-1	ACGAATAAACTCTCCTCCTACCACTAAG
		TATACACAGCCCAGTCTAAGGCCC
	rs6983267-2	GGGACGAATAAACTCTCCTCCT
		CTCCCTCCCCCACATAAAAT
LacZ mouse	Cloning	ACTGAAGCTTCATGGCAGAGGTCTTGAAAGCATGTATAA
		CAGTGTCGACGACAAACCTCAGAGGGATGACACTTCTTA
Myc in situ	Primers for probe	ATTTAGGTGACACTATAGAAGAGTCAGTGGTCTTTCCCTACCCG
		TAATACGACTCACTATAGGGAGATCGTTTCCTCAATAAGTCCTTTTC

Nature Genetics: doi:10.1038/ng.406

Supplementary Note

Study subjects

Finnish samples: The Finnish CRC / normal tissue samples were chosen from a series collected since 1994 and had been obtained after informed consent and Ethical Review Board approval (Ethics committees of Department of Medical Genetics, University of Helsinki and Helsinki University Central Hospital). This sample series is described in detail in previous reports[43,44]. Haploblock analysis was based on SNP data from 265 anonymous Finnish control individuals.

UK samples: Imputed data were generated on two case-control series: **CORGI**: 619 CRC cases (279 males, 340 females) ascertained through the Colorectal Tumour Gene Identification (CoRGI) consortium. All had at least one first-degree relative affected by CRC. Controls (N=932; 422 males, 510 females) were spouses or partners unaffected by cancer and without a personal family history (to 2nd degree relative level) of colorectal neoplasia. All cases and controls were of white UK ethnic origin. **NSCCG1**: 2,863 CRC cases (1,196 males, 1,667 females; mean age at diagnosis 59.3 years; SD ± 8.7) ascertained through two ongoing initiatives at the Institute of Cancer Research/Royal Marsden Hospital NHS Trust (RMHNHST) from 1999 onwards - The National Study of Colorectal Cancer Genetics (NSCCG) and the Royal Marsden Hospital Trust/Institute of Cancer Research Family History and DNA Registry. A total of 2,838 healthy individuals were recruited as part of ongoing National Cancer Research Network genetic epidemiological studies, NSCCG (1,219), the Genetic Lung Cancer Predisposition Study (GELCAPS) (1999-2004; n=911), and the Royal Marsden Hospital Trust/Institute of Cancer Research Family History and DNA Registry (1999-2004; n=708). These controls (1,136 males, 1,702 females; mean age 59.8 years; SD ± 10.8) were the spouses or unrelated friends of patients with malignancies. None had a personal history of malignancy at time of ascertainment. All cases and controls were British and of European descent, and there were no obvious differences in the demography of cases and controls in terms of place of residence within the UK.

Nature Genetics: doi:10.1038/ng.406