Supplementary Appendix File

Comprehensive assessment of variation at the Transforming Growth Factor Beta Type I Receptor locus and colorectal cancer predisposition

Luis G Carvajal-Carmona[1], Mike Churchman[1], Carolina Bonilla[2,3], Axel Walther[1], Jeremie H. Lefèvre[2,3], David Kerr[3], Malcolm Dunlop[4], Richard Houlston[5], Walter Bodmer[2,3], Ian Tomlinson[1]

[1]Molecular and Population Genetics Laboratory, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, United Kingdom

[2]Cancer and Immunogenetics Laboratory, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DZ, United Kingdom

[3]Department of Clinical Pharmacology, University of Oxford, OX3 7DQ, United Kingdom

[4]Colon Cancer Genetics Group, Medical Research Council, Human Genetics Unit, Western General Hospital, Edinburgh EH4 2XU, United Kingdom

[5]Section of Cancer Genetics, Institute of Cancer Research, Sutton SM2 5NG, United Kingdom

Correspondence to: Luis G Carvajal-Carmona, Molecular and Population Genetics Laboratory, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, United Kingdom. Tel: +44 (0) 1865 287500. Fax: +44 (0) 1865 287501. Email: luis@well.ox.ac.uk

Classification: Biomedical Sciences, Genetics

Supplementary Table 1. Association between CRC risk and *TGFBR1*(chr9:100,907,233-100,956,294) haplotypes defined by tagging SNPs. Haplotype blocks were assigned by the solid spine method as implemented in Haploview and a block chr9:100,897,000-100,973,999 contained *TGFBR1*. These eight haplotypes are defined by alleles at the tagging SNPs rs7034462, rs10819634, rs12686783, rs10988705, rs10739778, rs20226881, rs1571590, rs10512263 and rs7850895. Haplotypes 3 (TCTGCGATC) and 7 (CCCGCTGTT), shown in bold, contain the 6A allele at rs11466445. The VQ58 samples were typed with the Illumina Hap300 array. This array does not contain the SNPs rs10988705, rs10739778, rs20226881 and rs1571590. Therefore, haplotypes 1, 5 and 6 could not be distinguished from each other in the study, and the frequencies presented for haplotype 1 represent a pool of these three haplotypes. We also defined the haplotypes using two additional methods (confidence intervals and four-gamete rule) incorporated into Haploview and failed to detect any significant association between haplotypes in this region and CRC risk.

Haplotype	Corgi			Scotland			VQ58[3]		
	Frequency		P-	Frequency		P-	Frequency		P-
	Cases	Controls	value	Cases	Controls	value	Cases	Controls	value
CCCGAGATT	0.523	0.537	0.391	0.523	0.527	0.816	0.617	0.608	0.502
CTCGCTGTT	0.188	0.173	0.212	0.191	0.183	0.541	0.188	0.183	0.658
TCTGCGATC	0.081	0.085	0.613	0.090	0.080	0.277	0.095	0.088	0.422
CCCGAGACT	0.084	0.072	0.167	0.066	0.074	0.325	0.070	0.075	0.553
CCCGCGATT	0.045	0.058	0.074	0.057	0.059	0.771			
CCCAAGATT	0.043	0.048	0.524	0.045	0.045	0.971			
CCCGCTGTT	0.017	0.012	0.221	0.016	0.018	0.595	0.018	0.020	0.497
TCTGCTATT	0.014	0.011	0.368	0.010	0.011	0.694	0.009	0.014	0.075

Supplementary Table 2. Tests of gene-gene interaction at SNPs in the TGF-beta pathway. Genotype frequencies at rs11466445 and rs4779584 (A), rs961253 (B), rs4444235 (C) and rs4939827 (D), where 0= common homozygote, 1=heterozygote, 2=rare homozygote. Data from the 3 case series have been pooled for this analysis. Analysis of each sample series individually also produced no evidence of gene-gene interactions.

A)

		rs4779584		Total
rs11466445	0	1	2	
0	1453	788	142	2383
1	318	167	13	498
2	85	44	9	138
Total	1856	999	164	3019
			χ^2_4=9.46	P=0.05

B)

		rs961253		Total
rs11466445	2	1	0	
0	557	1119	707	2383
1	112	219	167	498
2	35	64	39	138
Total	704	1402	913	3019
			χ^2_4=3.41	P=0.49

C)

		rs4444235		Total
rs11466445	2	1	0	
0	575	1193	613	2381
1	108	246	144	498
2	37	63	37	137
	720	1502	794	3016
Total			χ^2_4=3.70	P=0.45

D)

		rs4939827		Total
rs11466445	2	1	0	
0	439	1182	761	2382
1	90	231	176	497
2	29	65	44	138
	558	1478	981	3017
Total			χ^2_4=2.99	P=0.56

Supplementary Table 3. *TGFBR1* primers used for rare variant screening.

Exon	Primer name	Primer sequence (5'-3')
1	TGFBR1-1F	CTCCGAGCAGTTACAAAGGG
	TGFBR1-1R	AAAGAGCAGGAGCGAGCC
2	TGFBR1-2F	AGAAATTGTGTGATAATAGGATCAAG
	TGFBR1-2R	CTTGCCTCTAAACGGAATGAG
3	TGFBR1-3F	GGGTCACTCATTAGTGCCTATC
	TGFBR1-3R	ACATTAAGAGATTTTAGGAATGCTATC
4	TGFBR1-4F	GCAGTGTGTGACTCAGGATTG
	TGFBR1-4R	TATAGGCATGAGCCACCAGG
5	TGFBR1-5F	TTTGGGTTGGGAGAAGAGAC
	TGFBR1-5R	TTTAAGCTGAGTTTCAGCAATG
6	TGFBR1-6F	CCAAATATGGCAGTAAGGGG
	TGFBR1-6R	TCGGTGACATCCTGTTTCAG
7	TGFBR1-7F	GGTGATCTTTTAATGCCTTGG
	TGFBR1-7R	TCCATGGCCCTTTCAATG
8	TGFBR1-8F	TCCAGACCAATGGAAAATGG
	TGFBR1-8R	ACACTGCTGCAAAAGGAAGC

Supplementary Table 4. Association between genotyped (in bold) and imputed SNPs at the *TGFBR1*-containing haplotype block and CRC risk. SNPs without the rs prefix have been recently identified by the 1000genome project and their code refers to the chromosomal (in the case, located on chromosome 9) and physical location (number after the "-") of the SNP. More details about these SNPs can be found at www.1000genomes.org). SNPs with nominal significant p-values in each one of the studies are highlighted in bold and red.

Supplementary Table 4 is available at: http://www.well.ox.ac.uk/~luis/

Supplementary Table 5. Association between SNPs proximal (A) and distal to the TGFBR1 haplotype block and CRC risk. SNPs without the rs prefix have been recently identified by the 1000genome project and their code refers to the chromosomal and physical location (number after the "-") of the SNP. More details about these SNPs can be found at www.1000genomes.org. The distal SNP rs410180 is highlighted in bold and red in part B of the table. NI refers to non-informative imputation. SNPS with overall p values <0.01 are highlighted in bold

Supplementary Table 4 is available at: http://www.well.ox.ac.uk/~luis/

Supplementary Table 6. Association between SNPs at and around TGFB1 , TGFB2, TGFB3 , TGFBR2 and TGFBR3 and CRC risk. SNPs without the rs prefix have been recently identified by the 1000genome project and their code refers to the chromosomal and physical location (number after the "-") of the SNP. More details about these SNPs can be found at www.1000genomes.org. No SNP at or around TGFB2, TGFB3 or TGFBR3 had overall p values <0.01. Only 20 SNPs at TGFB1 (n=7) and TGFBR2 (n=13) had overall p values <0.01. These latter SNPs are highlighted in bold

Supplementary Table 4 is available at: http://www.well.ox.ac.uk/~luis/