Identification of candidate susceptibility genes for colorectal cancer through eQTL analysis

Adrià Closa, David Cordero, Xavier Solé, Marta Crous-Bou, Laia Paré-Brunet, Rebeca Sanz-Pamplona, Antoni Berenguer, Susana Aussó, Adriana López-Dóriga, Henar Alonso, Victor Moreno.

Unit of Biomarkers and Susceptibility, Cancer Prevention and Control Program, IDIBELL – Catalan Institute of Oncology and CIBERESP. Barcelona, Spain.

Introduction

- To date genome-wide association studies have identified 26 SNPs in 23 susceptibility loci for colorectal cancer (CRC). Most of these SNPs are located in intergenic positions and are considered just markers since their functional roles are generally unknown.
- The identification of the relevant genes responsible for these associations is important, since they may be considered targets for developing new strategies for prevention or therapy.

Objectives

- Identify candidate genes responsible for CRC risk susceptibility using cis and trans-eQTL analysis in two series of samples, one of healthy colonic mucosa and other of normal mucosa adjacent to colon cancer.
- We have also analyzed the effect in tumor tissue, but these are not used for discovery, since as well known, the gene expression profiles in tumors are highly altered that may introduce both false positive and false negative results.

Materials & Methods

- Normal mucosa from 100 patients with colon cancer and 50 healthy donors that underwent colonoscopy have been included in the COLONOMICS project (www.colonomics.org).
- Gene expression data was generated with the Affymetrix Human Genome U219 Array Plate platform. Genotypes were extracted from the Affymetrix Genome-Wide Human SNP 6.0 array
- In total, 26 GWAS SNPs plus 4 additional risk SNPs were analyzed.
- SNP calling had been performed with the Corrected Robust Linear Model with Maximum Likelihood Classification (CRLMM). Genotypes for 18 of GWAS SNP were not available in the array, and were imputed using IMPUTE2 (v2.2.2) software after haplotyping with PHASEIT (v1.ESHG).
- We reduced the expression data to one unique value for gene using principal component analysis (PCA).

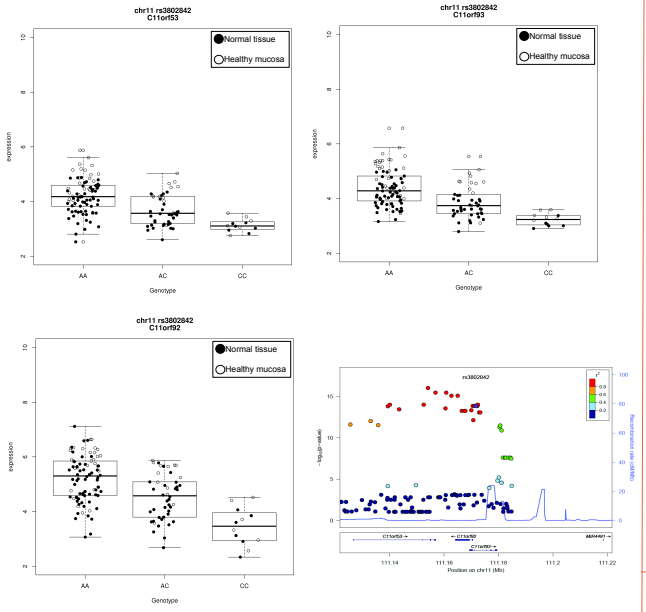
Results

The analysis of 30 GWAS SNPs identified three loci with five candidate genes (rs3802842 11q23.1: C11orf53, C11orf93, C11orf92; rs7136702 12q13.1: DIP2B; and rs5934683 Xp22.3: SHROOM2).

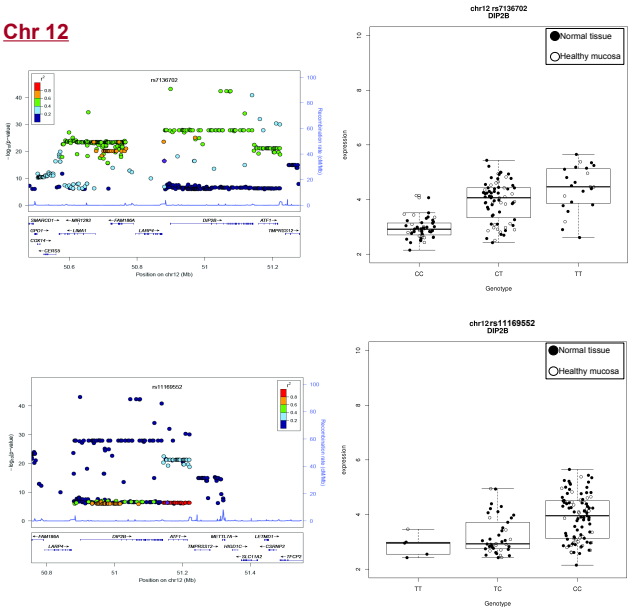
For chromosome 11 we analyze other 27 SNP that are in high LD (>0.8 R-Squared) with the rs3802842 and calculated the partial correlation with a summary expression value for the three orf genes derived from a PCA. SNP rs7130173 was the most significant in the region and a conditional analysis showed that it dominated the eQTL association.

For chromosome 12 a similar analysis identified rs61927768, which is located 40bp from the TSS of DIP2B, as the most relevant SNP in the region (p-value 2,2e-16).

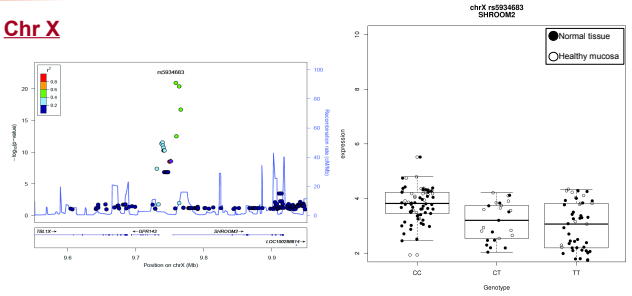
Chr 11



Chr 12



Chr X



Locus	SNP	Gene Symbol	Gene Name	r	P-value	Bonferroni
11q23.1	rs3802842	C11orf92	Chromosome 11 open reading frame 92	-0,57	1,30E-016	Y
11q23.1	rs3802842	C11orf53	Chromosome 11 open reading frame 53	-0,59	8,20E-018	Y
11q23.1	rs3802842	C11orf93	Chromosome 11 open reading frame 93	-0,62	4,40E-021	Y
12q13.12	rs11169552	DIP2B	DIP2 disco-interacting protein 2 homolog B (Drosophila)	0,41	1,10E-007	Y
12q13.12	rs7136702	DIP2B	DIP2 disco-interacting protein 2 homolog B (Drosophila)	0,63	5,38E-008	Y
Xp22.3	rs5934683	SHROOM2	Shroom family member 2	-0,47	1,50E-010	Y

Table 1. Summary table from significant QTL which pass Bonferroni correction.

Conclusions

We have identified candidate genes in three GWAS loci that are strong eQTLs. These findings are relevant since open the path for further functional studies that may reveal intervention strategies for preventing CRC.