Weak association between the $TGFBR1^*$6A variant and colorectal cancer: a family-based association study on non-syndromic family members from Australia and Spain.

Jason P Ross$^{1,2}$, Linda J Lockett$^{1,2}$, Bruce Tabor$^{1,3}$, Ian W Saunders$^{1,4}$, Graeme P Young$^{5}$, Finlay Macrae$^{6}$, Ignacio Blanco$^{7}$, Gabriel Capella$^{7}$, Glenn S Brown$^{1,2}$, Trevor J Lockett$^{1,2}$ and Garry N Hannan$^{1,2}$

$^1$CSIRO Preventative Health Flagship

$^2$CSIRO Animal, Food and Health Sciences, North Ryde, NSW, Australia

$^3$CSIRO Computational Informatics, North Ryde, NSW, Australia

$^4$CSIRO Computational Informatics, Glen Osmond, SA, Australia

$^5$Flinders Centre for Cancer Prevention and Control, Flinders University, Bedford Park, Adelaide, SA, Australia

$^6$Department of Medicine, University of Melbourne, and Colorectal Medicine and Genetics, The Royal Melbourne Hospital, Melbourne, VIC, Australia.

$^7$Cancer Genetic Counseling Program and Translational Research Laboratory, Institut Català d'Oncologia-IDIBELL and University of Barcelona, L'Hospitalet de Llobregat, 08907 Barcelona, Spain

$^\S$ Corresponding author

Email addresses:

JPR: jason.ross@csiro.au

LJL: linda.lockett@csiro.au

BT: bruce.tabor@csiro.au

IWS: ian.saunders@csiro.au
Abstract

Background

Genome-wide linkage studies have identified the 9q22 chromosomal region as linked with colorectal cancer (CRC) predisposition. A candidate gene in this region is transforming growth factor β receptor 1 (TGFBR1). Investigation of TGFBR1 has focused on the common genetic variant rs11466445, a short exonic deletion of nine base pairs which results in truncation of a stretch of nine alanine residues to six alanine residues in the gene product. While the six alanine (*6A) allele has been reported to be associated with increased risk of CRC in some population based study groups this association remains the subject of robust debate. To date, reports have been limited to population-based case-control association studies, or case-control studies of CRC families selecting one affected individual per family. No study has yet taken advantage of all the genetic information provided by multiplex CRC families.

Methods

We have tested for an association between rs11466445 and risk of CRC using several family-based statistical tests in a new study group comprising members of non-
syndromic high risk CRC families sourced from three familial cancer centres, two in Australia and one in Spain.

**Results**

We report a finding of a nominally significant result using the pedigree-based association test approach (PBAT; \( p = 0.028 \)), while other family-based tests were non-significant, but with a \( p \)-value < 0.10 in each instance. These other tests included the Generalised Disequilibrium Test (GDT; \( p = 0.085 \)), parent of origin GDT Generalised Disequilibrium Test (GDT-PO; \( p = 0.081 \)) and empirical Family-Based Association Test (FBAT; \( p = 0.096, \) additive model). Related-person case-control testing using the “More Powerful” Quasi-Likelihood Score Test did not provide any evidence for association (\( M_{QLS} \); \( p = 0.41 \)).

**Conclusions**

The results suggest there is only marginal evidence for an association between the \( TGFBR1^*6A \) allele and CRC risk in these families. This weak support for an increase in risk in CRC predisposed families is in agreement with recent meta-analyses of case-control studies, which estimate only a modest increase in sporadic CRC risk among \( 6^*A \) allele carriers.

**Keywords**

\( TGFBR1, 6^*A, \) rs11466445, colorectal, cancer, hereditary

**Background**

Several genome-wide studies [1-3] have provided evidence for significant genetic linkage between a chromosomal region on 9q22 and an increased risk of colorectal cancer (CRC). A further study confirmed this linkage signal and fine-mapped the
association to a region centred around 98.15 Mb [4]. Biologically, this chromosomal region houses several interesting candidate CRC susceptibility genes including \textit{PTCH1}, \textit{XPA}, \textit{GALNT12} and \textit{TGFBR1} [5]. Follow up efforts have particularly focused on \textit{TGFBR1} (hg19 coordinates, chr9:101.87-101.92 Mb), but with largely inconclusive results [3, 5-7].

The transforming growth factor \(\beta\) receptor type 1 (\textit{TGFBR1}) gene is an attractive candidate as TGF-\(\beta\) signalling plays an important role in the control of a range of biological functions associated with colon carcinogenesis including tissue homeostasis, angiogenesis, inflammation, proliferation and cellular differentiation and has and has also been implicated in both the suppression and promotion of CRC (see [1] for a recent review). On binding of the TGF-\(\beta\) ligand to TGFBR1, this serine/threonine protein kinase-containing receptor forms a heteromeric complex with type II TGF-\(\beta\) receptors thereby transducing the TGF-\(\beta\) signal from the cell surface to the cytoplasm. A common variant of \textit{TGFBR1}, rs11466445 (heterozygote frequency 0.211; dbSNP135), contains a deletion of three GCG triplets from the sequence of exon 1, resulting in the expression of a mutant receptor protein with six consecutive alanine (\textit{TGFBR1}*6A) rather than nine consecutive alanine (\textit{TGFBR1}*9A) residues. This is a hypomorphic mutation encoding a TGFBR1 variant protein with reduced TGF-\(\beta\) growth inhibition-signalling activity. The \textit{TGFBR1}*6A allele has been proposed to act as a low-penetrance susceptibility allele for a number of malignancies [8], perhaps acting by decreasing \textit{TGFBR1} allelic expression. Allele specific expression (ASE) of TGFBR1 in peripheral blood lymphocytes has been observed, with decreased expression associated with the *6A allele and two other SNPs in linkage disequilibrium [9]. Another study examined SNPs in the 3’ untranslated region of TGFBR1 and found that 29 of 138 patients with MSI-negative CRC showed
ASE, with 14 of the 29 (48%) having a *6A/*9A genotype and clear enrichment of ASE in familial cases [10].

Although some studies have suggested that the TGFBR1*6A allele confers an elevated risk of colorectal cancer [5, 8, 11], most studies have not found such an association [12-17]. A recent large meta-analysis of rs11466445 and colorectal cancer risk assessed nine association studies totalling 6,765 CRC patients and 8,496 unrelated controls and found that heterozygous *6A/*9A carriers showed a significantly increased risk of CRC with a pooled odds ratio (OR) of 1.12 (95% CI=1.02–1.23; p = 0.013) compared to homozygous *9A/*9A carriers [18]. A further recent meta-analysis, which included 15 subgroups (7,154 case and 8,851 controls), did not find an association with CRC with overall significance (OR=1.085, 95% CI=0.963, 1.222; additive model), but instead found a significant association with breast and ovarian cancer. The difference from the previous meta-analysis was the exclusion of one study and the inclusion of two further studies [19]. One of the included studies genotyped rs11466445 in a Spanish cohort somewhat enriched for familial cancer, with ~15% of cases having an affected first-degree relative and found it to be borderline significant with diagnosis of CRC (p = 0.0491; 515 cases, 515 controls) [5]. In the context of familial CRC in particular, two studies have examined families with genetic predisposition [15, 20]. In both studies, a case-control design was used - drawing on only one affected member from each family and comparing this group with unrelated controls. In each instance, TGFBR1*6A was not found to be associated with an increased familial colorectal cancer risk. Interestingly, a further study found evidence that the TGFBR1*6A allelic frequency is higher amongst familial CRC patients with mismatch-repair (MMR) negative disease [21].
There have been no reports to date that have explored the likelihood of an association of \textit{TGFBR1}^*6A with hereditary CRC using any family-based association test (FBAT) [22-24], or family-based case-control test designed for related individuals [25]. The family of FBATs examine associations within family groups and so are robust to population stratification, a known confounder of case-control studies [24]. It has been suggested this robustness comes at some cost. Simulations show that classical FBATs are less powerful than case-control tests [24, 26], as the latter examine between-family associations instead of exclusively within-family associations. Counter to this argument, the groups of affected relatives sampled from multiplex families should have more power to detect an association due to the higher than expected frequency of susceptibility alleles, compared with affected individuals having sporadic disease [25]. It is also possible to use quasi-likelihood score (QLS) tests, an alternative class of tests to FBATs with different theoretical underpinnings. As opposed to within-family tests, these are between-family case-control tests that can account for the correlation between individuals in families [25].

We recently completed a new genome-wide linkage study [27] using non-syndromic CRC families from three distinct regions in Australia and Spain. One of the linkage regions of interest identified in that study was located on chromosome 9q, proximal to the previously reported 9q22 linkage region, which contains the \textit{TGFBR1} locus. We genotyped an expanded set of families for rs11466445 and used FBATs and the “More Powerful” Quasi-Likelihood Score Test (MQLS) to test for association with diagnosis of colorectal neoplasia (i.e. either colorectal adenocarcinoma or advanced adenoma). We report that after applying several family-based association tests we only found a nominally significant result using the PBAT rapid algorithm (p = 0.028), with another three FBAT algorithms all non-significant, but each yielding a p-value <
There was no evidence of an association using the MQLS case-control model ($p = 0.41$).

**Methods**

**Family members**

A total of 414 individuals (172 males and 242 females), from 146 CRC families were recruited from clinics in Melbourne, Adelaide and Barcelona and informed consent was obtained from all participants. We restricted our study to non-syndromic high risk CRC families, defined as those containing at least one affected person who has one or more first-degree affected relatives and where the known causal mutations had been excluded. In each case, the diagnosis was confirmed by medical and pathology reports. FAP and MUTYH were excluded clinically and HNPCC or Lynch syndrome was excluded by testing for microsatellite instability (MSI) (as measured by tumour-associated length variation in microsatellites BAT-25 and BAT-26) and/or immunohistochemistry indicating loss of hMLHI, hMSH2, hMSH6 and hPMS2 encoded proteins. Affected status was defined as diagnosis with either colorectal adenocarcinoma (CA) or one or more advanced adenomas (AA), where AA was defined as three or more synchronous or metachronous adenomas and/or adenoma(s) with villous morphology, and/or with severe dysplasia, and/or diameter $\geq 10$ mm. Diagnoses were confirmed by pathology reports.

Unaffected individuals were family members who were either over 70 years of age with no history of CA or AA or were 50 years of age or older and had, within the last 5 years, recorded a colonoscopy result negative for neoplasia. As the age of onset is fairly late with a mean age of onset is 55.4 years (Table 1), the cohort is mostly sibships with missing parental genotypes. However, there is inclusion of some extended pedigrees of up to four generations (including non-genotyped founders).
containing parent–child, avuncular or cousin pairs. We reclassified 11 young “unaffected” people and those with previous detection of colorectal polyps as “unknown” in accordance with our previous work [27, 28]. Of these 11 people, three were heterozygous *6A/*9A genotype, six had the common *9A/*9A genotype and two the rare *6A/*6A genotype. When affecteds are misclassified as unaffecteds, family-based tests that make use of discordant information lose power[29], so it is sensible to reclassify particularly young unaffecteds as having unknown phenotype. All people in the study had their age at blood draw recorded.

Genotyping

The TGFBR1 rs11466445 variant status was determined by PCR amplification using primers Fwd 5'-GAGGCGAGGTTTGCTGGGGTGAGG-3' and Rev 5'-CATGTTTGAGAAAGAGCAGGAGCG-3'. PCR amplification was performed in a 25 µL reaction containing 50 ng genomic DNA using the Platinum Taq DNA polymerase with the addition of 3 × enhancer solution and followed the manufacturer’s protocol for GC-rich fragments (Invitrogen). Amplified fragments were separated by electrophoresis on a 10% polyacrylamide gel (Biorad) post-stained with gel-red (Jomar Diagnostics). Genotypes were assigned according to fragment sizes. A product size of 121 bp corresponded to the most common allele, *9A, whereas a product size of 112 bp corresponded to the *6A allele (Fig. 1).

Statistical analyses

Testing for deviation from Hardy-Weinberg equilibrium and Mendelian inconsistencies was performed using Pedstats [30]. The Generalised Disequilibrium Test (GDT) V0.1.1 software [24] was used to test for association in dichotomous relative pairs with identity-by- descent (IBD) statistics estimated using Merlin V1.1.2 [31] Generalised Family-Based Association Tests (FBAT) were undertaken in the
FBAT V2.0.4 beta software [22] and PBAT version 3.6 software [23]. FBAT was set to calculate empirical variance estimates and to use a null hypothesis of linkage and no association and p-values were generated from the asymptotic Normal distribution. For PBAT, we used the rapid algorithm, a null hypothesis of linkage and no association with sandwich variance estimation and p-values were generated using an empirical permutation-based method with 10,000 replicates. Sandwich estimation was also used to estimate the correlation between members of larger pedigrees. For time-to-onset analysis, the Wilcoxon Logrank FBAT statistic was examined.

For case-control testing the “More Powerful” Quasi-Likelihood Score Test (MQLS) was used [25]. The MQLS, an improvement on the quasi-likelihood score test WQLS [32], is a case-control test for allelic association that conditions on the pedigree structure using unconditional corrected variance to account for the relatedness amongst individuals. The MQLS can incorporate unaffected controls and controls of unknown affection state. It also makes use of the affection state of relatives with missing genotype data by using their affection status to weight the family. The rationale being that an affected person who has additional affected relatives is more likely to be carrying a genetic predisposition.

**Accounting for linkage**

The use of null hypotheses of “linkage and no association” in the FBAT and PBAT software was conservative. While the families show genetic linkage with cancer diagnosis in a region of chromosome 9 (9q33.3–9q34.3; non-parametric LOD=2.24) with a 1-LOD support interval of ~127.97–140.0 Mb[27], this does not cover the location of the TGFBR1 locus at 101.9 Mb and this region is only weakly linked with CRC. At the SNP rs928180, which is in the TGFBR1 intragenic region, the non-parametric (Snull) LOD score is 0.293. By using IBD information it is possible to
control for linkage using the GDT. Unlike the FBATs, the MQLS case-control test does not control for linkage and allows both linkage and association to contribute to the test statistic.

**Results**

**Genotyping and quality control**

We found 315, 95 and four people to be homozygous for the rs11466445 *9A allele, heterozygous and homozygous for the *6A allele, respectively, with a *6A allele frequency (AF) of 0.124. The four people carrying the *6A/*6A genotype were dispersed across two families, each having one discordant pair (one affected and one unaffected individual). An exact test found the genotype to be in Hardy-Weinberg equilibrium (all individuals, p=0.3687; 126 unrelated individuals, p=1.0) and there were no observed Mendelian inconsistencies. The allele frequencies of the *6A allele in the affected and unaffected family members (affected family member, AF=0.117, unaffected family member, AF= 0.130) were slightly higher than observed in a case-control British study of hereditary CRC (913 cases, AF=0.096; 828 controls, AF=0.100) [15] and a further Swedish Caucasian cohort with hereditary non-polyposis colorectal cancer (HNPCC) and non-HNPCC hereditary CRC patients (83 HNPCC + 179 non-HNPCC cases, AF=0.107; controls, AF=0.106) [20].

**Family-based association testing**

In the first instance, we tested for an association with colorectal neoplasia using the Generalised Disequilibrium Test (GDT). Given the large differences in pedigree sizes in this present study and the high number of possible intra-pedigree discordant pairings between genotyped people, the generalised relative pairs weighted by family size approach implemented by the GDT software, provides a good fit with the data. One caveat of using discordant pairs, however, is that in complex disease some people
inheriting a risk allele do not develop the disease, or develop it rather late in life and this needs to be taken into account. As some of the pedigrees are multi-generational, we used inheritance by descent (IBD) data to inform the GDT analysis. Testing for association between the rs11466445 *6A allele and colorectal neoplasia in 208 discordant relative pairs by the GDT algorithm produces a p-value of 0.085 (Table 2). Inclusion of gender as a covariate did not change the p-value. While this result is not significant at a 5% level, given the borderline p-value and to avoid false negative results, we further tested the association using other family-based association methods that construct a test with different assumptions and/or make use of different groupings of related people within the data. For this, we ran a parent of origin GDT test and also the tests implemented in the FBAT and PBAT software.

The GDT software allows analysis to be constrained to only examine discordant parent-child pairs (GDT-PO) and ignore unaffected sibling data. As relatively few parent-child pairs were genotyped in this study, the test will have much reduced power. However, given the age of the parents, the result should be more robust to misspecification of phenotype. This parent of origin test for the 15 parent-child pairs in the study was consistent with the full GDT result (p = 0.081; Table 2).

Next, we tested the association using the Family-based association test (FBAT), a statistic that examines the covariance between phenotype and allele transmission (Mendelian residuals) from parents to offspring. For the FBAT statistic, informative families are those with at least one parent heterozygous for the two TGFBRI alleles and having affected offspring. The use of only affected offspring in the association statistic makes it robust to phenotype misspecification of affected people as unaffected. In the case of a missing parent, or parents, the test conditions on the sufficient statistic for the genotype distribution in each family; where a parent
genotype is expressed as a set of likelihoods conditioned upon known offspring genotype(s). The design of the FBAT necessitates that extended pedigrees are split into nuclear families, which can introduce bias due to correlation. The FBAT also requires specification of the genetic model. Considering there are only four homogyzous *6A carriers we did not test the recessive genetic model. As the variant falls in an area of weak genetic linkage, FBAT empirical variance estimates were used to control for correlation amongst sibling genotypes within pedigrees. The FBAT result was non-significant (Table 2). Given the large number of missing parents in the current study and only having 22 (additive model) or 23 (dominant model) informative nuclear families, there is some reliance upon the sufficient statistic and large sample theory. Regardless, the p-value for the *6A allele under an additive model (p = 0.096) is close to that obtained with the GDT (p=0.081), which uses a robust measure not dependent upon large sample theory.

Finally, we tested for an association under an additive model with the *6A allele using the FBAT implemented in the PBAT software. The rapid algorithm in PBAT tests only the minor alleles and offers the ability to generate p-values using a robust Monte Carlo permutation based method instead of the asymptotic Normal distribution [23]. It also offers time to onset analyses with the same empirical p-values. The PBAT rapid algorithm differs from FBAT in that extended pedigrees are broken up into clusters of trios who share the same parents. The association was found to be nominally significant under an additive model (p = 0.0278; 10,000 permutations) with a null hypothesis of linkage and no association (with robust sandwich variance estimates). We also tested for an association between the *6A allele and age of CRC diagnosis, but found no evidence (additive model, FBAT-Wilcoxon, p = 0.150, null hypothesis – linkage, no association with sandwich variance).
Case-control testing

We used the “More Powerful” Quasi-Likelihood Score Test ($M_{QLS}$) which accounts for relatedness between subjects using a corrected variance. Unlike the FBATs, the $M_{QLS}$ can make use of the genotyping information of the 24 singletons in the study and can use the people with unknown affection status as controls. The result was insignificant (Table 2), with a $p$-value of 0.41 (180 cases, 137 controls) and specifying a disease prevalence of 0.05. The result was highly insensitive to specifying other disease prevalence values and setting prevalence to 0.001, 0.1 and 0.2 gave $p$-values of 0.39, 0.41 and 0.44, respectively. Unlike FBATs, the $M_{QLS}$ is not robust to population heterogeneity and will inflate type I error rates (the incorrect rejection of a true null hypothesis) in instances of population stratification. Given the convincingly non-significant result we did not investigate this further.

Discussion

Testing with the rapid PBAT algorithm gave a nominally significant result under an additive genetic model ($p = 0.028$). Under the GDT, GDT-PO and FBAT approaches we did not find a significant association, but all the $p$-values were consistently borderline, with $p$-values < 0.10. Unlike the FBAT approaches, we found no evidence of an association using $M_{QLS}$, a case-control method that corrects for relatedness amongst subjects ($p = 0.41$).

The GDT examines transmission disequilibrium between pairs of discordant relatives, while the FBAT and PBAT algorithms examine transmission disequilibrium from parents to affected offspring. The closeness in $p$-values between all family-based methods demonstrates the finding is not particularly sensitive to different assumptions underpinning these various algorithms. More broadly, given the difference in informative pairs/families in each FBAT method and nature of the algorithm, the
general agreement across methods suggests this marginal evidence of an association is not a spurious result. However, the nominally significant PBAT result should perhaps be treated with some caution. While the p-value was generated using a robust empirical Monte-Carlo based method, it is possible the partitioning of people into clusters of trios may inflate type I error. As the 22 informative nuclear families are broken into 26 informative clusters of trios, there is a degree of correlation between some clusters that is unaccounted for by the approach. Conversely, there is reason to think that such correlation may not greatly influence the p-value. The similarity in p-values between the GDT and the FBAT, which also splits extended pedigrees, suggests the difference in handling of extended pedigree structure between the methods did not overly affect the association test result in this instance.

In essence, a method (PBAT) which examines transmission to affected relatives but breaks pedigrees up for computational reasons is significant, while a method (GDT) that examines discordant relative pairs that does not adjust pedigree structure is non-significant. It is unclear how the different pairings or pedigree structure between these methods is contributing to the different in p-value. In simulations the GDT test was found to have more power over the FBAT in the majority of nuclear family and extended pedigree structures tested [24]. However, nuclear families with two missing parents (which includes most of this present cohort) were not simulated, so it is possible the FBAT implemented in the PBAT software is more powerful in this scenario and may help explain the lower p-value.

All reported association studies of rs11466445 have been of a case-control design [18]. Even studies that have gathered affected cases from families with an inherited predisposition have used a case-control design, with one case selected from each
family and compared to unrelated controls [15, 20]. To our knowledge, this present study is the first to examine the association between rs11466445 and colorectal cancer using family-based association statistics or case-control methods for related individuals. Our study is designed to examine the association of rs11466445 with colorectal neoplasia diagnosis within families predisposed to CRC. The within-family approach frees the analysis from concerns about population stratification.

Collectively, current evidence suggests the *6A allele is a relatively minor contributor to CRC prevalence. A modest 12% and 8.5% increase in CRC risk, respectively, was found by two meta-analyses across large populations [18, 19]. The first meta-analysis found the TGFBR1 *6A allele to be significantly associated with CRC while the latter did not. The authors of a recent review considering the *6A allele association and ASE studies together concluded that the effect of the allele on CRC predisposition is, at best, very subtle [33].

**Conclusions**

Our finding here of marginal evidence for an association with 6*A in CRC predisposed families supports the conclusions of recent meta-analyses and a review, which find the effect of the 6*A allele on CRC risk is modest. Our findings of weak evidence for association, together with the weak linkage signal in the region of the TGFBR1 locus, suggests that rs11466445 does not contribute significantly to the collective genetic predisposition towards CRC in these families.

**List of abbreviations**

6*A, six alanine allele; AA, advanced adenoma; ASE, Allele specific expression; CA, colorectal adenocarcinoma; CRC, colorectal cancer; FBAT, Family-Based
Association Test; GDT, Generalised Disequilibrium Test; GDT-PO, parent of origin
Generalised Disequilibrium Test; HNPCC, hereditary non-polyposis colorectal
cancer; IBD, identity-by-descent; M_{QLS} , “More Powerful” Quasi-Likelihood Score
Test; MSI, microsatellite instability; PBAT, pedigree-based association test; QLS,
 quasi-likelihood score.

**Competing Interests**

The authors declare no competing interests.

**Authors' contributions**

LJL performed all the genotyping and experimental lab work, prepared data for
analysis and helped with the manuscript; GSB performed experimental lab work; JPR,
BT and IWS analysed the data; GPY, FM, IB and GC coordinated the original
collection of the samples. JPR, BT and GNH wrote the manuscript; TJL contributed
programme support and insightful critique; GNH conceived and designed the current
study. All authors participated in data interpretation and critical revision of the
manuscript. All authors read and approved the final manuscript.

**Acknowledgements**

We thank the families for their participation. We also thank Dr Mike Buckley and Dr
Peter Molloy for critically reviewing this manuscript. This work is supported within
CSIRO by the CSIRO Preventative Health National Research Flagship and at the ICO
by contract/grant sponsor: Asociación Española Contra el Cáncer and PI10/00748.
References


Figures

Figure 1 - Genotyping example.

In an electrophoresis gel, the *TGFBR1*6A allele migrates as a 112 bp species and the *TGFBR1*9A allele migrates as a 121 bp species. Examples of homozygotes and heterozygotes of the two alleles are shown.
# Tables

Table 1 - Participant characteristics and demographics.

<table>
<thead>
<tr>
<th>Participant Characteristics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of individuals</strong></td>
<td></td>
</tr>
<tr>
<td>Total including founders</td>
<td>759</td>
</tr>
<tr>
<td>Genotyped subjects</td>
<td>414</td>
</tr>
<tr>
<td>Flinders Centre for Cancer Prevention and Control</td>
<td>202</td>
</tr>
<tr>
<td>The Royal Melbourne Hospital</td>
<td>165</td>
</tr>
<tr>
<td>Institut Català d'Oncologia</td>
<td>33</td>
</tr>
<tr>
<td>Peter MacCallum Cancer Centre</td>
<td>14</td>
</tr>
<tr>
<td><strong>Family structures</strong></td>
<td></td>
</tr>
<tr>
<td>Number of families</td>
<td>147</td>
</tr>
<tr>
<td>Singletons</td>
<td>24</td>
</tr>
<tr>
<td>2 generations (nuclear)</td>
<td>87</td>
</tr>
<tr>
<td>3+ generations (extended)</td>
<td>36</td>
</tr>
<tr>
<td>Trios</td>
<td>25</td>
</tr>
<tr>
<td><strong>Traits of genotyped subjects</strong></td>
<td></td>
</tr>
<tr>
<td>Total Affected</td>
<td>180</td>
</tr>
<tr>
<td>Colorectal adenocarcinoma</td>
<td>145</td>
</tr>
<tr>
<td>Advanced adenoma(s)</td>
<td>35</td>
</tr>
<tr>
<td>Unaffected relatives</td>
<td>223</td>
</tr>
<tr>
<td>Unknown affection</td>
<td>11</td>
</tr>
<tr>
<td>Relative Pairs</td>
<td>208</td>
</tr>
<tr>
<td>Sib</td>
<td>169</td>
</tr>
<tr>
<td>Cousin</td>
<td>1</td>
</tr>
<tr>
<td>ParentChild</td>
<td>15</td>
</tr>
<tr>
<td>Avuncular</td>
<td>19</td>
</tr>
<tr>
<td>Other</td>
<td>4</td>
</tr>
<tr>
<td>Female</td>
<td>58.45%</td>
</tr>
<tr>
<td>Age, mean +/- SD</td>
<td>56.4 ± 12.6</td>
</tr>
<tr>
<td>Affected Age, mean +/- SD</td>
<td>55.4 ± 12.5</td>
</tr>
<tr>
<td>Unaffected Age, mean +/- SD</td>
<td>57.2 ± 12.7</td>
</tr>
</tbody>
</table>
Table 2 - Association Results.

<table>
<thead>
<tr>
<th>Model</th>
<th>Allele</th>
<th>Allele Freq</th>
<th>Informative units</th>
<th>P-value*</th>
<th>Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GDT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allelic GDT</td>
<td>6</td>
<td>0.124</td>
<td>208</td>
<td>0.085</td>
<td>1.724</td>
</tr>
<tr>
<td><strong>Parent of origin GDT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allelic Parent-child pairs</td>
<td>6</td>
<td>0.124</td>
<td>15</td>
<td>0.081</td>
<td>1.746</td>
</tr>
<tr>
<td><strong>FBAT (empirical)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Additive Nuclear families</td>
<td>6</td>
<td>0.115</td>
<td>22</td>
<td>0.096</td>
<td>1.665</td>
</tr>
<tr>
<td>Additive Trios</td>
<td>9</td>
<td>0.885</td>
<td>22</td>
<td>0.096</td>
<td>-1.665</td>
</tr>
<tr>
<td>Dominant Nuclear families</td>
<td>6</td>
<td>0.115</td>
<td>23</td>
<td>0.193</td>
<td>1.301</td>
</tr>
<tr>
<td>Recessive Nuclear families</td>
<td>9</td>
<td>0.885</td>
<td>23</td>
<td>0.193</td>
<td>-1.301</td>
</tr>
<tr>
<td><strong>PBAT (rapid)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Additive Trios</td>
<td>6</td>
<td>0.114</td>
<td>26</td>
<td><strong>0.028</strong></td>
<td>-</td>
</tr>
<tr>
<td><strong>MQLS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allelic Cases/Controls</td>
<td>6</td>
<td>0.124</td>
<td>180/137</td>
<td>0.41</td>
<td>0.82</td>
</tr>
</tbody>
</table>

*- All p-values rounded to 3 decimal places.
TGFβR1 Genotypes

Figure 1