

## ARTICLE

# Identification of a prostate cancer susceptibility gene on chromosome 5p13q12 associated with risk of both familial and sporadic disease

Liesel M FitzGerald<sup>1</sup>, Briony Patterson<sup>1</sup>, Russell Thomson<sup>1</sup>, Andrea Polanowski<sup>1</sup>, Stephen Quinn<sup>1</sup>, Jesper Brohede<sup>2</sup>, Timothy Thornton<sup>3</sup>, David Challis<sup>4</sup>, David A Mackey<sup>1,5,6</sup>, Terence Dwyer<sup>1,7</sup>, Simon Foote<sup>1</sup>, Garry N Hannan<sup>8,9</sup>, James Stankovich<sup>1,10</sup>, James D McKay<sup>1,11</sup> and Joanne L Dickinson<sup>\*,1</sup>

<sup>1</sup>Genetics Unit, Menzies Research Institute, University of Tasmania, Hobart, Australia; <sup>2</sup>Karolinska Institutet, KI-Alzheimer Disease Research Centre, Stockholm, Sweden; <sup>3</sup>Department of Statistics, University of California, Berkeley, CA, USA; <sup>4</sup>Histopathology Department, Royal Hobart Hospital, Hobart, Australia; <sup>5</sup>Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, East Melbourne, Australia; <sup>6</sup>Department of Ophthalmology, Royal Children's Hospital, Melbourne, Australia; <sup>7</sup>Murdoch Children's Research Institute, Melbourne, Australia; <sup>8</sup>CSIRO Preventative Health National Research Flagship, North Ryde, New South Wales, Australia; <sup>9</sup>CSIRO Molecular and Health Technologies, North Ryde, New South Wales, Australia; <sup>10</sup>Bioinformatics Division, Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia; <sup>11</sup>International Agency for Research on Cancer, Lyon, France

Genetic heterogeneity is a difficulty frequently encountered in the search for genes conferring susceptibility to prostate cancer. To circumvent this issue, we selected a large prostate cancer pedigree for genome-wide linkage analysis from a population that is genetically homogeneous. Selected cases and first-degree relatives were genotyped with Affymetrix 10K SNP arrays, identifying a 14 Mb haplotype on chromosome 5 (5p13–q12) inherited identical-by-descent (IBD) by multiple cases. Microsatellite genotyping of additional deceased case samples confirmed that a total of eight cases inherited the common haplotype ( $P=0.0017$ ). Re-sequencing of eight prioritised candidate genes in the region in six selected individuals identified 15 SNPs segregating with the IBD haplotype, located within the *ITGA2* gene. Three of these polymorphisms were selected for genotyping in an independent Tasmanian data set comprising 127 cases with familial prostate cancer, 412 sporadic cases and 319 unaffected controls. Two were associated with prostate cancer risk: rs3212649 (OR = 1.67 (1.07–2.6),  $P=0.0009$ ) and rs1126643 (OR = 1.52 (1.01–2.28),  $P=0.0088$ ). Significant association was observed in both familial and sporadic prostate cancer. Although the functional SNP remains to be identified, considerable circumstantial evidence, provided by *in vivo* and *in vitro* studies, supports a role for *ITGA2* in tumour development. *European Journal of Human Genetics* (2009) 17, 368–377; doi:10.1038/ejhg.2008.171; published online 1 October 2008

**Keywords:** genetic; susceptibility; prostate; cancer; risk

\*Correspondence: Dr JL Dickinson, Menzies Research Institute, Private Bag 23, Hobart, Tasmania 7001, Australia.  
Tel: +61 3 62267622; Fax: +61 3 62267704;  
E-mail: jo.dickinson@utas.edu.au  
Received 6 June 2008; revised 25 July 2008; accepted 29 August 2008;  
published online 1 October 2008

## Introduction

In many Western countries prostate cancer (MIM176807) is the most commonly diagnosed cancer after skin cancer.<sup>1</sup> Familial aggregation of the disease indicates that it has a genetic component.<sup>2,3</sup> Evidence presented by the International Consortium for Prostate Cancer Genetics (ICPCG)<sup>4</sup>

highlighted five suggestive linkage regions supporting putative susceptibility loci reported earlier, including one on chromosome 5q.<sup>5</sup> More recently, several genome-wide association studies (GWAS) in large independent case-control data sets have provided compelling evidence for multiple new susceptibility loci. Common sequence variants on chromosomes 8q24 and 17q have been significantly and independently associated with prostate cancer risk and these findings have been replicated in several populations.<sup>6–10</sup> Subsequently, evidence for additional loci on chromosomes 2, 3, 6, 7, 8, 10, 11, 17 and X<sup>10–12</sup> has been provided. Although many of these SNPs are located in or close to known genes, a functional role for these SNPs in prostate cancer is yet to be demonstrated.

Although complex diseases are probably influenced by a mixture of common and rare genetic variants, there is evidence that rarer variants play a larger role in cancer than in other diseases.<sup>13,14</sup> Large multiplex pedigrees are a powerful approach for identifying rare variants. They can also be used to help address the problem of genetic heterogeneity and are therefore potentially advantageous when studying a complex disease such as prostate cancer. Here, we present the linkage analysis of an extended Tasmanian pedigree (PcTas9) in which 25 prostate cancer cases can be traced back seven generations to a common founding pair. Evidence for a prostate cancer susceptibility locus on chromosome 5 (5p13–q12) is presented, adding to the mounting evidence that this region harbours a putative prostate cancer susceptibility gene.<sup>4,5,15–18</sup> In addition, although the recently published GWAS have not highlighted specific candidate susceptibility SNPs on chromosome 5, Thomas *et al*<sup>12</sup> noted the likelihood that chromosome 5 may also harbour prostate cancer susceptibility loci.

Resequencing within the region of interest on chromosome 5 identified sequence variants that co-segregated with disease in this large pedigree. Genotyping of these variants in an independent data set, comprising both familial and sporadic prostate cancer cases and age-matched controls, revealed two polymorphisms that are significantly associated with prostate cancer risk in the wider Tasmanian population.

## Subjects and methods

### Study subjects and preparation

Ethics approval for this study was obtained from the Southern Tasmanian Human Research Ethics Committee and written informed consent was obtained from all participating individuals. Using a genealogical database at the Menzies Research Institute (MRI) and the records of the Tasmanian Cancer Registry (TCR), families with multiple cases of prostate cancer were identified. All families had at least two affected close relatives; with the largest family comprising 37 affected men, and 10 families with more

than 11 cases. Blood samples were available for 131 familial cases in total.

A seven-generation pedigree was selected to conduct genome-wide linkage analysis, comprising 25 patients diagnosed with histologically confirmed prostate cancer (average age of diagnosis 72, range 50–89) (Figure 1). A total of 16 DNA samples were obtained from PcTas9 prostate cancer patients; 9 were extracted from blood samples and 7 from paraffin-embedded prostate tumour tissue. In addition, 57 DNA samples were available from relatives. DNA was extracted from whole blood or buccal mucosa swabs using the Nucleon Bacc3 (Amersham Biosciences AB) and PureGene DNA Isolation Kits (Gentra Systems), respectively. Seven paraffin-embedded tumour blocks for the individuals shown in Figure 1 were obtained from two pathology laboratories located in Southern Tasmania. Using Method E as described by Sato *et al*,<sup>19</sup> DNA was isolated from selected tissue sections comprising a significant proportion of normal tissue.

A prostate cancer case-control study is also being conducted concurrently by the MRI and is recruiting sporadic prostate cancer cases and unaffected controls. Blood samples, serum samples, physical measures and environmental exposure data are being collected from participating individuals. A total of 412 sporadic prostate cancer cases were identified from the TCR and recruited. Eligible cases were men under the age of 70 years diagnosed with histologically confirmed cancer of the prostate during the period 1996–2005. Controls were randomly selected from the electoral roll. Eligible controls (319 in total) were age-matched within 5-year age groups to the sporadic cases and self-reported as unaffected with prostate cancer.

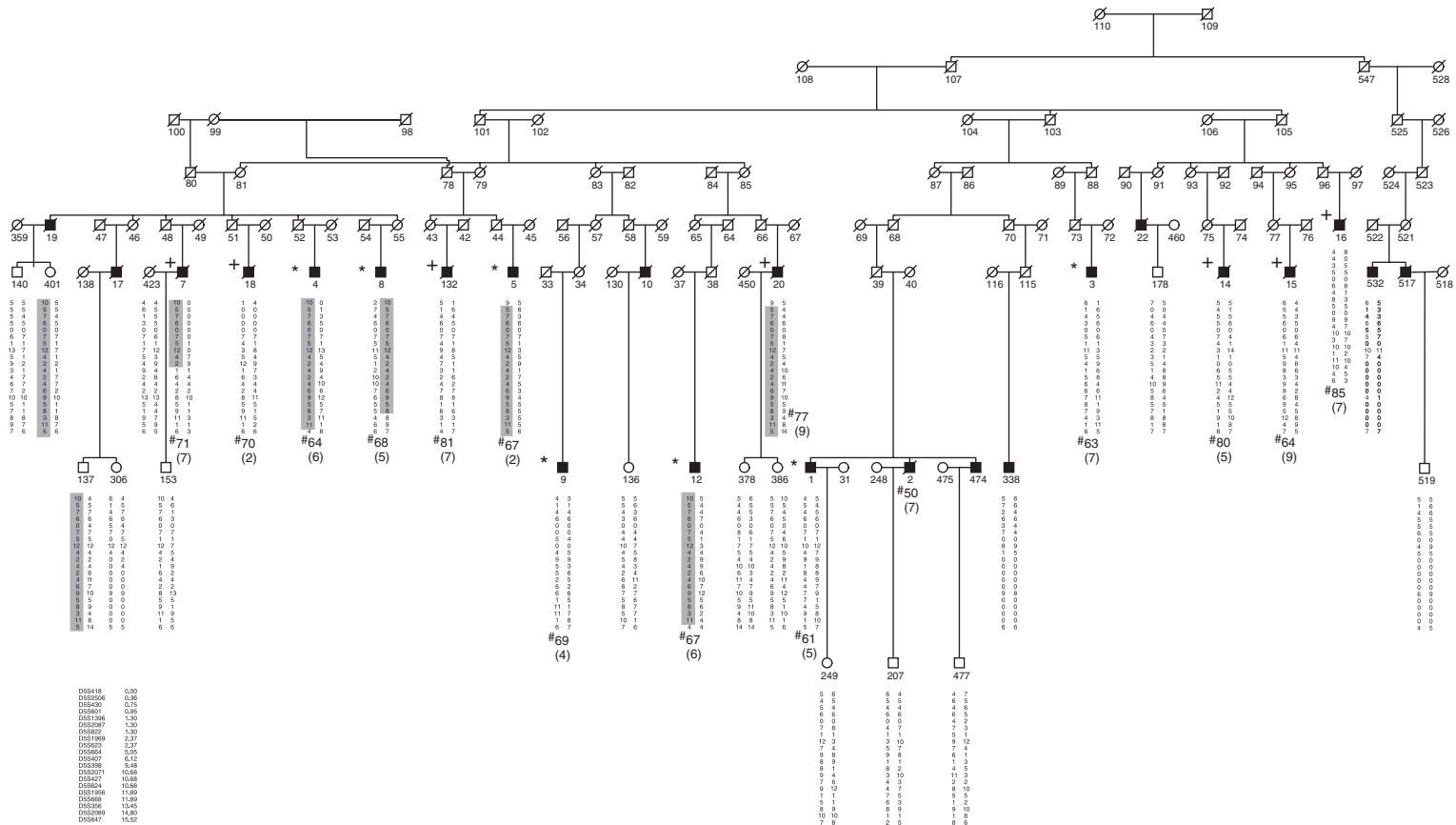
### Genotyping

Seven prostate cancer patients in family PcTas9 were genotyped with Affymetrix Human Mapping 10K arrays (indicated with an asterisk in Figure 1), along with six first-degree relatives to provide phase information. Genotyping was performed as described by Kennedy *et al*<sup>20</sup> using 250 ng of genomic DNA. The mean genotype call rate was 95%.

Twenty microsatellites across 5p13–q12 (chosen from the Genome Database – <http://www.gdb.org/> – see Figure 1) were genotyped in all 73 available DNA samples from PcTas9 using standard fluorescent PCR techniques and an ABI PRISM 310 Genetic Analyzer (Applied Biosystems). Microsatellite allele frequencies were estimated from 32 Tasmanian samples comprising spouses of affected cases and unaffected family members.

### Linkage analysis

PEDCHECK<sup>21</sup> was used to identify Mendelian errors in the genotype data. The 10K SNP data were analysed using the genetic map and Caucasian allele frequencies provided by Affymetrix.<sup>20</sup> A total of 954 markers in strong linkage disequilibrium (LD) with other markers were removed from



**Figure 1** PcTas9 prostate cancer pedigree showing the segregating haplotype across 5p13–q12. Only cases and individuals providing genotyping information pertinent to the cases are presented due to pedigree size. Two cases are not represented here (9.6–brother of 9.132; 9.13–uncle of 9.1, 9.2 and 9.474). The affected status of ‘older’ generations is unknown as the Tasmanian Cancer Registry only lists patients’ records since 1978. Shaded boxed areas represent the shared haplotype. Individuals genotyped with the Affymetrix 10K array are indicated by an asterisk (\*). Individuals genotyped from paraffin-embedded tissue are indicated by a cross (+). # Denotes age at diagnosis, followed by Gleason score presented in parentheses.

the data set. The family was too large for exact multipoint linkage analysis with the Lander–Green algorithm,<sup>22</sup> and the Markov chain Monte Carlo linkage algorithm in Simwalk2<sup>23</sup> was slow to converge with this number of markers. Linkage analysis was performed using a recently published method that breaks pedigrees into smaller subunits for inference of patterns of identical-by-descent (IBD) sharing between case pairs.<sup>24</sup> We have performed simulations, some using the PcTas9 pedigree (Figure 1), to confirm the accuracy of the pedigree-splitting approach for dense marker sets.<sup>24</sup>

Microsatellite haplotypes were inferred using Simwalk2.<sup>23</sup> Genotyping so many microsatellites allowed the determination of patterns of IBD sharing with virtually complete certainty – hence we assumed that patterns of IBD sharing were known with complete certainty for assessing significance. The significances of  $NPL_{\text{pairs}}$  scores (SNP data) and patterns of IBD sharing (microsatellite data) were computed by simulating fully informative markers on the complete pedigree under the hypothesis of no linkage (gene dropping).<sup>23</sup> Nominal (single-point) *P*-values were calculated using  $10^8$  simulations of a single marker, and genome-wide and ICPCG-region-wide corrected *P*-values were calculated using  $10^5$  simulations of the entire genome or the ICPCG linkage regions respectively, with recombinations simulated using the Haldane map function. The suggestive linkage threshold plotted in Figure 2 corresponds to the value of  $NPL_{\text{pairs}}$  expected to occur on average once in a genome-wide scan with fully informative markers.<sup>25</sup>

### Prioritising candidate genes

Over 90 genes in the region spanned by the chromosome 5 haplotype were prioritised for sequencing using the

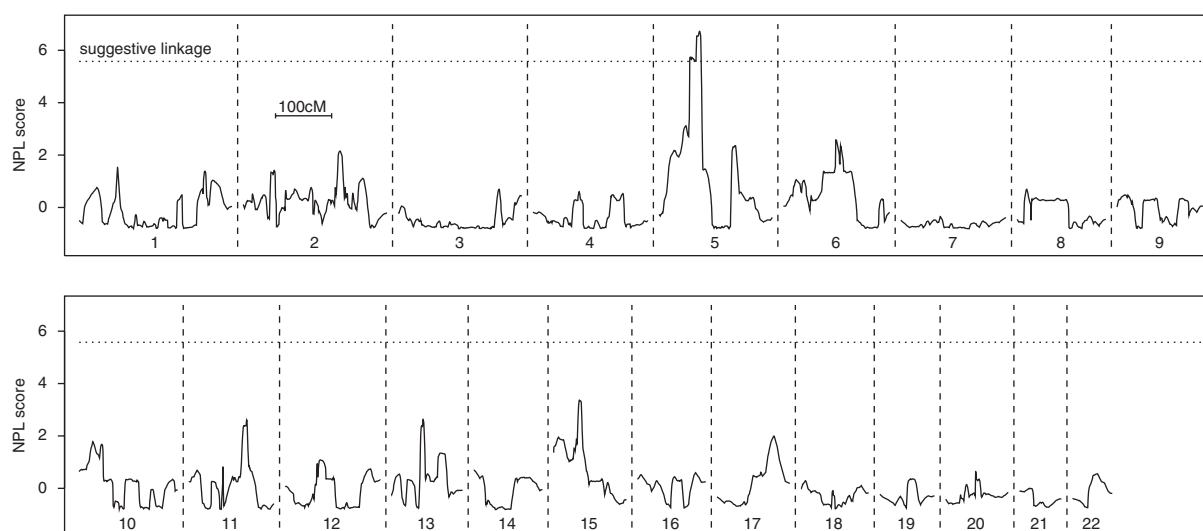
program ‘GeneSniffer’ (www.genesniffer.org; Autogen Limited, Australia). The program builds a list of ranked candidate genes within the specified region using publicly available human and mouse genetic databases, and a supplied list of key words (available on request).

### Sequencing candidate genes

The coding regions of eight candidate genes including all exon/intron boundaries and 500bp upstream of the transcription start site were resequenced in PcTas9.4, 9.5, 9.8 and 9.12, all of whom carried the chromosome 5 haplotype, and in two unrelated controls. Resequencing was performed using PCR primers designed using the software packages ‘Exon Primer’ (<http://ihg.gsf.de/cgi-bin/snps/seq1.pl>) and ‘Primer 3’ (<http://frodo.wi.mit.edu>). Primer sequences and conditions are available on request. Following PCR amplification, samples were sequenced on an ABI310 Genetic Analyzer (Applied Biosystems) in accordance with the manufacturer’s instructions.

### Genotyping of selected sequence variants

All familial prostate cancer cases (excluding PcTas9.4, 9.5, 9.8 and 9.12), sporadic cases and controls were genotyped for three sequence variants – rs28095, rs1126643 and rs3212649. Allele-specific PCR was used to detect the rs28095 polymorphism (herein referred to as C-52T), using forward primers 5'-aatcaggaggggcgggct-3', 5'-aatcaggaggggcgggccc-3' and reverse primer 5'-gcgctgggttgcagagtt-3' (Promega). Rs1126643 (herein referred to as C807T) was amplified using primers 5'-gatgccttaagctaccggc-3' and 5'-taacttcccagctgcctc-3' (Promega). Genotype was resolved by digestion with *Hyp188 I*. Rs3212649 (herein referred to as 3'UTR in/del) was amplified using primers, 5'FAM-gcaactacagaagtggaagtc-3' and



**Figure 2** Results of the Affymetrix 10K array genome-wide scan. An NPL score of 5.58 (suggestive linkage threshold expected to occur once by chance in a genome-wide scan with fully informative markers<sup>41</sup>) is indicated by a horizontal dashed line.

5'-tctgtggcaactttgatga-3' (Promega). Genotype was resolved by sizing on an ABI310 Genetic Analyser. The insertion allele was 137bp and the deletion allele was 134bp. Detailed conditions are available on request. For all three SNPs, 15% of all samples were re-genotyped to validate genotyping quality.

### Statistical analysis to detect association for selected sequence variants

The association analysis excluded all PcTas9 samples with the known chromosome 5 haplotype (cases PcTas9.4, 9.5, 9.8 and 9.12). The  $M_{QLS}$  test<sup>26</sup> was used to test for differences in allele frequencies at polymorphisms C-52T, C807T and 3'UTR in/del between familial and sporadic prostate cancer cases and controls. This test allows for the non-independence of the genotypes of related individuals, regardless of the complexities of the relationships, and also exploits the fact that cases with affected relatives are generally more likely to carry disease susceptibility alleles than cases without affected relatives.

We then performed further exploratory analyses, making no allowance for the non-independence of the familial cases. Logistic regression was used to calculate genotypic odds ratios for the two polymorphisms showing significant association, both adjusted and unadjusted for genotypes at the other polymorphism. Haploview 3.2 ([www.broad.mit.edu/mpg/haploview/](http://www.broad.mit.edu/mpg/haploview/)) was used to examine LD between the polymorphisms, and to perform haplotypic tests of association.

### Results

The results of the non-parametric linkage analysis of available, high-quality DNA using the 10K SNP genome-wide scan are displayed in Figure 2. There was one suggestive linkage peak at 5p13–q12 ( $NPL_{pairs} = 6.72$ ;  $P = 0.005$  using gene dropping),<sup>23</sup> where there was evidence that four out of seven cases had inherited a common haplotype. Along the rest of the genome, there was no evidence anywhere of more than three cases inheriting a common haplotype.

Microsatellite genotyping was possible in 15 of the remaining 17 cases, where SNP genotyping was not possible. This provided data on the 5p13–q12 region in PcTas9 prostate cancer patients and their children in a total of 23 of the 25 cases. Of these 23 cases, it was confirmed that 8 shared a common haplotype (including 4 cases genotyped with SNPs). Individuals 9.4, 9.5, 9.8 and 9.12 were confirmed as sharing the common IBD haplotype by both SNP and microsatellite genotyping; individuals 9.7 and 9.20 were confirmed by microsatellite genotyping alone; and for individuals 9.6 and 9.13 haplotype sharing was inferred by genotyping offspring with microsatellites. Twelve individuals did not carry this same haplotype, and the sharing status of three cases (individuals 9.10, 9.22,

9.517) could not be determined (Figure 1). Given this configuration of genotyped and un-genotyped cases and relatives, the probability of observing a common haplotype shared by eight or more cases by chance at a single locus is  $P = 0.0017$ . The chromosome 5 haplotype extends across 14Mb between markers D5S2506 and D5S664 (Figure 1).

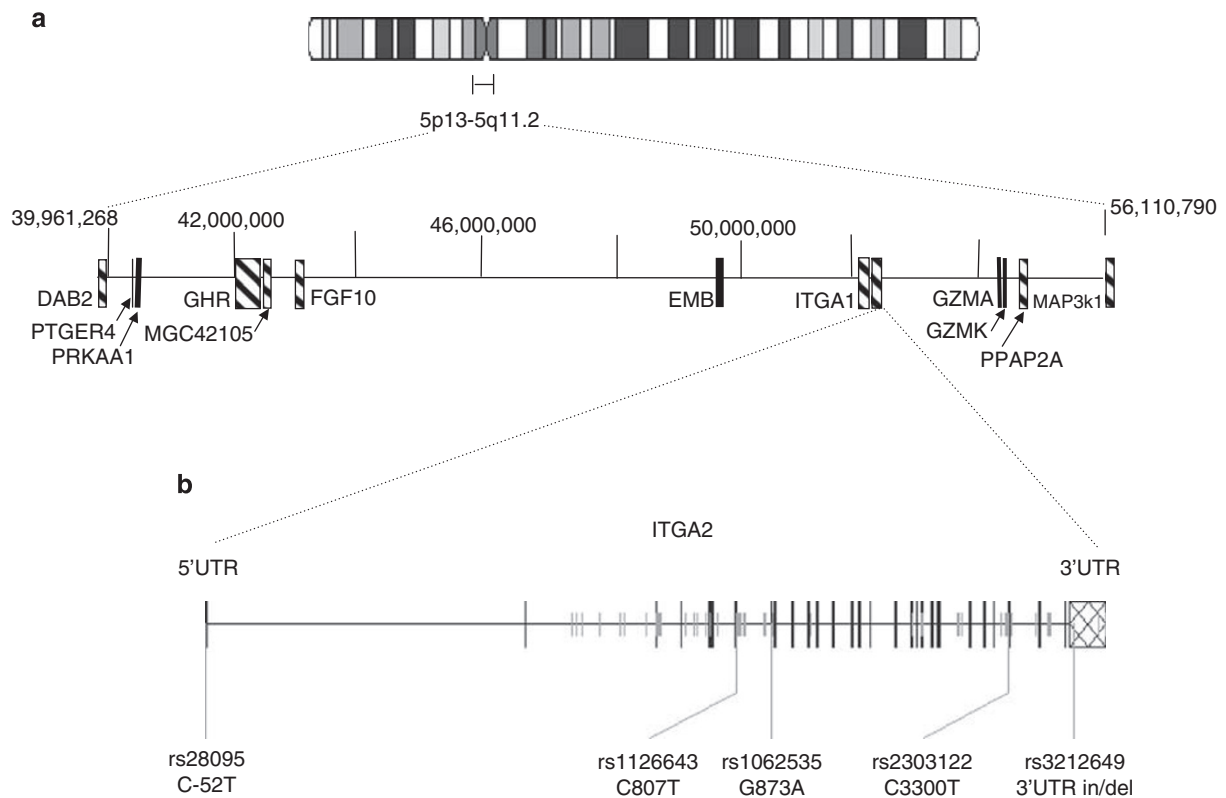
The eight patients carrying the chromosome 5 haplotype are clustered into a single branch of the pedigree descendant from individuals 9.101 and 9.102. The probability of identifying the same haplotype in eight or more affected descendants of these individuals is  $P = 0.0003$ . Although this linkage is not of genome-wide significance, this region overlaps substantially with one of five suggestive linkage regions identified in a large study conducted by the ICPCG.<sup>4</sup> Gene-dropping simulations of the 1-LOD drop intervals in these five regions showed that the probability of observing segregation in eight or more cases in one of these regions by chance is  $P = 0.043$ .

The mean age at diagnosis for PcTas9 cases sharing the chromosome 5 haplotype was 69.0 years compared to 71.6 years for those cases not sharing the haplotype ( $P = 0.53$ , *t*-test). Similarly, there was no significant difference in Gleason scores between cases sharing the risk haplotype and those that did not (mean of 5.83 and 5.75, respectively;  $P = 0.95$ , *t*-test).

Candidate genes within the chromosome 5 interval were prioritised using the bioinformatics tool, GeneSniffer ([www.genesniffer.org](http://www.genesniffer.org); Autogen Limited). Eight genes were resequenced from the GeneSniffer prioritised list. A diagrammatic representation of the 14Mb region of chromosome 5 including the selected genes is shown in Figure 3a.

The coding regions, intron/exon boundaries and promoter sequences of *FGF10*, *GHR*, *DAB2*, *PPAP2A*, *PTGER4*, *EMB* and *GZMA* were resequenced in four PcTas9 cases carrying the common chromosome 5 haplotype (PcTas9.4, 9.8, 9.5, 9.12 in Figure 1), and two unrelated controls. In these genes no variants were detected that were unique to the identified haplotype (data not shown). In contrast, within the *ITGA2* gene 15 polymorphisms were detected that segregated with the chromosome 5 haplotype (Table 1). The *ITGA2* gene (hg18 location chr5: 52320913–52426366) comprises 30 exons and spans 105288 bp (Figure 3b). Although 10 of the identified SNPs lie in intronic regions of *ITGA2*, the remaining 5 are located in the promoter (C-52T; rs28095), exon 7 (C807T; rs1126643), exon 8 (G873A; rs1062535), exon 27 (C3300T; rs2303122) and the 3'UTR of exon 30 (3'UTR in/del; rs3212649; Table 1).

Three of the SNPs in *ITGA2* (C-52T, C807T and the 3'UTR in/del) were selected for further examination using association testing in an independent data set. The PcTas9 individuals used to identify these SNPs were excluded from this analysis.



**Figure 3** Diagrammatic representation of the 14 Mb region of interest on chromosome 5. (a) Shows the location and relative size of prioritised genes for resequencing. (b) Shows an expanded view of the *ITGA2* gene. Exons are represented as tall black rectangles. Short grey lines represent SNPs identified from the HapMap and SeattleSNPs databases that are in strong LD ( $r^2 > 0.72$ ) with 3'UTR in/del. The relative positions of the five exonic SNPs discussed in the text are shown.

**Table 1** *ITGA2* SNPs identified by resequencing four PcTas9 haplotype carriers and two controls – at all SNPs the haplotype carriers share an allele that does not occur in either control

SNP ID	Location	Cases				Controls	
		9.4	9.8	9.5	9.12	9.122	9.127
rs26679	chr5:52,319,682	CG	CG	CG	CC	GG	GG
rs26680	chr5:52,319,893	TC	TC	TC	TT	CC	CC
rs28095 (C-52T)	chr5:52,320,624	TC	TC	TC	TT	CC	CC
rs3212441	chr5:52,373,415	TC	TC	TT	TT	CC	CC
rs1363192	chr5:52,373,590	GT	GT	GG	GG	TT	TT
rs1126643 (C807T)	chr5:52,382,876	TC	TC	TT	TT	CC	CC
rs1833558	chr5:52,383,068	AG	AG	AA	AA	GG	GG
rs2974987	chr5:52,386,749	AG	AG	GG	GG	AA	AA
rs1062535 (G873A)	chr5:52,386,920	AG	AG	AA	AA	GG	GG
rs2303127	chr5:52,391,361	CT	CT	CC	CC	TT	TT
rs2303126	chr5:52,401,669	GA	GA	GA	AA	GG	GG
rs2287871	chr5:52,403,213	CT	CT	CC	CC	TT	TT
rs984966	chr5:52,404,429	AT	AT	AA	AA	TT	TT
rs2303122 (C3300T)	chr5:52,414,784	TC	TC	TT	TT	CC	CC
rs3212649 (3'UTRin/del)	chr5:52,422,409	AAC/-	AAC/-	(-/-)	(-/-)	AACx2	AACx2

The C-52T SNP was selected for its potential functional role in regulation of the *ITGA2* gene. Located 52 bases upstream of the transcription start site, it is positioned between two tandem Sp1/Sp3-binding elements and the

presence of the T allele has been shown, by *in vitro* studies, to reduce *ITGA2* transcription.<sup>27</sup> The C807T polymorphism lies within exon 7, and whilst it does not alter the amino-acid sequence of the protein, there is circumstantial

evidence that it is associated with altered expression of *ITGA2*.<sup>28</sup> Functional modelling software identifies a predicted enhancer splice element (ESE) at C807T that is disrupted in the presence of the T allele (<http://pupasuite.bioinfo.cipf.es/>).

The synonymous G873A polymorphism in exon 8 is in strong LD with C807T ( $r^2=0.97$  in the HapMap CEU population of North-West European ancestry; [www.hapmap.org](http://www.hapmap.org)) and was therefore not genotyped.

The remaining two polymorphisms, C3300T and 3'UTR in/del, have no known or putative function. The C3300T is perfectly correlated with the 3'UTR in/del as reported in the PGA European Panel from the SeattleSNP database ( $r^2=1$ , <http://pga.gs.washington.edu/>) and consequently only the 3'UTR in/del was genotyped.

Thus the three SNPs were genotyped in 127 cases ascertained as part of other large pedigrees (ie, cases with strong evidence of family history), 412 sporadic cases and 319 controls. After adjusting for multiple testing using the Bonferroni method, the whole-sample unadjusted *P*-values quoted below remain significant.

No association with disease was seen with the C-52T SNP for sporadic cases ( $P=0.36$ ), familial cases ( $P=1.0$ ) or when all cases were combined ( $P=0.49$ ). In contrast, the C807T polymorphism was significantly associated with prostate cancer ( $P=0.0088$ ; Table 2) for combined prostate cancer

cases and unaffected controls. There was a more significant difference in allele frequencies between familial cases and controls ( $P=0.020$ ) than between sporadic cases and controls ( $P=0.070$ ; Table 2). The 3'UTR deletion allele showed a significant association with prostate cancer ( $P=0.0009$ ; Table 2) in combined prostate cancer cases and unaffected controls. This significant association was observed in both familial cases and controls ( $P=0.0018$ ) and between sporadic cases and controls ( $P=0.015$ ; Table 2).

Our data suggest that the 3'UTR deletion allele and the 807T allele both confer risk in a dominant manner, with similar estimated odds ratios for heterozygous and homozygous carriers of these alleles (Table 3). Odds ratios are higher in familial cases than in sporadic cases for both polymorphisms. There is strong LD between the two polymorphisms ( $r^2=0.72$ ;  $D'=0.96$ ) with the 3'UTR deletion allele almost always occurring with the 807T allele. In the absence of very large sample sizes, this strong LD makes it difficult to separate the individual effects of these alleles, and using a haplotypic test for association, the results were inconclusive (data not shown). However, after adjusting for the 3'UTR in/del and assuming a dominant disease model, the association between the C807T polymorphism and disease disappears (Table 3). This suggests that the risk from the 807T allele may only occur because of LD between the two SNPs.

**Table 2** Results of testing the 3'UTR 3 bp insertion (I)/deletion (D), and C807T polymorphisms for significant differences in allele frequencies between cases and controls using the  $M_{QLS}$  statistic<sup>26</sup>

Data set	3'UTR D allele frequency (95% CI)	3'UTR D P-value	807T allele frequency (95% CI)	C807T P-value
Controls ( $n=319$ )	0.315 (0.279–0.351)		0.373 (0.336–0.410)	—
Sporadic cases ( $n=412$ )	0.376 (0.343–0.409)	0.0150	0.420 (0.386–0.454)	0.0696
Familial cases ( $n=127$ )	0.421 (0.352–0.490)	0.0018	0.461 (0.391–0.531)	0.0198
All cases ( $n=539$ )	0.385 (0.355–0.414)	0.00090	0.428 (0.397–0.458)	0.0088

Abbreviation: CI, confidence intervals.

**Table 3** Genotypic OR<sup>a</sup> calculated for the 3'UTR insertion (I)/deletion (D) and C807T polymorphisms using logistic regression

Genotype	Sporadic cases	Familial cases	Total cases	Controls	OR_sporadic (95% CI)	OR_familial (95% CI)	OR_total (95% CI) <sup>a</sup>	OR_total (95% CI) <sup>b</sup>
<b>3'UTR in/del</b>								
I/I	158	38	196	155	1	1	1	1
I/D	198	67	265	127	1.53 (1.12–2.09)	2.15 (1.36–3.42)	1.65 (1.22–2.22)	1.84 (1.08–3.15)
D/D	56	22	78	37	1.49 (0.93–2.38)	2.43 (1.28–4.58)	1.67 (1.07–2.60)	
<b>C807T</b>								
C/C	137	34	171	131	1	1	1	1
C/T	204	65	269	138	1.41 (1.02–1.95)	1.82 (1.12–2.93)	1.49 (1.10–2.03)	0.88 (0.51–1.52)
T/T	71	28	99	50	1.36 (0.88–2.10)	2.16 (1.19–3.92)	1.52 (1.01–2.28)	

Abbreviations: CI, confidence interval; OR, odds ratios.

<sup>a</sup>Odds ratios represent increase in odds of disease compared to individuals with I/I or C/C genotypes.

<sup>b</sup>Odds ratios adjusted for the presence/absence of the disease-susceptibility allele at the other SNP (T or I).

Associations between clinical characteristics of disease (age at diagnosis, survival and Gleason score) and the risk genotypes (3'UTR deletion and 807T) were examined in both sporadic and familial data sets. Overall, there was little evidence to suggest there was a difference in age at diagnosis between sporadic cases carrying at least one risk 3'UTR deletion allele (age range 44–69 years, mean = 61.8 years) compared to those with no risk allele (age range 47–69 years, mean = 62.5 years;  $P=0.31$ , *t*-test). Similarly, no significant association between the risk 3'UTR deletion allele and age at diagnosis was observed in familial cases ( $P=0.27$ , *t*-test). Survival analysis of both familial and sporadic cases indicated that there was no decrease in overall survival associated with carrying the 3'UTR deletion allele ( $P=0.11$ , log-rank test).

For sporadic cases where Gleason score was available ( $n=92$ ), there was no significant difference between cases carrying at least one risk 3'UTR deletion allele (mean Gleason score 5.41) and those not carrying a risk allele (mean Gleason score 5.26;  $P=0.66$ , *t*-test). Similar results were obtained for familial cases ( $n=50$ ), with a mean Gleason score of 5.57 for those carrying at least one risk allele compared to a mean of 5.25 with no risk allele ( $P=0.67$ ).

Owing to the LD between the C807T and the 3'UTR in/del SNP, almost identical results were obtained when associations between the 807T risk allele and age at diagnosis, survival and Gleason score were examined in both prostate cancer data sets (data not shown).

It remains to be determined which of the SNPs identified in *ITGA2* are functionally significant, or whether another, as yet unidentified SNP, is involved. Samples of European ancestry from the HapMap and SeattleSNPs databases were examined to search for other candidate SNPs in LD with the 3'UTR in/del and C807T SNPs. An examination of these databases revealed 30 and 14 SNPs, respectively, in LD ( $r^2>0.72$ ) with C3300T (ie, LD greater than that observed between 3'UTR in/del and C807T). A review of these SNPs identified six SNPs within the 3'UTR in significant LD, with the 3'UTR in/del polymorphism. Use of modelling software that predicts miRNA-binding sequences revealed that two of these SNPs (rs6880055 and rs57674800) are contained within predicted miRNA-target motifs (<http://regrna.mbc.nctu.edu.tw/>). The presence of the alternative genotype at these sites is predicted to disrupt miRNA binding to the corresponding mRNA sequence. No SNP outside the *ITGA2* gene was correlated with an  $r^2>0.35$ .

## Discussion

Suggestive evidence of linkage at 5p13–q12 was obtained following analysis of a genome-wide scan using Affymetrix 10K SNP arrays in a large Tasmanian prostate cancer pedigree. Subsequent microsatellite genotyping of further

cases confirmed and extended this finding. Eight related cases shared a common haplotype ( $P=0.0017$ ). Nominal evidence for linkage to chromosome 5 has been provided by several previous studies,<sup>5,15–17</sup> including the ICPG study of 1233 prostate cancer families.<sup>4</sup> Through the use of a gene prioritisation tool, resequencing and a follow-up association study, two polymorphisms within the *ITGA2* gene, 3'UTR in/del and C807T, were found to be associated with prostate cancer risk in independent datasets comprising familial cases, sporadic cases and controls.

The *ITGA2* gene encodes for the  $\alpha 2\beta 1$  integrin receptor, a cell adhesion molecule, responsible for interaction and mediation of signalling events with the extracellular matrix (ECM).<sup>29</sup> The  $\alpha 2\beta 1$  integrin receptor is expressed on many epithelial cell types, and in normal prostate tissue expression is restricted to basal epithelial cells.<sup>30,31</sup> High levels of  $\alpha 2\beta 1$  integrin expression characterise prostate cancer stem cells,<sup>32</sup> and upregulation of  $\alpha 2\beta 1$  integrin may potentially explain the observation that over 80% of prostate cancer metastases are to bone,<sup>29,31</sup> as collagen is the main component of bone ECM and the preferred ligand for  $\alpha 2\beta 1$  integrin.<sup>33</sup>

Candidate gene association studies have also linked the *ITGA2* gene with cancer risk. Two separate studies of breast cancer and oral cancer have reported an association of the C807T *ITGA2* polymorphism with disease risk.<sup>34,35</sup> In a study comprising 500 sporadic breast cancer cases and 500 controls, the 807C–1648G haplotype was found to decrease risk compared to non-carriers, and higher grade breast tumours were significantly associated with the 807T/T genotype.<sup>34</sup> In addition, the 807T allele was significantly associated with increased risk of oral cancer ( $P<0.001$ ).<sup>35</sup> Studies conducted in platelets have demonstrated that the 807T allele is associated with a two-fold higher density of the *ITGA2* receptor on the platelet cell surface and may function to alter mRNA splicing as it is contained within a predicted ESE. However, it remains to be determined whether this polymorphism is directly influencing gene expression.<sup>36</sup>

Given the evidence presented here, the 3'UTR in/del polymorphism remains of particular interest, as it may be important in regulating *ITGA2* mRNA. Although we have not identified a functional role for the 3'UTR in/del polymorphism, we have identified two SNPs (rs6880055 and rs57674800) in high LD with this SNP. These SNPs lie within predicted miRNA-binding sites and the presence of the alternative genotype at these sites is predicted to disrupt miRNA binding to this sequence. Thus, presence of one or both of the alternate alleles at these SNPs may result in loss of miRNA-mediated transcriptional regulation. It is also possible that other SNPs in LD with the 3'UTR in/del identified here are causally related to prostate cancer.

Although variants within *ITGA2* have been significantly associated with other cancers, to our knowledge this is the first reported association between *ITGA2* and prostate



cancer risk. The cancer genetic markers of susceptibility (CGEMS) prostate cancer genome-wide association scan ([www.caintegrator.nci.gov/cgems/browse.do](http://www.caintegrator.nci.gov/cgems/browse.do)) reported no significant associations between prostate cancer and the 1569 SNPs genotyped within the 14 Mb interval identified here. However, attempts to replicate significant associations of risk variants to previously identified prostate cancer susceptibility loci frequently fail to support the original findings (*ELAC2*:<sup>37,38</sup>*MSR1*:<sup>39</sup>*RNASEL*:<sup>40</sup>). Reasons for this have been widely discussed and include: the heterogeneity of genes contributing to prostate cancer risk in different populations; different selection criteria for each of the study populations examined; and the variable contribution of environmental factors interacting with multiple genetic factors.

In summary, we have presented evidence for a prostate cancer susceptibility gene on chromosome 5p13–q12, with two *ITGA2* polymorphisms showing a significant association with prostate cancer risk in the Tasmanian population. Although the functional SNP is still to be identified, considerable circumstantial evidence, provided by *in vivo* and *in vitro* studies, supports a role for this gene in tumour development. Replication of this result in independent data sets and studies elucidating the role of this gene will be of fundamental interest to the field of prostate cancer biology.

### Acknowledgements

We thank Annette Banks, Kris Hazelwood, Sally Inglis and Annette Edwards for their assistance in creating genealogical records and/or sample collection; and Melanie Bahlo, Terence Speed and David Goldgar for useful suggestions and discussions about methods of linkage analysis. The Royal Hobart Hospital Research Foundation, The Cancer Council of Tasmania, CaPCure, The Mazda Foundation and Perpetual Charitable Trusts and the Australian Cancer Research Fund supported this study. Dr Dickinson is a Cancer Council of Tasmania Research Fellow. We are also greatly indebted to the participants of the TasPac study, the TCR staff, the Tasmanian Urologists and Pathologists, and the wider Tasmanian clinical and research community.

### References

- Parkin DM: Global cancer statistics in the year 2000. *Lancet Oncol* 2001; **2**: 533–543.
- Gronberg H, Damber L, Damber JE, Iselius L: Segregation analysis of prostate cancer in Sweden: support for dominant inheritance. *Am J Epidemiol* 1997; **146**: 552–557.
- Schaid DJ, McDonnell SK, Blute ML, Thibodeau SN: Evidence for autosomal dominant inheritance of prostate cancer. *Am J Hum Genet* 1998; **62**: 1425–1438.
- Xu J, Dimitrov L, Chang BL *et al*: A combined genomewide linkage scan of 1,233 families for prostate cancer-susceptibility genes conducted by the international consortium for prostate cancer genetics. *Am J Hum Genet* 2005; **77**: 219–229.
- Wiklund F, Gillanders EM, Albertus JA *et al*: Genome-wide scan of Swedish families with hereditary prostate cancer: suggestive evidence of linkage at 5q11.2 and 19p13.3. *Prostate* 2003; **57**: 290–297.
- Amundadottir LT, Sulem P, Gudmundsson J *et al*: A common variant associated with prostate cancer in European and African populations. *Nat Genet* 2006; **38**: 652–658.
- Gudmundsson J, Sulem P, Manolescu A *et al*: Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet* 2007; **39**: 631–637.
- Haiman CA, Patterson N, Freedman ML *et al*: Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet* 2007; **39**: 638–644.
- Yeager M, Orr N, Hayes RB *et al*: Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet* 2007; **39**: 645–649.
- Eeles RA, Kote-Jarai Z, Giles GG *et al*: Multiple newly identified loci associated with prostate cancer susceptibility. *Nat Genet* 2008; **40**: 316–321.
- Gudmundsson J, Sulem P, Rafnar T *et al*: Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. *Nat Genet* 2008; **40**: 281–283.
- Thomas G, Jacobs KB, Yeager M *et al*: Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet* 2008; **40**: 310–315.
- Risch N: The genetic epidemiology of cancer: interpreting family and twin studies and their implications for molecular genetic approaches. *Cancer Epidemiol Biomarkers Prev* 2001; **10**: 733–741.
- Wang WY, Barratt BJ, Clayton DG, Todd JA: Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet* 2005; **6**: 109–118.
- Smith JR, Freije D, Carpten JD *et al*: Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search. *Science* 1996; **274**: 1371–1374.
- Goddard KA, Witte JS, Suarez BK, Catalona WJ, Olson JM: Model-free linkage analysis with covariates confirms linkage of prostate cancer to chromosomes 1 and 4. *Am J Hum Genet* 2001; **68**: 1197–1206.
- Hsieh CL, Oakley-Girvan I, Balise RR *et al*: A genome screen of families with multiple cases of prostate cancer: evidence of genetic heterogeneity. *Am J Hum Genet* 2001; **69**: 148–158.
- Slager SL, Zarfes KE, Brown WM *et al*: Genome-wide linkage scan for prostate cancer aggressiveness loci using families from the University of Michigan Prostate Cancer Genetics Project. *Prostate* 2006; **66**: 173–179.
- Sato Y, Sugie R, Tsuchiya B, Kameya T, Natori M, Mukai K: Comparison of the DNA extraction methods for polymerase chain reaction amplification from formalin-fixed and paraffin-embedded tissues. *Diagn Mol Pathol* 2001; **10**: 265–271.
- Kennedy GC, Matsuzaki H, Dong S *et al*: Large-scale genotyping of complex DNA. *Nat Biotechnol* 2003; **21**: 1233–1237.
- O'Connell JR, Weeks DE: PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998; **63**: 259–266.
- Abecasis GR, Cherny SS, Cookson WO, Cardon LR: Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002; **30**: 97–101.
- Sobel E, Lange K: Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. *Am J Hum Genet* 1996; **58**: 1323–1337.
- Thomson R, Quinn S, McKay J *et al*: The advantages of dense marker sets for linkage analysis with very large families. *Hum Genet* 2007; **121**: 459–468.
- Lander E, Kruglyak L: Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 1995; **11**: 241–247.
- Thornton T, McPeck MS: Case-control association testing with related individuals: a more powerful quasi-likelihood score test. *Am J Hum Genet* 2007; **81**: 321–337.
- Jacquelin B, Tarantino MD, Kritzik M *et al*: Allele-dependent transcriptional regulation of the human integrin alpha2 gene. *Blood* 2001; **97**: 1721–1726.

- 28 Kunicki TJ: The influence of platelet collagen receptor polymorphisms in hemostasis and thrombotic disease. *Arterioscler Thromb Vasc Biol* 2002; **22**: 14–20.
- 29 Koistinen P, Heino J: Integrins in Cancer Cell Invasion; in Heino J (ed): *Cancer Invasion*. Austin, USA, 2006, Vol. 19. Available online at: <http://www.landesbioscience.com>
- 30 Knox JD, Cress AE, Clark V *et al*: Differential expression of extracellular matrix molecules and the alpha 6-integrins in the normal and neoplastic prostate. *Am J Pathol* 1994; **145**: 167–174.
- 31 Knudsen BS, Miranti CK: The impact of cell adhesion changes on proliferation and survival during prostate cancer development and progression. *J Cell Biochem* 2006; **99**: 345–361.
- 32 Collins AT, Berry PA, Hyde C, Stower MJ, Maitland NJ: Prospective identification of tumorigenic prostate cancer stem cells. *Cancer Res* 2005; **65**: 10946–10951.
- 33 Bonkhoff H, Stein U, Remberger K: Differential expression of alpha 6 and alpha 2 very late antigen integrins in the normal, hyperplastic, and neoplastic prostate: simultaneous demonstration of cell surface receptors and their extracellular ligands. *Hum Pathol* 1993; **24**: 243–248.
- 34 Langsenlehner U, Renner W, Yazdani-Biuki B *et al*: Integrin alpha-2 and beta-3 gene polymorphisms and breast cancer risk. *Breast Cancer Res Treat* 2006; **97**: 67–72.
- 35 Vairaktaris E, Yapijakis C, Derka S *et al*: Association of platelet glycoprotein Ia polymorphism with minor increase of risk for oral cancer. *Eur J Surg Oncol* 2006; **32**: 455–457.
- 36 Kunicki TJ, Kritzik M, Annis DS, Nugent DJ: Hereditary variation in platelet integrin alpha 2 beta 1 density is associated with two silent polymorphisms in the alpha 2 gene coding sequence. *Blood* 1997; **89**: 1939–1943.
- 37 Camp NJ, Tavtigian SV: Meta-analysis of associations of the Ser217Leu and Ala541Thr variants in ELAC2 (HPC2) and prostate cancer. *Am J Hum Genet* 2002; **71**: 1475–1478.
- 38 Severi G, Giles GG, Southey MC *et al*: ELAC2/HPC2 polymorphisms, prostate-specific antigen levels, and prostate cancer. *J Natl Cancer Inst* 2003; **95**: 818–824.
- 39 Sun J, Hsu FC, Turner AR *et al*: Meta-analysis of association of rare mutations and common sequence variants in the MSR1 gene and prostate cancer risk. *Prostate* 2006; **66**: 728–737.
- 40 Nupponen NN, Wallen MJ, Ponciano D *et al*: Mutational analysis of susceptibility genes RNASEL/HPC1, ELAC2/HPC2, and MSR1 in sporadic prostate cancer. *Genes Chromosomes Cancer* 2004; **39**: 119–125.
- 41 Lander ES, Schork NJ: Genetic dissection of complex traits. *Science* 1994; **265**: 2037–2048.