Functional analysis of *RAD51B* and *RAD23B* SNPs using Circular Chromosome Conformation Capture (4C) in human prostate cell lines

C. Cieza-Borrella[1], S. Wakerell[1], Q. Karlsson[1], R.A. Eeles[1,2], Z. Kote-Jarai[1]

Introduction

Prostate Cancer (PrCa) is the most frequently diagnosed cancer among men in developed countries and one of the most heritable solid tumours with 58% of cases estimated to be due to genetic factors. The effectiveness of PSA screening remains controversial and new molecular biomarkers are needed to improve application of treatment. Findings have led to a mixed model of common and rare variants, with variable effect sizes, to explain PrCa genetic origin. Rare mutations in *BRCA2* and *BRCA1* give rise to moderately elevated risk, whereas Genome-wide Association Studies (GWAS) have identified over 100 common, low penetrance PrCa susceptibility variants, some near or within *RAD51B* and *RAD23B* DNA repair genes[1,2,3].

RAD51B is a member of the RAD51 protein family with a relevant role in the homologous recombination repair (HRR) pathway of double stranded DNA breaks. Its overexpression has been found to cause cell cycle delay and apoptosis. Variations in the gene sequence have been associated with breast, ovarian, endometrial, colorectal, head and neck cancer and acute leukaemia. RAD23B is involved in the nucleotide excision repair (NER) system and although its role is still not fully understood, polymorphisms in this gene have been associated with breast cancer susceptibility.

Chromatin is highly organised in topologically associating domains (TADs) comprising 100kb to 1Mb clusters of locally interacting DNA regions. The intra- and inter- dynamic of these domains allows the interaction of different regions involving regulatory elements such as promoters, enhancers and silencers, which consequently leads to the regulation of gene expression and recombination[4].

Aim

The aim of this study is to understand the mechanism behind the association between single nucleotide polymorphisms (SNPs), lying in non-coding regions near or within *RAD51B* and *RAD23B*, and prostate cancer risk by analysing their interactions with the whole genome using Circular Chromosome Conformation Capture (4C).

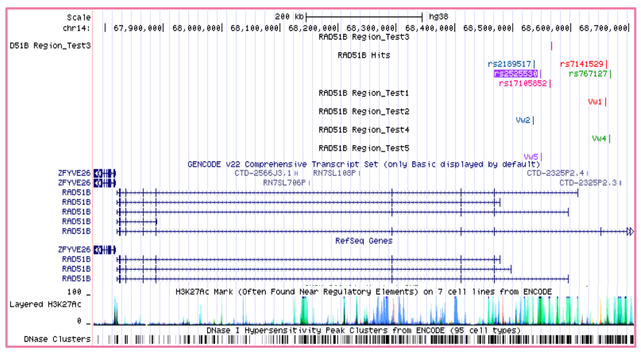


Figure 1: Location of the analysed *RAD51B* SNPs and viewpoints selected by the 4C Primer Designer

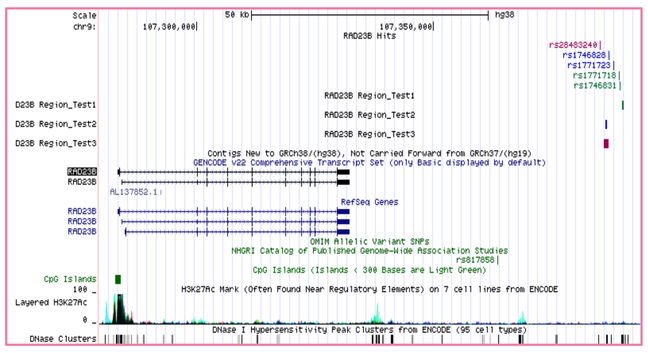


Figure 2: Location of the analysed *RAD23B* SNPs and viewpoints selected by the 4C Primer Designer

Cell line	Species	Tissue	Type	Androgens Sensitive	rs7141529	rs2189517	rs767127	rs2525530	rs1771718	rs1771723	rs28483240	rs1746831	rs1746828
LNCaP	Human	Prostate	Adenocarcinoma	YES	G/G	G/A	C/C	C/T	A/A	T/T	T/T	G/G	T/T
DuCaP	Human	Prostate	Adenocarcinoma	YES	G/A	G/A	C/T	C/T	A/A	T/T	T/T	G/G	T/C
PC3	Human	Prostate	Adenocarcinoma	NO	G/A	A/A	C/T	T/T	A/A	T/T	T/T	G/G	T/T
PNT1a	Human	Prostate	Normal	NO	G/G	G/A	C/C	C/T	A/A	T/T	T/T	G/G	T/T
RWPE1	Human	Prostate	Normal	YES	G/A	G/A	C/T	C/T	A/G	T/T	T/T	G/G	T/C

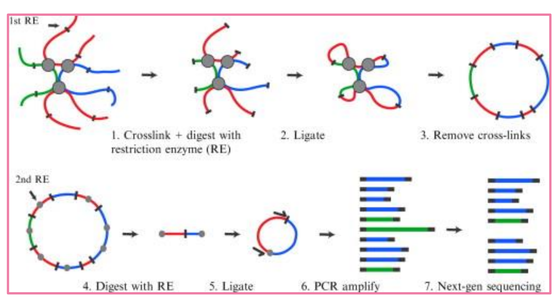
Table 1. Cell lines characteristics and genotypes of the polymorphisms analysed in *RAD51B* and *RAD23B* genes.

Materials and Methods

- **We analysed 9 SNPs associated with PrCa risk according to previous GWAS and fine mapping studies[1,2] (table 1): rs7141529 (Original SNP) was described in an initial GWAS[1] and remained the most significantly associated *RAD51B* variant after stepwise logistic regression analysis[2]; rs2189517, a *RAD51B* intronic variant not correlated with the Original SNP and rs2525530, also intronic and showing the highest LD with rs2189517; rs767127, in complete LD with the Original SNP and found by our collaborators to be interacting with *RAD51B* promoter; rs1771718, *RAD23B* "European hit" and 1771723, rs28483240, rs1746831 and rs1746828, all in highest LD with the European hit.**
- **Five human epithelial prostate cell lines were selected for our study: LNCaP, DuCaP, PC3, PNT1a and RWPE-1. Genotypes for each SNP were provided by the PRACTICAL Consortium Oncoarray project and confirmed by Sanger sequencing (table 1).**
- **Design of the 4C assay was carried out using 4C Primer Designer for 4C viewpoints program (Nobrega lab-Human Genetics Department-University of Chicago, USA). The proximity in the genome of SNPs rs1771718 and rs1746831 and SNPs rs1771723 and rs1746828 made it possible to extrapolate the results for each pair from the same viewpoint (figures 1 and 2).**
- **4C libraries were prepared following previously described protocols[5,6] (figure 3a) and optimising the conditions according to the characteristics of the selected viewpoints and cell lines. High complexity PCRs (6nt barcodes and adaptors attached) were set over our 4C templates (figure 3b). Once we obtained the desired result, PCRs were quantified and quality checked by fragment size analysis using Agilent Bioanalyzer 2100, qPCR and sequencing on a single flow cell in a MiSeq (Illumina). 4C libraries were subsequently sequenced on an HiSeq 2500 (RAPID SR 100 cycles, Illumina).**
- **Undigested and self-ligated reads were removed and valid reads were assigned to restriction fragments. Bioinformatics analysis was performed using the R program *4C-ker*[7]. *4C-ker* is a Hidden-Markov Model based pipeline which utilises adaptive window sizes to correct for differences in signal coverage in near-bait and far-cis regions. *4C-ker* quality control thresholds are a minimum of 1 million total reads, more than 40% of reads in cis regions and over 40% fragment coverage in the 200kb region around the bait, for a 4bp cutter.**

4C assay

3a)



3b)

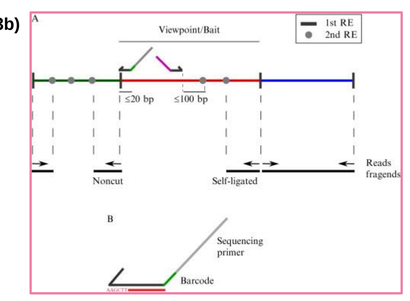


Figure 3: 4C technology: a) Overview of the 4C procedure; b) Details of the 4C PCR strategy[5].

Results

Our results show a high number of interactions between the analysed SNPs and the rest of the genome. Focusing on *cis* associations, it is worth mentioning the significant interactions between the promoter of *FUT8*, whose overexpression has been associated with PrCa, and rs767127 and rs2189517 in almost all cell lines with the exception of RWPE-1 (figure 4). We found additional interactions between the rs2189517 variant and *EVL/YY1* in PC3, PNT1a and RWPE-1 (figure 5). These genes might warrant further analysis, as *EVL* encodes a protein that exhibits RAD51 protein binding, DNA binding and DNA-annealing activities and the overexpression of YY1 transcription factor has been correlated with prostate tumour progression. We observed this same significant interaction in all cell lines except LNCaP, in which it was also detected but not significantly, when analysing rs767127. This SNP also presented a high number of interactions in PC3 and DuCaP with *FOS*, which is associated with tumour progression. Moreover, the rs767127 interaction with *RAD51B* promoter, previously observed by our collaborators, was additionally confirmed in all cell lines (figure 6). We also found remarkable the significant interactions detected between *DICER1* and rs767127 and rs2189517 in PC3 and DuCaP.

Regarding *RAD23B* SNPs, significant interactions were found between two genes closely located on chromosome 9: *PTCH1* (*9q22.32*) and *XPA* (*9q22.23*), and all the *RAD23B* SNPs included in our study, although variations in the type of cells showing these contacts was observed (figures 7 and 8). Overexpression of *PTCH1*, involved in the *hedgehog* pathway and acting as tumour suppressor gene, has been described to be associated with progression and metastases in PrCa. On another hand, *XPA* encodes a zinc finger protein involved in the nucleotide excision repair (NER) pathway whose family members are involved in PrCa development.

One of the main limitations of this work has been the high amount of self-ligated product observed in certain 4C libraries (rs7141529 and rs2525530), which was translated into a decrement in usable read counts, below the thresholds required for analysis. We are also planning further analysis using other bioinformatics programs such as *fourSig*, in order to elucidate the nature of observed interactions.

Figure 4

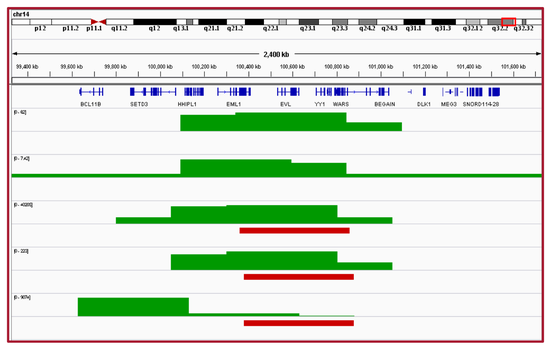


Figure 5

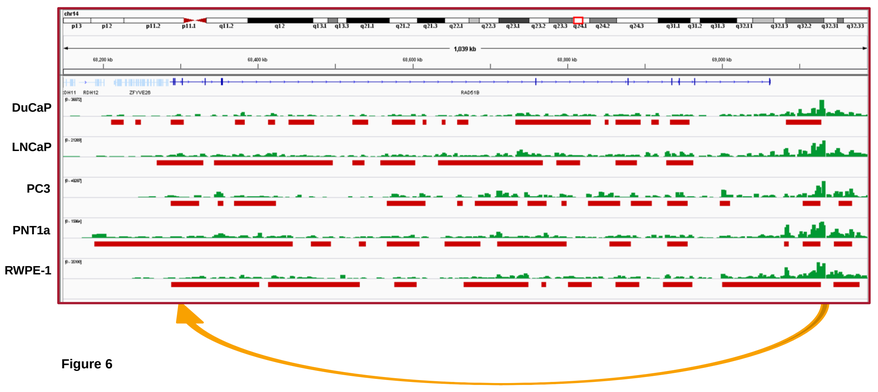


Figure 6

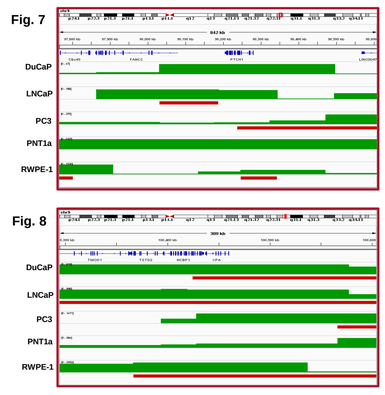


Figure 4: Zoomed in image detailing the significant interactions observed between rs2189517 and the promoter of *FUT8.* In RWPE-1 cell line the significance lies only within the rest of the gene; Figure 5: Zoomed in image representing the significant interactions between rs2189517 and *EVL/YY1 loci* in PC3, PNT1a and RWPE-1 cell lines. DuCap and LNCaP still showed interactions not reaching any significance; Figure 6: Representation of the interaction observed for the viewpoint selected for the study of the rs767127 SNP and *RAD51B* promoter region; Figure 7: Zoomed in image showing the significant interactions observed between the viewpoint selected for the study of rs1771723 and rs1746828 SNPs and *PTCH1* in PC3 and RWPE-1; Figure 8: Zoomed in image highlighting the significant interactions detected between the viewpoint selected for the study of the rs1771718 and rs1746831 SNPs and *XPA* gene in DuCap, LNCaP and RWPE-1 cell lines. Red boxes represent the regions found as significant by the *4C-ker* program whereas the green bars are the normalised read counts.

CONCLUSIONS

Our 4C analysis showed potentially important interactions in human prostate cell lines of high interest. Here we highlight the interactions between three *RAD51B* intronic SNPs and *FUT8*, *EVL/YY1*, *FOS* and *DICER1* genes, as well as the interaction between rs7141529 and *RAD51B* promoter. The interactions observed between the *RAD23B* index SNPs and *PTCH1* and *XPA* genes might be important showing association between this *locus* and prostate cancer development. Further analysis, comprising of 3C, will be carried out to validate our results.

AFFILIATIONS

1. The Institute of Cancer Research, 123 Old Brompton Road, London, UK
2. The Royal Marsden NHS Foundation Trust, Fulham Road, London, UK

REFERENCES

1. Eeles RA *et al*., *Nat Genet*, 2013
2. Olama AA *et al*., *Hum Mol Genet*, 2015
3. Saunders EJ *et al., BJC*, 2016
4. Ea V *et al*., *Genes,* 2015
5. Ven de Werken HJ *et al*., *Methods Enzymol,* 2012
6. Splinter E *et al*., *Methods,* 2012
7. Raviram R *et al*., *PLoS Comput Biol,* 2016

ACKNOWLEDGEMENTS

- PRACTICAL Consortium
- Richard and Debbi Burston

CONTACT

Clara Cieza-Borrella (Postdoctoral Fellow)

Division of Genetics and Epidemiology

The Institute of Cancer Research, London, UK

clara.cieza-borrella@icr.ac.uk