ORIGINAL ARTICLE

A Y chromosomal influence on prostate cancer risk: the multi-ethnic cohort study

S Paracchini, C L Pearce, L N Kolonel, D Altshuler, B E Henderson, C Tyler-Smith

....

See end of article for authors' affiliations
....

Correspondence to:
Dr C Tyler-Smith, The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK; cts@sanger.ac.uk
....

Background: A Y chromosomal role in prostate cancer has previously been suggested by both cytogenetic findings and patterns of Y chromosomal gene expression. We took advantage of the well established and stable phylogeny of the non-recombining segment of the Y chromosome to investigate the association between Y chromosomal DNA variation and prostate cancer risk.
Methods: We examined the distribution of 116 Y lineages in 930 prostate cancer cases and 1208 controls from four ethnic groups from a cohort study in Hawaii and California.
Results: One lineage, found only among the Japanese group in our study, was associated with a statistically significant predisposition to prostate cancer (odds ratio (OR) = 1.63; 95% confidence interval (CI) 1.07 to 2.47), and, in particular, to high severity disease in younger individuals (OR = 3.89; 95% CI 1.34 to 11.31).
Conclusions: This finding suggests that a Y chromosomal factor contributes significantly to the development of prostate cancer in Japanese men.

P rostate cancer is probably influenced by both genetic and environmental factors, but few of these have been identified. A Y chromosomal role in prostate cancer has been suggested, initially because the loss of the Y chromosome is the most common chromosomal change detected in prostate cancer tissue.[1][2] Although the Y is lost in some normal tissues from elderly men,[3] studies have shown Y loss in the prostate to be confined to malignant prostate epithelium; it has not been detected in adjacent apparently healthy tissue.[2][4] Gene expression analysis in prostate cancer has revealed aberrant patterns for some Y chromosomal genes.[5] In particular, *SRY*, the sex determining gene on the Y chromosome, is down regulated in prostate cancer. Recently, SRY has been shown to be a negative regulator of the androgen receptor (AR), suggesting that its loss may increase AR activity and consequently stimulate cancer growth.[6]

The human Y chromosome does not recombine for almost its entire length, allowing binary variants to be organised into a well characterised and stable phylogeny,[7–9] thereby providing a uniquely powerful resource for association studies.[10] In the same way, any mutation predisposing to or protecting against prostate cancer will be in complete linkage disequilibrium with all other binary markers (mostly single nucleotide polymorphisms), and will fall on a particular lineage (haplogroup) of the Y phylogeny (fig 1), regardless of physical distance between markers. Therefore, lineages carrying a predisposing or protective factor will be found at a higher frequency among cases or controls, respectively. Owing to the strong geographical differentiation of the Y chromosome haplogroup distribution, such a lineage specific association is expected to be geographically specific and, therefore, may be confined to a single population in a study.

To investigate a possible role of the Y chromosome in prostate cancer, we compared the frequency of 116 Y haplogroups in four ethnic groups from a large cohort study in Hawaii and Los Angeles.

SUBJECTS AND METHODS

Study population
This case–control study was nested within the Hawaii-Los Angeles Multiethnic Cohort (MEC) study. The study population included 930 incident prostate cancer cases and 1208 male controls from the African-American, white, Latino, and Japanese subjects enrolled in the cohort (table 1).

Details of the MEC study have been published previously.[11] Briefly, over 200 000 men and women between the ages of 45 and 75 years and residing in Hawaii and California completed a questionnaire that included data on demographic, lifestyle, and health characteristics, as well as a comprehensive dietary survey. This cohort is broadly similar to the general populations of those ethnic groups in Los Angeles and Hawaii.[11]

Participants in the MEC are followed for incident cancers. Incident case ascertainment is completed by computer linkage of the cohort with the Surveillance, Epidemiology, and End Results cancer registries in Hawaii and Los Angeles, as well as with the California Cancer Registry. Incident prostate cancer cases and a random sample of male controls in the MEC were contacted by telephone and asked to provide a blood sample. The overall participation rate for blood collection was 72% for cases and 69% for controls. The men who agreed to participate in the blood collection provided written informed consent following study approval by both the University of Hawaii and the University of Southern California Institutional Review Boards.

Genotyping
All samples were provided to the laboratory for genotyping, blinded as to case or control status and to ethnicity using a unique identifier. The samples were typed for 118 Y chromosomal binary markers, defining 116 haplogroups (lineages), using a method described previously, where single base primer extension products are analysed by mass spectrometry.[12] Although 118 markers were genotyped, several do not define unique haplogroups, but are required for assignment to the correct branch of the tree during the

....

Abbreviations: MEC, Hawaii-Los Angeles Multiethnic Cohort Study
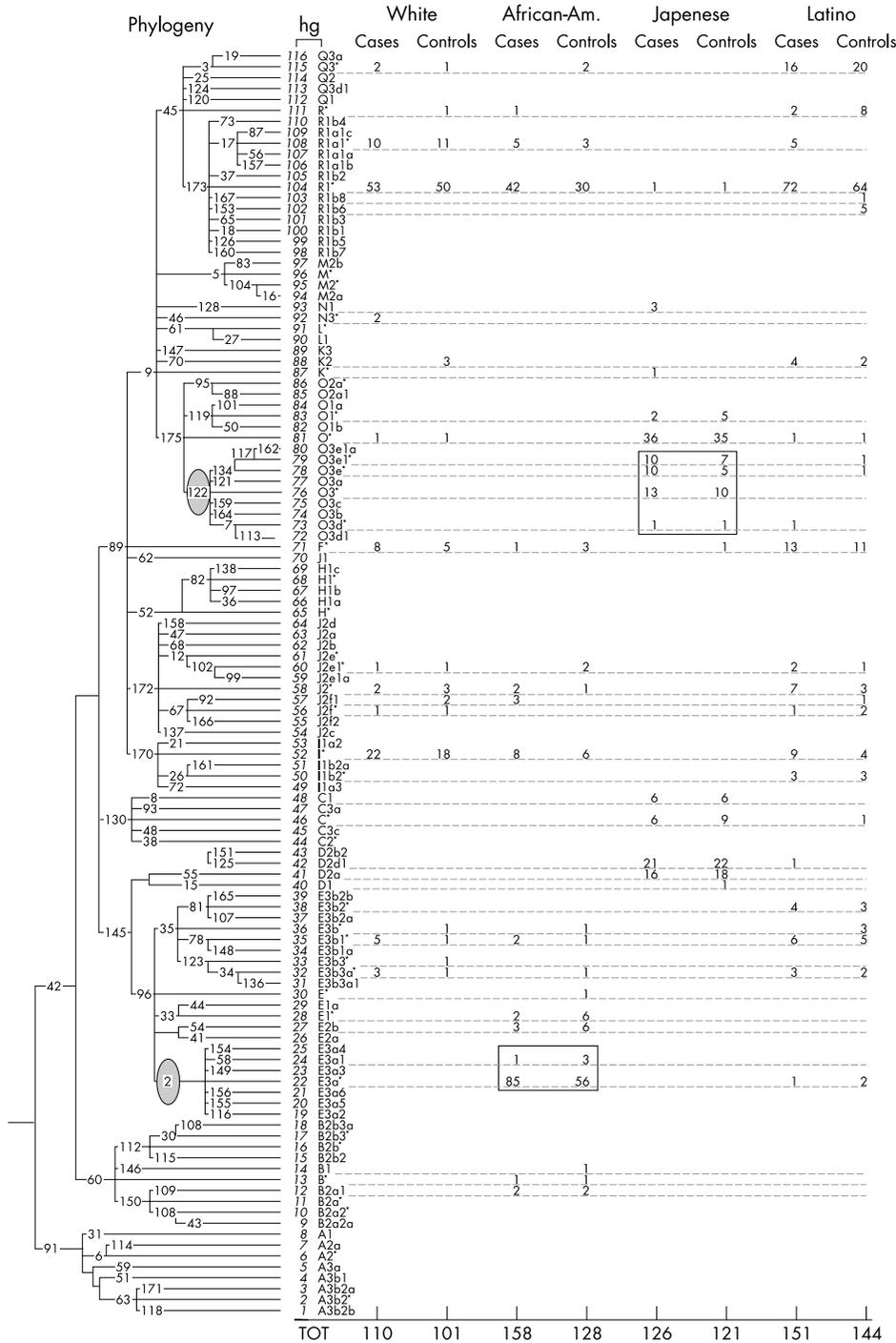
Figure 1 Y haplogroup distribution in prostate cancer cases and controls from four populations. The tree on the left shows the Y chromosomal phylogeny defined by 118 markers (numbers on the tree, with the 'M' prefix omitted) identifying 116 haplogroups (italic numbers) for which the equivalent name according to the Y Chromosome Consortium nomenclature[17] is given. The frequency of the haplogroups found in the pilot analysis is shown by ethnic origin and case–control status. A grey shadowed oval and a box indicate, respectively, the lineages and the haplogroup frequencies that showed a significant association in the pilot study.

early stages of the hierarchical genotyping strategy. The analysis of additional Japanese cases and controls was performed using the Sequenom mass spectrometry system according to the manufacturer's specifications (Sequenom, San Diego, CA, USA). The PCR primer sequences were: M122F: 5'-AGCGGATAACAGTCACTTGCTCTGTGTTAG-3', M122R: 5'-AGCGGATAACCTCTACTTAGTTGCCTTTTGG-3'. The sequence of the oligonucleotide used in the primer extension reaction was: M122: 5'-TCAGATTTTCCCC TGAGAGC-3'.

Blind replicate samples of 2.2% were included. The genotypes matched in 100% of these tests, confirming the reproducibility of the typing. A total of 1039 samples were

screened for the complete set of markers (fig 1). Another 795 African-Americans (288 cases and 507 controls) were typed for only one marker (M02) to determine whether they belonged to the E3a lineage as this was statistically significantly (p = 0.05) associated with prostate cancer risk in this ethnic group after genotyping 158 cases and 128 controls. An additional 304 (97 cases and 207 controls) Japanese samples were genotyped for one marker (M122) to determine if they belonged to the O3 lineage (grouping O3d*, O3, O3e, O3e1* lineages) because this lineage was statistically significantly (p = 0.03) associated with prostate cancer risk in this ethnic group after genotyping 126 cases and 121 controls.

Statistical analyses

To test whether the controls were properly matched to the cases, we used Rousset's test of population differentiation,[13] which is contained in the Arlequin package.[14] We found no significant overall difference between cases and controls within each ethnic group, indicating that they were well matched, while the four ethnic groups were significantly different from one another.

The associations between haplotype and risk of prostate cancer were analysed at both the level of individual Y haplogroups and at each node along the phylogenetic tree, using standard case–control approaches (fig 1 and supplementary table 1). All statistical analyses were conducted using the SAS statistical package (version 8.0; SAS Institute Inc., Cary, NC, USA). All reported P values are two sided. We considered age as a potential confounding and modifying factor.

Prostate tumours were classified according to disease severity using a combination of stage and Gleason grade. Low severity was defined as tumours localised to the prostate with a Gleason grade of <8. High severity was defined as tumours that had regional extension or distant metastases, regardless of grade, and tumours that were localised to the prostate, but had a Gleason grade of 8 or higher.

RESULTS

The characteristics of the study population are shown in table 1. Forty one percent of the prostate cancer cases were classified as high severity and 12.7% reported a first degree relative with prostate cancer, compared with 8.3% of the controls.

Of the 116 haplogroups defined by the 118 markers used in this study, 41 were observed in our sample (fig 1). The distributions of the haplogroups varied markedly among the four ethnic groups.

With the exception of one lineage (E3a) in the African-Americans and one (O3) in the Japanese, no association with prostate cancer was observed with any of the lineages or groupings of lineages. Initially, lineage E3a, which is of African origin, was associated with increased risk of prostate cancer in the African-Americans (age adjusted odds ratio (OR) = 1.58, 95% confidence interval (CI) 1.0 to 2.51, p = 0.05); however, after increasing our sample size the risk was attenuated and was not statistically significant

Table 1 Descriptive characteristics of the study population†

Characteristic	Controls (n = 1208)	Cases (n = 930)
Ethnicity‡		
White	101	110
African-American	635	446
Japanese	328	223
Latino	144	151
Disease severity§		
Low	–	491
High	–	383
Age (years)		
<65	425	257
65+	783	672
First degree family history		
Yes	101	119
No	997	721

†Numbers in table do not always sum to total due to missing data.
‡Self reported.
§Low severity = local stage and Gleason grade <8; high severity = local stage+Gleason grade 8 and/or regional/metastatic stage.

Table 2 Distribution of the O3 lineage versus all other lineages combined in Japanese men

	O3 lineage	
Disease status	All ages (n (%))	<65 (n (%))
Controls	56 (17.1)	28 (16.7)
All cases	56 (25.1)	14 (35.9)
Low severity†	28 (22.6)	6 (27.3)
High severity†	26 (28.0)	7 (43.8)

†Numbers do not sum to all cases owing to missing severity data

(age adjusted OR = 1.25, 95% CI 0.98 to 1.59, p = 0.08).

In the initial set of Japanese samples, we observed an increased risk associated with the O3 lineage (age adjusted OR = 2.12, 95% CI 1.07 to 4.22, p = 0.03; supplementary table 1), and the magnitude of this association was increased after stratifying by age and disease severity (supplementary table 2), suggesting that genotyping additional samples was warranted. Because age appeared to be acting as a modifier, all additional analyses are presented for both unadjusted and age stratified samples. The risk observed independently in the additional samples was consistent with our original observation (supplementary table 2). The modifying effects of age and disease severity were also observed again in the additional samples. The frequency of the O3 lineage is shown in table 2.

Overall, the risk of prostate cancer associated with the O3 lineage in Japanese in this study was 1.63 (95% CI 1.07 to 2.47, p = 0.02) (table 3). This risk was modified by both age and disease severity; Japanese men under the age of 65 years carrying lineage O3 were 2.8 times as likely to have prostate cancer compared with men belonging to the other lineages (95% CI 1.30 to 6.05, p = 0.009) whereas men 65 years and older were only 1.39 times as likely to have prostate cancer if they carried this lineage (table 3).

DISCUSSION

In this study, we found that the Y chromosome lineage O3, present almost exclusively in our Japanese sample, was associated with a significantly increased risk of prostate cancer, which depended upon age at diagnosis and disease severity. The strong geographical clustering of Y chromosome haplogroups explains why this association was only found in the Japanese men; only the men with the O3 lineage will carry the putative mutation and this lineage is found exclusively in populations of Asian origin.[8] The other ethnic groups in our study do not carry this lineage at significant frequency and so cannot show the association.

The marker M122 that defines the O3 lineage lies within an intron of the predicted protein coding gene *CYorf15B* (http://www.ensembl.org/Homo_sapiens/mapview?chr = Y), but seems unlikely to influence gene expression or prostate cancer development directly. Instead, M122 probably marks a lineage that carries changes elsewhere on the chromosome that alter gene sequence, copy number, expression pattern, or maintenance of the chromosome. Because complete linkage disequilbrium extends for nearly the entire length of the Y chromosome owing to the lack of recombination, M122 does not provide any information on the physical location of the putative variant.

The significance of this finding in our Japanese men must be considered in light of the multiple hypotheses that have been tested within this group. The purpose of this study is

Table 3 Odds ratios for prostate cancer risk with the Y chromosome O3 lineage in the Japanese

Disease status	Unadjusted			Age<65 years			Age 65 years+		
	OR	95% CI	p value	OR	95% CI	p value	OR	95% CI	p value
All cases	1.63	1.07 to 2.47	0.02	2.80	1.30 to 6.05	0.009	1.39	0.82 to 2.38	0.22
Low severity	1.42	0.85 to 2.36	0.18	1.88	0.68 to 5.21	0.23	1.30	0.70 to 2.42	0.41
High severity	1.89	1.10 to 3.22	0.02	3.89	1.34 to 11.31	0.01	1.54	0.80 to 2.99	0.20

hypothesis generation, and therefore no p value correction has been applied. Moreover, the assumption of independence cannot be applied to the tests we performed because the Y phylogeny makes them strictly correlated, and therefore a Bonferroni correction would be too conservative. Nevertheless, we attempted to minimise the probability of a type I error by increasing the sample size and analysing the two samples separately to determine if an association of similar magnitude was observed. This was indeed the case for the Japanese association with lineage O3.

One potential explanation for the finding in the Japanese could be population stratification. Although this is generally more of a concern in admixed populations,[15] we nevertheless used two approaches to investigate whether stratification could account for our findings. Firstly, Y haplotypes provide powerful information about geographical origin and can be used to assess admixture in cases and controls. Using published data on Y haplogroup distributions,[9] we could assign >98% of the Japanese Y chromosomes to an Asian origin, with no significant difference between cases and controls, indicating no detectable male mediated admixture in this ethnic group and no differential effect between cases and controls. Secondly, using a previously described method,[15] a set of 36 high frequency unlinked autosomal loci was used to test for stratification in these samples and no significant stratification was detected (Freedman *et al*, unpublished observations). Our results cannot be explained by gross levels of population stratification.

We did not find any statistically significant associations with prostate cancer risk for the other Y chromosome lineages in the other ethnic groups in this study, but this may be due to low power for detecting small effects in common lineages or moderate effects with rarer lineages.

Our results suggest that the O3 Y chromosome lineage found in Japanese men is associated with increased risk of prostate cancer; specifically, young Japanese men carrying the Y lineage O3 have nearly a fourfold increased risk of developing high severity prostate cancer. This finding should be followed up in an independent Japanese or other Asian population sample where the lineage is common. If it is confirmed in additional populations, a systematic evaluation of the genetic changes in this lineage will be warranted. For example, the hypothesis that increased loss of the Y chromosome is the important difference, either because the absence of SRY leads to increased AR activity[6] or through some other mechanism, could be tested by comparing the loss rate of the Y in O3 lineage cells with that in other lineages. In addition, the availability of the near complete sequence of the euchromatic portion of the Y [16] reveals that it codes for only 27 distinct proteins, 18 of which are encoded by unique genes and the remaining nine by gene families with between two and ~20–40 copies. It would therefore be possible to compare the sequence and copy number of each of these genes in the O3 and other lineages. Identification of the basis for the increased susceptibility would be an important advance in our understanding of prostate cancer.

. . . .

Authors' affiliations

S Paracchini[1], C Tyler-Smith[2], Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

Celeste L Pearce, Brian E Henderson, Department of Preventive Medicine and Zilkha Neurogenetic Institute, University of Southern California Keck School of Medicine, Los Angeles, CA, USA

Laurence N Kolonel, Cancer Research Center, University of Hawaii, Honolulu, Hawaii, USA

David Altshuler, Departments of Genetics and Medicine, Harvard Medical School and Massachusetts General Hospital, Wellman 901, Boston, MA 02114, USA and Program in Medical and Population Genetics, Whitehead/MIT Center for Genome Research, One Kendall Square, Building 300, Cambridge, MA 02139, USA

Present addresses: [1]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; [2]The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

REFERENCES

1 **Brothman AR**, Maxwell TM, Cui J, Deubler DA, Zhu XL. Chromosomal clues to the development of prostate tumors. *Prostate* 1999;**38**:303–12.
2 **Jordan JJ**, Hanlon AL, Al-Saleem TI, Greenberg RE, Tricoli JV. Loss of the short arm of the Y chromosome in human prostate carcinoma. *Cancer Genet Cytogenet* 2001;**124**:122–6.
3 **Elfving P**, Cigudosa JC, Lundgren R, Limon J, Mandahl N, Kristoffersson U, Heim S, Mitelman F. Trisomy 7, trisomy 10, and loss of the Y chromosome in short-term cultures of normal kidney tissue. *Cytogenet Cell Genet* 1990;**53**:123–5.
4 **Aly MS**, Cin PD, Van de Voorde W, van Poppel H, Ameye F, Baert L, Van den Berghe H. Chromosome abnormalities in benign prostatic hyperplasia. *Genes Chromosomes Cancer* 1994;**9**:227–33.
5 **Lau YF**, Zhang J. Expression analysis of thirty one Y chromosome genes in human prostate cancer. *Mol Carcinog* 2000;**27**:308–21.
6 **Yuan X**, Lu ML, Li T, Balk SP. SRY interacts with and negatively regulates androgen receptor transcriptional activity. *J Biol Chem* 2001;**276**:46647–54.
7 **Jobling MA**, Tyler-Smith C. Fathers and sons: the Y chromosome and human evolution. *Trends Genet* 1995;**11**:449–56.
8 **Jobling MA**, Tyler-Smith C. The human Y chromosome: an evolutionary marker comes of age. *Nature Rev Genet* 2003;**4**:598–612.
9 **Underhill PA**, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ. Y chromosome sequence variation and the history of human populations. *Nat Genet* 2000;**26**:358–61.
10 **Jobling MA**, Tyler-Smith C. New uses for new haplotypes: the human Y chromosome, disease and selection. *Trends Genet* 2000;**16**:356–62.
11 **Kolonel LN**, Henderson BE, Hankin JH, Nomura AM, Wilkens LR, Pike MC, Stram DO, Monroe KR, Earle ME, Nagamine FS. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. *Am J Epidemiol* 2000;**151**:346–57.
12 **Paracchini S**, Arredi B, Chalk R, Tyler-Smith C. Hierarchical high-throughput SNP genotyping of the human Y chromosome using MALDI-TOF mass spectrometry. *Nucleic Acids Res* 2002;**30**:e27.
13 **Rousset F**, Raymond M. Testing heterozygote excess and deficiency. *Genetics* 1995;**140**:1413–19.

14 **Schneider S**, Kueffer JM, Excoffier L. Arlequin: A software for population genetics data analysis. Version 2.000. Geneva: Genetics and Biometry Lab, Dept of Anthropology, University of Geneva, 2000.

15 **Reich DE**, Goldstein DB. Detecting association in a case-control study while correcting for population stratification. *Genet Epidemiol* 2001;**20**:4–16.

16 **Skaletsky H**, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, Chinwalla A, Delehaunty A, Delehaunty K, Du H, Fewell G, Fulton L, Fulton R, Graves T, Hou SF, Latrielle P, Leonard S, Mardis E, Maupin R, McPherson J, Miner T, Nash W, Nguyen C, Ozersky P, Pepin K, Rock S, Rohlfing T, Scott K, Schultz B, Strong C, Tin-Wollam A, Yang SP, Waterston RH, Wilson RK, Rozen S, Page DC. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 2003;**423**:825–37.

17 **Y Chromosome Consortium**. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 2002;**12**:339–48.