Supplementary Table 1

DNA sequencing statistics.

Tumours were targeted at a depth of 50X and had a real median depth of 62X (range 52.6-92.3X). A median of 86% of the exome was covered at depths greater than 50X (65.8 - 95.2%). Morphologically normal tissues and blood controls were targeted to a sequence depth of 30X with a median real depth of 34X (range 30.7-38.8X). A median of 81% of the exome was covered at depths greater than 30X (74.2 – 86.5%). A median of 1.8% of the exome was not covered, a median of 95.3% of reads were unique and there was a median of 8.8% unmapped pairs.

							Percentage coverage at read depths						
ID	Target coverage	Gbp Seq	UM Pairs	Unique	Phys X	Seq X	1+	11+	21+	31+	41+	51+	101+
0006_Blood	30x	118.83	8.50	95.93	47.27	33.52	98.22%	97.17%	93.65%	82.15%	47.32%	11.97%	0.08%
0006_N	30x	122.93	10.23	97.25	57.20	34.46	98.25%	97.17%	93.33%	79.79%	59.13%	40.17%	0.91%
0006_T1	50x	231.87	10.71	94.96	98.49	62.34	98.08%	97.33%	96.66%	95.23%	92.30%	86.71%	17.69%
0006_T2	50x	225.77	9.39	94.52	96.11	62.01	98.29%	97.76%	97.25%	96.05%	93.14%	87.36%	15.11%
0006_T3	50x	223.25	8.60	95.37	100.70	62.55	98.09%	97.24%	96.44%	94.77%	91.59%	86.30%	8.10%
0006_T4	50x	206.91	9.14	95.42	87.73	57.53	98.33%	97.79%	97.21%	95.65%	92.01%	84.12%	11.09%
0007_Blood	30x	108.31	8.29	96.69	46.71	30.73	98.03%	96.69%	91.96%	74.65%	36.70%	8.08%	0.07%
0007_N	30x	109.56	8.98	97.62	48.58	31.24	98.30%	97.05%	91.91%	74.17%	49.60%	28.30%	0.16%
0007_T1	50x	224.97	10.01	94.59	98.44	61.14	98.17%	97.41%	96.56%	94.72%	91.38%	85.76%	4.60%
0007_T2	50x	189.37	8.68	96.10	85.08	53.34	98.23%	97.56%	96.78%	94.64%	89.33%	78.14%	8.80%
0007_T3	50x	228.58	8.27	94.68	99.13	63.75	98.30%	97.76%	97.28%	96.18%	93.78%	89.05%	19.61%
0007_T4	50x	230.28	8.64	95.03	106.71	64.09	98.25%	97.57%	97.06%	95.94%	93.53%	88.93%	16.41%
0007_T5	50x	210.12	8.78	95.20	88.30	58.48	98.30%	97.67%	97.09%	95.63%	92.45%	85.91%	6.16%
0008_Blood	30x	123.45	8.48	96.58	52.59	35.06	98.23%	97.23%	94.19%	83.79%	58.37%	25.85%	0.13%
0008_N	30x	137.11	9.36	97.17	61.12	38.81	98.32%	97.50%	95.19%	86.54%	69.41%	50.84%	1.81%
0008_T1	50x	221.14	8.75	87.77	87.62	56.90	98.09%	97.02%	95.06%	90.46%	80.09%	65.82%	0.26%
0008_T2	50x	203.19	8.22	87.95	74.38	52.57	98.22%	97.61%	96.78%	94.03%	90.06%	78.11%	0.26%
0008_T3	50x	340.00	9.46	93.42	145.33	92.27	98.47%	97.92%	97.60%	97.18%	96.46%	95.22%	66.91%

Supplementary Table 2

Clinical characteristics of prostate cancers at initial diagnosis

a Initial Diagnosis.

Case Ref	PSA at diagnosis	Clinical Stage at diagnosis	Pathological Stage	Gleason Score	Gleason Sum	Progression
Case 6	7	T1NxMx	T3aN0Mx	3+4	7	Alive and relapse free at 43 months
Case 7	10.1	T1NxMx	T3aN0Mx	3+4	7	Alive and relapse free at 42 months
Case 8	6.7	T1N0Mx	T3aNxMx	3+4	7	Alive and relapse free at 36 months

b Gleason of samples selected for DNA sequencing.

Sample	SangerID	Gleason
6_T1	PD7445a	3+4=7
6_T2	PD7445c	3+4=7
6_T3	PD7445d	3+4=7
6_T4	PD7445e	3+3=6
6_N	PD7445f	
6_Blood	PD7445b	
7_T1	PD7446a	3+3=6
7_T2	PD7446c	4+3=7
7_T3	PD7446d	3+3=6
7_T4	PD7446e	3+4=7
7_T5	PD7446f	3+4=7
7_N	PD7446g	
7_Blood	PD7446b	
8_T1	PD7447a	3+4=7
8_T2	PD7447c	4+3=7
8_T3	PD7447d	3+3=6
8_N	PD7447e	
8_Blood	PD7447b	

Supplementary Table 3

A list of potential prostate cancer driver genes.

A list of potential prostate cancer driver genes compiled from Grasso *et al.*[6], Garraway *et al.*[7] and the ICGC DCC 16 release. Mutations were classed as potential driver mutations if they were recurrent and made a coding change or occurred within a splice site. For the ICGC dataset the mutation had to have high functional impact and appear in three or more donors. This gave 5542 potential driver genes. Out of these 91 genes were affected by coding mutations or mutations occurring in the splice site in the complex men dataset and are shown in this table. Mutations in 44 genes occurred in two samples within a patient. The fact that there were no potential drivers found in more than two samples suggests that these drivers are likely to be late metastatic drivers. *DCC* was the only potential driver gene that occurred in more that one patient. No genes were mutated independently in more that one sample from the same patient i.e. convergent evolution was not found. Well known cancer genes such as *ATM*, *KIT*, and *PTEN* were mutated. A number of potential driver genes were observed in morphologically normal tissue: in 7_N we detected *BCAT1* (Garraway), *CHPF2* (Grasso), & *FAT2* (Grasso, & Garraway) and in 6_N we found *RYR3* (Grasso, Garraway and ICGC).

Samples	Gene	P.Description	Type	num samples	num hits	source
0008_T1, 0008_T2	*DCC*	p.V963I	missense	2	2	Garraway
0006_T1, 0006_T2	*DCC*	p.Y341H	missense	2	2	Garraway
0007_T1, 0007_T2	*ABCF3*	p.R269W	missense	2	1	Tomlins, Garraway
0007_T1, 0007_T2	*ADAMTS18*	p.R1014H	missense	2	1	Garraway
0006_T1, 0006_T2	*AGAP2*	p.?	splice	2	1	Tomlins, Garraway
0006_T1, 0006_T2	*ANKRD17*	p.T1972A	missense	2	1	Tomlins
0006_T1, 0006_T2	*ANKRD50*	p.R324fs*7	frameshift_variant	2	1	Tomlins, Garraway
0008_T1, 0008_T2	*ATG9A*	p.C122Y	missense	2	1	Tomlins
0008_T1, 0008_T2	*ATM*	p.L1439P	missense	2	1	Tomlins, Garraway, ICGC
0006_T1, 0006_T4	*CALCRL*	p.W399C	missense	2	1	Tomlins
0007_T4, 0007_T5	*CCDC105*	p.A292T	missense	2	1	Tomlins
0008_T1, 0008_T2	*CEACAM1*	p.E490K	missense	2	1	Tomlins
0006_T1, 0006_T4	*CHSY3*	p.R527C	missense	2	1	Tomlins
0006_T1, 0006_T4	*CNGA4*	p.R213C	missense	2	1	Tomlins, Garraway
0006_T1, 0006_T2	*EPB41L3*	p.A921T	missense	2	1	Garraway, ICGC
0006_T1, 0006_T4	*FBN2*	p.G721S	missense	2	1	Tomlins, Garraway
0008_T1, 0008_T2	*FLG*	p.R3907C	missense	2	1	Tomlins, Garraway
0006_T1, 0006_T4	*FLNB*	p.N1285S	missense	2	1	Tomlins, Garraway
0006_T1, 0006_T4	*G6PC*	p.E319K	missense	2	1	Garraway
0007_T1, 0007_T2	*HIC1*	p.P411L	missense	2	1	Garraway
0006_T1, 0006_T2	*HIST1H2BJ*	p.V112E	missense	2	1	Garraway
0006_T1, 0006_T4	*KCNK9*	p.A320T	missense	2	1	Tomlins
0007_T4, 0007_T5	*KCTD8*	p.R407H	missense	2	1	Tomlins, Garraway
0007_T1, 0007_T2	*KIT*	p.Q79K	missense	2	1	Tomlins
0006_T1, 0006_T2	*LCA5*	p.S32C	missense	2	1	Garraway
0006_T1, 0006_T4	*MEGF10*		SPLICE_REGION_VAR insertion	2	1	Tomlins
0008_T1, 0008_T2	*MIA3*	p.P1170S	missense	2	1	Garraway
0008_T1, 0008_T2	*MYH2*	p.R1755H	missense	2	1	Tomlins, Garraway
0008_T1, 0008_T2	*MYH7*	p.R1420Q	missense	2	1	Tomlins, Garraway
0006_T1, 0006_T4	*MYO1F*	p.L191V	missense	2	1	Tomlins
0008_T1, 0008_T2	*ODZ3*	p.Y2318D	missense	2	1	Tomlins, Garraway
0007_T1, 0007_T2	*OR5H6*	p.L71F	missense	2	1	Tomlins
0006_T1, 0006_T2	*PCDH11X*	p.R1188*	nonsense	2	1	Garraway
0008_T1, 0008_T2	*PHF10*	p.A71G	missense	2	1	Garraway
0007_T4, 0007_T5	*PIPOX*	p.I316T	missense	2	1	Garraway
0006_T1, 0006_T2	*ROS1*	p.T2045K	missense	2	1	Tomlins, Garraway
0006_T1, 0006_T2	*RPGRIP1*	p.G917R	missense	2	1	Tomlins
0007_T1, 0007_T2	*SF3B1*	p.K700E	missense	2	1	Tomlins, Garraway
0006_T1, 0006_T4	*SKIV2L2*	p.G930fs*30	frameshift_variant	2	1	Tomlins
0008_T1, 0008_T2	*SMCHD1*	p.G68D	missense	2	1	Garraway

0006_T1, 0006_T2	*SORBS1*	p.S1230L	missense	2	1	Garraway
0006_T1, 0006_T2	*TNNT3*	p.R99H	missense	2	1	Garraway
0007_T1, 0007_T2	*UBR4*	p.R450G	missense	2	1	Garraway
0006_T1, 0006_T4	*ZC3H13*	p.E754fs*28	frameshift_variant	2	1	Tomlins, Garraway
0007_T1, 0007_T2	*ZNF208*	p.T1001S	missense	2	1	Garraway
0008_T1	*AASDH*	p.E182K	missense	1	1	Garraway
0007_T2	*ABI3BP*	p.R717*	nonsense	1	1	Tomlins, Garraway
0006_T3	*ADAP2*	p.M77I	missense	1	1	Tomlins
0006_T1	*ASTN1*	p.?	splice	1	1	Tomlins, Garraway
0007_N	*BCAT1*	p.L276M	missense	1	1	Garraway
0006_T3	*CEP110*	p.R1431C	missense	1	1	Tomlins
0006_T2	*CHD5*	p.R471Q	missense	1	1	Garraway
0007_N	*CHPF2*	p.R470L	missense	1	1	Tomlins
0006_T3	*DCAF8L1*	p.R152Q	missense	1	1	Tomlins
0006_T3	*DGKG*	p.D634N	missense	1	1	Tomlins
0006_T1	*DNPEP*	p.E215V	missense	1	1	Garraway
0007_T1	*FAM135B*	p.A300T	missense	1	1	Tomlins, Garraway
0007_N	*FAT2*	p.S4308T	missense	1	1	Tomlins, Garraway
0007_T5	*FREM2*	p.V1477A	missense	1	1	Tomlins, Garraway
0006_T3	*GALNT13*	p.W128*	nonsense	1	1	Garraway
0006_T2	*GBP7*	p.D97G	missense	1	1	Tomlins
0006_T3	*GLUD1*	p.G72R	missense	1	1	Garraway
0007_T5	*KCNJ4*	p.T131M	missense	1	1	Tomlins
0006_T4	*KIF19*	p.?	essential splice	1	1	Garraway
0006_T2	*KIF1A*	p.M1484V	missense	1	1	Tomlins, Garraway
0007_T5	*KIF2B*	p.R36C	missense	1	1	Tomlins
0006_T4	*KLHL11*	p.H482N	missense	1	1	Tomlins
0006_T3	*KPNA7*	p.A440E	missense	1	1	Tomlins
0006_T4	*LRP4*	p.P946A	missense	1	1	Tomlins, Garraway, ICGC
0006_T2	*MEIS2*	p.R131C	missense	1	1	Tomlins
0006_T2	*MYC*	p.P177R	missense	1	1	Tomlins
0006_T2	*OGDH*	p.H670N	missense	1	1	Tomlins
0006_T2	*PCDHA2*	p.R47C	missense	1	1	Tomlins, Garraway
0006_T4	*PCDHA3*	p.R65W	missense	1	1	Garraway
0006_T4	*PCDHB11*	p.Y279F	missense	1	1	Garraway
0007_T5	*PCSK2*	p.G366S	missense	1	1	Tomlins
0006_T1	*PDS5A*	p.?	splice	1	1	Garraway
0006_T4	*PLEC*	p.V2825E	missense	1	1	Tomlins, Garraway
0006_T1	*PTEN*	p.I253N	missense	1	1	Tomlins, Garraway, ICGC
0006_T1	*RP1*	p.S440L	missense	1	1	Tomlins, Garraway
0006_T1	*RYR2*	p.V3597A	missense	1	1	Tomlins, Garraway, ICGC
0006_N	*RYR3*	p.A525S	missense	1	1	Tomlins, Garraway, ICGC
0007_T2	*SDHA*	p.M1V	missense	1	1	Garraway
0006_T2	*SEMA3D*	p.R294H	missense	1	1	Tomlins, Garraway
0006_T1	*SETX*	p.I2150T	missense	1	1	Tomlins, Garraway
0007_T3	*SH3RF2*	p.R286C	missense	1	1	Tomlins
0006_T1	*SLC13A3*	p.M510I	missense	1	1	Tomlins
0006_T2	*SLC22A2*	p.E527K	missense	1	1	Tomlins
0006_T3	*SPAG17*	p.L476I	missense	1	1	Tomlins, Garraway
0007_T3	*STAB1*	p.I1704T	missense	1	1	Tomlins, Garraway
0007_T2	*TEKT3*	p.R437C	missense	1	1	Garraway
0006_T4	*ZNF236*	p.F976L	missense	1	1	Tomlins, Garraway

ICGC PROSTATE GROUP Membership (in addition to those named in the author list).

Steve Bova
Freddie Hamdy
Yong-Jie Lu
Anthony Ng
Yongwei Yu
Hongwei Zhang

DISCUSSION OF SOMATIC MOSAICISM

The recent publication by Holstege *et al*[1] raises the possibility that the mutation that we observed in morphologically normal prostate tissue may arise through somatic mosaicism. Here we show that the processes occurring in morphologically normal prostate and those reported by Holstege *et al*[1] are distinct.

There are a number of considerations:

1. First, the rate of mutation in human cells is thought to be around 1-2.5 mutation per cell division[2,3]. At this rate of mutation around 200-500 cell divisions would have had to occur in the single progenitor cell that gives rise to the mutated clones of cells that we observe in morphologically normal prostate (we observed 518 mutations in morphologically normal tissue from Case 6 and 454 mutations in Case 7).
2. The samples of DNA that we prepare from morphologically normal prostate are around 20-30 micrograms, a portion of which (1-2 micrograms) was subject to DNA sequencing. This means that our DNA samples are prepared from a minimum of about 4,000,000 cells (6pg of DNA per cell, assuming 100% DNA yield). To generate this tissue sample a single somatic prostate cell containing 500 mutations would have to undergo a minimum of 20-24 additional doublings: possibly many more since it would have to grow out against competition from other cells in the prostate.
3. In our manuscript we argued that selection would be involved in generating the clone of morphologically normal cells containing high mutational burden, but it is theoretically possible that the clone could arise by somatic mosaicism. However, even if somatic mosaicism is involved, the overall process would still have to be accompanied by a high rate of cell division and/or high (per cell division) rate of mutation. In a model involving somatic mosaicism the clone of cells could then arise without selection through genetic drift (or from an origin in prostate stem cells) only in the context of high rates of cell division, a property that is documented to be absent in morphologically normal prostate tissue[4], and/or high mutation rate. There is no evidence to support either of these possibilities in normal prostate development. Our work highlights the presence of high mutation rates in morphologically normal prostate tissue for the first time and will prompt future studies to provide clearer insights into the mechanisms and the effects on pathogenesis.
4. We have compared our findings with data obtained by Holstege *et al*[1] who examined the total white blood cell DNA from a 115-year old woman: in contrast to morphologically normal prostate it is well documented that hematopoietic cells have a high rate of cell turnover[5]. They found evidence for somatic mosaicism with the blood sample containing approximately 424 somatic mutations. By comparison no verifiable mutations were detected in similarly analysed normal brain tissue. Hematopoietic stem cells are thought to renew once or twice per year giving rise to

multi-potent progenitor that through hematopoiesis yields diverse blood cell types[5]. A rate of 2.5 accumulated somatic mutations per cell division and 200 doublings (~2 per year) would account for the figure of 424 mutations.

5. Critically in the study presented by Holstege *et al*[1] there was a high level of attrition of telomeres in the white blood cells, compared to intermediate length of telomeres in most other tissue, and long telomeres in non-dividing tissue (brain), consistent with the differences in cell turnover in these tissues. When we examined telomere length using the TelSeq tool[6] we found telomere lengths of 6.3kb in morphologically normal tissue from Patient 6, and 6.2kb in Patient 7. The telomeres in the corresponding cancers were slightly longer that in morphologically normal tissue in Patient 6 and the same in Patient 7. We concluded that the somatic mosaicism observed in white blood cells from the study of Holstege *et al*[1] was distinct from the phenomena that we were observing in the morphologically normal prostate; based on the absence of high levels of telomere attrition, and on the higher mutation rate observed in prostate, a tissue believed to be relatively quiescent. Also men in our study are younger than the patient presented by Holstege *et al*) (59 and 71 for cases 7 and 6 respectively, compared to 115).

References

1. Holstege, H. *et al.* Somatic mutations found in the healthy blood compartment of a 115-yr-old woman demonstrate oligoclonal hematopoiesis. *Genome Res.* **24,** 733–742 (2014).
2. Zong, C., Lu, S., Chapman, A. R. & Xie, X. S. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* **338,** 1622–1626 (2012).
3. Nussbaum, R., McInnes, R. R. & Willard, H. F. Thompson & Thompson Genetics in Medicine - Robert Nussbaum, Roderick R. McInnes, Huntington F Willard - Google Books. (2007).
4. Mucci, N. R. *et al.* Expression of nuclear antigen Ki-67 in prostate cancer needle biopsy and radical prostatectomy specimens. *J. Natl. Cancer Inst.* **92,** 1941–1942 (2000).
5. Catlin, S. N., Busque, L., Gale, R. E., Guttorp, P. & Abkowitz, J. L. The replication rate of human hematopoietic stem cells in vivo. *Blood* **117,** 4460–4466 (2011).
6. Ding, Z. *et al.* Estimating telomere length from whole genome sequence data. *Nucleic Acids Res.* **42,** e75 (2014).

REFERENCES FOR SUPPLEMENTARY FIGURES AND TABLES

1. Clark, J. *et al.* Complex patterns of ETS gene alteration arise during cancer development in the human prostate. *Oncogene* **27,** 1993–2003 (2008).
2. Nik-Zainal, S. *et al.* The life history of 21 breast cancers. *Cell* **149,** 994–1007 (2012)
3. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149,** 979–993 (2012).
4. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep* **3,** 246–259 (2013).
5. Baca, S. C. *et al.* Punctuated evolution of prostate cancer genomes. *Cell* **153,** 666–6 77 (2013).
6. Grasso, C. S. *et al.* The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487,** 239–243 (2012).
7. Barbieri, C. E. *et al.* Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat. Genet.* **44,** 685–689 (2012).