nature genetics

Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue

Colin S Cooper[1–3,27,28], Rosalind Eeles[1,4,27,28], David C Wedge[5,27], Peter Van Loo[5–7,27], Gunes Gundem[5], Ludmil B Alexandrov[5], Barbara Kremeyer[5], Adam Butler[5], Andrew G Lynch[8], Niedzica Camacho[1], Charlie E Massie[9], Jonathan Kay[9], Hayley J Luxton[9], Sandra Edwards[1], Zsofia Kote-Jarai[1], Nening Dennis[4], Sue Merson[1], Daniel Leongamornlert[1], Jorge Zamora[5], Cathy Corbishley[10], Sarah Thomas[4], Serena Nik-Zainal[5], Manasa Ramakrishna[5], Sarah O'Meara[5], Lucy Matthews[1], Jeremy Clark[3], Rachel Hurst[3], Richard Mithen[11], Robert G Bristow[12–14], Paul C Boutros[12,15,16], Michael Fraser[13,14], Susanna Cooke[5], Keiran Raine[5], David Jones[5], Andrew Menzies[5], Lucy Stebbings[5], Jon Hinton[5], Jon Teague[5], Stuart McLaren[5], Laura Mudie[5], Claire Hardy[5], Elizabeth Anderson[5], Olivia Joseph[5], Victoria Goody[5], Ben Robinson[5], Mark Maddison[5], Stephen Gamble[5], Christopher Greenman[17], Dan Berney[18], Steven Hazell[4], Naomi Livni[4], the ICGC Prostate Group[19], Cyril Fisher[4], Christopher Ogden[4], Pardeep Kumar[4], Alan Thompson[4], Christopher Woodhouse[4], David Nicol[4], Erik Mayer[4], Tim Dudderidge[4], Nimish C Shah[9], Vincent Gnanapragasam[9], Thierry Voet[20], Peter Campbell[5], Andrew Futreal[5], Douglas Easton[21], Anne Y Warren[22,27], Christopher S Foster[23,24,27,28], Michael R Stratton[5], Hayley C Whitaker[9,27], Ultan McDermott[5,27,28], Daniel S Brewer[1,3,25,27,28] & David E Neal[9,26–28]

Genome-wide DNA sequencing was used to decrypt the phylogeny of multiple samples from distinct areas of cancer and morphologically normal tissue taken from the prostates of three men. Mutations were present at high levels in morphologically normal tissue distant from the cancer, reflecting clonal expansions, and the underlying mutational processes at work in morphologically normal tissue were also at work in cancer. Our observations demonstrate the existence of ongoing abnormal mutational processes, consistent with field effects, underlying carcinogenesis. This mechanism gives rise to extensive branching evolution and cancer clone mixing, as exemplified by the coexistence of multiple cancer lineages harboring distinct *ERG* fusions within a single cancer nodule. Subsets of mutations were shared either by morphologically normal and malignant tissues or between different *ERG* lineages, indicating earlier or separate clonal cell expansions. Our observations inform on the origin of multifocal disease and have implications for prostate cancer therapy in individual cases.

Prostate cancer is commonly multifocal[1], although the origin of multifocal disease remains controversial. Analyses of patterns of allele loss have suggested the independence of most individual foci[2,3]. However, such studies cannot exclude the presence of common underlying mutations not detected by the methods used. Recent attempts to unravel the origins of multifocal disease using high-resolution genome technologies have also led to conflicting data, with different authors concluding either that all foci in a single prostate are related[4] or that all foci are unrelated[5]. To gain further insights into the mechanism of prostate cancer development—particularly the origin of multifocal disease—we selected three representative prostate cancers (**Fig. 1** and **Supplementary Fig. 1**) that had been mapped for *ERG* status using the break-apart FISH method[6,7]. Twelve cancer samples and three samples designated as morphologically normal prostate on the basis of central pathology review were analyzed using paired-end, massively parallel DNA sequencing of complete genomes to generate comprehensive catalogs of genetic alterations. (For coverage statistics, see **Supplementary Table 1**. For 3D representations of each prostate and clinical characteristics, see, respectively, **Supplementary Fig. 2** and **Supplementary Table 2**.) Prostates were named according to their Cancer Research UK project designations: cases 6, 7 and 8.

Somatic mutations, not present in cancer and blood samples, were observed at notable levels in morphologically normal prostate tissue distant from cancer in case 6 (518 substitutions) and case 7 (454 substitutions) (**Supplementary Fig. 3**). Some of these mutations might have potential functional significance (**Table 1**). The presence of substitution mutations in morphologically normal prostate tissue was confirmed in validation DNA sequencing experiments to an

A full list of author affiliations appears at the end of the paper.

Received 29 September 2014; accepted 21 January 2015; published online 2 March 2015; corrected after print 5 May 2015; doi:10.1038/ng.3221

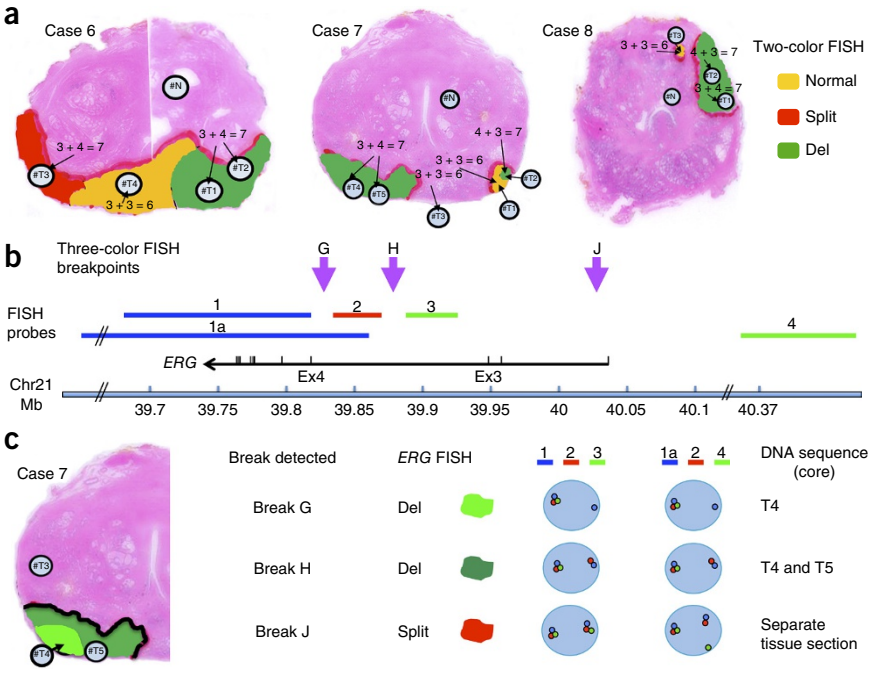

Figure 1 Prostate samples chosen for genome-wide sequencing. (**a**) *ERG* rearrangements determined by FISH. Case 7 was a multifocal cancer containing two separate foci (T1/T2/T4/T5 and T3). Case 8 was also designated as a multifocal cancer (with nodules T1/T2 and T3). Yellow, unrearranged normal *ERG* gene; red, *ERG* gene split but both 3′ and 5′ ends retained; green, *ERG* gene rearranged but only its 3′ end retained (Del). (**b**,**c**) Three-color FISH used to distinguish different *ERG*-locus translocation breakpoints in case 7. (**b**) Position of the three FISH probes: probe 1 (blue: 1, BAC RP11-164E1, and 1a, BACs RP11-95G19, RP11-720N21 and CTD-2511E13) was labeled in Aqua (Kreatech 415 Platinum Bright), probe 2 (red; fosmid G248P80319F5, 37 kb) was labeled with Cy3 and probe 3 (green: 3, fosmid G248P86592E2, 38.5 kb; and 4, BACs RP11-372O17, RP11-115E14 and RP11-729O4) was labeled with fluorescein isothiocyanate. Ex, exon; Chr, chromosome. The purple arrows represent the positions of *ERG* breakpoints detected in these experiments. For the precise positions of *ERG* breakpoints G and H, see **Table 2**. (**c**) Left, tumor areas with *ERG* locus breaks G and H are indicated in light and dark green, respectively. Break J was found in an adjacent prostate section not shown in this figure. Right, representations of the *ERG* FISH patterns. Original FISH images are show in **Supplementary Figure 1**. "Split" denotes that 5′ and 3′ *ERG* signals were separated but retained in the cell.

average read depth of 10,000. Substitutions were present in an estimated ~48% and ~42% of cells in morphologically normal samples from case 6 and case 7, respectively (**Supplementary Fig. 3b**), demonstrating clonal expansions of cells within morphologically normal prostate tissue, in agreement with studies using mitochondrially encoded enzyme cytochrome *c* oxidase as a marker[8].

Aiming to understand the subclonal architecture and phylogeny of the tumors, we initially constructed phylogenetic trees on the basis of copy numbers (**Supplementary Figs. 4** and **5** and **Supplementary Data Set 1**) and substitution data. We adapted our previously developed Bayesian Dirichlet process to identify clusters of substitutions in *n* dimensions[9], where *n* was the number of samples from a given case, such that shared and distinct subclones could be identified between related samples (**Fig. 2** and **Supplementary Fig. 6**). To further explore the fine details and confirm the main features of the phylogeny tree and clonal structure, we sequenced a selection of substitutions from each potential relationship between samples to an average read depth of 10,000 in independent DNA sequencing analyses, which verified 279 mutations across all samples. This provided us with our final integrated phylogenetic trees (**Fig. 2a–c**) and final list of somatic point mutations (**Supplementary Data Set 2**). The structure of these trees was also supported by verified insertions, deletions and breakpoints (**Supplementary Data Sets 3** and **4**). The single cancer mass from case 6 contained three independent cancer clones represented by samples 6_T2, 6_T3 and 6_T4 (**Fig. 2a**), with a single verified substitution linking 6_T1/6_T2 and 6_T3. Case 7 contained at least three independent cancer lineages: one (7_T3) representing the smaller cancer nodule, and two (7_T1/7_T2 and 7_T4/7_T5) present in the larger cancer mass (**Fig. 2b**). Ten mutations were common to the morphologically normal prostate sample and to cancer samples 7_T1 and 7_T2, and three mutations joined 7_T4/7_T5 to the separate multifocal lesion 7_T3. These observations show that case 7 contained at least two clones of cells that existed before the formation of the distinct cancer lineages. Case 8 contained two cancer lineages represented by 8_T1/8_T2 and 8_T3 (**Fig. 2c**), with 43 substitutions shared between the three tumor samples 8_T1, 8_T2 and 8_T3, 8 of which

Table 1 Mutations and clonal expansions in morphologically normal tissue

Sample	Description	Gene	Protein description	Type	Reads (%)	Total number of reads	MA predicted functional impact	ANNOVAR significant algorithms
0006#N	Chr9:g.131115799G>A	*SLC27A4*	p.V435I	Missense	13.79	58	Low	1
0006#N	Chr14:g.20389481C>T	*OR4K5*	p.T239M	Missense	13.25	83	High	4
0006#N	Chr15:g.33873844G>T	*RYR3*	p.A525S	Missense	33.33	48	Medium	
0006#N	Chr4:g.88766379C>G	*MEPE*	p.S120*	Nonsense	20.83	24	NA	2
0007#N	Chr5:g.150885254A>T	*FAT2*	p.S4308T	Missense	23.4	47	Low	5
0007#N	Chr7:g.150934857G>T	*CHPF2*	p.R470L	Missense	17.24	58	Medium	5
0007#N	Chr8:g.24192995G>A	*ADAM28*	p.D470N	Missense	17.78	45	Neutral	2
0007#N	Chr12:g.24989522G>T	*BCAT1*	p.L276M	Missense	26.47	34	Medium	

Point mutations present in exons with an indication of functional significance. Missense and nonsense mutations detected and visually confirmed in adjacent morphologically normal tissue were tested for functional impact using MutationAssessor.org[27] (MA) and wANNOVAR[28] services. *OR4K5* was excluded as a candidate because of the potential to overcall mutations in genes encoding very large proteins[29]. As none of the mutations had a high MA value, we considered that epigenetic changes might be a more likely driver of clonal expansion. NA, not applicable.

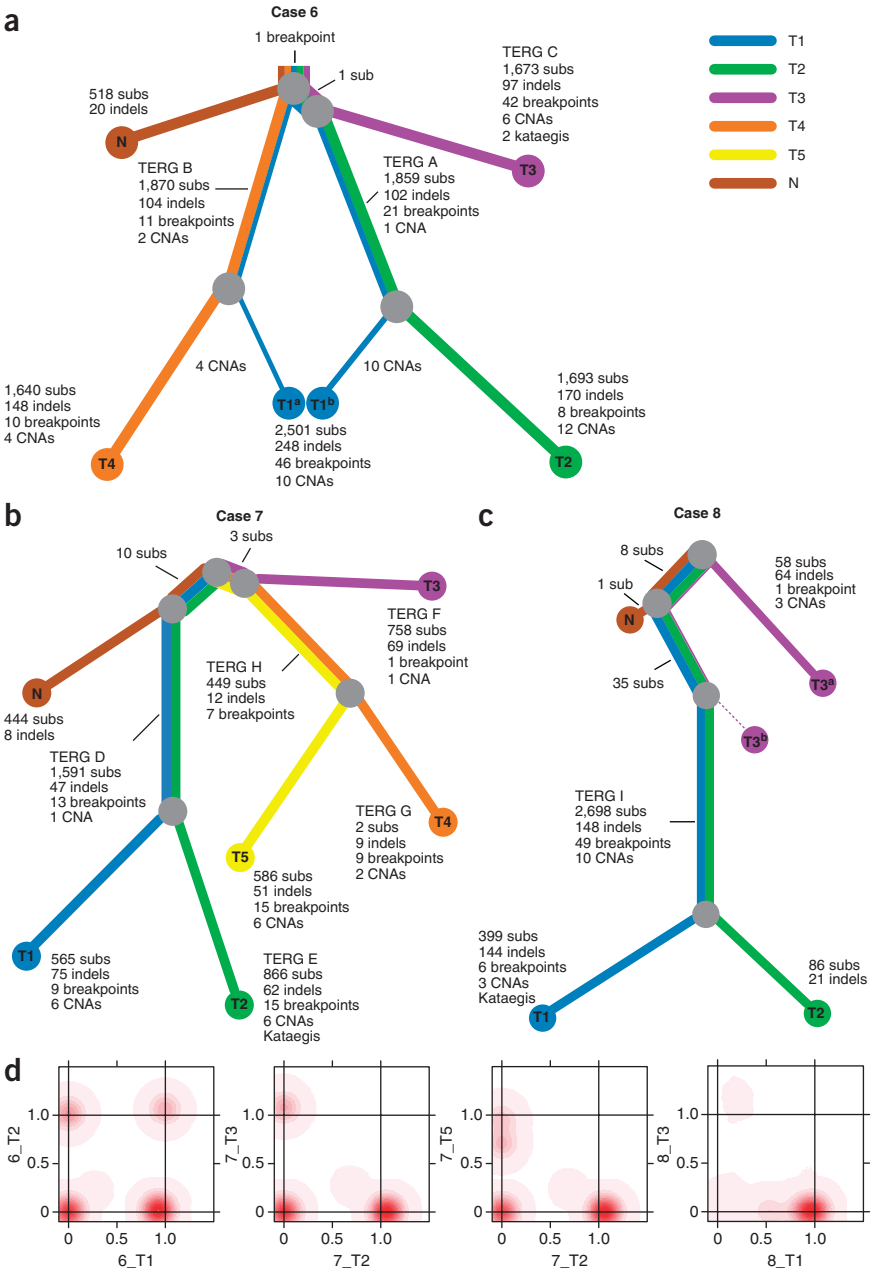

Figure 2 Phylogenies of multifocal prostate cancers. (**a**–**c**) Phylogenies revealing the relationships between sample clones for each case. Each line is associated with a clone from a particular sample. The length of each line is proportional to the weighted quantity of variations on a logarithmic scale. The thickness of a line indicates the proportion of the sample made up of that clone (48% and 52% for 6_T1a and 6T1b, respectively; 88% and 12% for 8_T3a and 8_T3b, respectively). The minor clone of 8_T3b had no detected unique variants. 8_T3 contained 43 mutations present as a 12% subclone (T3a) shared with 8_T1/8_T2. In validation experiments 8_T3 did not contain any of the five *ERG* and *TMPRSS2* rearrangements present in 8_T1/8_T2 (**Table 2**) or mutations that were unique to 8_T1/8_T2 (10,000 depth), which indicated that it represents an earlier clone of 8_T1/8_T2 seeded into tissue sample 8_T3. The various *TMPRSS2-ERG* translocations are indicated by their TERG I.D.s (**Table 2**). Sub, substitution; indel, insertion/deletion; CNA, copy-number alteration. (**d**) Example 2D density plots showing the posterior distribution of the fraction of cells bearing a mutation in two samples. The fraction of cells was modeled using a Bayesian Dirichlet process. These plots illustrate samples that had shared clonal mutations (6_T1/6_T2) and branched (unrelated) mutations (7_T2/T_T3). There are two examples of samples with a subclone: 7_T2/7_T5 had a peak at (0,0.72) that represented subclonal mutations in 72% of cells in 7_T5 that occurred only in this sample, after divergence from the other samples. Similarly, 8_T1/8_T3 had a peak at (0.54,0) representing subclonal mutations in 54% of cells in T1 only.

were also present in distant, morphologically normal sample 8_N.

Complex patterns of *ERG* alteration were observed in samples from case 6 and case 7 (**Fig. 3**); each main lineage contained at least one and in some cases two unique *TMPRSS2-ERG* fusions with distinct breakpoint locations in *TMPRSS2* and *ERG* (**Fig. 2** and **Table 2**). The presence of multiple distinct *TMPRSS2-ERG* fusions was demonstrated by direct PCR across the breakpoint and by an *ERG* break-apart FISH assay (**Table 2**, **Fig. 1b**,**c** and **Supplementary Fig. 1**). In this respect *TMPRSS-ERG* fusions could be considered similar to the convergent gene alterations observed in kidney cancer, where distinct alterations of genes such as *SETD2*, *PTEN* and *KDM5C* have been observed in different parts of the same cancer[10]. A deletion on chromosome 8 exhibited a very similar pattern of alterations (**Supplementary Fig. 7**), but we did not see convergent evolution for other potential driver genes (**Supplementary Table 3**). Where two *TMPRSS2-ERG* fusions existed in a single lineage, we were unable to determine whether these fusions coexisted at any time in the same cell as reported previously[11] and as implied by the phylogenic tree. However, the FISH assay (**Fig. 1b**,**c**) demonstrated that in sample 7_T4 the two *TMPRSS2-ERG* fusions were present in distinct cell populations when the cancer sample was taken. Moreover, an additional, separate *ERG* breakpoint was detected in a region of the cancer that had not been sampled in the DNA sequencing studies (TERG J). The occurrence of several *TMPRSS2-ERG* fusions in a single cancer mass was consistent with previous FISH-based studies reporting multiple E26 transformation–specific fusions in a low proportion of individual cancer foci[11]. *ERG* alterations are believed to represent a relatively early event in cancer development, in agreement with their occurrence in prostatic intraepithelial neoplasia[6], but our observations suggest that they might not always be present at the very first cellular expansion. Mutations shared either between different *ERG* lineages or between cancer and morphologically normal tissue might represent earlier clonal-cell expansions on the same lineage (**Fig. 2a**–**c**). Alternatively, they could represent separate clones of cells within which multiple independent cancer lineages developed.

Recently, we identified 21 distinct mutational signatures from 7,042 samples across 30 different cancer types[12]. The contribution of mutational processes was calculated for prostate cancer as previously

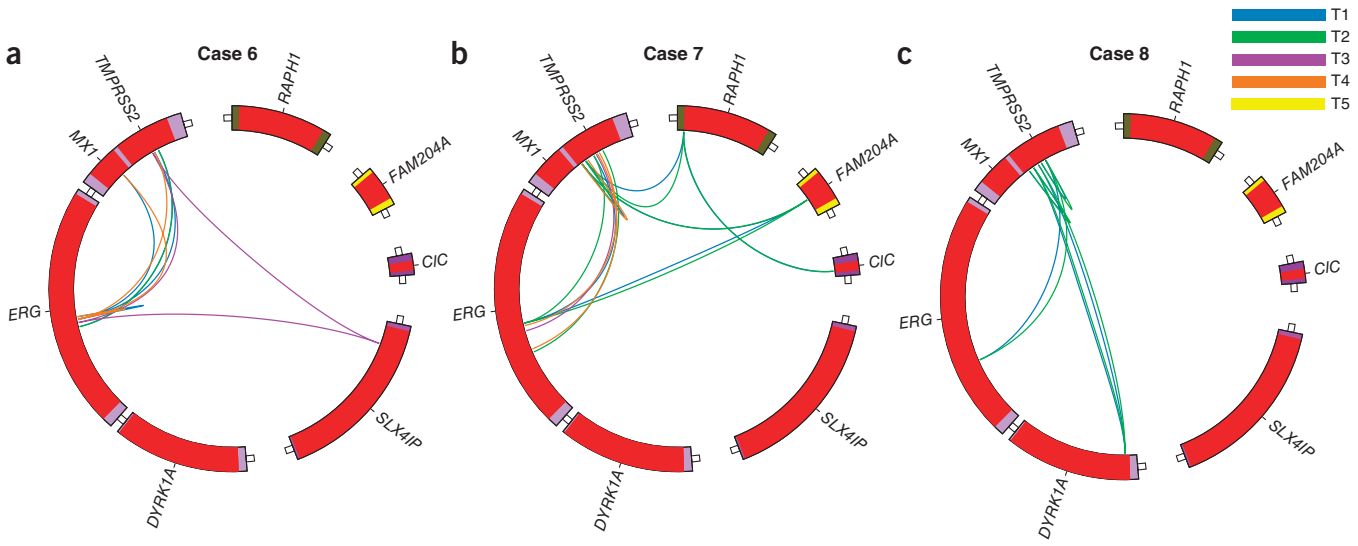

Figure 3 Patterns of *ERG* alterations. (**a**–**c**) Circos plots highlighting *ERG* rearrangements present in cases 6 (**a**), 7 (**b**) and 8 (**c**). Each color represents a different cancer sample (see legend).

described[12,13] (**Fig. 4**). A signature (designated signature 1A in ref. 12) associated with spontaneous deamination of 5-methylcytosine at CpG sequences explained ~50% of our mutations. Two additional signatures with unknown etiology, designated signature 5 and signature 8, best explained the remaining somatic mutations. Signature 5, present in all prostate samples, may reflect an endogenous mutational process[12]. Signature 8, present in two cancer samples from a single cancer nodule, is characterized by weak C>A strand bias. Critically, these observations show that the same mutational processes, giving rise to signatures 1a and 5, are detected both in cancer and in matched morphologically normal prostate tissue. We identified clustering of C>T and C>G mutations, referred to as kataegis[14], and complex, interdependent translocations and deletions called chromoplexy[15] in some cancer lineages (**Supplementary Figs. 8** and **9**).

Next-generation sequencing technologies have been used to identify critical genetic processes in prostate cancer development[15–19].

Table 2 Patterns of *ERG* alterations

	Donor			Middle		Acceptor						TERG
Samples	Chr	Position	Strand	Type	Sequence	Chr	Position	Strand	Breakpoint	Genes	Verification	I.D.
6_T1, 6_T2	21	39867180	+	Homology	T	21	42877104	+	Deletion	*ERG-TMPRSS2*	CS and P (6_T1); V (6_T1, 6_T2)	A
6_T1, 6_T4	21	39877208	+	Homology	T	21	42871170	+	Deletion	*ERG-TMPRSS2*	P (6_T1); V (6_T1, 6_T4)	B
6_T1, 6_T4	21	39877355	–	Homology	CC	21	42819405	–	Insertion	*ERG-MX1*	CS and P (6_T1); V (6_T1, 6_T4)	
6_T1, 6_T4	21	39877745	+	NTS	CAT	21	39880855	+	Deletion	*ERG-ERG*	CS and P (6_T1); V (6_T1, 6_T4)	
6_T3	20	10441211	–	Homology	G	21	39872887	+	Translocation	*SLX4IP-ERG*	CS, P and V (6_T3)	
6_T3	20	10441429	+	Homology	GT	21	42868518	–	Translocation	*SLX4IP-TMPRSS2*	CS, P and V (6_T3)	
6_T3	21	39872930	+	Exact	—	21	42868510	+	Deletion	*ERG-TMPRSS2*	CS, P and V (6_T3)	C
7_T1, 7_T2	1	205613440	+	Homology	C	21	42857784	–	Translocation	*_-TMPRSS2*	V (7_T1, 7_T2)	
7_T1, 7_T2	2	204298424	–	Homology	A	21	42849002	+	Translocation	*RAPH1-TMPRSS2*	V (7_T1, 7_T2)	
7_T1, 7_T2	2	204298476	+	Exact	—	19	42797705	+	Translocation	*RAPH1-CIC*	P (7_T1); V (7_T1, 7_T2)	
7_T1, 7_T2	10	120084722	–	Homology	TG	21	42842154	+	Translocation	*FAM204A-TMPRSS2*	CS and P (7_T1); V (7_T1, 7_T2)	
7_T1, 7_T2	10	120084747	+	Homology	AC	21	39872234	+	Translocation	*FAM204A-ERG*	CS and P (7_T2); V (7_T1, 7_T2)	
7_T1, 7_T2	21	39872152	+	Homology	A	21	42861527	+	Deletion	*ERG-TMPRSS2*	CS and P (7_T1); V (7_T1, 7_T2)	D
7_T1, 7_T2	21	42842403	+	Exact	—	21	42848506	–	Inversion_+	*TMPRSS2-TMPRSS2*	CS and P (7_T1); V (7_T1, 7_T2)	
7_T2	21	39831266	+	Homology	AAAC	21	42875633	+	Deletion	*ERG-TMPRSS2*	CS, P and V (7_T2)	E
7_T3	21	39861568	+	NTS	TA	21	42865303	+	Deletion	*ERG-TMPRSS2*	CS, P and V (7_T3)	F
7_T4	21	39835734	+	Homology	G	21	42867100	+	Deletion	*ERG-TMPRSS2*	CS, P and V (7_T4)	G
7_T4	21	42841552	–	Homology	GGCT	21	42851963	+	Inversion_–	*TMPRSS2-TMPRSS2*	CS, P and V (7_T4)	
7_T4, 7_T5	21	39868722	+	Exact	—	21	42870051	+	Deletion	*ERG-TMPRSS2*	CS and P (7_T4); V (7_T4, 7_T5)	H
8_T1, 8_T2	21	38745261	+	Homology	T	21	42851601	–	Inversion_+	*DYRK1A-TMPRSS2*	P (8_T1); V (8_T1, 8_T2)	
8_T1, 8_T2	21	38745286	–	Homology	A	21	42859198	–	Insertion	*DYRK1A-TMPRSS2*	CS and P (8_T1); V (8_T1, 8_T2)	
8_T1, 8_T2	21	39831518	+	Exact	—	21	42870497	–	Inversion_+	*ERG-TMPRSS2*	CS (8_T1); P and V (8_T1, 8_T2)	I
8_T1, 8_T2	21	42844460	–	Homology	T	21	42851648	+	Inversion_–	*TMPRSS2-TMPRSS2*	V (8_T1, 8_T2)	
8_T1, 8_T2	21	42863787	–	Homology	G	21	42870663	+	Inversion_–	*TMPRSS2-TMPRSS2*	CS and P (8_T1); V (8_T1, 8_T2)	

Position and structure of each *ERG* breakpoint and related rearrangements. The position and structure of the breakpoint were determined, in the majority of cases, by capillary sequencing using custom-designed PCR across the rearrangement breakpoint as previously described[30] ("CS" in "Verification" column) and/or by *in silico* reconstruction using local *de novo* assembly in Brass phase 2. Verification by sizing PCR products across the breakpoint using gel electrophoresis was also done ("P"). All breakpoints were visually verified ("V") to ensure the presence of discordant reads and checked to make sure that they did not occur in repeat regions. Chr, chromosome.

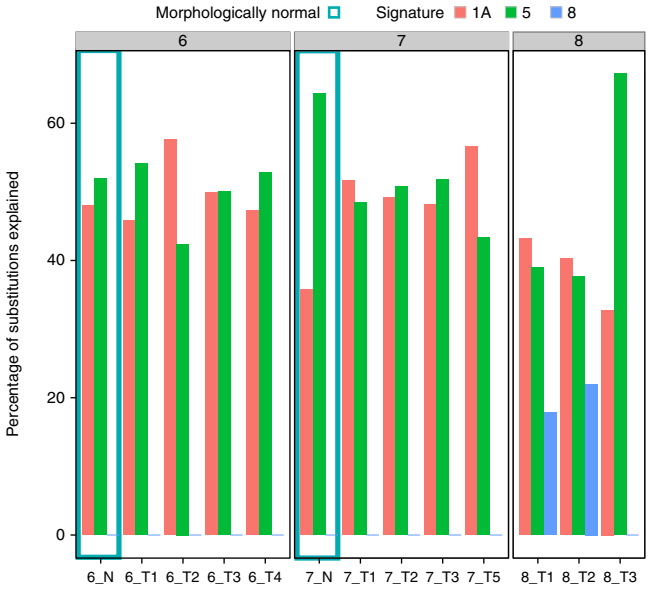

Figure 4 Relative contributions of mutational signatures to the total mutation burden of each sample. The mutational spectra of each sample, as defined by the triplets of nucleotides around each substitution, were deconvoluted into mutational processes using 22 distinct signatures determined from 7,042 cancers as described previously[12,13]. The signature designations (1a, 5 and 8) match those reported previously[12]. For samples 7_T4 and 8_N, there were too few mutations for the contributions of the mutational signatures to be identified accurately.

Our results demonstrate the presence of clonal expansions or fields of cells in morphologically normal prostate that provide a background against which prostate cancer develops. A recent study on a 115-year-old woman identified 424 point mutations, thought to result from somatic mosaicism, in the rapidly dividing tissue blood, but no mutations were detected in brain tissue[20]. The presence of mutations in blood was accompanied by telomere attrition that was not observed in other tissues. Prostate is considered a relatively quiescent tissue[21], and we found that the telomeres in morphologically normal tissue from cases 6 and 7 had not undergone attrition, being of comparable length to telomeres in adjacent cancer. The processes at work in morphologically normal prostate therefore appear to be distinct from those reported for blood (see the **Supplementary Note** for a full discussion). Whether the clones of cells observed in morphologically normal prostate are generated by a pathological process or are the product of somatic mosaicism involving unexpectedly high mutation rates, the resulting clonal fields of cells may influence cancer development and/or contribute to multifocality and the presence of multiple cancer lineages in a single cancer mass. Evidence for a field effect in prostate cancer is also supported by studies demonstrating tumor-like alterations in cytomorphology, gene expression and epigenetics in adjacent, morphologically normal tissue, as well as the presence of multifocal disease in a high proportion of prostates. Field effects have also been proposed for oral cancer[22], head and neck cancer[23] and breast cancer[24]. Our results have implications for the use of cancer focal therapy when targeting a single nodule of cancer in the prostate[25,26] and for potential chemotherapeutic approaches. We propose that (i) focal therapy may be curative only if surrounding clonal-cell populations within morphologically normal tissue are also ablated, and (ii) cancer heterogeneity may hinder therapeutic targeting and biomarker investigation.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. EGA: EGAD00001000689.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This work was funded by Cancer Research UK (grant C5047/A14835), the Dallaglio Foundation and the Wellcome Trust. We also acknowledge support from the Bob Champion Cancer Trust, the Orchid Cancer Appeal, the RoseTrees Trust, the North West Cancer Research Fund, Big C, the King family, the Grand Charity of Freemasons, and the Research Foundation Flanders (FWO). We thank D. Holland from the Infrastructure Management Team and P. Clapham from the Informatics Systems Group at the Wellcome Trust Sanger Institute. We acknowledge the Biomedical Research Centre at the Institute of Cancer Research and the Royal Marsden NHS Foundation Trust, supported by the National Institute for Health Research. We acknowledge support from the National Cancer Research Prostate Cancer: Mechanisms of Progression and Treatment (PROMPT) collaborative (grant G0500966/75466). We thank the National Institute for Health Research, Hutchison Whampoa Limited and the Human Research Tissue Bank (Addenbrooke's Hospital), the Cancer Research UK Cambridge Research Institute Histopathology, the *In-situ* Hybridisation Core Facility, the Genomics Core Facility Cambridge and the Cambridge University Hospitals Media Studio.

AUTHOR CONTRIBUTIONS

C.S.C., R.E. and D.E.N. are senior principal investigators who designed and coordinated the study. C.S.F. is a senior principal investigator and histopathology lead. D.S.B. and U.M. are senior principal investigators for this project and bioinformatics project coordinators. D.E., A.F. and M.R.S. are senior principal investigators for this project. D.C.W. and P.V.L. had overall responsibility for data analysis. A.Y.W. is a histopathology lead. G.G. performed chromoplexy analysis. L.B.A. analyzed mutational signatures. H.C.W. was a principal investigator for this particular project who also carried out data analysis and tissue collection. A.B. and S.O'M. are coordinators of the DNA mutation–analysis pipeline. C.E.M. was involved in data analysis and formulation of the manuscript structure. P.C., B.K., J.Z., S.N.-Z. and A.G.L. were involved in data analysis and interpretation. N.D., S.E., L. Matthews and S. Merson completed tissue collection and FISH analysis of DNA preparations. N.C., C.G., M.R. and Z.K.-T. carried out data analysis. D.L. performed data validation. J.K. and H.J.L. collected tissue and performed DNA extractions. S.T. obtained patient consent, collected blood and carried out blood DNA preparations. J.C. and R.H. performed FISH analysis. R.M. and T.V. were involved in data interpretation. R.G.B., P.C.B. and M.F. were involved in determining the overall study design. S.C., K.R., D.J., A.M., L.S., J.H., J.T., S. McLaren, L. Mudie, C.H., E.A., O.J., V. Goody, B.R., M.M. and S.G. ran the data mutational analysis pipeline. C.F., C.C., D.B., N.L. and S.H. completed histopathology and tissue collection. C.O., P.K., A.T., C.W., D.N., E.M., T.D., N.C.S. and V. Gnanapragasam were responsible for tissue collection.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests. Details are available in the online version of the paper.

1. Andreoiu, M. & Cheng, L. Multifocal prostate cancer: biologic, prognostic, and therapeutic implications. *Hum. Pathol.* **41**, 781–793 (2010).
2. Cheng, L. *et al.* Evidence of independent origin of multiple tumors from patients with prostate cancer. *J. Natl. Cancer Inst.* **90**, 233–237 (1998).
3. Kobayashi, M. *et al.* Molecular analysis of multifocal prostate cancer by comparative genomic hybridization. *Prostate* **68**, 1715–1724 (2008).
4. Boyd, L.K. *et al.* High-resolution genome-wide copy-number analysis suggests a monoclonal origin of multifocal prostate cancer. *Genes Chromosom. Cancer* **51**, 579–589 (2012).
5. Lindberg, J. *et al.* Exome sequencing of prostate cancer supports the hypothesis of independent tumour origins. *Eur. Urol.* **63**, 347–353 (2013).
6. Clark, J. *et al.* Complex patterns of ETS gene alteration arise during cancer development in the human prostate. *Oncogene* **27**, 1993–2003 (2008).
7. Attard, G. *et al.* Duplication of the fusion of TMPRSS2 to ERG sequences identifies fatal human prostate cancer. *Oncogene* **27**, 253–263 (2008).
8. Gaisa, N.T. *et al.* Clonal architecture of human prostatic epithelium in benign and malignant conditions. *J. Pathol.* **225**, 172–180 (2011).

9. Nik-Zainal, S. *et al.* The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).
10. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
11. Svensson, M.A. *et al.* Testing mutual exclusivity of ETS rearranged prostate cancer. *Lab. Invest.* **91**, 404–412 (2011).
12. Alexandrov, L.B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
13. 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).
14. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).
15. Baca, S.C. *et al.* Punctuated evolution of prostate cancer genomes. *Cell* **153**, 666–677 (2013).
16. Berger, M.F. *et al.* The genomic complexity of primary human prostate cancer. *Nature* **470**, 214–220 (2011).
17. Grasso, C.S. *et al.* The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487**, 239–243 (2012).
18. Barbieri, C.E. *et al.* Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat. Genet.* **44**, 685–689 (2012).
19. Weischenfeldt, J. *et al.* Integrative genomic analyses reveal an androgen-driven somatic alteration landscape in early-onset prostate cancer. *Cancer Cell* **23**, 159–170 (2013).
20. 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).
21. 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).
22. Slaughter, D.P., Southwick, H.W. & Smejkal, W. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. *Cancer* **6**, 963–968 (1953).
23. Leemans, C.R., Braakhuis, B.J.M. & Brakenhoff, R.H. The molecular biology of head and neck cancer. *Nat. Rev. Cancer* **11**, 9–22 (2011).
24. Dworkin, A.M., Huang, T.H.-M. & Toland, A.E. Epigenetic alterations in the breast: implications for breast cancer detection, prognosis and treatment. *Semin. Cancer Biol.* **19**, 165–171 (2009).
25. Karavitakis, M., Ahmed, H.U., Abel, P.D., Hazell, S. & Winkler, M.H. Tumor focality in prostate cancer: implications for focal therapy. *Nat. Rev. Clin. Oncol.* **8**, 48–55 (2011).
26. Tareen, B., Godoy, G. & Taneja, S.S. Focal therapy: a new paradigm for the treatment of prostate cancer. *Rev. Urol.* **11**, 203–212 (2009).
27. Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* **39**, e118 (2011).
28. Chang, X. & Wang, K. wANNOVAR: annotating genetic variants for personal genomes via the web. *J. Med. Genet.* **49**, 433–436 (2012).
29. Lawrence, M.S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
30. Campbell, P.J. *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat. Genet.* **40**, 722–729 (2008).

[1]Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK. [2]Department of Biological Sciences, University of East Anglia, Norwich, UK. [3]Norwich Medical School, University of East Anglia, Norwich, UK. [4]Royal Marsden NHS Foundation Trust, London and Sutton, UK. [5]Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton, UK. [6]Human Genome Laboratory, Department of Human Genetics, VIB and KU Leuven, Leuven, Belgium. [7]Cancer Research UK London Research Institute, London, UK. [8]Statistics and Computational Biology Laboratory, Cancer Research UK Cambridge Research Institute, Cambridge, UK. [9]Urological Research Laboratory, Cancer Research UK Cambridge Research Institute, Cambridge, UK. [10]Department of Histopathology, St. Georges Hospital, London, UK. [11]Institute of Food Research, Norwich Research Park, Norwich, UK. [12]Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada. [13]Department of Radiation Oncology, University of Toronto, Toronto, Ontario, Canada. [14]Princess Margaret Cancer Centre–University Health Network, Toronto, Ontario, Canada. [15]Informatics and Bio-Computing, Ontario Institute for Cancer Research, Toronto, Ontario, Canada. [16]Department Pharmacology & Toxicology, University of Toronto, Toronto, Ontario, Canada. [17]School of Computing Sciences, University of East Anglia, Norwich, UK. [18]Department of Molecular Oncology, Barts Cancer Centre, Barts and the London School of Medicine and Dentistry, London, UK. [19]A full list of members and affiliations is provided in the **Supplementary Note**. [20]Laboratory of Reproductive Genomics, Department of Human Genetics, KU Leuven, Leuven, Belgium. [21]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK. [22]Department of Histopathology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. [23]Department of Histopathology, University of Liverpool, Liverpool, UK. [24]HCA Pathology Laboratories, London, UK. [25]The Genome Analysis Centre, Norwich, UK. [26]Department of Surgical Oncology, University of Cambridge, Addenbrooke's Hospital, Cambridge, UK. [27]These authors contributed equally to this work. [28]These authors jointly supervised this work. Correspondence should be addressed to C.S.C. (colin.cooper@icr.ac.uk), R.E. (ros.eeles@icr.ac.uk) or D.E.N. (den22@medschl.cam.ac.uk).

ONLINE METHODS

Sample selection and fluorescence *in situ* hybridization. Samples for analysis were collected from prostatectomy patients at Addenbrooke's Hospital (see **Supplementary Table 2**). The study was approved by the Trent Multicentre Research Ethics Committee. Informed consent was obtained for all patients. Prostates were sliced and processed as described previously[31]. In brief, a single 5-mm slice of the prostate was selected for research purposes, and 4-mm or 6-mm cores were taken from the slice and frozen. Frozen cores were mounted vertically and sectioned transversely to create one 5-µm frozen section for H&E staining and six 50-µm sections for DNA preparation. The presence of or complete absence of cancer was confirmed independently by three pathologists in a central pathology review of the 5-µm H&E-stained tissue slice immediately adjacent to tissue slices used for DNA preparation. The *ERG* FISH break-apart assay for assessing *ERG* gene rearrangement was performed as described previously[6], both (i) on whole-mount formalin-fixed sections of tissue immediately adjacent to the research slice and (ii) on the frozen slices of tissue immediately adjacent to the samples selected for DNA sequencing that had been initially subjected to H&E staining. In all cases, the *ERG* statuses determined by these two methods (**Fig. 1**) were consistent.

DNA sequencing. *Samples and massively parallel sequencing*. DNA was extracted from 18 samples from three subjects: 12 prostate cancer samples; 3 adjacent, morphologically normal prostate samples; and 3 matched bloods. Paired-end genome-wide sequencing (GWS) of the samples was performed at Illumina, Inc. Paired-end libraries were manually generated from 1 µg of genomic DNA using the Illumina Paired End Sample Prep Kit (catalog no. PE-102-1002). Fragmentation was performed with Covaris E220. After end repair, A-tailing and adaptor ligation as per the instructions in the Sample Prep Kit, libraries were manually size-selected using agarose gel electrophoresis, targeting 300-bp inserts. Adaptor-ligated libraries were PCR amplified for ten cycles and purified through a second agarose gel electrophoresis. Final libraries were checked for quality control on an Agilent Bioanalyzer and quantified by qPCR and/or picogreen fluorimetry. Samples were clustered with Illumina v1.5 flow cells using the Illumina cBot with the TruSeq Paired End Cluster Kit v3. Flow cells were sequenced as 100 base paired-end (non-indexed) reads on the Illumina HiSeq2000 using TruSeq SBS chemistry v3 to a target depth of 50× for the tumor samples and 30× for adjacent, morphologically normal and blood samples. The Burrows-Wheeler Aligner was used to align the sequencing data from each lane to the GRCh37 reference human genome[32]. Lanes that passed quality control were merged into a single, well-annotated sample BAM file with duplicate reads removed. These data have been submitted to the European Genome-Phenome Archive (EGAD00001000689).

Mutation calling: substitutions. CaVEMan (Cancer Variants through Expectation Maximization), an in-house bespoke algorithm developed at the Sanger Institute, was used for calling somatic substitutions. CaVEMan utilizes a Bayesian expectation-maximization algorithm: given the reference base, copy-number status and fraction of aberrant tumor cells present in each cancer sample, CaVEMan generates a probability score for potential genotypes at each genomic position. A 'somatic' probability of 95% or more was applied as a cutoff. Further post-processing filters were applied to eliminate false positive calls arising from genomic features that generate mapping errors and systematic sequencing artifacts. In addition to the standard filters applied in the Sanger pipeline, we designed project-specific filters to improve the positive predictive value of our callers on the basis of results from visual inspection and calling of many hundreds of variants. Visual inspection involved checking that the variant was in at least three reads but not in any reads of control; that there was no strand bias or correlation of the reads containing the variant and read quality; and that the variant was not in a location where indels were also detected, in a poorly mapped region or in a repeat region. Substitutions that were found in the GWS data of more than 2.5% of a batch of 465 normal non-malignant samples from a range of tissue types were also removed. Additional visual verification across all samples for a subject was performed for all non-intronic gene substitutions; all substitutions in adjacent, morphologically normal samples; potential 'field effect' substitutions; substitutions shared between neoplastic and adjacent, morphologically normal samples; and the rare predicted substitutions that apparently violated the inferred phylogeny.

Mutation calling: insertions/deletions. Insertions and deletions in the tumor, morphologically normal and matched-blood control genomes were called using modified Pindel version 0.2.0 on the NCBI37 genome build[33]. As with the substitutions, all standard Sanger pipeline filters were applied, as well as a custom filter built on the basis of results from visual calling of identified variants. Indels that were detected by Pindel in more than two samples from a series of hundreds of malignant non-prostate tissues were also removed. If an indel detected by Pindel that did not pass the filters was found in another sample for that subject and did pass all filters in that detection, it was also included. Of those indels that passed all filters, for each sample, up to 100 variants were validated by capillary sequencing. In addition, visual verification across all samples for a subject was performed for all indels occurring within genes; all indels in adjacent, morphologically normal samples; potential field effect indels; those indels that were not supported by the phylogeny; and a sampling of variants from each phylogeny relationship.

Mutation calling: structural variants. Brass (Breakpoints via assembly), an in-house bespoke algorithm developed at the Sanger Institute, was used to detect structural variants. In Brass phase 1, discordant read pairs are detected and integrated to find regions of interest. These regions of interest are removed if they were found in the matched-blood normal sample, were detected as germline in PCR validation of any other sample, have a low numbers of reads supporting them or appear to be in a 'difficult' region of the genome. For a subset of regions, validation was performed by gel electrophoresis PCR using custom-designed PCR primers across the rearrangement breakpoint as previously described[34], and for those products that gave a band, the precise location and nature of the breakpoint was determined by standard Sanger capillary sequencing methods. In cases where the PCR experiments failed, Brass phase 2 was applied to the remaining predicted somatic structural variants. Phase 2 gathers reads around the region, including half-unmapped reads, and performs a local *de novo* assembly using Velvet[35]. Identifiable breakpoints had a distinctive De Bruijn graph pattern and allowed the breakpoint to be regenerated down to base-pair resolution. Any breakpoints where an exact location could not be determined were removed. To ensure that breakpoints shared between samples from a single subject were picked up, we performed *in silico* and PCR cross-sample experiments. All breakpoints reported were visually verified to ensure the presence of discordant reads and checked to ensure that they were not in repeat regions.

To detect rearrangements involved in chromoplexy, a recently described process that generates chained rearrangements, we applied ChainFinder[15]. We used default parameters, selecting the rearrangements from 57 prostate genomes as background. As input copy-number data, we used data derived from Affymetrix SNP 6.0 arrays and processed using ASCAT[36]. As input structural variants, for each subject we combined all high-confidence breakpoints detected in all samples from that subject. One chained event was manually filtered, as it combined somatic rearrangements present in separate subpopulations in different samples and thus could not have occurred as one chromoplexy event.

Mutation calling: copy number. The Battenberg algorithm was used to detect clonal and subclonal somatic copy-number alterations (CNAs) and estimate ploidy and tumor content from the next-generation sequencing data as previously described[9]. Briefly, germline heterozygous SNPs were phased using Impute2, and a- and b- alleles were assigned. Data were segmented using piecewise constant fitting[37], and subclonal copy-number segments were identified via *t*-test as those with deviations in the b-allele frequencies compared to the values that would be expected when all cells had a common copy number in that segment. Ploidy and tumor content were estimated with the same method used by ASCAT[36].

Construction of phylogenetic trees. For each subject, phylogenetic trees were constructed separately using (i) CNAs and (ii) point mutations. Clonal and subclonal CNAs were identified using the previously described Battenberg algorithm[9]. This method achieves high sensitivity for the detection of CNAs found in small proportions of cells by phasing heterozygous SNPs into parent-specific haplotype blocks. Joint analysis of SNPs within these blocks, rather than single SNPs, allows for the resolution of CNAs found in ~5% of cells, with 30× sequencing depth. Matching of copy number and rearrangement breakpoints, supported by visual inspection of allele frequency and logR plots,

was used to identify CNAs common to multiple samples. Point mutations were analyzed using an adaptation of a previously described Bayesian Dirichlet process. Mutations within each sample were modeled as deriving from an unknown number of subclones, each of which was present in an unknown fraction of tumor cells and contributed an unknown proportion of all somatic mutations, with all the unknown parameters jointly estimated. In order to identify clusters of mutations common to two or more samples, we extended the Dirichlet process into two dimensions, with the fraction of tumor cells bearing a mutation in each of a pair of samples jointly estimated from the number of reads observed in each sample. The presence of clusters of unique or shared mutations could be inferred from the position of the peaks in the resulting 2D probability density.

Dirichlet process clustering. We used a previously developed Bayesian Dirichlet process to model clusters of clonal and subclonal point mutations, which allowed us to infer the number of subclones, the fraction of cells within each subclone and the number of mutations within each clone[36]. Within this model, the number of reads bearing the ith mutation, y_i, is drawn from a binomial distribution,

$$y_i \sim \mathrm{Bin}(N_i, \zeta_i \pi_i), \text{ with } \pi_i \sim \mathrm{DP}(\alpha P_0)$$

where N_i is the total number of reads at the mutated base and ζ_i is the expected fraction of reads that would report a mutation present in 100% of tumor cells at that locus. $\pi_i \in (0, 1)$, the fraction of tumor cells carrying the ith mutation, is modeled as coming from a Dirichlet process.

We used the stick-breaking representation of the Dirichlet process,

$$\omega_h = V_h \prod_{l<h} (1 - V_l), \text{ with } V_h \sim Beta(1, \alpha)$$

where ω_h is the weight of the hth mutation cluster (i.e., the proportion of all somatic mutations specific to that cluster). This model was extended into n dimensions, where n is the number of related samples, with the number of mutant reads obtained from each sample modeled as an independent binomial distribution, each with an independent π drawn with a Dirichlet process from a base distribution U(0,1). Gibbs sampling implemented in R, version 2.11.1, was used to estimate the posterior distribution of the parameters of interest. The Markov chain was run for 500 iterations, of which the first 100 were discarded. In order to plot the mutation density, we treated each possible pair of related samples separately. The median of the density was estimated from π_h, weighted by the associated value of ω_h, using a bivariate Gaussian kernel implemented in the R library KernSmooth. Median values were then plotted with the R function 'levelplot', using a color palette graduated from white (low probability of a mutation) to red (high probability of a mutation).

Targeted PCR and MiSeq sequencing of selected mutations and structural variants. PCR primers for somatic substitutions and indels were designed using Primer-Z[38], with known SNPs and human repeats masked. All amplicons were designed to be a maximum of 500 bp, and all variants of interest were checked to make sure they were within a read generated on a 2 × 250-bp MiSeq run. DNA was amplified using the Phusion HotStart II DNA polymerase kit (Thermo Fisher Scientific) and a thermocycler. DNA was denatured at 98 °C for 30 s and then underwent 30 cycles of denaturing at 98 °C for 10 s, annealing at 65 °C for 20 s and extension at 72 °C for 20 s. Products were incubated at 72 °C for 5 min before being cooled to 4 °C. All PCR products were analyzed using 96-well 2% agarose E-gels with ethidium bromide (Life Technologies). If no detectable band was present, these reactions were repeated using an annealing temperature of 60 °C. We pooled 2 µl of PCR mixture for each sample of DNA. Pooled DNA was diluted 1:10 and tagged with an individual barcode (Fluidigm) using the Expand High Fidelity PCR System (Roche) according to the manufacturer's protocol (Access Array System for Illumina Systems User Guide). DNA was denatured at 98 °C for 1 min and then subjected to 15 cycles of denaturing at 98 °C for 15 s, annealing at 60 °C for 30 s and extension at 72 °C for 1 min. Products were incubated at 72 °C for 3 min before being cooled to 4 °C. Barcoded PCR samples were pooled for each subject and analyzed using a 2100 Bioanalyzer (Agilent) to determine the average size of the PCR library and by KAPA SYBR FAST qPCR (Anachem) to determine the library concentration. We analyzed 2 nmol of each sample using MiSeq (Illumina).

The average sequencing depth across all mutations assessed within each subject varied between 4,900 (in 8_T1) and 16,600 (in 7_T4). However, for around one-fifth of the targeted mutations within each subject, the average coverage across all samples from that subject was much lower (200 or less). Many of these low-coverage mutations had mutant allele frequencies that were very different from the values obtained from GWS. These PCRs were considered to have failed and were not included in subsequent analysis.

Because of the very high coverage, a low rate of sequencing errors was observed for most mutations. This manifested as a small percentage of aberrant reads, peaked close to zero and rapidly decaying exponentially with allele fraction. We evaluated the rate of these errors by considering those samples in which no mutant reads were reported in GWS. For this purpose, only mutations that were identified in samples that were previously identified as phylogenetically related were included, in order to filter out low-quality or questionable calls. Allele frequencies f_s were converted to mutation copy numbers n_{mut} as previously described[39].

$$n_{\mathrm{mut}} = f_s \frac{1}{\rho}[\rho n^t_{\mathrm{locus}} + n^n_{\mathrm{locus}}(1-\rho)]$$

where ρ, n^t_{locus} and n^n_{locus} are, respectively, the tumor purity, the locus-specific copy number in the tumor cells and the locus-specific copy number in the blood normal cells, inferred from the Battenberg algorithm. Mutation copy numbers correspond to the percentage of cells bearing a mutation multiplied by the number of chromosomal copies bearing the mutation and are more informative than raw allele frequencies, as they are adjusted for tumor ploidy and normal cell contamination. The distribution of misreads was then found to be similar for the different subjects, with average reported mutation copy numbers of 0.0059 ± 0.0072, 0.0032 ± 0.0070 and 0.0037 ± 0.0035 in subjects 6, 7 and 8, respectively. The highest reported mutation copy number for these mutations was 0.041. This value was therefore used as a threshold for distinguishing between mutations present in a small proportion of cells and misreads arising from sequencing errors. It should be noted that a mutation copy number of 0.041 corresponds to an allele frequency of ~1% for most mutations, as most mutations occur in diploid regions of the genome and the average tumor content across the samples was less than 50%.

For samples 6_T2, 6_T3 and 6_T4, it was apparent that nearly all mutations that were present in 6_T1 were identified at allele fractions slightly greater than the threshold used to exclude artifacts (corresponding to a mutation copy number of ~0.05). As these mutations were exclusively those present in 6_T1, it appeared that 'contamination' of these three samples by 6_T1 occurred at some point during the PCR experiment, although whether this contamination was physical or the result of bleed-through of tags used in multiplexing is unknown. Assessment of GWS data, which involved checking the allele frequency of mutations identified only in 6_T1 in samples 6_T2, 6_T3 and 6_T4, indicated that there might have been some intermixing of the cells from 6_T1 with 6_T2, corresponding to a much lower percentage of cells (1.8%) and possibly arising from growth of cells from 6_T1 into the region sampled in 6_T2. Further, no evidence for intermixing of 6_T1 with 6_T3 or 6_T4 was found in GWS data. For this reason, mutations apparently present in the PCR experiment in 6_T2, 6_T3 and 6_T4 and identified in 6_T1 in both GWS and PCR were considered to be validated only if they fell above a higher threshold, set to a mutation copy number of 0.2, that excluded mutant reads arising from the contamination of these samples.

Mutational signatures. The mutational spectra, as defined by the triplets of nucleotides around each mutation, of each sample were deconvoluted into mutational processes as described[12,13].

Clustering of mutations. We investigated regional clustering of substitution mutations by constructing 'rainfall' plots in which the distance between each somatic substitution and the substitution immediately before it was plotted for each mutation. This was achieved exactly as described previously[9].

doi:10.1038/ng.3221

31. Warren, A.Y. *et al.* Method for sampling tissue for research which preserves pathological data in radical prostatectomy. *Prostate* **73**, 194–202 (2013).
32. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
33. Ye, K., Schulz, M.H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865–2871 (2009).
34. Campbell, P.J. *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat. Genet.* **40**, 722–729 (2008).
35. Zerbino, D.R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829 (2008).
36. Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proc. Natl. Acad. Sci. USA* **107**, 16910–16915 (2010).
37. Nilsen, G. *et al.* Copynumber: efficient algorithms for single- and multi-track copy number segmentation. *BMC Genomics* **13**, 591 (2012).
38. Tsai, M.-F. *et al.* PrimerZ: streamlined primer design for promoters, exons and human SNPs. *Nucleic Acids Res.* **35**, W63–W65 (2007).
39. Stephens, P.J. *et al.* The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486**, 400–404 (2012).

Corrigendum: Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue

Colin S Cooper, Rosalind Eeles, David C Wedge, Peter Van Loo, Gunes Gundem, Ludmil B Alexandrov, Barbara Kremeyer, Adam Butler, Andrew G Lynch, Niedzica Camacho, Charlie E Massie, Jonathan Kay, Hayley J Luxton, Sandra Edwards, Zsofia Kote-Jarai, Nening Dennis, Sue Merson, Daniel Leongamornlert, Jorge Zamora, Cathy Corbishley, Sarah Thomas, Serena Nik-Zainal, Manasa Ramakrishna, Sarah O'Meara, Lucy Matthews, Jeremy Clark, Rachel Hurst, Richard Mithen, Robert G Bristow, Paul C Boutros, Michael Fraser, Susanna Cooke, Keiran Raine, David Jones, Andrew Menzies, Lucy Stebbings, Jon Hinton, Jon Teague, Stuart McLaren, Laura Mudie, Claire Hardy, Elizabeth Anderson, Olivia Joseph, Victoria Goody, Ben Robinson, Mark Maddison, Stephen Gamble, Christopher Greenman, Dan Berney, Steven Hazell, Naomi Livni, the ICGC Prostate Group, Cyril Fisher, Christopher Ogden, Pardeep Kumar, Alan Thompson, Christopher Woodhouse, David Nicol, Erik Mayer, Tim Dudderidge, Nimish C Shah, Vincent Gnanapragasam, Thierry Voet, Peter Campbell, Andrew Futreal, Douglas Easton, Anne Y Warren, Christopher S Foster, Michael R Stratton, Hayley C Whitaker, Ultan McDermott, Daniel S Brewer & David E Neal

***Nat. Genet.* 47, 367–372 (2015); published online 2 March 2015; corrected after print 5 May 2015**

In the version of this article initially published, author Manasa Ramakrishna was omitted from the author list. The error has been corrected in the PDF and HTML versions of this article.