Sequencing of prostate cancers identifies new cancer genes, routes of progression and drug targets.

David C. Wedge[1,2,34*], Gunes Gundem[2,3,34], Thomas Mitchell[2],[4,5,34], Dan J. Woodcock[1], Inigo Martincorena[2], Mohammed Ghori[2], Jorge Zamora[2], Adam Butler[2], Hayley Whitaker[6], Zsofia Kote-Jarai[7], Ludmil B. Alexandrov[2], Peter Van Loo[2,8], Charlie E. Massie[5], Stefan Dentro[2,8], Anne Y. Warren[9], Clare Verrill[10], Dan M. Berney[11], Nening Dennis[12], Sue Merson[7], Steve Hawkins[5], William Howat[9], Yong-Jie Yu[11], Adam Lambert[13], Jonathan Kay[6], Barbara Kremeyer[2], Katalin Karaszi[13], Hayley Luxton[6], Niedzica Camacho[7,3], Luke Marsden[13], Sandra Edwards[7], Lucy Matthews[13], Valeria Bo[14], Daniel Leongamornlert[7], Stuart McLaren[2], Anthony Ng[15], Yongwei Yu[16], Hongwei Zhang[16], Tokhir Dadaev[7], Sarah Thomas[12], Douglas F. Easton[17,33], Mahbubl Ahmed[7], Elizabeth Bancroft[7,12], Cyril Fisher[12], Naomi Livni[12], David Nicol[12], Simon Tavaré[14], Pelvender Gill[13], Christopher Greenman[18], Vincent Khoo[12], Nicholas Van As[12], Pardeep Kumar[12], Christopher Ogden[12], Declan Cahill[12], Alan Thompson[12], Erik Mayer[12], Edward Rowe[12], Tim Dudderidge[12], Vincent Gnanapragasam[4,19], Nimish C. Shah[4], Keiran Raine[2], David Jones[2], Andrew Menzies[2], Lucy Stebbings[2], Jon Teague[2], Steven Hazell[12], CAMCAP study group, Johann de Bono[7], Gerthardt Attard[7], William Isaacs[20], Tapio Visakorpi[21], Michael Fraser[22], Paul C Boutros[23,24,25], Robert G Bristow[22,24,26], Paul Workman[7], Chris Sander[27], The TCGA consortium[28], Freddie C. Hamdy[13], Andrew Futreal[2], Ultan McDermott[2], Bissan Al-Lazikani[7,35], Andrew G. Lynch[14,33,35], G. Steven Bova[21,20,33,35], Christopher S. Foster[29,30,33,35], Daniel S. Brewer[7,18,31,33,35], David Neal[5,19,33,35], Colin S. Cooper[7,18,33,35], and Rosalind A. Eeles[7,12,33,35*]

[1] Oxford Big Data Institute, University of Oxford, Old Road Campus, Oxford, OX3 7LF, UK

[2] Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, UK

[3]Memorial Sloan-Kettering Cancer Center, NY 10065, New York, USA

[4]Department of Urology, Addenbrooke's Hospital, Cambridge, CB2 0QQ, UK

[5]Uro-Oncology Research Group, Cancer Research UK, Cambridge Institute, Cambridge, CB2 0RE, UK

[6]Molecular Diagnostics and Therapeutics Group, University College London WC1E 6BT

[7]The Institute Of Cancer Research, London, SW7 3RP, UK

[8]Cancer Genomics, The Francis Crick Institute, London, NW1 1AT, UK

[9]Department of Histopathology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, CB2 0QQ, UK

[10]Oxford University Hospitals NHS Trust, John Radcliffe Hospital, Oxford, OX3 9DU, UK

[11]Centre for Molecular Oncology, Barts Cancer Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, E1 2AD, UK

[12]Royal Marsden NHS Foundation Trust, London and Sutton, SM2 5PT, UK

[13]The University of Oxford, Oxford, OX1 2JD, UK

[14]Statistics and Computational Biology Laboratory, Cancer Research UK Cambridge Institute, Cambridge, CB2 0RE, UK

[15]The Chinese University of Hong Kong, Shatin, NT, Hong Kong, China

[16]Second Military Medical University, Shanghai, China 20043

[17]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, CB1 8RN, UK

[18]Norwich Medical School, University of East Anglia, Norwich, NR4 7TJ, UK

[19]Department of Surgical Oncology, University of Cambridge, Addenbrooke's Hospital, Cambridge, CB2 0QQ, UK

[20]Johns Hopkins School of Medicine, Baltimore, MD 21205, USA

[21]Institute of Biosciences and Medical Technology, BioMediTech, University of Tampere and Fimlab Laboratories, Tampere University Hospital, Tampere, FI-33520, Finland

[22]Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada M5G 2M9

[23]Ontario Institute for Cancer Research, Toronto, Ontario, Canada M5G 0A3

[24]Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada M5G 1L7

[25]Department of Pharmacology and Toxicology, University of Toronto, Toronto, Ontario, Canada M5S 1A8

[26] Department of Radiation Oncology, University of Toronto, Toronto, Canada

[27]cBio Center, Dana-Farber Cancer Institute & Harvard Medical School, Boston, MA 02215, USA

[28]The TCGA consortium, National Cancer Institute @ NIH, Bethesda, MD20892, USA

[29]University of Liverpool, Liverpool, UK

[30]HCA Laboratories, London, WC1E 6JA, UK

[31]Earlham Institute, Norwich, NR4 7UH

[32]School of Mathematics and Statistics/School of Medicine, University of St. Andrews, KY16 9SS, UK

[33]Joint PIs of CRUK-Prostate Cancer ICGC Project

[34]Joint first authors made an equal contribution to this paper

[35]Joint last authors made an equal contribution to this paper

*Correspondence should be addressed to: D.C.W (david.wedge@bdi.ox.ac.uk) & R.A.E (Rosalind.eeles@icr.ac.uk)

Abstract

Prostate cancer (PCa) represents a significant clinical challenge because it is difficult to predict outcome and advanced disease is often fatal. We sequenced the whole genomes of 112 primary and metastatic PCa samples. From joint analysis of these cancers with those from previous studies, 930 cancers in total, we identified evidence for 22 novel putative coding driver genes, as well as evidence for *NEAT1* and *FOXA1* acting as drivers through non-coding mutations. Through the temporal dissection of aberrations, we identified driver mutations specifically associated with steps in the progression of PCa, for example establishing loss of *CHD1* and *BRCA2* as early events in cancer development of *ETS* fusion negative cancers. Computational chemogenomic (CanSAR) analysis of PCa mutations identified eleven targets of current drugs, eight of investigational drugs and fifty four compounds that may be active and should be considered candidates for future clinical trials.

INTRODUCTION

Prostate cancer is the most common solid cancer in men (diagnosed in 12%) and often fatal (9% of male cancer deaths). It is difficult to manage clinically due to a poor current understanding of what dictates its highly variable natural history, and of what underlies the development of castration-resistant disease[1]. Extensive data on the structure of prostate cancer genomes have been published[2-6], including work from our own consortium[7-10]. These studies have identified a number of genetically distinct subgroups, including cancers with *ERG*, *ETV1*, *ETV4*, *FLI1*, *SPOP*, *FOXA1* and *IDH1* alterations. Overlapping with these categories, cancers may have alterations in PI3K and DNA repair pathways, with the latter significantly over-represented in advanced disease[4]. However, we have relatively limited understanding of the ordering of genetic events with the exception that ETS gene alteration appears to represent an early event, whilst mutations of *AR* are later, sometimes convergent, events, occurring in advanced and metastatic disease. Indeed, we have very little understanding of the evolution of mutational processes, the various genetic paths that cancers traverse on their way to progression, the levels of heterogeneity at different stages of development or the effect of these factors on clinical outcome.

Gene status has been used in studies designed to improve the poor predictive value of conventional clinical markers (PSA, Gleason sum, stage) and to develop disease management strategies. For example, genetic alteration of *BRCA1/2*[11], *PTEN* deletion[12], amplification of *AURKA* together with the *MYCN* gene[13], and coordinated loss of *MAP3K7* and *CHD1*[14] have been reported to have prognostic value. A number of commercial prognostic tests based on gene expression profiles are also available [Prolaris™ [15], Decipher™ [16] and OncotypeDx™ [17]] and a classification framework has been proposed[18]. Improvements in the treatment of castration-resistant disease have been made through

better targeting of AR regulation using abiraterone[19] and enzalutamide[20], whilst PARP inhibitors are effective against cancers harbouring *BRCA1/2* mutations and other defects in DNA repair pathways[21]. However, a significant recent improvement in the treatment of newly diagnosed advanced prostate cancer has been achieved through using docetaxel in combination with hormone therapy, a re-tasking of conventional therapy[22].

In the present study, we use previously unpublished whole genome DNA sequencing data in combination with published data to provide new insights into the mechanism of progression of prostate cancer to lethal disease, and to design novel molecular-based strategies for drug targeting.

RESULTS

We whole genome sequenced cancerous and matched normal samples from 87 primary prostate cancers from the UK and 5 from Shanghai, China together with 10 hormone-naïve prostate metastases and 10 castration-resistant metastases from the USA. Analysis (see Online Methods) reveals insights into the nature and order of acquisition of driver alterations, heterogeneity of genomic heterogeneity in primary and metastatic cancers, changes in mutational signatures during progression, and potential drug targets. In addition, we identified coding and non-coding drivers by combining single nucleotide variants (SNVs) and small insertions/deletions (indels) within our dataset with those from TCGA[4] (425 samples), the COSMIC database[23] (243 samples) and Stand Up to Cancer[24] (SU2C-PCF, 150 samples) to give a combined dataset, hereafter referred to as the 'joint dataset', comprising 710 primary cancers and 220 metastases. Supplementary Table 1 summarises the genes affected in both our study and the joint dataset.

For the 112 cancer-normal pairs in our cohort, we identified 392,753 substitutions, 54,952 small insertion/deletions (indels) and 10,921 chromosomal rearrangements (Fig. 1). The mean genome-wide substitution rate across the whole dataset was 1.23/Mb, with a significant difference in mutational burden between the primary (0.99) and metastatic (2.30) samples (P=4.4x10^{-15}, Online Methods). Moreover, within the metastatic subset, mutation burden was higher in men treated with androgen deprivation therapy (ADT or CRPC) with metastases than the treatment-naïve cases (2.98 vs 1.61, P=0.015). There were also significantly more rearrangements in metastatic than in primary samples (P=0.0059), whilst the proportion of breakpoints attributed to a chromoplexy-like event[25] was indistinguishable between the two groups. Within the metastatic group, the ADT samples had more rearrangements than did the hormone-naïve (P=0.027), with no difference in the proportion of chromoplexy-like events (Fig. 1).

Genes of interest were identified through a comprehensive set of analyses to identify: excess of non-synonymous mutations in coding regions; excess missense mutations within a gene, indicative of an oncogenic driver; excess of mutations in non-coding regions; regions with an excess of structural variants in either ETS+ or ETS- cancers; and regions with recurrent copy number aberrations in either ETS+ or ETS- cancers. Overall, we identified 73 genes with evidence for involvement in prostate cancer development (Fig. 2, Table 1, Supplementary Table 2). Based on a literature search, each gene was assigned a high, medium or low level of previous supporting evidence (Table 1, Supplementary Table 2). In addition to 22 genes with little or no previous evidence of involvement in prostate cancer (Table 1, ‘low’ previous evidence), we provide corroborating evidence for 8 further genes previously lacking strong evidence of driving prostate cancer (Table 1, ‘medium’ previous evidence).

Coding drivers

We identified 28 genes with an excess of non-synonymous coding mutations, five of which are previously unknown drivers in prostate cancer (Supplementary Table 2). *TBL1XR1* was enriched in truncating SNVs and indels and is also located in a genomic region enriched for rearrangements in ETS+ cancers (chr3: 172-179Mb) (Fig. 3). These rearrangements result in loss of heterozygosity (LOH) or, in one case, homozygous deletion, suggesting a cancer suppressor role for this gene. Another significantly mutated gene primarily affected by truncating mutations was *ZMYM3*, which encodes a component of CoREST, a transcriptional repressor complex including *REST* (RE-1 silencing transcription factor) and involved in suppression of neuronal differentiation-related genes in non-nervous tissues[26]. In addition, two further CRPC samples from the SU2C-PCF study[24] had nonsense mutations and one sample within our study had a 70kb exonic deletion in *REST*.

Two other genes with recurrent truncating mutations were *IL6ST* and *CASZ1* (Fig. 3). The latter is a putative cancer suppressor in neuroblastoma[27] while the former encodes glycoprotein 130, the signal-transducing subunit of the interleukin 6 (IL6) receptor. The pattern of mutations we observe in the joint dataset for *IL6ST* is dominated by truncating events. Moreover, this gene is located in a genomic region recurrently rearranged in ETS+ cancers, resulting in either LOH or homozygous deletion (four cases of each), suggesting a cancer suppressive role. *TBX3*, which has previously been reported to harbor mutations in breast cancer[28], exhibited a mixed pattern of mutations with mostly missense mutations and two cancers harbouring truncating events.

Analysis of missense mutations identified mutations in seven further genes, of which two are newly reported (Supplementary Table 2). *CNOT3* exhibited mutation hotspots in two amino acid positions, p.E20K (4/932 samples) and p.E70K (5/932 samples), as well as a nonsense

mutation in a single sample (Fig. 3). *CNOT3* has a known cancer suppressive function in T-cell acute lymphoblastic leukaemia[29]. Enrichment for missense mutations was identified in *RPL11* a ribosomal protein and a putative cancer suppressor upstream of the MDM2/TP53 pathway[30]. In contrast to previous studies, the enrichment for missense mutations in both *CNOT3* and *RPL11* suggests oncogenic, rather than tumor suppressor, roles in prostate cancer.

A comparison between coding mutations in the metastatic and primary samples within the joint dataset identified enrichment in metastases for mutations in *TP53, AR, KMT2C, KMT2D, RB1, APC, BRCA2, CDK12, ZFHX3, CTNNB1, PIK3CB* (Supplementary Table 2), confirming previous studies[3,24].

Non-coding drivers

Analysis of the non-coding components of the genome identified two regions with statistically significant enrichment for mutations. *NEAT1*, a lncRNA recently reported to be associated with PCa progression[31], was mutated in 13/112 ICGC cases with a significant over-representation in patients with metastatic disease (6/20 metastases vs. 7/91 primaries, Fisher exact test, P=0.012, Fig. 3). Interestingly, out of the metastatic cases *NEAT1* mutations were found only in patients that had undergone ADT, consistent with the link between high *NEAT1* expression and resistance to AR-targeting therapies[31]. Notably, two of these six cases had two separate *NEAT1* mutations. The *FOXA1* promoter also had significant evidence of selection. This gene modulates AR-regulated transcriptional signalling[32] and has been found to harbor recurrent coding mutations in previous studies[5]. In our series, we identified 14 samples with coding and 6 samples with non-coding mutations, with two samples (PD14721a and PD12813a) bearing both a coding and a non-coding mutation. Interestingly, we also identified mutations in the *FOXP1* promoter, a gene with known cancer-suppressive effect in

prostate tumorigenesis[33], in three samples, but this was insufficient to reach statistical significance.

Structural variant enrichment in ETS+ and ETS- cancers

The density of rearrangements varies across the genome as a result of a variety of factors including chromatin state, GC content, gene density, replication timing and repetitive sequence. In order to remove the effect of these various factors, we segmented inter-breakpoint distance across the genome separately in ETS+ and ETS- cancers and identified regions with differential enrichment for rearrangements between the two subtypes. The functional importance of many of these regions was supported by an excess of truncating mutations or CNAs.

In addition to regions previously identified as enriched for rearrangements in ETS+ cancers (*FOXP1*, *RYBP*, *SHQ1*, *PTEN*, and *TP53*)[34-37], two unreported regions were identified. The region chr5:55-59Mb covers the genes *PPAP2A*, *PDE4D*, *MAP3K1* and *IL6ST* (Fig. 3). In *IL6ST* we also detected significant enrichment for coding mutations, suggesting this is the main target of the aberrations. In chr3:171:178Mb, *TBL1XR1* is similarly enriched for both rearrangements and truncating mutations.

In ETS- cancers, we confirmed a previously reported enrichment for rearrangements containing *CHD1*[(38,39)]. A target of enriched rearrangements in the region chr1:149-158Mb is likely *ETV3*. In 5/9 cancers, *ETV3* was exclusively affected by these events (4 LOH by deletion and 1 by translocation). Additionally, one cancer had a truncating mutation (p.R413fs*3) and two had missense mutations (p.A73V and p.L37Q). In total, 12 patients had localised alteration, 10 of whom had ETS- cancers. In addition, eleven tumors, 8 of them ETS-, had LOH covering *ETV3*. Moreover, within the joint dataset, there are four cancer samples with truncating mutations in this gene. In contrast to *ETV4*, the nature of variants in

ETV3 is indicative of a tumor suppressive role in PCa. Manual inspection of the recurrently rearranged region chr3:76-84Mb identified *ROBO1* and *ROBO2* as possible targets (Fig. 3). In total 16/112 samples had an event affecting one or other of these genes, and in four samples both were affected. Previously implicated in pancreatic ductal adenocarcinoma[40], these two genes have not been previously reported in the context of PCa.

Events enriched at chr6:80-114Mb indicate that *ZNF292* is a possible target. 11/112 patients (5 ETS+ and 6 ETS-) had loss of at least one chromosome copy and in two patients there was a homozygous loss specifically targeting *ZNF292*. Moreover, the joint dataset contained 5/932 samples with a truncating mutation, further suggesting a cancer suppressive function for this gene in PCa. Another gene affected by recurrent rearrangements on 6q was *SENP6*, a small ubiquitin-like modifier (SUMO)-specific protease that removes SUMO polypeptides from conjugated proteins[41], and possibly plays a role in *AR* function[42]. Of note, 4/5 rearrangements in this region affected *SENP6* only, leading to a significant reduction in expression (Supplementary Fig. 1). Finally, located at chr6:126Mb, the nuclear receptor co-activator *NCOA7* was altered in six samples, with one sample having homozygous loss.

Further regions enriched in ETS- cancers were chr2:133-144Mb (*LRP1B*), chr8:112-114 (*CSMD3*) and chr8:40-41Mb (*MYST3*). The first two genes are very large and fall within reported fragile sites[43]. Nevertheless, preferential enrichment of breakpoints in ETS- cancers may suggest either that underlying structure, such as AR binding sites or nucleosome structure, or epistatic interactions between ETS fusion and other rearrangements affect the occurrence of rearrangements at these loci. Samples containing structural variants affecting MYST3 were found to have significantly reduced RNA expression (Supplementary Fig. 1).

Timing of copy number aberrations

In order to identify routes to progression in PCa, we developed a novel approach to order the occurrence of copy number aberrations by combining information on: the clonality of copy number aberrations; timing relative to whole genome duplication; timing of homozygous deletions relative to neighboring hemizygous losses. Information from all tumors was combined using a Bradley-Terry model, to give the most likely ordering of events. By applying a set of logical rules (see Online Methods), we deciphered the temporal ordering of the subclonal CNAs within each cancer. In general, homozygous deletions appear late in oncogenesis, corroborating previous findings that homozygous deletions are associated with more advanced disease[44-46]. Clear differences emerge in the evolution of PCa in the ETS+ and ETS- subsets. Where present, the deletion between the TMPRSS2 and ERG genes in ETS+ cancers was an early (generally clonal) event, as was gain of chr8q within the locus 112 – 137Mb (Fig. 4a). The earliest homozygous deletions in ETS+ cancers include chr5: 55Mb-59Mb, corroborating the rearrangements targeting *PPAP2A*, *PDE4D*, *MAP3K1* and *IL6ST*, and chr10:89Mb-90Mb, which covers *PTEN* (Figs. 3 and 4a).

In ETS- cancers losses at chr5:60–100Mb (*CHD1* and *RGMB*), chr13:32-91Mb (which includes *BRCA2*, *RB1* and *FOXO1*), and chr6:73-120Mb are followed by losses at chr2:124-142Mb, then by gains at chr3:100-187Mb, and then whole chromosome gain of chr7 (Fig. 4b). Loss of *CHD1* has been previously implicated in the initiation of ETS- prostate cancers, preventing ERG re-arrangement in the prostate[38] and our data confirm the exclusivity between ETS positivity and homozygous loss of *CHD1* (Fig. 4c).

In both ETS-positive and ETS-negative cancers, whole genome duplication (WGD) was correlated with the loss of chromosomal segments at: chr1:94Mb, chr2:140Mb, chr12:12Mb, chr16:85Mb and chr17:7Mb (Fig. 4c). From timing analysis, these losses appear to occur co-

synchronously with WGD in most cases. Gains at chr8:101Mb occurred prior to WGD, chr3:131Mb occurred synchronously, and gains at chr7:88Mb tended to follow WGD.

Timing point mutations and indels

Point substitutions and indels were clustered according to their cancer cell fraction (CCF) using a Bayesian Dirichlet process[47]. The proportion of substitutions identified as subclonal showed considerable variation across cancers, but was significantly higher in primary than metastatic samples (Fig. 5a, P=0.022, Wilcoxon rank sum test), as was the proportion of subclonal indels (P=0.00033) and the fraction of the genome with subclonal copy number aberrations (P=0.0037, Supplementary Fig. 2). This is apparent evidence for a bottleneck in acquiring metastatic potential rather than a response to treatment, since levels of heterogeneity in untreated metastases are no lower than in androgen-deprived metastases (Fig. 5a).

The levels of heterogeneity observed in substitutions and indels were correlated (Fig. 5a, Pearson r = 0.57, $P=2.3\times10^{-9}$). Higher levels of heterogeneity were observed amongst indels than substitutions ($P=2.4\times10^{-8}$). However, it cannot be ruled out that variant calling of indels may have greater sensitivity for low allele frequency variants than calling of point substitutions.

Driver mutations were identified as clonal or subclonal according to the cluster to which they were assigned, with 84 classified as clonal and 22 (21%) as subclonal. Our power to detect subclonal mutations is limited by sequencing depth and the real number of subclonal driver mutations is likely to be much higher. The driver mutations identified as subclonal include two mutations in *APC* in the same sample, PD14713a. Interestingly, this cancer has undergone clonal loss of one copy of chr5q, followed by mutations in *APC* in 2 different subclones (Fig. 5b and Supplementary Fig. 3), suggesting convergent evolution. Five other

samples each have two subclonal drivers: PD12808a has a missense mutation in *ZNF292* and an essential splice site mutation in *SMAD2*; PD13401a has a nonsense mutation in *PPP1R3A* and a mutation in the promoter of *NEAT1*; PD13402a has a nonsense mutation in *USP34* and an essential splice site mutation in *ABI3BP* (Fig. 5b); PD12820a has a missense mutation in *USP48* and an essential splice site mutation in *ASXL2*; PD13389a has a frameshift mutation in *PHF12* and an essential splice site mutation in *TBX3* (not shown).

Subclonal mutations are also seen in several common drivers including one in *TP53* (PD13339a) and one in *PTEN* (PD12840a). On the other hand, *SPOP* was mutated in 10 samples, always clonally and always in ETS- tumors (Fig. 2).

Mutational signatures

Analysis of the mutational signatures by non-negative matrix factorisation (NMF) revealed that, in addition to the ubiquitous 'clock-like' signatures 1 and 5, there was presence of the previously described signatures 2, 3, 8, 13 and 18[48]. Signature-3-positive samples were enriched for germline/somatic mutations in *BRCA1/2* genes (4/6 samples) as reported previously[48] (Fig. 1). However, the presence of high levels of microhomology (MH)-mediated deletions was even more strongly correlated with the presence of *BRCA* mutations (6/6 samples). Separating the mutations into early clonal, late clonal and subclonal epochs, as described in Online Methods, revealed that the proportion of signature 1 mutations decreases over time, suggesting an increase of cancer-associated mutagenic processes relative to innate processes in normal cells ($P=2.2\times10^{-16}$, test for trend in proportions).

Signature 13, previously associated with the activity of the AID/APOBEC family of cytidine deaminases, was over-represented in advanced disease, 45% (9/20) in metastases vs. 14% (14/92) in primaries, (Fisher exact test, $P=5.6\times10^{-3}$). Similarly, signature 18, which has been previously associated with failure of base excision repair and to the accumulation of

mutations from 8-Oxoguanine damage[49], was enriched in advanced disease, 40% (8/20) in metastases vs. 11% (10/92) in primaries (Fisher exact test, P=3.8x10^{-3}). In a recent report of 560 breast cancer whole-genomes, signature 8 correlated with DNA damage repair deficiency[50]. Androgen signalling is known to positively regulate multiple genes involved in DNA repair[51,52], while androgen deprivation impairs DNA double-strand break repair[53]. In support of these previous reports, we have found that the proportion of mutations assigned to signature 8 is consistently higher amongst later appearing (subclonal) populations of cells (55% ± 24%) than earlier (clonal) populations (28% ± 12%) (*t*-test, P=1.3×10^{-4}, Supplementary Table 3). The proportion of metastases with evidence for the action of signature 8 was higher than that for primary tumors, although not reaching statistical significance (8/20 metastases, 25/92 primaries, Fisher exact test P=0.28). Increased prevalence of DNA-damage related genes in metastatic prostate cancer as well as the observations made in this study warrant an extensive study of mutational signatures in therapy-naïve disease and CRPC in a larger dataset to explore the relevance of check-point inhibition as an alternative therapy for advanced prostate cancer.

Clinical correlates

CDH12 and *ANTXR2* alterations were significantly associated with time to biochemical recurrence (Benjamin-Hochberg adjusted P = 0.0060 (*CDH1*) & 0.012 (*ANTXR2*), HR = 9.3 & 7.7, Cox regression model, Fig. 6), and were significant predictors of biochemical recurrence independent of cofactors Gleason, PSA at prostatectomy, and pathological T-stage (P = 0.00061 (*CDH1*) & 0.0015 (*ANTXR2*), HR = 7.3 & 6.5, Cox regression model, Supplementary Table 4). A Cox regression model containing a combination of *CDH12, ANTXR2, SPOP, IL6ST, DLC1 & MTUS1* mutations was determined to be an optimal predictor of time to biochemical recurrence and was a significant improvement over a baseline model of Gleason, PSA at prostatectomy, and pathological T-stage (model χ^2 test, P

= 0.00053). The number of mutational signatures identified in a cancer was negatively correlated with time to biochemical recurrence in prostatectomy patients (P = 0.014, HR = 3.0; Cox proportional hazards model on number of processes greater than 3, Supplementary Fig. 4) and is an independent predictor (P = 0.0061, HR = 3.6; Cox proportional hazards model). The number of substitutions detected was also an independent prognostic biomarker (P=0.031, HR=1.005; Cox proportional hazards model). The numbers of both samples and events within this study are modest and further analysis of larger cohorts is required to establish firmly these findings.

Druggable targets in the prostate cancer disease network

A key opportunity arising from systematic analyses of cancer genomics is the early objective identification of therapeutic intervention strategies. To this end, we applied established chemogenomic technologies using the canSAR knowledgebase[54] to map and pharmacologically annotate the cellular network of the prostate disease genes we identified in this study. We derived the network using curated protein-protein and transcriptional interaction data. We included the protein products of the genes identified in this study and other key proteins that either directly interact with several of these proteins or directly affect their function (see Online Methods and Supplementary Fig. 5 for details). This resulted in a focussed prostate network of 163 proteins. We annotated each of these proteins with drug information and annotated them based on multiple assessments of 'druggability': the likelihood of the protein to be amenable to small molecule drug intervention. (Table 2 and Supplementary Table 5). We find that PCa driver genes are embedded in a highly druggable cellular network that contains eleven targets of approved therapies and eight targets of investigational drugs. As well as the Androgen Receptor (AR) and the Glutocorticoid Receptor (GR), the network contains targets of drugs

approved for other indications, several of which (e.g. BRAF, ESR1, RARA, RXRA, HDAC3) are under clinical investigation for PCa.

Eight proteins within the prostate network are targets of drugs currently in clinical trials. In particular, the ataxia-telangiectasia mutated (ATM) inhibitor AZD-0156, currently in Phase 1 for safety assessment, is a likely candidate for exploration in PCa due to the recently described role of DNA damage repair, particularly in advanced PCas[21,55]. The network highlights targets of PI3 Kinase pathway inhibitors (PI3K, AKT1) that are undergoing clinical investigation in PCa, as well as IDH1 and MDM2 drug targets.

To give an indication of the potential of these drugs, we analysed the most recent systematic drug sensitivity data (GDSC- [56]- http://www.cancerrxgene.org/). Eighteen drugs acting on our network were tested in GDSC on prostate cancer cell lines. Of these, 5 showed significant effect on growth inhibition in at least one cell line, and all 18 showed weak activity in at least one cell line (Supplementary Table 6). However, to validate fully the potential of these drugs, extensive drug sensitivity testing needs to be performed in disease-relevant cancer models that correctly reflect the patient population.

Potential future opportunities for PCa therapy are also highlighted by 15 proteins that are under active chemical biology or drug discovery investigation (Table 2). These include the Inositol 1,4,5-trisphosphate receptor type 2 (IP3R2) and Menin (MEN1), a component of the MLL/SET1 histone methyltransferase complex. Mice with MEN1 mutations develop PCa[57] and recent data have shown that menin expression is involved in CRPC[58]. A further 50 of the proteins are predicted to be druggable and therefore potentially amenable to drug discovery. These include the known PCa protein SPOP, the transcription activator BRG1 (SMARCA4); CDK12; and the CREB binding protein CREBBP.

In summary, we find that 74 of the 163 proteins central to the prostate disease network are either targets of existing drugs or have the potential to be targeted in the future. To maintain an up-to-date-view of this analysis, we provide a link to a live-page in canSAR

http://cansar.icr.ac.uk/cansar/publications/sequencing_prostate_cancers_identifies_new_cancer_genes_routes_progression_and_drug_targets/ (link via google chrome)

DISCUSSION

The analysis of whole genome sequence data from 112 prostate cancers has revealed many of the genetic factors underlying the processes of carcinogenesis, progression, metastasis and the acquisition of drug resistance. Supporting evidence has been provided for thirty candidate driver genes with limited or no previous support, including the non-coding drivers *NEAT1* and *FOXA1*.

Through the timing of genomic aberrations, we have a picture of the possible routes to progression in PCa. Most driver mutations may occur either clonally or subclonally, but mutations in *SPOP* and ETS-fusions occur early in cancer development and are exclusively clonal. Whereas the gain of 8q and ETS fusion appear to be sufficient to drive a dominant clonal expansion, ETS- cancers typically need a combination of large-scale losses, acquired over an extended period of time. Known cancer drivers are frequently observed subclonally and two competing drivers are seen in several cancers. Less heterogeneity of aberrations is observed in metastatic than primary cancers, likely resulting from a bottleneck in achieving metastatic potential.

We observe changes in the mutational processes operative upon cancers during progression. Signature 8 was operative to a greater extent at later stages, and was enriched in subclonal expansions, while signatures 13 and 18 were enriched in metastatic cancers. Cancers with

germline or somatic *BRCA1/BRCA2* mutations were enriched for signature 3, demonstrating the effect of double-strand repair defects throughout cancer evolution.

Survival analysis reveals that losses of *CDH12* and *ANTXR2* result in poorer recurrence-free survival. We identify 84 PCa associated proteins that are either targets for currently available drugs or new potential targets for therapeutic development.

Analysis of the whole-genome sequences of over a hundred prostate cancers has started to reveal the complex evolutionary pathways of these cancers. The early acquisition of individual driver aberrations including ETS-fusions and whole genome duplications strongly affects the acquisition of subsequent aberrations. Acquisition of individual mutations affects both the subsequent likelihood of metastasis and response to treatment. Network analyses using the candidate driver genes here identified, in addition to previously known drivers, targets that can be exploited for immediate clinical investigation with existing drugs as well as targets for new drug discovery, giving potential for the results of genome analysis to be translated rapidly into therapeutic innovation and patient benefit.

ONLINE METHODS

Patient Cohorts, Samples and Ethics

92 cancer samples from prostatectomy patients treated at The Royal Marsden NHS Foundation Trust, London, at the Addenbrooke's Hospital, Cambridge, at Oxford University Hospitals NHS Trust, and at Changhai Hospital, Shanghai, China were collected as described previously[59,60]. Clinical details for the patients are shown in Supplementary Table 7. Ethical approval was obtained from the respective local ethics committees and from The Trent Multicentre Research Ethics Committee. All patients were consented to ICGC standards

https://icgc.org/. 20 men from PELICAN (Project to ELIminate lethal CANcer)[61], an integrated clinical-molecular autopsy study of metastatic prostate cancer, were the subjects of the current study. Subjects consented to participate in the Johns Hopkins Medicine IRB-approved study between 1995 and 2005. (Supplementary Table 7). A17 had a germline *BRCA1* mutation, as previously reported[62].

DNA preparation and DNA sequencing

DNA from whole blood samples and frozen tissue was extracted and quantified using a ds-DNA assay (UK-Quant-iT™ PicoGreen® dsDNA Assay Kit for DNA) following the manufacturer's instructions with a Fluorescence Microplate Reader (Biotek SynergyHT, Biotek). Acceptable DNA had a concentration of at least 50ng/μl in TE (10mM Tris/1mM EDTA), was between 1.8-2.0 with an OD 260/280. WGS was performed at Illumina, Inc. (Illumina Sequencing Facility, San Diego, CA USA) or the BGI (Beijing Genome Institute, Hong Kong), as described previously, to a target depth of 50X for the cancer samples and 30X for matched controls[59].

The Burrows-Wheeler Aligner (BWA) was used to align the sequencing data to the GRCh37 reference human genome[63]. Sequencing data have been deposited at the European Genome-phenome Archive (EGAS00001000262).

Variant Calling Pipeline

Substitutions, insertions and deletions were detected using the Cancer Genome Project Wellcome Trust Sanger Institute pipeline as described previously[59]. In brief, substitutions were detected using CaVEMan with a cut-off 'somatic' probability of 95%. Post-processing filters were applied. Insertions and deletions were called using a modified version of Pindel[64]. Variant allele frequencies of all indels were corrected by local realignment of unmapped

reads against the mutant sequence. Structural variants were detected using Brass[59]. A positive ETS status was assigned if a breakpoint between *ERG*, *ETV1* or *ETV4* and previously reported partner DNA sequences was detected.

Data availability

Sequencing data that support the findings of this study have been deposited in the European Genome-phenome Archive with the accession code EGAS00001000262 (https://www.ebi.ac.uk/ega/studies/EGAS00001000262). See Supplementary Table 7 for sample specific EGA accession codes.

Code availability

The CGP pipeline may be downloaded from https://github.com/cancerit/dockstore-cgpwgs. The Battenberg pipeline may be downloaded from https://github.com/Wedge-Oxford/battenberg.

Mutation burdens

Mutation burdens were compared between primary and metastatic samples and between ADT and hormone-naïve samples using a negative binomial generalised linear model (GLM), implemented with the R package *MASS*. Sample type was found to be an independent predictor of number of SNVs, as was age at time of sampling.

Timing of copy number events

We developed a novel approach to order the occurrence of copy number aberrations by combining three sources of information:

- Clonality of copy number aberrations
- Timing relative to whole genome duplication

- Timing of homozygous deletions relative to neighboring hemizygous losses.

Information from all tumors was combined using a Bradley-Terry model, to give the most likely ordering of events during progression of PCa.

The Battenberg algorithm was used to detect clonal and subclonal somatic copy-number alterations (CNAs) and to estimate ploidy and cancer content from the next-generation sequencing data as previously described[65]. Briefly, germline heterozygous SNPs were phased using Impute2, and a- and b- alleles were assigned. Data were segmented using piecewise constant fitting[66] and subclonal copy-number segments were identified via a *t*-test as those with b-allele frequencies that differed significantly from the values expected of a clonal copy number state. Ploidy and cancer purity were estimated with the same method used by ASCAT[67].

In this cohort, we defined WGD samples as those that had an average ploidy greater than 3. For tumors that had not undergone WGD, gains were defined as those regions that had at least one allele with copy number greater than 1, while losses were defined as those segments that undergone LOH. For tumors that had undergone WGD, losses were called in those segments with at least one allele with copy number of less than 2, whereas gains were called for those with an allelic copy number greater than 2. An extension of this logic was used for subclonal copy number segments – the evolving cellular fraction was always defined as that which deviated away from overall ploidy (defined as 2 for non-WGD samples and 4 for WGD samples). For example, if 75% of cells within a non-WGD tumor have a copy number of 3 + 1 at a given genomic loci, with the remaining 25% of cells having a copy number of 2 + 1, then we assume there has been clonal gain to 2 + 1, and then a subclone containing 75% of cells has undergone a further gain.

Three independent approaches were used to extract evolutionary data from each cancer sample. The first involved ordering clustered sub-clonal cancer fractions, the second used implicit ordering of clonal HDs in relation to losses, and the third estimated the relative timing of whole genome duplication. The logical arguments used within each approach were considered in turn:

1. Battenberg algorithm-derived estimates for the cellular fraction and standard deviation of each subclonal aberration were input to a Markov Chain Monte Carlo hierarchical Bayesian Dirichlet process to group linked events together in an unsupervised manner. This defined clusters of different cell populations, each present at a calculated cancer cell fraction. The pigeonhole principle was then used to determine the hierarchical relationship between these clusters. Using this process, gains, losses and HDs were ordered with the following caveat to ensure that only independent events are ordered: if there was a clonal and subclonal gain (or loss) at the same locus, then only the clonal or initial gain (or loss) was ordered.

2. Homozygous deletions have implicitly occurred after loss of heterozygosity at the same locus.

3. The parsimony principle was used to define the allele counts that correspond to early and late changes in relation to WGD. For losses, if the minor allele copy number equals 0, then the loss occurred prior to WGD. Otherwise the loss occurred after WGD. Regarding gains, if the major allele copy number is twice or greater than ploidy, then the gain occurred prior to WGD. Otherwise, the gain occurred after WGD.

The above arguments allow us to gain insights into the order of copy number events within each individual tumor sample. To establish a consensus order across a cohort of tumor samples requires the ordering data to be integrated across all samples. As specific copy

number events (location of breakpoints and the individual copy number states) tend to be unique to individual samples, we defined reference copy number segments that occurred recurrently. These were then used to build an overall contingency table.

The reference genomic segments were defined as regions that were recurrently aberrant. Regions of significant recurrence (false detection rate (FDR), $P < 0.05$) were determined by performing 100,000 simulations, placing the copy number aberrations detected from each sample in random locations within the genome. The process was repeated for gains, LOH and HDs and the randomly generated copy number landscape compared to that arising from this cohort provided significance levels. Each significantly aberrant region was initially segmented using all breakpoints from all the events that contributed to that region. For instance, the significantly enriched region for LOH: chr8: 0-44Mb contains over 300 breakpoints drawn from from all the samples which contain LOH at chromosome 8p. We computed significantly recurrent regions and reference segments for both ETS+ and ETS- sample subgroups.

Performing pair-wise comparisons between all segmented results using the Bradley-Terry method described below proved computationally expensive and therefore the total number of segments used in the pairwise comparison was rationalised by grouping reference segments to make combined segments of minimum length 1 MB.

We then considered each tumor sample in turn. If any copy number event overlapped the reference genomic segments and was ordered in relation to any other event (that also overlapped regions of significance), those overlapped reference segments were ordered in comparison to other overlapped reference segments. In addition to these reference segments, the TMRPSS2-ERG deletion was ordered more stringently by considering only those segments that could result in the gene fusion, and not merely overlap the locus. In this

manner, a contingency table of contests was constructed, using reference genomic segments as the variables. We built contingency tables for both ETS+ and ETS- tumor samples to determine whether their evolutionary trajectory differed significantly. .

An implementation of the Bradley-Terry model of pairwise comparison in R[68] with bias reduced maximum likelihood estimated the ability or overall order of each individual reference segment.

Subclonal Analysis

The fraction of each cancer genome with subclonal copy number aberrations was calculated as the total amount of the genome with subclonal CNA, as identified by the Battenberg algorithm, divided by the total amount of the genome that had copy number aberrations. One sample (PD13397a, Supplementary Table 8) was identified as having very low cellularity, as it had a completely flat copy number profile and only 411 identified SNVs. Since CNAs could not called in this sample, it was not possible to adjust allele frequencies to CCFs and this sample was excluded from subclonality analysis. Point substitutions and indels were separately clustered using a Bayesian Dirichlet process, as previously described[47]. Clonal variants are expected to cluster at a CCF close to 1.0. However, in 18 tumors (Supplementary Table 8), there was no cluster in the range [0.95,1.05]. The likely cause of a shift in CCF is inaccuracy in copy number calling and these samples therefore failed quality control and were excluded from subclonality analysis. From Markov Chain Monte Carlo sampling carried out within the Dirichlet process model, the posterior probability of each variant having a CCF below 0.95 was estimated. Variants with a probability above 80% were designated as ‘subclonal’, those with probability below 20% were designated ‘clonal’ and those with intermediate probabilities were designated as ‘uncertain’. The fraction of subclonal variants

used in Fig. 5 and Supplementary Fig. 2 was then calculated after excluding uncertain variants.

Mutational Spectra

The mutational spectra, defined by the triplets of nucleotides around each mutation of each sample were deconvoluted into mutational processes as previously described[48,69]. Clonal and subclonal variants were separated, as defined above. Further separation of clonal mutations was performed for mutations in genomic regions that had undergone copy number gains. These mutations were classified as 'early' or 'late' depending whether their observed allele frequencies were more likely to indicate their presence on 2 or 1 chromosome copies, respectively, as assessed by binomial probability. Assignment of mutations to mutational signatures was carried out on each subset of mutations (early, late, clonal, subclonal), as well as on all mutations from each sample (Supplementary Table 3).

Clinical survival analyses

A Cox regression model was fitted to 71 features: every gene with mutations (breakpoints, subs or indels) with a potential functional impact (missense, nonsense, start-lost, inframe, frameshift, or occurred in a non-coding transcript) or a CNA highlighted by the copy number aberration analysis that occurred in three of more prostatectomy patients. The endpoint was biochemical recurrence. *P*-values were adjusted for multiple testing using the Benjamini-Hochberg method. Multivariate analyses were performed on all genes found to be significant using discretised Gleason (6, 7, 8 or 9), pathological T-stage (T2, T3) and PSA at prostatectomy as cofactors. Gene selection for the optimal predictor of time to biochemical recurrence was determined using Lasso[70], a shrinkage and selection method for linear regression, starting with all genes that had a significant association with time to biochemical recurrence. Standard algorithms were used for survival analyses and statistical associations.

Identifying novel oncogenes

The joint dataset was compiled from the aggregation of variants called within our samples with 3 other datasets, yielding a total of 930 samples, comprised of 710 primary and 220 metastatic samples:

- **TCGA**[4], 425 primary cancer samples, whole exome sequencing with SureSelect Exome v3 baits on Illumina HiSeq 2000, average coverage ~100X
- **COSMIC database**[22], 243 samples, curated set of mutations from several sources, http://cancer.sanger.ac.uk/cosmic
- **Stand Up to Cancer**[23] (SU2C-PCF), 150 metastatic castrate resistant samples, paired-end, whole exome sequencing with SureSelect Exome v4 baits on Illumina HiSeq2000, average coverage ~160X

To identify coding and non-coding drivers from substitutions and indels, we used two previously described methods[50]. Coding drivers on the joint dataset (930 cancers) were identified using dNdScv, a dN/dS method designed to quantify positive selection in cancer genomes. dNdScv models somatic mutations in a given gene as a Poisson process. Inferences on selection are carried out separately for missense substitutions, truncating substitutions (nonsense and essential splice site mutations) and indels, and then combined into a global P-value per gene. Non-coding recurrence was studied using NBR. Both dNdScv and NBR model the variation of the mutation rate across the genome using a negative binomial regression with covariates. First, Poisson regression is used to obtain maximum-likelihood estimates for the 192 rate parameters (r_j) describing each of the possible trinucleotide substitutions in a strand-specific manner. $r_j = n_j/L_j$, where n_j is the total number of mutations observed across samples of a given trinucleotide class (j) and L_j is the number of available sites for each trinucleotide. These rates are used to estimate the total number of mutations

across samples expected under neutrality in each element considering the mutational signatures active in the cohort and the sequence of the elements ($E_h = \Sigma_j r_j L_{j,h}$). This estimate assumes no variation of the mutation rate across elements in the genome. Second, a negative binomial regression is used to refine this estimate of the background mutation rate of an element, using covariates and E_h as an offset. Both methods identify genes or non-coding regions with higher than expected mutation recurrence, correcting for gene length, sequence composition, mutation signatures acting across patients and for the variation of the mutation rate along the genome. A QQ-plot confirmed that *P*-values obtained from this method in this cohort were not subject to inflation and consequent over-calling of driver genes (Supplementary Fig. 6).

Chromoplexy, characterized by highly clustered genomic breakpoints that occur in chains and are sometimes joined by deletion bridges, has been shown to be prevalent in PCa[25]. To identify rearrangement drivers, we first used ChainFinder[25] to account for any bias towards regions with chromoplexy and identified 'unique' rearranged regions per sample taking the mid-point between all the breakpoints ChainFinder assigns to the same chromoplexy event. Next, separately aggregating the ICGC samples with and without ERG fusions, we calculated inter-breakpoint distance and performed piecewise constant fitting (PCF)[66] to identify genomic regions which were recurrently rearranged in multiple samples. Rearranged regions with potential functional impact were identified using two criteria: a minimum 3-fold difference in the number of SVs per MB of ERG+ and ERG- samples; region contains at least one gene with multiple samples with truncating events, i.e. homozygous deletion, stop codon, frameshift indel or essential splice site mutation. In addition, several identified regions were significantly enriched for LOH in either ETS+ or ETS- samples, from copy number analysis (see above). The variants identified in key regions are depicted in Fig. 3.

Chemogenomics annotation of the prostate cancer network

To construct the network, we used the 82 protein products of the genes identified in this study (hereon referred to as Prostate Proteins) to seed a search for all possible interacting proteins in the canSAR interactome[54]. This interactome contains merged and curated data from the IMeX consortium[71], Phosphosite (http://www.phosphosite.org/), and other databases. It includes:

1) interactions where there were more than two publications reporting experiments demonstrating the binary interaction between the two proteins

2) interactions where there is 3D protein structural evidence of a direct complex

3) interactions where there are at least two publications reporting that one protein is a substrate of the other

4) interactions where there are at least two papers reporting that one protein is the product of a gene under the direct regulatory control of the other

It excludes the following:

A) interactions that were inferred from a large immunoprecipitation experiment without follow-up to demonstrate the specific binary interaction

B) interactions inferred from text mining

C) interactions inferred from co-occurrence in publications or from gene expression correlation.

The initial prostate cancer seeded network resulted in a large collection of 3366 proteins that have some experimental evidence of interacting with at least one Prostate Protein. When we added extra proteins into the network, we wanted to ensure that we only add proteins that are

more likely to function primarily through interaction with the proteins in the network rather than just be generic major hubs. To this end, we carried out the following steps: Starting with the input (prostate protein) list, we obtained all possible first neighbours. We then computed, for each new protein, the proportion of its first neighbours that are in the original input list. To define the proteins that are most likely to function through our network, we calculated the chances of these proportions occurring in a random network. We did this by randomising our interactome 10,000 times and computing how often the observed proportions can be achieved by chance (empirical p-value). We corrected the p-values for multiple testing and retained only proteins that have corrected FDR p-values less than 0.05. (Supplementary Fig. 5). We performed network minimisation to maintain only proteins that are strongly connected to more than one Prostate Protein or whose only connection is to one of the Prostate Proteins. We identified a Prostate Cancer network of 163 proteins. Using canSAR's Cancer Protein Annotation Tool (CPAT)[72], we annotated the 163 proteins with pharmacological and druggability data. We labelled proteins that are: 1) targets of approved drugs; 2) targets of drugs under clinical investigation, 3) targets of preclinical or discovery stage compounds that are active at concentrations equal to or less than 100 nM against the protein of interest 4) proteins that we predict to be druggable using our structural druggability prediction protocols[72-75] but that have few or no published active inhibitors – these are potential targets for future drug discovery.

ACKNOWLEDGEMENTS

We acknowledge support from Cancer Research UK C5047/A22530, C309/A11566, C368/A6743, A368/A7990, C14303/A17197 and the Dallaglio Foundation, The NIHR support to The Biomedical Research Centre at The Institute of Cancer Research and The

Royal Marsden NHS Foundation Trust; Cancer Research UK funding to The Institute of Cancer Research and the Royal Marsden NHS Foundation Trust CRUK Centre; the National Cancer Research Institute (National Institute of Health Research (NIHR) Collaborative Study: “Prostate Cancer: Mechanisms of Progression and Treatment (PROMPT)” (grant G0500966/75466); the Li Ka Shing foundation (DCW, DJW). The Academy of Finland and Cancer Society of Finland (GSB). We thank the National Institute for Health Research, Hutchison Whampoa Limited, University of Cambridge, the Human Research Tissue Bank (Addenbrooke’s Hospital) which is supported by the NIHR Cambridge Biomedical Research Centre, The Core Facilities at the Cancer Research UK Cambridge Institute, Orchid and Cancer Research UK, Dave Holland from the Infrastructure Management Team & Peter Clapham from the Informatics Systems Group at the Wellcome Trust Sanger Institute. DMB is supported by Orchid. We also acknowledge support from The Bob Champion Cancer Research Trust, The Masonic Charitable Foundation and The King Family. PW is a Cancer Research Life Fellow. We acknowledge core facilities provided by CRUK funding to the CRUK ICR Centre, the CRUK Cancer Therapeutics Unit and support for canSAR C35696/A23187 and NIHR finding to the Biomedical Research Centre at Royal Marsden NHS Foundation Trust and Institute of Cancer Research. The authors would like to thank those men with prostate cancer and the subjects who have donated their time and their samples to the Cambridge, Oxford, The Institute of Cancer Research, John Hopkins, and University of Tampere BioMediTech Biorepositories for this study. We also would like to acknowledge support of the research staff in S4 who so carefully curated the samples and the follow-up data (Jo Burge, Marie Corcoran, Anne George, and Sara Stearn).

FIGURE LEGENDS

Figure 1. Mutational landscape of prostate cancers. From top-to-bottom: mutation status of DNA repair genes, ETS fusion status and sample type; proportion of mutations assigned to each signature[48]; number of somatic point substitutions identified in each sample; proportion of small insertions/deletions associated with microhomology or repetitive regions; number of insertions, deletions and complex insertions/deletions in each sample; total number of structural variants in each sample, separated into inversions, translocations, deletions and tandem duplications. Sample ordering is reported in Supplementary Table 7.

Figure 2. Landscape of driver genes in prostate cancer. Genes were identified using three different methods: upper panel shows genes that have undergone genetic aberration in at least 6 samples; middle panel shows genes with aberrations enriched in either ERG+ or ERG- cancers (Fisher exact test for *PTEN*, *TP53*, *SPOP*, 3p13, *PDE4D*, *PPAP2A*; *ROBO1* and *ROBO2* are in a region enriched for SVs in ETS- tumors; *IL6ST* is in a region enriched for SVs in ETS+ tumors); lower panel shows genes enriched in metastatic samples (Fisher exact test). DDR = DNA damage response, 'hemi.loss' = loss of heterozygosity resulting from copy number change, 'homo.loss' = homozygous deletion resulting from copy number aberration, 'two allele loss + sub/indel' indicates genes in triploid regions bearing aberrations of all 3 gene copies. Sample ordering is reported in Supplementary Table 7.

Figure 3. Putative novel driver genes. Putative drivers are shown in red and genomic aberrations are displayed as: missense SNVs – circles; nonsense SNVs – open triangles; essential splice site mutations – open squares; indels – closed squares; non-coding mutations – closed triangles; simple SV - yellow cross; chromoplexy event – blue cross; region enriched for loss of heterozygosity, with height proportional to the number samples containing LOH -

pink shading; region enriched for homozygous deletions, with height proportional to the number of samples containing homozygous deletion – blue shading.

Figure 4. Temporal evolution of copy number aberrations in ETS+ and ETS- prostate cancer. For (a) ETS+ cancers, and (b) ETS- cancers: Left: The landscape of copy number aberrations with genomic loci plotted against fraction of cancers. Loss-of-heterozygosity is depicted in blue, homozygous deletions in black, gains in red, *TMPRSS2-ERG* deletion in brown and whole genome duplication in green. Right: The temporal evolution of significantly recurrent ($p < 0.05$) copy number aberrations by genomic loci over time (mean with 95% confidence intervals, log precedence relative to arbitrary reference). Lower values indicate earlier events (c) Pairwise associations among copy number aberrations. Recurrently aberrant regions with a false discovery rate < 0.1 are shown. Associations are indicated by odds ratio (OR) with brown colors depicting mutually exclusive events and blue-green colors depicting correlated events. Genomic loci annotated by: type of aberration (G=gain, L=loss, HD=homozygous deletion); chromosome; median position in Mb. For focal events the putative target genes are annotated.

Figure 5. Heterogeneity and subclonal mutations. (a) Metastatic tumors have less heterogeneity than primary tumors, whether assessed from SNVs or indels. Each dot represents a different sample, colored by sample type. x-axis = fraction of point substitutions that are subclonal, y-axis = fraction of indels that are subclonal, contour lines calculated using R package kde2d. (b) Samples with multiple subclonal mutations in driver genes. Fraction of cancer cells carrying mutation is shown as grey histogram for all mutations and as red ovals for mutations in known driver genes. Mutations are clustered using a Dirichlet process as previously described[47], with thick plum-colored lines indicating fitted distribution and pale blue regions indicating 95% posterior confidence intervals. Peaks with a subclonal fraction close to 1 are clonal, whereas peaks at lower subclonal fractions indicate subclonal mutations.

Figure 6. Clinical outcome. Kaplan-Meier plots for biochemical recurrence. Kaplan-Meier plots of recurrent mutated genes where there is a significant correlation with time to biochemical recurrence after prostatectomy ($P < 0.05$; Cox regression model; Benjamin-Hochberg multiple testing correction). Clinical information was available for 89 prostatectomy samples with WGS data, with a median follow up of 1108 days in which biochemical recurrence occurred in 26 patients. The mutations in both genes consisted of a frameshift deletion in one sample and structural variants in the remaining samples.

Table 1. Putative driver genes. Genes were identified in our study using several methods, detailed in the last column: dN/dS; enrichment for SVs or CNAs in ETS+ or ETS- cancers; enrichment for truncating mutations or homozygous deletions, clinical correlation. From a PubMed literature search, prior evidence for each gene being a driver of prostate cancer was classified as 'low' if the gene has not been previously reported as playing a role in prostate cancer tumorigenesis or progression. Isolated alterations may have been observed or biological evidence for importance may have been presented as indicated in the prior evidence column. Prior evidence was classified as 'medium' for genes reported previously as playing a role in prostate carcinogenesis or progression but currently lacking statistical support based on genetic alterations. Evidence considered included presence of multiple genetic alterations, SNP associations, and known cancer genes in other tissues. The high confidence genes are those that are widely accepted to represent cancer genes and to be altered in prostate cancer: this would include genes such as *HRAS*, *SPOP*, *IDH1* etc. In each case there are two or more of the following: statistical verification of higher incidence, biological experiments, clinical correlations, confirmation in multiple studies, recognition as cancer genes in other cancer types. dN/dS = non-synonymous: synonymous ratio, calculated for all SNVs and indels; dN/dS (missesnse) = non-synonymous: synonymous ratio, calculated for missense SNVs only; SV = structural variant; CNA = copy number aberration; SNV =

single nucleotide variant; indel = small insertion/deletion; ETS = E26 transformation-specific.

Table 2. Drug targets identified from CanSAR analysis. Proteins in bold typeface are derived from genes identified as prostate drivers in this study or proteins that have a significant known interaction with these proteins.

SUPPLEMENTARY FIGURE AND TABLE LEGENDS

Supplementary Figure 1. RNA expression of novel driver genes. Data are taken from the CamCaP study, which includes 40 samples from this study. All genes identified as novel within this study and bearing structural variants in at least 2 samples are shown. P-values are from Wilcoxon rank sum test.

Supplementary Figure 2. Heterogeneity of point substitutions and copy number aberrations. Each dot represents a different sample, colored by sample type. x-axis = amount of each genome that has subclonal CNA divided by the total amount of each genome that is copy-number aberrant, y-axis = fraction of point substitutions that are subclonal, contour lines calculated using R package kde2d.

Supplementary Figure 3. Multiple driver mutations in *APC*. (a) Mutations in *APC* gene in PD14713a are mutually exclusive, indicating that they have occurred in separate subclonal populations. Each blue or yellow string represents a forward or backward read, with somatic variants shown in red. Mutations in the left-hand and right-hand red boxed regions never occur on the same reads. (b) PD14713a is haploid on chromosome 5q, so the mutual exclusivity of APC mutations cannot be explained by occurrence in different chromosome

copies. Purple lines show total copy number, blue lines show minor allele specific copy number, called by the Battenberg algorithm.

Supplementary Figure 4. Prognostic biomarkers. (a) Kaplan-Meier plot of the number of mutational processes detected vs time to biochemical recurrence (b) Correlation between the number of processes detected and the number of substitutions detected. There was a significant positive correlation (r = 0.31, *P* = 0.0027; Spearman's correlation). The number of mutational signatures identified in a cancer was negatively correlated with time to biochemical recurrence in prostatectomy patients (*P* = 0.014, HR = 3.0; Cox proportional hazards model on number of processes greater than three. The number of substitutions detected was also an independent prognostic biomarker (*P* = 0.031, HR = 1.0005; Cox proportional hazards model).

Supplementary Figure 5. Processes used in the canSAR analysis.

Supplementary Figure 6. dN/dS analysis. QQ plot of the *P*-values derived from dN/dS analysis show no sign of inflation. Four genes (of 20,184) had p-values = 0 and are not shown on this plot.

Supplementary Table 1. Summary of sample specific genetic aberrations. Worksheet 1 reports the number of genetic aberrations identified in each gene this study, separated by type: essential splice site, frameshift, inframe, missense, nonsense, silent, stop-lost, homozygous deletion, rearrangement. We also report the number of samples with multiple hits resulting in homozygous loss. Worksheet 2 reports the number of genetic aberrations identified in each gene across the joint dataset. Worksheet 3 reports the samples within our dataset that bear homozygous losses in each gene. Worksheet 4 contains results of dNdScv analysis to identify coding drivers. Novel drivers are shown with red text. Significant q-values (FDR < 0.1) from analysing either missense SNVs or all SNVs and indels are shown

with green shading. Effect sizes (dN/dS values) are reported separately for missense, nonsense, essential splice site and indel variants. Worksheet 5 contains results of NBR analysis to identify non-coding drivers. Significant regions (FDR < 0.1) are shown with blue / purple shading.

Supplementary Table 2. Classification of Driver Genes. See Table 1.

Supplementary Table 3. Mutational signature analysis. For each sample, the number of somatic mutations attributed to each signature is shown in worksheet 1. Worksheet 2 reports the number of mutations assigned to each signature, broken down into temporal categories (early, late, clonal, subclonal) as described in Online Methods.

Supplementary Table 4. Survival analyses. Cox regression model results for the presence of recurrent in genes with time to biochemical recurrence after prostatectomy as endpoint. Multivariate analysis was performed taking into account cofactors Gleason (6-9), PSA at prostatectomy, and pathological T-stage (T2, T3). Clinical information was available for 89 prostatectomy samples with WGS data, with a median follow up of 1108 days in which biochemical recurrence occurred in 26 patients. Red background shading indicates features that have a significant association with outcome in both univariate analyses after multiple testing correction and multivariate analyses. Dark grey shading indicates features that are only significant without multiple testing correction. Light grey shading indicates features that are significant in univariate but not in multivariate analyses.

Supplementary Table 5. Druggability analysis. Genes identified using the CanSAR software. Genes are color-coded: bright green = target of an approved drug; dark green = target of an investigational drug; yellow = target that is being investigated chemically; red = no chemical information in public databases, but predicted to be druggable using our structure-based method.

Supplementary Table 6. Drug sensitivity data. Of the drugs identified through CanSAR analysis, 18 are reported in the Genomics of Drug Sensitivity in Cancer database. Of these, 5 showed significant effect on growth inhibition in at least one cell line, and all 18 showed weak activity in at least one cell line.

Supplementary Table 7. Clinical and Molecular Details of Cancers Subject to DNA Sequencing. The order of samples in Fig. 1 and Fig. 2 are in the columns *fig1_order* and *fig2_order*.

Supplementary Table 8. Details of samples included and excluded from the subclonal analysis.

REFERENCES

1. Attard, G. *et al.* Prostate cancer. *Lancet* **387**, 70-82 (2016).
2. Weischenfeldt, J. *et al.* Integrative genomic analyses reveal an androgen-driven somatic alteration landscape in early-onset prostate cancer. *Cancer Cell* **23**, 159-70 (2013).
3. Grasso, C.S. *et al.* The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487**, 239-43 (2012).
4. Cancer Genome Atlas Research, The Molecular Taxonomy of Primary Prostate Cancer. *Cell* **163**, 1011-25 (2015).
5. Barbieri, C.E. *et al.* Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet* **44**, 685-9 (2012).
6. Berger, M.F. *et al.* The genomic complexity of primary human prostate cancer. *Nature* **470**, 214-20 (2011).
7. Lalonde, E. *et al.* Tumour genomic and microenvironmental heterogeneity for integrated prediction of 5-year biochemical recurrence of prostate cancer: a retrospective cohort study. *Lancet Oncol* **15**, 1521-32 (2014).
8. Cooper, C.S., Eeles, R., Wedge, D.C. & Van Loo, P. Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. **47**, 367-72 (2015).
9. Boutros, P.C. *et al.* Spatial genomic heterogeneity within localized, multifocal prostate cancer. *Nat Genet* **47**, 736-45 (2015).
10. Gundem, G. *et al.* The evolutionary history of lethal metastatic prostate cancer. *Nature* **520**, 353-7 (2015).
11. Castro, E. *et al.* Effect of BRCA Mutations on Metastatic Relapse and Cause-specific Survival After Radical Treatment for Localised Prostate Cancer. *Eur Urol* **68**, 186-93 (2015).
12. Kluth, M. *et al.* Concurrent deletion of 16q23 and PTEN is an independent prognostic feature in prostate cancer. *Int J Cancer* **137**, 2354-63 (2015).

13. Mosquera, J.M. *et al.* Concurrent AURKA and MYCN gene amplifications are harbingers of lethal treatment-related neuroendocrine prostate cancer. *Neoplasia* **15**, 1-10 (2013).
14. Rodrigues, L.U. *et al.* Coordinate loss of MAP3K7 and CHD1 promotes aggressive prostate cancer. *Cancer Res* **75**, 1021-34 (2015).
15. Cuzick, J. *et al.* Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study. *Lancet Oncol* **12**, 245-55 (2011).
16. Klein, E.A. *et al.* Decipher Genomic Classifier Measured on Prostate Biopsy Predicts Metastasis Risk. *Urology* **90**, 148-52 (2016).
17. Bostrom, P.J. *et al.* Genomic Predictors of Outcome in Prostate Cancer. *Eur Urol* **68**, 1033-44 (2015).
18. Luca, B.-A. *et al.* DESNT: A Poor Prognosis Category of Human Prostate Cancer. *European Urology Focus*.
19. Ryan, C.J. *et al.* Abiraterone acetate plus prednisone versus placebo plus prednisone in chemotherapy-naive men with metastatic castration-resistant prostate cancer (COU-AA-302): final overall survival analysis of a randomised, double-blind, placebo-controlled phase 3 study. *Lancet Oncol* **16**, 152-60 (2015).
20. Loriot, Y. *et al.* Effect of enzalutamide on health-related quality of life, pain, and skeletal-related events in asymptomatic and minimally symptomatic, chemotherapy-naive patients with metastatic castration-resistant prostate cancer (PREVAIL): results from a randomised, phase 3 trial. *Lancet Oncol* **16**, 509-21 (2015).
21. Mateo, J. *et al.* DNA-Repair Defects and Olaparib in Metastatic Prostate Cancer. *N Engl J Med* **373**, 1697-708 (2015).
22. James, N.D. *et al.* Addition of docetaxel, zoledronic acid, or both to first-line long-term hormone therapy in prostate cancer (STAMPEDE): survival results from an adaptive, multiarm, multistage, platform randomised controlled trial. *Lancet* **387**, 1163-77 (2016).
23. Forbes, S.A. *et al.* COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res* **43**, D805-11 (2015).
24. Robinson, D. *et al.* Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215-28 (2015).
25. Baca, S.C. *et al.* Punctuated evolution of prostate cancer genomes. *Cell* **153**, 666-77 (2013).
26. Svensson, C. *et al.* REST mediates androgen receptor actions on gene repression and predicts early recurrence of prostate cancer. *Nucleic Acids Res* **42**, 999-1015 (2014).
27. Liu, Z. *et al.* CASZ1, a candidate tumor-suppressor gene, suppresses neuroblastoma tumor growth through reprogramming gene expression. *Cell Death Differ* **18**, 1174-83 (2011).
28. Fischer, K. & Pflugfelder, G.O. Putative Breast Cancer Driver Mutations in TBX3 Cause Impaired Transcriptional Repression. *Front Oncol* **5**, 244 (2015).
29. De Keersmaecker, K. *et al.* Exome sequencing identifies mutation in CNOT3 and ribosomal genes RPL5 and RPL10 in T-cell acute lymphoblastic leukemia. *Nat Genet* **45**, 186-90 (2013).
30. Sasaki, M. *et al.* Regulation of the MDM2-P53 pathway and tumor growth by PICT1 via nucleolar RPL11. *Nat Med* **17**, 944-51 (2011).
31. Chakravarty, D. *et al.* The oestrogen receptor alpha-regulated lncRNA NEAT1 is a critical modulator of prostate cancer. *Nat Commun* **5**, 5383 (2014).
32. Yang, Y.A. & Yu, J. Current perspectives on FOXA1 regulation of androgen receptor signaling and prostate cancer. *Genes Dis* **2**, 144-151 (2015).
33. Takayama, K. *et al.* Integrative analysis of FOXP1 function reveals a tumor-suppressive effect in prostate cancer. *Mol Endocrinol* **28**, 2012-24 (2014).
34. Krohn, A. *et al.* Recurrent deletion of 3p13 targets multiple tumour suppressor genes and defines a distinct subgroup of aggressive ERG fusion-positive prostate cancers. *J Pathol* **231**, 130-41 (2013).

35. Carver, B.S. *et al.* Aberrant ERG expression cooperates with loss of PTEN to promote cancer progression in the prostate. *Nat Genet* **41**, 619-24 (2009).
36. King, J.C. *et al.* Cooperativity of TMPRSS2-ERG with PI3-kinase pathway activation in prostate oncogenesis. *Nat Genet* **41**, 524-6 (2009).
37. Kluth, M. *et al.* Clinical significance of different types of p53 gene alteration in surgically treated prostate cancer. *Int J Cancer* **135**, 1369-80 (2014).
38. Burkhardt, L. *et al.* CHD1 is a 5q21 tumor suppressor required for ERG rearrangement in prostate cancer. *Cancer Res* **73**, 2795-805 (2013).
39. Liu, W. *et al.* Identification of novel CHD1-associated collaborative alterations of genomic structure and functional assessment of CHD1 in prostate cancer. *Oncogene* **31**, 3939-48 (2012).
40. Biankin, A.V. *et al.* Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* **491**, 399-405 (2012).
41. Heun, P. SUMOrganization of the nucleus. *Curr Opin Cell Biol* **19**, 350-5 (2007).
42. Kaikkonen, S. *et al.* SUMO-specific protease 1 (SENP1) reverses the hormone-augmented SUMOylation of androgen receptor and modulates gene responses in prostate cancer cells. *Mol Endocrinol* **23**, 292-307 (2009).
43. Smith, D.I., Zhu, Y., McAvoy, S. & Kuhn, R. Common fragile sites, extremely large genes, neural development and cancer. *Cancer Lett* **232**, 48-57 (2006).
44. Taylor, B.S. *et al.* Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11-22 (2010).
45. Williams, J.L., Greer, P.A. & Squire, J.A. Recurrent copy number alterations in prostate cancer: an in silico meta-analysis of publicly available genomic data. *Cancer Genet* **207**, 474-88 (2014).
46. Chen, Z. *et al.* Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis. *Nature* **436**, 725-30 (2005).
47. Bolli, N. *et al.* Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat Commun* **5**, 2997 (2014).
48. Alexandrov, L.B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415-21 (2013).
49. Pilati, C. *et al.* Mutational signature analysis identifies MUTYH deficiency in colorectal cancers and adrenocortical carcinomas. *J Pathol* (2017).
50. Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47-54 (2016).
51. Polkinghorn, W.R. *et al.* Androgen receptor signaling regulates DNA repair in prostate cancers. *Cancer Discov* **3**, 1245-53 (2013).
52. Goodwin, J.F. *et al.* DNA-PKcs-Mediated Transcriptional Regulation Drives Prostate Cancer Progression and Metastasis. *Cancer Cell* **28**, 97-113 (2015).
53. Tarish, F.L. *et al.* Castration radiosensitizes prostate cancer tissue by impairing DNA double-strand break repair. *Sci Transl Med* **7**, 312re11 (2015).
54. Tym, J.E. *et al.* canSAR: an updated cancer research and drug discovery knowledgebase. *Nucleic Acids Res* **44**, D938-43 (2016).
55. Leongamornlert, D. *et al.* Frequent germline deleterious mutations in DNA repair genes in familial prostate cancer cases are associated with advanced disease. *Br J Cancer* **110**, 1663-72 (2014).
56. Yang, W. *et al.* Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* **41**, D955-61 (2013).
57. Seigne, C. *et al.* Characterisation of prostate cancer lesions in heterozygous Men1 mutant mice. *BMC Cancer* **10**, 395 (2010).
58. Malik, R. *et al.* Targeting the MLL complex in castration-resistant prostate cancer. *Nat Med* **21**, 344-52 (2015).

59. Cooper, C.S. *et al.* Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. *Nat Genet* (2015).
60. Mao, X. *et al.* Distinct genomic alterations in prostate cancers in Chinese and Western populations suggest alternative pathways of prostate carcinogenesis. *Cancer Res* **70**, 5207-12 (2010).
61. Liu, W. *et al.* Copy number analysis indicates monoclonal origin of lethal metastatic prostate cancer. *Nat Med* **15**, 559-65 (2009).
62. Nickerson, M.L. *et al.* Somatic alterations contributing to metastasis of a castration-resistant prostate cancer. *Hum Mutat* **34**, 1231-41 (2013).
63. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-95 (2010).
64. 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**, 2865-71 (2009).
65. Nik-Zainal, S. *et al.* The life history of 21 breast cancers. *Cell* **149**, 994-1007 (2012).
66. Nilsen, G. *et al.* Copynumber: Efficient algorithms for single- and multi-track copy number segmentation. *BMC Genomics* **13**, 591 (2012).
67. Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci U S A* **107**, 16910-5 (2010).
68. Firth, D. & Turner, H.L. Bradley-Terry models in R: the BradleyTerry2 package. *Journal of Statistical Software* **48**(2012).
69. 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**, 246-59 (2013).
70. Shin, S., Fine, J. & Liu, Y. Adaptive Estimation with Partially Overlapping Models. *Stat Sin* **26**, 235-253 (2016).
71. Orchard, S. *et al.* Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. *Nat Methods* **9**, 345-50 (2012).
72. Patel, M.N., Halling-Brown, M.D., Tym, J.E., Workman, P. & Al-Lazikani, B. Objective assessment of cancer genes for drug discovery. *Nat Rev Drug Discov* **12**, 35-50 (2013).
73. Bulusu, K.C., Tym, J.E., Coker, E.A., Schierz, A.C. & Al-Lazikani, B. canSAR: updated cancer research and drug discovery knowledgebase. *Nucleic Acids Res* **42**, D1040-7 (2014).
74. Mitsopoulos, C., Schierz, A.C., Workman, P. & Al-Lazikani, B. Distinctive Behaviors of Druggable Proteins in Cellular Networks. *PLoS Comput Biol* **11**, e1004597 (2015).
75. Workman, P. & Al-Lazikani, B. Drugging cancer genomes. *Nat Rev Drug Discov* **12**, 889-90 (2013).
76. Rudnicka, C. *et al.* Overexpression and knock-down studies highlight that a disintegrin and metalloproteinase 28 controls proliferation and migration in human prostate cancer. *Medicine (Baltimore)* **95**, e5085 (2016).
77. Zhang, H. *et al.* FOXO1 inhibits Runx2 transcriptional activity and prostate cancer cell migration and invasion. *Cancer Res* **71**, 3257-67 (2011).
78. Malinowska, K. *et al.* Interleukin-6 stimulation of growth of prostate cancer in vitro and in vivo through activation of the androgen receptor. *Endocr Relat Cancer* **16**, 155-69 (2009).
79. FitzGerald, L.M. *et al.* Identification of a prostate cancer susceptibility gene on chromosome 5p13q12 associated with risk of both familial and sporadic disease. *Eur J Hum Genet* **17**, 368-77 (2009).
80. Zhao, W., Cao, L., Zeng, S., Qin, H. & Men, T. Upregulation of miR-556-5p promoted prostate cancer cell proliferation by suppressing PPP2R2A expression. *Biomed Pharmacother* **75**, 142-7 (2015).

81. Parray, A. *et al.* ROBO1, a tumor suppressor and critical molecular barrier for localized tumor cells to acquire invasive phenotype: study in African-American and Caucasian prostate cancer models. *Int J Cancer* **135**, 2493-506 (2014).
82. Daniels, G. *et al.* TBLR1 as an androgen receptor (AR) coactivator selectively activates AR target genes to inhibit prostate cancer growth. *Endocr Relat Cancer* **21**, 127-42 (2014).
83. Jones, S. *et al.* Somatic mutations in the chromatin remodeling gene ARID1A occur in several tumor types. *Hum Mutat* **33**, 100-3 (2012).
84. Collart, M.A., Kassem, S. & Villanyi, Z. Mutations in the NOT Genes or in the Translation Machinery Similarly Display Increased Resistance to Histidine Starvation. *Front Genet* **8**, 61 (2017).

Author contribution

RAE, CSC, DN, DSB, CSF, SB, AGL, PW, BA, DCW, FCH and DFE designed the study.

RAE, CSC, DCW and DB wrote the paper, and all other authors contributed to revisions.

ZKJ, HW, CEM, DN, VG, AGL, RAE, FCH, SB, AYW, CSF, CV, DMB, ND, SM, SH, WH, Y-JL, AL, JK, KK, HL, LM, SE, LMatthews, AN, YY, HZ, ST, EB, CF, NL, SH, DNicol, PG, VK, NVA, PK, CO, DC, AT, EM, ER, TD, NCS, coordinated sample collection, pathology review and processing.

DCW, GG, TM, IM, DJW, DSB, MG, JZ, AB, LGB, SD, BK, NC, VB, DL, SM, TD, MA, STavare, CG, KR, DG, AM, LS, JT, AF, UM, supported, directed and performed the analyses.

CS and the TCGA, JdeB and GA provided data for the meta-analysis.

TABLE 1

gene	Mutation type(s)	Previous evidence	Prior evidence	Evidence in our study
ADAM28	SV, CNA	low	[76]biological evidence	SVs and CNA in ETS+
ANTXR2	SV, SNV/indel	low	none	clinical correlation
ASH1L	SV, SNV/indel	low	[25]	truncating mutations, SVs in ETS-
CDH12	SV	low	none	clinical correlation
FOXO1	CNA	low	[77]biological evidence	CNA in ETS-
IL6ST	SV	low	[78]biological evidence	dN/dS, SVs and CNA in ETS+, clinical correlation
LCE2B	SNV/indel	low	none	dN/dS (missense)
MAP3K1	SV, CNA	low	none	SVs, CNA in ETS+
MYST3	SV	low	[25]	SVs in ETS-, RNA expression
NCOA7	SV	low	none	SVs in ETS-
NDST4	SNV/indel	low	none	dN/dS (missense)
NEAT1	non-coding	low	[31]biological evidence	non-coding
PDE4D	SV	low	[79]SNP data	SVs and CNA in ETS+
PPAP2A	SV	low	[79]SNP data	SVs and CNA in ETS+
PPP2R2A	SV	low	[80]biological evidence	SVs and CNAs in ETS+
ROBO1	SV	low	[81]biological evidence	SVs in ETS+
ROBO2	SV	low	[25]	SVs in ETS+
RPL11	SNV/indel	low	[25]	dN/dS (missense)
SENP6	SV	low	[42]biological evidence	enriched SVs, RNA expression
TBL1XR1	SNV/indel,SV	low	[82]known AR co-regulator biological evidence	dN/dS
USP28	SV, CNA, SNV/indel	low	none	SVs, CNA, SNV/indel
ZNF292	SV, CNA SNV/indel,	low	[25]	enriched SVs, homozygous deletions, truncating mutations
ARID1A	SNV/indel	medium	[83]	dN/dS
CASZ1	SNV/indel	medium	COSMIC, TCGA and SU2C	dN/dS
CNOT3	SNV/indel	medium	[84]Mut. in leukemia	dN/dS (missense)
LRP1B	SV, CNA	medium	[79]SNP data	SVs and CNA in ETS-
PIK3R1	SNV/indel	medium	[24]	dN/dS
RGMB	CNA	medium	[38]deletions	CNA in ETS-
TBX3	SNV/indel	medium	known breast cancer gene	dN/dS
ZMYM3	SNV/indel	medium	COSMIC SU2C	dN/dS

Table 2

Target of approved drug
***AR, BRAF**, ESR1, HDAC3, **KCNH2**, MAP2K1, NR3C1, RARA, RARB, RARG, RXRA*
Target of investigational drug
AKT1, ATM, LRRK2**, MDM2, **PDE4D, PIK3CA, PIK3CB, TP53
Target being investigated chemically
*AHR, BRCA1, **CTNNB1, HRAS, IDH1, ITPR1, ITPR2**, JUN, **MAP3K1**, MEN1, **NCOR1, NCOR2**, NR4A1, **PIK3R1, PPP2R2A***
Predicted target by structure-based method
***ANTXR2, APC**, ARNT, **ASH1L, BRCA2**, CBFA2T2**, CDH12, CDK12**, **CHD1**, CREBBP, **DLC1, DOCK10, ERG, ESCO1, ETV3, FOXA1**, FOXG1, **FOXO1**, FOXO4, **FOXP1**, GATA1, GATA2, HDGF, HNF4A, **IL6ST**, **KAT6A**, KDM4A**, KDM6A**, **KMT2C, KMT2D, NEDD4L, NKX3-1**, PIAS1, PIAS2, **PTEN, RB1, RNF43**, SKI, **SMAD2**, SMAD3**, SMAD4**, SMARCA4, SPDEF, **SPOP**, TBL1X, **TBL1XR1**, **TBX3**, TP73, **ZBTB16**, ZHX2*

FIG. 1

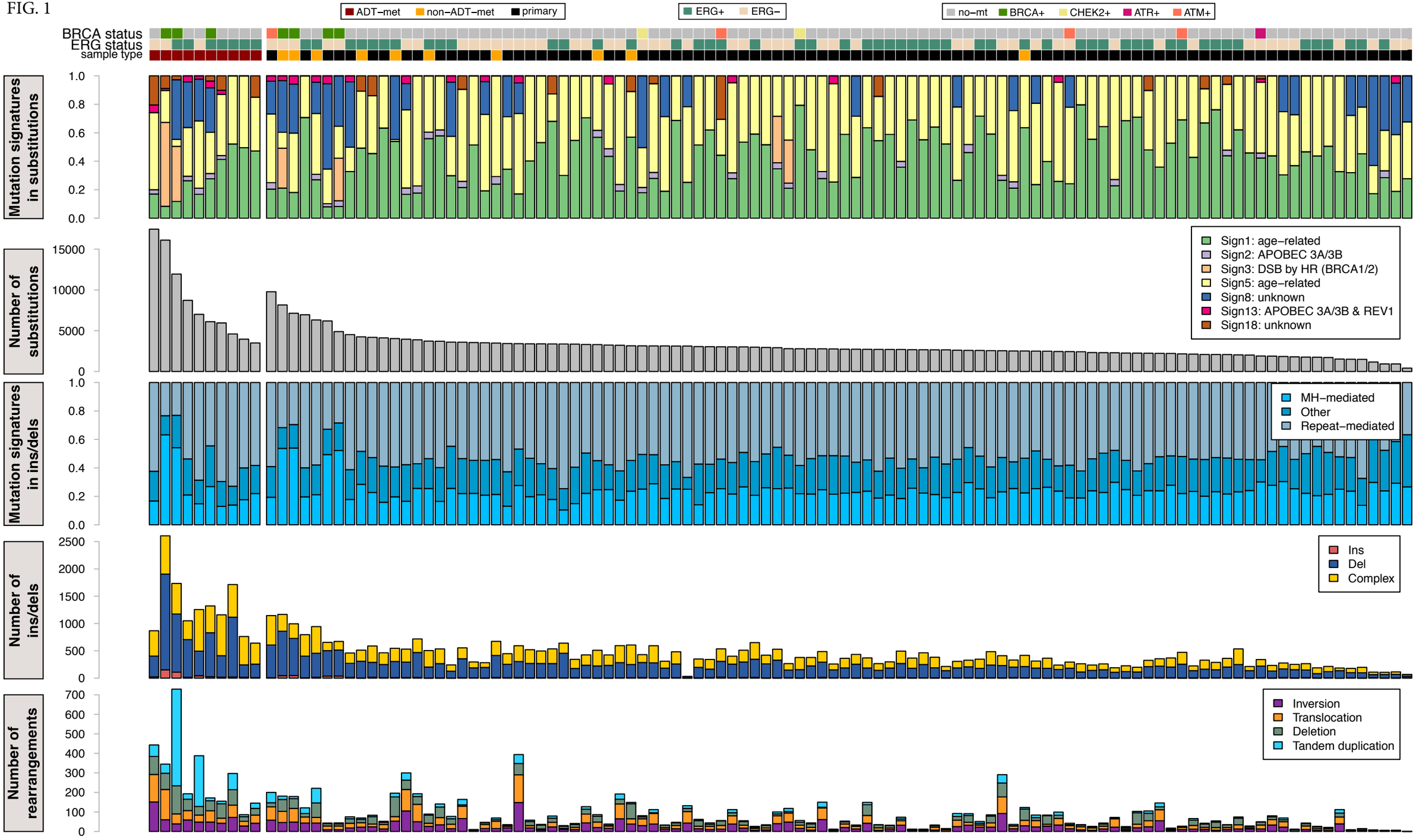
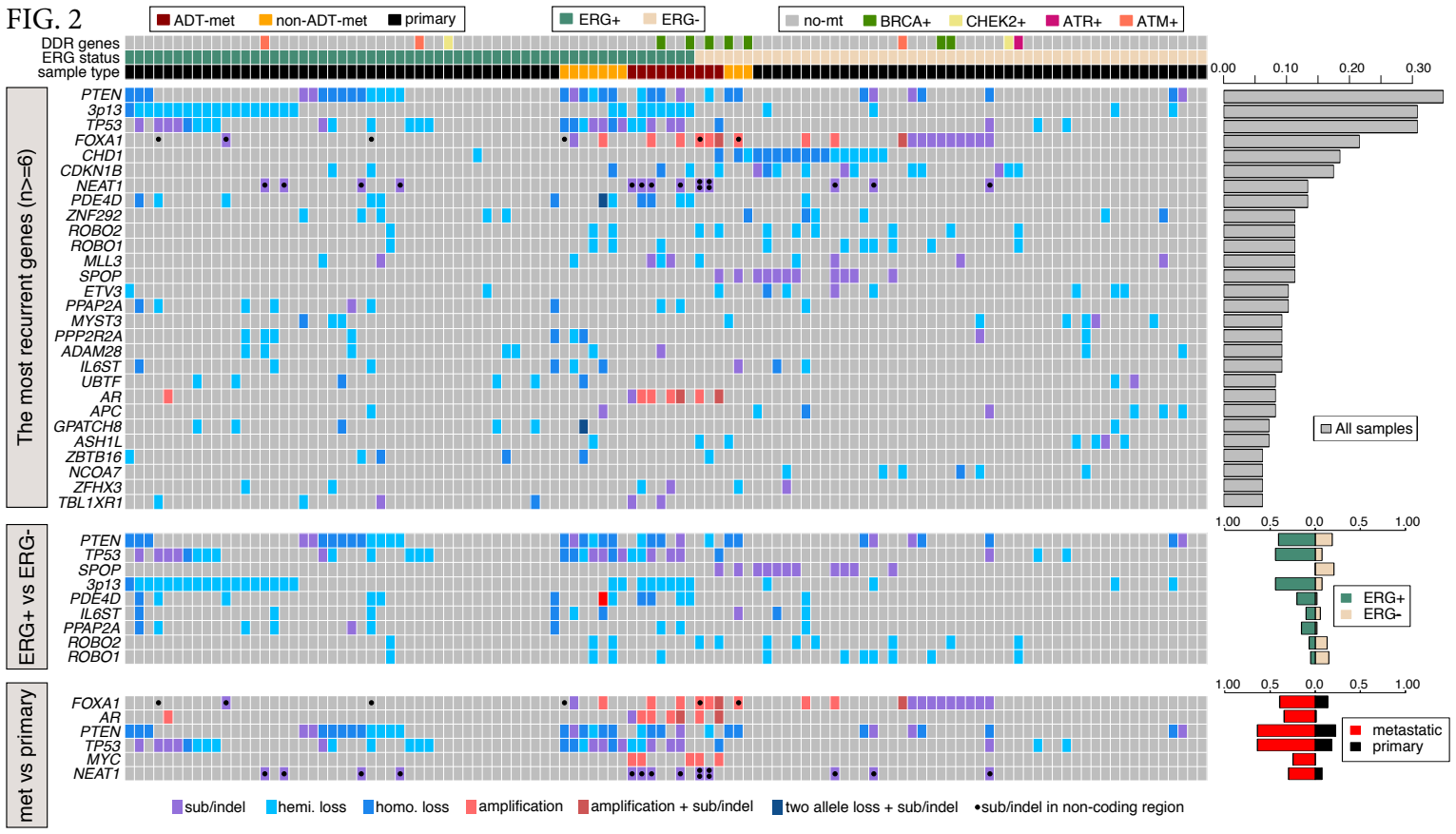
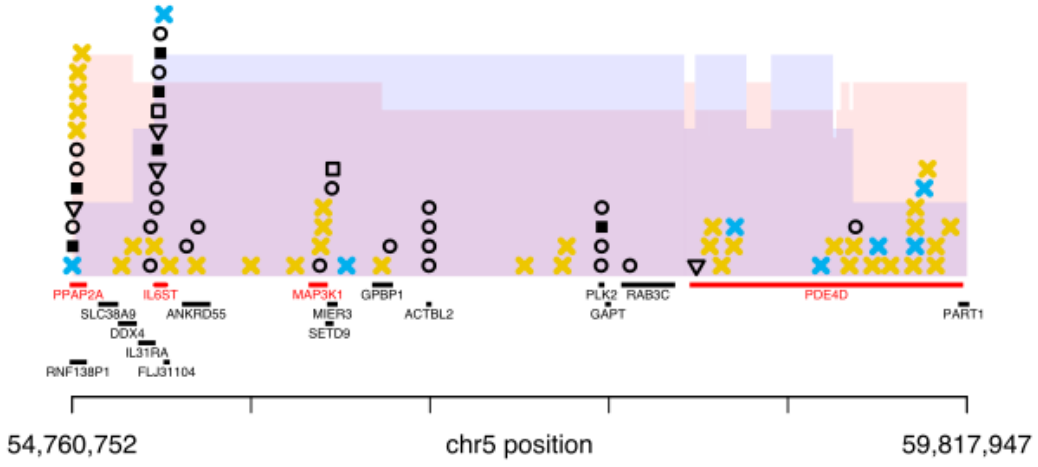
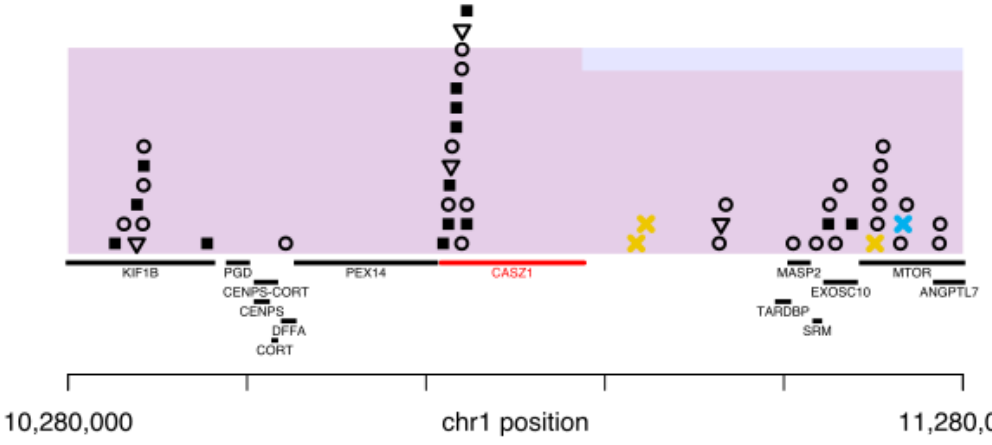
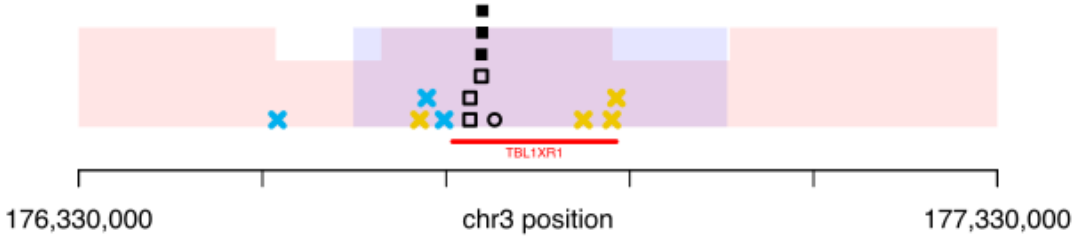

ADT−met
non−ADT−met
primary
ERG+
ERG−
no−mt
BRCA+
CHEK2+
ATR+
ATM+
BRCA status
ERG status
sample type
Mutation signatures in substitutions
1.0
0.8
0.6
0.4
0.2
0.0
Number of substitutions
15000
10000
5000
0
Sign1: age−related
Sign2: APOBEC 3A/3B
Sign3: DSB by HR (BRCA1/2)
Sign5: age−related
Sign8: unknown
Sign13: APOBEC 3A/3B & REV1
Sign18: unknown
Mutation signatures in ins/dels
MH−mediated
Other
Repeat−mediated
Number of ins/dels
2500
2000
1500
1000
500
0
Ins
Del
Complex
Number of rearrangements
700
600
500
400
300
200
100
0
Inversion
Translocation
Deletion
Tandem duplication

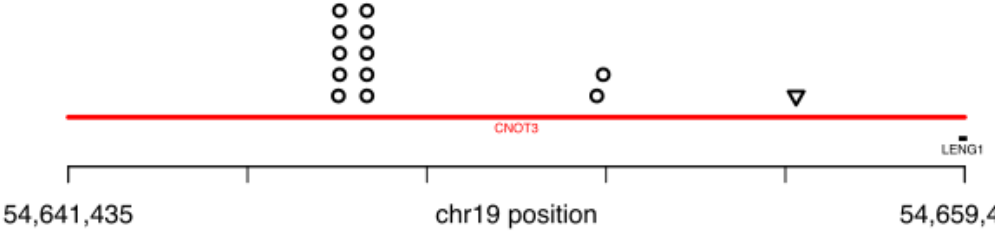
FIG. 2

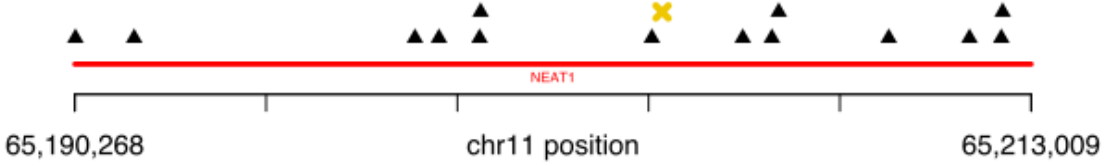

ADT-met
non-ADT-met
primary
ERG+
ERG-
no-mt
BRCA+
CHEK2+
ATR+
ATM+
DDR genes
ERG status
sample type
0.00
0.10
0.20
0.30
The most recurrent genes (n>=6)
PTEN
3p13
TP53
FOXA1
CHD1
CDKN1B
NEAT1
PDE4D
ZNF292
ROBO2
ROBO1
MLL3
SPOP
ETV3
PPAP2A
MYST3
PPP2R2A
ADAM28
IL6ST
UBTF
AR
APC
GPATCH8
ASH1L
ZBTB16
NCOA7
ZFHX3
TBL1XR1
All samples
ERG+ vs ERG-
PTEN
TP53
SPOP
3p13
PDE4D
IL6ST
PPAP2A
ROBO2
ROBO1
1.00
0.5
0.0
0.5
1.00
ERG+
ERG-
met vs primary
FOXA1
AR
PTEN
TP53
MYC
NEAT1
metastatic
primary
sub/indel
hemi. loss
homo. loss
amplification
amplification + sub/indel
two allele loss + sub/indel
•sub/indel in non-coding region

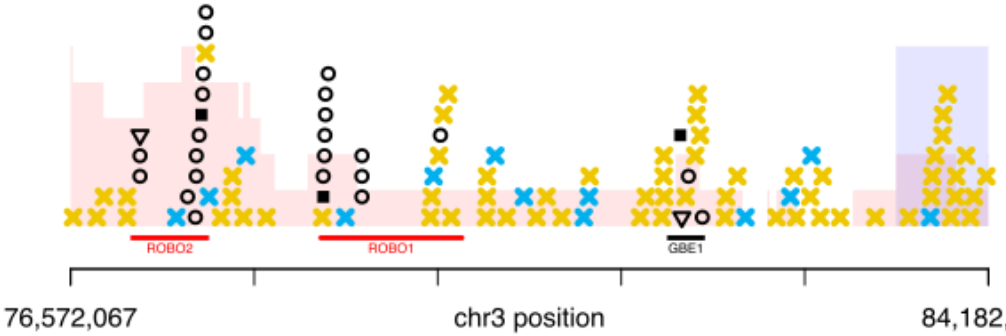

PPAP2A
IL6ST
MAP3K1
GPBP1
PLK2
RAB3C
PDE4D
SLC38A9
ANKRD55
MIER3
ACTBL2
GAPT
PART1
DDX4
SETD9
IL31RA
RNF138P1
FLJ31104
54,760,752
chr5 position
59,817,947

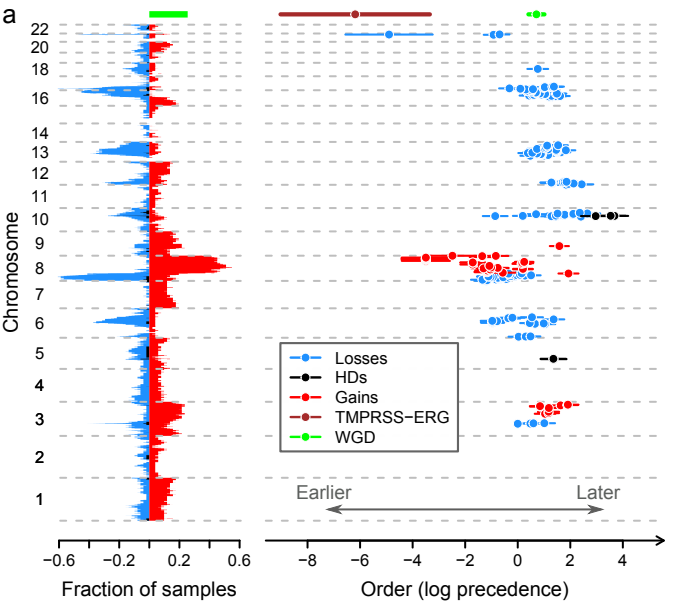

KIF1B
PGD
PEX14
CASZ1
MASP2
MTOR
CENPS-CORT
EXOSC10
ANGPTL7
CENPS
TARDBP
DFFA
SRM
CORT
10,280,000
chr1 position
11,280,000

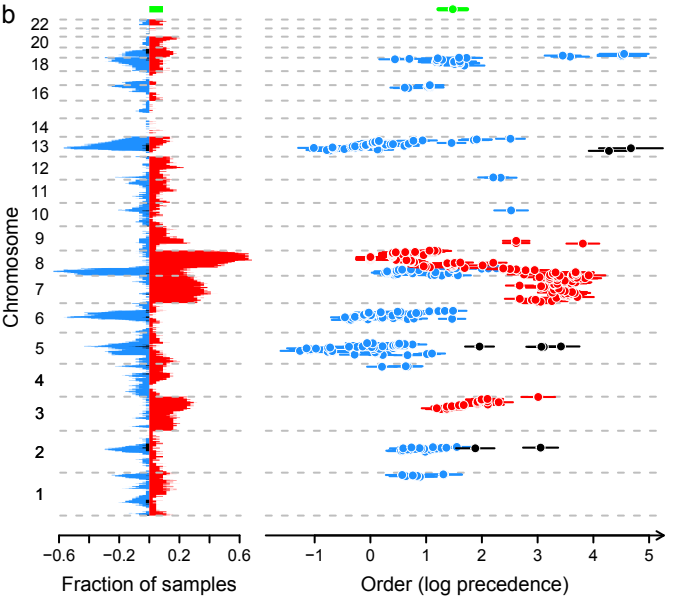

TBL1XR1
176,330,000
chr3 position
177,330,000


CNOT3
LENG1
54,641,435
chr19 position
54,659,446


NEAT1
65,190,268
chr11 position
65,213,009


ROBO2
ROBO1
GBE1
76,572,067
chr3 position
84,182,020


a
Chromosome
Losses
HDs
Gains
TMPRSS−ERG
WGD
Earlier
Later
Fraction of samples
Order (log precedence)


b
Chromosome
Fraction of samples
Order (log precedence)

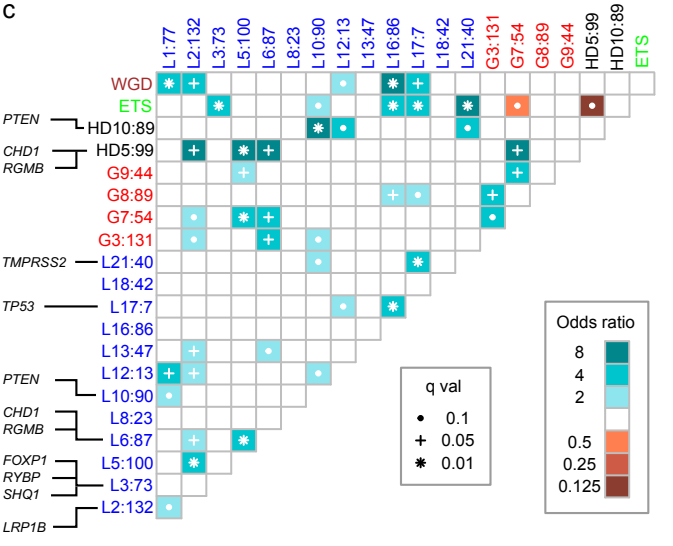
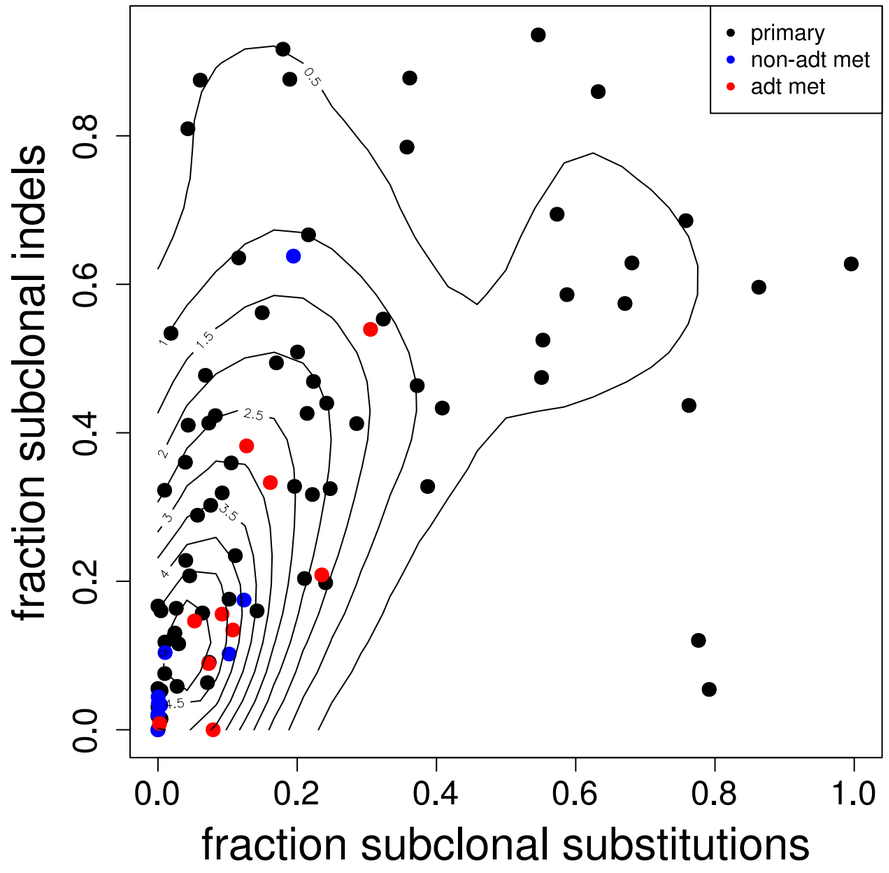
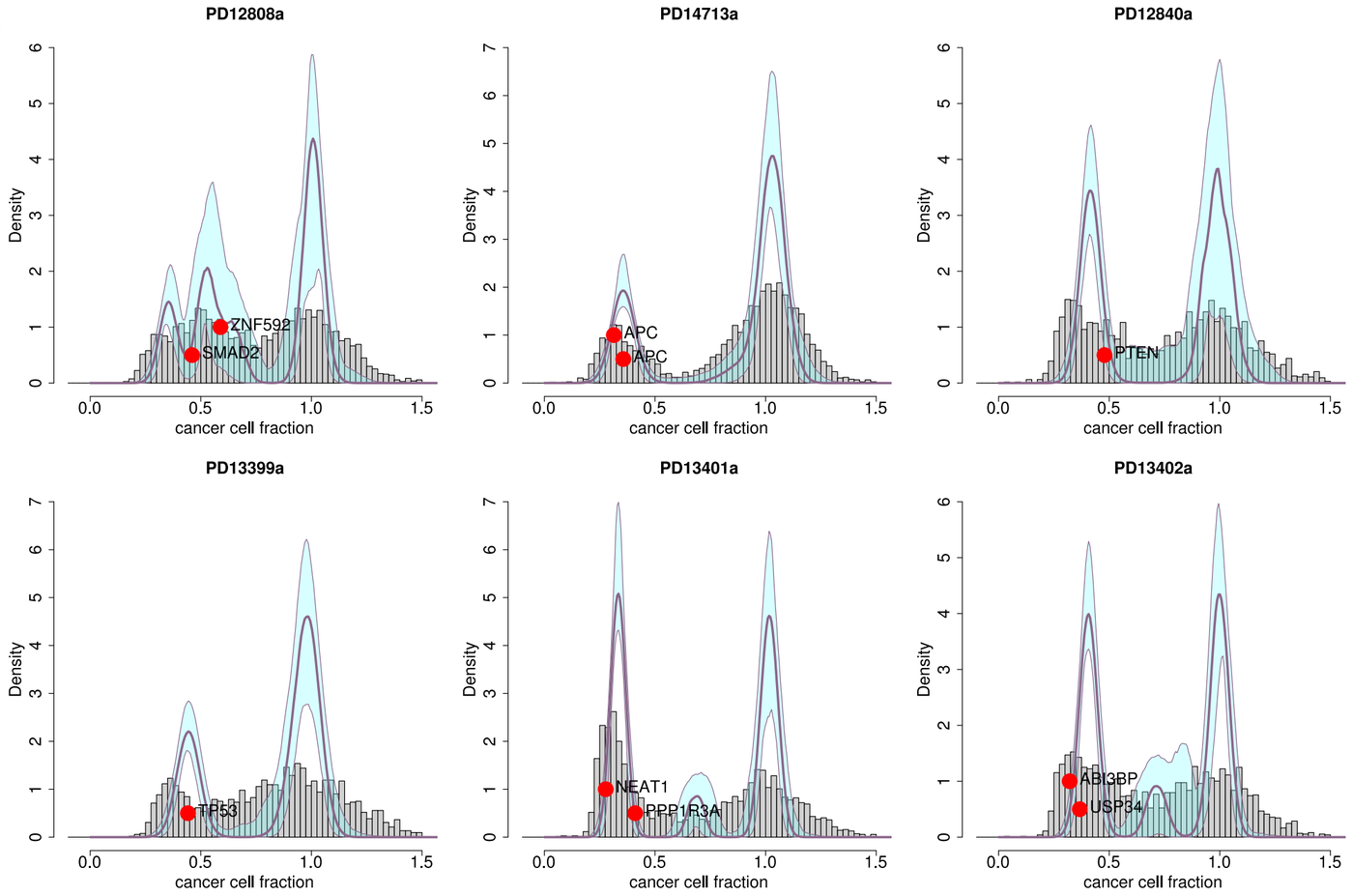

c
L1:77
L2:132
L3:73
L5:100
L6:87
L8:23
L10:90
L12:13
L13:47
L16:86
L17:7
L18:42
L21:40
G3:131
G7:54
G8:89
G9:44
HD5:99
HD10:89
ETS
WGD
HD10:89
HD5:99
G9:44
G8:89
G7:54
G3:131
L21:40
L18:42
L17:7
L16:86
L13:47
L12:13
L10:90
L8:23
L6:87
L5:100
L3:73
L2:132
PTEN
CHD1
RGMB
TMPRSS2
TP53
PTEN
CHD1
RGMB
FOXP1
RYBP
SHQ1
LRP1B
q val
0.1
0.05
0.01
Odds ratio
8
4
2
0.5
0.25
0.125

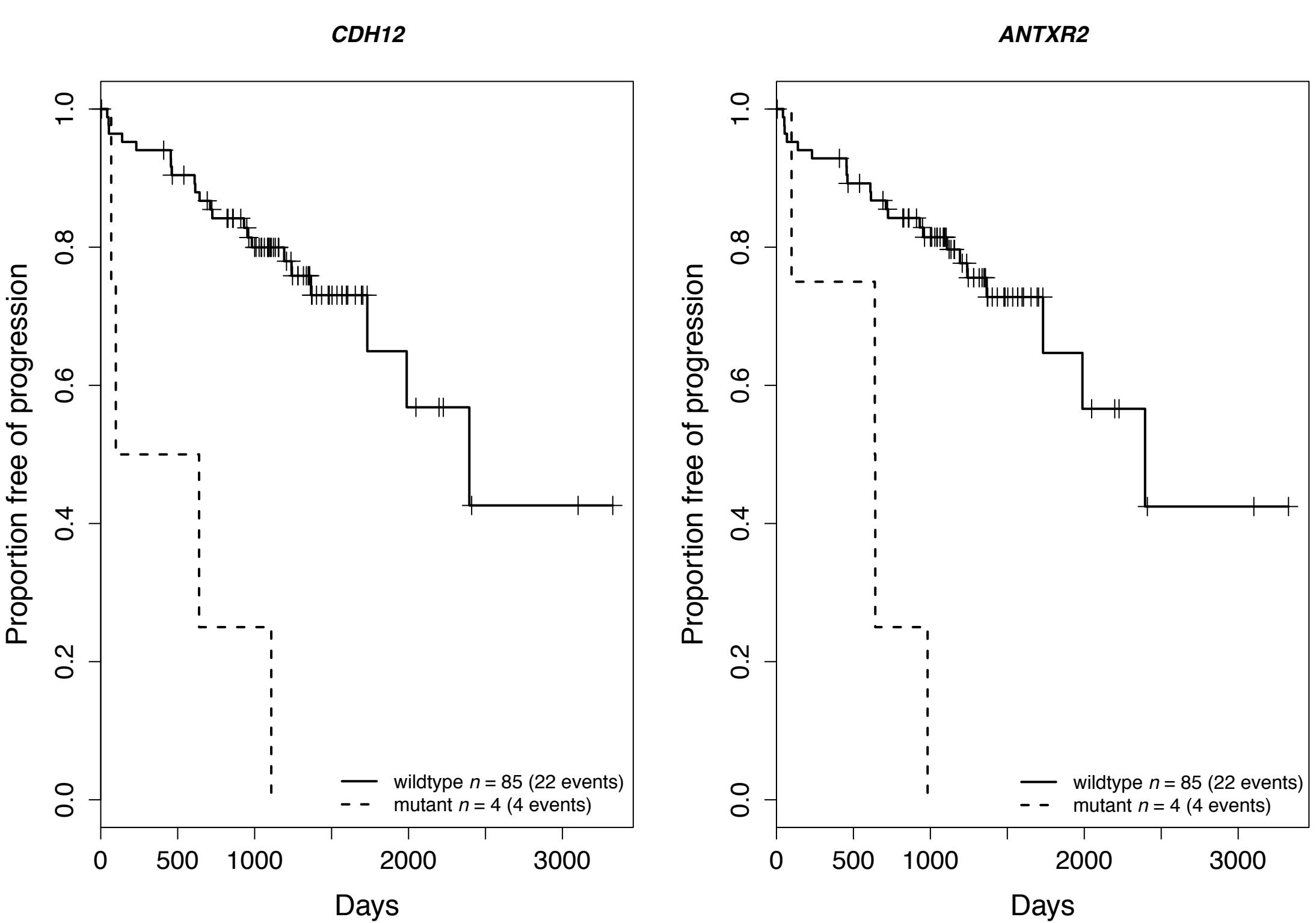
Figure 5

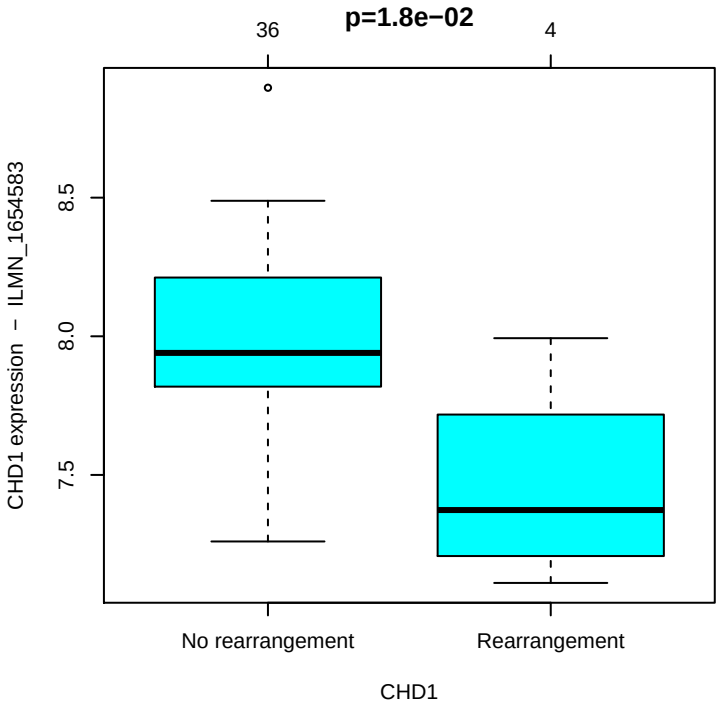
a

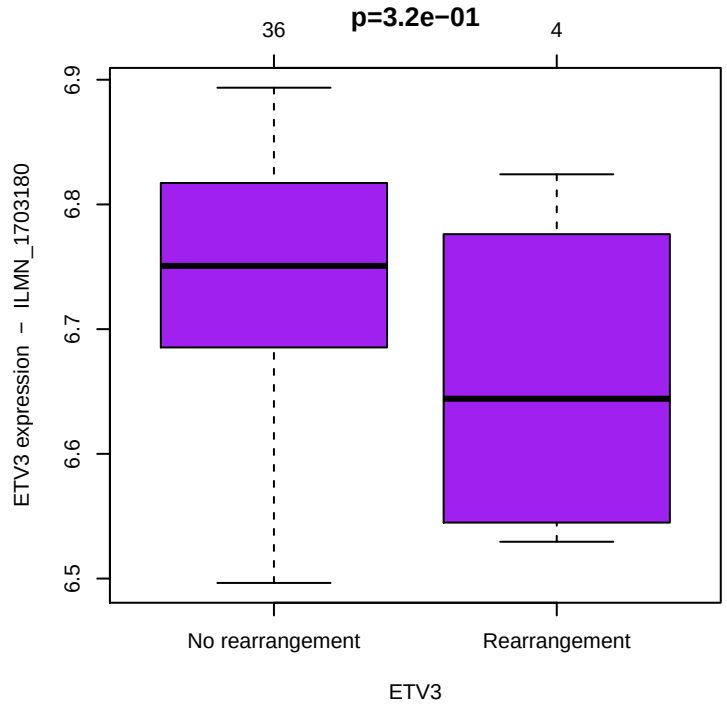

primary
non-adt met
adt met
fraction subclonal indels
fraction subclonal substitutions
0.0
0.2
0.4
0.6
0.8
1.0
0.5
1
1.5
2
2.5
3
3.5
4
4.5

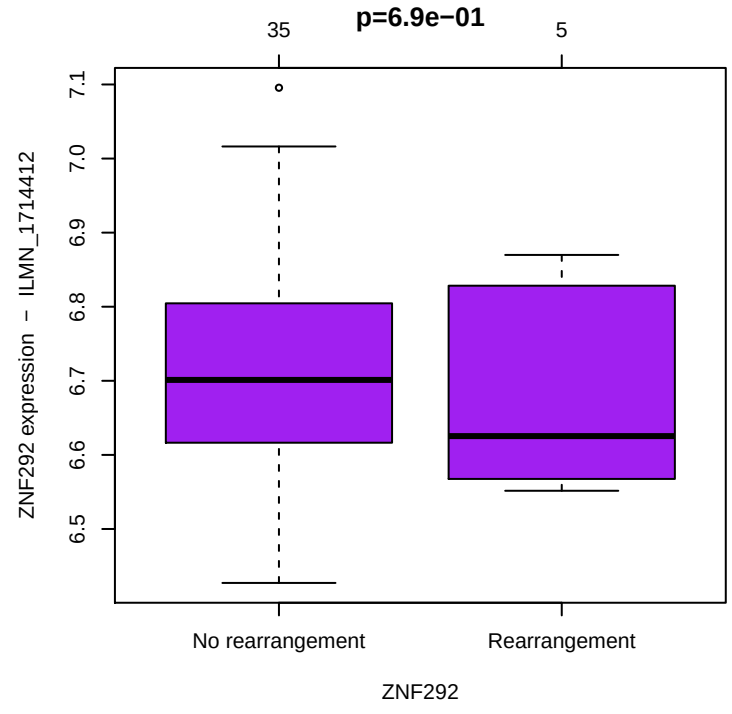
b

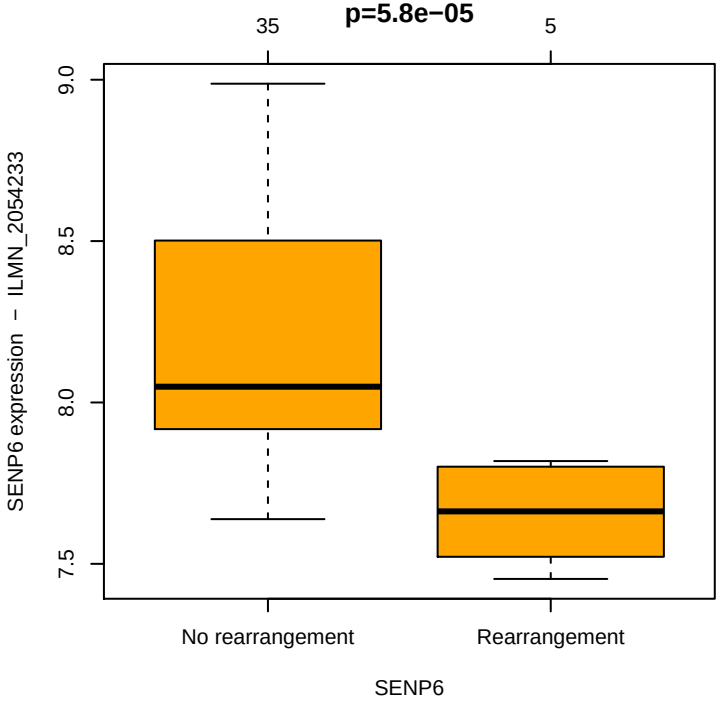

PD12808a
ZNF592
SMAD2
PD14713a
APC
APC
PD12840a
PTEN
PD13399a
TP53
PD13401a
NEAT1
PPP1R3A
PD13402a
ABI3BP
USP34
Density
cancer cell fraction
0.0
0.5
1.0
1.5

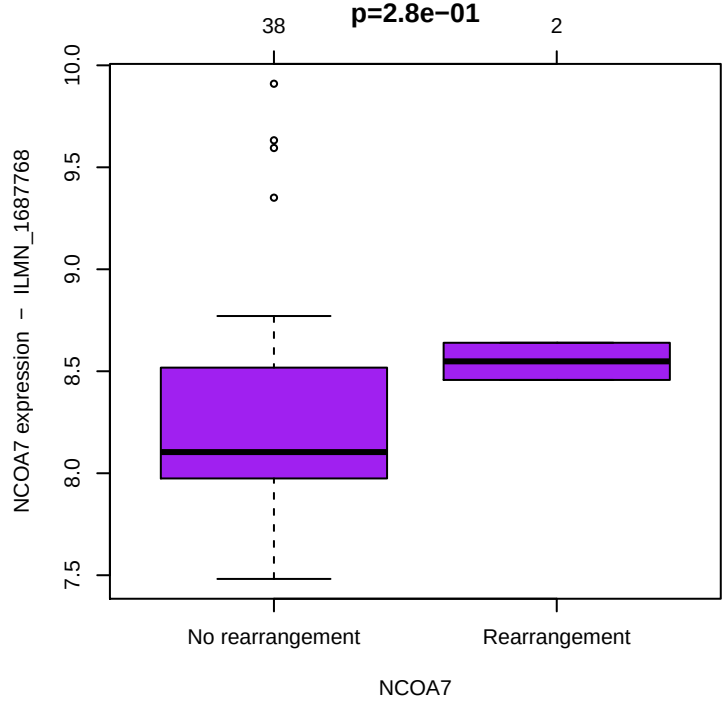

CDH12
ANTXR2
Proportion free of progression
1.0
0.8
0.6
0.4
0.2
0.0
0
500
1000
2000
3000
Days
wildtype n = 85 (22 events)
mutant n = 4 (4 events)

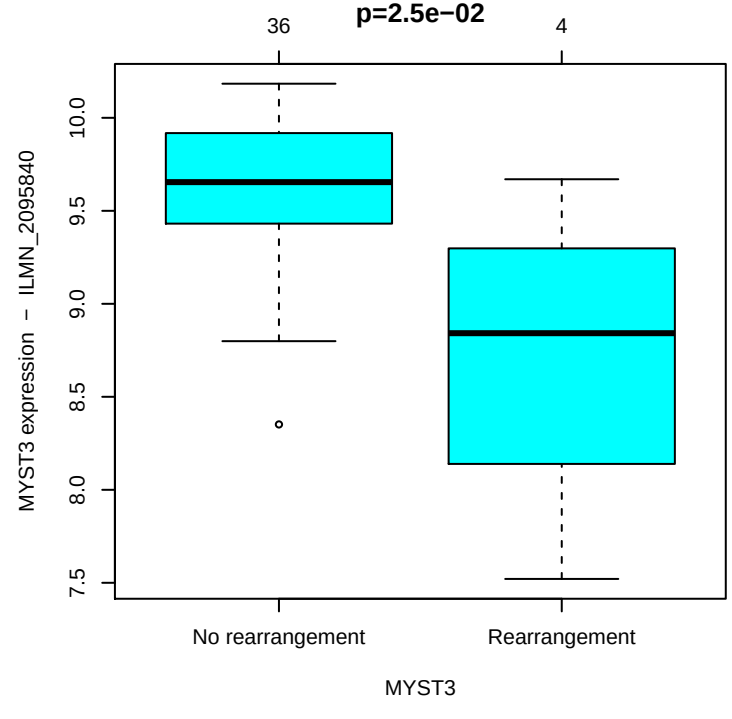

p=1.8e−02
36
4
CHD1 expression – ILMN_1654583
8.5
8.0
7.5
No rearrangement
Rearrangement
CHD1


p=3.2e−01
36
4
ETV3 expression – ILMN_1703180
6.9
6.8
6.7
6.6
6.5
No rearrangement
Rearrangement
ETV3

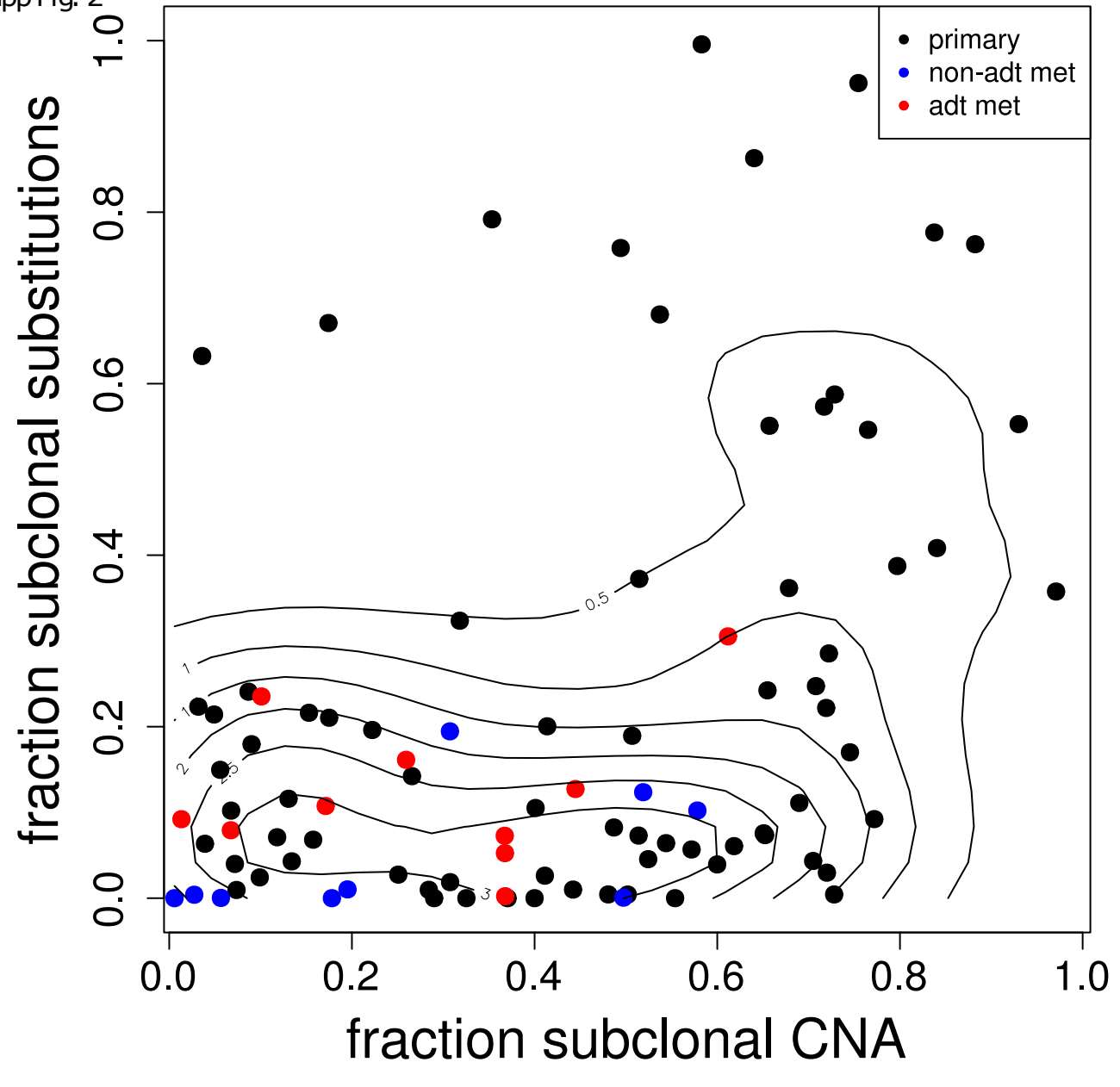

p=6.9e−01
35
5
ZNF292 expression – ILMN_1714412
7.1
7.0
6.9
6.8
6.7
6.6
6.5
No rearrangement
Rearrangement
ZNF292


p=5.8e−05
35
5
SENP6 expression – ILMN_2054233
9.0
8.5
8.0
7.5
No rearrangement
Rearrangement
SENP6


p=2.8e−01
38
2
NCOA7 expression – ILMN_1687768
10.0
9.5
9.0
8.5
8.0
7.5
No rearrangement
Rearrangement
NCOA7

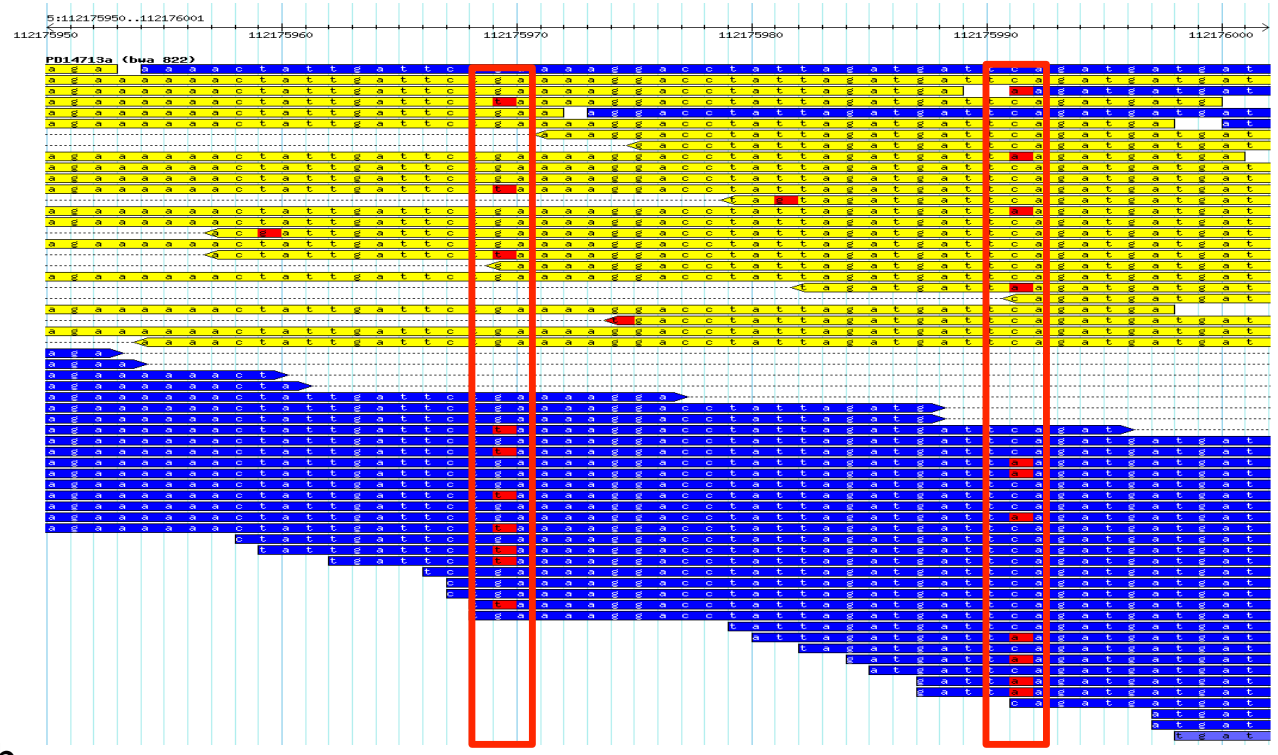

p=2.5e−02
36
4
MYST3 expression – ILMN_2095840
10.0
9.5
9.0
8.5
8.0
7.5
No rearrangement
Rearrangement
MYST3

Supp Fig. 2

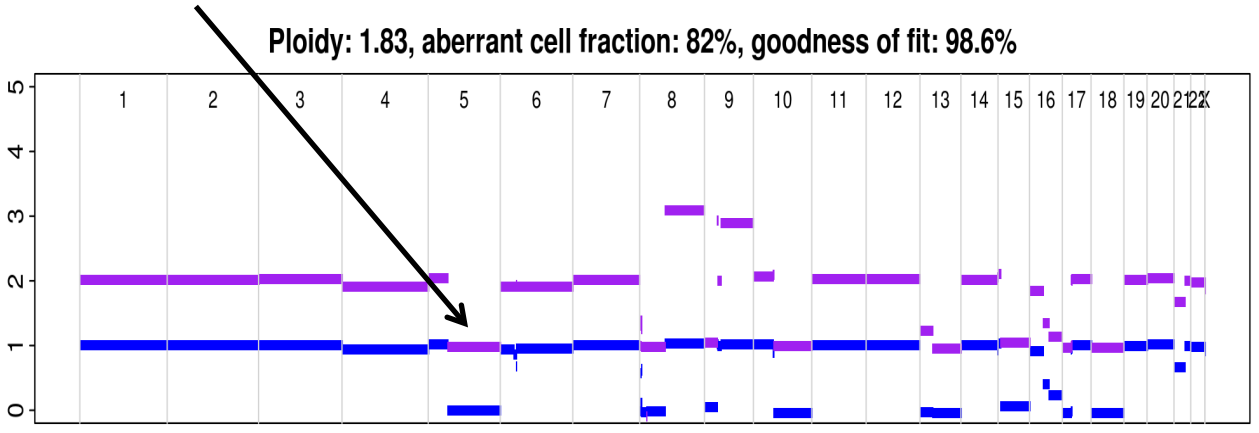

primary
non-adt met
adt met
fraction subclonal substitutions
fraction subclonal CNA
0.0
0.2
0.4
0.6
0.8
1.0
0.5
1
.1
2
.5
3

SUPP FIG 3

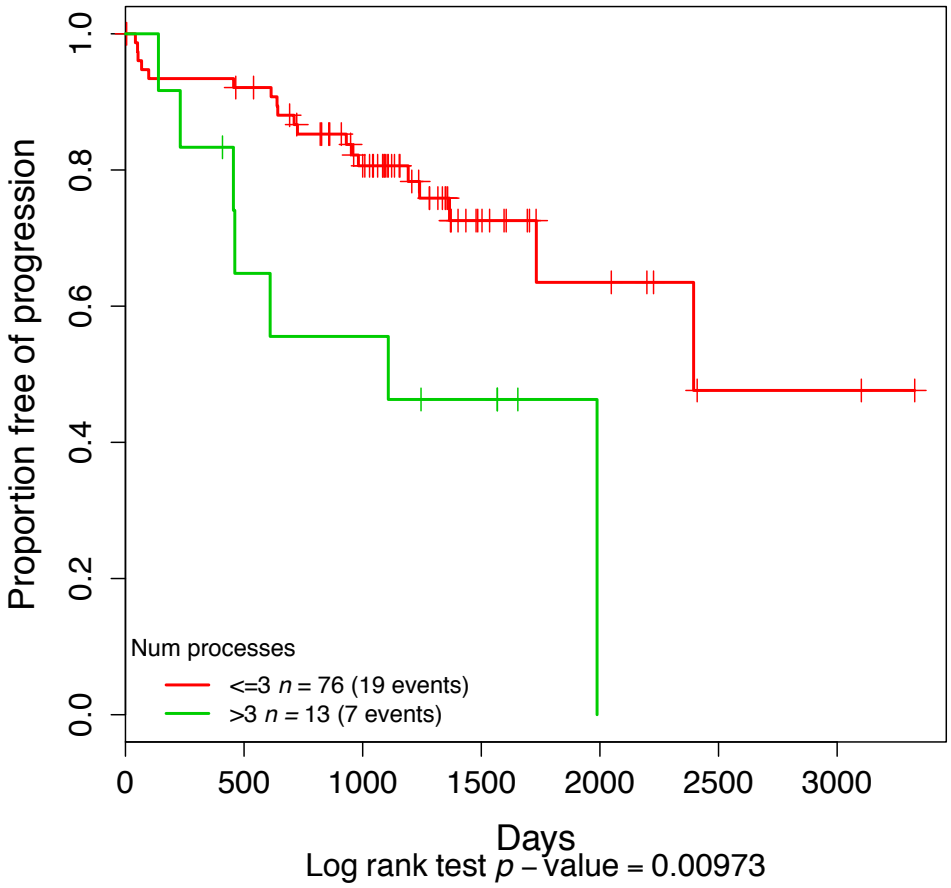
a


5:112175950..112176001
112175950
112175960
112175970
112175980
112175990
112176000
PD14713a (bwa 822)

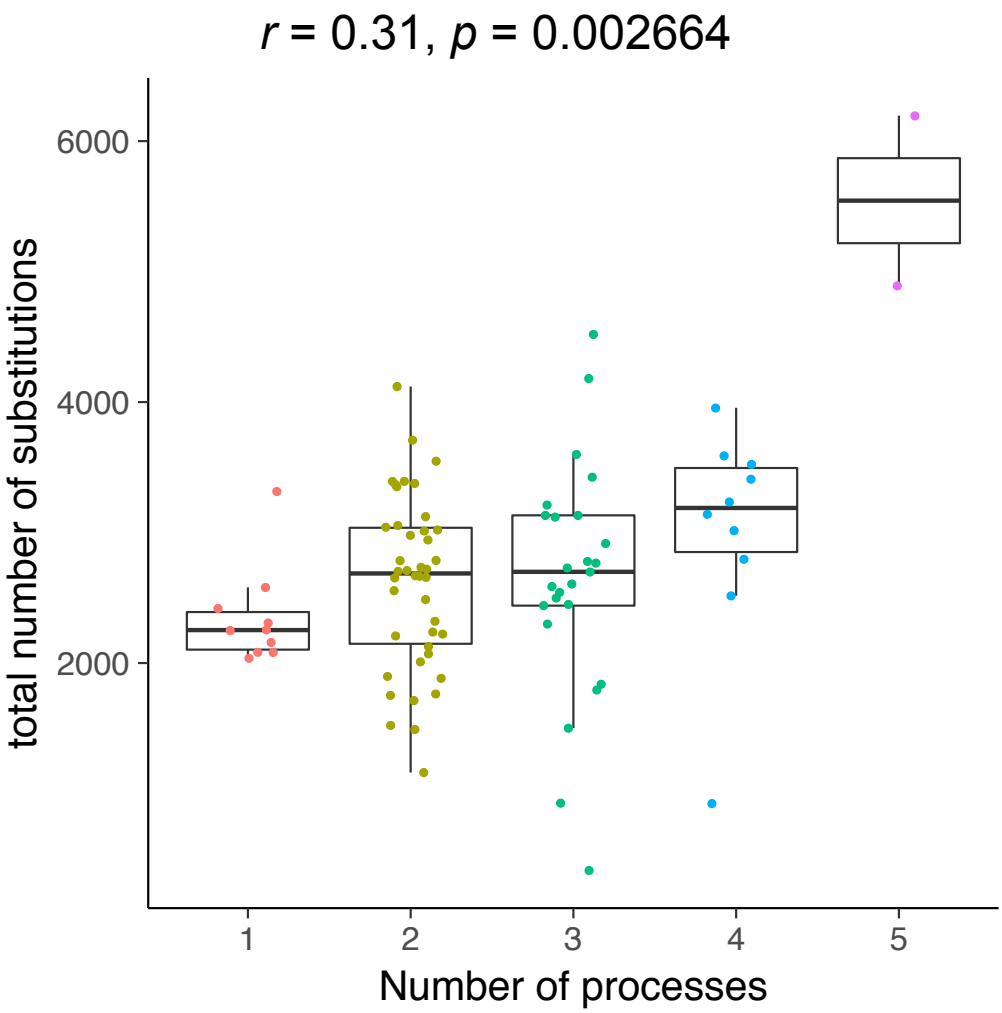
b


Ploidy: 1.83, aberrant cell fraction: 82%, goodness of fit: 98.6%
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X
0 1 2 3 4 5

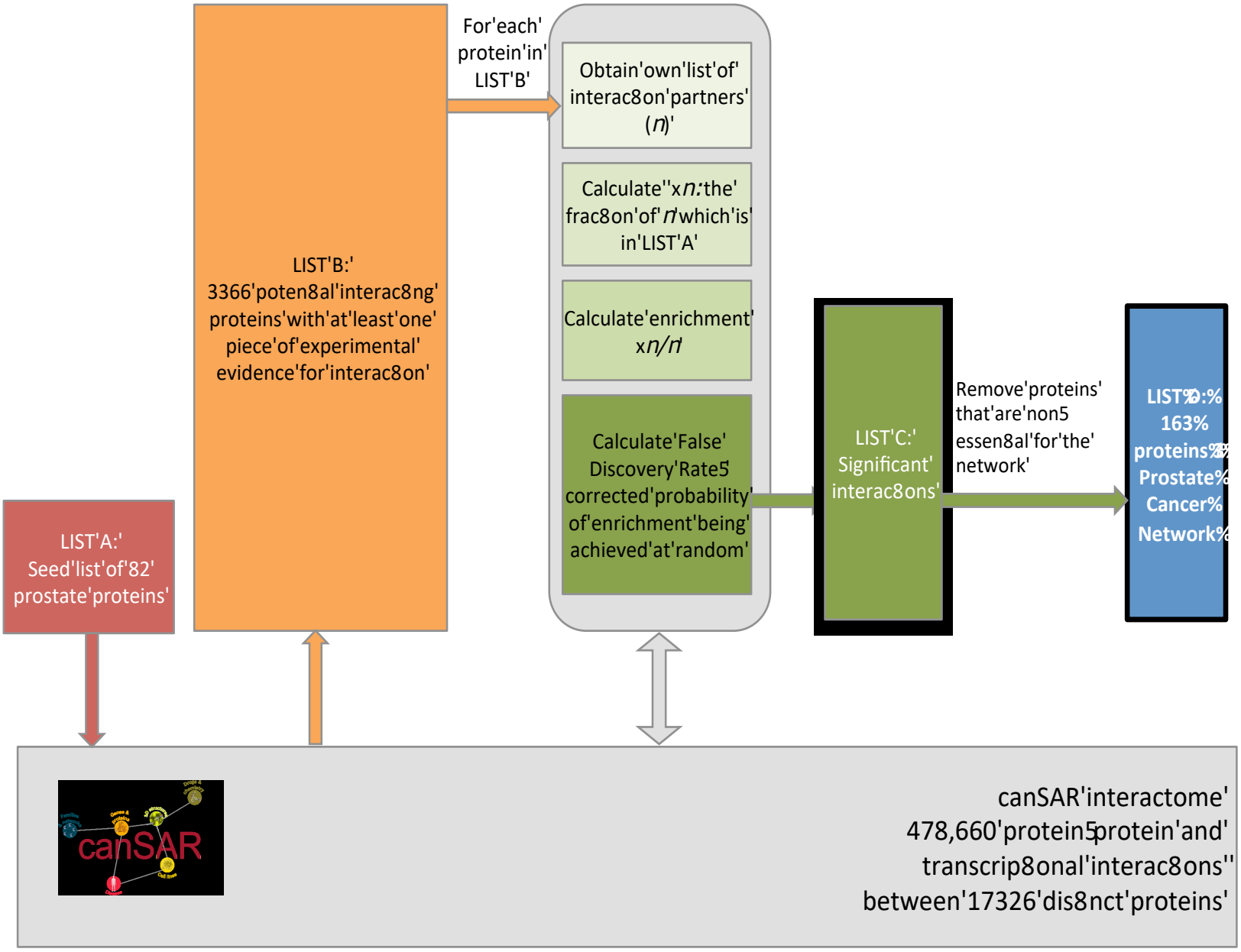
a

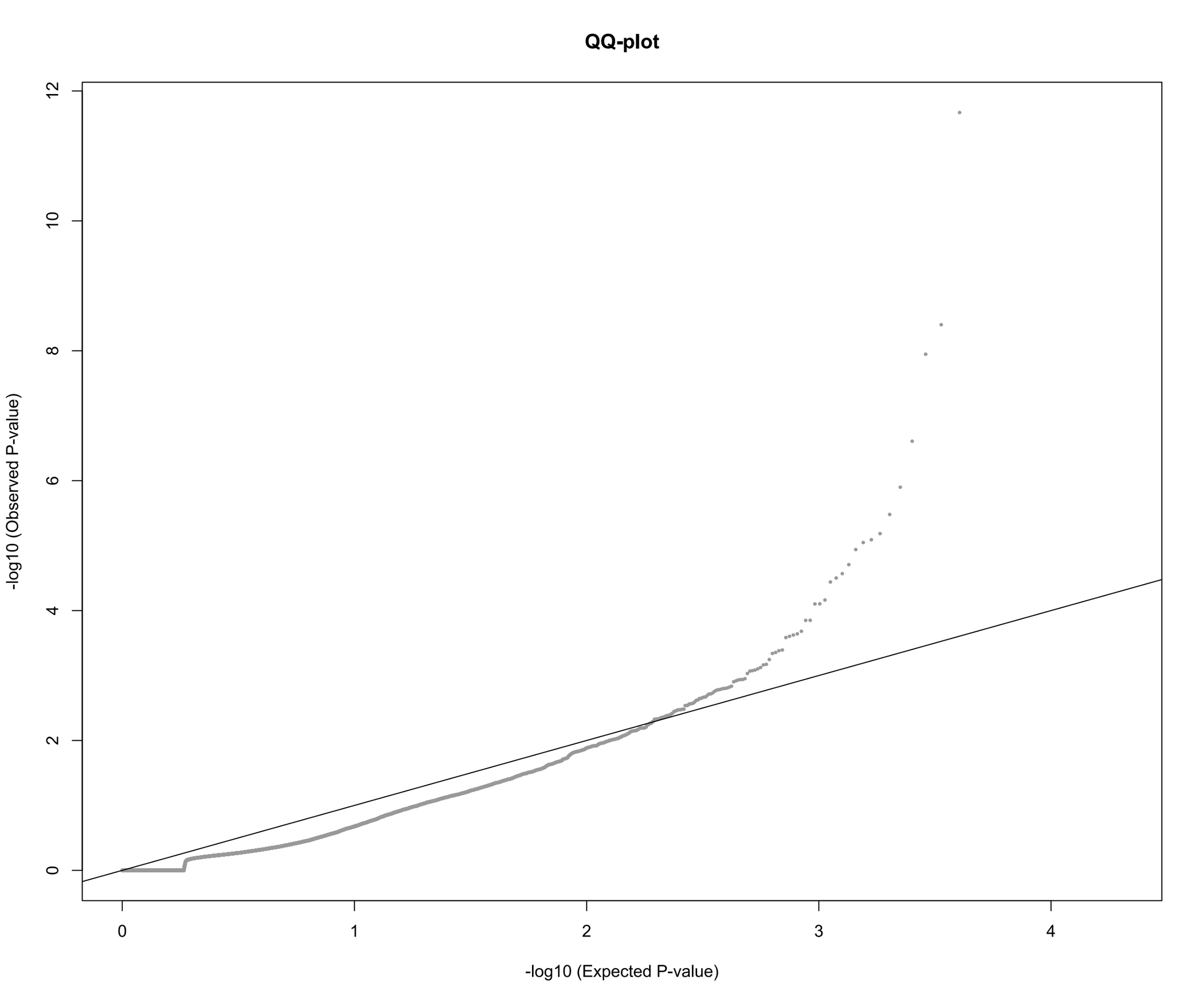

Proportion free of progression
0.0
0.2
0.4
0.6
0.8
1.0
Num processes
<=3 n = 76 (19 events)
>3 n = 13 (7 events)
0
500
1000
1500
2000
2500
3000
Days
Log rank test p – value = 0.00973

b


r = 0.31, p = 0.002664
total number of substitutions
2000
4000
6000
1
2
3
4
5
Number of processes

SUPP FIG. 5


LIST A: Seed list of 82 prostate proteins
LIST B: 3366 potential interacting proteins with at least one piece of experimental evidence for interaction
For each protein in LIST B
Obtain own list of interaction partners (n)
Calculate x n: the fraction of n which is in LIST A
Calculate enrichment x n/n
Calculate False Discovery Rate corrected probability of enrichment being achieved at random
LIST C: Significant interactions
Remove proteins that are non essential for the network
LIST D: 163 proteins Prostate Cancer Network
canSAR
canSAR interactome 478,660 protein protein and transcriptional interactions between 17326 distinct proteins


QQ-plot
-log10 (Observed P-value)
-log10 (Expected P-value)
0
2
4
6
8
10
12
0
1
2
3
4

SUPPLEMENTARY TABLE 1

ensg	gene	source	ess_spl	frames	infram	missen	nonser	silent	start-lc	stop-lo
ENSG000001	ERG	cgp	0	0	0	1	0	0	0	0
ENSG000001	PTEN	cgp	1	1	1	6	0	0	0	0
ENSG000001	TP53	cgp	0	6	0	6	1	0	0	0
ENSG000001	CDKN1B	cgp	0	0	0	0	3	0	0	0
ENSG000001	PDE4D	cgp	0	0	0	0	1	0	0	0
ENSG000001	ROBO1	cgp	0	0	0	0	0	2	0	0
ENSG000001	ROBO2	cgp	0	0	0	2	0	1	0	0
ENSG000001	ZNF292	cgp	0	0	0	1	0	0	0	0
ENSG000001	FOXA1	cgp	0	3	6	5	0	0	0	0
ENSG000000	PPAP2A	cgp	0	1	0	0	0	0	0	0
ENSG000001	ETV3	cgp	0	1	0	2	0	0	0	0
ENSG000000	ADAM28	cgp	0	1	0	0	0	2	0	0
ENSG000001	IL6ST	cgp	0	1	0	0	0	0	0	0
ENSG000000	MYST3	cgp	0	0	0	0	1	0	0	0
ENSG000002	PPP2R2A	cgp	1	0	0	0	0	0	0	0
ENSG000001	AR	cgp	0	0	0	3	0	0	0	0
ENSG000001	APC	cgp	0	1	0	0	1	1	0	0
ENSG000001	ASH1L	cgp	0	1	0	0	0	0	0	0
ENSG000001	DLC1	cgp	0	0	0	1	0	0	0	0
ENSG000001	NCOR2	cgp	0	0	0	1	0	0	0	0
ENSG000001	CDH12	cgp	0	1	0	0	0	0	0	0
ENSG000000	KMT2C	cgp	0	2	1	3	2	1	0	0
ENSG000001	SENP6	cgp	0	0	0	1	0	0	0	0
ENSG000001	ANTXR2	cgp	0	1	0	0	0	0	0	0
ENSG000000	ARID4B	cgp	0	0	0	2	0	0	0	0
ENSG000001	ASXL2	cgp	1	0	0	1	0	0	0	0
ENSG000001	DOCK10	cgp	0	0	0	2	0	0	0	0
ENSG000001	RB1	cgp	0	0	0	2	0	0	0	0
ENSG000001	ZFHX3	cgp	0	2	0	1	0	0	0	0
ENSG000001	TBL1XR1	cgp	2	0	0	1	0	0	0	0
ENSG000001	ATM	cgp	0	1	0	0	1	0	0	0
ENSG000001	BRCA2	cgp	0	0	1	0	0	1	0	0
ENSG000001	CDK12	cgp	0	2	0	0	0	0	0	0
ENSG000001	KDM6A	cgp	0	2	0	1	1	0	0	0
ENSG000001	NCOR1	cgp	0	0	0	1	0	0	0	0
ENSG000001	ARID1A	cgp	0	1	0	0	1	0	0	0
ENSG000001	CTNNB1	cgp	0	0	1	1	0	0	0	0
ENSG000001	SMAD2	cgp	1	0	0	1	0	0	0	0

ENSG000001	SMAD4	cgp	0	1	0	0	0	0	0	0
ENSG000001	ZMYM3	cgp	0	1	0	1	0	0	0	0

homod	n_rearr_icgc	chr	start	end	strand
0	56	21	39751949	40033704	-1
13	28	10	89622870	89731687	1
1	19	17	7565257	7590856	-1
0	15	12	12867992	12875305	1
3	13	5	58264865	59817947	-1
0	11	3	78646390	79816965	-1
0	11	3	77089294	77699114	1
3	11	6	87862551	87973914	1
0	10	14	38059189	38064239	-1
1	10	5	54720682	54830878	-1
0	9	1	157090983	157108266	-1
0	8	8	24151580	24212726	1
3	8	5	55230925	55290821	-1
0	8	8	41799567	41909505	-1
2	8	8	26149007	26230196	1
0	7	X	66763874	66950461	1
0	6	5	112043195	112203279	1
0	6	1	155305052	155532324	-1
1	6	8	13162116	13372429	-1
0	6	12	124808957	125052010	-1
0	5	5	21750777	22853731	-1
1	5	7	151832007	152133628	-1
0	5	6	76311225	76427997	1
0	4	4	80898662	80994477	-1
0	4	1	235330210	235490802	-1
1	4	2	25960557	26101385	-1
0	4	2	225629807	225907330	-1
3	4	13	48877887	49056122	1
0	4	16	72816784	73082274	-1
1	3	3	176738542	176915048	-1
0	2	11	108175402	108186638	1
2	2	13	32889611	32973347	1
0	2	17	37617739	37690800	1
0	2	X	44732423	44971847	1
1	2	17	15933408	16118874	-1
0	1	1	27022522	27108601	1
0	1	3	41236328	41301587	1
0	1	18	45359466	45457517	-1

0	1	18	48556583	48611415	1
0	1 X		70459474	70474996	-1

multiple_hit_all_number	multiple_h	multiple_hit_mt_and_sv_number
0	0	0
7	5	5
12	12	1
1	0	1
1	1	1
0	0	0
0	0	0
0	0	0
4	0	4
0	0	0
0	0	0
1	1	0
1	1	0
1	1	0
0	0	0
2	0	2
2	2	0
0	0	0
1	1	0
0	0	0
0	0	0
0	0	0
1	1	0
0	0	0
0	0	0
1	1	1
0	0	0
2	2	1
0	0	0
0	0	0
2	1	1
2	2	0
0	0	0
1	0	1
1	1	0
0	0	0
0	0	0
0	0	0

1	1	0
0	0	0

gene	source	ess_splice	frameshift	inframe	missense	nonsense	silent	start-lost
ADAM28	tcga	0	1	1	0	0	1	0
ERG	cgp	0	0	0	1	0	0	0
ANTXR2	tcga	0	0	0	0	1	0	0
ANTXR2	SU2C-PCF	0	2	0	0	0	0	0
PTEN	cgp	1	1	1	6	0	0	0
APC	SU2C-PCF	0	7	0	1	3	0	0
APC	tcga	0	2	0	2	5	0	0
APC	COSMIC	0	7	0	0	4	1	0
TP53	cgp	0	6	0	6	1	0	0
AR	tcga	0	0	0	3	0	0	0
AR	SU2C-PCF	0	0	1	26	0	6	0
AR	COSMIC	0	0	0	5	0	0	0
CDKN1B	cgp	0	0	0	0	3	0	0
ARID1A	COSMIC	0	0	1	2	1	0	0
ARID1A	tcga	0	0	3	3	1	1	0
ARID1A	SU2C-PCF	0	1	0	4	0	0	0
PDE4D	cgp	0	0	0	0	1	0	0
ARID4B	tcga	0	1	0	1	0	2	1
ARID4B	SU2C-PCF	0	0	0	1	0	0	0
ROBO1	cgp	0	0	0	0	0	2	0
ASH1L	tcga	0	3	0	2	3	1	0
ASH1L	SU2C-PCF	0	0	0	0	1	2	0
ASH1L	COSMIC	0	1	0	2	1	0	0
ROBO2	cgp	0	0	0	2	0	1	0
ASXL2	COSMIC	0	0	0	2	2	0	0
ASXL2	tcga	0	1	0	1	0	0	0
ASXL2	SU2C-PCF	0	1	0	3	0	2	0
ZNF292	cgp	0	0	0	1	0	0	0
ATM	SU2C-PCF	0	3	0	1	4	0	0
ATM	tcga	0	3	0	16	0	2	0
ATM	COSMIC	0	3	0	5	0	0	0
FOXA1	cgp	0	3	6	5	0	0	0
BRCA2	tcga	0	4	0	3	0	2	0
BRCA2	COSMIC	0	0	0	1	1	1	0
BRCA2	SU2C-PCF	0	8	0	2	0	0	0
PPAP2A	cgp	0	1	0	0	0	0	0
CDH12	SU2C-PCF	0	0	0	1	0	0	0
CDH12	COSMIC	0	0	0	1	0	1	0
CDH12	tcga	0	0	0	4	1	0	0
ETV3	cgp	0	1	0	2	0	0	0

CDK12	tcga	1	3	0	6	2	1	0
CDK12	COSMIC	0	2	0	1	0	0	0
CDK12	SU2C-PCF	0	5	0	2	2	3	0
ADAM28	cgp	0	1	0	0	0	2	0
IL6ST	cgp	0	1	0	0	0	0	0
CDKN1B	SU2C-PCF	0	2	0	0	0	0	0
CDKN1B	tcga	0	4	0	0	1	0	0
CDKN1B	COSMIC	0	3	0	1	0	1	0
CHD1	tcga	0	3	0	3	1	2	0
CHD1	COSMIC	0	0	0	2	1	0	0
CHD1	SU2C-PCF	0	2	0	2	1	0	0
CTNNB1	COSMIC	0	0	0	4	0	1	0
CTNNB1	tcga	0	0	0	11	0	0	0
CTNNB1	SU2C-PCF	0	0	0	6	0	0	0
MYST3	cgp	0	0	0	0	1	0	0
DLC1	tcga	0	0	0	3	0	1	0
DLC1	COSMIC	0	0	0	2	0	1	0
DLC1	SU2C-PCF	0	0	0	5	0	1	0
PPP2R2A	cgp	1	0	0	0	0	0	0
DOCK10	COSMIC	0	0	0	1	0	0	0
DOCK10	SU2C-PCF	0	0	0	2	1	1	0
DOCK10	tcga	0	0	0	4	2	1	0
AR	cgp	0	0	0	3	0	0	0
APC	cgp	0	1	0	0	1	1	0
ERG	tcga	0	0	0	0	0	1	0
ERG	SU2C-PCF	0	0	0	1	0	0	0
ASH1L	cgp	0	1	0	0	0	0	0
ETV3	tcga	0	1	0	1	1	0	0
ETV3	COSMIC	0	0	0	0	1	0	0
DLC1	cgp	0	0	0	1	0	0	0
FOXA1	SU2C-PCF	0	5	4	9	0	1	0
FOXA1	tcga	0	4	5	12	0	2	0
FOXA1	COSMIC	0	1	0	5	0	0	0
FOXP1	tcga	0	1	1	0	2	2	0
FOXP1	COSMIC	0	0	1	1	0	0	0
NCOR2	cgp	0	0	0	1	0	0	0
IL6ST	tcga	1	2	0	2	0	0	0
IL6ST	COSMIC	0	0	0	1	2	0	0
KDM6A	tcga	3	3	0	2	1	1	0
KDM6A	COSMIC	0	3	0	5	0	1	0
KDM6A	SU2C-PCF	0	2	0	2	0	0	0

CDH12	cgp	0	1	0	0	0	0	0
KMT2C	SU2C-PCF	0	2	0	12	8	5	0
KMT2C	tcga	1	4	0	8	5	3	0
KMT2C	COSMIC	0	2	0	7	1	3	0
KMT2C	cgp	0	2	1	3	2	1	0
MAP3K1	tcga	0	0	0	1	0	0	0
MAP3K1	SU2C-PCF	0	0	0	0	0	1	0
MTUS1	SU2C-PCF	0	0	0	1	0	0	0
MTUS1	tcga	0	0	1	1	1	0	0
SENP6	cgp	0	0	0	1	0	0	0
MYST3	tcga	0	0	0	3	0	0	0
MYST3	COSMIC	0	0	0	1	0	0	0
NCOA7	SU2C-PCF	0	0	0	1	1	0	0
NCOR1	SU2C-PCF	0	1	1	0	0	0	0
NCOR1	tcga	0	3	1	3	1	2	0
NCOR1	COSMIC	0	0	0	2	1	1	0
ANTXR2	cgp	0	1	0	0	0	0	0
ARID4B	cgp	0	0	0	2	0	0	0
NCOR2	COSMIC	0	0	0	0	1	1	0
NCOR2	SU2C-PCF	0	1	0	7	1	9	0
NCOR2	tcga	0	1	1	2	1	0	0
NKX3-1	COSMIC	0	0	0	2	0	0	0
NKX3-1	tcga	0	1	0	4	0	0	0
ASXL2	cgp	1	0	0	1	0	0	0
PDE4D	SU2C-PCF	0	0	0	1	0	0	0
PIK3R1	COSMIC	0	1	0	0	0	0	0
PIK3R1	SU2C-PCF	0	0	1	1	0	0	0
PIK3R1	tcga	0	0	0	1	0	0	0
DOCK10	cgp	0	0	0	2	0	0	0
PPAP2A	tcga	0	0	0	1	1	0	0
PPAP2A	SU2C-PCF	0	1	0	3	0	0	0
PPAP2A	COSMIC	0	0	1	0	0	0	0
RB1	cgp	0	0	0	2	0	0	0
ZFHX3	cgp	0	2	0	1	0	0	0
PTEN	SU2C-PCF	0	5	0	1	3	0	0
PTEN	COSMIC	0	2	1	5	3	0	0
PTEN	tcga	1	7	1	4	1	0	0
TBL1XR1	cgp	2	0	0	1	0	0	0
RB1	COSMIC	0	3	0	0	1	0	0
RB1	tcga	1	1	0	1	0	1	0
RB1	SU2C-PCF	0	3	0	0	1	0	0

ATM	cgp	0	1	0	0	1	0	0
ROBO1	SU2C-PCF	0	0	0	5	0	4	0
ROBO1	tcga	0	1	0	5	0	1	0
BRCA2	cgp	0	0	1	0	0	1	0
ROBO2	COSMIC	0	0	0	3	0	1	0
ROBO2	SU2C-PCF	0	1	0	2	0	0	0
ROBO2	tcga	0	0	0	4	1	0	0
RYBP	tcga	1	0	0	4	0	0	0
CDK12	cgp	0	2	0	0	0	0	0
SENP6	tcga	0	0	0	0	0	1	0
SHQ1	COSMIC	0	0	0	1	0	0	0
SMAD2	COSMIC	0	0	0	1	1	0	0
SMAD2	tcga	0	1	0	1	0	0	0
KDM6A	cgp	0	2	0	1	1	0	0
SMAD4	tcga	0	0	0	4	0	0	0
SMAD4	COSMIC	0	0	0	1	1	0	0
SMAD4	SU2C-PCF	0	0	0	0	0	1	0
NCOR1	cgp	0	0	0	1	0	0	0
ARID1A	cgp	0	1	0	0	1	0	0
TBL1XR1	COSMIC	0	1	0	0	0	0	0
TBL1XR1	SU2C-PCF	0	1	0	0	0	0	0
TBL1XR1	tcga	1	1	0	0	0	0	0
CTNNB1	cgp	0	0	1	1	0	0	0
TP53	COSMIC	0	5	0	31	3	0	0
TP53	SU2C-PCF	0	14	4	33	6	0	0
TP53	tcga	6	7	0	33	0	0	0
USP28	COSMIC	0	0	0	5	0	0	0
USP28	SU2C-PCF	0	3	0	0	1	0	0
USP28	tcga	0	0	1	0	1	0	0
ZBTB16	SU2C-PCF	0	0	0	1	1	1	1
ZBTB16	tcga	0	0	0	2	0	0	0
SMAD2	cgp	1	0	0	1	0	0	0
ZFHX3	COSMIC	0	3	0	5	2	4	0
ZFHX3	SU2C-PCF	0	4	0	4	2	4	0
ZFHX3	tcga	0	6	0	5	3	5	0
ZMYM3	COSMIC	0	3	0	4	0	0	0
ZMYM3	SU2C-PCF	0	1	0	1	0	0	0
ZMYM3	tcga	1	4	0	3	1	0	0
SMAD4	cgp	0	1	0	0	0	0	0
ZMYM3	cgp	0	1	0	1	0	0	0
ZNF292	COSMIC	0	0	0	3	0	0	0

ZNF292	SU2C-PCF	0	1	0	0	1	1	0
ZNF292	tcga	0	1	0	2	2	1	0

stop-lost	chr	start	end	strand	ensg
0	8	24151580	24212726	1	ENSG00000042980
0	21	39751949	40033704	-1	ENSG00000157554
0	4	80898662	80994477	-1	ENSG00000163297
0	4	80898662	80994477	-1	ENSG00000163297
0	10	89622870	89731687	1	ENSG00000171862
0	5	112043195	112203279	1	ENSG00000134982
0	5	112043195	112203279	1	ENSG00000134982
0	5	112043195	112203279	1	ENSG00000134982
0	17	7565257	7590856	-1	ENSG00000141510
0	X	66763874	66950461	1	ENSG00000169083
0	X	66763874	66950461	1	ENSG00000169083
0	X	66763874	66950461	1	ENSG00000169083
0	12	12867992	12875305	1	ENSG00000111276
0	1	27022522	27108601	1	ENSG00000117713
0	1	27022522	27108601	1	ENSG00000117713
0	1	27022522	27108601	1	ENSG00000117713
0	5	58264865	59817947	-1	ENSG00000113448
0	1	235330210	235490802	-1	ENSG00000054267
0	1	235330210	235490802	-1	ENSG00000054267
0	3	78646390	79816965	-1	ENSG00000169855
0	1	155305052	155532324	-1	ENSG00000116539
0	1	155305052	155532324	-1	ENSG00000116539
0	1	155305052	155532324	-1	ENSG00000116539
0	3	77089294	77699114	1	ENSG00000185008
0	2	25960557	26101385	-1	ENSG00000143970
0	2	25960557	26101385	-1	ENSG00000143970
0	2	25960557	26101385	-1	ENSG00000143970
0	6	87862551	87973914	1	ENSG00000188994
0	11	108175402	108186638	1	ENSG00000149311
0	11	108175402	108186638	1	ENSG00000149311
0	11	108175402	108186638	1	ENSG00000149311
0	14	38059189	38064239	-1	ENSG00000129514
0	13	32889611	32973347	1	ENSG00000139618
0	13	32889611	32973347	1	ENSG00000139618
0	13	32889611	32973347	1	ENSG00000139618
0	5	54720682	54830878	-1	ENSG00000067113
0	5	21750777	22853731	-1	ENSG00000154162
0	5	21750777	22853731	-1	ENSG00000154162
0	5	21750777	22853731	-1	ENSG00000154162
0	1	157090983	157108266	-1	ENSG00000117036

0	17	37617739	37690800	1	ENSG00000167258
0	17	37617739	37690800	1	ENSG00000167258
0	17	37617739	37690800	1	ENSG00000167258
0	8	24151580	24212726	1	ENSG00000042980
0	5	55230925	55290821	-1	ENSG00000134352
0	12	12867992	12875305	1	ENSG00000111276
0	12	12867992	12875305	1	ENSG00000111276
0	12	12867992	12875305	1	ENSG00000111276
0	5	98190908	98262240	-1	ENSG00000153922
0	5	98190908	98262240	-1	ENSG00000153922
0	5	98190908	98262240	-1	ENSG00000153922
0	3	41236328	41301587	1	ENSG00000153922
0	3	41236328	41301587	1	ENSG00000153922
0	3	41236328	41301587	1	ENSG00000153922
0	8	41799567	41909505	-1	ENSG00000083168
0	8	13162116	13372429	-1	ENSG00000164741
0	8	13162116	13372429	-1	ENSG00000164741
0	8	13162116	13372429	-1	ENSG00000164741
0	8	26149007	26230196	1	ENSG00000221914
0	2	225629807	225907330	-1	ENSG00000135905
0	2	225629807	225907330	-1	ENSG00000135905
0	2	225629807	225907330	-1	ENSG00000135905
0	X	66763874	66950461	1	ENSG00000169083
0	5	112043195	112203279	1	ENSG00000134982
0	21	39751949	40033704	-1	ENSG00000157554
0	21	39751949	40033704	-1	ENSG00000157554
0	1	155305052	155532324	-1	ENSG00000116539
0	1	157090983	157108266	-1	ENSG00000117036
0	1	157090983	157108266	-1	ENSG00000117036
0	8	13162116	13372429	-1	ENSG00000164741
0	14	38059189	38064239	-1	ENSG00000129514
1	14	38059189	38064239	-1	ENSG00000129514
0	14	38059189	38064239	-1	ENSG00000129514
0	3	71003844	71633140	-1	ENSG00000114861
0	3	71003844	71633140	-1	ENSG00000114861
0	12	124808957	125052010	-1	ENSG00000196498
0	5	55230925	55290821	-1	ENSG00000134352
0	5	55230925	55290821	-1	ENSG00000134352
0	X	44732423	44971847	1	ENSG00000147050
0	X	44732423	44971847	1	ENSG00000147050
0	X	44732423	44971847	1	ENSG00000147050

0	5	21750777	22853731	-1	ENSG00000154162
0	7	151832007	152133628	-1	ENSG00000055609
0	7	151832007	152133628	-1	ENSG00000055609
0	7	151832007	152133628	-1	ENSG00000055609
0	7	151832007	152133628	-1	ENSG00000055609
0	5	56111401	56191979	1	ENSG00000095015
0	5	56111401	56191979	1	ENSG00000095015
0	8	17501304	17658426	-1	ENSG00000129422
0	8	17501304	17658426	-1	ENSG00000129422
0	6	76311225	76427997	1	ENSG00000112701
0	8	41799567	41909505	-1	ENSG00000083168
0	8	41799567	41909505	-1	ENSG00000083168
0	6	126111909	126253176	1	ENSG00000111912
0	17	15933408	16118874	-1	ENSG00000141027
0	17	15933408	16118874	-1	ENSG00000141027
0	17	15933408	16118874	-1	ENSG00000141027
0	4	80898662	80994477	-1	ENSG00000163297
0	1	235330210	235490802	-1	ENSG00000054267
0	12	124808957	125052010	-1	ENSG00000196498
0	12	124808957	125052010	-1	ENSG00000196498
0	12	124808957	125052010	-1	ENSG00000196498
0	8	23536206	23540402	-1	ENSG00000167034
0	8	23536206	23540402	-1	ENSG00000167034
0	2	25960557	26101385	-1	ENSG00000143970
0	5	58264865	59817947	-1	ENSG00000113448
0	5	67511548	67597649	1	ENSG00000145675
0	5	67511548	67597649	1	ENSG00000145675
0	5	67511548	67597649	1	ENSG00000145675
0	2	225629807	225907330	-1	ENSG00000135905
0	5	54720682	54830878	-1	ENSG00000067113
0	5	54720682	54830878	-1	ENSG00000067113
0	5	54720682	54830878	-1	ENSG00000067113
0	13	48877887	49056122	1	ENSG00000139687
0	16	72816784	73082274	-1	ENSG00000140836
0	10	89622870	89731687	1	ENSG00000171862
0	10	89622870	89731687	1	ENSG00000171862
0	10	89622870	89731687	1	ENSG00000171862
0	3	176738542	176915048	-1	ENSG00000177565
0	13	48877887	49056122	1	ENSG00000139687
0	13	48877887	49056122	1	ENSG00000139687
0	13	48877887	49056122	1	ENSG00000139687

0	11	108175402	108186638	1	ENSG00000149311
0	3	78646390	79816965	-1	ENSG00000169855
0	3	78646390	79816965	-1	ENSG00000169855
0	13	32889611	32973347	1	ENSG00000139618
0	3	77089294	77699114	1	ENSG00000185008
0	3	77089294	77699114	1	ENSG00000185008
0	3	77089294	77699114	1	ENSG00000185008
0	3	72423744	72911065	-1	ENSG00000163602
0	17	37617739	37690800	1	ENSG00000167258
0	6	76311225	76427997	1	ENSG00000112701
0	3	72798428	72911065	-1	ENSG00000144736
0	18	45359466	45457517	-1	ENSG00000175387
0	18	45359466	45457517	-1	ENSG00000175387
0	X	44732423	44971847	1	ENSG00000147050
0	18	48556583	48611415	1	ENSG00000141646
0	18	48556583	48611415	1	ENSG00000141646
0	18	48556583	48611415	1	ENSG00000141646
0	17	15933408	16118874	-1	ENSG00000141027
0	1	27022522	27108601	1	ENSG00000117713
0	3	176738542	176915048	-1	ENSG00000177565
0	3	176738542	176915048	-1	ENSG00000177565
0	3	176738542	176915048	-1	ENSG00000177565
0	3	41236328	41301587	1	ENSG00000153922
0	17	7565257	7590856	-1	ENSG00000141510
1	17	7565257	7590856	-1	ENSG00000141510
0	17	7565257	7590856	-1	ENSG00000141510
0	11	113668598	113746256	-1	ENSG00000048028
0	11	113668598	113746256	-1	ENSG00000048028
0	11	113668598	113746256	-1	ENSG00000048028
0	11	113931288	114121397	1	ENSG00000109906
0	11	113931288	114121397	1	ENSG00000109906
0	18	45359466	45457517	-1	ENSG00000175387
0	16	72816784	73082274	-1	ENSG00000140836
0	16	72816784	73082274	-1	ENSG00000140836
0	16	72816784	73082274	-1	ENSG00000140836
0	X	70459474	70474996	-1	ENSG00000147130
0	X	70459474	70474996	-1	ENSG00000147130
0	X	70459474	70474996	-1	ENSG00000147130
0	18	48556583	48611415	1	ENSG00000141646
0	X	70459474	70474996	-1	ENSG00000147130
0	6	87862551	87973914	1	ENSG00000188994

0	6	87862551	87973914	1	ENSG00000188994
0	6	87862551	87973914	1	ENSG00000188994

gene	source	chr	start	end	strand	n_rearr_icgc
ERG	cgp	21	39751949	40033704	-1	56
PTEN	cgp	10	89622870	89731687	1	28
TP53	cgp	17	7565257	7590856	-1	19
CDKN1B	cgp	12	12867992	12875305	1	15
PDE4D	cgp	5	58264865	59817947	-1	13
ROBO1	cgp	3	78646390	79816965	-1	11
ROBO2	cgp	3	77089294	77699114	1	11
ZNF292	cgp	6	87862551	87973914	1	11
FOXA1	cgp	14	38059189	38064239	-1	10
PPAP2A	cgp	5	54720682	54830878	-1	10
ETV3	cgp	1	157090983	157108266	-1	9
ADAM28	cgp	8	24151580	24212726	1	8
IL6ST	cgp	5	55230925	55290821	-1	8
MYST3	cgp	8	41799567	41909505	-1	8
PPP2R2A	cgp	8	26149007	26230196	1	8
AR	cgp	X	66763874	66950461	1	7
APC	cgp	5	112043195	112203279	1	6
ASH1L	cgp	1	155305052	155532324	-1	6
DLC1	cgp	8	13162116	13372429	-1	6
NCOR2	cgp	12	124808957	125052010	-1	6
CDH12	cgp	5	21750777	22853731	-1	5
KMT2C	cgp	7	151832007	152133628	-1	5
SENP6	cgp	6	76311225	76427997	1	5
ANTXR2	cgp	4	80898662	80994477	-1	4
ARID4B	cgp	1	235330210	235490802	-1	4
ASXL2	cgp	2	25960557	26101385	-1	4
DOCK10	cgp	2	225629807	225907330	-1	4
RB1	cgp	13	48877887	49056122	1	4
ZFHX3	cgp	16	72816784	73082274	-1	4
TBL1XR1	cgp	3	176738542	176915048	-1	3
ATM	cgp	11	108175402	108186638	1	2
BRCA2	cgp	13	32889611	32973347	1	2
CDK12	cgp	17	37617739	37690800	1	2
KDM6A	cgp	X	44732423	44971847	1	2
NCOR1	cgp	17	15933408	16118874	-1	2
ARID1A	cgp	1	27022522	27108601	1	1
CTNNB1	cgp	3	41236328	41301587	1	1
SMAD2	cgp	18	45359466	45457517	-1	1
SMAD4	cgp	18	48556583	48611415	1	1
ZMYM3	cgp	X	70459474	70474996	-1	1

rearr_icgc_su	ensg	homodel_icg	multiple_hit_	multiple_hit_	multiple_hit_	multiple_hit_
PD9168a<TM	ENSG000001	0	1	PD13412a	1	PD13412a
PD11331a<sn	ENSG000001	PD12808a,PD	5	PD12844a,PD	5	PD13392a,PD
PD11329a<sn	ENSG000001	PD11329d	1	PD14728a	12	PD12818a,PD
PD11332a<sn	ENSG000001	0	1	PD12809a	0	
PD11330a<sn	ENSG000001	PD14734a,PD	1	PD14728a	1	PD14728a
PD11332a<sn	ENSG000001	0	0	-	0	
PD11328a<tr	ENSG000001	0	0	-	0	
PD12817a<sn	ENSG000001	PD14728a,PD	0	-	0	
PD11328a<ta	ENSG000001	0	4	PD12337a,PD	0	
PD11332a<tr	ENSG000000	PD14734a	0	-	0	
PD12337a<sn	ENSG000001	0	0	-	0	
PD12829a<sn	ENSG000000	0	0	-	1	PD11332a
PD12817a<sn	ENSG000001	PD14734a,PD	0	-	1	PD14726a
PD12808a<sn	ENSG000000	0	0	-	1	PD12842a
PD12829a<sn	ENSG000002	PD13412c,PD	0	-	0	
PD12845a<cc	ENSG000001	0	2	PD12337a,PD	0	
PD12809a<sn	ENSG000001	0	0	-	2	PD14728a,PD
PD11328a<tr	ENSG000001	0	0	-	0	
PD12836a<sn	ENSG000001	PD13408a	0	-	1	PD13388a
PD14728a<sn	ENSG000001	0	0	-	0	
PD11328a<cc	ENSG000001	0	0	-	0	
PD11328a<cc	ENSG000000	PD13412a	0	-	0	
PD12838a<sn	ENSG000001	0	0	-	1	PD11329a
PD13390a<sn	ENSG000001	0	0	-	0	
PD12836a<sn	ENSG000000	0	0	-	0	
PD11331a<sn	ENSG000001	PD14706a	1	PD11331a	1	PD11331a
PD11332a<sn	ENSG000001	0	0	-	0	
PD11335a<sn	ENSG000001	PD12834a,PD	1	PD11335a	2	PD11329a,PD
PD11330a<tr	ENSG000001	0	0	-	0	
PD13387a<1k	ENSG000001	PD14717a	0	-	0	
PD9169a<inv	ENSG000001	0	1	PD13401a	1	PD12843a
PD11335a<sn	ENSG000001	PD14731a,PD	0	-	2	PD12829a,PD
PD14722a<sn	ENSG000001	0	0	-	0	
PD11328a<sn	ENSG000001	0	1	PD11328a	0	
PD13380a<sn	ENSG000001	PD14725a	0	-	1	PD12838a
PD14728a<sn	ENSG000001	0	0	-	0	
PD11328a<ta	ENSG000001	0	0	-	0	
PD14706a<sn	ENSG000001	0	0	-	0	
PD14727a<in	ENSG000001	0	0	-	1	PD11335a
PD11333a<sn	ENSG000001	0	0	-	0	

multiple_hit_	multiple_hit_all_summary
1	PD13412a
7	PD12844a,PD13387a,PD12840a,PD14718a,PD11334a,PD13392a,PD14724a
12	PD14728a,PD12818a,PD12806a,PD11331a,PD9168a,PD14730a,PD14727a,PD1133
1	PD12809a
1	PD14728a
0	
0	
0	
4	PD12337a,PD13406a,PD14726a,PD11328a
0	
0	
1	PD11332a
1	PD14726a
1	PD12842a
0	
2	PD12337a,PD11334a
2	PD14728a,PD14713a
0	
1	PD13388a
0	
0	
0	
1	PD11329a
0	
0	
1	PD11331a
0	
2	PD11335a,PD11329a
0	
0	
2	PD13401a,PD12843a
2	PD12829a,PD14722a
0	
1	PD11328a
1	PD12838a
0	
0	
0	
1	PD11335a
0	

3a,PD11334a,PD14713a,PD12845a,PD11332a

GENE	SYN	MIS	NON	SPLICE	INDEL	dNdS_mis	dNdS_NON
TP53	0	105	10	6	33	147.432737	253.692241
SPOP	0	94	0	0	3	141.013749	0
PTEN	0	16	7	6	17	25.4268052	136.30721
FOXA1	3	30	0	0	29	9.81109971	0
CDKN1B	1	1	4	0	9	1.6831457	107.081529
MLL2	5	24	12	1	23	2.13464222	17.1632806
APC	2	4	14	0	14	0.70166849	26.8214594
CDK12	4	10	4	1	13	1.9532142	10.1645074
MLL3	13	39	16	0	15	1.4316258	7.27393243
KDM6A	1	6	2	3	8	2.59995615	9.88863037
ZFHX3	14	15	7	1	15	0.6685532	5.64165771
ATM	2	22	5	0	10	3.56247943	9.55277643
RB1	1	3	2	2	5	1.51157057	9.86031155
CHD1	2	7	3	3	5	1.52378026	7.19824641
ZMYM3	0	5	1	1	6	3.43593539	11.8402402
TBL1XR1	0	1	0	2	4	1.12232122	0
AR	6	34	0	0	1	5.90940047	0
ARID1A	1	9	3	0	6	2.64331568	16.3404042
BRCA2	4	7	1	0	13	0.80289319	1.69629446
IL6ST	0	4	2	1	3	3.15735344	20.6534624
CTNNB1	1	22	0	0	0	10.7914381	0
CASZ1	1	6	2	0	9	1.81674498	19.3498255
RNF43	1	4	3	0	2	2.07003675	22.399243
PIK3CA	3	23	0	0	1	5.71483253	0
SMAD2	0	3	1	1	1	4.33979878	14.7113888
FOXP1	2	1	2	1	3	0.39154464	9.62607296
PIK3R1	0	2	0	2	2	1.50442841	0
TBX3	0	9	0	0	3	5.0925344	0
BRAF	0	11	0	0	1	8.04169209	0
IDH1	0	9	0	0	0	10.9808339	0
HRAS	0	6	0	0	0	14.8899243	0
CNOT3	1	12	1	0	0	5.3551647	8.93810103
NDST4	1	13	0	0	1	4.71827098	0
RPL11	0	6	0	0	1	18.3361561	0
LCE2B	0	6	0	0	0	20.6936968	0

dNdS_SPLICE	dNdS_INDEL	pMIS	pALL	qMIS	qALL
222.714656	249.191884	0	0	0	0
0	23.90303	0	0	0	0
208.43008	126.399663	3.55E-10	0	0	0
0	195.094571	1.05E-09	2.47E-07	0	0
0	141.209958	0.12944	6.53E-06	0.16	0
4.99600593	12.979872	0.013501	3.97E-09	0.157	0
0	15.6104792	0.077096	1.13E-08	0.157	0
11.1802402	27.2814769	0.035413	0.000404	0.157	0
0	9.47480798	0.040913	1.26E-06	0.157	0
26.6982637	17.2254931	0.038754	7.88E-05	0.157	0
11.491477	12.9629448	0.062368	0.003805	0.157	0
0	9.88737324	0.001329	0.000141	0.06	0
14.1298896	15.8337048	0.086896	0.00151	0.157	0
14.2079143	8.82794742	0.06582	0.000228	0.157	0
18.6788792	13.3435247	0.032762	0.007936	0.157	0.001
38.2370089	22.9809546	0.095498	0.008221	0.157	0.001
0	3.40510979	1.39E-07	1.96E-05	0	0.003
0	8.22957819	0.027956	0.003337	0.157	0.003
0	11.9662882	0.086593	0.951295	0.157	0.003
16.6133537	9.99043049	0.029148	0.00079	0.157	0.003
0	0	7.60E-08	1.15E-05	0	0.005
0	15.9572425	0.093542	0.023145	0.188	0.013
0	7.92852346	0.057405	0.00265	0.157	0.029
0	2.8414227	6.78E-06	0.000568	0.001	0.047
34.3258381	6.43665368	0.020019	0.001868	0.157	0.068
6.61571033	13.2009916	0.072952	0.074323	0.157	0.068
36.4268504	8.32750722	0.097486	0.009524	0.157	0.072
0	12.5614972	0.003977	0.086703	0.131	0.081
0	3.90194564	0.000172	0.008109	0.01	0.277
0	0	4.39E-05	0.002694	0.003	0.347
0	0	6.01E-05	0.003482	0.004	0.383
0	0	0.000688	0.004104	0.036	0.43
0	3.52154498	0.001038	0.032659	0.05	0.617
0	16.2156234	1.61E-05	0.001172	0.026	0.679
0	0	8.49E-06	0.000686	0.017	1

region	exp_subs	exp_indels	obs_subs	obs_indels	local_t	local_t_ind
Exon_FOXA1	0.288113	0.001691	5	8	0.915694	10.03441
Exon_TP53	0.18919	0.001598	7	3	0.55651	5.325007
Exon_SPOP	0.138802	0.001392	11	0	0.700511	6.543536
Exon_PTEN	0.128824	0.001406	7	2	0.619405	5.688691
UTR3_FOXA1	0.147241	0.002362	3	5	0.915694	10.03441
Exonflanks_ARSD	0.157829	0.00236	8	0	0.526025	2.565645
ENSG00000245532 (NEAT1	3.890444	0.049559	9	6	0.604275	5.619608
Exon_CDK12	0.539239	0.005222	0	3	0.578958	7.817554
ENSG00000182021	2.401588	0.03194	13	0	0.893604	22.28641
ENSG00000229178	0.53852	0.005144	1	3	0.866064	6.453148
Exon_TBX3	0.403789	0.002589	0	2	0.953099	6.451043
Exon_CDKN1B	0.117888	0.000858	3	0	0.647826	7.143755
ENSG00000258414	2.100838	0.032824	3	5	0.982415	10.03441
UTR3_PDE4DIP	0.900626	0.015101	5	2	0.682246	14.82315
ENSG00000264734	1.706188	0.023407	9	1	1.042655	9.322698
ENSG00000264698	0.009584	0.000157	2	0	1.0041	7.806753
ENSG00000238385	0.011969	0.0002	2	0	0.930626	10.59962
ENSG00000235887	0.225929	0.002884	2	2	1.295134	33.94427
ENSG00000261466	1.088951	0.013304	11	0	1.984319	44.83325
ENSG00000232527	3.988412	0.050662	12	5	1.246748	31.93861
ENSG00000249388	3.531467	0.044789	6	4	0.783395	8.080765
Exon_EIF1AX	0.049635	0.000505	2	0	0.554568	2.488115
Exon_DDX6	0.160903	0.001684	1	1	0.67603	6.913211
Exon_TPM4	0.191569	0.001464	1	1	0.615317	5.897858
Exon_MDM2	0.166635	0.001914	3	0	0.814391	6.131522
UTR5_CREBRF	0.018882	0.000345	1	1	0.779091	5.996343
Exon_TFF3	0.072721	0.000654	4	0	0.599197	4.369566
UTR3_POM121L7	0.089106	0.001246	4	0	0.694481	10.1365
UTR3_PRRC2C	0.169462	0.002724	4	1	1.051891	9.221754
Promoter_STIL	0.189626	0.002708	2	1	0.621343	6.540899
Promoter_FOXP1	0.547081	0.007785	0	3	0.784195	6.510009
Promoter_FOXA2	0.121902	0.001576	4	0	1.051928	7.740274

pval_subs	pval_indels	pval_both	qval_subs	qval_indels	qval_both	exclude_ge	cv_predicte
0.000319	6.05E-05	3.62E-07	0.537846	0.044211	0.001833	TRUE	0.314365
8.83E-07	0.000177	3.68E-09	0.005959	0.044211	2.49E-05	TRUE	0.148177
1.07E-11	1	2.80E-10	2.16E-07	1	4.80E-06	TRUE	0.124167
8.00E-08	0.00023	4.74E-10	0.00081	0.044211	4.80E-06	TRUE	0.106928
0.00113	0.000144	2.70E-06	1	0.029464	0.026358	TRUE	0.156821
2.10E-08	1	3.92E-07	0.000424	1	0.00792	FALSE	0.124463
0.070491	0.000405	0.000327	1	0.005583	0.035152	FALSE	2.400613
1	0.000351	0.003142	1	0.044211	0.299535	TRUE	0.431183
0.00059	1	0.00498	0.901242	1	0.036212	FALSE	1.840881
0.394703	0.000256	0.00103	1	0.005583	0.035152	FALSE	0.404354
1	0.000321	0.002901	1	0.044211	0.299535	TRUE	0.456058
0.000885	1	0.007109	0.716933	1	0.390017	TRUE	0.100452
0.340483	0.000383	0.001298	1	0.005583	0.035292	FALSE	1.72122
0.011521	0.000538	8.05E-05	1	0.029464	0.1519	TRUE	0.815375
0.001956	0.000563	1.62E-05	1	0.005583	0.034651	FALSE	1.462461
5.80E-05	1	0.000624	0.310037	1	0.035152	FALSE	0.007981
9.03E-05	1	0.000931	0.321596	1	0.035152	FALSE	0.009433
0.026002	0.000249	8.38E-05	1	0.005583	0.034859	FALSE	0.234011
1.13E-05	1	0.00014	0.121131	1	0.034859	FALSE	1.890893
0.019598	0.000428	0.000106	1	0.005583	0.034859	FALSE	3.983905
0.199386	0.000441	0.000909	1	0.005583	0.035152	FALSE	2.492293
0.001966	1	0.014217	1	1	0.763467	TRUE	0.038806
0.141141	0.000416	0.000631	1	0.044211	0.299535	TRUE	0.140721
0.164168	0.000394	0.000688	1	0.044211	0.299535	TRUE	0.15841
0.00225	1	0.01597	1	1	0.850812	TRUE	0.165587
0.01825	1.06E-05	3.17E-06	1	0.004252	0.030395	FALSE	0.017694
1.26E-05	1	0.000155	0.063923	1	0.258294	FALSE	0.059245
1.37E-05	1	0.000167	0.26772	1	0.181396	FALSE	0.081361
0.000152	0.000438	1.16E-06	1	0.029464	0.022699	FALSE	0.198432
0.024531	0.000593	0.000176	1	0.054112	0.402543	TRUE	0.169028
1	0.00049	0.004223	1	0.054112	0.402543	TRUE	0.546891
0.000115	1	0.001162	0.681941	1	0.402543	TRUE	0.147135

cv_predicte	pval_subs_	pval_indels	pval_both_	qval_subs_	qval_indels	qval_both_	qval_both_
0.015625	0.000255	1.84E-16	0	0.420988	3.73E-12	0	0
0.011529	6.03E-08	8.19E-07	1.56E-12	0.000407	0.00829	1.58E-08	5.58E-10
0.010704	3.60E-13	1	1.07E-11	7.27E-09	1	7.19E-08	2.54E-09
0.010336	7.18E-09	8.19E-05	1.71E-11	7.26E-05	0.414184	8.67E-08	3.06E-09
0.024018	0.001191	2.01E-10	7.20E-12	1	3.92E-06	1.40E-07	5.13E-09
0.014794	1.63E-09	1	3.46E-08	3.29E-05	1	0.000698	0.000673
0.36631	0.0029	1.54E-05	8.02E-07	1	0.084275	0.008568	0.008568
0.042939	1	3.98E-05	0.000443	1	0.268395	1	0.063262
0.446832	2.53E-06	1	3.51E-05	0.027004	1	0.18752	0.18752
0.039252	0.328037	1.58E-05	6.82E-05	1	0.084275	0.242755	0.242755
0.019813	1	0.000297	0.002709	1	0.750953	1	0.32233
0.006812	0.000429	1	0.003758	0.511003	1	1	0.383354
0.287283	0.252791	4.68E-05	0.000146	1	0.166643	0.390034	0.390034
0.199246	0.006834	0.019098	0.001298	1	1	1	0.462711
0.199359	0.000118	0.177736	0.000247	0.252156	1	0.527597	0.527597
0.00126	3.43E-05	1	0.000387	0.135176	1	0.63947	0.63947
0.001791	4.79E-05	1	0.000525	0.135176	1	0.63947	0.63947
0.063045	0.024838	0.002231	0.000599	1	1	0.63947	0.63947
0.441212	5.06E-05	1	0.000551	0.135176	1	0.63947	0.63947
1.025565	0.004631	0.009469	0.000484	1	1	0.63947	0.63947
0.363764	0.057831	0.001146	0.000704	1	1	0.683568	0.683568
0.003135	0.001081	1	0.008465	0.948785	1	1	0.755504
0.013206	0.127053	0.013072	0.012291	1	1	1	0.819501
0.010881	0.141278	0.01079	0.011412	1	1	1	0.819501
0.014403	0.001713	1	0.012625	1	1	1	0.819501
0.002554	0.017234	0.000453	9.96E-05	1	0.23587	0.954002	0.911757
0.004488	3.46E-06	1	4.69E-05	0.017491	1	0.189952	0.916242
0.012737	7.66E-06	1	9.79E-05	0.149348	1	0.63621	0.919425
0.026499	0.000218	0.026105	7.44E-05	0.808592	1	0.63621	0.919425
0.023011	0.019765	0.022293	0.003845	1	1	1	0.952062
0.066035	1	0.000461	0.004	1	1	1	0.952062
0.014382	0.000206	1	0.001954	1	1	1	0.952062

source
Noncoding_CDS
Noncoding_CDS
Noncoding_CDS
Noncoding_CDS
Noncoding_UTR3
Noncoding_ExonFlanks
Noncoding_ncRNA
Noncoding_CDS
Noncoding_ncRNA
Noncoding_ncRNA
Noncoding_CDS
Noncoding_CDS
Noncoding_ncRNA
Noncoding_UTR3
Noncoding_ncRNA
Noncoding_ncRNA
Noncoding_ncRNA
Noncoding_ncRNA
Noncoding_ncRNA
Noncoding_ncRNA
Noncoding_ncRNA
Noncoding_CDS
Noncoding_CDS
Noncoding_CDS
Noncoding_CDS
Noncoding_UTR5
Noncoding_CDS
Noncoding_UTR3
Noncoding_UTR3
Noncoding_Promoters
Noncoding_Promoters
Noncoding_Promoters

SUPPLEMENTARY TABLE 2

previous evidence	gene	mutation_type	prior_evidence	evidence_in_our_study	tumor suppressor / oncogene?
low	ADAM28	SV, CNA	1 biological evidence	SVs and CNA in ETS+	tumor suppressor
low	ANTXR2	SV,SNV/indel	none	clinical correlation	tumor suppressor
low	ASH1L	SV,SNV/indel	2	truncating mutations, SVs in ETS-	tumor suppressor
low	CDH12	SV	none	clinical correlation	tumor suppressor
low	FOXO1	CNA	3 biological evidence	CNA in ETS-	tumor suppressor
low	IL6ST	SV	4 biological evidence	dN/dS, SVs and CNA in ETS+, clinical c	tumor suppressor
low	LCE2B	SNV/indel	none	dN/dS (missense)	tumor suppressor
low	MAP3K1	SV, CNA	none	SVs, CNA in ETS+	tumor suppressor
low	MYST3	SV	2	SVs in ETS-, RNA expression	tumor suppressor
low	NCOA7	SV	none	SVs in ETS-	tumor suppressor
low	NDST4	SNV/indel	none	dN/dS (missense)	oncogene
low	NEAT1	Non-coding	5 biological evidence	non-coding	oncogene
low	PDE4D	SV	6 SNP data	SVs and CNA in ETS+	tumor suppressor
low	PPAP2A	SV	6 SNP data	SVs and CNA in ETS+	tumor suppressor
low	PPP2R2A	SV	7 biological evidence	SVs and CNAs in ETS+	tumor suppressor
low	ROBO1	SV	8 biological evidence	SVs in ETS+	tumor suppressor
low	ROBO2	SV	2	SVs in ETS+	tumor suppressor
low	RPL11	SNV/indel	2	dN/dS (missense)	oncogene
low	SENP6	SV	9 biological evidence	enriched SVs, RNA expression	tumor suppressor
low	TBL1XR1	SNV/indel,SV	10 known AR co-regulator, biolo	dN/dS	tumor suppressor
low	USP28	SV, CNA, SNV/indel	none	SVs, CNA, SNV/indel	tumor suppressor
low	ZNF292	SV,SNV/indel, CNA	2	enriched SVs, homozygous deletions, tr	tumor suppressor
medium	ARID1A	SNV/indel	11	dN/dS	tumor suppressor
medium	ASXL2	SNV/indel	TCGA, COSMIC and SU2C	Not identified in this study	tumor suppressor
medium	CASZ1	SNV/indel	COSMIC, TCGA and SU2C	dN/dS	tumor suppressor
medium	CNOT3	SNV/indel	12 mutated in leukemia	dN/dS (missense)	oncogene
medium	DOCK10	SNV/indel,SV	TCGA, COSMICand SUC2C	Not identified in this study	tumor suppressor
medium	LRP1B	SV, CNA	6	SVs and CNA in ETS-	tumor suppressor
medium	PIK3R1	SNV/indel	13	dN/dS	tumor suppressor
medium	RGMB	CNA	14	CNA in ETS-	tumor suppressor
medium	SMAD4	SNV/indel	15	Not identified in this study	tumor suppressor

medium	SMAD4	SNV/indel	16	Not identified in this study	tumor suppressor
medium	TBX3	SNV/indel	known breast cancer gene	dN/dS	tumor suppressor
medium	ZMYM3	SNV/indel	COSMIC SU2C	dN/dS	tumor suppressor
high	AKT1	SNV/indel	known prostate cancer gene	Not identified in this study	
high	APC	SV,SNV/indel	known prostate cancer gene	dN/dS, enriched in metastases	
high	AR	SV,SNV/indel	known prostate cancer gene	dN/dS, enriched in metastases	
high	ARID4B	SV,SNV/indel, CNA	TCGA	LOH	
high	ATM	SNV/indel	17 18	dN/dS	
high	BRAF	SNV/indel	19	dN/dS (missense)	
high	BRCA2	SNV/indel,SV, CNA	17 18	dN/dS, enriched in metastases, CNA in	
high	CDK12	SNV/indel	known prostate cancer gene	dN/dS, enriched in metastases	
high	CDKN1B	SV,SNV/indel	known tumor supressor gene	dN/dS	
high	CHD1	SV	known prostate cancer gene	dN/dS, SVs and CNA in ETS-	
high	CSMD3	SV	20	SVs in ETS-	
high	CTNNB1	SNV/indel	known prostate cancer gene	dN/dS, enriched in metastases	
high	DLC1	SNV/indel	known prostate cancer gene	clinical correlation	
high	ERG	SV	known prostate cancer gene	enriched SVs and CNA	
high	ETV3	SV, CNAs, SNV/indel	16	SVs, CNAs, SNS/indel in ETS-	tumor supressor
high	FOXA1	SV,SNV/indel, non-coding	known prostate cancer gene	dN/dS, non-coding	oncogene
high	FOXP1	SV	known prostate cancer gene	dN/dS, SVs in ETS+	
high	HRAS	SNV/indel	Known cancer gene	dN/dS (missense)	
high	IDH1	SNV/indel	known cancer gene	dN/dS (missense)	
high	KDM6A	SNV/indel,SV	known prostate cancer gene	dN/dS	
high	KMT2C	SNV/indel	known prostate cancer gene	enriched in metastases	
high	KMT2D	SNV/indel	known prostate cancer gene	enriched in metastases	
high	MLL2	SNV/indel	16	dN/dS	
high	MLL3	SNV/indel	known cancer gene	dN/dS	
high	MTUS1	SNV/indel	known prostate cancer gene	clinical correlation	
high	NCOR1	SNV/indel	known prostate cancer gene	known prostate cancer gene	
high	NCOR2	SNV/indel,SV	known AR regulator	known AR regulator	
high	NKX3-1	SV	known prostate cancer gene	enriched SVs	
high	PIK3CA	SNV/indel	known prostate cancer gene	dN/dS	
high	PIK3CB	SNV/indel	known prostate cancer gene	enriched in metastases	

high	PTEN	SV,SNV/indel	known prostate cancer gene	dN/dS, CNA in ETS+
high	RB1	SNV/indel	known prostate cancer gene	dN/dS, enriched in metastases
high	RNF43	SNV/indel	18	dN/dS
high	RYBP	SV	known prostate cancer gene	SVs in ETS+
high	SHQ1	SV	known prostate cancer gene	SVs in ETS+
high	SMAD2	SNV/indel	known cancer gene	dN/dS
high	SPOP	SNV/indel	known prostate cancer gene	dN/dS, clinical correlation
high	TP53	SNV/indel,SV	known prostate cancer gene	dN/dS, enriched in metastases, SVs in ETS+
high	ZBTB16	SV	reported SVs in 25808865	SVs
high	ZFHX3	SNV/indel,SV	known prostate cancer gene	dN/dS, enriched in metastases

1 Rudnicka, C. *et al.* Overexpression and knock-down studies highlight that a disintegrin and metalloproteinase 28 controls proliferation and migratic
2 Baca, S.C. *et al.* Punctuated evolution of prostate cancer genomes. *Cell* **153**, 666-77 (2013).
3 Zhang, H. *et al.* FOXO1 inhibits Runx2 transcriptional activity and prostate cancer cell migration and invasion. *Cancer Res* **71**, 3257-67 (2011).
4 Malinowska, K. *et al.* Interleukin-6 stimulation of growth of prostate cancer in vitro and in vivo through activation of the androgen receptor. *Endocr*
5 Chakravarty, D. *et al.* The oestrogen receptor alpha-regulated lncRNA NEAT1 is a critical modulator of prostate cancer. *Nat Commun* **5**, 5383 (20
6 FitzGerald, L.M. *et al.* Identification of a prostate cancer susceptibility gene on chromosome 5p13q12 associated with risk of both familial and spor
7 Zhao, W., Cao, L., Zeng, S., Qin, H. & Men, T. Upregulation of miR-556-5p promoted prostate cancer cell proliferation by suppressing PPP2R2A e
8 Parray, A. *et al.* ROBO1, a tumor suppressor and critical molecular barrier for localized tumor cells to acquire invasive phenotype: study in African
9 Kaikkonen, S. *et al.* SUMO-specific protease 1 (SENP1) reverses the hormone-augmented SUMOylation of androgen receptor and modulates ger
10 Daniels, G. *et al.* TBLR1 as an androgen receptor (AR) coactivator selectively activates AR target genes to inhibit prostate cancer growth. *Endocr*
11 Jones, S. *et al.* Somatic mutations in the chromatin remodeling gene ARID1A occur in several tumor types. *Hum Mutat* **33**, 100-3 (2012).
12 Collart, M.A., Kassem, S. & Villanyi, Z. Mutations in the NOT Genes or in the Translation Machinery Similarly Display Increased Resistance to Hist
13 Gao, D. *et al.* Organoid cultures derived from patients with advanced prostate cancer. *Cell* **159**, 176-87 (2014).
14 Burkhardt, L. *et al.* CHD1 is a 5q21 tumor suppressor required for ERG rearrangement in prostate cancer. *Cancer Res* **73**, 2795-805 (2013).
15 Ding, Z. *et al.* SMAD4-dependent barrier constrains prostate cancer growth and metastatic progression. *Nature* **470**, 269-73 (2011).
16 Grasso, C.S. *et al.* The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487**, 239-43 (2012).
17 Mateo, J. *et al.* DNA-Repair Defects and Olaparib in Metastatic Prostate Cancer. *N Engl J Med* **373**, 1697-708 (2015).
18 Robinson, D. *et al.* Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215-28 (2015).
19 Taylor, B.S. *et al.* Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11-22 (2010).
20 Berger, M.F. *et al.* The genomic complexity of primary human prostate cancer. *Nature* **470**, 214-20 (2011).

on in human prostate cancer. *Medicine (Baltimore)* **95**, e5085 (2016).

Relat Cancer **16**, 155-69 (2009).
)14).
radic disease. *Eur J Hum Genet* **17**, 368-77 (2009).
expression. *Biomed Pharmacother* **75**, 142-7 (2015)
-American and Caucasian prostate cancer models. *Int J Cancer* **135**, 2493-506 (2014).
ne responses in prostate cancer cells. *Mol Endocrinol* **23**, 292-307 (2009).
Relat Cancer **21**, 127-42 (2014)

idine Starvation. *Front Genet* **8**, 61 (2017).

Supplementary Table 3

Sample Name	Signature1	Signature2	Signature3	Signature5	Signature8	Signature1
PD11328a	1174	302	0	3314	2053	165
PD11329a	2456	161	0	2560	0	170
PD11330a	2294	276	0	2983	2799	368
PD11331a	1651	0	0	1320	0	0
PD11332a	1688	229	0	1767	1909	290
PD11333a	1960	0	0	1999	0	0
PD11334a	2400	0	0	2204	0	0
PD11335a	1346	0	9493	3622	0	223
PD12337a	2959	530	0	9450	0	940
PD12806a	1511	0	0	1044	0	0
PD12807a	531	113	0	1257	615	0
PD12808a	1479	0	0	776	2075	189
PD12809a	875	132	0	1954	0	0
PD12810a	562	116	0	879	1584	0
PD12811a	582	0	0	1926	733	169
PD12812a	496	132	0	1510	3715	342
PD12813a	1340	0	0	1445	0	0
PD12814a	1762	0	0	1218	0	0
PD12815a	498	0	0	1025	0	0
PD12816a	663	0	0	642	490	0
PD12817a	773	156	0	2322	0	0
PD12818a	1288	0	0	1033	0	0
PD12819a	1165	0	0	1044	0	0
PD12820a	2212	0	0	575	0	0
PD12821a	990	0	0	1497	0	0
PD12822a	909	82	0	959	0	0
PD12823a	174	0	0	368	334	47
PD12824a	837	124	0	1909	0	145
PD12825a	1384	0	0	1270	0	0
PD12826a	200	0	0	230	731	0
PD12828a	1610	0	0	1403	0	0
PD12829a	967	110	0	1622	0	0
PD12830a	1011	0	0	2358	0	0
PD12831a	1266	0	0	772	0	0
PD12832a	1470	0	0	1196	0	0
PD12834a	774	136	0	1869	0	0
PD12835a	1885	0	0	1157	0	0
PD12836a	2144	0	0	978	0	0
PD12837a	1521	0	0	1424	0	0
PD12838a	2338	0	0	977	0	0
PD12840a	783	0	0	1173	774	0
PD12841a	802	67	0	945	0	44
PD12842a	1803	0	0	1590	0	0
PD12843a	1995	444	0	4717	2255	367
PD12844a	686	194	0	3003	0	0
PD12845a	4921	0	0	2034	0	0
PD13380a	401	198	1463	1098	1732	0
PD13381a	1824	0	0	1724	0	0
PD13382a	1179	0	0	1264	983	0
PD13383a	591	0	0	1427	481	0

PD13384a	826	0	0	1058	0	0
PD13386a	766	0	0	986	0	0
PD13387a	1610	0	0	1124	0	0
PD13388a	2609	0	0	1510	0	0
PD13389a	1727	0	0	993	0	0
PD13390a	910	0	0	1217	0	0
PD13391a	1545	0	0	710	0	0
PD13392a	674	0	0	490	327	0
PD13393a	1193	146	0	1248	0	0
PD13394a	612	149	0	2451	0	0
PD13395a	678	0	0	1892	811	142
PD13396a	1842	0	0	828	0	0
PD13397a	114	0	0	164	133	0
PD13398a	560	0	0	816	463	0
PD13399a	1074	0	0	936	0	0
PD13400a	264	45	0	264	355	0
PD13401a	1337	0	0	762	0	0
PD13402a	594	0	0	1642	896	0
PD13404a	1486	0	0	823	0	0
PD13405a	677	88	0	1777	0	0
PD13406a	590	0	0	1313	537	0
PD13407a	588	101	845	1262	0	0
PD13408a	662	178	0	2173	728	215
PD13409a	1587	0	0	495	0	0
PD13410a	1492	0	0	667	0	0
PD13411a	1900	0	0	1698	0	0
PD13412a	1401	0	4626	587	5006	0
PD14706a	2299	0	0	648	0	0
PD14707a	1476	0	0	843	0	0
PD14708a	1395	0	0	501	0	0
PD14709a	634	0	0	1706	0	110
PD14710a	823	0	0	941	0	0
PD14711a	798	0	0	1425	0	0
PD14712a	866	0	0	847	0	0
PD14713a	1406	174	0	1471	0	185
PD14714a	1833	0	0	1519	0	0
PD14716a	1590	0	0	1113	0	0
PD14717a	1850	0	0	730	0	0
PD14718a	1011	124	953	829	0	0
PD14719a	1702	0	0	955	0	0
PD14720a	704	0	0	1912	0	151
PD14721a	2090	0	0	1698	0	0
PD14722a	1724	0	2287	912	2960	274
PD14724a	2178	0	0	53	1802	0
PD14725a	1599	0	0	916	0	0
PD14726a	833	184	0	2464	0	0
PD14727a	1790	0	0	1011	0	0
PD14728a	1706	242	0	2699	1390	282
PD14729a	2082	170	0	1473	0	0
PD14730a	1872	162	0	1261	0	0
PD14731a	1290	0	0	2977	2456	414
PD14732a	922	0	0	1088	0	0
PD14733a	480	0	0	605	417	0

PD14734a	2146	154	0	1409	0	0
PD14735a	1927	0	0	491	0	0
PD14736a	1597	0	0	653	0	0
PD6204a_Illumina	524	101	0	1675	0	0
PD9168a	1570	0	0	1484	0	0
PD9169a	787	0	0	1656	677	0
PD9746a	1076	0	0	994	1372	157
PD9749a	1365	0	0	2029	0	0
PD9751a	695	0	0	1342	570	0

Signature1	Totalmutations
0	7008
603	5950
0	8720
524	3495
223	6106
0	3959
0	4604
1423	16107
3569	17448
0	2555
0	2516
0	4519
171	3132
0	3141
0	3410
0	6195
0	2785
0	2980
0	1523
0	1795
338	3589
0	2321
0	2209
0	2787
0	2487
122	2072
0	923
0	3015
0	2654
0	1161
0	3013
0	2699
0	3369
0	2038
0	2666
0	2779
0	3042
0	3122
0	2945
0	3315
0	2730
40	1898
0	3393
0	9778
0	3883
0	6955
0	4892
0	3548
0	3426
0	2499

0	1884
0	1752
0	2734
0	4119
0	2720
0	2127
0	2255
0	1491
0	2587
0	3212
0	3523
0	2670
0	411
0	1839
229	2239
0	928
922	3021
0	3132
0	2309
0	2542
0	2440
0	2796
0	3956
0	2082
0	2159
583	4181
324	11944
430	3377
390	2709
188	2084
0	2450
0	1764
0	2223
0	1713
0	3236
0	3352
0	2703
0	2580
0	2917
0	2657
0	2767
454	4242
0	8157
0	4033
0	2515
0	3481
344	3145
0	6319
0	3725
0	3295
0	7137
0	2010
0	1502

0	3709
0	2418
0	2250
0	2300
0	3054
0	3120
0	3599
0	3394
0	2607

SampleName	Signature1	Signature2	Signature3	Signature5	Signature8	Signature13
PD11328a_clonal	917	203	0	2403	1355	116
PD11328a_early	31	9	0	124	70	9
PD11328a_late	74	29	0	255	273	15
PD11328a_subclonal	77	68	0	366	355	18
PD11329a_clonal	1329	85	0	1299	0	72
PD11329a_early	256	41	0	174	0	10
PD11329a_late	520	2	0	391	0	24
PD11329a_subclonal	306	19	0	865	0	33
PD11330a_clonal	604	70	0	758	681	88
PD11330a_early	1102	82	0	735	884	64
PD11330a_late	429	101	0	989	873	137
PD11330a_subclonal	0	3	0	0	7	9
PD11331a_clonal	469	0	0	300	0	0
PD11331a_early	791	0	0	634	0	0
PD11331a_late	205	0	0	124	0	0
PD11331a_subclonal	46	0	0	102	0	0
PD11332a_clonal	310	0	0	271	138	44
PD11332a_early	1064	80	0	901	693	76
PD11332a_late	300	115	0	501	819	135
PD11332a_subclonal	76	24	0	88	247	3
PD11333a_clonal	1626	0	0	1444	0	0
PD11333a_early	17	0	0	39	0	0
PD11333a_late	7	0	0	0	0	0
PD11333a_subclonal	99	0	0	221	0	0
PD11334a_clonal	1380	0	0	1285	0	0
PD11334a_early	12	0	0	0	0	0
PD11334a_late	7	0	0	0	0	0
PD11334a_subclonal	530	0	0	661	0	0
PD11335a_clonal	1031	0	5965	2710	0	194
PD11335a_early	47	0	102	95	0	18
PD11335a_late	0	0	320	83	0	0
PD11335a_subclonal	54	0	794	147	0	0
PD12337a_clonal	951	212	0	2687	0	174
PD12337a_early	6	1	0	85	0	8
PD12337a_late	5	0	0	24	0	6
PD12337a_subclonal	128	14	0	422	0	14
PD12806a_clonal	887	0	0	595	0	0
PD12806a_early	1	0	0	0	0	0
PD12806a_subclonal	190	0	0	114	0	0
PD12807a_clonal	38	11	0	84	66	0
PD12807a_subclonal	39	25	0	395	165	0
PD12808a_clonal	588	0	0	382	556	51
PD12808a_late	0	0	0	0	2	0
PD12808a_subclonal	467	0	0	295	1081	94
PD12809a_clonal	9	1	0	41	0	0
PD12809a_early	8	2	0	2	0	0
PD12809a_late	3	0	0	0	0	0
PD12809a_subclonal	49	5	0	142	0	0
PD12810a_clonal	383	87	0	580	828	0

PD12810a_subclonal	0	2	0	51	465	0
PD12811a_clonal	315	0	0	819	194	56
PD12811a_subclonal	39	0	0	463	272	47
PD12812a_clonal	24	5	0	73	110	4
PD12812a_early	17	1	0	21	5	0
PD12812a_late	4	5	0	0	50	6
PD12812a_subclonal	0	0	0	340	1555	152
PD12813a_clonal	1230	0	0	1296	0	0
PD12813a_subclonal	15	0	0	53	0	0
PD12814a_clonal	1435	0	0	1052	0	0
PD12814a_subclonal	78	0	0	104	0	0
PD12815a_clonal	411	0	0	827	0	0
PD12815a_late	0	0	0	1	0	0
PD12815a_subclonal	2	0	0	33	0	0
PD12816a_clonal	66	0	0	44	37	0
PD12816a_early	115	0	0	13	18	0
PD12816a_late	463	0	0	519	358	0
PD12817a_clonal	572	152	0	1857	0	0
PD12817a_early	0	0	0	2	0	0
PD12817a_subclonal	77	1	0	94	0	0
PD12818a_clonal	707	0	0	537	0	0
PD12818a_late	0	0	0	4	0	0
PD12818a_subclonal	39	0	0	17	0	0
PD12819a_clonal	1123	0	0	923	0	0
PD12819a_early	0	0	0	1	0	0
PD12820a_clonal	634	0	0	251	0	0
PD12820a_subclonal	1025	0	0	235	0	0
PD12821a_clonal	648	0	0	1062	0	0
PD12821a_subclonal	97	0	0	204	0	0
PD12822a_clonal	71	4	0	34	0	0
PD12822a_subclonal	22	9	0	52	0	0
PD12823a_late	12	0	0	0	0	0
PD12823a_subclonal	154	0	0	334	295	33
PD12824a_clonal	639	81	0	1382	0	111
PD12824a_early	50	7	0	216	0	5
PD12824a_late	54	8	0	114	0	4
PD12824a_subclonal	22	3	0	0	0	2
PD12825a_clonal	1342	0	0	1136	0	0
PD12825a_late	0	0	0	1	0	0
PD12826a_clonal	0	0	0	0	3	0
PD12826a_early	40	0	0	0	26	0
PD12826a_late	128	0	0	225	633	0
PD12828a_clonal	379	0	0	368	0	0
PD12828a_early	837	0	0	756	0	0
PD12828a_late	277	0	0	207	0	0
PD12828a_subclonal	12	0	0	0	0	0
PD12829a_clonal	811	85	0	1308	0	0
PD12829a_subclonal	68	13	0	178	0	0
PD12830a_clonal	517	0	0	1498	0	0
PD12830a_subclonal	298	0	0	341	0	0

PD12831a_clonal	713	0	0	431	0	0
PD12831a_early	363	0	0	190	0	0
PD12831a_late	101	0	0	67	0	0
PD12831a_subclonal	0	0	0	18	0	0
PD12832a_clonal	895	0	0	601	0	0
PD12832a_subclonal	216	0	0	159	0	0
PD12834a_clonal	495	73	0	1041	0	0
PD12834a_early	7	3	0	40	0	0
PD12834a_late	16	0	0	63	0	0
PD12834a_subclonal	81	25	0	300	0	0
PD12835a_subclonal	1224	0	0	795	0	0
PD12836a_clonal	83	0	0	12	0	0
PD12836a_early	1452	0	0	636	0	0
PD12836a_late	422	0	0	263	0	0
PD12836a_subclonal	2	0	0	1	0	0
PD12837a_clonal	221	0	0	228	0	0
PD12837a_subclonal	282	0	0	490	0	0
PD12838a_clonal	2114	0	0	720	0	0
PD12838a_early	4	0	0	0	0	0
PD12838a_late	5	0	0	1	0	0
PD12838a_subclonal	28	0	0	0	0	0
PD12840a_clonal	281	0	0	300	177	0
PD12840a_subclonal	95	0	0	525	398	0
PD12841a_clonal	571	39	0	619	0	0
PD12841a_early	2	0	0	0	0	0
PD12841a_late	2	0	0	5	0	0
PD12841a_subclonal	126	24	0	186	0	38
PD12842a_clonal	1311	0	0	1069	0	0
PD12842a_late	1	0	0	0	0	0
PD12842a_subclonal	187	0	0	394	0	0
PD12843a_clonal	660	132	0	1589	579	137
PD12843a_early	868	186	0	1902	987	128
PD12843a_late	378	110	0	989	574	83
PD12844a_clonal	7	15	0	191	0	0
PD12844a_early	476	100	0	1651	0	0
PD12844a_late	169	77	0	941	0	0
PD12844a_subclonal	0	0	0	16	0	0
PD12845a_clonal	1094	0	0	476	0	0
PD12845a_early	1088	0	0	511	0	0
PD12845a_late	1888	0	0	657	0	0
PD12845a_subclonal	2	0	0	23	0	0
PD13380a_clonal	309	140	966	1043	1283	0
PD13380a_subclonal	6	11	191	0	172	0
PD13381a_clonal	413	0	0	429	0	0
PD13381a_subclonal	956	0	0	839	0	0
PD13382a_clonal	877	0	0	898	716	0
PD13382a_subclonal	28	0	0	49	73	0
PD13383a_clonal	412	0	0	1028	243	0
PD13383a_subclonal	44	0	0	89	187	0
PD13384a_clonal	508	0	0	754	0	0

PD13384a_subclonal	208	0	0	136	0	0
PD13386a_clonal	322	0	0	509	0	0
PD13386a_subclonal	1	0	0	108	0	0
PD13387a_clonal	810	0	0	722	0	0
PD13387a_subclonal	28	0	0	71	0	0
PD13388a_clonal	467	0	0	209	0	0
PD13388a_subclonal	793	0	0	584	0	0
PD13389a_clonal	1344	0	0	838	0	0
PD13389a_late	0	0	0	2	0	0
PD13389a_subclonal	139	0	0	33	0	0
PD13390a_clonal	648	0	0	1000	0	0
PD13390a_early	20	0	0	9	0	0
PD13390a_late	54	0	0	0	0	0
PD13390a_subclonal	22	0	0	11	0	0
PD13391a_clonal	1326	0	0	670	0	0
PD13391a_early	42	0	0	7	0	0
PD13391a_late	83	0	0	0	0	0
PD13391a_subclonal	0	0	0	2	0	0
PD13392a_clonal	0	0	0	0	1	0
PD13392a_subclonal	108	0	0	49	111	0
PD13393a_early	1	0	0	0	0	0
PD13393a_late	3	0	0	0	0	0
PD13393a_subclonal	402	56	0	467	0	0
PD13394a_clonal	428	107	0	1807	0	0
PD13394a_early	0	4	0	0	0	0
PD13394a_late	0	0	0	4	0	0
PD13394a_subclonal	14	9	0	89	0	0
PD13395a_clonal	122	0	0	136	98	5
PD13395a_early	0	0	0	0	2	0
PD13395a_late	8	0	0	16	11	1
PD13396a_clonal	1140	0	0	569	0	0
PD13396a_subclonal	284	0	0	90	0	0
PD13398a_clonal	420	0	0	609	198	0
PD13398a_early	1	0	0	0	3	0
PD13398a_late	5	0	0	3	1	0
PD13398a_subclonal	3	0	0	0	35	0
PD13399a_clonal	341	0	0	344	0	0
PD13399a_subclonal	192	0	0	210	0	0
PD13400a_clonal	233	38	0	240	312	0
PD13400a_late	0	1	0	2	0	0
PD13401a_clonal	500	0	0	420	0	0
PD13401a_early	6	0	0	12	0	0
PD13401a_late	32	0	0	5	0	0
PD13401a_subclonal	461	0	0	215	0	0
PD13402a_clonal	340	0	0	579	145	0
PD13402a_early	0	0	0	1	0	0
PD13402a_late	4	0	0	2	0	0
PD13402a_subclonal	103	0	0	702	484	0
PD13404a_clonal	23	0	0	23	0	0
PD13404a_subclonal	1050	0	0	474	0	0

PD13405a_clonal	550	68	0	1433	0	0
PD13405a_early	6	0	0	15	0	0
PD13405a_late	0	3	0	19	0	0
PD13405a_subclonal	1	0	0	51	0	0
PD13406a_clonal	267	0	0	649	160	0
PD13406a_early	0	0	0	0	8	0
PD13406a_late	2	0	0	1	4	0
PD13406a_subclonal	24	0	0	95	62	0
PD13407a_subclonal	541	89	717	1156	0	0
PD13408a_clonal	524	116	0	1222	31	96
PD13408a_early	11	0	0	57	0	6
PD13408a_late	15	1	0	17	33	11
PD13408a_subclonal	38	37	0	602	594	82
PD13409a_clonal	1498	0	0	420	0	0
PD13409a_late	0	0	0	5	0	0
PD13410a_clonal	1211	0	0	492	0	0
PD13410a_subclonal	85	0	0	68	0	0
PD13411a_clonal	2	0	0	10	0	0
PD13411a_early	411	0	0	339	0	0
PD13411a_late	928	0	0	540	0	0
PD13411a_subclonal	316	0	0	561	0	0
PD13412a_clonal	1116	0	3499	0	3856	0
PD13412a_early	45	0	150	0	183	0
PD13412a_late	24	0	100	0	430	0
PD13412a_subclonal	161	0	520	0	1167	0
PD14706a_clonal	1929	0	0	633	0	0
PD14706a_subclonal	110	0	0	13	0	0
PD14707a_clonal	881	0	0	440	0	0
PD14707a_subclonal	321	0	0	190	0	0
PD14708a_clonal	1192	0	0	480	0	0
PD14708a_subclonal	38	0	0	10	0	0
PD14709a_clonal	33	0	0	84	0	6
PD14709a_subclonal	371	0	0	1141	0	72
PD14710a_clonal	387	0	0	538	0	0
PD14710a_subclonal	256	0	0	268	0	0
PD14711a_clonal	659	0	0	1265	0	0
PD14711a_subclonal	80	0	0	0	0	0
PD14712a_clonal	548	0	0	509	0	0
PD14712a_late	2	0	0	0	0	0
PD14712a_subclonal	34	0	0	47	0	0
PD14713a_clonal	809	36	0	853	0	73
PD14713a_early	32	0	0	35	0	2
PD14713a_late	102	6	0	10	0	10
PD14713a_subclonal	165	142	0	396	0	83
PD14714a_clonal	218	0	0	233	0	0
PD14714a_early	0	0	0	1	0	0
PD14714a_subclonal	817	0	0	635	0	0
PD14716a_clonal	954	0	0	823	0	0
PD14716a_subclonal	307	0	0	204	0	0
PD14717a_clonal	1214	0	0	547	0	0

PD14717a_subclonal	384	0	0	118	0	0
PD14718a_clonal	844	103	824	694	0	0
PD14718a_subclonal	61	3	56	49	0	0
PD14719a_clonal	1077	0	0	524	0	0
PD14719a_subclonal	304	0	0	222	0	0
PD14720a_clonal	649	0	0	1649	0	120
PD14720a_late	0	0	0	12	0	3
PD14720a_subclonal	0	0	0	6	0	0
PD14721a_clonal	318	0	0	382	0	0
PD14721a_early	1251	0	0	966	0	0
PD14721a_late	433	0	0	275	0	0
PD14722a_clonal	1286	0	1115	632	1998	149
PD14722a_early	28	0	98	27	44	1
PD14722a_late	36	0	113	20	112	10
PD14722a_subclonal	85	0	390	73	330	58
PD14724a_clonal	297	0	0	0	177	0
PD14724a_early	1184	0	0	141	831	0
PD14724a_late	510	0	0	0	633	0
PD14724a_subclonal	0	0	0	0	39	0
PD14725a_clonal	496	0	0	284	0	0
PD14725a_subclonal	43	0	0	67	0	0
PD14726a_clonal	257	62	0	716	0	0
PD14726a_subclonal	62	12	0	176	0	0
PD14727a_clonal	11	0	0	8	0	0
PD14727a_early	1017	0	0	669	0	0
PD14727a_late	677	0	0	265	0	0
PD14727a_subclonal	0	0	0	12	0	0
PD14728a_clonal	1211	106	0	1669	852	134
PD14728a_early	59	26	0	7	22	0
PD14728a_late	106	57	0	132	40	58
PD14728a_subclonal	0	15	0	221	203	71
PD14729a_clonal	360	38	0	212	0	0
PD14729a_early	575	38	0	430	0	0
PD14729a_late	1045	80	0	725	0	0
PD14730a_clonal	249	29	0	110	0	0
PD14730a_early	1064	88	0	713	0	0
PD14730a_late	469	32	0	366	0	0
PD14730a_subclonal	0	0	0	1	0	0
PD14731a_clonal	224	0	0	245	325	43
PD14731a_early	689	0	0	1568	794	174
PD14731a_late	312	0	0	1021	1206	159
PD14731a_subclonal	0	0	0	0	4	0
PD14732a_clonal	327	0	0	618	0	0
PD14732a_subclonal	195	0	0	156	0	0
PD14733a_clonal	193	0	0	449	170	0
PD14733a_subclonal	174	0	0	88	190	0
PD14734a_clonal	1813	117	0	1067	0	0
PD14734a_early	0	1	0	0	0	0
PD14734a_late	0	0	0	2	0	0
PD14734a_subclonal	42	20	0	74	0	0

PD14735a_clonal	1036	0	0	356	0	0
PD14735a_subclonal	311	0	0	60	0	0
PD14736a_clonal	12	0	0	14	0	0
PD14736a_subclonal	347	0	0	153	0	0
PD6204a_Illumina_ea	0	0	0	8	0	0
PD6204a_Illumina_lat	0	2	0	0	0	0
PD6204a_Illumina_sul	0	38	0	0	0	0
PD9168a_early	855	0	0	747	0	0
PD9168a_late	677	0	0	693	0	0
PD9169a_clonal	525	0	0	1666	0	0
PD9169a_subclonal	78	0	0	179	0	0
PD9746a_early	212	0	0	307	161	14
PD9746a_late	794	0	0	644	1142	147
PD9749a_clonal	836	0	0	1457	0	0
PD9749a_subclonal	42	0	0	203	0	0
PD9751a_clonal	458	0	0	1027	338	0
PD9751a_subclonal	1	0	0	51	91	0

Signature18
0
0
0
0
224
35
124
242
0
0
0
0
156
120
108
36
33
69
52
15
0
0
0
0
0
0
0
0
973
15
65
60
561
16
0
0
0
0
0
0
0
0
0
0
0
0
0
0
33
0

0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
203
0
56
0
0
0
0
0
0
0
0
0
0
0
13
40
13
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0

0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
41
99
0
0
76
6
22
658
0
0
0
0
0
0

0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
65
215
215
286
0
12
38
287
70
132
184
135
26
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0

0
0
0
0
0
0
0
0
65
271
110
0
0
0
0
0
0
0
0
0
0
0
0
0
177
170
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0

0
0
0
0
0
0
0
0
0
137
51
0
0
70
29
0
0

SUPPLEMENTARY TABLE 4

	Cox model *p*-value	Cox model adjusted *p*-value (BH)	Hazard ratio	Cox multivariate model *p*-value	Cox multivariate model hazard ratio
CDH12	0.000085	0.006	9.3	0.00061	7.3
ANTXR2	0.00033	0.012	7.7	0.0015	6.5
SPOP	0.0042	0.081	3.9	0.0016	4.8
IL6ST	0.0046	0.081	4.8	0.0014	6.2
ZFHX4	0.015	0.22	3.8	0.021	3.6
NCOA7	0.019	0.23	4.4	0.012	5
DLC1	0.031	0.26	3.8	0.11	2.8
PPAP2A	0.032	0.26	2.9	0.037	2.9
SENP6	0.034	0.26	3.7	0.1	2.8
MTUS1	0.037	0.26	3.7	0.078	3
LOH, chr 13:46Mb	0.04	0.26	2.4	0.081	2.2
GRM1	0.05	0.29	4.4	0.077	3.8
PDE4D	0.061	0.33	2.8	0.028	3.4
CHD1	0.069	0.33	2.3	0.04	2.6
Gain, chr 9:43Mb	0.07	0.33	2.2	0.08	2.1
LOH, chr 1:235Mb	0.095	0.38	2.4	0.16	2.1
MTMR7	0.096	0.38	2.8	0.19	2.3
CDKN1B	0.097	0.38	2.1	0.087	2.3
MAP3K1	0.14	0.52	2.5	0.11	2.7
ROBO1	0.15	0.55	2.4	0.26	2
SLC4A4	0.16	0.56	2.4	0.32	1.9
LOH, chr 6:88Mb	0.21	0.68	1.7	0.063	2.4
ETV3	0.28	0.74	2	0.45	1.6
LOH, chr 8:23Mb	0.29	0.74	1.6	0.4	1.5
LOH, chr 2:141Mb	0.29	0.74	1.8	0.39	1.7
ETS_status	0.3	0.74	0.64	0.24	0.6
AMBRA1	0.3	0.74	2.2	0.48	1.7
IFI16	0.34	0.74	2	0.55	1.6
LOH, chr 12:12Mb	0.34	0.74	1.5	0.46	1.4
ACPP	0.35	0.74	2	0.32	2.1
LOH, chr 5:100Mb	0.35	0.74	1.5	0.25	1.7
LOH, chr 3:72Mb	0.36	0.74	0.63	0.23	0.5
PPP2R2A	0.38	0.74	1.7	0.53	1.5
RBM7	0.38	0.74	1.9	0.63	1.4
RYBP	0.38	0.74	0.62	0.4	0.63
LOH, chr 16:79Mb	0.41	0.74	0.66	0.49	0.67
NCOR2	0.43	0.74	1.8	0.29	2.2
USP28	0.44	0.74	1.8	0.72	1.3
PPP4R2	0.45	0.74	0.46	0.42	0.43
FOXA1	0.45	0.74	1.5	0.16	2.1
TP53	0.45	0.74	1.4	0.19	1.9
ERG	0.46	0.74	0.75	0.28	0.64
ADAM28	0.47	0.74	1.6	0.86	1.1
NKX3.1	0.47	0.74	1.6	0.86	1.1
LOH, chr 1:77Mb	0.47	0.74	1.5	0.56	1.4
LOH, chr 10:92Mb	0.48	0.75	0.65	0.25	0.44
HD, chr 10:90Mb	0.52	0.78	0.62	0.23	0.33
NEAT1	0.53	0.78	1.6	0.61	1.5

Gain, chr 3:130Mb	0.55	0.8	1.4	0.6	1.4
FOXP1	0.58	0.8	1.3	0.57	1.3
TGFBR2	0.58	0.8	0.56	0.47	0.47
ZBTB16	0.64	0.87	1.4	0.95	1
LOH, chr 17:7Mb	0.66	0.88	1.2	0.44	1.5
ZNF292	0.73	0.93	1.2	0.9	0.93
SHQ1	0.73	0.93	0.83	0.69	0.8
LOH, chr 18:73Mb	0.73	0.93	1.2	0.82	1.1
TCF12	0.8	0.99	1.2	0.85	1.2
ROBO2	0.82	0.99	1.2	0.92	0.94
ARHGAP15	0.82	0.99	1.2	0.86	0.87
Gain, chr 8:89Mb	0.87	1	0.93	0.92	1
HD, chr 5:98Mb	0.89	1	1.1	0.83	1.2
APC	0.92	1	1.1	0.75	1.3
Gain, chr 7:53Mb	0.93	1	1	0.9	1.1
MYST3	0.93	1	1.1	0.51	0.55
ASH1L	0.94	1	0.93	0.77	0.74
PTEN	0.96	1	1	0.44	0.67
TBX18	0.96	1	0.96	0.79	0.81
WGD	0.98	1	1	0.9	1.1
TBL1XR1	0.99	1	0.99	0.89	0.86
UBTF	1	1	1.3E-08	1	3.3E-09
GPATCH8	1	1	1.3E-08	1	3.2E-09

SUPPLEMENTARY TABLE 5

UniprotAccesstion	Gene Symbol	Source	Subcellular_l	long_name	node_label
P10275	AR	input_list	Nucl	Androgen rec	AR
P15056	BRAF	input_list	Cyto	Serine/threor	BRAF
P03372	ESR1	additional_er	Nucl	Estrogen rece	ER-alpha
O15379	HDAC3	additional_er	Nucl	Histone deac	HDAC3
Q12809	KCNH2	input_list	Cell membrar	Potassium vo	Kv11.1
Q02750	MAP2K1	additional_cc	Cyto	Protein Kinase	
P04150	NR3C1	additional_er	Nucl	Glucocorticoi	GR
P10276	RARA	additional_er	Nucl	Retinoic acid	RARA
P10826	RARB	additional_er	Nucl	Retinoic acid	RARB
P13631	RARG	additional_er	Nucl	Retinoic acid	RARG
P19793	RXRA	additional_er	Nucl	Retinoic acid	RXRA
P31749	AKT1	input_list	Cyto	RAC-alpha se	Akt1
Q13315	ATM	input_list	Nucl	Serine-protei	ATM
Q5S007	LRRK2	input_list	Cell membrar	Leucine-rich r	LRRK2
Q00987	MDM2	additional_cc	Nucl	E3 ubiquitin-protein ligase	
Q08499	PDE4D	input_list	Cyto	cAMP-specifi	PDE4D
P42336	PIK3CA	input_list	Cyto	Phosphatidyli	PIK3CA
P42338	PIK3CB	input_list	Cyto	Phosphatidyli	PIK3CB
P04637	TP53	input_list	Nucl	Cellular tumo	p53
P35869	AHR	additional_er	Cyto	Aryl hydrocar	AHR
P38398	BRCA1	additional_er	Nucl	Breast cancer	BRCA1
P35222	CTNNB1	input_list	Cyto	Catenin beta-	CTNNB1
P01112	HRAS	input_list	Cell membrar	GTPase HRas	HRas
O75874	IDH1	input_list	Cyto	Isocitrate del	IDH1
Q14643	ITPR1	input_list	Org Memb	Inositol 1,4,5-	IP3R1
Q14571	ITPR2	input_list	Org Memb	Inositol 1,4,5-	IP3R2
P05412	JUN	additional_er	Nucl	Transcription	Jun
Q13233	MAP3K1	input_list	Cyto	Mitogen-acti	MEKK1
O00255	MEN1	additional_er	Nucl	Menin	MEN1
O75376	NCOR1	input_list	Nucl	Nuclear recep	N-CoR1
Q9Y618	NCOR2	input_list	Nucl	Nuclear recep	SMRT
P22736	NR4A1	additional_er	Nucl	Nuclear recep	Nur77
P27986	PIK3R1	input_list	Cyto	Phosphatidyli	PIK3R1
P63151	PPP2R2A	input_list	Cyto	Serine/threor	PPP2R2A
P58335	ANTXR2	input_list	Cell membrar	Anthrax toxin	ANTXR2
P25054	APC	input_list	Cell membrar	Adenomatou	APC
P27540	ARNT	additional_er	Nucl	Aryl hydrocar	ARNT
Q9NR48	ASH1L	input_list	Nucl	Histone-lysin	ASH1L
P51587	BRCA2	input_list	Nucl	Breast cancer	BRCA2
O43439	CBFA2T2	additional_er	Nucl	Protein CBFA	CBFA2T2

P55289	CDH12	input_list	Cell membrar	Cadherin-12	CDH12
Q9NYV4	CDK12	input_list	Nucl	Cyclin-depen	CDK12
O14646	CHD1	input_list	Nucl	Chromodoma	CHD-1
Q92793	CREBBP	additional_er	Cyto	CREB-binding	CBP
Q96QB1	DLC1	input_list	Cyto	Rho GTPase-a	DLC1
Q96BY6	DOCK10	input_list	Cell membrar	Dedicator of	DOCK10
P11308	ERG	input_list	Nucl	Transcription	ERG
Q5FWF5	ESCO1	input_list	Nucl	N-acetyltrans	ESCO1
P41162	ETV3	input_list	Nucl	ETS transloca	ETV3
P55317	FOXA1	input_list	Nucl	Hepatocyte n	FOXA1
P55316	FOXG1	additional_er	Nucl	Forkhead box	FOXG1
Q12778	FOXO1	input_list	Nucl	Forkhead box	FOXO1A
P98177	FOXO4	additional_er	Nucl	Forkhead box	FOXO4
Q9H334	FOXP1	input_list	Nucl	Forkhead box	FOXP1
P15976	GATA1	additional_er	Nucl	Erythroid tra	GATA1
P23769	GATA2	additional_er	Nucl	Endothelial tr	GATA2
P51858	HDGF	additional_er	Cyto	Hepatoma-de	HDGF
P41235	HNF4A	additional_er	Nucl	Hepatocyte n	HNF4 alpha
P40189	IL6ST	input_list	Cell membrar	Interleukin-6	gp130
Q92794	KAT6A	input_list	Nucl	Histone acety	MYST3
O75164	KDM4A	additional_er	Nucl	Lysine-specifi	JMJD2A
O15550	KDM6A	input_list	Nucl	Lysine-specifi	UTX
Q8NEZ4	KMT2C	input_list	Nucl	Histone-lysin	MLL3
O14686	KMT2D	input_list	Nucl	Histone-lysin	MLL4
Q96PU5	NEDD4L	input_list	Cyto	E3 ubiquitin-p	NEDD4L
Q99801	NKX3-1	input_list	Nucl	Homeobox pr	NKX3-1
O75925	PIAS1	additional_er	Nucl	E3 SUMO-prc	PIAS1
O75928	PIAS2	additional_er	Nucl	E3 SUMO-prc	PIAS2
P60484	PTEN	input_list	Cyto	Phosphatidyli	PTEN
P06400	RB1	input_list	Nucl	Retinoblastor	Rb
Q68DV7	RNF43	input_list	Cell membrar	E3 ubiquitin-p	RNF43
P12755	SKI	additional_er	Nucl	Ski oncogene	SKI
Q15796	SMAD2	input_list	Cyto	Mothers agai	SMAD2
P84022	SMAD3	additional_er	Cyto	Mothers agai	SMAD3
Q13485	SMAD4	input_list	Cyto	Mothers agai	SMAD4
P51532	SMARCA4	additional_er	Nucl	Transcription	SMARCA4
O95238	SPDEF	additional_er	Nucl	SAM pointed	SPDEF
O43791	SPOP	input_list	Nucl	Speckle-type	SPOP
O60907	TBL1X	additional_er	Nucl	F-box-like/WI	TBL1X
Q9BZK7	TBL1XR1	input_list	Nucl	F-box-like/WI	TBL1XR1
O15119	TBX3	input_list	Nucl	T-box transcr	TBX3
O15350	TP73	additional_er	Cyto	Tumor protei	p73
Q05516	ZBTB16	input_list	Nucl	Zinc finger an	PLZF
Q9Y6X8	ZHX2	additional_er	Nucl	Zinc fingers a	ZHX2

Q9NY61	AATF	additional_er	Nucl	Protein AATF	AATF
Q9UKQ2	ADAM28	input_list	Cell membrar	Disintegrin ar	ADAM28
Q08117	AES	additional_	connectivity	Amino-termi	AES
P05067	APP	additional_	connectivity	Amyloid beta	APP
O14497	ARID1A	input_list	Nucl	AT-rich intera	ARID1A
Q68CP9	ARID2	input_list	Nucl	AT-rich intera	ARID2
Q4LE39	ARID4B	input_list	Nucl	AT-rich intera	ARID4B
Q76L83	ASXL2	input_list	Nucl	Putative Poly	ASXL2
O95352	ATG7	input_list	Cyto	Ubiquitin-like	ATG7
P54253	ATXN1	additional_	connectivity	Ataxin-1	ataxin-1
Q8WY36	BBX	additional_er	Nucl	HMG box trar	HBP2
Q86V15	CASZ1	input_list	Nucl	Zinc finger pr	CASZ1
Q9ULB5	CDH7	additional_	connectivity	Cadherin-7	CDH7
P46527	CDKN1B	input_list	Nucl	Cyclin-depen	p27Kip1
P17676	CEBPB	additional_er	Nucl	CCAAT/enhar	C/EBP-beta
Q9NZN8	CNOT2	additional_er	Nucl	CCR4-NOT tra	NOT2
O75175	CNOT3	input_list	Nucl	CCR4-NOT tra	NOT3
Q6UUV9	CRTC1	additional_er	Nucl	CREB-regulat	TORC1
Q96PT4	DUX3	additional_er	Nucl	Putative doul	DUX3
O00716	E2F3	additional_er	Nucl	Transcription	E2F3
Q7Z589	EMSY	additional_er	Nucl	BRCA2-intera	EMSY
O75593	FOXH1	additional_er	Nucl	Forkhead box	FOXH1
O43524	FOXO3	additional_er	Cyto	Forkhead box	FOXO3A
O00168	FXYD1	additional_	connectivity	Phospholemr	PLM
P10071	GLI3	additional_er	Nucl	Transcription	GLI3
Q96IK5	GMCL1	additional_	connectivity	Germ cell-les	GMCL1
Q9UKJ3	GPATCH8	input_list	Cyto	G patch dom	GPATCH8
Q13227	GPS2	additional_er	Nucl	G protein pat	GPS2
P68431	HIST1H3A|HI	additional_cc	Nucl	Histone H3.1	H3
Q5VVH5	IRAK1BP1	additional_er	Cyto	Interleukin-1	IRAK1BP1
P13612	ITGA4	additional_	connectivity	Integrin alpha	ITGA4
Q92993	KAT5	additional_	connectivity	Histone acety	Tip60
Q9NZR2	LRP1B	input_list	Cell membrar	Low-density l	LRP1B
P43364	MAGEA11	additional_	connectivity	Melanoma-as	MAGE-A11
Q03112	MECOM	additional_er	Nucl	MDS1 and EV	EVI1
Q13465	MECOM	additional_er	Nucl	MDS1 and EV	MDS1
Q02078	MEF2A	additional_er	Nucl	Myocyte-spe	MEF2A
Q03111	MLLT1	additional_er	Nucl	Protein ENL	MLLT1
Q9ULD2	MTUS1	input_list	Cyto	Microtubule-	MTUS1
P01106	MYC	additional_er	Nucl	Myc proto-or	Myc
Q8NI08	NCOA7	input_list	Nucl	Nuclear recep	NCOA7
Q12857	NFIA	additional_er	Nucl	Nuclear facto	NFI-A
Q13952	NFYC	additional_er	Nucl	Nuclear trans	NFYC
O43929	ORC4	additional_	connectivity	Origin recogn	ORC4L

O14494	PLPP1|PPAP2	input_list	Cell membrar	Phospholipid	PPAP2A
P29590	PML	additional_er	Nucl	Protein PML	PML
P67775	PPP2CA	additional_	connectivity	Serine/threor	PPP2CA
P17612	PRKACA	additional_cc	Cyto	cAMP-depen	PKACA
O75400	PRPF40A	additional_cc	Nucl	Pre-mRNA-pr	FNBP3
P28749	RBL1	additional_er	Nucl	Retinoblastor	Rb-like 1
Q9NVW2	RLIM	input_list	Nucl	E3 ubiquitin-p	RNF12
Q9Y6N7	ROBO1	input_list	Cell membrar	Roundabout I	ROBO1
Q9HCK4	ROBO2	input_list	Cell membrar	Roundabout I	ROBO2
P62913	RPL11	input_list	Nucl	60S ribosoma	RPL11
Q13950	RUNX2	additional_er	Nucl	Runt-related	AML3
Q8N488	RYBP	input_list	Nucl	RING1 and YY	RYBP
Q9GZR1	SENP6	input_list	Nucl	Sentrin-speci	SENP6
Q6PI26	SHQ1	input_list	Cyto	Protein SHQ1	SHQ1
Q96ST3	SIN3A	additional_er	Nucl	Paired amphi	SIN3A
O75182	SIN3B	additional_er	Nucl	Paired amphi	SIN3B
O15105	SMAD7	additional_er	Cyto	Mothers agai	SMAD7
Q12824	SMARCB1	additional_er	Nucl	SWI/SNF-rela	SMARCB1
P08047	SP1	additional_er	Nucl	Transcription	SP1
P11831	SRF	additional_er	Nucl	Serum respor	SRF
Q08945	SSRP1	additional_er	Nucl	FACT comple	SSRP1
O75528	TADA3	additional_er	Nucl	Transcription	TADA3L
P05549	TFAP2A	additional_er	Nucl	Transcription	AP-2 alpha
O43294	TGFB1I1	additional_er	Extracellular	Transforming	Hic-5
P51668	UBE2D1	additional_	connectivity	Ubiquitin-cor	UBE2D1
P17480	UBTF	input_list	Nucl	Nucleolar tra	UBF
Q96RU2	USP28	input_list	Nucl	Ubiquitin carl	USP28
P25490	YY1	additional_er	Nucl	Transcription	YY1
O15391	YY2	additional_er	Nucl	Transcription	YY2
O95365	ZBTB7A	additional_er	Nucl	Zinc finger an	FBI1
P37275	ZEB1	additional_er	Nucl	Zinc finger E-I	TCF8
Q15911	ZFHX3	input_list	Nucl	Zinc finger hc	ATBF1
Q9UKY1	ZHX1	additional_cc	Nucl	Zinc fingers a	ZHX1
O60481	ZIC3	additional_er	Nucl	Zinc finger pr	ZIC3
Q14202	ZMYM3	input_list	Nucl	Zinc finger M	ZNF261

	Druggability_	protein	Synonyms	Transcriptional_class	description
1	DT	AR		TF	Androgen rec
1	DT	BRAF			Serine/threor
1	DT	ER-alpha	ER	TF	Estrogen rece
1	DT	HDAC3	HD3	activator/repressor	Histone deac
1	DT	Kv11.1	ERG-1|Eag-related protein 1|Ether-a-go-		Potassium vo
1	DT	Mek1	MP2K1_HUMAN, MAP2K1, MEK1, PRKM		Dual specifici
1	DT	GR	GR	TF	Glucocorticoi
1	DT	RARA	RAR-alpha	TF	Retinoic acid
1	DT	RARB	RAR-beta	regulation	Retinoic acid
1	DT	RARG	RAR-gamma	regulation	Retinoic acid
1	DT	RXRA		regulation	Retinoic acid
2	CT	Akt1	PKB|PKB alpha		RAC-alpha se
2	CT	ATM	A-T mutated		Serine-protei
2	CT	LRRK2			Leucine-rich r
2	CT	MDM2	MDM2		E3 ubiquitin-p
2	CT	PDE4D			cAMP-specifi
2	CT	PIK3CA	PI3-kinase subunit alpha|PI3K-alpha|PI3		Phosphatidyli
2	CT	PIK3CB	PI3-kinase subunit beta|PI3K-beta|PI3Kb		Phosphatidyli
2	AC-CT	p53		TF	Cellular tumo
3	AC	AHR	Ah receptor|	TF	Aryl hydrocar
3	AC	BRCA1		activator/repressor	Breast cancer
3	AC	CTNNB1		TF	Catenin beta-
3	AC	HRas			GTPase HRas
3	AC	IDH1	IDH		Isocitrate deh
3	AC	IP3R1	IP3R 1|InsP3R1|Type 1 InsP3 receptor		Inositol 1,4,5-
3	AC	IP3R2	IP3R 2|InsP3R2|Type 2 InsP3 receptor		Inositol 1,4,5-
3	AC	Jun	AP1	TF	Transcription
3	AC	MEKK1	MEK kinase 1|MEKK 1		Mitogen-activ
3	AC	MEN1		activator/repressor	Menin
3	AC	N-CoR1	N-CoR|N-CoR	activator/repressor	Nuclear recep
3	AC	SMRT	N-CoR2|SMR	activator/repressor	Nuclear recep
3	AC	Nur77	Nur77	regulation	Nuclear recep
3	AC	PIK3R1	PI3-kinase regulatory subunit alpha|PI3-I		Phosphatidyli
3	AC	PPP2R2A			Serine/threor
4	Dr	ANTXR2	CMG-2		Anthrax toxin
4	Dr	APC	Protein APC		Adenomatous
4	Dr	ARNT	ARNT protein	TF	Aryl hydrocar
4	Dr	ASH1L	huASH1		Histone-lysine
4	Dr	BRCA2			Breast cancer
4	Dr	CBFA2T2		activator/repressor	Protein CBFA2

4	Dr	CDH12	BR-cadherin|N-cadherin 2		Cadherin-12
4	Dr	CDK12	CDC2-related protein kinase 7|CrkRS|hC		Cyclin-depen
4	Dr	CHD-1	CHD-1		Chromodoma
4	Dr	CBP		activator/repressor	CREB-binding
4	Dr	DLC1	DLC-1|StARD12		Rho GTPase-a
4	Dr	DOCK10			Dedicator of
4	Dr	ERG		TF	Transcription
4	Dr	ESCO1	ECO1 homolog 1|EFO1p|ESO1 homolog		N-acetyltrans
4	Dr	ETV3	PE-1		ETS transloca
4	Dr	FOXA1	HNF-3-alpha|	TF	Hepatocyte n
4	Dr	FOXG1	BF-1|BF-2|BF	TF	Forkhead box
4	Dr	FOXO1A		TF	Forkhead box
4	Dr	FOXO4		TF	Forkhead box
4	Dr	FOXP1	MFH	TF	Forkhead box
4	Dr	GATA1	GATA-1|GF-1	TF	Erythroid tran
4	Dr	GATA2		TF	Endothelial tr
4	Dr	HDGF	HDGF|HMG-	regulation	Hepatoma-de
4	Dr	HNF4 alpha	HNF-4-alpha|	TF	Hepatocyte n
4	Dr	gp130	IL-6 receptor subunit beta|IL-6R subunit		Interleukin-6
4	Dr	MYST3	MYST-3		Histone acety
4	Dr	JMJD2A		activator/repressor	Lysine-specifi
4	Dr	UTX			Lysine-specifi
4	Dr	MLL3	Lysine N-methyltransferase 2C		Histone-lysine
4	Dr	MLL4	Lysine N-met	regulation	Histone-lysine
4	Dr	NEDD4L			E3 ubiquitin-p
4	Dr	NKX3-1		TF	Homeobox pr
4	Dr	PIAS1	GBP	activator/repressor	E3 SUMO-pro
4	Dr	PIAS2	ARIP3|DIP|M	activator/repressor	E3 SUMO-pro
4	Dr	PTEN			Phosphatidyli
4	Dr	Rb	Rb	activator/repressor	Retinoblastor
4	Dr	RNF43			E3 ubiquitin-p
4	Dr	SKI		activator/repressor	Ski oncogene
4	Dr	SMAD2	MAD homolo	TF	Mothers agai
4	Dr	SMAD3	MAD homolo	TF	Mothers agai
4	Dr	SMAD4	MAD homolo	activator/repressor	Mothers agai
4	Dr	SMARCA4	BAF190A	activator/repressor	Transcription
4	Dr	SPDEF	Prostate-spec	TF	SAM pointed
4	Dr	SPOP			Speckle-type
4	Dr	TBL1X		activator/repressor	F-box-like/WD
4	Dr	TBL1XR1		activator/repressor	F-box-like/WD
4	Dr	TBX3	T-box protein	TF	T-box transcri
4	Dr	p73		TF	Tumor protei
4	Dr	PLZF		activator/repressor	Zinc finger an
4	Dr	ZHX2	AFP regulator	TF	Zinc fingers a

AATF		TF	Protein AATF
ADAM28	ADAM 28|MDC-L|eMDC II		Disintegrin ar
AES	Amino enhan	activator/repressor	Amino-termir
APP	AICD-50|AICI	TF	Amyloid beta
ARID1A	ARID domain-containing protein 1A|BAF		AT-rich intera
ARID2	ARID domain-containing protein 2|BAF2(		AT-rich intera
ARID4B	ARID domain-containing protein 4B|Sin3		AT-rich intera
ASXL2			Putative Polyc
ATG7	APG7-like|hAGP7		Ubiquitin-like
ataxin-1			Ataxin-1
HBP2		TF	HMG box trar
CASZ1		TF	Zinc finger pr
CDH7			Cadherin-7
p27Kip1			Cyclin-depenc
C/EBP-beta	C/EBP beta |I	TF	CCAAT/enhar
NOT2		activator/repressor	CCR4-NOT tra
NOT3			CCR4-NOT tra
TORC1	TORC-1|Tran	TF	CREB-regulate
		TF	Putative doub
E2F3	E2F-3	TF	Transcription
EMSY		regulation	BRCA2-intera
	Fast-1|Fast-2	TF	Forkhead box
FOXO3A		TF	Forkhead box
PLM			Phospholemn
GLI3	GLI3-190|GLI	TF	Transcriptiona
GMCL1			Germ cell-less
GPATCH8			G patch doma
GPS2	GPS-2	activator/repressor	G protein pat
H3			Histone H3.1
IRAK1BP1	IRAK1-bindin	regulation	Interleukin-1
ITGA4			Integrin alpha
Tip60	HIV-1 Tat interactive protein|Tip60		Histone acety
LRP1B	LRP-1B|LRP-DIT		Low-density li
MAGE-A11	CT1.11	activator/repressor	Melanoma-as
EVI1	EVI-1	activator/repressor	MDS1 and EV
MDS1		TF	MDS1 and EV
MEF2A		TF	Myocyte-spec
MLLT1		regulation	Protein ENL
MTUS1			Microtubule-a
Myc	bHLHe39	TF	Myc proto-on
NCOA7		activator/repressor	Nuclear recep
NFI-A	CTF|NF-I/A|N	TF	Nuclear facto
NFYC	NF-YC	TF	Nuclear trans
ORC4L			Origin recogn

PPAP2A	PAP-2a|PAP2a		Phospholipid
PML		TF	Protein PML
PPP2CA	PP2A-alpha|RP-C		Serine/threor
PKACA	PKA C-alpha		cAMP-depen
FNBP3	HIP-10		Pre-mRNA-pr
Rb-like 1	p107	activator/repressor	Retinoblastor
RNF12	R-LIM	activator/repressor	E3 ubiquitin-p
ROBO1			Roundabout h
ROBO2			Roundabout h
RPL11			60S ribosoma
AML3	CBF-alpha-1|	TF	Runt-related
RYBP	APAP-1|DED-	activator/repressor	RING1 and YY
SENP6			Sentrin-specif
SHQ1			Protein SHQ1
SIN3A		activator/repressor	Paired amphi
SIN3B		activator/repressor	Paired amphi
SMAD7	MAD homolo	TF	Mothers agai
SMARCB1	BAF47|hSNF5	activator/repressor	SWI/SNF-rela
SP1		TF	Transcription
SRF	SRF	TF	Serum respor
SSRP1	FACT 80 kDa s	initiation_complex	FACT complex
TADA3L	ADA3-like pro	TF	Transcription
AP-2 alpha	AP-2|AP2-alp	TF	Transcription
Hic-5	Hic-5	activator/repressor	Transforming
UBE2D1	SFT		Ubiquitin-con
UBF	UBF-1	TF	Nucleolar tra
USP28			Ubiquitin carl
YY1	YY-1	TF	Transcription
YY2	YY-2	TF	Transcription
FBI1	FBI-1|POK er	TF	Zinc finger an
TCF8	TCF-8	activator/repressor	Zinc finger E-b
ATBF1	ZFH-3	TF	Zinc finger ho
ZHX1		TF	Zinc fingers a
ZIC3		TF	Zinc finger pr
ZNF261			Zinc finger M

ec_number	location	go_terms	family	cansar_link
eptor	Nucleus, Cytoplasm		nuclear horm	<a href="http://cansar.icr.ac.uk/cansar/r
2.7.11.1	Nucleus, Cytc	cell body, cyt	TKL Ser/Thr p	<a href="http://cansar.icr.ac.uk/cansar/r
eptor	Nucleus, Cytoplasm, Cell m		nuclear horm	<a href="http://cansar.icr.ac.uk/cansar/r
3.5.1.98	Nucleus, Cytc	cytoplasm, cy	histone deace	<a href="http://cansar.icr.ac.uk/cansar/r
ltage-gated cl	Cell membrar	cell surface, p	potassium ch	<a href="http://cansar.icr.ac.uk/cansar/r
2.7.12.2	Cytoplasm, cy	cytoplasm, cy	STE Ser/Thr p	<a href="http://cansar.icr.ac.uk/cansar/r
d receptor	Cytoplasm, Nucleus, Mitocl		nuclear horm	<a href="http://cansar.icr.ac.uk/cansar/r
receptor alph	Nucleus, Cytc	actin cytoskel	nuclear horm	<a href="http://cansar.icr.ac.uk/cansar/r
receptor beta	Nucleus, Nuc	cytoplasm, nu	nuclear horm	<a href="http://cansar.icr.ac.uk/cansar/r
receptor gam	Nucleus	integral comp	nuclear horm	<a href="http://cansar.icr.ac.uk/cansar/r
receptor RXR-	Nucleus	axon, nuclear	nuclear horm	<a href="http://cansar.icr.ac.uk/cansar/r
2.7.11.1	Cytoplasm, N	cell-cell junct	AGC Ser/Thr p	<a href="http://cansar.icr.ac.uk/cansar/r
2.7.11.1	Nucleus, Cytc	cytoplasmic, I	PI3/PI4-kinas	<a href="http://cansar.icr.ac.uk/cansar/r
2.7.11.1	Membrane, C	amphisome, a	TKL Ser/Thr p	<a href="http://cansar.icr.ac.uk/cansar/r
6.3.2.-	Nucleus, nucl	cytoplasm, cy	MDM2/MDM	<a href="http://cansar.icr.ac.uk/cansar/r
3.1.4.53	Apical cell me	apical plasma	cyclic nucleot	<a href="http://cansar.icr.ac.uk/cansar/r
2.7.1.153		cytosol, lame	PI3/PI4-kinas	<a href="http://cansar.icr.ac.uk/cansar/r
2.7.1.153	Cytoplasm, N	cytosol, inter	PI3/PI4-kinas	<a href="http://cansar.icr.ac.uk/cansar/r
r antigen p53	Cytoplasm, Nucleus, Nucle		p53 family	<a href="http://cansar.icr.ac.uk/cansar/r
bon receptor	Cytoplasm, N	cytoplasm, cytosolic aryl hy		<a href="http://cansar.icr.ac.uk/cansar/r
6.3.2.-	Nucleus, Chrc	BRCA1-A complex, BRCA1-E		<a href="http://cansar.icr.ac.uk/cansar/r
1	Cytoplasm, N	adherens jun	beta-catenin	<a href="http://cansar.icr.ac.uk/cansar/r
	Cell membrane, Cell memb		Ras family	<a href="http://cansar.icr.ac.uk/cansar/r
1.1.1.42	Cytoplasm, P	cytoplasm, cy	isocitrate and	<a href="http://cansar.icr.ac.uk/cansar/r
trisphosphate	Endoplasmic	calcineurin co	InsP3 receptc	<a href="http://cansar.icr.ac.uk/cansar/r
trisphosphate	Endoplasmic	cell cortex, er	InsP3 receptc	<a href="http://cansar.icr.ac.uk/cansar/r
factor AP-1	Nucleus	cytosol, nucle	bZIP family	<a href="http://cansar.icr.ac.uk/cansar/r
2.7.11.25		cytoplasm, cy	STE Ser/Thr p	<a href="http://cansar.icr.ac.uk/cansar/r
	Nucleus	chromatin, cleavage furrow		<a href="http://cansar.icr.ac.uk/cansar/r
tor corepress	Nucleus	membrane, n	N-CoR nuclea	<a href="http://cansar.icr.ac.uk/cansar/r
tor corepress	Nucleus	histone deace	N-CoR nuclea	<a href="http://cansar.icr.ac.uk/cansar/r
tor subfamily	Cytoplasm, N	cytoplasm, nu	nuclear horm	<a href="http://cansar.icr.ac.uk/cansar/r
nositol 3-kinase regulatory		cell-cell junct	PI3K p85 subu	<a href="http://cansar.icr.ac.uk/cansar/r
ine-protein phosphatase 2		cytosol, nucle	phosphatase	<a href="http://cansar.icr.ac.uk/cansar/r
receptor 2	Cell membrar	endoplasmic	ATR family	<a href="http://cansar.icr.ac.uk/cansar/r
s polyposis co	Cell junction,	adherens jun	adenomatous	<a href="http://cansar.icr.ac.uk/cansar/r
bon receptor	Nucleus	cytoplasm, nucleoplasm, nu		<a href="http://cansar.icr.ac.uk/cansar/r
2.1.1.43	Nucleus, Cell	bicellular tigh	Histone-lysin	<a href="http://cansar.icr.ac.uk/cansar/r
type 2 suscep	Nucleus, Cytc	BRCA2-MAGE-D1 complex,		<a href="http://cansar.icr.ac.uk/cansar/r
2T2	Nucleus	nucleus	CBFA2T famil	<a href="http://cansar.icr.ac.uk/cansar/r

	Cell membrar	integral component of mer		<a href="http://cansar.icr.ac.uk/cansar/r
2.7.11.22	Nucleus, Nuc	cyclin K-CDK1	CMGC Ser/Th	<a href="http://cansar.icr.ac.uk/cansar/r
3.6.4.12	Nucleus, Cytc	cytoplasm, nu	SNF2/RAD54	<a href="http://cansar.icr.ac.uk/cansar/r
2.3.1.48	Cytoplasm, N	cytoplasm, histone acetyltr		<a href="http://cansar.icr.ac.uk/cansar/r
ctivating prot	Cytoplasm, C	caveola, cortical actin cytos		<a href="http://cansar.icr.ac.uk/cansar/r
cytokinesis protein 10		extracellular (	DOCK family	<a href="http://cansar.icr.ac.uk/cansar/r
al regulator El	Nucleus, Cytc	cytoplasm, in	ETS family	<a href="http://cansar.icr.ac.uk/cansar/r
2.3.1.-	Nucleus, Chrc	chromatin, nu	acetyltransfe	<a href="http://cansar.icr.ac.uk/cansar/r
tion variant 3	Nucleus	nuclear chror	ETS family	<a href="http://cansar.icr.ac.uk/cansar/r
uclear factor 1	Nucleus	microvillus, nucleolus, nucl		<a href="http://cansar.icr.ac.uk/cansar/r
protein G1	Nucleus	nucleus		<a href="http://cansar.icr.ac.uk/cansar/r
protein O1	Cytoplasm, N	cytoplasm, cytosol, mitoch		<a href="http://cansar.icr.ac.uk/cansar/r
protein O4	Cytoplasm, N	cytoplasm, cytosol, nucleop		<a href="http://cansar.icr.ac.uk/cansar/r
protein P1	Nucleus	nucleoplasm, nucleus		<a href="http://cansar.icr.ac.uk/cansar/r
nscription fact	Nucleus	nucleoplasm, nucleus, tran		<a href="http://cansar.icr.ac.uk/cansar/r
anscription fa	Nucleus	cytosol, mast cell granule, r		<a href="http://cansar.icr.ac.uk/cansar/r
erived growth	Cytoplasm, N	cytoplasm, ex	HDGF family	<a href="http://cansar.icr.ac.uk/cansar/r
uclear factor 4	Nucleus	cytoplasm, nu	nuclear horm	<a href="http://cansar.icr.ac.uk/cansar/r
receptor subu	Cell membrar	ciliary neurot	type I cytokin	<a href="http://cansar.icr.ac.uk/cansar/r
2.3.1.48	Nucleus, Nuc	Golgi apparat	MYST (SAS/M	<a href="http://cansar.icr.ac.uk/cansar/r
1.14.11.-	Nucleus	cytoplasm, nu	JHDM3 histor	<a href="http://cansar.icr.ac.uk/cansar/r
1.14.11.-	Nucleus	histone meth	UTX family	<a href="http://cansar.icr.ac.uk/cansar/r
2.1.1.43	Nucleus	histone meth	Histone-lysin	<a href="http://cansar.icr.ac.uk/cansar/r
2.1.1.43	Nucleus	histone meth	Histone-lysin	<a href="http://cansar.icr.ac.uk/cansar/r
6.3.2.-	Cytoplasm	cytoplasm, cytosol, extrace		<a href="http://cansar.icr.ac.uk/cansar/r
otein Nkx-3.1	Nucleus	intracellular,	NK-3 homeob	<a href="http://cansar.icr.ac.uk/cansar/r
6.3.2.-	Nucleus spec	nuclear speck	PIAS family	<a href="http://cansar.icr.ac.uk/cansar/r
6.3.2.-	Nucleus spec	nuclear speck	PIAS family	<a href="http://cansar.icr.ac.uk/cansar/r
3.1.3.16	Cytoplasm, N	apical plasma membrane, c		<a href="http://cansar.icr.ac.uk/cansar/r
ma-associated	Nucleus	chromatin, nu	retinoblaston	<a href="http://cansar.icr.ac.uk/cansar/r
6.3.2.-	Cell membrar	endoplasmic	ZNRF3 family	<a href="http://cansar.icr.ac.uk/cansar/r
	Nucleus	centrosome,	SKI family	<a href="http://cansar.icr.ac.uk/cansar/r
nst decapenta	Cytoplasm, N	activin respor	dwarfin/SMA	<a href="http://cansar.icr.ac.uk/cansar/r
nst decapenta	Cytoplasm, N	cytoplasm, cy	dwarfin/SMA	<a href="http://cansar.icr.ac.uk/cansar/r
nst decapenta	Cytoplasm, N	activin respor	dwarfin/SMA	<a href="http://cansar.icr.ac.uk/cansar/r
3.6.4.-	Nucleus	extracellular :	SNF2/RAD54	<a href="http://cansar.icr.ac.uk/cansar/r
domain-conta	Nucleus	nucleus	ETS family	<a href="http://cansar.icr.ac.uk/cansar/r
POZ protein	Nucleus, Nuc	Cul3-RING ub	Tdpoz family	<a href="http://cansar.icr.ac.uk/cansar/r
D repeat-cont	Nucleus	nucleoplasm,	WD repeat EE	<a href="http://cansar.icr.ac.uk/cansar/r
D repeat-cont	Nucleus	nucleoplasm,	WD repeat EE	<a href="http://cansar.icr.ac.uk/cansar/r
iption factor T	Nucleus	nucleus		<a href="http://cansar.icr.ac.uk/cansar/r
n p73	Nucleus, Cytc	cell junction,	p53 family	<a href="http://cansar.icr.ac.uk/cansar/r
d BTB domain	Nucleus	cytosol, nucle	krueppel C2H	<a href="http://cansar.icr.ac.uk/cansar/r
nd homeobox	Nucleus	cytoplasm, nu	ZHX family	<a href="http://cansar.icr.ac.uk/cansar/r

	Nucleus, nucl	centrosome,	AATF family	<a href="http://cansar.icr.ac.uk/cansar/r
3.4.24.-	Cell membrar	extracellular region, integra		<a href="http://cansar.icr.ac.uk/cansar/r
ıal enhancer c	Nucleus	nucleus	WD repeat G	<a href="http://cansar.icr.ac.uk/cansar/r
A4 protein	Membrane, Membrane, cla		APP family	<a href="http://cansar.icr.ac.uk/cansar/r
ıctive domain-	Nucleus	nBAF complex, npBAF com		<a href="http://cansar.icr.ac.uk/cansar/r
ıctive domain-	Nucleus	BAF-type complex, nucleop		<a href="http://cansar.icr.ac.uk/cansar/r
ıctive domain-	Nucleus, Cytc	cytoplasm, nucleoplasm, nu		<a href="http://cansar.icr.ac.uk/cansar/r
comb group p	Nucleus	nucleoplasm	Asx family	<a href="http://cansar.icr.ac.uk/cansar/r
modifier-acti	Cytoplasm, Pı	axoneme, cyt	ATG7 family	<a href="http://cansar.icr.ac.uk/cansar/r
	Cytoplasm, N	cytoplasm, in	ATXN1 family	<a href="http://cansar.icr.ac.uk/cansar/r
ıscription fact	Nucleus	cytoplasm, nucleoplasm		<a href="http://cansar.icr.ac.uk/cansar/r
otein castor h	Nucleus	cytoplasm, intracellular me		<a href="http://cansar.icr.ac.uk/cansar/r
	Cell membrar	integral component of mer		<a href="http://cansar.icr.ac.uk/cansar/r
dent kinase in	Nucleus, Cytc	Cul4A-RING E	CDI family	<a href="http://cansar.icr.ac.uk/cansar/r
ıcer-binding p	Nucleus, Cytc	CHOP-C/EBP	bZIP family	<a href="http://cansar.icr.ac.uk/cansar/r
ınscription co	Cytoplasm, N	CCR4-NOT co	CNOT2/3/5 fa	<a href="http://cansar.icr.ac.uk/cansar/r
ınscription co	Cytoplasm, N	CCR4-NOT co	CNOT2/3/5 fa	<a href="http://cansar.icr.ac.uk/cansar/r
ed transcriptic	Cytoplasm, N	cytoplasm, de	TORC family	<a href="http://cansar.icr.ac.uk/cansar/r
ole homeobox	Nucleus	nucleus	paired home	<a href="http://cansar.icr.ac.uk/cansar/r
factor E2F3	Nucleus	cytoplasm, nu	E2F/DP famil	<a href="http://cansar.icr.ac.uk/cansar/r
cting transcrip	Nucleus	nucleoplasm		<a href="http://cansar.icr.ac.uk/cansar/r
protein H1	Nucleus	activin responsive factor co		<a href="http://cansar.icr.ac.uk/cansar/r
protein O3	Cytoplasm, cy	cytoplasm, cytosol, membr		<a href="http://cansar.icr.ac.uk/cansar/r
nan	Membrane	chloride chan	FXYD family	<a href="http://cansar.icr.ac.uk/cansar/r
al activator GI	Nucleus, Cytc	axoneme, cili	GLI C2H2-typ	<a href="http://cansar.icr.ac.uk/cansar/r
s protein-like	Nucleus matr	nuclear envelope, nuclear r		<a href="http://cansar.icr.ac.uk/cansar/r
ain-containing	protein 8			<a href="http://cansar.icr.ac.uk/cansar/r
hway suppressor	2	nucleoplasm, transcriptiona		<a href="http://cansar.icr.ac.uk/cansar/r
	Nucleus, Chromosome		histone H3 fa	<a href="http://cansar.icr.ac.uk/cansar/r
receptor-asso	Cytoplasm, N	cytoplasm, in	IRAK1BP1 fan	<a href="http://cansar.icr.ac.uk/cansar/r
a-4	Membrane	cell surface, e	integrin alpha	<a href="http://cansar.icr.ac.uk/cansar/r
2.3.1.48	Nucleus, Nuc	NuA4 histone	MYST (SAS/M	<a href="http://cansar.icr.ac.uk/cansar/r
ipoprotein rec	Membrane	integral comp	LDLR family	<a href="http://cansar.icr.ac.uk/cansar/r
sociated antig	Nucleus, Cytc	cytoplasm, mitochondrion,		<a href="http://cansar.icr.ac.uk/cansar/r
I1 complex lo	Nucleus, Nuc	aggresome, cytoplasm, cyt		<a href="http://cansar.icr.ac.uk/cansar/r
I1 complex locus protein		MDS1		<a href="http://cansar.icr.ac.uk/cansar/r
cific enhancer	Nucleus	cytoplasm, nu	MEF2 family	<a href="http://cansar.icr.ac.uk/cansar/r
	Nucleus	cytoplasm, nucleolus, nucle		<a href="http://cansar.icr.ac.uk/cansar/r
associated tur	Mitochondric	extracellular	MTUS1 famil	<a href="http://cansar.icr.ac.uk/cansar/r
ıcogene prote	Nucleus, nucl	cytosol, nucleolus, nucleop		<a href="http://cansar.icr.ac.uk/cansar/r
otor coactivat	Nucleus	intracellular,	OXR1 family	<a href="http://cansar.icr.ac.uk/cansar/r
r 1 A-type	Nucleus	cell junction,	CTF/NF-I fam	<a href="http://cansar.icr.ac.uk/cansar/r
cription facto	Nucleus	CCAAT-bindir	NFYC/HAP5 s	<a href="http://cansar.icr.ac.uk/cansar/r
ition complex	Nucleus	actin cytoskel	ORC4 family	<a href="http://cansar.icr.ac.uk/cansar/r

3.1.3.4	Cell membrar	extracellular (	PA-phosphata	<a href="http://cansar.icr.ac.uk/cansar/r
	Nucleus, Nuc	cytoplasm, cytosol, early er		<a href="http://cansar.icr.ac.uk/cansar/r
3.1.3.16	Cytoplasm, N	chromosome	PPP phosphat	<a href="http://cansar.icr.ac.uk/cansar/r
2.7.11.11	Cytoplasm, C	calcium chan	AGC Ser/Thr p	<a href="http://cansar.icr.ac.uk/cansar/r
ocessing factc	Nucleus spec	cytoplasm, m	PRPF40 famil	<a href="http://cansar.icr.ac.uk/cansar/r
na-like protei	Nucleus	nucleoplasm, retinoblaston		<a href="http://cansar.icr.ac.uk/cansar/r
6.3.2.-	Nucleus	cytoplasm, cy	RNF12 family	<a href="http://cansar.icr.ac.uk/cansar/r
nomolog 1	Cell membrar	axon, cell sur	ROBO family	<a href="http://cansar.icr.ac.uk/cansar/r
nomolog 2	Membrane	axolemma, ce	ROBO family	<a href="http://cansar.icr.ac.uk/cansar/r
l protein L11	Nucleus, nucl	cytoplasm, cy	ribosomal pro	<a href="http://cansar.icr.ac.uk/cansar/r
transcription f	Nucleus	cytoplasm, nuclear chroma		<a href="http://cansar.icr.ac.uk/cansar/r
1-binding pro	Nucleus, Cytc	cytoplasm, nucleoplasm		<a href="http://cansar.icr.ac.uk/cansar/r
3.4.22.68	Nucleus	cytoplasm, nu	peptidase C48	<a href="http://cansar.icr.ac.uk/cansar/r
homolog	Cytoplasm, cy	cytoplasm, cy	SHQ1 family	<a href="http://cansar.icr.ac.uk/cansar/r
pathic helix pr	Nucleus, Nuc	cytoplasm, intercellular bri		<a href="http://cansar.icr.ac.uk/cansar/r
pathic helix pr	Nucleus	autosome, cytoplasm, nucl		<a href="http://cansar.icr.ac.uk/cansar/r
nst decapenta	Nucleus, Cytc	centrosome,	dwarfin/SMA	<a href="http://cansar.icr.ac.uk/cansar/r
ted matrix-ass	Nucleus	nBAF comple	SNF5 family	<a href="http://cansar.icr.ac.uk/cansar/r
factor Sp1	Nucleus, Cytc	cytoplasm, nu	Sp1 C2H2-typ	<a href="http://cansar.icr.ac.uk/cansar/r
ıse factor	Nucleus	cytoplasm, nuclear chroma		<a href="http://cansar.icr.ac.uk/cansar/r
subunit SSRF	Nucleus, Nuc	chromosome	SSRP1 family	<a href="http://cansar.icr.ac.uk/cansar/r
al adapter 3	Nucleus	Ada2/Gcn5/A	NGG1 family	<a href="http://cansar.icr.ac.uk/cansar/r
factor AP-2-a	Nucleus	centrosome,	AP-2 family	<a href="http://cansar.icr.ac.uk/cansar/r
growth factor	Cell junction,	cytoplasm, cy	paxillin family	<a href="http://cansar.icr.ac.uk/cansar/r
2.3.2.23	Cytoplasm	cytoplasm, cy	ubiquitin-con	<a href="http://cansar.icr.ac.uk/cansar/r
nscription fact	Nucleus, nucl	fibrillar center, nucleolus, n		<a href="http://cansar.icr.ac.uk/cansar/r
3.4.19.12	Nucleus, nucl	cytoplasm, nu	peptidase C19	<a href="http://cansar.icr.ac.uk/cansar/r
al repressor p	Nucleus matr	Ino80 comple	YY transcripti	<a href="http://cansar.icr.ac.uk/cansar/r
factor YY2	Nucleus	cytoplasm, nu	YY transcripti	<a href="http://cansar.icr.ac.uk/cansar/r
d BTB domain	Nucleus	nucleus		<a href="http://cansar.icr.ac.uk/cansar/r
box-binding h	Nucleus	cytoplasm, nu	delta-EF1/ZFH	<a href="http://cansar.icr.ac.uk/cansar/r
meobox prote	Nucleus, Cytc	cytoplasm, intracellular me		<a href="http://cansar.icr.ac.uk/cansar/r
nd homeobox	Nucleus	nucleoplasm,	ZHX family	<a href="http://cansar.icr.ac.uk/cansar/r
otein ZIC 3	Nucleus, Cytc	cytoplasm, nu	GLI C2H2-typ	<a href="http://cansar.icr.ac.uk/cansar/r
YM-type prote	Nucleus	cytoplasm, nucleoplasm		<a href="http://cansar.icr.ac.uk/cansar/r

nearest_drug	number_of_3	number_of_3	number_of_c	number_of_c	ligand_drugg	screened_cor
100.00|P102	106	91	47	47	95	0
100.00|P150	60	57	50	51	93	1413
100.00|P033	193	182	105	105	94	5689
100.00|O153	1	0	0	0	93	1769
100.00|Q128	10	0	0	0	96	12236
100.00|Q027	41	41	38	39	92	1798
100.00|P041	47	21	13	14	94	3419
100.00|P102	24	22	3	9	88	606
100.00|P108	24	22	5	9	94	567
100.00|P136	24	22	1	9	88	512
100.00|P197	83	60	43	47	89	812
32.41|O1492	38	30	15	30	95	3479
nolecular-targ	0	0	0	0	84	137
33.59|P4268	2	2	0	0	94	802
nolecular-targ	83	53	8	8	82	1426
72.23|P2781	75	72	37	70	95	2627
nolecular-targ	21	14	18	18	97	3777
nolecular-targ	2	2	2	2	96	1483
nolecular-targ	157	45	20	21	78	616
nolecular-targ	1	0	0	0	75	159
nolecular-targ	27	0	0	0	98	36
nolecular-targ	35	0	20	21	59	44
nolecular-targ	179	171	105	105	78	91
nolecular-targ	21	21	10	10	59	74
28.50|P2181	7	3	4	5	91	46
23.56|P2181	3	1	0	2	89	6
nolecular-targ	7	0	0	0	89	356
28.85|O6067	0	0	0	0	70	136
nolecular-targ	28	23	15	15	70	68
nolecular-targ	8	1	0	0	78	38
nolecular-targ	12	1	0	0	80	75
34.83|P1979	17	9	15	15	83	8
nolecular-targ	42	2	5	5	84	192
nolecular-targ	1	0	1	1	78	2
nolecular-targ	4	0	1	1	66	
nolecular-targ	20	0	1	1	94	0
nolecular-targ	22	2	1	1	59	11
nolecular-targ	7	6	5	5	84	
nolecular-targ	2	0	0	0	84	0
nolecular-targ	8	0	1	2	73	

nolecular-targ	0	0	0	0	59	
32.23|O6067	5	4	4	4	59	
nolecular-targ	10	0	4	4	83	
nolecular-targ	84	34	19	23	93	135
nolecular-targ	8	0	1	2	95	
nolecular-targ	0	0	0	0	97	
nolecular-targ	9	0	0	0	67	48
nolecular-targ	1	1	1	1	83	
nolecular-targ	0	0	0	0	59	
nolecular-targ	2	0	0	0	59	
nolecular-targ	0	0	0	0	59	
nolecular-targ	8	0	0	0	91	7
nolecular-targ	8	0	0	0	59	0
nolecular-targ	5	0	1	1	59	0
nolecular-targ	15	0	0	1	59	0
nolecular-targ	16	0	1	1	59	0
nolecular-targ	4	0	2	2	59	
43.46|P1979	7	2	3	3	86	18
nolecular-targ	8	1	3	3	82	3
nolecular-targ	9	2	3	3	59	0
nolecular-targ	65	56	51	57	68	51
nolecular-targ	14	12	14	14	82	10
nolecular-targ	7	3	2	3	97	1
nolecular-targ	5	3	1	3	93	1
nolecular-targ	24	1	4	8	59	
nolecular-targ	1	0	0	0	59	
nolecular-targ	1	0	0	0	59	
nolecular-targ	3	0	1	1	59	
nolecular-targ	6	4	3	3	59	0
nolecular-targ	15	1	8	8	81	10
nolecular-targ	1	0	1	1	59	
nolecular-targ	3	0	0	1	68	
nolecular-targ	9	0	1	2	59	3
nolecular-targ	12	0	2	3	59	3
nolecular-targ	7	0	2	2	59	0
nolecular-targ	8	3	1	2	86	48
nolecular-targ	2	0	0	0	59	
nolecular-targ	15	0	1	1	59	
nolecular-targ	4	0	0	1	59	
nolecular-targ	4	0	1	1	59	
nolecular-targ	1	0	1	1	63	
nolecular-targ	22	0	7	9	59	
nolecular-targ	2	0	0	0	59	0
nolecular-targ	2	0	0	0	73	

nolecular-targ	0	0	0	0	59	
nolecular-targ	0	0	0	0	59	
nolecular-targ	2	0	0	0	59	
nolecular-targ	121	9	4	4	87	424
nolecular-targ	3	0	0	0	91	
nolecular-targ	0	0	0	0	71	
nolecular-targ	1	0	0	0	59	
nolecular-targ	0	0	0	0	71	
nolecular-targ	0	0	0	0	89	2
nolecular-targ	6	0	2	2	72	
nolecular-targ	1	0	0	0	71	
nolecular-targ	0	0	0	0	59	
nolecular-targ	2	0	0	1	59	
nolecular-targ	3	0	0	0	59	
nolecular-targ	12	0	0	0	88	
nolecular-targ	4	0	0	0	59	
nolecular-targ	4	0	0	0	59	
nolecular-targ	0	0	0	0	59	0
nolecular-targ	0	0	0	0	59	
nolecular-targ	0	0	0	0	59	
nolecular-targ	3	0	0	0	59	
nolecular-targ	0	0	0	0	75	
nolecular-targ	10	0	0	0	91	3
nolecular-targ	1	0	0	0	59	
nolecular-targ	2	0	0	0	59	
nolecular-targ	0	0	0	0	59	
nolecular-targ	0	0	0	0	59	
nolecular-targ	1	0	0	0	59	
nolecular-targ	218	65	3	3	59	
nolecular-targ	0	0	0	0	63	
27.64|P0851	6	3	3	3	93	1660
nolecular-targ	3	1	1	1	59	25
nolecular-targ	0	0	0	0	86	
nolecular-targ	0	0	0	0	59	
nolecular-targ	0	0	0	0	59	0
nolecular-targ	0	0	0	0	59	0
nolecular-targ	9	1	0	0	59	
nolecular-targ	5	0	0	0	59	0
nolecular-targ	0	0	0	0	72	
nolecular-targ	7	0	0	0	59	36
nolecular-targ	0	0	0	0	59	
nolecular-targ	0	0	0	0	60	
nolecular-targ	3	0	0	0	59	
nolecular-targ	0	0	0	0	59	

nolecular-targ	0	0	0	0	59	
nolecular-targ	5	0	0	0	59	0
nolecular-targ	16	6	1	1	65	59
29.29|O1492	188	158	162	162	84	1931
nolecular-targ	10	0	0	0	95	
nolecular-targ	4	0	0	0	73	
nolecular-targ	0	0	0	0	90	
29.41|P1136	6	0	0	0	85	
26.75|P2180	6	0	0	0	59	
nolecular-targ	2	0	0	0	75	
nolecular-targ	23	0	0	0	59	1
nolecular-targ	2	0	0	0	59	
nolecular-targ	0	0	0	0	59	0
nolecular-targ	3	0	0	0	59	
nolecular-targ	9	0	0	0	84	
nolecular-targ	8	0	0	0	71	
nolecular-targ	7	0	0	0	59	
nolecular-targ	1	0	0	0	59	0
nolecular-targ	5	0	0	0	59	37
nolecular-targ	3	0	0	0	59	
nolecular-targ	1	0	0	0	59	
nolecular-targ	0	0	0	0	59	
nolecular-targ	0	0	0	0	59	
nolecular-targ	0	0	0	0	59	
nolecular-targ	53	0	0	3	64	
nolecular-targ	7	0	0	0	59	
nolecular-targ	2	0	0	0	72	1
nolecular-targ	3	1	0	0	59	
nolecular-targ	1	0	0	0	59	
nolecular-targ	2	0	0	0	59	
nolecular-targ	1	0	0	0	91	
nolecular-targ	4	0	0	0	68	
nolecular-targ	4	0	0	0	59	
nolecular-targ	2	0	0	0	59	
nolecular-targ	0	0	0	0	63	

screened_compounds_active at 50 nM or better
0
323
864
312
357
251
1345
101
115
81
184
605
2
141
326
600
873
359
44
21
1
1
7
3
19
2
14
2
13
14
27
1
67
1
0
0
0

113

0

0

0

931
0

0
0

0

0

0
2
68

0

0

0
0

0

UniprotAccesstion	Uniprot accedssion of the main gene product
Gene Symbol	HUGO gene Symbol
Source	The surce of the protein in the networ: input_list: prot
Subcellular_localisation	Summary term indicating main subcellular localisation
long_name	Full name of protein
node_label	Label as shown in the network
Druggability_Class	Druggability class: DT: Target of approved drug; CT: Tar
protein	Protein product symbol
Synonyms	Other synomyms
Transcriptional_class	For transcription factors, subclassificaton
description	Full description of prorein product
ec_number	For enzymes, the Enzyme Classification
location	Expanded detail of subcellular localisation
go_terms	Gene Ontology terms
family	Protein family
cansar_link	Link to live report in canSAR for this protein
nearest_drug_target	List of closest homologous drug targets
number_of_3d_structure within 75% hor	The number of 3D structures within 75% homology.
number_of_3d_structure_with_small_m	The number of 3D protein stryctures that have been sc
number_of_druggable_3d_structures wit	The number of druggable structures identified in canS
number_of_druggable_3d_structures wit	The number of druggable structures identified in canS
ligand_druggability_percentile_rank	Rank of druggability based on chemical information alc
screened_compounds	The number of published compouneds tested
screened_compounds_active at 50 nM or	The number of published compounds showing 50nM a

ein product of gene identified as a key altered gene in this study; additional_connectivity: key di

·get of drug under clinical investgation; AC: Target of compound in preclinical or discovery phase

)lved in complex with a small molecule ligand - this could be either a cofactor or inhibitor
AR that are within 90% homology of the protein - this is in effect direct evidence for the protein c
AR that are within 75% homology of the protein - this is a sufficiently close homology that the dru
)ne - 10% indicates it is in the top most druggable proteins based on chemical rather than protei

ctivity or better

rect interacting protein; additional_enriched: key enriched transcription factor (see online meth

exploration; Dr: none of the above but predicted druggable using canSAR

containing a druggable cavity
uggability mapping by homology is highly reliable
n structure information

ods)

SUPPLEMENTARY TABLE 6

Cell Line		
22RV	**Drug and targets**	22RV
22RV	AKT inhibitor VIII Target: AKT1, AKT2, AKT3	-1.03
22RV	AS605240 Target: PI3Kgamma	ns
22RV	AZD6482 Target: PI3Kbeta	ns
22RV	AZD8055 Target: MTORC1, MTORC2	-0.23
BPH-1	CP466722 Target: ATM	-0.30
BPH-1	Dactolisib Target: PI3K (Class 1), MTORC1, MTORC2	-0.43
BPH-1	GSK690693 Target: AKT1, AKT2, AKT3	ns
BPH-1	Idelalisib Target: PI3Kdelta	ns
DU-145	KU-55933 Target: ATM	ns
DU-145	MK-2206 Target: AKT1, AKT2	ns
LNCaP-Clone-FGC	Nutlin-3a (-) Target: MDM2	-0.84
LNCaP-Clone-FGC	Omipalisib Target: PI3K (class 1), MTORC1, MTORC2	ns
LNCaP-Clone-FGC	OSI-027 Target: MTORC1, MTORC2	ns
LNCaP-Clone-FGC	PI-103 Target: PI3Kalpha, DAPK3, CLK4, PIM3, HIPK2	ns
LNCaP-Clone-FGC	Pictilisib Target: PI3K (class 1)	ns
LNCaP-Clone-FGC	PIK-93 Target: PI3Kgamma	ns
LNCaP-Clone-FGC	Serdemetan Target: MDM2	ns
LNCaP-Clone-FGC	ZSTK474 Target: PI3K (class 1)	ns
LNCaP-Clone-FGC		
LNCaP-Clone-FGC		
LNCaP-Clone-FGC		
LNCaP-Clone-FGC		
	Data from the Genomics of Drug Sensitivity in Cancer (GDSC)- public d	
	values are shown as z-scores for drugs that act on targets within the Pc	
	prostate cancer cell lines in GDSC.	
LNCaP-Clone-FGC		
LNCaP-Clone-FGC		
LNCaP-Clone-FGC		
LNCaP-Clone-FGC		
LNCaP-Clone-FGC		
LNCaP-Clone-FGC		
LNCaP-Clone-FGC		
LNCaP-Clone-FGC		
PC-3		
PC-3		
PC-3		
PC-3		
PC-3		
PC-3		
PC-3		
PC-3		
PC-3		
PC-3		
PC-3		
PC-3		
PC-3		

PWR-1E
PWR-1E
PWR-1E
VCAP
VCAP
VCAP
VCAP
VCAP
VCAP
VCAP
VCAP
VCAP
VCAP
VCAP
VCAP

Cell line, Sensitivity in Z-score					
BPH-1	DU-145	LNCaP-Clone-FGC	PC-3	PWR-1E	VCAP
-1.07	ns	-1.17	-0.36	ns	-1.00
ns	ns	ns	ns	ns	-0.99
ns	ns	-1.46	ns	-0.28	ns
ns	ns	-0.78	-0.33	ns	ns
-0.31	ns	-0.54	ns	ns	-0.58
ns	ns	-0.63	-2.14	ns	ns
ns	ns	-3.31	-0.56	ns	-1.48
ns	ns	-2.07	ns	ns	ns
-0.21	-2.56	ns	-0.60	-0.60	-0.11
ns	-1.28	-1.44	-0.42	ns	ns
-0.41	ns	-1.86	ns	ns	ns
ns	ns	-0.63	-0.30	ns	-1.36
ns	ns	-1.22	-1.17	ns	-0.27
ns	ns	ns	-0.82	ns	-1.12
ns	ns	-0.91	-0.51	-0.18	ns
ns	ns	-0.20	-0.59	ns	-0.61
ns	ns	-2.33	ns	ns	-1.68
ns	ns	ns	-0.70	ns	-1.14

not sensitive			marginal		sensitive

lataset provided by the Sanger Instiute (http://www.cancerrxgene.org). Sensitivity
a druggable disease network identified in our study, and that have been tested on

SUPPLEMENTARY TABLE 7

fig1_order	fig2_order	Sanger ID	ICGC sample ID	ICGC Donor ID
1	62	PD12337a	0096_CRUK_PC_0096_M1	0096_CRUK_PC_0096
2	61	PD11335a	0022_CRUK_PC_0022_M1	0022_CRUK_PC_0022
3	59	PD13412a	0095_CRUK_PC_0095_M1	0095_CRUK_PC_0095
4	54	PD11330a	0017_CRUK_PC_0017_M1	0017_CRUK_PC_0017
5	60	PD11328a	0015_CRUK_PC_0015_M1	0015_CRUK_PC_0015
6	56	PD11332a	0019_CRUK_PC_0019_M1	0019_CRUK_PC_0019
7	53	PD11329a	0016_CRUK_PC_0016_M1	0016_CRUK_PC_0016
8	58	PD11334a	0021_CRUK_PC_0021_M1	0021_CRUK_PC_0021
9	57	PD11333a	0020_CRUK_PC_0020_M1	0020_CRUK_PC_0020
10	55	PD11331a	0018_CRUK_PC_0018_M1	0018_CRUK_PC_0018
11	74	PD12843a	0060_CRUK_PC_0060_T1	0060_CRUK_PC_0060
12	63	PD14722a	0098_CRUK_PC_0098_M1	0098_CRUK_PC_0098
13	65	PD14731a	0107_CRUK_PC_0107_M1	0107_CRUK_PC_0107
14	5	PD12845a	0062_CRUK_PC_0062_T1	0062_CRUK_PC_0062
15	50	PD14728a	0104_CRUK_PC_0104_M1	0104_CRUK_PC_0104
16	85	PD12812a	0029_CRUK_PC_0029_T1	0029_CRUK_PC_0029
17	86	PD13380a	0063_CRUK_PC_0063_T1	0063_CRUK_PC_0063
18	23	PD12808a	0025_CRUK_PC_0025_T1	0025_CRUK_PC_0025
19	46	PD14721a	0097_CRUK_PC_0097_M1	0097_CRUK_PC_0097
20	25	PD13411a	0094_CRUK_PC_0094_T1	0094_CRUK_PC_0094
21	30	PD13388a	0071_CRUK_PC_0071_T1	0071_CRUK_PC_0071
22	47	PD14724a	0100_CRUK_PC_0100_M1	0100_CRUK_PC_0100
23	67	PD13408a	0091_CRUK_PC_0091_T1	0091_CRUK_PC_0091
24	78	PD12844a	0061_CRUK_PC_0061_T1	0061_CRUK_PC_0061
25	51	PD14729a	0105_CRUK_PC_0105_M1	0105_CRUK_PC_0105
26	45	PD14734a	0110_CRUK_PC_0110_T1	0110_CRUK_PC_0110
27	18	PD9746a	0009_CRUK_PC_0009_T1	0009_CRUK_PC_0009
28	71	PD12817a	0034_CRUK_PC_0034_T1	0034_CRUK_PC_0034
29	102	PD13381a	0064_CRUK_PC_0064_T1	0064_CRUK_PC_0064
30	87	PD13395a	0078_CRUK_PC_0078_T1	0078_CRUK_PC_0078
31	64	PD14726a	0102_CRUK_PC_0102_M1	0102_CRUK_PC_0102
32	103	PD13382a	0065_CRUK_PC_0065_T1	0065_CRUK_PC_0065
33	69	PD12811a	0028_CRUK_PC_0028_T1	0028_CRUK_PC_0028
34	83	PD9749a	0012_CRUK_PC_0012_T1	0012_CRUK_PC_0012
35	101	PD12842a	0059_CRUK_PC_0059_T1	0059_CRUK_PC_0059
36	17	PD14706a	0113_CRUK_PC_0113_T1	0113_CRUK_PC_0113
37	12	PD12830a	0047_CRUK_PC_0047_T1	0047_CRUK_PC_0047
38	109	PD14714a	0122_CRUK_PC_0122_T1	0122_CRUK_PC_0122
39	100	PD12838a	0055_CRUK_PC_0055_T1	0055_CRUK_PC_0055
40	52	PD14730a	0106_CRUK_PC_0106_M1	0106_CRUK_PC_0106
41	90	PD14713a	0120_CRUK_PC_0120_T1	0120_CRUK_PC_0120
42	68	PD13394a	0077_CRUK_PC_0077_T1	0077_CRUK_PC_0077
43	49	PD14727a	0103_CRUK_PC_0103_M1	0103_CRUK_PC_0103
44	92	PD12810a	0027_CRUK_PC_0027_T1	0027_CRUK_PC_0027
45	66	PD12809a	0026_CRUK_PC_0026_T1	0026_CRUK_PC_0026
46	88	PD13402a	0085_CRUK_PC_0085_T1	0085_CRUK_PC_0085
47	13	PD12836a	0053_CRUK_PC_0053_T1	0053_CRUK_PC_0053

48	73	PD9169a	0005_CRUK_PC_0005_T1	0005_CRUK_PC_0005
49	6	PD9168a	0004_CRUK_PC_0004_T1	0004_CRUK_PC_0004
50	9	PD12835a	0052_CRUK_PC_0052_T1	0052_CRUK_PC_0052
51	15	PD13401a	0084_CRUK_PC_0084_T1	0084_CRUK_PC_0084
52	77	PD12824a	0041_CRUK_PC_0041_T1	0041_CRUK_PC_0041
53	98	PD12828a	0045_CRUK_PC_0045_T1	0045_CRUK_PC_0045
54	8	PD12814a	0031_CRUK_PC_0031_T1	0031_CRUK_PC_0031
55	99	PD12837a	0054_CRUK_PC_0054_T1	0054_CRUK_PC_0054
56	110	PD14718a	0126_CRUK_PC_0126_T1	0126_CRUK_PC_0126
57	76	PD13407a	0090_CRUK_PC_0090_T1	0090_CRUK_PC_0090
58	34	PD12820a	0037_CRUK_PC_0037_T1	0037_CRUK_PC_0037
59	11	PD12813a	0030_CRUK_PC_0030_T1	0030_CRUK_PC_0030
60	72	PD12834a	0051_CRUK_PC_0051_T1	0051_CRUK_PC_0051
61	111	PD14720a	0128_CRUK_PC_0128_T1	0128_CRUK_PC_0128
62	19	PD13387a	0070_CRUK_PC_0070_T1	0070_CRUK_PC_0070
63	82	PD12840a	0057_CRUK_PC_0057_T1	0057_CRUK_PC_0057
64	22	PD13389a	0072_CRUK_PC_0072_T1	0072_CRUK_PC_0072
65	10	PD14707a	0114_CRUK_PC_0114_T1	0114_CRUK_PC_0114
66	32	PD14716a	0124_CRUK_PC_0124_T1	0124_CRUK_PC_0124
67	24	PD12829a	0046_CRUK_PC_0046_T1	0046_CRUK_PC_0046
68	39	PD13396a	0079_CRUK_PC_0079_T1	0079_CRUK_PC_0079
69	36	PD12832a	0049_CRUK_PC_0049_T1	0049_CRUK_PC_0049
70	7	PD14719a	0127_CRUK_PC_0127_T1	0127_CRUK_PC_0127
71	35	PD12825a	0042_CRUK_PC_0042_T1	0042_CRUK_PC_0042
72	79	PD9751a	0014_CRUK_PC_0014_T1	0014_CRUK_PC_0014
73	105	PD13393a	0076_CRUK_PC_0076_T1	0076_CRUK_PC_0076
74	43	PD14717a	0125_CRUK_PC_0125_T1	0125_CRUK_PC_0125
75	4	PD12806a	0023_CRUK_PC_0023_T1	0023_CRUK_PC_0023
76	89	PD13405a	0088_CRUK_PC_0088_T1	0088_CRUK_PC_0088
77	84	PD12807a	0024_CRUK_PC_0024_T1	0024_CRUK_PC_0024
78	48	PD14725a	0101_CRUK_PC_0101_M1	0101_CRUK_PC_0101
79	75	PD13383a	0066_CRUK_PC_0066_T1	0066_CRUK_PC_0066
80	3	PD12821a	0038_CRUK_PC_0038_T1	0038_CRUK_PC_0038
81	108	PD14709a	0116_CRUK_PC_0116_T1	0116_CRUK_PC_0116
82	81	PD13406a	0089_CRUK_PC_0089_T1	0089_CRUK_PC_0089
83	29	PD14735a	0111_CRUK_PC_0111_T1	0111_CRUK_PC_0111
84	21	PD12818a	0035_CRUK_PC_0035_T1	0035_CRUK_PC_0035
85	16	PD13404a	0087_CRUK_PC_0087_T1	0087_CRUK_PC_0087
86	NA	PD6204a	0001_CRUK_PC_0001_T1	0001_CRUK_PC_0001
87	38	PD13391a	0074_CRUK_PC_0074_T1	0074_CRUK_PC_0074
88	1	PD14736a	0112_CRUK_PC_0112_T1	0112_CRUK_PC_0112
89	2	PD13399a	0082_CRUK_PC_0082_T1	0082_CRUK_PC_0082
90	80	PD14711a	0118_CRUK_PC_0118_T1	0118_CRUK_PC_0118
91	33	PD12819a	0036_CRUK_PC_0036_T1	0036_CRUK_PC_0036
92	31	PD13410a	0093_CRUK_PC_0093_T1	0093_CRUK_PC_0093
93	104	PD13390a	0073_CRUK_PC_0073_T1	0073_CRUK_PC_0073
94	28	PD14708a	0115_CRUK_PC_0115_T1	0115_CRUK_PC_0115
95	40	PD13409a	0092_CRUK_PC_0092_T1	0092_CRUK_PC_0092
96	27	PD12822a	0039_CRUK_PC_0039_T1	0039_CRUK_PC_0039
97	26	PD12831a	0048_CRUK_PC_0048_T1	0048_CRUK_PC_0048

98	112	PD14732a	0108_CRUK_PC_0108_T1	0108_CRUK_PC_0108
99	93	PD12841a	0058_CRUK_PC_0058_T1	0058_CRUK_PC_0058
100	91	PD13384a	0067_CRUK_PC_0067_T1	0067_CRUK_PC_0067
101	107	PD13398a	0081_CRUK_PC_0081_T1	0081_CRUK_PC_0081
102	95	PD12816a	0033_CRUK_PC_0033_T1	0033_CRUK_PC_0033
103	41	PD14710a	0117_CRUK_PC_0117_T1	0117_CRUK_PC_0117
104	37	PD13386a	0069_CRUK_PC_0069_T1	0069_CRUK_PC_0069
105	42	PD14712a	0119_CRUK_PC_0119_T1	0119_CRUK_PC_0119
106	94	PD12815a	0032_CRUK_PC_0032_T1	0032_CRUK_PC_0032
107	44	PD14733a	0109_CRUK_PC_0109_T1	0109_CRUK_PC_0109
108	20	PD13392a	0075_CRUK_PC_0075_T1	0075_CRUK_PC_0075
109	97	PD12826a	0043_CRUK_PC_0043_T1	0043_CRUK_PC_0043
110	14	PD13400a	0083_CRUK_PC_0083_T1	0083_CRUK_PC_0083
111	96	PD12823a	0040_CRUK_PC_0040_T1	0040_CRUK_PC_0040
112	106	PD13397a	0080_CRUK_PC_0080_T1	0080_CRUK_PC_0080

PELICAN ID	EGA Accession	Source
A21-19888-LRib5MassMet	EGAN00001202474	PELICAN
A34-19956-LiverMet3	EGAN00001202462	PELICAN
A17-20205-Abd Paraaortic LN Met	EGAN00001202468	PELICAN
A29-19932-RSuperficialInguinalLNMetA1	EGAN00001202438	PELICAN
A10-19916-RIliacLNMet	EGAN00001202422	PELICAN
A32-19941-RRib8Met1-11	EGAN00001202447	PELICAN
A22-19921-LHumerusBoneMarrowMet	EGAN00001202427	PELICAN
A24-19951-RAxillaryLNMet	EGAN00001202457	PELICAN
A12-19947-MediastinalLNMet	EGAN00001202453	PELICAN
A31-19935-RSubduralMet	EGAN00001202441	PELICAN
	EGAN00001250305	NDS (Oxford)
PB133-20219-NAD Pelvic LN MET	EGAN00001375298	PELICAN
PB635-20237-NAD Pelvic LN Met	EGAN00001375314	PELICAN
	EGAN00001250309	NDS (Oxford)
PB421-20231-NAD Pelvic LN MET	EGAN00001375308	PELICAN
	EGAN00001250247	ICR (Sutton)
	EGAN00001250311	CRI (Cambridge)
	EGAN00001250239	ICR (Sutton)
PB50-20217-NAD Pelvic LN MET	EGAN00001375296	PELICAN
	EGAN00001195729	ICR (Sutton)
	EGAN00001195703	CRI (Cambridge)
PB273-20223-NAD Pelvic LN MET	EGAN00001375300	PELICAN
	EGAN00001195725	ICR (Sutton)
	EGAN00001250307	NDS (Oxford)
PB522-20233-NAD Pelvic LN MET	EGAN00001375310	PELICAN
	EGAN00001250367	ICR (Sutton)
	EGAN00001250379	Shanghai
	EGAN00001250257	ICR (Sutton)
	EGAN00001195693	CRI (Cambridge)
	EGAN00001195711	CRI (Cambridge)
PB399-20227-NAD Pelvic LN MET	EGAN00001375304	PELICAN
	EGAN00001195695	CRI (Cambridge)
	EGAN00001250245	ICR (Sutton)
	EGAN00001250381	Shanghai
	EGAN00001250303	CRI (Cambridge)
	EGAN00001250335	CRI (Cambridge)
	EGAN00001250283	ICR (Sutton)
	EGAN00001250351	CRI (Cambridge)
	EGAN00001250297	CRI (Cambridge)
PB556-20235-NAD Pelvic LN Met	EGAN00001375312	PELICAN
	EGAN00001250349	CRI (Cambridge)
	EGAN00001195709	CRI (Cambridge)
PB402-20229-NAD Pelvic LN MET	EGAN00001375306	PELICAN
	EGAN00001250243	ICR (Sutton)
	EGAN00001250241	ICR (Sutton)
	EGAN00001250327	ICR (Sutton)
	EGAN00001250293	CRI (Cambridge)

	EGAN00001250377	Shanghai
	EGAN00001250375	Shanghai
	EGAN00001250291	CRI (Cambridge)
	EGAN00001195717	ICR (Sutton)
	EGAN00001250271	ICR (Sutton)
	EGAN00001250279	ICR (Sutton)
	EGAN00001250251	ICR (Sutton)
	EGAN00001250295	CRI (Cambridge)
	EGAN00001250357	CRI (Cambridge)
	EGAN00001195723	ICR (Sutton)
	EGAN00001250263	ICR (Sutton)
	EGAN00001250249	ICR (Sutton)
	EGAN00001250289	CRI (Cambridge)
	EGAN00001250361	CRI (Cambridge)
	EGAN00001195701	CRI (Cambridge)
	EGAN00001250299	CRI (Cambridge)
	EGAN00001195705	CRI (Cambridge)
	EGAN00001250337	CRI (Cambridge)
	EGAN00001250353	CRI (Cambridge)
	EGAN00001250281	ICR (Sutton)
	EGAN00001250321	CRI (Cambridge)
	EGAN00001250287	ICR (Sutton)
	EGAN00001250359	CRI (Cambridge)
	EGAN00001250273	ICR (Sutton)
	EGAN00001250383	Shanghai
	EGAN00001250319	CRI (Cambridge)
	EGAN00001250355	CRI (Cambridge)
	EGAN00001250235	ICR (Sutton)
	EGAN00001250331	ICR (Sutton)
	EGAN00001250237	ICR (Sutton)
PB344-20225-NAD Pelvic LN MET	EGAN00001375302	PELICAN
	EGAN00001250313	CRI (Cambridge)
	EGAN00001250265	ICR (Sutton)
	EGAN00001250341	CRI (Cambridge)
	EGAN00001195721	ICR (Sutton)
	EGAN00001250369	ICR (Sutton)
	EGAN00001250259	ICR (Sutton)
	EGAN00001250329	ICR (Sutton)
	EGAN00001250373	ICR (Sutton)
	EGAN00001250317	CRI (Cambridge)
	EGAN00001250371	ICR (Sutton)
	EGAN00001195715	ICR (Sutton)
	EGAN00001250345	CRI (Cambridge)
	EGAN00001250261	ICR (Sutton)
	EGAN00001195727	ICR (Sutton)
	EGAN00001250315	CRI (Cambridge)
	EGAN00001250339	CRI (Cambridge)
	EGAN00001250333	ICR (Sutton)
	EGAN00001250267	ICR (Sutton)
	EGAN00001250285	ICR (Sutton)

EGAN00001250363	ICR (Sutton)
EGAN00001250301	CRI (Cambridge)
EGAN00001195697	CRI (Cambridge)
EGAN00001250323	CRI (Cambridge)
EGAN00001250255	ICR (Sutton)
EGAN00001250343	CRI (Cambridge)
EGAN00001195699	CRI (Cambridge)
EGAN00001250347	CRI (Cambridge)
EGAN00001250253	ICR (Sutton)
EGAN00001250365	ICR (Sutton)
EGAN00001195707	CRI (Cambridge)
EGAN00001250275	ICR (Sutton)
EGAN00001250325	ICR (Sutton)
EGAN00001250269	ICR (Sutton)
EGAN00001195713	CRI (Cambridge)

Sample type	Sample organ	Age at tumour collection
adt-met	left rib 5	70
adt-met	liver	66
adt-met	abdominal paraaortic lymph node	64
adt-met	right superficial inguinal lymph node	82
adt-met	right iliac lymph node	74
adt-met	right rib 8	72
adt-met	left humerus bone marrow	48
adt-met	right axillary lymph node	69
adt-met	mediastinal lymph node	68
adt-met	right subdural	67
primary	Prostate	88
non-adt-met	Pelvic Lymph Node	70
non-adt-met	Pelvic Lymph Node	64
primary	Prostate	83
non-adt-met	Pelvic Lymph Node	66
primary	Prostate	55
primary	Prostate	68
primary	Prostate	67
non-adt-met	Pelvic Lymph Node	69
primary	Prostate	59
primary	Prostate	61
non-adt-met	Pelvic Lymph Node	53
primary	Prostate	65
primary	Prostate	70
non-adt-met	Pelvic Lymph Node	56
primary	Prostate	64
primary	Prostate	74
primary	Prostate	72
primary	Prostate	59
primary	Prostate	66
non-adt-met	Pelvic Lymph Node	65
primary	Prostate	61
primary	Prostate	68
primary	Prostate	70
primary	Prostate	63
primary	Prostate	62
primary	Prostate	70
primary	Prostate	66
primary	Prostate	57
non-adt-met	Pelvic Lymph Node	65
primary	Prostate	52
primary	Prostate	66
non-adt-met	Pelvic Lymph Node	56
primary	Prostate	68
primary	Prostate	66
primary	Prostate	61
primary	Prostate	63

primary	Prostate	61
primary	Prostate	65
primary	Prostate	57
primary	Prostate	61
primary	Prostate	65
primary	Prostate	68
primary	Prostate	67
primary	Prostate	69
primary	Prostate	68
primary	Prostate	71
primary	Prostate	50
primary	Prostate	72
primary	Prostate	68
primary	Prostate	73
primary	Prostate	58
primary	Prostate	66
primary	Prostate	61
primary	Prostate	53
primary	Prostate	66
primary	Prostate	59
primary	Prostate	55
primary	Prostate	56
primary	Prostate	61
primary	Prostate	59
primary	Prostate	74
primary	Prostate	59
primary	Prostate	63
primary	Prostate	61
primary	Prostate	57
primary	Prostate	71
non-adt-met	Pelvic Lymph Node	60
primary	Prostate	70
primary	Prostate	64
primary	Prostate	62
primary	Prostate	73
primary	Prostate	50
primary	Prostate	56
primary	Prostate	63
primary	Prostate	61
primary	Prostate	54
primary	Prostate	62
primary	Prostate	54
primary	Prostate	62
primary	Prostate	65
primary	Prostate	63
primary	Prostate	60
primary	Prostate	55
primary	Prostate	66
primary	Prostate	68
primary	Prostate	58

primary	Prostate	56
primary	Prostate	60
primary	Prostate	66
primary	Prostate	51
primary	Prostate	68
primary	Prostate	54
primary	Prostate	64
primary	Prostate	57
primary	Prostate	65
primary	Prostate	66
primary	Prostate	49
primary	Prostate	57
primary	Prostate	65
primary	Prostate	60
primary	Prostate	44

PSA at tumour collection	Gleason	Pathological t stage	Biochemical recurrence
99	5+4	T4	
40	5+4	T4	
8.5	4+3	T2	1
10.6	4+3	T3	0
6.7	4+5	T3	0
10.9	3+4	T3	1
16.8	3+4	T3	1
16.2	4+3	T3	1
107	5+4	T3	
5.8	3+4	T3	1
51.1	3+4	T2	0
4.4	3+4	T2	0
12	3+4	T3	0
8.6	3+4	T3	1
12.8	3+4	T3	0
5	3+4	T3	1
35.7	3+4	T2	0
15.1	3+4	T3	0
6.87	3+4	T3	1
7.1	4+3	T2	0
8.38	4+3	T3	0
4.5	4+3	T3	1
7	4+3	T2	0
7	3+5	T3	1
18.3	3+3	T3	1
12.1	4+3	T2	1
11	3+4	T2	1
8.5	3+4	T3	1

9.27	4+3	T2	0
7.38	3+4	T2	0
6	3+3	T3	0
7.5	3+4	T3	0
7.2	4+3	T2	1
10.4	3+4	T3	0
8.2	3+4	T2	0
15.8	3+3	T3	0
7.07	4+3	T3	0
11	3+4	T3	1
8.9	3+4	T2	0
8.65	4+3	T2	0
4.8	4+3	T3	1
5.83	3+4	T2	0
5.75	3+4	T3	0
9.7	3+4	T3	0
5.4	3+4	T3	1
13	4+3	T2	0
5.63	3+3	T3	0
17	4+3	T3	0
6.1	3+4	T2	0
9.1	3+4	T2	0
7.63	3+4	T3	0
12.4	3+4	T2	0
97	4+3	T2	0
7.3	3+4	T3	0
9.61	3+4	T3	0
7.8	3+4	T2	1
11.4	3+4	T2	0
10.7	3+4	T3	0
20.74	3+5	T2	0
3.8	3+4	T3	0
11.13	3+4	T3	0
13.6	4+3	T2	0
10.31	3+4	T3	0
7.25	4+3	T3	1
8.6	3+4	T2	0
6.9	3+4	T2	0
6.07	3+4	T3	0
4	3+4	T3	0
42.83	4+3	T3	1
8.8	4+3	T3	1
11.9	3+3	T2	0
7.8	3+4	T2	0
6.3	3+4	T3	1
7	3+4	T3	0
12.1	4+3	T3	1
5.43	4+3	T3	0
10	4+3	T3	1

5.2	3+4	T2	0
11.9	3+4	T3	0
10.3	3+4	T3	0
8.3	3+4	T2	0
3.2	3+4	T2	0
7.1	3+4	T3	0
7.57	3+4	T3	0
5.4	3+4	T3	0
8	3+4	T2	1
28	3+4	T3	0
9	3+4	T3	0
5.9	3+4	T2	0
7.8	3+4	T3	0
7.7	3+4	T2	0
4.3	3+3	T3	0

Follow up	Sequence coverage	Cellularity	Ploidy	ETS positive	Battenberg status
	31.168	0.8879469	2.6129	no	yes
	55.6349	0.794	2.1832	no	yes
	56.2832	0.8976817	2.1898	yes	yes
	60.873	0.327	4.9185	yes	yes
	72.0131	0.93	2.2937	no	yes
	55.7188	0.824	3.8892	yes	yes
	63.8777	0.8	2.804	yes	yes
	70.0629	0.753	2.1154	yes	yes
	53.8279	0.772	2.3603	yes	yes
	54.3952	0.896	3.9017	yes	yes
	68.2543	0.1327743	4.4868	no	yes
	69.3941	0.6562474	2.482	no	yes
	57.8208	0.6227155	3.8177	no	yes
	62.7755	0.7075184	4.1396	yes	yes
	63.7608	0.5190688	1.99	yes	yes
231	57.5147	0.6299532	2.1753	no	yes
1246	49.4257	0.81098	2.332	no	yes
1596	81.5937	0.7315538	2.0555	yes	yes
	50.8255	0.7068444	3.8775	yes	yes
1240	103.611	0.2282811	3.3818	yes	yes
42	56.2595	0.7146315	2.0582	yes	yes
	68.2616	0.7575697	3.8226	yes	yes
461	54.8458	0.8876603	2.6087	no	yes
	56.0296	0.7863649	4.2441	no	yes
	56.8823	0.3518549	3.7638	yes	yes
982	58.4852	0.6130416	2.4175	yes	yes
1503	46.5969	0.3269423	3.9782	yes	yes
1568	75.8161	0.6947852	2.3655	no	yes
861	54.8251	0.4886241	2.1483	no	yes
139	48.833	0.2928916	2.4505	no	yes
	88.0622	0.7522268	1.9434	no	yes
1092	47.9624	0.5435194	2.1736	no	yes
1108	66.0939	0.6430273	1.9862	no	yes
3102	48.1899	0.5504923	2.1605	no	yes
1281	71.0824	0.54395	2.0972	no	yes
639	63.7698	0.6309904	2.2266	yes	yes
1282	55.8186	0.5664677	1.9983	yes	yes
1042	63.1209	0.3536998	2.1131	no	yes
98	57.2926	0.3979055	2.3503	no	yes
	61.7085	0.6257645	3.966	yes	yes
1567	67.4508	0.8243116	1.922	no	yes
53	55.3672	0.6742534	2.1377	no	yes
	65.0003	0.4602516	4.2688	yes	yes
1988	61.7287	0.4855316	2.1614	no	yes
1192	65.272	0.5514199	2.142	no	yes
725	60.2786	0.7430892	2.1295	no	yes
1732	56.7202	0.3411305	4.2647	yes	yes

1703	49.5837	0.8253155	2.249	no	yes
1485	46.2282	0.3846632	5.0003	yes	yes
2410	60.3372	0.5061917	2.0067	yes	yes
1103	98.7059	0.5693289	2.0877	yes	yes
455	49.0305	0.6237579	2.5298	no	yes
1435	76.8805	0.5132007	3.8533	no	yes
1605	78.7609	0.8038008	2.0895	yes	yes
1694	57.1343	0.4592892	2.1574	no	yes
962	66.1064	0.57163	2.1208	no	yes
610	60.1169	0.2768668	2.2644	no	yes
1336	65.1874	0.5866868	2.1763	yes	yes
1478	70.0784	0.5939939	2.0956	yes	yes
2395	69.3372	0.5405621	2.2087	no	yes
949	60.1544	0.2859	2.3336	no	yes
1029	51.2779	0.6461844	2.0507	yes	yes
1370	60.462	0.6957791	2.1585	no	yes
931	48.978	0.5814945	2.1414	yes	yes
1154	55.7188	0.5405533	2.1403	yes	yes
1009	61.0135	0.6229427	2.0737	yes	yes
1373	69.5849	0.6335656	2.0119	yes	yes
1351	82.082	0.6003663	2.0062	yes	yes
910	69.1117	0.5525659	2.1295	yes	yes
1000	61.5416	0.532787	2.0207	yes	yes
2226	56.837	0.246433	2.788	yes	yes
5	49.384	0.5004422	2.2818	no	yes
1281	63.9073	0.52523	2.1076	no	yes
821	63.5043	0.6283428	2.0392	yes	yes
711	59.0325	0.8284161	2.0282	yes	yes
692	57.1309	0.7711552	2.1273	no	yes
409	60.2388	0.5998768	2.0839	no	yes
	70.1226	0.6600904	2.0513	yes	yes
1207	57.9645	0.4302704	2.4245	no	yes
2048	53.9335	0.5288179	2.0402	yes	yes
1063	65.5659	0.2962172	2.5969	no	yes
857	56.2484	0.729429	2.0892	no	yes
722	54.2026	0.5492461	2.0896	yes	yes
955	56.6229	0.4112793	2.0381	yes	yes
1045	60.2229	0.6117868	2.0872	yes	yes
2198	27.2141	0.48167	2.2952	no	yes
1317	57.8395	0.1895289	2.615	yes	yes
540	55.5093	0.5191865	1.9407	yes	yes
68	51.7051	0.5805523	2.0717	yes	yes
614	59.7155	0.6382734	2.1451	no	yes
3327	78.7577	0.100931	2.5521	yes	yes
1694	103.225	0.4181382	2.1426	yes	yes
642	50.8752	0.4556534	2.3492	no	yes
1108	63.5958	0.6915431	2.0345	yes	yes
1366	50.5219	0.2197996	2.5013	yes	yes
1535	74.9148	0.2442819	2.125	yes	yes
51	72.9105	0.6091219	3.3681	yes	yes

1084	53.1289	0.4181697	2.3438	no	yes
1348	58.0269	0.6306792	2.4705	no	yes
1134	50.9903	0.7880881	2.1128	no	yes
465	57.265	0.378287	2.1498	no	yes
1402	76.7009	0.2285832	3.3951	no	yes
1086	69.9879	0.5311579	2.1287	yes	yes
1095	57.1239	0.3999533	2.1649	yes	yes
1122	59.1506	0.3920645	2.148	yes	yes
456	50.1649	0.3646071	2.3496	no	yes
827	56.6248	0.6186164	2.1761	yes	yes
1358	61.8726	0.4433831	2.0948	yes	yes
1731	58.3241	0.1927425	4.7141	no	yes
1157	60.5998	0.1530631	2.4647	yes	yes
1654	57.0562	0.902111	3.4678	no	yes
1236	58.9659			no	No

no.subs	no.indels	Complex	Del	Ins	no.svs	Inversion	Translocation	Deletion
17448	867	23	381	463	443	151	141	92
16107	2606	152	###	701	345	60	155	83
11944	1734	109	###	560	729	39	51	144
8720	1050	16	690	344	193	59	48	58
7008	1256	44	450	762	388	48	36	44
6106	1322	25	806	491	172	48	58	51
5950	1159	18	393	748	155	42	26	74
4604	1714	21	###	598	297	72	63	79
3959	762	16	227	519	86	28	24	24
3495	641	10	246	385	145	42	42	34
9778	1145	8	599	538	200	58	68	21
8157	1167	45	814	308	181	46	57	61
7137	995	47	680	268	179	48	70	50
6955	796	6	397	393	121	43	31	18
6319	944	16	440	488	221	43	29	74
6195	655	37	466	152	43	9	19	12
4892	672	39	475	158	44	9	19	11
4519	460	13	258	189	75	18	24	23
4242	512	3	306	203	69	22	23	17
4181	589	16	273	300	83	24	14	38
4119	464	9	240	215	52	13	23	11
4033	546	10	295	241	196	53	23	101
3956	531	10	284	237	300	105	110	48
3883	717	12	457	248	193	50	89	40
3725	505	3	200	302	89	26	16	42
3709	534	5	261	268	141	34	37	54
3599	242	2	125	115	77	14	23	28
3589	554	9	346	199	164	66	63	7
3548	296	7	181	108	10	2	3	3
3523	283	4	192	87	47	15	21	9
3481	673	12	403	258	69	15	37	8
3426	448	9	244	195	35	18	8	5
3410	593	11	293	289	394	148	142	58
3394	520	6	263	251	58	18	12	21
3393	497	7	261	229	36	8	5	20
3377	535	9	258	268	79	23	24	21
3369	641	7	449	185	30	11	5	11
3352	344	9	168	167	41	10	11	15
3315	426	10	226	190	127	44	45	25
3295	510	4	219	287	86	20	28	30
3236	425	4	230	191	38	16	12	3
3212	594	7	277	310	192	65	76	24
3145	607	8	241	358	149	38	27	77
3141	430	11	275	144	81	30	35	10
3132	593	6	274	313	112	38	54	4
3132	309	5	178	126	33	12	8	12
3122	482	9	250	223	47	14	17	13

3120	32	0	24	8	132	43	59	16
3054	350	5	192	153	62	14	13	29
3042	342	10	155	177	44	10	16	11
3021	536	19	288	229	56	12	19	23
3015	401	7	249	145	112	38	42	16
3013	511	11	285	215	55	17	15	19
2980	651	13	333	305	26	6	9	5
2945	420	4	257	159	34	8	10	13
2917	529	6	323	200	100	40	40	10
2796	266	9	139	118	118	48	46	8
2787	378	5	151	222	56	16	16	15
2785	378	7	217	154	44	7	14	18
2779	487	13	280	194	150	61	59	4
2767	258	7	145	106	11	0	4	6
2734	363	7	180	176	53	20	18	11
2730	434	5	250	179	33	13	13	4
2720	370	4	182	184	149	42	40	55
2709	266	5	138	123	35	6	13	15
2703	298	8	177	113	32	10	8	12
2699	495	14	254	227	73	32	23	9
2670	367	6	175	186	12	4	5	1
2666	464	7	242	215	11	3	3	4
2657	382	6	183	193	33	7	15	8
2654	214	6	131	77	15	2	7	5
2607	307	7	180	120	92	25	33	19
2587	328	8	199	121	91	32	15	35
2580	347	5	174	168	51	22	11	9
2555	485	8	231	246	85	22	22	35
2542	410	6	224	180	291	92	85	71
2516	332	10	192	130	36	14	14	1
2515	417	7	205	205	123	29	28	58
2499	313	3	181	129	84	23	33	15
2487	365	2	178	185	84	27	10	44
2450	239	7	152	80	99	33	34	16
2440	294	3	179	112	42	14	16	5
2418	263	2	111	150	35	8	16	9
2321	239	4	138	97	26	6	12	4
2309	260	9	145	106	43	13	14	11
2300	190	0	107	83	72	22	43	2
2255	226	3	117	106	25	6	7	7
2250	199	2	111	86	103	19	18	59
2239	326	11	200	115	105	24	31	38
2223	355	10	209	136	146	55	56	15
2209	369	9	194	166	17	5	5	5
2159	472	5	250	217	27	10	11	3
2127	224	5	132	87	66	12	19	27
2084	329	5	159	165	42	11	10	17
2082	288	8	146	134	54	6	14	28
2072	362	5	198	159	35	10	10	14
2038	535	8	241	286	74	13	21	34

2010	213	4	133	76	28	13	4	8
1898	340	8	212	120	50	13	20	15
1884	233	4	126	103	80	33	28	14
1839	277	11	158	108	69	22	28	10
1795	252	5	150	97	11	2	0	8
1764	265	10	140	115	43	10	18	12
1752	188	5	78	105	26	7	8	9
1713	214	7	112	95	21	5	9	6
1523	185	1	136	48	112	39	35	20
1502	178	7	93	78	8	2	3	2
1491	198	3	95	100	14	2	5	7
1161	110	8	57	45	8	3	1	2
928	106	3	63	40	5	1	1	3
923	114	4	65	45	6	3	1	0
411	68	10	30	28	1	0	0	1

Tandem.duplication
59
47
495
28
260
15
13
83
10
27
53
17
11
29
75
3
5
10
7
7
5
19
37
14
5
16
12
28
2
2
9
4
46
7
3
11
3
5
13
8
7
27
7
6
16
1
3

14
6
7
2
16
4
6
3
10
16
9
5
26
1
4
3
12
1
2
9
2
1
3
1
15
9
9
6
43
7
8
13
3
16
7
2
4
5
5
5
7
12
20
2
3
8
4
6
1
6

3
2
5
9
1
3
2
1
18
1
0
2
0
2
0

SUPPLEMENTARY TABLE 8

Sanger ID	Battenberg status	Included_in_subclonal_analysis
PD11328a	yes	yes
PD11329a	yes	yes
PD11330a	yes	yes
PD11331a	yes	yes
PD11332a	yes	yes
PD11333a	yes	yes
PD11334a	yes	yes
PD11335a	yes	yes
PD12337a	yes	yes
PD13412a	yes	yes
PD14721a	yes	yes
PD14722a	yes	no
PD14724a	yes	yes
PD14725a	yes	yes
PD14726a	yes	yes
PD14727a	yes	yes
PD14728a	yes	yes
PD14729a	yes	yes
PD14730a	yes	yes
PD14731a	yes	yes
PD12806a	yes	yes
PD12807a	yes	yes
PD12808a	yes	yes
PD12809a	yes	yes
PD12810a	yes	yes
PD12811a	yes	yes
PD12812a	yes	yes
PD12813a	yes	yes
PD12814a	yes	yes
PD12815a	yes	yes
PD12816a	yes	yes
PD12817a	yes	yes
PD12818a	yes	yes
PD12819a	yes	no
PD12820a	yes	yes
PD12821a	yes	yes
PD12822a	yes	no
PD12823a	yes	no
PD12824a	yes	yes
PD12825a	yes	yes
PD12826a	yes	no
PD12828a	yes	yes
PD12829a	yes	yes
PD12830a	yes	yes
PD12831a	yes	yes
PD12832a	yes	yes
PD12834a	yes	yes

PD12835a	yes	yes
PD12836a	yes	no
PD12837a	yes	yes
PD12838a	yes	yes
PD12840a	yes	yes
PD12841a	yes	yes
PD12842a	yes	yes
PD12843a	yes	no
PD12844a	yes	yes
PD12845a	yes	yes
PD13380a	yes	yes
PD13381a	yes	yes
PD13382a	yes	yes
PD13383a	yes	no
PD13384a	yes	yes
PD13386a	yes	yes
PD13387a	yes	yes
PD13388a	yes	yes
PD13389a	yes	yes
PD13390a	yes	yes
PD13391a	yes	no
PD13392a	yes	no
PD13393a	yes	yes
PD13394a	yes	yes
PD13395a	yes	yes
PD13396a	yes	yes
PD13397a	no	no
PD13398a	yes	yes
PD13399a	yes	yes
PD13400a	yes	no
PD13401a	yes	yes
PD13402a	yes	yes
PD13404a	yes	no
PD13405a	yes	yes
PD13406a	yes	yes
PD13407a	yes	no
PD13408a	yes	yes
PD13409a	yes	no
PD13410a	yes	yes
PD13411a	yes	no
PD14706a	yes	yes
PD14707a	yes	yes
PD14708a	yes	yes
PD14709a	yes	no
PD14710a	yes	yes
PD14711a	yes	yes
PD14712a	yes	yes
PD14713a	yes	yes
PD14714a	yes	yes
PD14716a	yes	yes

PD14717a	yes	yes
PD14718a	yes	yes
PD14719a	yes	yes
PD14720a	yes	no
PD14732a	yes	no
PD14733a	yes	yes
PD14734a	yes	yes
PD14735a	yes	yes
PD14736a	yes	yes
PD6204a_Illumina	yes	yes
PD9168a	yes	yes
PD9169a	yes	yes
PD9746a	yes	yes
PD9749a	yes	yes
PD9751a	yes	yes