Supplementary Data

Supplementary Materials and Methods

1. Bioinformatics processing:

i. Whole genome sequencing

Paired-end sequencing reads (151bp) were aligned against the human reference genome GRCh37/hg19 using Burrow-Wheeler Aligner (bwa, v0.7.7)[1]. The aligned reads were sorted, merged and filtered using SAMtools (v0.1.19)[2]. Picard tools (v1.96) (http://picard.sourceforge.net) was used to sort, mark and remove the duplicates. Local realignment of reads around INDELs and base recalibration were performed using Genome Analysis Toolkit (GATK, v3.1.1)[3].

ii. Copy-number analysis

Copy number variations were predicted using the Bioconductor package HMMCopy (v1.8.0)[4]. A 150-kb window was used. Reads in each window were normalised by GC-content and mappability. A Hidden Markov Model based algorithm was used to classify and annotate segments into different copy-number states: 0 (homozygous deletion), 1 (heterozygous deletion), 2 (neutral), 3 (gain), 4 (amplification) or 5 (high-level amplification). Tumor and normal samples were compared to germline (blood) DNA data for 5/6 cases, and tumor samples were compared to normal samples in 1 case (Case 3).

iii. Whole exome sequencing somatic mutation detection and INDEL detection

Somatic SNVs were detected by MuTect (v1.1.5)[5] using the high confidence filters. Raw Mutect calls can be found in WES data (Supplementary file 2.1). Mutect calls filtered (Supplementary file 2.2) by exonic location and functional annotation (synonymous, nonsynonymous, stopgain and stoploss mutations) can be found in Supplementary file 2 data. Strelka (v1.0.14)[6] was used to detect INDELs. ANNOVAR (version 2015Mar22)[7] was used to annotate the results. Germline mutations were detected by using the HaplotypeCaller module in GATK (v3.1.1) and Ensembl Variant Effect Predictor (version 79) was used to annotate them. Tumor and normal samples were compared to germline (blood) DNA data for 5/6 cases, and tumor samples were compared to normal samples in 1 case (Case 3).

iv. RNA Sequencing

Paired-end sequencing reads (101bp) were aligned to the human reference genome GRCh37/hg19 using MapSplice (v 2.1.6)[8] with fusion gene detection mode enabled. Gene annotation was taken from Ensembl build 74. The featureCounts function from the Bioconductor package Rsubread (v1.16.1)[9] was used to count reads mapped to each gene. Gene expression was reported as Reads Per Kilobase Mapped (RPKM) values. Given that sequencing libraries were prepped with three different methods (total RNA, polyA-stranded and polyA-unstranded), different settings were used to count reads. Cases 1 and 5 were a total RNA-seq run and had strand specific read counting, Cases 2, 3 and 6 had reversely stranded read counting, and Case 4 reads were counted in non-stranded mode

v. Methylation

Methylation analyses were performed in R (v3.3.0). Bioconductor package ChAMP [10] was used for raw file processing and QC. Default parameters set by ChAMP were used for normalizations and QC of the probes.. ChAMP uses minfi [11,12] for reading the raw files. Probes with detection p-value > 0.01 were removed [13].

2. Fluorescence in situ hybridization (FISH) analysis

TMPRSS2-ERG rearrangement was assessed by using a three-colour probe system (Zyto*Light* SPEC ERG/TMPRSS2 TriCheckTM Probe), following manufacturer's instruction. The analysis was performed on >100 Nuclei in the tumor regions surrounding each core, by using a Zeiss Axio Imager M1 microscope and Metasystems ISIS FISH capture software (version 5.5).

3.1 cfDNA and germline DNA isolation

Double spun plasma was obtained from blood collected in Cell-Free DNA BCT tubes (STRECK). cfDNA was isolated from 4 or 6 ml of plasma double spun with QIAsymphony Circulating DNA kit (Qiagen) by manufacturer protocols. Germline DNA was extracted with Qiagen DNA blood extraction kit collected in EDTA tubes.

3.2 Library preparation and sequencing:

Up to 25 ng of cfDNA and germline DNA was used to prepare libraries using Accel-NGS 2S DNA Library kit (Swift Biosciences, Ann Arbor, MI) with some modifications as previously described [14] Target capture was prepared with SureSelect Custom panel (Agilent Technologies). All libraries were quantified using the Library Quantification Kit (Kapa Biosystems, Wilmington, MA) and sequenced using Illumina MiSeq and NextSeq500.

Supplementary Tables and Figure Legends

Supplementary Table 1 – Case core additional details. 'Cellularity prior to macro-dissection (%)' indicates assessment of tumour cellularity on H&E stained frozen sections from individual cores, before marked tumour areas were macro-dissected.

Supplementary Table 2 – Case core fusion events.

Supplementary Figure 1 – Tissue collection and DNA/RNA extraction. Simplified cartoon showing (A) prostate sampling process and (B) individual core processing and DNA/RNA extraction method.

Supplementary Figure 2 – Correlation between percent genome alteration (PGA) and mpMRI visibility. Histogram showing the PGA values for mpMRI visible and invisible tumour cores. Data are represented as average±SD.

Supplementary Figure 3 – TMPRSS2-ERG fusion events for Case 1-6. (A) Schematic representation of the experimental design used to evaluate TMPRSS2-ERG rearrangements. (B) TMPRSS2-ERG rearrangements by 21q22.2 deletion is indicated by co-localization of the orange and blue probes and loss of the green probe. (C) Quantification of TMPRSS2-ERG rearrangements: no-rearrangement (N), TMPRSS2-ERG (TE) by 21q22.2 deletion, not available (n/a). The analysis has been performed on >100 Nuclei in the tumour regions surrounding each core.

Supplementary Figure 4 – circulating-free DNA (cfDNA) analysis. Genome-wide copy number analysis of patient-derived cfDNA.

Supplementary Figure 5 – Principal component analysis of global gene expression levels for Case 1-6. The distance between tumour (red) and benign (green) cores reflects differences in gene expression profile.

Supplementary Figure 6 – Multidimensional scaling plot analysis (MDS) of prostate specimens from Case 1-6. MDS plots analysis based on the 2-D projection of the Euclidean distances calculated between samples using the 5000 most variable CpG positions. Tumour cores are represented in red while benign cores are represented in green.

Supplementary References

[1] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754–60. doi:10.1093/bioinformatics/btp324.

[2] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078–9. doi:10.1093/bioinformatics/btp352.

[3] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297–303. doi:10.1101/gr.107524.110.

[4] Ha G, Roth A, Lai D, Bashashati A, Ding J, Goya R, et al. Integrative analysis of genome-wide loss of heterozygosity and monoallelic expression at nucleotide resolution reveals disrupted pathways in triple-negative breast cancer. Genome Res 2012;22:1995–2007. doi:10.1101/gr.137570.112.

[5] Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 2013;31:213–9. doi:10.1038/nbt.2514.

[6] Saunders CT, Wong WSW, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics 2012;28:1811–7. doi:10.1093/bioinformatics/bts271.

[7] Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164. doi:10.1093/nar/gkq603.

[8] Wang K, Singh D, Zeng Z, Coleman SJ, Huang Y, Savich GL, et al. MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Res 2010;38:e178. doi:10.1093/nar/gkq622.

[9] Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res 2013;41:e108. doi:10.1093/nar/gkt214.

[10] Morris, T. J., Butcher, L. M., Feber, A., Teschendorff, A. E., Chakravarthy, A. R., Wojdacz, T. K., and Beck, S. (2014). Champ: 450k chip analysis

methylation pipeline pg - 428-30. Bioinformatics, 30(3), 428-30.

[11] Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD and Irizarry RA (2014). Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA Methylation microarrays. Bioinformatics, 30(10), pp. 1363-1369. doi: 10.1093/bioinformatics/btu049.

[12] Jean-Philippe Fortin, Timothy Triche, Kasper Hansen. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array. bioRxiv 065490; doi: https://doi.org/10.1101/065490

[13] Zhou W, Laird PW and Shen H: Comprehensive characterization, annotation and innovative use of Infinium DNA Methylation BeadChip probes. Nucleic Acids Research 2016

[14] Gremel G, Lee RJ, Girotti MR, et al. Distinct subclonal tumour responses to therapy revealed by circulating cell-free DNA. *Annals of Oncology*. 2016;27(10):1959-1965. doi:10.1093/annonc/mdw278.