Comprehensive molecular classification of localized prostate adenocarcinoma reveals a tumour subtype predictive of non-aggressive disease

A Kamoun, G Cancel-Tassin, G Fromont, N Elarouci, L Armenoult, M Ayadi, J Irani, X Leroy, A Villers, G Fournier, et al.

Original Article

Comprehensive molecular classification of localized prostate adenocarcinoma reveals a tumour subtype predictive of a non-aggressive disease

A. Kamoun[1,*], G. Cancel-Tassin[2,3], G. Fromont[2,4], N. Elarouci[1], L. Armenoult[1], M. Ayadi[1], J. Irani[5], X. Leroy[6], A. Villers[7], G. Fournier[8], L. Doucet[9], S. Boyault[10], L. Brureau[11,12], L. Multigner[12], A. Diedhiou[13], M. Roupret[2,3,14], E. Compérat[2,3,15], P. Blanchet[11,12], A. de Reyniès[1,†], O. Cussenot[2,3,16,†]

†jointly supervised this study

*Dr Aurélie Kamoun, Tumour Identity Card Program (CIT), French League Against Cancer, 14 rue Corvisart, Paris, 75013, France, +33 (0) 1 53 55 25 58, aurelie.kamoun@ligue-cancer.net

1 Tumour Identity Card Program (CIT), French League Against Cancer, Paris, France
2 Research Center for Prostatic Pathologies (CeRePP), Paris, France.
3 Clinical Research Group n°5 ONCOTYPE-URO, Pierre and Marie Curie University, Paris, France
4 Department of Pathology, Tours University Hospital, Tours, France
5 Department of Urology, Miletrie Hospital, Poitiers, France
6 Department of Pathology, Lille University Research Hospital, Lille, France
7 Department of Urology, Lille University Research Hospital, Lille, France
8 Department of Urology, Cavale Blanche Hospital, Brest, France
9 Department of Pathology, Cavale Blanche Hospital, Brest, France
10 Translational Research and Cancers Genomic Innovation Department, Synergie Lyon Cancer Foundation, Lyon, France
11 Department of Urology, Pointe-à-Pitre/Abymes University Hospital, Pointe-à-Pitre, Guadeloupe (French West Indies)
12 Inserm U1085 – IRSET, Rennes, France
13 Department of Pathology, Pointe-à-Pitre/Abymes University Hospital, Pointe-à-Pitre, Guadeloupe (French West Indies)
14 Department of Urology, Pitié-Salpêtrière Hospital, Paris, France
15 Department of Pathology, Tenon Hospital, Paris, France
16 Department of Urology, Tenon Hospital, Paris, France

ABSTRACT

Background

Management of localized prostate cancer is a major clinical challenge since most of these cancers won't evolve but a majority of patients will still undergo a life-changing radical surgery. Molecular studies have shown that prostate cancer can be classified according to their genomic alterations but none of the published prostate cancer molecular classifications could identify a subtype corresponding to non-evolutive tumours.

Materials and methods

Multi-omics molecular profiling was performed on post-radical prostatectomy material from a cohort of 130 patients with localized prostate cancer. We used unsupervised classification techniques to build a comprehensive classification of prostate tumours based on three molecular levels: DNA copy number, DNA methylation, and mRNA expression. Merged data from our cohort and The Cancer Genome Atlas (TCGA) cohort were used to characterize the resulting tumour subtypes. We measured subtype-associated risks of biochemical relapse using Cox regression models and survival data from five cohorts including the two aforementioned.

Results

We describe three prostate cancer molecular subtypes associated with specific molecular characteristics and different clinical outcomes. Particularly, one subtype was strongly associated with the absence of biochemical recurrence. We validated this finding on 746 samples from five distinct cohorts ($P = 3.41 \times 10^{-8}$, $N = 746$ tumour samples), and showed that our subtyping approach outperformed the most popular prognostic molecular signatures to accurately identify a subset of patients with a non-evolutive disease. We provide a set of 36 transcriptomic biomarkers to robustly identify this subtype of non-evolutive cases whose prevalence was estimated to 22% of all localized prostate cancer tumours.

Conclusion

At least 20% of patients with localized prostate cancer can be accurately predicted to have a non-evolutive disease on the basis of their molecular subtype. Those patients should not undergo immediate surgery and rather be placed under active surveillance.

Keywords: prostate cancer, genomics, molecular classification, prognosis, subtypes, biomarkers

Key Message

Identifying prostate cancer patients with indolent tumours remains a challenge. We report a molecular classification which isolates a subgroup of patients with a non-evolutive disease. This subgroup accounts for 22% of patients in a meta-analysis of 746 localized prostate adenocarcinoma. For clinical transfer, we provide a list of 36 robust biomarkers to accurately identify these tumours.

INTRODUCTION

Prostate cancer (PCa) is the most frequent cancer diagnosed in men, accounting for 15% of all male cancer cases diagnosed. Regardless of this incidence, the associated death rate remains relatively low since about 10% of men diagnosed with prostate cancer will die of their disease. According to randomized studies on localized PCa management (PIVOT, PROTECT, SPCG-4), 20% of cancers will progress to metastatic state and death after more than ten years of follow-up. These studies suggest that a majority of cancers diagnosed at a localized stage are exposed to overtreatment.

In clinical practice, prognosis is principally based on D'Amico classification[1] or CAPRA score[2], combining histological criteria such as Gleason score[3], PSA (Prostate-Specific Antigen) levels, and TNM classification. However, these tools fail to accurately identify patients whose prostate cancer will not progress from those whose tumour will relapse and progress, even after local curative therapy such as prostatectomy.

The democratization of genomic profiling has considerably help to characterize PCa molecular heterogeneity and find relevant biomarkers. More importantly, molecular prognosis tools developed by academic or commercial institutions have been proved to increase the accuracy of risk predictions, demonstrating the utility of using information at the molecular level to adapt patient care[4]. The most recent molecular classification of prostate adenocarcinoma published by The Cancer Genome Atlas (TCGA) identified several subtypes of tumours, characterized by specific driver mutations or transcript fusions[5]. Yet, this characterization leaves 30% of PCa unclassified and no significant associations with relapse or prognosis was

established. Also, the lack of consensus among published molecular classifications and prognostic molecular signatures makes it difficult to use the molecular characteristics within clinical routine[6]. A further understanding into PCa molecular heterogeneity is therefore necessary to generate efficient tools and markers that could be used by every clinician.

Here we report a classification of PCa into three molecular subtypes with distinct genomic, transcriptomic, epigenomic and clinical features. This classification was revealed after considering sample purity, and reproduced on TCGA data for validation. Using 5 independent cohorts, we show that one molecular subtype is strongly predictive of BCR-free survival, and provide a list of robust surrogate biomarkers.

METHODS

Patients and samples included in CIT cohort

The CIT (Cartes d'Identité des Tumeurs®) retrospective cohort study included 130 patients with localized PCa from CHU Poitiers, CHU Pointe-à-Pitre/Abymes, GH Pitié-Salpétrière, CHU Brest, CHU Lille and Tenon hospital. All patients provided written informed consent consistent with local Research Ethics Board guidelines. The study protocol (CeRePP-PROGENE) was approved by the CPP Ile-de-France IV Institutional Review Board (IRB00003835).

195 samples were included in the CIT cohort: 130 tumours and 65 normal adjacent samples. An expert urological pathologist centrally reviewed all tumour specimens. Biochemical relapse was reported when patient PSA levels rose above 0.2 ng/mL followed by another increase after radical prostatectomy.

Sample details are given in Supplementary Table S1.

Procedures

CIT samples were profiled on mRNA expression arrays (E-MTAB-6128), DNA methylation arrays (E-MTAB-6131), and SNP arrays (E-MTAB-6126). Profiling protocols are described in Supplementary Methods. TCGA data were downloaded from Broad GDAC Firehose (doi:10.7908/C11G0KM9).

Samples tumour purity was estimated using an *in silico* method from DNA methylation arrays (Supplementary Methods). Molecular subtypes of prostate

adenocarcinoma tumours were identified by unsupervised classifications of mRNA and DNA methylation data using consensus clustering methods (Supplementary Methods). Subtypes were confirmed on TCGA data by using the same approach. We built an mRNA predictor of these subtypes and used a deconvolution method (Supplementary Methods) to label the samples in 3 additional public datasets, and measure subtypes association with prognosis.

Statistical analysis

Association between molecular subtypes and categorical variables were measured by Fisher's exact or Chi-square tests. Associations with continuous variables were evaluated by Kruskal-Wallis tests or ANOVA.

Relapse-free survival analyses were performed on patients from 5 independent cohorts considering BCR as the relapse event. We built multivariate Cox models integrating molecular subtyping and clinical risk factors, stratified on each cohort (separate baseline hazard functions were fit for each strata). We constructed Kaplan-Meier curves and used log-rank tests to compare patient groups. All statistical and bioinformatics analyses were performed with R software environment (version 3.3.2).

RESULTS

Combined analysis of mRNA expression and DNA methylation reveals 3 distinct PCa molecular subtypes

We analyzed 130 primary prostate adenocarcinoma samples and 65 adjacent normal prostate samples referred henceforth as the Carte d'Identité des Tumeurs® (CIT) cohort. These samples were all profiled on DNA methylation and SNP arrays, and 101 of them were also profiled on mRNA arrays as described in Supplementary Table S1. In order to avoid biases from normal cells contamination and define "pure" molecular subtypes of prostate tumours, we first restricted the analysis to 63 samples containing more than 50% of tumour cells as estimated through their DNA methylation profiles (Figure 1, Supplementary Figure S1). Consensus hierarchical clustering of mRNA expression and DNA methylation data were remarkably consistent ($P<10^{-15}$, Fisher) and were combined into three stable molecular subtypes S1, S2 and S3, almost perfectly matching mRNA clusters, and hereafter referred to as CIT subtypes.

We validated this classification system using the same approach on public data from TCGA Prostate Adenocarcinoma (PRAD) cohort. De novo unsupervised classification of mRNA and DNA methylation profiles was performed on 253 TCGA samples with more than 50% of tumour cells. The resulting classes were similarly combined into three stable subtypes with consistent mRNA and DNA methylation patterns ($P<10^{-50}$, Fisher), almost perfectly matching mRNA clusters, and were strongly correlated with initial CIT subtypes (Pearson's correlation from 0.5 to 0.7; Supplementary Figure S2), thereby validating the reliability of CIT subtypes.
We compared this three-class CIT system with the classification results published by TCGA in 2015[5] and observed a high consistency ($P<10^{-38}$, Fisher) with their original subtypes and consensus classes (Supplementary Figure S2). Nevertheless, original TCGA mRNA classes were significantly associated with tumour sample purity ($P<10^{-14}$, Fisher), which probably biased the class identification and may explain the differences with CIT subtypes.

Comprehensive molecular characterization of S1, S2 and S3 subtypes.
We built a transcriptomic classifier (Supplementary Methods) to predict CIT subtypes labels in CIT and TCGA samples that were not included in the classification process described above. We then used pooled data from CIT and TCGA cohorts to further characterize CIT subtypes using clinical data, mRNA signatures, copy number and mutation data (Figure 2).
The analysis of mRNA and copy number data revealed a strong association of CIT subtypes with *ERG* fusion status: S1 and S2 subtypes strongly over-expressed the transcriptomic fusion signature (Fus+) as defined in Setlur *et al*[7] *(*$P<10^{-20}$, t-test), and the typical losses of *TMPRSS2* genomic locus were over-represented in both subtypes ($P<10^{-43}$, Fisher). In contrast, S3 tumours showed neither transcriptomic nor genomic marks of the fusion (Fus-).

While sharing the Fus+ molecular pattern, S1 and S2 subtypes showed distinct clinical and genomic properties. 83% of S1 samples fell into ISUP group 3 or higher and harboured numerous significant losses of genomic loci (Supplementary Figure S3, Supplementary Table S2). This subgroup was specifically associated with frequent deletions of *PTEN* (67%, $P<10^{-33}$). Mutations in *TP53* gene were frequently found in S1 tumours (22%) and significantly associated with this subtype ($P<10^{-4}$,

Fisher). Unlike S1 tumours, S2 tumours were enriched in low Gleason scores (33% of ISUP group 1) and harboured few genomic alterations. However, *RYBP* locus (3p13) was more frequently lost in S2 (45%) relatively to S1 tumours (27%) (*P*=0.002, Fisher), and *TMPRSS2* mutations were only found in S2 tumours (5%). Consistent with the differences in their genomic profiles, S1 tumours transcriptomic profile showed a clear inactivation of p53 and *PTEN* pathways relatively to S2 tumours, together with higher proliferation signals and a diminished androgen response ($P<10^{-13}$, t-test).

S3 tumours were all negative for the Fus- signature. Our results confirmed the previously reported associations of Fus- tumours with *SPOP* mutations (28%, $P<10^{-18}$, Fisher), *FOXA1* mutations (15%, $P<10^{-6}$, Fisher) and frequent losses of chromosome arms 2q, 5q, and 6q[8]. *CHD1* losses were found in 40% of S3 tumours ($P<10^{-5}$, Fisher). The loss of *ZNF292* on 6q was observed in 65% of S3 tumours ($P<10^{-13}$, Fisher), ranking as the most frequently deleted locus in those tumours. As for arm 2q, we identified a deletion peak encompassing *SPOPL* in 33% of S3 tumours ($P<10^{-8}$, Fisher). Finally, S3 tumours were also significantly associated with mutations of *KDM6A* (5.4%; *P*=0.004, Fisher) and *BRAF* (4.2%; *P*=0.017, Fisher), which are both related to epigenetic modifications in cancers.

S2 subtype is strongly associated with the absence of biochemical recurrence after radical prostatectomy in 4 distinct cohorts.

Strikingly, among patients with S2 tumour, none had had a biochemical recurrence (BCR) in CIT cohort and only one had recurred in TCGA cohort. In order to assess the significance of this association, we predicted CIT subtypes in 3 additional cohorts with available mRNA and clinical data (Taylor *et al*[9], Ross-Adams *et al*[10], Fraser *et al*[11]),totalling 841 patients with primary prostate adenocarcinoma (Figure 3; Supplementary Table S3). Subtype clinical features and associations with *ERG* fusion in each cohort were consistent with the features observed in the CIT and TCGA discovery cohorts. A oooled analysis of the 5 cohorts revealed a strong association of CIT subtypes with BCR-free survival ($P<10^{-10}$, log-rank), particularly for S2 subtype relatively to other subtypes ($P<10^{-7}$, log-rank). CIT subtypes were more predictive of BCR-free survival than any of the other unsupervised molecular subtypes identified by TCGA, Taylor *et al*, and Ross-Adams *et al* (Supplementary

Figure S4). Multivariate analysis including ISUP group, tumour stage, and PSA confirmed that S2 subtype was an independent prognostic factor of BCR-free survival (*P*=0.0019, likelihood; Figure 3B). Taken alone, S2 vs non-S2 subtyping approach achieved a positive predictive value of 95.73% for the absence of BCR, which was the best score among significant predictive factors analysed.

S2 subtyping is a promising predictive tool for suspected indolent PCa

We compared S2 subtype predictive power with other molecular approaches, including marker panels from Prolaris®, OncotypeDX®, and Decipher® assays. These tools are used by clinicians to better identify patients with aggressive tumours from patients who are not likely to progress. We used published data [4, 12, 13] to assign Prolaris-like, OncotypeDX-like and Decipher-like scores to the 841 samples and compare their predictive power with S2 subtyping. As expected, we observed that S2 subtype was significantly associated with lower scores in the 3 systems as compared to S1 and S3 subtypes ($P<10^{-13}$; Supplementary Figure S5). The scores were then discretized to define prognostic groups of low aggressive cases. Predictions based on Proveri Inc prognostic biomarkers[14], and the molecular signature used by Irshad *et al* to identify patients with a non-evolutive disease[15] were added to the comparison (Supplementary Methods). We finally demonstrated that S2 subtyping performed the best to identify PCa without BCR, with a smaller BCR-associated hazard ratio than any low risk group defined with the other approaches (Figure 3C).

S2 subtype can be accurately identified with a list of 37 biomarkers

We aimed to identify a set of surrogate biomarkers to accurately diagnose S2 tumours in clinical routine. ERG-positive tumours can be detected with either immunohistochemistry[16] or a simple urine based test[17], both methods achieving very high specificity. We therefore focused on providing clinicians with biomarkers to isolate S2 tumours from S1 tumours. Using the 5 cohorts and an iterative procedure to get a stable list of biomarkers (Supplementary Methods), we identified a robust set of 36 discriminant genes between S1 and S2 subtypes (Supplementary Table S4). We checked the discriminative power of this geneset (CIT36) by building simple linear classifiers and achieved 81% of specificity and 90% of sensitivity

(Supplementary Methods). As expected, CIT36 predictions were strongly associated with BCR among ERG-positive tumours ($P<10^{-5}$, Figure 3D).

We searched for previously published evidence associating each of these genes with prostate cancer prognosis and found PubMed publications of interest for 22 of them, thereby strengthening the relevance of this gene set as well as S2 subtype prognosis value. We therefore propose 37 biomarkers (ERG-positive marker + CIT36 signature) as a diagnostic tool for non-evolutive cases of PCa.

DISCUSSION

This study proposes a comprehensive molecular classification of prostate adenocarcinoma integrating several molecular levels. Using an in silico method to discard low purity tumour samples, we could define three molecular subtypes of prostate tumours S1, S2 and S3, showing distinct features at the genomic, transcriptomic, and epigenetic levels. We have shown that our classification is not only consistent with the most recent TCGA classification[5] but also contains a strong prognostic power that was lacking in their study. Here, we successfully combined molecular classification and prognosis implications on a large number of tumours including a new multi-omics cohort and four distinct major PCa cohorts.

Genomic *ERG* fusion is the molecular alteration most frequently observed in PCa (~50% of tumours) and is associated with clearly different genomic and molecular characteristics as shown in previously published studies[18, 19]. The present classification respects the separation between Fus- and Fus+ tumours and confirms reported association of Fus- tumours with *SPOP* mutations and genomic losses of 5q21, 6q14, and 2q22 regions[8]. Notably, frequent losses of *ZNF292* (6q14) and *SPOPL*(2q22) identify these genes as new candidate tumour suppressors (TS) in Fus- tumours. *ZNF292* has already been described as a TS gene in colorectal cancer[20] and *SPOPL* is an *SPOP* paralog sharing 85% sequence identity[21], suggesting a possible TS role for *SPOPL* as well.

More importantly, this study identifies two distinct molecular subtypes within Fus+ tumours (S1 and S2), at the transcriptomic and genomic level. Apart from *TMPRSS2-ERG* fusion and region 3p13 losses, S2 tumours show considerably less genomic alterations than S1 tumours, suggesting that patients with these tumours may present an earlier stage of the disease. Indeed, ETS fusion is a starter event for tumorigenesis but not sufficient for full neoplastic transformation[22]. S2 may then correspond to preneoplastic lesions. This also suggests that the 3p13 loss may be an early event in prostate carcinogenesis.

Our work shows that these molecular differences between Fus+ subtypes S1 and S2 have significant implications at the clinical level, highlighting the importance of using molecular tests in clinical diagnostic routine to adapt patients' care. Molecular subtyping may be questioned to deal with PCa intratumoural heterogeneity. But

although this heterogeneity was confirmed by our deconvolution approach (Supplementary Figure S6), our results suggest that patient prognosis still depends on the major subtype.

Overtreatment of patients with prostate tumours is unfortunately medically acknowledged. Medical studies agree that at least 20% of PCa are non-evolutive diseases and that patients could live with it without benefit of immediate radical treatment[23]. Our work suggests that patients with S2 tumours (here, 20% of patients) may correspond to those putative indolent cases who could reasonably be handled with active surveillance rather than undergoing radical surgery. We showed that S2 subtyping was a valuable tool to identify patients with an indolent disease, and outperformed the prognostic power of previously published classifications and molecular biomarkers. Consequently, we propose a list of 37 surrogate biomarkers (CIT36 + ERG fusion positive marker) for the development of a new molecular test. CIT36 gene set results from the analysis of five independent cohorts of prostate tumour samples whose mRNA expression was measured through five different technologies including RNA-seq and 4 microarray technologies. We therefore assume that this gene set is preserved from common overfitting issues that may partly account for the lack of successful clinical transfer of previous set of biomarkers. Moreover, we showed that many of our biomarkers have been reported to be associated with PCa prognosis, including 11 down-regulated genes in S2 that are found up-regulated in the bad prognosis subtype described in Tandefelt *et al*[24]. This set of markers identified through 5 distinct cohorts therefore constitutes a robust basis for developing a routine molecular assay and identifying S2 tumours which are not likely to evolve to a higher stage of the disease.

ACKNOWLEDGEMENTS

This work is part of the French national program Cartes d'Identité des Tumeurs ® (CIT) (http://cit.ligue-cancer.net). The results presented in this study are partly based upon data generated by The Cancer Genome Atlas (TCGA) pilot project established by the NCI and NHGRI. Information about TCGA and the investigators and institutions that constitute the TCGA research network can be found at http://cancergenome.nih.gov/. We are grateful for the patients who participated in this study. We thank Julie Godet, Cécile Gaffory, and Valérie Ondet for excellent technical assistance. We also thank Remy Nicolle for English proofreading and useful advice.

FUNDING

This work was supported by the French League Against Cancer. No grant number is applicable.

DISCLOSURE

The authors have declared no conflicts of interest.

REFERENCES

1. D'Amico AV, Whittington R, Malkowicz SB et al. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. JAMA 1998; 280(11):969–974.

2. Cooperberg MR, Pasta DJ, Elkin EP et al. The University of California, San Francisco Cancer of the Prostate Risk Assessment score: a straightforward and reliable preoperative predictor of disease recurrence after radical prostatectomy. J. Urol. 2005; 173(6):1938–1942.

3. Gleason DF, Mellinger GT. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. J. Urol. 1974; 111(1):58–64.

4. Cuzick J, Swanson GP, Fisher G 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. 2011; 12(3):245–255.

5. Cancer Genome Atlas Research Network. The Molecular Taxonomy of Primary Prostate Cancer. Cell 2015; 163(4):1011–1025.

6. Lalonde E, Ishkanian AS, Sykes J et al. Tumour genomic and microenvironmental heterogeneity for integrated prediction of 5-year biochemical recurrence of prostate cancer: a retrospective cohort study. Lancet Oncol. 2014; 15(13):1521–1532.

7. Setlur SR, Mertz KD, Hoshida Y et al. Estrogen-dependent signaling in a molecularly distinct subclass of aggressive prostate cancer. J. Natl. Cancer Inst. 2008; 100(11):815–825.

8. Barbieri CE, Baca SC, Lawrence MS et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nat. Genet. 2012; 44(6):685–689.

9. Taylor BS, Schultz N, Hieronymus H et al. Integrative genomic profiling of human prostate cancer. Cancer Cell 2010; 18(1):11–22.

10. Ross-Adams H, Lamb AD, Dunning MJ et al. Integration of copy number and transcriptomics provides risk stratification in prostate cancer: A discovery and validation cohort study. EBioMedicine 2015; 2(9):1133–1144.

11. Fraser M, Sabelnykova VY, Yamaguchi TN et al. Genomic hallmarks of localized, non-indolent prostate cancer. Nature 2017; 541(7637):359–364.

12. Knezevic D, Goddard AD, Natraj N et al. Analytical validation of the Oncotype DX prostate cancer assay - a clinical RT-PCR assay optimized for prostate needle biopsies. BMC Genomics 2013; 14:690.

13. Erho N, Crisan A, Vergara IA et al. Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy. PloS One 2013; 8(6):e66855.

14. Jia Z, Rahmatpanah FB, Chen X et al. Expression Changes in the Stroma of Prostate Cancer Predict Subsequent Relapse. PLOS ONE 2012; 7(8):e41371.

15. Irshad S, Bansal M, Castillo-Martin M et al. A molecular signature predictive of indolent prostate cancer. Sci. Transl. Med. 2013; 5(202):202ra122.

16. Chaux A, Albadine R, Toubaji A et al. Immunohistochemistry for ERG Expression as a Surrogate for TMPRSS2-ERG Fusion Detection in Prostatic Adenocarcinomas. Am. J. Surg. Pathol. 2011; 35(7):1014–1020.

17. Koo KM, Wee EJH, Mainwaring PN, Trau M. A simple, rapid, low-cost technique for naked-eye detection of urine-isolated TMPRSS2:ERG gene fusion RNA. Sci. Rep. 2016; 6:srep30722.

18. Demichelis F, Setlur SR, Beroukhim R et al. Distinct genomic aberrations associated with ERG rearranged prostate cancer. Genes. Chromosomes Cancer 2009; 48(4):366–380.

19. Geybels MS, Alumkal JJ, Luedeke M et al. Epigenomic profiling of prostate cancer identifies differentially methylated genes in TMPRSS2:ERG fusion-positive versus fusion-negative tumors. Clin. Epigenetics 2015. doi:10.1186/s13148-015-0161-6.

20. Takeda H, Wei Z, Koso H et al. Transposon mutagenesis identifies genes and evolutionary forces driving gastrointestinal tract tumor progression. Nat. Genet. 2015; 47(2):142–150.

21. Choo K-B, Chuang T-J, Lin W-Y et al. Evolutionary expansion of SPOP and associated TD/POZ gene family: impact of evolutionary route on gene expression pattern. Gene 2010; 460(1–2):39–47.

22. Tomlins SA, Laxman B, Varambally S et al. Role of the TMPRSS2-ERG gene fusion in prostate cancer. Neoplasia N. Y. N 2008; 10(2):177–188.

23. Hamdy FC, Donovan JL, Lane JA et al. 10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer. N. Engl. J. Med. 2016; 375(15):1415–1424.

24. Gasi Tandefelt D, Boormans JL, van der Korput HA et al. A 36-gene signature predicts clinical progression in a subgroup of ERG-positive prostate cancers. Eur. Urol. 2013; 64(6):941–950.

FIGURE LEGENDS

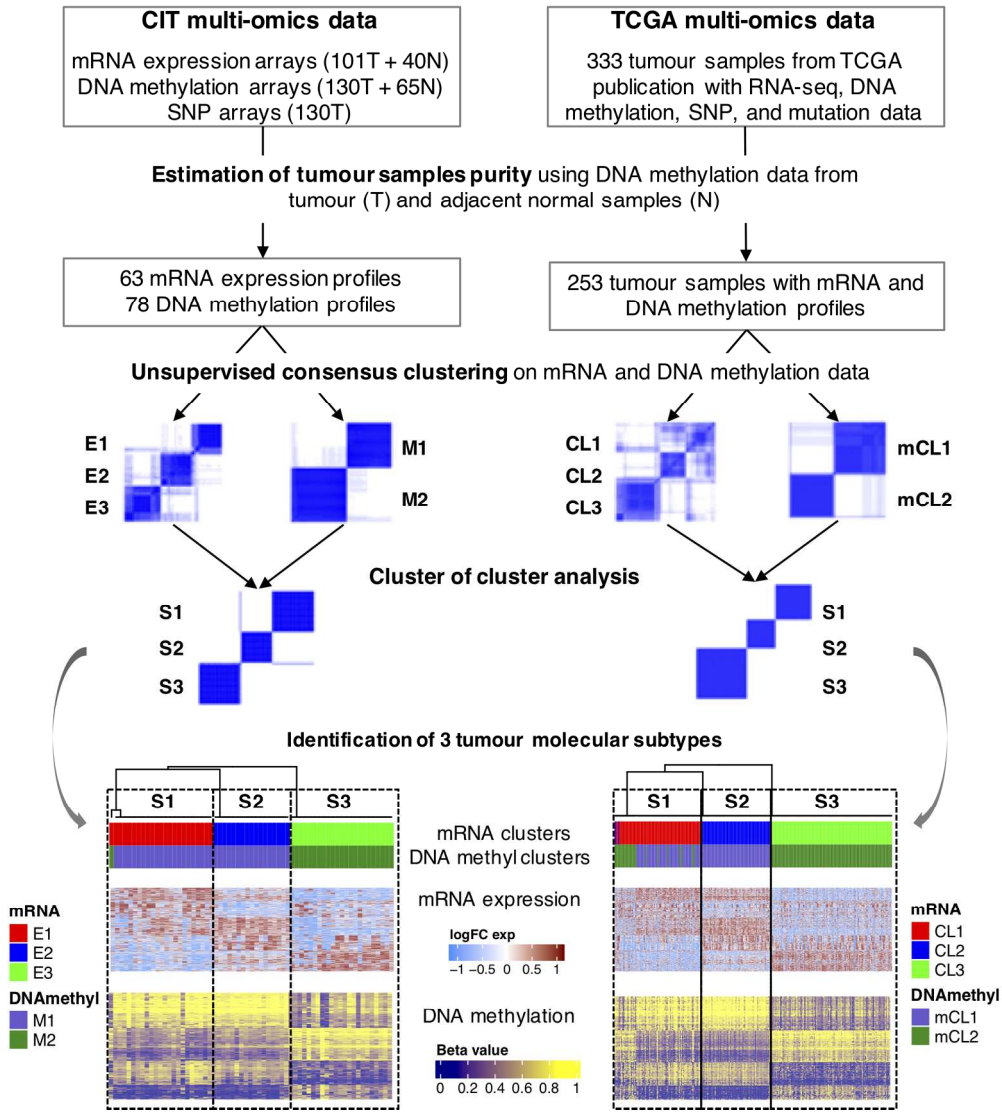
Figure 1: **Identification of three tumour molecular subtypes in CIT and TCGA cohorts**. We used CIT multi-omics data (130 samples of localized prostate adenocarcinoma and 65 samples of adjacent normal samples) to build an integrative molecular classification of prostate cancer. DNA methylation arrays from tumour and normal samples were used to assess the percentage of tumour cells in each sample. Only tumour samples containing more that 50% of tumour cells were selected to perform unsupervised consensus, both on mRNA and DNA methylation data. The resulting classes of each molecular layer were combined into integrative molecular subtypes using a cluster of cluster approach. The exact same process was similarly applied on TCGA published data from 333 patients. In both cohorts the analysis revealed 3 molecular subtypes almost perfectly matching the mRNA level clusters. The 3-class systems obtained in each cohort were highly similar (Supplementary Figure S1). Heatmaps displaying the most differentially expressed features at the mRNA expression and DNA methylation level illustrates the similarity between CIT and TCGA 3-subtype systems.

Figure 2: **Clinical and molecular characterization of tumour subtypes using CIT and TCGA cohorts**. 510 samples from CIT and TCGA cohorts were assigned one of S1, S2 or S3 subtypes. **A) Clinical repartition of ISUP groups within subtypes.** The pies show the proportion of ISUP groups within each subtype. **B) Molecular hallmarks of S1, S2 and S3 subtypes.** mRNA expression data was available for 493 tumour samples. We searched for differentially enriched gene sets between subtypes as described in Supplementary Methods. Single sample GSEA was performed to get an activity measure of each gene set for each sample, and resulting scores were centered and scaled for each cohort (CIT and TCGA). The mRNA signatures panel shows a selection of five gene sets which are differentially activated between the three subtypes. The activation levels in 53 adjacent normal samples (N) are displayed as a visual reference. The genomic alterations and exome data panels highlight genomic losses and mutation patterns of putative key driver genes whose alterations are significantly associated to the molecular classification. The copy number profiles of 504 tumour samples grouped by subtype was analysed by GISTIC2 to identify regions and genes that were significantly altered in each subtype

(Supplementary Table 2). We used TCGA exome data from 371 tumour samples to identify significant subtype specific mutations. Mutation profiles are displayed for the top 10 genes that were detected as significantly mutated by MutSig algorithm and whose mutation rates were significantly different between subtypes.
For each gene alteration (copy number loss or mutation) we used Fisher's exact tests to assess the association with subtypes. Stacked barplots on the right side shows the proportions of tumours in each subtype harbouring either a loss or a mutation of the genes.

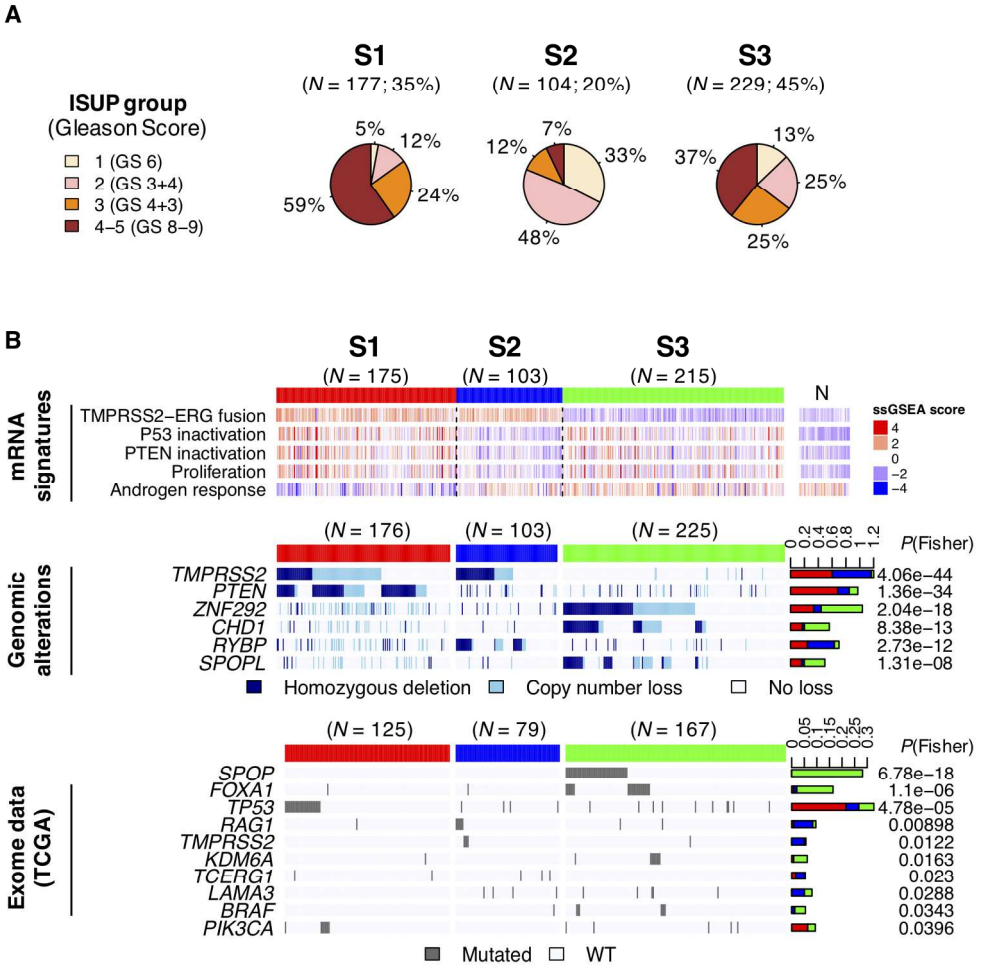
Figure 3: **Association of molecular subtypes with patient prognosis**. **A) Meta-analysis of BCR-free survival .** Five cohorts were used to study the association of CIT subtypes with prognosis. Clinical characteristics associated with CIT subtypes in each cohort are reported on the upper panel. *ERG* fusion status was assigned using transcriptomic signature from Setlur et al[7]. Kaplan-Meier plots show the evolution of patients' BCR-free survival, stratified by molecular subtypes. We used data from 746 patients with available clinical follow-up data, and generated Kaplan-Meier curves to estimate BCR-free survival. We used log-rank tests to evaluate the significance of differences between patient survival distributions when comparing S1, S2, and S3 subtypes (left panel) or S2 relatively to S1 and S3 subtypes (right panel). **B) Multivariate recurrence-free survival analysis using Cox regression model.** Regression analysis was performed on 729 samples with complete annotations for the included factors, and stratified on cohorts. For each predictive factor, we reported the hazard ratio (HR) associated with BCR, as well as the corresponding confidence intervals and likelihood test p-values. We also reported Fisher's exact test p-values and Positive Predictive Values (PPV) associated with each factor used to predict the absence of biochemical recurrence. **C) Prognostic power of S2 subtyping relatively to other molecular approaches.** We compared the prognostic significance of S2 subtype to low-risk groups defined by discretization of Prolaris-like, OncotypeDX-like, and Decipher-like scores, and to low-risk subtypes identified through the reproduction of Irshad *et al* or Jia *et al* (Proveri) computational methods (Supplementary Methods). Prolaris, OncotypeDx and Decipher low risk groups were defined by identifying cut-off values that best discriminated patients who had a biochemical recurrence from patients who did not. Resulting cut-offs used for discretization were 3, 31, and 1.7 for Prolaris-like, OncotypeDX-like and Decipher-like

scores. Hazard ratios refer to the relative risk of BCR in a multivariate Cox model including ISUP group (1-2 vs 3-4), tumour stage (T2 vs T3-T4), and PSA level (below or above 4ng/mL) from 729 tumour samples. **D) Association of CIT36 predictions with BCR-free survival in ERG fusion positive tumours.** The 36 genes listed in Supplementary Table S4 were used to build a predictor (CIT36) of S2 subtype among ERG-positive tumours as described in Supplementary Methods. The Kaplan-Meier plot shows patients' BCR-free survival evolution, stratified by the resulting predictions on the 5 cohorts.



Identification of three tumour molecular subtypes in CIT and TCGA cohorts. We used CIT multi-omics data (130 samples of localized prostate adenocarcinoma and 65 samples of adjacent normal samples) to build an integrative molecular classification of prostate cancer. DNA methylation arrays from tumour and normal samples were used to assess the percentage of tumour cells in each sample. Only tumour samples containing more that 50% of tumour cells were selected to perform unsupervised consensus, both on mRNA and DNA methylation data. The resulting classes of each molecular layer were combined into integrative molecular subtypes using a cluster of cluster approach. The exact same process was similarly applied on TCGA published data from 333 patients. In both cohorts the analysis revealed 3 molecular subtypes almost perfectly matching the mRNA level clusters. The 3-class systems obtained in each cohort were highly similar (Supplementary Figure S1). Heatmaps displaying the most differentially expressed features at the mRNA expression and DNA methylation level illustrates the similarity between CIT and TCGA 3-subtype systems.

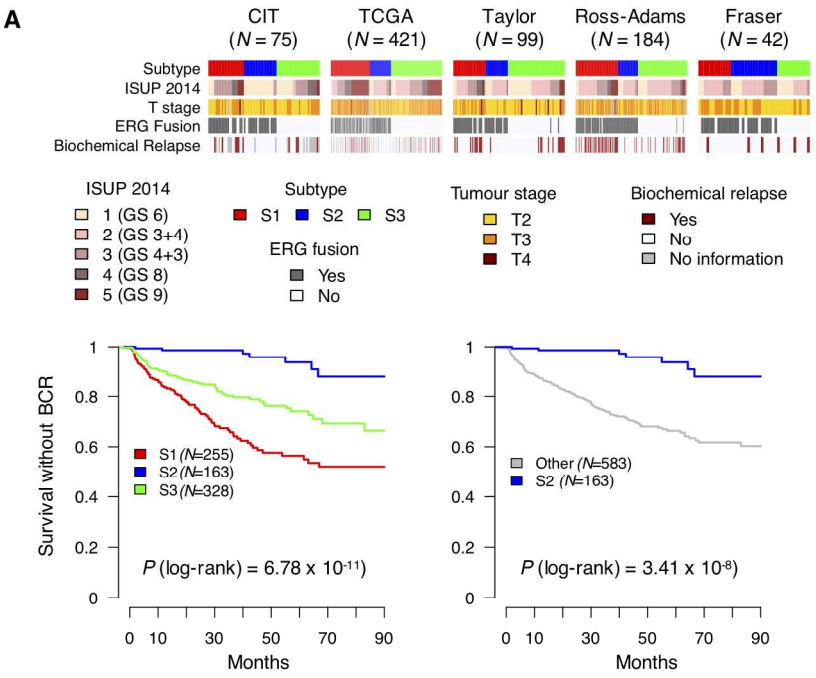
177x196mm (300 x 300 DPI)



Clinical and molecular characterization of tumour subtypes using CIT and TCGA cohorts. 510 samples from CIT and TCGA cohorts were assigned one of S1, S2 or S3 subtypes. A) Clinical repartition of ISUP groups within subtypes. The pies show the proportion of ISUP groups within each subtype. B) Molecular hallmarks of S1, S2 and S3 subtypes. mRNA expression data was available for 493 tumour samples. We searched for differentially enriched gene sets between subtypes as described in Supplementary Methods. Single sample GSEA was performed to get an activity measure of each gene set for each sample, and resulting scores were centered and scaled for each cohort (CIT and TCGA). The mRNA signatures panel shows a selection of five gene sets which are differentially activated between the three subtypes. The activation levels in 53 adjacent normal samples (N) are displayed as a visual reference. The genomic alterations and exome data panels highlight genomic losses and mutation patterns of putative key driver genes whose alterations are significantly associated to the molecular classification. The copy number profiles of 504 tumour samples grouped by subtype was analysed by GISTIC2 to identify regions and genes that were significantly altered in each subtype (Supplementary Table 2). We used TCGA exome data from 371 tumour samples to identify significant subtype specific mutations. Mutation profiles are displayed for the top 10 genes that were detected as significantly mutated by MutSig algorithm and whose mutation rates were significantly different between subtypes.

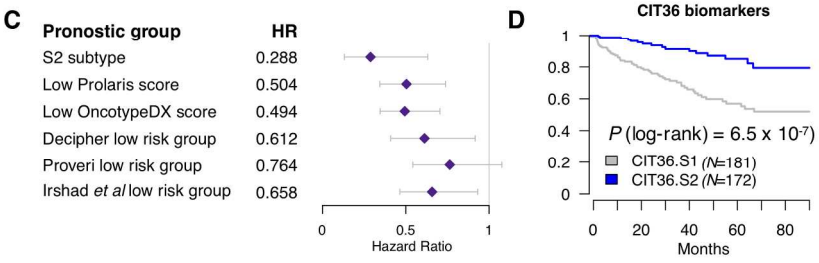
For each gene alteration (copy number loss or mutation) we used Fisher's exact tests to assess the association with subtypes. Stacked barplots on the right side shows the proportions of tumours in each subtype harbouring either a loss or a mutation of the genes.

175x172mm (300 x 300 DPI)



B Multivariate Cox model of BCR-free survival (*N* =729)

Predictive Factor of BCR-free survival	Hazard Ratio	Confidence Interval	*P* (Likelihood)	Association with BCR-free survival *P* (Fisher)	Positive Predictive Value (PPV)
Subtype : S2	0.29	0.13 - 0.63	**1,89E-03**	**2,42E-10**	**95,73%**
ISUP : 1-2	0.26	0.17 - 0.38	**2,78E-11**	**5,04E-11**	69,75%
T stage : T2	0.38	0.25 - 0.58	**4,39E-06**	**4,39E-10**	89,85%
PSA level < 4ng/ml	0.65	0.31 - 1.37	0,257	**7,52E-05**	74,12%



Association of molecular subtypes with patient prognosis. A) Meta-analysis of BCR-free survival . Five cohorts were used to study the association of CIT subtypes with prognosis. Clinical characteristics associated with CIT subtypes in each cohort are reported on the upper panel. ERG fusion status was assigned using transcriptomic signature from Setlur et al[7]. Kaplan-Meier plots show the evolution of patients' BCR-free survival, stratified by molecular subtypes. We used data from 746 patients with available clinical follow-up data, and generated Kaplan-Meier curves to estimate BCR-free survival. We used log-rank tests to evaluate the significance of differences between patient survival distributions when comparing S1, S2, and S3 subtypes (left panel) or S2 relatively to S1 and S3 subtypes (right panel). B) Multivariate recurrence-free survival analysis using Cox regression model. Regression analysis was performed on 729 samples with complete annotations for the included factors, and stratified on cohorts. For each predictive factor, we reported the hazard ratio (HR) associated with BCR, as well as the corresponding confidence intervals and likelihood test p-values. We also reported Fisher's exact test p-values and Positive Predictive Values (PPV) associated with each factor used to predict the absence of biochemical recurrence. C) Prognostic power of

S2 subtyping relatively to other molecular approaches. We compared the prognostic significance of S2 subtype to low-risk groups defined by discretization of Prolaris-like, OncotypeDX-like, and Decipher-like scores, and to low-risk subtypes identified through the reproduction of Irshad et al or Jia et al (Proveri) computational methods (Supplementary Methods). Prolaris, OncotypeDx and Decipher low risk groups were defined by identifying cut-off values that best discriminated patients who had a biochemical recurrence from patients who did not. Resulting cut-offs used for discretization were 3, 31, and 1.7 for Prolaris-like, OncotypeDX-like and Decipher-like scores. Hazard ratios refer to the relative risk of BCR in a multivariate Cox model including ISUP group (1-2 vs 3-4), tumour stage (T2 vs T3-T4), and PSA level (below or above 4ng/mL) from 729 tumour samples. D) Association of CIT36 predictions with BCR-free survival in ERG fusion positive tumours. The 36 genes listed in Supplementary Table S4 were used to build a predictor (CIT36) of S2 subtype among ERG-positive tumours as described in Supplementary Methods. The Kaplan-Meier plot shows patients' BCR-free survival evolution, stratified by the resulting predictions on the 5 cohorts.

170x241mm (300 x 300 DPI)