The genomic landscape of metastatic castration-resistant prostate cancers using whole genome sequencing reveals multiple distinct genotypes with potential clinical impact

Lisanne F. van Dessel[1,a], Job van Riet[2,3,a], Minke Smits[4], Yanyun Zhu[5,6], Paul Hamberg[7], Michiel S. van der Heijden[8,9,10], Andries M. Bergman[5,10], Inge M. van Oort[11], Ronald de Wit[1], Emile E. Voest[8, 10], Neeltje Steeghs[8, 10], John W.M. Martens[1,8], Stefan Sleijfer[1,8], Edwin Cuppen[12], Wilbert Zwart[5,6,13], Harmen J.G. van de Werken[2,3], Niven Mehra[4,b], Martijn P. Lolkema[1,8,b,*]

[1] Department of Medical Oncology, Erasmus MC Cancer Institute, University Medical Center Rotterdam, Dr. Molewaterplein 40, 3015 GD, Rotterdam, the Netherlands.

[2] Cancer Computational Biology Center, Erasmus MC Cancer Institute, University Medical Center Rotterdam, Dr. Molewaterplein 40, 3015 GD, Rotterdam, the Netherlands.

[3] Department of Urology, Erasmus MC Cancer Institute, University Medical Center Rotterdam, Dr. Molewaterplein 40, Rotterdam, the Netherlands.

[4] Department of Medical Oncology, Radboud University Nijmegen Medical Center, Geert Grooteplein Zuid 10 6525 GA Nijmegen, the Netherlands.

[5] Division on Oncogenomics, The Netherlands Cancer Institute, Plesmanlaan 121 1066 CX Amsterdam, the Netherlands

[6] Oncode Institute, the Netherlands.

[7] Department of Internal Medicine, Franciscus Gasthuis & Vlietland, Kleiweg 500 3045 PM Rotterdam, the Netherlands.

[8] Center for Personalized Cancer Treatment, Rotterdam, the Netherlands.

[9] Division of Molecular Carcinogenesis, The Netherlands Cancer Institute, Plesmanlaan 121 1066 CX Amsterdam, the Netherlands.

[10] Department of Medical Oncology, The Netherlands Cancer Institute, Plesmanlaan 121 1066 CX Amsterdam, the Netherlands.

[11] Department of Urology, Radboud University Nijmegen Medical Center, Geert Grooteplein Zuid 10 6525 GA Nijmegen, the Netherlands.

[12] Center for Molecular Medicine and Oncode Institute, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX, Utrecht, the Netherlands.

[13] Laboratory of Chemical Biology and Institute for Complex Molecular Systems, Department of Biomedical Engineering, Eindhoven University of Technology, PO Box 513, Eindhoven

[a,b] These authors contributed equally

[*] Corresponding author

Address for correspondence: Martijn P. Lolkema. Department of Medical Oncology, Erasmus MC Cancer Institute, Room NT-537, P.O. Box 2040, 3000 CA, Rotterdam, The Netherlands. Tel.: +31 10 704 19 06. E-mail: m.lolkema@erasmusmc.nl

Keywords: prostate cancer, metastasis, cancer genomics, homologous recombination deficiency

Abstract

Here we present whole-genome sequencing (WGS) analysis of fresh-frozen metastatic biopsies from 197 castration-resistant prostate cancer patients. Using hierarchical unsupervised clustering based on genomic aberrations only, we defined eight different clusters. We detected four distinct and potentially clinically relevant genotypes harboring unique genomic features, including: 1) Microsatellite Instability; 2) Homologous Recombination Deficiency (HRD) with enriched genomic deletions and *BRCA2* aberrations; 3) tandem duplication phenotype associated with biallelic *CDK12* mutations; and 4) a subgroup enriched for chromothripsis events. Our data suggest that classifying patients using WGS characteristics may improve classification of HRD patients. Moreover, we confirmed that important regulators of *AR*-mediated signaling are located in non-coding regions. Using ChIP sequencing data, we showed that the amplified *AR* and *MYC* promoter regions contain open chromatin and bind AR, suggesting a role in AR mediated biology. Thus, high-resolution WGS may be used to improve patient stratification.

47 **Main text**

48 Prostate cancer is known to be a notoriously heterogeneous disease and the genetic basis for
49 this interpatient heterogeneity is poorly understood. The ongoing development of new therapies for
50 metastatic prostate cancer that could potentially only be effective in molecularly defined subgroups,
51 further increases the need for accurate patient classification and stratification[1–3]. Comprehensive
52 genomic analyses in primary prostate cancer was able to classify 74% of analyzed patients into seven
53 predefined subtypes based on ETS fusions and mutations in *SPOP*, *FOXA1* and *IDH1*[4]. More
54 recently, whole-genome sequencing (WGS) of metastatic prostate cancer demonstrated that
55 structural variations (SVs) arose from specific alterations such as *CDK12*$^{-/-}$ and *BRCA2*$^{-/-}$ genotypes[5–
56 7]. As example, the predominance of tandem duplications was strongly associated with biallelic *CDK12*
57 mutations. Thus, WGS enables the identification of patterns of DNA aberrations (i.e. 'genomic scars')
58 that may profoundly improve classification of tumors that share a common etiology and that may be
59 targeted using different therapies.

60 We analyzed fresh-frozen metastatic tumor samples and matched blood samples from 197
61 castration-resistant prostate cancer (CRPC) patients using WGS generating to date the largest WGS
62 dataset for mCRPC and combine it with AR ChipSeq data (**Figure 1a).** Clinical details on biopsy site,
63 age and previous treatments of the included patients are described in **figure 1** and **supplementary**
64 **table 2**. An overview of the sequencing quality is provided in **supplementary figure 1**. The median
65 tumor mutational burden on coding regions (TMB) was 2.54/Mb in our mCRPC cohort; this is roughly
66 twice as high as compared to primary prostate cancer and furthermore, 14 patients had high TMB
67 (>10)), which in other tumor types has been associated with a high sensitivity to check-point inhibitors
68 (**supplementary figure 2**)[8]. We analyzed the somatic genomic aberrations and found a median of
69 6621 single-nucleotide variants (SNVs; IQR: 5048-9109), 1008 small insertions and deletions (InDels;
70 IQR: 739-1364), 55 multi-nucleotide variants (MNVs; IQR: 34-86) and 224 SVs (IQR: 149-370) per
71 patient (**supplementary figure 3a-c**). We observed a highly complex genomic landscape consisting
72 of multiple driver mutations and structural rearrangements in our cohort. We confirmed that known
73 key driver genes of prostate cancer were enriched for nonsynonymous mutations, including, *TP53,*
74 *AR, FOXA1, SPOP, RB1* and *PTEN* (**Figure 2** and **supplementary figure 3d-e**). Distinct amplified
75 genomic regions included 8q, 10q and Xq; deleted regions affected 8p, 10q, 13q and 17p
76 (**supplementary figure 3d**). In addition to large-scale chromosomal copy-number alterations, we

could also further pinpoint narrow genomic regions targeted with recurrent copy-number alterations which could potentially reveal important genes within or near the proximity of these events (**supplementary table 3**).

TMPRSS2-ERG gene fusions were the most common fusions in our cohort (*n* = 84 (91.3% of all ETS fusions); **Figure 2** and **supplementary figure 4**) and is comparable to localized prostate cancer[4,9]. The predominant deletion site was located upstream of the second exon of *ERG*, which preserves its ETS domain in the resulting fusion gene. In 42 patients (21.3%) we observed regional hypermutation ("kataegis") (**Figure 2** and **supplementary figure 5**); this seems to be comparable to kataegis rates in primary prostate cancer and thus is not an obvious driving force in metastatic progression[9].

Several studies have shown that metastatic disease significantly differs from localized prostate cancer and that disease progression towards CRPC is mainly driven by increased androgen receptor signaling[10,11]. In-depth analysis of the AR-pathway revealed that aberrant AR signaling occurred in up to 80% of our patients. In 57.3% of patients both *AR* and the *AR*-enhancer (located about 631 kB upstream of the *AR* gene[6]) were affected (**Figure 3a**). In an additional 6.6% and 14.7% of patients only *AR* alterations or *AR*-enhancer amplification occurred, respectively. Concurrent amplification of *AR* and *AR*-enhancer was not necessarily of equal magnitude, which resulted in differences in copy number enrichment of these loci (**Figure 3b**). ChIP-seq data of two mCRPC patients and prostate cancer cell-lines (LNCaP and VCaP) revealed active enhancer regions (H3K27ac), coupled with actively bound *AR and FOXA1,* at the detected amplification peaks, which was found to be enriched in CRPC settings (**Figure 3c**). This indicates that *AR*-enhancer amplification could be associated with increased AR-signaling for this genomic region, which is supported by previous studies demonstrating that this amplification ultimately resulted in significantly elevated expression of *AR* itself[6,7]. Furthermore, a recurrent focal amplification at a non-coding area was observed at 8q24.21 near *PCAT1*. This locus bears similar epigenetic characteristics to the *AR*-enhancer with regard to H3K27ac and, to a lesser extent, binding of AR and/or FOXA1 (**Figure 3c**). This locus could represent a somatically-acquired putative enhancer affecting *MYC* expression (**Figure 3d**), however functional follow-up studies should be performed to further this hypothesis[12]. In addition, *PCAT1* is a long non-coding RNA which is known to be upregulated in prostate cancer and negatively regulates *BRCA2* expression while positively affecting *MYC* expression[13,14]. These data

107 show that most prostate cancers reactivate the AR pathway either directly or indirectly when
108 progressing toward (m)CRPC.

109 Our comprehensive WGS data and sample size enabled us to perform unsupervised
110 clustering to identify genomic scars that can define subgroups of mCRPC patients. We clustered our
111 genomic data using total number of SVs, relative frequency of SV categories, TMB and tumor ploidy.
112 This analysis defined eight distinct subgroups (**Figure 4-5** and **supplementary figure 6-8**): A)
113 Microsatellite Instability (MSI) signature with high TMB and associated with mismatch repair
114 deficiency; B) Tandem duplications (>100 kb) phenotype associated with biallelic *CDK12* inactivation;
115 D) Homologous Recombination Deficiency (HRD) features with many (>100kb) deletions and
116 association with (somatic) mutations in BRCAness-associated genes; F) chromothripsis; C, E, G, H);
117 non-significant genomic signature without any currently known biological association. **Table 1**
118 summarizes the key features of each subgroup. Cluster A and B represent previously identified
119 genomic subgroups[5,7,10,15] and in both groups only a minority of the patients was allocated to these
120 subgroups without a specific mutation in the corresponding genes. Interestingly, 2 out of 13 patients in
121 cluster B (Tandem duplications (>100 kb) phenotype) did not show a bi-allelic (somatic) mutation in
122 *CDK12,* suggesting that tandem duplications may arise in patient without *CDK12* mutations. Cluster D
123 shows significant features of HRD, specifically biallelic *BRCA2* inactivation, mutational signature 3,
124 enrichments of deletions (<100 kb) and is supported by high HR-deficiency scores (CHORD)[16,17]
125 (**supplementary figure 6**). Although this is a known association[7,18], our clustering analysis potentially
126 refines patient classification, as 32% of this subgroup (7/22) does not have a defining biallelic *BRCA2*
127 (somatic) mutation of which four of these patients show at least one (deleterious) aberration in other
128 BRCAness-related genes[19]. In addition, 4 patients in other clusters show non-synonymous mutations
129 in *BRCA2* without corresponding genomic scars, mutational signature and/or HR-deficiency scores
130 (**figure 4**, **supplementary figure 9**). Only a single sample in cluster A harbored a *BRCA2* mutation
131 with known pathogenic effect (p.T3030fs; RCV000031792)[20] and was identified with a HR-deficiency
132 score which was clustered based on MSI-profile in our analysis. These patients might be deemed
133 false-negative or false-positive when using FDA-approved assays (BRCAnalysis™ and
134 FoundationFocus™) to predict response to poly(ADP-ribose) polymerase inhibitors (PARPi) or
135 alkylating drugs like platinum compounds based on the presence of *BRCA* mutations. In cluster F, we
136 detect significantly more chromothripsis events in comparison to the other clusters (80% vs 20%).

137 However, the overall frequency of chromothripsis (23.3%) was comparable with previous findings[7,9].
138 We failed to reproduce a previous finding suggesting chromothripsis to be associated with inversions
139 and p53 inactivation in prostate cancer[7]. Apart from the chromothripsis events, no clear gene
140 aberration was associated with this cluster. In the remaining patients, we could not identify a distinct
141 genomic signature or biologic rationale (cluster C, E, G, H). In cluster C, conjoint aberrations of
142 *BRCA1* and *TP53* were observed in one patient with a high HR-deficiency prediction score (CHORD),
143 which is known to lead to a small tandem duplication phenotype (<100 kb)[21]. Two other patients within
144 cluster C display a weak CHORD scoring associated with HR-deficiency, however no additional
145 evidence was found for a BRCA1 loss-of-function within these patients.

146 The classification of patients using WGS has the advantage of being, in theory, more precise
147 and less prone to bias compared to analyses using targeted panels consisting of a limited number of
148 genes. Overall, our study describes the complete genomic landscape of metastatic CRPC and
149 confirms the central role of *AR* signaling in this disease. We identify distinct CRPC subgroups based
150 on phenotypic characteristics encompassing genomic signatures, including MSI, BRCAness and
151 CDK12 inactivity, which may be clinically relevant.[5,19] Moreover, we show the added value of WGS-
152 based unbiased clustering in identifying additional patients with genomic scars who are eligible for
153 specific therapies and we could classify patients even if WGS (or our methodology) did not find
154 conclusive evidence for a bi-allelic mutation in the proposed gene-of-interest.

155 This study also showed that a large population of mCRPC patients do not fall into an as-of-yet
156 clinically-relevant or biologically-clear genotype and further research can help elucidate the oncogenic
157 driver events and provide new therapeutic options. In addition, further analysis using whole-genome
158 sequencing data allows us to gain more insight into the role of non-coding regions of the genome in
159 prostate cancer.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available in the supplementary information file.

References

1. Mullane, S. A. & Van Allen, E. M. Precision medicine for advanced prostate cancer. *Current Opinion in Urology* (2016). doi:10.1097/MOU.0000000000000278
2. Ciccarese, C. *et al.* Prostate cancer heterogeneity: Discovering novel molecular targets for therapy. *Cancer Treatment Reviews* (2017). doi:10.1016/j.ctrv.2017.02.001
3. Shtivelman, E., Beer, T. M. & Evans, C. P. Molecular pathways and targets in prostate cancer. *Oncotarget* (2014). doi:10.18632/oncotarget.2406
4. Cancer Genome Atlas Research Network *et al.* The Molecular Taxonomy of Primary Prostate Cancer. *Cell* **163,** 1011–25 (2015).
5. Wu, Y.-M. *et al.* Inactivation of CDK12 Delineates a Distinct Immunogenic Class of Advanced Prostate Cancer Article Inactivation of CDK12 Delineates a Distinct Immunogenic Class of Advanced Prostate Cancer. *Cell* **173,** 1770–1782.e14 (2018).
6. Viswanathan, S. R. *et al.* Structural Alterations Driving Castration-Resistant Prostate Cancer Revealed by Linked-Read Genome Sequencing. *Cell* **174,** 433–447.e19 (2018).
7. Quigley, D. A. *et al.* Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer. *Cell* **174,** 758–769.e9 (2018).
8. Campbell, P. J. *et al.* Pan-cancer analysis of whole genomes. *bioRxiv* 162784 (2017). doi:10.1101/162784
9. Fraser, M. *et al.* Genomic hallmarks of localized, non-indolent prostate cancer. *Nature* **541,** 359–364 (2017).
10. Robinson, D. *et al.* Integrative clinical genomics of advanced prostate cancer. *Cell* **161,** 1215–1228 (2015).
11. Grasso, C. S. *et al.* The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487,** 239–243 (2012).
12. Mazrooei, P. *et al.* Somatic Mutations and Risk-Variants Converge on Cis-Regulatory Elements to Reveal the Cancer Driver Transcription Regulators in Primary Prostate Tumors.

SSRN Electron. J. (2018). doi:10.2139/ssrn.3245213

13. Prensner, J. R. *et al.* The Long Non-Coding RNA PCAT-1 Promotes Prostate Cancer Cell Proliferation through cMyc. *Neoplasia* **16,** 900–908 (2014).
14. Prensner, J. R. *et al.* PCAT-1, a long noncoding RNA, regulates BRCA2 and controls homologous recombination in cancer. *Cancer Res.* **74,** 1651–1660 (2014).
15. Pritchard, C. C. *et al.* Complex MSH2 and MSH6 mutations in hypermutated microsatellite unstable advanced prostate cancer. *Nat. Commun.* **5,** (2014).
16. Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534,** 47–54 (2016).
17. Polak, P. *et al.* A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. *Nat. Genet.* (2017). doi:10.1038/ng.3934
18. Davies, H. *et al.* HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat. Med.* **23,** 517–525 (2017).
19. Lord, C. J. & Ashworth, A. BRCAness revisited. *Nat. Rev. Cancer* **16,** 110–120 (2016).
20. Landrum, M. J. *et al.* ClinVar: Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* **42,** (2014).
21. Menghi, F. *et al.* The Tandem Duplicator Phenotype Is a Prevalent Genome-Wide Cancer Configuration Driven by Distinct Gene Mutations. *Cancer Cell* **34,** 197–210.e5 (2018).
22. Priestley, P. *et al.* Pan-cancer whole genome analyses of metastatic solid tumors. *bioRxiv* 415133 (2018). doi:10.1101/415133
23. Broad Institute. Picard tools. *https://broadinstitute.github.io/picard/* (2016). Available at: https://broadinstitute.github.io/picard/%5Cnhttp://broadinstitute.github.io/picard/.
24. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536,** 285–291 (2016).
25. Casper, J. *et al.* The UCSC Genome Browser database: 2018 update. *Nucleic Acids Res.* **46,** D762–D769 (2018).
26. Forbes, S. A. *et al.* COSMIC: Somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45,** D777–D783 (2017).
27. Tamborero, D. *et al.* Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med.* **10,** 25 (2018).

28. Griffith, M. *et al.* CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat. Genet.* **49,** 170–174 (2017).

29. Martincorena, I. *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **171,** 1029–1041.e21 (2017).

30. Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12,** R41 (2011).

31. Harrow, J. *et al.* GENCODE: The reference human genome annotation for the ENCODE project. *Genome Res.* **22,** 1760–1774 (2012).

32. Huber, W., Toedling, J. & Steinmetz, L. M. Transcript mapping with high-density oligonucleotide tiling arrays. *Bioinformatics* **22,** 1963–1970 (2006).

33. Gel, B. & Serra, E. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* **33,** 3088–3090 (2017).

34. Blokzijl, F., Janssen, R., van Boxtel, R. & Cuppen, E. MutationalPatterns: Comprehensive genome-wide analysis of mutational processes. *Genome Med.* (2018). doi:10.1186/s13073-018-0539-0

35. Meijer, T. G. *et al.* Functional ex vivo assay reveals homologous recombination deficiency in breast cancer beyond BRCA gene defects. *Clin. Cancer Res.* (2018). doi:10.1158/1078-0432.CCR-18-0063

36. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500,** 415–421 (2013).

37. Gaujoux, R. & Seoighe, C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* **11,** (2010).

38. Cortés-ciriano, I., Lee, J., Xi, R., Jain, D. & Jung, Y. L. Comprehensive analysis of chromothripsis in 2 , 658 human cancers using whole-genome sequencing. (2018).

39. Hahsler, M., Hornik, K. & Buchta, C. Getting Things in Order : An Introduction to the R Package seriation. *J. Stat. Softw.* **25,** 1–27 (2008).

40. Abhishek A. Singh, Karianne Schuurman, Ekaterina Nevedomskaya, Suzan Stelloo, Simon Linder, Marjolein Droog, Yongsoo Kim, Joyce Sanders, Henk van der Poel, Andries M Bergman, Lodewyk FA Wessels, W. Z. Optimized ChIP-seq method facilitates transcription factor profiling in human tumors. *Life Sci. Alliance* **In Press,**

41. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).

42. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** (2008).

43. Kumar, V. *et al.* Uniform, optimal signal processing of mapped deep-sequencing data. *Nat. Biotechnol.* **31,** 615–622 (2013).

44. Stelloo, S. *et al.* Endogenous androgen receptor proteomic profiling reveals genomic subcomplex involved in prostate tumorigenesis. *Oncogene* **37,** 313–322 (2018).

Acknowledgements

This publication and the underlying study have been made possible partly on the basis of the data that Hartwig Medical Foundation and the Center of Personalized Cancer Treatment (CPCT) have made available to the study. We would like to thank the local principal investigators of all contributing centers for their help with patient enrollment (listed in **supplementary table 1**). We would also like to thank Tesa M. Severson for her help with the computational analyses of the ChIP-seq data, Suzan Stelloo for providing ChIP-seq results on cell lines and Arne van Hoeck for providing the CHORD (HR-deficiency) prediction scores. In addition, we would like to thank the Barcode for Life foundation for making this research possible.

Grants

This work was supported in parts by a KWF-Alpe d'HuZes project [NKI 2014-7080], a grant from Astellas Pharma [Lolkema/NL-72-RG-11] and a Johnson & Johnson grant [212082PCR3014].

Author contributions

LFVD, JVR, MS, MPL and HJGVDW wrote the manuscript, which all authors critically reviewed. JVR and HJGVDW performed the bioinformatics analyses. LFVD, MPL and NM managed clinical data assessment. PW and WZ performed ChIP-seq experiments and data analyses. NM, MPL, PH, AMB, MSVDH, IMVO, and RDW are clinical contributors. MSVDH, EV, NS, JWMM and SS are part of the CPCT-02 study. EPJGC coordinated the sequencing of samples and contributed to the bioinformatics analyses.

Competing interests

All authors declare no competing interests.

Additional information

Supplementary information is available for this paper in the supplementary information file.

285 **FIGURE LEGENDS**

286 **Figure 1 - Overview of study design and patient cohort (*n* = 197).**

287 (a) **Flowchart of patient inclusion.** From the CPCT-02 cohort, patients with metastatic prostate
288 cancer were selected. Patients were excluded if data from metastatic samples were not
289 available and if clinical data analysis showed that patients had hormone-sensitive or neuro-
290 endocrine prostate cancer or unknown disease status at the time of analysis.

291 (b) **Overview of the biopsy sites.** Number of biopsies per metastatic site analyzed with WGS.

292 (c) **Age of patients at biopsy.** Bee-swarm boxplot with notch of the patient age distribution.

293

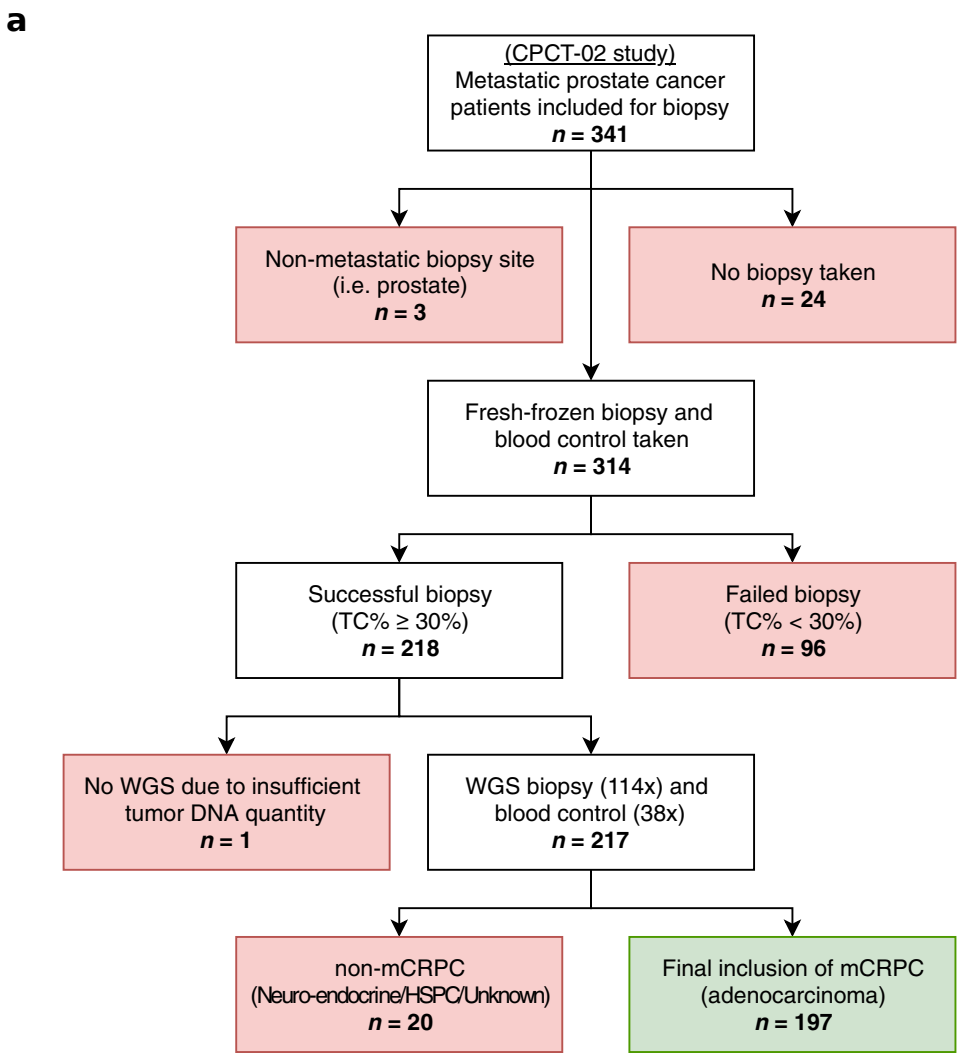
294 **Figure 2 - mCRPC remains a genetically heterogeneous disease but shows several recurrent**
295 **somatic alterations in key genes affecting several oncogenic pathways.**

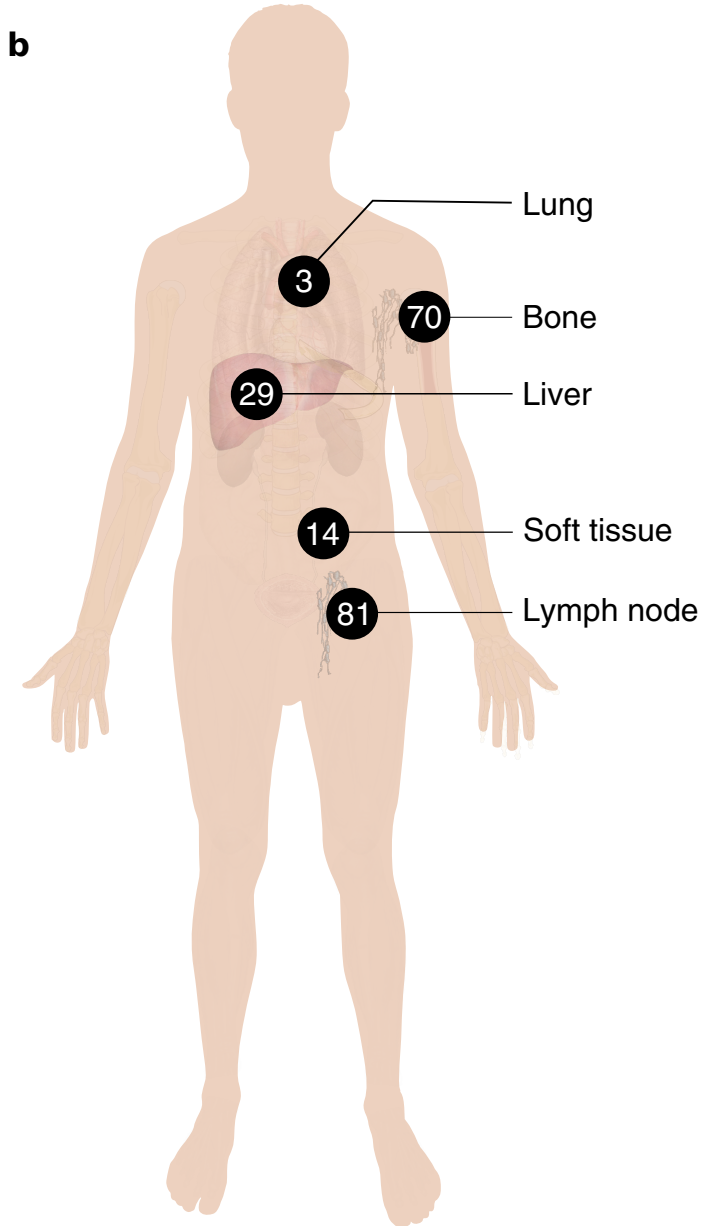
296 Based on dN/dS (q ≤ 0.1) and GISTIC2 focal peak (q ≤ 0.1) criteria, we show the genes and
297 focal genomic foci which are most often recurrently mutated, amplified or deleted in our
298 mCRPC cohort of 197 patients. The upper track (top bar plot) displays tumor mutational
299 burden (TMB) per SNV, InDel and MNV category on coding regions (square root scale).
300 Samples are sorted based on mutual-exclusivity of the depicted genes and foci. The heatmap
301 displays the type of mutation(s) per sample, (light-)green or (light-)red backgrounds depict
302 copy-number aberrations whilst the inner square depicts the type of (coding) mutation(s).
303 Relative proportions of mutational categories (coding mutations (SNV, InDels and MNV), SV,
304 deep gains (high-level amplifications resulting in many additional copies) and deep deletions
305 (high-level losses resulting in (near) homozygous losses) per gene and foci are shown in the
306 barplot next to the heatmap. Narrow GISTIC2 peaks covering ≤ 3 genes were reduced to
307 gene-level rows if one of these genes is present in the dN/dS (q ≤ 0.1) analysis or is a known
308 oncogene or tumor-suppressor. For GISTIC2 peaks covering multiple genes, only deep
309 amplifications and deep deletions are shown. Recurrent aberrant focal genomic foci in gene
310 deserts are annotated with their nearest gene. Significance scores (-1*log10(q)) of the dN/dS
311 and GISTIC2 analysis are shown on the outer-right barplots; bars in the GISTIC2 significance
312 plot are colored red if these foci were detected as a recurrent focal deletion and green if
313 detected as a recurrent focal gain. Per sample, the presence of (predicted) ETS fusions,

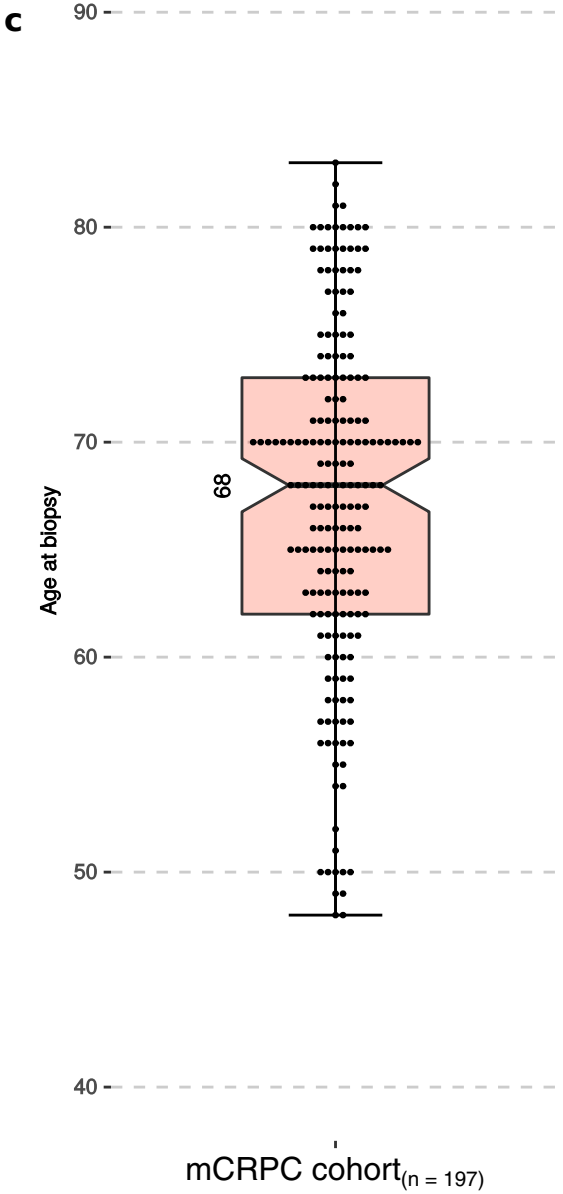
314 kataegis, chromothripsis, CHORD predictive score (HR-deficiency), MSI status and biopsy
315 location are shown as bottom tracks
316
317 **Figure 3 - Whole Genome Sequencing reveals novel insight into the various molecular (non-**
318 **coding) aberrations affecting AR regulation.**
319 **(a)** Mutational overview of top recurrently-mutated genes affecting AR regulation and their
320 putative enhancer foci (as detected by GISTIC2). The first track represents the number of
321 genomic mutations per Mb (TMB) per SNV, InDels and MNV category (square root scale).
322 Samples are sorted based on mutual-exclusivity of the depicted genes and foci. The heatmap
323 displays the type of mutation(s) per sample, (light-)green or (light-)red backgrounds depict
324 copy-number aberrations whilst the inner square depicts the type of (coding) mutation(s).
325 Relative proportions of mutational categories (coding mutations (SNV, InDels and MNV), SV,
326 deep gains and deep deletions) per gene and foci are shown in the barplot next to the
327 heatmap. The presence of (predicted) ETS fusions, kataegis, chromothripsis, CHORD
328 predictive score (HR-deficiency), MSI status and biopsy location are shown as bottom tracks.
329 **(b)** Overview of the copy-number deviations between putative enhancer and gene regions for *AR*
330 and *MYC*. Samples were categorized as enhancer- or gene-enriched if enhancer-to-gene
331 ratio deviated >1 studentized residual (residual derived from a linear model without the
332 respective observation) from a 1:1 ratio.
333 **(c)** Copy-number and ChIP-seq profiles surrounding the *AR* and *PCAT1/MYC* gene loci (with
334 1.25 additional Mb up-/downstream). The upper track displays the selected genomic window
335 and the overlapping genes. The first and second track display the aggregated mean copy-
336 number (per 1000bp window) of the enhancer- and gene-enriched samples, respectively.
337 These profiles identify distinct amplified regions (indicated by red asterisk) in proximity to the
338 respective gene bodies. The 3th to 8th tracks represent AR ChIP-seq profiles (mean read-
339 coverage per 1Mb windows) in two mCRPC patients, LNCaP and LNCaP with R1881
340 treatment, VCaP and bicalutamide-resistant VCaP. The 9th to 11th tracks represent FOXA1
341 ChIP-seq profiles (mean read-coverage per 1Mb windows) in two mCRPC patients and
342 LNCaP with R1881 treatment. The 12th to 14th tracks represent H3K27ac ChIP-seq profiles

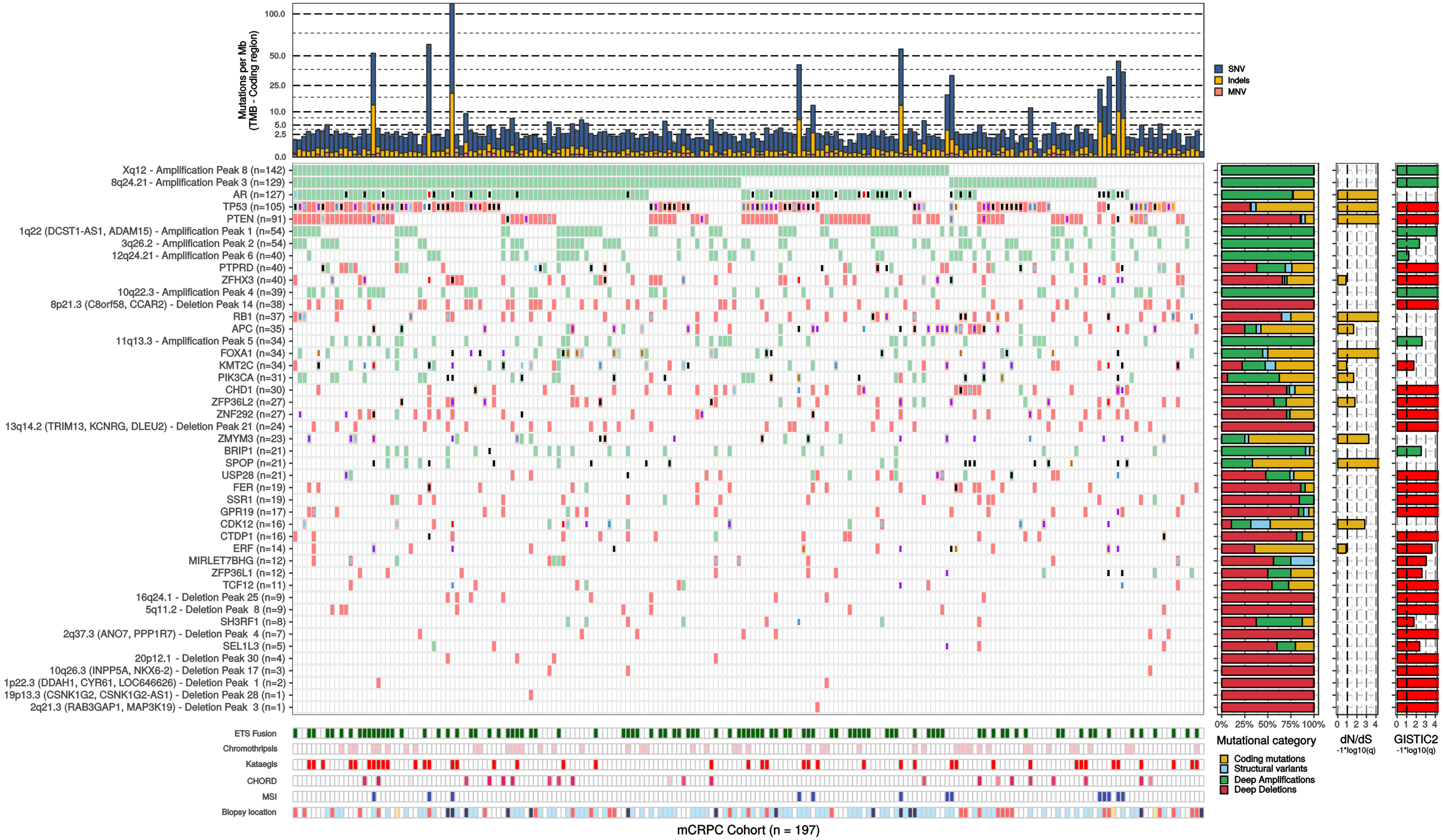
343 (mean read-coverage per 1Mb windows) in two mCRPC patients and LNCaP with R1881
344 treatment reflecting active enhancer regions.
345 ChIP-seq peaks (MACS/MACS2; q < 0.01) are shown as black lines per respective sample.
346
347 **Figure 4 - Unsupervised clustering of mCRPC reveals distinct genomic phenotypes.**
348 **(a) Dendogram of unsupervised clustering (Pearson correlation; ward.D) with optimal leaf**
349 **ordering.** Top eight clusters are highlighted and denoted based on order of appearance (left
350 to right): A to H. Y-axis displays clustering distance (Pearson; ward.D).
351 **(b) Number of genomic mutations per Mb (TMB) of SNV, InDels and MNV categories.** All
352 genome-wide somatic mutations were taken into consideration (square root scale).
353 **(c) Absolute number of unique structural variants per sample.** Cumulative frequency of
354 inversions, tandem duplication, deletions, insertions and translocations.
355 **(d) Relative frequency per structural variant category, Tandem Duplications and Deletions**
356 **are subdivided into >100kb and <100kb categories.** This track shows if an enrichment for
357 particular category of (somatic) structural variant can be detected, which in turn, can be
358 indicative for a specific mutational aberration.
359 **(e) Relative genome-wide ploidy status, ranging from 0 to ≥7 copies.** This track shows the
360 relative percentage of the entire genome which is (partially) lost (ploidy < 2/diploid) or
361 amplified (> 2/diploid).
362 **(f) Relative contribution to mutational signatures (COSMIC) summarized per proposed**
363 **etiology.** This track displays the proposed etiology of each SNV based on their mutational
364 contexts.
365 **(g) Relative frequency of SNV mutational changes.**
366 **(h) HR-deficient prediction score as assessed by CHORD.** The binary prediction score of
367 CHORD (ranging from 0 to 1) is shown, in which higher scores reflect more evidence for HR-
368 deficiency in a given sample.
369 **(i) MSI status as determined using a stringent threshold of MSI characteristics[22].**
370
371 **Figure 5 - Distinct molecular phenotypes in mCRPC are enriched by mutually-exclusive**
372 **aberrations in key pathways or large-scale somatic events.**

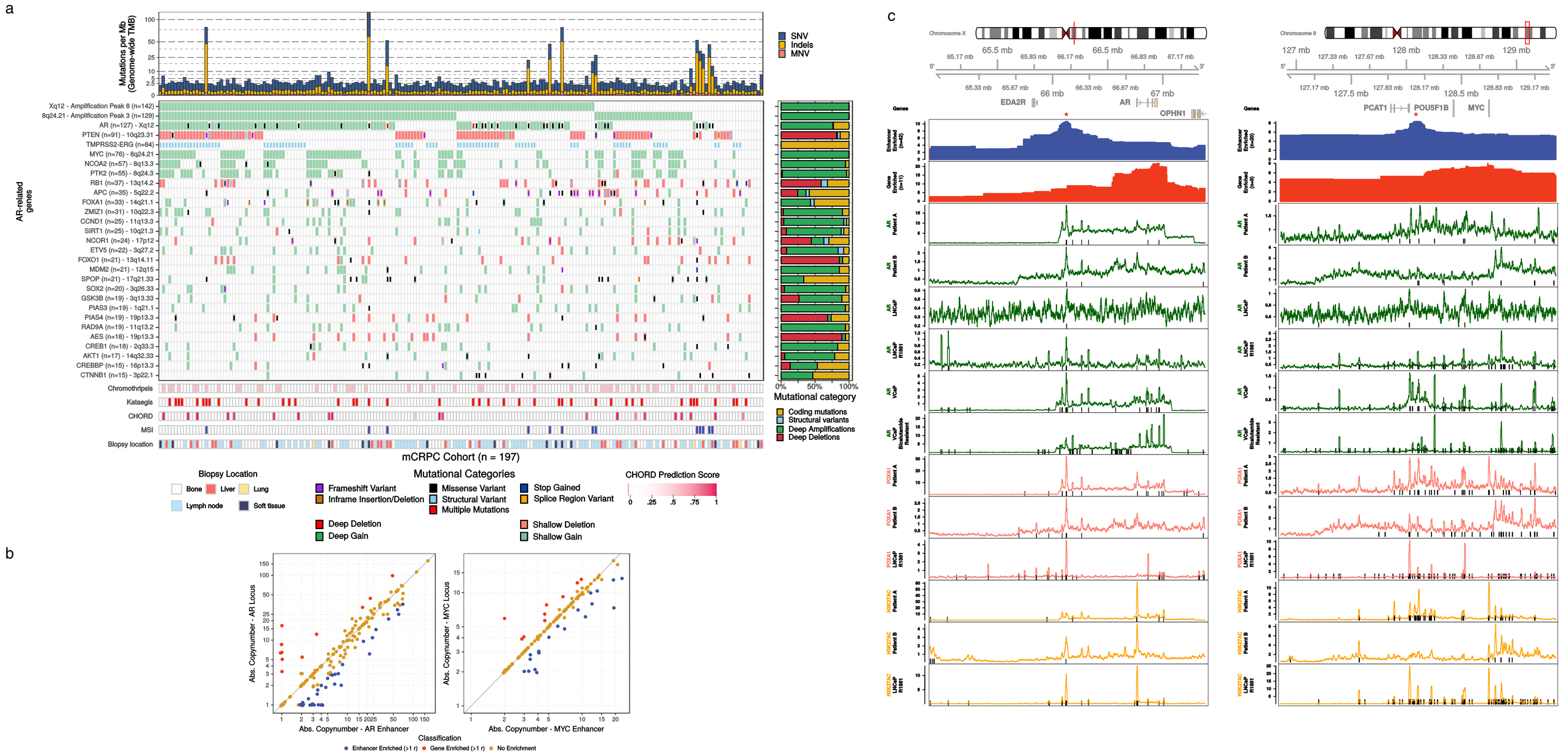
373 (a) **Cluster-specific enrichment of mutated genes, chromothripsis, gene fusions and**
374 **kataegis (Fisher's Exact Test ≤ 0.05).** Percentages to the left of the black line represent the
375 relative mutational frequency in mCRPC samples which are not present in the respective
376 cluster, whilst the percentages to the right of the black line represent the relative mutational
377 frequency present in the samples from the tested cluster.
378 (b) **Genomic overview with biologically-relevant genes in the clusters with mutational**
379 **enrichment of genes or large-scale events (A, B, D and F).** The first track represents the
380 number of genomic mutations per Mb (TMB) per SNV, InDels and MNV category (square-root
381 scale). The second track represents the absolute number of unique structural variants per
382 sample. The third track represents the relative frequency per structural variant category,
383 Tandem Duplications and Deletions are subdivided into >100kb and <100kb categories. The
384 fourth track represents relative genome-wide ploidy status, ranging from 0 to ≥7 copies. The
385 fifth track represents the relative contribution to mutational signatures (COSMIC) summarized
386 per proposed etiology. The sixth track displays somatic mutations in the relevant genes found
387 in at least one cluster. The lower tracks represent presence of ETS fusions, chromothripsis,
388 kataegis, HR-deficiency prediction scores and MSI status based on a threshold of MSI
389 characteristics.
390
391 **Table 1 – Cluster characteristics**
392 Overview of the distinctive characteristics for each cluster (A-H).

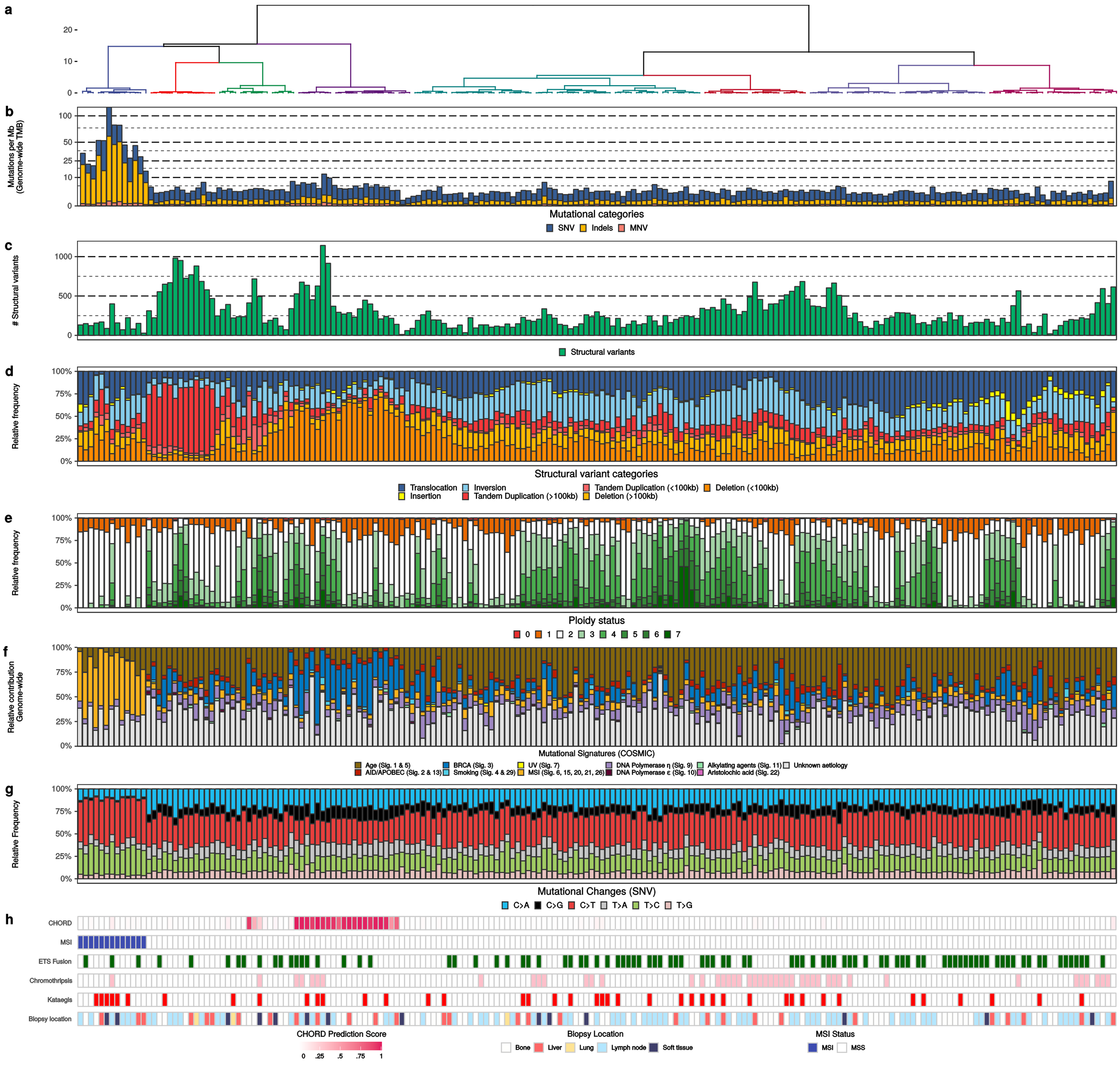

a
(CPCT-02 study)
Metastatic prostate cancer patients included for biopsy
n = 341
Non-metastatic biopsy site (i.e. prostate)
n = 3
No biopsy taken
n = 24
Fresh-frozen biopsy and blood control taken
n = 314
Successful biopsy (TC% ≥ 30%)
n = 218
Failed biopsy (TC% < 30%)
n = 96
No WGS due to insufficient tumor DNA quantity
n = 1
WGS biopsy (114x) and blood control (38x)
n = 217
non-mCRPC (Neuro-endocrine/HSPC/Unknown)
n = 20
Final inclusion of mCRPC (adenocarcinoma)
n = 197

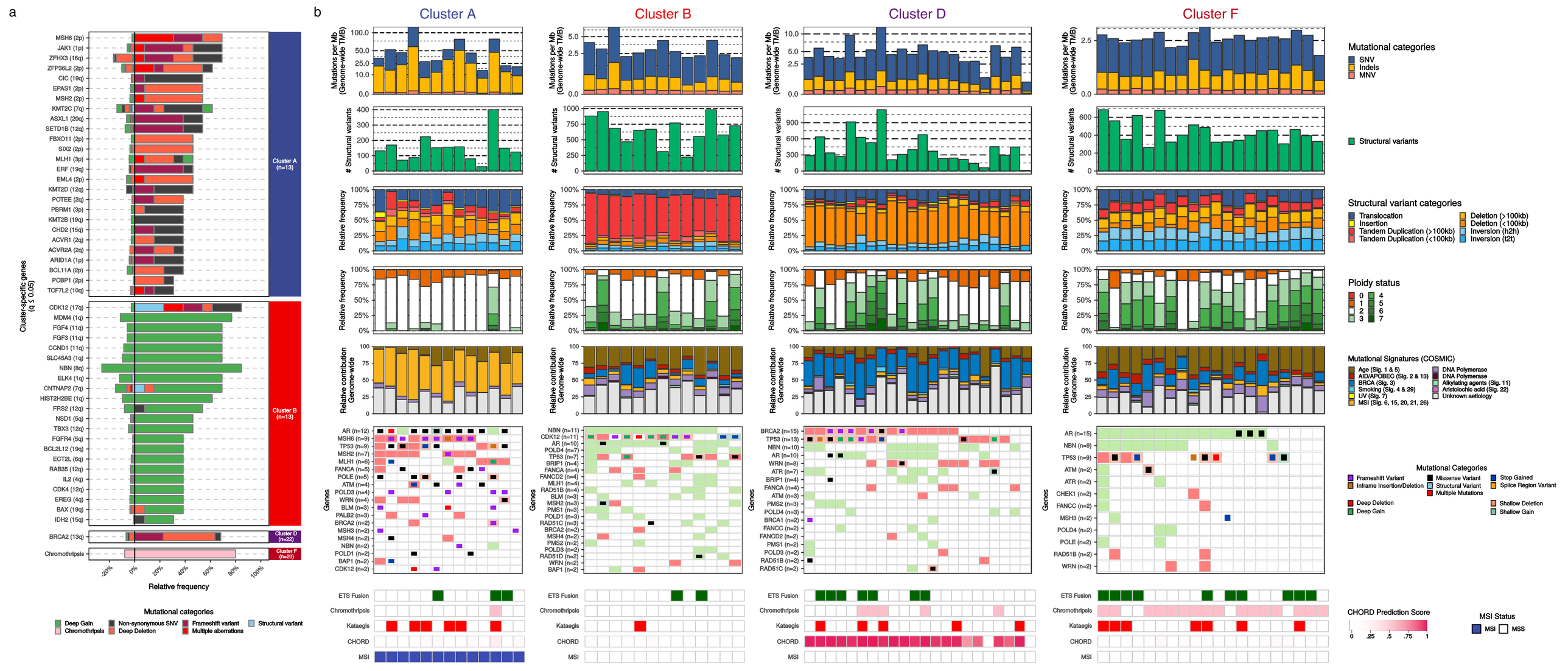

b
Lung
3
Bone
70
Liver
29
Soft tissue
14
Lymph node
81


c
90
80
70
60
50
40
Age at biopsy
68
mCRPC cohort(n = 197)


Mutations per Mb (TMB - Coding region)
100.0
50.0
25.0
10.0
5.0
2.5
0.0
SNV
Indels
MNV
Xq12 - Amplification Peak 8 (n=142)
8q24.21 - Amplification Peak 3 (n=129)
AR (n=127)
TP53 (n=105)
PTEN (n=91)
1q22 (DCST1-AS1, ADAM15) - Amplification Peak 1 (n=54)
3q26.2 - Amplification Peak 2 (n=54)
12q24.21 - Amplification Peak 6 (n=40)
PTPRD (n=40)
ZFHX3 (n=40)
10q22.3 - Amplification Peak 4 (n=39)
8p21.3 (C8orf58, CCAR2) - Deletion Peak 14 (n=38)
RB1 (n=37)
APC (n=35)
11q13.3 - Amplification Peak 5 (n=34)
FOXA1 (n=34)
KMT2C (n=34)
PIK3CA (n=31)
CHD1 (n=30)
ZFP36L2 (n=27)
ZNF292 (n=27)
13q14.2 (TRIM13, KCNRG, DLEU2) - Deletion Peak 21 (n=24)
ZMYM3 (n=23)
BRIP1 (n=21)
SPOP (n=21)
USP28 (n=21)
FER (n=19)
SSR1 (n=19)
GPR19 (n=17)
CDK12 (n=16)
CTDP1 (n=16)
ERF (n=14)
MIRLET7BHG (n=12)
ZFP36L1 (n=12)
TCF12 (n=11)
16q24.1 - Deletion Peak 25 (n=9)
5q11.2 - Deletion Peak 8 (n=9)
SH3RF1 (n=8)
2q37.3 (ANO7, PPP1R7) - Deletion Peak 4 (n=7)
SEL1L3 (n=5)
20p12.1 - Deletion Peak 30 (n=4)
10q26.3 (INPP5A, NKX6-2) - Deletion Peak 17 (n=3)
1p22.3 (DDAH1, CYR61, LOC646626) - Deletion Peak 1 (n=2)
19p13.3 (CSNK1G2, CSNK1G2-AS1) - Deletion Peak 28 (n=1)
2q21.3 (RAB3GAP1, MAP3K19) - Deletion Peak 3 (n=1)
ETS Fusion
Chromothripsis
Kataegis
CHORD
MSI
Biopsy location
0% 25% 50% 75% 100%
Mutational category
Coding mutations
Structural variants
Deep Amplifications
Deep Deletions
0 1 2 3 4
dN/dS
-1*log10(q)
0 1 2 3 4
GISTIC2
-1*log10(q)
mCRPC Cohort (n = 197)
Biopsy Location
Bone
Liver
Lung
Lymph node
Soft tissue
Mutational Categories
Frameshift Variant
Inframe Insertion/Deletion
Missense Variant
Structural Variant
Multiple Mutations
Stop Gained
Splice Region Variant
Deep Deletion
Deep Gain
Shallow Deletion
Shallow Gain
CHORD Prediction Score
0
.25
.5
.75
1


a
Mutations per Mb (Genome-wide TMB)
SNV
Indels
MNV
AR-related genes
Xq12 - Amplification Peak 8 (n=142)
8q24.21 - Amplification Peak 3 (n=129)
AR (n=127) - Xq12
PTEN (n=91) - 10q23.31
TMPRSS2-ERG (n=84)
MYC (n=76) - 8q24.21
NCOA2 (n=57) - 8q13.3
PTK2 (n=55) - 8q24.3
RB1 (n=37) - 13q14.2
APC (n=35) - 5q22.2
FOXA1 (n=33) - 14q21.1
ZMIZ1 (n=31) - 10q22.3
CCND1 (n=25) - 11q13.3
SIRT1 (n=25) - 10q21.3
NCOR1 (n=24) - 17p12
ETV5 (n=22) - 3q27.2
FOXO1 (n=21) - 13q14.11
MDM2 (n=21) - 12q15
SPOP (n=21) - 17q21.33
SOX2 (n=20) - 3q26.33
GSK3B (n=19) - 3q13.33
PIAS3 (n=19) - 1q21.1
PIAS4 (n=19) - 19p13.3
RAD9A (n=19) - 11q13.2
AES (n=18) - 19p13.3
CREB1 (n=18) - 2q33.3
AKT1 (n=17) - 14q32.33
CREBBP (n=15) - 16p13.3
CTNNB1 (n=15) - 3p22.1
Chromothripsis
Kataegis
CHORD
MSI
Biopsy location
mCRPC Cohort (n = 197)
Mutational category
Coding mutations
Structural variants
Deep Amplifications
Deep Deletions
Biopsy Location
Bone
Liver
Lung
Lymph node
Soft tissue
Mutational Categories
Frameshift Variant
Inframe Insertion/Deletion
Missense Variant
Structural Variant
Multiple Mutations
Stop Gained
Splice Region Variant
CHORD Prediction Score
Deep Deletion
Deep Gain
Shallow Deletion
Shallow Gain
b
Abs. Copynumber - AR Locus
Abs. Copynumber - AR Enhancer
Abs. Copynumber - MYC Locus
Abs. Copynumber - MYC Enhancer
Classification
Enhancer Enriched (>1 r)
Gene Enriched (>1 r)
No Enrichment
c
Chromosome X
Chromosome 8
Genes
EDA2R
AR
OPHN1
PCAT1
POU5F1B
MYC
Enhancer Enriched (n=42)
Gene Enriched (n=11)
Enhancer Enriched (n=20)
Gene Enriched (n=3)
AR Patient A
AR Patient B
AR LNCaP
AR LNCaP R1881
AR VCaP
AR VCaP Bicalutamide Resistant
FOXA1 Patient A
FOXA1 Patient B
FOXA1 LNCaP R1881
H3K27AC Patient A
H3K27AC Patient B
H3K27AC LNCaP R1881


a
b
Mutations per Mb (Genome-wide TMB)
Mutational categories
SNV Indels MNV
c
Structural variants
Structural variants
d
Relative frequency
Structural variant categories
Translocation Inversion Tandem Duplication (<100kb) Deletion (<100kb)
Insertion Tandem Duplication (>100kb) Deletion (>100kb)
e
Relative frequency
Ploidy status
0 1 2 3 4 5 6 7
f
Relative contribution Genome-wide
Mutational Signatures (COSMIC)
Age (Sig. 1 & 5) BRCA (Sig. 3) UV (Sig. 7) DNA Polymerase η (Sig. 9) Alkylating agents (Sig. 11) Unknown aetiology
AID/APOBEC (Sig. 2 & 13) Smoking (Sig. 4 & 29) MSI (Sig. 6, 15, 20, 21, 26) DNA Polymerase ε (Sig. 10) Aristolochic acid (Sig. 22)
g
Relative Frequency
Mutational Changes (SNV)
C>A C>G C>T T>A T>C T>G
h
CHORD
MSI
ETS Fusion
Chromothripsis
Kataegis
Biopsy location
CHORD Prediction Score
0 .25 .5 .75 1
Biopsy Location
Bone Liver Lung Lymph node Soft tissue
MSI Status
MSI MSS


a
b
Cluster A
Cluster B
Cluster D
Cluster F
Cluster-specific genes (q ≤ 0.05)
Relative frequency
Mutational categories
Deep Gain
Chromothripsis
Non-synonymous SNV
Deep Deletion
Frameshift variant
Multiple aberrations
Structural variant
Mutations per Mb (Genome-wide TMB)
Structural variants
Relative frequency
Relative contribution Genome-wide
Genes
ETS Fusion
Chromothripsis
Kataegis
CHORD
MSI
Mutational categories
SNV
Indels
MNV
Structural variants
Structural variant categories
Translocation
Insertion
Tandem Duplication (>100kb)
Tandem Duplication (<100kb)
Deletion (>100kb)
Deletion (<100kb)
Inversion (h2h)
Inversion (t2t)
Ploidy status
Mutational Signatures (COSMIC)
Age (Sig. 1 & 5)
AID/APOBEC (Sig. 2 & 13)
BRCA (Sig. 3)
Smoking (Sig. 4 & 29)
UV (Sig. 7)
MSI (Sig. 6, 15, 20, 21, 26)
DNA Polymerase
DNA Polymerase
Alkylating agents (Sig. 11)
Aristolochic acid (Sig. 22)
Unknown aetiology
Mutational Categories
Frameshift Variant
Inframe Insertion/Deletion
Missense Variant
Structural Variant
Multiple Mutations
Stop Gained
Splice Region Variant
Deep Deletion
Deep Gain
Shallow Deletion
Shallow Gain
CHORD Prediction Score
MSI Status
MSI
MSS

	Cluster A
Number of patients (n; % of cohort)	13 (6.6)
Tumor mutational burden (CDS)	36.88
InDel/SNV ratio	0.99
Number of structural variants	149
Main structural variant category or differentiating category	None
Ploidy status	1.92
Main mutational signature	MSI
Top 3 cluster-specific abberations (% of cluster)	MSH6 (69.2) JAK1 (69.2) CIC (58.3)
ETS-fusions (n)	3
Chromothripsis (n)	1
Kataegis (n)	6

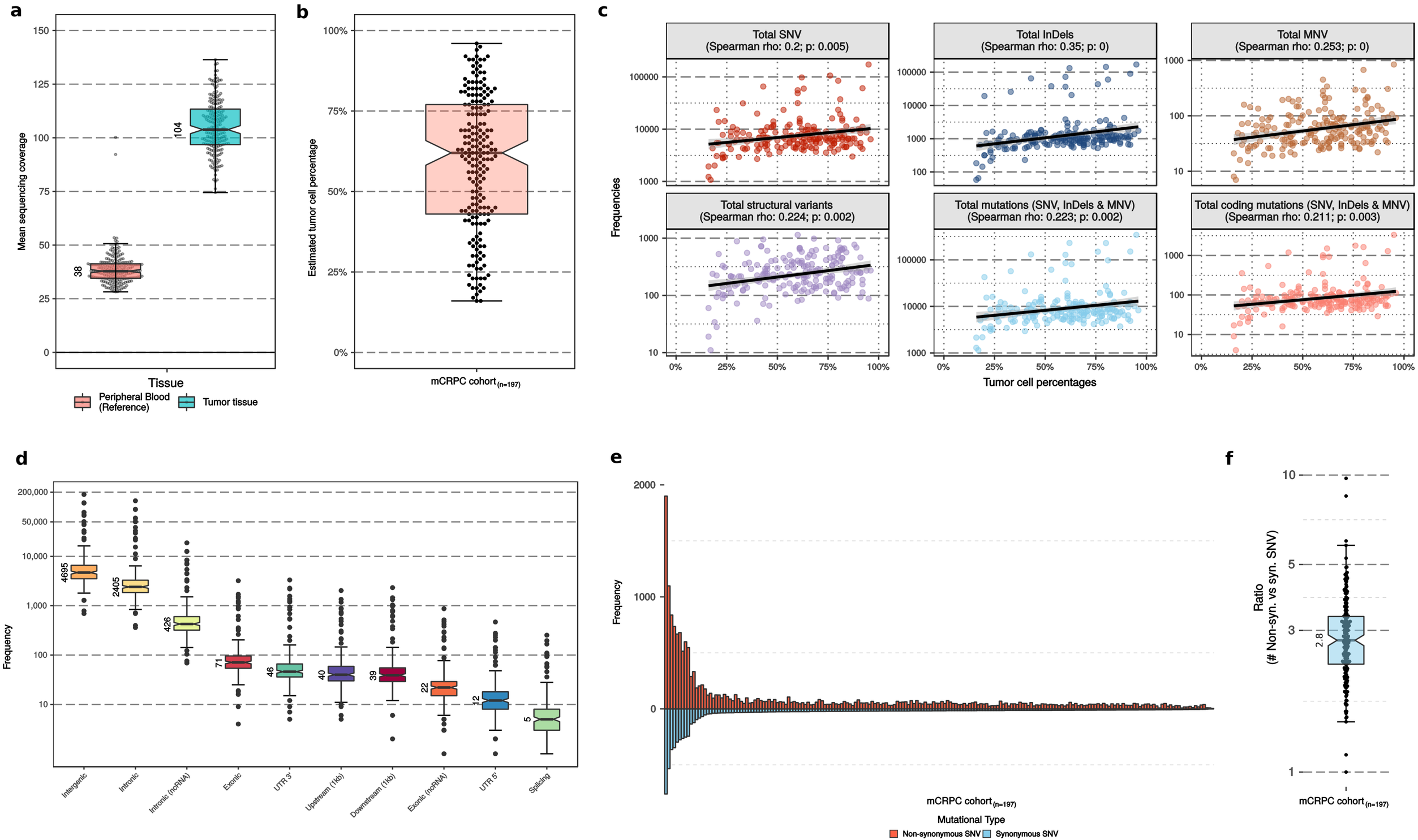
All numbers are median of the cluster, unless otherwise indicated.
CDS, coding sequence

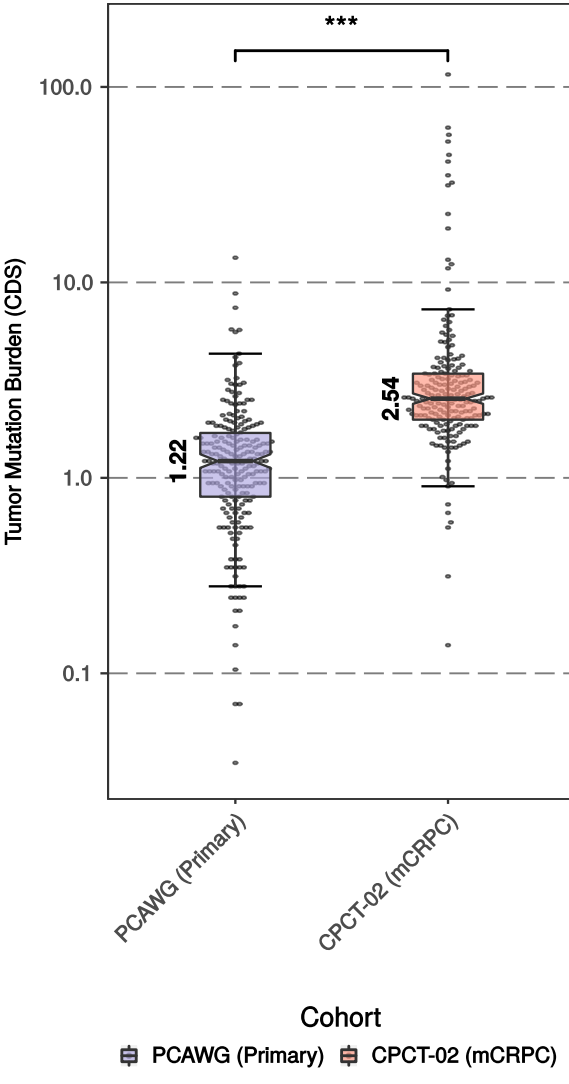
Table 1: Cluster characteristics

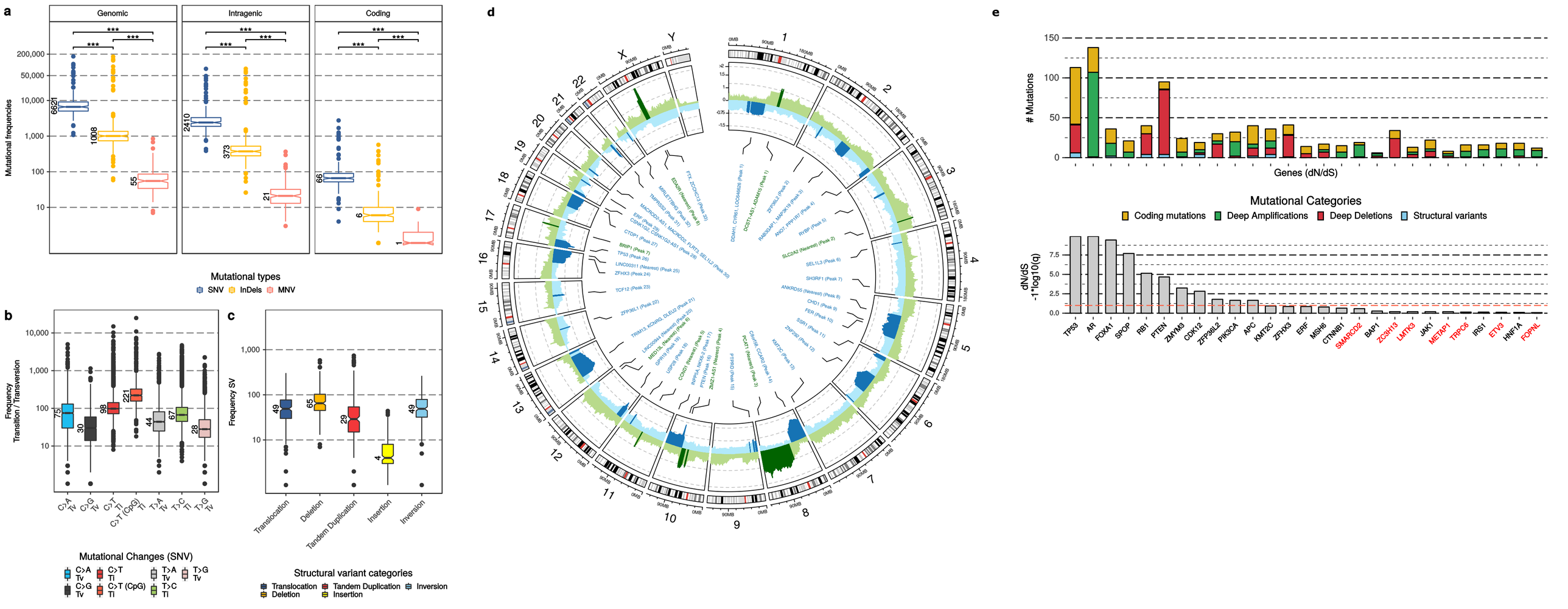
Cluster B	Cluster C	Cluster D
13 (6.6)	15 (7.6)	22 (11.2)
2.44	3.00	4.39
7.07	6.73	7.28
669	237	323
Tandem duplications (>100 kb)	Tandem duplications (<100 kb)	Deletions (>100kb)
2.39	3.19	2.16
N/A	N/A	BRCA
CDK12 (84.6) FGF3 (69.2) FGF4 (69.2)	None	BRCA2 (68.2)
2	7	7
0	1	5
1	2	5

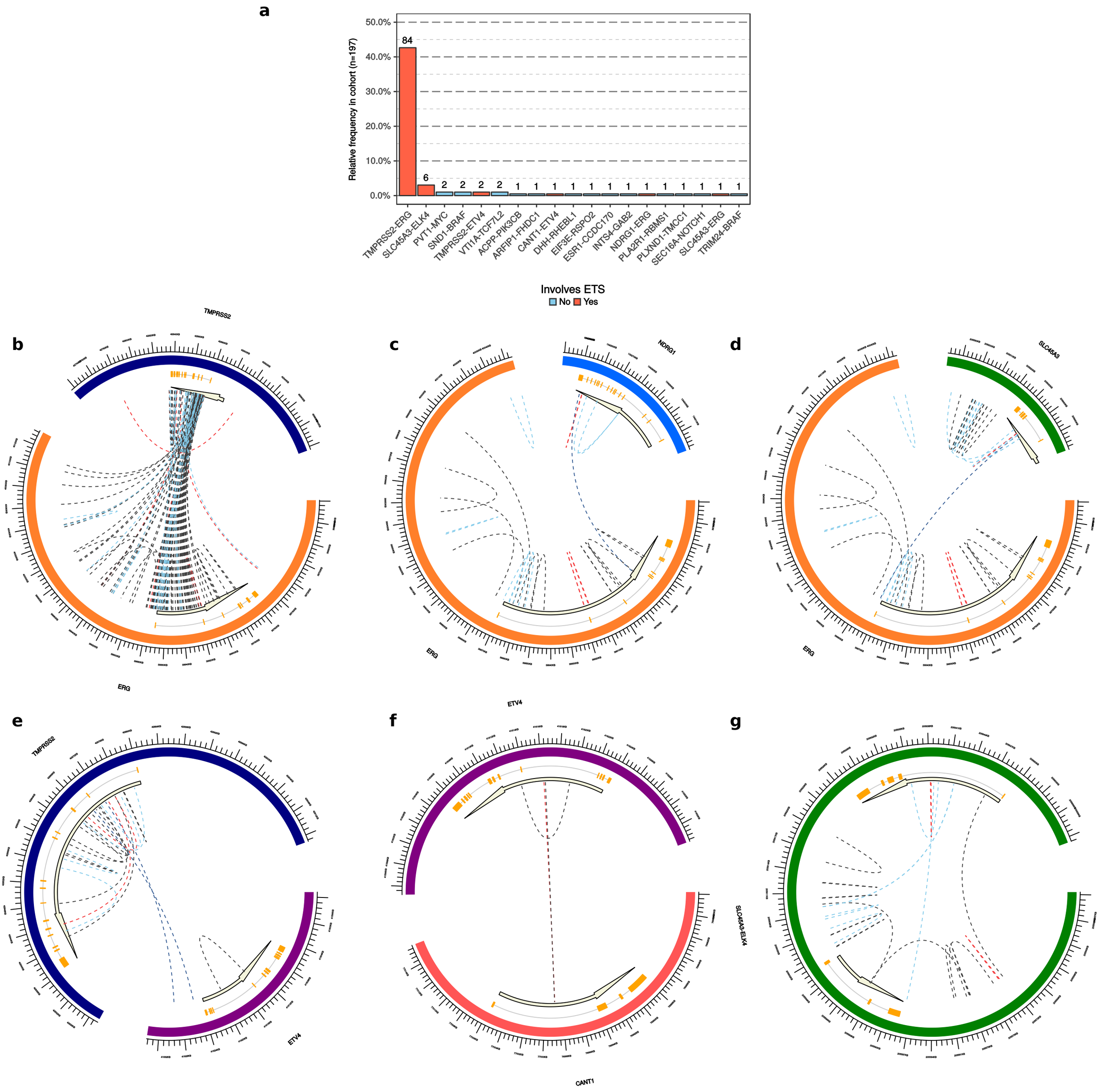
Cluster E	Cluster F	Cluster G
55 (27.9)	20 (10.15)	34 (17.3)
2.12	2.51	2.12
7.13	6.15	6.13
178	399.5	221.5
None	None	None
3.24	3.35	2.98
N/A	N/A	N/A
None	Chromothripsis (80)	None
25	10	23
8	16	8
13	7	5

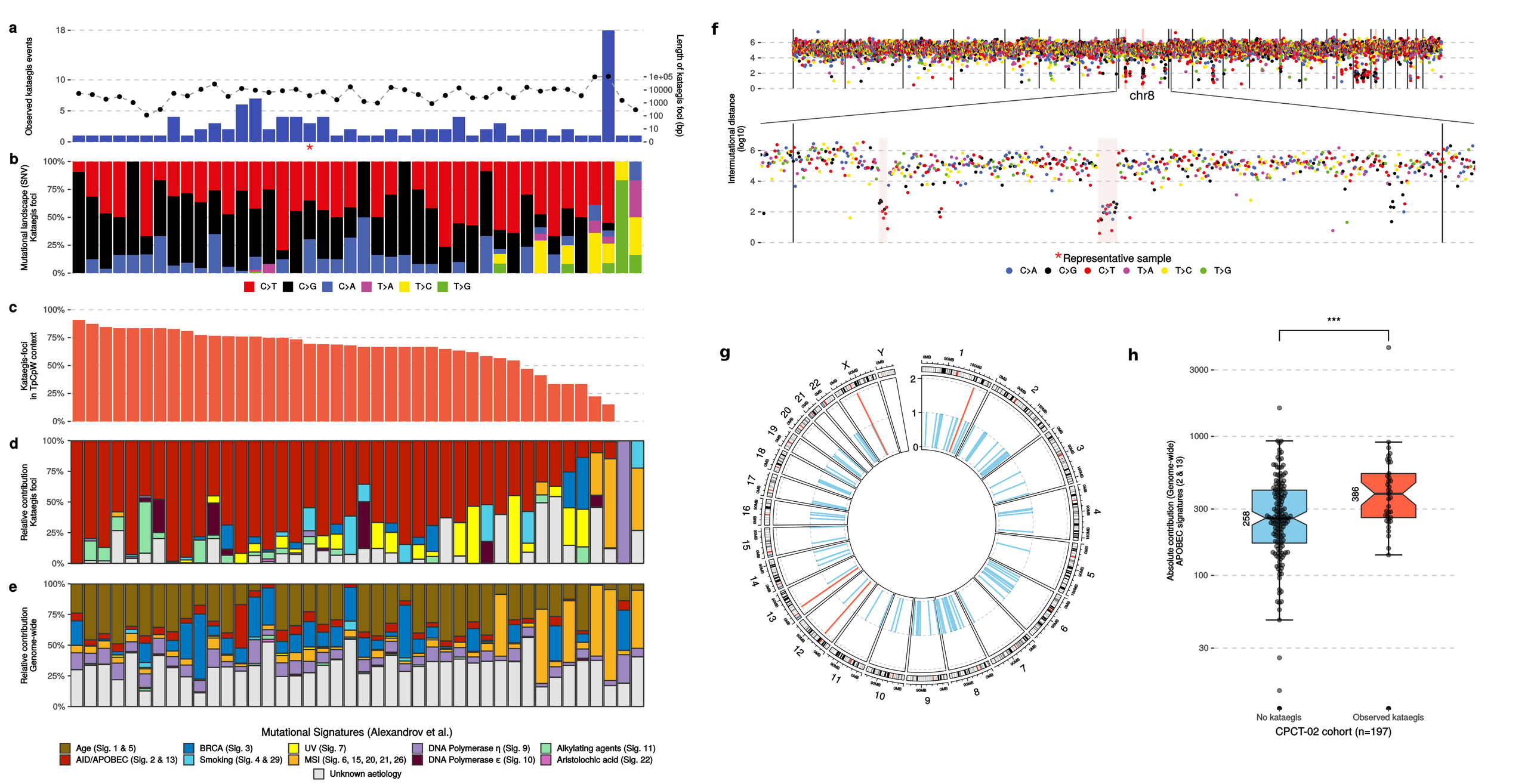
Cluster H
25 (12.7)
2.30
5.81
201
Insertions
1.97
N/A
None
16
7
3

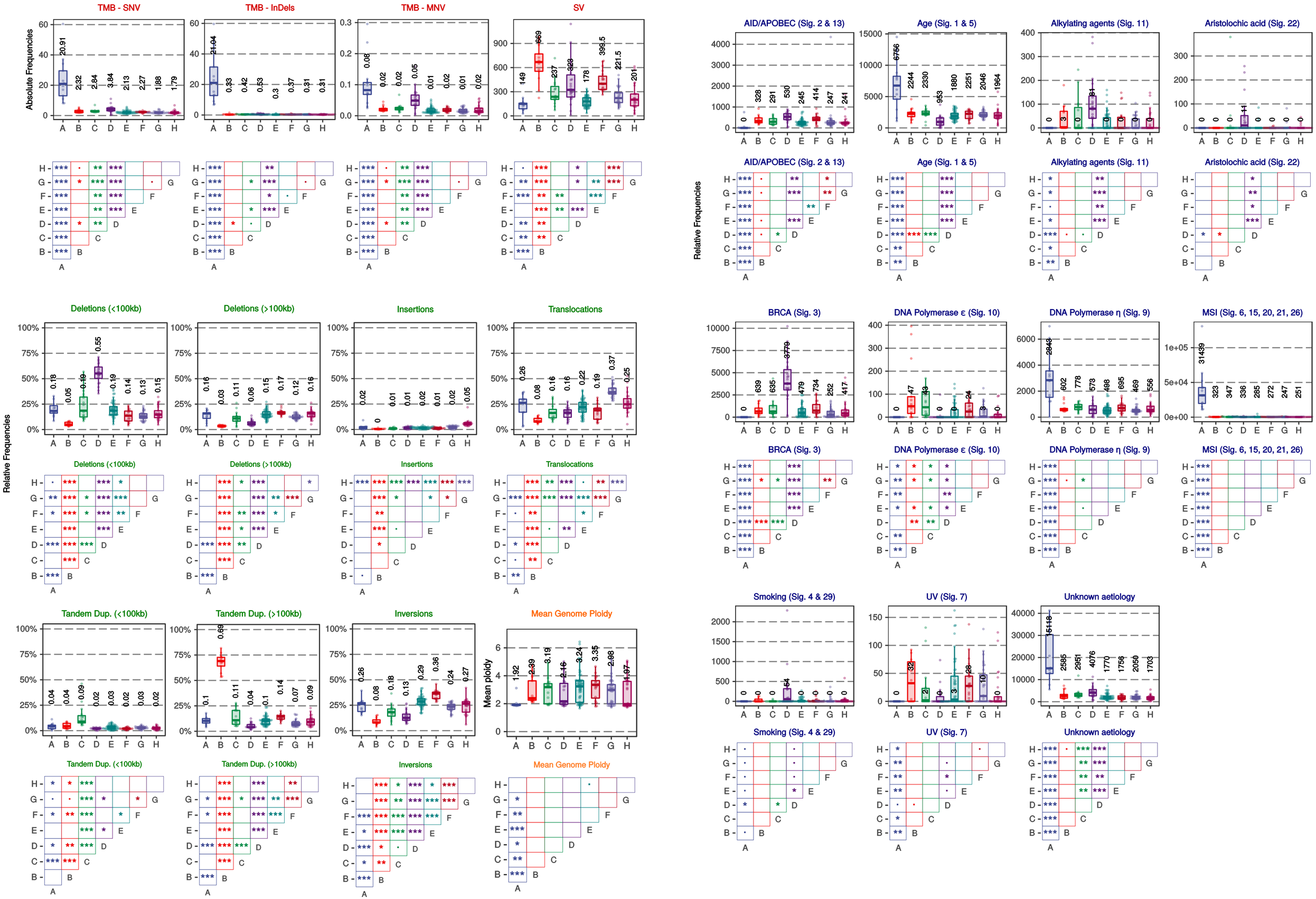

a
Mean sequencing coverage
150
125
100
75
50
25
0
104
38
Tissue
Peripheral Blood (Reference)
Tumor tissue
b
Estimated tumor cell percentage
100%
75%
50%
25%
0%
mCRPC cohort (n=197)
c
Total SNV (Spearman rho: 0.2; p: 0.005)
Total InDels (Spearman rho: 0.35; p: 0)
Total MNV (Spearman rho: 0.253; p: 0)
Total structural variants (Spearman rho: 0.224; p: 0.002)
Total mutations (SNV, InDels & MNV) (Spearman rho: 0.223; p: 0.002)
Total coding mutations (SNV, InDels & MNV) (Spearman rho: 0.211; p: 0.003)
Frequencies
Tumor cell percentages
d
Frequency
200,000
50,000
10,000
1,000
100
10
4695
2405
426
71
46
40
39
22
12
5
Intergenic
Intronic
Intronic (ncRNA)
Exonic
UTR 3'
Upstream (1kb)
Downstream (1kb)
Exonic (ncRNA)
UTR 5'
Splicing
e
Frequency
2000
1000
0
mCRPC cohort (n=197)
Mutational Type
Non-synonymous SNV
Synonymous SNV
f
Ratio (# Non-syn. vs syn. SNV)
10
5
3
1
2.8
mCRPC cohort (n=197)

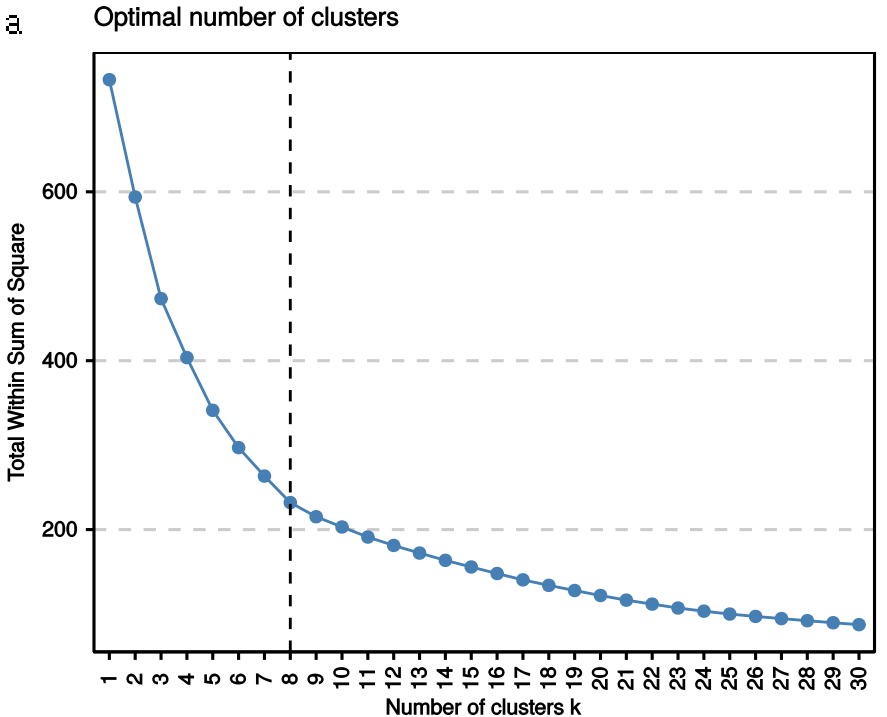

***
100.0
10.0
1.0
0.1
Tumor Mutation Burden (CDS)
1.22
2.54
PCAWG (Primary)
CPCT-02 (mCRPC)
Cohort
PCAWG (Primary)
CPCT-02 (mCRPC)

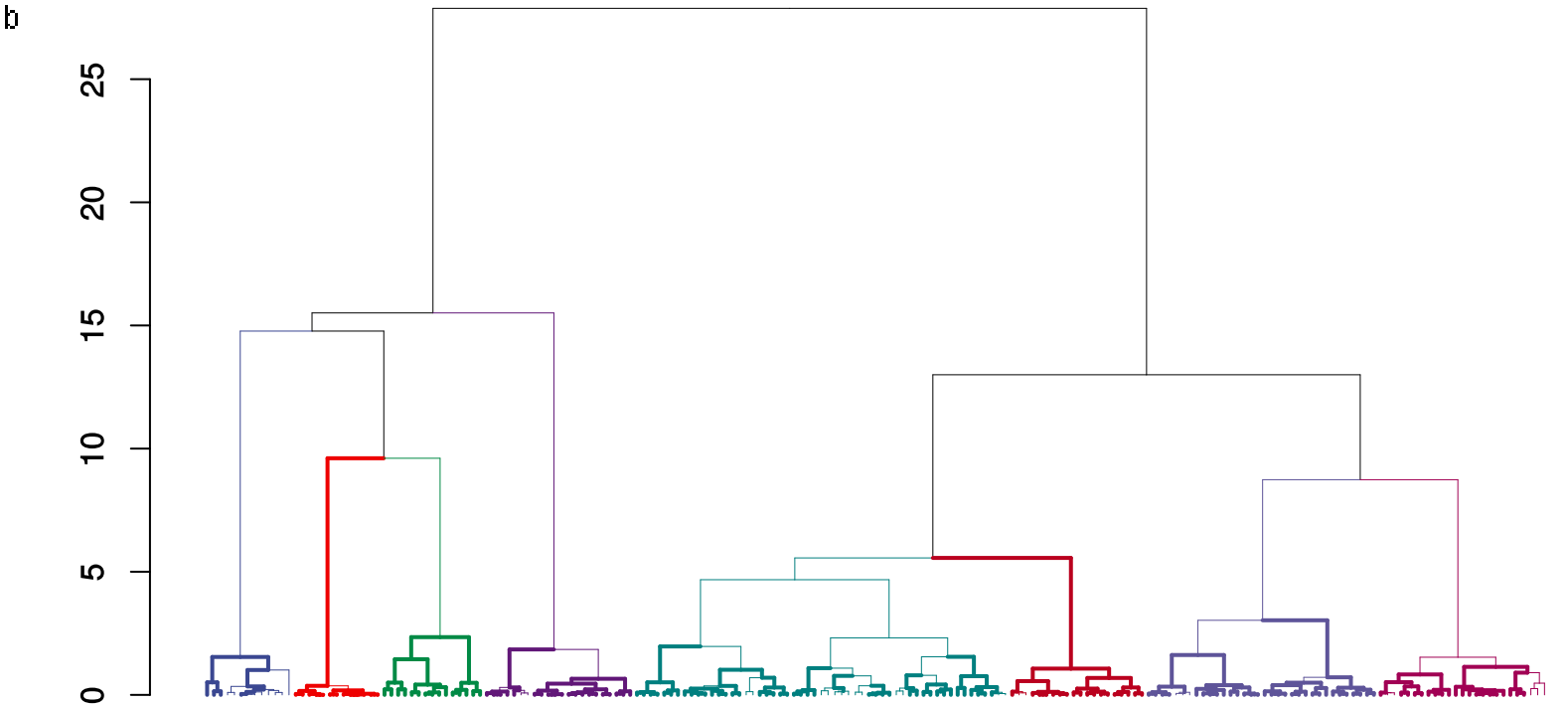

a
Genomic
Intragenic
Coding
Mutational frequencies
Mutational types
SNV
InDels
MNV
b
Frequency
Transition / Transversion
Mutational Changes (SNV)
c
Frequency SV
Translocation
Deletion
Tandem Duplication
Insertion
Inversion
Structural variant categories
d
e
Mutations
Genes (dN/dS)
Mutational Categories
Coding mutations
Deep Amplifications
Deep Deletions
Structural variants
dN/dS
-1*log10(q)
TP53
AR
FOXA1
SPOP
RB1
PTEN
ZMYM3
CDK12
ZFP36L2
PIK3CA
APC
KMT2C
ZFHX3
ERF
MSH6
CTNNB1
SMARCD2
BAP1
ZC3H13
LMTK3
JAK1
METAP1
TRPC6
IRS1
ETV3
HNF1A
FOPNL


a
Relative frequency in cohort (n=197)
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
84
6
2
2
2
2
1
1
1
1
1
1
1
1
1
1
1
1
1
TMPRSS2-ERG
SLC45A3-ELK4
PVT1-MYC
SND1-BRAF
TMPRSS2-ETV4
VTI1A-TCF7L2
ACPP-PIK3CB
ARFIP1-FHDC1
CANT1-ETV4
DHH-RHEBL1
EIF3E-RSPO2
ESR1-CCDC170
INTS4-GAB2
NDRG1-ERG
PLA2R1-RBMS1
PLXND1-TMCC1
SEC16A-NOTCH1
SLC45A3-ERG
TRIM24-BRAF
Involves ETS
No
Yes
b
TMPRSS2
ERG
c
NDRG1
ERG
ETV4
d
SLC45A3
ERG
e
TMPRSS2
ETV4
f
CANT1
g
SLC45A3-ELK4

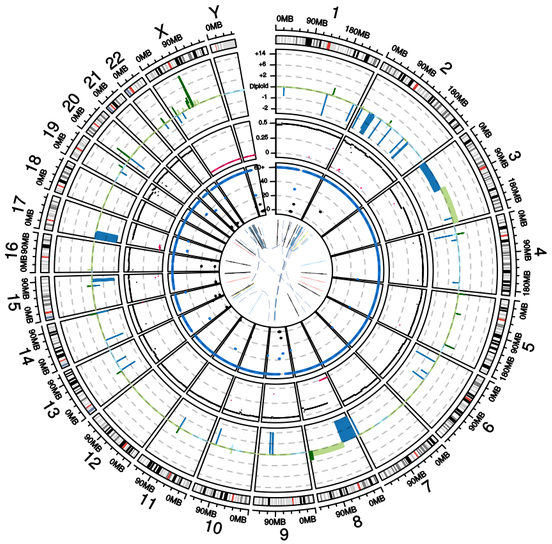

a
Observed kataegis events
Length of kataegis foci (bp)
b
Mutational landscape (SNV)
Kataegis foci
C>T
C>G
C>A
T>A
T>C
T>G
c
Kataegis-foci in TpCpW context
d
Relative contribution
Kataegis foci
e
Relative contribution
Genome-wide
Mutational Signatures (Alexandrov et al.)
Age (Sig. 1 & 5)
AID/APOBEC (Sig. 2 & 13)
BRCA (Sig. 3)
Smoking (Sig. 4 & 29)
UV (Sig. 7)
MSI (Sig. 6, 15, 20, 21, 26)
DNA Polymerase η (Sig. 9)
DNA Polymerase ε (Sig. 10)
Alkylating agents (Sig. 11)
Aristolochic acid (Sig. 22)
Unknown aetiology
f
chr8
Intermutational distance (log10)
Representative sample
C>A
C>G
C>T
T>A
T>C
T>G
g
h
***
Absolute contribution (Genome-wide) APOBEC signatures (2 & 13)
258
386
No kataegis
Observed kataegis
CPCT-02 cohort (n=197)


TMB - SNV
TMB - InDels
TMB - MNV
SV
Absolute Frequencies
Relative Frequencies
AID/APOBEC (Sig. 2 & 13)
Age (Sig. 1 & 5)
Alkylating agents (Sig. 11)
Aristolochic acid (Sig. 22)
Deletions (<100kb)
Deletions (>100kb)
Insertions
Translocations
BRCA (Sig. 3)
DNA Polymerase ε (Sig. 10)
DNA Polymerase η (Sig. 9)
MSI (Sig. 6, 15, 20, 21, 26)
Tandem Dup. (<100kb)
Tandem Dup. (>100kb)
Inversions
Mean Genome Ploidy
Mean ploidy
Smoking (Sig. 4 & 29)
UV (Sig. 7)
Unknown aetiology

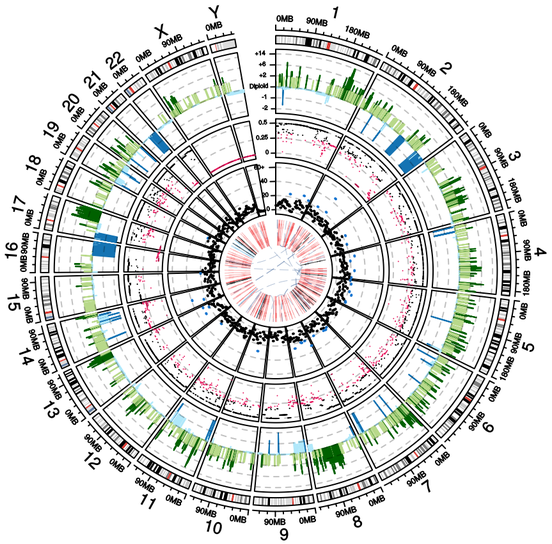

a
Optimal number of clusters
Total Within Sum of Square
600
400
200
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
Number of clusters k


b
25
20
15
10
5
0

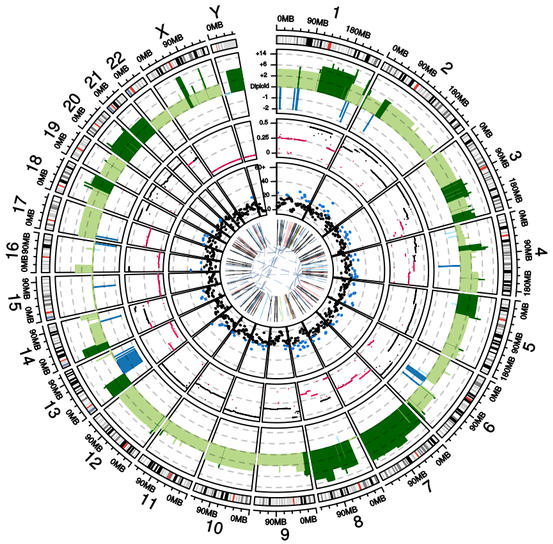
Cluster A
n = 13



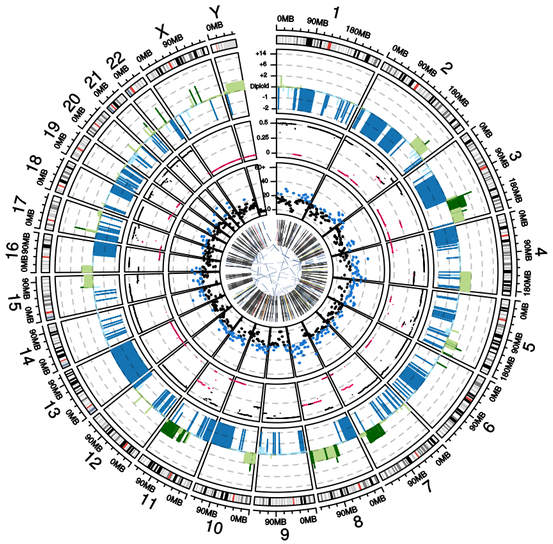
Cluster B
n = 13

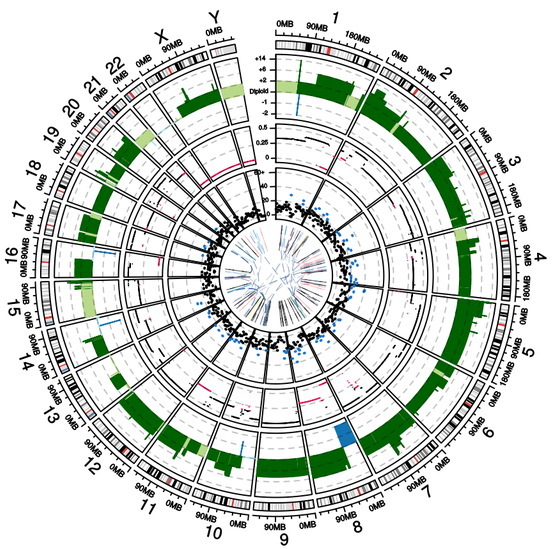


Cluster C
n = 15

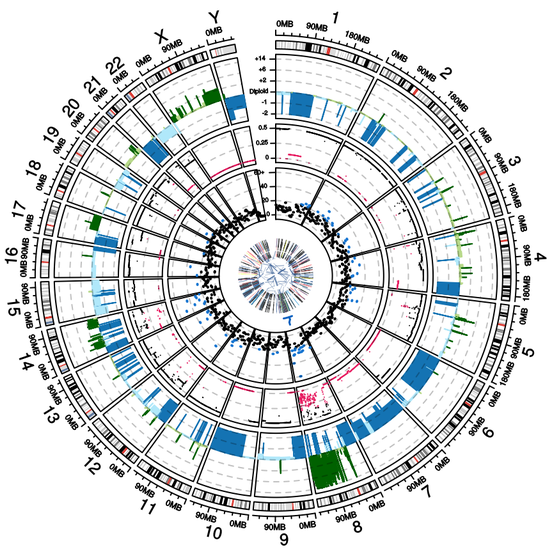


Cluster D
n = 22

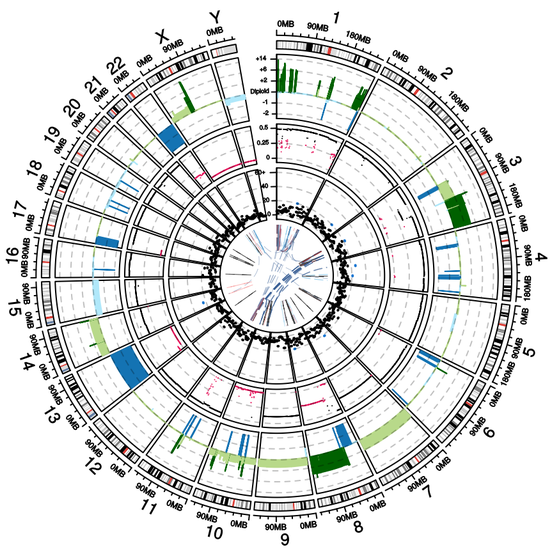




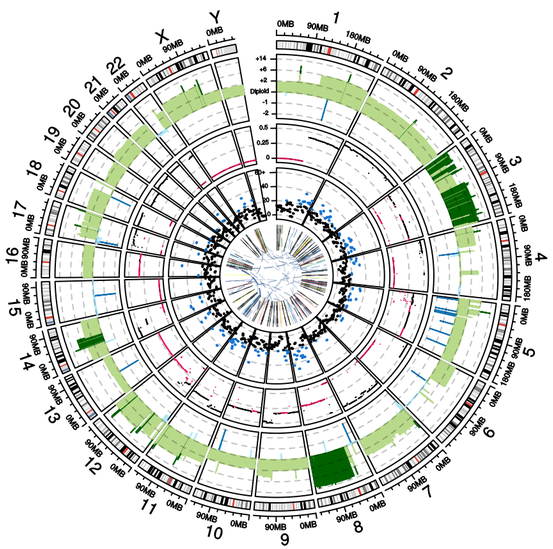
Cluster E
n = 55



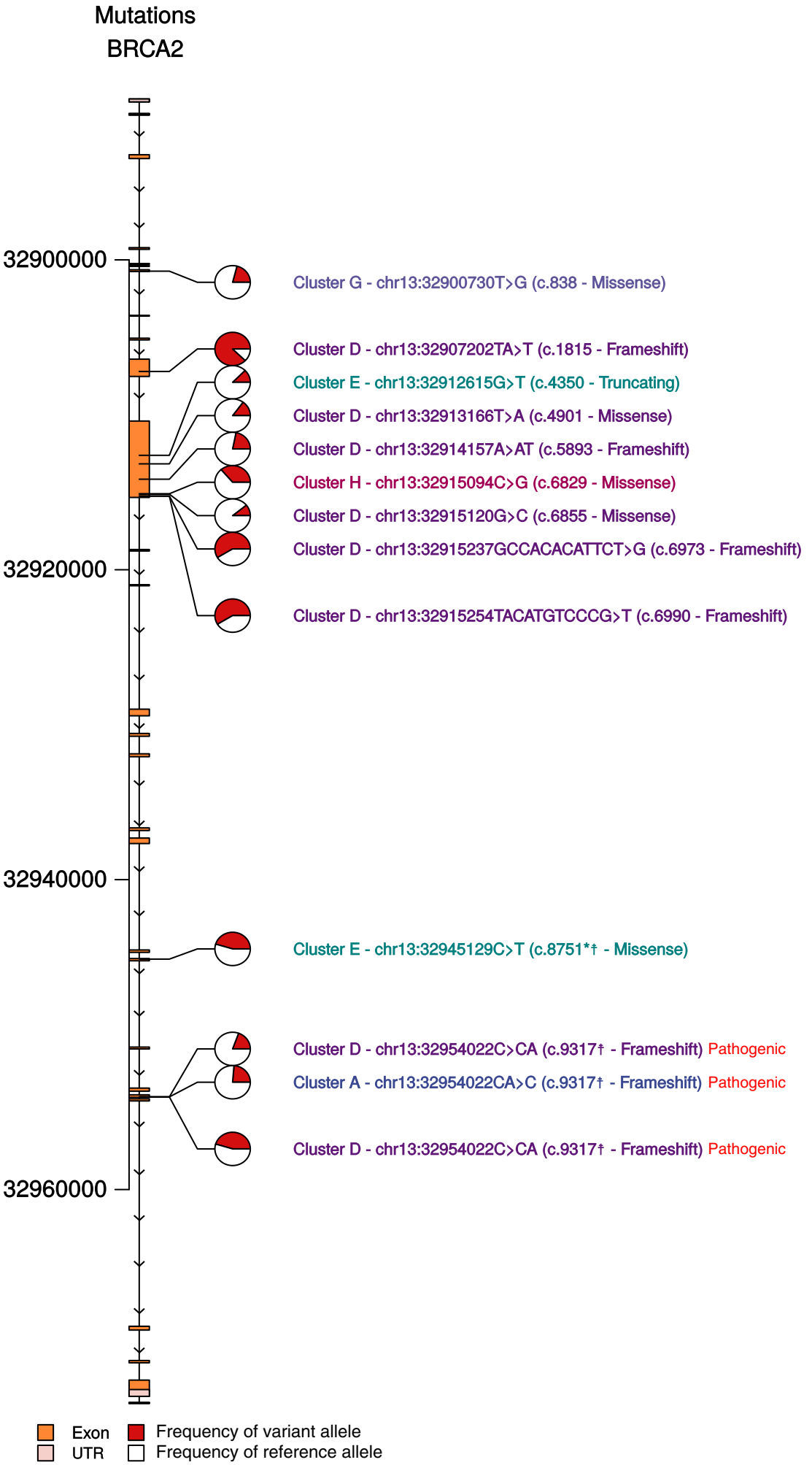
Cluster F
n = 20

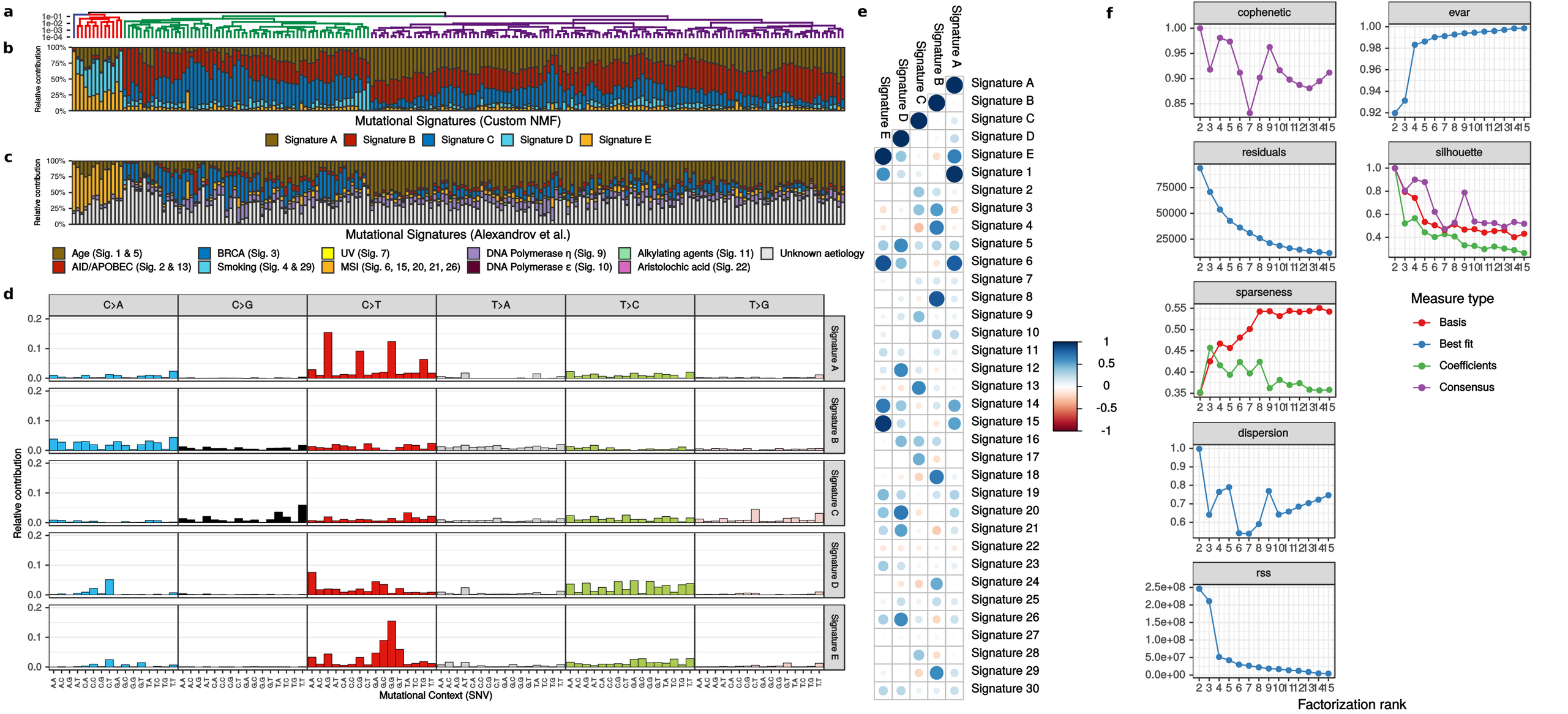


Cluster G
n = 34



Cluster H
n = 25


Mutations
BRCA2
32900000
32920000
32940000
32960000
Cluster G - chr13:32900730T>G (c.838 - Missense)
Cluster D - chr13:32907202TA>T (c.1815 - Frameshift)
Cluster E - chr13:32912615G>T (c.4350 - Truncating)
Cluster D - chr13:32913166T>A (c.4901 - Missense)
Cluster D - chr13:32914157A>AT (c.5893 - Frameshift)
Cluster H - chr13:32915094C>G (c.6829 - Missense)
Cluster D - chr13:32915120G>C (c.6855 - Missense)
Cluster D - chr13:32915237GCCACACATTCT>G (c.6973 - Frameshift)
Cluster D - chr13:32915254TACATGTCCCG>T (c.6990 - Frameshift)
Cluster E - chr13:32945129C>T (c.8751*† - Missense)
Cluster D - chr13:32954022C>CA (c.9317† - Frameshift) Pathogenic
Cluster A - chr13:32954022CA>C (c.9317† - Frameshift) Pathogenic
Cluster D - chr13:32954022C>CA (c.9317† - Frameshift) Pathogenic
Exon
UTR
Frequency of variant allele
Frequency of reference allele


a
1e-01
1e-02
1e-03
1e-04
b
Relative contribution
100%
75%
50%
25%
0%
Mutational Signatures (Custom NMF)
Signature A
Signature B
Signature C
Signature D
Signature E
c
Relative contribution
Mutational Signatures (Alexandrov et al.)
Age (Sig. 1 & 5)
AID/APOBEC (Sig. 2 & 13)
BRCA (Sig. 3)
Smoking (Sig. 4 & 29)
UV (Sig. 7)
MSI (Sig. 6, 15, 20, 21, 26)
DNA Polymerase η (Sig. 9)
DNA Polymerase ε (Sig. 10)
Alkylating agents (Sig. 11)
Aristolochic acid (Sig. 22)
Unknown aetiology
d
C>A
C>G
C>T
T>A
T>C
T>G
Relative contribution
Mutational Context (SNV)
e
Signature A
Signature B
Signature C
Signature D
Signature E
Signature 1
Signature 2
Signature 3
Signature 4
Signature 5
Signature 6
Signature 7
Signature 8
Signature 9
Signature 10
Signature 11
Signature 12
Signature 13
Signature 14
Signature 15
Signature 16
Signature 17
Signature 18
Signature 19
Signature 20
Signature 21
Signature 22
Signature 23
Signature 24
Signature 25
Signature 26
Signature 27
Signature 28
Signature 29
Signature 30
1
0.5
0
-0.5
-1
f
cophenetic
evar
residuals
silhouette
sparseness
dispersion
rss
Measure type
Basis
Best fit
Coefficients
Consensus
Factorization rank

Supplementary table 1: Participating centers

Organization	Local principal investigator
Radboud UMC, Nijmegen	Carla van Herpen
Erasmus MC, Rotterdam	Martijn Lolkema
Franciscus Gasthuis & Vlietland, Rotterdam	Paul Hamberg
NKI-AVL, Amsterdam	Neeltje Steeghs
Isala, Zwolle	Jan Willen de Groot
Martini Ziekenhuis, Groningen	Johan van Rooijen
Medisch Centrum Leeuwarden	Hiltje de Graaf
Maastricht UMC, Maastricht	Vivianne Tjan-Heijnen
Noordwest Ziekenhuisgroep, Alkmaar	Mathijs Hendriks
UMC Utrecht, Utrecht	Els Witteveen
Amphia Ziekenhuis, Breda	Bert Jan ten Tije
Reinier de Graaf Gasthuis, Delft	Annelie Vulink
Treant Zorggroep, Hoogeveen	Sophia van den Boogerd
Zuyderland Medisch Centrum, Geleen	Frans Erdkamp
ETZ Elisabeth, Tilburg	Laurens Beerepoot
Leids Universitair Medisch Centrum, Leiden	Hans Gelderblom
Maasstad Ziekenhuis, Rotterdam	Rineke Leys
Meander Medisch Centrum, Amersfoort	Haiko Bloemendal
St. Antonius Ziekenhuis, Utrecht	Maartje Los
VUmc, Amsterdam	Henk Verheul
ZGT, Almelo	Esther Siemerink

Included patients for this study (n)
91
38
20
15
3
3
3
3
3
3
2
2
2
2
1
1
1
1
1
1
1

lartini Ziekenhuis (n=3)
3)
)
ZGT (n=1)
2)

Supplementary table 2: Patient characteristics		
Patients (n=197)		
	n	**%**
Age at biopsy		
Median	68	
Range (min-max)	48-83	
Prior ADT		
Yes	197	100.0
Drug-based	181	91.9
Surgery-based (orchiectomy)	3	1.5
With Docetaxel	6	3.0
No clear documentation of ADT type	7	3.6
Prior systemic therapy (other than ADT)		
0 previous treatments	27	13.7
≥ 1 previous treatments	170	86.3
1 previous treatment	45	22.8
2 previous treatments	69	35.0
3 previous treatments	31	15.7
4 previous treatments	19	9.6
5 previous treatments	6	3.0
Type of prior systemic therapy (other than ADT)		
Hormonal therapy only	20	10.2
Chemotherapy only	37	18.8
Radionucleotide therapy only	4	2.0
Immunotherapy only (Dendritic cell therapy)	4	2.0
Targeted therapy only	0	0.0
Hormonal and chemotherapy	68	34.5
Hormonal and radionucleotide therapy	3	1.5
Chemotherapy and radionucleotide therapy	3	1.5
Hormonal and immunotherapy	3	1.5
Chemotherapy and immunotherapy	3	1.5
Hormonal, chemotherapy and radionucleotide therapy	15	7.6
Hormonal, chemotherapy and immunotherapy	4	2.0
Hormonal, radionucleotide and immunotherapy	2	1.0
Hormonal, chemotherapy and targeted therapy (Olaparib)	2	1.0
Hormonal, chemotherapy, radionucleotide and immunotherapy	1	0.5
Unknown at time of analysis	1	0.5
Prior radiotherapy		
Yes (curative radiotherapy of the prostate and/or palliative radiotherapy of metastases)	117	59.4
No	77	39.1
Unknown at time of analysis	3	1.5
Started therapy after biopsy for whole-genome sequencing		
Yes	138	70.1
Hormonal therapy	53	26.9

Chemotherapy	56	28.4
Radionucleotide therapy	12	6.1
Immunotherapy (Pembrolizumab)	6	3.0
Targeted therapy	3	1.5
Combinational therapy	6	3.0
Other*	2	1.0
No	19	9.6
Unknown at time of analysis	40	20.3
Biopsy site		
Liver	29	14.7
Lymph node	81	41.1
Bone	70	35.5
Lung	3	1.5
Soft tissue/Other**	14	7.1
*Boneregulating agent		
**Soft tissue/other: (sub)cutis, muscle, peritoneum, kidney, bladder, adrenal gland		

Sheet G - Exclusive mutations

Gene	Cluster	mutationsInCluster	nomutationsInCluster	mutationsInOtherCluster
CDK12	Cluster B	11	2	5
BRCA2	Cluster D	15	7	12
MSH6	Cluster A	9	4	4
Chromoth	Cluster F	16	4	30
JAK1	Cluster A	9	4	11
FGF3	Cluster B	9	4	11
FGF4	Cluster B	9	4	11
MDM4	Cluster B	10	3	21
CIC	Cluster A	7	6	5
CCND1	Cluster B	9	4	16
EPAS1	Cluster A	7	6	6
MSH2	Cluster A	7	6	7
SLC45A3	Cluster B	9	4	18
FBXO11	Cluster A	6	7	5
ELK4	Cluster B	9	4	22
NSD1	Cluster B	6	7	6
BCL2L12	Cluster B	5	8	3
FGFR4	Cluster B	5	8	3
HIST2H2I	Cluster B	8	5	18
ASXL1	Cluster A	7	6	12
SIX2	Cluster A	6	7	7
ZFP36L2	Cluster A	8	5	19
NBN	Cluster B	11	2	48
ARID1A	Cluster A	6	7	8
ERF	Cluster A	6	7	8
MLH1	Cluster A	6	7	8
CNTNAP2	Cluster B	9	4	28
POTEE	Cluster A	5	8	4
SETD1B	Cluster A	7	6	14
TBX3	Cluster B	6	7	9
FRS2	Cluster B	7	6	15
ACVR1	Cluster A	5	8	5
CHD2	Cluster A	5	8	5
KMT2B	Cluster A	5	8	5
PBRM1	Cluster A	5	8	5
ZFHX3	Cluster A	9	4	31
ECT2L	Cluster B	5	8	5
PCBP1	Cluster A	4	9	2
IDH2	Cluster B	4	9	2
BCL11A	Cluster A	6	7	11
TCF7L2	Cluster A	5	8	6
CDK4	Cluster B	5	8	6
IL2	Cluster B	5	8	6
RAB35	Cluster B	5	8	6
KMT2C	Cluster A	8	5	26
EML4	Cluster A	6	7	12
KMT2D	Cluster A	6	7	12
ACVR2A	Cluster A	5	8	7
BAX	Cluster B	5	8	7
EREG	Cluster B	5	8	7

Sheet G - Exclusive mutations

nomutationsInOtherCluster	p	p.adj
179	9.91E-13	7.29E-09
163	1.63E-10	5.99E-07
180	4.62E-10	1.13E-06
147	1.84E-08	3.39E-05
173	9.44E-08	9.93E-05
173	9.44E-08	9.93E-05
173	9.44E-08	9.93E-05
163	4.77E-07	4.39E-04
179	5.75E-07	4.70E-04
168	1.04E-06	7.63E-04
178	1.21E-06	8.10E-04
177	2.35E-06	1.33E-03
166	2.28E-06	1.33E-03
179	8.98E-06	4.42E-03
162	9.00E-06	4.42E-03
178	1.74E-05	8.00E-03
181	2.76E-05	1.07E-02
181	2.76E-05	1.07E-02
166	2.64E-05	1.07E-02
172	2.99E-05	1.09E-02
177	3.12E-05	1.09E-02
165	3.66E-05	1.22E-02
136	3.89E-05	1.24E-02
176	5.29E-05	1.44E-02
176	5.29E-05	1.44E-02
176	5.29E-05	1.44E-02
156	4.87E-05	1.44E-02
180	5.99E-05	1.57E-02
170	6.50E-05	1.65E-02
175	8.53E-05	2.09E-02
169	9.25E-05	2.20E-02
179	1.16E-04	2.30E-02
179	1.16E-04	2.30E-02
179	1.16E-04	2.30E-02
179	1.16E-04	2.30E-02
153	1.00E-04	2.30E-02
179	1.16E-04	2.30E-02
182	1.63E-04	3.08E-02
182	1.63E-04	3.08E-02
173	1.97E-04	3.42E-02
178	2.04E-04	3.42E-02
178	2.04E-04	3.42E-02
178	2.04E-04	3.42E-02
178	2.04E-04	3.42E-02
158	2.49E-04	4.07E-02
172	2.86E-04	4.49E-02
172	2.86E-04	4.49E-02
177	3.38E-04	4.97E-02
177	3.38E-04	4.97E-02
177	3.38E-04	4.97E-02

Test used:
Two-sided Fisher's Exact Test