Racial differences in genomic features and new trans-ancestry prognosis subtypes in bladder cancer

Baifeng Zhang (zhangbaifeng1@gmail.com)
The University of Hong Kong https://orcid.org/0000-0003-1966-2285
Peilin Jia
Jiayin Wang
Xi'an Jiaotong University
Guangsheng Pei
The University of Texas Health Science Center at Houston, Houston https://orcid.org/0000-0003-1804-7598
Changxi Wang
Shimei Pei
Xiangchun Li
Tianjin Medical University Cancer Institute and Hospital
Shuai Ding
Zhongming Zhao
The University of Texas Health Science Center at Houston https://orcid.org/0000-0002-3477-0914
Xin Yi
Yi Huang

Article

Keywords:

Posted Date: December 27th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1157888/v1

Racial differences in genomic features and new trans-ancestry prognosis subtypes in bladder cancer

Baifeng Zhang[1,2,3*], Peilin Jia[4], Jiayin Wang[5], Guangsheng Pei[4], Changxi Wang[3], Shimei Pei[3], Xiangchun Li[6], Shuai Ding[7], Zhongming Zhao[4], Xin Yi[3], Yi Huang[3,5,8*]

[1]Departments of Clinical Oncology, The University of Hong Kong-Shenzhen Hospital, Shenzhen, 518053, China

[2]Departments of Clinical Oncology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China

[3]Geneplus-Beijing, Beijing 102206, China

[4]Center for Precision Health, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA

[5]Department of Computer Science and Technology, School of Electronic and Information Enginering, Xi'an Jiaotong Unviersity, Shaanxi 710049, China

[6]Department of Epidemiology and Biostatistics, Tianjin Medical University Cancer Institute and Hospital, Tianjin 300060, China

[7] School of Management, Ministry of Education Key Laboratory of Process Optimization and Intelligent Decision-Making, Hefei University of Technology, Hefei, 230009, China

[8]Luohu people's hospital, Shenzhen, 518000, China

*Correspondence: zhangbaifeng1@gmail.com;huangyi@geneplus.org.cn

ABSTRACT

The incidence and survival of bladder cancer vary greatly among different populations but the influence of the associated molecular features and evolutionary processes on its clinical treatment and prognostication remains unknown. Here, we analyze the genomic architectures of over 500 bladder cancer patients from Asian/Black/White populations. We identify novel association between *AHNAK* mutations and APOBEC-a mutational signature whose activities vary substantially across populations. All significantly mutated genes but only half of arm-level somatic copy number alterations (SCNAs) are enriched with clonal events, indicating large-scale SCNAs as rich sources fostering bladder cancer clonal diversities. The prevalence of *TP53* and *ATM* clonal mutations as well as the associated burden of SCNAs is significantly higher in Whites/Blacks than in Asians. We identify a trans-ancestry prognostic subtype of bladder cancer characterized by: enrichment of non-muscle-invasive patients and muscle-invasive patients with good prognosis, increased *CREBBP/FGFR3/HRAS/NFE2L2* mutations, decreased intra-tumor heterogeneity and genome instability and activated tumor microenvironment.

INTRODUCTION

Bladder cancer is the sixth most common type of adult cancer with almost 430, 000 new cases diagnosed per year[1]. Bladder cancer is classified into non-muscle invasive bladder cancer (NMIBC) and muscle-invasive bladder cancer (MIBC) according to their histopathological features and clinical behaviors. NMIBC, the major subtype of bladder cancer that accounts for 70% to 80% of all newly diagnosed cases, has better overall survival but high recurrence rate of 31-78% while the rest of patients with muscle-invasive or

metastatic diseases exhibit quite low survival rate[2,3]. The genetic or molecular basis underlying such great disparities in clinical outcomes between NMIBC and MIBC remains largely unclear.

NMIBC and MIBC are supposed to develop through different evolutionary paths on distinct genetic background. The genetic architecture of MIBC has become increasingly clear with the accomplishment of several large-scale genome sequencing projects in large cohorts of MIBC patients[4–6]. Characterization of the genetic causes of NMIBC by large-scale approaches had only been performed in a relatively small number of patients[7,8]. These studies identified a number of new driver events, such as truncating mutations of the chromatin remodeling genes (*UTX*, *MLL-MLL3*, *ARID1A* and others) and recurrent fusions involving *FGFR3* and *TACC3*, in addition to the well-known TP53-RB1 and FGFR3-Ras pathways that are mutated frequently in MIBC and NMIBC, respectively[7,8]. Bladder cancer is a heterogeneous group of diseases with distinct subtypes exhibiting different mutational signatures and expression features[4]. However, these studies primarily focused on the somatic events with elevated mutation frequencies among different patients or subtypes[5–8].

The degrees of intra- and inter-tumor heterogeneities can be assessed by analyzing the genomic architectures of individual bladder cancer patients. One previous study analyzing the clonality of MIBC in a small number of pre and post-chemotherapy samples revealed that the extensive and dynamic clonal evolution found in MIBC patients may lead to chemotherapy resistance[9]. It was expected that the genomic architectures would have great influence on the clinical outcomes of bladder cancer and somatic driver events acquired at different evolutionary stages would also have different prognostic values. However, few of the genetic prognostic biomarkers for bladder cancer had been developed utilizing their clonal or sub-

clonal states in large cohorts of patients with different genetic background, despite of the well-documented disparities in bladder cancer risk and survival across races/ethnics in epidemiological studies[10]. Whites are more than twice likely to develop bladder cancer than Blacks while Asian Americans have lower risk of developing bladder cancer than Whites and Blacks[11]. The clinical outcomes of bladder cancer also exhibit racial disparities with shorter survival observed in Blacks[10]. Moreover, these previous prognostic studies also focused only on single level of molecular data and no study had characterized the overall genomic architectures of individual bladder cancer patients based on both large-scale somatic copy number alterations (SCNAs) and single nucleotide variants (SNVs) simultaneously to stratify distinct prognostic subtypes. In the study, we tried to investigate that with disparities in genetic background, life history and mutagens exposure, whether bladder cancer patients from different populations might show different mutational signatures and therefore further lead to discrepancies in mutational features and genomic architectures.

To more comprehensively understand the genetic diversities of bladder cancer, we collect large-scale genomic data from The Cancer Genome Atlas (Cohort 1: TCGA-BLCA)[4,6] and the Chinese population (Cohort 2: Chinese-BLCA)[7,8]. Cohort 1 consists of MIBC patients from three different racial populations (White, Black and Asian) and Cohort 2 consists of both MIBC and NMIBC patients from Asia only. We use the MIBC or NMBIC patients from Cohort 2 separately or jointly to test or validate some of the findings observed in Cohort 1. In total, our study cohorts include 468 MIBC and 37 NMIBC patients. We compare the differences in mutational signatures among Asian, Black and White populations and further characterize the association between driver gene mutations and the activities of mutational signatures. We perform prognostic analysis of MIBC in Cohort 1 patients with survival information available and further cluster the MIBC and NMBIC patients according to their

molecular patterns. After quantifying the cancer cell fractions (CCFs) of all potential prognostic somatic events, we categorize the bladder cancer patients into distinct prognostic subtypes showing differences in intratumor heterogeneity, genome instability, metastatic ability and immune features.

RESULTS

Racial differences in mutational signatures

To dissect the somatic mutation landscapes of the 505 bladder cancer patients in our study, Mutect2[12] and ReCapSeg[13] were used to call somatic mutations and allelic SCNAs with the whole-exome sequencing data. After further filtering steps (Online Methods), 136, 346 somatic mutations (132, 100 SNVs and 4, 246 short insertions/deletions (Indels)) and 21, 779 SCNA segments were generated for downstream analysis. We then inferred the CCFs of SNVs and SCNAs simultaneously for each patient. By combining all somatic events from all patients, 11.2% of SNVs, 30.5% of Indels and 41.1% of SCNAs were identified as sub-clonal events (Supplementary Table 1).

The mutational signatures document the evolutionary history of interactions between multiple exogenous and endogenous mutational processes which consequently result in the accumulation of somatic mutations in tumor cells. Bladder cancer patients from different populations may have distinct genetic background, life history or mutagen exposure. To investigate whether there were any racial differences in mutational signatures during bladder cancer evolution, we first divided our study cohorts into three populations of Asian (140), Black (23) and White (324) origins. We extracted the mutational signatures from SNVs with Bayesian non-negative matrix factorization (NMF). We identified five signatures in Asian,

two signatures in Black and four signatures in White (**Fig. 1a** and **Supplementary Fig. 1**). Through clustering analysis of the cosine similarities and Pearson correlation values between these signatures and the 30 Sanger signatures with reported roles in mutagenesis[14–16], the five mutational signatures identified in our cohorts could be grouped into categories resembling the APOBEC-a, APOBEC-b, ERCC2 & AGE, ARISTOLOCHIC and MMR Sanger signatures (cosine similarities between 0.68 and 0.97, **Fig. 1a**, **Supplementary Fig. 2** and **Supplementary Data 1**). APOBEC-a and APOBEC-b signatures, accounting for the majority of mutations in bladder cancer, appeared to be more likely to occur early during bladder cancer development[4]. The ERCC2 & AGE signature included one signature mediated by *ERCC2* mutations[17] and another signature correlated with the age of cancer diagnosis. ARISTOLOCHIC signature was related to exposures to aristolochic acid. The last mutational signature showed relatively low similarity (cosine similarity of 0.68) to the MMR signature which had not been reported in bladder cancer previously.

Of these five signatures, the activities of both APOBEC-a and ERCC2 & AGE signatures were the highest in White and the lowest in Asian, regardless of the clonal/sub-clonal states of mutations ($\boldsymbol{P} < 0.001$ & FDR< 0.1, **Fig. 1b**). The ARISTOLOCHIC and MMR-like signatures occurred only in Asian and were both enriched in the sub-clonal state ($\boldsymbol{P} < 0.001$ & FDR< 0.1, **Fig. 1b** and **Supplementary Fig. 3**). These observations might suggest the presence of population variations in mutational processes operating at different evolution stages during bladder cancer development. To further assess the potential clinical association of these mutational signatures, we performed unsupervised clustering of the patients in Cohorts 1 and 2 according to their mutational signature activities and identified four mutational signature clusters (MSig1 to MSig4, **Fig. 1c**). Different clusters of patients showed varied activities of different mutational signatures (**Fig. 1d**) and were associated with

distinct clinical outcomes (**Fig. 1e**, ***P*** = 0.002). Cluster MSig2 showed enrichment in Asian patients (**Fig. 1c**) and had high MMR-like signature activity (***P*** < 0.001 & FDR< 0.1, **Fig. 1d**). Prognostic analysis of Cohort 1 revealed that MIBC samples in cluster MSig2 tended to have the shortest overall survival (**Fig. 1e**) and this observation had to be further investigated in future studies as the number of samples with survival information available in cluster MSig2 was quite small in our study. Patients in the MSig3 and MSig4 clusters had high levels of APOBEC and *ERCC2* & AGE signature mutagenesis (**Fig. 1d**), respectively. Multivariate Cox regression analysis of Cohort 1 patients showed that those patients in the MSig3 and MSig4 clusters had better survival than those belonging to MSig1 cluster (**Fig. 1f**).

The Asian cohort in our study was consisted of patients from both Europe/USA areas (Cohort 1: MIBC) and Asia areas (Cohort 2: MIBC and NMIBC) and we compared the relative contribution of the mutational signatures among these three Asian patient groups and found that the activities of APOBEC-a and -b signatures were relatively higher in MIBC patients from Cohort 2 than the other two groups of Asian patients. Both MIBC and NMIBC patients from Cohort 2 showed significantly high levels of MMR-like signature activities compared with MIBC samples from Cohort1 (**Supplementary Fig. 3**). Nevertheless, more studies and samples were needed to validate the observed racial differences in the activity of MMR-like signature and we could not rule out the potential influence of study batch effect on this observation due to the following reasons: (i) the MMR-like signature with minor exposures could only be observed in Cohort 2 generally; (ii) most of the mutations contributed to the MMR-like signature were subclonal; (iii) the MMR-like signature had a relatively low level of similarity to any known COSMIC signatures; and (iv) little previous evidence suggested the presence of this signature in bladder cancer (**Supplementary Fig. 4**).

176

177 **Gene mutations potentially associated with signature activities**

178 To explore whether there were any genes whose mutations were associated with the
179 mutational signature activities, we performed signature enrichment analysis for those genes
180 mutated at elevated frequencies with permutation tests which could control the inflation that
181 an increase in the overall mutation burden in each sample or in the total number of mutations
182 in each gene usually correlated with increased signature activity. To determine the
183 significance level of associations between the elevated mutation frequency of each gene and
184 the increased activity of each signature, the mutation burden attributed to a given signature
185 from tumors that harbored non-silent mutations in the gene was compared with tumors that
186 did not by combining the two study cohorts (Online Methods).

187

188 Consistent with previous reports[17], we showed that *ERCC2* mutation state and smoking were
189 significantly correlated with ERCC2 & AGE signature activity (FDR < 0.1, $P < 0.001$, **Fig**
190 **2a, b**). Interestingly, we also found that *ERCC2* mutation state was correlated with the
191 activity of APOBEC-b signature (FDR < 0.1), with an increase of 75 APOBEC-b mutations
192 in *ERCC2*-mutant patients (**Fig. 2b**). Notably, we identified that the APOBEC-a signature
193 activity, an indicator of good clinical outcome in bladder cancer[4], was associated with
194 *AHNAK* mutations which appeared to act as a tumor suppressor through potentiating the
195 TGFβ signalling in a previous study at the protein level[18] (**Fig. 2a**). The median numbers of
196 APOBEC-a mutations in patients with and without non-silent *AHNAK* mutations were 235
197 and 62, respectively (**Fig. 2b**, $P < 0.001$). Correlation between *HRAS* mutation states and the
198 MMR-like signature activity was also noticed in our study (**Fig. 2a, b**).

199

We further validated that whether racial differences in mutation frequencies of several key genes or pathways may partially account for the racial differences in mutational signature activities among Asian/Black/White populations by combining Cohorts 1 and 2. The frequency of *AHNAK* mutations was the highest in the White patients who also showed the highest APOBEC-a signature activity (9.6% in White vs 4.3% in Black & 6.4% in Asian) while the frequency of *HRAS* mutations was the highest in the Asian patients (14.3% in Asian vs 0.0% in Black & 3.4% in White) (**Fig. 2c**). At the pathway level, genes involved in the DNA polymerase pathway and other conserved DNA damage response genes also had the highest mutation frequencies in White population (**Supplementary Fig. 5a**).

To determine whether low-frequency mutations in genes involved in the DNA damage responses were linked to mutational signature activities, we performed permutation tests for the DNA damage response pathways from curated databases (Online Methods). We observed that mutations of the nucleotide excision repair (NER) pathway which includes the *ERCC2* gene were significantly associated with the activities of both ERCC2 & AGE and APOBEC-b signatures (**Supplementary Fig. 5b**). Mutations in the DNA polymerase pathway were identified to be associated with the APOBEC-a signature activity. Additionally, mutations of a group of genes defined as other conserved DNA damage response genes were significantly associated with the mutagenesis of three signatures, APOBEC-a, APOBEC-b and ERCC2 & AGE (**Supplementary Fig. 5b, c**).

We next compared the expression levels of *AHNAK*, *ERCC2* and *HRAS* among the three racial populations in Cohort 1. We observed that the expression levels of both *AHNAK* and *ERCC2* were significantly lower in Asian than those in White population ($P < 0.05$ and $P < 0.001$, respectively) (**Supplementary Fig. 6)**, which provided further evidence supporting

that racial differences in the activities of *AHNAK* and *ERCC2* might contribute to the observed racial differences in the prevalence of APOBEC-a and ERCC2 & AGE signatures in bladder cancer. However, *HRAS* expression showed no significant difference among different race groups (**Supplementary Fig. 6)**.

Racial disparities in cancer gene mutations and SCNAs

To depict the racial disparities in molecular features among different racial groups, we first identified the significantly mutated genes (SMGs) by combining all samples from the two study cohorts and then compared the prevalence of mutations in SMGs among different race groups. A total of 57 SMGs were identified by MutSig2CV ($q < 0.1$) (**Supplementary Data 2)** and 26 of them were cancer census genes that were expressed in over 75% of TCGA patients (supported by at least three read counts) and were kept for downstream analysis[19]. Five of the 26 SMGs had not been reported as SMGs in bladder cancer in previous studies[6,20–22]: *CDKN1B* (1.2%), *ITK* (1.8%), *PRDM2* (3.1%), *FOXA1* (2.7%) and *BAP1* (1.8%). A total of 34 frequent arm-level SCNAs occurred in at least 30% of patients by combining the two study cohorts.

The prevalence of clonal mutations in serval SMGs varied substantially among different race groups. Both Black and White patients showed higher fractions of clonal *TP53* mutations than Asian patients from Cohorts 1 and 2. The frequencies of clonal mutations in *HRAS* in Asian patients from Cohorts 1and 2 were consistently higher than Whites. Compared with Asian patients, clonal mutations in *ATM* and *ARID1A* showed higher prevalence in Blacks and Whites, respectively. The frequency of *NRAS* mutations was higher in Blacks than Whites and three genes, including *FGFR3, ELF3* and *TSC1,* showed higher frequencies of clonal mutations in Whites than in Asian patients from Cohort 2. Interestingly, *FGFR3* clonal

mutations were observed to occur more frequently not only in Asian patients in Cohort 1 than in Cohort 2. *FGFR3* mutations are the early driver events during NMIBC evolution and we also observed that *FGFR3* clonal mutations occurred more frequently in NMIBC than in MIBC in Cohort 2 patients[23] (**Fig. 3a, b**).

We next compared the overall burden of SCNAs at the chromosome/arm/focal levels across different racial groups (Online Methods). NMIBC patients generally had a lower burden of SCNAs at the chromosome/arm levels than the MIBC patients. The overall arm- or chromosome-level SCNA scores in Asian MIBC patients were lower than those in White or Black MIBC patients (**Fig. 3c**). This observation was consistent with the finding of low prevalence of clonal mutations in the *TP53* and *ATM* gatekeeper genes governing DNA repair in Asian MIBC patients. We further categorized each chromosome/arm level SCNA into clonal or subclonal event and found that the majority of arm-level deletions either in clonal or subclonal states showed substantial inter-population variations in their prevalence. A large number of chromosomes/arms showed higher prevalence of clonal deletions in White/Black MIBC patients than in the Asian MIBC patients. We also observed a large number of chromosomes/arms showing higher prevalence of subclonal SCNAs in Whites than in Asian patients from Cohort 2 (**Fig. 3d**). Moreover, we observed that only quite a few number of clonal arm-level SCNAs, including del(3p) and amp(20), had higher prevalence but a number of subclonal SCNAs showed significantly lower prevalence in Asian MIBC patients from Cohort 2 than in Asian patients from Cohort 1 (**Fig. 3e**).

It has been reported that bladder cancer patients with different genetic backgrounds may be presented with different disease stages at the time of diagnosis[10]. In our study cohorts, we also observed that MIBC patients of Asian origins had higher prevalence of low-stage disease

while Blacks and Whites had higher prevalence of high-stage disease (**Supplementary Fig. 7**). To correct for the potential influence of heterogeneities in disease stages on the racial disparities in genomic features of MIBC, we further analyzed the racial differences in the mutation landscapes and SCNA profiles in low-staged (staged I & II) MIBC patients due to the lack of sufficient numbers of high-staged patients in the Asian cohorts. Again, the prevalence of clonal mutations in *TP53* and *ATM* was significantly higher in Whites and Blacks than in Asians, respectively, among MIBC patients with low-stage disease (**Supplementary Fig. 8**). The prevalence of clonal mutations in *NFE2L2* was the highest in low-staged Black patients and the fractions of White patients harboring clonal mutations in *FGFR3*, *TSC1* and *CDKN2A* were significantly higher than those of Asian patients from Cohort 2. Clonal mutations in *HRAS* also showed significantly higher prevalence in Asians than in Whites among the low-staged patients. We also confirmed that the prevalence of *FGFR3* clonal mutations was significantly higher in Asians from Cohort 1 than in Asians from Cohort 2 among the low-stage MIBC patients. The trends of higher prevalence of clonal deletions and subclonal SCNAs in Whites or Blacks were also observed among patients with low stage MIBC. Differences in the frequencies of clonal deletions of several chromosome arms including del(3p) and subclonal SCNAs affecting a number of other arms were also observed between Asians from Cohorts 1 and 2 among the low-staged patients.

Genomic architectures and subtypes of bladder cancer

Our study included a large number of bladder cancer patients with different genetic backgrounds and we then tried to perform molecular subtyping of bladder cancer by integrating multiple somatic events according to their CCFs in all samples by combining Cohorts 1 and 2. We inferred the genomic architectures of bladder cancer by focusing on the 26 SMGs with reported roles in the Cancer Gene Census database and the 34 frequent arm-

level SCNAs altered in at least 30% of patients as defined above. To explore the contribution of these SMGs and frequent arm-level SCNAs during bladder cancer evolution, we estimated the CCF of each SMG and SCNA in each sample. Overall, the median CCFs of SMGs were higher than those of SCNAs by combining the two study cohorts (**Fig. 4a**). All the 26 SMGs were found to be enriched with clonal non-silent mutations (FDR < 0.1) while only about 50% (17/34) and 15% (5/34) of the frequent arm-level SCNAs were significantly enriched with clonal and sub-clonal events (FDR < 0.1), respectively (**Fig. 4a**). These data suggested that heterogeneities in the genomic architectures of bladder cancer were more likely to arise from diversity in the clonity of arm-level SCNAs.

Few previous studies evaluated the prognostic values of individual somatic events according to their clonal or subclonal states in bladder cancer. Nevertheless, somatic alterations affecting each individual SMG and SCNA usually could only be detected in a minor fraction of samples and stratification of them into clonal or subclonal events would further reduce the numbers of patients harboring alterations in prognosis analysis. We did not observe significant differences in the clinical outcomes of different racial groups in Cohort 1, which may be partially due to the lack of sufficient number of Asian and Black patients with their survival information available (**Supplementary Fig. 9**). Thus, we further evaluated the prognostic value of the SMGs and frequent SCNA events according to their clonal or sub-clonal states by ignoring the racial origins of the patients to increase the sample sizes. We combined the different racial groups in Cohort 1 together in Kaplan-Meier analysis and multivariate Cox regression analysis. In total, we identified only four individual events with potential prognostic value in Cohort 1 (**Table 1, Supplementary Data 3**). Cohort 1 patients with *KRAS* mutations, all of which were clonal, had poor clinical outcome while clonal

deletions of either 21q or 12p and sub-clonal deletions of 18p were associated with shortened overall survival in Cohort 1.

We next tried to stratify whether there were any subtypes of bladder cancer whose overall genetic architectures had great influence on their clinical outcomes. We performed NMF analysis of the CCFs of the 26 SMGs and 34 arm-level SCNAs by combining the two study cohorts. In total, we identified two molecular subgroups (termed as clusters A and B), whose clinical outcomes were divergent. Because long-term survival information was not available for the patients from Cohort 2, we performed survival analysis for all the MIBC samples in the TCGA dataset. TCGA MIBC patients in cluster A had worse survival rate than those in cluster B (**Fig. 4b**) (HR = 1.64, 95% CI: 1.14 ~ 2.36; $P = 0.007$). Multivariate analysis further proved that the genomic architectures could predict the clinical outcomes of bladder cancer patients independently by adjusting for age, gender, TNM stage and mutational signature (**Fig. 4c**). Further comparison of these two genomic subtypes with the molecular subtypes defined by TCGA at different expression levels showed that genomic cluster B tended to be clustered with the molecular subtypes mRNA_Luminal_papillary, lncRNA_3, miRNA_3 and RPPA_1 in Cohort 1 patients[20] (**Supplementary Fig. 10**). It had been shown previously that bladder cancer patients exhibiting the Luminal_papillary signature at the mRNA level had low risk of disease progression[20]. Patients belonging to either the miRNA subtype 3 which was correlated with the lncRNA_3 signature or the RPPA cluster 1 showed the best survival in previous TCGA analysis[20]. Overall, our analysis at the genomic level provided further evidence complementary to the previously defined molecular subtypes of bladder cancer at the expression levels.

Molecular features of prognostic subtypes defined by genomic architectures

To understand which factors contributed to the distinct prognosis of bladder cancer subtypes defined by genomic architectures, we compared the prevalence of 60 potential driver events (26 SMGs and 34 arm-level SCNAs) between clusters A and B. We showed that four SMGs (*CREBBP*, *FGFR3*, *HRAS* and *NFE2L2*) but none of the frequent arm level SCNAs had significantly higher prevalence of genetic aberrations in cluster B patients who had better prognosis than in cluster A patients (FDR < 0.1, **Fig. 4a** and **Supplementary Data 4**). On the other hand, three SMGs (*TP53*, *RB1* and *PRDM2*) and about 91% of the frequent arm level SCNAs were enriched with alterations in cluster A patients who showed poor prognosis (FDR < 0.1, **Fig. 4a** and **Supplementary Data 4**). Notably, of the SMGs mutated more frequently in cluster B than in cluster A, three showed race-specific enrichment of clonal mutations among the low-staged patients with *FGFR3*, *HRAS* and *NFE2L2* mutations being enriched in Whites/Asians from Cohort 2, Asians from the two study cohorts and Blacks, respectively (**Fig. 4a)**. These observations indicated that trans-ancestry analysis of the genomic architecture of bladder cancer may be effective in screening out the trans-ancestry prognostic subtypes.

We inferred the potential temporal ordering of the acquirements of SMGs and frequent arm-level SCNAs in clusters A and B patients during bladder cancer evolution (Online Methods). Among cluster A patients, 25 of 26 SMGs and 21 of 34 arm-level SCNAs were acquired at the early and intermediate/late evolution stages, respectively (**Fig. 4a)**. Among cluster B patients, 17 of 24 SMGs and 27 of 34 arm-level SCNAs were acquired at early and intermediate/late evolution stages, respectively (**Fig. 4a)**. These data further supported that the clonal diversities of both clusters A and B patients were largely contributed by the large-scale SCNAs acquired at late evolution stages. Several SCNAs, including del(9), del(17p),del(8p) and del(11p), were early clonal events shared by patients in both clusters A

and B while arm-level SCNAs, including del(4q), del(15q), del(10q),del(11q) and del(12), were acquired at the early and late stages in clusters A and B, respectively (**Fig. 4a)**.

To explore whether there were any other differences in the overall genomic landscapes between clusters A and B patients, we firstly compared their non-silent somatic mutation burdens and found no significance difference. The extent of intra-tumor heterogeneities as measured by MATH values and Shannon entropy of CCFs in cluster A patients was significantly higher than that of cluster B patients ($P< 0.001$, **Fig. 4a and Supplementary Fig. 11**). We next estimated the overall burden of SCNAs at the chromosome/arm/focal levels for each patient and showed that cluster A patients had significant higher burden of large-scale SCNAs (arm & chromosome levels) but not focal SCNAs ($P < 0.001$, **Fig. 4d**). Whole-genome doubling (WGD) events were prevalent in cluster A and patients with Asian origin were enriched in cluster B (both $P < 0.001$). To be more specific, although our cohorts included only a small number of NMIBC patients from Cohort 2, cluster B were enriched with Cohort 2 NMIBC patients (51.0% in cluster B vs. 22.9% in cluster A, $P < 0.005$) and Cohort 1 MIBC patients who were less likely to develop metastatic disease (52.6% in cluster B vs. 33.3% in cluster A, $P < 0.001$, **Fig. 4a**). Further analysis of the overall survival for Cohort 1 MIBC patients showed that patients in cluster A had worse survival rate than those in cluster B after adjusting for WGD, racial origins as well as other clinical and genomic parameters (**Fig. 4c**). These observations suggested that it was feasible to distinguish the heterogeneous MIBC patients into two subtypes with distinct clinical outcomes based on their inherited genomic architectures.

To further study whether high intra-tumor heterogeneity and chromosomal instability of cluster A patients had influence on the tumor microenvironment, we compared the relative

expression levels of the immune signature genes among two genomic prognostic clusters using ssGSEA (**Fig. 4e**). In Cohort 1, the infiltrating levels of Th17 cells and CD8+ T cells were observed to be significantly higher in cluster B patients ($P < 0.001$ and FDR < 0.1, **Fig. 4e**). Previous studies proved that Th17 cells as subsets of T helper lymphocytes can mediate the antitumor immune responses through interacting with effector CD8+ T cells[24]. Cluster A patients had significantly higher levels of infiltrating Th2 cells and Macrophages (Both $P < 0.001$ and FDR < 0.1, **Fig. 4e**), which had been proved to facilitate tumor growth and metastasis leading to the reduced survival in the cancer[25–27]. These observations on differences in tumor microenvironment between clusters A and B were also validated independently in the Chinese cohort (**Fig. 4e**).

DISCUSSION

Bladder cancer, a heterogeneous group of diseases with high morbidity and mortality all over the world, still lacks effective treatments and prognostic indicators. Copious epidemiological evidence demonstrates the existence of great racial disparities in the incidence and prognosis of bladder cancer. However, there were few studies investigating the molecular features of bladder cancer patients across different racial populations. Mutational signatures are reflective of previous exposures to various exogenous and endogenous mutagenic processes during the lifetime of cancer evolution. The observed enrichment of clonal mutations in APOBEC-a signature in our study further confirmed the tumorigenic roles of this signature during the early stages of bladder cancer development[28–30].

Notably, we also observed that the activities of several mutational signatures varied substantially among different racial populations. Racial differences in exposures to

exogenous factors related to life history may account for some of the observed racial differences in mutational signatures. The ARISTOLOCHIC signature could only be detected in some Asian patients, especially those from Cohort 2. Herbs or products containing aristolochic acids are widely consumed in some areas in Asia and the ARISTOLOCHIC signature had been detected in several human cancers, including bladder cancer, renal cell carcinoma and liver cancer, in previous studies in Asian patients[31–33]. Our observation that the ARISTOLOCHIC signature showed significant enrichment of subclonal mutations indicated that this life history related signature was less likely to play important roles during the early development of bladder cancer. In terms of the newly identified MMR-like signature which also significantly enriched with subclonal mutations, more evidence was needed to further exclude the potential influence of study batch effects on the observation of this signature in a single study cohort.

Interestingly, our study also suggested the possible roles of racial differences in the prevalence of gene mutations in mediating the racial disparities in the activities of mutational signatures. It had been shown that mutations in *ERCC2* were associated with the activities of COSMIC signature 5 in bladder cancer previously[17]. We showed that mutations in *ERCC2* and other SMGs or pathways were associated with the activities of mutational signatures in bladder cancer. APOBEC-a signature showed the highest activities in White patients and we provided preliminary evidence showing that the high APOBEC-a signature activity was associated with the high prevalence of somatic mutations in *AHNAK* in patients from the White population. Future studies are needed to characterize the biology function of the associations between mutations in SMGs or pathways and mutational signature activities to establish the potential of utilizing the mutated genes such as *AHNAK* as biomarkers for the clinical management of bladder cancer patients.

Previous studies had shown that high APOBEC-a signature activity was indicative of good prognosis[4,7,8]. In our analysis, we showed that clustering of the patients based on their mutational signatures could stratify the patients into different prognostic subgroups. Batch effects may have some influence on interpreting the results of cluster analysis of mutational signatures in patients from different study sources. However, prognostic analysis was performed for patients from Cohort 1 only in our study, which means that all the results of survival analysis were less likely to be affected by batch effects. Although survival information was not available for the NMIBC patients from Cohort 2, it had been well known that NMIBC patients usually can survive with the disease for quite a long time. We assumed that the observed clustering of a subset of Cohort 1 MIBC patients who had better survival with NMIBC patients from Cohort 2 in prognostic analysis cannot be explained by batch effects solely.

Trans-ancestry analysis of the genomic features of bladder cancer revealed that all the SMGs were enriched with clonal mutations but a large number of arm- or chromosome-level SCNAs were subclonal events. Thus, clonal defects of the SMGs may play pivotal roles in the early evolution stages of bladder cancer. Clonal mutations in SMGs (such as *TP53* and *ATM*) regulating the DNA repair machinery further lead to the accumulation of an abundant of SCNAs which provides the rich resources fostering the clonal diversity and inter-population heterogeneity of bladder cancer. The prevalence of *TP53* and *ATM* clonal mutations varied greatly among Whites/Blacks/Asians, which may contribute to the diversified overall burden of SCNAs as well as the inter-populational differences in the prevalence of individual clonal/subclonal SCNA events. Interestingly, when we performed trans-ancestry molecular subtyping based on the genomic architectures of patients from

different populations, we identified a trans-ancestry prognostic subtype that were enriched with Asian NMIBC patients and White/Black/Asian MIBC patients who tend to harbor defects in driver genes, such as *FGFR3*, *HRAS* and *NFE2L2*, showing significant racial differences in mutation prevalence among low-staged patients. These findings demonstrated the feasibility of identifying trans-ancestry prognostic or even therapeutic subtypes based on their genomic architectures, despite of the fact that the genomic features of bladder cancer varied greatly among different populations.

The long-term survival rates for the MIBC patients are generally much lower than the NMIBC patients who usually suffer from disease recurrence but rarely die of bladder cancer[23]. To reduce the risk of developing metastatic disease, MIBC patients are routinely treated with cystectomy of the bladder, which is much more radical and invasive than transurethral resection of bladder tumor currently used for treating NMIBC. Thus, it is imperative to identify molecular markers or subtypes to predict the prognosis of MIBC for guiding the clinical choice of treatments to improve the life-quality of bladder cancer patients. Our study cohorts were mainly composed of MIBC patients from two studies cohorts with different racial backgrounds and a small number of NMIBC patients from a single cohort were also used as the external control for stratifying the MIBC patients into prognostic subtypes. Only a few clonal or sub-clonal somatic events, including SMGs and arm-level SCNAs, were potentially associated with the prognosis of MIBC in our study. Previous studies showed evidence suggesting that a subset of MIBC patients shared similar molecular features with the NMIBC patients whose genetic landscapes were characterized by high prevalence of mutations in the FGFR3-Ras pathway and low burden of SCNAs[34]. Molecular subtypes of MIBC exhibiting similarities to NMIBC at the mRNA expression level had been reported previously but the genetic causes underlying the dynamic changes at

expression levels between subtypes are not clear[4]. We analyzed the genomic architectures of MIBC and NMIBC patients and selected the clonal/sub-clonal states of 60 potential driver events as features of the prognostic subtypes. We identified two prognostic clusters of MIBC, with the shorter overall survival subtype showing higher burden of large-scale SCNAs and WGD, higher prevalence of mutations in *TP53* and *RB1*, higher levels of intra-tumor heterogeneity as well as suppressed tumor microenvironment and higher rate of developing metastatic disease. These findings highlighted the possibility of directing the clinical management of MIBC patients based on their genomic architectures.

ONLINE METHODS

Data sources and sample information

Raw whole-exome sequencing, RNA-seq data and clinical information on the Chinese bladder cancer patients were downloaded from the Sequence Read Archive (SRA063495)[7]. Whole exome sequencing bam files, gene expression read counts and clinical information in the TCGA-BLCA cohort were downloaded from the Genomic Data Commons (GDC) data portal (http://gdc-portal.nci.nih.gov). All bladder cancer patients were classified into the NMIBC or MIBC subtype based on their TNM stages. Patients who had metastatic diseases at the time of biopsy or within one year after biopsy were identified as metastatic and patients who were followed up for at least one year and did not show metastatic diseases during the entire follow up period were identified as non-metastatic. Data on WGD were downloaded from the Genomic Data Commons (GDC) website (https://gdc.cancer.gov/about-data/publications/panimmune)[35]. All bladder cancer patients were classified into Asian, black and White populations based on their racial origins (**Supplementary Table 2 and 3**).

Identification of SNVs and SCNAs

Bam files for the tumor and matched normal samples were processed as input. Somatic mutations were called by using Mutect2 software[12] and further filtered with a read depth of at least 10x in the germline and tumor samples, a maximum of two variant supporting reads in the germline, a minimum tumor variant allele frequency of 10% and a maximum germline variant allele frequency of 2%. Allelic SCNAs were called with the ReCapSeg software following the Broad's GATK documentation (http://gatkforums.broadinstitute.org/categories/recapseg-documentation).

Estimation of intratumor genetic heterogeneity

To estimate the level of intratumor genetic heterogeneity in bladder cancer, we calculated each tumor's MATH value from the median absolute deviation (MAD) and the median of its mutant-allele fractions at somatically mutated loci: MATH=100*MAD/median[36].

RNA-seq analysis and estimation of immune cell score

Bam files for paired samples were processed to quantify the transcript abundance levels by using kallisto[37]. We then used tximport[38] R package to convert transcript level value into gene level. We used signature genes from previous study[39] to infer the infiltration levels of different types of immune cells using ssGSEA[40]. Normalized RNA-seq datasets were provided as the inputs for the execution with 'gsva (data, list of signatures, method = "ssgsea")'[41].

Mutation significance analysis

MutSig2CV algorithm[42,43] was used to identify the SMGs which were further filtered with $P < 0.05$ & $q < 0.1$. A SMG should express in above 75% of the TCGA-BLCA patients and

549 with at least three read counts. Also, they should be reported as the census of human cancer
550 genes (https://cancer.sanger.ac.uk/census, version: 4_08_47_12_2018).
551
552 **Mutation clonality analysis**
553 We used the updated version of ABSOLUTEv1.2 algorithm[44,45] with allelic copy number and
554 mutation data as inputs to infer the ploidies and CCFs of tumor cells. SNVs and SCNAs were
555 defined as clonal if the probabilities of observing CCFs ≥ 0.95 were greater than 0.5
556 ($Pr_{(CCF >=0.95)} >0.5$) or sub-clonal otherwise [46,47]. To assess whether a specific gene or arm
557 level event was enriched with clonal or sub-clonal mutations, we used permutation tests
558 repeated by 10000 times to obtain a p value of clonal enrichment by dividing the times when
559 the observed clonal/sub-clonal ratio was greater than the expected ratio by 10000. We applied
560 the method of constructing the potential temporal order of mutation acquisitions during tumor
561 evolution[48].
562
563 **Mutational signature analysis**
564 We applied a Bayesian variant of NMF (BayesNMF) with the functions of optimal inference
565 for the number of signatures and *de novo* signature discovery to identify the mutational
566 signatures[17]. Comparison of the signatures extracted from three populations (Asian, Black
567 and White) and 30 COSMIC signatures was performed using the standard hierarchial
568 clustering R package with a distance measurement of 'cosine' similarity or 'pearson'
569 correlation. The MSig clustering analysis was performed using a standard hierarchical
570 clustering in R, with a 'euclidian' distance for the signature activity matrix and a 'ward.D'
571 linkage tree. By using the Elbow method to look at the total within-cluster sum of square
572 (wss) as a function of the number of clusters, the number of MSig clusters ($K = 4$) was
573 chosen so that adding another cluster does not improve much better the total wss.

574

575 **Signature enrichment analysis**

576 To search for genes whose mutation statuses were associated with the activity of a specific
577 signature, we applied a permutation statistical test[17] to compare, for each gene, the signature
578 activities between samples with and without mutations in the gene. Firstly, to remove the
579 inflation in the number of mutations in each gene associated with the elevated background
580 mutation rates, we randomly produced a total of 10^4 permutated mutation matrix in which the
581 total counts of gene-specific and sample-specific mutations were kept the same as those
582 observed in our patient cohorts, following an approach described by Strona *et al*[49]. We used
583 the one-tailed Wilcoxon rank-sum *P* value to compare the signature activity between mutant
584 and wild type samples of a specific gene. To get the final *P* value for a given gene, we
585 calculated the fraction of permutated matrix with a test *P* value equal to or more extreme than
586 the observed matrix (Sum of $P_{observed} \geq P_{random}$ times/the total number of permutations). To
587 increase the computational efficiency, we analyzed only 174 genes with a non-silent mutation
588 frequency of 5% or above, 26 SMGs, the APOBEC family member gene sets and the human
589 DNA repair gene sets that had been reported to be associated with some mutational signature
590 activities (downloaded from https://www.mdanderson.org/documents/Labs/Wood-
591 Laboratory/human-dna-repair-genes.html). We corrected for multiple-hypothesis testing
592 using the Benjamini-Hochberg procedure and used FDR $q < 0.1$ as the significance threshold.

593

594 **Calculation of the SCNA scores**

595

596 We calculated the SCNA scores according to a previous study with minor modifications [50].
597 Our calculation of SCNA was based on the integer allelic copy number calls generated by
598 ABSOLUTE [51] which took the tumor purities into consideration. The absolute copy number

of each segment was determined and a chromosomal arm event would be called out if the cumulative length of the deletions (genomic segments with minor alleles of 0 copy) or amplifications (genomic segments with major alleles of 2 or more copies) on this arm was greater than 50% of the chromosome arm. Chromosome level events were distinguished from arm level events when both arms of a chromosome had the same copy number changes (in sign). After excluding the arm- and chromosome-level copy number changes, the remaining segments with SCNAs were defined as focal changes. Based on the copy number change sign associated with each segment *x* of a chromosome, arm and focal event *t*, the assigned score *S* were defined as:

$$Sx = \begin{cases} +2 & one\ allele\ had\ amplification\ and\ the\ another\ one\ had\ deletion \\ +1 & both\ the\ major\ and\ minor\ alleles\ had\ amplifications\ or\ deletions \\ +1 & one\ of\ major\ and\ minor\ allele\ was\ in\ normal\ state \\ 0 & both\ major\ and\ minor\ alleles\ were\ in\ normal\ state \end{cases}$$

The sum score of deletions or amplifications at the chromosome (*ChromL*), arm (*ArmL*), and focal levels (*FocalL*), were calculated separately:

$$ChromL = \sum_{t\ \epsilon\ Chrom} Sx$$

$$ArmL = \sum_{t\ \epsilon\ Arm} Sx$$

$$FocalL = \sum_{t\ \epsilon\ Focal} Sx$$

The *ChromL*, *ArmL* and *FocalL* scores of each tumor sample were normalized to the mean and standard deviation calculated among all samples. In addition, we divided the summed score of all SCNA segments by the total length of genomic segments to get "total normalized SCNA level" score for each tumor sample, which represented the integrated level of SCNAs among the genome[52].

Unsupervised clustering based on genomic architecture

We used the CCFs of 26 SMGs and frequent arm level SCNAs to create a numerical matrix (samples as columns) and then applied the NMF 'lee' algorithm to perform molecular subtyping of bladder cancer. After manual inspection we chose the K = 2 solution and reported two clusters (clusters A vs B).

Statistical analysis

Two-sided Mann-Whitney and Fisher's exact tests were performed with the R functions Wilcox.test and chisq.test to generate the empirical *P* values, respectively. *P* values were adjusted for multiple hypothesis tests using the R function p.adjust with the "fdr" option

Survival analysis

Asian bladder cancer patients from Asia cohort were excluded in survival analysis due to the lack of some key clinical information. Chi-squre test statistics in Kaplan-Meier curves were computed using log-rank tests. *P* values were also calculated from multivariate Cox proportional-harzards regression models using the R package "survival".

Data availability

Any relevant data are available from the authors upon reasonable request. The raw sequencing data used in this manuscript are all publicly available (SRA063495 and Genomic Data Commons (GDC) data portal (http://gdc-portal.nci.nih.gov)). The data produced by the analysis in this manuscript are summarised in the Supplementary Tables and Supplementary Data.

Code availability

All custom code used in this work is available from the corresponding authors upon reasonable request.

Competing interests: The authors declare no competing interests.

Author contributions

B.F.Z. and Y.H. conceived the study. B.F.Z. and X.C.L. performed the bioinformatics analysis. G.S.P., P.L.J. and Z.M.Z. downloaded and processed TCGA-BLCA data. C.X.W., S.D. and S.N.P. performed GSEA analysis. B.F.Z. and Y.H. wrote the manuscript. B.F.Z., J.Y.W., X.Y. and Y.H. revised the manuscript.

Funding: This study was supported by Shenzhen municipal government (JCYJ20160429093033251)

REFERENCES

1. Ferlay, J. *et al.* Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J. cancer* **136**, E359–E386 (2015).
2. Sylvester, R. J. *et al.* Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials. *Eur. Urol.* **49**, 466–477 (2006).
3. Fernandez-Gomez, J. *et al.* Predicting nonmuscle invasive bladder cancer recurrence and progression in patients treated with bacillus Calmette-Guerin: the CUETO scoring model. *J. Urol.* **182**, 2195–2203 (2009).
4. Robertson, A. G. *et al.* Comprehensive molecular characterization of muscle-invasive bladder cancer. *Cell* **171**, 540-556. e25 (2017).
5. Hedegaard, J. *et al.* Comprehensive Transcriptional Analysis of Early-Stage Urothelial Carcinoma. *Cancer Cell* (2016) doi:10.1016/j.ccell.2016.05.004.
6. Weinstein, J. N. *et al.* Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* **507**, 315–322 (2014).
7. Guo, G. *et al.* Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation. *Nat. Genet.* **45**, 1459 (2013).
8. Gui, Y. *et al.* Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat. Genet.* **43**, 875 (2011).
9. Faltas, B. M. *et al.* Clonal evolution of chemotherapy-resistant urothelial carcinoma. *Nat. Genet.* **48**, 1490–1499 (2016).
10. Yee, D. S., Ishill, N. M., Lowrance, W. T., Herr, H. W. & Elkin, E. B. Ethnic differences in bladder cancer survival. *Urology* **78**, 544–549 (2011).

694 11. Gandomani, H. S. *et al.* Essentials of bladder cancer worldwide: incidence, mortality
695 rate and risk factors. *Biomed. Res. Ther.* **4**, 1638 (2017).
696 12. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and
697 heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
698 13. Lichtenstein, L., Woolf, B., MacBeth, A., Birsoy, O. & Lennon, N. ReCapSeg:
699 Validation of somatic copy number alterations for CLIA whole exome sequencing.
700 (2016).
701 14. Health, N. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 1–
702 108 (2013).
703 15. Alexandrov, L. B. & Stratton, M. R. Mutational signatures: the patterns of somatic
704 mutations hidden in cancer genomes. *Curr. Opin. Genet. Dev.* **24**, 52–60 (2014).
705 16. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R.
706 Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.*
707 **3**, 246–259 (2013).
708 17. Kim, J. *et al.* Somatic ERCC2 mutations are associated with a distinct genomic
709 signature in urothelial tumors. *Nat. Genet.* (2016).
710 18. Lee, I. H. *et al.* Ahnak functions as a tumor suppressor via modulation of TGFβ/Smad
711 signaling pathway. *Oncogene* **33**, 4675 (2014).
712 19. Futreal, P. A. *et al.* A census of human cancer genes. *Nat. Rev. cancer* **4**, 177 (2004).
713 20. Robertson, A. G. *et al.* Comprehensive Molecular Characterization of Muscle-Invasive
714 Bladder Cancer. *Cell* **171**, 540-556.e25 (2017).
715 21. Gui, Y. *et al.* Frequent mutations of chromatin remodeling genes in transitional cell
716 carcinoma of the bladder. *Nat. Genet.* **43**, 875–878 (2011).
717 22. Guo, G. *et al.* Whole-genome and whole-exome sequencing of bladder cancer
718 identifies frequent alterations in genes involved in sister chromatid cohesion and

segregation. *Nat. Genet.* **45**, 1459–1463 (2013).

23. Soria, F., Moschini, M., Korn, S. & Shariat, S. F. How to optimally manage elderly bladder cancer patients? *Translational Andrology and Urology* vol. 5 683–691 (2016).

24. Asadzadeh, Z. *et al.* The paradox of Th17 cell functions in tumor immunity. *Cell. Immunol.* (2017).

25. De Monte, L. *et al.* Intratumor T helper type 2 cell infiltrate correlates with cancer-associated fibroblast thymic stromal lymphopoietin production and reduced survival in pancreatic cancer. *J. Exp. Med.* **208**, 469–478 (2011).

26. Ostuni, R., Kratochvill, F., Murray, P. J. & Natoli, G. Macrophages and cancer: from mechanisms to therapeutic implications. *Trends Immunol.* **36**, 229–239 (2015).

27. Walker, J. A. & McKenzie, A. N. J. T H 2 cell development and function. *Nat. Rev. Immunol.* **18**, 121 (2018).

28. McGranahan, N. *et al.* Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Sci. Transl. Med.* **7**, 283ra54-283ra54 (2015).

29. Hedegaard, J. *et al.* Comprehensive Transcriptional Analysis of Early-Stage Urothelial Carcinoma. *Cancer Cell* **30**, 27–42 (2016).

30. Lamy, P. *et al.* Paired exome analysis reveals clonal evolution and potential therapeutic targets in urothelial carcinoma. *Cancer Res.* **76**, 5894–5906 (2016).

31. Poon, S. L. *et al.* Mutation signatures implicate aristolochic acid in bladder cancer development. *Genome Med.* **7**, 38 (2015).

32. Hoang, M. L. *et al.* Aristolochic acid in the etiology of renal cell carcinoma. *Cancer Epidemiol. Biomarkers Prev.* **25**, 1600–1608 (2016).

33. Ng, A. W. T. *et al.* Aristolochic acids and their derivatives are widely implicated in liver cancers in Taiwan and throughout Asia. *Sci. Transl. Med.* **9**, (2017).

744 34. Felsenstein, K. M. & Theodorescu, D. Precision medicine for urothelial bladder
745 cancer: Update on tumour genomics and immunotherapy. *Nat. Rev. Urol.* **15**, 92–111
746 (2018).
747 35. Thorsson, V. *et al.* The immune landscape of cancer. *Immunity* **48**, 812-830. e14
748 (2018).
749 36. Mroz, E. A. & Rocco, J. W. MATH, a novel measure of intratumor genetic
750 heterogeneity, is high in poor-outcome classes of head and neck squamous cell
751 carcinoma. *Oral Oncol.* **49**, 211–215 (2013).
752 37. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-
753 seq quantification. *Nat. Biotechnol.* **34**, 525 (2016).
754 38. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq:
755 transcript-level estimates improve gene-level inferences. *F1000Research* **4**, (2015).
756 39. Şenbabaoğlu, Y. *et al.* Tumor immune microenvironment characterization in clear cell
757 renal cell carcinoma identifies prognostic and immunotherapeutically relevant
758 messenger RNA signatures. *Genome Biol.* **17**, 231 (2016).
759 40. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven
760 cancers require TBK1. *Nature* **462**, 108 (2009).
761 41. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for
762 microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
763 42. Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21
764 tumour types. *Nature* **505**, 495 (2014).
765 43. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new
766 cancer-associated genes. *Nature* **499**, 214 (2013).
767 44. Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human
768 cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).

45. Huang, Y. *et al.* Clonal architectures predict clinical outcome in clear cell renal cell carcinoma. *Nat. Commun.* **10**, 1–10 (2019).

46. Lohr, J. G. *et al.* Widespread genetic heterogeneity in multiple myeloma: Implications for targeted therapy. *Cancer Cell* **25**, 91–101 (2014).

47. Landau, D. A. *et al.* Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* **152**, 714–726 (2013).

48. Wang, J. *et al.* Tumor evolutionary directed graphs and the history of chronic lymphocytic leukemia. *Elife* **3**, e02869 (2014).

49. Strona, G., Nappo, D., Boccacci, F., Fattorini, S. & San-Miguel-Ayanz, J. A fast and unbiased procedure to randomize ecological binary matrices with fixed row and column totals. *Nat. Commun.* **5**, 4114 (2014).

50. Davoli, T., Uno, H., Wooten, E. C. & Elledge, S. J. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science (80-.).* **355**, eaaf8399 (2017).

51. Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).

52. Davoli, T., Uno, H., Wooten, E. C. & Elledge, S. J. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science (80-.).* **355**, eaaf8399 (2017).

Figure Legends

Figure 1. Mutational signatures of bladder cancer among Asian-140, Black-23 and White-324 populations. (a) Unsupervised hierarchical clustering of signatures identified in our cohorts and the 30 signatures described by COSMIC. (b) Differences in the mutational signature activities among three populations. (c) Unsupervised hierarchical clustering of 487 samples (MSig clusters) based on the number of mutations assigned to the five mutational processes. (d) Enriched signature features of the four MSig clusters. (e) Kaplan-Meier survival analysis of the four MSig clusters. (f) Multivariate Cox regressions analysis of survival by including MSig clusters, racial origins, age, gender and TNM stage.

Figure 2. Gene mutations associated with signature activities. (a) Mutation enrichment analysis identifies an association between four somatically mutated genes and signature activities. (b) Comparison of the estimated numbers of signature mutations in tumors with wild-type versus mutant *AHNAK*/*ERCC2*/*HRAS*. (c) Different mutation frequencies of the three genes among Asian/Black/White populations.

Figure 3. Racial difference in SMG mutations and SCNAs. (a) The prevalence of mutations in SMGs across races. (b) SMGs showing significantly different mutation frequencies in different populations. (c) The overall burden of SCNAs in different populations. (d) The prevalence of SCNAs across race groups. (e) Chromosome- or arm-level SCNAs showing great racial disparities.

Figure 4. Prognostic subtypes of bladder cancer with distinct genomic architectures. (a) Landscape of genomic architectures in 505 bladder cancer samples. (b) Kaplan-Meier

819 survival analysis of the two clonal subtypes. (c) Multivariate Cox regressions analysis of
820 survival by including clonal subtype, racial origins, WGD, MSig cluster, age, gender and
821 TNM stage. (d) Differences in the genome-wide/chromosome/arm and focal levels of SCNAs
822 scores between two subtypes. (e) Immune features of the two subtypes.

Figures

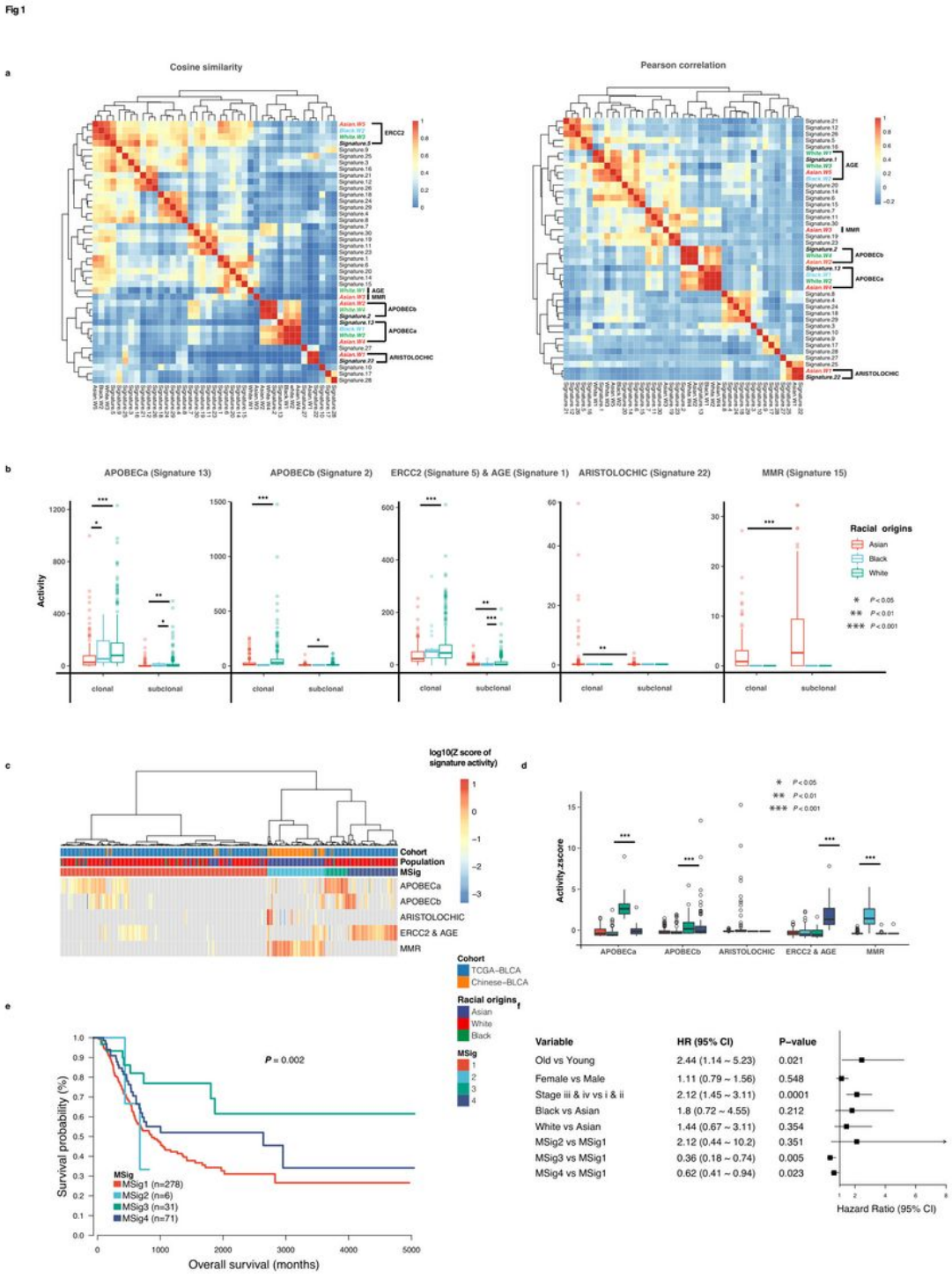

Figure 1

Mutational signatures of bladder cancer among Asian-140, Black-23 and White-324 populations. (a) Unsupervised hierarchical clustering of signatures identified in our cohorts and the 30 signatures described by COSMIC. (b) Differences in the mutational signature activities among three populations. (c)

Unsupervised hierarchical clustering of 487 samples (MSig clusters) based on the number of mutations assigned to the five mutational processes. (d) Enriched signature features of the four MSig clusters. (e) Kaplan-Meier survival analysis of the four MSig clusters. (f) Multivariate Cox regressions analysis of survival by including MSig clusters, racial origins, age, gender and TNM stage.

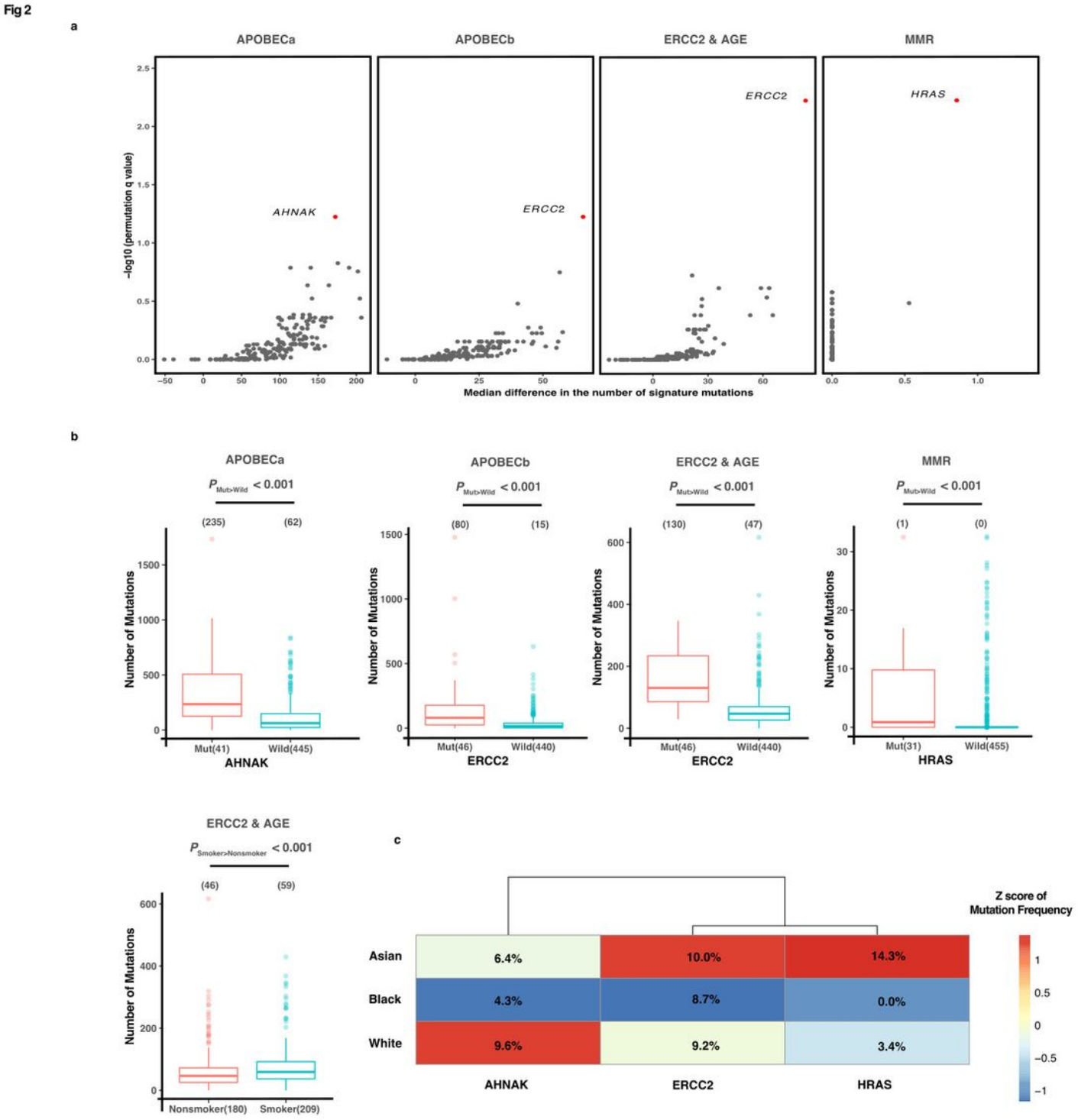

Figure 2

Gene mutations associated with signature activities. (a) Mutation enrichment analysis identifies an association between four somatically mutated genes and signature activities. (b) Comparison of the estimated numbers of signature mutations in tumors with wild-type versus mutant AHNAK/ERCC2/HRAS. (c) Different mutation frequencies of the three genes among Asian/Black/White populations.

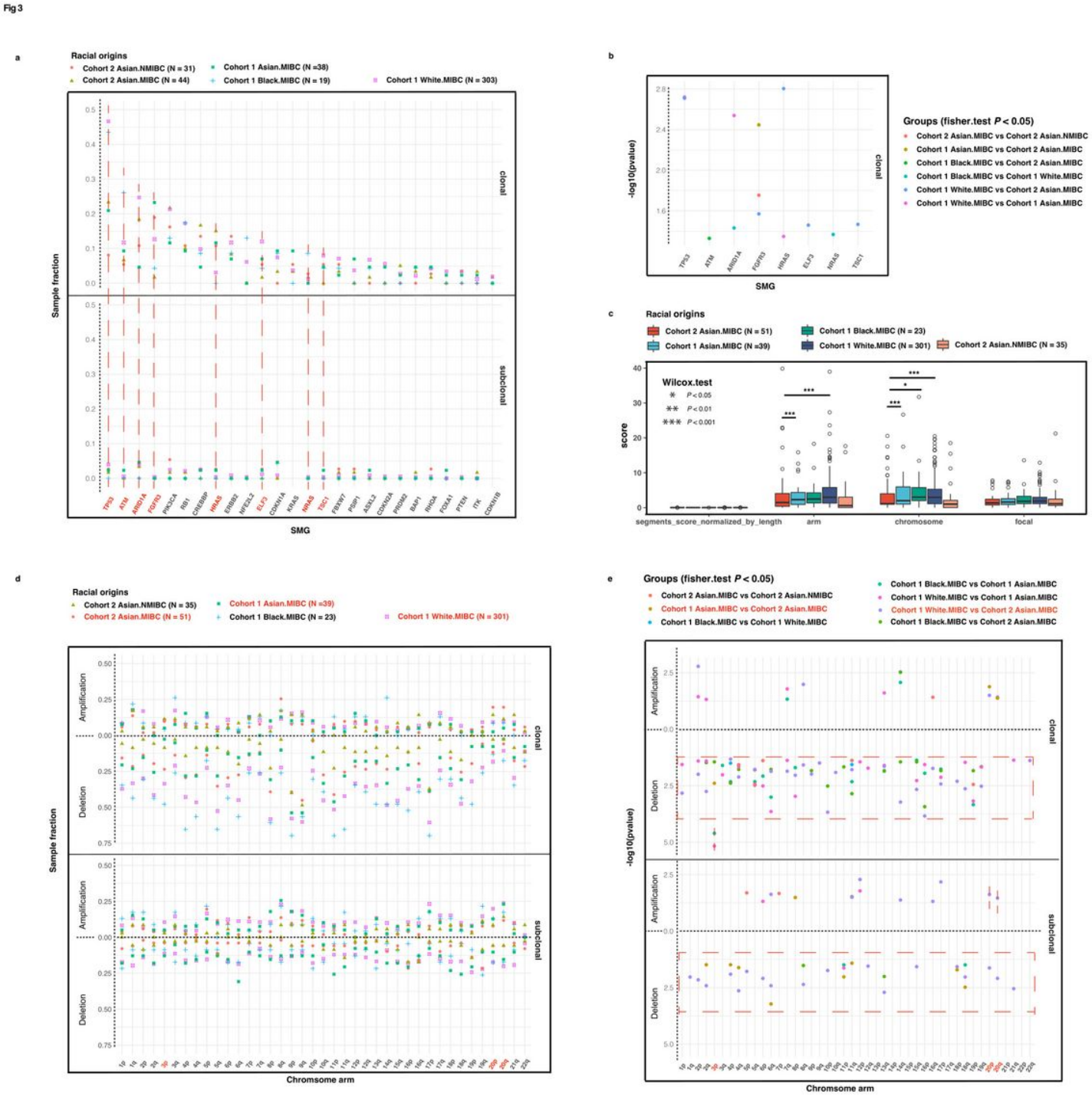

Figure 3

Racial difference in SMG mutations and SCNAs. (a) The prevalence of mutations in SMGs across races. (b) SMGs showing significantly different mutation frequencies in different populations. (c) The overall burden of SCNAs in different populations. (d) The prevalence of SCNAs across race groups. (e) Chromosome- or arm-level SCNAs showing great racial disparities.

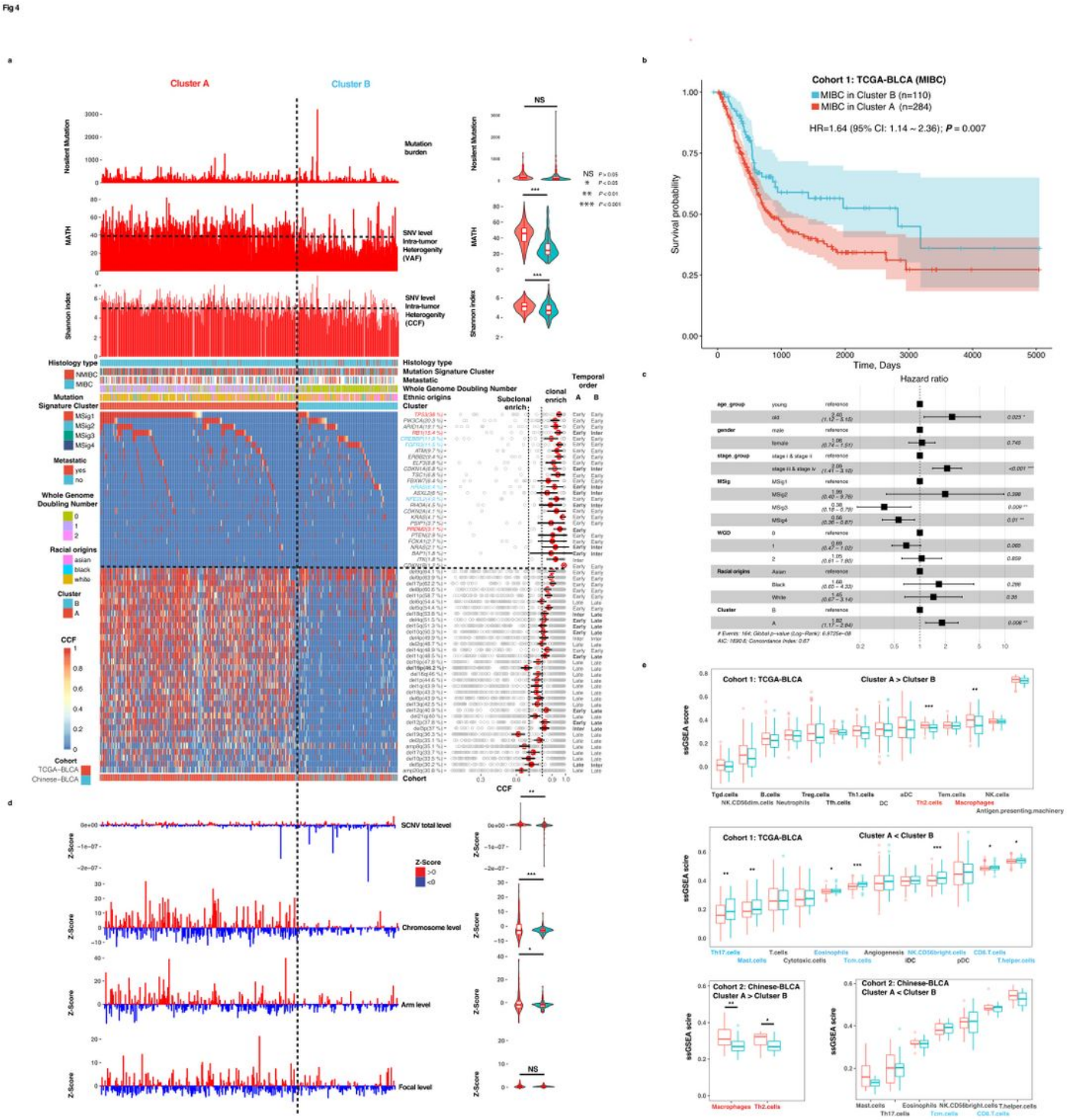

Figure 4

Prognostic subtypes of bladder cancer with distinct genomic architectures. (a) Landscape of genomic architectures in 505 bladder cancer samples. (b) Kaplan-Meier survival analysis of the two clonal subtypes. (c) Multivariate 819 Cox regressions analysis of survival by including clonal subtype, racial origins, WGD, MSig cluster, age, gender and TNM stage. (d) Differences in the genome-wide/chromosome/arm and focal levels of SCNAs scores between two subtypes. (e) Immune features of the two subtypes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1.pdf
- SupplementaryData14.zip
- Supplementarymaterials.pdf