MUTATION

Macroscopic somatic clonal expansion in morphologically normal human urothelium

Ruoyan Li[1]*, Yiqing Du[2]*, Zhanghua Chen[1]*, Deshu Xu[1]*, Tianxin Lin[3]*, Shanzhao Jin[1], Gongwei Wang[4], Ziyang Liu[1], Min Lu[5], Xu Chen[3], Tao Xu[2]†, Fan Bai[1]†

Knowledge of somatic mutation accumulation in normal cells, which is essential for understanding cancer development and evolution, remains largely lacking. In this study, we investigated somatic clonal events in morphologically normal human urothelium (MNU; epithelium lining the bladder and ureter) and identified macroscopic clonal expansions. Aristolochic acid (AA), a natural herb-derived compound, was a major mutagenic driving factor in MNU. AA drastically accelerates mutation accumulation and enhances clonal expansion. Mutations in MNU were widely observed in chromatin remodeling genes such as *KMT2D* and *KDM6A* but rarely in *TP53*, *PIK3CA*, and *FGFR3*. *KMT2D* mutations were found to be common in urothelial cells, regardless of whether the cells experience exogenous mutagen exposure. Copy number alterations were rare and largely confined to small-scale regions, along with copy-neutral loss of heterozygosity. Single AA-associated clones in MNU expanded to a scale of several square centimeters in size.

Over the course of their life span, cells inevitably acquire somatic mutations that mainly result from unrepaired or incorrectly repaired DNA replication errors that occur during cell division (*1*, *2*). Although most somatic mutations in normal cells do not have any phenotypic consequence, mutations that affect essential genes, especially those related to cell proliferation and death, may trigger mutant clonal expansions (*3*). A well-recognized example is human cancer, in which progressive accumulation of somatic mutations drives clonal expansions and the eventual malignant transformation of cells (*4*). Although genomic sequencing of various human malignancies has revolutionized our understanding of the molecular and genetic bases of cancer development and evolution (*5*–*7*), little is known about the patterns and driving factors of somatic mutations in normal cells before malignant transformation. Recent studies have shed light on the mutational landscapes of different normal tissues, including skin epidermis (*8*), esophageal tissue (*3*, *9*), colorectal tissue (*10*), liver (*11*), endometrial epithelium (*12*), bronchial epithelium (*13*), brain (*14*, *15*), embryonic tissue (*16*), and blood cells (*17*), thus contributing to our understanding of mutation rates, driver genes, and mutagenic driving forces in normal cells (*18*, *19*). In particular, previous studies have highlighted the critical roles of the aging-related endogenous mutational process in normal cells, evidenced by the positive correlation between mutation load and age (*3*, *9*, *15*). Also, ultraviolet light, as an exogenous mutagenic factor, has been reported to trigger mutagenesis in normal skin cells and induce skin cancer onset (*8*, *20*, *21*). Whether other underlying mutational processes, both endogenous and exogenous, operate early in normal cells warrants further investigation.

The urothelium is the epithelium that lines the urinary bladder and ureters. It is classified as a transitional epithelium because its properties lie between stratified squamous and simple nonstratified epithelia (*22*). It is highly regenerative in response to damage, thus guaranteeing its barrier function (*23*). Given its direct contact with urine, the urothelium is continually exposed to an array of potentially carcinogenic metabolic products and environmental factors that can cause tissue damage and pose genotoxic stress to urothelial cells. Under these conditions, the urothelium may accumulate somatic mutations through recurrent cell turnover. In this study, using a combination of laser-capture microdissection and exome sequencing, we systematically investigated somatic mutant clonal events in morphologically normal urothelium (MNU), including both bladder and ureter urothelium, from 120 patients with urothelial cell carcinoma (UCC).

Somatic mutations in MNU tissues

In total, we sequenced 161 MNU samples from 120 UCC patients with radical cystectomy or nephroureterectomy (table S1). Urothelium layers of each sample were dissected from consecutive tissue sections using laser-capture microdissection to provide a urothelial surface area of ~2 mm^2 (fig. S1A). Independent pathological examinations confirmed that MNU samples (125 from bladder and 36 from ureter), which were extracted far from tumors, were free of histological changes (fig. S1B). DNA from white blood cells of each patient was sequenced as the germline comparator. We also sequenced 126 tumors (93 bladder, 17 ureter, and 16 renal pelvis tumors) from the 120 patients. On average, we obtained 138-fold, 129-fold, and 138-fold coverage depth of target regions in UCC, MNU, and blood samples, respectively (table S2). Overall, the median mutational burden of UCC was higher than those of prostate, breast, and kidney clear cell carcinomas and comparable to The Cancer Genome Atlas (TCGA) bladder cancer data (fig. S2A and tables S3 and S4). Unexpectedly, while the median mutational burden was low, the overall mutational burden of MNU displayed a wide range (fig. S2A and tables S4 and S5). Several urothelium samples were even hypermutated (for example, sample P65U had >6000 mutations). This finding illustrates that detectable somatic mutations have accumulated in some MNU samples.

Next, we combined our cohort (including both UCC and MNU samples) with a bladder cancer cohort (Chinese population) (n = 99 individuals) (*24*) to catalog significantly mutated genes (SMGs). We identified 19 SMGs with significant recurrent mutation rates, including canonical cancer genes such as *TP53*, *ARID1A*, and *PIK3CA* (table S6). Mutations in these genes have high clonalities in tumors (fig. S2B). All 19 SMGs identified here have been reported by TCGA as potential driver genes in bladder cancer (*25*, *26*). To further investigate the occurrence of mutations in putative driver genes in MNU, we focused on both the 19 SMGs and nine additional genes that were reported by TCGA as potential driver genes and were frequently, but not significantly, mutated in our cohort (e.g., *ATM*, *KMT2C*, and *FAT1*) (Fig. 1A and tables S7 and S8). These 28 genes recapitulated key pathways (e.g., cell cycle and p53 pathways) that have been implicated in urothelial tumorigenesis (fig. S2, C and D). Overall, we found that ~37% of MNU samples had a somatic mutation in at least one of the 28 putative driver genes (Fig. 1A and tables S7 and S8). Meanwhile, we found that 28 MNU samples shared origins with their paired tumors from the same patients (Fig. 1A and table S9). When we excluded these samples, MNU with mutations in *KMT2D* (16/133, 12.0%), *KDM6A* (15/133, 11.3%), *ATM* (11/133, 8.3%), *CREBBP* (11/133, 8.3%), *FAT1* (12/133, 9.0%), and *KMT2C* (10/133, 7.5%) remained widely observed. Although *TP53* was the second most frequently mutated gene in UCC, *TP53* mutations were relatively rare in MNU (5/133, 3.8%) (Fig. 1A). Notably, the mutation rates of *FGFR3* (0/133, 0.0%) and *PIK3CA* (1/133, 0.8%) were lower than those of *CREBBP*, *ATM*, and *KMT2C* in MNU (Fig. 1A). This observation was substantially different from that in UCC, suggesting

[1]Biomedical Pioneering Innovation Center (BIOPIC), School of Life Sciences, Peking University, Beijing, China. [2]Department of Urology, Peking University People's Hospital, Beijing, China. [3]Department of Urology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China. [4]Department of Pathology, Peking University People's Hospital, Beijing, China. [5]Department of Pathology, School of Basic Medical Sciences, Peking University Third Hospital, Peking University Health Science Center, Beijing, China.
*These authors contributed equally to this work.
†Corresponding author. Email: xutao@pkuph.edu.cn (T.X.); fbai@pku.edu.cn (F.B.)

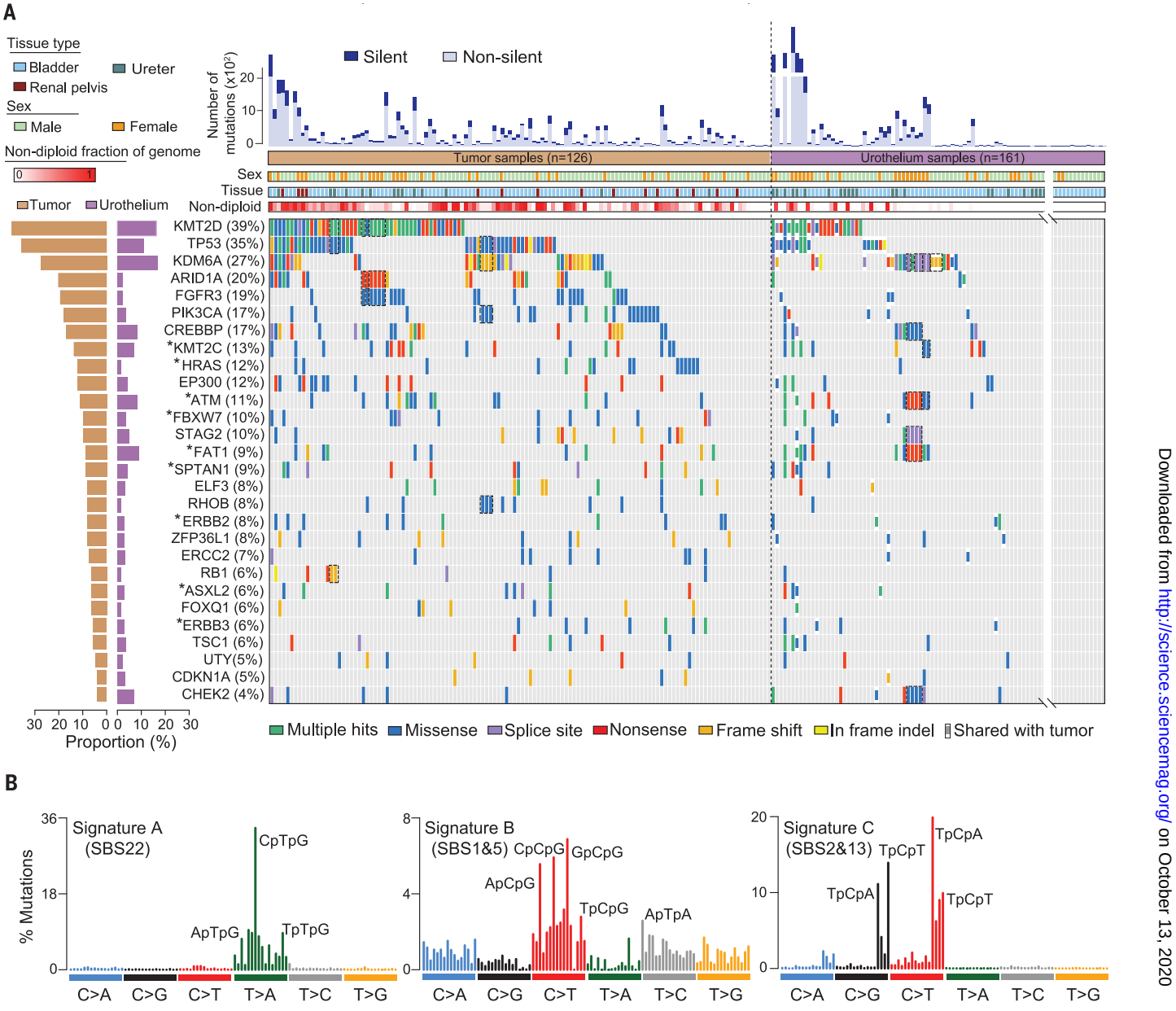

Fig. 1. Somatic mutations and mutational signatures. (**A**) Mutational landscapes of UCC and MNU samples showing mutations in putative driver genes, including 19 SMGs and nine frequently, but not significantly, mutated genes (indicated with asterisks), ordered by their mutation frequency in UCC (percent of tumors with mutations in each gene is in parentheses). Black dashed boxes show that mutations are shared in samples from the same patient. indel, insertion or deletion. (**B**) Mutational spectrum of the three de novo mutational signatures extracted by the SigProfilerExtractor analysis. Representative 3–base pair (bp) mutational contexts are labeled. Corresponding COSMIC signatures are labeled in parentheses.

that different molecular mechanisms underlie early clonal expansion and final cancer development. Putative driver genes in MNU, such as *KMT2D* and *FAT1*, are also frequently mutated in normal skin and esophageal tissues (*3*, *8*, *9*). However, we did not observe enrichment of *NOTCH1* mutations in MNU, although it has been reported as the most frequently mutated gene in skin and esophageal tissues. This observation may reflect intrinsic biological differences among various cell types.

Widespread mutagenesis related to aristolochic acid in MNU

To explore the underlying mutagenic driving factors, we used a nonnegative matrix factorization algorithm on the MNU and UCC samples to extract potential mutational signatures (table S10). We identified three mutational signatures through the de novo extraction (Fig. 1B, fig. S3A, and table S11). Signature B closely resembled the Catalogue of Somatic Mutations in Cancer (COSMIC) signature SBS1 and SBS5 (Fig. 1B and fig. S3B). Signature C exhibited dominant C>G and C>T substitutions in the 5′-TpCpA-3′ and 5′-TpCpT-3′ context and largely conformed to COSMIC SBS2 and SBS13 (Fig. 1B and fig. S3B), which are associated with the activity of APOBEC cytidine deaminases.

Signature A displayed predominant T>A transversions with conspicuous biases in the local sequence context and a markedly high proportion in the 5′-CpTpG-3′ context (Fig. 1B and fig. S3C). This signature matched COSMIC SBS22 with the underlying etiological factor being aristolochic acid (AA), a natural

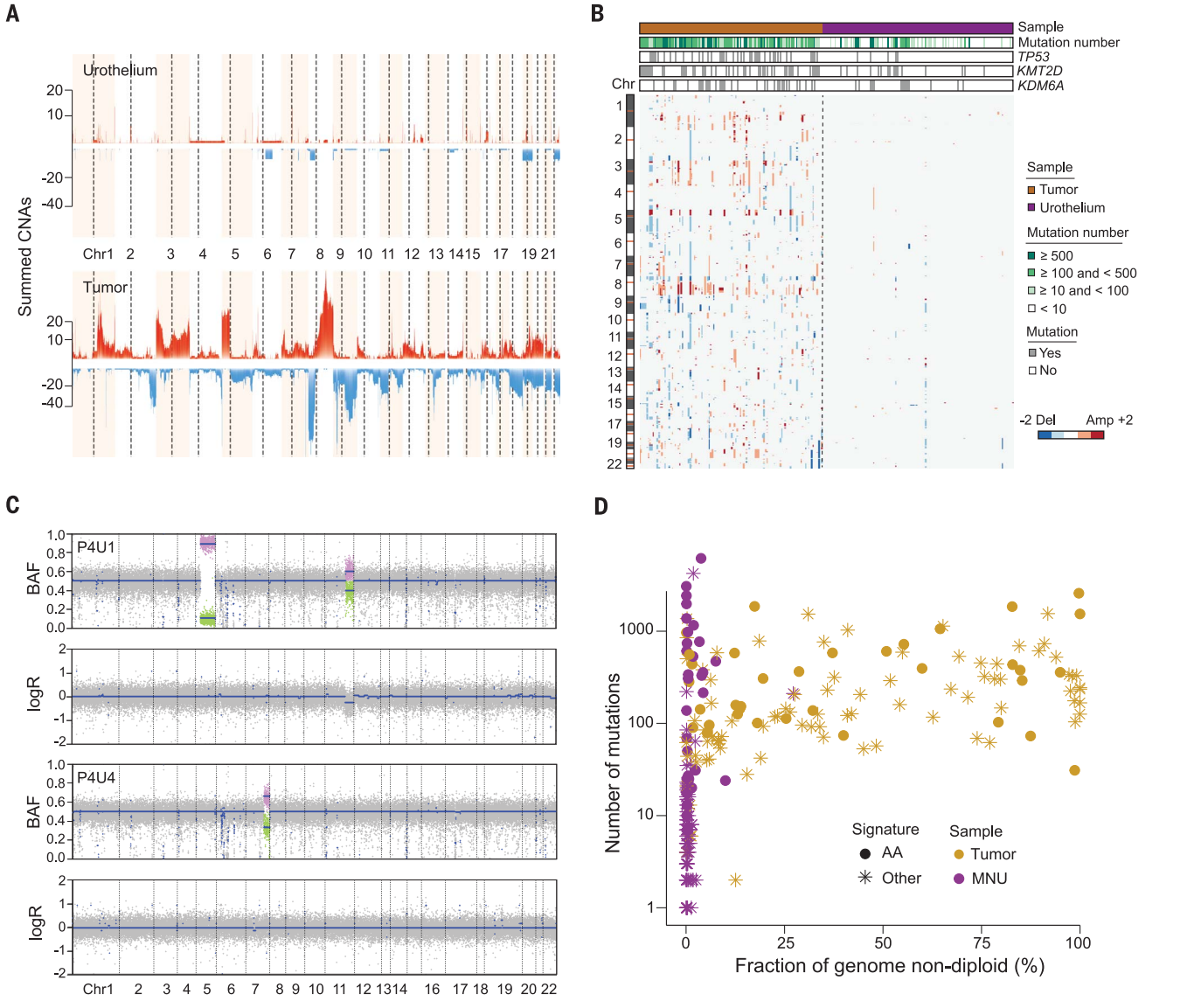

Fig. 2. CNAs in UCC and MNU samples. (**A**) Stacked mountain plots comparing summed CNAs in UCC and MNU samples. Red stacks represent amplifications, and blue stacks represent deletions. Chr, chromosome. (**B**) Comparison of CNAs in UCC and MNU samples. The mutational burdens and mutation statuses of driver genes are indicated. Del, deletion; Amp, amplification. (**C**) CNA log ratios (logR) and B allele frequency (BAF) of two representative MNU samples showing typical copy number statuses and copy-neutral LOH. Blue lines represent the fitted values of logR and BAFs calculated by the circular binary segmentation algorithm. (**D**) The relationship between somatic mutation loads and CNAs. UCC ("Tumor") and MNU samples were assigned to one of the two groups according to their mutational signatures: AA-associated and non-AA-associated ("Other").

herb-derived compound that is known as a notorious mutagen (*27*–*33*) (fig. S3B). Our finding demonstrates that AA mutagenesis is prevalent in normal human tissues (fig. S3, D and E, and table S11), although it was reported in noncancerous tissues in a patient with alcohol-related liver disease (*11*). Widespread AA mutagenesis in MNU was further confirmed using another mutational signature analysis approach (figs. S4 and S5A). AA-associated samples, both tumors and MNU, exhibited significantly higher mutation numbers ($P < 0.001$, Wilcoxon rank-sum test), demonstrating the strong mutagenic effect of AA (fig. S5B). We also found that AA mutagenesis was more prevalent in females than in males (fig. S5C, $P < 0.001$, Fisher's exact test). This gender bias has been reported in upper tract urothelial carcinoma, but the underlying mechanism is unclear (*34*). Our findings demonstrate that AA exposure poses strong genotoxic stress to urothelial cells and widely triggers mutagenesis in normal urothelium.

Copy number alterations in MNU tissues

We assessed copy number alterations (CNAs) in MNU and UCC samples using exome sequencing data (fig. S6). Overall, we observed marked differences in CNAs between tumors and MNU. As expected, tumors harbored extensive CNAs across the whole genome, with recurrent CNA regions consistent with TCGA data (e.g., chromosome 5p, 8q, and 3p amplifications; and chromosome 8p and 9p deletions) (Fig. 2A). However, CNAs were rare across the genomes of MNU samples, even those with

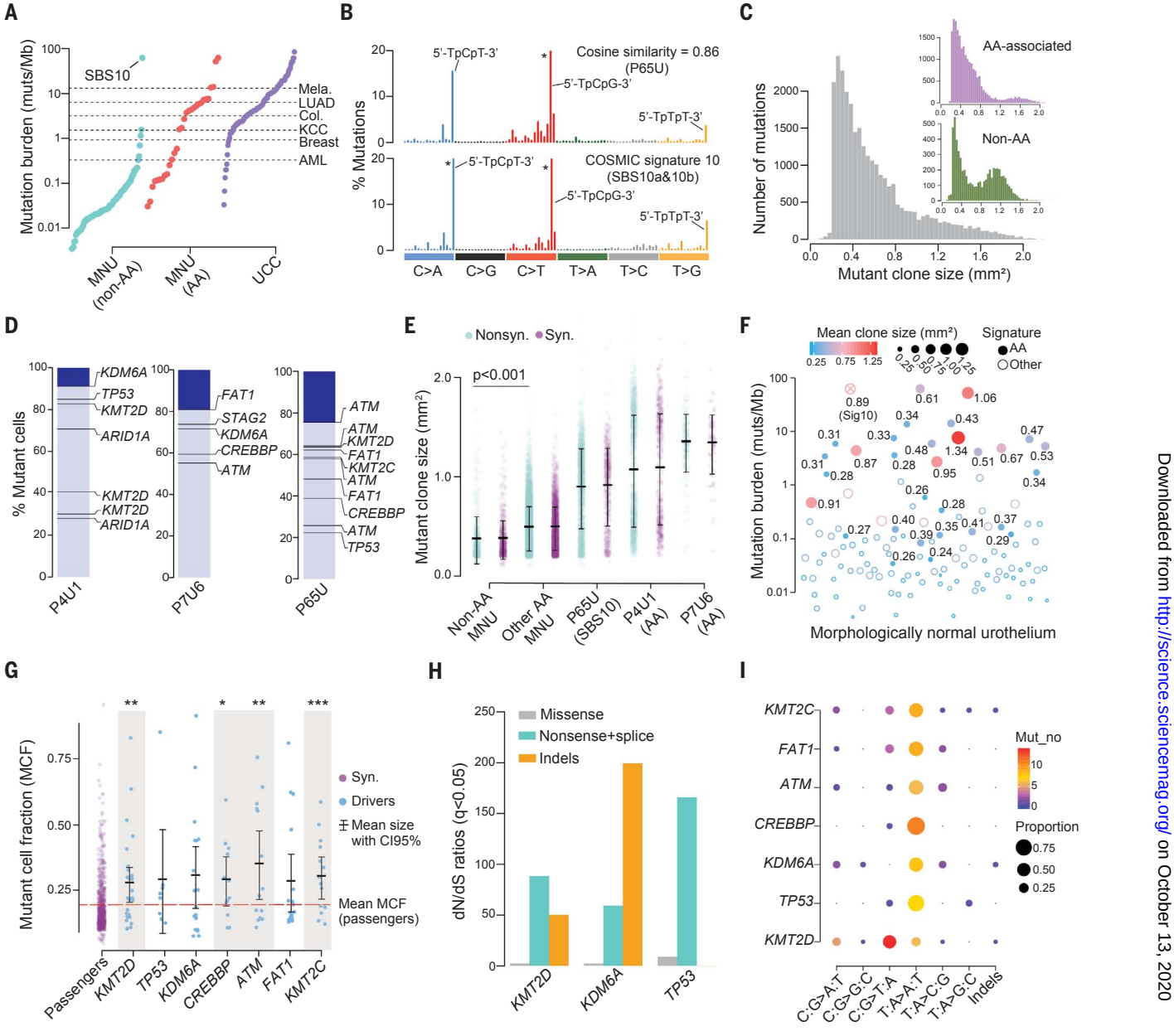

Fig. 3. Mutational burden and mutant clone expansion in MNU.
(**A**) Comparison of mutational burdens (muts/Mb, mutations per megabase) in MNU (both AA-associated and non-AA-associated) and UCC samples. Median mutational burdens of six other cancer types are indicated by dashed lines. AML, acute myeloid leukemia; KCC, kidney clear cell carcinoma; Col., colorectal carcinoma; LUAD, lung adenocarcinoma; Mela., melanoma. SBS10 resembles the COSMIC mutational signature 10 (SBS10a and 10b). (**B**) Comparison of mutational contexts of sample P65U and COSMIC SBS10a and 10b. Representative 3-bp mutational contexts are labeled. (**C**) Distributions of mutant clone sizes of mutations in AA-associated and non-AA-associated MNU samples. The *y* axis represents the number of mutations. (**D**) Bar plots displaying the co-occurrence of driver mutations in the same clones deduced on the basis of the pigeonhole principle. (**E**) Comparison of clone sizes among different MNU samples. Nonsyn., nonsynonymous; Syn., synonymous. (**F**) Mutational burdens and average mutant clone sizes in MNU samples. The average mutant clone sizes of AA-associated samples are labeled. (**G**) Comparison of clone sizes between putative driver and passenger mutations. Wilcoxon rank-sum test was used. **P* < 0.05, ***P* < 0.01, ****P* < 0.001. CI95%, 95% confidence interval. (**H**) dN/dS ratios for the three genes under significant positive selection in MNU. (**I**) Mutational spectra of putative driver mutations. Mut_no, mutation number.

mutational burdens comparable to tumors and with mutations in *TP53*, *KMT2D*, and *KDM6A* (Fig. 2, A and B). For example, sample P6U5 had 1978 mutations and harbored driver mutations in *KMT2D* and *TP53* but displayed no obvious CNAs across its genome (fig. S6; fraction of nondiploid genome: 0.1%). CNAs in MNU were sporadic and largely confined to small-scale genomic regions, along with copy-neutral loss of heterozygosity (LOH) (Fig. 2C and fig. S6). We further explored how accumulation of somatic mutations and CNAs coordinate in UCC and MNU tissues. Notably, we found that some MNU tissues, especially those exposed to AA, had mutational burdens similar to or even higher than those of tumors, but the vast majority of their genomes remained diploid (Fig. 2D). This finding implies that acquisition of CNAs occurs late in clonal expansion in the urothelium and that genomic stability is a choke point for the final malignant transformation.

Mutational burden and mutant clone expansion in MNU

To gain deeper insights into the mutational burden and mutant clonal expansion in MNU, we used the variant allele fractions of somatic mutations to estimate the mutant cell fraction (MCF) and clone size in MNU samples. Overall, the mutational burden was markedly different between AA-associated and non-AA-associated MNU (Fig. 3A). Mutational burdens in the AA-associated MNU ranged broadly, with the median (2.2 mutations per Mb) being higher than that in breast cancer (0.9 mutations per Mb) and kidney clear cell carcinoma (1.5 mutations per Mb). The mutational burden of non-AA-associated MNU was more than one order of magnitude lower than that of AA-associated MNU (Fig. 3A). P65U stood out among non-AA-associated MNU with an extraordinarily high mutational burden (~62 mutations per Mb) (Fig. 3A). We further explored the mutational signature of this hypermutator and found that it conformed to the COSMIC SBS10a and SBS10b (cosine similarity = 0.86) (Fig. 3B and fig. S7, A and B). The mutational process underlying this signature has been implicated in altered activity of the central DNA polymerase POLE (*35*, *36*). We detected a canonical mutation in the *POLE* proofreading domain (Pro^{286}→Arg) that could lead to this mutational signature in the sample (fig. S7C). To our knowledge, there have been no previous reports of a mutational process associated with mutated *POLE* occurring in normal human tissues.

Next, we characterized mutant clonal expansion in MNU tissues. Overall, the distribution of mutant clone sizes exhibited a long tail (Fig. 3C and table S12). Upon further dividing samples into AA-associated and non-AA-associated MNU, we observed a bimodal distribution of mutant clone sizes in both groups (Fig. 3C). The peak corresponding to larger clone sizes in AA-associated MNU was due mainly to two samples, P4U1 and P7U6, while in the non-AA-associated distribution, the peak corresponding to larger clone sizes was almost exclusively due to sample P65U. This observation demonstrates that drastic clonal expansions have occurred in some individual MNU tissues. One possible explanation for this observation is that urothelial cells from the same clone have acquired more than one driver mutation, which confers proliferative and competing advantages. To test this hypothesis, we deduced the co-occurrences of driver mutations in those clones using the pigeonhole principle (*37*). Indeed, more than one driver mutation was nested in single clones in each of the above three samples (P4U1, P7U6, and P65U) (Fig. 3D). For example, mutations in *KDM6A*, *TP53*, *KMT2D*, and *ARID1A* were simultaneously acquired by a single clone in P4U1 (Fig. 3D). Compared with non-AA-associated samples, AA-associated MNU had a significantly larger median mutant clone size ($P < 0.001$, Wilcoxon rank-sum test), even when we excluded the three samples with drastic clonal expansions (Fig. 3E). Additionally, we found that AA-associated MNU displayed greater mutational burdens (Fig. 3, A and F). Altogether, these data indicate that AA exposure considerably accelerates somatic mutation accumulation and enhances clonal expansions in normal urothelium.

Positive selection of somatic mutations provides the necessary fuel for clonal expansion. We compared the clone size of driver mutations with those of synonymous mutations in non-AA-associated MNU, which were considered as passengers under neutral selection (table S13). As expected, clones with driver mutations were larger than those with passengers (Fig. 3G). However, statistically significant differences were observed only in *KMT2D*, *CREBBP*, *ATM*, and *KMT2C*, not in canonical driver genes such as *TP53* (Fig. 3G, $P < 0.05$, Wilcoxon rank-sum test). This observation is similar to previous findings and is likely attributable to putative passengers co-occurring with driver mutations in individual clones being hijacked by positive clonal selection (*8*). We next estimated genes under positive selection using a context-dependent dN/dS model (dN/dS is the ratio of the rate of substitution at nonsilent sites versus silent sites) (*38*). Genes with a significant global dN/dS ratio included *KMT2D*, *KDM6A*, and *TP53*, which are the top three recurrently mutated genes in UCC (Fig. 3H and table S14).

A predominance of T:A>A:T transversions was observed in most mutations in driver genes, suggesting that AA mutagenesis in MNU can explain the occurrence of most driver mutations observed in the current study (Fig. 3I). This was further confirmed by analyzing the probability of each mutational signature underlying the driver mutations (table S15). This finding rationalizes our observation that AA exposure largely boosts mutant clone sizes in MNU (Fig. 3, E and F). Unexpectedly, mutations in *KMT2D* were dominant for C:G>T:A transitions rather than T:A>A:T transversions (Fig. 3I). Nearly half of the *KMT2D* mutations (13/28) occurred in non-AA-associated MNU. Even in AA-associated MNU, ~60% of the *KMT2D* mutations were not T:A>A:T transversions, which contrasted greatly with other driver genes (fig. S7, D to F), although these *KMT2D* mutations were still most likely caused by AA mutagenesis (table S15). This finding implies that *KMT2D* mutations may be widely carried by urothelial cells through both intrinsic (e.g., SBS2, SBS13, and SBS5) and exogenously triggered mutational processes (e.g., SBS22). Mutations in this gene may be essential for clonal expansion in urothelial cells, regardless of whether they experience exogenous mutagen exposure.

Competitive mutant clones evolve independently under AA exposure

We sequenced two tumors and three MNU samples from the ureter tract of patient P4. Somatic mutations harbored by the five samples were different from each other, indicating that they evolved independently, as reflected by the phylogenetic tree (Fig. 4A). All five samples displayed clear AA-associated mutational signatures (fig. S8A). Given the different sampling sites, we concluded that AA-triggered mutational processes can spread throughout the entire ureter tract (Fig. 4A). Similar results were also observed in other patients (fig. S8, B and C). Forming competing clones in the ureter tract, each sample independently accumulated driver mutations that were most likely triggered by AA mutagenesis (table S15). These driver mutations may confer competitive advantages on these clones. Among the three MNU samples, we observed putative driver mutations convergent in *KDM6A*, suggesting that mutations in this gene were widespread in MNU and important for early mutant clonal evolution (Fig. 4A). Interestingly, in P4U1, we identified an obvious bimodal distribution of MCFs of somatic mutations (Fig. 4B). The smaller peak, with driver mutations in *KMT2D* and *ARID1A*, was estimated to be a subclone originating from the major clone (the larger peak) on the basis of the pigeonhole principle. Given that most driver mutations in P4U1 were caused by AA mutagenesis (table S15), this observation suggests that mutant subclones originate and evolve in MNU under AA mutagenic stress.

A single AA-associated clone in MNU can expand to a scale of several square centimeters in size

We have demonstrated that AA mutagenesis drives mutant clonal expansion in MNU. However, to what scale an AA-associated mutant clone can expand remains to be elucidated. We sequenced six MNU samples from patient P7 that were extracted from different sites in the bladder (fig. S9A). The mutational burdens of these samples ranged from 1.7 to 7.6 mutations per Mb (fig. S9B). Somatic mutations detected in these samples were entirely different from those in the patient's tumor, indicating their independent clonal origins (fig. S9C). Additionally, we found that the mutational spectra largely matched the AA-associated signature, which was ubiquitous among the six samples (Fig. 4C).

Next, we compared somatic mutations and their MCFs among the six MNU samples from patient P7. Except for P7U3, the other five MNU samples shared 161 somatic mutations with variable MCFs (Fig. 4D), demonstrating that those five MNU samples may originate from the large expansion of a single mutant clone. In this case, we observed five independent *KDM6A* mutations, further demonstrating that

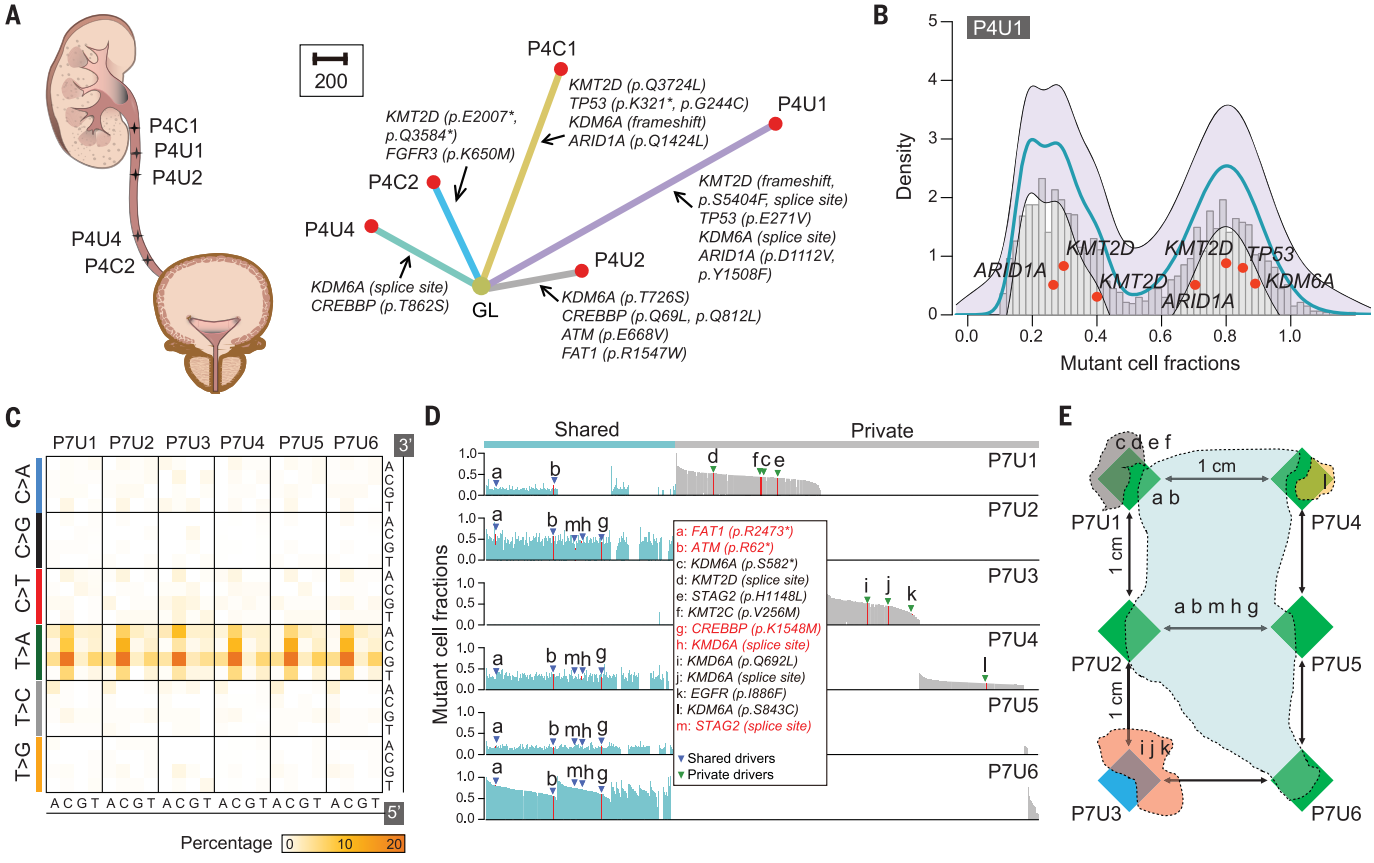

Fig. 4. Mutant clonal evolution in MNU exposed to AA. (**A**) (Left) Sampling sites in patient P4. (Right) Phylogenetic tree depicting the clonal relationships of five samples from P4. Putative driver mutations are labeled on the branches. Asterisks indicate stop codons. Scale bar, 200 mutations. (**B**) Clustering of clone sizes in P4U1 using a Dirichlet process. The blue line indicates the fitted distribution, and the purple region represents the 95% posterior confidence interval. Red dots indicate mutations in putative driver genes. (**C**) Heatmaps showing the proportion of each mutation type within 96 mutational contexts. (**D**) MCFs of somatic mutations in the six MNU samples. Putative driver mutations are highlighted by red vertical lines. (**E**) Schematic displaying the area of different mutant clones, as inferred by mutation overlapping and clustering results. Lowercase letters indicating driver mutations correspond to the table in (D). The differently colored regions represent different mutant clones.

mutations in this gene are widely carried by urothelial cells (Fig. 4D). We next clustered somatic mutations into mutant clones using a Dirichlet process (fig. S10). A single mutant clone with putative driver mutations in *FAT1* and *ATM* was shared by the five MNU samples. This shared clone seemingly derived from a small clone in P7U1 which progressively evolved and acquired additional driver mutations in *CREBBP*, *KDM6A*, and *STAG2* (Fig. 4D and fig. S10A). Another possibility is that an independent mutant clone in P7U1 intermingled with the large clone, which could explain the low MCFs of the shared mutations in P7U1 (Fig. 4D and fig. S10A). A competing mutant clone in P7U3 originated independently and evolved in parallel, acquiring two mutations in *KDM6A* and one in *EGFR* (Fig. 4D and fig. S10A). Taken together and combined with the sampling distances, our results reveal that a single AA-associated mutant clone can expand massively to a scale of several square centimeters in size (Fig. 4E). We observed similar results in another patient sample (fig. S11).

Discussion

Accumulation of mutations in somatic cells has long been implicated in various pathological processes, including human cancer (*2*). However, how and in what patterns somatic mutations occur and drive clonal expansion in normal cells remain largely uncharacterized. Recent genome sequencing studies have revealed landscapes of somatic mutations in various normal tissues, thus broadening our knowledge of mutagenesis in somatic cells (*3*, *8*–*12*, *14*–*17*). Although previous studies identified some mutations with low allele frequencies in normal-appearing urothelium in a limited number of bladder cancer patients (*39*, *40*), our study depicts a comprehensive mutational landscape of human normal urothelium from UCC patients, especially under exogeneous mutagen exposure, and a study by Lawson *et al.* published in this issue investigated somatic mutations in normal bladder urothelium mostly from cancer-free individuals (*41*).

Overall, we observed variable numbers of somatic mutations in MNU using a relatively large sampling size and moderate sequencing depth. We found that macroscopic mutant clones originated in at least some MNU tissues. Acquisition of putative driver mutations in MNU may explain why some mutant clones can expand to a detectable size. Our mutational signature analysis revealed an underlying prevalence of AA mutagenesis in MNU, demonstrating that the mutational process triggered by AA can occur widely in normal human somatic cells in vivo. As AA is prevalent in traditional herbal medicine in Asia (*42*–*45*), our results may reflect the specificity of Chinese and Asian populations. AA exposure boosts somatic mutation accumulation and clonal expansions in MNU; the latter most likely because the mutagenic effects of AA widely cause driver mutations in MNU, facilitating clonal expansion through positive clonal selection. Unexpectedly, we identified drastic expansion of single AA-associated clones to a scale of several square centimeters in size. However, nailing down a precise size of AA-implicated mutagenesis and AA-triggered clonal expansion in urothelium

requires further investigation using denser sampling. A previous study used *N*-butyl-*N*-(4-hydroxybutyl)nitrosamine (BBN) to trigger bladder carcinoma in mice and found that a single basal stem cell could proliferate and colonize the entire urothelium (*46*). Here, we reported multiple independent clones originating in human urothelium under AA exposure. These differences in findings could be explained by the fact that tumorigenesis in humans is considerably more complicated than in mice or by the different degrees of mutagenic stress caused by BBN versus AA.

Chromatin remodeling–related genes, such as *KMT2D* and *KDM6A*, were frequently mutated in MNU, whereas canonical driver genes in UCC, such as *PIK3CA* and *FGFR3*, were rarely mutated. This suggests that epigenetic dysfunction may be critical for early clonal expansion in human urothelium. *TP53* had a clearly low mutation rate. A total of five MNU samples (independent clonal origin) harbored *TP53* mutations, four of which were AA-associated and one of which was the hypermutator P65U (COSMIC SBS10). This finding suggests that either (i) *TP53* is rarely mutated in early clonal evolution in MNU unless there are certain strong mutagenic driving forces or (ii) mutant clones with *TP53* mutations are too small to be detected. The latter possibility can be examined using ultra-sensitive and deep sequencing strategies (*47*). Although the overwhelming mutagenicity of AA caused most of the driver mutations in our study, *KMT2D* mutations were widely observed in other MNU lacking the AA signature, which suggests that mutations in this gene widely occur in urothelial cells, regardless of whether they experience exogenous mutagen exposure.

Unlike somatic mutations, CNAs are relatively rare and less extensive in MNU than in tumors. Even in those MNU with obvious clonal expansion (such as P65U), copy number remained diploid across the entire genome. Similar results were observed in previous studies of normal skin and esophageal tissues (*8*, *9*). These findings may reveal a universal principle in somatic clone expansion, namely that genomic instability is fundamental for malignant transformation.

One possible clinical implication of our study is that more radical treatment strategies may be appropriate in AA-associated UCC patients because their normal-appearing urothelium may harbor high mutational burdens and have undergone mutant clonal expansions. Such conditions may substantially contribute to tumor relapse.

REFERENCES AND NOTES

1. I. Martincorena, P. J. Campbell, *Science* **349**, 1483–1489 (2015).
2. R. A. Risques, S. R. Kennedy, *PLOS Genet.* **14**, e1007108 (2018).
3. I. Martincorena *et al.*, *Science* **362**, 911–917 (2018).
4. C. Tomasetti, B. Vogelstein, G. Parmigiani, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1999–2004 (2013).
5. M. R. Stratton, P. J. Campbell, P. A. Futreal, *Nature* **458**, 719–724 (2009).
6. M. Greaves, C. C. Maley, *Nature* **481**, 306–313 (2012).
7. K. A. Hoadley *et al.*, *Cell* **173**, 291–304.e6 (2018).
8. I. Martincorena *et al.*, *Science* **348**, 880–886 (2015).
9. A. Yokoyama *et al.*, *Nature* **565**, 312–317 (2019).
10. H. Lee-Six *et al.*, *Nature* **574**, 532–537 (2019).
11. S. F. Brunner *et al.*, *Nature* **574**, 538–542 (2019).
12. L. Moore *et al.*, *Nature* **580**, 640–646 (2020).
13. K. Yoshida *et al.*, *Nature* **578**, 266–272 (2020).
14. T. Bae *et al.*, *Science* **359**, 550–555 (2018).
15. M. A. Lodato *et al.*, *Science* **359**, 555–559 (2018).
16. Y. S. Ju *et al.*, *Nature* **543**, 714–718 (2017).
17. H. Lee-Six *et al.*, *Nature* **561**, 473–478 (2018).
18. K. Yizhak *et al.*, *Science* **364**, eaaw0726 (2019).
19. P. E. García-Nieto, A. J. Morrison, H. B. Fraser, *Genome Biol.* **20**, 298 (2019).
20. D. E. Brash *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 10124–10128 (1991).
21. A. Ziegler *et al.*, *Nature* **372**, 773–776 (1994).
22. S. C. Baker, J. Southgate, in *Electrospinning for Tissue Regeneration*, L. A. Bosworth, S. Downes, Eds. (Woodhead Publishing Series in Biomaterials, Woodhead Publishing, 2011), pp. 225–241.
23. C. Varley *et al.*, *Exp. Cell Res.* **306**, 216–229 (2005).
24. G. Guo *et al.*, *Nat. Genet.* **45**, 1459–1463 (2013).
25. A. G. Robertson *et al.*, *Cell* **171**, 540–556.e25 (2017).
26. Cancer Genome Atlas Research Network, *Nature* **507**, 315–322 (2014).
27. J. Michl, M. J. Ingrouille, M. S. Simmonds, M. Heinrich, *Nat. Prod. Rep.* **31**, 676–693 (2014).
28. U. Mengs, *Arch. Toxicol.* **61**, 504–505 (1988).
29. A. W. T. Ng *et al.*, *Sci. Transl. Med.* **9**, eaan6446 (2017).
30. Y. Du *et al.*, *Eur. Urol.* **71**, 841–843 (2017).
31. S. L. Poon *et al.*, *Genome Med.* **7**, 38 (2015).
32. L. B. Alexandrov *et al.*, *Nature* **500**, 415–421 (2013).
33. L. B. Alexandrov *et al.*, *Nature* **578**, 94–101 (2020).
34. C. H. Chen *et al.*, *Int. J. Cancer* **133**, 14–20 (2013).
35. Cancer Genome Atlas Network, *Nature* **487**, 330–337 (2012).
36. C. Kandoth *et al.*, *Nature* **497**, 67–73 (2013).
37. S. Nik-Zainal *et al.*, *Cell* **149**, 994–1007 (2012).
38. I. Martincorena *et al.*, *Cell* **171**, 1029–1041.e21 (2017).
39. M. B. H. Thomsen *et al.*, *Sci. Rep.* **7**, 11702 (2017).
40. T. Majewski *et al.*, *Cell Rep.* **26**, 2241–2256.e4 (2019).
41. A. R. J. Lawson *et al.*, *Science* **370**, 75–82 (2020).
42. L. Vaclavik, A. J. Krynitsky, J. I. Rader, *Food Addit. Contam. Part A Chem. Anal. Control Expo. Risk Assess.* **31**, 784–791 (2014).
43. H. Y. Yang, P. C. Chen, J. D. Wang, *BioMed Res. Int.* **2014**, 569325 (2014).
44. T. P. Cheung, C. Xue, K. Leung, K. Chan, C. G. Li, *Clin. Toxicol.* **44**, 371–378 (2006).
45. M. N. Lai, S. M. Wang, P. C. Chen, Y. Y. Chen, J. D. Wang, *J. Natl. Cancer Inst.* **102**, 179–186 (2010).
46. K. Shin *et al.*, *Nat. Cell Biol.* **16**, 469–478 (2014).
47. J. J. Salk *et al.*, *Cell Rep.* **28**, 132–144.e3 (2019).
48. R. Li *et al.*, supplementary of UTUC-Normal, Version V1, Zenodo (2020); https://doi.org/10.5281/zenodo.3966801.
49. R. Li *et al.*, supplementary of UTUC-Normal 2, Version V1, Zenodo (2020); https://doi.org/10.5281/zenodo.3726413.

ACKNOWLEDGMENTS

We thank all the patients for their consent and participation in this study. We thank X. M. Wang from Peking University for the helpful discussion on mutational signature analysis. We thank T. S. Sun from Peking University for helping create the bladder cartoon. **Funding:** This work was financially supported by the National Key Research and Development Program (2018YFA0902802), the National Science and Technology Major Project (2018ZX10302205 and 2019YFC1315702), the National Natural Science Foundation of China (31722003, 31770925, and 81802533), the Guangdong Province Key Research and Development Program (2019B020226002), and the Beijing Municipal Science and Technology Commission (Z191100006619010). R.L. was funded by the BoYa Postdoctoral Fellowship of Peking University. **Author contributions:** F.B., T.X., R.L., Y.D., Z.C., and D.X. designed the experiments. R.L. led the data analysis with help from Z.C. and Z.L. Y.D., T.L., and X.C. collected the samples with help from G.W. Y.D., M.L., and G.W. performed pathological examinations. Z.C. and D.X. performed experiments with help from S.J. and Y.D. R.L. and F.B. wrote the manuscript with help from Y.D. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The sequencing raw data generated in this study have been deposited in the Genome Sequence Archive (GSA) in BIG Data Center (https://bigd.big.ac.cn), Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, under accession number HRA000137 (https://bigd.big.ac.cn/gsa-human/s/A88e89NG). Data from the Chinese bladder cancer study that were used in the MutSigCV analysis are deposited in the Sequence Read Archive (SRA) (www.ncbi.nlm.nih.gov/sra) under accession number SRA063495 (*24*). The input data for Sequenza, ReCapSeq, ABSOLUTE, mutational signature analysis, Dirichlet process analysis, and VCF files of all somatic mutations; the raw outputs from Sequenza, ABSOLUTE, and Dirichlet process analysis; and the code for Dirichlet process analysis are all available at Zenodo (*48*, *49*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6512/82/suppl/DC1
Materials and Methods
Figs. S1 to S11
Tables S1 to S15
References (*50*–*65*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

30 December 2019; accepted 4 August 2020
10.1126/science.aba7300

Science

Macroscopic somatic clonal expansion in morphologically normal human urothelium

Ruoyan Li, Yiqing Du, Zhanghua Chen, Deshu Xu, Tianxin Lin, Shanzhao Jin, Gongwei Wang, Ziyang Liu, Min Lu, Xu Chen, Tao Xu and Fan Bai

Science **370** (6512), 82-89.
DOI: 10.1126/science.aba7300

Genetic profiles of the bladder

Depending on the environment of the individual, the human bladder can be exposed to carcinogens as they are flushed through the body. Lawson *et al.* and Li *et al.* examined the genetic composition of laser-dissected microbiopsies from normal and cancer cells collected from the urothelium, a specialized epithelium lining the lower urinary tract (see the Perspective by Rozen). These complementary studies identified the mutational landscape of bladder urothelium through various sequencing strategies and identified high mutational heterogeneity within and between individuals and tumors. Both studies identified mutational profiles related to specific carcinogens such as aristolochic acid and the molecules found in tobacco. These studies present a comprehensive description of the diverse mutational landscape of the human bladder in health and disease, unraveling positive selection for cancer-causing mutations, a diversity of mutational processes, and large differences across individuals.

Science, this issue p. 75, p. 82; see also p. 34