

Mutational Context and Diverse Clonal Development in Early and Late Bladder Cancer

Iver Nordentoft,^{1,6} Philippe Lamy,^{1,6} Karin Birkenkamp-Demtröder,^{1,6} Karey Shumansky,² Søren Vang,¹ Henrik Hornshøj,¹ Malene Juul,¹ Palle Villesen,³ Jakob Hedegaard,¹ Andrew Roth,² Kasper Thorsen,¹ Søren Høyer,⁴ Michael Borre,⁵ Thomas Reinert,¹ Niels Frstrup,⁵ Lars Dyrskjöt,¹ Sohrab Shah,² Jakob Skou Pedersen,^{1,7} and Torben F. Ørntoft^{1,7,*}

¹Department of Molecular Medicine, Aarhus University Hospital, Brendstrupgaardsvej 100, 8200 Aarhus, Denmark

²BC Cancer Research Centre, 675 West 10th Avenue, Vancouver, BC V5T4E6, Canada

³Department of Bioinformatic Research, Aarhus University, Universitetsparken, 8000 Aarhus, Denmark

⁴Department of Pathology, Aarhus University Hospital, North Road Ringgade 8000 Aarhus, Denmark

⁵Department of Urology, Aarhus University Hospital, Brendstrupgaardsvej 100, 8200 Aarhus, Denmark

⁶Co-first author

⁷Co-senior author

*Correspondence: orntoft@ki.au.dk

<http://dx.doi.org/10.1016/j.celrep.2014.04.038>

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

SUMMARY

Bladder cancer (or urothelial cell carcinoma [UCC]) is characterized by field disease (malignant alterations in surrounding mucosa) and frequent recurrences. Whole-genome, exome, and transcriptome sequencing of 38 tumors, including four metachronous tumor pairs and 20 superficial tumors, identified an APOBEC mutational signature in one-third. This was biased toward the sense strand, correlated with mean expression level, and clustered near breakpoints. A > G mutations were up to eight times more frequent on the sense strand ($p < 0.002$) in [ACG]AT contexts. The patient-specific APOBEC signature was negatively correlated to repair-gene expression and was not related to clinicopathological parameters. Mutations in gene families and single genes were related to tumor stage, and expression of chromatin modifiers correlated with survival. Evolutionary and subclonal analyses of early/late tumor pairs showed a unitary origin, and discrete tumor clones contained mutated cancer genes. The ancestral clones contained *Pik3ca/Kdm6a* mutations and may reflect the field-disease mutations shared among later tumors.

INTRODUCTION

Bladder cancer is the fifth most common cancer in western countries and is costly to treat because it usually involves multiple recurrences over several years. Approximately 75% of patients have a single or multiple non-muscle-invasive urothelial cell cancers (UCCs), and 10%–15% of these patients progress. The remaining 25% have a muscle-invasive cancer as the first lesion (Babjuk et al., 2013). The molecular events that lead to formation of field cancerization in urothelium (Majewski

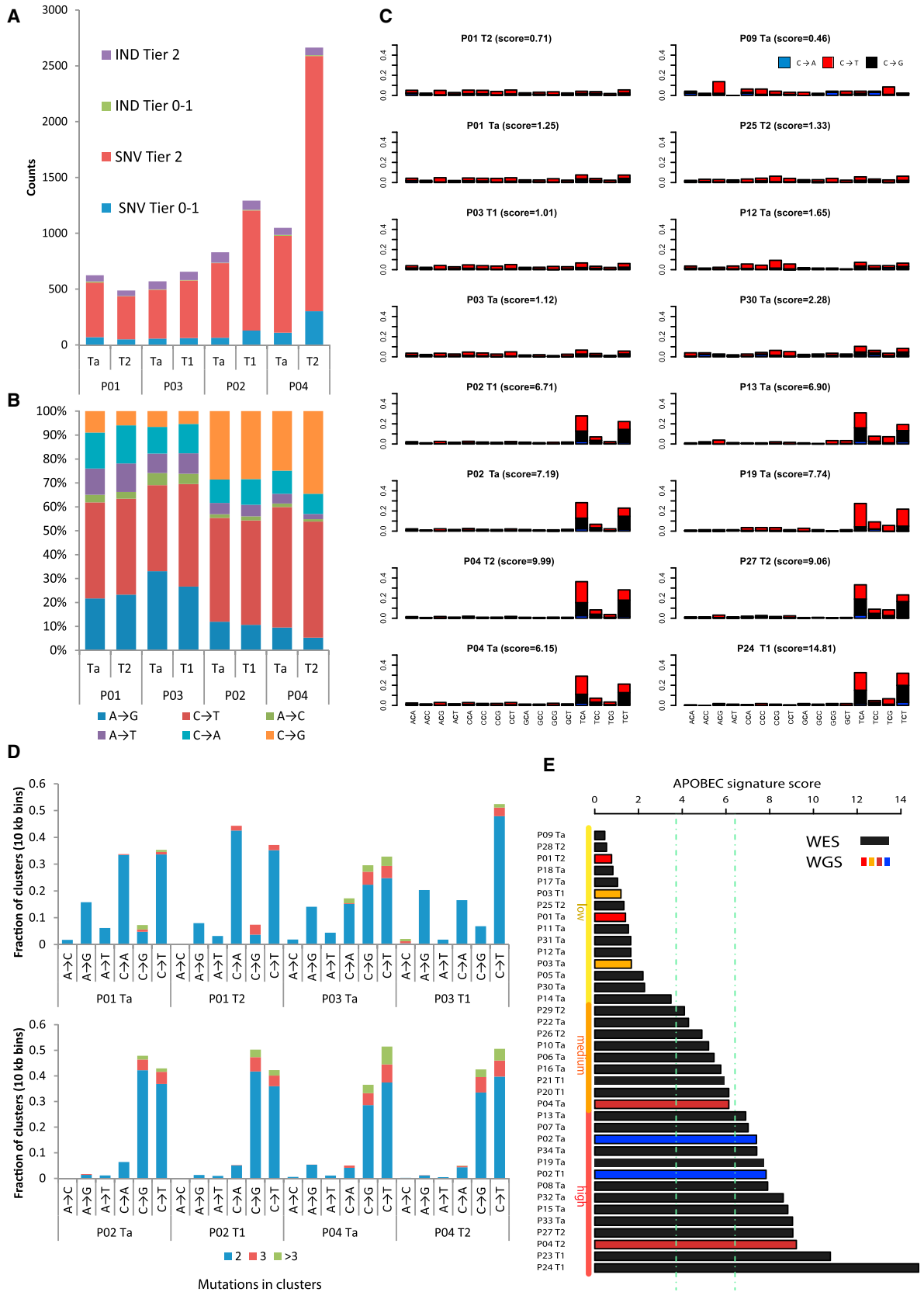
et al., 2008), disease progression, and muscle invasion are poorly understood, leaving few options for rational treatment.

Recently, the mutational events in mainly muscle-invasive UCC have been described (Alexandrov et al., 2013; Balbás-Martínez et al., 2013; Gui et al., 2011; Guo et al., 2013; Iyer et al., 2013; Kandoth et al., 2013a; Ross et al., 2014). Apart from the well-known frequent mutations in tumor suppressors and oncogenes, mutations in many chromatin-modifying genes, as well as in *Stag2*, were frequent. Invasive UCC has a relatively high prevalence of somatic mutations that reflect an extensive heterogeneity in UCC, with mutations scattered over many different groups of gene functions, in contrast to, e.g., colorectal cancer, where *Apc*, *Tp53*, and *Kras* dominate (Kandoth et al., 2013a). Insight into non-muscle-invasive, low-grade UCC by massively parallel sequencing has only been reported in a few exomes (Balbás-Martínez et al., 2013), and whole-genome data have not yet been reported for this tumor type.

Mutagenesis in UCC was recently suggested to reflect the effect of the APOBEC family of cytidine deaminases (Burns et al., 2013b; Roberts et al., 2013). Mainly the enzyme APOBEC3B provides a frequent mutational signature (Burns et al., 2013a). The APOBEC enzyme has a mutational specificity for the motif TCW, where W is T or A. This trinucleotide motif showed a highly significant frequency of mutations in UCC, where C mutated to T or G (Burns et al., 2013a; Roberts et al., 2013). The signature was related to frequent mutations and the presence of strand-coordinated clusters of cytosine mutations (Burns et al., 2013b). However, no detailed relation to individual UCC tumor characteristics has been presented.

It is important to define cancer cell subpopulations to further our understanding of alterations that occur among clonal subpopulations during disease progression (Aparicio and Caldas, 2013; Kandoth et al., 2013b). In UCC, characterization of cellular subpopulations by sequencing is needed to gain insight into the field disease of the mucosa and the population dynamics of progression.

In this work, we sought to improve our understanding of genomic alterations in low-grade and -stage UCCs, and clonal composition in superficial and invasive cancers, with a special



(legend on next page)

emphasis on the relation between the APOBEC signature and genomic and clonal changes. We performed whole-genome sequencing (WGS) of pairs of a superficial noninvasive tumor (Ta) and a later mucosa- or muscle-invasive tumor (T1/T2) from four patients, together with matched normal blood samples (paired-WGS set; [Table S1A](#)). The metachronous tumor pairs were separated by 323–997 days and were also subjected to deep-exome sequencing to gain insight into cellular subpopulations ([Shah et al., 2012](#)). These data were supplemented with whole-exome sequencing (WES) on a new set of 30 mainly low-stage tumors (20 Ta, 5 T1, 5 T2–4, WES set; [Tables S1B and S1C](#)). RNA sequencing (RNA-seq) and SNP arrays provided further data on molecular alterations.

We observed a strong APOBEC signature in one-third of cases, which was present in both early and late tumors from single individuals. The *ApoBec3b* gene transcript was positively correlated to high stage and increased in lymph node metastases. We found frequent mutations in epigenetic modifiers and cell membrane/cell adhesion proteins. The former group mutated early in the disease course and at protein level EP300 provided an independent prognostic biomarker. Neither single nucleotide variants (SNVs) nor clinicopathological parameters were related to the APOBEC signature. The subclonal architecture showed an ancestral common clone, a progression clone, and a nonaggressive clone in all patients with paired tumors. Many clones contained both mutated tumor suppressors and oncogenes. The UCC ancestral clones, defining the field disease, were driven mainly by *Pik3ca* and *Kdm6a* mutations, which thus represent potential therapy targets.

RESULTS

Mutations in the Paired-WGS and WES Sets

From each tumor of the paired-WGS set, we generated ~2,200 million paired sequences of 101 basepairs (bp), resulting in a genome-wide average read depth of $49 \times (35.5\text{--}65.7)$ ([Figure S1A](#); [Table S1D](#); [Supplemental Experimental Procedures](#)). This was supplemented with deep whole-exome sequencing to a depth of $311 \times (181\text{--}405)$, which was mainly used for cell population studies in the context of clonal evolution ([Table S1E](#)). We identified 13,443–56,140 SNVs and 2,887–5,253 indels per sample across the whole genome, and categorized these by strength of evidence (from Cat 1 [high frequency and high

score] to Cat 3 [low frequency and low score]), and by expected functional impact (Tier 0: loss of function; Tier 1: missense; Tier 2–4: synonymous and noncoding; see [Supplemental Experimental Procedures](#); [Tables S2A–S2C](#)). In the analyses presented below, we focus on the Cat 1–2 SNVs.

In protein-coding regions of the WGS set, we found 50–302 SNVs with a protein-level effect (Tiers 0 and 1) per sample ([Figure 1A](#); [Table S2A](#)), with a median mutation rate of 1.8 SNVs/Mb (1.33–8.05). In neutral regions, away from protein-coding genes and evolutionarily conserved elements, the mutation rate was 3.1 SNVs/Mb excluding the sample P04 T2, which is an outlier with a much elevated mutation rate of 9.2 SNVs/Mb. Within protein-coding regions, we separately estimated the synonymous mutation rate (dS = 3.7 SNVs/Mb) and the non-synonymous mutation rate (dN = 2.7 SNVs/Mb), accounting for the number of relevant sites. A dN/dS ratio (0.74) showed that purifying selection overall dominated in protein-coding regions, although a relaxed selection pressure on some genes combined with driver mutations resulted in a much higher ratio than is observed in human evolution (dN/dS = 0.25) ([Scally et al., 2012](#)).

In the WES set, 30 tumors and matched germline DNA were sequenced to an average depth of $47 \times (31\text{--}126)$ ([Table S1F](#)). In the protein-coding target regions, we observed an average of 195 (26–779) SNVs and 23 indels (7–60) per sample. An average of 79 (12–316) genes was affected at the protein level, with 12 (1–34) genes inactivated by loss of function. The paired WGS samples similarly had an average of 15 (6–37) genes with loss-of-function mutations, in line with the rate found previously ([Gui et al., 2011](#)). Overall, a total of 2,537 genes were affected at the protein level and 402 were inactivated across all samples.

RNA-seq ([Figure S1B](#); [Table S1G](#)) and Sanger sequencing of PCR products validated >90% of SNVs and indels ([Supplemental Experimental Procedures](#); [Figures S1C–S1F](#); [Tables S2D–S2F and S3A](#)). The observed mutation rate is similar to that found in previous studies ([Table 1](#)), but lower than those reported in recent studies using different algorithms and mainly based on muscle-invasive UCC (6–8/Mb), which had a higher mutation rate ([Alexandrov et al., 2013](#); [Kandoth et al., 2013a](#)).

A Strong APOBEC Mutagenesis Pattern Is Observed in One Out of Three UCCs

To gain more detailed insight into the mutational process, we analyzed the 5' and 3' contexts of the mutated bases. We

Figure 1. Overview of Mutations and APOBEC Signature Score

- (A) Distribution of SNVs and indels (Tier 0–2, Cat 1–2) for the paired-WGS samples.
- (B) Mutation type profile for the paired-WGS samples. For each sample, we made a cumulative bar plot showing each type of transitions or transversions.
- (C) For each sample and for each trinucleotide with a cytosine as a middle base, the proportion of observed mutations was calculated. This was done on the neutral genomic regions for the WGS set taking into account only Cat 1 SNVs (left), and on the exome-capture regions for eight WES samples taking into account Cat 1–2 SNVs (right, color code for mutations upper right). Four WES samples with low APOBEC scores and four with high APOBEC scores were selected. The score (shown in parentheses) was defined as the normalized ratio of the number of mutations from C to G or T in TCW context to the number outside of TCW context.
- (D) Fraction of mutation clusters by mutation type for somatic mutations from eight WGS samples. Mutation clusters are defined as genomic windows of size 10 kb containing at least two mutations of the same type (A → C, A → G, A → T, C → A, C → G, or C → T). Colors indicate further grouping of the fractions into number of mutations.
- (E) An APOBEC score was calculated for each sample in the WES and WGS sets. The score reflects the strength or weakness of the APOBEC mutational (Cat 1–2) signature in the exome-capture genomic regions. The samples were further subdivided into three groups based on the 35th and 65th percentiles (see also [Figure S11](#) for the APOBEC mutational signature in the neutral genomic regions in the WGS set and the TCGA set). See also [Figure S1](#) and [Tables S1](#) and [S2](#).

Table 1. Mutations Found by Deep Sequencing of UCC in Five Studies and the TCGA Set

Hugo Symbol	Nonaggressive; Noninvasive, Low-Grade UCC (Ta Grades I and II)			Aggressive; High-Grade Atypia or Invasive UCC (Ta Grade III, T1–T4)						
	This Study	Balbás-Martínez et al. (2013) ^a	Guo et al. (2013)	This Study	Balbás-Martínez et al. (2013) ^a	Guo et al. (2013)	Ross et al. (2014)	Iyer et al. (2013)	TCGA	
	n = 20	n = 25	n = 5	n = 18	n = 51 PCR validation		n = 93	n = 35	n = 95	n = 126
					TagIII(8)				T1(11)	
					T1(32)				T2(15)	
					T2(9)				T3(33)	
					T3(2)				T4(32)	
									CIS(4)	
KDM6A*	13(65)	3(12)	3(60)	6(33.3)	7(13.7)	29(31.2)	10(29)			37(29.36)
FGFR3	8(40)	10(40)	1(20)	5(27.8)	4(7.8)	10(10.8)	4(11)	12(13)		17(13.49)
ARID1A	7(35)	3(12)	2(40)	3(16.7)	7(13.7)	13(14.0)	7(20)			42(33.3)
PIK3CA	5(25)	5(20)	2(40)	6(33.3)	4(7.8)	11(11.8)	9(26) ^b	17(18)		29(23)
EP300	5(25)	3(12)	2(40)	1(5.6)	4(7.8)	11(11.8)				22(17.46)
ELF3	5(25)	1(20)	0(0)	3(16.7)		6(6.45)				15(11.9)
STAG2	5(25) ^d	6(24)	0(0)	2(11.1) ^d	5(9.8)	11(11.8)				18(14.3)
MLL3	4(20)		0(0)	5(27.8)	1/12(8)	5(5.4)				35(27.8)
RARG	4(20)		0(0)	1(5.6)		4(4.3) ^e				10(7.9)
CREBBP	4(20)	4(16)	1(20)	4(22.2)	7(13.7)	14(15.1)				19(15.1)
RBM10	4(20)	1(20)	0(0)	3(16.7)	1(8)	3(3.2) ^e				8(6.3)
ZFP36L1	3(15)		0(0)	1(5.6)		9(9.7) ^e				8(6.3)
RANBP2	3(15)		0(0)	1(5.6)		3(3.2) ^e				11(8.7)
XIRP2	3(15)		1(20) ^e	1(5.6)		10(10.8) ^e				14(11.1)
MLL2	3(15) ^d	6(24)	0(0)	1(5.6) ^d	5(9.8)	1(1.1) ^e	1(2.9)			46(36.5)
HRAS	2(10)		1(20)	1(5.6)		12(12.9)	1(2.9)			7(5.6)
ASH1L	2(10)		0(0)	1(5.6)		2(2.15) ^e				12(9.5)
FAT1	2(10)			4(22.2)	2(16)					16(12.7)
DCHS2	2(10)		0(0)	2(11.1)	1(8)	6(6.45) ^e				8(6.3)
CACNA1D	2(10)	1(20)	0(0)	2(11.1)		1(1.1) ^e				4(3.2)
C1orf173	2(10)		0(0)	2(11.1)		2(2.15) ^e				7(5.6)
LRRC7	2(10)		0(0)	3(16.7)	1(8)	1(1.1) ^e				12(9.5)
VCAN	2(10)		0(0)	1(5.6)	1(8)	2(2.15) ^e				13(10.3)
ZFYVE26	2(10)		0(0)	1(5.6)		2(2.15) ^e				6(4.8)
TP53*	1(5)	2(8)	0(0)	10(55.6) ^f	9(17.6) ^f	24(25.8) ^f	19(54) ^f	32(34) ^f		77(61.1) ^f
RB1	1(5)	1(4)	1(20)	4(22.2)	3(5.9)	12(12.9)	6(17)	15(16) ^f		22(17.46) ^f
OSMR	1(5)		0(0)	0(0)		1(1.1) ^e				3(2.4)
PCDHA9	1(5)		0(0)	2(11.1) ^f		2(2.15) ^e				9(7.14) ^f
TSC1	1(5)		0(0)	3(16.7) ^f		5(5.4)	2(6)	7(7)		11(8.7)
NEB	1(5)			2(11.1) ^f	1(8)					19(15.1) ^f
OBSCN	1(5)			2(11.1) ^f	1(8)					15(11.9) ^f
PDZD2	1(5)		0(0)	3(16.7) ^f	2(16)	6(6.45)				8(6.3)
LGALS8	1(5) ^d			1(5.6) ^d						0(0)
ATM	1(5) ^d	4(16)	0(0)	0(0)	4(7.8)	5(5.4)				20(15.8)
MYCBP2	1(5) ^d	2(8)	0(0)	0(0)	2(3.9)	3(3.2)				11(8.7)
FANCA	1(5) ^d	0(0)	0(0)	0(0)	4(7.8)	2(2.15) ^e	1(2.9)			6(4.8)
CPAMD8	0(0)		0(0)	3(16.7) ^f		4(4.3) ^{e,f}				3(2.4) ^f
BRAF	0(0)	2(8)	0(0)	0(0)	5(9.8) ^f	4(4.3) ^f	2(6) ^f	2(2.1) ^{b,f}		3(2.4) ^f
CCND1							5(14) ^{c,f}	14(14.7) ^{c,f}		2(1.6) ^f

(Continued on next page)

Table 1. Continued

		Nonaggressive; Noninvasive, Low-Grade UCC (Ta Grades I and II)			Aggressive; High-Grade Atypia or Invasive UCC (Ta Grade III, T1–T4)				
This Study		Balbás-Martínez et al. (2013) ^a	Guo et al. (2013)	This Study	Balbás-Martínez et al. (2013) ^a	Guo et al. (2013)	Ross et al. (2014)	Iyer et al. (2013)	TCGA
n = 20		n = 25	n = 5	n = 18	n = 51 PCR validation		n = 93	n = 35	n = 95
Hugo	TagI(2)	TagI(14)	TagI(3)	TagIII(4)	TagIII(8)	T1(32)			T1(11)
Symbol	TagII(18)	TagII(11)	TagII(2)	T1(7)	T1(32)	T2(28)			T2(15)
				T2(7)	T2(9)	T3(21)			T3(33)
					TIS(2)	T4(12)	T4(35)		T4(32)
									CIS(4)
CCND3			0(0)			1(1.1) ^{e,f}	3(11) ^{c,f}		4(3.2) ^f
CDKN2A/B			0(0)			1(1.1) ^{e,f}	8(23) ^f	24(25.3) ^f	13(10.3) ^f
EGFR			0(0)			1(1.1) ^f	2(6) ^{b,f}		3(2.4) ^f
ERBB2	0(0)		0(0)	2(11.1) ^{d,f}		5(5.4) ^f	2(6) ^{b,f}	5(5.3) ^{c,f}	12(9.5) ^f
ERCC2	0(0)	5(20)	1(20)	0(0) ^f	1(2.0) ^f	6(6.45)			16(12.7) ^f
FGFR1		0(0)	0(0)		0(0)	1(1.1) ^{e,f}	5(14) ^{b,f}	5(5.3) ^{c,f}	5(4) ^f
INADL	0(0)			2(11.1) ^f					5(4) ^f
LPHN3*	0(0)			4(22.2) ^f	2(16)				7(5.6) ^f
MAPK8IP3	0(0)	3(12)	0(0)	0(0) ^f	2(3.9) ^f	1(1.1) ^{e,f}			4(3.2) ^f
MCL1							4(11) ^{b,f}		3(2.4) ^f
MDM2			0(0)			2(2.15) ^{e,f}	4(11) ^{c,f}	5(5.3) ^{c,f}	
NF1	0(0)	1(20)	1(20)	1(5.6) ^{d,f}	1(8)	6(6.45)	2(6) ^f	2(2.1) ^f	11(8.7) ^f
TNC	0(0)		0(0)	2(11.1) ^f		1(1.1) ^{e,f}			6(4.8) ^f
WNK1*	0(0)		0(0)	4(22.2) ^f	1(8)	2(2.15) ^{e,f}			10(7.9) ^f

Numbers in parentheses indicate %. * $p < 0.05$ comparing nonaggressive and aggressive tumors. See also Table S3.

^aSeventeen tumors were exome sequenced and the remaining tumors were subjected to targeted PCR-based sequencing. Figures in italic are from exome-seq screen only.

^bGene amplifications included as event.

^cAll events are gene amplifications.

^dNo validation performed; Tier 0-1, Cat 1-2.

^ePredicted, not validated.

^fMutation frequencies $\times 2$ or more in invasive/high grade compared with non-muscle-invasive low grade.

evaluated the enrichment of the observed contexts relative to what would be expected if mutations were context independent and just followed the genomic background (Nik-Zainal et al., 2012; Figure S1G). Overall, C > T mutations were the most common across all WGS tumors (Figure 1B). As expected, CpG sites, which when methylated can spontaneously deaminate to TpG, accounted for the majority of these and were consistently over-represented across all tumors with 5.8- to 14-fold enrichment (all $p < 10^{-61}$, binomial [BN] test). In addition, CpG to ApG were also consistently enriched, by up to 6-fold (2.5–6.0; $p = 10^{-2}$ to 10^{-30} , BN test; Figure S1H; Tables S4A–S4H).

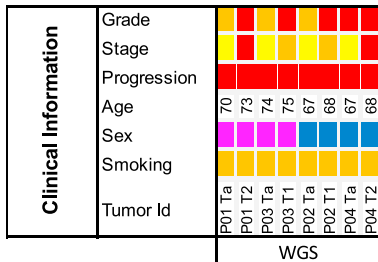
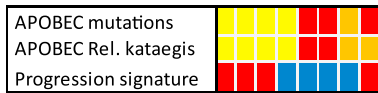
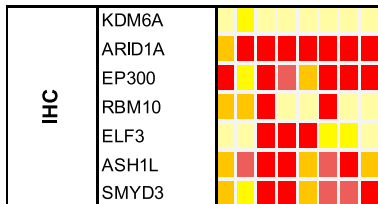
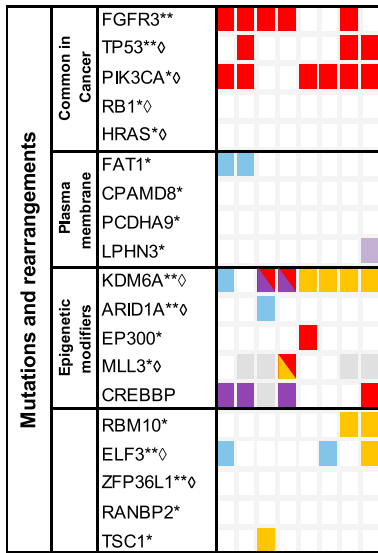
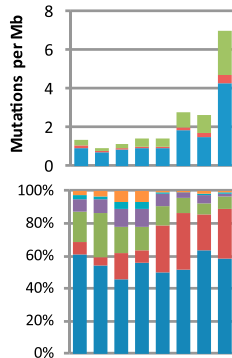
In patients P02 and P04, C > G transversions were more than four times as common as in the other two patients, P01 and P03 (Figure 1B). This was driven by a >15-fold higher C > G mutation rate at the 3' end of TpC dinucleotides ($p < 10^{-16}$, BN test; Tables S4A–S4H). In the other two patients (P01 and P03), the C > G transversions showed little or no increase in prevalence at TpC sites.

These context preferences correspond to recent descriptions of an APOBEC mutational signature (Burns et al., 2013b; Roberts et al., 2013). APOBEC3B mutagenesis leads to frequent C > T

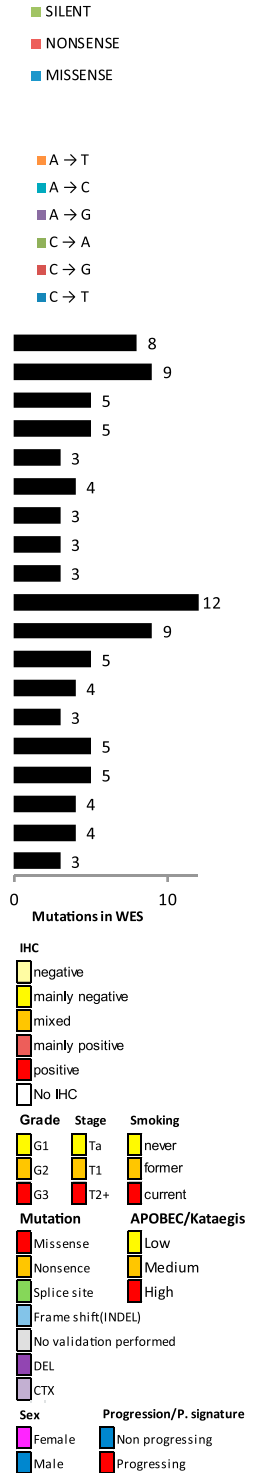
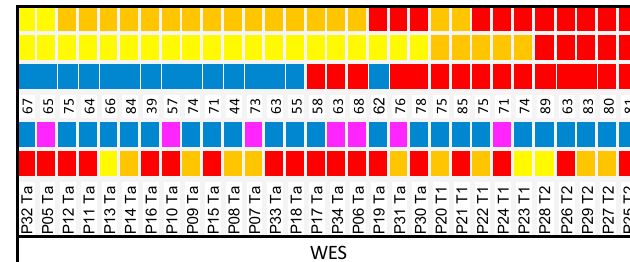
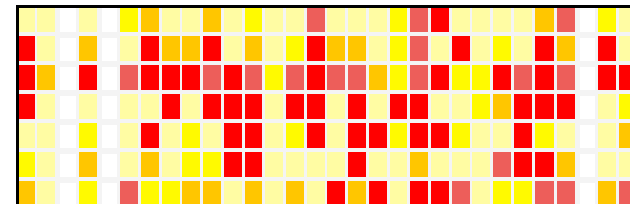
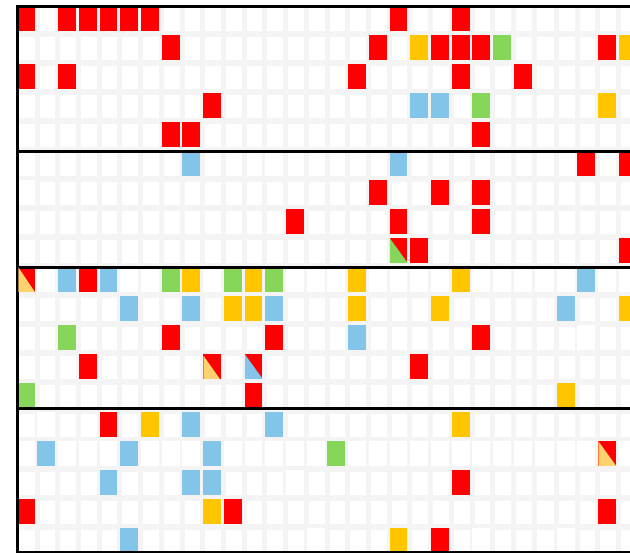
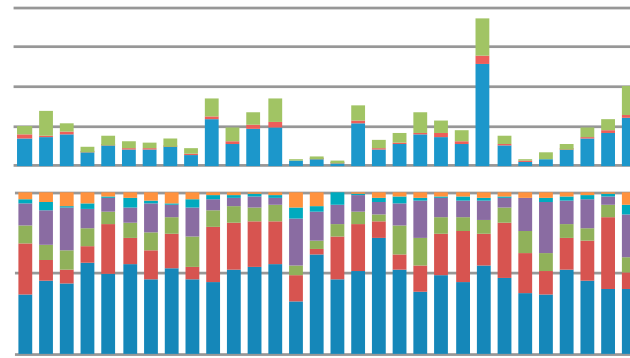
and C > G mutations at a TCW motif, where W is A or T (Roberts et al., 2013). We searched specifically for this mutational signature, eliminating overlap with the highly mutable CpG sequence, in the paired-WGS set. There was a clear enrichment of the APOBEC signature mutations in the four tumors from patients P02 and P04, in contrast to the four tumors from patients P01 and P03 (Figures 1C and 2A). Previously, this signature has been related to strand-coordinated clusters of cytosine mutations, defined as two or more mutation events per 10 kb window, and named “kataegis” (Alexandrov et al., 2013). Clusters with C > T and C > G mutations clearly dominated in the four tumors with the APOBEC signature (Figure 1D).

We searched for the APOBEC signature in our WES set and found a strong signature in 11/30 tumors, a medium signature in 8/30, and a low signature in 11/30 tumors (Figure 1E). Finally, we downloaded The Cancer Genome Atlas (TCGA) bladder cancer WGS set (<http://cancergenome.nih.gov/>) and found a strong APOBEC signature in 25%, a medium signature in 22%, and a low signature in 53% of the 95 tumors (Figure S1I). We conclude that about one-third of UCCs carry an APOBEC cytidine deaminase mutagenesis pattern. We examined the

A



B



(legend on next page)

exomes for APOBEC-related mutation clusters and found evidence for these in 5/26 cases, all overlapping with the APOBEC signature (Figures 2 and S1J; $p < 0.03$, including WGS-set patients). We conclude that clustered C > T and C > G mutations reflect the APOBEC mutagenesis in individual cases of UCC. Both the mutational signature and clustering were strikingly similar in early and late tumors from an individual patient, indicating a stable mutagenesis mechanism over time (Figures 1C, 1D, S1G, and S1H). Similarly, the number of mutations generally varied greatly between early and late samples; however, their composition was remarkably similar in the individual patient (Figures 1A and 1B; Tables S2E, S2F, and S4A–S4H).

Common Mutations in Coding Regions of Nonaggressive and Aggressive UCCs Are Independent of the APOBEC Signature

A total of 26 genes were significantly mutated in the WES set by high-impact SNVs or indels (Tier 0-1) after correction for various parameters, including gene size and background mutation rate (Figures 2B and S1F; Table S3A). The most frequently mutated genes were *Kdm6a* (47%), *Tp53* (32%), *Arid1a* (29%), *Fgfr3* (29%), *Rbm10* (18%), *Fat1* (15%), *Rb1* (15%), *Tsc1* (12%), and *Ranbp2* (12%), using both sets. We evaluated the generality of these findings by comparing them with recent results (Balbás-Martínez et al., 2013; Guo et al., 2013; Iyer et al., 2013; Ross et al., 2014), split into superficial mainly nonaggressive and mainly aggressive tumors (Table 1). All 26 genes were also found to be mutated in either of two large data sets on aggressive UCC containing 96 exomes (Guo et al., 2013) and 126 exomes (TCGA). A few publications have included up to five exomes on nonaggressive tumors (Balbás-Martínez et al., 2013; Guo et al., 2013). Here, we sequenced 20 of these and described the most frequently mutated genes in the aggressive and nonaggressive groups (Table 1). A significant difference was found for *Tp53* ($p < 0.003$), which was mostly mutated in aggressive UCC; *Lphn3* ($p < 0.04$) and *Wnk1* ($p < 0.04$), which were only mutated in aggressive UCC; and *Kdm6a*, which was mostly mutated in nonaggressive UCC ($p < 0.01$).

Most genes were mutated mainly by SNVs, except for *Elf3*, which showed indels in five out of eight cases (Figures 2 and S2C). More generally, the highest dN/dS ratio was observed in many cancer pathways (e.g., bladder, non-small cell, and pancreatic cancer pathways), and a strong degree of purifying selection was found in metabolic pathways (Table S3B). The gene mutations were uncorrelated with both the APOBEC signature and a previously published UCC mRNA progression

signature (Dyrskjøet et al., 2003), as was immunohistochemical staining for the proteins (Figure 2).

We identified 13 gene families that were significantly altered by inactivating SNVs and indels, after correction for gene size and numbers (Table S3C). When these families were grouped into related superfamilies, proteins related to epigenetic modification were most commonly altered, comprising four out of 13 families (chromatin demethylases, methyltransferases, acetyltransferases, and histone proteins; Table S3C). Another commonly mutated superfamily consisted of membrane protein/cell adhesion protein families involved in extracellular interactions (Table S3C). Additional adhesion proteins that were not included in the superfamily were also significantly mutated (LPHN3, FAT1, and INADL). The plasma membrane proteins were significantly more mutated in aggressive tumors ($p < 0.0002$; Figure 2B), and could be involved in the well-known reduced cohesion of high-grade bladder tumors.

To obtain a deeper insight into the prognostic importance of the epigenetic modifiers, we performed immunohistochemistry (IHC) for selected proteins (Figures 2 and 3). For KDM6A and ARID1A, a borderline significant relation to progression-free survival (PFS) (Figures 3A, S2A, S2D, and S2E; Table S5) was observed. The presence of mutations correlated with depleted protein for ARID1A ($p < 0.021$). The histone methyl transferases SMYD3 and ASH1L, and the acetyltransferase EP300 (Figures 3B–3G, S2A, S2B, and S2F) were all changed by at least one SNV/indel and/or copy number increase. Remarkably, strong ASH1L nuclear immunostaining was related to both PFS ($p < 0.024$), disease-specific survival (DSS; $p < 0.004$), and overall survival (OS; $p < 0.006$); negative staining of SMYD3 was related to PFS ($p < 0.046$) and DSS ($p < 0.046$); and positive staining of EP300 was related to PFS ($p < 0.0016$), DSS ($p < 0.029$), and OS ($p < 0.0068$) (Figures 3B–3F, S2A, and S2B; Table S5).

Somatic Rearrangements and Fusion Transcripts

Analysis of somatically acquired genomic rearrangements (Wang et al., 2011) revealed that the four tumors from patients P01 and P04 were most unstable (Figures S3A and S3B). Deletions and inter- and intrachromosomal rearrangements were dominating in all tumors with varying frequencies (40%–70% of rearrangements). Inversions showed only a relatively high frequency in two Ta tumors, and were not seen in the more invasive tumors (Figures S3A and S3B). In many cases, the number of breakpoints was higher in gene regions than in intergenic regions ($0.0009 < p < 0.001$; Figure S3C), probably pointing toward selection. Validation of somatic rearrangements by PCR and Sanger sequencing confirmed 81% of

Figure 2. Landscape of Genetic Variation and Protein Expression in Bladder Cancer

Data tracks (rows) facilitate the comparison of clinical, protein, and genomic data across cases with bladder cancer (columns) in the WGS set (A) and the WES set (B). Samples in the WES set are ordered according to tumor stage. Color codes are shown at bottom right. Upper panel: counts per Mb of mutations and their types. Middle panel: mutational landscape. Mutations were validated by RNA-seq and targeted deep sequencing. Genes followed by * have a p value < 0.01 in the WES; ** marks significance after Benjamini-Hochberg (BH) multiple testing correction in the WES; ◇ marks significance in the TCGA bladder cancer (BLCA) data set (MutSigCV & BH, $p < 0.01$). IHC panel: immunohistochemical staining of individual FFPE sections. APOBEC panel: samples were categorized into three different APOBEC signature groups (see Figure 1E for explanation). APOBEC-related kataegis was grouped according to the number of mutation clusters. The progression signature defines the likelihood of progression based on mRNA transcripts. Clinical sample information appears at the bottom. Right graph: counts of mutations per gene in the WES group.

See also Figure S4 and Table S4.

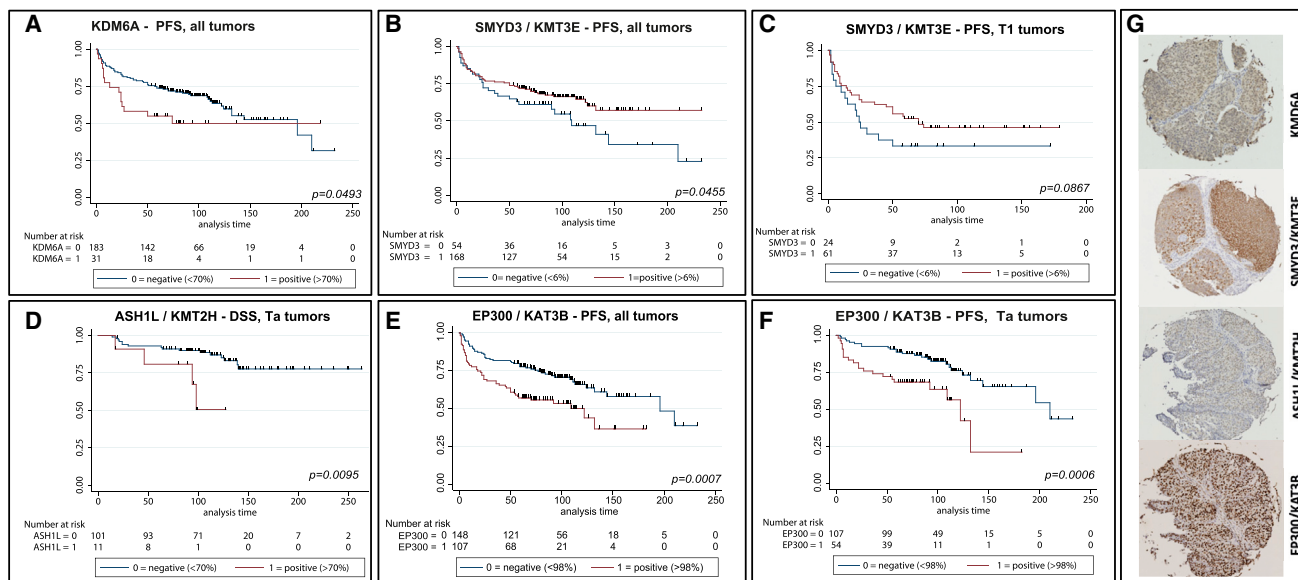


Figure 3. IHC Analysis of Epigenetic Modifier Proteins and Their Correlation to Clinical Endpoints

(A) Kaplan-Meier PFS estimates according to the expression level of KDM6A. (B and C) Kaplan-Meier PFS estimates according to the expression level of SMYD3/KMT3E for (B) all tumors and (C) T1 tumors only. (D) Kaplan-Meier DSS estimates according to the expression level of ASH1L/KMT2H with regard to Ta tumors only. (E and F) Kaplan-Meier PFS estimates according to the expression level of EP300/KAT3B for (E) all tumors and (F) Ta tumors only. (G) Examples of characteristic IHC staining patterns for these molecules. The significance levels are indicated in all graphs. See also Figure S2 and Table S5.

154 events (Figures S3D–S3F), in line with previous publications (Wang et al., 2011).

In a combined breakpoint analysis of both DNA-seq and RNA-seq data, we identified six significant fusion transcripts (Table S6A), all validated by Sanger sequencing. They included fusions involving the cancer-relevant genes *Ctnnb1*, *Mtss1*, *Polb*, and *Kdm6a* (Tables S6A and S6B), which could lead to truncated protein products. The *Ctnnb1* promoter was placed in front of the *Fam110a* gene in the in-frame fusion (Tables S6A and S6B), and could lead to increased levels of *Fam110a*, partly supported by the transcript data.

The fusions were confined to a specific tumor in five of six fusions. In four of six fusions, it was not present in the early tumor from the same individual, showing that they mostly arose relatively late in the development of the individual tumors.

Having defined the breakpoints, we then analyzed whether local hypermutability was related to breakpoint loci, as recently suggested (Drier et al., 2013). In all tumors, the mutation frequency increased toward the breakpoint for all mutations, which was most clearly seen within 1,000 nt from the breakpoint (Figure 4A). C > G and C > T mutations accounted for a large fraction of the mutations that were enriched for ($p < 0.03$), and the APOBEC context was dominating (Figure 4).

DNA Damage Response Pathways Anticorrelated with the APOBEC Signature

We studied pathways involved in bladder cancer using information from both the genome and transcripts. In the genome, SNVs

and indels involved a number of pathways mainly driven by *Rb1* and *Pik3ca* (Tables S7A–S7E), and, to a lesser degree, *Tp53* and *Atm* mutations.

When we compared transcriptome-defined pathways in late versus early tumor to identify pathways involved in progression, we found four pathways that were downregulated in both of the two non-APOBEC tumor pairs (“EIF2 signaling” [$p < 10^{-32}$ and $p < 10^{-39}$ for P01 and P03, respectively], “regulation of EIF4 and p70s6k signaling” [$p < 10^{-16}$ and $p < 10^{-9}$], “mitochondrial dysfunction” [$p < 10^{-4}$ and $p < 10^{-19}$], and “mTOR signaling” [$p < 10^{-13}$ and $p < 10^{-8}$]; Table S7F). No pathway changes were common to the two APOBEC tumor pairs.

In the WES-set, transcriptome canonical pathways were downregulated in the ten samples with a strong APOBEC signature compared with those with a weak signature. Interestingly, 11 of the 20 most significantly downregulated pathways ($p < 0.01$) reflected the DNA damage response, cell-cycle checkpoints, and double-strand break repair (DSB; Tables S7G–S7I). In concert with this, the *Stag2* transcript was strongly anticorrelated with the level of the APOBEC signature (Table S7G). *Stag2* and other sister chromatid cohesion and segregation molecules were recently reported to be inactivated in 11%–16% of UCCs (Balbás-Martínez et al., 2013; Guo et al., 2013). As the APOBEC signature is not correlated to stage and grade, the pathway correlations do not simply reflect an increased cell-cycle rate in high-stage tumors. We hypothesize that reduced DNA repair and cell-cycle checkpoint control may lead to increased chromosomal instability and recruit APOBEC3 enzymes to DSB sites

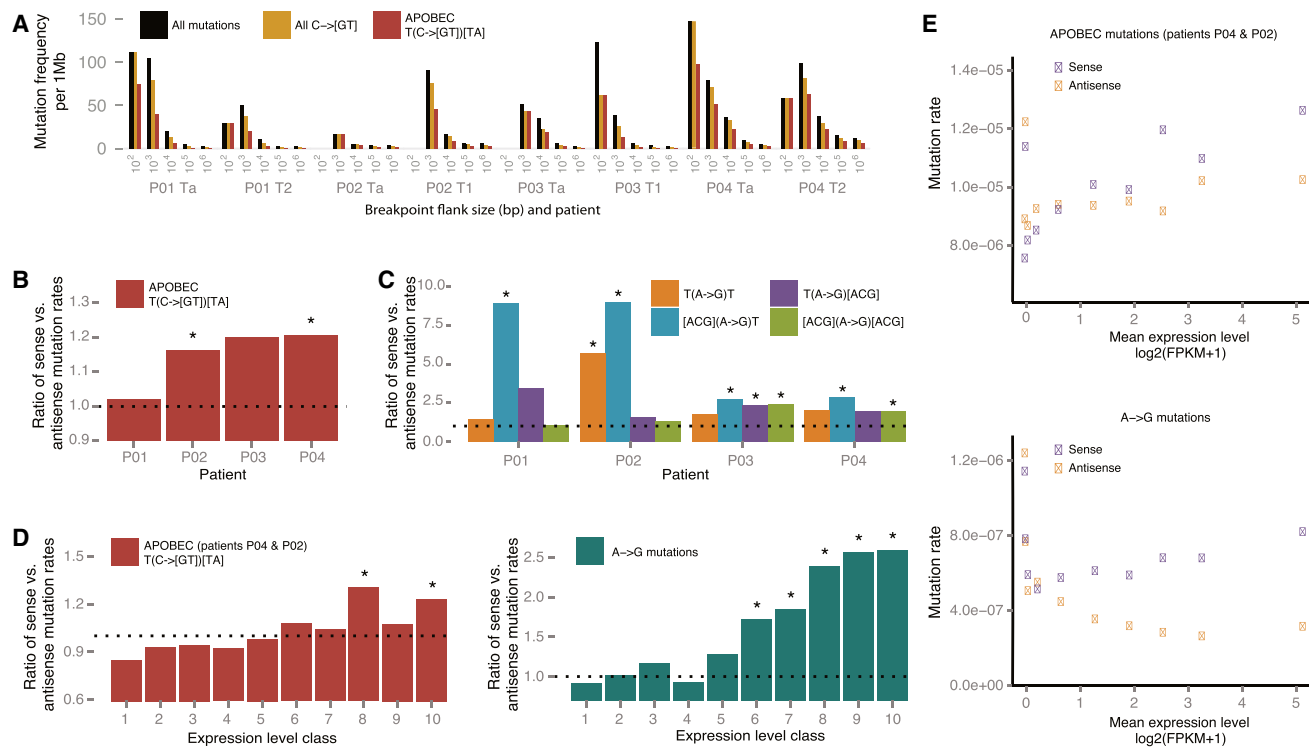


Figure 4. Mutational Patterns across Chromosomal Breakpoints and Transcribed Regions

(A) Elevated mutation frequency at breakpoints. The mutation frequency for all mutation types (black), C to G or T mutations (yellow), and APOBEC signature mutations (red) are shown for regions within a given distance (100 bp to 1 Mb) of a breakpoint.

(B) The APOBEC signature mutations are biased toward the sense strand of transcribed genes. Each bar represents the ratio of the sense to antisense mutation rate for the given mutation type among all possible sites across the top 25% most expressed genes. The bias is significant for both patients P02 ($p < 0.0074$) and P04 ($p < 10^{-5}$; Fisher's exact test).

(C) A → G mutations are also biased toward the sense strand, irrespective of mutational context. Same analysis as in (B). All ratios with asterisks are significant ($10^{-12} < p < 0.0164$).

(D) Both the APOBEC signature mutations and A → G mutations correlate positively with expression level. The strand bias is evaluated as in (B) for all genes divided into ten expression classes. Asterisks indicate significance ($p < 0.01$). For APOBEC, only results for patients P04 and P02, which showed significant bias in (B), are included.

(E) Strand-specific mutation rate as function of expression level across the ten gene classes.

See also [Table S8](#).

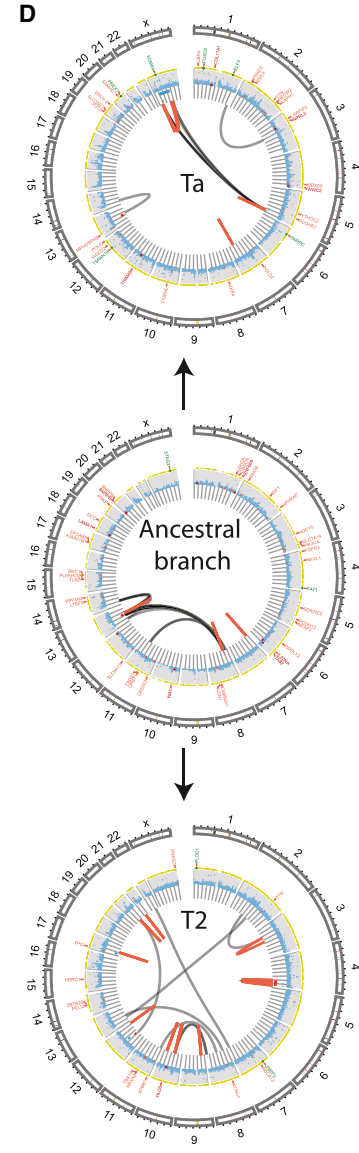
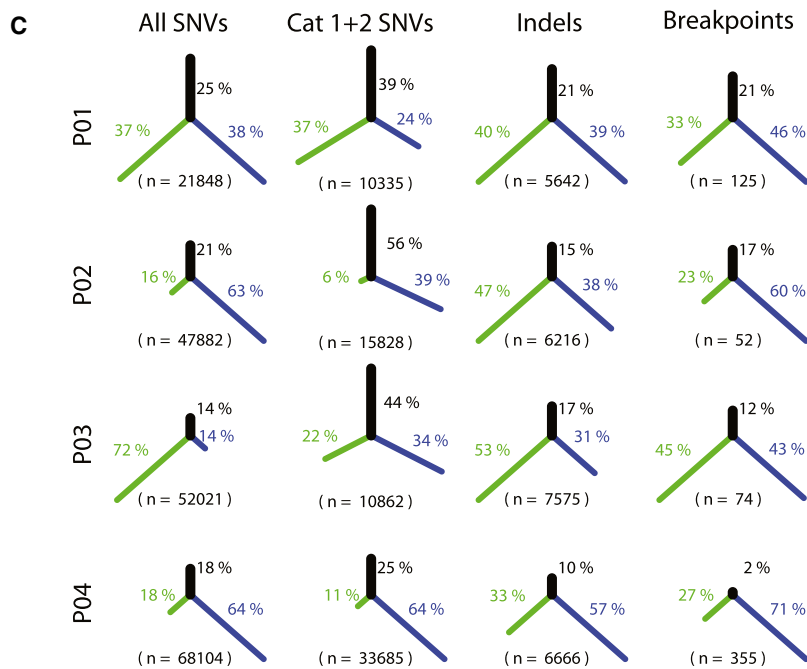
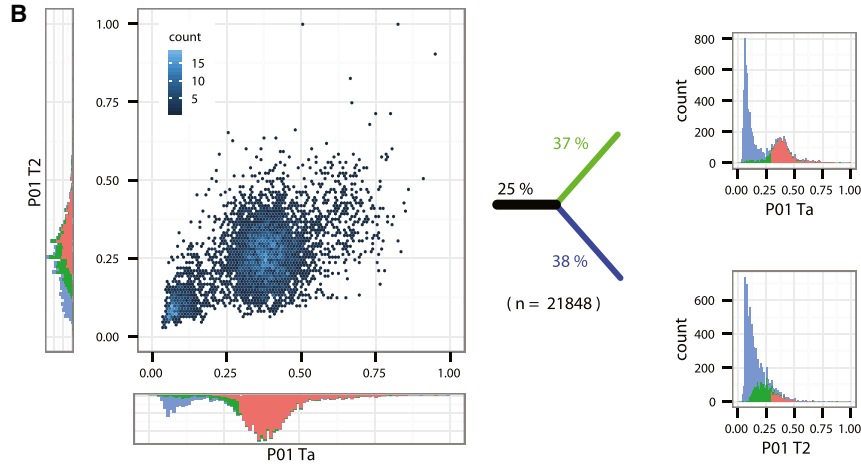
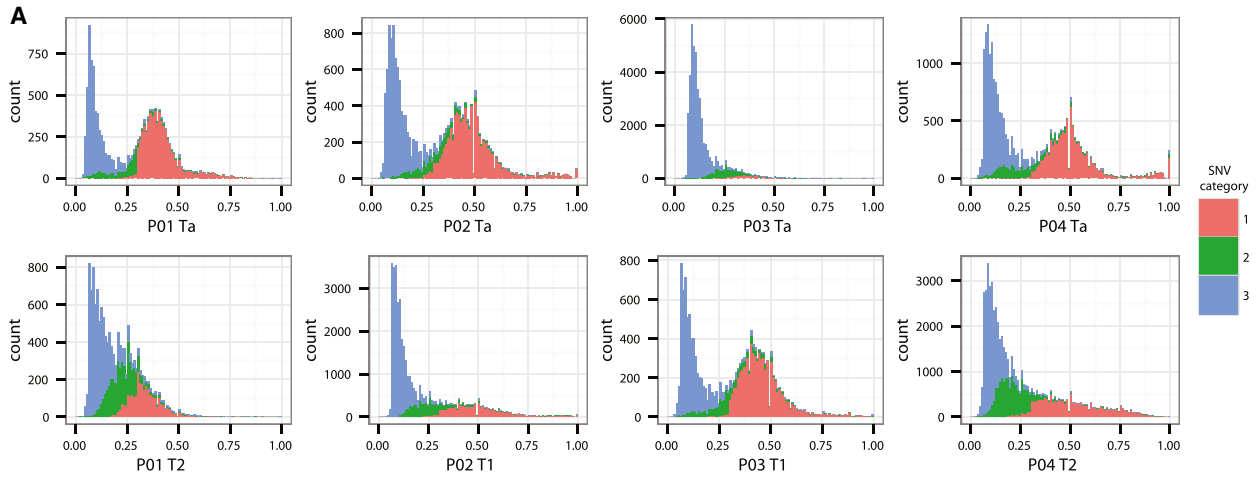
for repair (Nowarski et al., 2012). The APOBEC3 enzymes then in turn insert mutations via deamination, serving a dual role in preventing cell death and promoting mutations that drive the cancer (Nowarski and Kotler, 2013).

We analyzed the transcripts from the APOBEC3 family, *Apobec3a,b,c,d,f,g,h*, and *Apobec2* on three different microarray platforms across 177 samples using an in-house data set. *Apobec1* and *Apobec4* were not expressed in UCC. We found a correlation between an increased level of *Apobec3b* and increased stage ($p < 0.003$), and an increased level in lymph node metastases compared with the primary tumors ($p < 0.03$; Figure S4). In the 34 WES and WGS individual cases with RNA data, the transcripts did not correlate with the APOBEC signature at either the single or the combined level, corroborating a recent study (Roberts et al., 2013). Accordingly, there may be other enzymes involved, or a time issue may influence the measurements.

Finally, we asked whether the mutation pattern was related to gene expression, potentially through transcription-coupled

repair (TCR) processes, which act specifically on the antisense (transcribed) strand (Lans et al., 2010). For this purpose, we separately evaluated the sense and antisense strand-specific mutation rates of all possible mutation types, including nucleotide neighbor contexts ($n = 192$), across all highly expressed genes (top 25%) of each WGS patient (Table S8). The APOBEC signature mutations were significantly enriched on the sense strand (up to 1.20-fold, $p < 0.01$) in the two APOBEC patients (Figure 4B). Even more strikingly, A > G mutations were up to eight times more frequent on the sense strand ($p < 0.002$) in [ACG]AT contexts and showed a consistent bias irrespective of context across all four patients (Figure 4C). When we divided all genes into expression classes, we found that the sense strand bias correlated strongly with expression level for both APOBEC mutations (Spearman rho = 0.88; $p < 0.01$) and A > G mutations (rho = 0.96; $p < 10^{-16}$; Figure 4D).

To better understand the source of the bias, we finally plotted the mutation rate separately for the sense and antisense strands as a function of expression level (Figure 4E). For the APOBEC



(legend on next page)

mutations, both the sense and antisense mutation rates increase with expression level ($\rho = 0.71$; $p = 0.03$ versus $\rho = 0.41$; $p = 0.24$), but the sense-strand rate grows faster. This suggests that the APOBEC mutation rate overall increases with expression (Lawrence et al., 2013), but the effect is dampened on the antisense strand by TCR. In contrast, for the A > G mutations, the sense-strand rate is uncorrelated with expression level ($\rho = 0.006$; $p \approx 1$) and the antisense rate is strongly negatively correlated ($\rho = -0.95$; $p < 10^{-16}$), suggesting that the strand bias for the A > G mutations may be driven by TCR correcting A > G mutations on the antisense strand during transcription (Green et al., 2003).

Ancestral Cell Clones Define the Field Disease in UCC

The total number of SNVs generally increased from superficial tumors to invasive tumors, and their frequency distribution altered dramatically. Three of the four paired WGS-set patients had bimodal distributions in the early Ta tumor, with modes at the heterozygous allele ratio and at low frequency (Figure 5A). In contrast, late invasive tumors only showed the low frequency mode (Figure 5A), which could be due to extreme copy-number variation or to pervasively mutated subclones (Carter et al., 2012). Patient P03 showed a reversed pattern, which may imply that the sequence of tumor resection may not always be related to the age of the tumor (van Tilborg et al., 2000).

Assignment of mutations to the branches of the evolutionary trees showed that a high fraction of point mutations were unique to the early and late tumors (Figures 5B–5D, S3G, and S3H). Furthermore, very few rearrangements were in common between early and late tumors. This was most evident in patient P04, where almost no rearrangement was found in both tumors, although 25% of SNVs were in common (Figures 5C and S3H). This indicates that the ancestral cells were mainly hit by SNVs and developed into subpopulations that separately acquired rearrangements and SNVs. Furthermore, both loss-of-heterozygosity (LOH) and fusion transcripts were specific for early or late tumors (Tables S6C–S6E). This suggests that the field cancerization in UCC could be represented by the mutual ancestral changes, and that individual tumors, formed over time, have a divergent evolution with private alterations in addition to the ancestral ones.

To understand patterns of clonal selection related to progression, we analyzed the deep exomes from the paired-WGS set (181x–405x) and inferred the clonal population structure using PyClone (Shah et al., 2012; Roth et al., 2014). Estimating the cellular prevalence of a set of mutations provides insight into the relative timing of mutations in a tumor's evolutionary history and approximates the number of clones present in a tumor. PyClone assigns cellular prevalence estimates to individual

mutations while accounting for their copy-number states and the presence of regional LOH. Since tumor cellularity was estimated histologically to be >90% for our cases, we corrected for this number.

In general, groups of mutations with different distinct cellular prevalence were observed, indicating clonal subpopulations (Figures 6A–6D). All four patients exhibited a unitary ancestral origin, with primary and invasive tumor pairs sharing identical mutations at high cellular prevalence. The ancestral clones contained several mutations that are commonly mutated in cancer and/or are listed as tumor suppressors or oncogenes (e.g., *Fgfr3*, *Kdm6a*, *Pik3ca*, and *Tp53*; Figures 6A–6D; Vogelstein et al., 2013). We suggest that the ancestral clone is a major driver of the malignant process in general, and an obvious therapy target. In all cases, we also observed expansion of a clone in the progression sample that was absent or present at minor prevalence in the early tumor, with mutations in tumor inhibitors such as *Tp53*, *Mll3*, *Fbxw7*, and *Setd2* (Figures 6A–6D). This clone may be driving the progression. Furthermore, all samples had nonaggressive subclones that were present in the early Ta tumor, with mutations such as *Fgfr3*, *Mllt4*, and *H3f3a*, and more or less absent in the invasive tumor (Figures 6A–6F). Any relation to the APOBEC signature was not seen in this small material.

Pathway analysis of subpopulations showed *Cell Cycle: G2/M DNA Damage Checkpoint Regulation* to be enriched in the progression populations, and *FGF signaling* in populations with a *Pik3ca* mutation (Figure S5).

DISCUSSION

The APOBEC signature in one-third of tumors was patient specific, as it was present or absent in both early and late tumors from the same patient. Consequently, the mutation mechanism was conserved over time and could reflect a basic property of the urothelium, either due to a patient-specific genomic background with specific polymorphisms in DNA-editing enzymes or due to the impact of infectious agents that may have triggered the APOBEC enzymes, as suggested by Roberts et al. (2013). We searched for papilloma virus nucleotide sequences, as these have been related to UCC (Abol-Enein, 2008), but did not find any, in accord with the TCGA findings (<http://cancergenome.nih.gov/>). However, other agents could have been in play.

The *Apobec3b* expression level correlated to stage as in ovarian cancer (Leonard et al., 2013). The lack of correlation between the APOBEC signature and the APOBEC3 family of deaminases is interesting because it indicates that the APOBEC signature-creating machinery contains other enzymes or may fluctuate over time. Others also found a lack of correlation in

Figure 5. Mutational and Structural Events during Progression of UCC Based on Paired Sets of Tumors—Paired-WGS Set

- (A) Stacked histograms of SNV frequencies for the paired-WGS set. Categories are indicated by colors.
 (B) Common and unique SNVs in patient P01. Left, scatterplot showing frequencies of the ancestral SNVs in the Ta versus the T2-4 tumor. Right, stacked histograms of the unique SNVs in the Ta tumor (top) and the T2-4 tumor (bottom).
 (C) Evolution trees inferred by all SNVs, Cat 1-2 SNVs, indels, or breakpoints. For each tree, the number of events is written at the bottom. The percentage of events contributing to the ancestral branch (black), the Ta branch (green), and the T1/T2-4 branch (blue) is shown.
 (D) Branch-specific circo plot for patient P01. All SNVs, indels, and rearrangements were plotted in the same plot (see Figure S4H for details).
 See also Figure S3 and Table S6.

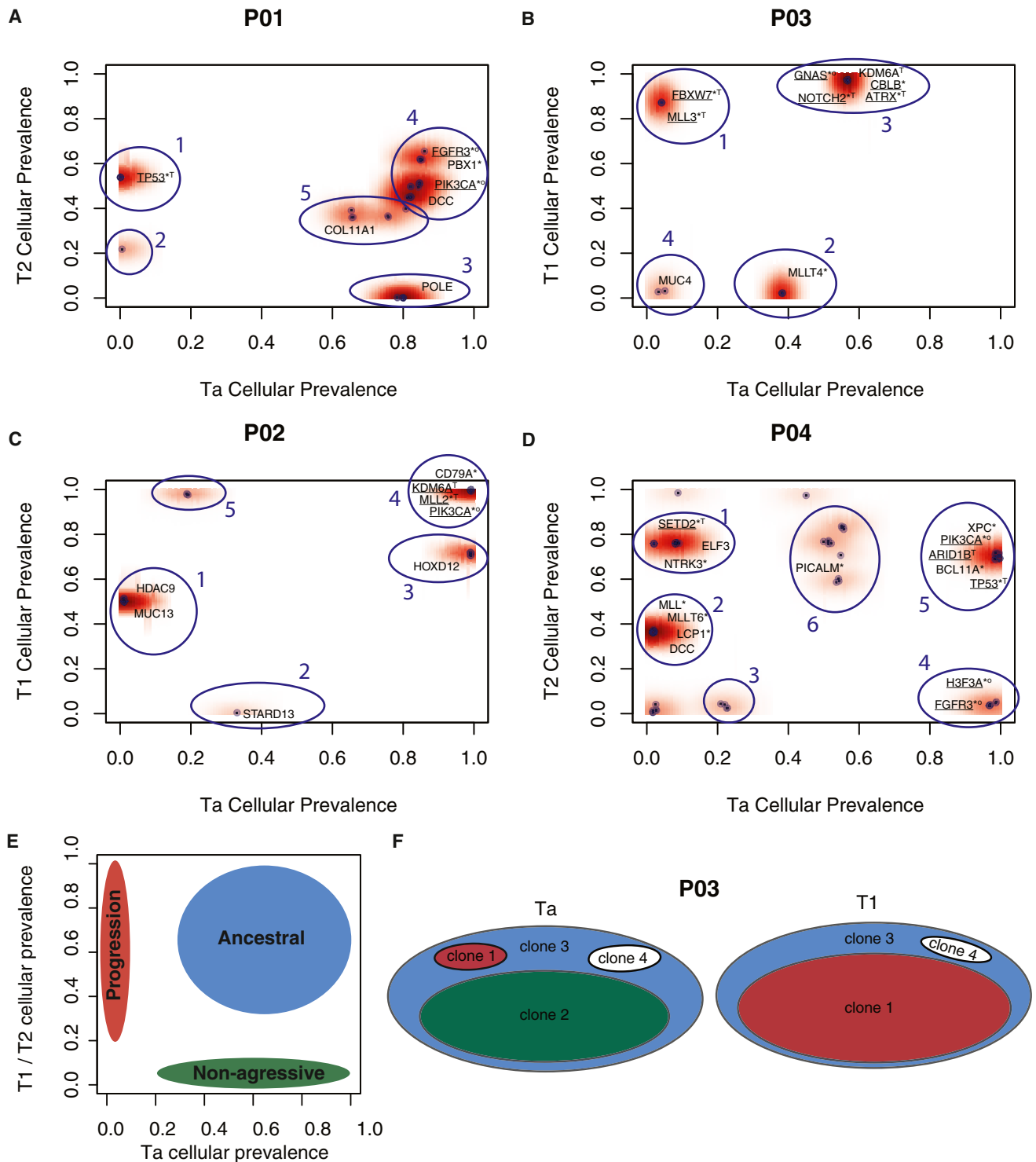


Figure 6. Subclonal Populations in the Paired-WGS Set

(A–D) Cellular prevalence of individual SNVs were inferred with PyClone and plotted (black dots) for both early-stage (Ta) and late-stage (T1/T2) samples from each patient in the paired-WGS set. Selected SNVs are named and the overall density is indicated (red shading). The enumerated SNV clusters reflect the clonal structure of the samples. Samples P01 T2 and P03 Ta have a lower maximal cellular prevalence than the other samples, suggesting tumor purity slightly lower than the histologically estimated 90%. Underlined genes are driver genes (either tumor suppressors (^T) or oncogenes (^C)) (Vogelstein et al., 2013). Genes marked with * are found in the Cosmic database.

(legend continued on next page)

individual cases, although they did find a general correlation for the whole population examined (Roberts et al., 2013).

The fidelity of the DNA strand is overseen by many molecules, and, surprisingly, the DNA cell-cycle checkpoints, as well as some DSB repair pathways and molecules, were anticorrelated to the APOBEC signature at the transcript level, including the recently reported *Stag2*. Without knowing what is the cause or effect at this point in time, we suggest that a low repair activity could lead to an increased number of breakpoints and maybe even a longer exposure time of single-stranded DNA (ssDNA) to editing or modifying enzymes, giving more time for formation of the mutation-based signature on these ssDNA regions. This fits into the finding of more signatures, and mutations in general, closer to breakpoints (Roberts et al., 2013). The lack of correlation between the APOBEC signature or enzyme transcript level and specific genes mutated, proteins expressed, or clinical outcome was unexpected. However, although the signature is very distinct, it may not provide specific properties to the cells. Much more research is needed before we can draw firm conclusions about this aspect. The A > G sense-strand bias may be a general phenomenon, as it has previously been observed in species evolution and coupled to TCR (Green et al., 2003).

We describe the mutational spectrum in low-grade and low-stage UCC because previous reports have almost exclusively focused on muscle-invasive cancers. *Kdm6a* showed a significantly higher mutation frequency in these tumors, and a number of mutations were either not present or were seen at a significantly lower frequency (e.g., *Tp53*, *Lphn3*, and *Wnk1*). A gene family that was frequently mutated in invasive UCC was plasma membrane proteins, mainly related to cell adhesion. This may explain the low cohesion of UCC cells, and offers an explanation for this phenomenon, which leads to exfoliation of multiple tumor cells into the urine of patients with UCC. The dN/dS ratio was <1, pointing toward a purifying selection; however, it was striking that the oncogenic pathways showed significantly more nonsynonymous mutations than expected, and the metabolic genes showed fewer than expected. We interpret this as indicating a selection for the malignant genotype, as mutations in oncogenes and tumor suppressors drive malignancy, whereas the metabolism needed for cell survival in general is protected and only synonymous mutations are tolerated in these genes.

UCC is a field disease that encompasses a large and variable part of the urothelium, mainly in patients who experience recurrent tumors and disease progression. Our data on the subclones present in each tumor pair showed one or two ancestral clones in each patient that theoretically could originate from coresected “normal” cells, but probably identify the field-disease mutations (see the Supplemental Discussion). These cells may have an almost normal phenotype in the light microscope, but are known to have profoundly altered transcriptomes (Dyrskjot et al., 2004;

Majewski et al., 2008). These cells undergo a divergent evolution, leading to private mutations in each tumor that over time develops from the cells in the field. Both the ancestral and the nonaggressive and progression clones contained mutated oncogenes and tumor suppressors. The mutated genes varied, except for *Pik3ca*, which was present in three out of four ancestral clones. Since mutated *Kdm6a* was present in two ancestral clones and supplemented *Pik3ca*, these two molecules might have been ideal as therapy targets (treatment by nilotinib and AKT inhibitor VIII, respectively) in the patients examined.

EXPERIMENTAL PROCEDURES

Sample Preparation

UCC samples were obtained fresh from resection, embedded in Tissue-Tek, and stored at -80°C . The inclusion criteria are defined in Supplemental Experimental Procedures. Blood samples were collected at the time of the patients' initial visit. Genomic DNA and total RNA were extracted from serial cryosections. The project was approved by the Central Denmark Region and National Committees on Health Research Ethics (file 1300174).

NGS Library Construction

For WGS, high-molecular-weight genomic DNA (100–400 ng) was fragmented prior to library construction using standard Illumina kits. For WES, exome capture was performed using a NimbleGen kit on libraries prepared with the Illumina TruSeq Kit, using either 1 μg genomic DNA or 2.5 μg whole-genome amplified (WGA) DNA. For whole-transcriptome RNA-seq, libraries for paired-end and indexed RNA-seq were prepared from rRNA-depleted total RNA (300–400 ng) using Ribo-Zero and ScriptSeq (Epicenter).

Sequencing

We employed standard sequencing kits (Illumina HiSeq2000) to generate 2 × 100 bp paired-end sequencing. All sequencing data were deposited to the European Genome-Phenome Archive (EGA).

Sequence Processing, Mutation, and Rearrangement Identification

Overlapping read pairs were joined using AdapterRemoval prior to mapping against the HG19. DNA samples were mapped using BWA, and RNA samples were analyzed using the Tuxedo Suite. For DNA samples, the alignments were recalibrated and realigned using the Picard (<http://picard.sourceforge.net/>) and GATK suites. nFuse was used to identify fusion transcripts combining the RNA-seq and whole-genome DNA-seq data. MuTect was used to call somatic mutations (SNVs) and the Somatic Indel Detector from GATK was used to call indels in the paired-WGS set and the WES set. CREST was used to detect structural rearrangements in the paired-WGS set.

Annotation and Validation of Mutations

The SNVs and indels were functionally annotated using snpEff (v2_1b) (Cingolani et al., 2012). Additional annotations were made using the UCSC Genome Browser database (see the Supplemental Experimental Procedures). SNVs and indels were put in three categories and five tiers. SNVs and genomic rearrangements were validated using one of the following: (1) Sanger sequencing on WGA DNA, (2) PCR-based targeted resequencing on MiSeq, (3) deep WES, and (4) RNA-seq. Fusion transcripts predicted by nFuse were validated by Sanger sequencing of cDNA amplified from RNA using the Ovation PicoSL WTA System V2 (NuGEN).

(E) SNVs found at high prevalence in both samples are ancestral (blue) and include the initial driver events. Clones specific for the Ta samples are generally thought to be nonaggressive (green), whereas the T1/T2-specific clones underlie tumor progression (red).

(F) Example of clonal structure interpretation based on the inferred frequencies (patient P03). Clone 1 is present at low frequency in the Ta sample, but dominates the T1 sample.

See also Figure S5 and Table S7.

SNP Array Data

DNA was labeled and hybridized to SNP Arrays 6.0 (Affymetrix). The R package *aroma.affymetrix* was used for preprocessing/probe summarization of the SNP6.0 data. The R package *Rsegg* (Lamy et al., 2011) was used to segment each tumor sample. Allelic imbalance was defined at heterozygous SNPs in the germline sample, and *Rsegg* was used to segment the samples from the paired-WGS set, except for patient P04.

IHC, Tissue Microarray, Microscopy, and Correlation to Clinical Data

IHC was performed on 4 μ m FFPE sections from the eight paired-WGS samples, 27 of the 30 WES samples, and on a tissue microarray (TMA) with 283 biopsies from primary, stage Ta and T1 urothelial bladder tumors with long-term follow-up data (Fristrup et al., 2013). TMA core images (Nanozoomer; Hamamatsu) were scored manually using VIS software (Visiopharm).

Ingenuity Pathway and Cluster Analysis

Data were analyzed with the use of IPA, build May 2013 (Ingenuity Systems, <http://www.ingenuity.com>). Hierarchical cluster analysis was performed using Cluster 3.0 software and visualized using Java tree-view software.

Parental Copy Number for the Paired-WGS Set

The HMMcopy suite of tools (Ha et al., 2012) was used to estimate the copy numbers for each of the paired-WGS samples. The copy-number information from tumor and normal samples were entered into the APOLLOH program for LOH estimation (Ha et al., 2012).

Statistical Inference of Tumor Cell Populations Using Deep-Exome Sequencing

PyClone 0.12.0 (Roth et al., 2014) was used to infer subclonal populations in the paired-WGS set.

Analysis of dN/dS Ratios

The dN/dS ratios were computed with *codeml* of the PAML software (Yang, 2007). The dN/dS ratios were evaluated genome wide and for gene sets defined by KEGG. Confidence intervals and significance levels were estimated by a Monte Carlo approach using 1,000 samples obtained by random sampling with replacement.

Identification of the Effect of Transcription on Mutation Patterns

Using the SNVs of the WGS set and the RNA-seq, mutations and possible mutable sites on the sense and antisense strands were counted for the 192 mutation types, including neighbor context. Differences were assessed using Fisher's exact test with Benjamini-Hochberg correction at a false-discovery rate of 0.05 (Figures 4B and 4C).

To evaluate the mutation intensity ratio and expression level, the entire gene set was ordered into ten bins based on expression level. The correlation was evaluated using Spearman's rank correlation test (Figures 4D and 4E).

Statistics

For statistical analysis, Fisher's exact test was used unless otherwise indicated.

ACCESSION NUMBERS

The sequencing data reported in this paper have been deposited to the EGA (<https://www.ebi.ac.uk/ega/>) under accession number EGAS00001000641. Access can be obtained by contacting the corresponding author, fulfilling the criteria of the EGA.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Discussion, Supplemental Experimental Procedures, five figures, and eight tables and can be found with this article online at <http://dx.doi.org/10.1016/j.celrep.2014.04.038>.

AUTHOR CONTRIBUTIONS

I.N., P.L., K.B.-D., J.S.P., and T.F.Ø. designed the research. K.B.-D., N.F., and L.D. performed the protein experiments. P.L., S.V., H.H., M.J., P.V., I.N., K.B.-D., and J.S.P. performed the bioinformatics data analysis. K.S., A.R., S.S., and P.L. performed the clonal analysis on the basis of PyClone. J.H., I.N., and L.D. performed the RNA data analysis. S.H. and M.B. performed clinical and pathological examinations of patients. T.R., L.D., K.B.-D., and N.F. collected the samples. I.N., K.B.-D., P.L., K.T., J.H., and S.V. generated and validated NGS data. T.F.Ø. wrote the manuscript with input from all of the authors.

ACKNOWLEDGMENTS

We thank technicians Lone Andersen, Hanne Steen, Gitte Glistrup Nielsen, and Pamela Celis for excellent technical support. We thank Adam Siepel and Melissa Jane Hubisz for defining neutral genomic regions. We thank TCGA for making data available. This work was supported by grants from the John and Birthe Meyer Foundation, the Lundbeck Foundation, the Danish Cancer Society, the Danish Research Council, the Danish Cancer Biobank, the European Union (UROMOL), and the University of Aarhus.

Received: August 9, 2013

Revised: April 2, 2014

Accepted: April 21, 2014

Published: May 15, 2014

REFERENCES

- Abol-Enein, H. (2008). Infection: is it a cause of bladder cancer? *Scand. J. Urol. Nephrol. Suppl.* 278, 79–84.
- Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.L., et al.; Australian Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MML-Seq Consortium; ICGC PedBrain (2013). Signatures of mutational processes in human cancer. *Nature* 500, 415–421.
- Aparicio, S., and Caldas, C. (2013). The implications of clonal genome evolution for cancer medicine. *N. Engl. J. Med.* 368, 842–851.
- Babjuk, M., Burger, M., Zigeuner, R., Shariat, S.F., van Rhijn, B.W., Compérat, E., Sylvester, R.J., Kaasinen, E., Böhle, A., Palou Redorta, J., and Rouprêt, M.; European Association of Urology (2013). EAU guidelines on non-muscle-invasive urothelial carcinoma of the bladder: update 2013. *Eur. Urol.* 64, 639–653.
- Balbás-Martínez, C., Sagrera, A., Carrillo-de-Santa-Pau, E., Earl, J., Márquez, M., Vazquez, M., Lapi, E., Castro-Giner, F., Beltran, S., Bayés, M., et al. (2013). Recurrent inactivation of STAG2 in bladder cancer is not associated with aneuploidy. *Nat. Genet.* 45, 1464–1469.
- Burns, M.B., Lackey, L., Carpenter, M.A., Rathore, A., Land, A.M., Leonard, B., Refsland, E.W., Kotandeniya, D., Tretyakova, N., Nikas, J.B., et al. (2013a). APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* 494, 366–370.
- Burns, M.B., Temiz, N.A., and Harris, R.S. (2013b). Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat. Genet.* 45, 977–983.
- Carter, S.L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P.W., Onofrio, R.C., Winckler, W., Weir, B.A., et al. (2012). Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* 30, 413–421.
- Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 6, 80–92.
- Drier, Y., Lawrence, M.S., Carter, S.L., Stewart, C., Gabriel, S.B., Lander, E.S., Meyerson, M., Beroukhim, R., and Getz, G. (2013). Somatic rearrangements across cancer reveal classes of samples with distinct patterns of DNA breakage and rearrangement-induced hypermutability. *Genome Res.* 23, 228–235.

- Dyrskjot, L., Thykjaer, T., Kruhoffer, M., Jensen, J.L., Marcussen, N., Hamilton-Dutoit, S., Wolf, H., and Orntoft, T.F. (2003). Identifying distinct classes of bladder carcinoma using microarrays. *Nat. Genet.* **33**, 90–96.
- Dyrskjot, L., Kruhoffer, M., Thykjaer, T., Marcussen, N., Jensen, J.L., Møller, K., and Orntoft, T.F. (2004). Gene expression in the urinary bladder: a common carcinoma in situ gene expression signature exists disregarding histopathological classification. *Cancer Res.* **64**, 4040–4048.
- Fristrup, N., Birkenkamp-Demtröder, K., Reinert, T., Sanchez-Carbayo, M., Segersten, U., Malmström, P.U., Palou, J., Alvarez-Múgica, M., Pan, C.C., Ulhøi, B.P., et al. (2013). Multicenter validation of cyclin D1, MCM7, TRIM29, and UBE2C as prognostic protein markers in non-muscle-invasive bladder cancer. *Am. J. Pathol.* **182**, 339–349.
- Green, P., Ewing, B., Miller, W., Thomas, P.J., and Green, E.D.; NISC Comparative Sequencing Program (2003). Transcription-associated mutational asymmetry in mammalian evolution. *Nat. Genet.* **33**, 514–517.
- Gui, Y., Guo, G., Huang, Y., Hu, X., Tang, A., Gao, S., Wu, R., Chen, C., Li, X., Zhou, L., et al. (2011). Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat. Genet.* **43**, 875–878.
- Guo, G., Sun, X., Chen, C., Wu, S., Huang, P., Li, Z., Dean, M., Huang, Y., Jia, W., Zhou, Q., et al. (2013). Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation. *Nat. Genet.* **45**, 1459–1463.
- Ha, G., Roth, A., Lai, D., Bashashati, A., Ding, J., Goya, R., Giuliany, R., Rosner, J., Oloumi, A., Shumansky, K., et al. (2012). Integrative analysis of genome-wide loss of heterozygosity and monoallelic expression at nucleotide resolution reveals disrupted pathways in triple-negative breast cancer. *Genome Res.* **22**, 1995–2007.
- Iyer, G., Al-Ahmadie, H., Schultz, N., Hanrahan, A.J., Ostrovskaya, I., Balar, A.V., Kim, P.H., Lin, O., Weinhold, N., Sander, C., et al. (2013). Prevalence and co-occurrence of actionable genomic alterations in high-grade bladder cancer. *J. Clin. Oncol.* **31**, 3133–3140.
- Kandoth, C., McLellan, M.D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J.F., Wyczalkowski, M.A., et al. (2013a). Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339.
- Kandoth, C., Schultz, N., Cherniack, A.D., Akbani, R., Liu, Y., Shen, H., Robertson, A.G., Pashtan, I., Shen, R., Benz, C.C., et al.; Cancer Genome Atlas Research Network (2013b). Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67–73.
- Lamy, P., Wiuf, C., Orntoft, T.F., and Andersen, C.L. (2011). Rseg—an R package to optimize segmentation of SNP array data. *Bioinformatics* **27**, 419–420.
- Lans, H., Martejijn, J.A., Schumacher, B., Hoeijmakers, J.H., Jansen, G., and Vermeulen, W. (2010). Involvement of global genome repair, transcription coupled repair, and chromatin remodeling in UV DNA damage response changes during development. *PLoS Genet.* **6**, e1000941.
- Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218.
- Leonard, B., Hart, S.N., Burns, M.B., Carpenter, M.A., Temiz, N.A., Rathore, A., Vogel, R.I., Nikas, J.B., Law, E.K., Brown, W.L., et al. (2013). APOBEC3B upregulation and genomic mutation patterns in serous ovarian carcinoma. *Cancer Res.* **73**, 7222–7231.
- Majewski, T., Lee, S., Jeong, J., Yoon, D.S., Kram, A., Kim, M.S., Tuziak, T., Bondaruk, J., Lee, S., Park, W.S., et al. (2008). Understanding the development of human bladder cancer by using a whole-organ genomic mapping strategy. *Lab. Invest.* **88**, 694–721.
- Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., Van Loo, P., Greenman, C.D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L.A., et al.; Breast Cancer Working Group of the International Cancer Genome Consortium (2012). Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993.
- Nowarski, R., and Kotler, M. (2013). APOBEC3 cytidine deaminases in double-strand DNA break repair and cancer promotion. *Cancer Res.* **73**, 3494–3498.
- Nowarski, R., Wilner, O.I., Cheshin, O., Shahar, O.D., Kenig, E., Baraz, L., Britan-Rosich, E., Nagler, A., Harris, R.S., Goldberg, M., et al. (2012). APOBEC3G enhances lymphoma cell radioresistance by promoting cytidine deaminase-dependent DNA repair. *Blood* **120**, 366–375.
- Roberts, S.A., Lawrence, M.S., Klimczak, L.J., Grimm, S.A., Fargo, D., Stojanov, P., Kiezun, A., Kryukov, G.V., Carter, S.L., Sakseena, G., et al. (2013). An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat. Genet.* **45**, 970–976.
- Ross, J.S., Wang, K., Al-Rohil, R.N., Nazeer, T., Sheehan, C.E., Otto, G.A., He, J., Palmer, G., Yelensky, R., Lipson, D., et al. (2014). Advanced urothelial carcinoma: next-generation sequencing reveals diverse genomic alterations and targets of therapy. *Mod. Pathol.* **27**, 271–280.
- Roth, A., Khattra, J., Yap, D., Wan, A., Laks, E., Biele, J., Ha, G., Aparicio, S., Bouchard-Côté, A., and Shah, S.P. (2014). PyClone: statistical inference of clonal population structure in cancer. *Nat. Methods* **11**, 396–398.
- Scally, A., Dutheil, J.Y., Hillier, L.W., Jordan, G.E., Goodhead, I., Herrero, J., Hobolth, A., Lappalainen, T., Mailund, T., Marques-Bonet, T., et al. (2012). Insights into hominid evolution from the gorilla genome sequence. *Nature* **483**, 169–175.
- Shah, S.P., Roth, A., Goya, R., Oloumi, A., Ha, G., Zhao, Y., Turashvili, G., Ding, J., Tse, K., Haffari, G., et al. (2012). The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **486**, 395–399.
- van Tilborg, A.A., de Vries, A., de Bont, M., Groenfeld, L.E., van der Kwast, T.H., and Zwarthoff, E.C. (2000). Molecular evolution of multiple recurrent cancers of the bladder. *Hum. Mol. Genet.* **9**, 2973–2980.
- Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., Jr., and Kinzler, K.W. (2013). Cancer genome landscapes. *Science* **339**, 1546–1558.
- Wang, J., Mullighan, C.G., Easton, J., Roberts, S., Heatley, S.L., Ma, J., Rusch, M.C., Chen, K., Harris, C.C., Ding, L., et al. (2011). CREST maps somatic structural variation in cancer genomes with base-pair resolution. *Nat. Methods* **8**, 652–654.
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591.