Supplementary data

Materials and Methods

Supplementary Table 1. Demographics of all participants of the Golestan Cohort Study (n=50,045).

Age (years)	**Number (%)**
mean ± standard deviation	52·1 ± 8·9
Sex	
Male	21,234 (42·4%)
Female	28,811 (57·6%)
Ethnicity	
Turkman	37,253 (74·4%)
Non-Turkman	12,792 (25·6%)
Residence	
Rural	40,011 (80·0%)
Urban	10,034 (20·1%)
Formal education	
No	35,118 (70·2)
Yes	14,927 (29·8)
Tobacco use	
Never	41,388 (82·7%)
Ever	8,657 (17·3%)
Opium use	
Never	41,537 (83·0%)
Ever	8,508 (17·0%)
Alcohol drinking	
Never	48,316 (96·6%)
Ever	1,729 (3·5%)

UroMuTERT assay and mutation analysis

A single amplicon of 147bp covering the two hotspot mutations in the *TERT* promoter region (C228T and C250T mutations) was designed for the NGS assay. The sequences of the forward and reverse primers for the *TERT*-NGS amplicon are 5'-CTTCCAGCTCCGCCTCCTCCGCGCGG-3' and 5'-AGCGCTGCCTGAAACTCGCGCC-3' respectively.

Amplification of 5 ng urinary DNA and 10ng buffy coat DNA was performed in a PCR reaction mix containing 5X custom Buffer, 100 nM forward and reverse primers and 0·04 U/mL of AccuStart HiFi Taq Polymerase (Quanta BioSciences). PCR condition was: 98ºC for 10 min, 50 cycles of 98ºC for 30sec, 70ºC for 30sec, and 72ºC for 40 sec followed by 72ºC for 5min and then kept at hold at 12ºC.

Barcoded libraries were prepared from the amplified products as previously described.[1] The libraries were purified using magnetic beads and size-specific amplicons were extracted using the gel purification method. Ion

OneTouch 2 system and the Ion PI Hi-Q OT2 200 Kit (Thermo Fisher Scientific) were used to perform the emulsion PCR of the 7µL of 100pM purified library. Finally, the sequencing reaction was performed on an Ion Proton System (Thermo Fisher Scientific) using Life Technologies' Ion PI™ Chip Kit v3 and Ion PI™ Hi-Q™ Sequencing 200 Kit (Thermo Fisher Scientific).

The data from the NGS run in the ion proton system was analysed using our in house developed Needlestack multi-sample variant caller pipeline (Github: https://github.com/IARCbioinfo/needlestack).[2] Needlestack models sequencing errors using a negative binomial regression at each target position and target variant. Then the mutations are detected as being outliers from the regression model. A p-value for being a mutation is calculated for each position and then transformed into q-values (QVAL) to account for multiple testing. A QVAL of 20 was used as a threshold for calling mutations and we present here the needlestack plots for each position where we detected mutations in the urinary DNA **(Supplementary Figure 1).**

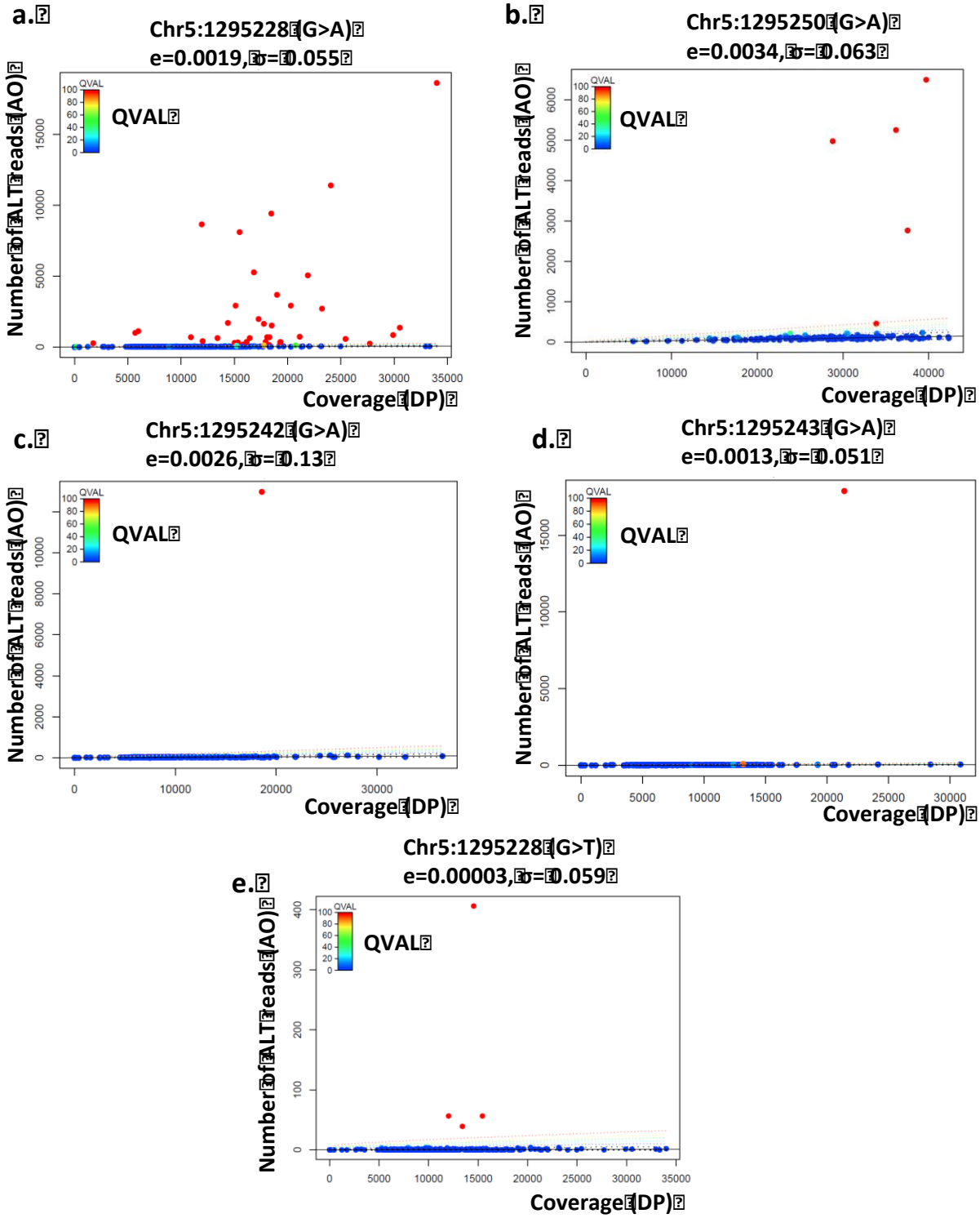
To control for potential low-allelic fraction amplification artifacts or sequencing errors, all DNA samples with evidence for a *TERT* mutation (QVAL>20) were re-sequenced as a technical duplicate. These samples, as well as sufficient wild type samples to ensure appropriate model fit were then reanalyzed using re-amplified, re-sequenced independent barcoded libraries and analyzed with Needlestack algorithm. Samples with confirmed QVAL>20 were retained after UroMuTERT analysis (**Figure 1**).

ddPCR assay and mutation analysis

We developed droplet digital PCR (ddPCR) assays for the four *TERT* promoter mutations observed in our study: C228T, C228A, CC242-243TT and C250T. The probes were custom designed probes from Biorad. The sequences of the probe are given in **Supplementary Table 2.**

All mutation positive samples after UroMuTERT analysis were considered for technical validation using droplet digital PCR (ddPCR). Ten wild type samples chosen at random were included for technical controls. We have developed and validated ddPCR assays for the two most common *TERT* promoter mutations: C228T and C250T. Briefly, a 22 µL reaction mix was prepared using 10ng of urinary DNA as a template, 11 µL of 2x ddPCR supermix-no dUTP (Biorad), 1·1 µL of 20x FAM and HEX probes for mutated and wildtype alleles, 1·1 µL of RsaI restriction enzyme (10 U/ µL) and 0·2 µL of 7-deaza-dGTP, Li-salt (2 µM). The droplets were generated using the autoDG droplet generator (Biorad). The PCR amplifications of the droplets were carried out separately for the C228T and C250T assays using the following PCR conditions: 95ºC for 10 min, 40 cycles of 94ºC for 30sec, ramp 2·5/sec, 54ºC for C228T assay (55ºC for C228A and CC242-243TT assays and 64ºC for C250T assay) for 30 sec followed by 98ºC for 10min and then kept at hold at 12ºC. The fluorescent intensity of each

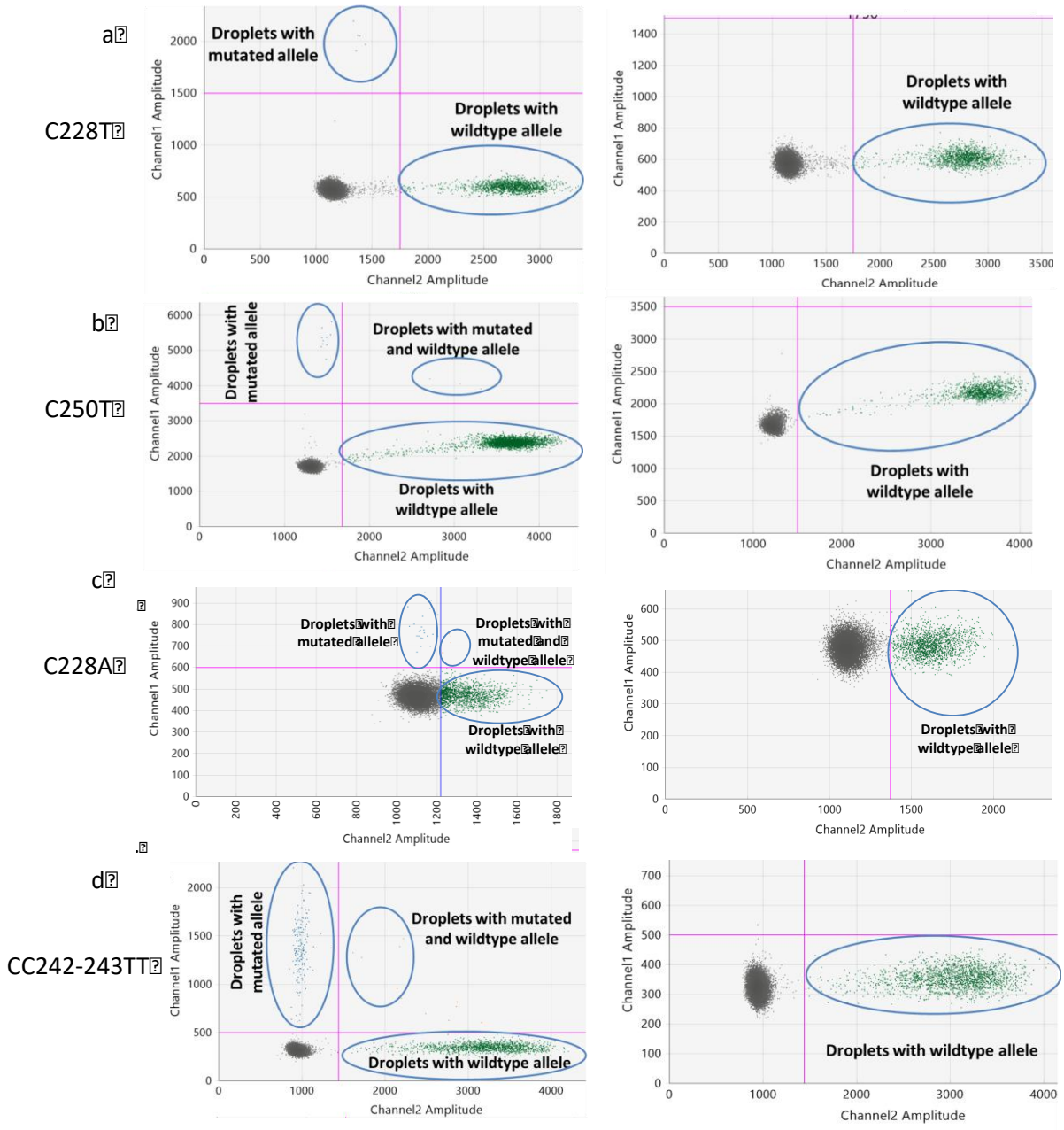
droplet was measured using the droplet reader QX200 (Biorad). Analysis of the ddPCR data was performed using the QuantaSoft™ Analysis Pro 1·0·596 software from Biorad (**Supplementary Figure 2)**. All the samples for this study were run in duplicates using 10ng of DNA derived from the urine samples, and the average fractional abundance of the duplicates was considered. The 2D amplitude plots from the QuantaSoft™ analysis pro software were analyzed by setting the threshold amplitudes for both the mutated and wild type channels. The 2D amplitude plots from the QuantaSoft™ analysis pro software were analyzed by setting the threshold amplitudes for both the mutated and wild type channels. The thresholds for the channel 1 (mutated probe) for the C228A, C228T, CC242-243TT and C250T probes were 700, 1500, 500 and 3500 respectively. For the channel 2 (wild type probe) the thresholds for C228A, C228T, CC242-243TT and C250T probes were 1250, 1750, 1440 and 2000 respectively. A minimum number of 5 or 6 positive droplets (mutated blue droplets above the Channel 1 thresholds) were considered to be positive in ddPCR for C228A/C250T and C228T/CC242-243TT assays respectively. All laboratory analyses were conducted blindly to the case or control status of the samples.



Supplementary Figure 1. Example of Needlestack plots for *TERT* promoter mutations detection. The Needlestack plots for the different types of *TERT* promoter mutations C228T (a), C250T (b), CC242-243TT (c and d), and C228A (e). The number of altered reads is plotted against the depth of sequencing (coverage). QVAL=20 is used as threshold to call mutations.

Supplementary Table 2. Probes for detecting *TERT* promoter mutations with ddPCR assays.

Mutation type	Primer/probe	Sequence (5' to 3')	Fluorescent Dye
TERT C228T	fw_primer	CCCTCCCGGGTCC	-
	rev_primer	CCGCGGAAAGGAAGG	-
	wt_probe	CGGAgGGGGCTGG	HEX_IowaBlack
	mut_probe	CCCGGAaGGGGCTG	FAM_IowaBlack
TERT C250T	fw_primer	CTTCACCTTCCAGCTCC	-
	rev_primer	GAGGGCCCGGAGG	-
	wt_probe	ACCCGGgAGGGGT	HEX_IowaBlack
	mut_probe	CCCGGaAGGGGTCG	FAM_IowaBlack
TERT C228A	fw_primer	CGCGGAAAGGAAGGG	-
	rev_primer	CCCCTCCCGGGTC	-
	wt_probe	CGGAgGGGGCTGG	HEX_IowaBlack
	mut_probe	CCCGGAtGGGGCTG	FAM_IowaBlack
TERT CC242-243TT	fw_primer	GAGGGCCCGGAGG	-
	rev_primer	CTTCACCTTCCAGCTCC	-
	wt_probe	CTGGGCCGGggAC	HEX_IowaBlack
	mut_probe	CCGGaaACCCGGGA	FAM_IowaBlack

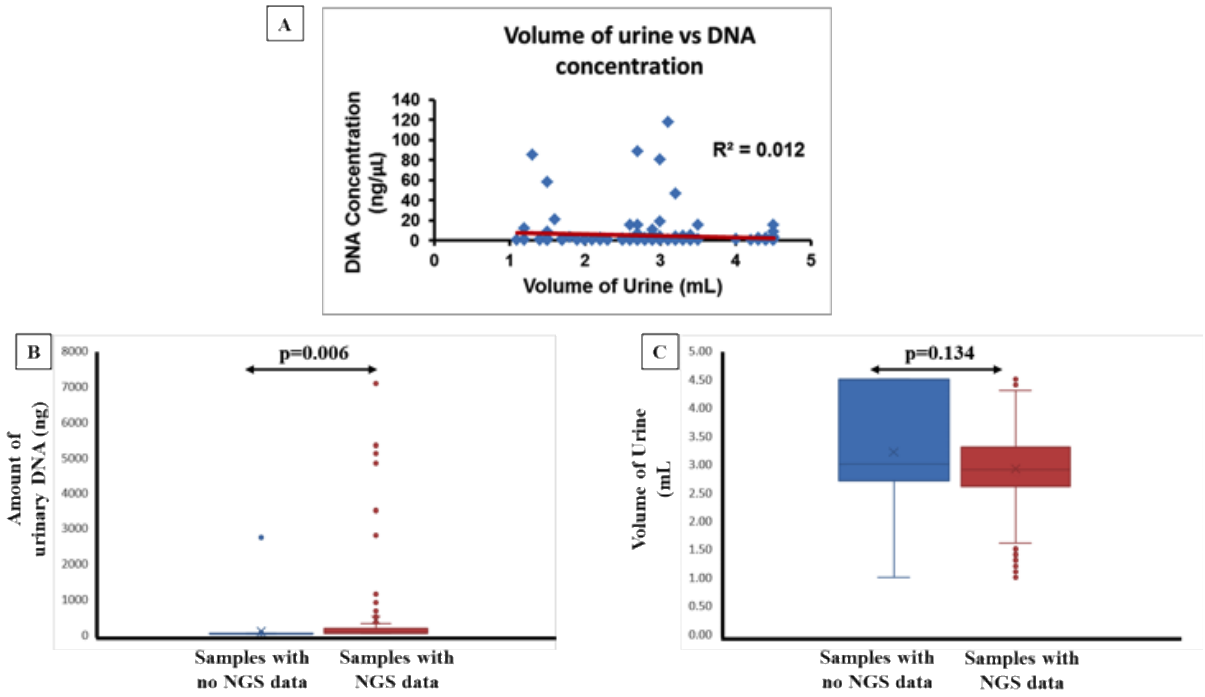


Supplementary Figure 2. 2D amplitude plots of *TERT* promoter mutation ddPCR assays in representative samples from the Golestan cohort. Assays testing for C228T (A), C250T (B), C228A (C) and CC242-243TT (D) mutations are displayed in four examples of mutated samples (left panels) and in four examples of wild-type samples (right panels). In the left panel, fluorescent probes (FAM) detect respective mutations as exemplified by the count of droplets with mutated alleles, while in the right panel, wild-type samples do not show any positive droplets (FAM) above the threshold line but show droplets with HEX fluorescence associated with wild-type probes. The pink lines are the thresholds for channel 1 (mutated probe) and channel 2 (wild-type probe) for the ddPCR mutation assays.

Results

Urinary DNA concentration

The median volume of urine in the samples analysed (n=188) was 2·9 mL (range: 1·1 - 4·5 mL) and the median DNA concentration was 0·8 ng/µL. The DNA was eluted in a 60µL volume. We obtained an average of 256 ng DNA (median 48 ng; range: 6ng – 7 µg). Twelve out of 188 urine samples did not yield any measurable amount of DNA. There was no correlation between the concentration of DNA achieved and the urine volume used (**Supplementary Figure 3**).



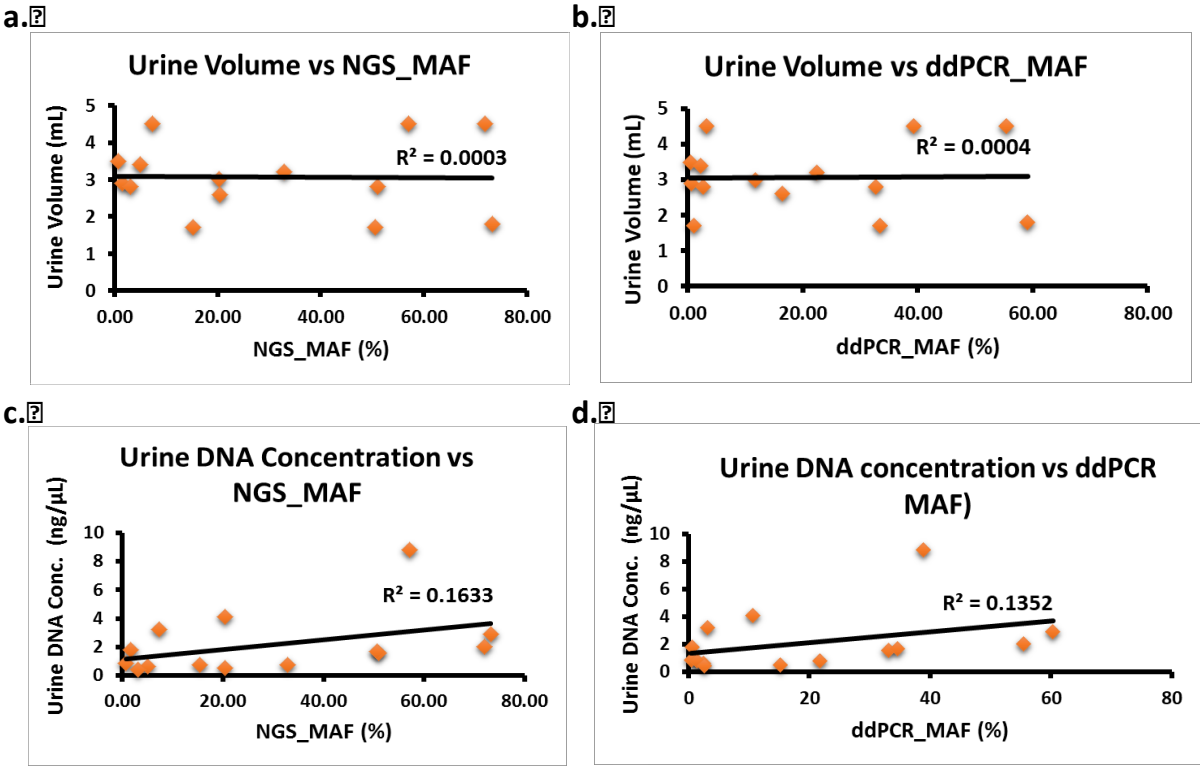
Supplementary Figure 3. Factors affecting urine DNA concentration and success of the UroMuTERT assay. Evaluation of the association between urine volume and DNA concentration after DNA isolation (A); the association between the successful NGS reads obtained after *TERT* mutation screening with the UroMuTERT assay and the amount of urinary DNA (B) and the volume of urine (C). Samples with poor or no data (N=57) and samples with data (N=131) are displayed.

Supplementary Table 3. Proportion of successful UroMuTERT assay according to the calendar year of urine sample collection.

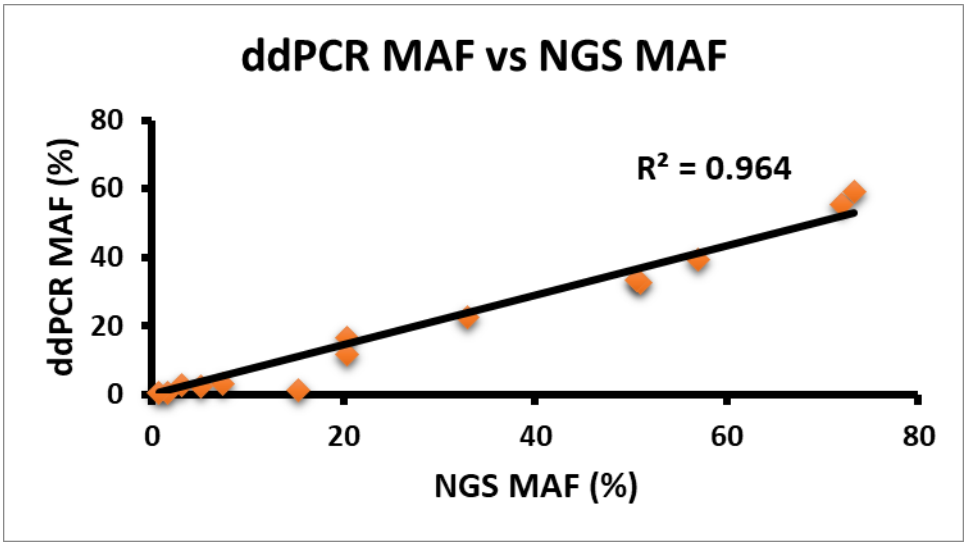
Year of Urine Sample Collection	Successful in UroMuTERT, Number (%)	Failed in UroMuTERT, Number (%)	Total, Number
2004	9 (30·00)	21 (70·00)	30
2005	21 (58·33)	15 (41·67)	36
2006	38 (86·36)	06 (13·64)	44
2007	43 (76·79)	13 (23·21)	56
2008	20 (90·91)	02 (9·09)	22
Total	131 (69·68)	57 (30·32)	188
P-value <0·001 for the difference between the groups.			

Detection of *TERT* promoter mutations in pre-diagnostic urine samples

The urine volume used in this study and the urinary DNA concentration were not associated with the mutant allelic fraction (MAF) of the *TERT* promoter mutations detected in these samples using either ddPCR or UroMuTERT (NGS) approaches (**Supplementary Figure 4**). The MAFs detected using the UroMuTERT (NGS) and ddPCR assays were highly correlated (r2=0·96), (**Supplementary Figure 5**), validating the quantitative measurement of the mutational load by the two methods.



Supplementary Figure 4. Correlation between urine volume and urine DNA concentration with the mutant allelic fraction (MAF) detected by Next-Generation Sequencing (NGS) assay (a and c respectively) or Digital Droplet PCR (ddPCR) approach (b and d respectively).



Supplementary Figure 5 Correlation between Mutant Allelic Fraction (MAF) detected by the Next-generation Sequencing (NGS) and Digital Droplet PCR (ddPCR) based methods.

Supplementary Table 4. Clinical and *TERT* mutational data of the bladder cancer cases from the Golestan Cohort.

Anonymized PID	Sex	Residence	Time from urine collection to BC Dx (Years)	TERT mutation status	Average NGS MAF (%)	Average ddPCR MAF (%)	Death	Death-Reason	TNM	Stage	Grade
BC16	Male	urban	10·3	C228T	7·3	3·3	1	Bladder Cancer	T3N1M0	3A	3
BC09	Male	rural	2·6	C228T	57·1	39·2	1	Bladder Cancer	T1N0M0	1	3
BC03	Male	rural	9·1	ND	ND	ND	1	Bladder Cancer	T2bN0M0	2	3
BC02	Female	urban	9·5	ND	ND	ND	1	Bladder Cancer	TxNxM1	4	
BC27	Male	urban	7·3	ND	ND	ND	0		T1N0M0	1	1
BC25	Male	urban	8·1	ND	ND	ND	0		T1N0M0	1	1
BC19	Male	urban	6·7	CC242-243TT	72·0	55·4	1	Bladder Cancer	TxNxM1	4	3
BC26	Male	urban	7·9	ND	ND	ND	0		T1N0M0	1	1
BC21	Female	urban	9·3	ND	ND	ND	0		TxNxM1	4	3
BC22	Male	rural	8·9	ND	ND	ND	0		NA		
BC15	Male	urban	9·8	C228A	3·1	2·7	1	Other Symptoms, Signs and Abnormal Findings, NOS	NA		
BC12	Female	rural	8·7	C228T	1·6	0·6	1	Bladder Cancer	T2-3NxM1	4	3
BC05	Male	rural	0·4	ND	ND	ND	1	Bladder Cancer	Tis	0	
BC14	Female	rural	7·5	ND	ND	ND	1	Bladder Cancer	T4bNxM0	4A	3
BC29	Female	rural	2·8	ND	ND	ND	0		T1N0M0	1	1
BC18	Female	urban	8·4	ND	ND	ND	0		NA		1
BC06	Male	rural	9·8	C228T	20·4	16·5	0		T4aN2M1a 4	4	3
BC24	Male	rural	7·3	C228T	0·7	0·5	0		T2N0M0	2	3
BC08	Male	rural	0·4	ND	ND	ND	1	Bladder Cancer	T2N0M0	2	3
BC17	Female	rural	8·7	C250T	20·3	12·3	0		NA	3	3
BC11	Male	rural	0·9	ND	ND	ND	1	Bladder Cancer	T1 OR T2	1	1
BC01	Male	rural	9·3	ND	ND	ND	1	Bladder Cancer	T2N0M0	2	3
BC10	Male	rural	5·6	C228T	5·0	2·2	1	Bladder Cancer	TxNxM1	4	3
BC04	Male	rural	2·9	C228T	50·6	33·4	1	Bladder Cancer	T2NxM1	4	3
BC20	Male	rural	4·2	C250T	15·3	3·2	1	Bladder Cancer	T3NxM0	3A	3
BC30	Male	rural	1·1	C228T	51·0	32·7	0		T1N0M0	1	1
BC28	Male	rural	2·8	C228T	32·9	22·4	0		T1N0M0	1	1
BC13	Female	rural	2·3	ND	ND	ND	1	Bladder Cancer	NA		
BC07	Male	rural	5·4	C228T	73·3	59·1	1	Bladder Cancer	T2N0M0	2	3
BC23	Female	rural	6·9	ND	ND	ND	0		NA		2

NA: Data not available.