Surveying mutation density patterns around specific genomic features

(Supplemental Materials)

Contact: Yan Guo, yanguo1978@gmail.com August 31st, 2022

Contents

Properties of an outdated replication origin dataset (Haradhvala et al. 2016) 2
Figure S1. Landscape of replication origins across 24 chromosomes 2
Figure S2. A/T/G/C content in vicinity of replication origins 3
MutDens application on enhancers in seven ICGC cohorts 4
Supplemental Figure S3. Mutation density patterns around enhancers 4
Properties of mutation data in liver cancer cohorts 5
Figure S4. Mutation burden statistics in two liver cancer cohorts. 5
Table S1. Population-specific liver cancer cohorts and mutation statistics 5
Figure S5. Imbalance between C>A and G>T mutation forms in LICA-CN 6
Comparison of mutation density between RIP and Origin 6
Figure S6. RIP shows higher mutation density than Origin in LMS-FR 6
Melanoma C>T transcriptional strand bias revealed in random subsets of samples 6
Figure S7. Transcriptional strand bias was revealed in ten random MELA-AU patients 6
Figure S8. Transcriptional strand bias was blurry in random single MELA-AU patients 8
Runtime analysis of MutDens 8
Figure S9. MutDens computational time scales linearly with position/mutation quantity 8
Option configuration file samples 9
Supplemental_Code_S6 (options_1reg.txt) 9
Supplemental_Code_S7 (options_2cht.txt) 10
Supplemental_Code_S8 (options_2reg.txt) 10

Additional, separate files

Supplemental_Code_S1.txt Primary R code

Supplemental_Code_S2.txt Secondary R code

Supplemental_Code_S3.txt Rmd file 1 (modality that entails input of one mutation file and one focal position set)

Supplemental_Code_S4.txt Rmd file 2 (MutDens can compare two sample cohorts on one set of focal positions)

Supplemental_Code_S5.txt Rmd file 3 (MutDens can compare two sets of focal positions for one sample cohort)

Properties of an outdated replication origin dataset (Haradhvala et al. 2016)

Figure S1. Landscape of replication origins across 24 chromosomes

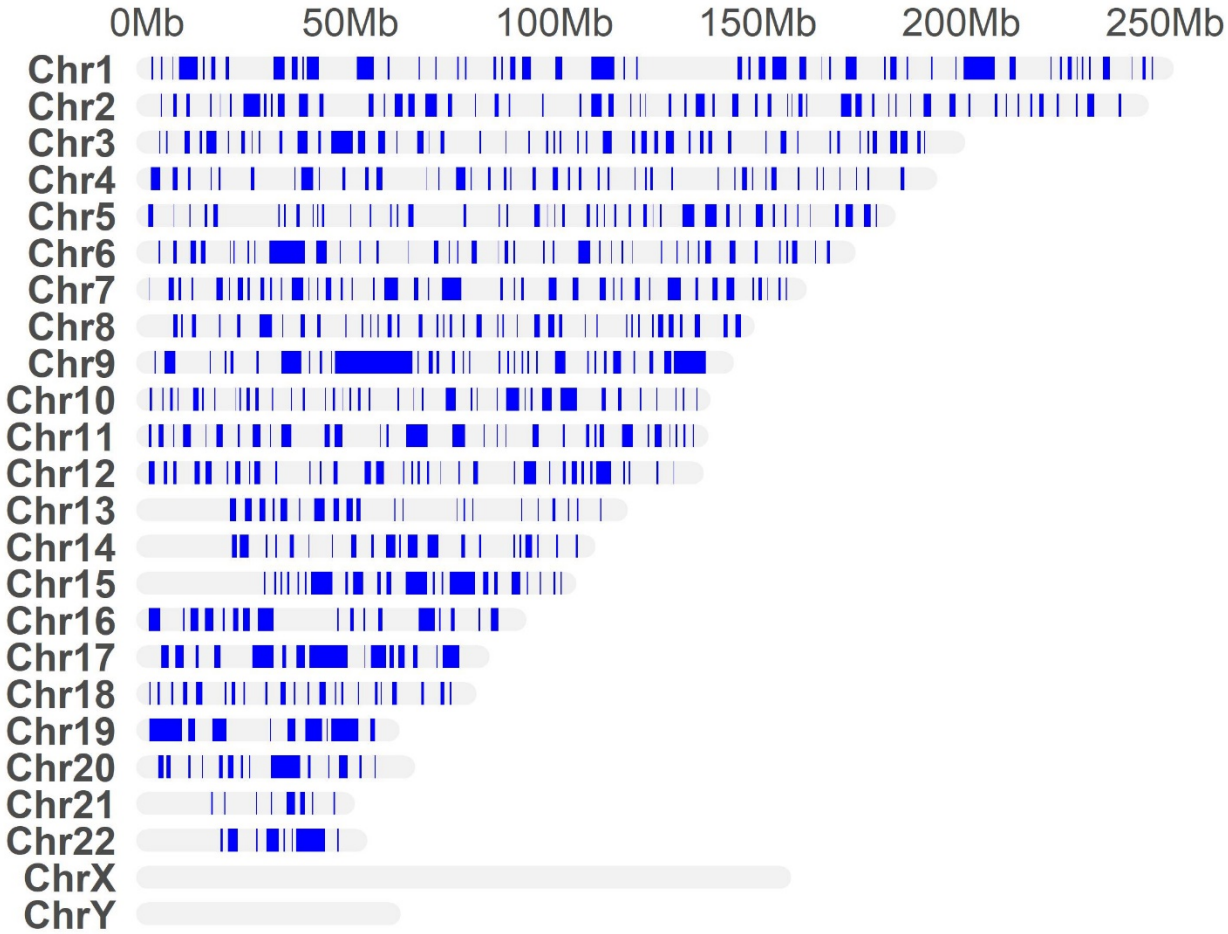
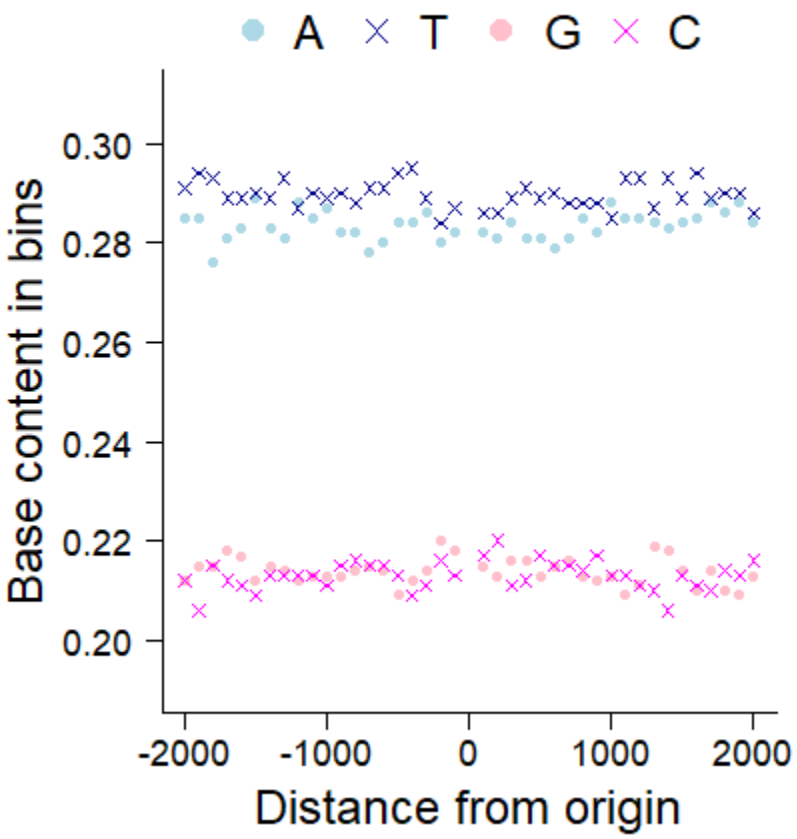
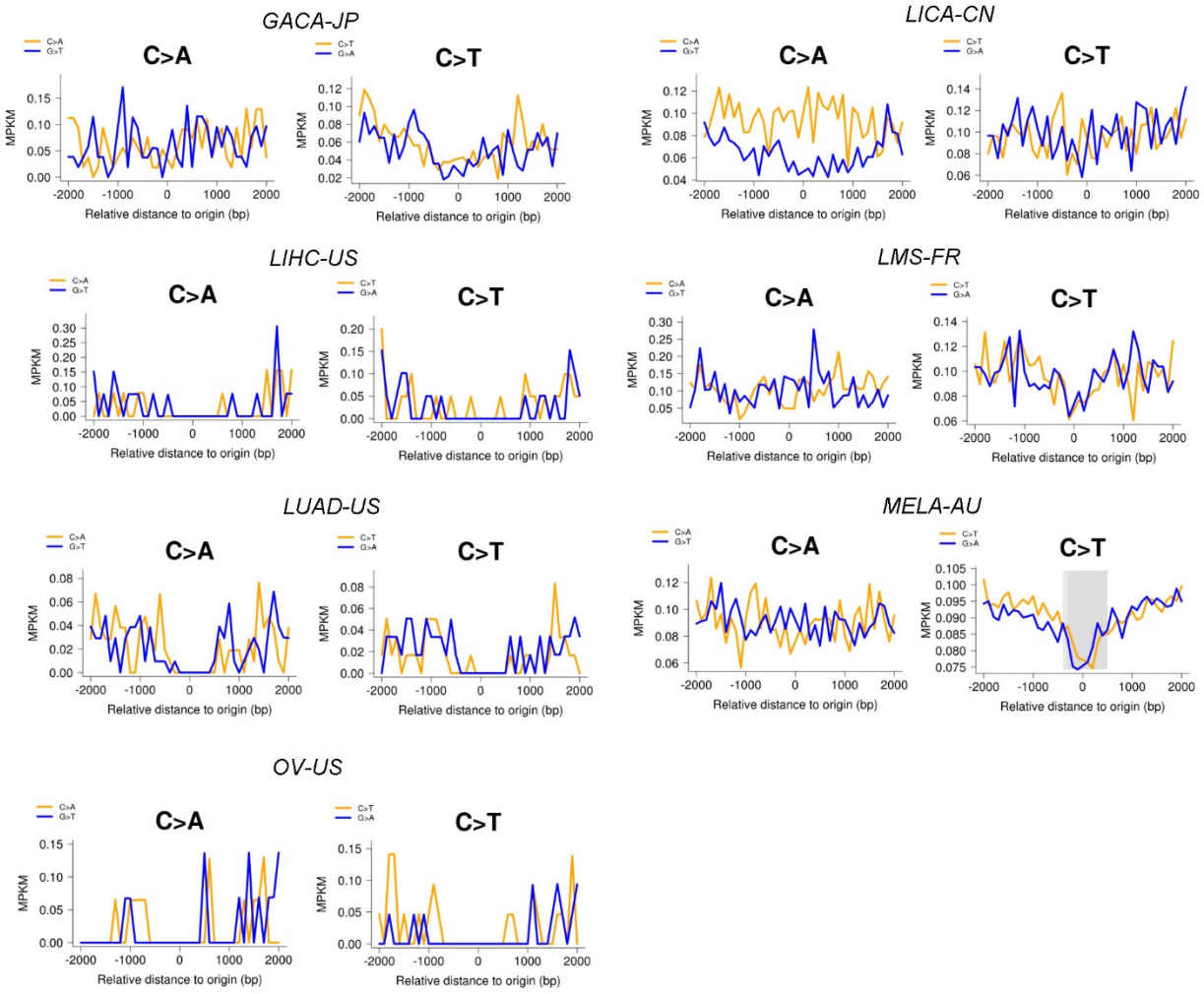


Figure S2. A/T/G/C content in vicinity of replication origins



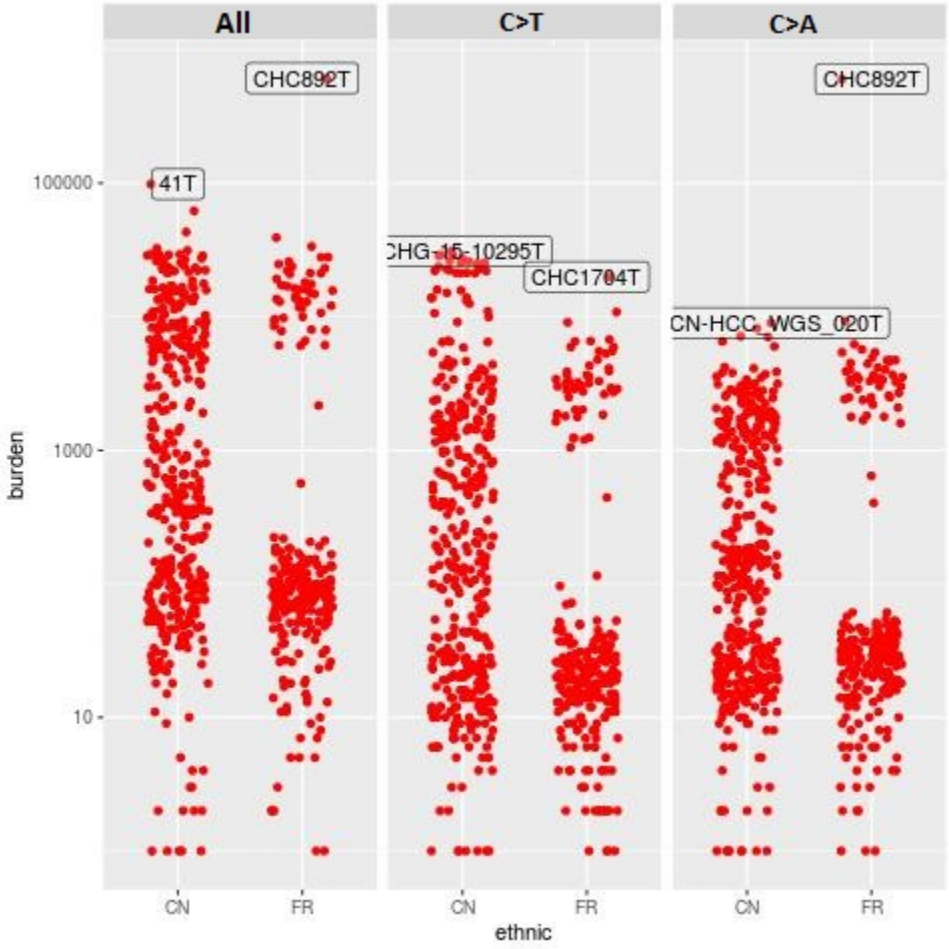
MutDens application on enhancers in seven ICGC cohorts

Supplemental Figure S3. Mutation density patterns around enhancers



Properties of mutation data in liver cancer cohorts

Figure S4. Mutation burden statistics in two liver cancer cohorts.



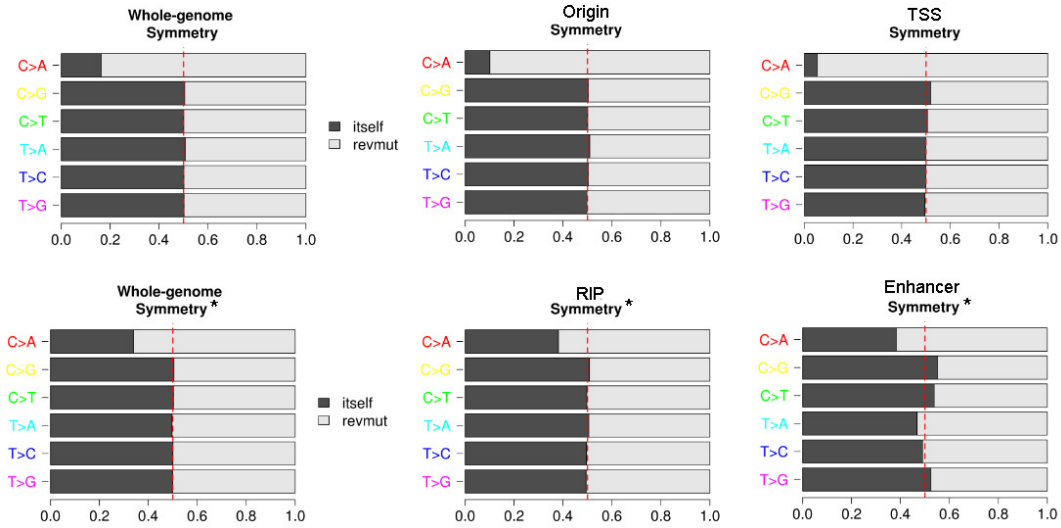
The single patients of the highest mutation burden in each perspective were identified with the patient code in the respective cohort. CN, LICA-CN; FR, LICA-FR.

Table S1. Population-specific liver cancer cohorts and mutation statistics

	Cohort	#patients	#mutations*	Median burden	Maximum burden
All mutations	CN	400	1,983,143	537	98,093
	FR	254	1,369,678	91	609,503
C>A	CN	390	888,142	173	30,924
	FR	250	193,505	21	19,897
C>T	CN	396	284,314	119	8,998
	FR	250	770,613	30	597,946

*Number of unique mutation observed in the whole cohort.

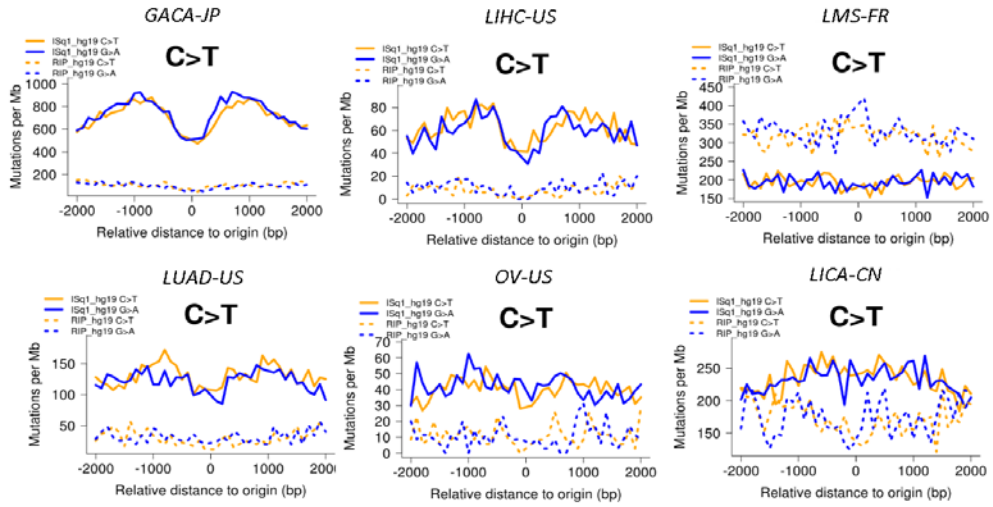
Figure S5. Imbalance between C>A and G>T mutation forms in LICA-CN



* The bottom three plots were based on a trimmed cohort with 40 hyper-mutated patients excluded (the top three plots used the whole cohort).

Comparison of mutation density between RIP and Origin

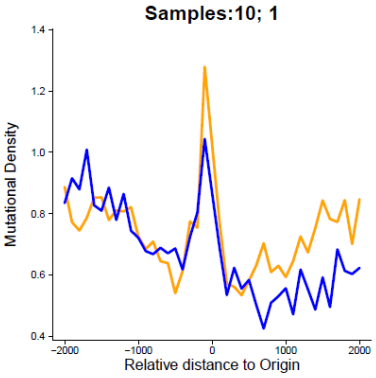
Figure S6. RIP shows higher mutation density than Origin in LMS-FR

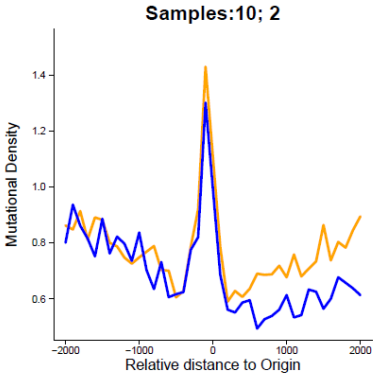


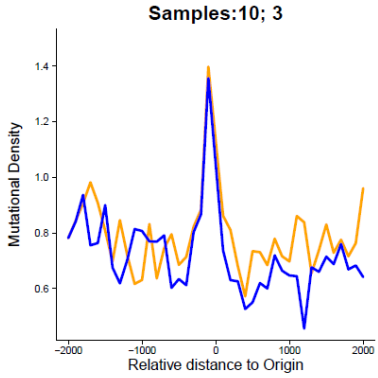
Melanoma C>T transcriptional strand bias revealed in random subsets of samples

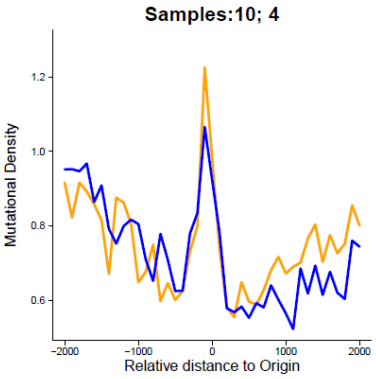
Figure S7. Transcriptional strand bias was revealed in ten random MELA-AU patients

Random selection of ten patients was repeated five times.


Samples:10; 1
Mutational Density
Relative distance to Origin


Samples:10; 2
Mutational Density
Relative distance to Origin


Samples:10; 3
Mutational Density
Relative distance to Origin


Samples:10; 4
Mutational Density
Relative distance to Origin

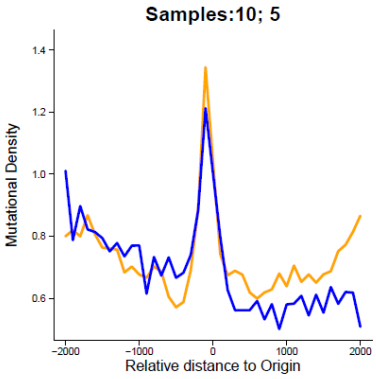
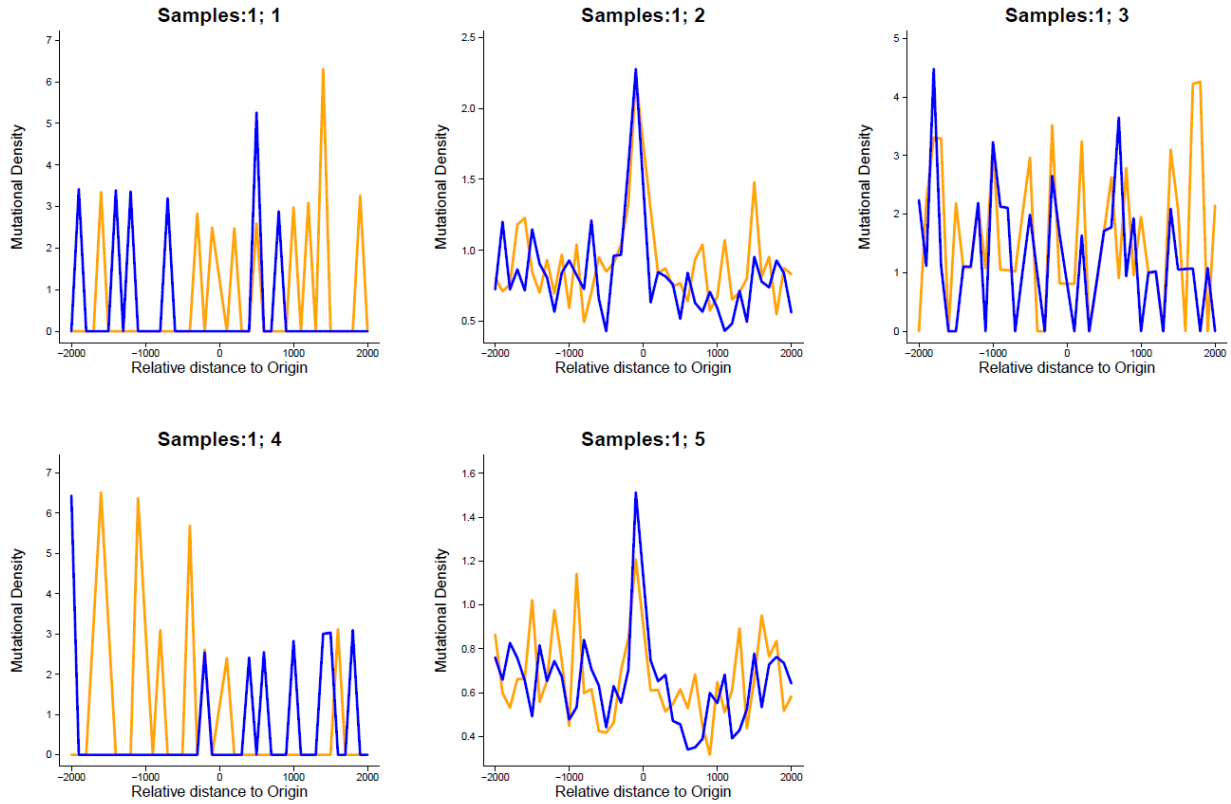

Samples:10; 5
Mutational Density
Relative distance to Origin

Figure S8. Transcriptional strand bias was blurry in random single MELA-AU patients

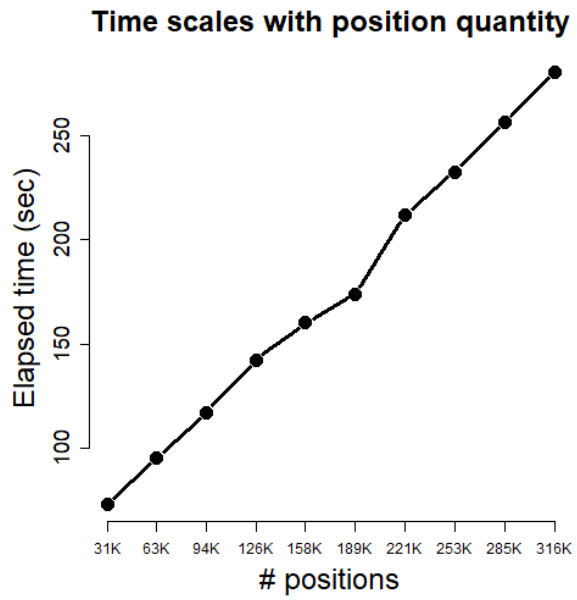
Random selection of a single patients was repeated five times.

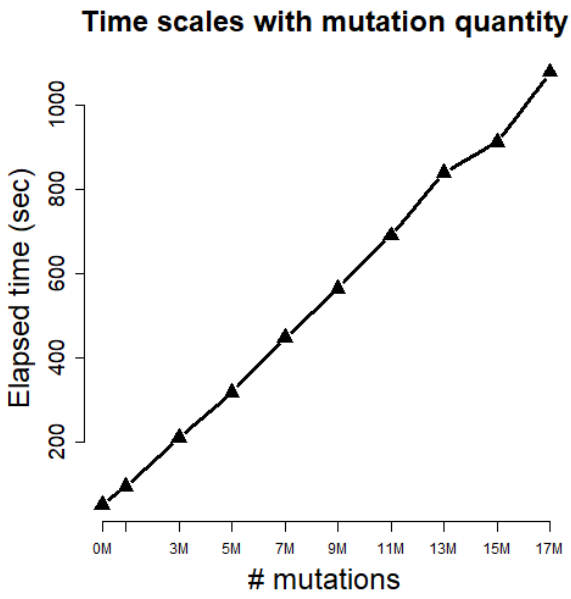


Runtime analysis of MutDens

Figure S9. MutDens computational time scales linearly with position/mutation quantity

Sessions were tested on a Linux Ubuntu work station with Intel Xeon CUP E5-2650 V4 @ 2.20GHz and 32 GB memory.





Option configuration file samples

Each option file must be tab-delimited. The following shows the contents for three option files: options_1reg.txt, options_2cht.txt, and options_2reg.txt. These three option files are required by Supplemental_Code_S3.txt, Supplemental_Code_S4.txt, and Supplemental_Code_S5.txt, respectively. For more related files (data files, example output, option explanation, etc.), please refer to https://github.com/hui-sheen/MutDens

Supplemental_Code_S6 (options_1reg.txt)

Key	Value
bsz	100
calcSrc	NULL
gFt	TSS
gn	hg19
inspan	2000
mutF	mutFiles/LICA-FR.off10.tsv
outerspan	7000
pointsF	data/proTSS_hg19.tsv
sbs	SBS6
shapeModel	pois

span	2000
triCont	FALSE

Supplemental_Code_S7 (options_2cht.txt)

Key	Value
bsz	100
calcSrc	NULL
gFt	TSS
gn	hg19
inspan	2000
mutF1	mutFiles/LICA-CN.off10.tsv
mutF2	mutFiles/LICA-FR.off10.tsv
outerspan	7000
pointsF	data/proTSS_hg19.tsv
sbs	SBS6
shapeModel	pois
span	2000
triCont	FALSE

Supplemental_Code_S8 (options_2reg.txt)

Key	Value
bsz	100
calcSrc	NULL
gFt1	ISq1
gFt2	eRNA
gn	hg19
inspan	2000
mutF	mutFiles/LICA-CN.off10.tsv
outerspan	7000
points1F	data/ISq1_hg19.tsv
points2F	data/eRNA_hg19.tsv

sbs SBS6

shapeModel pois

span 2000

triCont FALSE