Distinctive indel mutational signature in tumors carrying TOP2A p.K743N

Arnoud Boot[1,2*], Steven G. Rozen[1,2*]

[1] Cancer and Stem Cell Biology, Duke-NUS Medical School, 8 College Road, 169857, Singapore

[2] Centre for Computational Biology, Duke-NUS Medical School, 8 College Road, 169857, Singapore

* Correspondence to: AB: arnoud.boot@duke-nus.edu.sg, SGR: steve.rozen@duke-nus.edu.sg

Abstract

Topoisomerases are essential for genome stability. Here we link a rare recurrent hotspot mutation, p.K743N, in topoisomerase TOP2A to a unique pattern of small insertions and deletions (indels) in human cancers. This pattern consists mainly of duplications of 2 to 4 base pairs and deletions of 6 to 8 base pairs. These resemble indel mutational signatures ID17 and ID8, which until now had unknown causes. In total, we identified 7 tumors carrying TOP2A p.K743N, all of which showed both ID8 and ID17 characteristics. Furthermore, ID17 was absent from all 12,269 tumors lacking this mutation. We also report evidence of structural variation associated with TOP2A p.K743N. We propose that TOP2A p.K743 induces a novel mutational signature incorporating ID17 and features of ID8, which we call ID_TOP2A. This is the first report of topoisomerase-associated mutagenesis in human cancers, shedding further light on TOP2A's crucial role in genome maintenance and importance in cancer.

Main text

Type II topoisomerases are essential for genome maintenance. They untangle DNA by creating and then repairing double strand breaks to allow entangled DNA strands to pass

through each other and thus untangle[1-3]. During this process, the 5' ends of the breaks are bound to the topoisomerase with a phosphotyrosyl bond. After the breaks have been repaired, the phosphotyrosyl bond is cleaved, releasing the topoisomerase-DNA complex[4]. Humans have 2 genes encoding type II topoisomerases: *TOP2A* on chromosome 17 and *TOP2B* on chromosome 3 (reference [5]). TOP2A and TOP2B function similarly, but TOP2A is mainly expressed in proliferating cells, whereas TOP2B is also expressed in quiescent cells[6]. Topoisomerases have been widely studied as therapeutic targets, with several topoisomerase inhibitors, such as etoposide and doxorubicin, being commonly used chemotherapeutics[5,7]. Additionally, upregulation of topoisomerases in cancer has been shown to be a marker of poor prognosis[8,9]. However, to date, there have been no reports of somatically altered type II topoisomerases playing a role in human tumorigenesis.

To study whether defects in TOP2A or TOP2B contribute to carcinogenesis, we looked for hotspot mutations in 23,829 tumors[10]. TOP2B showed a clear hotspot at p.R651, and TOP2A showed several hotspots in various protein domains (Fig. 1a, 1b). TOP2B p.R651 and most recurrent TOP2A mutations were in hypermutated tumors (Fig. 1c, Supplementary Table S1). TOP2A p.K743N on the other hand, was observed in 4 gastric cancers (GCs) and 1 cholangiocarcinoma (CCA_TH_19), none of which were hypermutated. Strikingly, all GCs carrying TOP2A p.K743N showed more small insertions and deletions (indels) than single base substitutions. Examining the indel spectra of the 4 GCs showed a distinct pattern that was previously identified as indel mutational signature ID17, which is characterized by duplications of 2 to 4 base pairs (bp), not in repetitive sequences[10] (Fig. 1d). Among 12,273 tumors with indels analyzed by Alexandrov et al., TOP2A p.K743N and ID17 occurred only in these 4 GCs ($p = 1.06\times10^{-15}$, two-sided Fisher's exact test). Alexandrov et al. did not analyze indels in CCA_TH_19. In addition, COSMIC contained 2 additional tumors carrying TOP2A p.K743N, and WES was available for these[11,12]. Re-alignment and indel calling in CCH_TH_19 and the 2 tumors in COSMIC revealed ID17 in all three (Fig. 1d).

Hypothesizing that ID17 is related to DNA cleavage by TOP2A, and because TOP2A cleavage is enriched in highly transcribed regions[13], we examined ID17 mutagenesis in relation to transcriptional activity. Consistent with this, indel mutagenesis in CCA_TH_19 increased with transcriptional activity ($p = 6.13\times10^{-53}$, one-sided Cochran-Armitage test, Fig. 2a). Notably, in addition to duplications of 2 to 4 bp not in repetitive sequences, all other classes of indels were also positively correlated with transcriptional activity, suggesting that they also stemmed from TOP2A p.K743N mutagenesis (FDR < 0.05 for all classes, one-sided Cochran-Armitage test, Fig. 2b). As with CCA_TH_19, all whole-exome sequenced samples also showed higher density of indel mutagenesis in more highly transcribed regions ($p = 0.016$, one-sided sign-test, Fig. 2c).

We also noted that the pattern of ≥ 5 bp deletions in TOP2A p.K743N tumors closely resembled indel mutational signature ID8, which consists almost entirely of ≥ 5 bp deletions. In fact, in the indel spectra of the 4 p.K743N GCs, Alexandrov et al., assigned the ≥ 5 bp deletions to ID8[10]. Furthermore, the numbers of ID17-like mutations and ≥ 5 bp deletions were correlated in the p.K743N tumors. This suggested that the signature of TOP2A p.K743N might in fact be a combination of ID17-like mutations and ≥ 5 bp deletions. This raised two additional questions, because ID8 occurs in the majority of tumors lacking TOP2A p.K743N and occurs across most cancer types.

First, we considered whether ID8 might represent indels caused by TOP2A absent the p.K743N substitution. To investigate this, we compared the ≥ 5 bp deletions in the TOP2A p.K743N carriers to those in all WGS tumors with indels analyzed by Alexandrov et al[10]. These deletions were positively correlated with transcriptional activity in only 124 tumors (Supplementary Table S2). Most of these tumors showed ID8 mutagenesis (82.6%), and 70.2% were liver cancers (Supplementary Table S2, Supplementary Fig. S1). However, in the vast majority tumors with ID8, transcriptional activity and ID8 density were uncorrelated. Thus, in most tumors, ID8 mutagenesis seems unrelated to TOP2A activity.

Second, we considered whether the grouping together of all ≥ 5 bp deletions as part of categorizing indels might have obscured important differences in deletion size distributions stemming from different mutational processes. To investigate this, we compared the sizes of ≥ 5 bp deletions in TOP2A p.K743N tumors to the sizes of these deletions in tumors with many ID6 or ID8 mutations, as these two signatures consist mainly of ≥ 5 bp deletions. We separately analyzed ID8-high tumors with or without increased ID8 mutagenesis in highly transcribed regions (ID8_GE8pos and ID8_GEneg, respectively, Fig. 2d, Supplementary Table S2). We observed marked enrichment for 6 to 8 bp deletions in TOP2A p.K743N carriers as compared the other tumors ($p = 7.49\times10^{-6}$, 1.94×10^{-5} and 1.97×10^{-5} for ID6, ID8_GEpos and ID8_GEneg respectively, two-sided Wilcoxon rank sum tests).

To summarize, in addition to the duplications of 2 to 4 bp that were previously reported as ID17, TOP2A p.K743N also induces substantial numbers of 6 to 8 bp deletions, which previous deletion categorizations did not separate from other ID8 deletions[10]. We have shown that these 6 to 8 bp deletions correlate with transcriptional activity, as do all the other major classes of indels observed in TOP2A p.K743N carriers, which is consistent with known mechanistic properties of TOP2A. Thus, we define a novel indel mutational signature incorporating ID17 and features of ID8, which we call ID_TOP2A (Fig. 4, Supplementary Table S3).

As TOP2A creates and repairs double-strand breaks, and arrested TOP2A cleavage complexes have previously been shown to induce genomic rearrangements[14], we also investigated whether TOP2A p.K743N might induce large-scale, structural variants (SVs). Because SVs are strongly depleted in exons, and WGS was available only for CCA_TH_19, we compared the SVs previously reported for this tumor with the SVs in the other CCAs in the same cohort[15]. CCA_TH_19 had the highest SV-load, consisting mainly of deletions (41.5%) and inter-chromosomal translocations (27.6%, Fig. 3a). Similar to the indels, SVs were also enriched in highly transcribed regions ($p = 2.56\times10^{-12}$, one-sided Cochran-Armitage test, Fig. 3b). Enrichment in more highly transcribed regions was also observed for SV deletions,

inter-chromosomal translocations, intra-chromosomal translocations and insertions (FDRs 1.93×10^{-2}, 4.00×10^{-9}, 1.33×10^{-2} and 1.24×10^{-3} respectively, one-sided Cochran-Armitage tests, Fig. 3b). This suggests that the SVs observed in CCA_TH_19, are also associated with TOP2A activity.

Although we have only observed this pattern of indel mutagenesis in human tumors carrying the TOP2A p.K743N substitution, recent reports studying a TOP2 double-mutant (p.F1025Y and p.R1128G) in yeast (*Saccharomyces cerevisiae*) suggests that other TOP2A alterations might induce 2 to 4 bp duplications similar to those in ID_TOP2A[16]. As we observed in a human TOP2A p.K743N tumor, the yeast double mutants showed strong increases in SVs. This fits with previous observations that trapped TOP2 complexes in yeast result in increased SVs[14]. Human TOP2A p.K743N is located within the DNA cleavage domain of TOP2A, and neither yeast TOP2 p.F1025 nor p.R1128 is located within the orthologous domain. Furthermore, neither of the amino acids altered in TOP2 p.F1025 or p.R1128 is conserved in human TOP2A or TOP2B. Moreover, TOP2 overexpression also resulted in indels similar to the ID_TOP2A and the indels observed in the p.F1025 and p.R1128 double mutant[16]. Given that such different alterations, as well as overexpression of wild-type TOP2, generated similar patterns of indels suggests that still other alterations might result in similar mutagenesis. Therefore it is likely that as more human cancers are sequenced, more somatic TOP2A alterations that induce ID_TOP2A will be discovered.

In conclusion, we identified novel indel mutational signature ID_TOP2A, which is characterized by *de novo* duplications of 2 to 4 bp and deletions of 6 to 8 bp, and, we posit, is induced by TOP2A p.K743N substitutions.

Methods

Data sources

We used published mutation spectra from 23,829 tumors (https://www.synapse.org/#!Synapse:syn11726601)[10]. Variant calls for all 2780 WGS samples from the ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium and gene expression data for a subset of these were obtained from the ICGC data portal (https://dcc.icgc.org/releases/current/Projects/)[17]. Sequencing reads from samples 04-112 and 10T were kindly provided by Professors Peter S. Nelson and Fergus J. Couch[11,12]. Sequencing reads from CCA_TH_19, for which Alexandrov et al. had not previously analyzed indels, were downloaded from EGA (EGAS00001001653). SV calls for 70 whole-genome sequenced cholangiocarcinomas were obtained from supplementary information of the associated publication[15].

Reanalysis of raw sequencing data

Read alignment, variant calling and filtering were performed as described previously[18], except that reads were aligned to GRCh38.p7. For analysis of SV occurrence as a function of transcriptional activity, the transcriptional activity at the location of the SV breakpoint with the highest transcriptional activity was taken.

Mutational signature analysis

We used the classification for indel mutational signatures as proposed previously[10]; for details see https://www.synapse.org/#!Synapse:syn11801742. Mutational signatures were plotted using ICAMSv2.1.2.9000 (github.com/steverozen/ICAMS).

Correlation between transcriptional activity and mutagenesis in WGS data

For each tumor type, genes were assigned to 1 of 4 gene expression bins. Subsequently for every sample, variants were grouped by these same expression bins by linking the genomic position of the variant to the genes assigned to expression bins. Mutation density was reported as events per Mb to compensate for differences in size of the total gene expression bins. Cochran-Armitage test for trend was performed to determine statistical significance. As not all ICGC projects have RNA-sequencing data available, for some ICGC

cohorts we used RNA-sequencing data from similar cohorts. For details, see Supplementary Table S2.

Definition of ID6-high and ID8-high tumors in WGS data

For comparison of the size distributions of ≥ 5 bp deletions in ID6-high and ID8-high tumors, we classified tumors using the existing indel signature assignments[10]. ID6-high tumors were those with > the median of counts of ID6 mutations across tumors that had > 0 ID6 mutations. ID8-high tumors were selected analogously. In addition, in selecting ID8-high tumors, tumors with > 0 ID6 mutations were excluded.

Statistics

Statistical analyses were performed in R v.3.6.3; multiple testing correction was done according to Benjamini-Hochberg method using the p.adjust function. Cochran-Armitage tests were performed using the DescTools package[19].

Acknowledgements

We thank Sue Jinks-Robertson (Duke University Medical Center) for sharing preliminary data from her TOP2 yeast data that led us to the discovery of the link between ID17 and TOP2A, and Dmitry Gordenin (US National Institute of Environmental Health Sciences) for initiating discussions between our laboratory and Professor Jinks-Robertson's laboratory. We also thank professors Peter S. Nelson and Fergus J. Couch for kindly providing sequencing data. This work was supported by a Khoo Postdoctoral Fellowship Award (Duke-NUS-KPFA/2018/0027) to AB and MOH-000032/MOH-CIRG18may-0004 to SGR.

Author contributions

AB and SGR both contributed to the conception, analysis, writing, and editing of the manuscript.

Figure legends

Fig. 1: **TOP2A and TOP2B hotspot mutations and associated mutagenesis**. Recurrent amino acid substitutions in TOP2A (**a**) and TOP2B (**b**) generated using MutationMapper[20]. **c,** Mutagenesis profiles of TOP2A and TOP2B hotspot carriers as reported by Alexandrov et al. Ovals in top sub-panel show combined mutation load of indels, DBSs and SBSs; bars in bottom panel show the proportions of indels (red), DBSs (green) and SBSs (gray). **d,** Indel spectra of tumors carrying the TOP2A p.K743N substitution, and the ID8 and ID17 mutational signatures as reported by Alexandrov et al. Abbreviations: bp: base pair, DBS: Doublet base substitution, indels: small insertions and deletions, MH: microhomology, Mut: mutation, SBS: Single base substitution, WES: Whole exome sequencing, WGS: whole-genome sequencing.

Fig. 2: **Characteristics of TOP2A p.K743N associated indels**. **a,** Density of TOP2A p.K743N associated indels correlates with transcriptional activity in CCA_TH_19. This is consistent with the known enrichment of TOP2A activity in highly transcribed regions[13]. **b,** All major indel classes show increased density of indel mutagenesis with increased transcriptional activity in CCA_TH_19; *, **, ***, **** indicate Benjamini-Hochberg false discovery rates of < 0.05, < 0.01, < 0.001, and < 0.0001, based on p values from one-sided Cochran-Armitage tests for trend. **c,** All WES samples also show higher indel density in highly transcribed regions (p = 0.016, one-sided sign-test). Because of low indel counts in WES data, mutations were split into low (light gray) and high (dark gray) expression bins. **d,** Comparison of size distribution of ≥ 5 bp deletions in TOP2A p.K743N carriers, in tumors with high counts of ID6 deletions, in tumors with high counts of ID8 deletions with increasing mutation density in highly transcribed genes (ID8_GEpos), and in tumors with high counts of ID8 deletions without this characteristic

(ID8_GEneg). Please see Online Methods for details. Compared to each of the other groups of tumors, TOP2A p.K743N carriers were enriched for 6 to 8 bp deletions ($p < 0.0001$ for each of the pairwise comparisons, two-sided Wilcoxon rank sum tests). For ID6, ID8_GEneg and ID8_GEpos, means and 95% confidence intervals are plotted. For display purposes, all TOP2A p.K743N WES data was combined, statistical comparisons between the groups was performed using the individual data points. Abbreviations: bp: base pair, Del: deletion, Ins: insertion, indels: small insertions and deletions, MH: microhomology, WES: Whole exome sequencing.

Fig. 3: **Structural variants in CCAs**. **a,** Total numbers and composition of SVs in 70 CCAs[15]. **b**, In CCA_TH_19, all types of SVs except inversions correlate with transcriptional activity, which is similar to indels in this tumor; *, **, ***, **** indicate Benjamini-Hochberg false discovery rates of < 0.05, < 0.01, < 0.001, and < 0.0001, based on p values from two-sided Cochran-Armitage tests for trend. Abbreviations: Mb: megabase, NS: not significant, SV: structural variation.

Fig. 4: **The ID_TOP2A mutational signature**. Consensus indel signature observed in tumors carrying TOP2A p.K743N. Each bar shows the mean proportion of indels in a given indel class; error bars show standard error of the mean.

Supplementary Material

Sup.Fig. S1: **Correlation of ⩾ 5 bp deletions with transcriptional activity in publicly available tumor whole genome sequence**. **a,** Volcano plot of correlation between density of ≥ 5 bp deletions and transcriptional activity. Genomes were divided into 4 gene expression bins. The X-axis shows the ratio of the deletion density in the medium and high expression

bins versus deletion density in the lowest and low expression bins; the Y-axis show false discovery rates for increasing deletion density across the 4 expression bins (Benjamini-Hochberg false discovery rates based on two-sided Cochran-Armitage tests for trend). **b,** Total number of ≥ 5 bp deletions per tumor type. Underlying data is in Supplementary Table S2. Abbreviations: bp: base pair, FDR: false-discovery rate

Sup.Table S1: **Tumors carrying TOP2A or TOP2B hotspot mutations**.

Sup.Table S2: **ID8 mutagenesis (≥ 5 bp deletions) as a function of transcriptional activity in WGS data**

Sup.Table S3: **The ID_TOP2A indel mutational signature**

References

1. Vos, S.M., Tretter, E.M., Schmidt, B.H. & Berger, J.M. All tangled up: how cells direct, manage and exploit topoisomerase function. *Nat Rev Mol Cell Biol* **12**, 827-41 (2011).
2. Pendleton, M., Lindsey, R.H., Jr., Felix, C.A., Grimwade, D. & Osheroff, N. Topoisomerase II and leukemia. *Ann N Y Acad Sci* **1310**, 98-110 (2014).
3. Wang, J.C. Moving one DNA double helix through another by a type II DNA topoisomerase: the story of a simple molecular machine. *Q Rev Biophys* **31**, 107-44 (1998).
4. Nitiss, J.L. & Nitiss, K.C. Tdp2: a means to fixing the ends. *PLoS Genet* **9**, e1003370 (2013).
5. Pommier, Y., Leo, E., Zhang, H. & Marchand, C. DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. *Chem Biol* **17**, 421-33 (2010).
6. Sakaguchi, A. & Kikuchi, A. Functional compatibility between isoform alpha and beta of type II DNA topoisomerase. *J Cell Sci* **117**, 1047-54 (2004).
7. Liang, X. *et al.* A comprehensive review of topoisomerase inhibitors as anticancer agents in the past decade. *Eur J Med Chem* **171**, 129-168 (2019).
8. Heestand, G.M., Schwaederle, M., Gatalica, Z., Arguello, D. & Kurzrock, R. Topoisomerase expression and amplification in solid tumours: Analysis of 24,262 patients. *Eur J Cancer* **83**, 80-87 (2017).
9. Ren, L., Liu, J., Gou, K. & Xing, C. Copy number variation and high expression of DNA topoisomerase II alpha predict worse prognosis of cancer: a meta-analysis. *J Cancer* **9**, 2082-2092 (2018).
10. Alexandrov, L.B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94-101 (2020).
11. Kumar, A. *et al.* Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. *Nat Med* **22**, 369-78 (2016).
12. Murphy, S.J. *et al.* Genetic alterations associated with progression from pancreatic intraepithelial neoplasia to invasive pancreatic tumor. *Gastroenterology* **145**, 1098-1109 e1 (2013).

13. Yu, X. *et al.* Genome-wide TOP2A DNA cleavage is biased toward translocated and highly transcribed loci. *Genome Res* **27**, 1238-1249 (2017).
14. Gomez-Herreros, F. *et al.* TDP2 suppresses chromosomal translocations induced by DNA topoisomerase II during gene transcription. *Nat Commun* **8**, 233 (2017).
15. Jusakul, A. *et al.* Whole-Genome and Epigenomic Landscapes of Etiologically Distinct Subtypes of Cholangiocarcinoma. *Cancer Discov* **7**, 1116-1135 (2017).
16. Stantial, N. *et al.* Trapped Topoisomerase II initiates formation of de novo duplications via the nonhomologous end-joining pathway in yeast. *bioRxiv* (2020).
17. The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature* **578**, 82-93 (2020).
18. Boot, A. *et al.* Identification of novel mutational signatures in Asian oral squamous cell carcinomas associated with bacterial infections. *bioRxiv* (2019).
19. Signorell, A. *et al.* DescTools: Tools for descriptive statistics. 0.99.35 edn (2020).
20. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**, pl1 (2013).

Figure 1

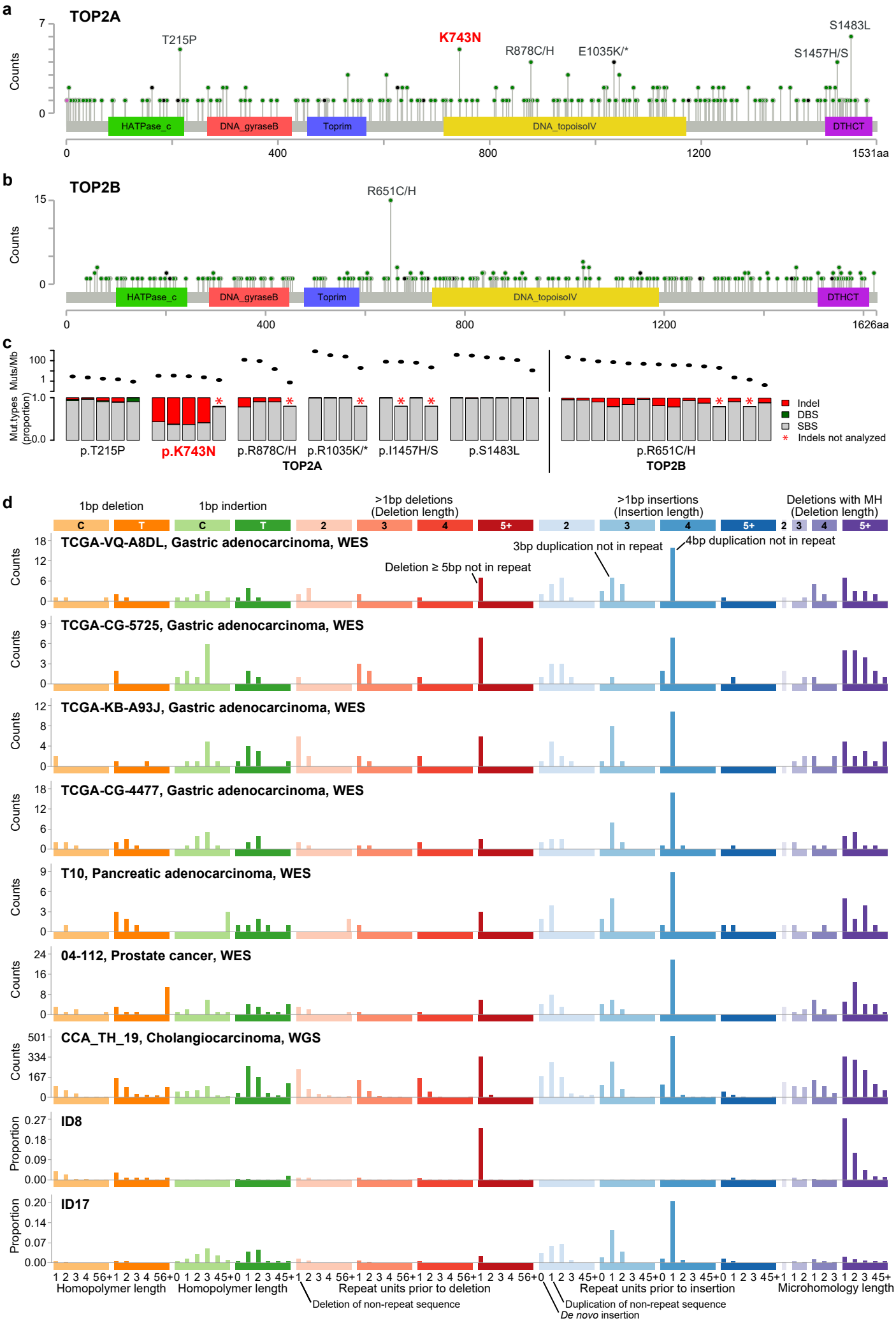
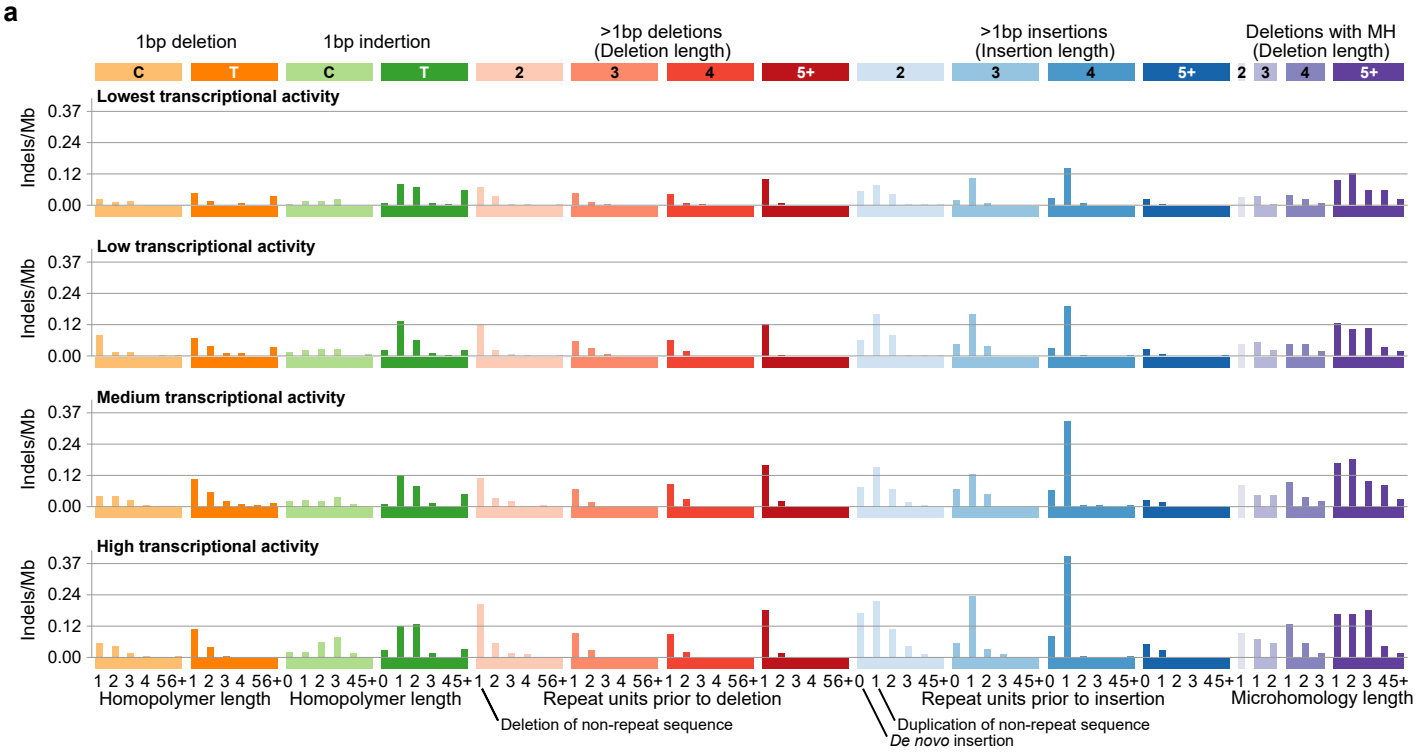
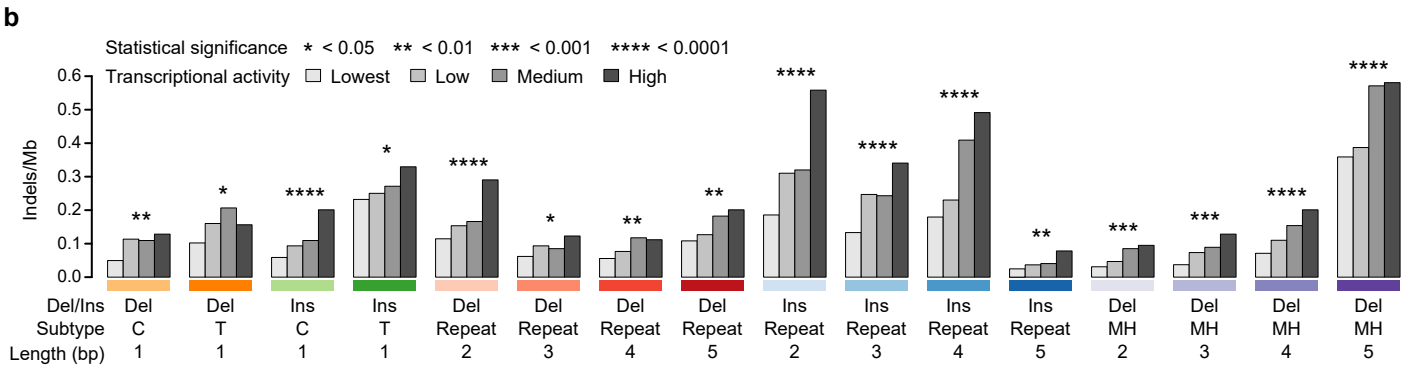
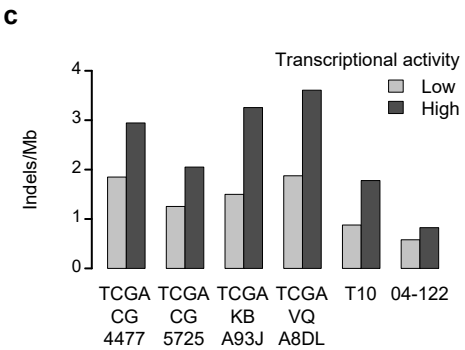

a
TOP2A
Counts
T215P
K743N
R878C/H
E1035K/*
S1483L
S1457H/S
HATPase_c
DNA_gyraseB
Toprim
DNA_topoisoIV
DTHCT
0
400
800
1200
1531aa
b
TOP2B
R651C/H
1626aa
c
Muts/Mb
Mut.types (proportion)
p.T215P
p.K743N
p.R878C/H
p.R1035K/*
p.I1457H/S
p.S1483L
TOP2A
p.R651C/H
TOP2B
Indel
DBS
SBS
Indels not analyzed
d
1bp deletion
1bp indertion
>1bp deletions (Deletion length)
>1bp insertions (Insertion length)
Deletions with MH (Deletion length)
TCGA-VQ-A8DL, Gastric adenocarcinoma, WES
3bp duplication not in repeat
4bp duplication not in repeat
Deletion ≥ 5bp not in repeat
TCGA-CG-5725, Gastric adenocarcinoma, WES
TCGA-KB-A93J, Gastric adenocarcinoma, WES
TCGA-CG-4477, Gastric adenocarcinoma, WES
T10, Pancreatic adenocarcinoma, WES
04-112, Prostate cancer, WES
CCA_TH_19, Cholangiocarcinoma, WGS
ID8
ID17
Proportion
Homopolymer length
Repeat units prior to deletion
Repeat units prior to insertion
Microhomology length
Deletion of non-repeat sequence
Duplication of non-repeat sequence
De novo insertion

Figure 2


a
1bp deletion
1bp indertion
>1bp deletions (Deletion length)
>1bp insertions (Insertion length)
Deletions with MH (Deletion length)
Lowest transcriptional activity
Low transcriptional activity
Medium transcriptional activity
High transcriptional activity
Indels/Mb
Homopolymer length
Repeat units prior to deletion
Repeat units prior to insertion
Microhomology length
Deletion of non-repeat sequence
Duplication of non-repeat sequence
De novo insertion


b
Statistical significance * < 0.05 ** < 0.01 *** < 0.001 **** < 0.0001
Transcriptional activity Lowest Low Medium High
Indels/Mb
Del/Ins
Subtype
Length (bp)


c
Transcriptional activity
Low
High
Indels/Mb
TCGA CG 4477
TCGA CG 5725
TCGA KB A93J
TCGA VQ A8DL
T10
04-122

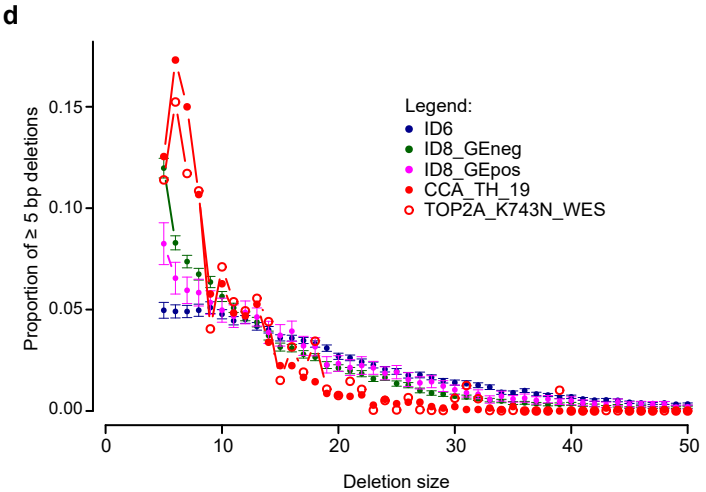

d
Legend:
ID6
ID8_GEneg
ID8_GEpos
CCA_TH_19
TOP2A_K743N_WES
Proportion of ≥ 5 bp deletions
Deletion size

Figure 3

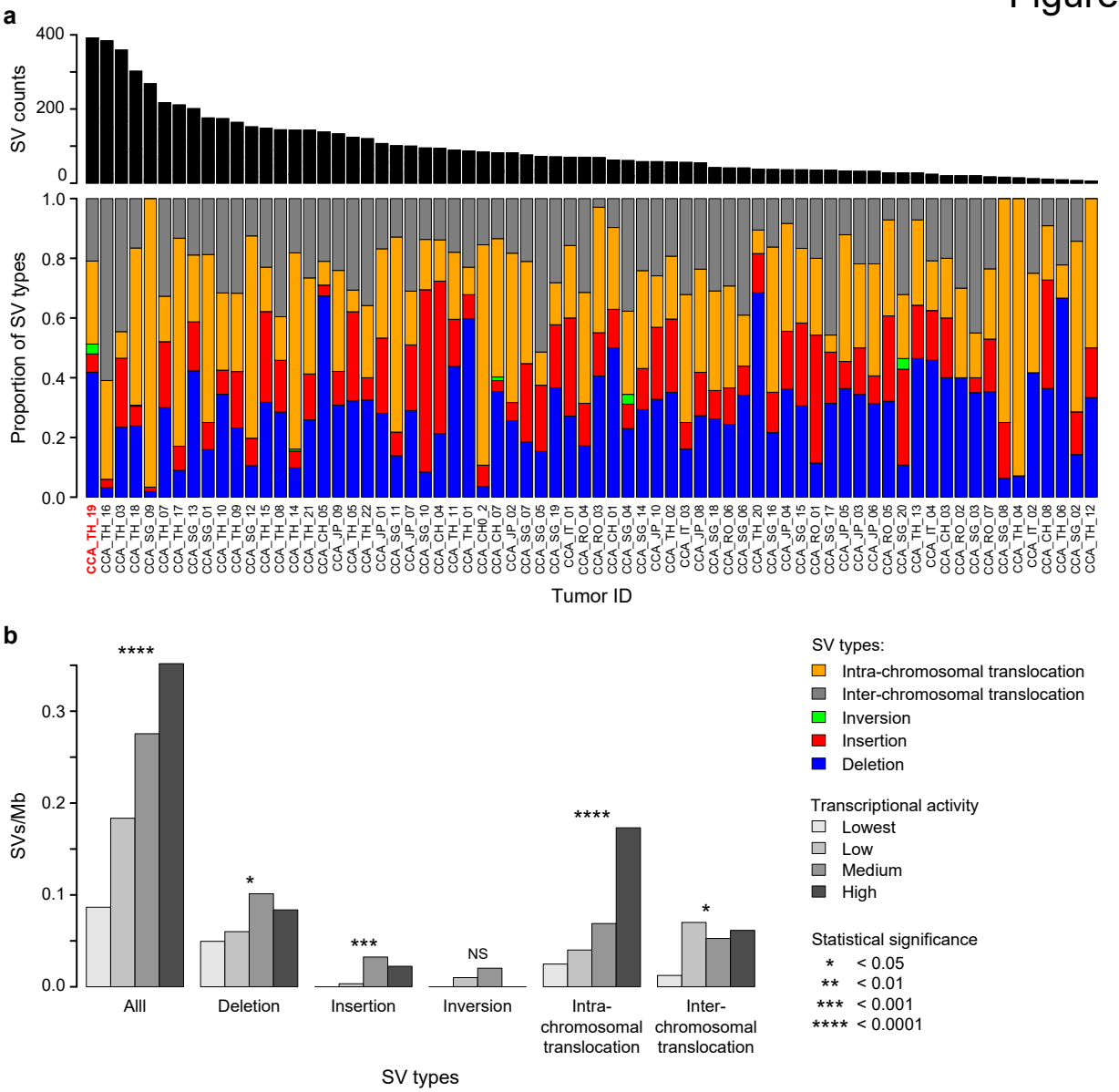
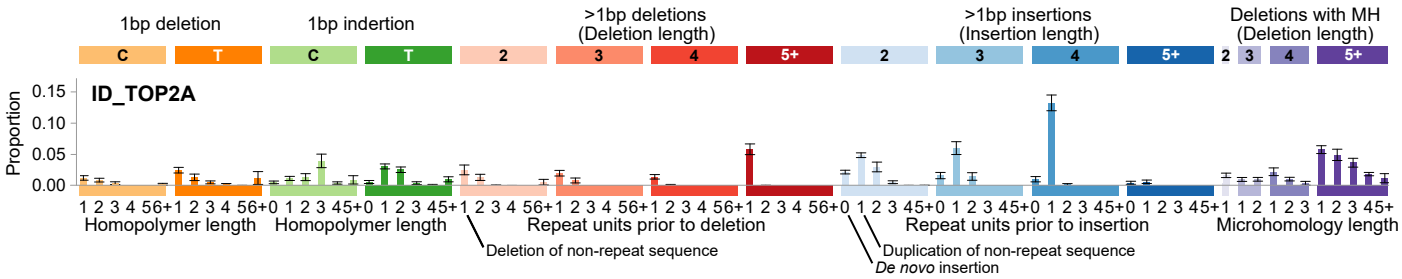

a
SV counts
400
200
0
Proportion of SV types
1.0
0.8
0.6
0.4
0.2
0.0
Tumor ID
b
SVs/Mb
0.3
0.2
0.1
0.0
****
*
***
NS
****
*
Alll
Deletion
Insertion
Inversion
Intra-chromosomal translocation
Inter-chromosomal translocation
SV types
SV types:
Intra-chromosomal translocation
Inter-chromosomal translocation
Inversion
Insertion
Deletion
Transcriptional activity
Lowest
Low
Medium
High
Statistical significance
* < 0.05
** < 0.01
*** < 0.001
**** < 0.0001

Figure 4


1bp deletion
1bp indertion
>1bp deletions
(Deletion length)
>1bp insertions
(Insertion length)
Deletions with MH
(Deletion length)
C
T
C
T
2
3
4
5+
2
3
4
5+
2
3
4
5+
ID_TOP2A
Proportion
0.15
0.10
0.05
0.00
1 2 3 4 56+1 2 3 4 56+0 1 2 3 45+0 1 2 3 45+1 2 3 4 56+1 2 3 4 56+1 2 3 4 56+1 2 3 4 56+0 1 2 3 45+0 1 2 3 45+0 1 2 3 45+0 1 2 3 45+1 1 2 1 2 3 1 2 3 45+
Homopolymer length
Homopolymer length
Repeat units prior to deletion
Repeat units prior to insertion
Microhomology length
Deletion of non-repeat sequence
Duplication of non-repeat sequence
De novo insertion