Heliyon

Received:
22 June 2018
Revised:
9 January 2019
Accepted:
19 February 2019

Cite as: Min Zhang, Da Yang, Barry Gold. Origin of mutations in genes associated with human glioblastoma multiform cancer: random polymerase errors versus deamination.
Heliyon 5 (2019) e01265.
doi: 10.1016/j.heliyon.2019.e01265

Origin of mutations in genes associated with human glioblastoma multiform cancer: random polymerase errors versus deamination

Min Zhang, Da Yang, Barry Gold*

Department of Pharmaceutical Sciences, University of Pittsburgh, 708 Salk Hall, 3501 Terrace Street, Pittsburgh PA 15261, USA

* Corresponding author.
E-mail address: goldbi@pitt.edu (B. Gold).

Abstract

The etiology of glioblastoma multiforme (GBM), the most serious form of brain cancer, remains obscure, although it has been proposed that cancer risk is a function of random polymerase errors that occur during stem cell division and the resulting mutations in oncogenes and tumor suppressor genes. Analysis of the 8 genes (PTEN, TP53, EGFR, PIK3R1, PIK3CA, NF1, RB1, IDH1) that are mutated in at least 5% of GBM tumors indicates a non-random mutation pattern that reflects a significant role for hydrolytic deamination at CpG sites. The formation of activating mutations in some genes, e.g., IDH1, where a very limited set of mutations are oncogenic, statistically cannot involve random mutagenesis due to polymerase errors that occur during each stem cell replication. Comparison of the in vitro misincorporation tendencies of three replicative polymerases and the "random" mutation pattern in a subset of genes indicates non-polymerase based pathways are involved. Analysis of the mutation patterns shows that chemical deamination that occurs at a slow rate at each CpG is favored over random polymerase errors by a factor of more than 10 million. Therefore, if a truncating nonsense mutation in a tumor suppressor, or an

https://doi.org/10.1016/j.heliyon.2019.e01265

activating missense mutation in an oncogene, can occur due to a C > T base substitution at a CpG sequence, it is highly favored over other mutation pathways.

Keywords: Genetics, Cancer research, Biochemistry

1. Introduction

The etiology of glioblastoma multiforme (GBM), the most serious form of brain cancer, is obscure. Attempts to assign specific causative factors responsible for driver mutations in critical oncogenes and tumor suppressor genes have generally not been informative [1, 2, 3]. It has been proposed based on compelling data that a dominant factor in the relative incidence of different cancers in different tissues, including GBM, is related to the low random DNA polymerase-based error rate that occurs each time a stem cell replicates its DNA [4]. The yield of mutations, and the corresponding lifetime cancer risk, is therefore predicted to be a function of the number of stem cells in a tissue and how often they replicate over a period of time. It was noted that for some cancers, but not GBM, that the correlation between the number of stem cell turnovers and cancer risk is not as strong, suggesting factors other than replication related errors. The polymerase miscoding error rate would be sensitive to the level of DNA damage from increased exposure of the genome to endogenous (e.g., RNOS) and exogenous (e.g., heterocyclic aromatic amines) promutagenic chemicals [5, 6] due to the recruitment of lower fidelity DNA repair and/or translesion synthesis polymerases.

In the current work, we have dissected the mutation patterns in 8 genes associated with 3 oncogenic pathways [2, 7, 8, 9], that are frequently mutated in GBM in order to determine whether the mutation pattern is, as predicted, based on random polymerase errors, or whether there is a level of specificity that can be attributed to other mutagenic pathways. The genetic alterations in PTEN, TP53, EGFR, PIK3R1, PIK3CA, NF1, RB1 and IDH1, which are "highly significantly associated" ($\geq$5% incidence, Table 1) with GBM [7, 8, 9], were analyzed. While it is not possible to correlate any environmental factor with brain cancer incidence, it is clear that a significant fraction of the critical mutations in genes associated with GBM are not derived from random errors introduced by DNA polymerases. As observed in other cancers, endogenous C deamination at gene body CpG, and at non-CpG sites, plays an important role in generating the mutations in many GBM oncogenes and tumor suppressor genes [10, 11, 12, 13, 14]. An unambiguous example of the role of deamination at CpG sites is detailed for the highly targeted mutation pattern in GBM tumors bearing activated IDH1 alleles where it is possible to compare the mutagenic rates derived from polymerase errors versus deamination. While the origin of the mutations may not be tightly correlated with polymerase errors, the source of the mutations is likely dependent on the number of stem cells and stem cell replications

https://doi.org/10.1016/j.heliyon.2019.e01265

Table 1. Missense and nonsense mutation profile in GBM associated suppressor genes and oncogenes.

Somatic mutation[a] (incidence in GBM)	PTEN (31.1%)	TP53 (28.9%)	EGFR (26.4%)	PIK3R1 (11.4%)	PIK3CA (10.3%)	NF1 (10.6%)	RB1 (9.2%)	IDH1 (4.8%)	Mean% ± s.d.[b]
C•G→T•A at CpG sites	17 (27.0%)[c]	25 (31.3%)	7 (8.2%)	0	4 (16.7%)	4 (25.0%)	5 (33.3%)	12 (92.3%)	29.3 ± 28.0
C•G→T•A at non-CpG sites	20 (31.7%)	17 (21.3%)	41 (48.2%)	5 (33.3%)	8 (33.3%)	8 (50.0%)	4 (26.7%)	0	30.4 ± 15.9
C•G→A•T	3 (4.8%)	15 (18.8%)	19 (22.4%)	1 (6.7%)	1 (4.2%)	2 (12.5%)	1 (6.7%)	0	9.5 ± 7.8
C•G→G•C	6 (9.5%)	5 (6.3%)	7 (8.2%)	1 (13.3%)	1 (4.2%)	1 (6.3%)	1 (6.7%)	1 (7.7%)	8.6 ± 4.5
T•A→A•T	3 (4.8%)	1 (1.3%)	1 (1.2%)	3 (20.0%)	2 (8.4%)	0	2 (13.3%)	0	7.8 ± 2.7
T•A→G•C	3 (4.8%)	5 (6.3%)	9 (10.6%)	1 (6.7%)	5 (20.8%)	0	2 (13.3%)	0	7.8 ± 7.0
T•A→C•G	10 (15.9%)	12 (15.0%)	1 (1.2%)	4 (26.7%)	3 (12.5%)	1 (6.3%)	0	0	9.7 ± 9.5
Total missense + nonsense mutations	63	80	85	15	24	16	15	13	14.7 ± 10.4

[a] Highly significantly mutated genes occurring in GBM tumors (% incidence in GBM tumors) [7,8].

[b] Random mutagenesis predicts 16.7% for each type of mutation.

[c] Percent of the total of missense and nonsense mutations.

https://doi.org/10.1016/j.heliyon.2019.e01265

where DNA is transiently single-stranded, and more exposed to solvent and reactive molecules, accelerating deamination and other DNA damaging reactions.

2. Methods

2.1. Patient data

There are 604 GBM samples in the TCGA Provisional glioblastoma multiforme database [7, 8]. There are 8 genes that are mutated in at least 5% of the tumors. In these 8 genes there are a total of 446 mutations of which 311 are a combination of missense and nonsense mutations. For each tumor sample, there is information on the gender and age at diagnosis.

2.2. Statistical analysis

Standard statistical tests were used to analyze the clinical and genomics data, including the Chi-square and t-test. Significance was defined as a P value of less than 0.05. Analyses were primarily performed using R Foundation for Statistical Computing (http://www.r-project.org/) and SPSS version 18 (SPSS Inc., Chicago, Illinois).

3. Results and discussion

3.1. Mutation spectra in key GBM driver genes

The analyses of the missense and nonsense mutations in 8 genes associated with GBM are outlined below. Splice site mutations, and deletion and insertion frame-shifts, are not included because it is often not possible to assign the specific base that is the origin of the mutation. For some genes, e.g., PTEN, these non-base substitutions account for ~10% and 22%, respectively, of the mutations in the GBM cohort [7, 8]. The remaining 68% being missense and nonsense mutations. For other genes, such as TP53, the contributions of splice site (7%) and frame shift (7%) mutations are lower. While the impact of nonsense mutations should be relatively independent of the nature of the stop codon and how it is generated, this is certainly not the case for missense mutations whose phenotype can vary with the nature and location of the amino acid substitution. Most, but not all, of the nonsense and missense mutations observed in the tumor samples are predicted to be oncogenic [7, 8].

3.1.1. PTEN

Mutation of PTEN, which has 403 codons, is the second most common oncogenic mutation in human cancers, and the most frequently mutated gene in GBM,

https://doi.org/10.1016/j.heliyon.2019.e01265

occurring in 30% of the tumors (Table 1) [7, 8]. There are 141 codons in PTEN that can be mutated to a stop codon by single base substitution. Of these 141 codons, 3 (2.1%) are CGA Arg codons. However, 11 of the 16 nonsense mutations (68.8%) in the GBM cohort are C•G→T•A transitions that occur at CGA codons (Table S1). This is >30-fold above what would be statistically anticipated based upon random polymerase mutations that can yield stop codons. Of the 47 missense mutations in the GBM cohort, 6 (12.8%) are at CGB (B = G, C, T) codons, which account for 1.0% of the 403 codons in the gene. In total, 27.0% of the PTEN nonsense and missense mutations occur at the 12 CGN (N = A, G, C, T) codons that account for only 3.0% of the total. This corresponds to a 9-fold increase above the statistically anticipated mutation frequency. Moreover, there are only 8 codons in PTEN that are mutated ≥2 times in different tumor samples in the GBM cohort, and of the 25 mutations that occur at these 8 sites, 18 (72.0%) are C•G→T•A transitions at Arg CGN codons. An example of selective generation of mutations at CGN sites in PTEN is that the R233 CGA codon is mutated in 5 samples, yet the adjacent R232 (AGG) codon is not mutated in any GBM tumor in the cohort. In addition, conversion of AGA Arg codons to TGA stop codons via T•A→A•T transversions is not observed even though this type of mutation is predicted for replicative polymerases (see discussion below) [15, 16, 17].

3.1.2. TP53

TP53 is mutated in 29% of the GBM samples (Table 1) [8]. There are 108 potential codons out of a total of 393 in TP53 that can be converted to a stop codon by a single base substitution. Of these, 4 (3.7%) are CGA Arg codons yet 3 of the 4 nonsense TP53 mutations observed in GBM are C•G→T•A transitions found at these codons. The 20-fold enrichment of TP53 nonsense mutations at CGA in GBM is similar to that reported in other tumor tissues [14, 18]. There are 20 Arg CG(T, C, G) codons in TP53 out of the total of 393 codons (5.1%). The CG(C,T, G) codons are targeted with 28 (36.4%) C•G→T•A transitions out of the total of 77 missense mutations, which is 7.1-fold higher than predicted based on a random mechanism (Table S2). These missense mutations are focused in the protein's DNA binding domain (codons 102−292) and assumed to be oncogenic. Of the 50 codons in TP53 that are mutated at least twice in the GBM cohort, 24 (54%) involve C•G→T•A mutations at CGN codons.

3.1.3. EGFR1

Missense mutations are most common in the extracellular domain of EGFR, with A89→$^{V}/_{D}$ and G598V predominating presumably due to their effect on increasing receptor auto-phosphorylation in the absence of ligand [19, 20]. These two hotspots are unique to GBM: lung adenocarcinomas also frequently have EGFR mutations

https://doi.org/10.1016/j.heliyon.2019.e01265

but the mutations are located in the kinase domain. There are 1210 amino acids in EGFR with 36 encoding Arg (CG[T/C/G]). This is 3.0% of all the codons and 7 (8.1%) of the 86 missense mutations occur at these codons. Of the 37 mutations that occur at the same codon in ≥2 tumor samples, 16 (43.2%) are at CG(T/C/G) sequences (Table S3). A truncating mutation in EGFR via a C•G→T•A mutation in a CGA codon would presumably be lethal [21, 22].

3.1.4. PIK3R1

Mutations in PIK3R1, which codes for the p85α protein that regulates the catalytic activity of p110α kinase encoded by PIK3CA, are observed in 15 of GBM samples. The protein has 724 amino acids. The 22 CGN codons account for 3.0% of the total but no mutations are found at these Arg codons. There are only 3 codons mutated more than once in PIK3R1: G376R (GGA→AGA), K379N (AAA→AAT) and $D560^{G}/_{H}$ (GAC→GGC/CAC) (Table S4), which are in the SH2 and iSH2 domains that interact with residues on the protein encoded by PIK3CA to inhibit kinase activity [23]. There are no CpG sequences in the SH2 and iSH2 interaction domains.

3.1.5. PIK3CA

PIK3CA encodes the p110α kinase with 1069 amino acids that is regulated by p85α. Mutated PIK3CA, which is observed in 10% of GBM, shows a loss of enzymatic regulation by p85α that is encoded by PIK3R1 (see above). PIK3CA has 28 CGN codons (2.6%), and 4 mutations occur at CGN, which is 16.7% of the total of 24 base substitutions. None of the CGN encoded amino acids are in the region associated with the regulatory domain of p110α (see above). Of the 10 mutations observed in ≥2 GBM tumors, 20% are at a CG(T/C/G) codon. The dominant missense mutations are at $E542^{K}/_{V}$ (GAA→AAA/GTA), E545K/A (GAG→AAG/GCG) and Q546K (CAG→AAG) (Table S5). This region is devoid of CGN codons.

3.1.6. NF1

The tumor suppressor NF1, which is mutated in 19 GBM samples, contains 2818 codons, of which 59 (2.1%) are CGN codons. Of the 12 nonsense mutations observed in NF1, 5 (41.7%) are C•G→T•A transitions that occur at CGA Arg codons, which is 19.9-fold above that statistically anticipated. Two of these mutations are at codon R192, which is the only codon mutated ≥2. Of the remaining 7 nonsense mutations, 3 are distributed at CAG Gln codons (Table S6). There are missense mutations at L844, D1849 and C622 in the cohort, but none of these are known to have an oncogenic phenotype [7, 8].

https://doi.org/10.1016/j.heliyon.2019.e01265

3.1.7. RB1

The tumor suppressor RB1, which has 928 codons, is mutated in 8% of GBM [8]. Of the 322 codons that can be converted into a stop codon by a single base substitution, 12 (3.7%) are CGA Arg codons. Out of the 12 nonsense mutations in the cohort, 4 (33.3%) are C•G→T•A transitions, which is 9-fold above that statistically predicted. Arg445 (CGA), which is mutated in 3 samples is the only codon mutated ≥2 (Table S7).

3.1.8. IDH1

IDH1, which has 414 codons, is mutated in 5% of GBM [8]. All of the 13 mutations are at R132 (CGT): 12 R132H (C•G→T•A) and 1 R132G (C•G→G•C) (Table S8). IDH1 is a unique oncogene since the mutations associated with cancer exquisitely switch the enzymology of the WT protein from catalyzing the sequential $NADP^+$-dependent oxidation of isocitrate (via oxalosuccinate) to α-ketoglutarate, to the NADPH-dependent reduction of α-ketoglutarate to 2-hydroxyglutarate [24]. IDH1 functions as a homodimer and mutant IDH1 acts as a dominant negative by combining with the WT allele product to afford a dysfunctional heterodimer of WT and mutant. The lower level of α-ketoglutarate, which is a co-factor in multiple enzymatic pathways, results in a diminished ability to oxidize and demethylate 5-methylcytosine (5-mC) causing a hypermethylated genome. In addition, the increased level of 2-hydroxyglutarate acts as a competitive inhibitor of many α-ketoglutarate dependent cellular reactions [25].

3.2. Nature of the mutations in GBM

The types of mutations that can occur are missense point mutations, nonsense point mutations, frame shift deletions and insertions, and splice site mutations. The frame shift deletion and insertion mutations are most likely to arise as a result of DNA replication, including replication that occurs during DNA repair. Accordingly, the frequency of frame shift mutations is expected to be lower in the brain, which has limited stem cell divisions, than in other tumors. A comparison of the ratios of truncating nonsense point mutations to truncating frame shift mutations in the tumor suppressor genes (PTEN, TP53, NF1 and RB1) commonly mutated in GBM versus the same genes in uterine endometrial carcinoma and malignant melanoma was performed (Table 2). There is no indication that frameshifts are less common relative to nonsense mutations in GBM than in ovarian tumors, or different from the ratio in all cancers.

3.3. Sequences flanking C•G→T•A and C•G→A•T mutations

3.3.1. C•G→T•A mutations

Transition C•G→T•A mutations at CpG sites are attributed to deamination of 5-mC [11]. C•G→T•A mutations at non-CpG sequences (Cp[T/C/A]) are also elevated in

https://doi.org/10.1016/j.heliyon.2019.e01265

Table 2. Ratios of frameshift deletions and insertions to nonsense mutations in tumor suppressor genes commonly mutated in GBM versus the ratios in all cancers, endometrial cancer and melanoma.

Tumor suppressor gene	Cancer	FS Del & Ins mutations	Nonsense mutations (NS)	Ratios of FS to NS
PTEN	All cancers[a]	247	147	1.68
	GBM	16	16	1.00
	Endometrial	73	53	1.38
	Melanoma	7	6	1.17
TP53	All cancers	677	685	0.99
	GBM	7	4	1.75
	Endometrial	6	7	0.86
	Melanoma	4	10	0.40
NF1	All cancers	145	18	0.77
	GBM	13	13	1.00
	Endometrial	0	11	-
	Melanoma	4	22	0.18
RB1	All cancers	146	188	0.78
	GBM	7	12	0.58
	Endometrial	1	8	0.13
	Melanoma	3	7	0.43

[a] Data for all cancers in based on relative percentages; data for GMB, uterine endometrial cancer and melanoma are the number of tumors with the specific mutations [7,8].

GBM oncogenes and tumor suppressors. Deamination of non-methylated C's yield dU•G mismatches, which are corrected by base excision repair to afford C•G→T•A mutations. If C→T mutations at CpG sites are subtracted from the total number of C→T missense and nonsense mutations for each gene (Table 1), 16.7% is the statistically predicted mutation incidence for each of the remaining 6 potential single base mutations. However, C→T mutations at Cp(T/C/A) sites account for 43.5, 30.9, 52.6, 33.3, 40.0, 66.7 and 40.0% of the missense and nonsense mutations in PTEN, TP53, EGFR, PIK3R1, PIK3CA, NF1 and RB1, respectively. In comparing C→T mutations at CpG versus Cp(T/C/A) sequences, it must be remembered that there are many more of the latter than the former. There are 202 C's in PTEN, a gene with 1212 nucleotides, that are not in CpG sequences versus 14 CpG sites (14.4-fold difference). For TP53, the ratio of Cp(T/C/A) to CpG is 7.5 (316/42). The ratios of C→T mutations at Cp(T/C/A) versus CpG sites is 1.1 and 0.7 for PTEN and TP53, respectively, making mutations at CpG sequences approximately 10-fold higher that predicted based on random deamination or polymerase errors.

The sequences flanking the C→T mutation sites (Table 3, Tables S9−S11) in PTEN, TP53 and EGFR are not consistent with enzymatic deamination by APOBEC or AID deaminases, which have a preference for 5′-T-C and 5′-$^{A}/_{T}$-$^{A}/_{G}$-C, respectively [26, 27, 28, 29, 30]. Therefore, mutations are most likely derived via non-enzymatic hydrolytic deamination of 5-mC [31].

https://doi.org/10.1016/j.heliyon.2019.e01265

It is possible that the G in the complementary strand at Cp(T/C/A) sequences is the initial site of misinsertion. Many endogenous mutagens (e.g., reactive oxygen species, free radicals, methylating agents, etc.), and exogenous mutagens (e.g., heterocyclic aromatic amines, polycyclic aromatic hydrocarbons, aflatoxins) preferentially react at G and also show modest sequence selectivity in adduct formation [32]. ROS generated by free radicals tends to mutate G in $G_{\geq 2}$ runs with the 5′-G being the most reactive [33, 34]. This would predict that C on the complementary strand with a 5′-C would be more extensively mutated. Peroxy radicals selectively react at G in 5′-NpGpC rather than CpG sequences [35], which is quite distinct from H_2O_2 mediated DNA oxidation [36]. Methylating agents, such as S-adenosylmethione (SAM) preferentially react at G runs rather than CpG sites to afford G→A mutations [37]. Polycyclic aromatic hydrocarbons, heterocyclic aromatic amines, fungal metabolites also tend to react at G runs and produce a mutational spectrum dominated by C•G→A•T transversions [38, 39, 40]. Using ligation mediated PCR, it has been shown that polycyclic aromatic amines selectively react at 5-mCpG sequences in cells but the G adduct produces G→T mutations [41]. C•G→T•A mutations at GpG•CpC sequences in PTEN, TP53 and EGFR account for 26.4 ± 10.6% of the mutations, which is close to the 25% predicted for NpC sequences (Table 3). The analysis for mutations at Cp(T/C/A) indicates that 34.0 ± 11.0% of the C•G→T•A mutations occurs at CpC, which is close to the 33% statistically predicted.

3.3.2. *C•G→A•T mutations*

C•G→A•T transversions are the second most common mutation, albeit observed significantly less frequently than the C•G→T•A mutations. Analysis of the sequences that flank all of the G→T mutations at CpN sites in PTEN, TP53, EGFR, PIK3R1, PIK3CA, NF1 and RB1 was performed (Table 4). Of the 34 C•G→A•T mutations in the 7 genes, 14 (41.2%) are at 5′-GGA sequences (Table

Table 3. Effect of flanking 5′- and 3′-base on C•G→T•A mutations at non CpG sites.

Sequence	PTEN	TP53	EGFR
ApC	7/16 (43.8%)	4/13 (30.8%)	2/41 (4.9%)
TpC	4/16 (25.0%)	1/13 (7.7%)	6/41 (14.6%)
CpC	3/16 (18.8%)	5/13 (38.5%)	9/41 (22.0%)
GpC	2/16 (12.5%)	3/13 (23.1%)	24/41 (58.5%)
CpA	8/16 (50.0%)	7/13 (53.8%)	17/41 (41.5%)
CpT	4/16 (25.0%)	4/13 (30.8%)	19/41 (46.3%)
CpC	3/16 (18.8%)	1/13 (7.7%)	1/41 (2.4%)

https://doi.org/10.1016/j.heliyon.2019.e01265

Table 4. Effect of flanking 5′- and 3′-base on C•G→A•T mutations at non CpG sites in PTEN, TP53, EGFR, PIK3R1, PIK3CA, NF1, RB1.

N	5′-N-G	G-N-3′
G	18	3
A	4	19
T	8	6
C	4	6
total	34	34

S12). This preference is dominated by the 11 G598V (G→T) mutations in EGFR and likely reflects selection versus mutation rates. Outside of EGFR, the flanking sequence has no effect on the mutation frequency suggesting random DNA damage and/or random misincorporation.

3.4. Comparison of error signatures of replicative DNA polymerases vs. GBM mutation patterns

To further probe whether random polymerase misincorporation errors are a major source of mutations in addition to the transitions at CpG sites, which are clearly not random, we compared mutation patterns in PTEN, TP53 and EGFR versus the *in vitro* fidelities of the mammalian replicative polymerases (Pol δ, Pol ε, Pol α). These 3 genes were selected because of the high percentage (>25%) of samples in the GBM cohort in which they are mutated. For the 3 polymerases, the biochemically determined fidelity rates for base substitution are very high at physiological dNTP concentrations; however, raising the dNTP concentration increased misinsertion rates [15]. The relative order of fidelity at different bases is shown in Fig. 1 (note that a statistical analysis of the error rates was not reported and there is only ~10-fold difference between the highest and lowest error rates).

The normalized misincorporation mutation spectra in the *lacZ* gene produced by the three calf thymus replicative polymerases is plotted in Fig. 2 [15]. The mutations are shown as base pairs changes (e.g., C→T + G→A = C•G→T•A). Also plotted are the normalized mutation frequencies in PTEN, TP53 and EGFR in the absence of C•G→T•A mutations at CpG sites. C•G→T•A mutations, excluding those at CpG sites, account 25−55% of the single base substitutions in the three genes. This is consistent with the misincorporation and mutagenicity data for the 3 polymerases where C•G→T•A transitions are the most dominant misincorporation error and the most observed mutation in the *LacZ* gene (Figs. 1 and 2). In these studies using a single-stranded template, there is the possibility that enhanced deamination of C in the template can give rise to a portion of the C→T mutations. The major differences between the predicted mutation fingerprint based on the three

https://doi.org/10.1016/j.heliyon.2019.e01265

Pol δ: (C→T) > (G→T) > (T→A) > (G→A) > (C→A) > (C→G) > (A→T) > (A→C) > (A→G) > (G→C) > (T→G) ~ (T→C)

Pol ε: (C→T) > (G→T) > (T→C) > (A→T) > (C→G) > (T→G) > (A→C) > (A→G) ~ (C→A) > (T→A) > (G→C) > (G→A)

Pol α: (G→A) ~ (A→T) > (T→A) > (T→C) > (C→A) > (C→T) > (G→C) > (A→G) > (G→T) > (T→G) > (C→G) > (A→C)

Mutation incidence in GBM: (C•G→T•A) > (T •A→C•G) > (C•G→A•T) > (C•G→G•C) > (T•A→A•T) > (T•A→G•C)

Fig. 1. In vitro mutation signatures for replicative DNA polymerases Pol δ, Pol ε and Pol α and the mutation fingerprint in the 8 GBM genes. Color codes for complementary base pairs.

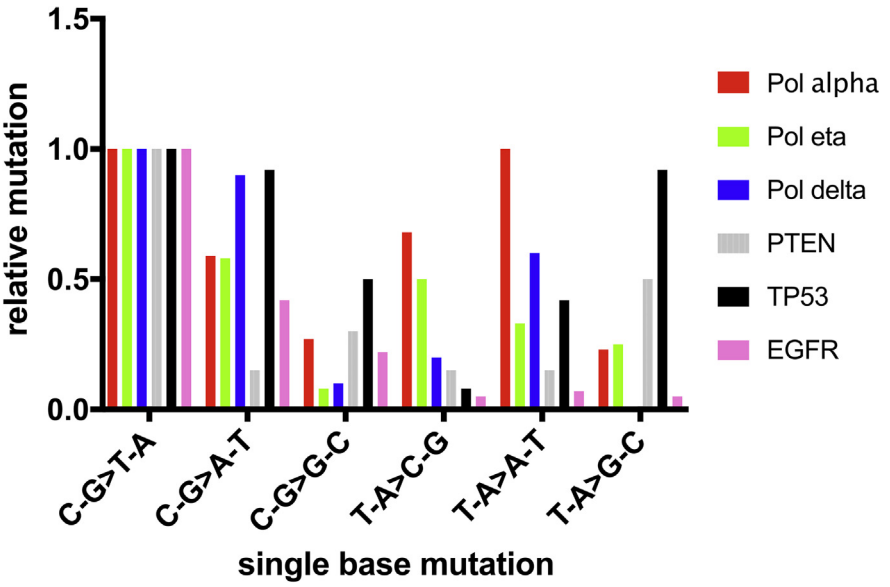

Fig. 2. Single base substitution mutation patterns produced in vitro by DNA polymerases α, ε and δ in *lacZ* [15] versus the GBM mutation patterns observed in PTEN, TP53 and EGFR [7,8].

polymerases and the GBM mutation patterns observed in the three genes are (i) the high incidence of T•A→G•C mutations in PTEN and TP53; (ii) the high level of C•G→G•C transversions in GBM; and (iii) lower than anticipated levels of T•A→A•T mutations. In terms of the latter, a transversion of T•A→A•T at Arg AGA codons would produce stop codons in tumor suppressors with the same effect as C•G→T•A transitions at CGA Arg codons. Transversion mutations at AGA are not observed in any of tumor suppressor genes analyzed, although there is no shortage of AGA codons. In fact, AGA and AGG Arg codons are also cold spots for missense mutations in all 8 genes. Because the GBM mutation profile in many of the driver genes is not random and not fully consistent with the infidelities of the replicative polymerases, it is likely that other more efficient mutagenic pathways are responsible for a significant fraction of the unaccounted cancer risk.

There is an important mutagenic role for error prone translesion polymerases that bypass DNA lesions that block replicative polymerases [42]. The recruitment of these polymerases, which leave distinct mutation patterns are dependent on the

https://doi.org/10.1016/j.heliyon.2019.e01265

nature of the blocking lesion. Depending on the blocking lesions, translesion synthesis (TLS) performed by Pol η, κ, ι, and ζ and Rev1 may be error-free or error-prone [43]. For abasic sites formed from depurination, TLS yields deletions and A misincorporation opposite the non-cognate lesion [44]. Accordingly, lesions on G that are converted to an abasic site will tend to show up at C•G→A•T transversions.

3.5. Quantitative comparison of DNA polymerase based mutations versus deamination

Analysis of the mutation spectrum in the IDH1 gene provides an insight into the potential role of random polymerase mutations versus other mechanisms of mutagenesis. The lifetime cancer risk for GBM is estimated to be 2.2×10^{-3} [45]. The lifetime cancer risk for an IDH1 mutant GBM is estimated to be 1.1×10^{-4}:

[(5% incidence of GBM with IDH1 mutations) × (2.2×10^{-3} lifetime risk for all GBM)]

Using the mutation rate of replicative polymerases as 7.6×10^{-10} mutations/stem cell division (including mismatch repair) [4], the statistical risk to acquire a random oncogenic C→T mutation at a specific base in a stem cell, e.g., R132 (CGT) in IDH1, is 5.1×10^{-19}/stem cell division:

[(7.6×10^{-10} mutations/stem cell division) × (4 target bases in the two IDH1 alleles/6×10^{9} bases/stem cell genome)].

Based on the estimate of 1.6×10^{8} astrocyte stem cells per brain [4], the statistical risk for an oncogenic mutation at one of the 4 target bases per brain is 8.1×10^{-11}:

[(5.1×10^{-19}) × (1.6×10^{8} stem cells per brain)].

A unique aspect of GBM is that virtually all stem cell divisions occur prior to birth so there is presumably little amplification of the mutation frequency due to a lifetime of stem cell divisions. Even if all stem cells divided once per year, and this mutation frequency of 8.1×10^{-11} is compounded over a 60-year period, statistically there would only be 4.9×10^{-9} stem cells with an IDH1 mutation in a human brain or 1 per 2×10^{8} humans. No doubt this calculated mutation frequency does not relate to cancer risk in a straightforward manner. It is also possible that the cells that give rise to GBM are not restricted to the stem cell population used in the calculation [4]. However, is seems unlikely that random errors can produce GBM's with activating IDH1 mutations, even if this was the sole mutation required, which is highly unlikely.

What other mechanism is consistent with the mutations observed in IDH1? As mentioned above, mutation patterns and mutation frequencies in tumor suppressors

https://doi.org/10.1016/j.heliyon.2019.e01265

and oncogenes have been attributed to hydrolytic deamination of 5-mC to yield C→T transitions at CpG sites [11]. The rate of hydrolytic deamination of 5-mC is reported to be between 4.7×10^{-4} and 1.8×10^{-5} deaminations/year per stem cell [46, 47]. Based on this rate, there would be 1.9×10^{-3} to 7.2×10^{-5} T•G mismatches/year or 1.1×10^{-1} to 4.3×10^{-3} T•G mismatches by age 60 distributed at the four 5-mC's at R132 CpG sites in both IDH1 alleles:

$(1.8 \times 10^{-5}$ deaminations at 5-mC/year) $\times$ 4(5-mC at R132 in both alleles of IDH1) $\times$ (60 years)

Due to efficient repair of T•G mismatches, it is conservatively estimated that by age 60 no more than 1% of the mismatches (1.1×10^{-3} to 4.3×10^{-5}) would have been fixed into C→T activating mutations at R132 in a brain stem cell. Modest changes in the rate of hydrolytic deamination of 5-mC in different sequences or in regions that are being actively transcribed have been ignored in this analysis, as have potential sequence-dependent repair rates. Regardless, hydrolytic deamination is orders of magnitude more efficient than polymerase errors in generating the mutations observed in IDH1.

3.6. Relationship between DNA replication and mutation rates

To reconcile the strong relationship between the number of stem cells and stem cell divisions with relative GBM cancer risk [4], there should be a correlation between deamination and stem cell division. Deamination rates are significantly different between single-stranded and double-stranded DNA due to increased exposure of the bases to H_2O in the former [46, 47]. The >100-fold increased rate of deamination in single-stranded versus double-stranded DNA is consistent with the transient formation of single-stranded DNA generated during DNA replication and transcription, being an important factor in oncogenesis.

3.7. Complementary relationship in GBM between oncogenes and tumor suppressor gene mutations

Another interesting feature of the IDH1 mutant GBM, which has been previously noted [48], is that 100% of the GBM tumors also harbor mutant TP53, while TP53 mutations occur in <24% of the remaining GBM tumors. The combination of mutant TP53 and IDH1 appears sufficient to initiate transformation since 3 of the 13 tumors with this pair of mutations share no other known oncogenic mutations. In 8 of the remaining 10 tumors, truncating mutations occur in the Alpha Thalassemia/Mental Retardation Syndrome X-Linked (ATRX) gene that is involved in chromatin remodeling, genetic stability and maintenance of telomeres in the absence of telomerase activity [49, 50]. Other genes in the Receptor Tyrosine

https://doi.org/10.1016/j.heliyon.2019.e01265

Kinase/Ras/MAP kinase signaling pathway (e.g., EGFR, PIK3CA and PIK3R) are mutated in 6 tumors.

In the RB1 mutant cohort with wild type IDH1, TP53 is mutated in 19 (70.4%) GBM tumors, and in 10 out of 15 tumors with nonsense mutations in RB1. In contrast, RB1 mutants are not observed in any of the GBM tumors with mutant TP53 that also have IDH1 mutations suggesting that mutant IDH1 somehow compensates for mutant RB1. However, RB1 expression is significantly higher in tumors with mutant IDH1 than in tumors with RB1 mutations ($P = 0.002$) or in normal tissue ($P = 0.03$) (Fig. 3a). RB1 methylation is similar in the IDH1 and RB1 mutants GBM (Fig. 3b). The concurrence of TP53 and RB1 mutations has been previously noted and implies that both tumor suppression pathways are important in GBM not bearing an IDH1 mutation [51]. How the loss of TP53 observed in GBM is related to IDH1 mutations is unclear, but it is possible that mutated IDH1 may be responsible for the previously observed hypermethylation of the RB1 promoter and reduced RB1 expression [52]. The effect of the IDH1 mutation on DNA methylation is also consistent with the clinical observation that the DNA methylase inhibitor 5-azacytidine has a therapeutic effect on GBM with IDH1 mutations [53].

3.8. Mutations and age at diagnosis

The sequence of mutagenic events leading to GBM is not known since the disease normally presents after the tumor is advanced preventing the genetic analysis of preneoplastic or early stage cancer. To probe this issue, we analyzed the age at diagnosis of GBM patients who harbored the mutations discussed above, although it is appreciated that the relationship between age at diagnosis and the age when the initial mutagenic event occurred are not directly related. It also should be pointed out that many GBM harbor mutations in more than one of the 8 genes analyzed. Consistent with previous analyses [54], the mean age of diagnosis in the cohort is virtually the same (62.0 ± 1.0 y) for all mutations, with 1 exception (Table 5). The exception is GBM cancers with an IDH1 missense mutation that have a mean age at diagnosis of 39.6 ± 15.7 y.

Analysis of the PTEN mutations in GBM tumors shows no significant difference in the age at diagnosis for males versus females: 66.3 ± 9.7 and 59.9 ± 13.3, respectively (Table S1). There is also no difference in the age at diagnosis for tumors with PTEN nonsense versus missense mutations. However, all 3 tumors in patients diagnosed at $\leq$40 y harbored C•G→T•A PTEN mutations. The same is true for all 6 tumors in patients diagnosed at the age of $\leq$50 y. At >50 y at diagnosis, C•G→T•A transitions account for 52.7% of all of the other 5 potential missense and nonsense mutations.

https://doi.org/10.1016/j.heliyon.2019.e01265

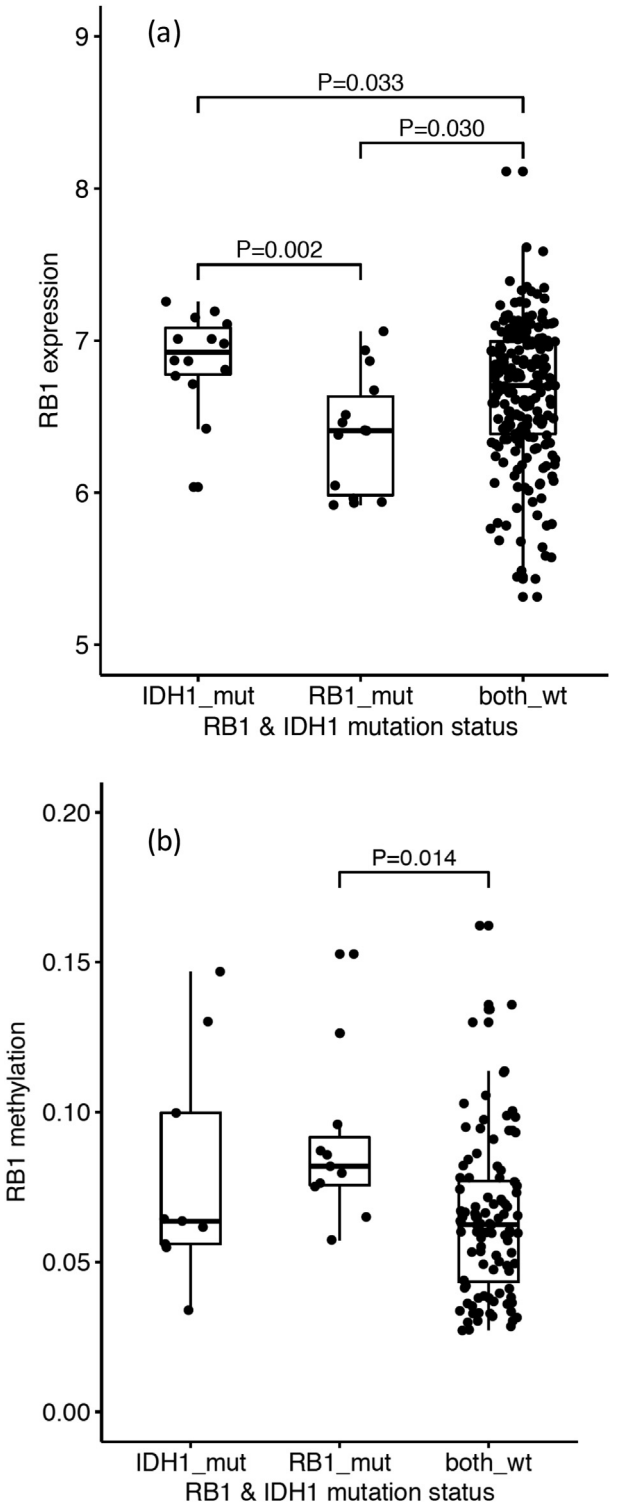

Fig. 3. Relationship between RB1 expression and gene methylation in GBM with wild type versus mutant IDH1.

https://doi.org/10.1016/j.heliyon.2019.e01265

Table 5. Relationship between mean age at diagnosis and mutations in GBM cohort [7,8].

	PTEN	TP53	EGFR	PIK3R1	PIK3CA	NF1	RB1	IDH1
Nonsense mutations	62.7 ± 13.8 (n = 14)	70.3 ± 5.5 (n = 3)	None observed	82 (n = 1)	72 (n = 1)	61.5 ± 9.5 (n = 12)	64.0 ± 15.6 (n = 12)	None observed
Missense mutations	62.5 ± 12.6 (n = 45)	60.1 ± 15.1 (n = 76)	60.6 ± 11.6 (n = 97)	61.2 ± 15.0 (n = 13)	61.7 ± 14.6 (n = 22)	66.8 ± 11.9 (n = 4)	59.3 ± 4.0 (n = 3)	39.6 ± 15.7 (n = 13)
Combined	62.5 ± 14.2 (n = 59)	60.5 ± 14.9 (n = 79)	60.6 ± 11.6 (n = 97)	62.6 ± 15.5 (n = 14)	62.1 ± 14.4 (n = 23)	62.5 ± 10.3 (n = 16)	63.1 ± 14.1 (n = 15)	39.6 ± 15.7 (n = 13)

As observed in the PTEN mutants, there is no difference in age at diagnosis between males and females with mutated TP53 (Table S2). There is no bias for C•G→T•A mutations occurring in patients diagnosed at different ages: ≤30 y (40.0%); ≤40 y (66.7%); ≤50 y (60.0%); ≤60 y (66.7%); >60 y (44.2%). The incidence of C•G→T•A mutations in EGFR as a function of age is as follows: ≤ 30 y (100%); ≤40 y (75.0%); ≤50 y (71.4%); ≤60y (45.5%); >60 y (44.6%) (Table S3). The mean age at diagnosis is 60.6 ± 11.8 y. In GBM with PIK3R1 mutations, there are only 6 cases diagnosed prior to age 60 and there is only 1 C•G→T•A transversion in this subset (Table S4). This is consistent with there being no CpG sites within the SH2 and iSH2 domains in the protein that are the target for oncogenic mutations. There are only 3 out of a total of 24 samples with PIK3CA mutations where the age at diagnosis for GBM is below 50 y (Table S5). Despite this, the mean age at diagnosis was 62.1, which is the same as for GBM with the other mutated genes. Only 1 out of the 16 tumors with NF1 mutations was there a patient diagnosed prior to age 50 (Table S6). The RB1 mutations in the cohort were predominantly (86.7%) in patients diagnosed over the age of 50 y and the few earlier diagnosed tumors did not have C•G→T•A mutations (Table S7). Overall, there are no significant differences between the different mutation types in the 7 genes discussed above (Tables S1−S7) and the age at diagnosis. Whether there is any difference in the type of mutation that occurs in the earliest stage of GBM development remains unknown.

The tumors with mutated IDH1 are diagnosed at a much earlier age than in GBM bearing the other mutations (Table S8). Four out of the 5 of the tumors diagnosed at 30 y or less bearing an IDH1 mutation have C•G→T•A transitions, with the remaining tumor having a C•G→G•C transversion. This transversion is the only non C•G→T•A mutation in the IDH1 cohort.

3.9. Potential approaches to limit GBM driver mutations

If the bulk of mutations are derived from chemical reactions (e.g., deamination, depurination, oxidation), or to mismatch insertions produced by DNA polymerases, the formation of base pair mismatches is unavoidable. However, the efficiency for the

https://doi.org/10.1016/j.heliyon.2019.e01265

conversion of a mismatch into a mutation will be affected by whether the initial mismatch is removed via repair or present during a round of DNA replication. The G•T mismatch, which is generated when 5-mC is deaminated, is a substrate for glycosylases, such as thymine-DNA glycosylase (TDG) that excises thymine in mismatches. After the initial step of cleaving the glycoside bond, the TDG enzyme remains bound to the DNA at the site of the excision. Release of the enzyme and replacement by AP endonuclease-1 (APEX1) and continuation of successful base excision repair, has been reported to be facilitated by sumoylation of TDG, which allows the glycosylase to escape and repair other mismatches [55, 56]. There is also a report that sumoylation is not necessary for TDG turnover [57]. Accordingly, molecules that facilitate the recycling of TDG by enhancing sumoylation, or by induction of some other structural allosteric change that would accelerate dissociation of the protein from the excision site, could improve the efficiency of G•T mismatch repair. The same is true for enhancing repair of G•dU mismatches that arise from deamination of C.

4. Conclusions

The mutation patterns in the GBM tumor suppressors and oncogenes are not random, and in many cases random mutagenesis cannot account for the sequence restrictive mutations in GBM oncogenes and tumor suppressor genes. Therefore, other more sequence-dependent and/or efficient mutagenic pathways such as deamination [31] and depurination [58,59] are likely to play a significant role in mutagenesis. Deamination at CGN Arg codons, which can yield 5 different outcomes (Fig. 4), can explain a significant fraction of the missense and nonsense mutations in many GBM driver genes. If deamination can account for the generation of the highly improbable mutations in IDH1 (4 base target in 6×10^9 bases), it surely must be important in the oncogenic mutations in other genes, assuming that the critical DNA sequences contain a CGN codon.

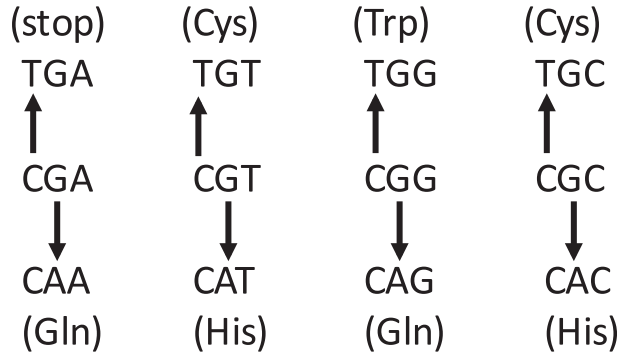

Fig. 4. Potential C•G→T•A deamination derived mutations at 5-mCpGpN codons (only coding strand shown).

https://doi.org/10.1016/j.heliyon.2019.e01265

Declarations

Author contribution statement

Min Zhang, Da Yang: Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Barry Gold: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interest statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2019.e01265.

References

[1] L.B. Alexandrov, S. Nik-Zainai, D.C. Wedge, et al., Signatures of mutation processes in human cancer, Nature 500 (2013) 415−421.

[2] The Cancer Genome Atlas Research Network, Comprehensive genomic characterization defines human glioblastoma genes and core pathways, Nature 455 (2008) 1061−1068.

[3] L.B. Alexandrov, M.R. Stratton, Mutational signatures: the patterns of somatic mutations hidden in cancer genomes, Curr. Opin. Genet. Dev. 24 (2014) 52−60.

[4] C. Tomasetti, B. Vogelstein, Variation in cancer risk among tissues can be explained by the number of stem cell divisions, Science 347 (2015) 78−81.

[5] A. Tubbs, A. Nussenzweig, Endogenous DNA damage as a source of genomic instability in cancer, Cell 168 (2017) 644−656.

[6] M. Petljak, L.B. Alexandrov, Understanding mutagenesis through delineation of mutational signatures in human cancer, Carcinogenesis 37 (2016) 531−540.

[7] J. Gao, B.A. Aksoy, U. Dogrusoz, G. Dresdner, B. Gross, S.O. Sumer, Y. Sun, A. Jacobsen, R. Sinha, E. Larsson, E. Cerami, C. Sander, N. Schultz, Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal, Sci. Signal. 6 (2013) pl1.

[8] E. Cerami, J. Gao, U. Dogrusoz, B.E. Gross, S.O. Sumer, B.A. Aksoy, A. Jacobsen, C.J. Byrne, M.L. Heuer, E. Larsson, Y. Antipin, B. Reva, A.P. Goldberg, C. Sander, N. Schultz, The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data, Cancer Discov. 2 (2012) 401−404.

[9] M.S. Lawrence, P. Stojanov, C.H. Mermel, J.T. Robinson, L.A. Garraway, T.R. Golub, M. Meyerson, S.B. Gabriel, E.S. Lander, G. Getz, Discovery and saturation analysis of cancer genes across 21 tumour types, Nature 505 (2014) 495−501.

[10] W.M. Rideout 3rd, G.A. Coetzee, A.F. Olumi, P.A. Jones, 5-Methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes, Science 249 (1990) 1288−1290.

[11] A.N. Magewu, P.A. Jones, Ubiquitous and tenacious methylation of the CpG site in codon 248 of the p53 gene may explain its frequent appearance as a mutational hot spot in human cancer, Mol. Cell Biol. 14 (1994) 4225−4232.

[12] S.J. Clark, J. Harrison, M. Frommer, CpNpG methylation in mammalian cells, Nat. Genet. 10 (1995) 20−27.

[13] D.N. Cooper, M. Mort, P.D. Stenson, E.V. Ball, N.A. Chuzhanova, Methylation-mediated deamination of 5-methylcytosine appears to give rise to mutations causing human inherited disease in CpNpG trinucleotides, as well as in CpG dinucleotides, Hum. Genomics 4 (2010) 406−410.

[14] B. Gold, Somatic mutations in cancer: stochastic versus predictable, Mutat. Res. 814 (2017) 37−46.

[15] D.C. Thomas, J.D. Roberts, R.D. Sabatino, T.W. Myers, C.K. Tan, K.M. Downey, A.G. So, R.A. Bambara, T.A. Kunkel, Fidelity of mammalian DNA replication and replicative DNA polymerases, Biochemistry 30 (1991) 11751−11759.

[16] M.W. Schmitt, Y. Matsumoto, L.A. Loeb, High fidelity and lesion bypass capability of human DNA polymerase δ, Biochimie 91 (2009) 1163−1172.

[17] D.A. Korona, K.G. Lecompte, Z.F. Pursell, The high fidelity and unique error signature of human DNA polymerase epsilon, Nucleic Acids Res. 39 (2011) 1763−1773.

https://doi.org/10.1016/j.heliyon.2019.e01265

[18] M. Zhang, Y. Yang, B. Gold, The Adenomatous Polyposis Coli (APC) mutation spectra in different anatomical regions of the large intestine in colorectal cancer, Mutat. Res. 810 (2018) 1−5.

[19] J.C. Lee, I. Vivanco, R. Beroukhim, J.H. Huang, W.L. Feng, R.M. DeBiasi, K. Yoshimoto, J.C. King, P. Nghiemphu, Y. Yuza, Q. Xu, H. Greulich, R.K. Thomas, J.G. Paez, T.C. Peck, D.J. Linhart, K.A. Glatt, G. Getz, R. Onofrio, L. Ziaugra, R.L. Levine, S. Gabriel, T. Kawaguchi, K. O'Neill, H. Khan, L.M. Liau, S.F. Nelson, P.N. Rao, P. Mischel, R.O. Pieper, T. Cloughesy, D.J. Leahy, W.R. Sellers, C.L. Sawyers, M. Meyerson, I.K. Mellinghoff, Epidermal growth factor receptor activation in glioblastoma through novel missense mutations in the extracellular domain, PLoS Med. 3 (2006) e485.

[20] I. Vivanco, H.I. Robins, D. Rohle, C. Campos, C. Grommes, P.L. Nghiemphu, S. Kubek, B. Oldrini, M.G. Chheda, N. Yannuzzi, H. Tao, S. Zhu, A. Iwanami, D. Kuga, J. Dang, A. Pedraza, C.W. Brennan, A. Heguy, L.M. Liau, F. Lieberman, W.K. Yung, M.R. Gilbert, D.A. Reardon, J. Drappatz, P.Y. Wen, K.R. Lamborn, S.M. Chang, M. Prados, H.A. Fine, A. Horvath, N. Wu, A.B. Lassman, L.M. DeAngelis, W.H. Yong, J.G. Kuhn, P.S. Mischel, M.P. Mehta, T.F. Cloughesy, I.K. Mellinghoff, Differential sensitivity of glioma- versus lung cancer-specific EGFR mutations to EGFR kinase inhibitors, Cancer Discov. 2 (2012) 458−471.

[21] M. Sibilia, E.F. Wagner, Strain-dependent epithelial defects in mice lacking the EGF receptor, Science 269 (1995) 234−238.

[22] D.W. Threadgill, A.A. Dlugosz, L.A. Hansen, T. Tennenbaum, U. Lichti, D. Yee, C. LaMantia, T. Mourton, K. Herrup, R.C. Harris, et al., Targeted disruption of mouse EGF receptor: effect of genetic background on mutant phenotype, Science 269 (1995) 230−234.

[23] C.H. Huang, D. Mandelker, O. Schmidt-Kittler, Y. Samuels, V.E. Velculescu, K.W. Kinzler, B. Vogelstein, S.B. Gabelli, L.M. Amzel, The structure of a human p110alpha/p85alpha complex elucidates the effects of oncogenic PI3Kalpha mutations, Science 318 (2007) 1744−1748.

[24] M.S. Waitkus, B.H. Diplas, H. Yan, Isocitrate dehydrogenase mutations in gliomas, Neuro Oncol. 18 (2016) 16−26.

[25] W. Xu, H. Yang, Y. Liu, Y. Yang, P. Wang, S.H. Kim, S. Ito, C. Yang, P. Wang, M.T. Xiao, L.X. Liu, W.Q. Jiang, J. Liu, J.Y. Zhang, B. Wang, S. Frye, Y. Zhang, Y. Xu, Q.Y. Lei, K.L. Guan, S.M. Zhao, Y. Xiong, Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α- ketoglutarate-dependent dioxygenases, Cancer Cell 19 (2011) 17−30.

 https://doi.org/10.1016/j.heliyon.2019.e01265

[26] S.A. Roberts, M.S. Lawrence, L.J. Klimczak, S.A. Grimm, D. Fargo, P. Stojanov, A. Kiezun, G.V. Kryukov, S.L. Carter, G. Saksena, S. Harris, R.R. Shah, M.A. Resnick, G. Getz, D.A. Gordenin, An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers, Nat. Genet. 45 (2013) 970−976.

[27] K. Chan, S.A. Roberts, L.J. Klimczak, J.F. Sterling, N. Saini, E.P. Malc, J. Kim, D.J. Kwiatkowski, D.C. Fargo, P.A. Mieczkowski, G. Getz, D.A. Gordenin, An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers, Nat. Genet. 47 (2−15) (2015) 1067−1072.

[28] S.K. Dickerson, E. Market, E. Besmer, F.N. Papavasiliou, AID mediates hypermutation by deaminating single stranded DNA, J. Exp. Med. 197 (2003) 1291−1296.

[29] R.C. Beale, S.K. Petersen-Mahrt, I.N. Watt, R.S. Harris, C. Rada, M.S. Neuberger, Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo, J. Mol. Biol. 337 (2004) 585−596.

[30] J.D. Salter, R.P. Bennett, H.C. Smith, The APOBEC protein family: united by structure, divergent in function, Trends Biochem. Sci. 41 (2016) 578−594.

[31] T. Lindahl, Instability and decay of the primary structure of DNA, Nature 362 (1993) 709−715.

[32] R. Rinne De Bont, N. van Larebeke, Endogenous DNA damage in humans: a review of quantitative data, Mutagenesis 19 (2004) 169−185.

[33] S. Kawanishi, Y. Hiraku, S. Oikawa, Mechanism of guanine-specific DNA damage by oxidative stress and its role in carcinogenesis and aging, Mutat. Res. 488 (2001) 65−76.

[34] Y. Margolin, V. Shafirovich, N.E. Geacintov, M.S. DeMott, P.C. Dedon, DNA sequence context as a determinant of the quantity and chemistry of guanine oxidation produced by hydroxyl radicals and one-electron oxidants, J. Biol. Chem. 283 (2008) 35569−35578.

[35] H. Rodriguez, M.R. Valentine, G.P. Holmquist, S.A. Akman, J. Termini, Mapping of peroxyl radical induced damage on genomic DNA, Biochemistry 38 (1999) 16578−16588.

[36] H. Rodriguez, R. Drouin, G.P. Holmquist, T.R. O'Connor, S. Boiteux, J. Laval, J.H. Doroshow, S.A. Akman, Mapping of copper/hydrogen peroxide-induced DNA damage at nucleotide resolution in human genomic

https://doi.org/10.1016/j.heliyon.2019.e01265

DNA by ligation-mediated polymerase chain reaction, J. Biol. Chem. 270 (1995) 17633−17640.

[37] J.A. Hartley, Mechanisms of DNA sequence selective modifications by alkylating agents, in: B. Pullman, J. Jortner (Eds.), Molecular Basis of Specificity in Nucleic Acid-Drug Interactions, The Jerusalem Symposia on Quantum Chemistry and Biochemistry, vol. 23, Springer, Dordrecht, 1990.

[38] Y. Hashimoto, K. Shudo, Sequence selective modification of DNA with mutacarcinogenic 2-amino-6-methyldipyrido[1,2-a:3′,2′-d]imidazole, Biochem. Biophys. Res. Commun. 116 (1983) 1100−1106.

[39] A. Dipple, C.A. Bigger, Mechanism of action of food-associated polycyclic aromatic hydrocarbon carcinogens, Mutat. Res. 259 (1991) 263−276.

[40] G.P. Pfeifer, P. Hainaut, On the origin of G–> T transversions in lung cancer, Mutat. Res. 526 (2003) 39−43.

[41] M.F. Denissenko, A. Pao, M. Tang, G.P. Pfeifer, Preferential formation of benzo[a]pyrene adducts at lung cancer mutational hotspots in P53, Science 274 (1996) 430−432.

[42] M.F. Goodman, Error-prone repair DNA polymerases in prokaryotes and eukaryotes, Annu. Rev. Biochem. 71 (2002) 17−50.

[43] T. Yagi, Y. Fujikawa, T. Sawai, T. Takamura-Enya, S. Ito-Harashima, M. Kawanishi, Error- prone and error-free translesion DNA synthesis over site-specifically created DNA adducts of aryl hydrocarbons (3-nitrobenzanthrone and 4-aminobiphenyl), Toxicol. Res. 33 (2017) 265−272.

[44] N. Sabouri, E. Johansson, Translesion synthesis of abasic sites by yeast DNA polymerase ε, J. Biol. Chem. 284 (2009) 31555−31563.

[45] Key Statistics for Colorectal Cancer in Cancer Facts & Figures, American Cancer Society, Atlanta, GA, 2017. http://www.cancer.org/cancer/colonandrectumcancer/detailedguide/colorectal-cancer-key-statistics (Accessed 21 June 2018).

[46] J.-C. Shen, W.M. Rideout III, P.A. Jones, The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA, Nucleic Acids Res. 6 (1994) 972−976.

[47] L.A. Frederico, T.A. Kunkel, B.R. Shaw, A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy, Biochemistry 29 (1990) 2532−2537.

https://doi.org/10.1016/j.heliyon.2019.e01265

[48] K. Ichimura, D.M. Pearson, S. Kocialkowski, L.M. Bäcklund, R. Chan, D.T. Jones, V.P. Collins, IDH1 mutations are present in the majority of common adult gliomas but rare in primary glioblastomas, Neuro Oncol. 11 (2009) 341−347.

[49] Y. Hu, G. Shi, L. Zhang, F. Li, Y. Jiang, S. Jiang, W. Ma, Y. Zhao, Z. Songyang, J. Huang, Switch telomerase to ALT mechanism by inducing telomeric DNA damages and dysfunction of ATRX and DAXX, Sci. Rep. 6 (2016) 32280.

[50] B. Philip, D.X. Yu, M.R. Silvis, C.H. Shin, J.P. Robinson, G.L. Robinson, A.E. Welker, S.N. Angel, S.R. Tripp, J.A. Sonnen, M.W. VanBrocklin, R.J. Gibbons, R.E. Looper, H. Colman, S.L. Holmen, Mutant IDH1 promotes glioma formation in vivo, Cell Rep. 23 (2018) 1553−1564.

[51] C.H. Coschi, A.L. Martens, K. Ritchie, S.M. Francis, S. Chakrabarti, N.G. Berube, F.A. Dick, Mitotic chromosome condensation mediated by the retinoblastoma protein is tumor-suppressive, Genes Dev. 24 (2010) 1351−1363.

[52] L.E. Huang, A.L. Cohen, H. Coleman, R.L. Jensen, D.W. Fults, W.T. Couldwell, IGFBP2 expression predicts IDH-mutant glioma patient survival, Oncotarget 8 (2017) 191−202.

[53] A. Borodovsky, V. Salmasi, S. Turcan, A.W. Fabius, G.S. Baia, C.G. Eberhart, J.D. Weingart, G.L. Gallia, S.B. Baylin, T.A. Chan, G.J. Riggins, 5-Azacytidine reduces methylation, promotes differentiation and induces tumor regression in a patient- derived IDH1 mutant glioma xenograft, Oncotarget 4 (2013) 1737−1747.

[54] H. Ohgaki, P. Dessen, B. Jourde, S. Horstmann, T. Nishikawa, P.L. Di Patre, C. Burkhard, D. Schüler, N.M. Probst-Hensch, P.C. Maiorka, N. Baeza, P. Pisani, Y. Yonekawa, M.G. Yasargil, M.U.M. Lütolf, P. Kleihues, Genetic pathways to glioblastoma: a population- based study, Cancer Res. 64 (2004) 6892−6899.

[55] H.D. Ulrich, SUMO modification: wrestling with protein conformation, Current Biol. 15 (2005) R257−R259.

[56] T.R. Waters, P.F. Swann, Thymine-DNA glycosylase and G to A transition mutations at CpG sites, Mutat. Res. 462 (2000) 137−147.

[57] D. McLaughlin, C.T. Coey, W.C. Yang, A.C. Drohat, M.J. Matunis, Characterizing requirements for small ubiquitin-like modifier (sumo) modification and binding on base excision repair activity of thymine-DNA glycosylase in vivo, J. Biol. Chem. 291 (2016) 9014−9024.

https://doi.org/10.1016/j.heliyon.2019.e01265

[58] J. Nakamura, J.A. Swenberg, Endogenous apurinic/apyrimidinic sites in genomic DNA of mammalian tissues, Cancer Res. 59 (1999) 2522−2526.

[59] M. Nakamura, Y. Yonekawa, P. Kleihues, H. Ohgaki, Promoter hypermethylation of the RB1 gene in glioblastomas, Lab. Invest. 81 (2001) 77−82.

https://doi.org/10.1016/j.heliyon.2019.e01265