Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA
Published: March 25, 2026
DOI: 10.46254/AP07.20260318

The Use of Blue Whale Proteins in DNA Reparation to Prevent Cancer

Santiago Arizpe
Tecnológico De
Monterrey Monterrey
Mexico, México
santiagoarizpe0818@gmail.com

Abstract

Blue whales exhibit a resistance to cancer despite their size and lifespan, and as numerous studies have found that when the size of organisms increases, the probability of acquiring tumors decreases. This trend happens due to enhanced DNA repair mechanisms. The research ran a comparative genomic analysis on key genes responsible for the reparation of genetic material in the process of BER. Using NIH databases and benchling, sequences of UNG DNA glycosylase, PNKP, APEX1, DNA polymerases, Lig III, PARP1 and XRCC1, the proteins were aligned and compared to human counterparts to identify amino acid substitutions and assess the functional impact on the proteins. Results were able to showcase that proteins exhibit a substantial difference from those present in humans and blue whales, with pairwise identities of as 64.00% in PNKP and 97.92% DNA glycosylase, blue whale proteins which have a lower pairwise identity; For example PNKP reaching an instability index of 50.65% compared to 35.34% in humans. While proteins with similar pairwise identities showcased a higher stable form of the protein. For example, the blue whale's UNG presents an instability index of 47.16% compared to 47.19% in humans. Although the isoforms of many of these genes present in blue whales present a higher instability, a high turnover reaction rate serves as a solution to this instability. Results suggest that blue whales have evolved to have an optimized BER pathway though protein-protein interactions. This work helps to pinpoint specific protein candidates for future wet lab studies aimed at reinforcing human DNA reparation mechanisms.

Keywords
Cancer, Genomic analysis, Cetaceans, Tumor suppressor genes.

1. Introduction

Blue whales, having an average lifespan of 90 years, and the lowest cancer rate in the animal kingdom, this is because of the ability the cells have to repair the DNA better than other animals, and even though they posses an enormous amount of cells, blue whales are not prone to cancer, which is incredible because they are the largest animal to ever live (Firsanov, D. *et al*, 2025). The reason why these blue whales are not prone to acquiring cancer may be because of Peto's paradox, which argues that cancerous tumors are not related to body size and or age, arguing that even if animals have 1000 times the size of humans, they must have a better cancer suppression system (Caulin et al, 2011).

Due to this advanced suppression system it is relevant that these systems are studied for implementation on humans. Via gene editing mechanisms to improve the preparation of cells in humans. Furthermore this is backed up by the investigation of Rangarajan *et al* (2004) who consistently found that the mechanisms to repair fibroblasts in mice are 2 while the same mechanisms in humans jump up to 5. With this information, blue whale repair pathways must include more layers of protection to prevent damage to their genetic material

1.1 Objectives

This research paper aims to create a computational proof of concept on the functional impact of using cetacean inspired mutations on the human protein Lig 1 present on the Base excision repair pathways. Specifically, the aim is to identify divergent residues on catalytic domains and model how these changes affect the structure of the protein, comparing

the mutated chains with the original Human chains. In order to reduce cancer risk by enhancing cell reparation modifying the to enhance the possibility to better the lifestyle of affected people, potentially adding to their lifespans, to achieve a whale like reparation system.

2. Literacy review

2.1 Peto's paradox

Peto's paradox can be thoroughly explained when comparing the cancer rate across mammals of greatly different sizes, because even though organisms with a bigger size should have a higher chance of having cancer, biological trends work quite oppositely. In a study developed and published by Zhou *et al* (2023) mice at 18 months of age have an 80% probability to acquire some type of cancerous tumor and a mortality rate reaching as high as 95% (Gorbunova, V,2014). Dogs have the present threat of devolving cancer in their lifetimes with an incidence in the early stages of their life of at least 25% and increasing to a 50% probability of having cancer when they reach an age of 10 or greater (American Veterinary Medical Association, n.d), and a mortality rate of 27%. Humans normally have a 20% risk of acquiring cancer in their lifetimes, with a rapidly growing risk of acquiring it in their lifetimes (world health organization, 2024). And a mortality rate of 17%. (world health organization, 2025) While elephants cancer acquisition rate is significantly lower, different sources suggest a different percentage, varying between 3% acquisition rate (Phang Nga Elephant Park, 2018) and a 5% acquisition rate (Reyes Castro *et al*, 2023), on the other hand mortality rate due to cancer in elephants is highly investigated, which according to Abegglen *et al* (2015) its at a 4.81% meaning that no matter if they get cancer elephants, have a low probability to die from it. Second to last, bowhead whales have a cancer rate of 3.3% and going as high as 4.7% (Vazquez, 2022); there is no data on the mortality rate caused by cancerous tumors of these whales. Lastly, blue whales have a near zero probability of getting cancer and a lower risk of dying from it. This relationship between the susceptibility of cancer and body size is displayed in graph 1. Which excludes the blue whale's data due to lack of information on the cancer rate of blue whales.

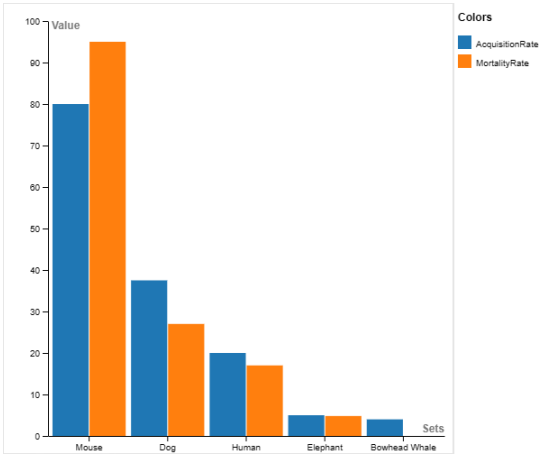

Figure 1. Comparative graph that shows relation between mortality rate, incident rate and species ordered by size

It is important to notice that elephants in contrast to blue whales, rely on a cell death strategy, having up to 20 copies of the TP53 tumor suppressor gene, while whales, appear to have a more efficient reparation system, focusing on maintaining genomic integrity rather than eliminating cells, on the study the research focused on more efficient reparation, rather than on the expression of the TP53 protein. This is because as seen in Figure 1, elephants may get cancer while whales don't.

2.2 Proteins responsible

Cancer being a problem in the genetic material raises the question of what can we do to prevent the breaking down and deterioration of normal cells. Organisms as previously stated have inner pathways meant to repair this damage. The pathway that will be analyzed throughout this paper is the following:

Base excision repair (BER) Base excision repair (Figure 2), this works by correcting lesions on the genetic code, these lesions include Deamination, Oxidation of the DNA molecules or methylation. It focuses mainly on small repairs that don't change the structure of the DNA (Krokan, H and Bjoras, M., 2013). BER uses mainly 10 proteins which are DNA glycosylase, polynucleotide kinase-phosphatase (PNKP), apurinic/apyrimidinic endonucle- ase 1 (APE1), DNA polymerases (POLB), ligase III (LigIII), ligase I (LigI),flap structure-specific endonuclease 1 (FEN1) , poly (ADP-ribose) polymerases, PARP1 and PARP2 (PARP), and X-ray repair cross-complementing protein 1 (XRCC1).

(Rathnaiah, G, 2025) In the case of the DNA polymerases, they play a fundamental role in the initiation of the BER, making cleaving the glucosidic bonds in the DNA and leaving apurinic/apyrimidinic (AP) sites (Krokan *et al*, 1997), these sites, are harmful because they may result in breakage of the DNA strands, the de-naturalization of it or the sudden stop of DNA replication (Amidon, K. *et al*, 1997). In this part APE1 becomes essential, This is due to the protein trimming the aforementioned AP sites from the DNA, allowing for the reparation of these mutated nucleotides, after the APE1 finishes trimming the AP sites. Either of two processes are carried out, the first one being short patch repair which mainly uses the proteins Lig III, Lig 1 and XRCC1, it consists of filling a single nucleotide gap, which is ligated, to the DNA (Krokan, H. *et al* 2013), and the second one being long patch repair, which mainly uses Lig 1 and FEN1, consists on the repair of a strand of DNA, typically being at a length of approximately 2-10 amino acids (Krokan, H. *et al* 2013).

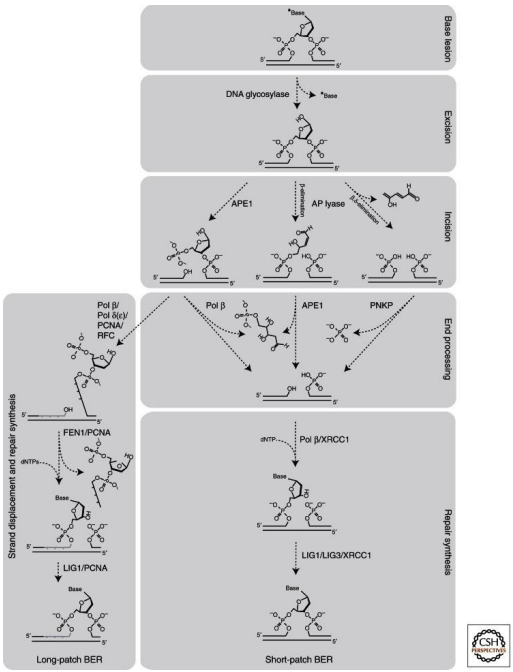

Figure 2. Figure that shows the BER Pathway. Adapted from Krokan 2013)

2.3 Gene editing techniques

Crispr Cas-9 is a gene editing technology that according to Redman, M (2016) makes it possible to correct errors in the genome, allowing for modification of the genome in the search of a better efficiency. Crispr works by using a guide RNA to identify the sequence and match it to a target gene to which the modified genetic code, uses the enzyme CAS-9 to cut the original genetic material in a set sequence, this allows for the foreign sequence to bind to the host DNA (Redman, M. 2016)

3. Methods

This work uses genetic modeling as a theoretical approach on how genetic modifications could change the way the cells in humans repair the genetic material.

3.1 Sequence selection and filtering

To find candidate proteins for modification, the blue whale orthologs were aligned to the corresponding human sequences, for instance, Human LigI isoform was aligned with blue whale LigI isoform 1 obtaining relevant information on their active sites and domains. By comparing the two matching protein sequences, a pairwise identity can be created, this identity is defined by formula 1

$$Identity = N_{matches}/L_{total} \quad (1)$$

Where N is the number of matches in both DNA sequences and L is the total length of the sequence that is being compared.

3.2 Sequence alignment and conservational analysis

Comparison was done in a computer based analysis of the protein sequences responsible for the BER pathway, which are LigIII, LigI, FEN1, PNKP, XECC1, APE1, DNA glycosylase, DNA polymerase and PARP1. (NCBI, 2025) The aforementioned proteins were retrieved from the databases of the National Institute Of Health (NIH), from the National Center for Biotechnology Information (NCBI). These databases offer a detailed description of proteins and their sequences, having annotations (NCBI, 2025) The sequences, were aligned using benchling, which is a tool that allows sequence analysis and alignment, allowing for bulk amino acid sequence alignment, also providing the pairwise identity of the selected amino acids. (Terron-Macias, V. *et al*, 2025), making the research progress much faster.

3.3 Sequence divergence analysis

PNKP and LIG 1 were selected for further analysis, the proteins were analyzed using the following bio-informatics algorithms:

- **Sorting Intolerant from Tolerant (SIFT):** which predicts whether an amino acid changes protein function based on the properties of the amino acids. Its prediction is based on whether or not the amino acid is conserved across multiple homologous or not, generating a score ranging from 0 to 1 with less than 0.05 being considered deleterious and greater or equal to 0.05 being considered benign or neutral.(Kumar P. et al, 2009; Ngak-Leng S. 2012)
- **BLOSUM64:** Is an analytical tool used to predict. whether an aminoacid tool is conservative or non conservative, utilizing matrixes for aminoacid sequence comparison (Anna Postovskaya *et al*)
- **UniProt:** UniProt is an algorithm that uses the databases of the UniProt Knowledge base (UniProtKB), it is used as a way to analyze proteins in order to obtain the annotations of said proteins and relevant data about them. (UniProt, 2025) In the research UniProt was used as a way to determine the active sites of the proteins.
- **InterPro:** Interpro uses the UniProtKB database along with the Protein Data banks in Europe, Gene Ontology and AlphaFold protein structure database. It uses these databases in order to predict the presence of domains, significant sites and classify proteins in their respective families. In the research, InterPro was used in order to determine where the active sites and domains are, in order to modify them later.

To determine if the analysis by alpha missense was correct, the data was compared to the sift data, achieving a higher validity index. The protein chains were particularly analyzed on chemically active and coding sites, in order to determine if these specific sequences could affect the reparation mechanisms inside cells, due to changes in the protein conformation or folding.

3.4 Mutational analysis

Having found set chemically active and coding sites a targeted substitution change on the human LigI was performed, this substitution was decided according to the following criteria:

- **Functional Impact:** which was determined by whether or not, the amino acid regions were on functional sites, like active sites or binding sites
- **Significance:** The chains were evaluated using the Uniprot and InterPro data, to determine their position, being either on a α-helix or β-sheet. Changes were determined as the ones that were considered deleterious and the amino acids around that area that were considered neutral or tolerated

3.5 Protein Modeling

To be able to compare the protein structure of the modified Human LigI protein with the original proteins, they were modeled using the algorithm Alpha Fold 3, was utilized to predict the structure of the proteins, this works because Alpha Fold 3, uses various benchmarks to determine the accuracy of the protein sequences, with a Predicted Local Distance Difference Test (pLDDT), which is a prediction of how the protein acts and looks on a real world environment, the score goes from 0-100, with scores higher than 70 exhibit a higher confidence and scores exceeding 90, exhibit a higher accuracy to proteins in real environments. It is also important to emphasize that a lower score may be given if the protein doesn't have a set structure, for instance in the linker chains. (Embl-Ebi. n.d.; Abramson, J. 2024)

3.6 Protein Model Comparison

To be able to know whether or not the modified protein sequences deviated a lot from the original sequences, the sequences modeled by AlphaFold, were analyzed using ChimeraX and PyMOL, in order to know the Root Mean Square Deviation (RMSD). The values $> 2Å$ represent a high conservation meanwhile values higher than that, represent a significant structural shift(Meng, *et al*, 2023; Rosignoli *et al*, 2022; Kufareva, *et al* 2012)

4. Data Collection

To ensure the reliability of the data, all sequences were reviewed thoroughly, to achieve this, only reviewed sequences from the Swiss-Prot databases were utilized for the analysis of the sequences, this was in order to obtain functional annotations and ensure that they were backed up by real life data. Also the isoforms of each sequence were verified as orthologs with BLASTp software. The ID belonging to all of these databases are listed in Table 1.

Table 1. UniProt and Ensembl IDs of human Base Excision Repair (BER) proteins

Protein	UniProt ID	Ensembl ID	Date Retrieved
PNKP	Q96T60	ENSG00000039650	2026-02-18
FEN1	P39748	ENSG00000168496	2026-03-03
LIG1	P49916	ENSG00000105486	2026-02-20
APE1	P27695	ENSG00000100823	2026-01-29
UNG	P13051	ENSG00000076248	2026-01-30
XRCC1	P18887	ENSG00000073050	2026-02-16
POLB	P06746	ENSG00000070501	2026-02-15
PARP1	P09874	ENSG00000143799	2026-02-16
LIG3	P49916	ENSG00000005156	2026-02-15

Note: All proteins are from *Homo sapiens* and involved in the Base Excision Repair (BER) pathway.

5. Results and Discussion

5.1 Numerical Results

5.1.1 Isoforms

To evaluate the differences between genomes and be able to determine candidates for gene editing, a comparative analysis was run on both protein sequences. Results are shown in Table 2. Identity varied across proteins, but notably, the proteins that are useful on both long and short patch repair displayed larger differences in pairwise identity than those whose primary function is in strand separation.

Table 2. Percentage identity of DNA repair protein isoforms.

Protein	**Isoform Identity(%)**
DNA glycosylase	Isoform :97.95%
FEN1	Isoform : 97.63%
APE1	Isoform 1: 83.66%
	Isoform 2: 94.08%
	Isoform 3: 94.08%
PARP1	Isoform: 94.08%
POLB	Isoform 1: 88.92%
	Isoform 2: 70.54%
PNKP	Isoform: 88.47%
LIG3	Isoform 1: 78.19%
	Isoform 2: 78.28%
	Isoform 3: 89.46%
	Isoform 4: 90.43%
	Isoform 5: 77.09%
	Isoform 6: 86.75%
	Isoform 7: 57.8%
XRCC1	Isoform: 84.8%
LIG1	Isoform 1: 85.25%
	Isoform 2: 79.65%
	Isoform 3: 82.05%

The percentage identity reveals that core proteins like FEN1 that are present before either long or short excision repair present a lower change on the pairwise identity. Meanwhile, the ones present in the repair, like Lig1 present lower pairwise identities, along with a higher number of isoforms.

5.1.2 Sequence Divergence

Based on the pairwise identity filtering, two proteins passed on to the analysis, these proteins being PNKP and LigI, exhibited a higher significant sequence variation between blue whales and humans. Justifying further investigation of these proteins. PNKP and Lig1 were analyzed by SIFT and BLOSUM (Table 3).

Table 3. Table that shows the number of variants classified as deleterious by SIFT score (< 0.05) and BLOSUM score (< 0), along with the total number and the ratio of deleterious and tolerated variants for each protein isoform.

Protein	**Isoform**	**Sift score (Amt. < 0.05)**	**BLOSUM (Amt. < 0)**	**Deleterious Amt.**	**Tolerated Amt.**	**Ratio (tol/del)**
PNKP	1	27	27	27	42	1.47
Lig 1	1	54	54	131	98	0.75
	2	11	43	54	68	1.26
	3	55	55	55	63	1.14

The sequences PNKP and Lig 1 were selected for sequence analysis based on their identity divergence. PNKP was selected as it exhibited the lowest Pairwise identity among the enzymes responsible for the priming part of the BER pathway. Lig 1 was chosen because of its high relevance on long excision repair, and has the lowest pairwise identity average of the Long excision. Long excision repair, that fixes complex genetic damage, ranging from 2 to 13 nucleotides. Lig 1 also demonstrated the highest genetic divergence from the proteins present on long excision repair, suggesting a higher divergence on protein structure and function.

5.2 Graphical Results

5.2.1 Protein Modeling

Sequences chosen for 3D protein modeling were the isoforms of Lig 1, due to the sift average of the proteins. PNKP did not enter the 3D modeling section. due to less of a difference between the amino acid sequences, because as seen in the sequence divergence section, less deleterious amino acids are present on the PNKP sequence.

Modeling was carried out using the GoogleDeepmind Alphafold 3 software, proteins were modified according to the data previously gathered, making sure that all of the changes in a range of 3 amino acids of the deleterious amino acids were changed, to prevent denaturalization of the proteins.

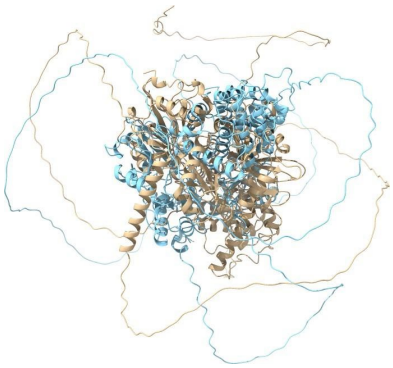

Figure 3. Figure that shows the proteins beige is the original and cyan is the modified

Figure 3, shows the superimposed LIG1 isoform 1 modified and non-modified proteins. The structure RMSD, when calculating the deviation a value of 0.354 Å is obtained, where in the protein sequences 5053 of 5053 atoms analyzed. The exceptionally low value suggests that the tertiary structure of the protein was highly conserved despite the low Sift scores, meaning that residue substitutions are compatible with human proteins. It is also important to mention that the modified protein (cyan), is more compact than the wild human protein (beige), also the linker proteins are accommodated in a more closed position.

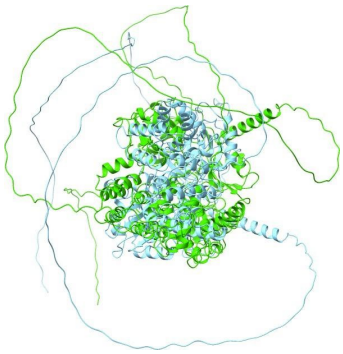

Figure 4. Figure that shows the proteins green is the original and cyan is the modified

Structural analysis for LIG1 Isoform 2 (Figure 4), revealed that the RSMD score was 0.256 Å , which is lower than the first protein, with 4924 atoms analyzed, achieving another unusually low value, taking into account the SIFT scores that were received beforehand. In the figure, it is also shown how changes really happened on the linker protein

chains, with little to no changes on the tertiary structure, however it is important to note that on the right side it appears to be an conformational shift of the same protein, this can be seen in the outermost θ helix where on the modified protein (cyan) is on the other side of the protein. Suggesting that blue whale inspired proteins, may induce to a local adjustment of the protein's outer structure

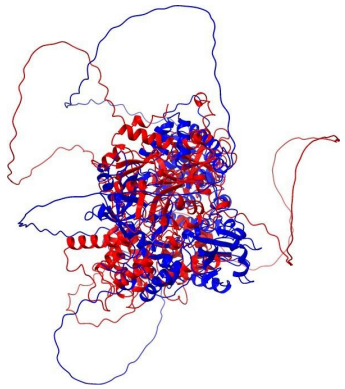

Figure 5. Figure that shows the proteins red is the original and blue is the modified

Figure 5, showcases LIG1 isoform 3 modified and non-modified proteins, where the RMSD score was 0.258 Å , which is similar to the score obtained by Isoform 2, with 4926 atoms analyzed, this makes sense, due to the ratios of the isoforms being closer, unlike isoform's 1 ratio. It is also relevant to mention that the modified structure (Blue) doesn't exhibit linker chains on the right side of the protein, instead presenting them on the left side, where the wild human isoform (Red), only presents 1 linker chain.

5.2.2 Proposed Improvements

An important improvement, that would benefit this research, is the use of real wet-lab data of the identified Lig1 and PNKP proteins. Analysis should be done in vitro, with purified enzymes as it is necessary to quantify if the cetacean inspired proteins truly affect the efficiency of DNA reparation compared with wild human counterparts. Also, the use of CRISPR-CAS9 should be implemented as it ensures that the proteins are produced by the patients cells, reducing the risk for an immunodegenerative disease. Also as a way to reduce immune response, computer based epitope predictions should be run, as a way to eliminate or at least mitigate the immune response to these cetacean inspired proteins. Lastly, further works should also implement other pathways as synergy of pathways needs to be achieved if a more efficient cancer suppression system is to be achieved

5.3 Validation

Study Limitations

This study gives a very limited scope, given that to achieve true cancer suppression, the other repair mechanisms should be studied, as \textit{Balaenoptera musculus} cancer suppression system, likely arises from the synergic interactions between all of the repair pathways present on the organism. If a full combination of the genetic sequences is to be achieved, the study should intend to reach the repair architecture of blue whales, this is in order to make DNA reparation pathways more efficient. This is to hopefully achieve a more effective cancer suppression system.

Application of the results is also limited, given that the research is theoretical, and the modifications proposed are not yet tested in a real life environment, this is because of the complexity of the human genome and the ethical implications of modifying it. However, this research serves as a proof of concept, which is a key step in the fight against cancer. Also, the research would is limited by the species selected, given that the blue whale is not the only species that has a low cancer rate, and it is possible that other species have better solutions to the problem of cancer, for instance, elephants have a higher number of copies of the TP53 gene, which is a tumor suppressor gene (Abegglen *et al*, 2015), this is a different solution to the problem of cancer than the one proposed by blue whales,

Another important limitation is that the the research faces the problem of the incompatibility of the blue whale proteins with the human proteins, given that the integration of said proteins may have detrimental consequences, with the risk of negative immune responses, the clinical application becomes more limited, however, the research works as a proof of concept, which is a key step on the fight against cancer. Also the research doesn't account for the complex inter pathway synergy, as pathways need to work with each other, analysis of the other pathways needs to be conducted

both as a computer based analysis and wet lab analysis, in order to be able to create an efficient repair pathway.

6. Conclusions

The research provides a successful in silico proof of concept for using cetacean inspired proteins as a way to reduce cancer risk as they show an amazing potential on the topic of reparation of the genetic material. While traditional tolerance prediction tools like SIFT classified many mutations as deleterious by human standards, structural modeling reveals that analyzed proteins have a stable tertiary structures with the RSMD values going as low as 0.256 Å. Suggesting that blue whales achieved a higher tolerance to mutational loads, allowing them to maintain structural integrity. The research is a critical step towards understanding how Peto's paradox works, by offering a proof of concept which is a key step in the fight against cancer. The aforementioned suggests that the protein Lig1 is a high plasticity candidate that can tolerate evolutive divergence without the risk of collapse. The study pinpoints what specific proteins in the BER pathway should be prime candidates for wet lab analysis. Ultimately providing the blueprint necessary to transition from theoretical in silico analysis to in vivo analysis, aiding in the engineering of more resilient human cells, able to repair the genetic code more efficiently.

References

Abegglen, L. M. et al. "Potential Mechanisms for Cancer Resistance in Elephants and Comparative Cellular Response to DNA Damage in Humans." *JAMA*, 314(17), pp. 1850–1860, 2015. DOI: 10.1001/jama.2015.13134.

Abramson, J. et al. "Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3." *Nature*, 630(8016), pp. 493–500, 2024. DOI: 10.1038/s41586-024-07487-w.

American Veterinary Medical Association. "Cancer in Pets." Available: https://www.avma.org/resources/pet-owners/petcare/cancer-pets. Accessed: [insert date], n.d.

Amidon, K. M., and Eichman, B. F. "Structural Biology of DNA Abasic Site Protection by SRAP Proteins." *DNA Repair*, 94, p. 102903, 2020. DOI: 10.1016/j.dnarep.2020.102903.

A*STAR Bioinformatics Institute. "SIFT - Predict Effects of Nonsynonymous/Missense Variants." Available: https://sift.bii.a-star.edu.sg/index.html. Accessed: March 1, 2026, 2025.

Caulin, A. F., and Maley, C. C. "Peto's Paradox: Evolution's Prescription for Cancer Prevention." *Trends in Ecology & Evolution*, 26(4), pp. 175–182, 2011. DOI: 10.1016/j.tree.2011.01.002.

Cheng, J. et al. "Accurate Proteome-Wide Missense Variant Effect Prediction with AlphaMissense." *Science*, 381(6664), eadg7492, 2023. DOI: 10.1126/science.adg7492.

European Bioinformatics Institute (EMBL-EBI). "Introducing Missense Variants and AlphaMissense — AlphaFold." Available: https://www.ebi.ac.uk/training/online/courses/alphafold/. Accessed: March 5, 2026, 2026.

European Bioinformatics Institute (EMBL-EBI). "pLDDT: Understanding Local Confidence — AlphaFold." Available: https://www.ebi.ac.uk/training/online/courses/alphafold/. Accessed: [insert date], n.d.

Firsanov, D. et al. "Evidence for Improved DNA Repair in the Long-Lived Bowhead Whale." *Nature*, 648(8094), pp. 717–725, 2025. DOI: 10.1038/s41586-025-09694-5.

Gorbunova, V. et al. "Comparative Genetics of Longevity and Cancer: Insights from Long-Lived Rodents." *Nature Reviews Genetics*, 15(8), pp. 531–540, 2014. DOI: 10.1038/nrg3728.

Jumper, J. et al. "Highly Accurate Protein Structure Prediction with AlphaFold." *Nature*, 596(7873), pp. 583–589, 2021. DOI: 10.1038/s41586-021-03819-2.

Krokan, H. E., and Bjørås, M. "Base Excision Repair." *Cold Spring Harbor Perspectives in Biology*, 5(4), a012583, 2013. DOI: 10.1101/cshperspect.a012583.

Krokan, H. E., Standal, R., and Slupphaug, G. "DNA Glycosylases in the Base Excision Repair of DNA." *Biochemical Journal*, 325(1), pp. 1–16, 1997. DOI: 10.1042/bj3250001.

Kufareva, I., and Abagyan, R. "Methods of Protein Structure Comparison." In: A. J. W. Orry and R. Abagyan (eds.), *Homology Modeling: Methods and Protocols*, pp. 231–257, 2012. DOI: 10.1007/978-1-61779-588-6_10.

Kumar, P., Henikoff, S., and Ng, P. C. "Predicting the Effects of Coding Non-Synonymous Variants on Protein Function Using the SIFT Algorithm." *Nature Protocols*, 4(7), pp. 1073–1081, 2009. DOI: 10.1038/nprot.2009.86.

Meng, E. C. et al. "UCSF ChimeraX: Tools for Structure Building and Analysis." *Protein Science*, 32(11), e4792, 2023. DOI: 10.1002/pro.4792.

National Center for Biotechnology Information. "About NCBI." Available: https://www.ncbi.nlm.nih.gov/home/about/. Accessed: February 25, 2026, 2025.

Phang Nga Elephant Park. "Why Elephants Rarely Get Cancer." Available: https://phangngaelephantpark.com/elephants-rarely-get-cancer/. Accessed: [insert date], 2018.

Postovskaya, A. et al. "tcrBLOSUM: An Amino Acid Substitution Matrix for Sensitive Alignment of Distant Epitope-Specific TCRs." *Briefings in Bioinformatics*, 26(1), bbae602, 2024. DOI: 10.1093/bib/bbae602.

Rangarajan, A. et al. "Species- and Cell Type-Specific Requirements for Cellular Transformation." *Cancer Cell*, 6(2), pp. 171–183, 2004. DOI: 10.1016/j.ccr.2004.07.009.

Rathnaiah, G., and Sweasy, J. B. "Protein–Protein Interactions in Base Excision Repair." *Biomolecules*, 15(6), p. 890, 2025. DOI: 10.3390/biom15060890.

Redman, M. et al. "What is CRISPR/Cas9?" *Archives of Disease in Childhood: Education and Practice Edition*, 101(4), pp. 213–215, 2016. DOI: 10.1136/archdischild-2016-310459.

Reyes Castro, L., Cimpean, I., and Castro Garcia, L. R. "Walking Through Elephant Cancer Resistance: What It Can Teach Us About Elephants, Genetics and Disease Defenses." *Journal of Student Research*, 12(4), 2023. DOI: 10.47611/jsrhs.v12i4.5377.

Rosignoli, S., and Paiardini, A. "Boosting the Full Potential of PyMOL with Structural Biology Plugins." *Biomolecules*, 12(12), p. 1764, 2022. DOI: 10.3390/biom12121764.

Sim, N.-L. et al. "SIFT Web Server: Predicting Effects of Amino Acid Substitutions on Proteins." *Nucleic Acids Research*, 40(W1), pp. W452–W457, 2012. DOI: 10.1093/nar/gks539.

Terron-Macias, V. et al. "BioSuiteT: A Unified Tool for Biological Sequence Analysis." *Applied Sciences*, 15(12), p. 6565, 2025. DOI: 10.3390/app15126565.

The UniProt Consortium. "UniProt: The Universal Protein Knowledgebase in 2025." *Nucleic Acids Research*, 53(D1), pp. D609–D617, 2025. DOI: 10.1093/nar/gkae1010.

Vazquez, J. M., Kraft, M., and Lynch, V. J. "A CDKN2C Retroduplication in Bowhead Whales is Associated with the Evolution of Extremely Long Lifespans and Altered Cell Cycle Dynamics." *bioRxiv*, 2022. DOI: 10.1101/2022.09.07.506958.

World Health Organization. "Cancer." Available: https://www.who.int/news-room/fact-sheets/detail/cancer. Accessed: [insert date], 2025.

World Health Organization. "Global Cancer Burden Growing, Amidst Mounting Need for Services." Available: https://www.who.int/news/item/. Accessed: [insert date], 2024.

Zhou, Y. et al. "Experimental Mouse Models for Translational Human Cancer Research." *Frontiers in Immunology*, 14, p. 1095388, 2023. DOI: 10.3389/fimmu.2023.1095388.

Biography

Santiago Arizpe joined the Tecnologico de Monterrey in June 2024, as a high school student who is in the International Baccalaureate in the first half of the program. He joined the biology olympiad in the summer of 2025, and won as a state champion, this is because he has a high interest in sciences and new technologies. He is currently enrolled in Math HL, Chemistry and Biology courses. Working on finding solutions via genetics in both the biomedical and energy related fields. Also he's currently working on researching how nervous. In 2023 he won an award for a biology competition in which he excelled, also showing expertise on anatomy related fields working with a doctor on a cardiac related issue. As a result of this, he started pursuing investigation projects.