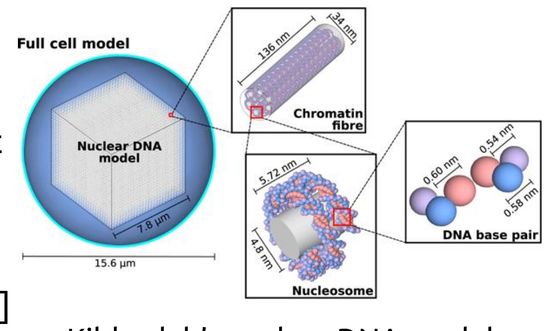
Implementation of Post-Irradiation DNA Repair Mutation Modelling In A Novel Whole-Genome Sequencing Simulator

Zhenlun Dai[1], John Kildea[1]

[1]*McGill University, Montreal, QC, H4A 3J1, Canada, Medical Physics Unit.*

MOTIVATION AND OBJECTIVES

- Ionizing radiation induces DNA damage, which have been shown to leave a distinct mutational signature on the DNA after repair mechanisms have run their course. These signature include base-substitutions, insertion-deletions, balanced inversions and chromosome aberrations. [1][2][3]
- The Kildea Lab employs a Monte Carlo simulation pipeline using TOPAS and its extension, TOPAS-nBio, to predict DNA double strand breaks caused by high-energy particles. [4][5] The DNA damage is then sequenced with RadiSeq, a sequencing simulator built in the lab. [6]
- However, the **current pipeline does not include the repair of DNA damage**.
- This project aims to integrate a DNA repair model, Medras-MC, before the genome is subjected to sequencing. [7]
- **Medras-MC**, in its base version, only outputs post-repair statistics. In this project, we **modified** it **to output the post-repair genome** containing structural variants.



Kildea lab's nuclear DNA model, built in TOPAS-nBio.
Image credit: James Manalad [5]

DNA Double Strand Break (DSB)

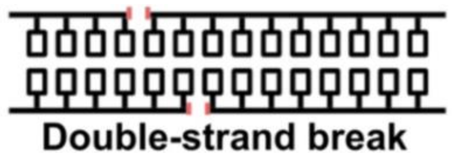
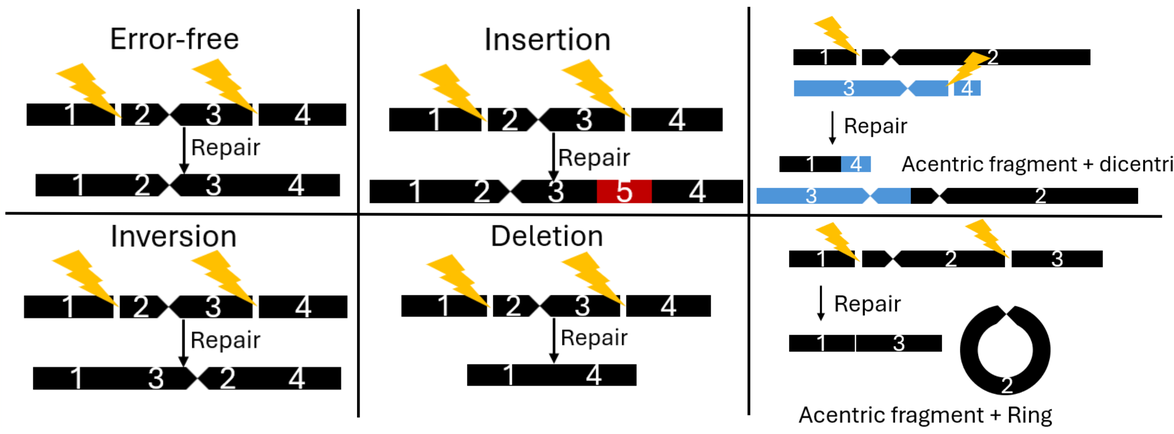


Image credit: [5]

- In this work, a double-strand break (DSB) is defined as two or more DNA backbone lesions occurring on separate strands of the DNA within 10 base-pairs (bp) distance.
- When a DSB occurs, the chromatin can be cleaved into two pieces, resulting in mutation, carcinogenesis, or cell death.

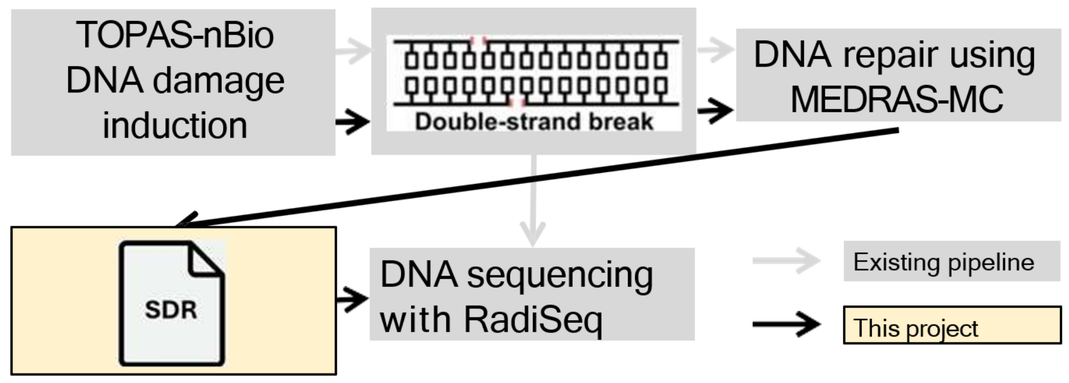
Structural Variants (SVs)



- When the broken ends of DNA fragments are not rejoined with their original pair, a structural variant (SV) can occur.
- Possible SVs includes insertions, inversions, deletions, dicentrics, and more.
- SVs are a hallmark of ionizing radiation damage on the chromosome.
- Some SVs may lead to formation of lethal chromosome aberrations, such as dicentrics and acentric fragments. Their presence disturbs cell mitosis and may cause massive loss of genetic information.
- This project aims to create a file format capable of storing complex SVs that can be later mapped to a genome for sequencing.

METHODS

- The DNA damages in DSBs from existing TOPAS-nBio DNA damage simulations are recorded in the Standard DNA Damage (SDD) file format. [8] The DNA damages are then used to perform DNA repair with the Python software package Mechanistic DNA Repair and Survival (Medras-MC). [7]
- Medras-MC rejoins all unpaired DSB ends across every chromosome and reconstructs the post-repair chromosomes as an intermediate step for its statistical calculations. In this project, the code was modified to pass the repaired DNA strands to a file making extension.
- The file stores the repaired DNA strands in a novel Standard DNA Repair (SDR) file that is inspired from the SDD file format.



RESULTS

- A single SDR file can store multiple cells.
- The SDR file has a master header for information common to all cells.
- A separate per-cell header exists to highlight information unique to the cell's repair results.
- A cell's data section stores the post-repair DNA strands that were incorrectly repaired.
- Each DNA strand is made of **fragments** from the intact DNA strands.

In each fragment, there is information on:

- The intact **strand's ID** where the fragment came from originally.
- The **start position** on the intact strand.
- The **end position** on the intact strand.
- If a **centromere** exists in this fragment.

Master header	Author Associated SDD file name Intact chromosome sizes
Cell 1 header	Cell ID Damaged chromosome sizes Intact strands ID Medras-MC output log
Cell 1 data	Cell 1 repaired DNA strand 1 Cell 1 repaired DNA strand 2

```
Author: Zhenlun Dai
Associated SDD File: n5MeV_inner_alpha_3_SDDOutput.t
Intact Chromosome Sizes: 46, 263.2, 258.1, 212.4, 20
121.3, 113.0, 106.9, 94.7, 83.5, 81.0, 68.0, 66.6, 5
155.5, 149.4, 144.0, 142.9, 140.8, 121.3, 113.0, 106

***end of master header***

***subheader - cell 1***
Cell ID: 1
Damaged Chromosome Sizes: {47: 149.75081461527725, 4
7.36966824774754e-05, 51: 2.2564984537914654}
Intact Strands ID: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38,
Medras-MC Log: ['0', '47', '0', '43', '9', '4', '0',

***data - cell 1***
0; 47; 7/89.3684422808057/89.36845702014217/0; 0;
0; 48; 7/111.47325367890996/111.47331447867299/0, 7/
111.47325367890996/0, 7/111.47281426244076/111.47281
0; 49; 7/71.65495272511848/71.65343549466824/0, 7/71
71.65343549466824/0, 7/71.8713584271327/71.871316972
0; 50; 4/0/70.20679426647001/0, 4/70.2072582300885/7
```

An example SDR file. The syntax was inspired by the SDD file format.

FUTURE WORK

- The RadiSeq sequencing simulator will be modified to take the SDR file and map the repaired chromosomes containing SVs onto the human reference genome.
- A misrepair can be identified in sequenced data when different ends of the fragment maps to different locations on the genome.
- By identifying these misrepairs, this project aims to find DNA inversions, insertions and deletions, as well as inter-chromosome misrepairs.
- This work seeks to contribute to our lab's overarching goal of identifying a genomic biomarker of exposure to ionizing radiation using single-cell sequencing, providing preliminary mutation and translocation information before investing in expensive sequencing kits.

REFERENCES

[1] SJ McMahon, J Schuemann, H Paganetti, KM Prise, Mechanistic modelling of DNA repair and cellular survival following radiation-induced DNA damage. Sci. Reports 6 (2016).
[2] Behjati S, et al. Mutational signatures of ionizing radiation in second malignancies. Nat Commun. 2016;7:12605.
[3] Youk et al., 2024. Quantitative and qualitative mutational impact of ionizing radiation on normal cells. Cell Genomics 4(2):100499. doi:10.1016/j.xgen.2024.100499
[4] Faddegon B, Ramos-Mendez J, Schuemann J, McNamara A, Shin J, Perl J, Paganetti H, The TOPAS Tool for Particle Simulation, a Monte Carlo Simulation Tool for Physics, Biology and Clinical Research, Physica Medica.
[5] Manalad, J., Montgomery, L., & Kildea, J. (2023). A study of indirect action's impact on simulated neutron-induced DNA damage. Physics in Medicine & Biology, 68(7), 075014.
[6] Mathew, F., Galarneau, L., & Kildea, J. (2025). RadiSeq: A single- and bulk-cell whole-genome DNA sequencing simulator for radiation-damaged cell models. Physics in Medicine & Biology, 70(17), 175001.
[7] SJ McMahon, KM Prise, A mechanistic DNA repair and survival model (medras): Applications to intrinsic radiosensitivity, relative biological effectiveness and dose-rate. Front. Oncol. 11 (2021).
[8] Schuemann, J., McNamara, A. L., Warmenhoven, J. W., Henthorn, N. T., Kirkby, K. J., Merchant, M. J., et al. (2018). A new standard DNA damage (SDD) data format. Radiation Research, 191(1), 76.

ACKNOWLEDGEMENTS

The authors would like to thank the members of the NICE-ROKS team of the Kildea Lab for their insights and assistance.

For questions, please contact: zhenlun.dai@mail.mcgill.ca