Computational Investigation of Single Nucleotide Driver Mutations and Tumor Evolution Using Chromatin Conformation Data

Ellie Rahm Kim[1], **Kadir Caner Akdemir**[2]

1) The University of Texas at Austin; 2) Department of Neurosurgery, MD Anderson Cancer Center

THE UNIVERSITY OF TEXAS MD Anderson Cancer Center

Introduction

- Cancer is a disease of uncontrolled cell growth caused by mutations in the cell's genome.
- Computational analysis of cancer genome sequencing data provides key insight into the underlying mechanisms of cancer initiation and progression.
- Frequency of mutations can be studied to track the evolutionary path of tumor samples. This allows physicians to make better prognosis for future patients[1].
- Characterizing known patterns in genome sequencing data allows inferences to be made about individual causes of tumor[2].
- These studies altogether form the cornerstone of precision cancer medicine studies and the development of personalized treatment for patients.
- Here, we explored a novel approach of employing 3D genome sequencing data obtained from the Hi-C assay to derive significant conclusions from tumor samples.

Technique

- BWA-MEM and Samtools were used for mapping, sorting, and indexing raw sequencing data from the Hi-C assay[3,4].
- Picard was used for filtering duplicates and marking read groups[5].
- MuTect was used for detecting meaningful somatic point mutations[6].
- Funcotator was used for analyzing the functions of detected variants and annotating the genes accordingly[6].
- deconstructSigs was used for identifying patterns in mutational spectra to reveal the possible extrinsic/intrinsic causes of tumor[7].
- Matplotlib was used for visualizing the results[8].

Fig. 1 | Workflow for computational analysis of cancer genome sequencing data[9]

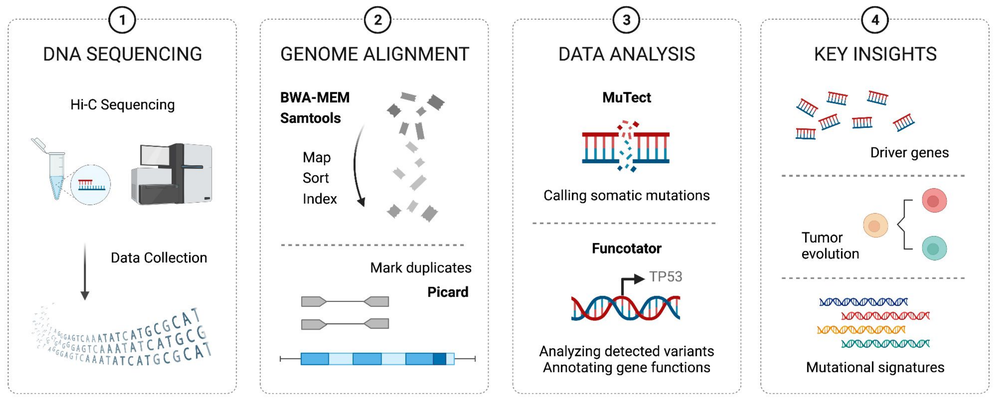
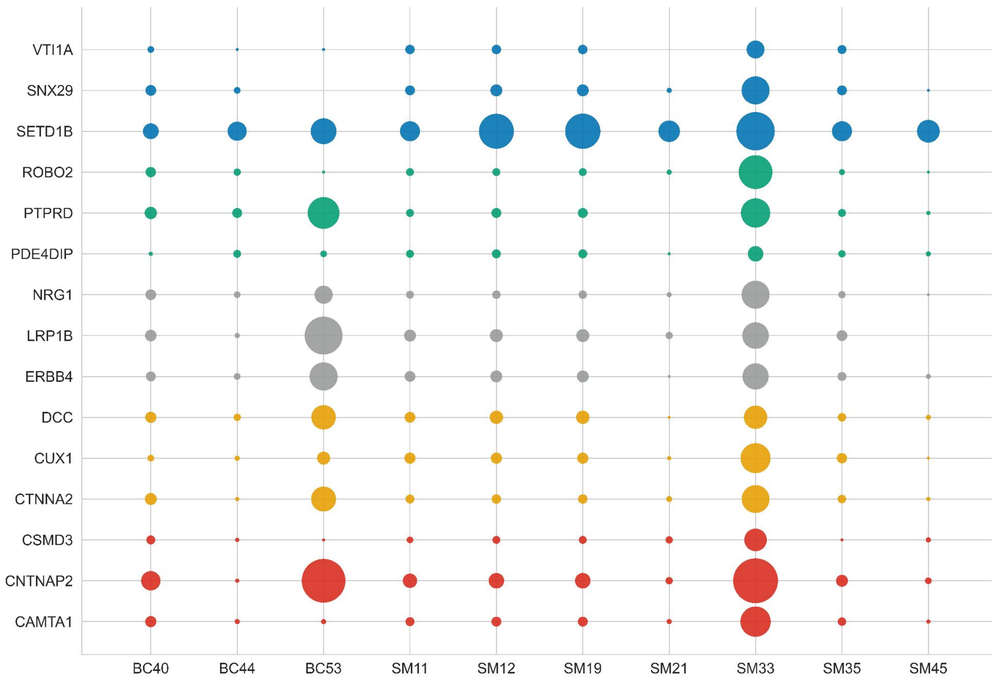


Fig. 2 | Mutational cancer driver genes



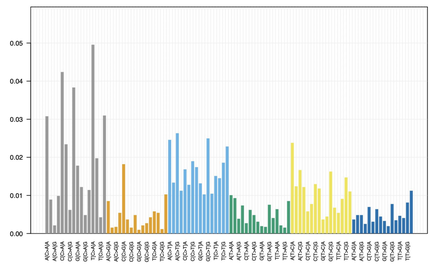
Based on the gene function analysis of the detected somatic mutations, these cancer genes were identified as drivers across the tumor samples. Number of mutations in each gene is represented as the size of the dots in the matrix.

Table. 1 | Sequenced tumor sample information

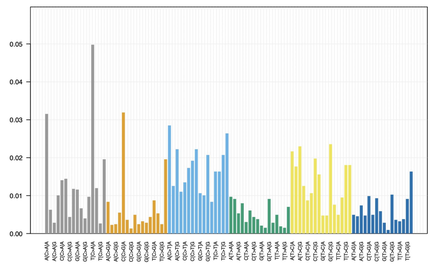
	BC40	BC44	BC53	SM11	SM12	SM19	SM21	SM33	SM35	SM45
Cancer Type	WDLPS	OS	OS	DDLPS	WDLPS	OS	DDLPS	OS	DDLPS	DDPLS
Raw Reads	483 M	370 M	441 M	466 M	581 M	5 M	386 M	551 M	455 M	952 M
Total Mutations	33,278	10,592	69,981	5,368	5,786	144	8,185	267,638	18,601	4,829
Missense	254	38	224	31	47	0	15	105	64	32
Nonsense	23	5	5	2	0	0	0	5	4	2
Intron	14,318	4,108	27,561	2,156	2,403	62	3,816	133,073	8,030	2,172
IGR	14,093	4,891	32,667	2,430	2,439	53	3,199	98,999	7,957	1,911
Silent	143	32	206	22	31	0	18	65	37	27
Splice Site	19	7	20	2	6	0	2	9	8	3
5' UTR	121	24	135	11	18	1	19	204	29	6
3' UTR	312	97	539	66	52	2	84	1,818	137	64

Fig. 3 | Single base substitution spectrum

a Sample SM11



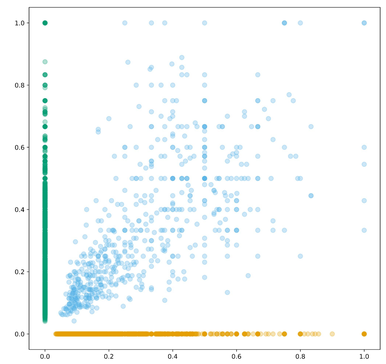
b Sample SM12



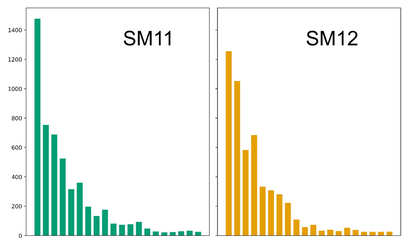
a | Sample SM11 was affected by tobacco smoking (SBS4) and normal somatic cell division (SBS5)[2]. **b** | Sample SM12 was affected by defective homologous recombination DNA damage repair (SBS3) and normal somatic cell division (SBS5)[2].

Fig. 4 | Variant allele frequency of somatic mutations

a



b



a | The presence of shared mutations with relatively small variant allele frequencies indicate that the tumor samples originated from the same phylogenetic structure[1]. **b** | Sample SM12 exhibits mutations with higher variant allele frequencies compared to sample SM11. Therefore, SM12 consists of more clonal mutations.

Results

- 10 tumor samples across 3 different cancer types were sequenced and analyzed. A total of 424,402 somatic mutations were detected in the samples.
- Among the 733 genes that are listed in the Cancer Gene Census (CGC), 653 genes had a variant present within the samples[10].
- We detected 15 genes as the most mutated genes in our sample set. SETD1B gene had a substantially high mutation rate compared to other genes, followed by CNTNAP2 gene.
- In-depth analysis of two different sarcoma samples from a single patient (i.e. sample SM11 and SM12) provided insight into tumor heterogeneity and clonal growth of cancer.

Conclusions

- Computational analysis of chromatin conformation data allowed novel conclusions to be drawn about the genetic basis of cancer.
- Further investigation on multiple tumor samples within a single patient will allow us to gain a better understanding of the evolutionary processes in cancer.

References

1) Korber et al. (2019). Current Opinion in Systems Biology, 16:1-9.
2) Alexandrov et al. (2020). Nature, 578:94-101.
3) Li et al. (2009). Bioinformatics, 25:1754-1760.
4) Li et al. (2009). Bioinformatics, 25:2078–2079.
5) Broad Institute. (2022). http://broadinstitute.github.io/picard/.
6) Van der Auwera et al. (2020). Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (1st Edition). O'Reilly Media.
7) Rosenthal et al. (2016). Genome Biol, 17:31.
8) Hunter et al. (2007). Computing in Science & Engineering, 9(3):90-95.
9) Created with BioRender.com.
10) Sondka et al. (2018). Nature Reviews Cancer, 18(11):696-705.