Abstract # 507

A Signature of Stress-Induced Mutagenesis in Cancer

Kimberly J. Bussey and Luis Cisneros

NantOmics, LLC, Tempe, Arizona

Background

- AID/APOBEC family of cytosine deaminases contribute to mutational clustering but fail to explain 50% of the clusters observed in cancer genomes[1]
- Stress-induced mutagenesis in bacteria occurs when double-strand breaks (DSB) in DNA happen in the context of additional cellular stress sufficient to initiate the SOS response[2]
  - DinB activity results in mutational clusters characterized by mutational density decaying as a function of the distance from the DSB, but remaining above background rates of mutagenesis up to 1 MB away
- In humans, the orthologous genes to DinB have become specialized for translesion synthesis (TLS)[3]
- The dysregulation of cell cycle and DNA repair that characterizes most tumors would logically increase the need for mutagenic break repair and TLS in cancer
- We hypothesized that stress-induced mutagenesis in cancer would result in peaked clusters of SNVs driven by TLS

Methods

- SNVs from whole-genome sequencing
  - 764 tumors from International Cancer Genomics Consortium (ICGC), release 19 with somatic structural variants[4]
  - 129 offspring from trios in the 1000 Genomes project[5]
- 500 independent samples of simulated data from random independent events, with uniform distribution across the genome, at seven different total SNVs counts (1000; 2500; 5000; 10,000; 25,000; 50,000; 100,000)
- Simulated SNVs were modeled as a one dimensional Poisson Point Process, with the assumption that events are stochastically independent and uniformly distributed in the genomic space
  - The number of events in a region of size X is a random variable with Poisson distribution: $P(N=n) = (\Lambda^n/n!)e^{-\Lambda}$ with $\Lambda=N_{SNV}(X/L)$ and L the genome length
  - Clustering properties of the mutation distribution depend of the average density of mutations
  - Clusters are defined as those SNVs with a less than 1% probability of representing a group of 3 or more SNVs with an inter-SNV distance of 25 kb or less
  - Clusters are not expected to be observed for samples with less than 1000 SNVs
  - The average inter-mutation distance is less than 25 kb for about 123,000 SNVs
  - These limits define a range on total SNVs for which the null hypothesis is useful as a comparison
- 21 base pairs of DNA centered on each SNV pulled from reference to determine sequence context for TLS, APOBEC, and AID
- SItH Score:
  - Measure of cluster shape, e.g. how does the density of SNVs vary as a function of the distance from the center of mass of a cluster
  - Center of mass of a cluster is taken as the likely region of DSB
  - Represents difference between the cumulative probability of encountering an SNV at a given distance within 250 kb of the center of mass of the observed data from a uniform distribution
    - SNVs are not necessarily part of the cluster that defined the center of mass
  - SItH score ranges from -1 to 1
    - Values below 0 indicate a region with more SNVs at the boundaries than near the center of mass
    - Values greater than 0 point to more SNVs closer to the center of mass

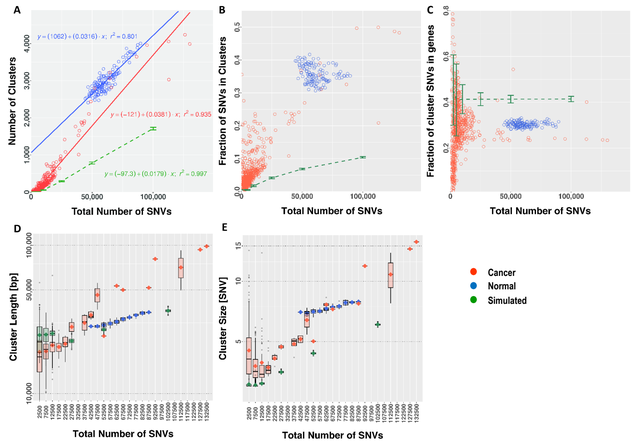

Figure 1. Somatic mutation is driven by SNVs found in clusters. As the total number of SNVs increases, so does the number of clusters and the fraction of SNVs that are part of clusters (A, B, D, and E). The fraction of SNVs in clusters that overlap genes remains constant, indicating there is a shift to mutation outside of genes (C).

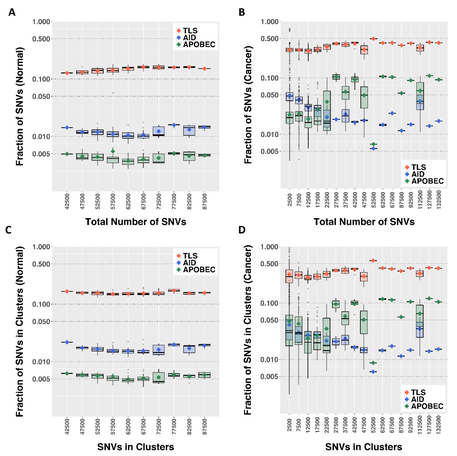

Figure 3. TLS contributes a greater number of SNVs overall and to clusters than either APOBEC or AID. TLS contributes ~ 15% of the SNVs found in normal (A) and ~45% of the SNVs in cancer (B). This amount remains constant even as the total number of SNVs increases. The number of SNVs in clusters with the TLS motif reflects the genome-wide distribution (C & D).

Results

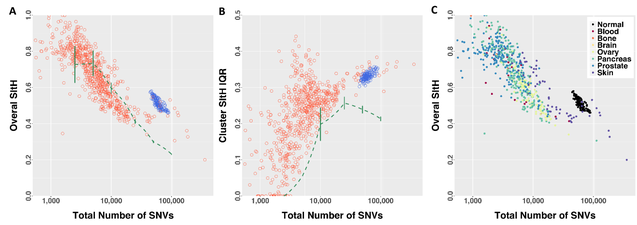

Figure 2. Clusters in cancer are more peaked in shape than normal or simulated data. The overall SItH score is higher in cancer (A) and shows a greater diversity (B) than would be predicted under an assumption of random, uniform mutation or compared to normal. SItH score varies by tissue of origin (C).

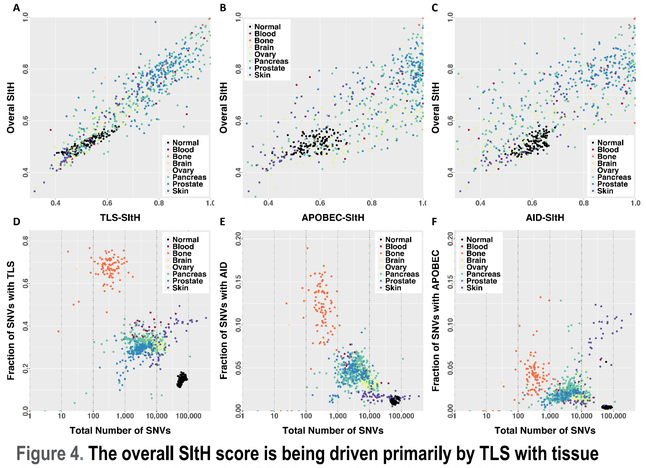

Figure 4. The overall SItH score is being driven primarily by TLS with tissue dependent contributions of APOBEC and AID. Motif specific SItH scores were computed using only SNVs assigned to that particular motif. TLS-SItH shows a good correlation (see Table 1 for specific coefficients for each motif) with overall SItH (A). The amount of each mechanism contributing to somatic mutation along with total SNV count is sufficient to separate tumors types from one another (D-F).

Table 1. Contribution of Mechanism to Overall SItH Score

Tissue of Origin	TLS SItH Coefficient	APOBEC SItH Coefficient	AID SItH Coefficient	Adjusted R-square
Blood	0.9845	NS	NS	0.8901
Bone	1.015	NA	NA	0.9978
Ovary	0.7196	0.0499	NS	0.8678
Pancreas	0.743	NS	NS	0.8493
Prostate	0.5724	NS	0.1083	0.7434
Skin	0.9311	NS	0.0948	0.9974

Coefficients were estimated from fitting an additive, linear model with the lm function in R:
Overall-SItH ~ TLS-SItH + APOBEC-SItH + AID-SItH
NS = Not Significant; NA = Not Applicable

Conclusions

- Somatic mutation in cancer is occurring in clusters
- Cancer has more peaked clusters with a greater diversity of cluster shape than either normal or simulated data
- The shape of the clusters in cancer is dependent on tissue of origin
- The majority of SNVs in clusters in both normal and cancer have a motif suggestive of TLS
- TLS contributes more to cluster shape than either APOBEC or AID
- Our results support the hypothesis that cancer is using a conserved mechanism of stress-induced mutagenesis to generate somatic mutation

Acknowledgements

We thank Paul Davies, Charley Lineweaver, Susan Rosenberg, and Bob Austin for insightful discussions into the role of genomic instability in cancer within an evolutionary context. We also acknowledge the work of the clinical collaborators, data analysis teams, and funders in generating the WGS data in the ICGC database, release 19.

References

1. Roberts SA, Sterling J, Thompson C, Harris S, Mav D, Shah R, et al. Clustered Mutations in Yeast and in Human Cancers Can Arise from Damaged Long Single-Strand DNA Regions. Mol Cell. 2012;46: 424–435. doi:10.1016/j.molcel.2012.03.030
2. Shee C, Gibson JL, Rosenberg SM. Two Mechanisms Produce Mutation Hotspots at DNA Breaks in Escherichia coli. Cell Rep. 2012;2: 714–721. doi:10.1016/j.celrep.2012.08.033
3. Waters LS, Minesinger BK, Wiltrout ME, D'Souza S, Woodruff RV, Walker GC. Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance. Microbiol Mol Biol Rev MMBR. 2009;73: 134–154. doi:10.1128/MMBR.00034-08
4. Hudson (Chairperson) TJ, Anderson W, Aretz A, Barker AD, Bell C, Bernabé RR, et al. International network of cancer genome projects. Nature. 2010;464: 993–998. doi:10.1038/nature08987
5. Khurana E, Fu Y, Colonna V, Mu XJ, Kang HM, Lappalainen T, et al. Integrative Annotation of Variants from 1092 Humans: Application to Cancer Genomics. Science. 2013;342: 1235587. doi:10.1126/science.1235587

To download this poster, please scan