Three-Dimensional Genome Architecture Predicts Chromosomal Alterations in Human Cancers

Geoff Fudenberg, Leonid Mirny

Massachusetts Institute of Technology

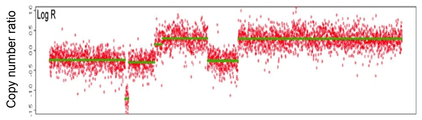
HST

Introduction

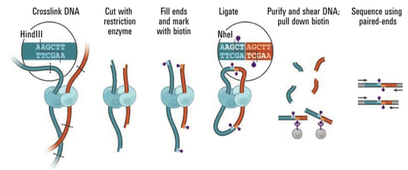
Genomic instability is a common theme in cancer. As characterized by Beroukhim et al. 2010, Somatic Copy-Number Alterations (SCNAs) occur with very high frequencies; these alterations (deletions and amplifications) affect driver genes and many more passenger genes. We developed a new heatmap-based method to analyze SCNAs, and find:

(i) genomic re-arrangements are influenced by 3D chromosomal architecture (HiC data: Lieberman-Aiden and van Berkum et al. 2009)

(ii) passenger SCNA events show signatures of purifying selection.

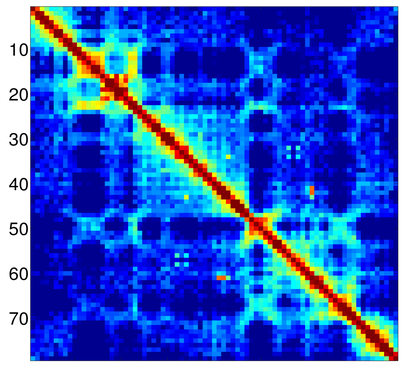


Illustrative SCNA probe profile showing segmentation of copy number profile vs. genomic location (Loo et al, 2010)



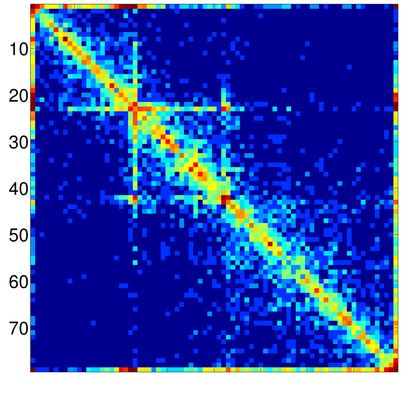
Illustratration of HiC method for determining 3D chromatin architecture (Lieberman-Aiden and van Berkm et al., 2009)

Constructing SCNA heatmaps provides a novel method for comparison with 3D architecture



Chr 17: HiC heatmap

HiC data represented as a matrix of counts of spatial interactions (contacts) between positions *i* and *j* as determined in GM06690 cell line.

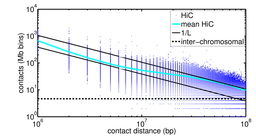


Chr 17: SCNA heatmap

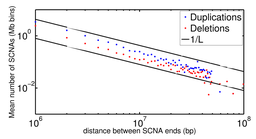
SCNA matrix constructed by counting the number of focal duplications or deletions that start at position *i* and ending at *j* on the same chromosome.

Our subsequent analysis exlcuded SNCAs which start/end within 1Mb of telomeres/centromeres.

Patterns of 3D architecture are Visible in SCNA data

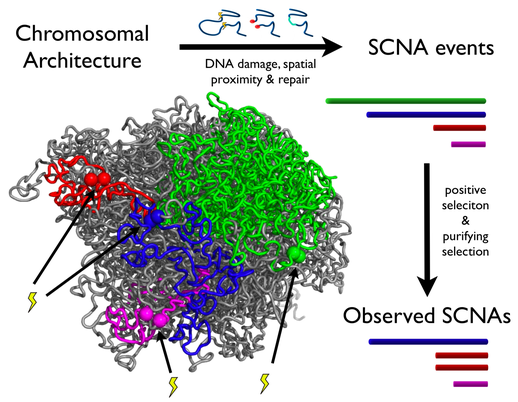


Contact probability scaling for HiC



Scaling for mean number of SCNA

Above: Scaling for mean number of SCNA is heuristically similar to *1/L* scaling observed in HiC data and predicted for a "fractal globule" state



Above: The observed distribution of SCNAs depends upon a combination of mutational and selective forces. Illustration shows four possible SCNA events whose lengths are determined by a fractal globular structure (coordinates M. Imakaev). Contact points are highlighted by spheres.

Purifying selection dictates a distribution of SCNA lengths

Population genetics provides the probability of fixation for a new mutation:

$$\pi_N(\triangle F) = \frac{\triangle F}{1 + \triangle F - (1 + \triangle F)^{-(N-1)}}$$

A simple model for purifying selection acting on SCNA assumes that the deleterious effect of a SCNA increases linearly with the length of the affected chromosomal region. This sort of purifying selection implies a roughly exponential length distribution for SCNAs:

$$P_{\mathrm{SCNA}}(L) \sim \mu(L)\exp(-L/L_0)$$

Poisson Loglikelihood Allows For Sensitive Comparison of Competing Models

We compare models which depend on: 3D architecture alone, purifying selection alone, or a combination of 3D architecture and purifying selection.

For a given model: $L(\mathrm{SCNA}|\lambda) = \sum_{|i-j|>2} -\lambda_{ij} + \mathrm{SCNA}_{ij}\log\lambda_{ij}$

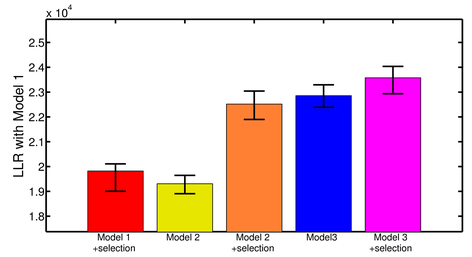
References

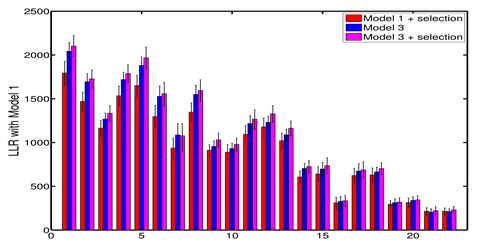
Beroukhim et. al., "The Landscape of Somatic Copy Number Alterations Across Human Cancers." Nature, 2010.

Lieberman-Aiden and van Berkum et al., "Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome." Science, 2009.

Loo et al., "Allele-Specific Copy Number Analysis of Tumors." PNAS, 2010.

1/L fits distribution of SCNA better than exponential





Comparison of Models across chromosomes.

Model 1: two ends of alteration drawn randomly from same chromosomal arm

Model 2: probability of alteration depends on probability of 3D contact between ends as given by HiC data

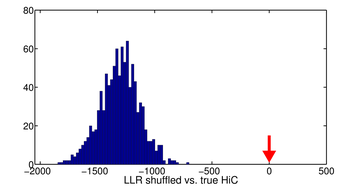
Model 3: same as 2, but with probability scaling as *1/L* , ie. according to the fractal globule model.

Each model is considered by itself and in combination with purifying selection. Errorbars from bootstrapping

Aggregate: theoretical contact probability between two loci explains length distribution of SCNA better than a simple model of purifying selection.

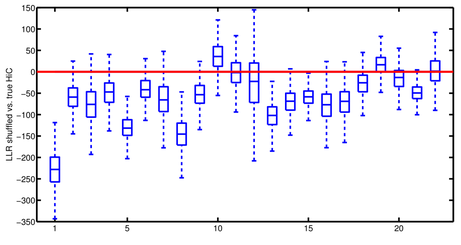
By chromosome: in particular, *1/L* fits relatively better for longer, gene-poor chromosomes.

True 3D architecture fits distribution of SCNAs significantly better than permuted data



Distributions of loglikelihood ratios L(SCNA|shuffled HiC) – L(SCNA|HiC) aggregated across chromosomes 1-22.

Since both SCNAs and HiC heatmaps have a strong distance dependence, we shuffle in a way which preserves overall distance dependence, but destroys the rest of the fine structure..



Distributions of loglikelihood ratios L(SCNA|shuffled HiC) - L(SCNA|HiC) for chromosomes 1-22.

HiC beats permuted HiC (ie. LLR < 0) for all chromosomes except chr 10, 19, and 22. HiC fits better with p <.001 for chrs 1,5,8,13.

Summary & Conclusions

- A mutation rate related to nuclear geometry explains the length distribution of SCNA better than a simple model of selection.
- Comparisons with permuted data demonstrates the significance of detailed nuclear structure for SCNA.
- Passenger mutations show signature of purifying selection

Future Directions:

- Develop better, and more detailed, evolutionary models for SCNA.
- Use 3D information to gain insight into mechanisms of SNCA and improve prediction of cancer causing genes.