

2013

Understanding The Molecular Mechanisms Underlying Complex Cancer Genome Rearrangements

Robert Hedlund
Lutz, FL

Vladimir N. Uversky
University of South Florida, vversky@health.usf.edu

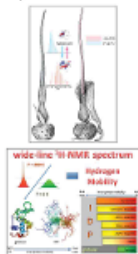
Follow this and additional works at: https://digitalcommons.usf.edu/mme_facpub

 Part of the [Medicine and Health Sciences Commons](#)

Scholar Commons Citation

Hedlund, Robert and Uversky, Vladimir N., "Understanding The Molecular Mechanisms Underlying Complex Cancer Genome Rearrangements" (2013). *Molecular Medicine Faculty Publications*. 448.
https://digitalcommons.usf.edu/mme_facpub/448

This Article is brought to you for free and open access by the Molecular Medicine at Digital Commons @ University of South Florida. It has been accepted for inclusion in Molecular Medicine Faculty Publications by an authorized administrator of Digital Commons @ University of South Florida. For more information, please contact digitalcommons@usf.edu.



Understanding the molecular mechanisms underlying complex cancer genome rearrangements

Robert Hedlund & Vladimir N Uversky

To cite this article: Robert Hedlund & Vladimir N Uversky (2013) Understanding the molecular mechanisms underlying complex cancer genome rearrangements, *Intrinsically Disordered Proteins*, 1:1, e25954, DOI: [10.4161/idp.25954](https://doi.org/10.4161/idp.25954)

To link to this article: <https://doi.org/10.4161/idp.25954>



Copyright © 2013 Landes Bioscience



Published online: 23 Aug 2013.



Submit your article to this journal [↗](#)



Article views: 523



View related articles [↗](#)

Understanding the molecular mechanisms underlying complex cancer genome rearrangements

Robert Hedlund¹ and Vladimir N. Uversky^{1,2,3*}

¹Department of Molecular Medicine and USF Health Byrd Alzheimer's Research Institute; Morsani College of Medicine; University of South Florida; Tampa, FL USA; ²USF Health Byrd Alzheimer's Research Institute; Morsani College of Medicine; University of South Florida; Tampa, FL USA; ³Institute for Biological Instrumentation; Russian Academy of Sciences; Moscow Region, Russia

In their recent paper, Malhotra et al. performed a comprehensive analysis of the structural variation breakpoints in 64 cancer genomes from 7 tumor types to find the prevalence and origins of complex genomic rearrangements.¹ Since this work represents a breakthrough in understanding of the molecular mechanisms underlying complex cancer genome rearrangements, we provide a brief overview of this outstanding study.

Structural variation (SV), which includes duplications, deletions, inversions, transpositions, and other genomic rearrangements, is an abundant and functionally important class of genetic variation in mammals that serves as a rich source of genetic diversity in mammals.² Furthermore, “spontaneous genomic rearrangements are a major source of genetic diversity in cancer and the cause of numerous human disorders. While most genome structural variants (SVs) can be readily categorized into the canonical forms—deletion, duplication, inversion, and translocation—there is growing evidence that a nontrivial fraction are complex genomic rearrangements (CGRs) composed of multiple clustered breakpoints that cannot be explained by a single DNA end-joining or recombination event.”¹ From this hypothesis, the authors perform a series of increasingly specific bioinformatics experiments to clarify those recombination events. Their results suggest greater importance of double strand breaks in tumorigenesis.

The authors analyzed 64 tumors from The Cancer Genome Atlas (TCGA)

including 12 basal-like breast cancers (BRCA), 3 colon adenocarcinomas (COAD), 18 glioblastomas (GBM), 6 lung adenocarcinomas (LUAD), 13 lung squamous cell carcinomas (LUSC), 11 ovarian cancers (OV), and 2 renal adenomas (READ). Tumor and matched normal samples, in the form of blood or normal solid tissue, were subjected to the Illumina paired-end sequencing by TCGA.

To identify structural variant (SV) breakpoints, they used HYDRA-MULTI, a new multisample version of their HYDRA paired-end mapping algorithm³ using population-based clustering. Read pairs from all 64 tumor samples and 65 normal samples were combined into a single clustering step, which enabled simultaneous measurement of the evidence for each breakpoint in each sample. This method and several filtering steps identified 6,179 somatic rearrangement breakpoints. SV breakpoints with distance < 1 Mb were classified as deletions, tandem duplications, or inversions based purely on their orientation.

Since DNA was not available, the authors used local de novo breakpoint assembly to assess the validation rate. They modified the SGA assembler to report all paths through the assembly string graph, rather than just a consensus contig. This allows for assembly of breakpoints present at relatively low (< 50%) allele frequencies within tumor cell populations, as the vast majority of somatic SVs are. Contigs exhibiting split alignments consistent with the original breakpoint prediction were judged to validate the call.

Keywords: cancer genome, genome rearrangements, Structural variation, breakpoint, chromothripsis

*Correspondence to: Vladimir N. Uversky; Email: vuversky@health.usf.edu

Submitted: 07/25/2013

Accepted: 07/29/2013

Published Online: 08/23/2013

<http://dx.doi.org/10.4161/idp.25954>

Commentary to: Malhotra A, Lindberg M, Faust GG, Leibowitz ML, Clark RA, Layer RM, Quinlan AR, Hall IM. Breakpoint profiling of 64 cancer genomes reveals numerous complex rearrangements spawned by homology-independent mechanisms. *Genome Res* 2013; 23:762-76; PMID:23410887; <http://dx.doi.org/10.1101/gr.143677.112>

The authors excluded small inversions (< 10 kb). Despite this filter, the assembly-based validation rate for inversions (54%) was substantially lower than for large-scale rearrangements (80.3%), tandem duplication (91.8%), and deletions (92.5%). Sets of 3 or more interconnected breakpoints, referred to as “breakpoint clusters” or simply “clusters,” were identified in 2 steps: (1) breakpoint loci were defined by merging calls whose mapping positions in the reference genome are within 100 kb of one another; and (2) loci that share breakpoint calls in common were chained together. Therefore, all the breakpoints in a cluster are interconnected and no farther than 100 kb from another breakpoint in the cluster.

As an independent test of accuracy, the authors assessed the relative number of somatic breakpoint calls in tumor vs. normal samples and found that calls private to a single sample are overwhelmingly enriched in tumors. Of the 6,502 breakpoint calls detected exclusively in one of the 129 data sets, 6,179 (95%) were observed in a single tumor, whereas a mere 323 (5%) were found in a single normal sample.

The authors performed a variety of control experiments. First, they conducted a Monte Carlo simulation by shuffling breakpoint positions within uniquely mappable regions of the reference genome, controlling for the size distribution and class of SV calls, and discovered a mean of 4.9 breakpoint clusters per iteration. This is in stark contrast to the 154 found among all the samples in the cancer tissue data.

A breakpoint cluster may result from a complex one-off mutational event that simultaneously generates multiple breakpoints, or from a series of simple mutations that occur in stepwise fashion. Although their simulations clearly demonstrate that breakpoint clusters are very rarely identified by chance under a model of random mutation, tumor genomes do not necessarily evolve through random processes. Breakpoint clusters may be generated by breakage-fusion bridge cycles that promote repeated rounds of mutation within a chromosome arm, or from progressive amplification of genes that confer fitness advantage. An informative feature

for inferring the most likely mutational scenario is the number of DNA copy number states associated (CNA) with a breakpoint cluster. One-off mutations caused by repair of multiple DNA breaks have limited ability to generate multiple copy number states because DNA breakage and ligation can only involve the small number of chromosomes inside of a cell at any given time, and most reported chromothripsis events involve 3 or fewer states (e.g., loss, gain, and unaltered). Replication-based mechanisms such as microhomology-mediated break-induced replication (MMBIR) can in theory generate an unlimited number of states in a one-off event, and there have been reports of triplication, but to the authors’ knowledge most if not all variants attributed to MMBIR also exhibit 3 or fewer states. In contrast, stepwise mutations often produce numerous copy number states due to the likelihood that new CNAs arise within older CNAs.

For a breakpoint cluster to be judged as a complex rearrangement resulting from a one-off mutation, the authors required that it exhibited no more than three copy number states and no more than 1 amplified copy number state exceeding 4 predicted copies, and that it was not a focal amplification composed of a single contiguous amplified region. Using these criteria, 97 of the 154 breakpoint clusters (63%) are consistent with being generated by a one-off mutational event. The authors hereafter refer to these as complex genomic rearrangements (CGRs). CGRs are found in 43 of 64 tumors (67%) and account for 13.6% of all somatic breakpoints

The most rigorous definition of chromothripsis relies on performing a Monte Carlo simulation for each putative event to test whether the observed rearrangement breakpoints, applied to an *in silico* chromosome in stepwise fashion and random order, produce significantly more CNA states than observed in the data. The authors used this simulation strategy to test the 13 extreme CGRs composed of 10 or more breakpoints. The authors find that seven of 13 extreme CGRs have significantly fewer CNA states than expected by chance under a stepwise model ($p < 0.05$). The 7 events are found in 7 different tumors, 5 of which are GBMs, and

the increased incidence in GBM samples remains statistically significant (Fisher’s exact; $p = 0.0157$).

Finally, inspection of rearrangement patterns across the entire spectrum of CGRs suggests that chromothripsis may be the most extreme manifestation of a common underlying mutational process. However, this difference in CGR severity as related to breakpoint number and density is, in the view of the authors, more likely to reflect differences in the severity of DNA damage events provoking rearrangement, not a distinct mechanism. These results indicate that highly complex CGRs often arise early during tumorigenesis or, alternatively, are often under strong selection and rise to high frequency. Either scenario implicates complex rearrangements as a functionally important form of tumor genome evolution.

The mechanism(s) of CGR formation is an unresolved question. There are 2 general models: (1) template switching at a DNA replication fork; and (2) chromothripsis, which involves chromosome shattering followed by nonhomologous or microhomology-mediated end-joining (NHEJ/MMEJ). There is evidence for each model, including sequencing of several hundred chromothripsis breakpoints in cancer genomes and analysis of 282 CGR breakpoints from germline genomes.

In practice, it is difficult to distinguish between template switching and end-joining because both mechanisms can use stretches of microhomology (2–10 bp) and both can lead to small-scale DNA insertions or rearrangements at the breakpoint. Moreover, despite the elegance of MMBIR for explaining certain CGR architectures (e.g., triplication), end-joining can, in principle, lead to any conceivable CGR architecture given a sufficient number of chromosomes and breaks. However, MMBIR has one strict requirement that end-joining does not: It requires microhomology. To the authors’ knowledge, no DNA polymerase can initiate template-directed synthesis without a primer.

If a CGR is generated by template switching, then it is reasonable to expect that all of the breakpoints for that CGR would exhibit microhomology. Thus, the authors next asked how many variants were composed solely of breakpoints

containing microhomology. Of the 134 breakpoint clusters for which 1 or more breakpoints were assembled, 97 (72.4%) have at least 1 flush breakpoint that appears to be derived from NHEJ, not MMBIR or MMEJ. This is true for 68.9% of stepwise rearrangements, 69.7% of mild one-off CGRs, and 100% of chromothripsis events. Thus, at most 27.6% of breakpoint clusters are consistent with being generated solely by microhomology-mediated mechanisms. Given that end-joining can also use microhomology through MMEJ and that MMEJ is thought to account for a nontrivial fraction of end-joining events, these data further argue that the contribution of replication-based mechanisms to CGR formation is minor.

These data indicate that complex rearrangements are an important aspect of cancer genome evolution. The authors identified chromothripsis events in an unbiased, automated fashion and found a significantly higher incidence in GBM (38.9%) relative to the other tumor types (8.7%). This definitively shows

that chromothripsis is a variable phenotype among tumor types. At present, it is unclear whether variable prevalence is due to differences in the frequency of specific transacting mutations, variable exposure to chromothripsis-causing mutagens, differences in the selective pressures faced by different cancers, and/or other unknown factors.

Since tumorigenesis is frequently linked to the dysfunctions of some key proteins and since many cancer-related proteins are intrinsically disordered,^{4,5} next logical step would be to move study from the genome to proteome level. It would be interesting to see how proteome changes due to the cancerous genome rearrangement and to find how disorder propensities of the involved proteins would be affected by the genome rearrangements described in this study.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References

1. Malhotra A, Lindberg M, Faust GG, Leibowitz ML, Clark RA, Layer RM, Quinlan AR, Hall IM. Breakpoint profiling of 64 cancer genomes reveals numerous complex rearrangements spawned by homology-independent mechanisms. *Genome Res* 2013; 23:762-76; PMID:23410887; <http://dx.doi.org/10.1101/gr.143677.112>
2. Zhang F, Gu W, Hurler ME, Lupski JR. Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet* 2009; 10:451-81; PMID:19715442; <http://dx.doi.org/10.1146/annurev.genom.9.081307.164217>
3. Quinlan AR, Clark RA, Sokolova S, Leibowitz ML, Zhang Y, Hurler ME, Mell JC, Hall IM. Genome-wide mapping and assembly of structural variant breakpoints in the mouse genome. *Genome Res* 2010; 20:623-35; PMID:20308636; <http://dx.doi.org/10.1101/gr.102970.109>
4. Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol* 2002; 323:573-84; PMID:12381310; [http://dx.doi.org/10.1016/S0022-2836\(02\)00969-5](http://dx.doi.org/10.1016/S0022-2836(02)00969-5)
5. Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys* 2008; 37:215-46; PMID:18573080; <http://dx.doi.org/10.1146/annurev.biophys.37.032807.125924>