



Aneuploidy meets network analysis: leveraging copy number alterations to identify molecular pathways disrupted in high-grade serous ovarian carcinomas

Yuchae Jung¹, Tae-Min Kim^{2,3}

¹Department of IT Engineering, Sookmyung Women's University, Seoul, Korea; ²Department of Medical Informatics, ³Cancer Research Institute, College of Medicine, The Catholic University of Korea, Seoul, Korea

Correspondence to: Tae-Min Kim, MD, PhD. Department of Medical Informatics, College of Medicine, The Catholic University of Korea, Seocho-Gu, Banpo-Ro 222, Seoul 06591, Korea. Email: tmkim@catholic.ac.kr.

Comment on: Delaney JR, Patel CB, Willis KM, *et al.* Haploinsufficiency networks identify targetable patterns of allelic deficiency in low mutation ovarian cancer. *Nat Commun* 2017;8:14423.

Submitted Apr 12, 2017. Accepted for publication Apr 17, 2017.

doi: 10.21037/tcr.2017.05.03

View this article at: <http://dx.doi.org/10.21037/tcr.2017.05.03>

Cancer is a genomic disorder that often involves the accumulation of various types of genomic alterations that play roles in disease development and progression (1,2). In solid tumors, somatic mutations (herein, point mutations and short indels) and SCNAs (somatic copy number alterations as chromosomal amplifications/deletions) comprise the majority of the genomic alterations in cancer genomes in terms of abundance and genomic fraction, respectively; several to tens of thousands of somatic mutations and SCNAs occupying >50% of the genome have been observed in solid tumor genomes. Aneuploidy is defined as the presence of an abnormal number of chromosomes, and this chromosome-level SCNA has long been recognized as a hallmark of cancer genomes (3). Unlike somatic mutations that can be identified only by sequencing or base pair-resolution genotyping, aneuploidy or chromosome-level SCNAs can be identified by microscopic examination (i.e., karyotyping), which has been refined with high-resolution genotyping techniques such as comparative genomic hybridization (CGH) and microarray-based CGH (array-CGH) to identify subchromosomal or focal SCNAs. Despite many years of research into SCNAs and aneuploidy in cancer, novel features associated with structural variations including SCNAs, such as chromothripsis (4) and chromoplexy (5), have only recently been identified in cancer genomes, suggesting that the biological relevance of SCNAs is not yet completely understood.

SCNAs can be as large as an entire chromosome and may include several hundreds to thousands of genes. Unlike somatic mutations for which the affected genes can be readily identified, the large size of SCNAs and the 'one-to-many' relationship between SCNAs and their related genes has hampered proper biological interpretation of SCNAs. SCNAs can be divided into two categories in terms of their size, i.e., arm-level and focal SCNAs (6). Canonical cancer-related genes such as oncogenes and tumor suppressor genes are enriched in focal SCNAs. Moreover, SCNAs with high copy number changes (e.g., high-level amplifications and homozygous deletions) that are more likely to be functional than single copy number changes are more commonly found in focal SCNAs than in arm-level SCNAs. Thus, it is challenging to uncover which genes in large, arm-level SCNAs are functionally relevant as cancer drivers since it is reasonable to assume that the majority of genes affected by large SCNAs are functionally neutral passengers. It has been proposed that the recurrent SCNAs in a given cohort (i.e., the genomic loci supported by recurrent SCNAs in a population) are likely to be the functional cancer drivers (7). The algorithm "GISTIC" (Genomic Identification of Significant Targets in Cancer) has been used to identify recurrent SCNAs and was employed in a landmark cancer genome analysis project, the Cancer Genome Atlas (TCGA). This frequency-based and data-driven strategy has a number of limitations, such that cancer drivers are often low- or moderate-frequency aberrations (8) and frequent

genomic alterations are often found in fragile regions of the genome that lack apparent biological significance (9). It is of note that the incorporation of knowledge-based information such as network topology (10,11) and other mutational features (12,13) may improve the identification of cancer drivers and thus perform better than data-driven methods, including frequency-based approaches.

Among the various types of solid tumors, high-grade serous ovarian carcinoma (OV) genomes are unique in that SCNAs predominate over somatic mutations and thus are expected to comprise the major cancer drivers (14). In a recent article by Delaney *et al.* (15), the SCNA profiles of hundreds of OV genomes were obtained from TCGA consortium (16) and analyzed to reveal SCNA-driven functional disturbances and the molecular pathways that are affected in OV genomes. The authors employed a novel method called HAPTRIG (haploinsufficient/triplosensitive gene) to identify recurrent aneuploidy or SCNA patterns with functional significance in OV genomes. In brief, 187 gene sets were obtained from the KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway database. To identify possible pairs of genes in each gene set, protein-protein interactions were examined using the BioGRID database (17). Each interaction was weighted according to type (either chromosomal amplification or deletion) and SCNA copy number (single copy number changes or high-level amplifications/homozygous deletions) as well as whether the interacting genes were known to be dosage-sensitive (e.g., haploinsufficient or haploproficient genes). The sum of the weighted edge scores was calculated and normalized for each pathway. An average normalized edge score was calculated across OV samples and further transformed into the permutation-based significance with adjustment for multiple testing. With this method, the authors identified a number of KEGG pathways with significantly higher HAPTRIG scores toward amplification or deletion in OV genomes. Among them, autophagy was ranked at the top for chromosomal deletions, along with other proteostasis pathways such as endoplasmic reticulum (ER) stress-, ubiquitin-mediated proteolysis-, and lysosome-related pathways. To functionally interpret the SCNA profiles of OV genomes with extensive aneuploidy, HAPTRIG exploits several data- or knowledge-level types of information: (I) protein-protein interactions from the BioGRID database; (II) curated functional gene sets in KEGG pathways; (III) SCNAs of individual OV genomes as a GISTIC output; and (IV) annotated dosage-sensitive genes. Below, we will briefly discuss the issues and concerns

related to these points.

The accumulation of a large quantity of genomics data has led to the emergence of systems biology in an effort to understand the basic principles that underlie a living organism by focusing on the interactions and relationships between various biological elements. Several such system- or network-based analysis methods have been proposed to identify cancer driver mutations. For example, Torkamani and Schork investigated rare but biologically relevant cancer driver mutations by identifying mutations that are co-expressed with frequent mutations within co-occurring gene modules (11). Instead of co-expression data, interaction data can be obtained from public databases containing various resources and types of interaction-based data (18). Methods that make use of such network-based mutation data perform better in identifying cancer driver genes than did those based on simple mutational frequencies (10). In addition, SCNAs and somatic mutations can be both employed to identify rare, but potential cancer drivers that are enriched in certain sub-networks (19). In contrast to these network-based models, HAPTRIG uses the sum of edge scores that are available in a protein-protein interaction database (BioGRID) to weight genes according to the degree or number of connections in a network of interest. In addition, HAPTRIG uses KEGG pathways as functional gene sets, including those involved in autophagy. Similar to gene set enrichment analysis (GSEA) (20) or other types of enrichment analysis, several hundred genes that perform a specific molecular function are collected and used for analysis without considering genetic hierarchies or pathway topology. If the topology of pathways and the hierarchy of genes in the functional gene set is provided, such information can be exploited, as previously shown (21). This analysis exploits pathway topology by examining perturbed genes as differentially expressed genes or those on SCNAs, which are given differential weights whether or not they are located in important positions in the pathway topology. Thus, it can capture actual perturbations in a given pathway, which is not feasible with conventional methods of enrichment analysis that treat all individual members of a set of genes equally. We assume that network or pathway topology may contain important attributes that can be exploited by genome analysis methods including HAPTRIG.

In the article by Delaney *et al.*, the authors performed GSEA analysis and compared the results with those of HAPTRIG. It is of note that autophagy-related genes showed a certain level of enrichment in the GSEA analysis,

but at lower levels than were shown by HAPTRIG. They reasoned that this discrepancy may have arisen because GSEA does not consider the interaction between genes or haploinsufficiency data, which are important attributes of HAPTRIG. However, a more detailed analysis or a comparison with the results of the other algorithms mentioned above may be required to clarify the discrepancy between the GSEA and HAPTRIG results. It is also of note that autophagy is only annotated in a limited number of subsets in KEGG and GO among the thousands of MSigDB gene sets (20) that are currently used as standards for functional gene set analysis, such as GSEA, suggesting that the current annotation of functional gene sets may be biased to well-studied molecular functions and can also lead to a biased functional interpretation of genomics data.

The SCNA profiles are generally provided as segmentation data in which individual chromosomes are divided into a number of segments, each of which has an identical copy number. Since segmentation data usually provide log₂ ratios as estimates of the copy number of given segments, it should be determined whether the log₂ values of segments or further processed absolute copy numbers are assigned to individual genes for gene-level analyses such as HAPTRIG. HAPTRIG uses the absolute copy number calls from GISTIC, such as -2, -1, 0, 1, and 2, representing homozygous deletions, single copy losses, neutral copy numbers, single copy gains, and high-level amplifications, respectively. Using the absolute copy numbers has several advantages, including that outlier values as well as noisy profiles can be effectively ignored or handled. To this end, the GISTIC algorithm that uses log₂ segment values as input have capped values (+1.5 and -1.5 for maximum log₂ ratios for amplifications and deletions, respectively) and threshold values (+0.1 and -0.1 for minimum log₂ ratios for amplifications and deletions, respectively) by default. However, it is also reasonable to expect that the processing of log₂ ratios into absolute copy numbers may accompany a certain level of information loss and potential errors. Recently, it was proposed that absolute copy numbers can be accurately estimated by considering the tumor purity and ploidy levels that are also estimated from SCNA profiles (22). The refined absolute copy numbers may be also exploited by SCNA analysis methods such as HAPTRIG as alternatives to GISTIC output. To deal with recently highlighted tumor heterogeneity or clonality issues, the SCNAs can be distinguished into clonal *vs.* subclonal alterations (22) and separately used by HAPTRIG to facilitate evolution-

associated functional interpretation of SCNAs.

One important resource used in HAPTRIG analysis is annotation of dosage-sensitive genes. Delaney *et al.* obtained a list of dosage-sensitive genes (e.g., haploinsufficient and haploproficient genes) from the yeast and mouse gene databases and used their human orthologs for the analysis. An additional resource, such as a list of haploinsufficient human genes curated by extensive text mining (23), can be also used or combined with the current list since the functional annotation of dosage-sensitive genes may be incomplete.

The role of autophagy in tumorigenesis is controversial and is thought to be context-dependent (24). The major finding of the study by Delaney *et al.* is that autophagy represents one of the major functions associated with chromosomal deletions in OV genomes. High HAPTRIG scores of autophagy-related pathways toward chromosomal deletions further indicate that genes annotated as haploinsufficient and the products of which frequently interact with other gene products are commonly observed in chromosomal deletions, including loci with single copy losses. It was previously shown that the number of dosage-sensitive genes associated with ubiquitination or proteasomal processes in yeast is high (25), but strong evidence for a relationship between autophagy and somatic SCNA was first demonstrated by Delaney *et al.* Those authors also revealed that SCNA-mediated disruption of proteostasis may be tumor type-specific by extending their analyses using a TCGA-PanCancer database encompassing >20 tumor types. The identification of key driver pathways such as autophagy can be turned into the development of therapeutic strategies with potential clinical relevance. The authors assumed that cancer cells that exhibit disruptions in autophagy may be sensitive to proteotoxic or autophagy-stressing drugs. To validate this assumption, they treated an OV cell line with chloroquine and nelfinavir to inhibit autophagy and induce ER stress, respectively. Interestingly, treatment of the cells with targeted agents such as rapamycin and dasatinib augmented the sensitivity of the cells to chloroquine and nelfinavir, highlighting the effect of synergistic drug combinations. Knockdown of potential target genes such as *LC3* and *BECN1*, as prioritized in HAPTRIG analysis, conferred sensitivity to cells undergoing autophagic stress, suggesting that targeted inhibition of these genes may also have clinical relevance in the treatment of disease.

Considering all of the abovementioned factors, the HAPTRIG method proposed by Delaney *et al.* can identify

the potential biological or clinical significance of extensive aneuploidy or frequent SCNAs in OV genomes. Genome-wide SCNA profiles gathered across hundreds of cancer genomes often seem random, but it is expected that some of them, if not all, may follow a nonrandom distribution suggesting that they are under selective pressure and represent potential cancer drivers. The functional interpretation of SCNAs is not easy, especially for OV genomes with prevalent SCNAs. However, incorporation of several sources of information, as done in HAPTRIG, may help to overcome this problem and facilitate the functional interpretation of SCNAs. HAPTRIG analysis revealed that autophagy may be one of the major cellular functions perturbed by recurrent SCNAs in OV genomes, and it also provided lines of evidence supporting the usefulness of inhibition of autophagy by chemical or targeted knockdown of *LC3* and *BECN1*. Their extensive analysis, which included PanCancer-scale HAPTRIG analysis of autophagy-related functions, provided additional insights into the roles of autophagy in different tumor types. We have witnessed an explosion in genomic data, especially that generated by new technologies including next generation sequencing. Unlike the genomics data generated from microarray-based platforms, the versatility of sequencing data facilitates its use for multiple purposes. For example, DNA sequencing data from whole-genome or -exome sequencing can be used to identify both somatic mutations and SCNAs. Thus, the recent explosion of DNA sequencing data on tumor genomes will lead to an enormous number of available cancer genome profiles with somatic mutations and SCNAs, for which the development of analytic methods geared toward advanced mechanistic insights into cancer genomes with potential clinical relevance will be the next hot issues in near future.

Acknowledgments

Funding: This study was supported by the National Research Foundation of Korea (2012R1A5A2047939) and the Basic Science Research Program through the National Research Foundation of Korea, funded by the Ministry of Science, ICT & Future Planning (2016M3C1B6929313).

Footnote

Provenance and Peer Review: This article was commissioned and reviewed by the Section Editor Zheng Li (Department of Gynecologic Oncology, The Third Affiliated Hospital

of Kunming Medical University (Yunnan Tumor Hospital), Kunming, China).

Conflicts of Interest: Both authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/tcr.2017.05.03>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Garraway LA, Lander ES. Lessons from the cancer genome. *Cell* 2013;153:17-37.
2. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011;144:646-74.
3. Albertson DG, Collins C, McCormick F, et al. Chromosome aberrations in solid tumors. *Nat Genet* 2003;34:369-76.
4. Stephens PJ, Greenman CD, Fu B, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* 2011;144:27-40.
5. Baca SC, Prandi D, Lawrence MS, et al. Punctuated evolution of prostate cancer genomes. *Cell* 2013;153:666-77.
6. Beroukhi R, Mermel CH, Porter D, et al. The landscape of somatic copy-number alteration across human cancers. *Nature* 2010;463:899-905.
7. Beroukhi R, Getz G, Nghiemphu L, et al. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc Natl Acad Sci U S A* 2007;104:20007-12.
8. Lawrence MS, Stojanov P, Mermel CH, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 2014;505:495-501.
9. Bignell GR, Greenman CD, Davies H, et al. Signatures

- of mutation and selection in the cancer genome. *Nature* 2010;463:893-8.
10. Cho A, Shim JE, Kim E, et al. MUFFINN: cancer gene discovery via network analysis of somatic mutation data. *Genome Biol* 2016;17:129.
 11. Torkamani A, Schork NJ. Identification of rare cancer driver mutations by network reconstruction. *Genome Res* 2009;19:1570-8.
 12. Chang MT, Asthana S, Gao SP, et al. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nat Biotechnol* 2016;34:155-63.
 13. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009;4:1073-81.
 14. Ciriello G, Miller ML, Aksoy BA, et al. Emerging landscape of oncogenic signatures across human cancers. *Nat Genet* 2013;45:1127-33.
 15. Delaney JR, Patel CB, Willis KM, et al. Haploinsufficiency networks identify targetable patterns of allelic deficiency in low mutation ovarian cancer. *Nat Commun* 2017;8:14423.
 16. The Cancer Genome Atlas consortium. Integrated genomic analyses of ovarian carcinoma. *Nature* 2011;474:609-15.
 17. Chatr-Aryamontri A, Oughtred R, Boucher L, et al. The BioGRID interaction database: 2017 update. *Nucleic Acids Res* 2017;45:D369-79.
 18. Klingstrom T, Plewczynski D. Protein-protein interaction and pathway databases, a graphical review. *Brief Bioinform* 2011;12:702-13.
 19. Leiserson MD, Vandin F, Wu HT, et al. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat Genet* 2015;47:106-14.
 20. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545-50.
 21. Tarca AL, Draghici S, Khatri P, et al. A novel signaling pathway impact analysis. *Bioinformatics* 2009;25:75-82.
 22. Carter SL, Cibulskis K, Helman E, et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol* 2012;30:413-21.
 23. Dang VT, Kassahn KS, Marcos AE, et al. Identification of human haploinsufficient genes and their genomic proximity to segmental duplications. *Eur J Hum Genet* 2008;16:1350-7.
 24. Mathew R, Karantza-Wadsworth V, White E. Role of autophagy in cancer. *Nat Rev Cancer* 2007;7:961-7.
 25. Delneri D, Hoyle DC, Gkargkas K, et al. Identification and characterization of high-flux-control genes of yeast through competition analyses in continuous cultures. *Nat Genet* 2008;40:113-7.

Cite this article as: Jung Y, Kim TM. Aneuploidy meets network analysis: leveraging copy number alterations to identify molecular pathways disrupted in high-grade serous ovarian carcinomas. *Transl Cancer Res* 2017;6(Suppl 3):S564-S568. doi: 10.21037/tcr.2017.05.03