Journal of
Data Mining in Genomics & Proteomics

Weier, J Data Mining Genomics Proteomics 2012, 3:3
DOI: 10.4172/2153-0602.1000e101

Editorial Open Access

Rags to Riches: Re-viewing the Tumor-specific Expression of Long Non-coding DNA and Satellite DNA Repeats in Humans in the Era of Next Generation Sequencing

Heinz-Ulli G. Weier*

Department of Cancer & DNA Damage Responses, Life Sciences Division, University of California-LBNL, Berkeley, USA

Research into the pathogenesis and progression of endocrine tumors has advanced in many ways in the last two decades. Yet, we still face major roadblocks in our attempts to find a cure for the most deadly disease, i.e., Cancer. Is it possible that research has been studying less significant parts of an equation which tips the tender equilibrium of balanced 'normal' protein synthesis and post-transcriptional modifications versus irreversible malignant changes, the latter driving the progression normal epithelium to preneoplastic lesions and ultimately to aggressively growing, invading and metastasizing carcinomas, while missing the most important drivers of malignant transformation? However, due to major advances in high throughput, next generation sequencing of DNA and RNA, the relevant epi-/genetic changes and their fundamental functional consequences have moved into the spotlight of temporary cancer research.

For example, with more than 200,000 new cases diagnosed every year, breast cancer remains the most frequent cancer in women in the U.S. Although breast cancer-related deaths have decreased over the last decade, this dreaded disease is the second most common cause or cancer-related deaths in women. Breast cancer has an unpredictable course and even after removal of the primary tumor, the risk of metastasis continues for 20 years or more [1]. Many patients would greatly benefit from detecting the disease in its earliest stages, when chances for a complete cure are still high.

Breast cancer is a complex disease and the cancer cells often show alteration in pathways ranging from signal transduction to DNA repair, drug response and apoptosis to survival in nutrient or oxygen deficient environments [2-7]. Our understanding of the disease seems very limited considering cofounding factors such as ethnicity in the age at onset or diagnosis of breast cancer [8]. Breast cancers progress through accumulation of genomic aberrations that enable development of cancer patho-physiological changes such as unlimited growth and metastasis.

Accumulated evidence has demonstrated that breast cancer is a complex and intrinsically heterogeneous disease in which patients may exhibit similar symptoms, but appear to have the same disease phenotype, for entirely different genetic reasons. Most published studies of breast cancer tumorigenesis have focused on the role of protein-coding genes during the onset and progression of the disease [9-12]. However, the role of genomic DNA sequences not coding for proteins, which may make up more than 80% of the human transcriptome, still needs to be determined.

Thus, an investigation of the role of large intergenic non-coding RNAs (lincRNAs; also known as 'long non-coding RNAs' or 'lncRNAs') [13,14] seems warranted, and an assessment of lincRNAs as potential biomarkers for the early detection of breast cancer as well as other malignant neoplasms such as prostate or thyroid may be in reach given the recent developments in genome research.

Recent discoveries showed thousands of DNA sequences in the human genome potentially coding for lincRNAs with individual sizes ranging from a few hundred to more than a hundred kb [14]. The expression of lincRNAs is strikingly tissue-specific and they are typically co-expressed with neighboring genes. It is well documented that lincRNAs play key roles in diverse biological processes such as gene dosage compensation, imprinting, chromatin remodeling, mRNA splicing and tumor metastasis [15-19]. For example, overexpression of the lincRNA 'HOTAIR' predicts tumor recurrence in hepatocellular cancer, and in breast cancer, it has been shown to remodel the chromatin state to promote cancer metastasis [15]. In prostate cancer, a recent analysis of RNA sequencing (RNAseq) data identified several non-coding RNA species associated with disease progression and unfavorable prognosis [19].

While the expression of particular lincRNA species change as normal tissues undergo malignant transformation, detailed knowledge might allow us to define biomarkers for early detection of cancer and, furthermore, design sensitive tests to predict the course of preneoplastic lesions or early stage tumors.

But, in the human genome, there are also vast regions comprised of short, simple DNA repeats, termed 'DNA satellites' due to their appearances as extra peaks in assays fractionating human genomic DNA. This type of DNA has long been thought to be a type of passenger or 'junk' DNA accumulated during the evolution of the human genome.

For a long time, our lab has been interested in highly repeated satellite DNA for its use as chromosome-specific DNA probes [20,21]. While evaluating cloned DNA probes, we typically found probes which lead to an unacceptable level of cross-hybridization. Some of them, however, showed a striking pattern of binding to chromosomes in meiosis [22]. Emphasizing the role of expressed satellite DNA, a recent RNAseq study using nexgen technology and deep sequencing detected transcripts from highly repeated DNA satellite sequences, and was able to demonstrate abnormal expression of satellite DNA sequences in human tumors [23].

Thus, no longer should non-coding DNA be considered a useless bystander or 'junk'. The studies of lincRNAs and differentially expressed

***Corresponding author:** Heinz-Ulli G. Weier, Department of Cancer & DNA Damage Responses, Life Sciences Division, E.O. Lawrence Berkeley National Lab, 1 Cyclotron Road, MS 977, Berkeley, CA 94720, USA, Tel: +1-510-486-5363; Fax: +1-510-486-5343; E-mail: ugweier@lbl.gov

Received December 21, 2012; **Accepted** December 21, 2012; **Published** December 28, 2012

Citation: Weier HUG (2012) Rags to Riches: Re-viewing the Tumor-specific Expression of Long Non-coding DNA and Satellite DNA Repeats in Humans in the Era of Next Generation Sequencing. J Data Mining Genomics Proteomics 3:e101. doi:10.4172/2153-0602.1000e101

satellite DNA sequences mentioned above ought to be considered to be even less than 'the tip of the iceberg'. Thorough cell biology research, bioinformatics and data mining of genome as well as transcriptome databases will provide much needed insight into the function of non-coding DNA in mammalian genomes, and that of breast and other endocrine tumor, in particular.

Acknowledgement

The skillful assistance of guests and staff of the Weier laboratory, LBNL, is gratefully acknowledged. This work was supported in parts by NIH grants CA123370, CA132815 and CA136685 carried out at the Earnest Orlando Lawrence Berkeley National Laboratory under contract DE-AC02-05CH11231.

References

1. Keys HM, Bakemeier RF, Savlov ED (1987) Breast Cancer, In Clinical Oncology, Rubin P (Ed), American Cancer Society: Rochester, New York.
2. Wotiz HH, Chattoraj SC, Kudisch M, Muller RE (1978) Impeding estrogens and the etiology of breast cancer. Cancer Res 38: 4012-4020.
3. Dickson RB, Lippman ME, Slamon D (1990) UCLA colloquium. New insights into breast cancer: the molecular biochemical and cellular biology of breast cancer. Cancer Res 50: 4446-4447.
4. Balz LM, Bartkowaik K, Andreas A, Pantel K, Niggemann B et al. (2012) The interplay of HER2/HER3/PI3K and EGFR/HER2/PLC-gamma1 signalling in breast cancer cell migration and dissemination. J Pathol 227: 234-244.
5. Helms MW, Kemming D, Contag CH, Pospisil H, Bartkowiak K, et al. (2009) TOB1 is regulated by EGF-dependent HER2 and EGFR signaling, is highly phosphorylated, and indicates poor prognosis in node-negative breast cancer. Cancer Res 69: 5049-5056.
6. Agelopoulos K, Buerger H, Brandt B (2008) Allelic imbalances of the egfr gene as key events in breast cancer progression--the concept of committed progenitor cells. Curr Cancer Drug Targets 8: 431-445.
7. Agelopoulos K, Kersting C, Korsching E, Schmidt H, Kuijper A, et al. (2007) Egfr amplification specific gene expression in phyllodes tumours of the breast. Cell Oncol 29: 443-451.
8. Amend K, Hicks D, Ambrosone CB (2006) Breast cancer in African-American women: differences in tumor biology from European-American women. Cancer Res 66: 8327-8830.
9. Stemke-Hale K, Gonzalez-Angulo AM, Lluch A, Neve RM, Kuo WL, et al. (2008) An integrative genomic and proteomic analysis of PIK3CA, PTEN, and AKT mutations in breast cancer. Cancer Res 68: 6084-6091.
10. Bose R, Kavuri SM, Searleman AC, Shen W, Shen D, et al. (2012) Activating HER2 Mutations in HER2 Gene Amplification Negative Breast Cancer. Cancer Discov.
11. Weitzel JN, Clague J, Martir-Negron A, Ogaz R, Herzog J, et al. (2012) Prevalence and Type of BRCA Mutations in Hispanics Undergoing Genetic Cancer Risk Assessment in the Southwestern United States: A Report From the Clinical Cancer Genetics Community Research Network. J Clin Oncol.
12. Shannon KM, Chittenden A (2012) Genetic testing by cancer site: breast. Cancer J 18: 310-319.
13. Guttman M, Amit I, Garber M, French C, Lin MF et al. (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 458: 223-227.
14. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, et al. (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev 25: 1915-1927.
15. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, et al. (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 464: 1071-1076.
16. Yang Z, Zhou L, Wu LM, Lai MC, Xie HY et al. (2011) Overexpression of long non-coding RNA HOTAIR predicts tumor recurrence in hepatocellular carcinoma patients following liver transplantation. Ann Surg Oncol 18: 1243-1250.
17. Niland CN, Merry CR, Khalil AM (2012) Emerging Roles for Long Non-Coding RNAs in Cancer and Neurological Disorders. Front Genet 3: 25.
18. Cui Z, Ren S, Lu J, Wang F, Xu W, et al. (2012) The prostate cancer-up-regulated long noncoding RNA PlncRNA-1 modulates apoptosis and proliferation through reciprocal regulation of androgen receptor. Urol Oncol.
19. Ren S, Peng Z, Mao JH, Yu Y, Yin C, et al. (2012) RNA-seq analysis of prostate cancer in the Chinese population identifies recurrent gene fusions, cancer-associated long noncoding RNAs and aberrant alternative splicings. Cell Res 22: 806-821.
20. Weier HU, Kleine HD, Gray JW (1991) Labeling of the centromeric region on human chromosome 8 by in situ hybridization. Hum Genet 87: 489-494.
21. Weier HU, Gray JW (1992) A degenerate alpha satellite probe, detecting a centromeric deletion on chromosome 21 in an apparently normal human male, shows limitations of the use of satellite DNA probes for interphase ploidy analysis. Anal Cell Pathol 4: 81-86.
22. Dozortsev D, Coleman A, Nagy P, Diamond MP, Ermilov A, et al. (2000) Nucleoli in a pronuclei-stage mouse embryo are represented by major satellite DNA of interconnecting chromosomes. Fertil Steril 73: 366-371.
23. Ting DT, Lipson D, Paul S, Brannigan BW, Akhavanfard S, et al. (2011) Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. Science 331: 593-596.