Journal of Mucosal Immunology Research

Nishimura, et al. J Mucosal Immunol Res 2020, 4:2

Short Communication Open Access

Data-driven Protein Networks of Early Lung Adenocarcinomas Harboring EGFR Mutations

Toshihide Nishimura[12*] and György Marko-Varga[3,4]

1Department of Translational Medicine Informatics, St. Marianna University School of Medicine, Kanagawa, Japan

2Department of Chest Surgery, St. Marianna University School of Medicine, Kanagawa, Japan

3Department of Biomedical Engineering, Lund University, Lund, Sweden

4Department of Translational Medicine, Lund University, Skane University Hospital Malmo, Malmo, Sweden

[*]**Corresponding author:** Dr. Toshihide Nishimura, Department of Translational Medicine Informatics, St. Marianna University School of Medicine, Kanagawa, Japan, E-mail: t-nisimura@marianna-u.ac.jp

Received date: August 28, 2020; **Accepted date:** September 14, 2020; **Published date:** September 21, 2020

Abstract

It remains unclear how epidermal growth factor receptor EGFR major driver mutations (L858R or Ex19del) affect downstream molecular networks and pathways, which would influence treatment outcomes of NSCLC patients. Our studies were conducted to unveil the influences of these mutations by assessing 36 tissue specimens of lung adenocarcinoma (Ex19del, nine; L858R, nine; no Ex19del/L858R, 18). Weighted correlation network analysis, together with analysis of the variance-based or over-representative test, identified core co-expressed modules and their hub proteins. Data-driven molecular networks obtained by weighted correlation analysis will put an important foundation on biological and medical research. Network proteogenomic approaches will change the treatment system and will help identify potential therapeutic targets and develop therapeutic strategies to improve patient outcomes.

Keywords: Lung adenocarcinoma; EGFR mutation; Data-driven network; Weighted correlation network analysis; Network proteogenomics; Mass spectrometry

Introduction

Lung cancer is the leading cause of cancer death worldwide and is characterized by multiple histologic types. Adenocarcinomas are most abundant among non-small cell lung cancer (NSCLC) subtypes. The new pathologic classification of lung adenocarcinoma [1] categorized invasive adenocarcinomas into nine subtypes according to the predominant histologic pattern: lepidic (LPA), acinar, papillary (PAP), micropapillary (MP), solid (SLD), and three variants. NSCLC subsets have been defined by recurrent driver mutations that occur in multiple oncogenes, including AKT1, ALK, BRAF, EGFR, HER2, KRAS, MEK1, MET, NRAS, PIK3CA, RET, and ROS1. Driver oncogenic mutations lead to constitutive activation of mutant signaling proteins that induce and sustain tumorigenesis. Clinical proteogenomics had already begun with the discovery that lung adenocarcinoma patients with mutated epidermal growth factor receptor (EGFR) received great benefit from EGFR-tyrosine kinase inhibitors (EGFR-TKIs). Osimertinib was recently recommended as first-line treatment for patients with EGFR-mutant NSCLC according to the FLAURA trial that reported significantly better PFS and OS with osimertinib than with first-generation EGFR-TKIs (gefitinib or erlotinib) [2, 3]. Precision medicine for lung cancer has been urgently demanded to help improve patient outcomes, including various tyrosine kinase inhibitors (TKIs) and immune checkpoint inhibitors, together with biomarkers to guide treatment strategies, detecting early stages of the disease, verifying status/stage, and response to treatment.

Somatic alterations, mutations, and fusions in lung cancer frequently affect cellular pathway activities involved in lung cancer subtypes [4]. Therefore, a pivotal challenge is to understand how the major driver mutations such as EGFR L858R and Ex19del affect disease-related downstream networks together with other upstream driver mutations crosstalk, which plays central roles in the context of lung cancer progression, malignancy, and outcome and/or resistance regarding TKI therapies.

Data-driven Protein Networks

Identification of data-driven disease-related molecular networks and their core molecules themselves are more important rather than existing pathways to unveil respective disease mechanisms, whereas GO annotations with existing pathways provide some functional pieces of knowledge. Recent advances in high-accuracy mass spectrometry (MS) have made proteomics more compatible with shotgun sequencing and quantitative analysis of disease-related proteins expressed in clinical specimens. Proteomic expression data obtained from such analyses can be used to extract key disease-related proteins and identify novel biomarkers and therapeutic targets. A laser microdissection (LMD) technique enables the collection of target cells of a certain type from sections of formalin-fixed paraffin-embedded (FFPE) cancer tissue.

Today, proteomics is strongly powered by mass spectrometry (MS) based methodologies, largely due to that modern mass spectrometers offer high mass resolution and accuracy required for correct protein identification, which potentially multiplies the sequencing speed above 100 Hz for single run proteome analysis at a depth of 6,000 proteins with no loss in sensitivity. Besides, a novel liquid chromatography (LC) system is capable of processing more than 200 samples per day. Those provide a novel, high-performance MS instrumentation equipped with high throughput and robust LC separation required in clinical MS-based proteomic analyses of a large cohort of clinical specimens.

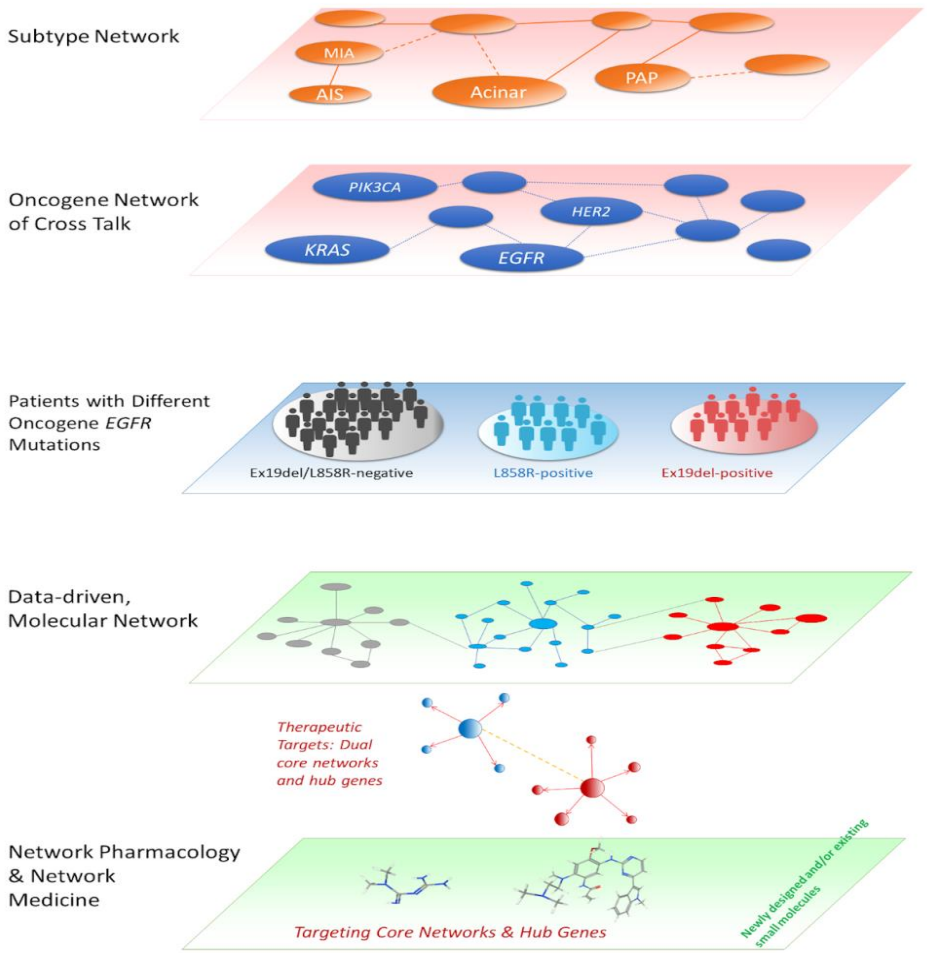

Figure 1: An illustration of disease-related, data-driven, networks-based challenge against lung adenocarcinoma subtypes. Lung adenocarcinomas are breakdown to key networks at various levels, such as subtype network, oncogene network in which diver oncogenes can interact mutually via cross-talk in some cases, and which affect subsequent cellular functional protein interaction networks via activation or inactivation of core module networks, resulting in aberrant signal transductions in a cancer cell. Identification of key networks significantly associated with a disease will help to develop effective therapeutic strategies and discover promising therapeutic targets and drug candidates that are in line with precision medicine.

Figure 1 illustrates our challenge against lung adenocarcinoma subtypes via network proteogenomics, in which identification of disease-related, data-driven networks is crucial. We have conducted an MS-based proteomic analysis of histologically well-defined lung adenocarcinomas (Ex19del, nine; L858R, nine; no Ex19del/L858R, 18) [5]. Weighted correlation network analysis, a more general term for weighted gene co-expression analysis (WGCNA), together with analysis of variance-based screening identified more than ten core-network modules and their hubproteins [5]. Pathways under the *EGFR* Ex19del mutation were found significantly to involve SUMOylation, epithelial and mesenchymal transition (EMT), ERK/mitogen-activated protein kinase signalling via phosphorylation and Hippo signalling, whereas those under the *EGFR L858R* mutation were characteristically related to cancer cell survival and death, including *VEGFA*. It was interestingly suggested that the PD-1/PD- L1 cancer immunotherapy pathway was not activated but the B7/CD28 immune-checkpoint pathways were activated, which is

consistent with clinical findings that most patients are in early-stage lung adenocarcinomas. The therapeutic effects of EGFR-TKIs in *EGFR* mutation-positive NSCLC were found to differ depending on the mutation subtype (Ex19del or L858R), regardless of the EGFR-TKI therapies received (osimertinib and standard TKIs) [2]. Our results might support that L858R-positive NSCLC harbors many miscellaneous compound mutations other than EGFR mutation [6], whereas Ex19del-positive NSCLC is more likely to be relatively pure concerning the oncogene mutation that drives proliferation and is mainly dependent on the EGFR pathway, which would result in a long PFS with EGFR-TKIs.

The first study of mutant proteomic analysis has been performed for clinical tissue specimens obtained from patients of lung adenocarcinoma with *EGFR* oncogenic driver mutations [7]. Surprisingly, the orthogonal partial least square-discriminant analysis (OPLS-DA) revealed the profound differences among the profiles of mutant proteins identified under the different EGFR mutation statuses, that was never seen before. This observation might evidence that cancer cells harboring L858R or Ex19del emerge from cellular origins different from L858R/Ex19del-negative cells. The weighted correlation network analysis of mutant proteomes, screened by over representative test, has shed new light on network dynamics of early lung adenocarcinomas with and without the mutated EGFR gatekeeper driver oncogenes. Upstream regulators and causal networks predicted using Ingenuity Pathway Analysis (IPA, http://www.ingenuity.com) software for disease-related mutant protein networks also confirmed a close link to *EGFR* mutation-positive cancers, mainly NSCLC.

Network Proteogenomics

Targeting disease-associated dual core-networks rather than targeting a single protein (gene) as in conventional approaches will greatly improve the outcomes of individual patients such as efficacy and safety. Hopkins [8] first proposed such a concept which aims to induce synthetic lethality by targeting dual hub molecules involved in different disease core networks (network pharmacology). Barabási, one of network medicine's pioneers stated [9] that "It is only a matter of time until these advances will start to affect medical practice as well, marking the emergence of a new field that may be aptly called network medicine." It will be then crucial to firstly capture data-driven molecular networks. In the short future, a new network science will rigorously integrate both proteomes and genomes of a disease subtype. Data-driven molecular networks will give one of the most important foundations in life science. Network proteogenomic approaches will make a big paradigm shift in biological and medical research and will change the treatment system. As an example, Connectivity MAP (CMAP) (https://portals.broadinstitute.org/cmap) is creating a genome-scale library of cellular signatures that catalogues transcriptional responses to chemical, genetic, and disease perturbation, which currently contains more than 1 Million gene profiles representing more than 1,300 compounds resulting from perturbations of multiple cell types [10]. CMAP can be utilized to predict existing small-molecule drugs for a specific disease by evaluating the correlation between those drugs and core network modules/hub genes. Our studies will be extended firstly to a large cohort, whereas the number of patients to be examined is often limited due to a difficulty in collecting a large number of homogeneous tumor-derived samples with well-defined histologies. Secondly, disease-related genomic expression and alteration analysis will be conducted simultaneously for the same samples.

References

1. Travis WD, Brambilla E, Riely GJ (2013)New pathologic classification of lung cancer: relevance for clinical practice and clinical trials. J Clin Oncol 10: 3992-1001.
2. Soria JC, Ohe Y, Vansteenkiste J, Reungwetwattana T, Chewaskulyong B, et al. (2018) Osimertinib in untreated EGFR-mutated advanced non-small-cell lung cancer. N Engl J Med 378: 113-125.
3. Ramalingam SS, Gray JE, Ohe Y, Cho BC, Zhou C, et al. (2019) Osimertinib vs comparator EGFR-TKI as first-line treatment for EGFRm advanced NSCLC (FLAURA): Final overall survival analysis. Ann Oncol 30: v914-915.
4. Shtivelman E, Hensing T, Simon GR, Dennis PA, Otterson GA, et al. (2014) Molecular pathways and therapeutic targets in lung cancer. Oncotarget 5: 1392-1433.
5. Nishimura T, Nakamura H, Yachie A, Hase T, Fujii K, et al. (2020) Disease-related cellular protein networks differentially affected under different EGFR mutations in lung adenocarcinoma. Sci Rep 10: 10881.
6. Kohsaka S, Nagano M, Ueno T, Suehara Y, Hayashi Y, et al. (2017) A method of high-throughput functional evaluation of EGFR gene variants of unknown significance in cancer. Sci Transl Med 9: eaan6566.
7. Nishimura T, Végvári Á, Nakamura H, Kato H, Saji H (2020) Mutant Proteomics of Lung Adenocarcinomas Harboring Different EGFR Mutations. Front Oncol 10:1494.
8. Hopkins AL. (2008) Network pharmacology: the next paradigm in drug discovery. Nature Chem Biol 4: 682-690.
9. Barabási A-L (2007) Network Medicine — From Obesity to the "Diseasome". N Engl J Med 357: 404-407.
10. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, et al. (2017) A Next Generation Connectivity Map: L1000 Platform And The First 1,000,000 Profiles Cell 171: 1437-1452.