Searching For Essential Cancer Proteins: Analysis Of Hypomutated Genes In Skin Melanoma

Mikhail Pyatnitskiy

Orekhovich Institute of Biomedical Chemistry, 119121, Moscow, Pogodinskaya str, 10,

Pirogov Russian National Research Medical University, 117997, Moscow, Russia,

mpyat@bioinformatics.ru

Dmitriy Karpov

Orekhovich Institute of Biomedical Chemistry, 119121, Moscow, Pogodinskaya str, 10,

Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, 119991, Moscow, Russia,

aleom@yandex.ru

Ekaterina Poverennaya

Orekhovich Institute of Biomedical Chemistry, 119121, Moscow, Pogodinskaya str, 10,
k.poverennaya@gmail.com

Andrey Lisitsa

Orekhovich Institute of Biomedical Chemistry, 119121, Moscow, Pogodinskaya str, 10,
andrey.lisitsa@ibmc.msk.ru

Sergei Moshkovskii

Orekhovich Institute of Biomedical Chemistry, 119121, Moscow, Pogodinskaya str, 10,

Pirogov Russian National Research Medical University, 117997, Moscow, Russia,

smosh@mail.ru

Results of cancer genome projects had literally revolutionized our understanding of cancer molecular biology and are especially promising for personalized anti-tumor therapy. Analysis of mutational landscape makes possible to disclose driver genes which provide tumor development and progression. Amino acid sequences of such driver proteins are affected by positive selection. Necessity of oncogenes and tumor suppressors to be mutated in most cancers inspires researchers to implement scoring systems to select hypermutated driver genes [1].

In contrast to these useful efforts, instead we focused on proteins whose corresponding genes are hypomutated in tumor, i.e. experience negative selection during cancer evolution and hence can be regarded as essential for cancer cell survival. We draw an analogy between

essential cancer proteins and well-known Abraham Wald's work on estimating the plane critical areas using data on survivability of aircraft encountering enemy fire. Wald reasoned that parts hit least on the returned planes are critical and should be protected more. Similarly we propose that genes essential for tumor cell should carry less high-impact mutations in cancer compared to polymorphisms found in normal cells (Figure 1).

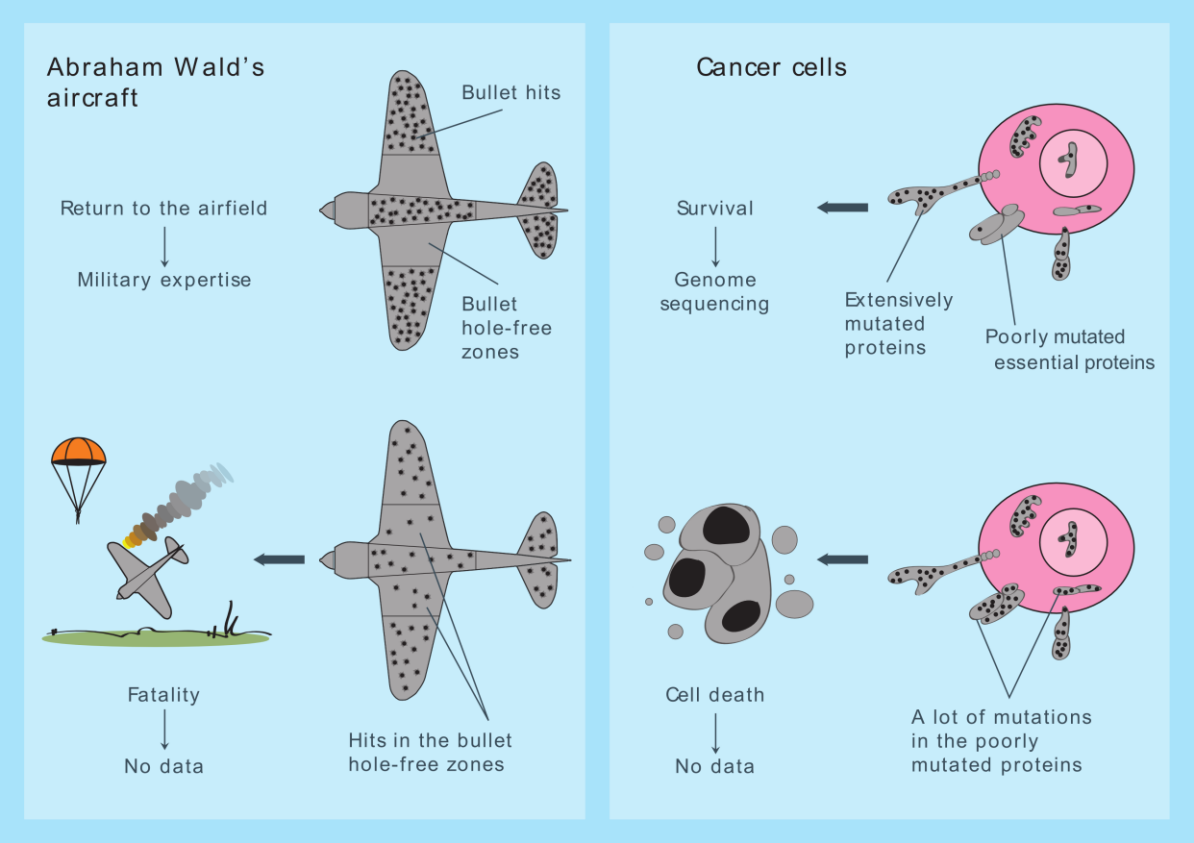

Figure 1. Analogy between bullet-free plane critical areas and hypomutated proteins essential for cancer.

We used data on mutations from the Cancer Genome Atlas and polymorphisms found in healthy humans (from 1000 Genomes Project, 1KG) to predict protein-coding genes essential for skin cutaneous melanoma (SKCM). These genes were selected according to several criteria including negative selection, expression in melanocytes and decrease in the proportion of high-impact mutations in cancer compared with normal cells.

In order to find genes experiencing negative selection during cancer evolution we used d_N/d_S

as an indicator of selective pressure [2] and chose d_N/d_S value=0.25 as a reasonable threshold. Next we removed genes with expression levels in the lowest 20% which can be regarded as non-expressing in melanoma cells. Finally we selected genes which are depleted by functionally important variants in cancer as compared to normal tissues, i.e. genes where fraction of mutations with high functional impact f for 1KG data was greater than the same fraction for cancer data, $f_{1KG} > f_{SKCM}$. Functional effect of non-synonymous substitutions was predicted via dbNSFP database [3]. After all steps we obtained 91 protein-coding genes designated hereinafter as "essential cancer proteins".

We classified essential proteins into several categories. Most represented categories included membrane transport proteins, such as ion channels and solute carriers, neural proteins of various functions, cell adhesion molecules, etc. In order to describe the resulting list more formally we performed functional enrichment analysis by Gene Ontology categories. We found significant trend towards membrane proteins, specifically, proteins of plasma membrane and cell periphery. We speculate that this could be a sign of immune system-driven negative selection of cancer neo-antigens. Another finding is significant overrepresentation of semaphorin receptors, which can mediate distinctive signalling cascades and are involved in various aspects of tumor development. Cytokine receptors CCR5 and CXCR1 were also identified as cancer essential proteins and this finding is confirmed by other studies.

Overall our goal was to illustrate the idea of detecting proteins whose sequence integrity and functioning is important for cancer cell survival. Hopefully, this prediction of essential cancer proteins may point to new targets for anti-tumor therapies.

1. M.S.Lawrence et al. (2014) Discovery and saturation analysis of cancer genes across 21 tumour types, *Nature*, **505**:495–501.

2. Z.Yang, J.P.Bielawski (2000) Statistical methods for detecting molecular adaptation. *Trends Ecol Evol*, **12**:496–503.

3. X.Liu et al (2013). dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Hum Mutat,* **34**;9:E2393–402.