Póster presentado en el Joint Meeting of XV International Conference on Intelligent Systems for Molecular Biology & VI European Conference on Computational Biology. Viena, 22-25 July 2007

A network of protein domains that are present in chimeric tyrosine-kinase proteins in cancer

I. Ortiz de Mendibil, J. L. Vizmanos, F. J. Novo
Department of Genetics
University of Navarra, Spain

motivation

Chromosomal translocations resulting in fusion genes that code for chimeric proteins is a very common feature of cancer. The presence of fusion proteins with tyrosine-kinase (TK) activity, in particular, has been implicated in many different types of tumours. These fusion TKs are generally made up by an oligomerization domain (encoded by one gene) together with a TK domain (encoded by the fusion partner gene). In this work, we have created a network of protein domains present in chimeric TK proteins in cancer.

material and methods

We have recently created **TICdb** (*http://www.unav.es/genetica/TICdb/*), a collection of gene-mapped translocation breakpoints in human tumors. Using the information contained in that database, it is possible to build a network of genes involved in oncogenic translocations that create TK fusion proteins (TK gene network, Figure 1).

Fig 1: Network of fused genes

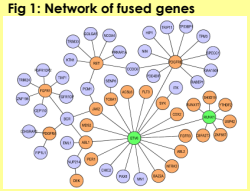
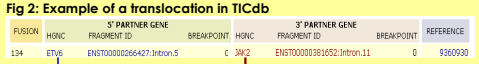


Fig 2: Example of a translocation in TICdb



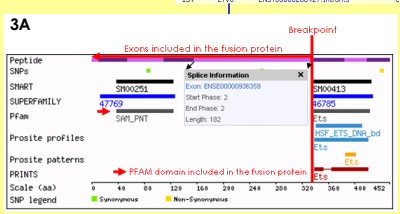
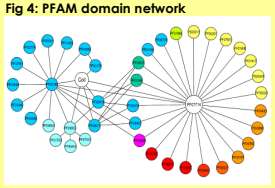


Fig 4: PFAM domain network



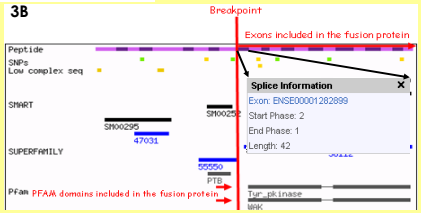


Fig 3A and 3B: "Protein Information View" from Ensembl. As we know which exons are included in the fusion gene, we can collect the PFAM domains present in the fusion protein.

Then, for each chromosomal translocation involving a gene in the TK gene network, we obtained from Ensembl the accession number of PFAM domains coded by the exons, from both partner genes, that are present in the fusion gene. This enabled us to record which PFAM domains are present in the same fusion protein (Figures 2, 3A and 3B). This information was then processed using Cytoscape 2.4.0 in order to construct a network of PFAM domains present in the same chimeric protein. We also recorded the reading frame of the exons flanking each translocation breakpoint, in order to verify whether fusion genes keep an intact reading frame.

results and discussion

The resulting network of PFAM domains is represented in Figure 5B and contains 43 nodes and 57 edges. The topology is clearly different to the original network of fused genes shown in Figure 5A, which is comprised by 58 nodes and 59 edges.

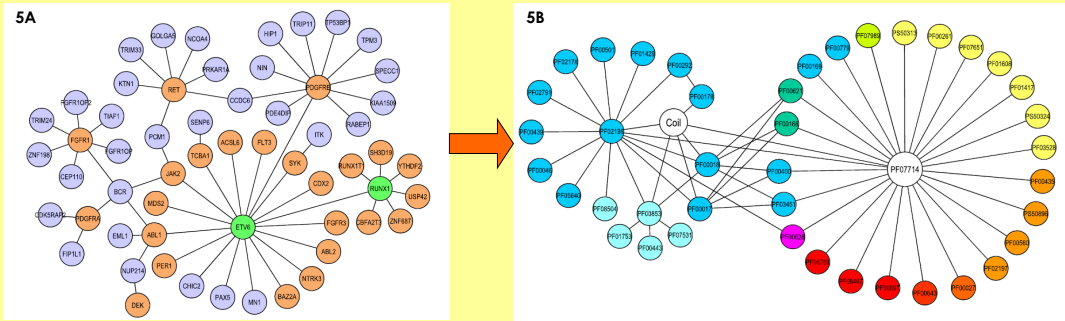


Figure 5. The TK gene network (panel A, left) and the corresponding protein domain network (panel B, right). The TK gene network includes 5 hubs (nodes with 5 or more edges), four of which correspond to known TK genes and the other is ETV6, a gene frequently rearranged in cancer. In the protein domain network, in contrast, we see two main hubs with more than 10 edges, which correspond to PF02198 (SAM_PNT domain of the ETV6 gene) and to PF07714 (the Tyr_pkinase domain present in all the TK genes of the network). Thus, transforming the gene fusion network into the protein domain network has not only reduced the complexity (grouping together domains that are present in different genes), but also identified the major mechanisms by which these translocations drive the oncogenic process.
In most cases the protein motif is completely preserved in the fusion protein, with a few exceptions in which the breakpoint fell within the motif. Additionally, only 9 fusions (6.5%) do not seem to keep an intact reading frame, underscoring that a complete fusion protein is necessary for cancer development.

acknowledgements

This work has been funded with the help of the Institute of Health Carlos III (FIS PI040037), the PIUNA Program of the University of Navarra and the Caja Navarra Foundation through the Program "You choose, you decide" (Project 10.830)

conclusion

The work presented here for the TK gene network is a proof of principle that the construction of a network of protein domains present in chimeric fusion proteins is a good tool to identify pathogenic mechanisms driving cancer:

- The combination of protein domains is not random, and reflects the selective pressures that favour chimeric proteins with oncogenic properties.
- The topology of the protein domain network shows which domains are more frequently involved in fusion proteins. Thus, the identification of potential pathogenic mechanisms is much more straightforward than using the gene fusion network.

In addition, the network of protein domains might be useful for the prediction of new partner genes fused to a known translocated gene, depending on the presence of specific protein domains similar to those already found in fusion proteins of the same gene. Furthermore, the pressure to maintain the reading frame and the integrity of the exons coding for the protein domains involved, might determine the location of the translocation breakpoint within specific genes.