Genomic-Scale Screening for Gene Fusions in Human Solid Tumors by Integrative Biomedical Informatics

Xiaosong Wang[1,2], Saravana M Dhanasekaran[1], John R. Presener[1], Bo Han[1], Nallasivam Palanisamy[1], Maureen A. Sartor[2], Gilbert S. Omenn[2] & Arul M. Chinnaiyan[1,2]

1. Michigan Center for Translational Pathology, 2. National Center for Integrative Biomedical Informatics, CCMB, University of Michigan, Ann Arbor, MI.

The recent discovery of recurrent fusions in prostate and lung cancers stimulated a search for gene fusions in solid tumors, and is expected to inspire drug developments. However, due to limited understanding of basic principles for the chromosome rearrangements, the identification of causal gene fusions from the background of non-specific chromosomal aberrations of solid tumor genomes remains challenging.

The common characteristics of fusion genes

Through integrative database mining, we analyzed the shared characteristics of established gene fusions in cancer [see http://portal.ncibi.org for tools and databases].

I. Fusion genes can be delineated by unbalanced breakpoints, which follow a stable genetic principle

Figure 1. The fusion breakpoint principle.
(a) Recurrent unbalanced TCF3-PBX1 fusion (n=17)
(b) Deletions generally remove the 3′ region of 5′fusion partners, and the 5′ region of 3′ partners. In contrast, amplifications generally amplify the 5′ region of 5′ partners, and the 3′ region of 3′ partners.
The left panel illustrates the application of our principle to inter-chromosomal translocations.
The right panel illustrates the inferred principle for intra-arm translocation.

II. Gene fusions often result in gain and loss of exonal expressions of 3′ and 5′ fusion partners

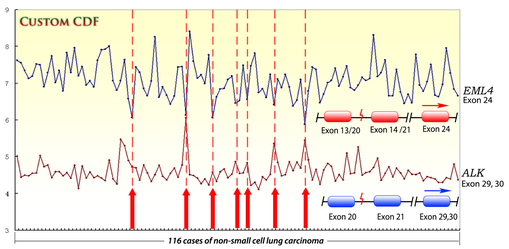

Figure 2. Synergistic loss of EML4 exon 24 and upregulation of ALK exon 29, 30 in 7/116 NSCLC patients. The exon expressions of EML4 and ALK are revealed by probe level analysis of the EXPO cancer gene expression data with custom CDF (www.ncibi.org). EML4-ALK fusion was reported in 6.7% of non-small cell lung cancer patients.

III. tumor specific genes are frequently involved in promoter -type fusions

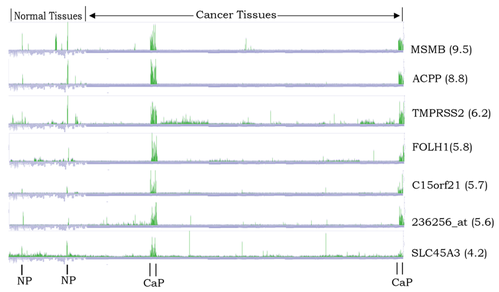

Figure 3. Identification of tumor specific genes in prostate cancer by tumor specific expression profile analysis.

This method identifies TMPRSS2, C15orf21, and SLC45A3 as top candidates, which were known ETS fusion partners

IV. The functional relevance of fusion genes in cancer can be quantified by concept signature score (ConSig score)

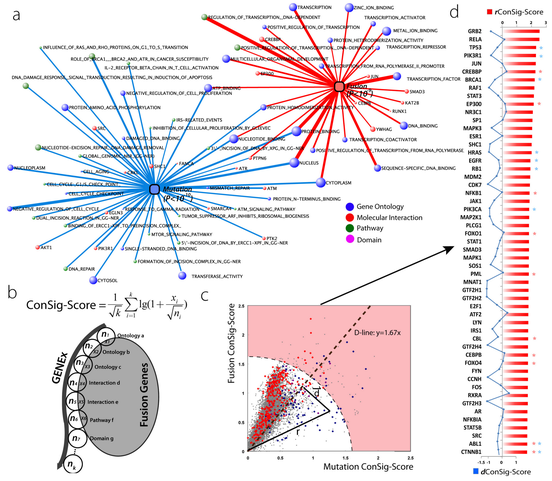

Figure 4. Molecular concept signatures and ConSig analysis for fusion and mutation genes.
(a) Enrichment analysis with the compendia of molecular concepts for fusion and mutation genes.
(b) The ConSig algorithm for Gene X based on fusion genes.
(c) Plotting the fusion and mutation ConSig-score against each other.
Note: r, rConSig-score; d, dConSig-score.
(d) Isolating the top 60 genes rated by rConSig-score.

The integrative model for new fusion discovery

Based on multiple evidence characterizing the fusion genes, we created an integrative translational bioinformatics model for the prediction of novel gene fusions in human solid tumors. Several comprehensive strategies were developed to address the possibility of recurrent fusions based on available evidence in different cancer types. We applied this approach to predict gene fusions based on public genomic, sequence and functional data, as well as our deep sequencing data.

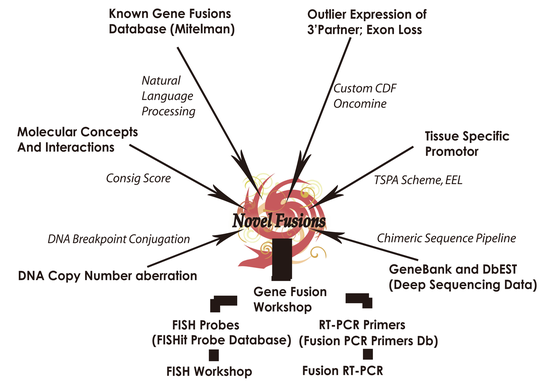

Figure 5: The integrative bioinformatics model for translating the high-throughput biological data into evidence of candidate gene fusions in cancer.

The output of the integrative model

An example of the integrative strategy -- new fusion discovery from the cancer expressed sequence tag data

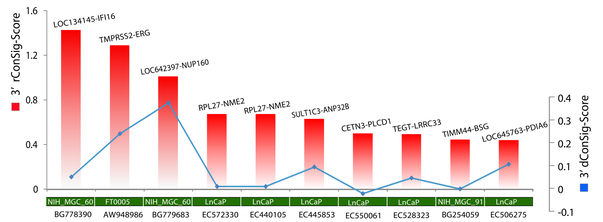

Figure 6. Ranking chimera ESTs from prostate cancer by rConSig-score of 3′ partner genes identifies known TMPRSS2-ERG fusion and novel RPL27-NME2 fusion chimeras

The computational results were validated by Regular or Quantitative Reverse Transcription PCR (qRT-PCR), Rapid Amplification of cDNA Ends (RACE) and Fluorescent in Situ Hybridization (FISH). The integrative approach was proved to be efficient. Up to 20-50% of candidates were validated by different assays.

Gene	Cancer Type	FISH/RT-PCR/RACE Results	Confirmation
Fusion 01	Breast cancer	FISH: Out of 49, 1×, 3×5′.	
Fusion 02	Prostate cancer	FISH: Out of 99, 5×.	
Fusion 03	Breast cancer	FISH: (-)	NA
Fusion 04	Prostate cancer	RACE: Fused to ETV1	
Fusion 05	Prostate cancer	FISH: Out of 81, 4×, 1×5′, 1×3′.	
Fusion 06	Prostate cancer	FISH: (-)	NA
Fusion 07	Prostate cancer	FISH: (-)	NA
Fusion 08	Breast cancer	FISH: (-)	NA
Fusion 09	Prostate cancer	FISH: (-)	NA
Fusion 10	Breast cancer	FISH: Out of 65, 2×3′.	NA
Fusion 11	Breast cancer	RT-PCR: MCF7(+)	
Fusion 12	Cervix cancer	RT-PCR: (-)	NA
Fusion 13	Prostate Cancer	RT-PCR: LnCaP(+)	
Fusion 14	Prostate Cancer	RT-PCR: DU145(+)	
Fusion 15	Prostate Cancer	qRT-PCR(+)	
Fusion 16	Prostate Cancer	qRT-PCR(+)	NA

Figure 7. The representative validation results of 16 fusion gene candidates by FISH, RACE, or RT-PCR. Represents confirmation on individual tissue sections; represents confirmation by sequencing of RT-PCR products. NA = incomplete.

This wok was supported by NIH grant #U54 DA021519