Meta-analysis of multiple microarray datasets reveals a novel genomic signature associated to invasive growth of epithelial cells and early breast cancer metastasis.

Lorenza D'Alessandro [(1)], Barbara Felice[(2)], Filippo Montemurro[(1)] and Enzo Medico [(1)]

[(1)] Institute for Cancer Research and Treatment, University of Torino Medical School, S.P.142, 10060 Candiolo (TO)
enzo.medico@ircc.it
[(2)] FIRC Institute of Molecular Oncology, via Adamello 16, 20139 Milano

Keywords. Microarray, meta-analysis, molecular signature

Introduction

HGF, also known as "Scatter Factor", is a mesenchymal cytokine that acts on epithelial and endothelial cells by promoting a highly integrated biological program, hereafter referred to as "invasive growth". This program involves coordinated control of basic cellular functions including dissociation and migration ("scattering"), invasion of extracellular matrix, proliferation, prevention of apoptosis and polarization [1]. As a consequence, complex developmental processes take place, such as branched morphogenesis of epithelia and angiogenesis. Oncogenic activation by overexpression or point mutation of the gene encoding the tyrosine kinase receptor for HGF, c-MET, is involved in the progression of tumors towards the invasive-metastatic phenotype [2]. To identify genes involved in Met-driven invasive growth, we explored the transcriptional response of mouse liver cells to HGF at different time points. Two different microarray platforms were adopted, consisting respectively of high-density spotted cDNAs (Incyte) and in-situ synthesized oligonucleotides (Affymetrix). Global exploration of 25'000 gene transcripts yielded over 1500 transcriptionally regulated sequences, corresponding to genes involved in the control of the basic biological functions underlying the invasive growth program: transcription, signal transduction, apoptosis, proliferation, cytoskeleton organization, motility and adhesion. Joint analysis of the data obtained by the two platforms allowed identification of genes with more consistent and reproducible regulation. Meta-analysis on genomic expression datasets obtained from breast carcinoma showed that expression of genes belonging to the HGF signature is correlated to cancer progression.

Results

1. Identification of HGF-regulated genes

We explored the transcriptional changes induced by HGF in the mouse embryo liver cell line MLP-29, that specifically responds to HGF by scattering and invasive growth [3]. Messenger RNA was prepared in two replicated experiments from cells that had been stimulated with HGF or with control medium, for 1h, 6h and 24h. To perform the widest transcriptome analysis, and for systematic data validation, the above-described set of mRNAs was used on two commercial microarray platforms: (1) Incyte Mouse Gem cDNA microarrays (8734 cDNAs); (2) four Affymetrix high-density oligonucleotide GeneChips (Mu11k A/B and MGU74 B/C), for a total of over 45'000 sequences explored. To evaluate redundancy, we matched the sequences to the Unigene database, defining a total of ~25'000 genes explored, of which over 6000 explored by both platforms, providing a wide

basis for systematic cross-validation. Statistical analysis and systematic cross-comparison allowed us to define a signature of 257 genes significantly regulated by HGF under both types of microarray platforms.

2. Meta-analysis on breast cancer microarray data

Having identified a 257-genes signature corresponding to genes regulated during HGF-induced invasive growth in mouse hepatocytes, we sought to verify if expression of those genes was associated a metastatic propensity of human cancer. We therefore downloaded the breast cancer dataset published by van 't Veer et al. [4]: http://www.rii.com/publications/2002/default.htm . In particular, we focused on microarray data on about 25000 genes obtained from (1) 34 patients that underwent surgery for breast cancer and showed a metastatic relapse within five years, and (2) 44 operated patients that showed a metastasis-free survival greater than 5 years. In the supplementary information we could also find sample features like: diameter of the cancer mass, the metastasis-free follow up time, the presence of metastasis within five years, tumor grade, angioinvasion, Erp and PRp, Lymphocytic Infiltrate and BRCA1 mutations. To intersect our mouse data with the human breast cancer dataset, we converted the mouse UNIGENE IDs of our HGF signature into the human UNIGENE IDs of the corresponding orthologues using the TIGR website, at the link: http://pga.tigr.org/tigr-scripts/magic/r1.pl . Using Excel, we intersected all the genes of the breast cancer dataset with the 257 HGF signature genes. We found that 211 genes of the signature were included in the breast cancer dataset. None of these was included in the 70-genes panel identified by van ‘t Veer and colleagues as correlated with poor prognosis.

3. Definition of a breast cancer prognostic signature

For each of the 211 genes of the HGF signature for which we found expression data in the breast cancer dataset, we calculated the Pearson correlation between gene expression and months of metastasis-free survival of the respective patient samples. We were able to select 20 genes with an individual correlation coefficient of at least +/- 0.3. We then calculated a single, virtual metagene by averaging the expression levels of the 20 genes (corrected for the positive/negative correlation). The correlation between the expression of this metagene and the months of metastasis-free survival was found to be of 0.7. When the 70-genes signature extracted by van’t Veer and colleagues was used to generate a metagene in the same manner, correlation of this metagene with metastasis-free survival was found to be of 0.63. Kaplan-Meier analysis conducted on the same data using the 20-genes HGF metagene as a predictor of good/intermediate/bad prognosis defined a subgroup of patients with 85% of probability to develop a metastasis within 2.5 years, which indicates a very poor prognosis ($p<0.001$) and the requirement for a more aggressive therapeutic strategy.

References

[1] P. M. Comoglio, L. Tamagnone and C. Boccaccio, *Plasminogen-related growth factor and semaphorin receptors: a gene superfamily controlling invasive growth*, Exp. Cell Res., 253:88-99, 1999

[2] M.F. Di Renzo, M. Olivero, T. Martone, A. Maffe, P. Maggiora, A.D. Stefani, G. Valente, S. Giordano, G. Cortesina and P.M. Comoglio, *Somatic mutations of the MET oncogene are selected during metastatic spread of human HNSC carcinomas*, Oncogene, 19:1547-1555, 2000

[3] E. Medico, A.M. Mongiovi, J. Huff, M.A. Jelinek, A. Follenzi, G. Gaudino, J.T. Parsons and P.M. Comoglio, *The tyrosine kinase receptors Ron and Sea control "scattering" and morphogenesis of liver progenitor cells in vitro*, Mol Biol Cell, 7:495-504, 1996

[4] L.J. van 't Veer, H. Dai, M.J. van de Vijver et al., *Gene expression profiling predicts clinical outcome of breast cancer,* Nature 415:530-536, 2002