



Project no.
503438

Project acronym

TRANSFOG

Project title

Translational and Functional Oncogenomics: from cancer-oriented genomic screenings to new diagnostic tools and improved cancer treatment

Instrument: **Integrated Project**

Area: **LSH-2002-2.2 Priority: 1.1.1.ii.b**

Title of report

Publishable final activity report

Period covered: **1st June 2007-30th November 2008**

Date of preparation: **12th February 2008**

Start date of project: **1st June 2004**

Duration: **54 months**

Project Coordinator: OECI-EEIG,
Organisation of European Cancer
Institutes,
European Economic Interest Grouping
11, Rue d'Egmont, 1000 Brussels (BE).
Prof. Guy Storme,
Phone: +32 2 512 0146
Email: guy.storme@az.vub.ac.be

Scientific Coordinator: IRCC,
Institute for Cancer Research and Treatment,
University of Torino Medical School,
S.P. 142, km 3.95, 10060 Candiolo (IT).
Dr Enzo Medico,
Phone: +390 119 933 234
Email: enzo.medico@ircc.it

Table of Contents

- 1. Project execution 3**
 - 1.1 Project objectives..... 4**
 - 1.2 Work performed and major results 6**
 - 1.3 Achievements of the project related to the state-of-the-art..... 9**
 - 1.4 Impact of the project on its industry or research sector 11**
- 2. Dissemination and use..... 13**
- References 16**

1. Project execution



<http://www.transfog.org>

TRANSFOG is an EC-founded FP6 project carried out by a Consortium of 20 Partners located in 9 EU member States, plus Switzerland and Israel.

1.	OECI	OECI-EEIG, Organisation of European Cancer Institutes, European Economic Interest Grouping.	Belgium
2.	IRCC	Fondazione Piemontese per la Ricerca sul Cancro – ONLUS.	Italy
3.	CNIO	Centro Nacional de Investigaciones Oncologicas.	Spain
4.	DKFZ	Deutsches Krebsforschungszentrum.	Germany
5.	NKI	Nederlands Kankerinstituut/ Antoni van Leeuwenhoek ziekenhuis	Netherlands
6.	UMCU	University Medical Center Utrecht.	Netherlands
7.	IFOM	Fondazione Italiana per la Ricerca sul Cancro, Istituto FIRC di Oncologia Molecolare.	Italy
8.	EMBL-EBI	European Molecular Biology Laboratory.	Germany
9.	FLEMING	Biomedical Sciences Research Centre “Alexander Fleming”	Greece
10.	FMI	Novartis Forschungsstiftung, Zweigniederlassung Friedrich Miescher Institute for Biomedical Research.	Switzerland
11.	AGENDIA	Agendia BV.	Netherlands
12.	IPP	Medical University of Innsbruck (MUI), Institute of Pathophysiology	Austria
13.	KI	Karolinska Institutet.	Sweden
14.	LICR	Ludwig Institute for Cancer Research -	Switzerland
15.			
16.	LNCIB	Consorzio Interuniversitario Biotecnologie.	Italy
17.	MDC	Max-Delbrück-Centrum für Molekulare Medizin (MDC) Berlin-Buch.	Germany
18.	WIS	The Weizmann Institute of Science.	Israel
19.	ALTA	ALTA Ricerca e Sviluppo in Biotecnologie Srl	Italy
20.	KCL	Kings's College London	UK

TRANSFOG is jointly managed by a Project Coordinator and a Scientific Coordinator.

Project Coordinator: **OECI** (Organisation of European Cancer Institutes),
Prof. Guy Storme,
c/o Fondation Universitaire, Rue d’Egmont 11, 1000 Brussels (BE).
Phone: +32 2 477 6147.
Email: guy.storme@az.vub.ac.be

Scientific Coordinator: **IRCC** (Institute for Cancer Research and Treatment),
Prof. Enzo Medico,
S.P. 142, km 3.95, 10060 Candiolo (IT).
Phone: +390 119 933 234.
Email: enzo.medico@ircc.it

1.1 Project objectives

Sooner or later during the development of most types of human cancer, primary tumour masses spawn pioneer cells that move out, infiltrate adjacent tissues, to then travel to distant sites to colonize new terrain in the body where, at least initially, nutrients and space are not limiting. These distant settlements of tumour cells, that is, metastases, are the most life-threatening aspects of the oncogenic process and account for 90% of human cancer deaths. Like the formation of the primary tumour mass, successful invasion and metastasis depend upon a critical balance between deregulated proliferation and inhibition of apoptosis. However, these alterations must act in concert with more subtle operational strategies, involving changes in the physical coupling of cells to their microenvironment and activation of proteases that degrade the extracellular matrix. The integration of all these cellular behaviors defines a complex, multi-step program of tumor-host interactions that is conventionally termed “invasive growth”¹. Identification of new genes involved in the invasive growth program and in-depth characterization of the complex networks governing embryonal epithelial morphogenesis and cancer metastatic progression are likely to provide new tools for personalized diagnosis and for targeted therapeutic approaches. While extensive analysis over the last two decades led to a deep insight into the control of cell proliferation and survival, and their alterations during cancer onset, still much remains to be clarified about the genetic lesions and alterations of cell signalling that lead to aberrant activation of invasive growth, cancer progression and metastasis.

TRANSFOG aims at the systematic identification and functional characterization of novel cancer genes with high potential diagnostic and therapeutic value. The main ambition of this project is to re-establish European leadership in the field of cancer genomics. To this aim, the project is integrating academic and industrial cutting edge groups to develop a joint effort towards systematic identification and functional characterization of novel cancer genes with high potential diagnostic and therapeutic value in breast, colon and lung cancer. The project is based on the scientific excellence of the applicants, who have made major recognized contributions to the field and on the construction of shared technological platforms. It includes: **i)** molecular profiling by DNA microarrays, and proteomics of human tumours and of cell-based model of oncogenic progression; screening of epigenetic alterations by MeDIP-on-CHIP; **ii)** Clinical validation by extensive microarray analysis, generation of monoclonal antibodies, digitalization and automatic evaluation of IHC analysis, generation of tissue microarrays containing human samples, TaqMan Low-Density Arrays; **iii)** in-depth analysis of protein-protein interactions and oncogenic signal transduction; **iv)** tumor mouse models; **v)** systematic gain- and loss-of function genetic screens in cellular models; **vi)** Negative feedback regulation of growth factor signalling in cancers of epithelial origin.

TRANSFOG brings together world-recognised competences and resources to reach the following, integrated research objectives:

- 1 Identification of novel cancer-related genes of high clinical-diagnostic potential, with a specific focus on progression and metastasis of colon, breast, lung cancer.
- 2 Set-up of technologies for systematic cancer gene functional analysis and identification of new molecular targets.
- 3 Systematic exploration of oncogenic/antioncogenic signalling pathways, epigenetic regulatory mechanisms to generate a comprehensive overview of the alterations in signalling and regulatory networks involved in cancer progression.

- 4 Development of tools for diagnostic validation of molecular signatures for cancers of high population impact, such as colon, breast and lung. This enables translation into clinical use of signatures obtained through the cancer-oriented genomic screenings performed by the Partners.
- 5 Establishment of a shared bioinformatic platform for functional oncogenomics data handling and standardization.

The TRANSFOG experimental pipeline consists of seven research components that synergistically enable streamlined translation of large-scale genomic screenings into high-impact contributions to cancer diagnosis and therapy. The seven activities, illustrated in Figure 1, are described in the following paragraphs together with a brief outline of the most significant results obtained.

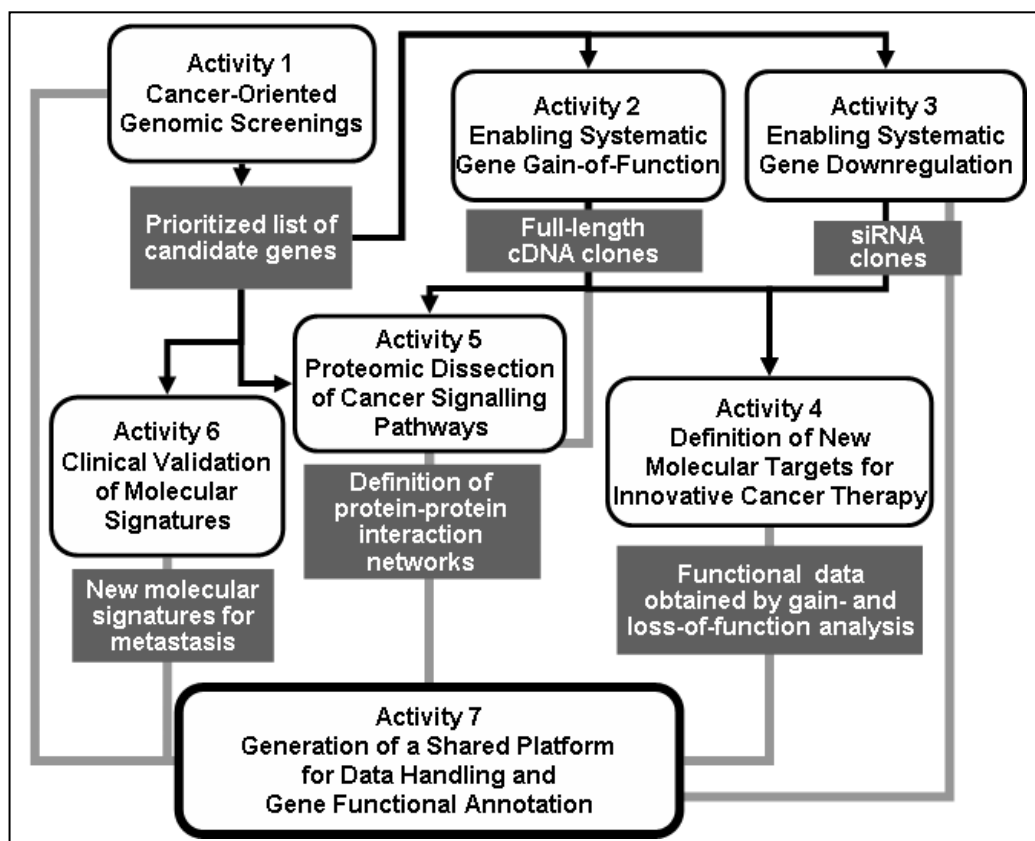


Figure 1. Flowchart of the seven research activities of the TRANSFOG pipeline (white boxes), of their outputs (gray boxes), dependencies (black arrows) and interconnections (gray lines). The main outcome of Activity 1, the list of candidate genes resulting from cancer-oriented screenings, informs Activities 2 and 3 for the production of tools for specific gene gain- and loss-of-function, respectively. These tools are in turn essential for Activities 4 and 5, which carry out systematic analysis of gene function and signal transduction. An important feed-back will come to Activity 1 from Activity 5, which will identify new binding partners for the proteins studied and therefore add them up to the novel candidate cancer gene list. Activity 1 also instructs Activity 6, with the definition of molecular signatures of potential clinical value to be validated. A key element to enable data standardization, sharing and mining is Activity 7, which will provide the TRANSFOG bioinformatics backbone.

1.2 Work performed and major results

1. Cancer-oriented genomic screenings in tumours and cell lines.

Recent works have shown that it is possible to exploit gene expression profiling of tumour samples to define sets of genes (signatures) whose expression correlates, positively or negatively, with metastasis-free survival, e.g. in breast cancer². It has also been found that a general signature associated with metastatic behaviour can be shared between solid tumours of different organs³, which indicates that common alterations of basic cellular functions and signalling pathways trigger metastatic progression of cancer. During the TRANSFOG project, genome-wide screenings by DNA microarrays, array-CGH, epigenetic and proteomics have been carried out by various partners and merged with the particular aim of identifying and prioritizing genes with a potential role in cancer metastasis (the “candidate genes”).

2. Development of enabling technologies for systematic gene gain-of-function.

One approach for functional characterization of candidate genes identified by Activity 1 is based on enabling the expression of their full-length (FL)-cDNAs in cells of interest. This required the assembly of a core FL-cDNA collection, exploiting the expertise and resources of the Partner DKFZ⁷. The functional characterization and validation of genes in functional genomic and proteomic applications requires comprehensive resources for the analysis of gene gain- and loss-of-function phenotypes. These are provided by ORFeome and RNAi resources, respectively, and the Transfog consortium, consequently, had two dedicated workpackages aiming to create and provide such resources. Workpackage 2 focused on the generation and dissemination of ORFeome resources, where the open reading frames of candidate genes were systematically cloned into the Gateway cloning system to generate a versatile resource for gain-of-function studies. Furthermore, expression constructs were generated that permit the overexpression of the ORF-encoded proteins in bacterial and mammalian cell systems.

3. Generation or acquisition of tools for RNA interference-based systematic gene loss-of-function analysis

A second way to analyze the function of genes is by inducing loss-of-function via RNA interference. In this view, a key effort of TRANSFOG has been the generation of a shared collection of hundreds of human shRNA constructs in a plasmid/retroviral expression system (which allows easy further transfer of the construct in the target cells of choice), mainly targeting genes of unknown function that emerge from TRANSFOG cancer-oriented genomic explorations described in Activity 1. The technology has been made available by Partner NKI, who set-up a methodology for systematic generation of retroviral shRNA vectors for functional screenings⁸. This approach is being flanked by the use of double-stranded siRNA oligonucleotides for the 100 top genes from the TRANSFOG prioritized gene list. Such siRNAs represent a high-quality reagent that can be used in standard transfection assays by all consortium members, allowing optimal comparison of data between laboratories. The use of

single-gene silencing RNA species will allow the identification of individual gene functions whereas combinatorial approaches will allow the characterization of polypeptides active in the same cellular pathways. Finally, regulatable RNAi vector systems have also been developed to avoid cell vitality problems deriving from stable silencing of essential genes⁹.

4. Gene functional characterization by cell-based assays and analysis in model organisms.

Modulation of growth, motility, survival, invasion, adhesion, morphogenesis, senescence, and other basic biological functions altered during tumour progression and metastasis have been analyzed by transduction of cultured cells with FL-cDNAs, shRNAs or siRNAs. Some of the proposed assays reached an adequate throughput for systematic gene functional analysis, allowing the identification of genes modulating the SRC proto-oncogene¹⁰ or the analysis of complex and combinatorial effects of multiple gene modulation¹¹. In other cases, low-throughput studies allowed detailed characterization of few genes. As an example, a transcriptional switch between two EGF transcriptional targets, the actin-binding proteins tensin and cten, was found to be essential for EGF-driven mammary cell migration¹².

Specific animal models were exploited to assess the therapeutic potential of targeting the MET proto-oncogene¹³, the cooperation between oncogenic KRAS activation and chronic pancreatitis to promote pancreatic cancer¹⁴, or the surprising dispensability of the cell cycle proteins Cdk2 and Cdk4 for mouse viability¹⁵. Finally, systematic gene functional analysis has also been conducted in lower vertebrates like zebrafish and drosophila¹⁶, with a particular care on accurate annotation of the result, for optimal cross-species comparison¹⁷.

5. Proteomic approaches to the study of signal transduction and protein-protein interactions.

Protein post translational modifications and interactions with other proteins play a key role in many biological processes related to cancer progression. However, a comprehensive view of the networks of interactions and of their dynamics in normal and cancer cell is still lacking, and technically challenging. To shed light on the candidates of interest from this point of view, the TRANSFOG partners have exploited mass spectrometry, Biacore biosensor and cell-based analysis. Although Biacore measurements were originally included in the plans to obtain quantitative and kinetic data, it became clear that this approach did not seem suitable for high numbers of proteins, and for that matter were minimized. The other approaches were successful. It should be noted that the number of identified protein-protein interactions using tagged proteins expressed in cell lines might not have been the highest, but here most actual technical progress has been made (knowledge about suitable tags, isolation procedures, sample handling for mass spec approaches). The level of throughput differed among groups for obvious reasons like the nature of proteins studied and the processes in which such proteins are involved.

6. Preliminary diagnostic validation of molecular cancer signatures.

Converting a molecular signature emerged from a cancer genomic screening into a validated tool of potential clinical utility is a demanding task. For instance, the platform originally used to define the signature (e.g. a certain type of microarray) may not be the most adequate for

subsequent capillary diffusion of the signature assay. A translational research phase has therefore been carried out, to re-assess the signatures of interest on new tumour samples and with other platforms (e.g. realtime PCR, Tissue microarrays, immunohistochemistry), and to make cross-comparisons between platforms available at different sites. Standardized procedures have been defined for the various platforms and for the management of both clinical and experimental data.

7. Generation of a common platform for data handling and gene functional annotation.

The TRANSFOG pipeline has to deal with a large variety of data coming from the various activities: (Activity 1) microarray data from different platforms and different organisms and related biological/clinical information, epigenetic data and differential proteomics data; Activities 2-3) Availability of vectors or reagents for specific gene overexpression and downmodulation; (Activity 4) Cell-based functional assays (from various organisms); phenotype descriptions (model organisms); clinical descriptions (preclinical proof-of-concept experiments with mouse xenografts and other mouse models); (Activity 5) protein-protein interaction data and networks from various organisms; (Activity 6) experimental and clinical data from signature validation experiments and from tissue microarrays. Capturing and representing all this information in a format appropriate for data mining is a complex task. Moreover, these data come from different laboratories and different organisms; therefore we needed to develop data communication standards and systematic orthologue analysis. All the data have been standardized and integrated to reach a comprehensive human/mouse/other organism genome annotation system, exploiting the Distributed Annotation System²⁵ (DAS, <http://www.biodas.org>), originally developed by Lincoln Stein, to provide a simple, flexible bioinformatics backbone.

DAS requires a reference server providing the framework to which all annotation will be anchored (usually the genomic sequence). In this project, the reference system is made by all the explored genes anchored to their position in the genomic sequence. Relative to the reference framework, multiple annotation servers can provide annotation for the reference objects. We adapted the system to include additional data types of our interest, e.g. the availability of full-length cDNAs, results of genomic screenings, conditional expression changes derived from microarray results, and functional or proteomic assays. DAS clients can connect to the reference server and any number of annotation servers and present an integrated view of the annotation of the reference objects. Figure 2 shows how single bits of annotation anchored to each gene can be added in multiple layers to provide a high-level, integrated view of experimental results from the project partners; experimental details remain accessible through hyperlinks to the database systems of the project partners or to the submitted data in public repositories.

A significant effort within this activity is also devoted to standardization of the research output. To this aim, the Partner EBI leads or participates to international standardization initiatives, such as the definition of a “Functional Genomics Experiment model”²⁶, of the “minimum information required for reporting a molecular interaction experiment”²⁷, or of the “minimum information about a proteomics experiment”²⁸.

1.3 Achievements of the project related to the state-of-the-art

I. Genome-wide screenings by DNA microarrays, array-CGH, epigenetic and proteomics have been carried out by 14 partners and finally merged with the particular aim of identifying and prioritizing genes with a potential role in cancer metastasis, the “candidate genes”. The TRANSFOG screenings concentrated on breast, lung and colon cancer, which altogether account for the majority of cancer deaths in the general population. Apart from tumours, screenings have also included cancer-oriented experimental models, like serine and tyrosine kinase receptor-driven transcriptional and proteomic responses^{4,5}, transcriptional responses to oncogenic Ras mutation⁶, ligand-induced in vitro epithelial morphogenesis and invasive growth, in vitro angiogenesis of endothelial cells. The aim was to obtain a genome-wide exploration of the basic mechanisms of cancer progression. By merging the results of the screenings, we could find “common” genes, i.e. genes emerging from more than one screening as associated to invasion and metastasis, and “specific” genes, whose expression is only altered in small subgroups or subtypes of tumours/metastases or cellular models. The relevant genes have been ranked for priority towards functional characterization and/or diagnostic validation, with the main priority criterion being their emergence in more than one screening.

II. The open reading frames (ORFs) of 224 genes and splice variants were cloned into Gateway-compatible entry vectors and then sequence verified. In total, 443 entry clones were generated, so that for most ORFs one clone was available that contained the natural stop codon, while a second clone was created without the stop codon, to allow for the expression of C-terminal fusion proteins. All clones were disseminated to the partners of the consortium to permit gene gain-of-function studies in cancer relevant context. The cloned ORFs have additionally been entered into the International ORFeome Collaboration, which aims to generate and provide a comprehensive resource of Gateway-compatible clones that shall be made available in academia and industry.

III. Efficient tools for single gene loss-of function analysis have been developed, and siRNAs targeting 100 genes have been acquired by the consortium and made available to partners.

IV. Functional characterization of hundreds of candidate genes by the various partners. As examples: (i) results of the functional analysis showed that HGF-driven invasive growth involves transcriptional downregulation of Arhgap12, a Rac1-specific GAP (Gentile et al, Oncogene. 2008 May 26. [Epub ahead of print]). (ii) A genome wide shRNA screen has been performed for genes whose suppression confers resistance to a natural compound inhibitor of the mTOR kinase (rapamycin) in breast cancer cells. This is clinically relevant, as rapamycin analogs are in phase 2 clinical testing for treatment of breast cancer. We have validated a novel gene whose suppression causes resistance to rapamycin in breast cancer. A manuscript describing these findings is being written. (iii) We also focused on the role of oxidative stress in regulation of the tumor suppressor proteins from the FOXO family. They have now gone on to demonstrate that the MDM2 protein functions as a E3 ligase, responsible for the (multi)mono-ubiquitination induced by oxidative stress. This modification results in activation of FOXO (Brenkman et al. (2008) PlosOne 3, e2819). At later stages, oxidative stress leads to the phosphorylation of a number of newly identified residues of FOXO. These phosphorylation sites make up an interaction interface for the isomerase PIN1. PIN1 enhances the activity of the deubiquitinating USP7 towards, thereby lowering the amount of (multi)mono ubiquitinated FOXO and its activity. Thus, PIN1 functions as a negative regulator of FOXO. Interestingly, a negative correlation in breast tumor samples was found between levels of

PIN1 and the FOXO target p27kip1 (Brenkman et al. (2008) Cancer Res. Accepted). (iv) Site-directed mutagenesis of mapped phosphorylation sites on the JNK-specific DUSP8 phosphatase has resulted in a delayed time-course activation of JNK by the tumor promoter arsenite, suggesting that phosphorylation is a negative regulatory mechanism for DUSP8; (v) Domain-swapping experiments and deletion mutagenesis have mapped the JNK-DUSP8 interaction site and the region conferring specificity between the two isoforms, JNK1 and JNK2, in terms of the differential regulation of their interaction by arsenite; (vi) RNAi experiments with a Drosophila JNK-specific DUSP, directed to various tissues with GAL4 drivers, have revealed important roles in the development of the nervous system; (vii) High-Mobility Group I(Y) interacting proteins have been identified and characterized, suggesting important regulatory roles in gene expression; (viii) The mechanism of action of the transcription factor NF- κ B has been deciphered, in terms of its ability to promote interchromosomal interactions that permit monoallelic gene expression.

V. Proteomics and protein-protein interaction research activity led to (i) a detailed characterization of the role of the c-Jun N-terminal kinase in the PDGF receptor pathway promoting cell migration¹⁸; (ii) the definition of post-translational modifications essential for the activity of the FOXO4 transcription factor¹⁹, and (iii) modulation of the activity and localization hypoxia-inducible factor-1alpha via MAPK-mediated phosphorylation²⁰. These results add a further level of complexity to the simplistic view of the signal transduction “pathway”, and rather highlight the real, “network” status of the intracellular signaling system. Much work is still ongoing in this field as a follow-up to the results obtained during the TRANSFOG Project.

VI. Diagnostic validation of genomic signatures and single genes. The main diagnostic problem that TRANSFOG is addressing is the prediction of the probability with which a primary carcinoma will give rise to metastasis. In the first phases of the project, careful assessment of the optimal data treatment and cross-comparison has been performed, dealing in particular with data clustering²¹, analysis and optimization of the predictive models²², and projection of data clusters across independent experimental datasets²³. In addition to complex multigene classifier validation, single gene analyses are also being conducted to explore new potentially useful diagnostic tools. As an example, specific mutations of the EGFR gene are associated to impaired ubiquitination and downregulation, potentially leading not only to oncogenic activation, but also to refractoriness to targeted anticancer treatments aimed at EGFR downregulation²⁴.

VII. Partner 8 (EMBL-EBI) created the TRANSFOG Distributed Annotation System (DAS) Portal (<https://www.ebi.ac.uk/transfog-srv/portal/>) to integrate data from consortium and public databases for human and computer analysis. The portal is used to search and visualise the prioritised candidate gene list generated in work package 1, as well as provide detailed information on each gene. The search engine provides summary information about candidate genes, including gene name, rank, description and chromosome. The detailed view includes the same information as the search view, with the addition of consortium and public functional annotation and database cross-references. The search engine provides semantic querying of TRANSFOG data based on ontologies such as GO, ECO and PSI:MI. For example, a search for the GO term 'apoptosis' (GO:0006915) automatically includes a search for descendants of the term (activated T cell apoptosis, anoikis, apoptotic program ...etc). Search results can be downloaded in XML and CSV formats for off-line analysis. All unpublished data is kept secure using password protection and 128-bit data encryption. The search engine is built on PHP and Solr (<http://lucene.apache.org/solr/>), an open source enterprise search server based on the Lucene Java search library. The detailed view is built on

Dasty2 (<http://www.ebi.ac.uk/dasty/>), a fast and interactive DAS client developed for the BioSapiens Network of Excellence project. The use of a semantic query engine and data encryption is novel in the field of DAS.

EMBL-EBI developed a DAS annotation server to facilitate the sharing of TRANSFOG data during and after the lifetime of the project (<https://www.ebi.ac.uk/transfog-srv/server/bantam.php/das/dsn>). The server consists of two files that can be installed with minimal effort, and is thus highly suitable for laboratories with limited bioinformatics support. The annotation server is written in PHP and uses General Feature Format (GFF) files to publish partners' prioritised gene lists, experimental data and functional annotation. GFF files can be created in a text editor, generated by Excel or exported from database systems. To enable partners to manage information about their own annotation servers without intervention from EMBL-EBI, a TRANSFOG-specific installation of the DAS Registry system (<http://www.dasregistry.org>) was created (<https://www.ebi.ac.uk/transfog-srv/dasregistry/>). EMBL-EBI also created an Ensembl DAS reference server based on Ensembl human gene identifiers to provide genes names and descriptions, and cross-references to UniProt and IPI (http://www.ebi.ac.uk/das-srv/genedas/das1/Homo_sapiens.Gene_ID.reference). The TRANSFOG annotation server has been designed to be very easy to install and maintain, particularly for small laboratories, and in this respect is an improvement on the two main DAS annotation frameworks, ProServer (<http://www.sanger.ac.uk/Software/analysis/proserver/>) and Dazzle (<http://www.derkholm.net/thomas/dazzle/>). The TRANSFOG annotation server only supports data in GFF files, and so is less suitable than either ProServer or Dazzle for laboratories using relational databases to store gene annotations.

EMBL-EBI developed method to represent protein interaction partners using the TARGET element of the DAS FEATURES response, and worked with Michael Ashburner at Cambridge University's Department of Genetics to ensure all experimental methods used in TRANSFOG are included in the Evidence Codes ontology (http://www.obofoundry.org/cgi-bin/detail.cgi?id=evidence_code).

EMBL-EBI created a Microsoft Excel spreadsheet (<https://www.ebi.ac.uk/transfog-srv/portal/downloads.html>) to create, edit, import and export GFF files to facilitate candidate gene ranking and functional annotation, and subsequent data sharing via the DAS annotation server. The spreadsheet offers a facility to automatically convert gene and protein identifiers to Ensembl gene IDs at the click of a button using Ensembl's BioMart web service (<http://www.biomart.org>). Assignment of ontology terms is made straightforward using editing tools built on EMBL-EBI's Ontology Lookup Service (<http://www.ebi.ac.uk/ontology-lookup/>). The spreadsheet is written in Visual Basic for Applications (VBA). The creation, editing and import/export of GFF files in an Excel spreadsheet is novel, as is the integration of BioMart web services for gene identifier mapping. Ontology term lookup using the OLS web service is based on work undertaken at EMBL-EBI for the ProteomeHarvest PRIDE submission spreadsheet (<http://www.ebi.ac.uk/pride/proteomeharvest/>).

1.4 Impact of the project on its industry or research sector

The goal of the present project is dual:

- (i) to develop innovative cancer-specific molecular signatures based upon in-depth analysis and understanding at the genomic level of tumour development and metastatic

spread, as a novel approach for the diagnosis and treatment of breast, lung, colon and possibly other epithelial cancer;

(ii) identification of key genes controlling basic biological functions involved in cancer progression, and potentially exploitable as new molecular targets for innovative therapies.

The results obtained in this project will open new perspectives in the diagnosis and treatment of cancer, with social impacts measurable as improved health, and reduction of health societal costs, including direct and indirect costs for patient assistance as well as loss of manpower.

2. Dissemination and use

Exploitable result 1

ColoPrint is a **novel microarray-based gene expression diagnostic test** for the prediction of distant metastases of early stage colon cancer.

This test can be used to select patients with early stage colon cancer for further chemotherapy, which is difficult at present based on conventional clinical/histopathological. This may lead to a better selection of the colon cancer patients that benefit from chemotherapy and hence to a better treatment of this disease.

Agendia has in the recent past commercialized a comparable test for breast cancer, including clearance of this test through all regulatory systems in the EU and US. We see no major hurdles in commercialization, other than the poor climate to raise funds for this purpose at the present time. Hence, we seek partners for commercialization of the product.

The next steps in making this test available would be to develop a custom microarray with the relevant genes, and validate the findings on the custom array. Then, regulatory approval must be filed in the EU and US.

Commercialization will be taken up by Agendia BV directly, as this company also has the sales channel for this purpose. Partnerships are welcome. Please contact Mr. Bas van der Baan, director EU sales Agendia BV. Amsterdam, email bas.vanderbaan@agendia.com.

Exploitable result 2

Knowledge on the protein-protein interactions of a number proteins (Pin1, MSP1, Borealin, Epac1), known to be involved in tumorigenesis. The partner involved in the exploitation is UMCU. His role and activities are: owner of IP and phospho-specific antibodies.

The result . phospho-specific antibodies might be exploited - directly (spin offs etc) or indirectly (licensing). Commercial value of antibodies is out weighted by the need to make knowledge available to the scientific community. Knowledge and antibodies are valuable for further research aimed at unravelling signal transduction pathways involved in tumorigenesis.

The results obtained will be followed up in more nationally oriented initiatives like Top Institute Pharma (TIP) and Center for Translational and Molecular Medicine (CTMM). TIP and CTMM are Dutch initiatives aimed at strengthening collaborative research efforts of academia and industry.

Due to the basic science character, UMCU chose for dissemination of data and protocols via publications in international scientific journals (see publications partner 6 UMCU), rather than filing patents. Furthermore, our work was made public at many international meetings. In addition, we were involved in organizing the CGC/NFK meeting voor patienten en partners, where presentatieons were given for an audience of Dutch cancer patients and relatives), March 10, 2007 (Utrecht, The Netherlands).

Exploitable result 3

A set of putative tumor suppressor genes induced by growth factors and having a negative regulatory function in normal mammary cells.

- *Application in cancer diagnosis and prognosis is envisaged, but will require additional characterization as well as selection of a shorter list of genes.*
- *Possible market application: a broad range of carcinoma (e.g., breast and colon cancer). Market application as biomarkers in cancer diagnosis and prognosis prediction, as well as markers of response to experimental drugs.*
- *The team is working on a laboratory prototype.*

Publication: Amit, I., Citri, A., Shay, T., Lu, Y., Katz, M., Zhang, F., Tarcic, G., Siwak, D., Lahad, J., Jacob-Hirsch, J., Amariglio, N., Vaisman, N., Segal, E., Rechavi, G., Alon, U., Mills, G.B., Domany, E., and Yarden, Y. A module of negative feedback regulators defines growth factor signaling. *Nature Genetics* 39(4):503-12. (2007)

Exploitable result 4

Identification of Cten as a marker of breast tumor metastasis and a monoclonal antibody to Cten

- *Identified an intracellular protein, Cten, involved in intracellular events related to invasive growth of breast cancer*
- *Demonstration on a cohort of breast cancer patients that Cten expression predicts metastasis to lymph nodes.*
- *Generation of a monoclonal antibody to Cten and confirmation of activity in tumor specimens*
- *The antibody may be developed by an industrial collaborator; negotiations have been started*

Publication: Katz M, Amit I, Citri A, Shay T, Carvalho S, Lavi S, Milanezi F, Lyass L, Amariglio N, Jacob-Hirsch J, Ben-Chetrit N, Tarcic G, Lindzen M, Avraham R, Liao YC, Trusk P, Lyass A, Rechavi G, Spector NL, Lo SH, Schmitt F, Bacus SS, Yarden Y. A reciprocal tensin-3-cten switch mediates EGF-driven mammary cell migration. *Nat Cell Biol.* 9(8):961-9. (2007).

Exploitable result 5

The researchers from IFOM have identified a molecular progression signature of breast cancer metastases, with respect to the primary tumor. This finding is relevant from a therapeutic viewpoint, if the development of metastasis-specific targeting strategies is to remain a viable option.

The simultaneous down-regulation of RERG, PGR1 and GATA3 genes also associated with a low ER protein concentration seems to be related to an unfavourable prognosis in tamoxifen treated breast cancer patients. Expression levels of these genes could be indicative for the presence of a tumor refractory to tamoxifen treatment.

Exploitable result 6

A new strategy for gene functional screening to identify molecular targets for cancer treatment, combining the use of cDNA expression libraries with DNA microarrays. Since it involves the use of different species as the source for expression libraries and for the target cells, we named it “xenoarray analysis”. After the transduction of a cDNA expression library from a given species into a target cell line of a different species, the abundance of exogenous cDNAs derived from the library can be directly monitored with standard expression arrays in the entire cell population, by exploiting orthologue diversity. In such

approach, sequence divergence between orthologue transcripts can be exploited as a “molecular bar-code” for species-specific hybridization on microarrays. Then, if a selective pressure is applied to the transduced population, enrichment of cDNAs conferring resistance to the selection can be directly monitored at once with expression arrays. The new procedure brings substantial advantages to the field of expression cloning, being faster, more systematic and more sensitive. After selecting transduced cells, Xenoarray analysis allows identification of enriched genes within a week, instead of the months typically required to follow-up single clones or to rescue proviruses and repeat the selection. Genes identified by shRNA-based screenings are not immediately exploitable as therapeutical targets, being genes whose loss is advantageous, not disadvantageous, to cancer cells. Conversely, the Xenoarray approach establishes a positive cause-effect relationship in identifying genes that confer resistance to specific selective stresses. Such genes are directly exploitable as potential therapeutical targets. The method has been patented (PCT: “Method for the identification of targets for cancer therapy) and published.

Publication: Martelli ML, Isella C, Mira A, Fu L, Cantarella D, Medico E. Exploiting orthologue diversity for systematic detection of gain-of-function phenotypes. BMC Genomics. 2008 May 29;9:254.

References

1. Comoglio PM, Trusolino L: Invasive growth: from development to metastasis. *J Clin Invest*, 109: 857-62, 2002.
2. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH: Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 415: 530-536, 2002.
3. Ramaswamy S, Ross KN, Lander ES, Golub TR: A molecular signature of metastasis in primary solid tumors. *Nat Genet*, 33: 49-54, 2003.
4. Amit I, Citri A, Shay T, Lu Y, Katz M, Zhang F, Tarcic G, Siwak D, Lahad J, Jacob-Hirsch J, Amariglio N, Vaisman N, Segal E, Rechavi G, Alon U, Mills GB, Domany E, Yarden Y: A module of negative feedback regulators defines growth factor signaling. *Nat Genet*, 39: 503-512, 2007.
5. Saridaki A, Panayotou G: Identification of growth factor-regulated proteins using 2D electrophoresis and mass spectrometry. *Growth Factors*, 23: 223-232, 2005.
6. Arena S, Isella C, Martini M, de Marco A, Medico E, Bardelli, A: Knock-in of oncogenic Kras does not transform mouse somatic cells but triggers a transcriptional response that classifies human cancers. *Cancer Res*, 67: 8468-8476, 2007
7. Wiemann S, Arlt D, Huber W, Wellenreuther R, Schleege S, Mehrle A, Bechtel S, Sauermann M, Korf U, Pepperkok R, Sultmann H, Poustka A: From ORFeome to biology: a functional genomics pipeline. *Genome Res*, 14: 2136-2144, 2004.
8. Berns K, Hijmans EM, Mullenders J, Brummelkamp TR, Velds A, Heimerikx M, Kerkhoven RM, Madiredjo M, Nijkamp W, Weigelt B, Agami R, Ge W, Cavet G, Linsley PS, Beijersbergen RL, Bernards R: A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature*, 428: 431-437, 2004.
9. Heinonen JE, Mohamed AJ, Nore BF, Smith CI: Inducible H1 promoter-driven lentiviral siRNA expression by Stuffer reporter deletion. *Oligonucleotides*, 15: 139-144, 2005.
10. Eichhorn PJ, Creighton MP, Wilhelmsen K, van Dam H, Bernards R: A RNA Interference Screen Identifies the Protein Phosphatase 2A Subunit PR55gamma as a Stress-Sensitive Inhibitor of c-SRC. *PLoS Genet*, 3: e218, 2007.
11. Sahin, O., Lökke, C., Korf, U., Sultmann, H., Poustka, A., Wiemann, S., Arlt, D: Combinatorial RNAi strategy: The next generation of quantitative protein network analysis. *Proc Natl Acad Sci USA*, 104: 6579-6584, 2007.
12. Katz M, Amit I, Citri A, Shay T, Carvalho S, Lavi S, Milanezi F, Lyass L, Amariglio N, Jacob-Hirsch J, Ben-Chetrit N, Tarcic G, Lindzen M, Avraham R, Liao YC, Trusk P, Lyass A, Rechavi G, Spector NL, Lo SH, Schmitt F, Bacus SS, Yarden Y: A reciprocal tensin-3-cten switch mediates EGF-driven mammary cell migration. *Nat Cell Biol*, 9: 961-969, 2007.
13. Corso S, Migliore C, Ghiso E, De Rosa G, Comoglio PM, Giordano S: Silencing the MET oncogene leads to regression of experimental tumors and metastases. *Oncogene*, 27: 684-693, 2008.
14. Guerra C, Schummacher AJ, Cañamero M, Grippo PJ, Verdaguer L, Pérez-Gallego L, Dubus P, Sandgren EP, Barbacid M: Chronic Pancreatitis Is Essential for Induction of

- Pancreatic Ductal Adenocarcinoma by K-Ras Oncogenes in Adult Mice. *Cancer Cell*, 11: 291-302, 2007.
15. Barrière C, Santamaría D, Cerqueira A, Galán J, Martín A, Ortega S, Malumbres M, Dubus P, Barbacid M: Mice thrive without Cdk4 and Cdk2. *Molecular Oncology*, 1: 72-83, 2007.
 16. Miskey, C., Izsvák, Z., Kawakami, K. and Ivics, Z: DNA transposons in vertebrate functional genomics. *Cell Mol Life Sci*, 62:629-641, 2005.
 17. Sims D, Bursteinas B, Gao Q, Zvelebil M, Baum B: FLIGHT: database and tools for the integration and cross-correlation of large-scale RNAi phenotypic datasets. *Nucleic Acids Res*, 34: D479-483, 2006.
 18. Amagasaki K, Kaneto H, Heldin CH, Lennartsson J: c-Jun N-terminal kinase is necessary for platelet-derived growth factor-mediated chemotaxis in primary fibroblasts. *J Biol Chem*, 281: 22173-22179, 2006.
 19. van der Horst, A., de Vries-Smits, AM., Brenkman, AB., van Triest, MH., van den Broek, N., Colland, F., Maurice, MM., Burgering, BM. FOXO4 transcriptional activity is regulated by monoubiquitination and USP7/HAUSP. *Nat Cell Biol*. 8, 1064-73 (2006).
 20. Mylonis I, Chachami G, Samiotaki M, Panayotou G, Paraskeva E, Kalousi A, Georgatsou E, Bonanou S, Simos G: Identification of MAPK phosphorylation sites and their role in the localization and activity of hypoxia-inducible factor-1alpha. *J Biol Chem*, 281: 33095-33106, 2006.
 21. Fu L, Medico E: FLAME, a novel fuzzy clustering method for the analysis of DNA microarray data. *BMC Bioinformatics*, 8: 3, 2007.
 22. Reid JF, Lusa L, De Cecco L, Coradini D, Veneroni S, Daidone MG, Gariboldi M, Pierotti MA: Limits of predictive models using microarray data for breast cancer clinical treatment outcome. *J Natl Cancer Inst*, Jun 97: 927-930, 2005.
 23. Lusa L, McShane LM, Reid JF, De Cecco L, Ambrogi F, Biganzoli E, Gariboldi M, Pierotti MA: Challenges in projecting clustering results across gene expression-profiling datasets. *J Natl Cancer Inst*, 99: 1715-1723, 2007.
 24. Shtiegman K, Kochupurakkal BS, Zwang Y, Pines G, Starr A, Vexler A, Citri A, Katz M, Lavi S, Ben-Basat Y, Benjamin S, Corso S, Gan J, Yosef RB, Giordano S, Yarden Y: Defective ubiquitinylation of EGFR mutants of lung cancer confers prolonged signaling. *Oncogene*, 26: 6968-6978, 2007.

WP7

25. Dowell RD, Jokerst RM, Day A, Eddy SR, Stein L: The distributed annotation system. *BMC Bioinformatics*, 2: 7, 2001.
26. Jones AR, Miller M, Aebersold R, Apweiler R, Ball CA, Brazma A, Degreef J, Hardy N, Hermjakob H, Hubbard SJ, Hussey P, Igra M, Jenkins H, Julian RK Jr, Laursen K, Oliver SG, Paton NW, Sansone SA, Sarkans U, Stoeckert CJ Jr, Taylor CF, Whetzel PL, White JA, Spellman P, Pizarro A: The Functional Genomics Experiment model (FuGE): an extensible framework for standards in functional genomics. *Nat Biotechnol*, 25: 1127-1133, 2007.
27. Orchard S, Salwinski L, Kerrien S, Montecchi-Palazzi L, Oesterheld M, Stümpflen V, Ceol A, Chatr-aryamontri A, Armstrong J, Woollard P, Salama JJ, Moore S, Wojcik J, Bader GD, Vidal M, Cusick ME, Gerstein M, Gavin AC, Superti-Furga G, Greenblatt J,

- Bader J, Uetz P, Tyers M, Legrain P, Fields S, Mulder N, Gilson M, Niepmann M, Burgoon L, De Las Rivas J, Prieto C, Perreau VM, Hogue C, Mewes HW, Apweiler R, Xenarios I, Eisenberg D, Cesareni G, Hermjakob H: The minimum information required for reporting a molecular interaction experiment (MIMIx). *Nat Biotechnol*, 25: 894-898, 2007.
28. Taylor CF, Paton NW, Lilley KS, Binz PA, Julian RK Jr, Jones AR, Zhu W, Apweiler R, Aebersold R, Deutsch EW, Dunn MJ, Heck AJ, Leitner A, Macht M, Mann M, Martens L, Neubert TA, Patterson SD, Ping P, Seymour SL, Souda P, Tsugita A, Vandekerckhove J, Vondriska TM, Whitelegge JP, Wilkins MR, Xenarios I, Yates JR 3rd, Hermjakob H: The minimum information about a proteomics experiment (MIAPE). *Nat Biotechnol*, 25: 887-893, 2007.