

LOCCANDIA: Lab-on-Chip Based Protein Profiling for Cancer Diagnosis

Blanca Jordan^{#1}, Manuel M Perez-Perez[#], Rainer Ossig[§], Sophie Kaforou[&], Emeline Mery^{*}, Lena Brunet-Errard^{*}, Caroline Paulus^{*}, Vangelis Kritsotakis[&], Manos Kalaitzakis[&], Laurent Gerfault^{*}, Carmen Reina[#], Dimitris Kafetzopoulos[&], George Potamias[&], Florence Ricoul^{*}, Manolis Tsiknakis[&], Jürgen Schneckeburger[§] and Pierre Grangeat^{*2}

[#]ATOS Research and Innovation, Biotechnology and Health Unit
C/ Albarracin 25, Madrid 28035, Spain

^{*}CEA-LETI, MINATEC, Département microTechnologies pour la Biologie et la Santé (DTBS)
17 Rue des Martyrs, F 38054 GRENOBLE Cedex 9, France

[&]FORTH, Foundation for Research and Technology – IMBB & ICS
Vassilika Vouton, Heraklion 71110, Greece

[§]Westfaelische Wilhelms-Universitaet Muenster, Department of Medecine B Gastrointestinal Molecular Cell Biology
Domagkstrasse 3A, Münster, D-48149, Germany

¹blanca.jordan@atosorigin.com

²pierre.grangeat@cea.fr

Abstract — The purpose of this paper is to present the project LOCCANDIA as a promising nano-technological application in the medical-proteomics for early pancreatic cancer detection. LOCCANDIA stands for Lab-On-Chip based protein profiling for CANcer DIAgnosis and it is collaborative and multidisciplinary effort that aims to integrate a full proteomic analysis chain from blood sample to diagnosis information using innovative technologies.

The objective of the project is to validate the application of plasma protein profiling for early pancreatic cancer diagnosis by means of developing an innovative nano-technology based (lab-on-a-chip) platform integrated in a full proteomics analysis chain, from blood sample preparation to the diagnosis information combining bio- and information related technologies.

I. INTRODUCTION

Gastric cancers, such as pancreatic cancers, are among the most frequently observed severe diseases in developed countries. These types of cancers are detected by expensive diagnostic imaging methods at a late stage resulting in poor prognosis and high mortality rate since the only effective therapy is an early resection of the tumours

The human plasma proteome holds the promise of a revolution in disease diagnosis and therapeutic monitoring. The plasma protein analysis aims to characterize the proteomic status of cells and in particular to define the degree of their disorder according to their expression level pattern. This is in particular highly relevant to the efforts that have been done in associating specific protein marker levels in patients' blood with the different cancer stages.

One major breakthrough comes from the utilization of multi-protein disease markers detection instead of single protein analytes and the detection of all the isoforms of the selected proteins. Recent advances in mass spectrometry coupled with lab-on-chip technologies and bioinformatics may revolutionize medical diagnosis and cancer screening.

II. LOCCANDIA OBJECTIVES

The main research outcomes of LOCCANDIA project will be: (1) An optimised chromatographic-electrospray lab-on-chip dedicated to protein profiling for cancer diagnosis, (2) An Integrated Clinic-Proteomics Environment supporting the integrated device and diagnosis, (3) A proof-of-concept of this innovative lab-on-chip technology and the associated analysis chain for cancer diagnosis.

This proof of concept will be validated with a set of ninety-six human samples, coming from adenocarcinoma patients, controls and confused cases. The expected outcome would overcome the actual state-of-the-art approach for diagnosis based on imaging analysis and single proteins measurements by Elisa (Enzyme Linked Immuno Sorbent Assay)..

III. MEDICAL USER REQUIREMENT

The general requirements for molecular marker measurement are: exact measurement of validated protein marker concentrations, known concentration range and maximal tolerated variation of the marker, time frame for the full analysis procedure should not exceed 96 h.

Improved molecular markers and marker detection methods should result in the early detection of tumours smaller than 1 cm. The integration of various diagnostic data should support

a reliable decision support system for diagnostics and the selection of treatment options. It should be achieved with a minimum of diagnostic markers and *in vivo* imaging. Molecular markers are preferred for cost reasons.

IV. LOCCANDIA BIO ASPECTS

Standards protocols for extraction of blood samples and affinity columns for protein isolation are used. The preparation of proteins to make synthetic proteins mixture, the production of the specific antibodies and the design of the affinity columns and the quality control of the proteins using MALDI-MS are the main goals of this module.

For the development of the analytical chain three proteins with different structural features reflecting the heterogeneity of human tumour biomarker have been selected: 1) PAP1 is a small secreted factor with an elevated expression in acute pancreatitis and pancreatic tumours. 2) CEA is a large glycoprotein and an established tumour marker. 3) p120 Catenin is an intracellular protein which promotes pancreas tumour growth. In particular, p120 can be expressed in more than 20 isoforms, within the project is planned to be able of detecting various p120 Catenin isoforms.

A. PAP1 Protein and Antibody Production

Human PAP1 protein has been produced as a recombinant protein in bacteria (*E. coli*). cDNA coding for the human PAP1 protein has been cloned and modified for adequate bacterial production of the full length sequence of human PAP1 protein and introduced into a procaryotic expression vector. Grown bacteria were harvested and broken to get access to the expressed proteins. After several washing procedures the isolated bacterial proteins were dissolved. Thereby isolated proteins were further refined using an affinity-purification column. Afterwards, buffer and solubilizing conditions for the purified PAP1 were tested and adjusted by means of dialysis. . PAP1 protein exhibiting the predicted molecular weight is produced in quantities of more than ten milligrams (Fig. 1, left panel)

In order to design and produce antibodies specific for PAP1 the protein sequence and structure were analysed. Primary and secondary structure analysis was performed to select peptide sequences which provide a high probability for hydrophilicity and antigenicity. Finally, peptide sequences of 16 amino acid residues in length were chosen that are likely to be presented on the proteins surface and exhibit low homology to other human proteins. Corresponding peptides were chemically synthesized and injected into chicken repeatedly to cause an immune response and production of specific antibodies against PAP1. Eggs were collected and antibodies were separated from the egg yolk using several steps of lipid/protein extraction. Subsequently, the isolated antibodies were tested for reactivity and specificity. Produced antibodies effectively recognize and bind to immobilized PAP1 protein (Fig. 1, center panel). Moreover, we succeed in

demonstrating the binding of anti-PAP1 antibodies to the protein in organelles of human pancreas tumour cells by immunofluorescence microscopy (Fig. 1, right image).

With the combination of recombinant PAP1 and specific PAP1 antibodies we have generated tools for the development of an affinity module for the specific isolation of PAP1 from human blood samples.

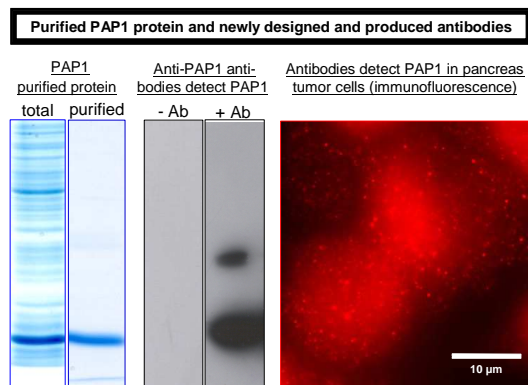


Figure 1 Total protein extract and purified recombinant PAP1 protein (left panel). Purified anti-PAP1 antibodies detect and bind to PAP1 protein (+Ab, center panel). Same protein extracts on filter membrane, incubated with non-specific antibodies show no reaction at all (-Ab, center panel). Anti-PAP1 antibodies detect native human PAP1 in pancreas tumour cells (right image). Anti-PAP1 antibody is able to detect native PAP1 produced by human tumour cells, an important prerequisite for the project

B. The affinity module

Protein isolation from plasma sample is performed by means of an affinity module. Particularly, the development of this module is based on the Immunoaffinity Chromatography (IAC), which exploits the interactions between an immobilized antibody and the analyte (antigen or hapten) to extract very selectively the analyte from a complex sample matrix. In our case, such a matrix consists of plasma sample that is known of its perplexity. Moreover, the enrolment of the immuno-adsorber (specific antibody raised against a specific antigen-protein) leads the sample pre-treatment to be reduced to a one-step solid-phase extraction and purification that can be in the order of thousand-fold.

The principal stages of the immuno-affinity module are the following: First, the antibodies are immobilized on a support matrix, which has been chosen to be the surface of magnetic beads. Further, the immuno-adsorber is loaded with the crude sample (plasma). The analyte (hapten) is retarded on the column by selective interactions between antibody and analyte. In a washing step, the non-interest molecules are washed out. The final step is the elution. In order to achieve a maximum enrichment of the analyte, the elution volume is very small and the elution agents that are recruited comprise of volatile buffers. The decision for the usage of the latter has been taken owing to the requirements of the digestion module for low salt

content of the sample. The properties of the elution buffers are defined according to the magnitude of the extraction, which should at least approach an optimum recovery.

The product data sheet provides several protocols for ligand coupling which have been evaluated in a semiquantitative manner (SDS-PAGE) in order to decide which one should serve better. According to the same data sheet the first protocol is less laborious and most of all gives higher yields but is not recommended when using ligands that contain carboxylic acid groups such as proteins, which react with the carbodiimide and cause polymerization of the ligand. In this case, excess of the latter should be applied to compensate for the loss due to the phenomenon of polymerization. The second one protocol seems to be suitable when limited amounts of the ligand are available but it requires a very fast wash of the beads in cold buffer after activation. The last protocol should be preferred when the ligand is in alkaline buffer or a buffer with high phosphate concentration.

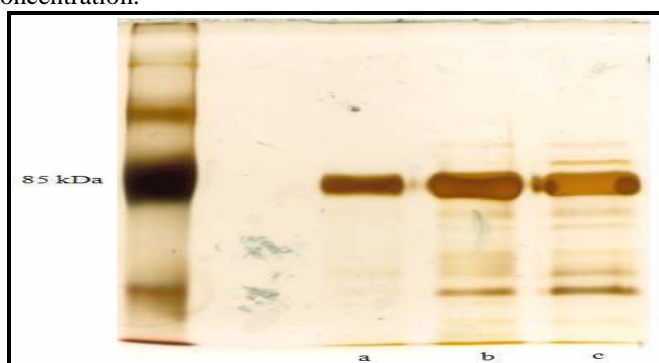


Figure 2: Silver staining results of an 8% SDS-PAGE: on the left is a prestained protein marker and on the right there are: a) one-step coupling protocol, b) two-step coupling protocol without NHS and c) two-step coupling protocol with NHS.

After coupling the anti-PAP1 antibodies with the beads, we proceeded with running the denatured ligand on an 8% SDS-PAGE. In particular, we have recruited half of the final volume of the coupled beads, which had been adjusted to 66 μ l with withphosphate-buffered saline pH 7.4. The visualization of the bands had been accomplished by a silver stain, Figure 2. The results indicate that the second protocol performs better in comparison with the other two. Our present efforts are focused on the evaluation of the PAP1 protein isolation from non-crude synthetic samples of different concentration in order to evaluate the dynamic range of the system to recover the protein.

V. LOCCANDIA NANO ASPECTS

Microfluidic chip platforms for manipulating liquid volumes in the nanoliter range are serious alternatives to current robotic systems. Automation will continue to be a bottleneck for standard laboratory approaches. In principle, lab-on-chip merges multiple functions onto one chip, attaining high sensitivity through nanoliter range volumes and the elimination of connections and valves that may introduce dead

volumes or leakages that produce unreliable results. Two integrated modules are being developed based on a micro-pillar technology on silicium wafer [1,2]: a digestion module and a nano-liquid chromatography-electrospray ionisation module.

A. Silicon Microchip for LC/MS Analysis of Peptides Mixtures

The silicon microfluidic devices, called E-spray chips, integrate both a reversed-phase separation column and a nano-electrospray emitter. Microchips are fabricated following standard steps of silicon microtechnology. The separation column is a perfectly ordered 2-dimension array of squared micro-pillars, directly etched in the silicon substrate, and the electrospray emitter is a planar nib-like nano-tip. Two kinds of reversed-phase coatings are grafted in our columns using (i) a "chip by chip" process in liquid phase with a C18-alkylated silane and (ii) a collective process in vapor phase with a C10-perfluorated silane. Analyses of standard tryptic digests of cytochrome c in hydrodynamic pumping mode have demonstrated a spray of good quality and effective separation performances of these microdevices with a higher retention capacity for C10-perfluorated ones. These new microchips can be produced at a very large scale with a mass production process from the microfabrication to the chemical treatment. An optimization of the nano-electrospray emitter geometrical parameters is under development in order to avoid overflows that sometimes occur around the nano-electrospray emitter and to enhance spray robustness. As well, an optimization of the analytical conditions, in terms of gradient composition and flow rates and a complete characterization of the separation column in terms of capacity, plate height, and sensitivity are currently performed. Various surfaces over volume ratios, micropillars shapes and gaps, column lengths and widths will then be tested to improve the column efficiency. Analysis of more complex peptides mixtures prepared in collaboration with LOCCANDIA partners will soon be considered.

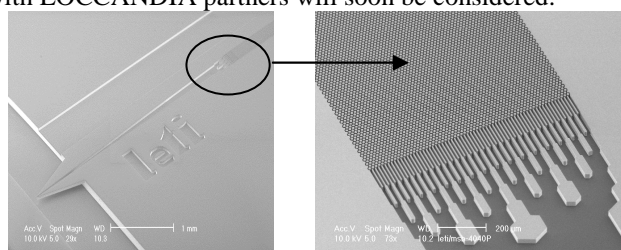


Figure 3: Nano-LC separation module. The separation column is a perfectly ordered two-dimension array of squared micro-pillars, directly etched in the silicon substrate (left). The ESI emitter is nib-like tip integrated directly on silicon at the end of the chromatographic column (right).

VI. LOCCANDIA INFO ASPECTS

In this part we focus on the presentation of the design of an integrated information management and analysis platform for a full-chain clinico-proteomic analysis. The various functional modules of such an information system are

discussed, and the architectural principles as well as the technological standards adopted for the implementation of the system are presented.

Info part of the project addresses the complexities of data generated from the analysed mixture and is focused on delivering methods and tools for improving the measurement reliability, and ultimately providing a robust and easy to use system. As a result, the LOCCANDIA Information Management System (LIMS) has been developed. LIMS is an innovative integrated clinico-proteomics computational environment that has been designed, combining standard-based informatics systems and best-of-breed computational modules, to support the LOCCANDIA integrated lab-on-chip based diagnostic device. LIMS will allow physicians to interpret the digital spectrogram data delivered by the mass spectrometer and to provide support to set their diagnosis.

Managing heterogeneous information requires the design and implementation of a “clinic-proteomics database”, which must adhere to strict rules ensuring clarity and quality. The common format should include all proteome data (e.g. peak lists), linked when possible with other characteristics obtained from fractionation together with the study protocol, sampling procedure and sample preparation, calibration, matching and quality control criteria, as well as clinical data. Such a database does not currently exist in the public domain.

Therefore, a database was designed and developed, by building on the outputs of existing standardization initiatives and was integrated into the LIMS. The LIMS architecture and the main integrated modules for reconstruction, classification, identification and visualization are described below.

A. The LIMS Architecture

LIMS is a web-based application, responsible for the storage, examination and manipulation of clinical and proteomic data. LIMS acts as a mediation platform for the integration and data. Its ultimate objective is to intelligently correlate clinical and proteomic information towards LOCCANDIA’s goal of early pancreatic cancer diagnosis. The application was designed and developed using the “three-tier” software architecture. This is a well known client-server architecture that clearly distinguishes the different modules that participate in a web-based application software.

Apart from the usual advantages of modular software with well defined interfaces, the three-tier architecture is intended to allow any of the three tiers to be upgraded or replaced independently as requirements or technology change.

B. The Reconstruction Module

As seen before, the diagnosis is based on the concentration estimation of a panel of targeted proteins chosen as markers of the disease. The profile reconstruction module realizes this quantification from LC-MS spectrograms, at the output of the analysis chain. The proposed approach is based on

chemometrics methods and is more global than peak extraction methods since it takes into account the whole 2D signal. It relies on a functional model linking the unknown molecular concentration profile and the spectrograms measurements. The chosen techniques are called PLS (Partial Least Square), NPLS, CLS (Classical Least Square) and PARAFAC (PARallel FACtor Analysis). It enables to model the variations of a given number of variables X for which the direct measurement is not possible (in our case, the proteins concentration), as a function of others variables Y easily measurable (the 2D spectrograms at the output of the analysis chain). The reconstruction module includes two steps: the pre-processing step (conversion and interpolation of the MS data, time delay correction, chemical noise suppression, data smoothing) and the processing step which corresponds to the estimation of the concentration of the targeted proteins.

The reconstruction module has been tested on experimental datasets obtained with different complexity of mixtures (water, urine, plasma) and has provided highly accurate results with less than 10% of error.

C. The Identification Module

This module is responsible for identifying the proteins and peptides that are not in the initial targeted panel but it is possible to be present in the measurement. This module is based on the Phenyx software, which is a software platform for proteomic mass spectrometry data analysis. Phenyx is capable of identifying and characterizing proteins and peptides from mass spectrometry data, utilizing the OLAV true probabilistic scoring system for applying MS/MS scoring schemes [14].

D. The Visualization Module

The visualization module provides a sophisticated way of visualizing mass spectrometer raw data, reconstructed profiles and any relevant information for data analysis. Implementation of this module is based on MSight software platform. MSight is the software for displaying and browsing any portion of the collected mass spectra.

Phenyx is powered by Aldente algorithm in order to identify proteins from peptide mass fingerprinting data. That implies that an unknown protein is digested with a proteinase of known cleavage specificity and the resulting peptides measurement by mass spectrometry. Phenyx can perform a comparison between the experimentally measured peptide masses and the theoretical peptides (in a protein sequence database).

E. The Classification Module

The classification module aims to predict patients outcomes based on the measurements of the set of selected proteins. Additionally, relevant clinical data will be also used to improve the accuracy and sensitivity of the method. The module has been developed in R language and implemented in the Java environment of LIMS. Final users have a panel of

tools based on Logistic regression and Support Vector Machines with nine different kernels. These tools will be validated with ninety six real cases to assess the better methodology for reaching the better patient diagnosis. One major advantage of the selected methods are their capability of continue learning thus, as more the classifier is used for predicting new samples as better would be the predictive outcome.

VII. CONCLUSIONS

Adenocarcinoma of the pancreas is the fourth leading cause of cancer deaths in Western Countries. By the time of diagnosis, at least 80% of tumours are unresectable. The overall five-year survival rate is 4 percent, and localized, resectable disease has only a 17 percent survival rate.

The earliest stages of this cancer like other types of tumour remain poorly understood. Therefore, diagnosis based on biomarkers research has found several pitfalls as lack of specificity and accuracy. Attempts on detecting new biomarkers or on using several biomarkers to gain accuracy and specificity should be done for improving early detection of this kind of malignancies

In vitro diagnostics state of the art and gold standard for end users are absolute and validated concentrations of molecular markers (e.g. ng/ml). The analysis of protein markers requires a period of 4 working days. For each molecular diagnostic marker is a concentration range for healthy individuals and diseases defined. Disease causes a shift in the concentration range of relevant molecular markers. These concentration ranges are overlapping for many markers, a single marker is therefore not indicative and sufficient for diagnosis.

The advent of new technological improvements based on developing of lab-on-chips components that are able to carry out operations at nano-litter scale providing higher sensitivity and specificity. Using this microsystem coupled with mass spectrometry leads to get better reproducibility of the experiments.

Mass spectrometry (MS) approaches are very attractive to detect protein panels and protein isoforms in a sensitive way. Information generated using MS needs to be managed and analysed to be useful for physicians. Improvements on using mathematical and statistical models for analysing complex and heterogeneous data will lead researches and doctors to extract knowledge, uncover potential 'hidden' patterns in the data and discover biomarker models of predictive power that would support them in doctors' clinical decisions. However, the application of MS to clinical diagnosis is still at its beginning. The need for new and relatively simple devices to allow for the translation of these research results to clinical practice is urgent.

LOCCANDIA project intends to detect the disease at tumour-earliest-stage by measuring circulating proteins by applying an innovative integrated lab-on-chip system coupled to mass spectrometer and supported by innovative IT developments. Early diagnostics, providing advanced protein profiling information enables patients/citizens to get the right information at the right time in the right context. This will allow them to improve their health by early detection of cancer before symptoms appear, by a personalized medicine through molecular fingerprints, and by ubiquitous access to the information through the development of point of care devices.

The big step offered by advanced protein profiling is to early detect and monitor a disease through an analysis of protein fingerprints in plasma probes. Early detection of disease by following simple molecular parameters is a main objective of modern medicine. The diagnostics of protein fingerprints might help also to specify the relevant chemotherapy treatment of cancers. There is a need of automatically operating integrated devices such as the lab-on-chip. LOCCANDIA is a starting point to allow clinical routine mass spectrometry based in vitro diagnosis. There is also a need for information technology to improve the ease-to-use and the reliability of this measurement process, and the access to complementary patient information.

The development of an innovative lab-on-chip and the associated analysis chain including biomedical informatics (BMI) technology will strengthen healthcare industries. Innovation in LOCCANDIA is based not only on the development of a full analytical chain that aims to improve the state-of-the-art but also on the seamless integration of a bio, nano and information processing stages for the development of a novel integrated diagnostic system. From the commercial point of view, the innovation brought by LOCCANDIA will be major marketing arguments to reinforce industry competitiveness

The successful achievements of LOCCANDIA objectives will lead to revolution the field of cancer diagnostic.

ACKNOWLEDGMENT

LOCCANDIA project is partially funded by the European Commission under the Six Framework Program. Contract number 034202.

REFERENCES

- [1] Calleri, et al. J. Chromatogr. A 2004, 1045, 99-109.
- [2] Combette, P.; Constantin, O.; Marchand, G.; Vinet, F.: FR pat. 0206084, 2002.
- [3] de Mello, A. J. Lab on a chip 2001, 1, 7N-12N.
- [4] Constantin, O.; Sarrut, N.: FR pat. 450173, 2004.
- [5] Gygi, et al. Nat. Biotechnol. 1999, 17, 994-999.
- [6] Iulia M. Lazar, J. G., Frantisek Foret, Mass Spectrometry Reviews 2006, 25, 573-594.
- [7] Jin, L. J.; Ferrance, J.; Sanders, J. C.; Landers, J. P. Lab on a Chip 2003, 3, 11-18.

- [8] Lion, N.; et al. H. *Electrophoresis* 2003, 24, 3533-3562.
- [9] Oda, Y., Huang, K., Cross, FR., Cowburn, D., Chait, B. *Proc. Natl. Acad. Sci. U.S.A* 1999, 96, 6591-6596
- [10] Yan, W.; Chen, S. S. *Briefings in functional genomics and proteomics* 2005, 4(1), 27-38.
- [11] Zhou, H., Ranish, JA., Watts, JD., Aebersold, R. *Nat. Biotechnol.* 2002, 20, 512-515.
- [12] Rosty C, Christa L, Kuzdzal S, Baldwin WM, Zahurak ML, Carnot F, Chan DW, Canto M, Lillemoe KD, Cameron JL, Yeo CJ, Hruban RH, Goggins M. Identification of hepatocarcinoma-intestine-pancreas/pancreatitis-associated protein I as a biomarker for pancreatic ductal adenocarcinoma by protein biochip technology. *Cancer Res.* 2002 Mar 15;62(6):1868-75.
- [13] Xie MJ, Motoo Y, Iovanna JL, Su SB, Ohtsubo K, Matsubara F, Sawabu N. Overexpression of pancreatitis-associated protein (PAP) in human pancreatic ductal adenocarcinoma. *Dig Dis Sci.* 2003 Mar;48(3):459-64.
- [14] Ni XG, Bai XF, Mao YL, Shao YF, Wu JX, Shan Y, Wang CF, Wang J, Tian YT, Liu Q, Xu DK, Zhao P. The clinical value of serum CEA, CA19-9, and CA242 in the diagnosis and prognosis of pancreatic cancer. *Eur J Surg Oncol.* 2005 Mar;31(2):164-9.
- [15] Molina LM, Díez M, Cava MT, Maestro ML, Ortega MD, Mendíz JG, Prego A, De Diego JA, Balibrea JL. Tumour markers in pancreatic cancer: a comparative clinical study between CEA, CA 19-9 and CA 50. *Int J Biol Markers.* 1990 Jul-Sep;5(3):127-32.
- [16] Mayerle J, Friess H, Büchler MW, Schneckeburger J, Weiss FU, Zimmer KP, Domschke W, Lerch MM Up-regulation, nuclear import, and tumour growth stimulation of the adhesion protein p120 in pancreatic cancer. *Gastroenterology.* 2003 Apr;124(4):949-60.
- [17] Seidel B, Braeg S, Adler G, Wedlich D, Menke A. E- and N-cadherin differ with respect to their associated p120ctn isoforms and their ability to suppress invasive growth in pancreatic cancer cells. *Oncogene.* 2004 Jul 15;23(32):5532-42.