Dissertation

submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

presented by

MSc Satoshi Okawa
born in: Tokyo, Japan
suggested oral-examination: 11/July/2013

Technical advances in mass spectrometry-based proteomics and its application to the characterisation of glioma-initiating cells

Referees: Dr. Lars Steinmetz
Prof. Bruce Edgar

Abstract

A typical workflow for current liquid chromatography coupled to mass spectrometry (LC-MS) based proteomics employs the mass, charge and peptide fragmentation pattern for peptide/protein identification. In this thesis I employ this technique for profiling of cellular proteomes applied here to the analysis of glioblastoma stem cells and neuronal stem cells to gain insight into the cancer signature in the former cell type. At the same time, I explore improvements for proteomic workflows by better exploiting experimental parameters produced in proteomic experiments. Specifically, peptide isotope patterns are not routinely used in the current proteomics workflows, yet they are available in any mass spectrum that is generated. Therefore, we explorered the use of isotope patterns in MS1 mass spectra to support peptide identification. We demonstrated that the relative isotope abundance (RIA) error is 4-5%, and that this is only modestly influenced by spectral intensity, resolution and the number of MS1 scans. The current RIA accuracy has limited discriminatory power at a proteome-wide scale. At the same time, the analysis was hampered by the difficulty in calculating FDRs, particularly in constructing proper decoy databases that are similar in size as the target database, yet different in molecular composition for all peptides considered. Alternative strategies to calculate FDRs will be required to address this issue for complex proteomes. Regardless, the utility of RIA may become relevant with future instrument developments, considering that even a relatively modest decrease in RIA error down to <1% strongly improves discriminatory power. Alternatively, at increased mass accuracy even current RIA accuracy levels may be sufficient to fit isotope patterns as a constraint in the peptide identification process as a parameter that comes for free in any MS-based proteomic experiment.

Glioblastma is the most severe form of brain tumour and there has been strong evidence that brain cancers arise from a small population of cells known as cancer stem cells (CSCs) within the tumour that exhibit normal stem cell characteristics, i.e., long-term self-renewal, longevity and capacity to differentiate into adult cells. To reveal the difference in global protein expression between malignant neural stem cells derived from adult gliomas (GNSs) and untransformed, karyotypically normal foetal neural stem cells [1], we performed mass spectrometric analyses of both total cell proteome and secreted proteome of these cells. This resulted in a total of ~7500 and ~2000 quantified proteins and 446 differentially expressed proteins (152 up-regulation, 294 down-regulation in GNSs) and 167 differentially expressed proteins (144 up-regulation, 23 down-regulation), respectively. After data analyses, several candidate proteins for surface markers that could distinguish between NSs and GNSs were experimentally validated using immunocytochemistry. Next, candidate GNS-secreted factors that could mediate tumourigenic

transformation of NSs were evaluated using time-lapse imaging and colony-forming assays. Both of these experiments produced some positive results, demonstrating the power of proteomics. More experiments are, however, necessary to solidify these findings.

Zusammenfassung

Der derzeit klassische Ablauf einer Peptid/Protein Identifikation mittels Flüssigkeitschromatography mit gekoppeltem Massenspektrometer (LC-MS) beruht auf Masse, Ladung und Fragmentierung der Peptide. In dieser Doktorarbeit wende ich diese Technik zur Charakterisierung von zellulären Proteomen an, um Gliomstammzellen und neuronale Stammzellen zu analysieren um so Einblicke in die Krebssignatur der frühen Zellarten zu erhalten. Parallel untersuche ich in proteomischen Arbeitsabläufen Verbesserungen durch bessere Ausnutzung von experimentellen Parametern, die in proteomischen Experimenten auftreten. Besonders die Isotopenverteilungen von Peptiden wird generell nicht in proteomischen Methoden verwendet, obwohl sie in jedem akquirierten Spektrum enthalen sind. Daher untersuchten wir die Verwendung von Isotopenmustern in MS1 Massenspektren, um die Peptididentifikation zu unterstützen. Wir belegten, dass der relative Isotopenhäufigkeits (RIA)-Fehler 4-5% beträgt und dieser nur gering von Spektrumintensität, Auflösung und Anzahl an MS1 Spektren abhängt. In kompletten Proteomanalysen hat die derzeitige RIA-Genauigkeit eine limitierende Trennschärfe. Mit einherging, dass die Analyse durch die Schwierigkeit der FDR Berechnung beeinträchtigt wurde, insbesonders bei der Erstellung richtiger Zufallsdatenbanken, die ähnlich in der Größe aber unterschiedlich in der molekularen Zusammensetzung aller enthaltener Peptide der Zieldatenbank sind. Alternative Strategien zur Berechnung der FDR werden vonnöten sein um dieses Problem in komplexen Proteomanalysen Herr zu werden. Trotzdem wird die Nützlichkeit von RIA mit künftigen instrumentellen Entwicklungen vielleicht relevant werden, wenn berücksichtigt wird dass eine nur geringe Senkung des RIA-Fehlers unter 1% eine starke Verbesserung der Trennschärfe bewirkt. Alternativ wäre mit Zunahme der Massengenauigkeit vielleicht sogar der derzeitige RIA Genauigkeitslevel ausreichend, um Isotopenmustern als Nebenbedingung in Peptididentifikationen einzubinden, da sie als Parameter „umsonst“ in jedem MS-basierden proteomischen Experiment vorhanden sind.

Das Glioblastom ist der häufigste bösartigste Gehirntumor und es gibt starke Hinweise, dass Gehirntumore aus einer geringen Population von Zellen, bekannt als Krebsstammzellen (CSCs) erstehen. Der Tumor weist normale Stammzell-Charakteristika auf, wie langfristige Selbsterneuerung, Langlebigkeit und die Fähigkeit sich in adulte Zellen zu differenzieren. Um die Unterschiede der globalen Proteinexpression zwischen bösartigen neuralen Stammzellen, die aus adulten Gliomen (GNSs) entstehen und unveränderten, karyotypischen normalen fötalen neuronalen Stammzellen [1] zu lüften, erbrachten wir massenspektrometrische Analysen von beidem, dem totalen Zellproteom und dem sekretierten Proteom dieser Zellen. Dies resultierte in ~7500 und ~2000 quantifizierten Proteinen und 446 unterschiedlich exprimierten Proteinen (152 hoch- und 294 herunterreguliert in GNSs) beziehungsweise 167 unterschiedlich exprimierten Proteinen (144 hoch- und 23 herunterreguliert). Nach der Datenanalyse konnten mehrere Proteinkandidaten als Oberflächenmarker zwischen NSs und GNSs unterschieden und mittels Immunozytochemie

validiert werden. Weiterhin wurden Kandidaten der GNS-sekretierten Faktoren, die eine tumorgenetische Veränderung der NSs vermitteln können, mittels Zeitraffer Aufnahmen und koloniebildende Assays evaluiert. Beide Experimente lieferten positive Ergebnisse, die die Bedeutung der Protoemics demonstiert. Allerdings sind weitere Experimente nötig, um diese Resultate zu festigen.

Table of Contents

Abbreviations

Abbreviation	Description
AHA	azidohomoalanine
CSC	cancer stem cell
DAPI	4',6-diamidino-2-phenylindole
DE	differentially expressed
DTT	dithiothreitol
EBI	Europian Bioinformatics Institute
ECM	extracellular matrix
ES	embryonic stem cell
FDR	false discovery rate
FWP	features with PSM
GBM	glioblastoma multiforme
GNS	glioma neural stem cell
GO	gene ontology
GSEA	gene set enrichment analysis
IAA	indoacetamide
IEF	isoelectric focusing
iPS	induced pluripotent stem cell
LDA	linear discriminant analysis
LS-MS/MS	liquid chromatography coupled to tandem mass spectrometry
MS	mass spectrometry
NPC	neural progenitor cell
NS	neural stem cell
OPC	oligodendrocyte progenitor cell
pSILAC	pulsed SILAC
PSM	peptide spectrum match
QDA	quadratic discriminant analysis
RIA	relative isotope abundance
SILAC	stable isotope labelling with amino acids in cell culture
SVM	support vector machine
TF	transcription factor
TSG	tumour suppressor gene
TT	tumour tissue
UCL	University College London

Chapter 1

1 Introduction

Cellular homoeostasis is regulated by genes, which encode RNAs, or transcripts, which are then translated into proteins. RNAs/proteins intricately associate with each other and rarely function in isolation. Protein-protein interactions can mediate cellular signalling and formation of protein complexes with diverse functionality. Protein-DNA/RNA interactions regulate transcriptional/transnational regulations of protein expression. For this reason, simultaneous identification/quantification of RNAs/proteins is crucial for the understanding of how these molecules collectively execute biological functions. Quantification is important since protein functions are governed by their concentrations in the cell and some proteins, such as transcription factors, can affect a phenotype even at very low concentrations and by subtle changes in concentration.

Microarray/RNA-seq and mass spectrometry (MS) are widely used for transcriptomics and proteomics, respectively, which can allow for identification and quantification of thousands of transcripts/proteins in a single sample. While transcriptomics is able to quantify a complete set of RNAs, MS-based proteomics is unable to identify a complete set of proteins of higher eukaryotes, such as human, in a reasonable time frame, due to high complexity and dynamic range of protein abundance within a cell. To make things worse, proteins are often “decorated” with post-translational modifications (PTMs), such as phosphorylation and ubiquitination, and confidently identifying these PTMs requires additional steps for sample preparation, fractionation and data analysis. A typical workflow for current MS-based proteomics consists of peptide separation by high-performance liquid chromatography (HPLC), peptide ionization, acquisition of a full mass spectrum (MS1), followed by fragmentation of selected precursor ions and acquisition of MS/MS spectra (MS2). These MS/MS spectra are then compared to masses generated *in silico* from candidate peptides derived from a target protein sequence database. Currently, however, about 50 % of all acquired MS2 spectra are not confidently identified due to poor spectral quality. Furthermore, even if peptides are detected by mass spectrometer, the majority of those peptides are not targeted for tandem MS and therefore remain unidentified [2]. A typical workflow for current LC-MS based proteomics employs the mass, charge and peptide fragmentation pattern for peptide/protein identification. However, LC-MS-based proteomics generates other parameters such as LC retention time and peptide isotope patterns, which are not routinely used in the current proteomics workflow.

Chapter 2 explores the use of isotope patterns in mass spectra for peptide identification. Because the isotope pattern are determined by the atomic composition of a peptide, it could be used as a signature of its sequence. To this end, first, the instrument accuracy of relative isotope abundance (RIA) measurement obtained from various proteomic datasets was examined. Then we set up a strategy to expore how the isotope patterns could aid in peptide identification either using RIA alone or by combining RIA and MS2-based database searches using *E. coli*, *S. cerevisiae* and human samples as model systems of various complexities. Furthermore, a theoretical framework describing the combined contributions of mass accuracy and RIA for confident peptide identification was investigated. The utility of RIA may become relevant with future instrument developments. Alternatively, at increased mass accuracy even current RIA accuracy levels may be sufficient for improving peptide identification.

Despite its limitation to the number of protein identification, MS-based proteomics has some advantages over transcriptomics. For example, the mRNA abundance does not always correlate with that of proteins due to post-transcriptional regulations. Since proteins are the core executors of biological processes, phenotypes should be more accurately inferred from proteins than from mRNAs. In addition, secreted proteins can only be accurately quantified on the protein level. Therefore, Chapter 3 describes a quantitative proteomic study to characterize cells that initiate a tumour, by comparing glioblastoma-derived stem cells to neural stem cells. This project is a collaboration with Dr. Steven Pollard (UCL) and Dr. Paul Bertone (EBI).

Glioblastma is the most severe form of brain tumour according to the World Health Organization. There is strong evidence that brain cancers arise from a small population of cells within the tumour that exhibit normal stem cell characteristics, i.e., long-term self-renewal, longevity and capacity to differentiate into adult cells. These cells are called cancer stem cells (CSCs), or tumour initiating cells, and unless these cells are eradicated, tumour will rise again. Since CSCs exist in a small population, in vitro culture techniques are required to study biological properties of these cells, and to design strategies for anti-cancer therapy. However, culturing these cells in conventional serum-containing media causes loss of multipotency and induction of differentiation [3]. In addition, repetitive passaging will lead to accumulation of de novo mutations that may not have any relevance to the original tumour [3]. To overcome these problems and keep cells in the multipotent state, Pollard et al [4] developed a protocol for culturing human glioma neural stem cells (GNSs), in a very homogeneous fashion without losing their tumourigenic capacity. To identify key proteins/pathways that are the hallmarks of malignant GNSs, here we aim to globally characterise the protein expression of these malignant GNSs by having normal, untransformed foetal neural stem cells [1] as the control. Both total cell- and secreted proteomics analyses are performed in order to identify key proteins/pathways in the cell and secreted auto-

/paracrine factors mediating tumourigenesis. Since the gene expression of individual gliomas differ considerably from one another [5], we used four NS lines and four GNS lines from different individuals, so as to minimize the effect of inter-individual variability. Furthermore, we obtained tag-seq data on the NS and GNS lines from Dr. Paul Bertone and glioma tissue microarray data from public databases and combined them with the proteomics data to gain further insight into the biology. Following the data analyses, we selected candidate proteins for markers for malignant phenotypes and conducted immunocytochemistry for validation. We also selected candidate proteins for auto/paracrine factors for mediating tumourigenesis by adding these proteins to the culturing media and observing phenotypic changes over time. Colony-forming assays were also made with addition of these candidate proteins to the media and the number of colonies was counted as a readout for proliferation. Finally, the prospects of the candidate proteins as well as the global picture of protein expression between the NSs and GNSs is discussed.

In Chapter 4 I discuss the two projects in more general terms and conclude by mentioning the outlook and future plan for the projects and potential implications of the projects for future research in general.

Chapter 2

2 Properties of isotope patterns and their utility for peptide identification in large-scale proteomic experiments

2.1 Introduction

2.1.1 Peptide isotope patterns are not used in the standard proteomics workflow

Mass spectrometry (MS) is widely used in proteomics to identify and quantify proteins in biological samples. A typical workflow for current MS-based proteomics consists of peptide separation by high-performance liquid chromatography (HPLC), peptide ionization, acquisition of a full mass spectrum (MS1), followed by fragmentation of selected precursor ions and acquisition of MS/MS spectra (MS2) [6]. These MS/MS spectra are then compared to masses generated *in silico* from candidate peptides derived from a target protein sequence database. Several search engines and scoring algorithms have been developed for this task [7] such as SEQUEST [8], Mascot [9], OMSSA [10], X!Tandem [11] and Andromeda [12]. Although search engines have been the gold standard in MS-based proteomics, additional parameters independent of MS2 information have been used to support peptide identification, such as peptide retention time, isoelectric point and accurate mass. Retention time prediction calculates the expected elution time for a peptide of a given sequence and for particular chromatographic conditions [13-17] to narrow down the number of peptide candidates expected in a retention time window. Similarly, when isoelectric focusing is used for first-dimension peptide separation, the predicted versus observed isoelectric point can be used to increase confidence in peptide identification [18]. In modern mass spectrometers, especially using TOF, Orbitrap and FT-ICR detectors, the mass of intact peptides can be determined with high accuracy in MS1, greatly restricting the number of peptides that need to be considered and contributing to low false discovery rates (FDR) in large-scale proteomics [19].

A parameter that has not been considered extensively for peptide identification, but one that is readily available in any high-resolution mass spectrometric experiment is the isotope pattern in MS1 spectra. The isotope pattern is determined by the atomic composition of the peptide (or any compound) which is reflected in the relative intensities of isotope peaks (with contributions mainly from ^{13}C, ^{15}N, ^{18}O). Therefore, this is a compound-specific parameter that can be calculated from

any given elemental composition. Isotope patterns have been used for various purposes such as charge state assignment [20], determination of monoisotopic mass and deisotoping [21], filtering of non-peptide features [22], estimation of protein turnover rate [23], and detection of peptides containing particular atoms such as mercury [24]. However, their use as a general constraint for peptide identification is largely unexplored. This is somewhat different in metabolomics where isotope patterns are employed routinely to compute possible atomic compositions, thereby reducing the number of candidate compounds [25-30]. A similar approach may become helpful in proteomics where it may serve as a parameter to support peptide identification. The primary challenge in the implementation of RIA as a constraint pertains to the complexity of the proteome, where it needs to be demonstrated that experimentally determined isotope patterns have the discriminatory power to distinguish closely related peptides. Therefore, in this study we have investigated the accuracy of relative isotope abundance in large-scale proteomic datasets and its dependence on various mass spectrometric parameters. In addition, we explored various strategies to quantify the discriminatory power of isotope patterns in the context of proteome analyses of various complexities (*E. coli*, S*accharomyces cerevisiae*, and human cells). Finally, we provide a theoretical framework describing the combined contributions of mass accuracy and relative isotope abundance for confident peptide identification.

2.2 Materials and methods

2.2.1 Sample preparation for mass spectrometry

E. coli and yeast were grown under standard conditions, lysed, and homogenized in 50 mM ammonium bicarbonate. HeLa cells were grown overnight in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal calf serum. The lysates were concentrated using Amicon Ultra Centrifugal Filters (0.5 ml, 3 kDa cutoff) (Millipore). Disulphide bonds were reduced with 100 mM dithiothreitol (DTT) and cysteine residues were alkylated with 200 mM iodoacetamide (IAA) followed by protein digestion with sequencing grade modified trypsin (Promega) overnight at 37°C. Peptides were acidified and desalted with C18 Stagetips (Empore 3M) [31] and then dried down by vacuum centrifugation. Human peptides were fractionated into 12 fractions on an Agilent 3100 OFFGEL Fractionator (settings as described by the manufacturer) using Immobiline DryStrips (ph 3-10 NL, 13 cm, GE Healthcare). Dried samples were resuspended in 360 ml H2O and diluted into 1.44 ml 1.25 x IEF stock solution (6% glycerol, 2% Ampholytes pH 3-10 (1:50)). Focusing was performed at a constant current of 50 mA with a maximum voltage of 8,000 V until 20 kVh. After

IEF, peptides were acidified, desalted, and dried as described above, and then reconstituted with 4% acetonitrile in 0.1% formic acid.

2.2.2 Liquid chromatography coupled to mass spectrometry (LC-MS/MS)

Peptides were analyzed by nanoflow LC coupled to an LTQ Orbitrap Velos (Thermo Fisher Scientific) using a Proxeon nanospray source. Reverse phase chromatography was performed with a nanoACQUITY UltraPerformance LC system (Waters) fitted with a trapping column (nanoAcquity Symmetry C18, 5 µm, 180 µm x 20 mm) and an analytical column (nanoAcquity BEH C18, 1.7 µm, 75 µm x 200 mm) directly coupled to the ion source. The mobile phases for LC separation were 0.1% (v/v) formic acid in LC-MS grade water (solvent A) and 0.1% (v/v) formic acid in ACN (solvent B). Peptides were separated at a constant flow rate of 300 nl/min with a 3 to 40% solvent B gradient (120 min for *E. coli* and yeast, 145 min for IEF fractions of HeLa cells). The MS1 scan was acquired in the Orbitrap from m/z 300 to 1,700 at a maximum filling time of 500 ms and10^6 ions. The resolution was set to 30,000 (at m/z 400) unless stated otherwise. Fragmentation was performed in the LTQ by collision induced dissociation, selecting up to 15 most intense ions (top15) at an isolation window of 2 Da, unless stated otherwise. Target ions previously selected for fragmentation were dynamically excluded for 30s with relative mass window of 10 ppm. A lock mass correction was applied using a background ion (m/z 445.12003).

2.2.3 Feature extraction

Raw data files were converted into mzXML files in centroid mode with MM file Conversion Tool available at http://sourceforge.net/projects/massmatrix/files/MM_File_Conversion.zip/download. We also tested the profile mode, however, this did not change the result. All raw mz- and intensity values were extracted from mzXML files with the mzR R package (http://www.bioconductor.org/packages/2.9/bioc/html/mzR.html) and processed with a C++ programme written in-house. The source codes (as well as those for the scripts described below) are available upon request. For each MS/MS spectrum, the corresponding MS1 chromatographic feature was extracted from all the scans in a window +/- 0.3 min relative to the time of the MS/MS event. Isotopic peaks were extracted based on the theoretical distance calculated from the charge, (i.e., 0.5 Th for a +2 ion) with 10 ppm mass accuracy. Only the first three isotopic peaks were extracted since the 4^{th} and 5^{th} peaks were often missing. Scans that had a missing value in the first three isotopic peaks were discarded.

2.2.4 Decoy database construction

To use the FDR approach [32] for MS/MS sequence search, the decoy database was obtained by the "decoy.pl" script from (http://www.matrixscience.com/help/decoy_help.html) with the standard reverse mode and "random" mode. The latter generates random sequences with the same average amino acid composition as the input database.

2.2.5 Peptide identification and mass recalibration

Mascot generic files were generated with MM File Conversion Tool from the raw file and peptide identification was carried out with Mascot (version 2.2.06). The precursor mass tolerance was set to 5 ppm, fragment ion mass tolerance was 0.6 Da, cystein-carbamidomethylation as a fixed modification and methionine-oxidation as a variable modification. The number of missed cleavages was set to 1. The FDR was computed as the number of decoy Mascot top hits over the number of target Mascot top hits. Based on this FDR, the q-value was assigned to each Mascot top hit. This q-value function was interpolated for Mascot runner-up hits. Each peptide spectrum match (PSM) was matched back to the corresponding MS1 chromatographic feature, indicated as a feature with peptide spectrum match (FWP). Mass recalibration was performed based on top Mascot hits with a robust linear regression function in the R package MASS.

2.2.6 Relative isotope abundance (RIA) measurement error calculation

For each FWP, only MS1 scans having an intensity value above 50% of the top intensity of that FWP were extracted and averaged. The theoretical isotope pattern of the identified peptide (including fixed cysteine-carbamidomethylation and variable methionine-oxidation) was generated with an R programme written in-house. The RIA measurement error for each isotopic peak was calculated

as: $$\text{RIA error } j\ (\%) = 100 \text{ x } \left(\frac{E j}{\sqrt{I1^2+I2^2+I3^2}} - \frac{T j}{\sqrt{i1^2+i2^2+i3^2}}\right)$$

where $T = \begin{pmatrix} i1 \\ i2 \\ i3 \end{pmatrix}$ and $E = \begin{pmatrix} I1 \\ I2 \\ I3 \end{pmatrix}$

with E being the experimental isotope abundance (intensity) of the three isotopic peaks, T the theoretical RIA, and j indicating the isotopic peak index. Each isotopic peak was normalized by the Euclidian norm of the three isotopic peaks and this difference between the experimental- and theoretical isotope patterns was converted into the percentage.

2.2.7 Assessing discriminating power of RIA

To examine the power of RIA to discriminate between target and decoy hits upon a Mascot search, the ratio of the RIA error was computed as mean(E – T1) / mean(T1 - D1), where E is the vector of the experimental isotope pattern, T1 is the vector of the theoretical isotope pattern of the corresponding top target hit, and D1 is the vector of the theoretical isotope pattern of the decoy hit that is most similar to that of the target hit. Therefore, the scalars both in the nominator and numerator indicate the mean RIA error of the first three isotope peaks. Only if the ratio mean(E - T1) / mean(T1 - D1) is < 1, the experimental isotope pattern can discriminate the target hit from the closest decoy hit.

Similarly, to test the power of RIA to discriminate between candidate peptides in a Mascot search for each top Mascot hit–runner-up hit combination, the RIA error was computed and the minimum RIA error was taken as mean(T1 – Tr), where T1 is the isotope pattern of the top Mascot hit and Tr is the isotope pattern of the runner-up hit that minimizes mean(T1 - Tr). Only if the ratio mean(E -T1) / mean(T1 - Tr) is < 1, the experimental isotope pattern can discriminate the top Mascot hit from the closest runner-up hit.

2.2.8 In silico digestion of proteomes

The E. coli and human proteomes were obtained from SwissProt (http://www.uniprot.org/downloads), the yeast proteome from the Saccharomyces Genome Database (http://www.yeastgenome.org/). In silico digestion of these proteomes was done with the proteogest.pl script from (http://www.utoronto.ca/emililab/proteogest.htm) allowing 1 missed cleavage, while cysteine carbamidomethylation and methionine-oxidation were used as a fixed and variable modification, respectively.

2.2.9 Reclassification of target and decoy hits

The RIA error was scored as the mean of residuals by fitting the first three isotopic peaks to the theoretical isotope patterns using the “lm” linear regression function in R. LDA and QDA were perfomed using the R package MASS. SVM was performed using the R package e1071. The posterior probability was used for scoring re-classified peptides.

2.2.10 Generation of theoretical isotopic fine structures and RIA error calculation

The theoretical isotopic fine structures were generated using the most abundant peaks, namely C12H1N14O16, C12H1N15O16, C13H1N14O16, C13H1N15O16, C14H1N14O16 and

C12H1N14O16S34 (if sulphur is absent this value is 0). The summed RIA difference, rather than the mean RIA difference, over the peaks was calculated in this analysis in order to compare the RIA differences conferred by different numbers of peaks (i.e., three for normal isotope patterns and six for isotopic fine structures) with the same RIA accuracy.

2.3 Result and discussion

2.3.1 Accuracy of RIA in proteomic data

If isotope patterns are to be used as a constraint in peptide identification, one of the defining features is the accuracy in the measurement of relative isotope abundance (RIA). Since it is unknown what RIA accuracy can be obtained from a typical proteome analysis, we analyzed a proteomic dataset collected using routine conditions. We used a yeast digest as a model system, analyzed by LC-MS/MS on an LTQ-Orbitrap Velos using a top15 method (i.e. 1 MS1 scan followed by up to 15 MS2 scans). To generate a reference list of confidently identified peptides, MS2 spectra were submitted to a Mascot database search only retaining the hits with q-value <=0.01. Matching the experimental to the corresponding theoretical isotope patterns of these peptides, the mean RIA error was 3.7%, 4.1% and 4.8% for the mono-isotopic, first and second isotope peak, respectively. This error depended on intensity to a limited extent (Figure 2-1 B), e.g. for the mono-isotopic peak ranging from 1.6% (± 2.5%) for the highest intensity bin (10^7) to 5.0% (± 5.4%) for the lower intensity bin ($10^{4.5}$). The better concordance for high-abundant ions was reported previously for Orbitrap [30] and FT-ICR [29], and is unlikely to be caused by wrong Mascot identifications since there was only a slight positive correlation between the precursor intensity and Mascot score (Figure 2-1 C). In order to evaluate the potential effect of chromatographic peak interference on RIA error, we calculated the RIA error from a digest of single protein (BSA) analyzed by LCMSMS over a 30 min gradient, thereby minimizing peak overlap. This resulted in 4-5 % RIA error (Figure 2-2), thus being very similar to the values obtained for complex proteomes (Table 2-1), indicating that RIA is minimally affected by sample complexity. We observed a tendency for underestimation of the intensity of the monoisotopic peak but an overestimation of the 2^{nd} isotopic peak (Figure 2-1 B). This was also observed for both *E. coli* and human samples (not shown), and for the BSA digest (Figure 2-2). The underestimation of ion intensities of compounds of similar mass was reported before as a result of isotopic beat effects in FT-ICR [33], however, the overestimation is, to our knowledge, previously unreported. Nevertheless, overall there seems to be a limited effect on the measured isotope distribution, since the RIA error is symmetrically distributed (Figure 2-1 D) with

only a small sign of divergence for a minority of cases (Figure 2-1 D, inset).

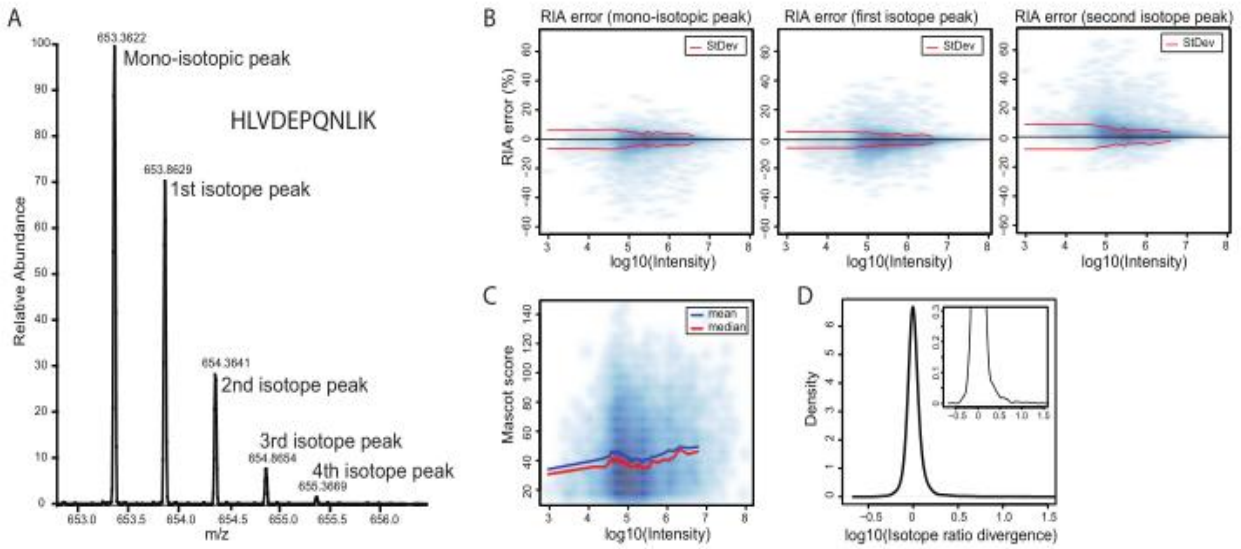


Figure 2-1. (A) Error in the relative isotope abundance (RIA) as a function of ion intensity, for the mono-isotopic peak and the first 2 isotope peaks across all features detected in a total yeast digest. Red lines indicate standard deviation of RIA error. **(B)** Mascot score as a function of ion intensity. The blue and red lines indicate the mean and median Mascot score, respectively, indicating that Mascot score is largely independent of peak intensity. **(C)** Distribution of the isotope ratio divergence comparing the experimental intensities of the mono-isotopic and 2nd isotope peaks to their theoretical values. The divergence is defined as (I0/I2)theory / (I0/I2)experimental, where I0 and I2 are the intensities of the mono-isotopic and second isotope peak, respectively, within a peptide isotope pattern. The distribution of these values shows a normal distribution with only a slight tailing to higher values (see inset).

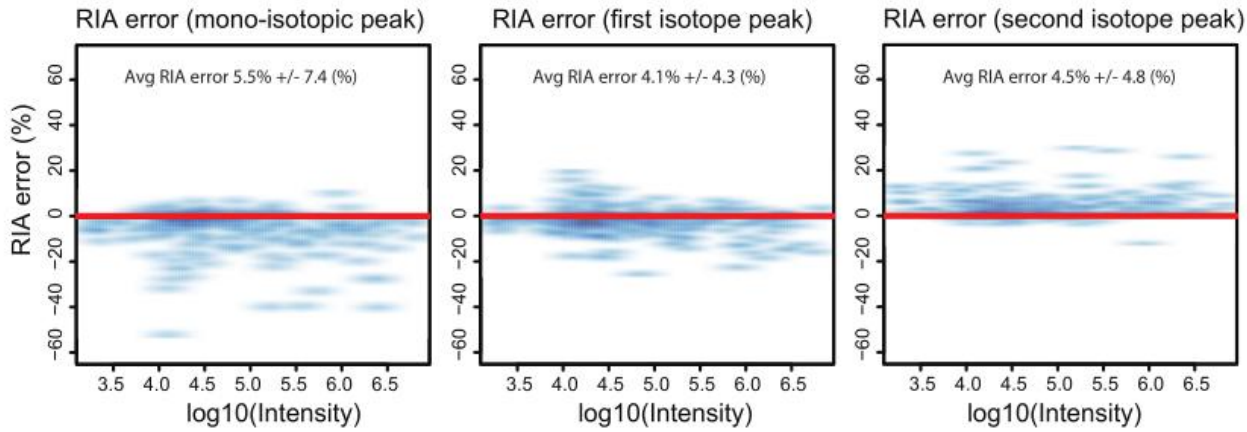


Figure 2-2. Error in the relative isotope abundance (RIA) as a function of ion intensity, for the mono-isotopic peak and the first 2 isotopic peaks across all features detected in a BSA digest.

2.3.2 Effect of mass resolution on RIA measurement

We next aimed to investigate in more detail how other experimental parameters impact the accuracy of the RIA measurement, including mass resolution. Reasoning that the scan time, and thus the resolution, of MS1 spectra may influence RIA accuracy, a yeast digest was analyzed at resolution

7,500, 15,000, 30,000, 60,000 and 100,000 (corresponding to MS1 scan times of 270, 360, 560, 950 and 1,710 milliseconds, respectively). The difference in RIA error between these conditions was relatively small (Table 2-1; Figure 2-3), although the error was the highest at the highest resolution (100,000). This is consistent with recent studies [26, 29, 30] showing a larger RIA error at high resolution in Oribtrap and FT-ICR.

The increased scan time for higher resolution inevitably decreases the number of both MS1 and MS2 spectra, peptide spectrum matches (PSMs), features with PSM (FWPs), and peptide and protein identifications (Table 2-1). Decreasing the resolution from 30,000 to 15,000 or 7,500 resulted in an increased number of spectra but the numbers of PSMs and FWPs were worse than at 30,000. Taken together, with this yeast data set the highest number of peptide identification was achieved at resolution 30,000 in LTQ-Orbitrap Velos with a mean RIA error of 4.2%.

Table 2-1. Evaluation of the error in relative isotope abundance and other mass spectrometric parameters across a range of resolutions. Applied to the analysis of a complete yeast digest by LC-MS/MS.

Resolution	Total spectra (features)	Peptides with spectrum match (PSM)	Features with PSM (FWP)	RIA error (%)	Unique peptide identificaitons	Unique protein identifications
7500	30388	12345	10627	4.45	5649	1212
15000	27609	12618	11877	4.16	6219	1330
30000	23559	13271	12758	4.17	6284	1311
60000	21118	12184	11808	4.30	5788	1233
100000	17203	9519	8794	5.56	4461	999

2.3.3 Correlation between RIA error and number of MS1 scans

Considering that chromatographic features almost always consist of multiple scans, we investigated the effect of the number of MS1 scans on the RIA error by varying the number of MS2 scans between MS1 scans from 3 (top3, resulting in a larger number of MS1 scans across a chromatographic peak) to 50 (top50, resulting in a fewer scans per peak) (Figure 2-4). As before, only yeast peptide identifications with q-value below 0.01 were used for the RIA error calculation. The results (Figure 2-4) indicate that the mean RIA error of 4.2% obtained in the top15 method can be modestly improved to 3% by increasing the number of MS1 scans from 7 to 22 in the top 3 method. Conversely, a decreased number of MS1 spectra per peak (in the top 30 and top 50 methods) slightly increased RIA error to 4.5% and 4.7%, respectively. In addition and as expected, increasing the number of MS1 scans comes at the cost of a reduced number of MS2 scans and thus

peptide identifications (Figure 2-4).

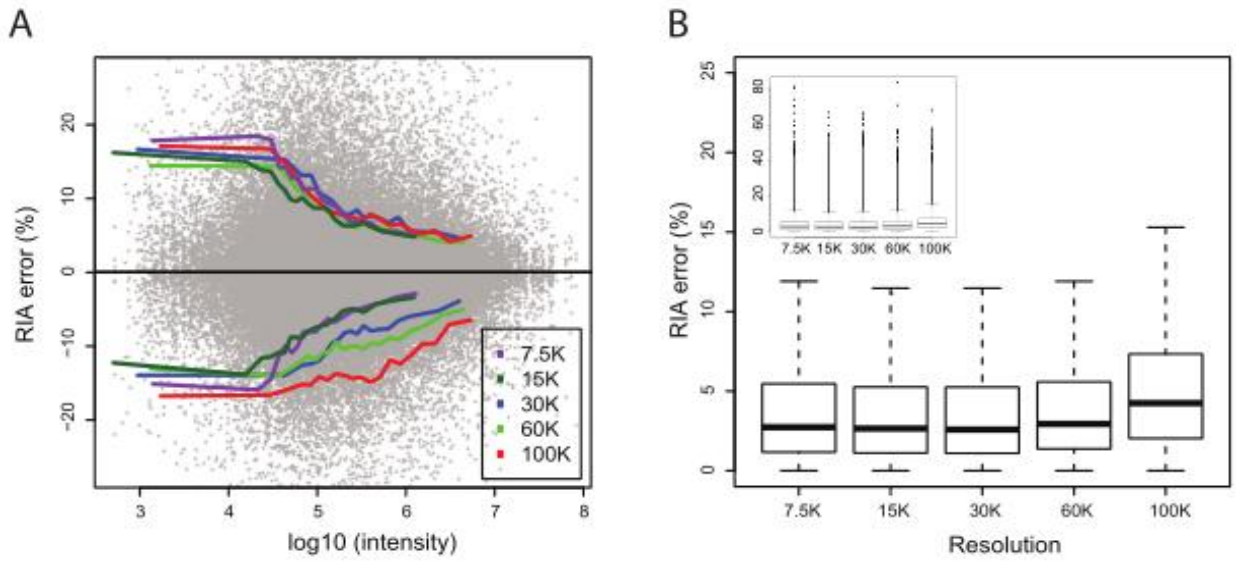


Figure 2-3. **(A)** RIA error as a function of ion intensity at different resolutions ranging from 7,500 to 100,000. Coloured lines indicate 95 percentiles of all isotope distributions, plotted in grey in the background. **(B)** RIA error at different resolutions across all intensities, showing that on average RIA error increases with increasing resolution. Inset: zoom-out of the same data, showing highly similar distribution of outliers.

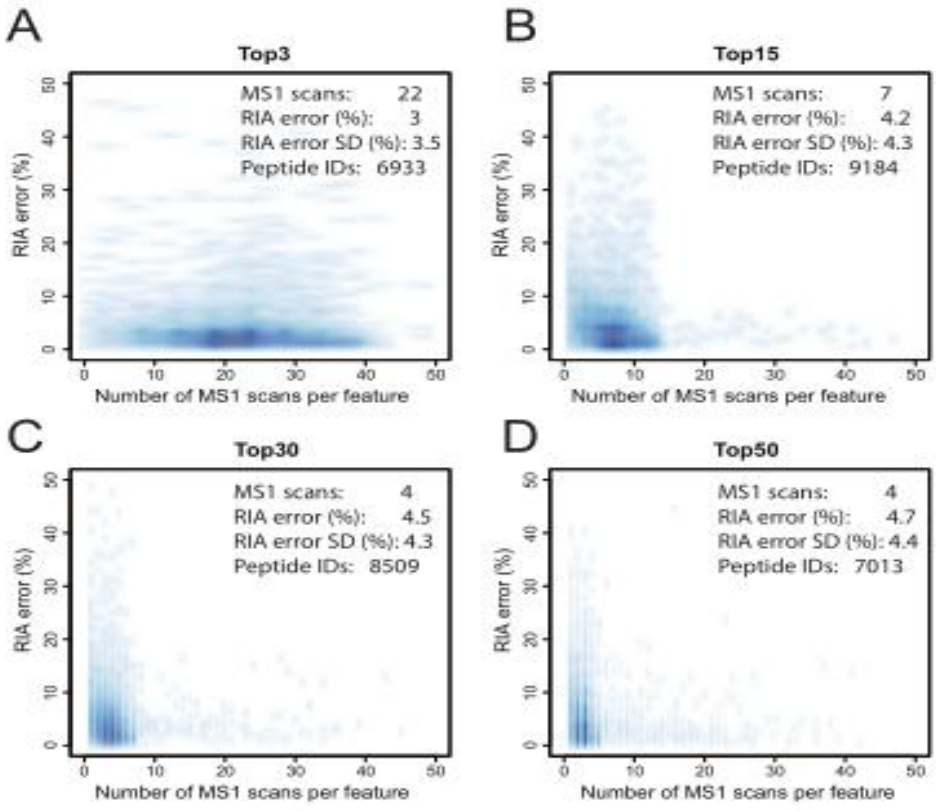


Figure 2-4. RIA error as a function of the number of MS1 scans per detected peptide feature. A yeast digest was analyzed at resolution 30,000 with different MS2 methods: (A) top3, having MS1 scans separated by 3 MS2 scans (B) top15, (C) top30, (D) top50. Numbers in the insets in each panel indicate that increasing the number of MS1 scans decreases the RIA error, but offsets the number of peptide identifications.

2.3.4 Use of isotope patterns as an independent scoring scheme

Aiming to explore several potential approaches to assess if isotope patterns may be used as a constraint for peptide identification, we first considered an ideal scenario where RIA is to be used as the sole parameter to identify a peptide. This requires that a given proteome do not contain peptides with identical atomic composition since these cannot be distinguished by isotope patterns. To determine the number of unique peptides that can be expected in proteomes of various complexities, we *in silico* digested the entire *E. coli*, yeast and human proteomes allowing one missed cleavage and methionine oxidation as a variable modification. *E. coli* had 262,722 unique peptide sequences, 252,032 of which were unique in their atomic composition (95.9%). For yeast these numbers were 480,317 and 362,885 (75.6%), and for the human proteome 1,692,023 and 1,030,190 (60.9%). The relatively high proportion of unique atomic compositions indicates that in principle a large proportion of these proteomes are accessible for analysis by isotope patterns, most notable for the less complex organisms. Therefore, we investigated if a global FDR approach was possible by using isotope patterns as the sole 'search engine'. For this to happen, there should not be common atomic compositions between the target and decoy databases. Table 2-2 shows that in 30-50% of the reverse database and 10-20% of the randomized database, atomic compositions were common to the target database. Therefore, it is clear that isotope patterns alone cannot be used to uniquely identify peptides in complex proteomes.

2.3.5 Difference in RIA error between Mascot target hits and decoy hits

We then investigated if isotope patterns may be used in conjunction with a classical peptide identification based on MS2-spectra. Specifically, we tested if RIA may help in reducing the number of decoy hits in a Mascot search, thereby improving FDR. We analyzed the proteomes of *E. coli*, yeast and HeLa cells by LC-MS/MS, performing Mascot searches against the respective target-randomized databases for these species. This procedure significantly reduced the number of decoy-to-target hit ratio in the peptide set compared to the full proteome analysis above. Among the confident Mascot top hits (q-value <= 0.01), decoy hits with identical atomic composition to a target hit were excluded. This resulted in 1,219 target hits and 9 decoy hits with unique atomic composition in *E. coli*; 5,885 target and 53 decoy hits in yeast, and 20,391 target and 180 decoy hits in HeLa cells. Based on these numbers, *E. coli* was excluded from the subsequent analysis since the number of decoy hits was insufficient for the proper FDR calculation. The RIA error distribution (Figure 2-5 A, B) showed no separation between the target and decoy hits in both yeast and HeLa cells, suggesting that an FDR approach on this Mascot result is not possible with the current RIA accuracy (4-5%). We also attempted to generate a decoy database without common atomic

compositions with the target database. However, we did not succeed evidenced from the strong bias towards the target hits upon Mascot search (Figure 2-6), making the proper FDR calculation unrealistic.

Table 2-2. Number of tryptic sequences in *E. coli*, yeast and human proteome target databases, and the proportion of molecular compositions shared with reverse and randomized decoy databases.

E. coli	**Reverse**	**Randomized**
Total sequences	3,030,939	3,137,134
Total common atomic compositions with target	1,582,388	889,570
Unique sequences	257,084	1,101,606
Unique common atomic compositions with target	137,138	117,854
Fraction of unique common atomic compositions (%)	53.3	10.7
Yeast	**Reverse**	**Randomized**
Total sequences	1,299,076	1,368,272
Total common atomic compositions with target	818,914	565,478
Unique sequences	470,012	521,331
Unique common atomic compositions with target	208,328	116,457
Fraction of unique common atomic compositions (%)	44.3	22.3
Human	**Reverse**	**Randomized**
Total sequences	4,669,151	4,953,655
Total common atomic compositions with target	3,208,613	2,603,995
Unique sequences	1,663,003	1,851,902
Unique common atomic compositions with target	595,296	380,153
Fraction of unique common atomic compositions (%)	35.8	20.5

We therefore examined the RIA accuracy that would be required to confidently discriminate between target and decoy hits upon a Mascot search. Figure 2-5 C shows that in about 80% (yeast) and 70% (human) of cases the experimental isotope patterns can discriminate the top target hits from the decoy hits (normalized RIA error <1). From the same plot, it can be estimated that to do so in 95% of cases would require four times higher RIA accuracy in yeast and about 15 times in human samples, i.e. corresponding to a tolerable RIA error of 1.25% and 0.35%, respectively.

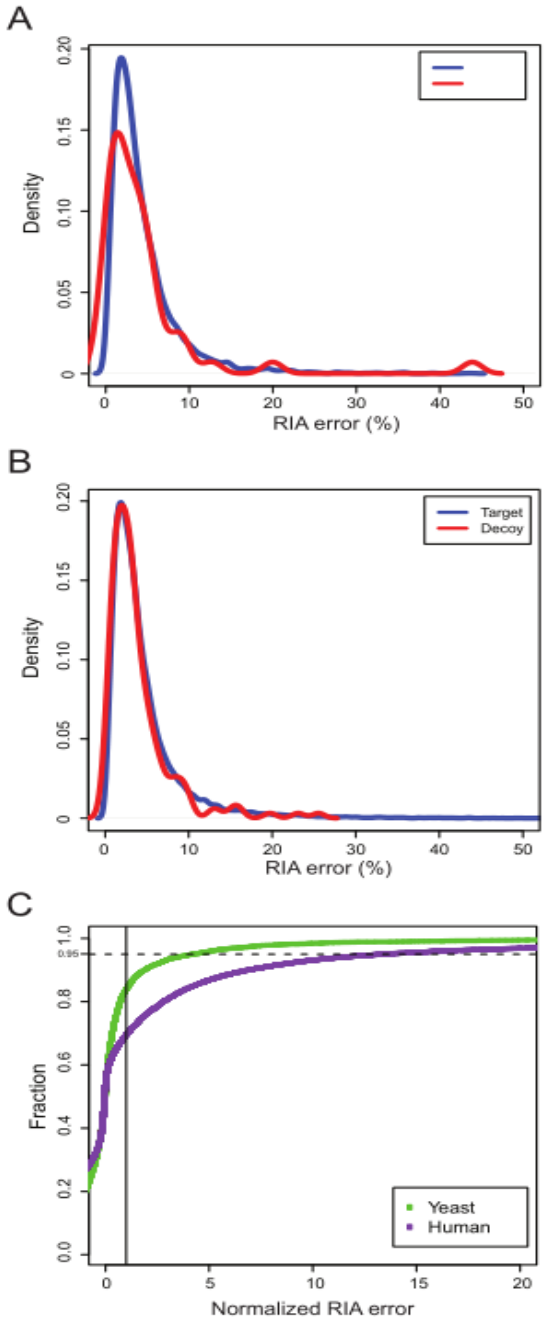


Figure 2-5. RIA error distribution of target and decoy hits in **(A)** yeast, and **(B)** HeLa cells. **(C)** Fraction of cases where experimental isotope pattern can distinguish between the top target hit and the closest decoy hit. The normalized RIA error is defined as RIA error(E – T1) to RIA / RIA error (T1 - D1), where E, T1 and D1 are experimental isotope pattern, the theoretical isotope pattern of its top target hit, and the theoretical isotope pattern of the closest decoy hit, respectively. The dashed line indicates 95% confidence.

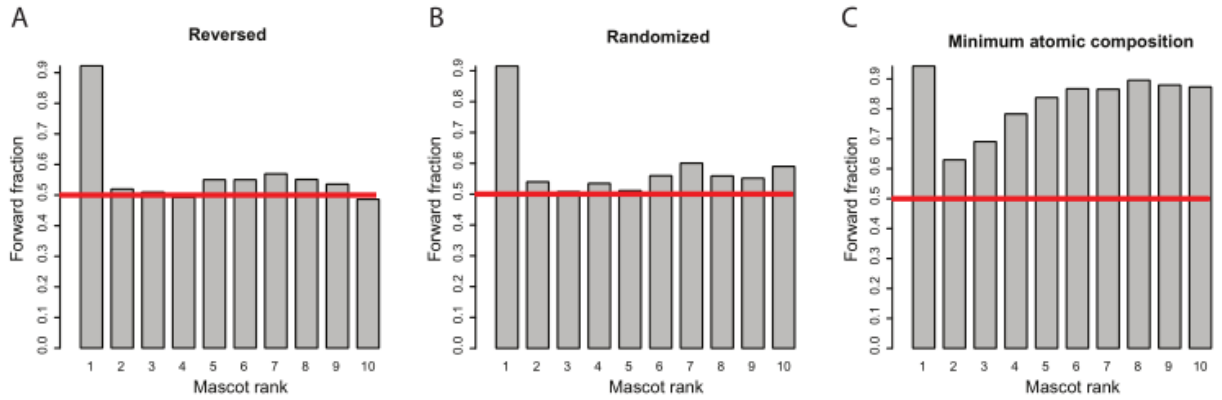


Figure 2-6. We also attempted to generate a decoy database without common atomic compositions with the target database. However, in order to make the number of peptides in the decoy database similar to that of the target database, the same atomic compositions had to be used repeatedly in the decoy database, resulting in a strong bias towards the target hits upon Mascot search, making the FDR unrealistic Mascot search bias towards a target or decoy database. **(A)** Reversed sequences. **(B)** Randomized sequences.**(C)** Decoy database with minimum atomic composition overlap with the target.

2.3.6 RIA error difference between top Mascot hits and runner-up hits

In an alternative approach, we investigated to what extent isotope patterns can discriminate between top Mascot hits and their runner-up hits, identified from the same MS1 feature with a lower Mascot score. We only retained the FWPs where runner-up hits had a different atomic composition than the top hit (373, 606 and 1217 in *E. coli*, yeast and human samples, respectively). The result, split into five intensity bins (Figure 2-7), indicates that in about 95% of the cases isotope patterns can distinguish between the top hits and runner-up hits (normalized RIA error <1) within the higher intensity bins, down to around 70% in the lowest intensity bin. This is unlikely due to incorrect Mascot identification since there was little difference in the Mascot score distribution among the intensity bins (Figure 2-8). Thus, with the current experimental procedure and instrument it was possible to a large degree to discriminate between top Mascot hits and runner-up hits by their isotope patterns especially when the intensity is high. However, in order to do so in at least 95% of the cases across all intensities, Figure 2-7 A-C indicate that the RIA accuracy has to be 4-5 times higher than the current level.

In proteomics it is common practice to assume that the top Mascot target hit for a particular spectrum is the “correct” peptide without considering lower scoring alternatives. However, it would be interesting to use isotope patterns for ranking hit sequences especially when the Mascot score is similar between the top hit and runner-up hit. To investigate this, we first composed a reference set of very confident Mascot identifications (Mascot score >= 70), and estimated the RIA error rate of experimental isotope patterns by interpolating the mean RIA error as a function of intensity. This subset covered the entire intensity range observed in the entire data set (Figure 2-9). Then the ratio

of this RIA error rate to half the theoretical RIA difference between the top hit and 1st runner-up hit (T – Tr) was taken (i.e., RIA error rate / (T -Tr)/2). Note that halving was necessary since the RIA error needs to be taken into account for both top hit and runner-up hit. If this ratio is < 1, the RIA error rate is small enough to tell the top hit and runner-up hit apart by the isotope pattern. Taking the FWPs mentioned above where a runner-up hit was found (i.e. 373 (*E. coli*), 606 (yeast) and 1,217 peptides (HeLa cells), Figure 2-7 D shows the fraction of the cases where these could be discriminated based on isotope patterns. This was the case for about 30% of *E. coli* cases and <10 % in yeast and HeLa cells, indicating that the RIA accuracy was insufficient to tell them apart.

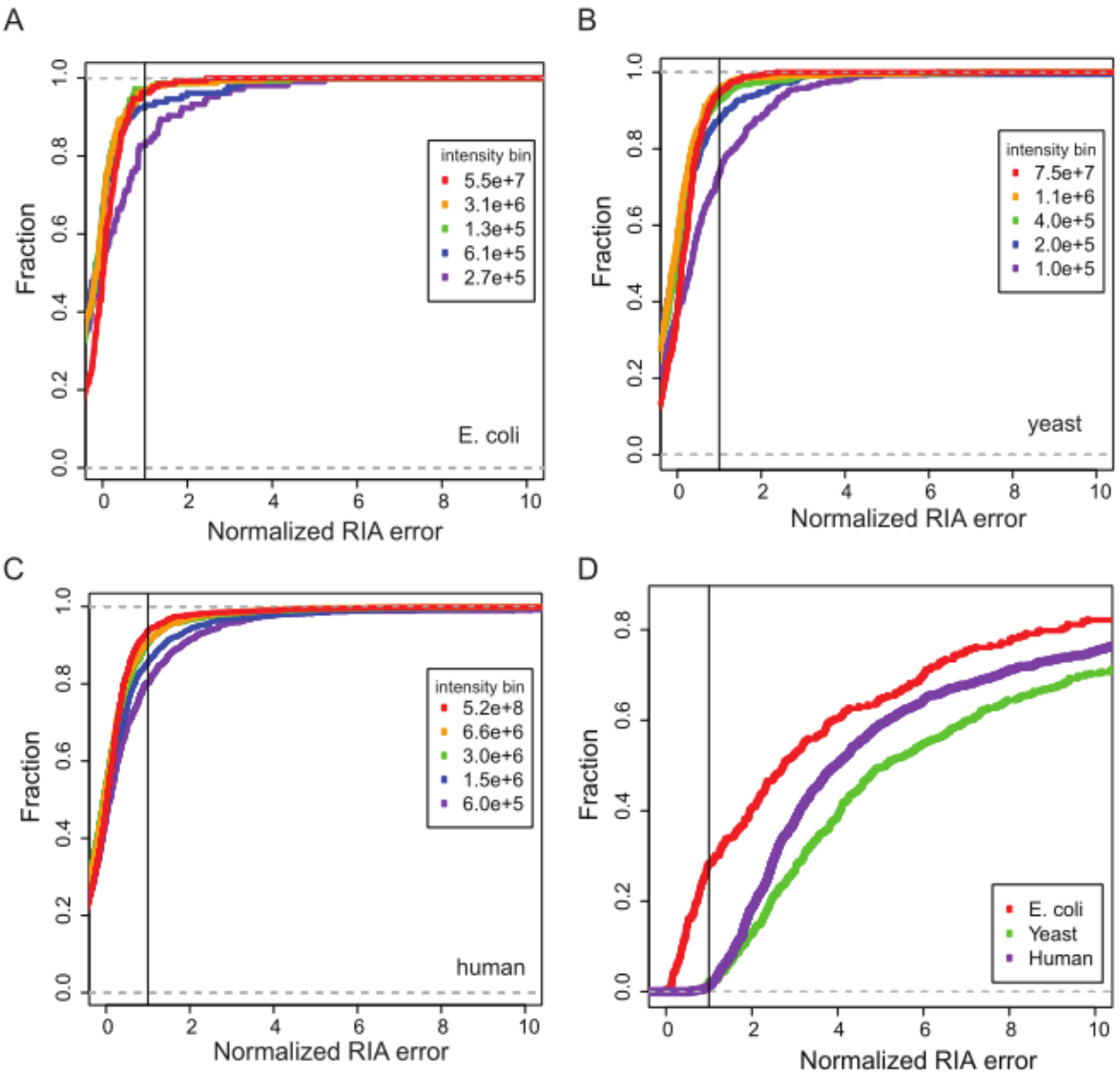


Figure 2-7. (A) Fraction of cases in *E. coli* where experimental isotope pattern can distinguish between the top target hit and the closest decoy runner-up hit as a function of the normalized RIA error. The normalized RIA error is defined as RIA error(E – T1) / RIA error(T1 - Tr), where E, T1 and Tr are the experimental isotope pattern, the theoretical isotope pattern of its top target, and the theoretical isotope pattern of the closest runner-up hit, respectively. The entire dataset was split by intensity into five bins. **(B)** Same as panel A, but for yeast. **(C)** Same as panel A, but for human.

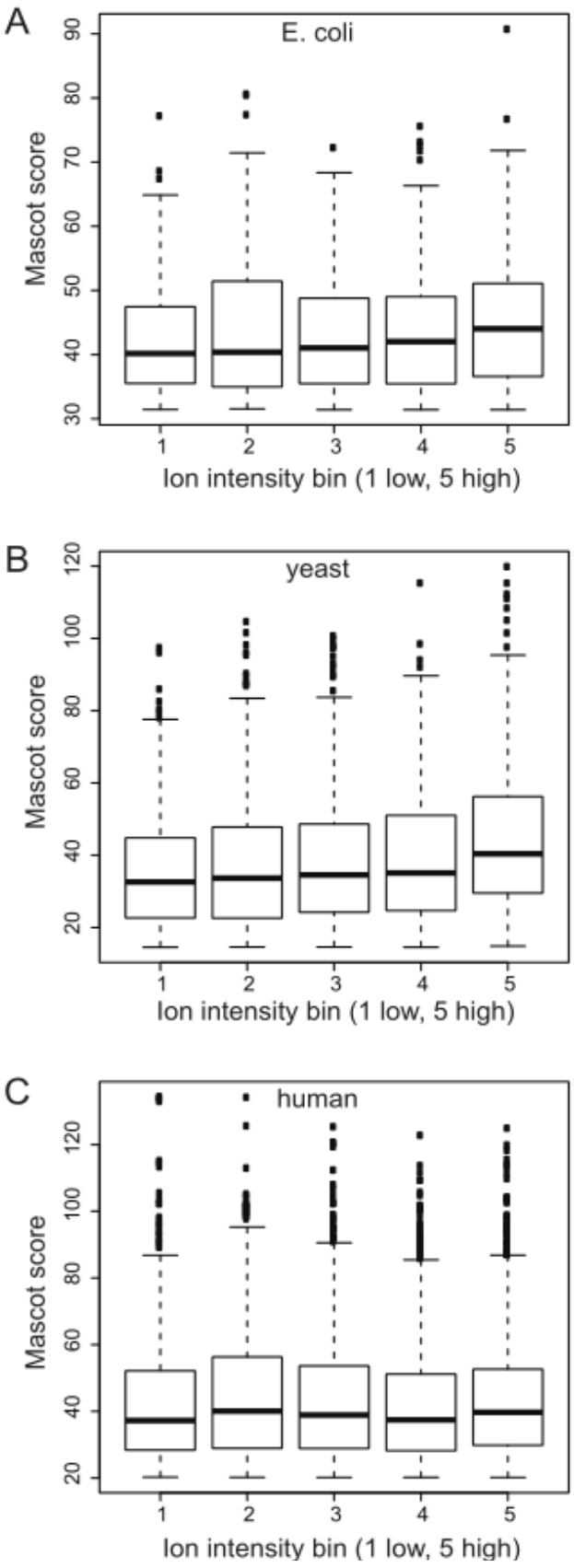


Figure 2-8. Mascot score distribution of the top Mascot target hits in different intensity bins in the proteome analysis of **(A)** *E. coli*, **(B)** yeast, and **(C)** human. Intensity bins for the respective organisms correspond with those indicated in Figure 1-6.

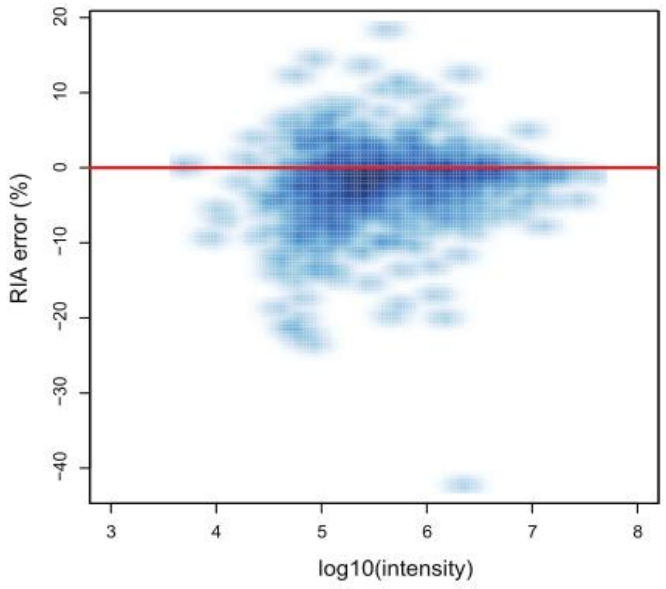


Figure 2-9. RIA error as a function of ion intensity of yeast sample, only considering peptides that were identified with very high confidence (Mascot score >=70). The mean RIA error over the three isotopic peaks was taken.

2.3.7 Re-classifying target and decoy hits using RIA did not improve peptide identification

Next, we investigated whether the RIA could contribute to improving peptide identification when combined with other parameters. The first three isotopic peaks to the theoretical isotope patterns were fitted with a linear model and the mean of the absolute value of the residuals was used as a score for how similar the experimental- and theoretical isotope patterns were to each other. Then this mean absolute residuals and the Mascot score were used to re-classify target and decoy hits by LDA, QDA and SVM (Figure 2-10 A, B, C, respectively). The posterior error probability was used as the score after re-classification. The number of target hits as a function of FDR is plotted in (Figure 2-10 D). The result showed that QDA and SVM performed much worse than the Mascot score alone and LDA was almost identical to the Mascot score, suggesting that increasing confidence in peptide identification in this two-dimensional space is difficult.

2.3.8 Required accuracy in mass and RIA for theoretical digest peptide isotope patterns

Since our data indicated that RIA accuracy is generally insufficient to be helpful in peptide identification at a proteomic scale, we took a theoretical approach to estimate the RIA accuracy that would be required to uniquely identify peptides in a given proteome. Therefore, the *E. coli*, yeast and human proteomes were *in silico* digested in order to calculate the atomic compositions and isotope distributions of the resulting peptides. We generated a representative dataset by selecting all peptides from 24 mass windows, each 1 Th-wide and 50 Th apart, ranging from 300-301 Th to

1450-1451 Th. This resulted in 7304, 10838 and 27087 peptides from *E. coli*, yeast, and human proteomes, respectively. Figure 2-11 A-C show that without considering RIA (i.e. RIA > 0.15), a larger proportion of peptides can be uniquely identified with increasing mass accuracy in the three organisms. For instance, at 3 ppm mass accuracy less than 20% of the peptides in *E. coli* can be uniquely identified (Figure 2-11 A), increasing to 40% and 65% at mass accuracies 1 and 0.5 ppm, respectively. In order to achieve more than 90% of unique peptide identification by mass alone, 0.1 ppm mass accuracy would be necessary for *E. coli* and yeast, and 0.05 ppm for human peptides.

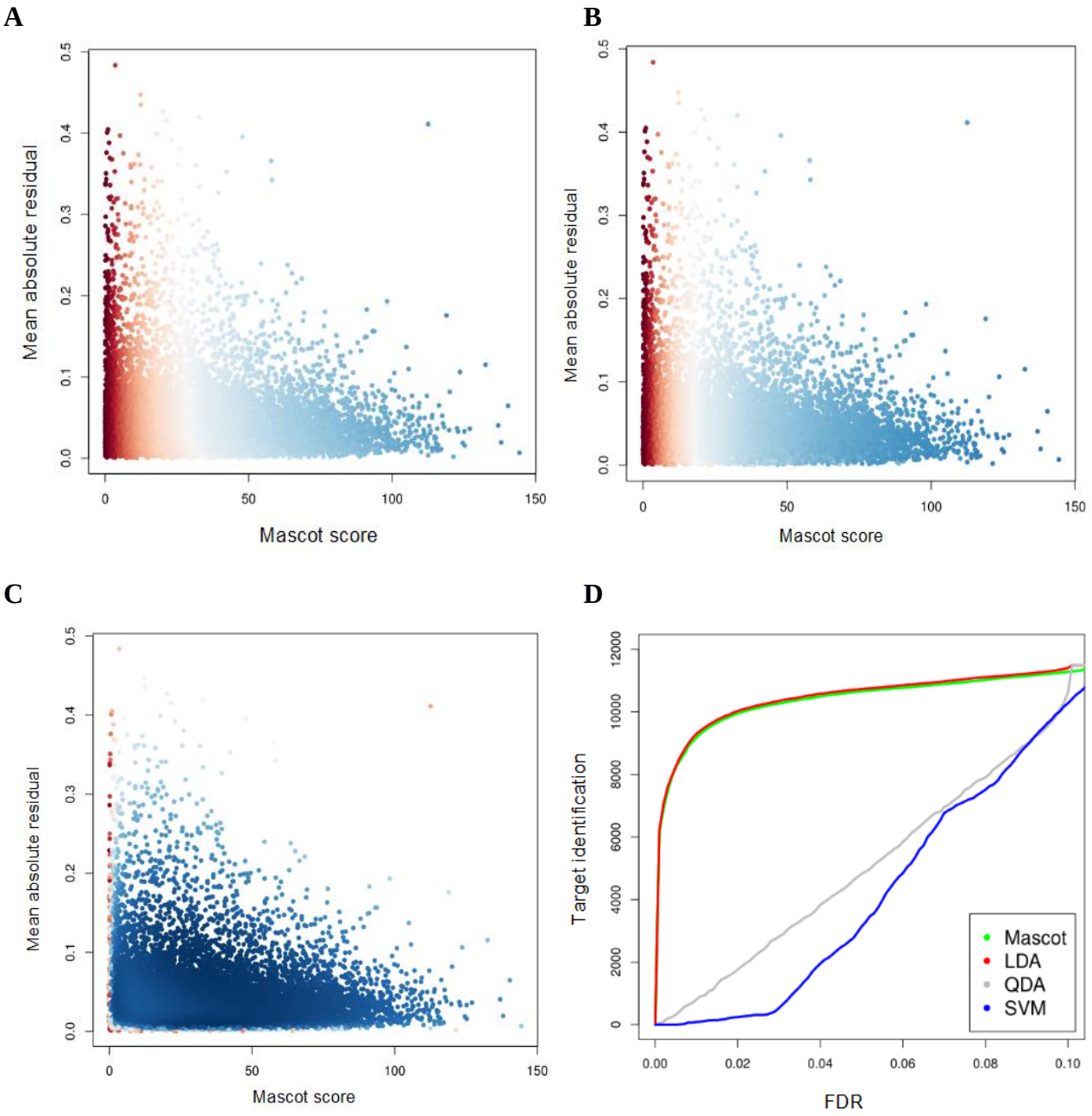


Figure 2-10. Target and decoy hits re-classified by Mascot score and mean absolute residual using **(A)** linear discriminant analysis (LDA), **(B)** quadratic discriminant analysis and **(C)** support vector machine (parameters). (D) Number of target hits false discovery rate (FDR) using Mascot socre and comibined score after re-classification by LDA, QDA and SVM. X-axis range from 0 to 0,1.

When the RIA is used as an additional filter, the number of unique identifications showed a modest increase between the RIA errors 5% and 10%, but increased sharply at RIA error below 2% (Figure 2-11 A-C). To achieve 95% unique peptide identification in *E. coli*, at 1 ppm mass accuracy about 0.5% RIA accuracy would be required (Figure 2-11 D), and even a better RIA in yeast (Figure 2-11 E) and human (Figure 2-11 F). The required RIA accuracy sharply decreases at 0.1 ppm to around 8%, 2%, and 1.5% in *E. coli*, yeast and human, respectively. It should be noted, however, that at this mass accuracy more than 90% of the peptides were already uniquely identified by mass alone in *E. coli* and yeast (Figure 2-11 A, B), and 80% in human (Figure 2-11 C). With a mass accuracy of 1-3 ppm, the RIA accuracy would need to be ~0.2%, i.e. 20-30 times higher than the current level (4-5%) in order to uniquely identify atomic compositions proteome-wide, independent of MS2. Conversely, with the current RIA accuracy, the mass accuracy would have to be at least better than 0.05 ppm (20-60 times better than our current accuracy). This analysis illustrates the combined contribution of mass and RIA accuracy to confidently identify peptides with unique atomic composition.

2.3.9 Required accuracy in mass and RIA for the isotopic fine structures of theoretical digest peptide

So far we have considered only the most dominant isotopic peaks assuming that these peaks consist of a mixture of different atomic elements (e.g. C, N, O, S). However, it has been known that a sufficiently high mass resolution could resolve contributions from these different elements known as isotopic fine structures produced by the slight differences in mass increase of isotopes of C, N, O and S [34, 35]. Miladinovic et al [36] were able to resolve isotopic fine structures of a few peptides using FT-ICR with resolving power 1,500,000 at m/z 1061. This resolution cannot be achieved by the current Orbitrap technology, and its maximum resolution (100,000) is insufficient to resolve isotopic fine structures. Therefore, we performed a theoretical study to assess whether isotopic fine structures may be more informative than normal isotope patterns to elucidate peptide identity. The result (Figure 2-12) indicates that in order to achieve 95% unique atomic composition identification, isotopic fine structures could relax the RIA accuracy requirement by 1.5 – 2 fold in comparison to the normal isotope patterns in all the three organisms. A similar trend was observed for the 99% unique identification rate (Figure 2-12), suggesting that resolving isotopic fine structures is in principle another constraint that could aid in peptide identification.

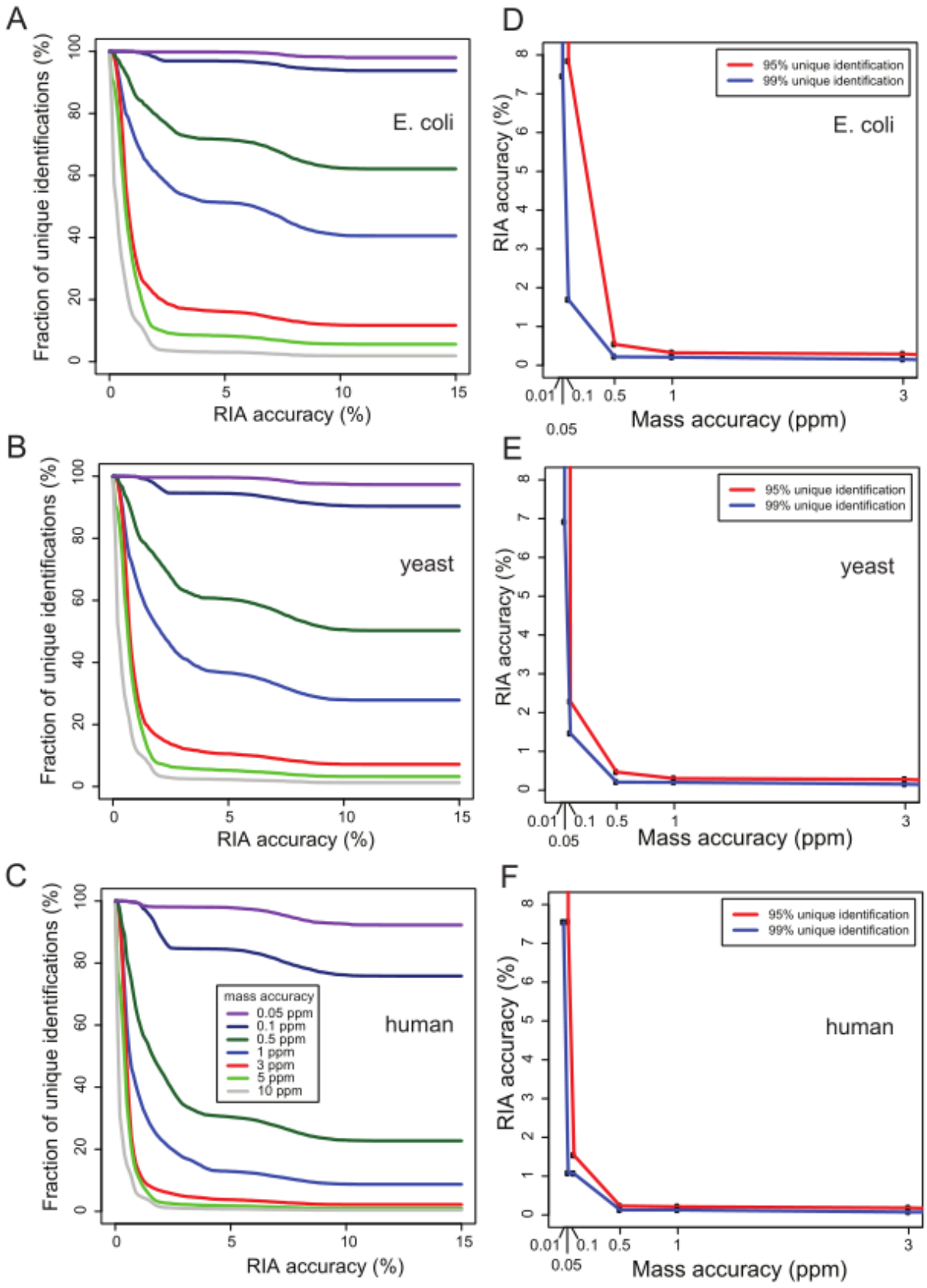


Figure 2-11. (A) The theoretical identification rate for unique atomic composition as a function of the RIA accuracy at different mass accuracies (coloured lines) in the *E. coli* proteome, **(B)** the yeast proteome and **(C)** the human proteome. **(D)** The RIA accuracy necessary for uniquely identifying 95% (red line) and 99% (blue line) of atomic composition at different mass accuracies in the *E. coli* proteome, **(E)** yeast proteome, and **(F)** human proteome.

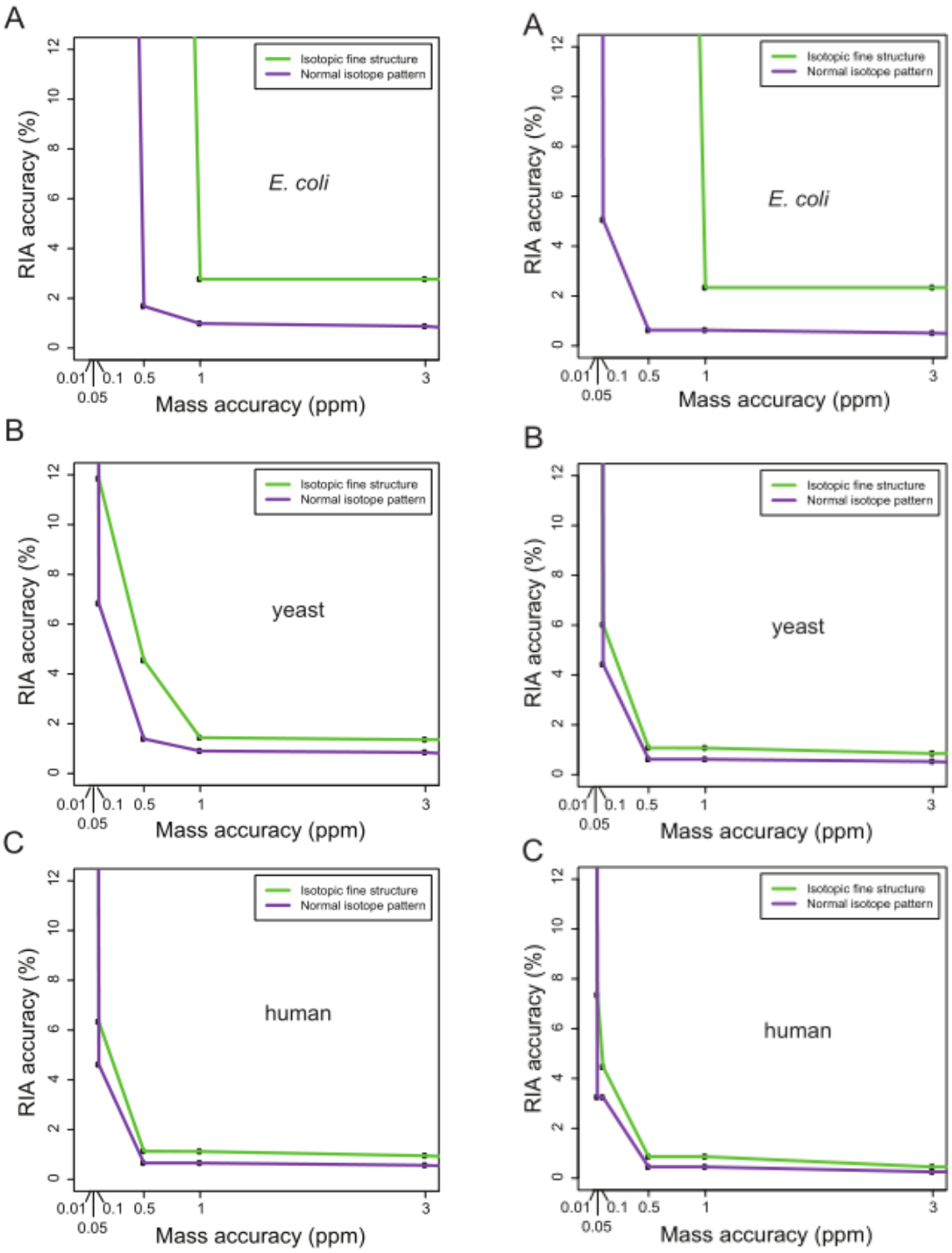


Figure 2-12. Required RIA performance to achieve 95% (A, B, C) and 99% (D, E, F) peptide identification rate comparing normal isotope patterns (green line) and isotopic fine structures (purple line) as a function of mass accuracy, applied to *in silico* digested proteomes of *E. coli* (A, D), yeast (B, E) and human (C, F). Note, the y-axis is the RIA accuracy (%) for the summed RIA difference over isotopic peaks, not the mean, since here the RIA difference conferred by different numbers of isotopic peaks (i.e., isotope patterns or isotopic fine structures) is evaluated.

2.4 Conclusion

In this study we have quantitatively examined the accuracy of RIA obtained from proteomic datasets. We demonstrate that the RIA error is 4-5%, and that this is only modestly influenced by spectral intensity, resolution and the number of MS1 scans. To assess the utility of isotope patterns as a discriminatory feature in peptide identification, we have tested a number of potential applications, either using RIA alone or in combination with MS2-based database searches, applied to proteomes of various complexities. The current RIA accuracy has limited discriminatory power at a proteome-wide scale. At the same time, it should be stated that this analysis was hampered by the difficulty in calculating FDRs, particularly in constructing proper decoy databases that are similar in size as the target database, yet different in molecular composition for all peptides considered. Alternative strategies to calculate FDRs will be required to address this issue for complex proteomes. Regardless, our analyses have shown that the limited utility of RIA in Orbitrap data is primarily due to the fact that it cannot compete with the strong discriminatory power of mass accuracy. The utility of RIA may become relevant with future instrument developments, considering that even a relatively modest decrease in RIA error down to <1% strongly improves discriminatory power (Figure 6). Alternatively, at increased mass accuracy even current RIA accuracy levels may be sufficient to fit isotope patterns as a constraint in the peptide identification process as a parameter that comes for free in any MS-based proteomic experiment.

Chapter 3

3 A comparative proteomics study between neural stem cells [1] and glioma neural stem cells (GNSs)

3.1 Introduction

3.1.1 Epigenetic models of tumour development

Knudson [37] was the first to propose the clonal genetic model of cancer, in which tumour development begins with genetic alterations in a single founding cell which then evolves and becomes a dominant population by undergoing clonal selection. However, epigenetic changes regulated by DNA methylation and non-coding RNAs can stably influence gene transcription and are often observed at the earliest stages of neoplasia within the altered tissue stem/progenitor cells. To take these epigenetics into account, Feinberg et al [38] postulated the epigenetic progenitor model of cancer development, which proposes that malignant transformation takes place in three steps. First, essential epigenetic disruptions of stem/progenitor cells make them epigenetically permissive to tumourigenesis. Second, genetic alterations occur in a tumour suppressor gene or an oncogene. Third, genetic and epigenetic changes occur, which enhance their ability to stably evolve the tumour phenotype. A major difference to the clonal genetic model is that the epigenetic events occur prior to genetic alterations, and are necessary for creating a polyclonal population that subsequently undergoes genetic alterations and transformation. It remains unclear whether epigenetic changes are necessary and sufficient for initiating and/or sustaining the malignancy. Teng et al. [39] showed that methylating the tumour suppressor genes RASSF1A and HIC1 in human mesenchymal stem cells by recruiting DNMTs to these loci resulted in the formation of stem-like cells that started malignant transformation. The resulting cells showed genome-wide changes in DNA methylation and altered TP53 function and DNMT inhibitors were able to reverse this phenotype. Thus, this is a direct demonstration that aberrant DNA methylation can directly lead to tumourigenesis. It is, however, still unknown if the DNA methylation, or other epigenetic mechanisms, is indeed employed for initiating tumours in vivo as the majority of experimental evidence is correlative in nature.

3.1.2 Glioblastoma multiform consists of heterogeneous cell types

Among primary adult brain tumours, glioblastoma multiform (GBM) (grade IV astrocytoma) is the most common and severe form [40] with a median survival time of only 15 months [41]. Efforts have been made to characterise GBM and the genetic aberrations and disrupted signalling pathways have been identified [42, 43]. However, individual tumours contain varying proportions of different cell types including both differentiated cells and ill-defined anaplastic cells. Thus, it is uncertain how these findings operate in different cell types. It also remains unclear how tumours are initiated and maintained.

3.1.3 Cancer stem cells are stem-like cells within tumour that can initiate a tumour

Among various cell types, a small population (0.01-1%) of stem-like cells within the tumour that exhibit self-renewing and differentiation capacity are called cancer stem cells (CSCs). The cancer stem cell hypothesis proposes that a tumour arises from these CSCs, as demonstrated for leukemia [44] for the first time. Analogous to normal stem cells differentiating into various committed cells, these stem-like cells are thought to give rise to a heterogeneous population of cells that constitute hierarchy of tumorigenic stem-like cells and their differentiated non-tumorigenic progeny. Therefore, CSCs are also called tumour initiating cells. CSCs in brain cancer were first isolated from adult- and child brain tumours using an NS marker CD133 [45, 46] and they were able to reproduce the brain tumour upon xenotransplantation [47]. The origin of brain tumour stem cells is unclear, and requires more investigations. They may arise via transformation of *bone fide* adult stem cells, committed progenitors or through de-differentiation of mature cells [48]. Since both normal stem cells and CSCs have the ability to differentiate into functional cells, terminally differentiating CSCs may suppress tumour development and malignant properties which often arise in later stages of cancer, possibly driven by epigenetic mechanisms [49]. Regardless of the cell of origin of the tumour initiating cells, it is imperative that we understand their biology in order to selectively target them. Thus, it is now an important goal in the field to identify proteins that are easily accessible for external stimuli, such as in the plasma membrane or extracellular matrix, as these are tractable therapeutic targets compared to intrinsic regulators such as transcription factors.

3.1.4 Serum culture and neurosphere culture for CSCs have been widely used but have some limitations

Since CSCs drive tumour growths, it is more accurate to study specifically these cells rather than the entire heterogeneous population of a primary tumour. Because CSCs often represent a

subpopulation of the tumour bulk, screening studies such as compound screening and omics analyses necessitate efficient derivation and propagation of these cells. Historically, investigators have made use of serum culture media to try and propagate GBMs in derivation of 'classic' cell lines such as U87. However, these cellular models do not provide a realistic model of the disease, primarily because serum irreversibly differentiates CSCs and resultant cells no longer have the capacity to form a tumour again when transplanted into immunosuppressed mice [3, 50]. In addition, repetitive passaging results in accumulation of extensive in vitro acquired de novo mutations [3]. It is likely the high levels of BMPs within serum are what drive differentiation of primary tumour cultures and restrict their expansion. To circumvent these problems, the serum-free neural stem cell culture conditions, such as the neurosphere culture methods, have been used for enrichment of brain tumour stem cells [45, 46, 50-52]. Neural stem/progenitor cells float in suspension culture in a serum-free media containing necessary growth factors, namely EGF and FGF. Though neurosphere culture has been successful in maintaining tumourigenic capacity of the tumour stem cells, there are some important limitations. The cultured cells can contain not only true stem cells but also partially-committed progenitor cells and even differentiated cells due to spontaneous differentiation and apoptosis [53-55]. The resultant sphere aggregation of different cell types provides a hurdle for meaningful population level analyses such as monitoring of stem cell behaviour and marker expression and omics studies. Furthermore, the cells in these spheres are unequally exposed to agents applied to the cells (growth factors, drugs, transfection reagents, etc.), limiting the rigorousness of these experiments.

3.1.5 Adherent monolayer culture enables expansion of pure normal- and glioma stem cells

To overcome these shortcomings of neurosphere culture, Pollard et al. [4] using the 'NS cell' culturing methodology reported in [56-58] derived a panel of adherent monolayer cell cultures of human glioma cells that display stem cell characteristics, which they termed the glioma neural stem (GNS) cells. Adherent monolayer culture employs laminin to attach stem cells to the surface of the flask, thereby reducing aggregation and detachment, facilitating expansion and propagation of a highly homogeneous stem cell populations. It also allows cells' uniform access to media, restricting the spontaneous differentiation and cell death that accompany the three dimension culture of neurospheres. The NSs were derived from human foetuses since adult NSs are difficult to obtain for ethical reasons, while the GNSs were derived from adult patients. The GNSs are tumour initiating upon intracranial transplantation into immunocompromised mice (NOD/SCID) using as few as 100 cells, whereas NSs never generated tumours even using 100,000 cells [4]. Furthermore, different GNS cell lines fall into different phenotypic classes which expressed different molecular markers of

distinct stem and progenitor subtypes and correlate with cell-of-origin (personal communication with S.P.).

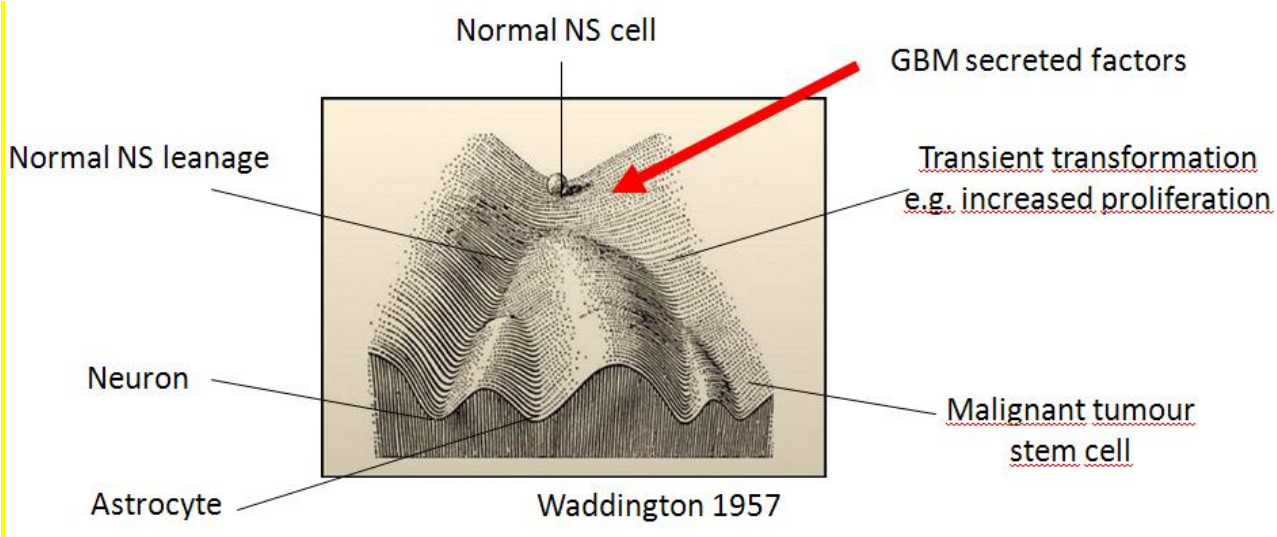
3.1.6 Auto-/Paracrine factors in microenvironment regulate stem- and cancer cells

It has been recognized that stem cells of various tissues are tightly regulated by the immediate microenvironment, or stem cell niche [59]. Though stem cell niches manifest themselves in diverse forms, their critical function is to provide extrinsic cues that sustain stem cell self-renewal. This can either be specific growth factors or signalling molecules, but also specific forms of extracelluar matrix (ECM). For instance, GBM secretome has been demonstrated to aid in the invasiveness of this tumour [61]. Glioma stem cells have been shown to secrete VEGF that directly supports the development of the local vasculature [60, 62]. It has recently been prposed that glioma stem cells can also directly contribute to the microvasculature through their transdifferentiation into vascular cells [63, 64], although the efficiency with which this occurs is low. Furthermore, endothelial cells secrete nitric oxide that induces Notch signalling in glioma cells [65] and only glioma stem cells, but not glioma non-stem cells, were dependent on nitric oxide synthase-2 [66]. Thus, glioblastoma stem cells are known to be regulated by factors secreted by the surrounding microenvironment, however, whether secreted factors from glioma cells could help normal cells turn into a malignant state remains elusive. Venugopal et al [67] showed using the neurosphere culture that GBM secretome was able to transiently enhance proliferation of normal neural progenitor cells (NPCs) and these NPCs acquired glioblastoma stem cell-like properties such as increased CD133, ALDH and IGFBP7 expressions and decreased differentiation. Growth factors such as EGF, VEGF and PDGF and their cognate receptors were also found to be up-regulated in these GBM secretome-treated NPCs. However, their mass spectrometry analysis of the GBM secretome could not definitively identify critical signalling factors. When these GBM secretome-treated NPCs were reverted back to normal conditions for seven days, their neoplastic properties disappeared. Thus, it is still uncertain whether prolonged exposure to GBM-secretome can epigenetically reset NS cells to acquire tumour specific disruptions that subsequently enable transformation.

3.1.7 Objectives

In the current study we have three objectives. First, we aim to globally characterize the differences in protein expression of both total cell proteome and secreted proteome between untransformed foetal NSs and malignant adult GNSs. By using genetically normal NSs as the reference control, we aim to derive a cancer signature present in GNSs as possible causes for tumourigenesis by

eliminating the properties related to stem cells present in both NS and GNSs. Second, we aim to identify surface markers that distinguish between NSs and GNSs, since proteins outside the plasma membrane can be more effectively targeted with drugs than those inside the cell. It is of particular interest to try and define cell surface molecules unique to the GNS cells, as these may be used in antibody based therapeutic approaches. Third, we investigate the effects of GNS-secreted factors on NSs. NSs can most likely be vulnerable to transformation due to their longevity and self-renewal capacity, the properties that GNSs also share. Three TGF-alpha family growth factors, EGF, VEGF and PDGF, have been shown to be over-expressed in glioma-conditioned media [67]. EGF is a growth factor we use to proliferate NSs in culture and VEGF has a role in angiogenesis [62]. However, it is still unclear whether these are the only auto-/paracrine factors mediating the entire process of tumourigenesis in vivo. In previous drug screening EGFR pathway inhibitors did not efficiently kill our GNSs, indicating that other pathways may be operating and that GNS secreted factors, other than EGF, VEGF and PDGF, may be involved in this process.



3.2 Materials and methods

3.2.1 Cell lines

1. Total cell analysis

The NS lines (CB660, CB192, CB152 and CTX985) were derived from human foetuses. GNS lines (G144, G166, G179 and G25) were derived from adult patients following local ethical board approval. The details on the GNS derivation is described in [4]. Briefly, tumours were dissociated

into single cells with Accutase (Sigma) or an enzyme cocktail for 15–20 min at 37°C. For those tumours with excess debris, cells were initially allowed to form spheres/aggregates in suspension culture, and these were then transferred to a fresh laminin-coated flask.

2. Secretome analysis

For the secretome experiment, CB660, U5 and CB11130 NS lines, and G179, G144 and G7 GNS lines were used.

3. Mouse cells

The mouse NS line ANS4 was derived from adult mouse. The GNS line IENS was obtained from the NS line by removing the Ink4a/Arf locus (-/-) followed by virus over-expression of EGFRViii.

3.2.2 Cell culturing

1. Total cell analysis

Both NS cells and GNSs cells were cultured in serum-free, Dulbecco's modified Eagle's medium supplemented with bovine serum albumin (Invitrogen), penicillin/streptomycin (PAA), N2 (PAA), B27 (PAA), EGF (Peprotech), FGF (Peprotech) and laminin (Sigma), as described previously [4, 58]. Medium was replaced every 3–5 days. Cells were grown to 60-70 % confluency harvested and pelleted down. The passage numbers are 20 for CB660, 27 for CB192, 15 for CB152, 10 for CTX985, 28 for G144, 18 for G166, 18 for G179 and 26 for G25.

2. Secretome analysis

For stable isotope-labelled amino acid labelling (SILAC) and AHA-labelling, cells were grown to 60-70% confluence. Then the cells were grown either in intermediate media containing 1ul/ml Lys-4, Arg-6 and 0.2 ul/ml AHA, or in heavy media containing 1ul/ml Lys-6, Arg-10 and 0.2 ul/ml AHA. Supernatant was collected after 24 hours and centrifuged at 4000 RPM for 10 min and the supernatant was collected for LS-MS/MS. The viability of the cells after AHA incorporation was assessed with ViCellar. The passage numbers are 10 for CB660, 10 for U5, 14 for CB11130, 40 for G144, 31 for G166 and 30 for G7.

3. Colony forming assay

For NS/GNS cell colony-forming assays, 1000 cells were plated on a 10 cm dish filled with 10 ml of culturing media. Factors of interest were added to the media at the same time. When conditioned media was added, 5 ml of the culturing media and 5 ml of the conditioned media were mixed to make the dilution 1:1. The number of colonies was counted 7-10 days later, depending on the

growth speed. The colony count was normalized to the +EGF/FGF control. The number of cells within each colony was not taken into account.

3.2.3 Peptide sample preparation

1. Total cell

The cell pellets were homogenized in RapiGest (Waters) followed by reduction of disulphide bonds using 100 mM dithiothreitol (DTT), alkylation using 100 mM iodoacetic acid (IAA) and protein digestion using sequencing grade modified trypsine (Promega) overnight at 37°C. Peptides were stable isotope-labelled via reductive dimethylation, as described in [69]. All NS lines were combined to form a common reference, to which each individual GNS cell line was compared. These samples are named NSpool-G166, NSpool-G144, NSpool-G25 and NSpool-G179. A common GNS cell line pool was also made and compared to the NS pool (NSpool-GNSpool). This sample was only used for increasing peptide identifications during the "match between runs" process in MaxQuant (see below). Reverse labelling was performed on an aliquot of each of these samples. Peptides fractionated into 12 fractions on Agilent 3100 OFFGEL Fractionator (settings as described by the manufacturer) using Immobiline DryStrips (pH 3 – 10 NL, 13 cm, GE Healthcare). Dried samples were resuspended in 360 ml H2O and diluted into 1.44 ml 1.25 x IEF stock solution (6% glycerol, 2% Ampholytes pH 3 – 10 (1:50)). Focusing was performed at a constant current of 50 mA with a maximum voltage of 8,000 V until 20 kVh. Then peptides were acidified, desalted with C18 StageTips (Empore 3M) [31], and reconstituted with 4% acetonitrile in 0.1% formic acid.

2. Secretome analysis

8 ml of each cell supernatant was filtered using Amicon Ultra Centrifugal Filters (3-kDa cutoff) to the end volume of ~250 µL. The proteins in this concentrated media were enriched using the Click-iT Protein Enrichment Kit (Invitrogen C10416) and applying the vendor's protocol with slight modifications; 100 µl of agarose resin slurry was used and the volume of all the reagents was halved. After washing the resin with 900 µL water, the concentrated media was diluted in 250 µL urea buffer, to which the catalyst solution was added and incubated for 16–20 h at room temperatures. Then the resin was washed with 900 µL water, and 0.5 mL SDS buffer and 0.5 µL 1 M dithiothreitol (DTT) (Bio-Rad) were added to the resin and vortexed at 70 °C for 15 min. The supernatant was aspirated and 3.7 mg iodoacetamide (IAA) (Bio-Rad) was added and incubated for 30 min in the dark. The resin was transferred to a spin column and washed with 20 mL of SDS buffer, 20 mL of 8 M urea in 100 mM Tris pH 8, 20 mL 20% isopropanol and 20 mL 20% acetonitrile. After dissolving the resin in the digestion buffer (100 mM Tris, pH 8, 2 mM CaCl2 and

10% acetonitrile), 0.5 µg trypsin (Promega) was added and incubated overnight at 37 °C. The peptide solution was collected and the resin was washed with 500 µL water. Both solutions were combined and acidified with 20 µL 10% CF3COOH. Peptides fractionated into 12 fractions on Agilent 3100 OFFGEL Fractionator was performed, as described above (Section 1). IEF fractions 1 and 3, 4 and 5, 6 and 9, 7, 8 and 12, 10 and 11 were combined prior to LS-MS/MS.

3.2.4 Liquid chromatography-tandem MS (LC-MS/MS)

Peptides were analysed by LC coupled to an LTQ Orbitrap Velos (Thermo Fisher Scientific) using a Proxeon nanospray source. Reverse phase chromatography was performed with a nanoACQUITY UltraPerformance LC system (Waters) fitted with a trapping column (nanoAcquity Symmetry C18, 5 µm, 180 µm x 20 mm) and an analytical column (nanoAcquity BEH C18, 1.7 µm, 75 µm x 200 mm) directly coupled to the ion source. The mobile phases for LC separation were 0.1% (v/v) formic acid in LC-MS grade water (solvent A) and 0.1% (v/v) formic acid in can (solvent B). Peptides were separated at a constant flow rate of 300 nl/min with a 3 to 40% solvent B gradient for 145 min for each IEF fraction. The MS1 scan was acquired in the Orbitrap from m/z 300 to 1,700 at a maximum filling time of 500 ms and 106 ions. The resolution was set to 30,000. Fragmentation was performed in the LTQ by collision induced dissociation, selecting up to 15 most intense ions (top15) at an isolation window of 2 Da, unless stated otherwise. Target ions previously selected for fragmentation were dynamically excluded for 30s with relative mass window of 10 ppm. MS/MS selection threshold was set to 2,000 ion counts. A lock mass correction was applied using a background ion (m/z 445.12003).

3.2.5 Data processing

Raw files were processed with MaxQuant [22] version (1.2.0.17) and the Andromeda search engine [12]. The MS/MS spectra were searched against the database containing the forward and reverse Human SwissProt databases and common contaminants. The precursor mass tolerance was set to 20 ppm for the first pass and 6 ppm for the 2^{nd} pass. The fragment mass tolerance was 0.5 Da. Quantification was done by DimethylLys0 + DimethylNter0 for the light labelling and DimethylLys4 + DimethylNter4 for the intermediate labelling. Unique- and razor peptides were used for quantification, using only unmodified peptides without discarding unmodified counterpart peptides. Cysteine-carbamidomethylation and methionine-oxidation were set for the fixed modification and variable modification, respectively. The minimum peptide length was set to 6, the maximum allowed miss-cleavage was 2 and the false discovery rate (FDR) was set to 0.01 for both peptide and protein identifications. Re-quantification, match between runs, and intensity based

absolute quantification (iBAQ) were also performed. After MaxQuant processing, reverse and contaminant proteins and proteins with only one peptide identification were discarded. MaxQuant-computed raw protein ratios were normalized with median and median absolute deviation (MAD) using the following scheme:

$$\bar{X}_j = \frac{X_j - median}{MAD}$$

where X_j is a raw protein ratio.

3.2.6 Significance test for differential expression (DE)

Significance test for DE protein groups was performed using the limma R package [70]. As limma assumes independent sample variances, the NSpool-GNSpool sample was not used for this analysis. Since the dependence of ratio variance on the intensity was observed, data was split into intensity bins with each containing 300 protein groups and significance test was applied to each bin. Multiple hypothesis testing correction was done with Benjamini-Hochberg FDR threshold of 0.05. Subsequently, protein groups with expression value less than 2-fold were discarded. For the secretome analysis, the fold-change filter was not applied.

Cell line-specific DE protein groups were defined as 1) those that were quantified at least in three out of the four cell lines including the cell line in question, and 2) those that had expression values less than +-2-fold in the cell line in question but more than 4-fold in the other three cell lines, or those that had expression values more than +-4-fold in the cell line in question but less than 2-fold in the other three cell lines in the reverse expression direction.

3.2.7 Prediction of secreted proteins

Computational prediction of secreted proteins was performed using SignalP (http://www.cbs.dtu.dk/services/SignalP/), UniProt keywords (http://www.uniprot.org/) “Signal”, “Secreted” and “Extracellular space”. If a protein is positive at least in one of these four criteria, that protein is considered secreted.

3.2.8 Gene set enrichment analysis

In each of the two experiments the DE proteins were subjected to a Fisher's exact test for chromosomes, Panther pathway [71], OMIM (http://omim.org/), CORUM [72], Gene Ontology Biological Processes (GO.BP) (http://www.geneontology.org/), Reactome Pathway [73], embryonic stem (ES) cell-signature gene sets [74], Signalling Pathway Impact Analysis [75] and Molecular

signature database [76]. Multiple test correction was done with Benjamini-Hochberg correction.

3.2.9 Transcriptome data

The tag-seq data was obtained from [77] and processed the same way. Glioblastoma microarray data, survival information and other related metadata were retrieved from The Cancer Genome Atlas (TCGA) and processed as described in [77].

3.2.10 Transcription factor annotation and their target genes

The "transcription factor" annotation was retrieved from Panther, MetaCore and UniProtKB. Transcription targets were retrieved from MetaCore (by Thomson Reuters) using "search for genes that transcriptionally directly interact downstream with the protein of interest" in Meta search programme.

3.2.11 Immunocytochemistry

Cells were fixed in 4% paraformaldehye (PFA) for 8 min and then washed/permeabilised using wash buffer (1× PBS+0.1% Triton-X 100). Blocking was carried out for 1 h using a blocking solution (wash buffer plus 3% goat serum and 1% BSA). Primary antibodies were incubated overnight at 4 °C at the appropriate dilution in blocking solution. Excess primary antibody was washed for 5 min (2 times) and then 15 min (2 times) and secondary antibodies were incubated for at least 1 h at room temperatures before another round of washing. Images were taken using an Olympus IX50 inverted fluorescent microscope with a DP-50 camera. DAPI was used as a nuclear counterstain. Primary antibodies used were: Tenascin-C (1:100, Sigma), Galectin-3 (1:100, abcam), THY1 (1:100, Millipore), CD9 (1:100, Millipore), TES (1:100, SigmaAb1), mouse Nestin (1:10, DSHB), human Nestin (1:500, R&D) and KI67 (1:1000, Labvision). Species-specific or Ig-subtype specific goat secondary antibodies were used throughout with either Alexa488 or Alexa594 fluorophores (Molecular Probes/Invitrogen).

3.2.12 Time lapse imaging and growth factor screening

For time-lapse imaging and generation of growth curves, we used the Incucyte system (Essen Instruments, USA). Cells were plated at 5%–30% confluence on 12-well plates (Falcon) in the standard culturing media, incubated for 1 day and then factors were added to the wells. Confluence readings were obtained at each time point. CellProfiler (http://www.cellprofiler.org/) was used to count the number of cells. The parameters used were; typical diameter of objects between 13 and 70

pixels, the three-class Otsu-Adaptive threshold method, clumped objects distinguished by shape and divided by intensity. The default settings were used for the other parameters. For cell tracking analysis, we processed image stacks using ImageJ and analyzed cell tracks using the MTrackJ Plugin (http://rsb.info.nih.gov/ij/). The factors used for screening were; Tenascin-C (Millipore), Midkine (PeproTech), CNTF (PeproTech), APOE3 (PeproTech), IGFBP3 (PeproTech), IGFBP4 (PeproTech) and CSF1 (PeproTech).

3.3 Results

3.3.1 446 protein groups were differentially expressed in total cell proteomes of GNS and NS cells

The total proteomes of four GNS lines (G166, G144, G25 and G179) were compared to the reference pool of NS cells (containing CB660, CB192, CB152 and CTX985) in a total of 120 LC-MS/MS runs. 7476 protein groups were quantified across the five total cell samples with two label-swapped technical replicates, each with a complete overlap of 6492 proteins (Table 3-1). The mean number of unique peptide identifications for each protein group was 11, and the mean sequence coverage was 17%. The mean sample correlation coefficient based on protein groups present in all the samples was 0.38. A significance test yielded 743 DE proteins, 719 of which were present in all the four GNS samples. The proteins with less than 2-fold change were discarded, leaving 464 DE protein groups, 446 proteins of which were found in all the four samples. 152 of the 446 proteins were up-regulated and 294 were down-regulated (Figure 3-1). These DE proteins were used in the subsequent analyses.

Table 3-1. Number of quantified protein groups in total cell experiment. Differential expression was defined as adjusted p-value <=0.05 and absolute log2(mean fold change) >= 1 (Up1 and Down1), and as adjusted p-value <=0.05 and absolute log2(fold change of each sample) >= 1 (Up2 and Down2). Complete overlap indicates number of proteins quantified in all the four samples. *NSpool-GNSpool was not used for the significance test since its variance is not independent of the other samples.

Sample	Quantified	DE	Up1	Down1	Up2	Down2
*CBall – Gall	7064	-	-	-	-	-
CBall – G166	6765	461	167	294	166	292
CBall – G144	7215	455	161	294	161	291
CBall – G25	6891	459	164	295	164	295
CBall – G179	7068	463	168	295	168	292
Complete overlap	6290	446	155	294	155	294
Total	7476	464	166	295	166	295

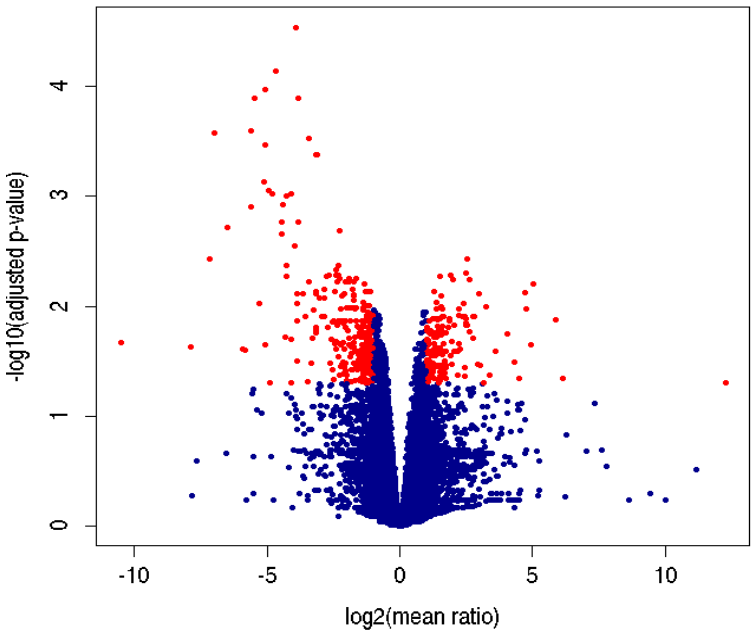


Figure 3-1. Volcano plot of protein ratios of total cell experiments. Red dots are DE proteins.

3.3.2 Galectin-3, Galectin-3-binding protein, L1CAM, GFAP were over-expressed, while integrin α6 and ALDH2 were under-expressed in our GNSs

Several glioma stem cell markers have been identified and used previously, the expression of which we examined across the four GNS cell lines used here. Candidate markers include CD133 (Prominin 1), CD44, CD15, L1CAM, A2B5, SSEA-1, Nestin, SOX2, MELK, CXCR4, Olig2, LGALS3 (galectin-3), LGALS3BP (galectin-3-binding protein), Musahi-1, BMI-1, ITGA6 (integrin α6), podoplanin, ALDH2 [78, 79]. The expression of TBR2, DLX2 as markers for transiently-amplifying cells and GFAP as a marker for astrocytes was also checked. The result showed that LGALS3, LGALS3BP, L1CAM and GFAP were differentially over-expressed, while ITGA6 and ALDH2 were under-expressed in GNSs (Figure 3-2). All the others were either not DE or not identified in this analysis. The pluripotency factors Nanog, OCT4 and KLF4 were not identified and SOX2 was not DE. None of the epithelial (CDH1) and mesenchymal markers (CDH3, VIM, FN1, Zeb2, FOXC2, SNAIL1, SNAIL2, TWIST1 and TWIST2) were DE (data not shown).

LGALS3 is an adult astrocyte stem cell marker [80] known to have many functions; one of them may promote cell migration that was induced by CSPG4 via α3β1 integrin [81]. Wei et al [82] found that the GBM-initiating cells markedly inhibited T-cell proliferation and activation, induced regulatory T cells and triggered T-cell apoptosis that was mediated by B7H1 (CD274?) and soluble LGALS3. These immunosuppressive properties were diminished on altering the differentiation of the GBM-initiating cells, suggesting the role of LGALS3 in immunosurveillance evasion. LGALS3BP was also up-regulated in our GNSs. It is an ECM protein that can bind to LGALS3 and several other proteins [83] and may promote cell adhesion [84]. L1CAM (neural cell adhesion

molecule L1) is a neuronal cell adhesion molecule. Cheng et al. [85] has shown that the DNA damage checkpoint response and radioresistance of GNSs is regulated in part by L1CAM through the activation of the ATM kinase pathway. The L1CAM expression patterns was similar to a stem cell marker CD133 [86-90], and was required for maintaining the growth and survival of CD133+ glioma cells both in vitro and in vivo [91]. CD133 is a five transmembrane glycoprotein whose function is still largely unknown. Although it has been used to isolate brain tumour stem cells from different brain tumours [46, 47], it was identified only in our G166 and G179 lines and not DE when compared to the NSs and also not on the mRNA level [77], suggesting that this marker may not be reliable in separating between NSs and GNSs. The absence of CD133 expression in glioma stem cells was also observed in Beier et al. [92] and Pollard et al. [4]. GFAP (glial fibrillary acidic protein) is an astrocyte marker but was also expressed in some GNS cell line [4]. Since the GFAP isoforms do not seem to be annotated in the uniprot database we used for the protein identification, it is uncertain whether GFAP α or δ/ε were over-expressed. ITGA6 (Integrin α6) is a receptor for the ECM protein laminin and its mRNA was highly expressed in embryonic-, neural- and heamatopoietic stem cells [93]. Lathia et al. [94] showed that ITGA6 was differentially up-regulated in CD133+ glioma stem cells in comparison to CD133- glioma stem cells, and regulated the CD133+ population. However, in our data ITGA6 was under-expressed in GNSs. A more discussion about integrins is found in section (3.3.3 and 3.3.4). ALDH2 (mitochondrial aldehyde dehydrogenase) was also down-regulated in our GNSs. It is a detoxifying enzyme that can oxidise intracellular aldehydes [95] and was identified as a CSC marker in hepatocytes [96], leukemia [97], lung [98], head and neck [99], pancreas [100] and ovary [101]. Chute et al. [102] showed that inhibition of ALDH promoted heamatopoietic stem cell self-renewal via reduction of retinoic acid activity, implying that down-regulation of ALDH in our GNSs may be contributing to their increased proliferation. Indeed, a positive correlation between ALDH expression and patient survival was observed in many cancers [103]. In summary, several known glioma stem cell markers were DE between NSs and GNSs, which could be used for distinguishing both cell types. However, heterogeneity between GNSs cells may compromise general applicability of some markers.

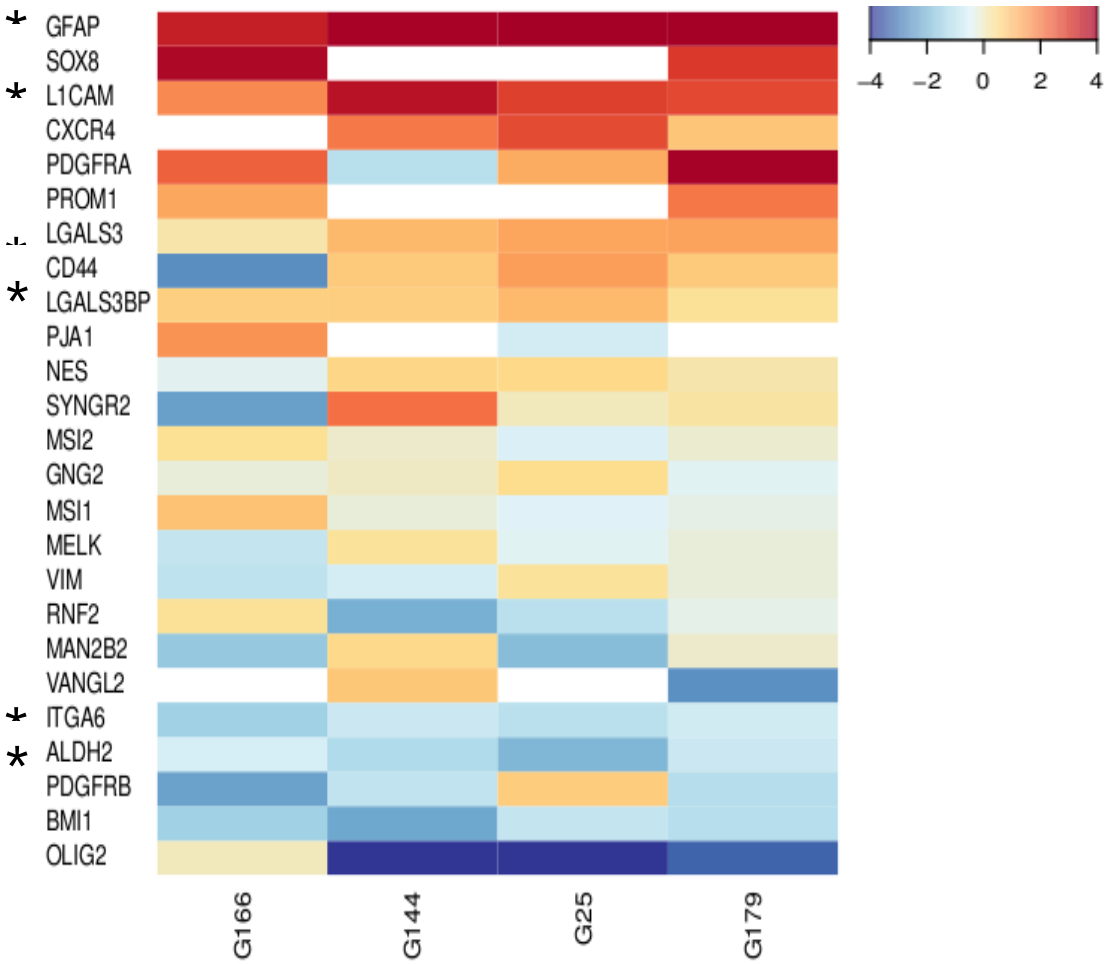


Figure 3-2 Heatmap of known NS/GNS marker expression. Colours indicate log2(fold change) of protein ratios.

3.3.3 Gene set enrichment analysis and signalling pathway impact analysis captured known chromosomal aberrations and putative tumour-associated processes

The 446 DE proteins completely overlapping across all the four comparisons were subjected to the GSEA and signalling pathway impact analysis [75] [75]. The result is summarized in Table 3-2. The over-representation of the glioma pathway (KEGG) reassures the quality of our data, as this pathway was expected to be disregulated and was also over-represented in the transcriptome study [77]. Over-representation of chromosomes 7 and 15 and under-representation of chromosome X were observed. Consistent with the GSEA, the protein expression of chromosome 7 appeared higher than the average, whereas that of chromosome 15 appeared lower in each of the cell lines (Figure 3-3). The gain of chromosome 7 and loss of 15 were previously reported [4, 77] and the under-representation of X is most likely due to the sample genders, where three out of the four GNS samples are male whereas NS samples are half male and half female. These findings suggest that

chromosomal aberrations influence not only mRNA- but also protein expression levels. The over-representation of Gene Ontology Biological Processes (GO.BP) “cell differentiation” and “neuron differentiation” is in line with the hypothesis that tumourigenesis starts with a block of differentiation and concomitant mitotic arrest followed by uncontrolled proliferation. Although many DE proteins bearing “cell proliferation” categories were found in our data, the GSEA did not result in cell proliferation-related categories, presumably because both NS and GNS cells were cultured in the proliferating conditions. Enrichment in GNSs of cell motility-related categories such as regulation of cell migration (GO.BP), cell junction organization (Reactome Pathway) and gap junction (KEGG) could be explained by the increased cell motility and invasive property of GNSs. Several integrin-related categories such as integrin signalling pathway (GO.BP, PantherPathway), cell-cell adhesion mediated by integrin (GO.BP), integrin-mediated cell adhesion (WikiPathway), integrin cell surface interactions (Reactome Pathway) were enriched. In addition, many integrin-related protein complexes were over-represented in CORUM and Reactome Complex. Integrins are heterodimer cell surface receptors that consist of α and β subunits, through which the extracellular matrix (ECM) modulates cell behaviour including cellular shape, motility and cell cycle progression [104]. Interactions between the extracellular matrix and the actin cytoskeleton commonly take place at focal adhesions on the cell surface that contain localized concentrations of integrins, signalling molecules and cytoskeletal elements (http://www.biocarta.com/pathfiles/h_integrinpathway.asp). In agreement with this, focal adhesion (WikiPathway, SPIA), ECM-receptor interaction (KEGG), ECM organization (GO.BP) and cell-ECM interactions (Reactome Pathway), and several cytoskeleton related processes such as cytoskeletal regulation by Rho GTPase, regulation of cytoskeletal remodelling and cell spreading by IPP complex components (Reactome Pathway) and regulation of actin cytoskeleton (WikiPathway, KEGG) were over-represented, suggesting the importance of ECM re-organization for tumourigenesis, a property that apparently is maintained even in cultured cells. Among ECM interactions, the L1CAM interactions (Reactome Pathway) was also over-represented. Apart from the integrin signalling pathway, several signalling pathways also appeared in the analysis, including MAPK signalling pathway (KEGG), ERK1 activation (Reactome Pathway) Negative regulation of TGF-beta receptor signalling pathway (GO.BP), positive regulation of calcium-mediated signalling (GO.BP), response to elevated platelet cytosolic Ca2+ (Reactome Pathway), signalling of hepatocyte growth factor receptor (WikiPathway), GnRH signalling pathway (KEGG). This indicates that disregulation of these signalling pathways could be associated with tumouriginicity. However, the expression levels alone do not tell if these pathways are activated or inhibited since signalling cascades also depend on post-translational modifications of proteins such as phosphorylation, sub-cellular locations of the molecular players and mutations on amino acid residues that could disrupt protein-protein interactions. The categories such as blood

vessel development (GO.BP), platelet activation, signalling and aggregation (Reactome Pathway) and VEGF signalling pathway (KEGG) strongly indicates the role of angiogensis in our GNS-mediated tumourigenesis. Endocytosis (GO.BP) is known to digest ECM components such as Cadherins and its over-representation in our GNSs may suggest an enhanced microvesicle transport of RNA and proteins, which was previously reported in glioma cells [105].

Taken together, our enrichment analyses captured known chromosomal aberrations and many biological pathways and processes that could be responsible for the malignancy. However, it is hard to tell from an enrichment analysis which of these enriched categories are the actual "drivers" for the tumouriginicity and which are just "bystanders". It should also be noted that a single protein may be present in multiple processes and it is difficult to distinguish between the effects of that particular protein on different pathways. Thus, we looked into further details on the DE proteins.

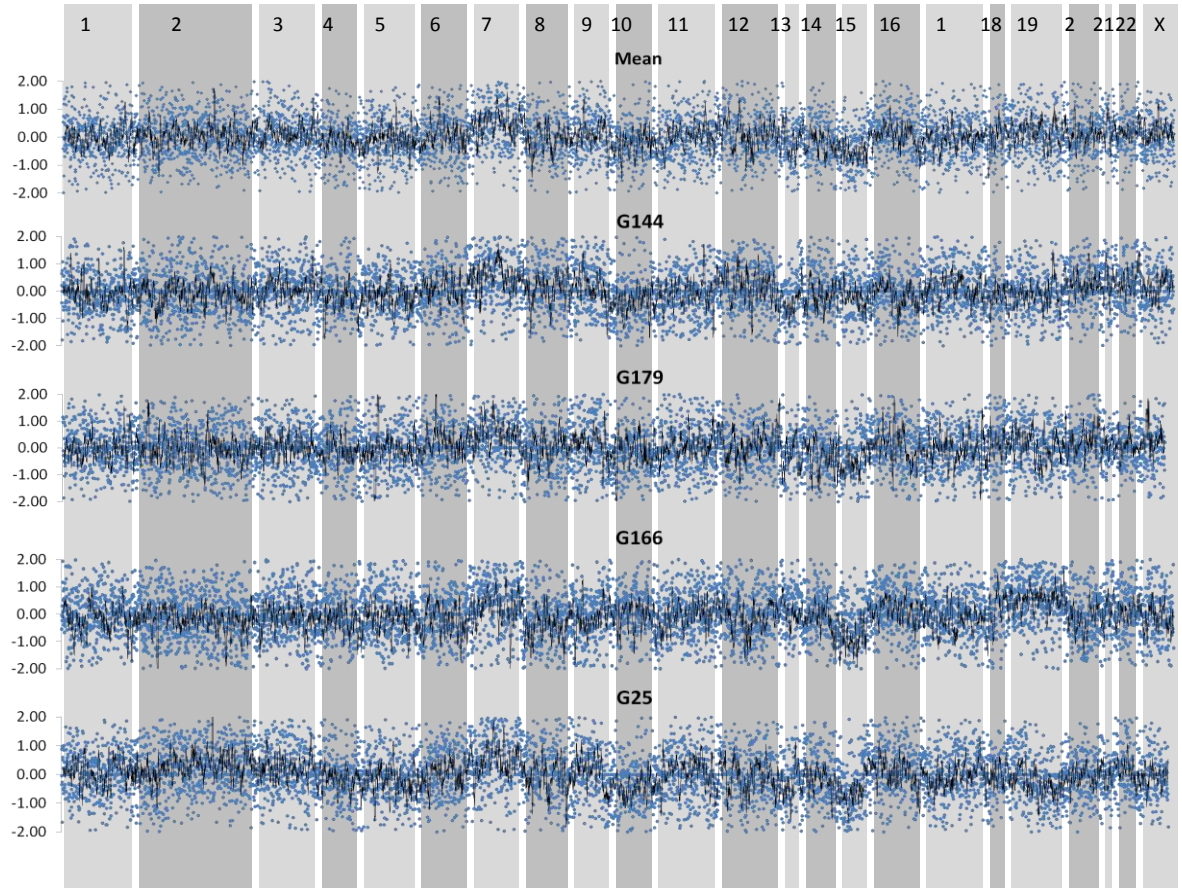


Figure 3-3. log2(Protein ratios) (blue dots) as a function of chromosomal locations. Solid black line is a moving average of 10 ratios. Chromosome number is indicated on top.

Table 3-2 A Selected categories enriched by DE proteins in chromosomes, OMIM, Gene Ontology, Panther Pathway, Reactome, WikiPathway, InterPro and CORUM. Annotation: total number of genes in category, DE genes: number of differentially expressed genes, oddsRatio: measure of over-representation (> 1 is over-representation, < 1 is under-representation), padj: Fisher's exact test p-value adjusted with Benjamini Hochberg correction.

	Differentially expressed (446 proteins)			
	Annotation	**DE genes**	**oddsRatio**	**padj**
A. Chromosome				
15	225	27	2.1	0.011
7	366	36	1.7	0.035
X	269	6	0.4	0.043
B. OMIM				
ESOPHAGEAL CANCER	3	3	16.8	0.023
F. GO.BP				
cell differentiation	1001	97	1.8	0.000
endocytosis	176	29	2.9	0.000
neuron differentiation	444	53	2.1	0.000
response to chemical stimulus	1103	100	1.7	0.001
integrin-mediated signaling pathway	31	9	4.9	0.004
NAD metabolic process	18	7	6.6	0.004
blood vessel development	187	24	2.2	0.011
isocitrate metabolic process	6	4	11.2	0.014
regulation of cell migration	146	20	2.4	0.015
multicellular organismal signaling	242	28	2.0	0.015
response to metal ion	108	16	2.5	0.019
cell-cell adhesion mediated by integrin	7	4	9.6	0.020
negative regulation of transforming growth factor beta receptor signaling pathway	17	5	5.0	0.040
positive regulation of calcium-mediated signaling	5	3	10.1	0.045
extracellular matrix organization	65	10	2.6	0.049
C. Panther pathway				
Integrin signalling pathway	113	26	4.0	0.000
Cytoskeletal regulation by Rho GTPase	45	9	3.4	0.023
D. Reactome pathway				
Cell-extracellular matrix interactions	11	10	15.5	0.000
Cell junction organization	34	13	6.6	0.000
L1CAM interactions	70	16	3.9	0.000
Regulation of cytoskeletal remodeling and cell spreading by IPP complex components	5	5	16.9	0.002
Response to elevated platelet cytosolic Ca2+	38	10	4.5	0.003
Localization of the PINCH-ILK-PARVIN complex to focal adhesions	3	4	22.5	0.004
Platelet activation, signaling and aggregation	101	16	2.7	0.008
Integrin cell surface interactions	45	10	3.8	0.009
Sema4D in semaphorin signaling	22	7	5.4	0.009
c-src mediated regulation of Cx43 function and closure of gap junctions	3	3	16.8	0.023
ERK1 activation	3	3	16.8	0.023
Phase 1 - Functionalization of compounds	11	4	6.1	0.046
E. Wikipathway				
Focal Adhesion	117	18	2.6	0.006
Integrin-mediated cell adhesion	70	12	2.9	0.017
Regulation of Actin Cytoskeleton	92	14	2.6	0.020
Signaling of Hepatocyte Growth Factor Receptor	27	7	4.4	0.020
G. GO.MF				
actinin binding	8	7	14.8	0.000
integrin binding	43	13	5.2	0.000
calmodulin binding	85	15	3.0	0.005
isocitrate dehydrogenase activity	5	4	13.5	0.010
DNA binding	817	29	0.6	0.021

Table 3-2 A (continued)

	Differentially expressed (446 proteins)			
	Annotation	**DE genes**	**oddsRatio**	**padj**
H. InterPro				
Znf_LIM	40	15	6.5	535991460478
CH-domain	54	11	3.5	0.009
EF_HAND_2	88	14	2.7	0.015
Spectrin/alpha-actinin	19	6	5.3	0.018
Isocitrate_DH_NAD	3	3	16.8	0.023
Myosin_head_motor_dom	21	6	4.8	0.024
EF_hand_Ca-bd	57	10	3.0	0.028
ZU5	4	3	12.6	0.035
EPS15_homology	10	4	6.7	0.039
I. CORUM				
40S ribosomal subunit, cytoplasmic	29	10	5.9	0.001
LRP-1-Alpha-2-M-annexin VI complex	3	4	22.5	0.004
40S ribosomal subunit, cytoplasmic	140	21	2.6	0.004
Polycystin-1 multiprotein complex (ACTN1, CDH1, SRC, JUP, VCL, CTNNB1, PXN, BC/	9	5	9.4	0.008
ITGAV-ITGB3-COL4A3 complex	5	4	13.5	0.010
ITGAV-ITGB5-SPP1 complex	5	4	13.5	0.010
ITGAV-ITGB6 complex	3	3	16.8	0.023
ITGA11-ITGB1 complex	3	3	16.8	0.023
ITGAV-ITGB3-SPP1 complex	4	3	12.6	0.035
ITGAV-ITGB3-ADAM15 complex	5	3	10.1	0.045
ITGA9-ITGB1-FIGF complex	1	2	33.6	0.050
J. Reactome complex				
Alpha 11 beta 1 integrin: Collagen type-I:Mg++ complex [plasma membrane]	4	4	16.9	0.006
L1:Integrin complex [plasma membrane]	4	4	16.9	0.006
Integrin alpha2beta1:Collagen I:Mg++ [plasma membrane]	4	4	16.9	0.006
Cx43:ZO-1:c-src hemi-channel [plasma membrane]	3	3	16.8	0.023
phospho-Y265 Cx43:ZO-1 gap junction [plasma membrane]	3	3	16.8	0.023
Cx43:ZO-1:c-src gap junction [plasma membrane]	3	3	16.8	0.023
PINCH-ILK-parvin complex [cytosol]	3	3	16.8	0.023
IDH3 complex [mitochondrial matrix]	3	3	16.8	0.023
Docked Cx43-containing transport vesicles [plasma membrane]	10	4	6.7	0.039
connexons in Golgi transport vesicle docked to microtubules [cytosol]	10	4	6.7	0.039
Calcium Bound Sarcomere Protein Complex [cytosol]	10	4	6.7	0.039
ADP:Calcium Bound Sarcomere Protein Complex [cytosol]	10	4	6.7	0.039
Inactive Sarcomere Protein Complex [cytosol]	10	4	6.7	0.039
ATP:Calcium Bound Sarcomere Protein Complex [cytosol]	10	4	6.7	0.039

Table 3-2 B Selected KEGG pathways from Signalling Pathway Impact Analysis of DE proteins. Annotation: total number of genes in category, DE genes: number of differentially expressed genes, padj: p-value from Signalling Pathway Impact Analysis [75].

Pathway	Annotation	DE genes	padj	Predicted status in GNS lines
Focal adhesion	130	24	0.000	Inhibited
Gap junction	53	14	0.001	Inhibited
Regulation of actin cytoskeleton	132	23	0.001	Inhibited
MAPK signaling pathway	127	15	0.017	Inhibited
Leukocyte transendothelial migration	52	9	0.043	Activated
GnRH signaling pathway	52	8	0.044	Inhibited
Glutamatergic synapse	53	7	0.044	Inhibited
VEGF signaling pathway	43	7	0.045	Inhibited
ECM-receptor interaction	52	9	0.066	Activated
Glioma	44	6	0.080	Inhibited
Dopaminergic synapse	69	11	0.080	Inhibited
Long-term potentiation	42	8	0.086	Inhibited

3.3.4 Differentially expressed proteins in Gene Ontology Biological Process "neuron differentiation" physically and transcriptionally interact with each other and some of them have no prior association with glioma

Given the hypothesis that a disrupted balance between self-renewal and differentiation initiates tumourigenesis, we more closely looked at the DE proteins in Gene Ontology Biological Process (GO.BP) neuron differentiation. The direct interaction (physical interaction and transcriptional regulation) among these DE proteins retrieved from MetaCore displayed a highly interconnected network (Figure 3-4), underlying the importance of this process to GNS malignancy. We investigated their prior tumour associations to tumour development (Table 3-3). Several integrins (αV, α6, β1) were present among these DE proteins belonging to GO.BP neuron differentiation. Since integrins are obligate heterodimers, these integrins likely form a complex and play a role in distinguishing between normal NSs and GNSs. In breast cancer integrin α6β1 and neuropilin-2 regulate the formation of focal adhesions [106]. Integrin α6 was discussed in section (3.3.2). Integrin beta-1 (ITGB1) is a heterodimeric receptor involved in cell-matrix and cell-cell adhesion, and has been implicated to play a crucial role in the maintenance of stemness by controlling the angle of cell division by interacting with astral microtubules that regulate centrosome positioning. Since the loss of asymmetric division could over-amplify the stem cell pool, disrupted integrin beta-1 functionality is thought to lead to tumourigenesis [107]. In fact, many structural proteins such as myosins (MRCK, MYO6, MYH9, MYH11, MYH14, MYRL2), tubulins (TUBB1, TUBB2, TUBB3), collagens (COL1A1, COL1A2, COL1A3), GJA1, Dystrophin and SPTBN2, were also among the DE proteins in the GO.BP neuron differentiation category, illustrating the possible roles in tumourigenesis played by these proteins. Disregulated signalling pathways are likely involved in this process, as indicated by the presence of kinases/phosphatases (ILK, LIMK1, MAP2Ks, MAPKs, SRC). Other connected components include Tenascin-C (TNC) is an ECM glycoprotein that is abundantly expressed in foetal NSs and vanishes as the organism matures and is absent in normal adult brains [108] but re-expressed upon injury or neoplasia. This agrees with our data where TNC was over-expressed in our GNSs. Tnc is thought to enhance the sensitivity of mouse embryonic NSs to EGF by increasing EGFR expression [109]. It also has putative EGFR binding domains [110]. Mouse oligodendrocyte progenitor cells (OPCs) proliferate less but migrate faster within the optic nerves of Tnc-deficient mice [111] and that cultured OPCs from Tnc-deficient mice display higher maturation rates [112]. Thus, up-regulation of TNC may contribute to sustained proliferation. THY1 (CD90) is a GPI-anchored, plasmamembrane protein localized to lipid rafts expressed on human fibroblasts, neurons, blood stem cells, endothelial cells and murine T cells [113, 114]. It is also a common marker for mesenchymal stem cells, multipotent mesenchymal stromal cells [115] and early stages of iPS reprogramming [116]. THY1 expression correlated with the tumourigenic potential of hepatocellular carcinoma cell lines and was suggested as a putative

liver CSC marker [117]. Furthermore, CD133$^+$ glioblastoma CSCs from primary cultures showed high levels of THY1 mRNA and were resistant to several chemotherapeutic agents [118]. It has no reported association with glioma and the over-expression of THY1 in our GNSs makes it a potential marker for GNS cells but not for normal NSs. LZTS1 (Leucine zipper putative tumour suppressor 1) has been shown be ubiquitously expressed in normal tissues and in uveal melanomas. The expression of this protein is silenced in rapidly metastasizing and metastatic tumour cells but has normal expression in slowly metastasizing or non-metastasizing tumour cells (RefSeq, Nov 2009). LZTS1 was shown to inhibit cancer cell growth through mitosis by activating CDK1 [119]. It is somewhat contradictory that this protein was over-expressed in GNSs, while CDK1 was under-expressed, however, this could be due to difference in the expression of other proteins between somatic cancer cells and cancer stem cells as well as between different tissue types. Alternatively, LZTS1 might have loss-of-function mutations and constitutively activated. EPHB2 (Ephrin type-B receptor 2) has been implicated in many cancers and in intestinal CSCs [120]. NCAM1 (neural cell adhesion molecule 1) is involved in neuron-neuron adhesion, neurite fasciculation and outgrowth of neurites but is also known to be dis-regulated in cancers including glioma. FKBP4 (Peptidyl-prolyl cis-trans isomerase) belongs to the immunophilin protein family, and is involved in immunoregulation and protein folding and trafficking. To our knowledge, no prior association with glioma has been reported.

Other “non-connected” components include CD166 antigen (ALCAM), which belongs to the Single-pass type 1 membrane protein that was shown to regulate long-term HSCs [121]. It is a plasma membrane protein that interacts homophilically and heterophilically with L1/NgCAM and CD6 [122-125]. ALCAM has been shown to be involved in neuronal cell adhesion [122], axon growth and navigation [125], migration [126] and differentiation [127]. ALCAM is selectively present on neurons carrying an axon and not found on neuroblasts or non-neuronal cells, however, its down-regulation in our GNSs may indicate an ALCAM's role in tumourigenesis. Taken together, DE proteins in GO.BP neuron differentiation appear to directly/transcriptionally interact with each other to a large degree, suggesting that these DE proteins are likely disrupting the neuron differentiation in our GNSs. In addition, we highlighted those that have known associations with glioma and those that do not. The latter can be novel candidates for glioma tumourigenesis.

Table 3-3 Differentially expressed proteins belonging to GO.BP neuron differentiation. Mean ratio: mean ratios over four protein samples. 'Association with glioma' and 'Association with other cancer' columns indicate PubMed IDs for pertinent reference. 'many' is put when > 5 references were found.

Chromosome	Gene name	Log2(mean ratio)	Association with glioma	Association with other cancer
3	ALCAM	-1.5	18941255	many
4	ANK2	-2.4	-	21042036
1	CAP1	-1.0	-	-
6	CAP2	-1.6	-	-
16	CBFB	-0.9	-	23160462, 22160378
10	CDK1	-1.4	many	many
17	COL1A1	-3.7	21072323	many
7	COL1A2	1.3	21325292, 18664619	many
2	COL3A1	-4.0	-	-
X	DMD	-1.8	many	many
10	DOCK1	-2.0	22323579, 22080864, 17671188	many
7	EGFR	-1.3	many	many
1	EPHB2	-2.7	many	many
10	FGFR2	-2.6	many	many
12	FKBP4	1.3	-	many
17	GFAP	4.8	many	many
6	GJA1	-2.1	22230665, 22131169, 20512920	many
2	GPC1	1.5	18417614	21996748, 18064304, 17016645, 18064304
5	GPRIN1	2.4	-	-
11	ILK	-1.6	many	many
2	ITGA6	-1.0	many	many
2	ITGAV	-1.3	many	many
10	ITGB1	-1.0	many	many
X	L1CAM	2.8	many	many
7	LIMK1	1.8	-	many
8	LZTS1	2.5	-	many
15	MAP2K1	-1.5	many	many
19	MAP2K2	-0.4	many	many
22	MAPK1	-0.7	many	many
16	MAPK3	-1.2	many	many
16	MYH11	-2.1	-	many
19	MYH14	-2.6	-	-
22	MYH9	-1.9	-	many
20	MYL9	-4.8	-	22898599, 21139803, 20818426, 20551518, 19198601, 17341888
6	MYO6	-1.5	-	20353999, 18543251
11	NCAM1	1.5	many	many
15	NEDD4	-1.6	22217575, 20332230, 18539596	many
15	NPTN	-1.5	-	22586443, 18568347
2	NRXN1	-2.4	-	23236287
4	PDLIM5	-2.1	16549780	22392539, 22454401
1	PHGDH	-1.4	19089318, 23229761	many
11	PICALM	-1.3	-	many
1	PPT1	1.1	-	many
13	RAP2A	-1.4	23093786, 23093778	many
19	RRAS	-1.2	-	many
2	SERPINE2	-3.3	8940166, 10037469, 11814314	many

Table 3-3 (continued)

Chromosome	Gene Symbol	log2(mean ratio)	Association with glioma	Association with neoplasia
5	SLC1A3	-1.8	many	many
11	SPTBN2	1.3	-	-
20	SRC		many	many
1	SRGAP2	-2.0	-	-
11	THY1	2.5	16354539	many
9	TNC	1.1	many	many
20	TUBB1		21807073	many
6	TUBB2A	-1.0	21807073	many
16	TUBB3	1.1	21807073	many
1	UBE4B	-1.3	-	many

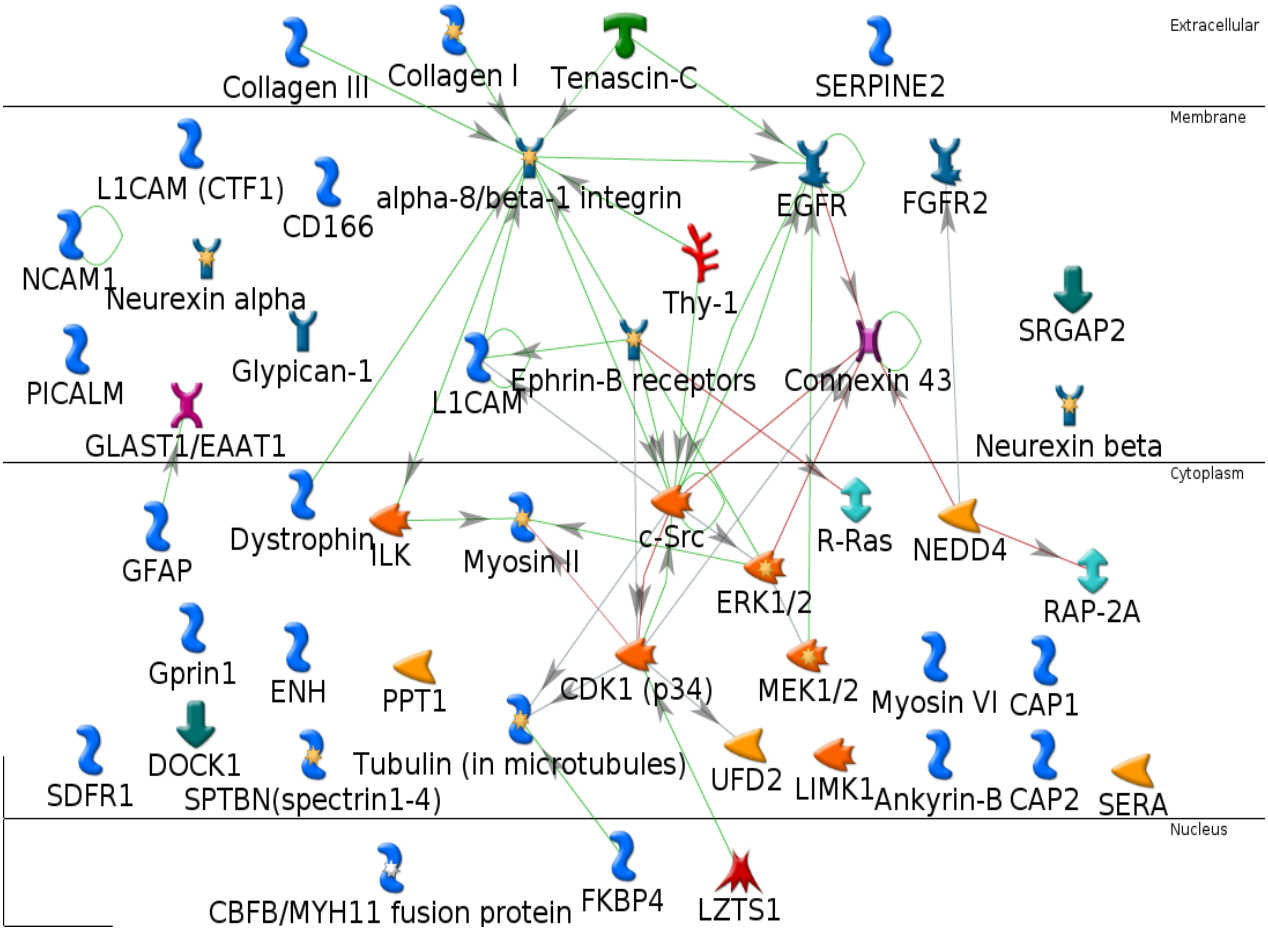


Figure 3-4. Differentially expressed (DE) proteins belonging to GO.BP neuron differentiation in MetaCore direct interaction.

3.3.5 Among 36 DE transcription factors/regulators, several had little prior associations with glioma and could be novel genes for further study

In our data we identified 36 transcription factors that were differentially expressed between GNS and NS cells, 19 of which were up-regulated and 17 down-regulated (Figure 3-5). 16 of them are known to have direct transcriptional target genes, ranging from many hundreds for TP53 and NFIC

to a few dozens or less for most of the others (Table 3-4). Furthermore, eight of them have been implicated in glioma while others have been associated with other cancers. FOXO3 (Forkhead box protein O3) was up-regulated in GNSs and had 238 known targets. It is a transcriptional activator that is known to trigger apoptosis in the absence of survival factors such as IGF1 [128-130] by causing oxidative stress in neuronal cells [131]. Notably, it is regulated by FOXG1 [132] and TP53 [133]. FOXG1 is a key transcription factor that is expressed in non-quiescent forebrain progenitor cells during development and into adulthood [134, 135], involved in the regulation of self-renewal [136] and a reprogramming factor from MEF to NPCs [137]. FOXG1 was over-expressed in GNSs on the mRNA level [77] (not identified in our proteomics data), suggesting that these two proteins may play important roles in maintenance of tumourigenicity. TP53 is a well-known tumour-suppressor gene (TSG) with numerous target genes. Its up-regulation in our GNSs may seem contradictory, although it is not uncommon that this protein undergoes loss-of-function mutations resulting in constitutive activation. So we assume, without proof, that this protein is probably not functioning in our GNSs. NFIC (nuclear factor 1 C-type) was up-regulated in GNSs and had 621 known direct transcriptional targets. Being a member of nuclear factor (NF) proteins, it is involved in driving astrocyte identity/differentiation (personal communication with S.P.). Thus, the observed up-regulation might be because the GNSs originated from astrocytes that de-differentiated and this protein was not silenced afterwards. SATB2 is a nuclear matrix attachment region (MAR) protein that can induce local chromatin-loop remodelling that is expressed in adult brain and to a lesser extent in foetal brain [138]. Our GNSs more highly expressed this protein than the NSs, due presumably to the difference between adult and foetal brains in our samples. HMGA2 is expressed predominantly during embryogenesis and was down-regulated in our GNSs and this was also the case on the mRNA level [77]. Low or absent mRNA expressions has also been observed in glioblastoma tissue in comparison to low-grade gliomas [139] and HMGA2 polymorphisms have been associated with mRNA-based survival time in glioblastoma [140]. Among those without known targets identified in our data, IGHMBP2 (immunoglobulin mu-binding protein 2) has been shown to be expressed in neuronal body and axon [141, 142], implicated as a ribosome-associated helicase [142] and was most highly expressed in our GNSs. This is in favour with the notion that mature neurons de-differentiated into GNSs. LZTS1 is a putative transcription factor with a leucine zipper domain present in the interaction network in GO.BP neuron differentiation discussed above (Figure 3-4). Although LZTS1 was shown to activate CDK1 by physically binding to it, it has no known transcriptional targets as well as no prior association with glioma (Table 3-3), suggesting that this protein is a novel gene for further studies in relation to glioma. Two over-expressed zinc finger proteins ZNF121 and ZNRF2 may be involved in the tumourigenicity but their functions are not well-studied. Finally, STAT3 was not among the 446 DE proteins since it did not pass our cut-

off criteria. Yet, it was consistently under-expressed in all the four GNS cell lines (average log2 fold change -0.98). STAT3 was implicated in maintenance of tumourigenic capacity of colon cancer stem cells [143] and in maintaining the stemness of glioma stem cells [144]. Since it was down-regulated in comparison to the normal NSs, STAT3 may play a role specifically in tumourigenesis, rather than stem cell-related properties, in GNSs. In summary, we identified 36 transcription factors/regulators, eight of which have been implicated in glioma and the others are either associated with other cancers or without any known association.

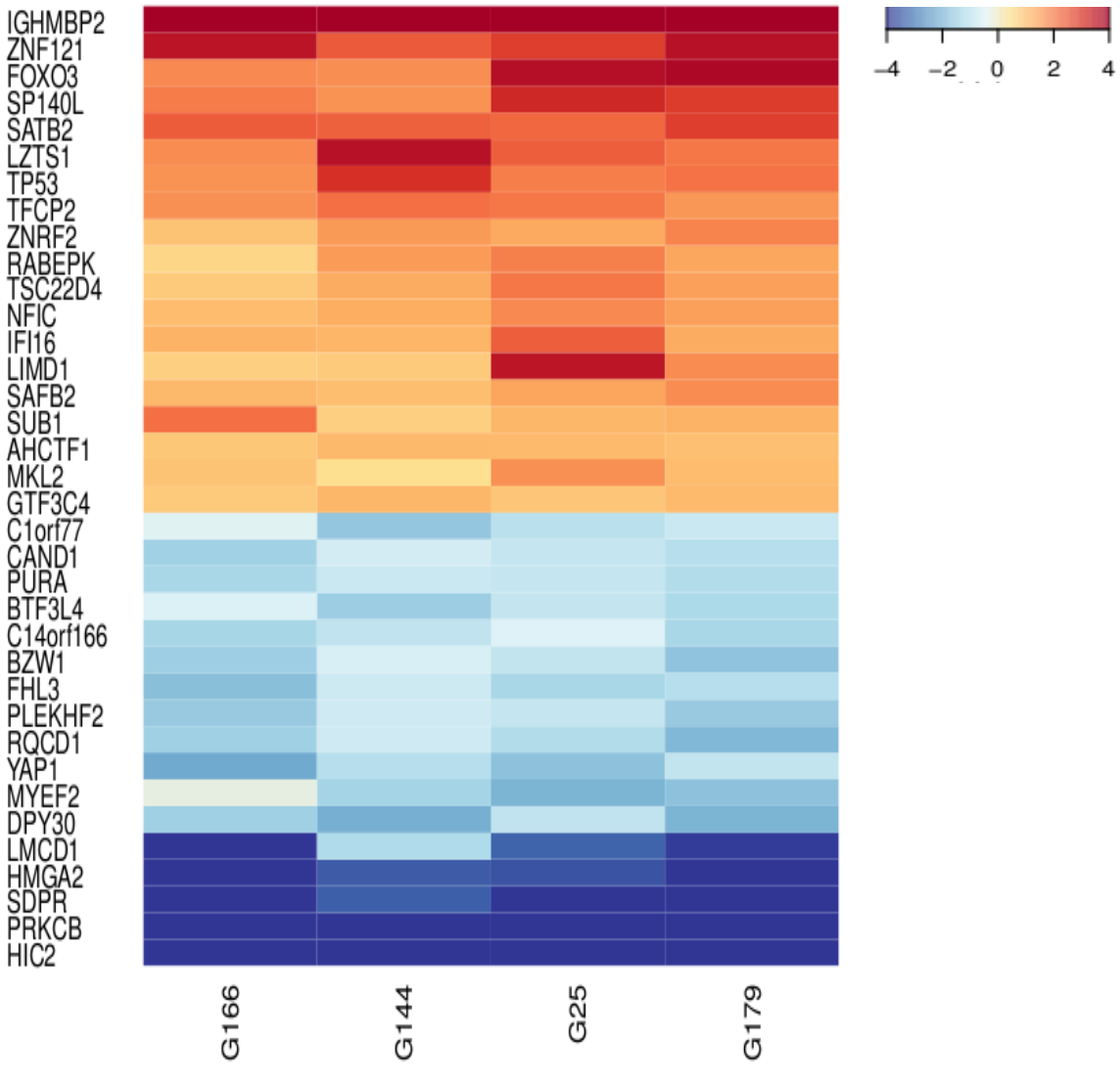


Figure 3-5 Heatmap of DE transcription factors. Column names indicate GNS line. Colours indicate log2(fold change) of protein ratios.

Table 3-4 Number of direct transcriptional target genes of 36 DE transcription factors and their associations with glioma and other cancer types. 'Target gene number' indicates number of known targets present in MetaCore. 'Identified target gene number' indicates number of targets identified in our data. 'Association with glioma' and 'Association with other cancer' columns indicate PubMed IDs for pertinent reference. 'many' is put when > 5 references were found.

Gene names	Target gene number	Identified target gene number	Association with glioma	Association with other cancer
IGHMBP2	2	0	-	16752224
ZNF121	0	0	-	-
FOXO3	238	114	23197693, 22782899	many
SATB2	18	3	-	many
LZTS1	0	0	-	many
SP140L	0	0	-	20056315
TP53	1423	600	many	many
TFCP2	216	96	-	many
LIMD1	0	0	-	many
IFI16	4	1	23387973	many
ZNRF2	0	0	-	-
TSC22D4	0	0	-	23307490
NFIC	621	260	19540848	many
RABEPK	0	0	-	many
SAFB2	0	0	-	19077293, 14587024, 12660241
SUB1	4	1	-	19086899
MKL2	3	0	-	22139079, 20607705, 1717086
AHCTF1	0	0	-	-
GTF3C4	0	0	-	-
CHTOP	1	0	-	-
CAND1	0	0	-	23019411, 17823919
C14orf166	0	0	-	19775290, 19152423
PURA	23	9	18927497, 15517862, 11748591	many
BTF3L4	0	0	-	-
BZW1	0	0	-	19446954
PLEKHF2	0	0	-	-
FHL3	0	0	-	many
RQCD1	0	0	-	20878056, 19724902
MYEF2	3	0	-	-
YAP1	19	9	21666501, 21267586, 19952108, 17114655	many
DPY30	19	5	-	23508102
LMCD1	0	0	-	21996735
HMGA2	18	6	22572881, 21360625, 20368557	many
SDPR	0	0	-	18422756, 17878531
PRKCB	0	0	10417813	many
HIC2	1	0	-	19420922, 17475218

3.3.6 The four cell lines exhibit heterogeneous protein expression patterns

Since 1186 out of 7476 proteins were not identified in all the four samples, we investigated the difference among the cell lines by examining completely overlapping, DE proteins in each line (see Materials and methods for the definition of DE proteins in this analysis). There were 136, 110, 59

and 79, DE proteins in G166, G144, G25 and G179 lines, respectively (Figure 3-6), showing the heterogeneity among them. The interpretation of this heterogeneity, however, is multitudes, ranging from genotypes to environmental effects.

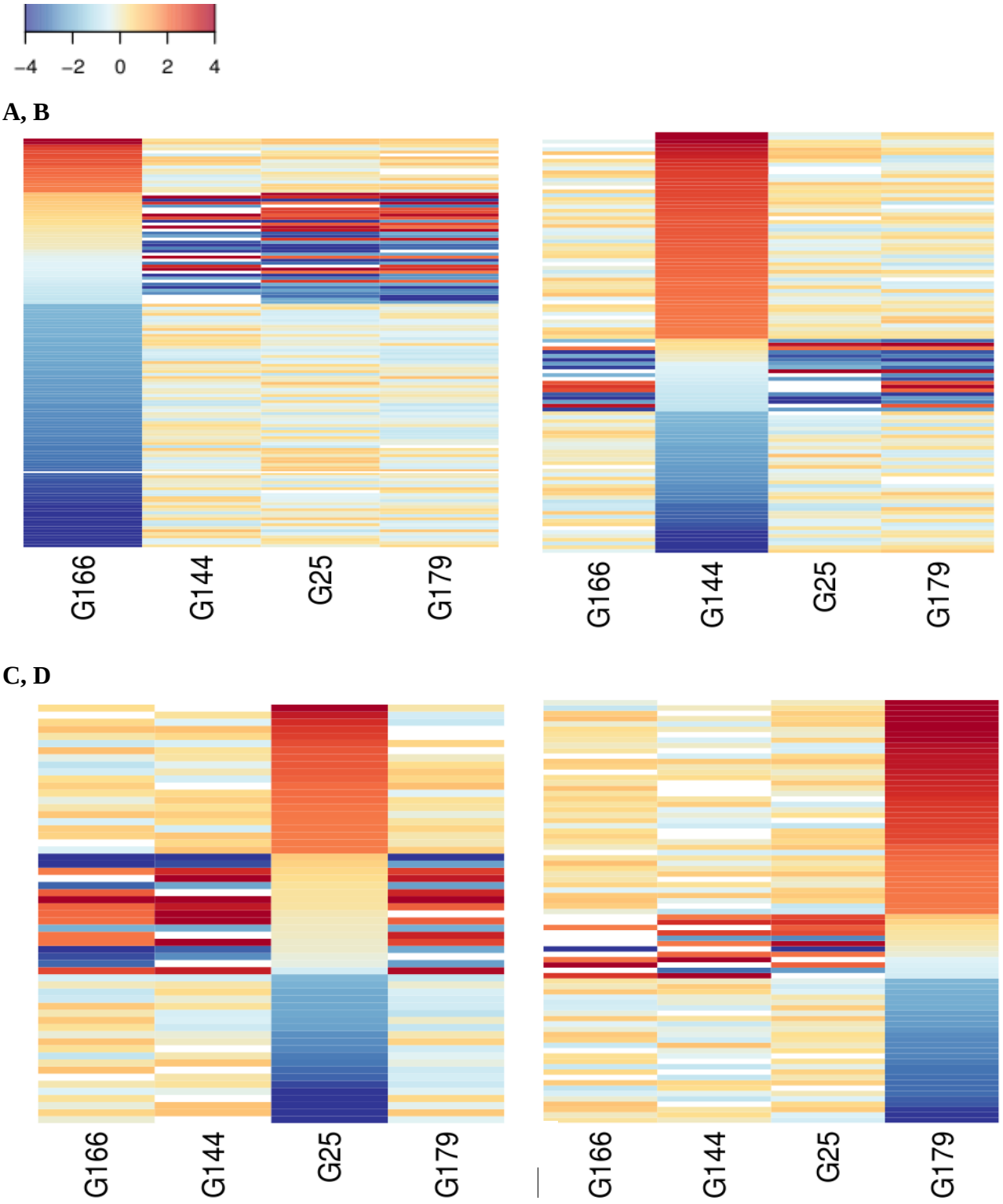


Figure 3-6. Heatmap of cell line specific DE proteins in (A) G166, (B) G144, (C) G25 and (D) G179. Colours indicate log2(fold change) of protein ratios.

3.3.7 Comparison with a study by Thirant et al. [145] shows proteins whose expression patterns are consistent with ours

Our data was also compared to the data by Thirant et al. [145], in which a proteomics analysis was conducted on four of their glioma stem cell (GSC) lines, tumour tissues (TTs) and normal NSs using 2D-gel electrophoresis followed by LC-MS/MS. They identified 108 proteins, 18 were over-expressed in the GSCs, 23 in the NSs and 19 were common to the GSCs and TTs. The overlap between these proteins and our 446 DE proteins was evaluated; four out of 18 proteins up-regulated in their GSCs were overlapped, two (NNMT, YWHAG) of which were also up-regulated in our GNSs. Similarly, eight out of 23 proteins up-regulated in their NSs were overlapped and five (FABP7 (BLBP), ACTR1A, PPP2R1A, CASP3, PEA15) of them were also up-regulated in our NSs. Finally, seven out of 23 proteins up-regulated in their GSCs and TTs were overlapped, two (GANAB, TPI1) of which were also up-regulated in our GNSs. These consistently overlapping proteins could be good markers for distinguishing between GNSs and NSs.

3.3.8 Comparison of proteomics data with transcriptome data increases the confidence of some DE proteins

Since a similar comparison between GNS and NS cells was performed recently by tag-seq data [77], we compared this to our proteomics data. The tag-count of genes without protein identification was overall lower than that of genes with protein identification (Figure 3-7 A), suggesting, as expected, that our proteomics data is biased against low abundance proteins. By superimposing the data (Figure 3-7 B), 41 proteins were DE on both levels (blue dots), 207 proteins were DE only on the mRNA level (red dots) and 405 proteins were DE only on the protein level (green dots). This relatively poor overlap of DE genes may be due partly to posttranscriptional regulation of protein expression but perhaps also to a slightly different set of cell lines used in the two experiments. Next, these overlapping DE genes/proteins were further classified into six categories; 1) up-regulated on both levels (Figure 3-8 A), 2) down-regulated on both levels (Figure 3-8 B), 3) down-regulated on the protein level and up-regulated on the mRNA level (Figure 3-8 C), 4) up-regulated on the protein level and down-regulated on the mRNA level (Figure 3-8 D), 5) DE on the protein level and not DE on the mRNA level (Figure 3-8 E) and 6) DE on the mRNA level and not DE on the protein level (Figure 3-8 F). There were 15, 19, 5, 2, 405 and 207 proteins in these categories, resepctively. GFAP, THY1, NCAM1, LGALS3 and TNC fall in category 1 and HMGA2 and TES in category 2; these were already discussed above. NNMT (category 1) is related to radio-resistance of CSCs [146] and possibly conferring radio-resistance to our GNSs too. CD9 (category 1) is a tetraspanin found in exosomes and has been shown to be able to modulate cell migration and tumour metastasis [147]. TAGLN (category 2) is an actin-binding protein of the calporin family, and has been

characterised as a tumour suppressor gene in non-brain tissues [148]. It is reasonable to assume that if a protein is DE both on mRNA and protein levels in the same direction, then the confidence that the differential expression of that protein and its influence on the cell is strong. Therefore, we selected LGALS3, CD9 and TES as unique candidates from this analysis for further follow-up studies (discussed below**)**. Finally, since categories 5 and 6 each contained many proteins, they were subjected to a GSEA (Table 3-5). Several ECM and signalling-related processes were enriched among the proteins, whereas immunity-related categories were enriched among the mRNAs.

Table 3-5. Selected categories of GSEA on (A). genes DE in only proteins (category 5), and (B) genes DE in only mRNAs (category 6).

A

DE in only protein
Chromosome 15
ITGAV-ITGB3 complexes
actin filament-based process
Platelet activation, signaling and aggregation
integrin-mediated signaling pathway
vesicle-mediated transport
aerobic respiration
Cell-extracellular matrix interactions
Signaling of Hepatocyte Growth Factor Receptor
ERK1 activation

B

DE in only mRNA
Interferon-gamma-mediated signaling pathway
reactive oxygen species metabolic process
antigen processing and presentation of peptide or polysaccharide antigen via MHC class I
lymphocyte activation
response to vitamin D
Ionotropic glutamate receptor pathway
Cadherin signaling pathway
MHC Class II bearing antigen peptide [plasma membrane]
Cytokine Signaling in Immune system
TCR signaling

3.3.9 136 DE proteins were significantly related to patient survival based on public microarray data on glioma tissues

A patient survival analysis was performed on the publicly available GBM and HGG data sets. 5579 out of 20768 genes were significantly (adjusted p-value <= 0.05) associated with the patient survival in HGG and GBM datasets. 136 out of 5579 genes were also among the 446 DE proteins. This proportion was not statistically significant (p-value = 0.199). The GSEA was done on these 139 proteins but no process was enriched.

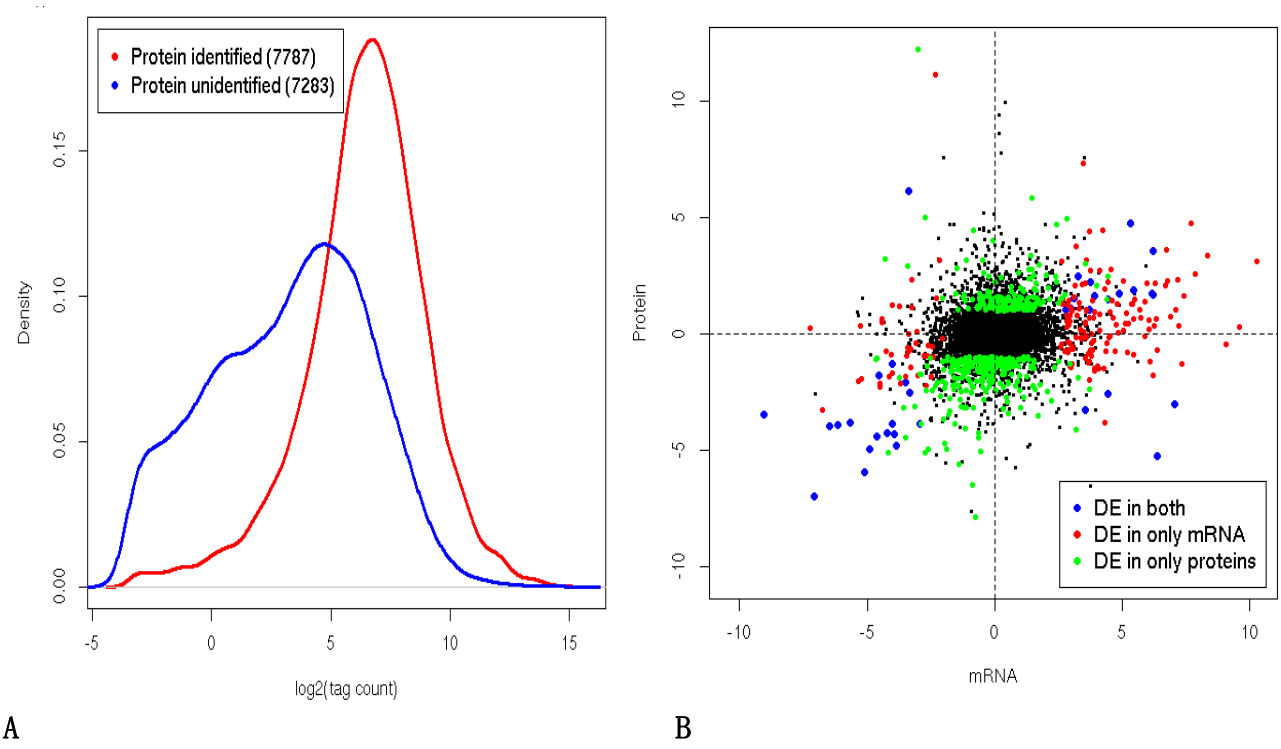
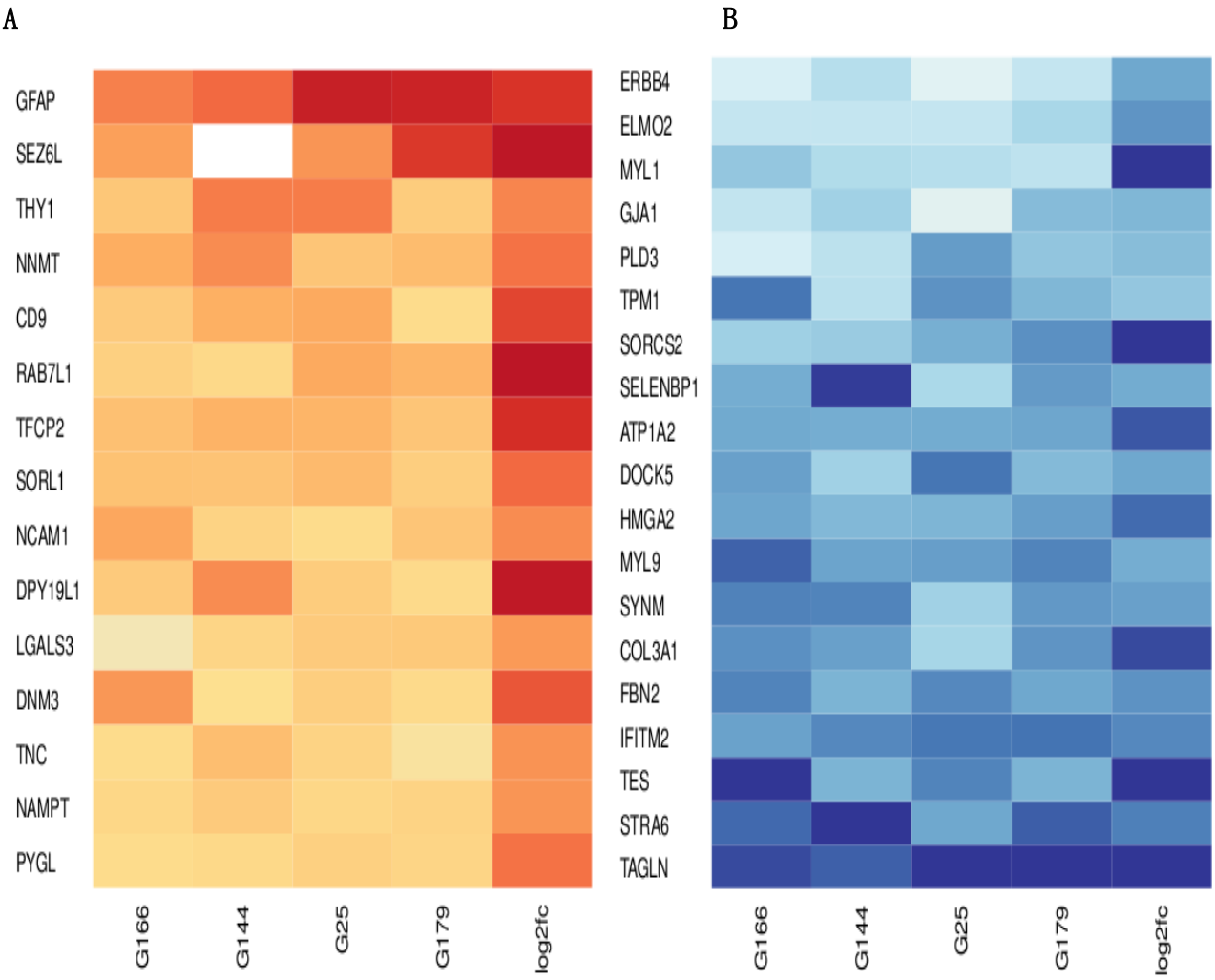


Figure 3-7. (A) Density distribution of tag-seq intensity of genes with corresponding protein identification (red) and without identification (blue). (B) Scatter plot of mRNAs (tag-seq) and protein log2 fold change.



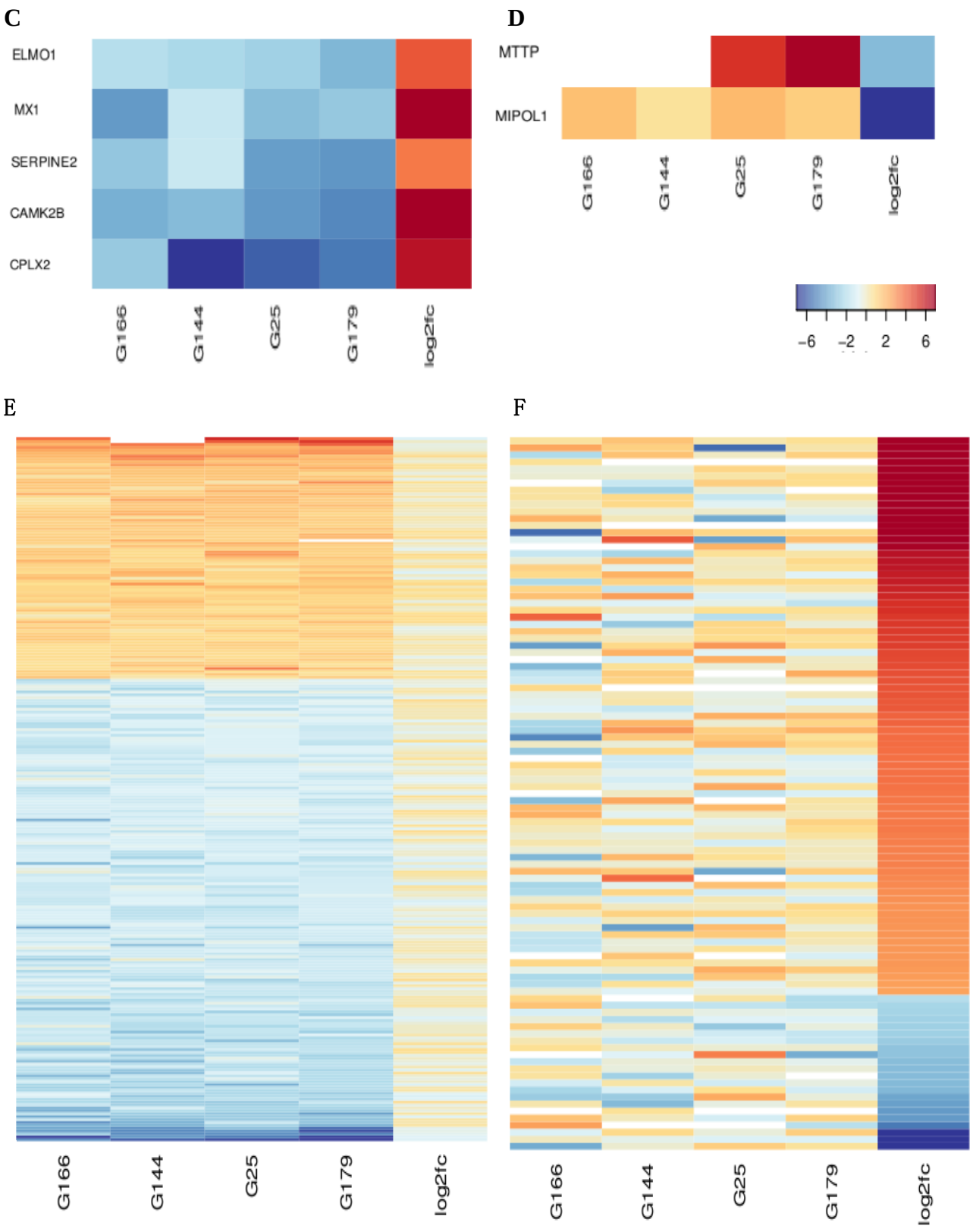


Figure 3-8. Heatmap of DE (adjusted p-value <= 0.05, mean log2(absolute fold-change) >= 1). proteins and mRNAs. The first four columns are protein expression on each GNS line and the last column (log2fc) is mean mRNA expression. Proteins and mRNAs were (A) both over-expressed, (B) both under-expressed, (C) under-expressed and over-expressed, respectively, and (D) over-expressed and under-expressed. (E) Proteins differentially expressed only on the protein level, and (F) genes differentially expressed only on mRNA level.

3.3.10 167 protein groups were DE in secretome

The secreted proteome analysis the reference NS sample (containing CB660, U5 and CB11130) and each GNS line (G166, G144 and G7) resulted in a total of 36 LC-MS/MS runs. The median number of unique peptide identifications for each protein was 5 and the median sequence coverage was 14.8 %. The mean sample correlation coefficient based on protein groups present in all samples was 0.6. 1718 protein groups were quantified across the three sample pairs and 595 of them were DE, and 389 of them were predicted to be secreted based on SignalP, UniProt keywords "Signal", "Secreted" and "Extracellular space" (see Materials and methods) (Table 3-6). Among the 389 protein groups that were predicted to be secreted, 167 were DE, 144 of them were up-regulated and 23 of them were down-regulated. The volcano plot also shows a bias of DE proteins towards over-expression (Figure 3-9), suggesting that GNSs secrete more proteins than NSs.

Table 3-6. Number of quantified protein groups in total cell.

Sample	Quantified	DE	Up	Down
NSall - G166	1931	666	383	283
NSall - G144	1939	673	391	282
NSal - G7	1926	679	398	281
Complete overlap	1718	595	319	276
Complete overlap, predicted secreted	389	167	144	23
Total	2093	731	448	283

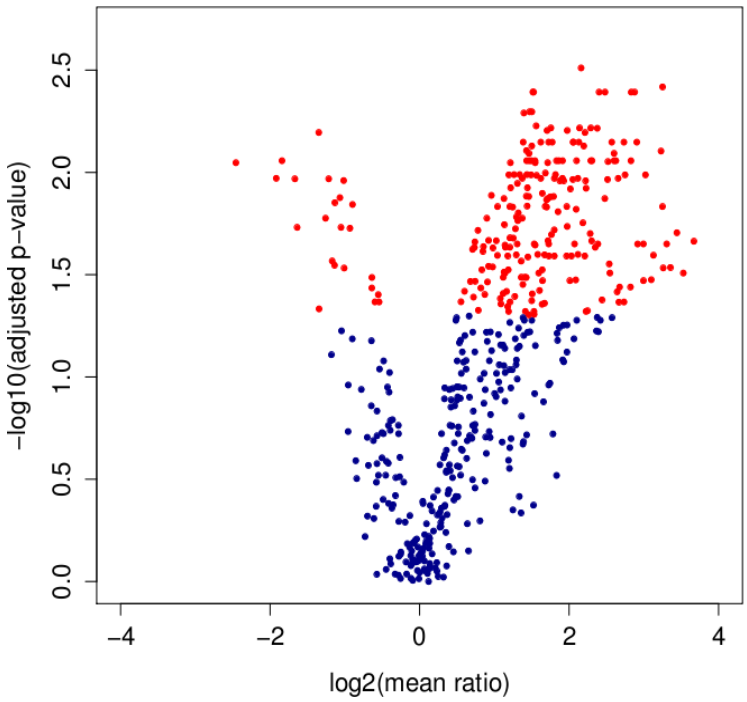


Figure 3-9. Volcano plot of DE secreted proteins.

3.3.11 Total cell- and secretome experiments resulted in disparate sets of DE proteins

The overlap of quantified protein groups between the two experiments is presented in (Figure 3-10 A). 9 % of the secretome protein groups (182) was unique to the secretome analysis. However, if only the DE proteins in both datasets were considered that were predicted to be secreted, the overlap was small (Figure 3-10 B) with only 12 protein groups in common. The overlap between the 182 proteins and 155 DE proteins from the secretome experiment that were predicted to be secreted (Figure 3-10 C) resulted in relatively a poor overlap (28 proteins). The possible reasons for this are that quantification of secreted proteins was not accurate in the total cell experiment, or due to the slightly different cell lines used between the two experiments, or due to the pulsed-SILAC methodology used in the secretome experiment which captures newly synthesized proteins, or the combination of these reasons. The Pearson correlation coefficient of the ratios of 1893 overlapping protein groups between total cell and secretome exepriments was 0.28 (Figure 3-10 D). The ratios of the 12 DE overlapping proteins appear to have a good correlation.

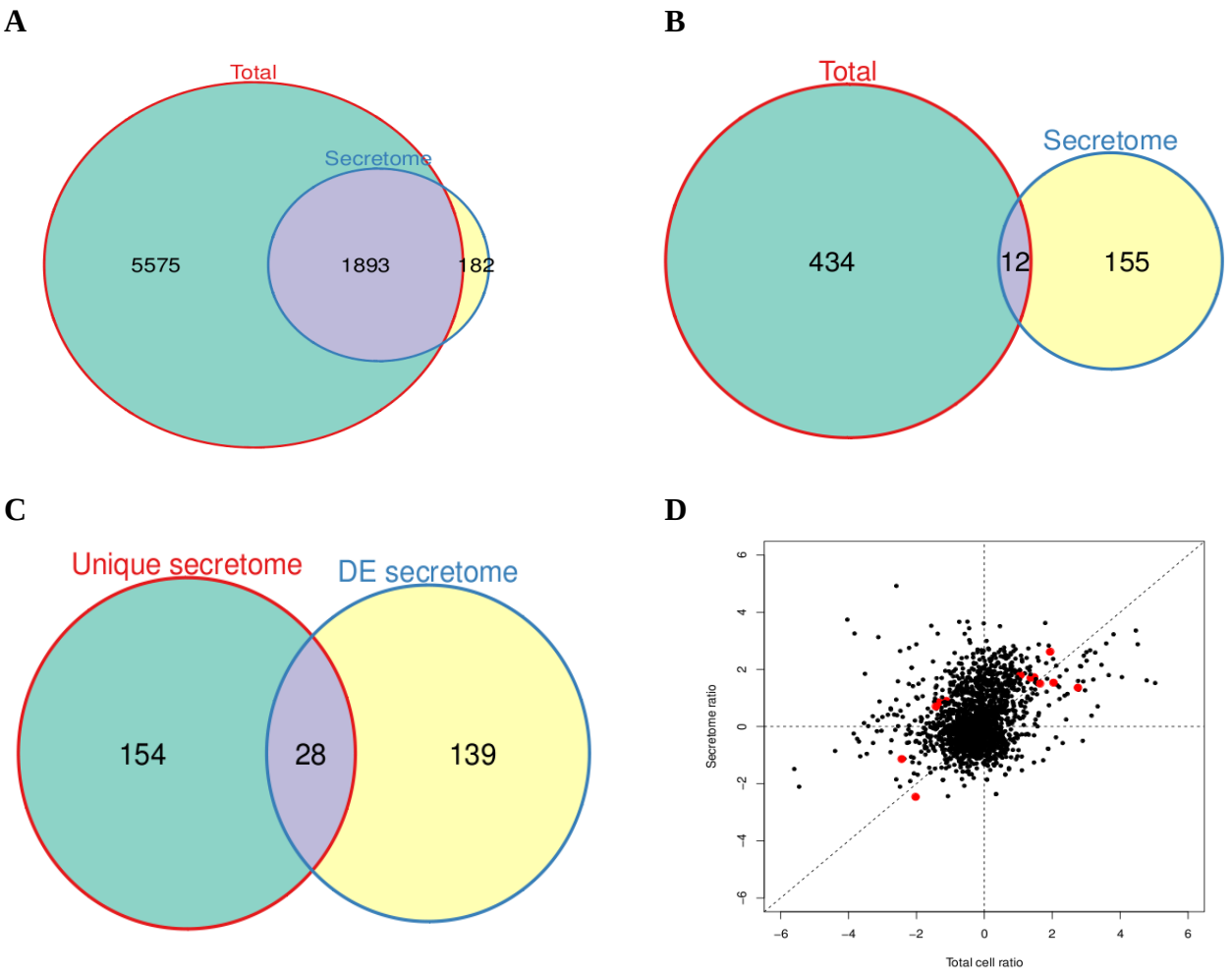


Figure 3-10. Venn diagram showing overlap between two experiments; (A) quantified protein groups and (B) DE total cell protein groups and DE secretome protein groups that were predicted to be secreted. (C) Overlap between 182 proteins unique to secretome experiment and 167 DE secretome proteins predicted to be secreted. (D) Scatter plot of ratios of 1893 overlapping protein groups between total cell and secretome exepriments. Large red dots are 12 DE overlapping protein groups.

3.3.12 Classification of DE secretory proteins revealed diverse functional categories including growth factors and cytokines and many of them did not have prior associations with glioma

64 and 167 DE protein groups that were predicted to be secreted in the total cell experiment and in the secretome experiments, respectively, were functionally categorized using UniProt keywords (Figure 3-11). In addition, their prior associations with glioma and neural stem cells were investigated by PubMed search (Table 3-7). In the secretome experiment the majority of the DE proteins were over-expressed, whereas in the total cell experiment the categories with abundant proteins tended to have more under-expressed ones and more over-expressed proteins were found in categories with a few proteins. The four most abundant terms were the same between the two experiments; Receptor / membrane protein, Disease mutation, Collagen, (Extracellular matrix) and Cell adhesion. The high abundance of receptor / membrane proteins suggests that many of these proteins are shed to the extracellular space. It is also noteworthy and expected that there were 18 proteases in the secretome experiment, while there was only three in the total cell experiment. This proportional discrepancy was observed only in this category. 21 proteins in total were proteases; A2M, APLP2, ADAM10, CTSD, LGMN, CPXM1, DPP7, TIMP1, TIMP2, CTSA, ERAP1, PEBP1, PLAT, CST3, C1S and C1R (over-expressed), SERP, IDE, PCSK9 and NRD1 (under-expressed), many of them are metalloproteinases and cathepsins. ADAM10 is a metalloproteinase, TIMP1 and TIMP2 are metalloproteinase inhibitors, and CTSD and CTSA are cathepsins. Metalloproteinases and cathepsins were also found to be up-regulated in U87 glioma conditioned media in relation to that of other glioma cell lines LN18 and U118 [61]. CST3 (cystatin-C) is an inhibitor of cysteine protease, and has been shown to regulate neural stem cells [1] and glial development [149] and to mediate differentiation of ESs into neural stem cells [150] and into neuronal cell [151]. It has also been implicated in glioma [152, 153], and was over-expressed in our GNSs, suggesting that regulation of cysteine proteases is a factor that can differentiate between NSs and GNSs. Many other proteases do not have any known associations with glioma and/or neural stem cells (Table 3-7). There were in total 10 proteins belonging to growth factor/growth factor binding; MDK, PRNP, CSF1, VEGFA, IGFBP4, NOV, LTBP4 (over-expression), IGFBP3 and IGFBP5 (under-expression), and two cytokines; CSF1 (also a growth factor) and GRN (over-expression). These will be further discussed below. No cadherin was differentially expressed.

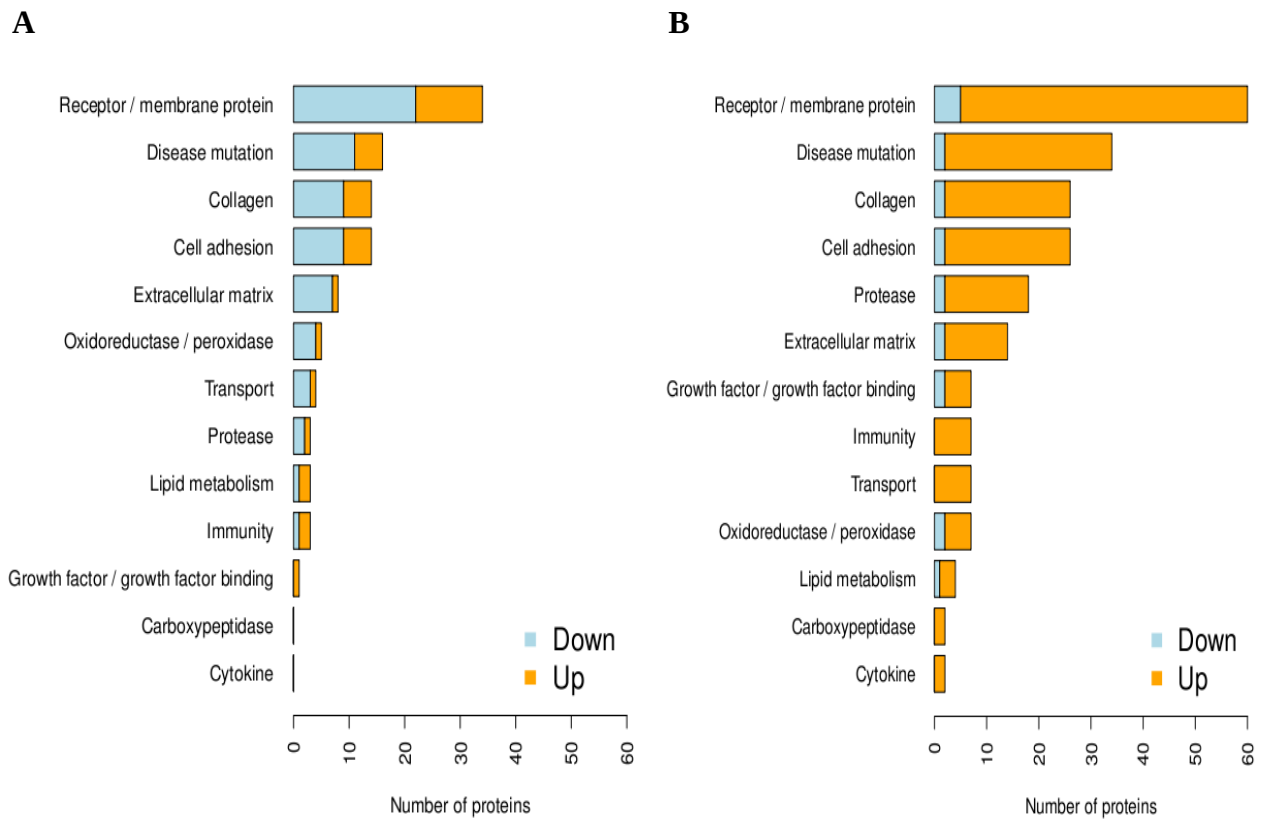


Figure 3-11 UniProtKB functional annotation of DE proteins predicted to be secreted. (A) Total cell experiment. (B) Secretome experiment.

3.3.13 Interaction between DE secreted proteins and receptors identified candidate factors

Next, all known interactions between our DE secretory proteins (from both total cell and secretome experiments) and identified receptors were retrieved from MetaCore (Table 3-8). We hypothesized that if either a secreted protein and/or its plasma membrane receptor is DE, they may have a functional significance to the cell. We further narrow the list down by focusing only on those that belonged to UniProt keywords “Growth factor”, “Growth factor binding”, “Cytokine” and “Transport”. Every interaction was manually checked if they were experimentally validated via corresponding literatures, which resulted in APOE, GRN, IGFBP3, IGFBP4, IGFBP5, MDK, NOV, PRNP and VEGFA as prime candidates for further studies and discussed below. The direct interactions between DE secretome and GO.BP neuron differentiation was also investigated, however, no new interaction was identified (data now shown).

Table 3-7 Log2(mean ratio), log10(intensity) and prior associations with glioma, neural stem cells and cancer of 167 DE proteins from secretome experiment that were predicted to be secreted. 'Association with glioma', 'Association with neural stem cell' and 'Association with other cancer' columns indicate PubMed IDs for pertinent references. 'many' is put when > 5 references were found.

Chromosome	Gene name	log2(ratio)	log10(intensity)	GBM survival Padj	Association with glioma	Association with neural stem cell	Association with cancer
15	ADAM10	1.13	7	0.032	many	21878106, 22466506, 23262104	many
1	AGT	2.67	9.45	0.813	many	-	many
15	ANXA2	1.64	8.56	0.54	16640645, 19351187, 21033036, 23082878, 23195957	19351187	many
19	APLP1	1.67	8.34	0.073	-	20049903	many
11	APLP2	1.09	8.5	NA	7616233	11987239, 18717733, 20049903	many
1	APOA1BP	1.04	7	0.533	-	-	18277965
19	APOE	1.93	8.17	0.627	many	21352230	many
6	ARG1	0.83	7.16	0.631	3016193	-	many
22	ARSA	1.51	6.27	0.06	15494095	16871228, 17080190, 18667806	many
8	ASAH1	1.28	7.63	0.079	15088070	-	many
20	ATRN	1.71	7.84	0.4	17085642	-	many
15	B2M	1.44	8.27	0.347	10378372	20716364, 23317542	many
19	BCAM	2.57	7.63	0.327	-	-	many
3	BCHE	2.05	7.68	0.18	20641589	17459421	many
3	BTD	1.21	7.77	0.651	11792359	-	many
12	C12orf10	-0.9	7.88	0.241	-	-	20811708
17	C1QL1	2.16	7.21	0.144	-	-	-
12	C1R	2.12	8.12	NA	-	-	15988036
12	C1S	2.1	7.95	0.714	-	-	15988036
8	CA2	0.98	6.8	0.437	many	-	many
11	CADM1	0.71	8.07	0.873	-	20871982, 21672091	many
5	CANX	1.39	7.33	0.577	17545629, 17878160	-	many
HSCHR6_MHC_DBB	CDSN	0.91	6.75	0.807	-	-	-
11	CHID1	1.43	7.83	0.185	-	-	-
13	CLN5	1.14	7.13	0.014	-	21235444	16955048, 19345705
1	CLSTN1	1.47	9.61	0.35	-	-	many
12	CLSTN3	1.8	8.45	0.091	-	-	18489135
1	COL11A1	1.68	7.97	0.842	19351187, 22537279	-	many
1	COPA	-0.54	8.34	0.305	21365010	-	many
20	CPXM1	1.5	7.21	0.009	-	-	-
5	CRHBP	-1.64	8.28	0.452	-	-	-
1	CSF1	2.31	8.03	0.622	2038877	-	many
20	CST3	1.88	9.06	0.716	12483523, 16153465	21417836, 1659563, 11144350	some
20	CTSA	1.72	8.08	0.647	-	-	many
11	CTSD	1.36	8.58	0.068	many	7020877, 18346466, 23365100	many
3	DAG1	1.26	8.39	0.092	many	many	many
11	DCHS1	0.79	6.57	NA	-	-	-
HSCHR6_MHC_DBB	DDR1	-0.6	6.82	NA	11126911, 15750623, 16234985, 16652150, 22879068	22008533	many
11	DKK3	1.43	9.05	0.341	18443132, 19847810, 18033687	-	many
3	DNAJB11	1.19	7.61	0.279	-	-	20418907
7	DNAJB9	1.14	6.67	0.401	-	-	many
19	DNASE2	2.05	7.81	0.266	-	-	-
9	DPP7	1.53	7.25	0.079	-	-	20817072
18	DSC1	0.92	7.67	0.669	16521483	-	many
18	DSG1	1.03	7.96	0.342	-	-	many
7	EPDR1	1.53	7.61	0.126	-	-	18374504, 22252855
5	ERAP1	1.79	7.54	NA	-	-	many
12	ERP29	1.62	7.39	0.982	21667264	-	many
7	FAM3C	1.44	7.83	0.133	-	-	many
4	FAT1	1.19	8.56	0.609	22986533	16865240	many
11	FAT3	1.47	7.62	0.715	-	21903076	many
22	FBLN1	1.25	8.95	0.964	-	-	many
2	FBLN7	2.28	6.43	NA	-	-	-
20	FLRT3	2.12	7.89	0.925	-	-	17091452, 17450523,

Table 3-7 (continued 1)

Chromosome	Gene name	log2(ratio)	log10(intensity)	GBM survival Padj	Association with glioma	Association with neural stem cell	Association with cancer
13	FREM2	2.62	9	0.233	22538188	-	22538188, 20629094, 16087869
6	FUCA2	1.37	8.03	0.047	-	-	19666478
17	GAA	1.73	6.26	0.007	15170390	-	many
18	GALNT1	0.85	6.9	0.033	-	-	many
10	GFRA1	-1.35	7.28	0.836	many	many	many
8	GGH	1.77	9.01	0.971	-	-	many
X	GLA	1.51	6.9	0.737	22740420, 2540282	-	many
16	GLG1	0.9	7.79	0.785	-	-	many
12	GNS	1.18	7.77	0.357	-	-	12932876
13	GPC6	1.08	6.92	0.855	-	-	many
17	GRN	1.43	7.69	0.111	10728698	17179653	many
15	HEXA	1.29	8.03	0.468	-	-	many
5	HEXB	1.41	8.46	0.611	-	19591217, 22367451, 23383290	many
13	HMGB1	-0.64	8.34	NA	many	21527633, 23137544,	many
4	HMGB2	-1.13	8.41	0.028	19240692	-	many
22	HMOX1	-1.26	7.42	0.893	many	many	many
21	HSPA13	1.22	7.82	0.357	-	-	-
HG299_PATCH	HYOU1	1.12	7.92	0.252	-	-	many
X	IDS	1.56	8.12	0.997	-	-	many
6	IGF2R	1.55	7.82	0.511	16582634, 18562769	16037066, 21047779, 23525019	many
7	IGFBP3	-1.84	8.73	0.603	many	21136151	many
17	IGFBP4	2.73	7.49	0.021	7683520, 12937144, 16586492	16809006	many
2	IGFBP5	-1.05	8.55	0.629	many	19772911, 20604680	many
3	IL1RAP	1.31	7.64	0.666	-	-	many
20	JAG1	1.34	8.19	0.522	2229617	many	many
11	JAM3	1.32	7.54	0.634	19795504	22114908	many
16	KARS	-0.64	7.89	NA	-	-	many
X	L1CAM	1.36	7.32	0.027	many	many	many
20	LAMA5	0.92	8.61	0.028	-	-	18506748
3	LAMB2	1.21	8.17	0.733	-	-	many
19	LDLR	1.21	8.14	0.945	17052361, 16679074, 22586629, 22059152	10375696	many
1	LEPRE1	1.18	6.5	0.439	-	-	22724020, 22955849
22	LGALS1	-1.67	8.49	0.745	many	many	many
14	LGALS3	1.8	7.71	0.854	many	21587270, 21693585	many
17	LGALS3BP	1.44	8.64	0.003	-	-	many
14	LGMN	1.46	7.73	0.968	-	-	16702559, 21237226
5	LMAN2	1.3	8.12	0.638	-	-	23161554
5	LOX	-1.92	8.48	0.923	17931358	-	many
2	LOXL3	2.1	8.17	0.381	-	-	many
1	LPHN2	-2.46	7.65	0.424	-	-	many
4	LPHN3	1.24	8.01	0.197	-	-	20668451, 3317273
1	LRP8	1.21	7.12	0.433	-	22407947	15615770, 22614235, 23142051
11	LTBP3	0.98	7.68	0.025	-	-	many
19	LTBP4	1.97	7.95	0.078	-	-	many
4	MANBA	1.77	7.87	0.654	-	-	17899454
22	MAPK1	-0.55	8.37	0.024	many	many	many
8	MATN2	1.45	8.03	0.222	16401863	-	many
11	MCAM	1.42	7.83	0.506	8616875	-	many
16	METRN	-1.07	7.03	0.094	-	15085178, 19130216, 22044868	many
15	MFAP1	-1.34	7.54	0.813	-	-	19377877
15	MFGE8	1.85	8.01	0.052	11085522	-	many
22	MIF	0.55	8.43	0.199	many	many	many
1	NFASC	1.58	7.76	0.94	-	-	8812479, 8921253
1	NID1	1.16	7.92	0.895	-	17569787, 21283688	many
14	NID2	2.6	8.12	0.837	21349332	-	many
16	NOMO2	1.54	8.12	0.008	-	-	-
8	NOV	2.07	6.67	0.959	18784988, 18004727, 17340618, 15213231, 11577170	19286457	many

Table 3-7 (continued 2)

Chromosome	Gene name	log2(ratio)	log10(intensity)	GBM survival Padj	Association with glioma	Association with neural stem cell	Association with cancer
1	NRD1	-1.01	7.68	0.231	-	-	21769958, 22653443
2	NRXN1	-1.14	7.89	0.979	-	23536886	20113834, 2323628, 23474816
19	NUCB1	0.96	8.84	0.023	-	-	-
11	NUCB2	1.37	7.75	0.942	-	-	12087473, 19351608, 21988594
11	OAF	1.43	6.95	0.018	-	-	-
12	OS9	1.98	7.93	0.744	-	-	some
5	PAM	0.82	8.27	0.385	9778036	-	many
13	PCDH17	1.32	7.96	0.564	-	-	many
13	PCDH9	1.68	7.42	0.028	2230079, 18828157	-	22150124
1	PCSK9	-1.17	7	0.3	-	-	many
12	PEBP1	1.82	8.93	0.943	22292035	17146836	many
8	PLAT	1.86	8.29	0.955	many	many	many
7	PLOD3	1.7	8.79	0.491	-	-	20687567, 22559327, 16322899
22	PLXNB2	1.2	7.27	0.392	-	-	-
15	PPIB	1.15	8.48	0.204	-	-	many
19	PRKCSH	2.06	7.51	0.001	-	-	many
20	PRNP	1.39	7.61	0.938	15274317, 17390034	many	many
3	PROS1	2.22	8.09	0.467	-	-	some
1	PTGFRN	1.55	7.97	0.925	-	-	many
1	PTPRF	-1.02	7.31	NA	-	-	many
3	PTPRG	0.78	7.82	0.523	9795134	21969550	many
6	PTPRK	0.68	7.75	0.026	-	20212451	many
19	PTPRS	1.04	8.06	0.053	-	16784531	many
7	PTPRZ1	2.35	9.44	0.164	many	18308476	many
3	PTX3	1.62	8.41	0.676	2148922	10457011, 14732407, 15135891, 17196845, 18929554	many
19	PVR	1.73	7.6	0.264	many	22529845	many
11	PVRL1	0.74	7.28	0.008	-	-	many
19	PVRL2	1.56	7.14	0.163	-	-	15254712, 16391793, 17696193, 22122800, 22997493
2	PXDN	0.73	8.35	NA	20063114	-	17330099, 20667089
1	QSOX1	1.83	8.89	0.023	-	-	many
11	RCN1	1.91	8.05	0.802	-	-	some
6	RNASET2	1.49	7.06	0.273	-	-	many
3	ROBO1	0.72	7.86	0.898	16636676, 17968499, 20008733, 21113198	22244746, 22433866	many
1	SDF4	1.5	8.33	0.028	-	-	16215274, 21949389
5	SEMA5A	1.27	7.25	0.908	20696765, 21706053	-	many
11	SERPING1	-0.93	7.89	0.608	-	-	many
17	SEZ6	2.13	7.37	0.652	-	-	16863507
11	SIAE	1.53	6.56	0.635	-	-	21803834
15	SORD	1.3	6.33	0.35	2144507	-	many
11	SORL1	1.5	7.29	0.291	1763461	-	22774576
1	SORT1	1.47	6.94	0.863	-	-	16540638
5	SPARC	0.87	9.74	0.889	many	many	many
X	SRPX	0.85	8.33	0.265	20964819	-	many
3	TFRC	1.71	6.96	0.095	19386095	20373404	many
1	THBS3	0.6	8.33	0.906	-	-	18452548, 17022822
X	TIMP1	1.7	9.33	0.778	22995409, 21327941, 20530493, 20332466	-	many
17	TIMP2	1.75	8.82	0.764	many	-	many
3	TIMP4	3.31	6.71	0.976	19062176	-	many
6	TNFRSF21	1.12	6.39	0.906	22802048	-	many
9	TXN	2.29	8.46	0.802	many	11565801	many
6	TXNDC5	2.02	8	0.31	-	-	many
6	ULBP2	1.53	7.78	NA	16891318, 19089914	-	many
5	VCAN	1.91	8.48	0.544	many	many	many
6	VEGFA	1.76	7.17	0.84	many	many	many
9	VLDLR	0.74	7.99	0.837	-	12586425, 16190894	many
1	YARS	-1.22	8.35	0.671	-	-	15577315

Table 3-8 DE secretome-receptor interactions retrieved from MetaCore. 'Symbol.from' is secretory proteins and 'Symbol.to' is recepors.

Symbol.from	Mechanism	symbol.to	Effect	References
A2M	Binding	LRP1	Activation	9349534;10652313;10815129;12194978;15053742;15910735;16149055; 16725309;16982616;17288987
ADAM10	Cleavage	HLA-A	Unspecified	17150042
ADAM10	Cleavage	PVRL1	Activation	20501653
ADAM10	Cleavage	NGFR	Unspecified	12843241;15701642
ADAM10	Cleavage	LRP4	Unspecified	20383322
ADAM10	Cleavage	PTPRK	Activation	16648485
ADAM10	Cleavage	PRNP	Unspecified	11477090;15975064;16263114;16824663;18951988;19564338
ADAM10	Cleavage	F11R	Activation	19258599
ADAM10	Cleavage	CLSTN1	Unspecified	19864413
ADAM10	Cleavage	APLP2	Activation	16279945
ADAM10	Cleavage	L1CAM	Activation	12475894;15814625;16199880;18064447;18289051;18762209; 18951988;19260824;20594269;21195665;21346732;22586143
ADAM10	Cleavage	CLSTN3	Unspecified	19864413
ADAM10	Cleavage	LRP8	Unspecified	17913923
ADAM10	Cleavage	EPHB2	Unspecified	17428795;18951988
ADAM10	Cleavage	SORT1	Activation	12419319;21730062
ADAM10	Cleavage	CDH2	Inhibition	21123580
APLP1	Binding	HMOX1	Inhibition	11144356
APLP2	Binding	HMOX1	Inhibition	11144356
APOE	Binding	LRP1	Activation	2266137;7615159;9124278;10815129;11421580;15863833;16401069; 17288987;17341585
APOE	Binding	LDLR	Activation	10075730;10357834;10889196;11067868;11890675;12036962;16630895; 16725309;17234631;17923100
APOE	Binding	VLDLR	Activation	11353330;11374859;11421580;11743951;11839845;12167620;12870663; 12966036;15319263;15863833;19116273
APOE	Binding	LRP8	Activation	8626535;11421580;12167620;12950167
APOE	Binding	SORL1	Activation	11557679;17326667
ASAH1	nscription regula	NR0B1	Inhibition	22927646
B2M	Binding	HLA-A	Activation	9162021;9427624;9605335;9774416;10064069;10428963; 10631933;11160214;16181333
B2M	Binding	FCGRT	Activation	4758346;8702683;10933786;11336709;12006623;12023961;12144784; 12162790;12242328;16002696;20936779
BCAM	Binding	ITGA4	Activation	17158232
C1S	Cleavage	SERPING1	Unspecified	17709141
C1S	Cleavage	IGFBP5	Inhibition	18930415
CADM1	Binding	JAM3	Activation	18055550
CANX	Binding	L1CAM	Inhibition	22222883
CANX	Binding	HLA-A	Activation	8006598;8943049;9551918;12788224;15494401;16181333; 17708944;18420789
COL18A1	Binding	ITGA5	Inhibition	12682293;14973128;17597104;18006826;19542224
COL18A1	Binding	GPC1	Activation	11336704;20936779
COL3A1	Binding	ITGA2	Activation	16043429
COL3A1	Binding	GPR56	Activation	21768377;22238662
COL3A1	Binding	ITGA1	Activation	16043429
CTSD	Cleavage	COL18A1	Unspecified	11119712
CTSD	Cleavage	CTSD	Activation	1812719;10508159
CTSD	Cleavage	SRI	Unspecified	20627866
CTSD	Cleavage	IGFBP3	Inhibition	9275067
CTSD	Cleavage	IGF2R	Inhibition	15258139;15518240;20541250;20936779
CTSD	Cleavage	IGFBP5	Unspecified	9275067
CTSD	Cleavage	CST3	Inhibition	2013314;22898924
CTSD	Cleavage	IGFBP4	Inhibition	9275067
DDR1	Phosphorylation	DDR1	Activation	10681566;16440311
DDR1	Phosphorylation	PTPN11	Unspecified	16337946;16611743;16626936;22057045
DNAJC3	Binding	VCAM1	Activation	16923392
EGFR	Phosphorylation	PTPRJ	Unspecified	
EGFR	Phosphorylation	TLR3	Activation	22810896
EGFR	Phosphorylation	FAS	Activation	12586732;15917250;16772302;17258167

Table 3-8 (continued 1)

Symbol.from	Mechanism	symbol.to	Effect	References
EGFR	Binding	GRB2	Activation	1322798;7506413;7510700;7518560;7527043;7798267;7925272;7929151; 8305738;8316835;8386805;8479540;8479541;8524223;8577724;8626525; 8626530;8647858;8662998;8810325;8887653;8940013;9050991;9363897; 9685397;9715408;9765228;9886492;10026169;10085134;10090597;10490623; 10635327;11208164;11287756;11297548;11412040;11726515;11823423;11853876; 11960376;12577067;14498832;14679214;14743216;15194809;15352158;15635092; 15657067;15782189;15784896;15982853;16055672;16099987;16273093;16477079; 16729043;16799092;16889899;17242169;17372273;17548515;17671194;17715395; 17971399;18174162;18358509;18562239;18721752;18793634;19278030;19289468; 19531065;19836242;20067773;20124286;20462955;20473329;20945942;21185312; 21258366;21258655;21278786;21278788;21356361;21596750;21706016;23027125
EGFR	Phosphorylation	EPS8	Activation	7532203;7532293;8404850;11099046;12127568
EGFR	Binding	GRB10	Activation	7731717;9006901;9506989;15901248;18721752
EGFR	Phosphorylation	EGFR	Activation	2543678;2552117;3138233;3494473;9335547;11894079;15708576; 16946702;17139251
EGFR	Phosphorylation	GAB1	Activation	8596638;9890893;10648629;10734310;11432805;11606067;11940581;12370245; 12628344;14668796;15231819;16185843;18046719;19359598;19651513; 21214269;21278788;23027125
EGFR	Binding	NCK1	Activation	1333047;8561895;8662998;9362449;10026169;11252954; 16273093;18721752
EPHB2	Binding	GRB2	Activation	16298995
EPHB2	Phosphorylation	L1CAM	Activation	9089215
FBLN1	Binding	ADAMTS1	Activation	16061471
FBN1	Binding	ITGA5	Activation	12807887;17158881
GALNT1	Glycosylation	DAG1	Unspecified	21937429
GALNT1	Glycosylation	COL18A1	Unspecified	21937429
GFRA1	Binding	NCAM1	Activation	12837245;12953054;17322291
GPC1	Binding	VEGFA	Activation	10196157
GRN	Binding	ADAMTS7	Inhibition	20506400
GRN	Binding	ADAMTS12	Inhibition	20506400
HMGB1	Binding	NR3C1	Activation	9671457;12006575
HMGB1	Binding	NR3C1	Activation	9033409;15808513
HSPG2	Binding	ITGA2	Activation	15240572;16882656;17197432
HYOU1	Transport	VEGFA	Unspecified	11358846;11435455;11435456;11771177;12445237
IGF2BP1	Binding	CD44	Activation	16541107;17101699
IGFBP3	Binding	RXRA	Inhibition	10874028;14715249;15935690;17644060;19324019
IGFBP3	Binding	LRP1	Activation	9252371;10037769;14597676
IGFBP4	Binding	LRP1	Activation	10037769
IGFBP4	Binding	LRP6	Inhibition	18528331
IGFBP5	Binding	PLAT	Activation	16505491
IGFBP5	Binding	LRP1	Activation	10037769
ITGAV	Binding	PLAUR	Activation	8548872;12297505
ITGB1	Binding	EGFR	Activation	9822606;18247373;21217148
JAM3	Binding	JAM3	Activation	11590146;11739175
L1CAM	Binding	ITGB1	Activation	12077189;16330023
L1CAM	Binding	ITGA3	Activation	16330023
L1CAM	Binding	CNTN1	Activation	7595520;18490510
L1CAM	Binding	PTPRZ1	Activation	7559574
L1CAM	Binding	L1CAM	Activation	12084815
LAMA5	Binding	ITGA6	Activation	12519075;17383963
LAMA5	Binding	ITGA3	Activation	12519075;17383963
LAMA5	Binding	ITGA6	Activation	12297042;12441134;15761669;16581764
LAMA5	Binding	BCAM	Activation	9642222;11133776;11319237;11507772;12244066;12921739; 17383963;17638854;20936779
LGALS3	Binding	ITGA3	Activation	15181153;19755493
LGALS3	Binding	EGFR	Activation	17889671;19940114;21258405
LGALS3	Binding	L1CAM	Activation	20124415
LGALS3	Binding	LGALS3BP	Activation	1917996;8390986;8813152;9501082;11146440;14758079;16189514; 16393961;16518858;19060903;21031433
LRP1	Binding	TIMP2	Inhibition	15489233
LRP1	Binding	A2M	Activation	1423505;2430968;2451858;7693397;10652313;12194978

Table 3-8 (continued 2)

Symbol.from	Mechanism	symbol.to	Effect	References
LSAMP	Binding	LSAMP	Activation	11984841
LTBP3	Binding	ITGAV	Activation	12358597
LTBP3	Binding	ITGAV	Activation	12358597
MDK	Binding	ITGA6	Activation	15466886
MDK	Binding	ITGA4	Activation	15466886
MDK	Binding	PTPRZ1	Inhibition	10212223;10706604;11340082;12573468;20936779
MDK	Binding	LRP1	Unspecified	10772929;12215536;12573468;17971413;20936779;21688265
MFGE8	Binding	ITGAV	Activation	14697347;16529932;17299048;17591687;21901532
MIF	Binding	CXCR4	Activation	17435771;18818421;19066630;20807568;20861157
NCAM1	Binding	NCAM1	Activation	14527396
NCAM1	Binding	PTPRZ1	Activation	7528221;7559574;9049255
NID1	Binding	PLXDC1	Activation	16574105
NOV	Binding	ITGA5	Activation	12695522
NOV	Binding	ITGAV	Activation	15611078
NRD1	Binding	ADAM17	Activation	16923819;19005493;20184396
PCSK9	Cleavage	LRP8	Inhibition	18039658;18675252
PCSK9	Cleavage	LDLR	Inhibition	17012247;17080197;17435765;17449864;17452316;18039658;18250299;18354137;18675252;18753623;19196236;21692990;22081141
PCSK9	Cleavage	VLDLR	Inhibition	18039658;18675252
PDIA3	anscription regulati	A2M	Unspecified	19995546
PDIA3	Covalent modificatio	ITGB5	Unspecified	17170699;19887585
PDIA3	Covalent modificatio	ADAM17	Unspecified	17170699;19887585
PDIA3	Covalent modificatio	ITGB1	Unspecified	17170699;19887585
PDIA3	Covalent modificatio	ADAM10	Unspecified	17170699;19887585
PDIA3	Binding	PRNP	Inhibition	15772339;19798432
PDIA3	Binding	CANX	Activation	8974399;9497314;9545232;10436013;11160214;12052826;14988724;15236594;16181333;16467570;19054761;20936779
PDIA3	Binding	HLA-A	Activation	9545232;9637923;9637924;15494401;16467570
PLAT	Binding	EGFR	Activation	17456763
PLAT	Binding	LRP1B	Activation	11384978;16725309
PLAT	Binding	LRP1	Activation	1502153;1502154;10632583;16303771;17170123;20936779
PLAT	Cleavage	IGFBP3	Unspecified	22778398
PPIB	Binding	BSG	Activation	11688976
PPIC	Binding	BSG	Activation	17483319;17700972
PRNP	Binding	NGFR	Activation	11489911;14625887
PRNP	Binding	NCAM1	Activation	11743735;14625887;15146195;15851519;17051207;19209230;19798432
PTPRF	Dephosphorylation	EPHB2	Inhibition	19047466
PTPRF	Dephosphorylation	PTPRF	Unspecified	11158333
PTPRF	Dephosphorylation	EGFR	Unspecified	1599438
PTPRK	Dephosphorylation	EGFR	Inhibition	15899872;16263724
PTPRS	Dephosphorylation	CDH2	Activation	17060446
PTPRS	Dephosphorylation	EGFR	Inhibition	10435588;10749673
PVR	Binding	PVRL3	Activation	12456712;12759359;16189514;16216929;16904340;17352739;21880730
ROBO1	Binding	ROBO2	Activation	12504588;16226035
SERPINE2	Binding	LRP1	Unspecified	12871303;17409116
SERPINE2	Binding	PLAT	Inhibition	15128599
SERPING1	Binding	C1R	Inhibition	3458172;6282262;11044372;20936779
SERPING1	Binding	C1S	Inhibition	3458172;3756141;6604523;9882449;11044372
SORL1	Binding	PLAUR	Activation	14764453
SORT1	Binding	NGFR	Activation	14985746;14985763;15626491
TIMP1	Binding	ADAM10	Inhibition	10818225
TIMP2	Binding	ADAM33	Inhibition	14676211;15949939
TIMP2	Binding	ITGA3	Activation	12887919;12968163
TIMP4	Binding	ADAM33	Inhibition	14676211;15949939
TIMP4	Binding	ADAM17	Inhibition	15713681
TNC	Binding	ITGA7	Activation	14715956
TNC	Binding	EGFR	Activation	11470832;16632194;17311283
TNC	Binding	ITGA9	Activation	7523411;8798654;10209034;16632194
TNC	Binding	PTPRZ1	Activation	7512960;7514167;7559574;16632194
TNC	Binding	ITGA8	Activation	7559467;9548928;16632194
TNC	Binding	ITGA2	Activation	7693733;16632194

Table 3-8 (continued 3)

Symbol.from	Mechanism	symbol.to	Effect	References
TXNDC5	Covalent modificatio	LAMB2	Unspecified	19887585
TXNDC5	Covalent modificatio	LDLR	Unspecified	19887585
VCAN	Binding	EGFR	Activation	16648628
VCAN	Binding	ITGB1	Activation	11805102;14978219;15126624;16045811;20936779
VCAN	Binding	CD44	Activation	10950950;11821431
VEGFA	Binding	ITGA9	Activation	17363377

The major findings from our proteomics data are surmmarized in Figure 3-12. We selected several candidates for further studies from the different analyses (Figure 3-13) and these will be discussed in the next sections.

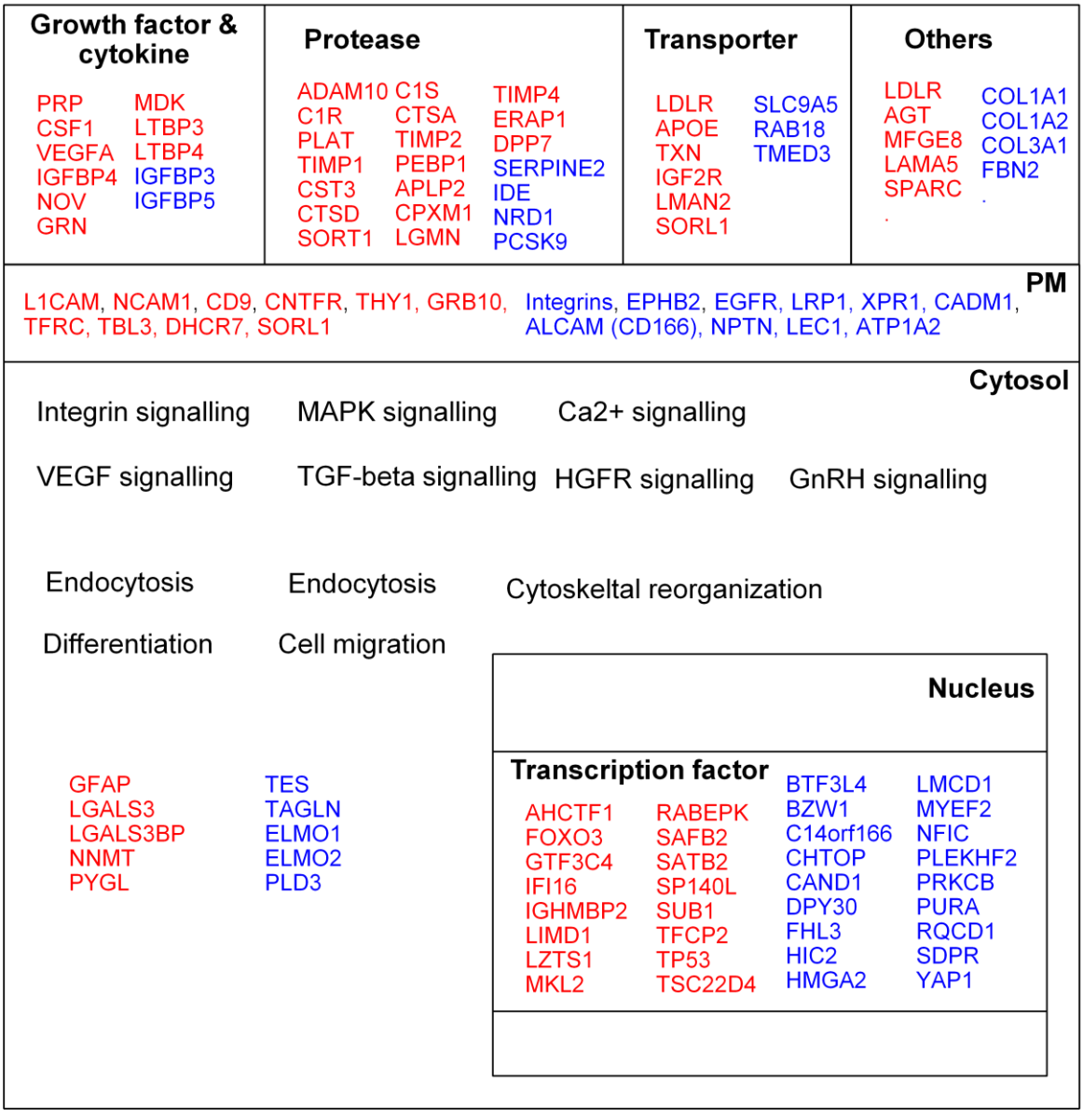


Figure 3-12 Picture summarizing major findings from the proteomic study.

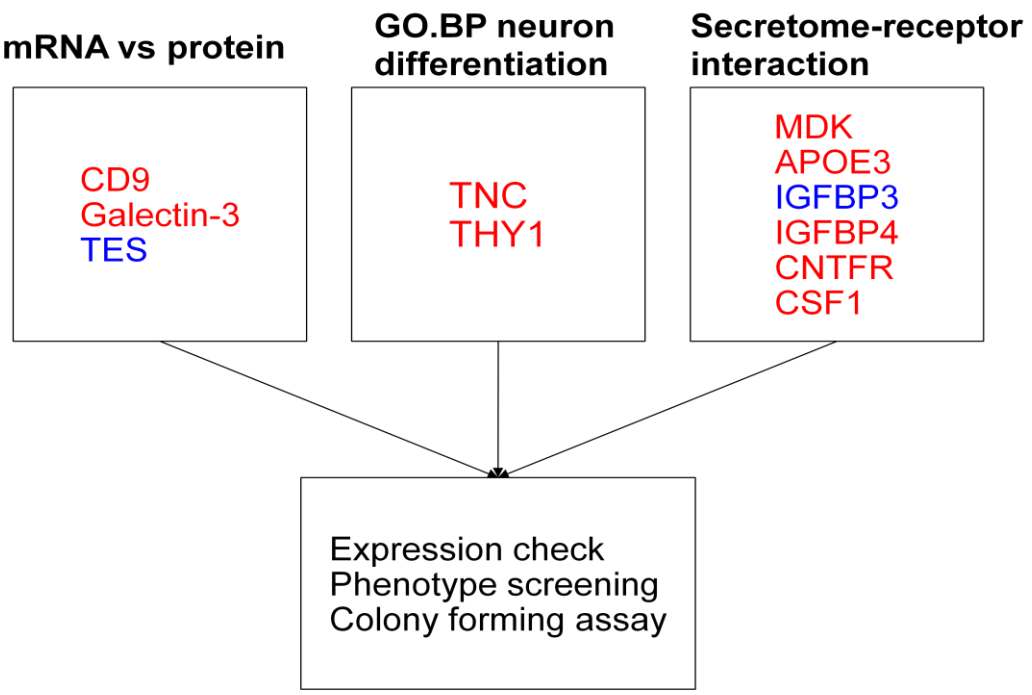
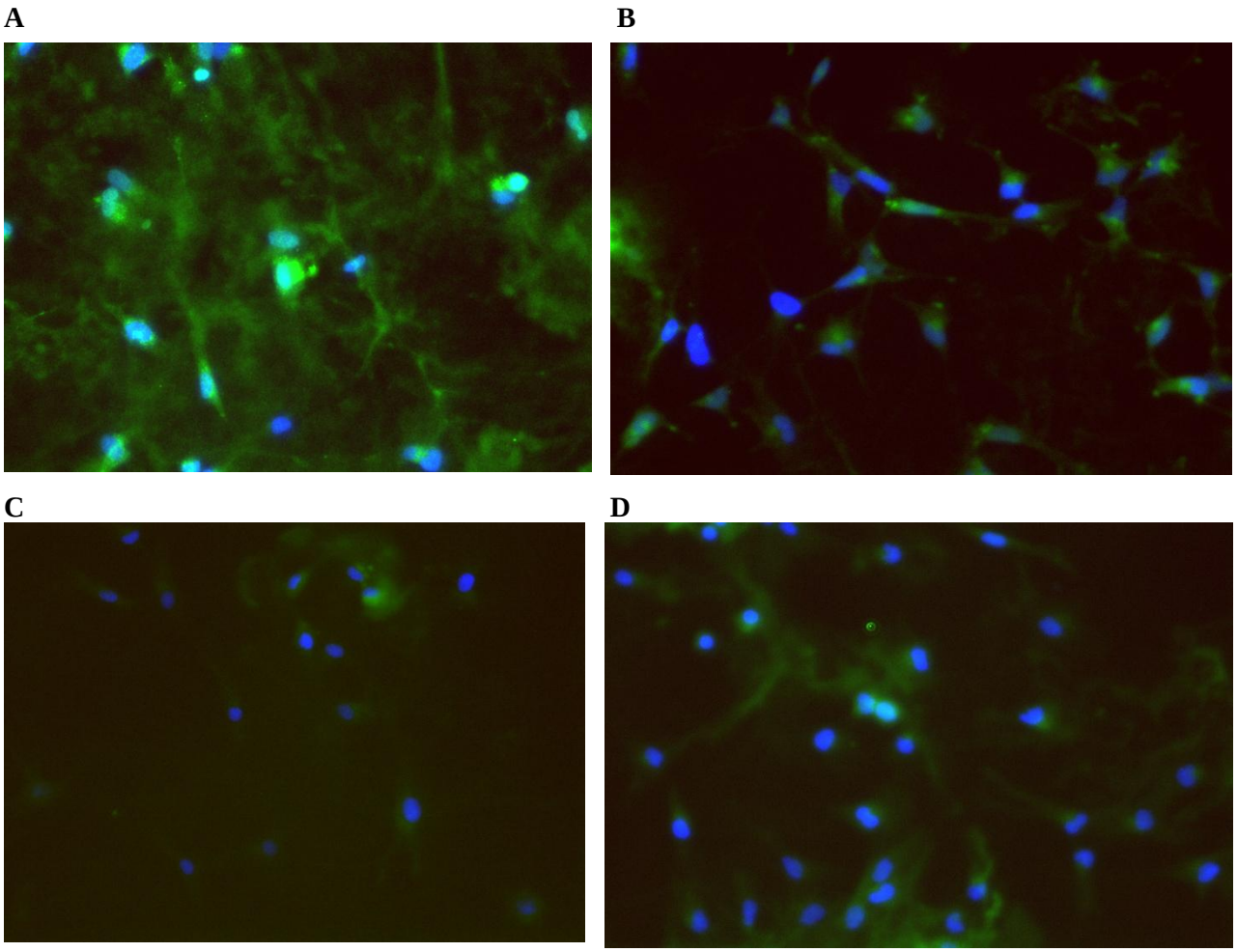


Figure 3-13 Diagram showing proteins to be further studied experimentally.

3.3.14 TNC and LGALS3 were expressed in both NSs and GNSs, whereas THY1 and CD9 were expressed only in GNSs and could be GNS-specific markers

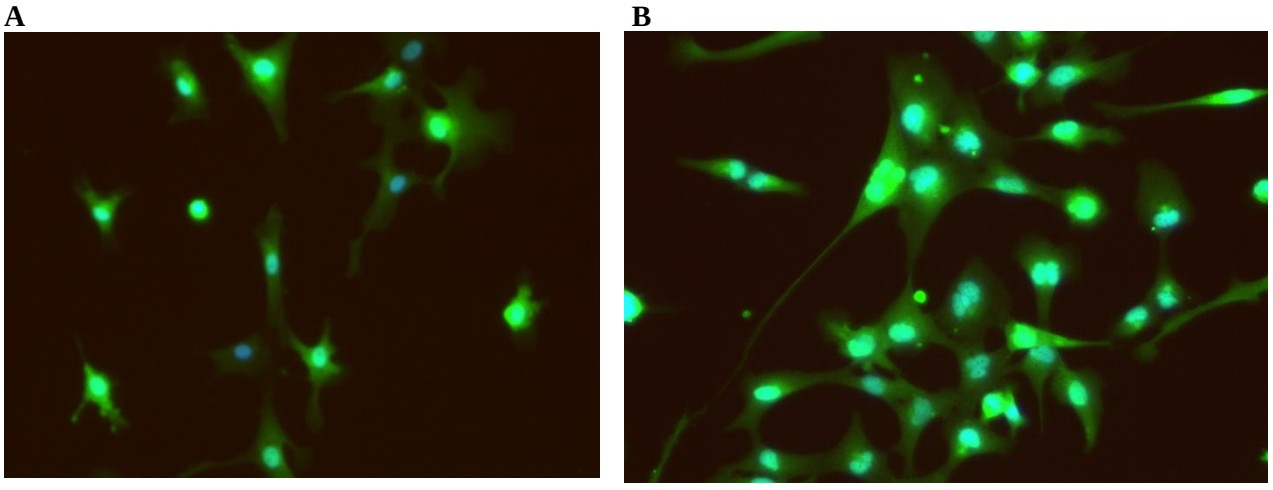
To validate our proteomics data and to identify potential surface markers that could distinguish between GNSs and NSs, we performed immunocytochemistry for TNC, LGALS3, THY1, CD9 and TES (Figure 3-14 , Figure 3-15, Figure 3-16, Figure 3-18, respectively). For this experiment, U7 and CB660 lines were used for the NSs and G7 and G166 lines were used for the GNSs, since the other cell lines were not available. TNC is an ECM protein implicated in guidance of migrating axons during development and in neuronal regeneration. It is absent in normal adult brains but re-expressed upon injury or neoplasia [109, 154]. Our proteomics data indicated that TNC was over-expressed in GNSs and belonged to the GO.BP neuron differentiation. However, TNC is known to be expressed also in foetal NSs. Cultured mouse neural stem/progenitor cells express all Tnc receptors [154]. Indeed, it was difficult to see an appreciable difference between the NSs and GNSs in our immunostaining.

LGALS3 was over-expressed in GNSs both on the mRNA and protein levels. It was shown to be an adult astrocyte stem cell marker [80] likely functional in cell migration/motility [81]. Our immunostaining showed that LGALS3 was expressed in both NSs (CB660) and GNSs (G166) to a similar degree. This may be because G166 had a low level of LGALS3 expression (Figure 3-15 Table). More immunostaining with the other GNS lines needs to be carried out to validate this. The expression was not confined to the cell surface but also within cells, which is consistent with the fact that LGALS3 is expressed also in the nucleus (http://www.uniprot.org/uniprot/P17931).



GNS line	G166	G144	G25	G179
log2 (TNC)	0.9	1.8	1.1	0.6

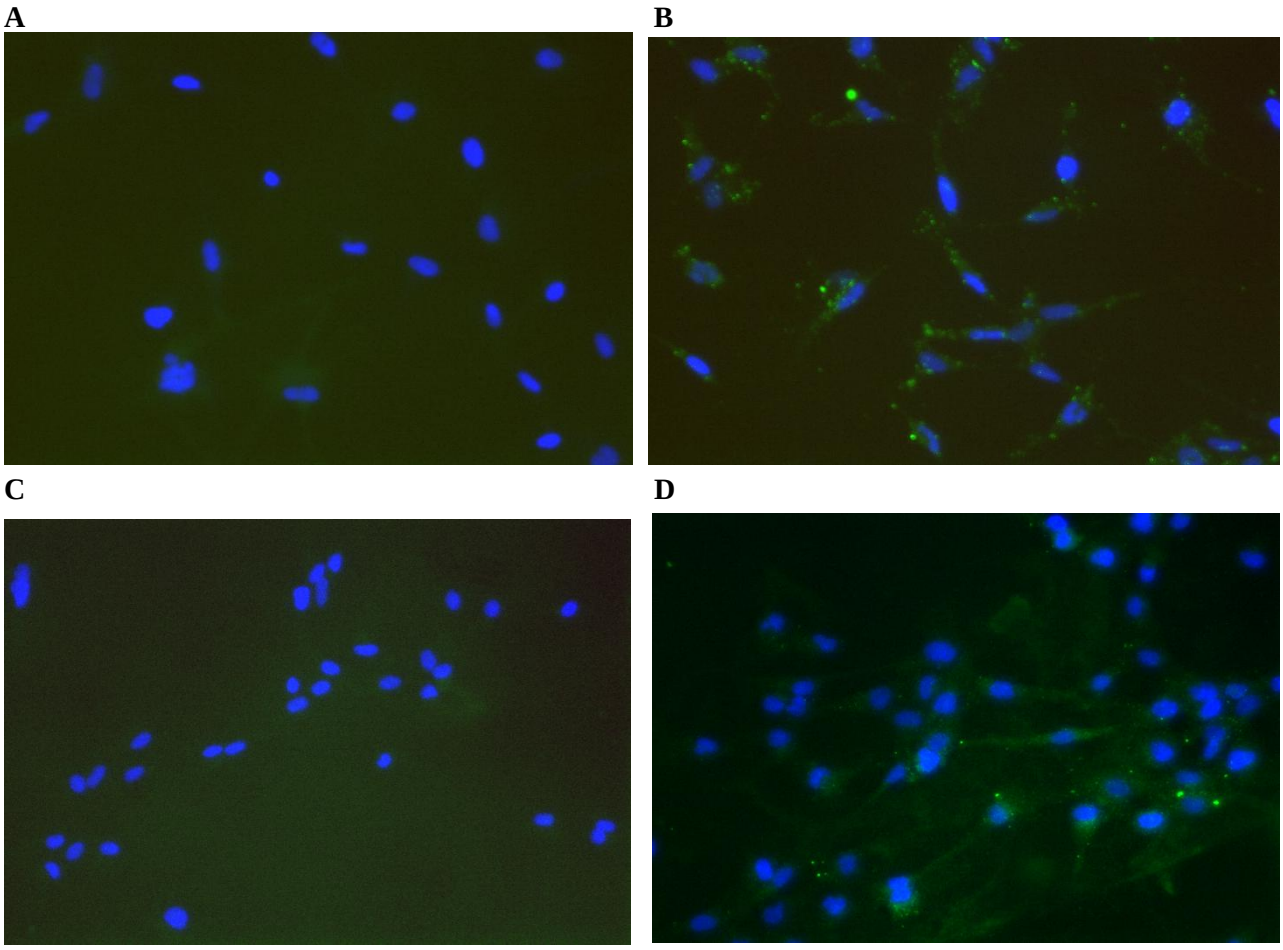
Figure 3-14 Immunostaining for TNC (green). Nuclei were stained with DAPI (blue) (A) U7 (NS). (B) G7 (GNS). (C) CB660 (NS). (D) G166 (GNS). Table shows log2(mean ratio) of TNC in each cell line from proteomics data.



GNS line	G166	G144	G25	G179
log2(LGALS3)	0.2	1.1	1.4	1.5

Figure 3-15 Immunostaining for LGALS3 (green). Nuclei were stained with DAPI (blue). (A) CB660 (NS). (B) G166 (GNS). Table shows log2(mean ratio) of TNC in each cell line from proteomics data.

THY1 has been characterised as a marker for many cell types including heamatopoietic stem cells [113], glioma stem cells [155] and early stages of iPS reprogramming from MEF [116]. It belongs to GO.BP neuron differentiation, and was over-expressed in our GNS. Our immunostaining did not show any THY1 expression on the NSs (U7 and CB660), while being expressed in many GNSs (G7 and G166), suggesting that it could be a reliable GNS-specific marker. We still need to validate this aspect by differentiating GNSs and examining THY1 expression on these cells. The observed expression could also be due to other reasons such as the age and patient background that is not related to glioma.

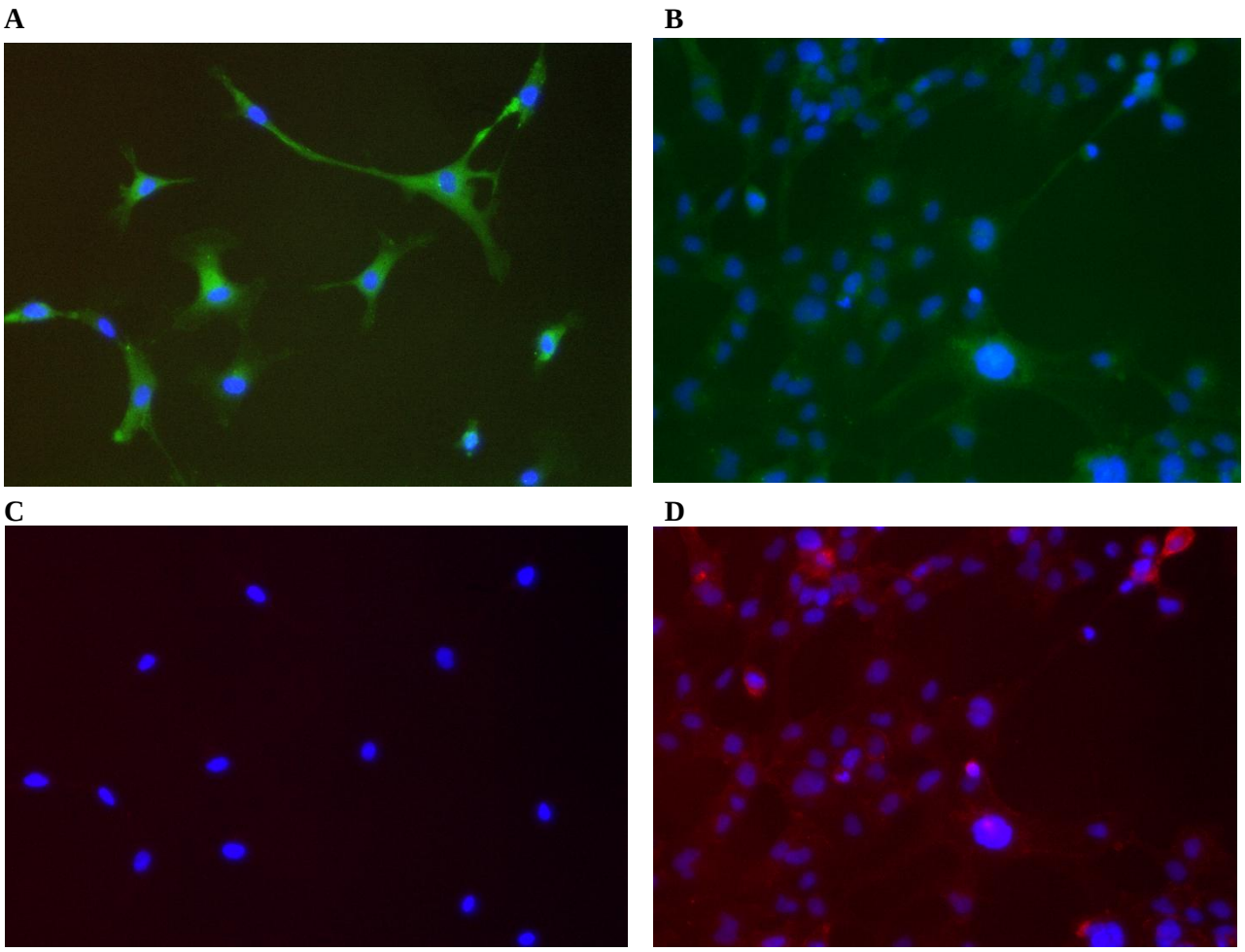


GNS line	G166	G144	G25	G179
log2(THY1)	1.5	3.5	3.5	1.4

Figure 3-16 Immunostaining for THY1 (green). Nuclei were stained with DAPI (blue). (A) U7 (NS). (B) G7 (GNS). (C) CB660 (NS). (D) G166 (GNS).

CD9 is a tetraspanin (transmembrane 4) on the cell surface with four hydrophobic domains, and was over-expressed in our GNSs. It can modulate cell adhesion and migration and tumour metastasis, and also trigger platelet activation and aggregation [147]. Co-immunostaining was carried out for these two proteins. In accordance with our tag-seq and proteomics data, TES was more highly

expressed in CB660 than in G7, whereas CD9 was totally absent in CB660 and weakly expressed in G7. Therefore, CD9 could be a marker only for GNSs but not for NSs, and vice versa for TES. More experiments are necessary to confirm this finding, as is the case for THY1.



GNS line	G166	G144	G25	G179
log2(CD9)	1.4	2.3	2.4	0.9
log2(TES)	-7.3	-3.6	-5	-3.6

Figure 3-17 Co-immunostaining for TES (green) for (A) CB660 (NS) and (B) G7 (GNS), and CD9 (red) for (C) CB660 (NS) and (D) G7 (GNS). Nuclei were stained with DAPI (blue).

Taken together, THY1 and CD9 could be GNS-specific markers, however, the markers could be the remnants of the original cells (e.g. neuron, astrocytes) that de-differentiated into GNSs and not silenced afterwards. To rule out this possibility, we need to make sure these proteins disappear after differentiating GNSs into adult cells as well as to test more GNS lines.

3.3.15 The IENS conditioned media could increase IENS cell proliferation but not ANS4 cells, and the ANS4 conditioned media did not proliferate either cell line

In the 2nd line of the follow-up experiment, we aimed to find GNS-secreted factors that could mediate tumourigenic transformation of NSs ex vivo. For this part of the project, ANS4 (mouse

normal NSs) and IENS (mouse GNSs) cell lines were also used since human cells often grow too slowly. As a preliminary step, the effect of ANS4- and IENS conditioned media was investigated by growing each cell line in both ANS4 and IENS media. The hypothesis here is that IENS conditioned media contains enough concentrations of secreted factors that could transform ANS4 cells. The result showed that the ANS4 conditioned media did not increase proliferation of either cell line, whereas the IENS conditioned media promoted proliferation of the IENS cells but not ANS4 cells (Figure 3-18). The increase in both confluency and cell count of the IENS cells exposed to autologous conditioned media appeared later and weaker than that of the EGF/FGF treatment. It is difficult to say if this increase was due to EGF (or related factors) and/or other factors since we could not identify EGF in our proteomics data (Table 3-9). The cellular morphology of IENS conditioned media-treated IENS cells did not seem to be different from that of EGF/FGF-treated IENS cells (Figure 3-19).

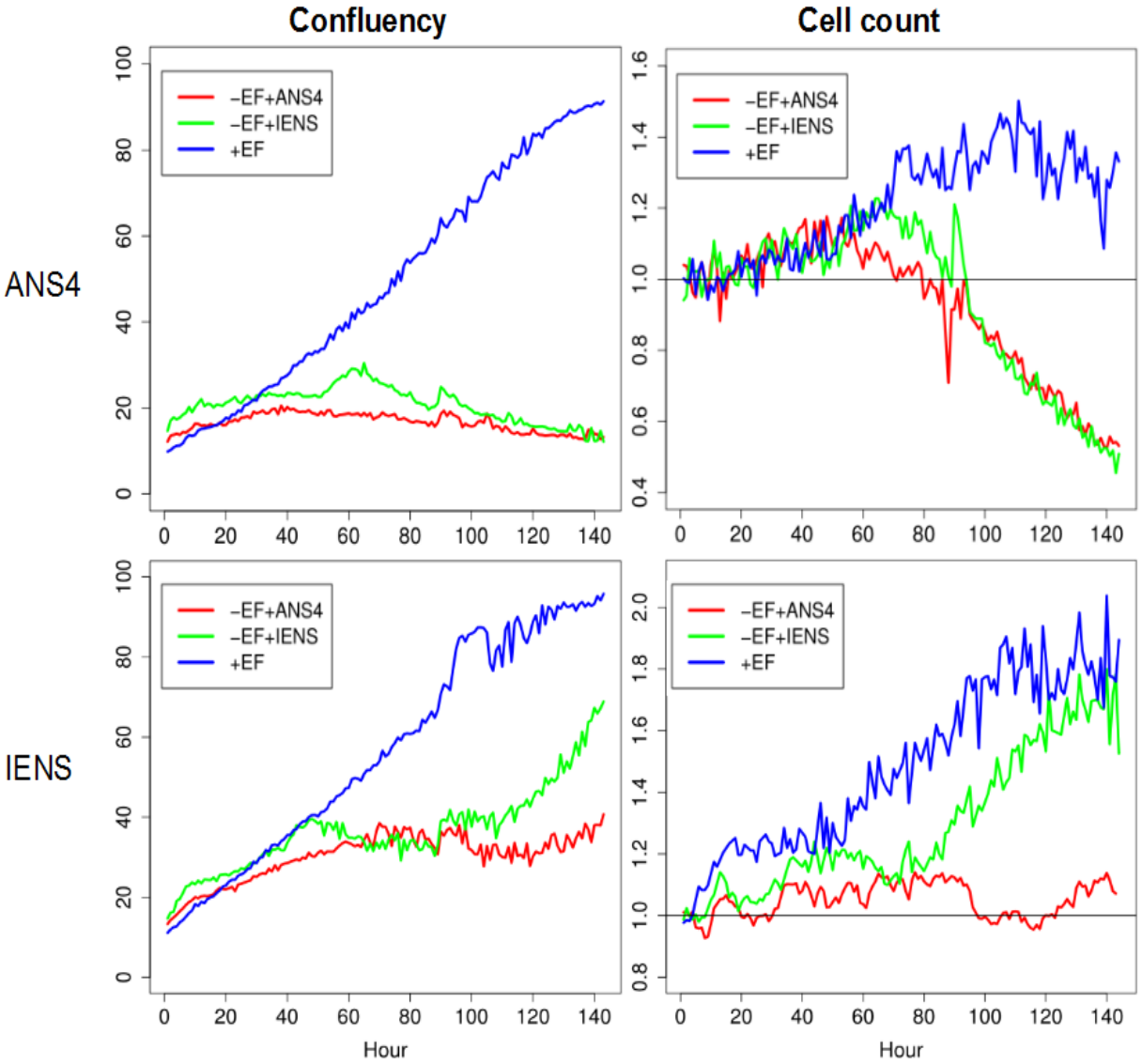


Figure 3-18 Confluency and cell count for cell line ANS4 and IENS treated with EGF/FGF (+EF) alone, ANS4 conditioned media without EGF/FGF (-EF+ANS4), and IENS conditioned media without EGF/FGF (-EF+IENS).

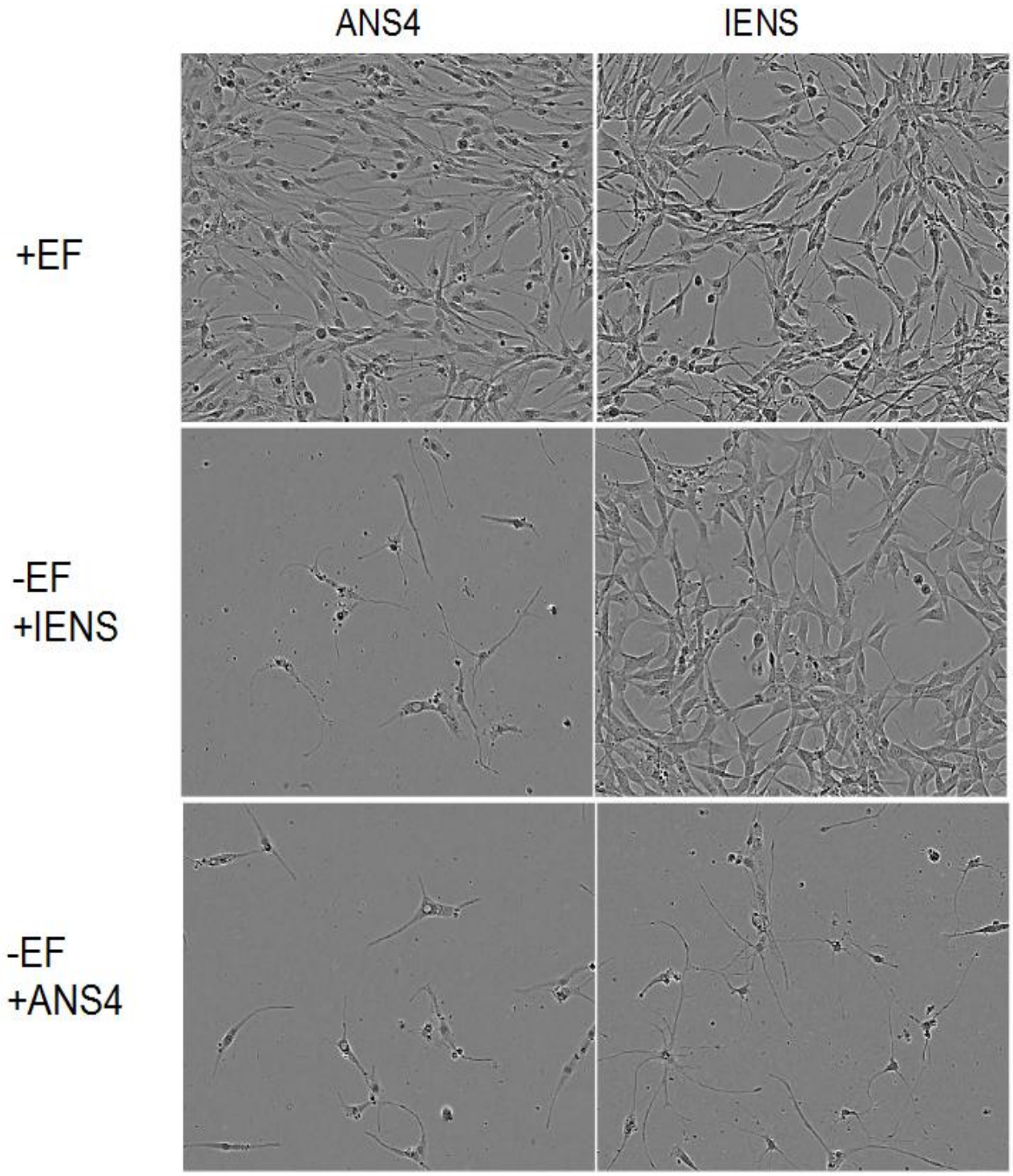


Figure 3-19 Phase-contrast images of ANS4 cell line at t = 144 h treated with EGF/FGF (+EF), without EGF/FGF +ANS4 conditioned media (-EF+ANS4), and without EGF/FGF +ANS4 conditioned media (-EF+IENS).

3.3.16 ANS4 cells stopped proliferating at EGF/FGF concentration below 0.1 ng/ml, whereas IENS cells continued proliferating in the absence of EGF/FGF

We hypothesized that the observed increase in IENS cell proliferation, but not ANS4 cells, in the IENS conditioned media was due to the difference in cell sensitivity to the media composition. Therefore, the sensitivity of ANS4 and IENS cells to EGF/FGF was investigated by exposing them

to a dilution series of these factors, since we already know how cells respond to them. The result (Figure 3-20) showed that the proliferation capability of ANS4 cells decreased at dilution factor 10E7 (the standard dilution factor is 10E5 (10 ng/ml)) and proliferation was almost completely ceased at 10E8. On the other hand, IENS cells continued proliferating even in the absence of EGF/FGF. This was, however, somewhat contradictory to the earlier experiment, where IENS cells did not proliferate in the -EF+ANS4 conditioned media (Figure 3-19), implying that ANS4 cells might secrete anti-proliferation/pro-differentiation factors. Further experiments are necessary to confirm this.

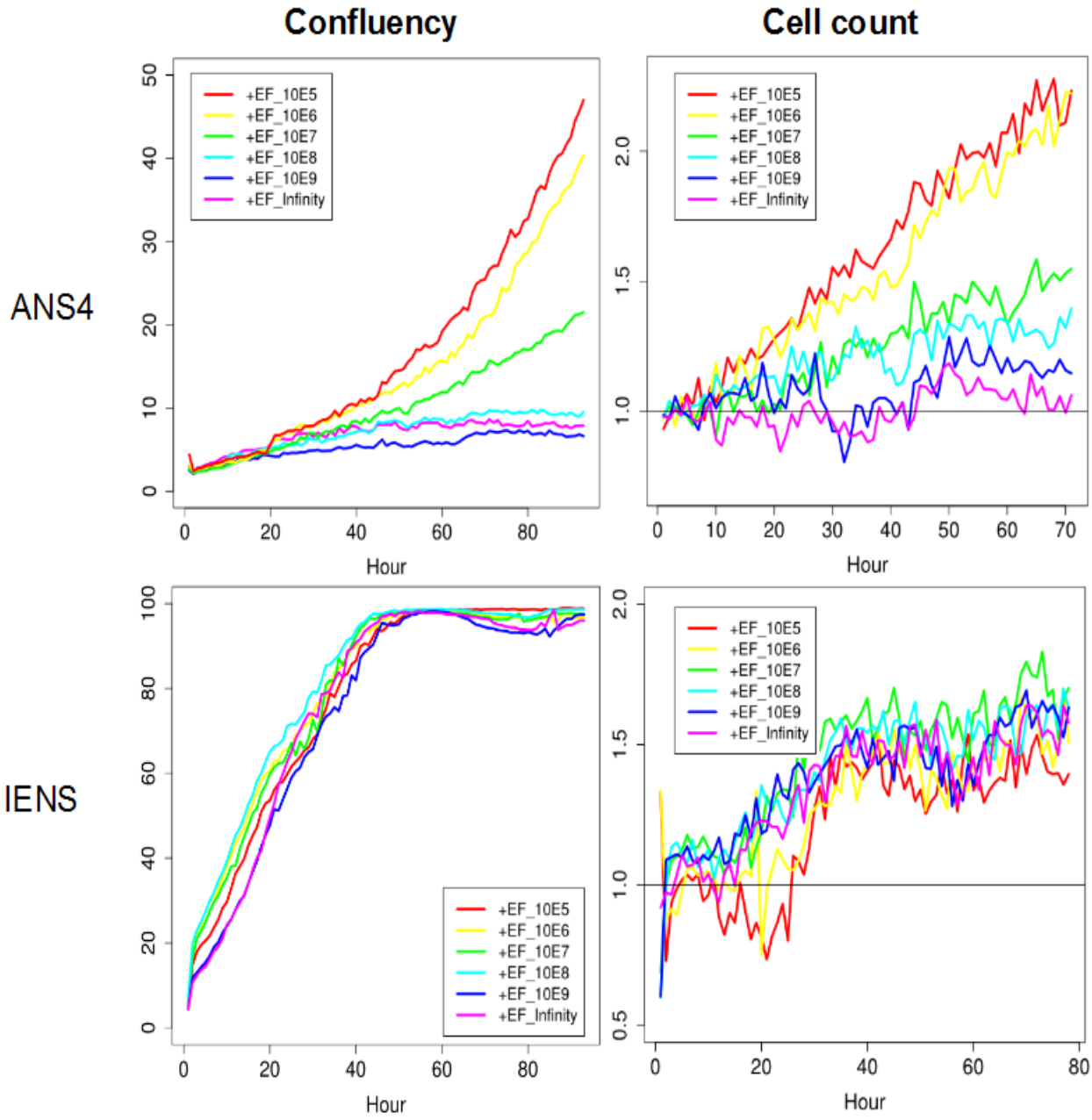


Figure 3-20 Confluency and cell count for cell lines ANS4 and IENS treated with EGF/FGF (+EF) at different doses. The standard dilution factor is 10E5. The dilution series was make up to 10E9.

3.3.17 TNC, MDK, CNTF, APOE, IGFBP3, IGFBP4 and CSF1 were chosen as candidate secreted factors for proliferation/tumourigenesis

Next, we aimed to test the effect of candidate secreted factors from our proteomics data. Venugopal et al. [67] observed using neurosphere culture that the expression of TGF-alpha family ligands, EGF, VEGF, PDGF and their cognate receptors, was higher than the control media, which suggests that these factors may be responsible for transient proliferation of their NSs. However, our proteomics analysis did not sufficiently identify these proteins (Table 3-9) and therefore their expression levels need to be quantified by experiments like ELISA. In a previous drug screen, EGFR pathway inhibitors did not efficiently kill GNSs, indicating that other pathways may be operating downstream and that GNS secreted factors, other than EGF, VEGF and PDGF, may induce/help tumourigenesis of NSs. In the Venugopal et al. [67] study, a mass spectrometry analysis of the GBM secretome did not definitively identify key secreted factors that may play a role in induction of transformation of NPCs. Here, we selected seven factors, which were DE in our total cell/secretome data, for testing their ability to transform NSs.

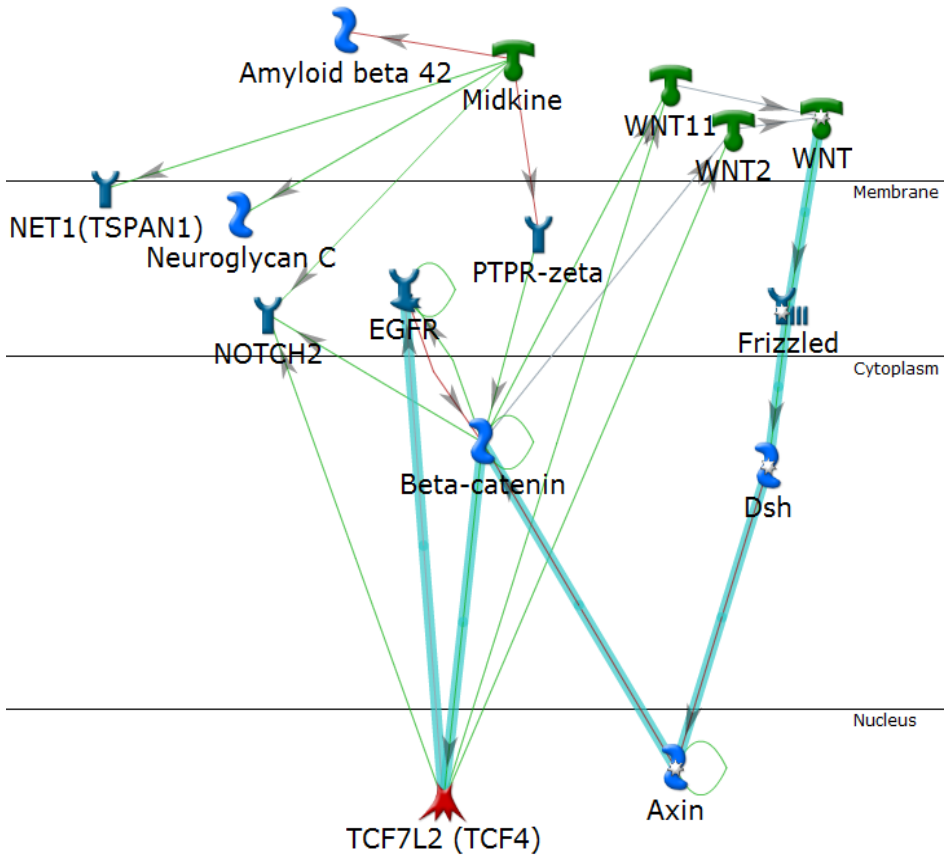
Tenascin-C (TNC) was one of the proteins in the GO.BP neuron differentiation category that is known to increase the sensitivity of embryonic NSs to EGF and FGF [109]. It was also shown to directly bind to EGFR [110]. So our question is whether TNC could induce an EGF-like effect, leading to increased proliferation. The 2nd candidate, Midkine (MDK), is more than 4-fold up-regulated in our GNSs. It is a pleiotrophin family growth factor that can activate anaplastic lymphoma kinase [58] and subsequent mitogen-activated protein kinase (MAPK) and PI3-kinase, which leads to cell proliferation and can be related to tumourigenesis (http://www.uniprot.org/uniprot/P21741). Care should be taken, however, since it is also involved in early foetal adrenal gland development and its differential expression could be due to the comparison between foetuses and adults. Networks of known interactions for the candidates were made in MetaCore. Although these networks consist of interactions identified in different cell types and conditions, they might give us some hint as to what could happen when the system is perturbed. The network for MDK (3-21 A) suggests that it could increase sensitivity to EGF by increasing EGFR expression. APOE3 is a secreted lipid transporter that binds to MEGF7, LRP1, APOER2, VLDLR, and can stimulate neurosphere formation via MAPK/ERK pathway. The network of known interactions (Figure 3-21 B) indicates that APOE might activate RBPJ and NOTCH1, which are known self-renewal regulators. APOE could also activate CNTF. Since CNTFR was up-regulated in our GNSs, CNTF was also chosen for the follow-up study. In the network CNTF could lead to the activation of JAK-STAT pathway. CNTF is also known to promote astrocyte differentiation and its receptor CNTFR was shown to be over-expressed in glioma stem cells by another study [156]. Since mutations were not found in its sequence, this receptor was apparently functional [156]. It is

thus contradictory that proliferating GNSs express more receptors for a pro-differentiation factor and this question needs to be addressed. Since APOE is known to activate CNTF, we hypothesized that APOE could have a similar effect to that of CNTF. IFGBP3 and IFGBP4 can prolong the half-life of insulin-like growth factors (IGFs), which has a growth promoting effect. Thus, our hypothesis is that IGFBP3 and IGFBP4 could increase the IGF effect, namely proliferation. The network analysis (Figure 3-21 C) suggests that these proteins could lead to activation of an oncogene c-Myc. CSF1 is a cytokine released from tumour cells and thought to stimulate macrophages to release EGF. It could activate VEGFR-3, whose signal could have a similar outcome as the EGFR signalling. Its putative network is shown in (Figure 3-21 D).

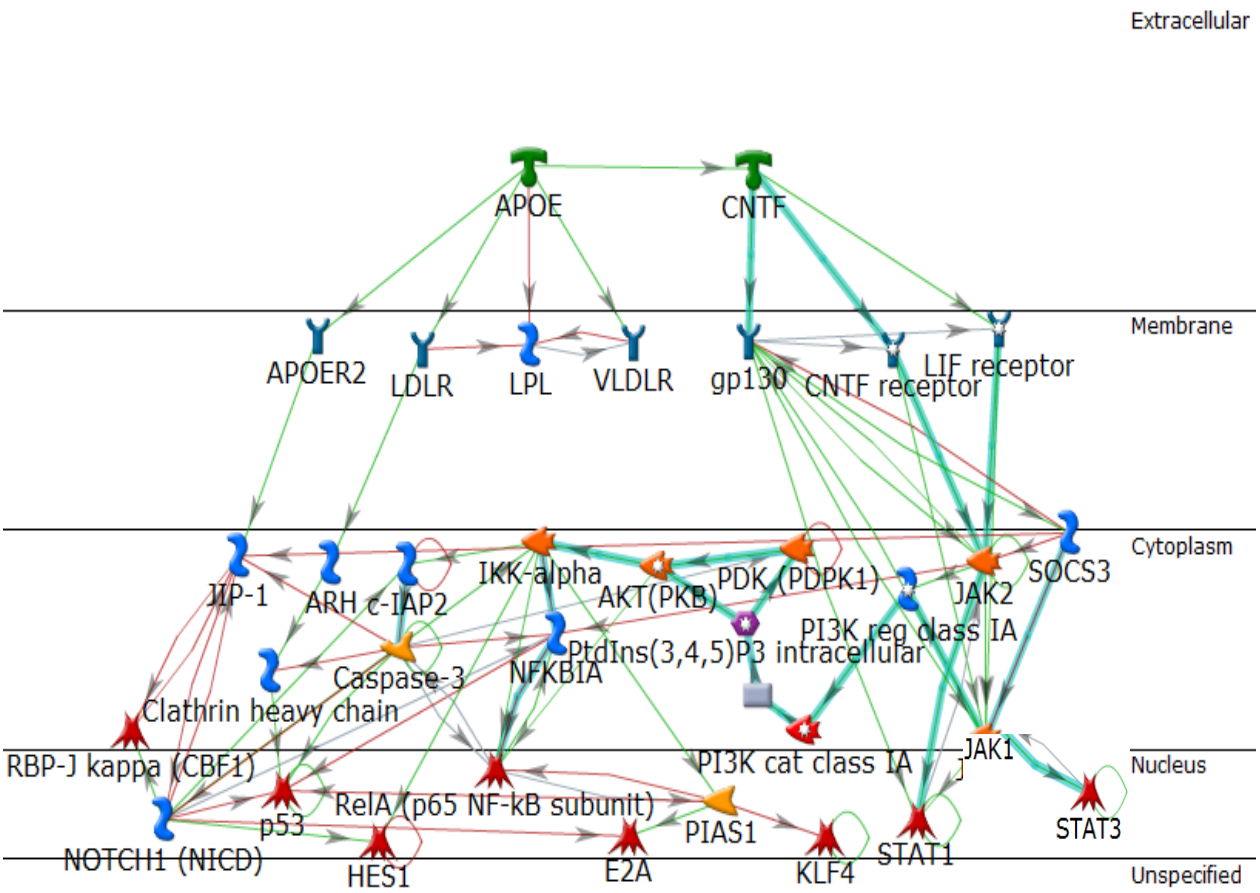
Table 3-9 Expression level of EGF, EGFR, VEGF, VEGFR, PDGF and PDGFR.

	G166	G144	G25	G179	G166.secre	G144.secre
EGF	-	-	-	-	-	-
EGFR	-0.99	-1.95	-0.71	-1.59	0.31	-0.42
VEGF(A)	-	-	-	-	2.36	0.78
VEGFR	-	-	-	-	-	-
PDGF	-	-	-	-	-	-
PDGFR(A)	2.38	-1.08	1.34	3.95	-	-
PDGFR(B)	-2.40	-0.98	0.78	-1.17	-	-

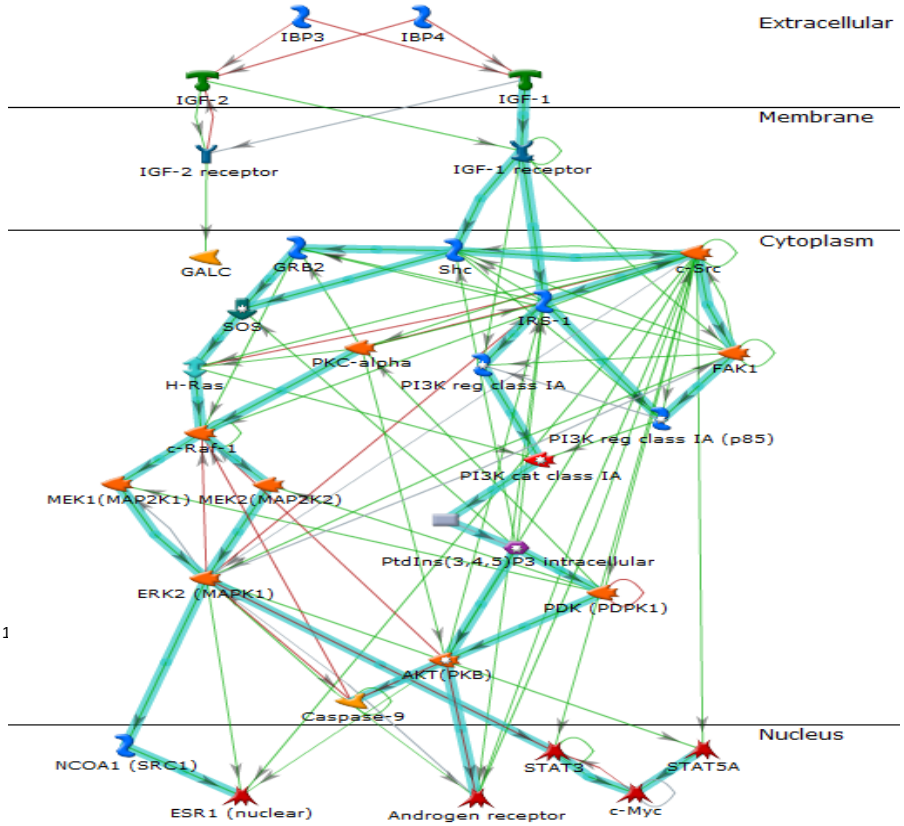
A



B


Extracellular
APOE
CNTF
Membrane
APOER2
LDLR
LPL
VLDLR
gp130
CNTF receptor
LIF receptor
Cytoplasm
JIP-1
ARH
c-IAP2
IKK-alpha
AKT(PKB)
PDK (PDPK1)
JAK2
SOCS3
PI3K reg class IA
PtdIns(3,4,5)P3 intracellular
Caspase-3
NFKBIA
Clathrin heavy chain
RBP-J kappa (CBF1)
JAK1
Nucleus
PI3K cat class IA
RelA (p65 NF-kB subunit)
p53
PIAS1
STAT3
NOTCH1 (NICD)
HES1
E2A
KLF4
STAT1
Unspecified

C


Extracellular
IBP3
IBP4
IGF-2
IGF-1
Membrane
IGF-2 receptor
IGF-1 receptor
Cytoplasm
GALC
GRB2
Shc
c-Src
SOS
IRS-1
PKC-alpha
FAK1
H-Ras
PI3K reg class IA
PI3K reg class IA (p85)
c-Raf-1
PI3K cat class IA
MEK1(MAP2K1)
MEK2(MAP2K2)
PtdIns(3,4,5)P3 intracellular
ERK2 (MAPK1)
PDK (PDPK1)
AKT(PKB)
Caspase-9
Nucleus
NCOA1 (SRC1)
STAT3
STAT5A
ESR1 (nuclear)
Androgen receptor
c-Myc

1

D

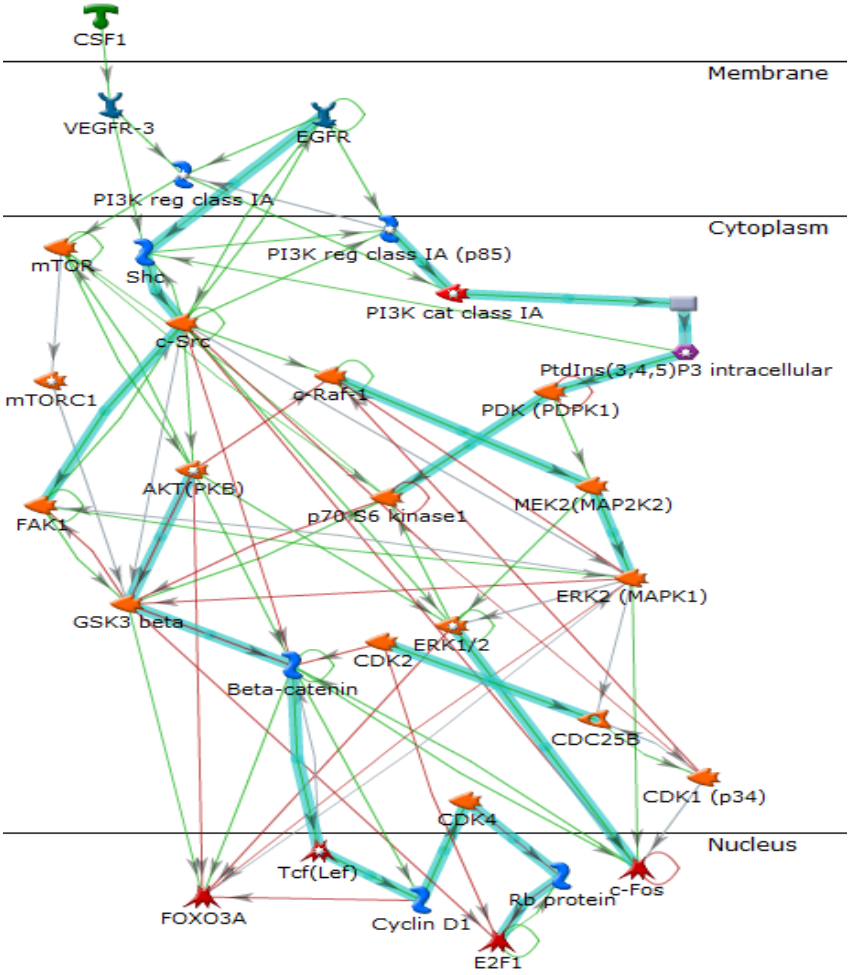


Figure 3-21 Networks of known interactions reconstructed in MetaCore for (A) MDK, (B) APOE, (C) IGFBP3 and IGFBP4, and (D) CSF1. Each factor was used as a seed and MetaCore function 'Auto expand' was used with 50 maximum number of nodes. Resultant networks were pruned by removing nodes not connecting down to nucleus. Thick arrows indicate well-established, canonical pathways.

3.3.18 Up to 100 ng/ml TNC, MDK and CNTF did not have any visible effect on mouse ANS4 and IENS cells in the presence of EGF/FGF

First, the effect of TNC, MDK and CNTF on cells at different doses (5, 10, 50 and 100 ng/ml) in the standard culturing conditions (i. e., in the presence of EGF/FGF) was evaluated. At the highest dose (100 ng/ml), these factors did not have a discernible effect on both ANS4 and IENS cells in terms of the confluency (Figure 3-22), cell count (Figure 3-22), morphology (Figure 3-23) and proliferation capacity (Figure 3-24). 5 ng/ml, 10 ng/ml and 50 ng/ml showed the same results (data not shown), suggesting that the factors were neither pro-proliferation, nor pro-differentiation, nor toxic to both NSs and GNSs in the presence of EGF/FGF.

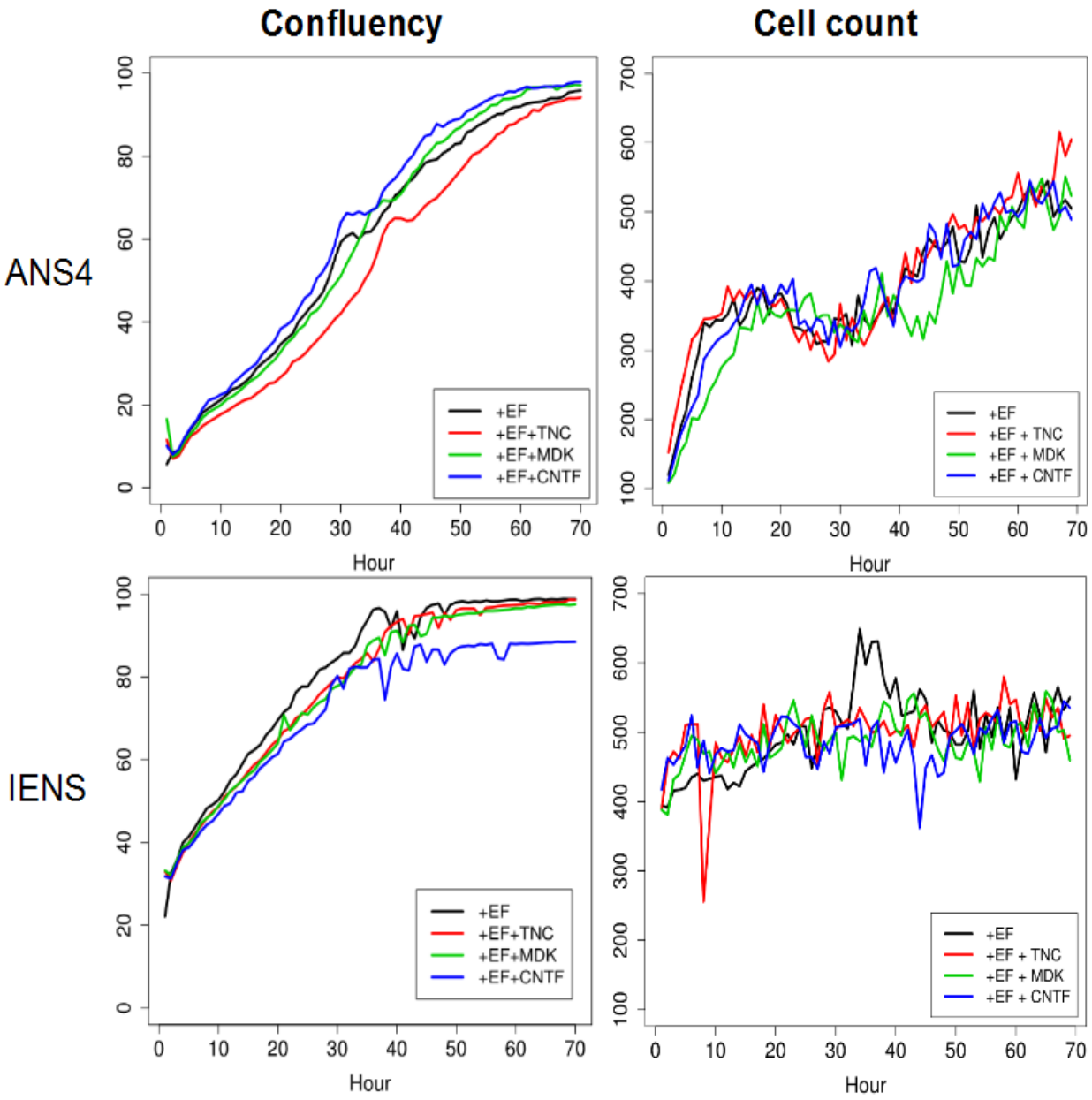


Figure 3-22 Confluency and number of cells counted every hour for 68 hours. ANS4 and IENS cells were treated with EGF/FGF (+EF) alone and in combination with 100 ng/ml TNC, 100 ng/ml MDK and 100 ng/ml CNTF.

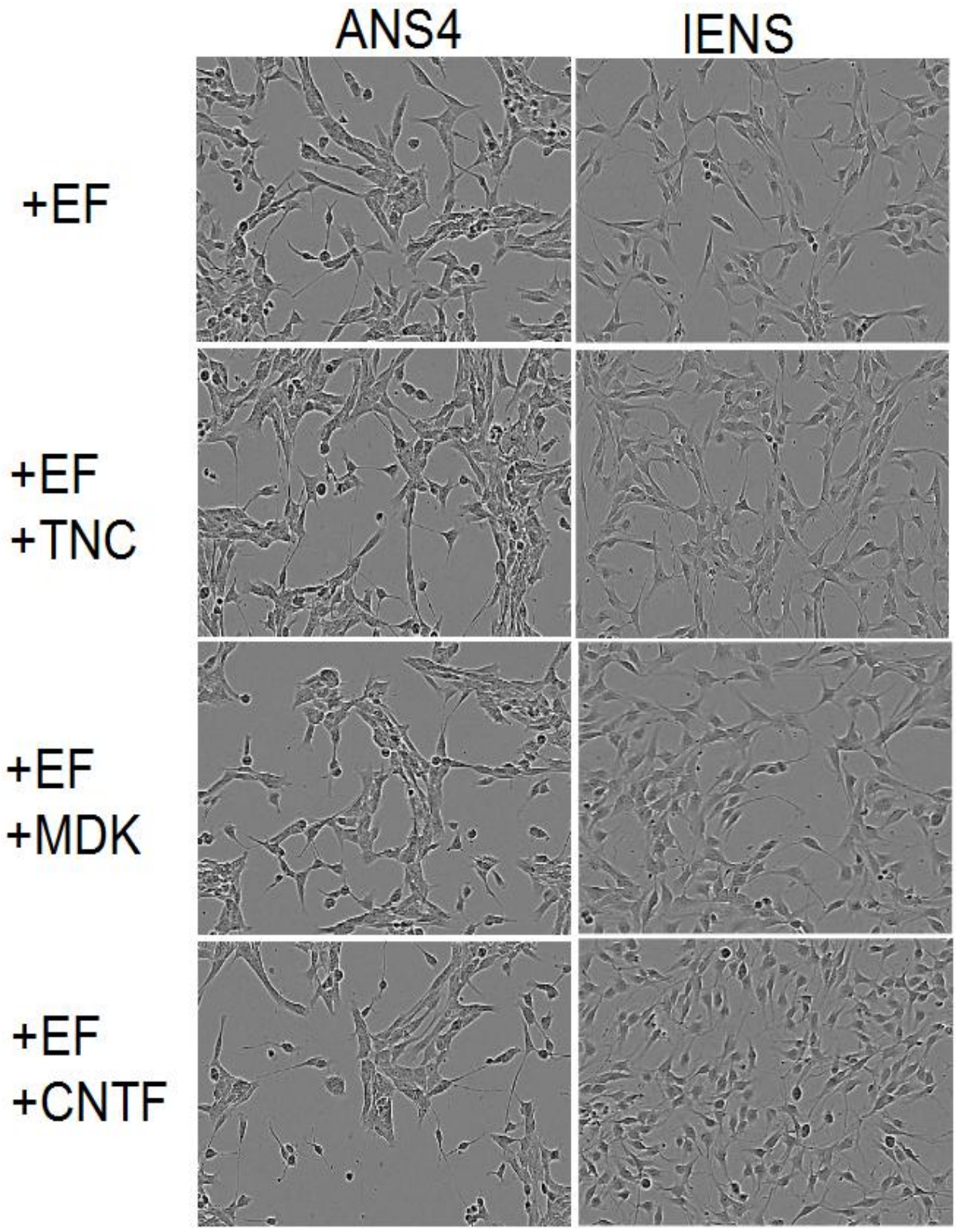


Figure 3-23 Phase-contrast images of ANS4, IENS and 223 mouse lines at t = 48 hours. ANS4 and IENS cells were treated with EGF/FGF (+EF) alone and in combination with 100 ng/ml TNC, 100 ng/ml MDK and 100 ng/ml CNTF.

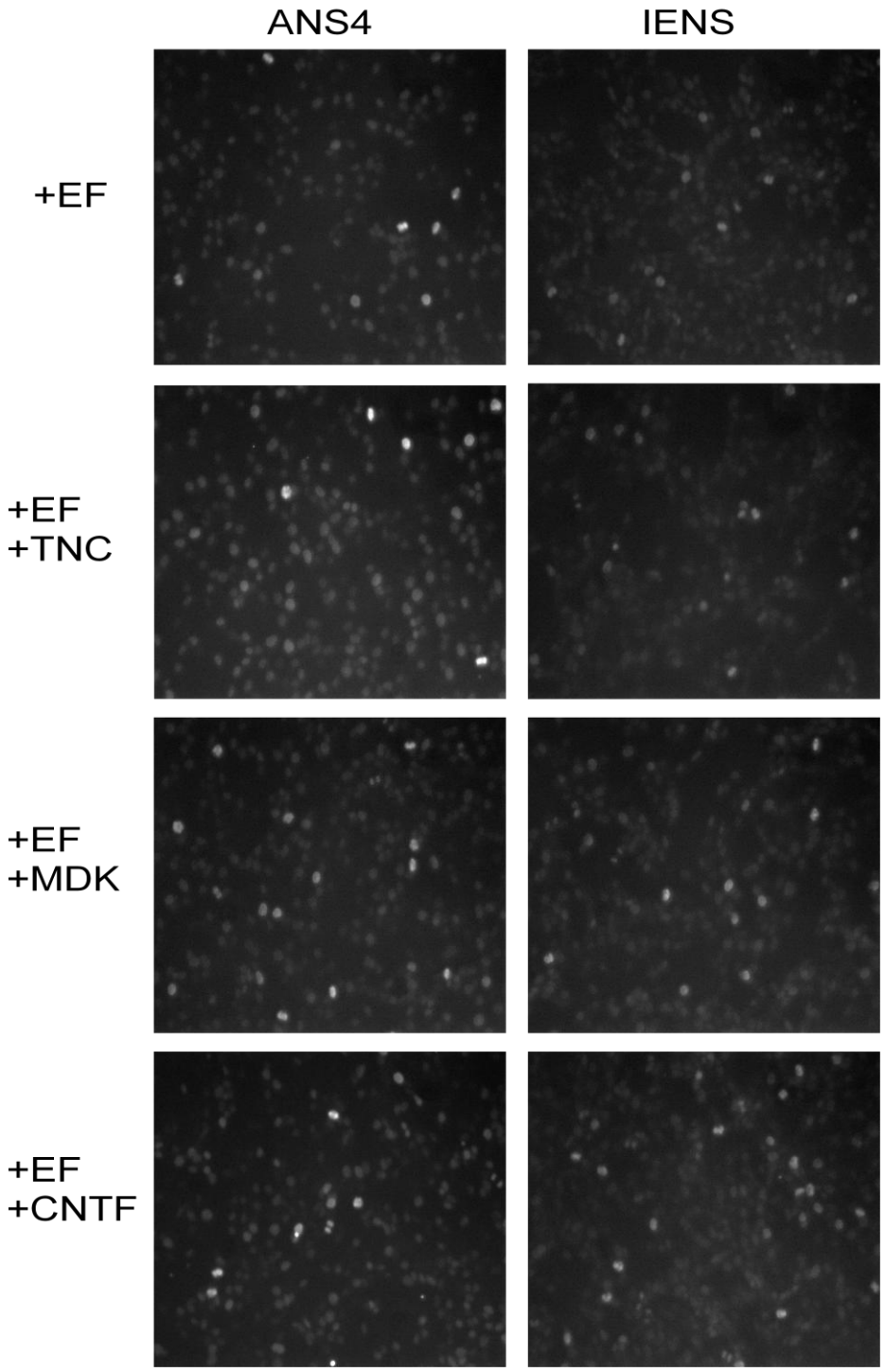


Figure 3-24 Immunocytochemistry for KI67 in cell lines ANS4 and IENS 72 hours post-treatment with EGF/FGF (+EF) alone and in combination with 100 ng/ml TNC, 100 ng/ml MDK and 100 ng/ml CNTF.

3.3.19 Effect of TNC, MDK, CNTF, APOE3, IGFBP3, IGFBP4 and CSF1 on ANS4 and CB660 cells in the absence of EGF/FGF

The effect of each of the seven candidate factors (TNC, MDK, CNTF, APOE3, IGFBP3, IGFBP4 and CSF1) on the ANS4 and CB660 cell lines was tested in the absence of EGF/FGF (-EF). The cell count, but not confluency, of ANS4 cells treated with MDK (-EF+MDK) increased at a similar rate to the +EF positive control up to time t = 80 h (Figure 3-25). Thus, the morphology of these ANS4

cells was checked at t = 77 h (Figure 3-26). Cells, however, looked more similar to the -EF negative control, indicating that MDK might promote proliferation to a certain degree but the effect is transient and not as strong as EGF. This effect was not observed in CB660. More replicates with different lines are needed to confirm this. The confluency of CNTF-treated ANS4 cells looked a little larger than the others but the cell count remained the same (Figure 3-25). The morphological comparison between -EF, -EF+CNTF, +EF and +EF+CNTF at t = 144 h (Figure 3-27) showed that -EF+CNTF cells appeared more roundish than the other three, contributing to the observed increased confluency. CNTF is known to induce astrocyte differentiation and EGF is probably a strong pro-proliferation factor that can suppress the CNTF differentiation effect. This CNTF effect was, however, apparently not observed in CB660 and the morphology seemed the same as the -EF negative control (Figure 3-28).

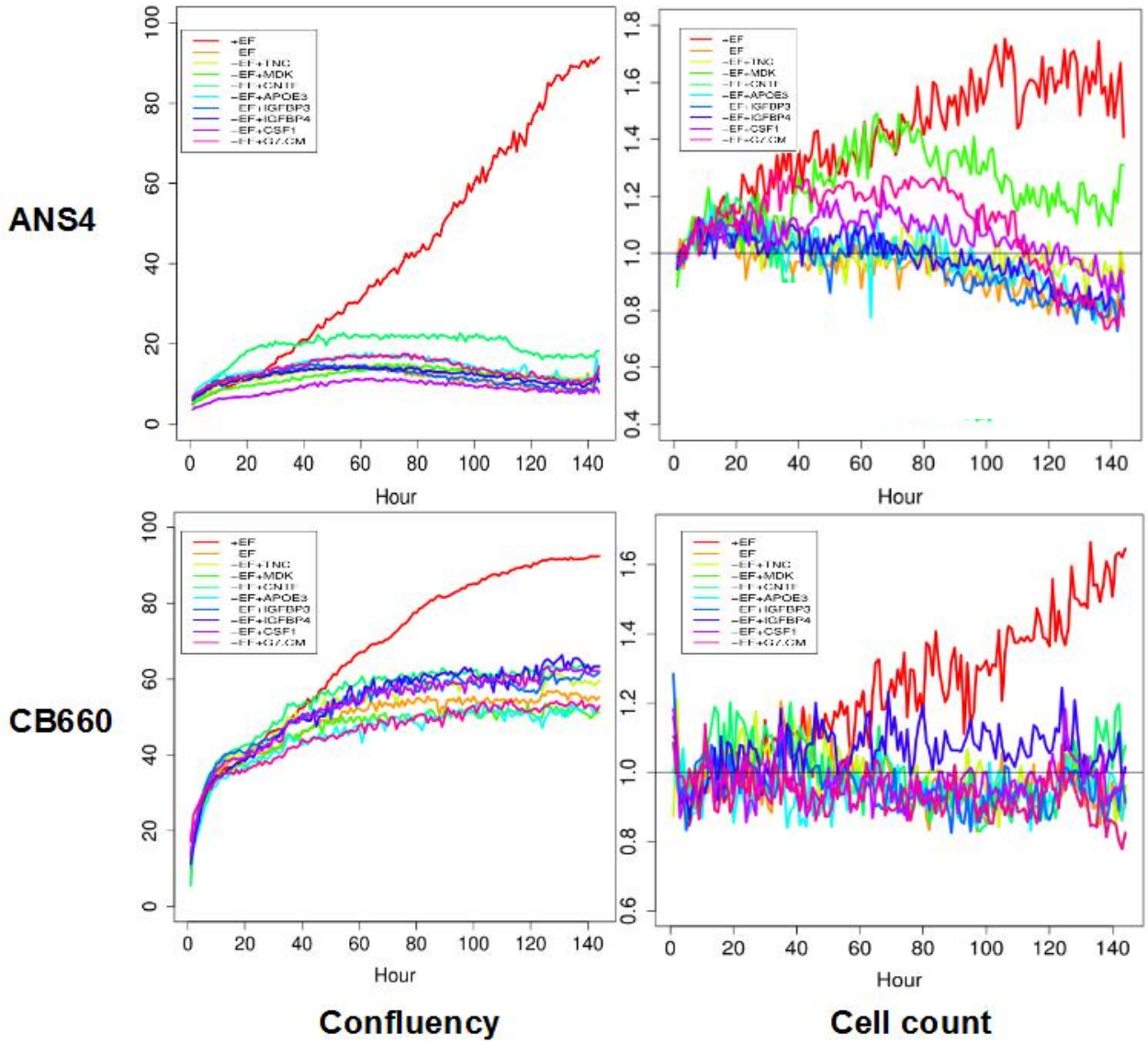


Figure 3-25 Confluency curve for cell lines ANS4, CB660 treated with EGF/FGF (+EF), without EGF/FGF (-EF), -EF+TNC, -EF+MDK, -EF+CNTF, -EF+APOE3, -EF+IGFBP3, -EF+IGFBP4, -EF+CSF1 and -EF+G7 conditioned media (-EF+G7.CM).

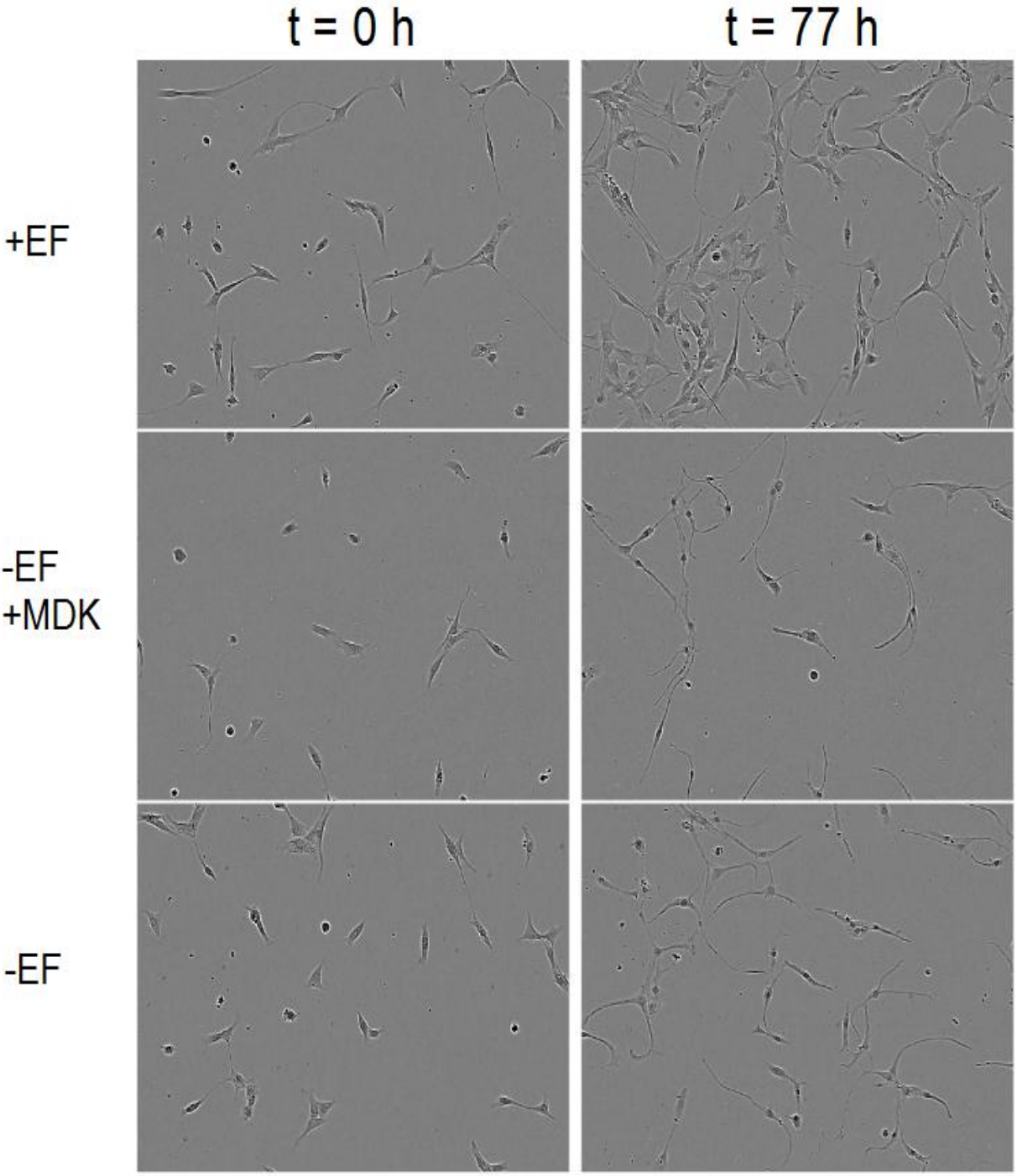


Figure 3-26 Phase-contrast images of ANS4 cell line at t = 0 and 77 h treated with EGF/FGF (+EF), without EGF/FGF (-EF) and -EF+MDK 10 ng/ml.

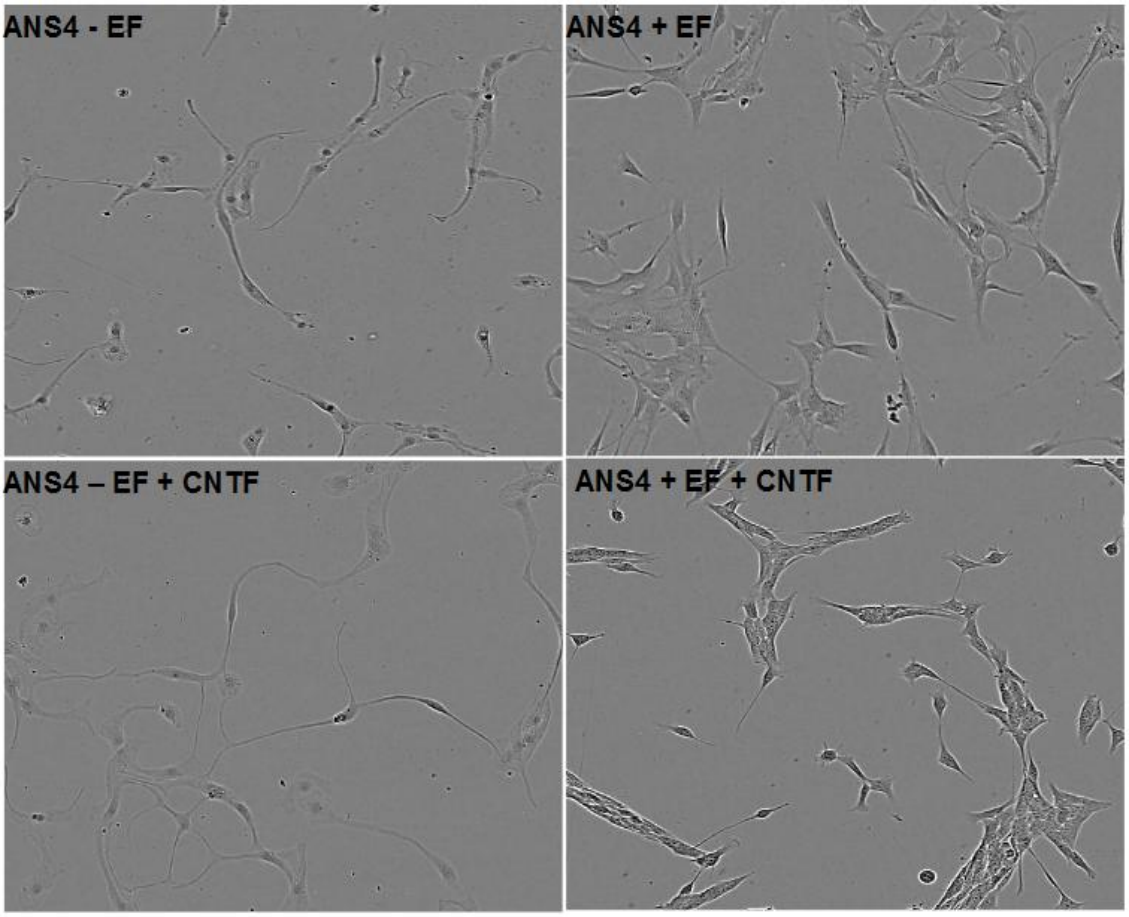


Figure 3-27 Phase-contrast images of of ANS4 cell line at t = 144 h treated without EGF/FGF (-EF), -EF+CNTF, with EGF/FGF (+EF) and +EF+CNTF 10 ng/ml.

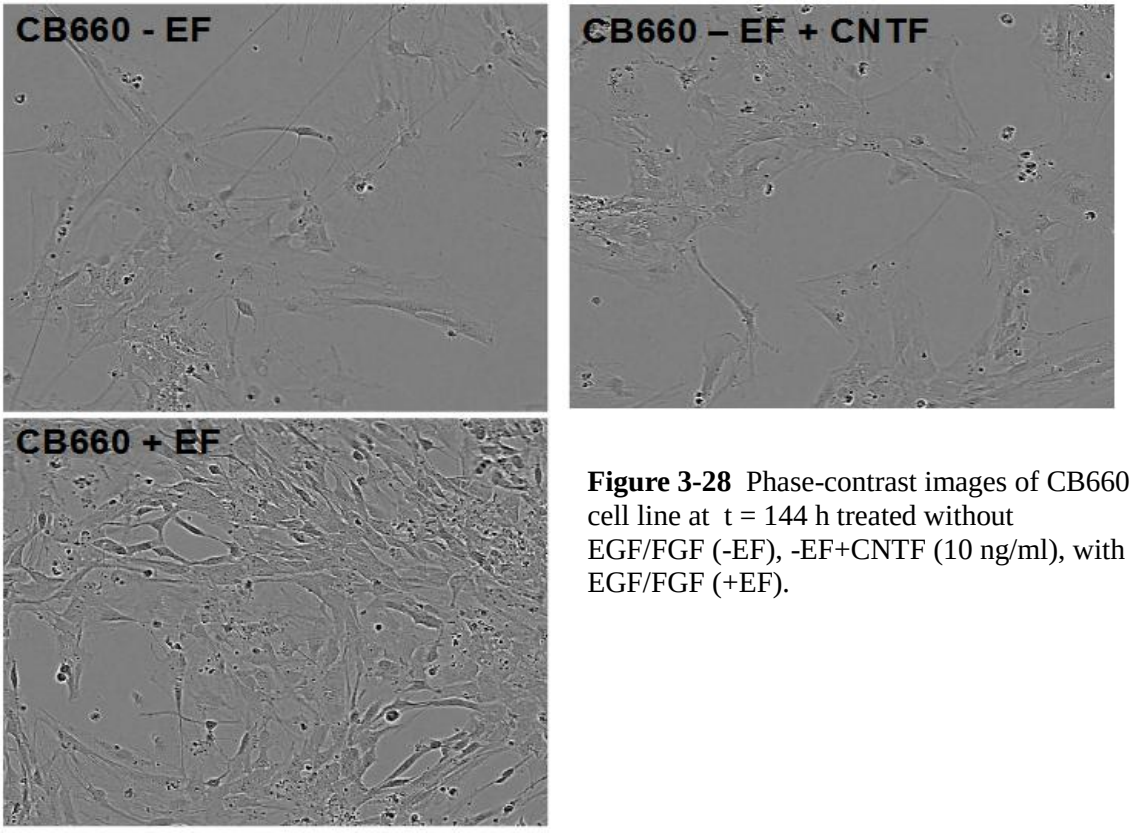


Figure 3-28 Phase-contrast images of CB660 cell line at t = 144 h treated without EGF/FGF (-EF), -EF+CNTF (10 ng/ml), with EGF/FGF (+EF).

3.3.20 Treating ANS4 cells with all the seven factors (TNC, MDK, CNTF, APOE3, IGFBP3, IGFBP4 and CSF1) simultaneously was similar to the CNTF treatment alone, and treating with six factors (TNC, MDK, APOE3, IGFBP3, IGFBP4 and CSF1) did not induce a visible difference

Next, all the above seven factors were added to the media at the same time and the effect on ANS4 cells was monitored. The results (Figure 3-29) and (Figure 3-30) showed that the seven factors resulted in cells similar to when they were treated with the CNTF alone. When only six factors excluding CNTF were added to the media, cells did not show an appreciable difference from the -EF negative control, suggesting that these factors do not have a synergistic effect on ANS4 cells in the absence of EGF/FGF.

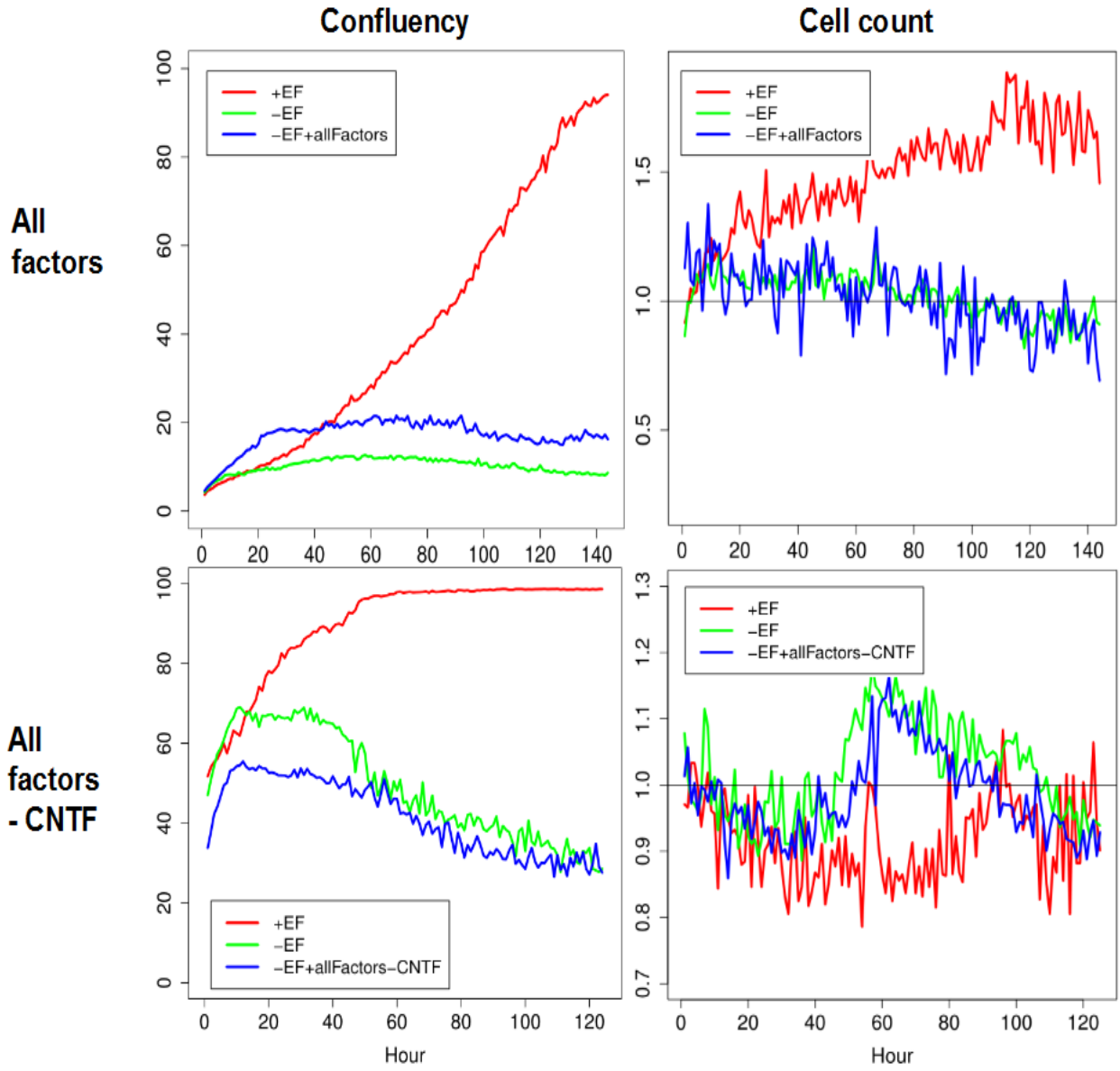


Figure 3-29 Confluency and cell count for cell line ANS4 treated with EGF/FGF (+EF) alone, without EGF/FGF (+EF) and in combination with , 100 ng/ml MDK and 100 ng/ml CNTF.

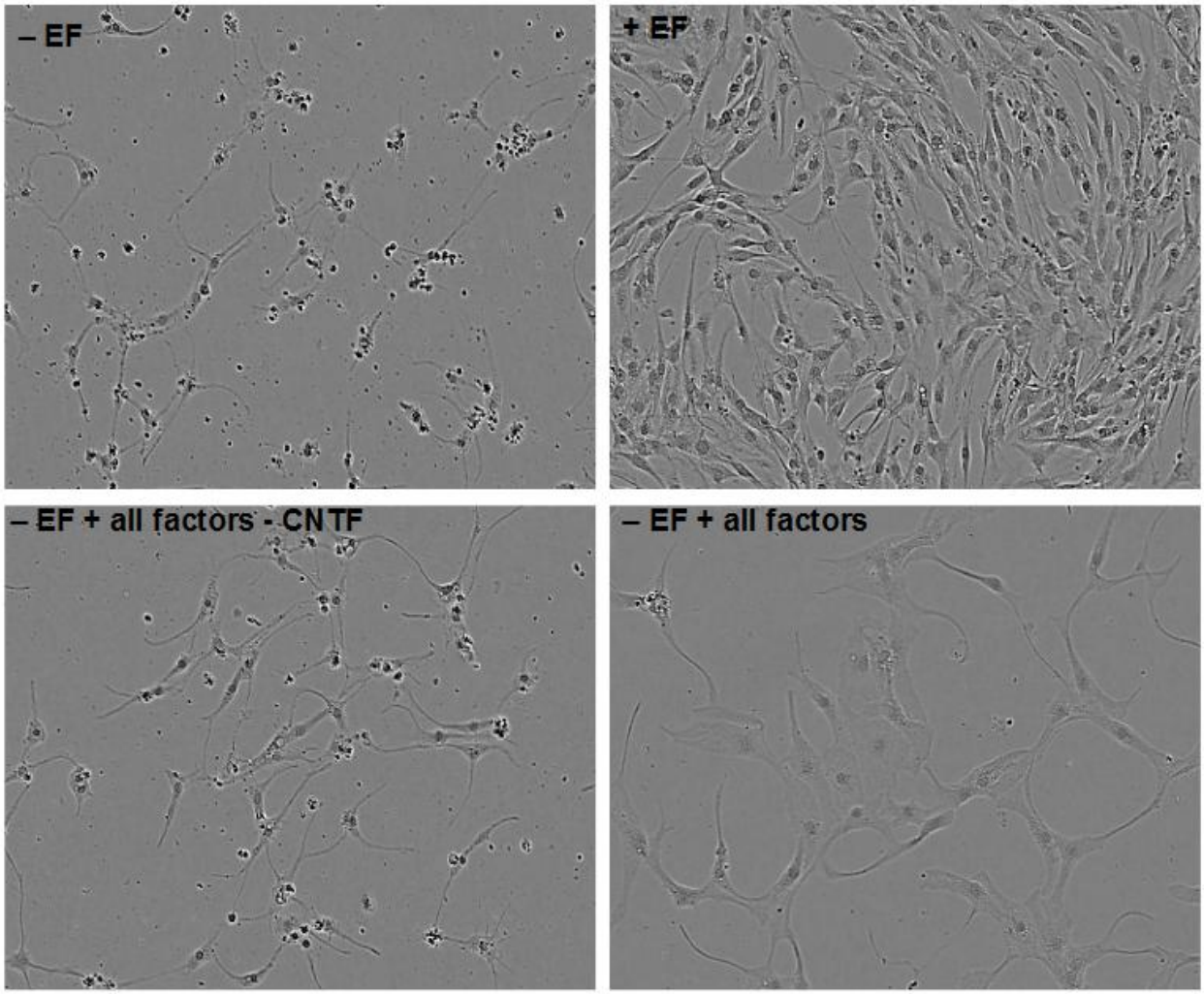


Figure 3-30 Phase-contrast images of ANS4 cell line at t = 144 h treated without EGF/FGF (-EF), -EF+all factors (TNC, MDK, CNTF, APOE3, IGFBP3, IGFBP4 and CSF1), with EGF/FGF (+EF) and -EF+all factors-CNTF (TNC, MDK, APOE3, IGFBP3, IGFBP4 and CSF1).

3.3.21 Some factors might have an effect on ANS4 and IENS cell colony formation

The effect of the seven factors and conditioned media in the presence of EGF/FGF was also investigated by the colony forming assay (Figure 3-31). TNC, MDK and IGFBP4 might enhance colony formation of ANS4 cells, whereas IGFBP3 might enhance colony formation of IENS cells. Since IGFBP4 was up-regulated in the human GNSs and IGFBP3 was up-regulated in NSs, differential use of IGFBP3 and 4 might be associated with tumourigenesis. Notably, CNTF strongly inhibited IENS colony formation, further validating that it is a pro-differentiation (i.e., anti-proliferation) factor. Since the number of replicates is still n=1, drawing any conclusion requires more experiments.

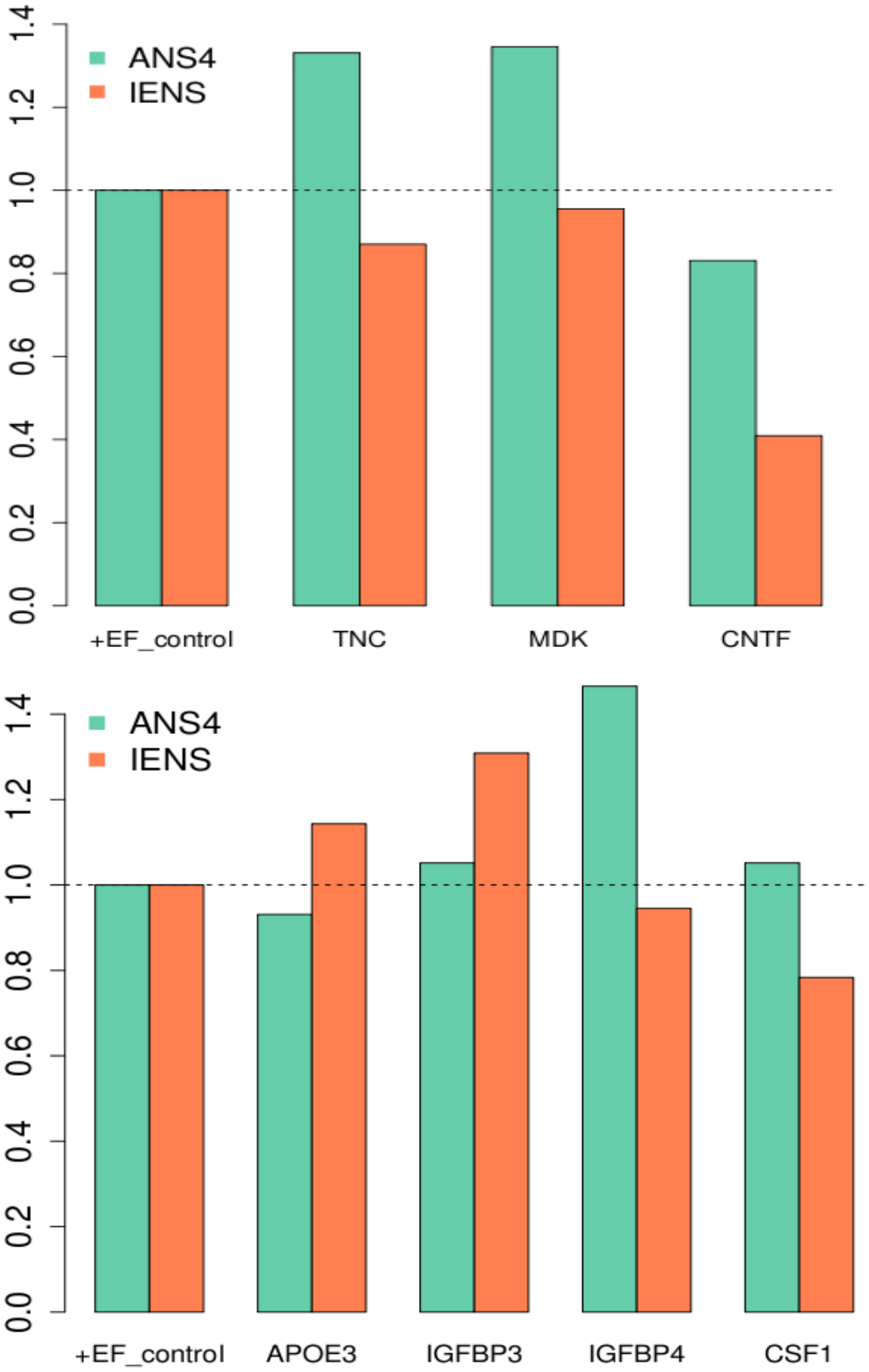


Figure 3-31 Normalized colony count of ANS4 and IENS cell lines at t = 7-10 days after treated with EGF/FGF (+EF) and TNC, MDK, CNTF, APOE3, IGFBP3, IGFBP4 or CSF1 conditioned media.

3.4 Discussion

To reveal the difference in global protein expression between untransformed, karyotypically normal foetal neural stem cells [1] and malignant neural stem cells derived from adult gliomas (GNSs), we performed mass spectrometry analyses of both total cell proteome and secreted proteome of these cells. This resulted in a total of ~7500 and ~2000 quantified proteins and 446 differentially expressed (DE) proteins (152 up-regulation, 294 down-regulation in GNSs) and 167 DE proteins (144 up-regulation, 23 down-regulation), respectively. Among the known NS and glioma stem cell markers, Galectin-3, Galectin-3-binding protein, L1CAM, GFAP were over-expressed in our GNSs, while integrin α6 and ALDH2 were under-expressed. CD133 has been used to isolate brain tumour stem cells from different brain tumours [46, 47] but it was identified only in our G166 and G179 lines, and was not DE when compared to the NSs and not on the mRNA level either [77], suggesting that this marker may not be a universal and reliable marker to distinguish GNSs from NSs.

The gene set enrichment analysis (GSEA) of the 446 DE proteins resulted in chromosome 7 and 15, which are almost always gained and lost in our GNSs, respectively, suggesting that chromosomal aberrations influence not only mRNA- but also protein expressions. In addition, several tumour-associated processes were enriched including those related to cell differentiation, cell motility, ECM interactions, focal adhesion, structural organization and cell signalling. These findings support the notion that cell-matrix and cell-cell adhesion play a crucial role in the maintenance of stem cell pools, for instance integrins and other surface proteins mediate signals whose disruption may lead to tumourigenesis. Furthermore, the enrichment of categories related to blood vessel development is in accordance with the importance of angiogenesis in CSC maintenance, as shown in squamous carcinoma [157] and glioma stem cells [158, 159]. Although the role of Notch, Wnt and sonic hedgehog (shh) signalling pathways in CSCs have been implicated [160, 161], they were not over-represented in our analysis, indicating that these pathways may be necessary for regulation of stem-like properties and/or transient proliferation but not for GNS tumourigenicity. In fact, to our knowledge no genetic studies have implicated shh pathways being activated in GBM and our mouse NS cells could grow without problem in a Notch inhibitor (personal communication with S.P.).

Since it is speculated that tumourigenesis starts with sustained differentiation and subsequent uncontrolled proliferation, we more closely looked at the enriched GO.BP “neuron differentiation” category. A network reconstructed only from the DE proteins belonging to this category was connected from the ECM to the nucleus, further supporting its relevance to

tumourigenesis. Some of the DE proteins have been implicated with glioma to varying degrees (integrins, tubulins, collagens, kinases, CDK1, EGFR, EPHB2, TNC, etc.) and others were not (myosins, CAP1, CAP2, FKBP4, LIMK1, LZTS1, NRXN1, THY1, etc.). The "cell proliferation"-related processes were not over-represented. This could be because both NS and GNSs were both cultured in proliferating conditions.

A transcription factor [116] can drive the expression of several hundreds of genes. There were 36 DE transcription factors/regulators in our data, 9 of which (FOXO3, TP53, IFI16, NFIC, PURA, YAP1, HMGA2 and PRKCB) have been implicated in glioma, while the others do not, including LZTS1 found in GO.BP neuron differentiation (Table 3-4). NFIC and IGHMBP2 have been shown to be expressed in astrocyte and neuron, respectively, possibly indicating that these mature cells could de-differentiate into GNSs and that the expression of these proteins are not silenced afterwards. This hypothesis of de-differentiation is not far-fetched, since FoxG1 has been shown to be a reprogramming factor from fibroblasts to neural progenitor cells (NPCs) capable of differentiation into neurons, astrocytes or oligodendrocytes [137] and since FOXG1 was strongly up-regulated in our GNSs on the mRNA level [77] (not identified in our proteomics data).

In the transcriptome comparison between NSs and GNSs, novel genes up-regulated in GNSs that had not been previously detected by microarray profiling of glioma tumour tissue samples were identified [77]. The comparison between this transcriptome data and our proteomics data showed a general agreement in expression direction. However, 34 gene/proteins were DE in the same direction on both levels, whereas 405 proteins were DE only on the protein level and 207 gene were DE only on the mRNA level. It is widely accepted that the expression level of transcripts do not always correlate with that of proteins, especially in higher eukaryotes such as human, and this relatively poor overlap of DE genes is due partly to posttranscriptional regulation of protein expression but perhaps also to a slightly different set of cell lines used in the two experiments. On the other hand, those that do agreen on the expression levels of both mRNA and protein are likely to have some significance to the cells.

In our secreted proteome analysis the majority of DE proteins were up-regulated (144 out of 167) (Figure 3-11). This trend was also observed in the comparison between cancer and normal cells using the AHA method [162]. On the other hand, the distribution of up- and down-regulated, secretory proteins in the total cell experiment was more or less equal (Figure 3-11). This discrepancy could be because the total cell experiment is capturing secretory proteins inside the cell and if they are more rapidly transported outside the cell, the overall abundance of these proteins in the cell at a given moment would remain the same. In both experiments the four most abundant protein classes by UniProt keywords were receptor / membrane protein, disease mutation, collagen and cell adhesion (Figure 3-11). The categories; protease and growth factor / growth factor binding,

were 6th and 8th most abundant in the secretome experiment, while they were 8th and 11th in the total cell experiment, respectively (Figure 3-11), possibly implying that these two classes in the secretome could particularly have a significant influence on the cells. Previous secretome studies have shown that IGFBP7 is a marker for GBM vessels, and that it was up-regulated in human brain endothelial cells upon GBM secretome exposure [163]. CHI3L1 is involved in the regulation of malignant transformation and local invasiveness of gliomas [164]. In our current data, both IGFBP7 and CHI3L1 were strongly expressed in both up and down directions, depending on cell llines. Thus, there was a high, inter-individual variability in what has been reported to be GBM-secretome signatures and these published results should be taken with caution.

After the global proteome characterisation, we aimed to identify surface markers that distinguish between NSs and GNSs, since proteins on the plasma membrane can be more effectively targeted by drugs than those inside the cell. Markers that are not expressed at all in NSs are particularly desirable, as normal NSs should not be killed when GNSs are treated with drugs targeting these markers. As glioblastoma stem cells are particularly resistant to radiation, identifying new drugs that can kill these cells is likely to be most important [68]. The immunocytochemistry of the five candidates (TNC, LAGLS3, THY1, CD9 and TES) revealed THY1 and CD9 being possible markers only expressed on GNSs, while TES was specific for NSs. Although THY1 is a marker for many stem cells and implicated in GO.BP neuron differentiation, its function in GNS malignancy is unclear. Because THY1 is also a marker for neurons, the possibility remains that GNSs arose from adult neurons and THY1 expression was just the remnants of it. To rule out this, we need to make sure the protein disappears after differentiating GNSs as well as to test more GNS lines.

The function of glioblastoma secretome in invasiveness of glioblastoma cells has been reported [61] and three TGF-alpha family growth factors, EGF, VEGF and PDGF, have been shown to be over-expressed in glioma-conditioned media, which increased proliferation of NSs [67]. It remains elusive, however, whether these are the only auto-/paracrine factors mediating the entire process of tumourigenesis in vivo. Our conditioned media experiment with ANS4 (mouse NSs) and IENS (moues GNSs) revealed that IENS-conditioned media only increased proliferation of IENS cells but not that of ANS4 cells (Figure 3-20). This is probably because ANS4 cells are less sensitive to the media composition. Indeed, the subsequent experiment showed that ANS4 cells did not proliferate at EGF/FGF concentration below 0.1 ng/ml, whereas IENS cells proliferated even without EGF/FGF (Figure 3-21). Since proliferation of IENS cells was inhibited in ANS4-conditioned media (Figure 3-20), ANS4 media may contain pro-differentiation factors at concentrations sufficient to suppress IENS proliferation. More experiments are necessary to establish this. Our investigation of the effect of seven candidate secreted factors TNC, MDK, CNTF, APOE, IGFBP4, CSF1 (over-expressed in GNSs) and IGFBP3 (under-expressed in GNSs)

on mouse NS and GNS cells revealed that CNTF could induce a morphological change (Figure 3-27), probably differentiation to astrocytes, as has been shown in NSs [165-167]. It is contradictory though that CNTFR was up-regulated in GNSs and Lu et al. [156] reported that CNTFR-alpha is a marker for glioma initiating cells and mutations were not common to this protein. Although CNTF did not cause any visible change to our IENS cells (data not shown), it has been reported to differentiate glioma initiating cells [156] which was supported by our colony-forming assay exhibiting a dramatic decrease in IENS colony count but not in ANS4 (Figure 3-31). None of the other six factors had any visible effect on confluency, cell count or morphology. However, in the colony-forming assay experiment, TNC, MDK and IGFBP4 might have enhanced colony formation of ANS4 cells, whereas IGFBP3 might have enhanced colony formation of IENS cells. Since IGFBP4 was up-regulated in the human GNSs and IGFBP3 was up-regulated in NSs, differential use of IGFBP3 and 4 might be associated with tumourigenesis. Our results also indicate that the colony-forming assay could be more sensitive to the media composition and more suitable for the current task than the time-lapse imaging. More replicates of the colony-forming assay are necessary for drawing a solid conclusion.

Conventional cancer cell cultures contain serum, which induces differentiation of CSCs, while blocking the cells' capacity to reconstitute the tumour in vivo [3]. In addition, repetitive passaging, which is a common practice for non-primary cancer cell lines, results in accumulation of de novo mutations of non-primary tumour origin [3]. The serum-free neurosphere culture for CSCs can circumvent these problems, however, it contains a heterogeneous mixture of self-renewing and differentiating cells. Our cell cultures are advantageous in that primary malignant glioma stem cells that can reconstitute the tumour in vivo can be directly compared to the normal, untransformed counterparts [4, 56, 58]. By performing both total cell- and secreted proteome analyses using our new secretome technique [162], we thus identified many proteins and processes both with and without prior associations with glioma biology. Several candidate proteins for surface markers and for tumourigenic transformation of NSs were further experimentally tested with some positive results. More experiments are, however, necessary to solidify these findings.

Chapter 4

4. Conclusion

In this doctoral study I carried out two projects. The 1st one was about the exploration of the use of isotope patterns in MS1 spectra for peptide identification. We demonstrated that the RIA error is 4-5%, and that this is only modestly influenced by spectral intensity, resolution and the number of MS1 scans. The current RIA accuracy has limited discriminatory power at a proteome-wide scale. At the same time, the analysis was hampered by the difficulty in calculating FDRs, particularly in constructing proper decoy databases that are similar in size as the target database, yet different in molecular composition for all peptides considered. Alternative strategies to calculate FDRs will be required to address this issue for complex proteomes. Regardless, the utility of RIA may become relevant with future instrument developments, considering that even a relatively modest decrease in RIA error down to <1% strongly improves discriminatory power. Alternatively, at increased mass accuracy even current RIA accuracy levels may be sufficient to fit isotope patterns as a constraint in the peptide identification process as a parameter that comes for free in any MS-based proteomic experiment.

In the 2nd project we conducted comparative proteomics analysis of karyotypically normal, untransformed neural stem cells [1] and malignant, adult glioma neural stem cells (GNSs). This resulted in a total of ~7500 and ~2000 quantified proteins and 446 differentially expressed (DE) proteins (152 up-regulation, 294 down-regulation in GNSs) and 167 DE proteins (144 up-regulation, 23 down-regulation), respectively. Many proteins and processes without prior associations with glioma biology as well as those with known associations were found. After data analyses, several candidate proteins for surface markers that could distinguish between NSs and GNSs were experimentally validated using immunocytochemistry. Then, candidate GNS-secreted factors that might mediate tumourigenic transformation of NSs were evaluated using time-lapse imaging and colony-forming assays. Both of these experiments produced some positive results, demonstrating the power of proteomics. More experiments are, however, necessary to solidify these findings. The next goals would be 1) to find drugs that could specifically kill only GNSs by targeting these surface markers, and 2) molecularly characterize the complete process of tumourigenesis from NSs to GNSs ex vivo.

BIBLIOGRAPHY

1. Taupin, P., et al., *FGF-2-responsive neural stem cell proliferation requires CCg, a novel autocrine/paracrine cofactor.* Neuron, 2000. **28**(2): p. 385-97.
2. Michalski, A., J. Cox, and M. Mann, *More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS.* J Proteome Res, 2011. **10**(4): p. 1785-93.
3. Lee, J., et al., *Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines.* Cancer Cell, 2006. **9**(5): p. 391-403.
4. Pollard, S.M., et al., *Glioma stem cell lines expanded in adherent culture have tumor-specific phenotypes and are suitable for chemical and genetic screens.* Cell Stem Cell, 2009. **4**(6): p. 568-80.
5. Phillips, H.S., et al., *Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis.* Cancer Cell, 2006. **9**(3): p. 157-73.
6. Domon, B. and R. Aebersold, *Options and considerations when selecting a quantitative proteomics strategy.* Nat Biotechnol, 2010. **28**(7): p. 710-21.
7. Sadygov, R.G., D. Cociorva, and J.R. Yates, 3rd, *Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book.* Nat Methods, 2004. **1**(3): p. 195-202.
8. Yates, J.R., 3rd, et al., *Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database.* Anal Chem, 1995. **67**(8): p. 1426-36.
9. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data.* Electrophoresis, 1999. **20**(18): p. 3551-67.
10. Geer, L.Y., et al., *Open mass spectrometry search algorithm.* J Proteome Res, 2004. **3**(5): p. 958-64.
11. Craig, R. and R.C. Beavis, *TANDEM: matching proteins with tandem mass spectra.* Bioinformatics, 2004. **20**(9): p. 1466-7.
12. Cox, J., et al., *Andromeda: a peptide search engine integrated into the MaxQuant environment.* J Proteome Res, 2011. **10**(4): p. 1794-805.
13. Klammer, A.A., et al., *Improving tandem mass spectrum identification using peptide retention time prediction across diverse chromatography conditions.* Anal Chem, 2007. **79**(16): p. 6111-8.
14. Krokhin, O.V., *Sequence-specific retention calculator. Algorithm for peptide retention prediction in ion-pair RP-HPLC: application to 300- and 100-A pore size C18 sorbents.* Anal Chem, 2006. **78**(22): p. 7785-95.
15. Moruz, L., D. Tomazela, and L. Kall, *Training, selection, and robust calibration of retention time models for targeted proteomics.* J Proteome Res, 2010. **9**(10): p. 5209-16.
16. Petritis, K., et al., *Improved peptide elution time prediction for reversed-phase liquid chromatography-MS by incorporating peptide sequence information.* Anal Chem, 2006. **78**(14): p. 5026-39.
17. Pfeifer, N., et al., *Statistical learning of peptide retention behavior in chromatographic separations: a new kernel-based approach for computational proteomics.* BMC Bioinformatics, 2007. **8**: p. 468.
18. Krijgsveld, J., et al., *In-gel isoelectric focusing of peptides as a tool for improved protein identification.* J Proteome Res, 2006. **5**(7): p. 1721-30.
19. Mann, M. and N.L. Kelleher, *Precision proteomics: the case for high resolution and high mass accuracy.* Proc Natl Acad Sci U S A, 2008. **105**(47): p. 18132-8.
20. Senko, M.W., S.C. Beu, and F.W. McLafferty, *Automated assignment of charge states from resolved isotopic peaks for multiply charged ions.* Journal of the American Society for Mass Spectrometry, 1995. **6**(1): p. 52–56.

21. Senko, M.W., S.C. Beu, and F.W. McLafferty, *Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions.* Journal of the American Society for Mass Spectrometry, 1995. **6**(4): p. 229–233.
22. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.* Nat Biotechnol, 2008. **26**(12): p. 1367-72.
23. Zhang, Y., et al., *Proteome scale turnover analysis in live animals using stable isotope metabolic labeling.* Anal Chem, 2011. **83**(5): p. 1665-72.
24. Polacco, B.J., et al., *Discovering mercury protein modifications in whole proteomes using natural isotope distributions observed in liquid chromatography-tandem mass spectrometry.* Mol Cell Proteomics, 2011. **10**(8): p. M110 004853.
25. Bocker, S., et al., *SIRIUS: decomposing isotope patterns for metabolite identification.* Bioinformatics, 2009. **25**(2): p. 218-24.
26. Erve, J.C., et al., *Spectral accuracy of molecular ions in an LTQ/Orbitrap mass spectrometer and implications for elemental composition determination.* J Am Soc Mass Spectrom, 2009. **20**(11): p. 2058-69.
27. Kind, T. and O. Fiehn, *Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm.* BMC Bioinformatics, 2006. **7**: p. 234.
28. Pluskal, T., T. Uehara, and M. Yanagida, *Highly accurate chemical formula prediction tool utilizing high-resolution mass spectra, MS/MS fragmentation, heuristic rules, and isotope pattern matching.* Anal Chem, 2012. **84**(10): p. 4396-403.
29. Weber, R.J., et al., *Characterization of isotopic abundance measurements in high resolution FT-ICR and Orbitrap mass spectra for improved confidence of metabolite identification.* Anal Chem, 2011. **83**(10): p. 3737-43.
30. Xu, Y., et al., *Evaluation of accurate mass and relative isotopic abundance measurements in the LTQ-orbitrap mass spectrometer for further metabolomics database building.* Anal Chem, 2010. **82**(13): p. 5490-501.
31. Rappsilber, J., M. Mann, and Y. Ishihama, *Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips.* Nat Protoc, 2007. **2**(8): p. 1896-906.
32. Elias, J.E. and S.P. Gygi, *Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry.* Nat Methods, 2007. **4**(3): p. 207-14.
33. Easterling, M.L., et al., *Isotope Beating Effects in the Analysis of Polymer Distributions by Fourier Transform Mass Spectrometry.* Journal of the American Society for Mass Spectrometry 1999. **10**(11): p. 1074-1082.
34. Miura, D., et al., *A strategy for the determination of the elemental composition by fourier transform ion cyclotron resonance mass spectrometry based on isotopic peak ratios.* Anal Chem, 2010. **82**(13): p. 5887-91.
35. Shi, S.D., C.L. Hendrickson, and A.G. Marshall, *Counting individual sulfur atoms in a protein by ultrahigh-resolution Fourier transform ion cyclotron resonance mass spectrometry: experimental resolution of isotopic fine structure in proteins.* Proc Natl Acad Sci U S A, 1998. **95**(20): p. 11532-7.
36. Miladinovic, S.M., et al., *On the utility of isotopic fine structure mass spectrometry in protein identification.* Anal Chem, 2012. **84**(9): p. 4042-51.
37. Knudson, A.G., Jr., *Mutation and cancer: statistical study of retinoblastoma.* Proc Natl Acad Sci U S A, 1971. **68**(4): p. 820-3.
38. Feinberg, A.P., R. Ohlsson, and S. Henikoff, *The epigenetic progenitor origin of human cancer.* Nat Rev Genet, 2006. **7**(1): p. 21-33.
39. Teng, I.W., et al., *Targeted methylation of two tumor suppressor genes is sufficient to transform mesenchymal stem cells into cancer stem/initiating cells.* Cancer Res, 2011. **71**(13): p. 4653-63.

40. Kleihues, P. and L.H. Sobin, *World Health Organization classification of tumors.* Cancer, 2000. **88**(12): p. 2887.
41. Stupp, R. and F. Roila, *Malignant glioma: ESMO clinical recommendations for diagnosis, treatment and follow-up.* Ann Oncol, 2009. **20 Suppl 4**: p. 126-8.
42. Furnari, F.B., et al., *Malignant astrocytic glioma: genetics, biology, and paths to treatment.* Genes Dev, 2007. **21**(21): p. 2683-710.
43. *Comprehensive genomic characterization defines human glioblastoma genes and core pathways.* Nature, 2008. **455**(7216): p. 1061-8.
44. Lapidot, T., et al., *A cell initiating human acute myeloid leukaemia after transplantation into SCID mice.* Nature, 1994. **367**(6464): p. 645-8.
45. Hemmati, H.D., et al., *Cancerous stem cells can arise from pediatric brain tumors.* Proc Natl Acad Sci U S A, 2003. **100**(25): p. 15178-83.
46. Singh, S.K., et al., *Identification of a cancer stem cell in human brain tumors.* Cancer Res, 2003. **63**(18): p. 5821-8.
47. Singh, S.K., et al., *Identification of human brain tumour initiating cells.* Nature, 2004. **432**(7015): p. 396-401.
48. Stiles, C.D. and D.H. Rowitch, *Glioma stem cells: a midterm exam.* Neuron, 2008. **58**(6): p. 832-46.
49. Shackleton, M., et al., *Heterogeneity in cancer: cancer stem cells versus clonal evolution.* Cell, 2009. **138**(5): p. 822-9.
50. Galli, R., et al., *Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma.* Cancer Res, 2004. **64**(19): p. 7011-21.
51. Ignatova, T.N., et al., *Human cortical glial tumors contain neural stem-like cells expressing astroglial and neuronal markers in vitro.* Glia, 2002. **39**(3): p. 193-206.
52. Yuan, X., et al., *Isolation of cancer stem cells from adult glioblastoma multiforme.* Oncogene, 2004. **23**(58): p. 9392-400.
53. Suslov, O.N., et al., *Neural stem cell heterogeneity demonstrated by molecular phenotyping of clonal neurospheres.* Proc Natl Acad Sci U S A, 2002. **99**(22): p. 14506-11.
54. Reynolds, B.A. and R.L. Rietze, *Neural stem cells and neurospheres--re-evaluating the relationship.* Nat Methods, 2005. **2**(5): p. 333-6.
55. Singec, I., et al., *Defining the actual sensitivity and specificity of the neurosphere assay in stem cell biology.* Nat Methods, 2006. **3**(10): p. 801-6.
56. Conti, L., et al., *Niche-independent symmetrical self-renewal of a mammalian tissue stem cell.* PLoS Biol, 2005. **3**(9): p. e283.
57. Pollard, S.M., et al., *Adherent neural stem (NS) cells from fetal and adult forebrain.* Cereb Cortex, 2006. **16 Suppl 1**: p. i112-20.
58. Sun, Y., et al., *Long-term tripotent differentiation capacity of human neural stem (NS) cells in adherent culture.* Mol Cell Neurosci, 2008. **38**(2): p. 245-58.
59. Moore, K.A. and I.R. Lemischka, *Stem cells and their niches.* Science, 2006. **311**(5769): p. 1880-5.
60. Gilbertson, R.J. and J.N. Rich, *Making a tumour's bed: glioblastoma stem cells and the vascular niche.* Nat Rev Cancer, 2007. **7**(10): p. 733-6.
61. Formolo, C.A., et al., *Secretome signature of invasive glioblastoma multiforme.* J Proteome Res, 2011. **10**(7): p. 3149-59.
62. Plate, K.H., et al., *Vascular endothelial growth factor is a potential tumour angiogenesis factor in human gliomas in vivo.* Nature, 1992. **359**(6398): p. 845-8.
63. Ricci-Vitiani, L., et al., *Tumour vascularization via endothelial differentiation of glioblastoma stem-like cells.* Nature, 2010. **468**(7325): p. 824-8.
64. Wang, Z.H., Y.X. Xue, and Y.H. Liu, *The modulation of protein kinase A and heat shock protein 70 is involved in the reversible increase of blood-brain tumor barrier permeability induced by papaverine.* Brain Res Bull, 2010. **83**(6): p. 367-73.
65. Charles, N., et al., *Perivascular nitric oxide activates notch signaling and promotes stem-*

like character in PDGF-induced glioma cells. Cell Stem Cell, 2010. **6**(2): p. 141-52.
66. Eyler, C.E., et al., *Glioma stem cell proliferation and tumor growth are promoted by nitric oxide synthase-2.* Cell, 2011. **146**(1): p. 53-66.
67. Venugopal, C., et al., *GBM secretome induces transient transformation of human neural precursor cells.* J Neurooncol, 2012. **109**(3): p. 457-66.
68. Bao, S., et al., *Glioma stem cells promote radioresistance by preferential activation of the DNA damage response.* Nature, 2006. **444**(7120): p. 756-60.
69. Boersema, P.J., et al., *Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics.* Nat Protoc, 2009. **4**(4): p. 484-94.
70. Smyth, G.K., *Linear models and empirical bayes methods for assessing differential expression in microarray experiments.* Stat Appl Genet Mol Biol, 2004. **3**: p. Article3.
71. Thomas, P.D., et al., *PANTHER: a library of protein families and subfamilies indexed by function.* Genome Res, 2003. **13**(9): p. 2129-41.
72. Ruepp, A., et al., *CORUM: the comprehensive resource of mammalian protein complexes--2009.* Nucleic Acids Res, 2010. **38**(Database issue): p. D497-501.
73. Matthews, L., et al., *Reactome knowledgebase of human biological pathways and processes.* Nucleic Acids Res, 2009. **37**(Database issue): p. D619-22.
74. Ben-Porath, I., et al., *An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors.* Nat Genet, 2008. **40**(5): p. 499-507.
75. Tarca, A.L., et al., *A novel signaling pathway impact analysis.* Bioinformatics, 2009. **25**(1): p. 75-82.
76. Liberzon, A., et al., *Molecular signatures database (MSigDB) 3.0.* Bioinformatics, 2011. **27**(12): p. 1739-40.
77. Engstrom, P.G., et al., *Digital transcriptome profiling of normal and glioblastoma-derived neural stem cells identifies genes associated with patient survival.* Genome Med, 2012. **4**(10): p. 76.
78. Gilbert, C.A. and A.H. Ross, *Cancer stem cells: cell culture, markers, and targets for new therapies.* J Cell Biochem, 2009. **108**(5): p. 1031-8.
79. Visvader, J.E. and G.J. Lindeman, *Cancer stem cells: current status and evolving complexities.* Cell Stem Cell, 2012. **10**(6): p. 717-28.
80. Beckervordersandforth, R., et al., *In vivo fate mapping and expression analysis reveals molecular hallmarks of prospectively isolated adult neural stem cells.* Cell Stem Cell, 2010. **7**(6): p. 744-58.
81. Fukushi, J., I.T. Makagiansar, and W.B. Stallcup, *NG2 proteoglycan promotes endothelial cell motility and angiogenesis via engagement of galectin-3 and alpha3beta1 integrin.* Mol Biol Cell, 2004. **15**(8): p. 3580-90.
82. Wei, J., et al., *Glioma-associated cancer-initiating cells induce immunosuppression.* Clin Cancer Res, 2010. **16**(2): p. 461-73.
83. Hellstern, S., et al., *Functional studies on recombinant domains of Mac-2-binding protein.* J Biol Chem, 2002. **277**(18): p. 15690-6.
84. Tinari, N., et al., *Glycoprotein 90K/MAC-2BP interacts with galectin-1 and mediates galectin-1-induced cell aggregation.* Int J Cancer, 2001. **91**(2): p. 167-72.
85. Cheng, L., et al., *L1CAM regulates DNA damage checkpoint response of glioblastoma stem cells through NBS1.* EMBO J, 2011. **30**(5): p. 800-13.
86. Miraglia, S., et al., *A novel five-transmembrane hematopoietic stem cell antigen: isolation, characterization, and molecular cloning.* Blood, 1997. **90**(12): p. 5013-21.
87. Weigmann, A., et al., *Prominin, a novel microvilli-specific polytopic membrane protein of the apical surface of epithelial cells, is targeted to plasmalemmal protrusions of non-epithelial cells.* Proc Natl Acad Sci U S A, 1997. **94**(23): p. 12425-30.
88. Yin, A.H., et al., *AC133, a novel marker for human hematopoietic stem and progenitor cells.* Blood, 1997. **90**(12): p. 5002-12.
89. Corbeil, D., et al., *The human AC133 hematopoietic stem cell antigen is also expressed in*

epithelial cells and targeted to plasma membrane protrusions. J Biol Chem, 2000. **275**(8): p. 5512-20.
90. Uchida, N., et al., *Direct isolation of human central nervous system stem cells.* Proc Natl Acad Sci U S A, 2000. **97**(26): p. 14720-5.
91. Bao, S., et al., *Targeting cancer stem cells through L1CAM suppresses glioma growth.* Cancer Res, 2008. **68**(15): p. 6043-8.
92. Beier, D., et al., *CD133(+) and CD133(-) glioblastoma-derived cancer stem cells show differential growth characteristics and molecular profiles.* Cancer Res, 2007. **67**(9): p. 4010-5.
93. Fortunel, N.O., et al., *Comment on " 'Stemness': transcriptional profiling of embryonic and adult stem cells" and "a stem cell molecular signature".* Science, 2003. **302**(5644): p. 393; author reply 393.
94. Lathia, J.D., et al., *Integrin alpha 6 regulates glioblastoma stem cells.* Cell Stem Cell, 2010. **6**(5): p. 421-32.
95. Sladek, N.E., *Human aldehyde dehydrogenases: potential pathological, pharmacological, and toxicological impact.* J Biochem Mol Toxicol, 2003. **17**(1): p. 7-23.
96. Ma, X., et al., *The differentiation of hepatocyte-like cells from monkey embryonic stem cells.* Cloning Stem Cells, 2008. **10**(4): p. 485-93.
97. Ran, D., et al., *Aldehyde dehydrogenase activity among primary leukemia cells is associated with stem cell features and correlates with adverse clinical outcomes.* Exp Hematol, 2009. **37**(12): p. 1423-34.
98. Jiang, F., et al., *Aldehyde dehydrogenase 1 is a tumor stem cell-associated marker in lung cancer.* Mol Cancer Res, 2009. **7**(3): p. 330-8.
99. Chen, Y.C., et al., *Aldehyde dehydrogenase 1 is a putative marker for cancer stem cells in head and neck squamous cancer.* Biochem Biophys Res Commun, 2009. **385**(3): p. 307-13.
100. Rasheed, Z.A., et al., *Prognostic significance of tumorigenic cells with mesenchymal features in pancreatic adenocarcinoma.* J Natl Cancer Inst, 2010. **102**(5): p. 340-51.
101. Silva, I.A., et al., *Aldehyde dehydrogenase in combination with CD133 defines angiogenic ovarian cancer stem cells that portend poor patient survival.* Cancer Res, 2011. **71**(11): p. 3991-4001.
102. Chute, J.P., et al., *Inhibition of aldehyde dehydrogenase and retinoid signaling induces the expansion of human hematopoietic stem cells.* Proc Natl Acad Sci U S A, 2006. **103**(31): p. 11707-12.
103. Alison, M.R., et al., *Finding cancer stem cells: are aldehyde dehydrogenases fit for purpose?* J Pathol, 2010. **222**(4): p. 335-44.
104. Hynes, R.O., *Integrins: bidirectional, allosteric signaling machines.* Cell, 2002. **110**(6): p. 673-87.
105. Skog, J., et al., *Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers.* Nat Cell Biol, 2008. **10**(12): p. 1470-6.
106. Goel, H.L., et al., *Neuropilin-2 regulates alpha6beta1 integrin in the formation of focal adhesions and signaling.* J Cell Sci, 2012. **125**(Pt 2): p. 497-506.
107. Véronique, M.-S., *The Stem Cell Niche: The Black Master of Cancer.* Cancer Stem Cells Theories and Practice, 2011.
108. Joester, A. and A. Faissner, *The structure and function of tenascins in the nervous system.* Matrix Biol, 2001. **20**(1): p. 13-22.
109. Garcion, E., et al., *Generation of an environmental niche for neural stem cell development by the extracellular matrix molecule tenascin C.* Development, 2004. **131**(14): p. 3423-32.
110. von Holst, A., *Tenascin C in stem cell niches: redundant, permissive or instructive?* Cells Tissues Organs, 2008. **188**(1-2): p. 170-7.
111. Garcion, E., A. Faissner, and C. ffrench-Constant, *Knockout mice reveal a contribution of the extracellular matrix molecule tenascin-C to neural precursor proliferation and migration.* Development, 2001. **128**(13): p. 2485-96.

112. Garwood, J., et al., *The extracellular matrix glycoprotein Tenascin-C is expressed by oligodendrocyte precursor cells and required for the regulation of maturation rate, survival and responsiveness to platelet-derived growth factor.* Eur J Neurosci, 2004. **20**(10): p. 2524-40.
113. Craig, W., et al., *Expression of Thy-1 on human hematopoietic progenitor cells.* J Exp Med, 1993. **177**(5): p. 1331-42.
114. Haeryfar, S.M. and D.W. Hoskin, *Thy-1: more than a mouse pan-T cell marker.* J Immunol, 2004. **173**(6): p. 3581-8.
115. Dominici, M., et al., *Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement.* Cytotherapy, 2006. **8**(4): p. 315-7.
116. Stadtfeld, M., et al., *Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse.* Cell Stem Cell, 2008. **2**(3): p. 230-40.
117. Yang, Z.F., et al., *Significance of CD90+ cancer stem cells in human liver cancer.* Cancer Cell, 2008. **13**(2): p. 153-66.
118. Liu, G., et al., *Analysis of gene expression and chemoresistance of CD133+ cancer stem cells in glioblastoma.* Mol Cancer, 2006. **5**: p. 67.
119. Ishii, H., et al., *FEZ1/LZTS1 gene at 8p22 suppresses cancer cell growth and regulates mitosis.* Proc Natl Acad Sci U S A, 2001. **98**(18): p. 10374-9.
120. Merlos-Suarez, A., et al., *The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse.* Cell Stem Cell, 2011. **8**(5): p. 511-24.
121. Jeannet, R., et al., *Alcam regulates long-term hematopoietic stem cell engraftment and self-renewal.* Stem Cells, 2013. **31**(3): p. 560-71.
122. Tanaka, H., et al., *Molecular cloning and expression of a novel adhesion molecule, SC1.* Neuron, 1991. **7**(4): p. 535-45.
123. DeBernardo, A.P. and S. Chang, *Heterophilic interactions of DM-GRASP: GRASP-NgCAM interactions involved in neurite extension.* J Cell Biol, 1996. **133**(3): p. 657-66.
124. van Kempen, L.C., et al., *Molecular basis for the homophilic activated leukocyte cell adhesion molecule (ALCAM)-ALCAM interaction.* J Biol Chem, 2001. **276**(28): p. 25783-90.
125. Buhusi, M., et al., *ALCAM regulates mediolateral retinotopic mapping in the superior colliculus.* J Neurosci, 2009. **29**(50): p. 15630-41.
126. Heffron, D.S. and J.A. Golden, *DM-GRASP is necessary for nonradial cell migration during chick diencephalic development.* J Neurosci, 2000. **20**(6): p. 2287-94.
127. Stephan, J.P., et al., *Distribution and function of the adhesion molecule BEN during rat development.* Dev Biol, 1999. **212**(2): p. 264-77.
128. Brunet, A., et al., *Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor.* Cell, 1999. **96**(6): p. 857-68.
129. Brunet, A., et al., *Protein kinase SGK mediates survival signals by phosphorylating the forkhead transcription factor FKHRL1 (FOXO3a).* Mol Cell Biol, 2001. **21**(3): p. 952-65.
130. Lehtinen, M.K., et al., *A conserved MST-FOXO signaling pathway mediates oxidative-stress responses and extends life span.* Cell, 2006. **125**(5): p. 987-1001.
131. Hagenbuchner, J., et al., *FOXO3-induced reactive oxygen species are regulated by BCL2L11 (Bim) and SESN3.* J Cell Sci, 2012. **125**(Pt 5): p. 1191-203.
132. Seoane, J., et al., *Integration of Smad and forkhead pathways in the control of neuroepithelial and glioblastoma cell proliferation.* Cell, 2004. **117**(2): p. 211-23.
133. Renault, V.M., et al., *The pro-longevity gene FoxO3 is a direct target of the p53 tumor suppressor.* Oncogene, 2011. **30**(29): p. 3207-21.
134. Dou, C.L., S. Li, and E. Lai, *Dual role of brain factor-1 in regulating growth and patterning of the cerebral hemispheres.* Cereb Cortex, 1999. **9**(6): p. 543-50.
135. Shen, L., et al., *FoxG1 haploinsufficiency results in impaired neurogenesis in the postnatal hippocampus and contextual memory deficits.* Hippocampus, 2006. **16**(10): p. 875-90.

136. Fasano, C.A., et al., *Bmi-1 cooperates with Foxg1 to maintain neural stem cell self-renewal in the forebrain.* Genes Dev, 2009. **23**(5): p. 561-74.
137. Lujan, E., et al., *Direct conversion of mouse fibroblasts to self-renewing, tripotent neural precursor cells.* Proc Natl Acad Sci U S A, 2012. **109**(7): p. 2527-32.
138. Dobreva, G., J. Dambacher, and R. Grosschedl, *SUMO modification of a novel MAR-binding protein, SATB2, modulates immunoglobulin mu gene expression.* Genes Dev, 2003. **17**(24): p. 3048-61.
139. Akai, T., et al., *High mobility group I-C protein in astrocytoma and glioblastoma.* Pathol Res Pract, 2004. **200**(9): p. 619-24.
140. Liu, Y., et al., *Polymorphisms of LIG4, BTBD2, HMGA2, and RTEL1 genes involved in the double-strand break repair pathway predict glioblastoma survival.* J Clin Oncol, 2010. **28**(14): p. 2467-74.
141. Grohmann, M., et al., *Characterization of differentiated subcutaneous and visceral adipose tissue from children: the influences of TNF-alpha and IGF-I.* J Lipid Res, 2005. **46**(1): p. 93-103.
142. Guenther, U.P., et al., *IGHMBP2 is a ribosome-associated helicase inactive in the neuromuscular disorder distal SMA type 1 (DSMA1).* Hum Mol Genet, 2009. **18**(7): p. 1288-300.
143. Lin, L., et al., *STAT3 signaling pathway is necessary for cell survival and tumorsphere forming capacity in ALDH(+)/CD133(+) stem cell-like human colon cancer cells.* Biochem Biophys Res Commun, 2011. **416**(3-4): p. 246-51.
144. Guryanova, O.A., et al., *Nonreceptor tyrosine kinase BMX maintains self-renewal and tumorigenic potential of glioblastoma stem cells by activating STAT3.* Cancer Cell, 2011. **19**(4): p. 498-511.
145. Thirant, C., et al., *Differential proteomic analysis of human glioblastoma and neural stem cells reveals HDGF as a novel angiogenic secreted factor.* Stem Cells, 2012. **30**(5): p. 845-53.
146. D'Andrea, F.P., et al., *Cancer stem cell overexpression of nicotinamide N-methyltransferase enhances cellular radiation resistance.* Radiother Oncol, 2011. **99**(3): p. 373-8.
147. Ikeyama, S., et al., *Suppression of cell motility and metastasis by transfection with human motility-related protein (MRP-1/CD9) DNA.* J Exp Med, 1993. **177**(5): p. 1231-7.
148. Prasad, P.D., J.A. Stanton, and S.J. Assinder, *Expression of the actin-associated protein transgelin (SM22) is decreased in prostate cancer.* Cell Tissue Res, 2010. **339**(2): p. 337-47.
149. Hasegawa, A., et al., *Regulation of glial development by cystatin C.* J Neurochem, 2007. **100**(1): p. 12-22.
150. Kato, T., et al., *A neurosphere-derived factor, cystatin C, supports differentiation of ES cells into neural stem cells.* Proc Natl Acad Sci U S A, 2006. **103**(15): p. 6019-24.
151. de Azevedo-Pereira, R.L., et al., *Cysteine proteases in differentiation of embryonic stem cells into neural cells.* Stem Cells Dev, 2011. **20**(11): p. 1859-72.
152. Konduri, S.D., et al., *Modulation of cystatin C expression impairs the invasive and tumorigenic potential of human glioblastoma cells.* Oncogene, 2002. **21**(57): p. 8705-12.
153. Nakabayashi, H., M. Hara, and K. Shimuzu, *Clinicopathologic significance of cystatin C expression in gliomas.* Hum Pathol, 2005. **36**(9): p. 1008-15.
154. von Holst, A., et al., *Neural stem/progenitor cells express 20 tenascin C isoforms that are differentially regulated by Pax6.* J Biol Chem, 2007. **282**(12): p. 9172-81.
155. He, J., et al., *CD90 is identified as a candidate marker for cancer stem cells in primary high-grade gliomas using tissue microarrays.* Mol Cell Proteomics, 2012. **11**(6): p. M111 010744.
156. Lu, J., et al., *CNTF receptor subunit alpha as a marker for glioma tumor-initiating cells and tumor grade: laboratory investigation.* J Neurosurg, 2012. **117**(6): p. 1022-31.
157. Beck, B., et al., *A vascular niche and a VEGF-Nrp1 loop regulate the initiation and stemness of skin tumours.* Nature, 2011. **478**(7369): p. 399-403.

158. Hamerlik, P., et al., *Autocrine VEGF-VEGFR2-Neuropilin-1 signaling promotes glioma stem-like cell viability and tumor growth.* J Exp Med, 2012. **209**(3): p. 507-20.
159. Yao, X., et al., *Vascular Endothelial Growth Factor Receptor 2 (VEGFR-2) Plays a Key Role in Vasculogenic Mimicry Formation, Neovascularization and Tumor Initiation by Glioma Stem-like Cells.* PLoS One, 2013. **8**(3): p. e57188.
160. Cerdan, C. and M. Bhatia, *Novel roles for Notch, Wnt and Hedgehog in hematopoesis derived from human pluripotent stem cells.* Int J Dev Biol, 2010. **54**(6-7): p. 955-63.
161. Takebe, N., et al., *Targeting cancer stem cells by inhibiting Wnt, Notch, and Hedgehog pathways.* Nat Rev Clin Oncol, 2011. **8**(2): p. 97-106.
162. Eichelbaum, K., et al., *Selective enrichment of newly synthesized proteins for quantitative secretome analysis.* Nat Biotechnol, 2012. **30**(10): p. 984-90.
163. Pen, A., et al., *Glioblastoma-secreted factors induce IGFBP7 and angiogenesis by modulating Smad-2-dependent TGF-beta signaling.* Oncogene, 2008. **27**(54): p. 6834-44.
164. Ku, B.M., et al., *CHI3L1 (YKL-40) is expressed in human gliomas and regulates the invasion, growth and survival of glioma cells.* Int J Cancer, 2011. **128**(6): p. 1316-26.
165. Hughes, S.M., et al., *Ciliary neurotrophic factor induces type-2 astrocyte differentiation in culture.* Nature, 1988. **335**(6185): p. 70-3.
166. Johe, K.K., et al., *Single factors direct the differentiation of stem cells from the fetal and adult central nervous system.* Genes Dev, 1996. **10**(24): p. 3129-40.
167. Hermanson, O., K. Jepsen, and M.G. Rosenfeld, *N-CoR controls differentiation of neural stem cells into astrocytes.* Nature, 2002. **419**(6910): p. 934-9.

Acknowledgement

I would like to acknowledge all the people who helped me in some way or another during my PhD. First and foremost, I would like to thank my supervior Jeroen for hist continuous support throughout my study. In the similar line I am thankful for all the current lab members, Chris, Daniel, Ita, Jenny, Mandy, Sina, Sophia and Tim, and the former lab members Katrin, Markus, Sergey and Sonja. My thank also goes to the EMBL proteomics core facility members, Joanna, Kristina and Stefan, and former Gavin's lab member, Jan, and current Gavin's lab member, Marco. I have a lot of thank to Bernd Fischer in Wolfgang Huber's lab for his valuable help with statistics and R. Outside Heidelberg, Paul Bertone (EBI) helped initiate my 2nd project as well as gave me valuable advice during the project. Pär Engström in Bertone's lab also provided me with the transcriptome data, scripts for the analyses and some hints for the project direction. I would like to thank a lot our principal collaborator for my 2nd project, Steven Pollard (UCL), for giving me important biological insights and letting me do experiments in his lab. In the same line, I thank Christine Ender in Pollard's lab for performing experiments for me and giving me useful advice. All the Pollard's lab members were very nice to me during my stay there. Finally, I would like to acknowledge my thesis advisory comittee members, Paul Bertone, Bruce Edgar (DKFZ) and Lars Steinmetz.