Multivariate analyses of proteomic and metabolomic patterns in brain tumors

Carl Wibom

Departments of Radiation Sciences, Oncology
Pharmacology and Clinical Neuroscience, Neurosurgery
Chemistry
Umeå University, Sweden
2009

Copyright © Carl Wibom
ISBN: 978-91-7264-828-9
ISSN: 0346-6612-1281
Front cover: Ida Holmgren, Print & Media
Printed by: Print & Media, Umeå University:2006923
Umeå, Sweden 2009

To my family

Contents

1. LIST OF PAPERS 7

2. ABSTRACT 8

3. POPULÄRVETENSKAPLIG SAMMANFATTNING PÅ SVENSKA 10

4. ABBREVIATIONS AND NOTATIONS 13

4.1 Abbreviations 13

4.2 Notations 14

5. INTRODUCTION 15

5.1 Glioblastoma 15

5.1.1 Primary and secondary glioblastoma *17*

5.1.2 Tumor stem cells *18*

5.1.3 Neovascularization *18*

5.1.4 Invasiveness *19*

5.1.5 Glioblastoma treatment *20*

5.2 Meningioma 21

5.2.1 NF2-gene *23*

5.2.2 Vascular endothelial growth factor *23*

5.2.3 Sex hormones *24*

5.2.4 Meningioma treatment *24*

5.3 Proteomics & Metabolomics 25

5.4 Biomarkers 27

5.4.1 Brain tumor biomarkers *28*

6. AIMS OF THE STUDY 30

7. METHODS 31

7.1 Samples and sampling 31

7.1.1 Animal model *31*

7.1.2 Meningioma samples *31*

7.1.3 Microdialysis samples *32*

7.2 Instrumentation 34

7.2.1 Mass spectrometry *34*

7.2.2 Surface Enhanced Laser Desorption/Ionization *36*

7.2.3 Gas Chromatography *38*

7.3 Data analysis using multivariate projection methods 40

7.3.1 Principal Component Analysis *42*

7.3.2 Partial Least Squares *44*

7.3.3 Orthogonal Partial Least Squares *46*

7.3.4 Validation *47*

7.4 Analysis of SELDI-MS spectra 49
7.4.1 Data pre-treatment *50*
7.4.2 Detecting features of interest *50*
7.4.3 Analysis and validation of extracted features *52*
7.5 Analysis of GC-MS data 54
7.5.1 Deconvolution and data pre-treatment *55*

8. RESULTS AND DISCUSSION 59
8.1 Sample characterization by means of non-targeted screening methods 59
8.2 Analysis of proteomic and metabolomic patterns 61
8.2.1 Protein expression data generated by SELDI-MS *61*
8.2.2 Metabolic data generated by GC-MS *62*
8.3 Treatment-induced changes in protein and metabolite composition 63
8.3.1 Radiotherapy treatment *63*
8.3.2 Experimental angiogenesis treatment *66*
8.4 Classification of benign meningioma growth behavior 68

9. CONCLUSIONS 69

10. REFERENCES 71

11. ACKNOWLEDGEMENTS 85

1. List of papers

This thesis is based on the following papers, referred to in the text by their roman numerals in bold font:

I. **Wibom C**, Pettersson F, Sjöström M, Henriksson R, Johansson M, Bergenheim AT (2006). Protein expression in experimental malignant glioma varies over time and is altered by radiotherapy treatment. *British Journal of Cancer, 94, 1853-63.*

II. **Wibom C**, Surowiec I, Mörén L, Bergström P, Johansson M, Antti H, Bergenheim AT. Metabolomic patterns in glioblastoma and changes during radiotherapy – a clinical microdialysis study. *Manuscript.*

III. **Wibom C**, Sandström M, Henriksson R, Antti H, Johansson M, Bergenheim AT. Vandetanib alters protein pattern in malignant glioma and normal brain in the BT4C rat glioma model. *Submitted manuscript.*

IV. **Wibom C**, Mörén L, Aarhus M, Knappskog PM, Lund-Johansen M, Antti H, Bergenheim AT (2009). Proteomic profiles differ between bone invasive and noninvasive benign meningiomas of fibrous and meningothelial subtype. *Journal of Neuro-oncology, 94, 321-331.*

Paper **IV** was reprinted with kind permission from Springer Science+Business Media. For paper **I** no permission was needed.

2. Abstract

Glioblastoma multiforme (GBM) is the most common primary brain tumor. Given the current standard of care, the prognosis for patients diagnosed with this disease is still poor. There consequently exists a need to improve current treatments, as well as to develop new ones. Many obstacles however need to be overcome to facilitate this effort and one of these involves the development of improved methods to monitor treatment effects. At present, the effects of treatment are typically assessed by radiological means several months after its initiation, which is unsatisfactory for a fast growing tumor like GBM. It is however likely that treatment effects can be detected on a molecular level long before radiological response, especially considering many of the targeted therapies that are currently being developed. Biomarkers for treatment efficacy may be of great importance in the future individualization of brain tumor treatment.

The work presented herein was primarily focused on detecting early effects of GBM treatment. To this end, we designed experiments in the BT4C rat glioma model in which we studied effects of both conventional radiotherapy and an experimental angiogenesis inhibitor, vandetanib. Brain tissue samples were analyzed using a high throughput mass spectrometry (MS) based screening, known as Surface Enhanced Laser Desorption/Ionization - Time of Flight - Mass Spectrometry (SELDI-TOF-MS). The vast amounts of data generated were subsequently analyzed by established multivariate statistical methods, such as Principal Component Analysis (PCA), Partial Least Squares (PLS), and Orthogonal Partial Least Squares (OPLS), developed for analysis of large and complex datasets. In the radiotherapy study we detected a protein spectrum pattern clearly related to tumor progression. We notably observed how this progression pattern was hampered by radiotherapy. The vandetanib study also revealed significant alterations of protein expression following treatment of different durations, both in tumor tissue and in normal brain contralateral to the tumor.

In an effort to further elucidate the pathophysiology of GBM, particularly in relation to treatment, we collected extracellular fluid (ECF) samples from 11 patients diagnosed with inoperable GBM. The samples were collected by means of stereotactic microdialysis, both from within the contrast enhancing tumor and the brain adjacent to tumor (BAT). Samples were collected longitudinally from each patient in a time span of up to two weeks, during which the patient received the first five fractions of radiotherapy. The ECF samples were then analyzed by Gas Chromatography Mass Spectrometry (GC-MS) to screen them with respect to concentrations of low molecular weight compounds (metabolites). Suitable multivariate analysis strategies enabled us to

extract patterns of varying metabolite concentrations distinguishing between samples collected at different locations in the brain as well as between samples collected at different time points in relation to treatment.

In a separate study, we also applied SELDI-TOF-MS and multivariate statistical methods to unravel possible differences in protein spectra between invasive and non-invasive WHO grade I meningiomas. This type of tumor can usually be cured by surgical resection however sometimes it grows invasively into the bone, ultimately causing clinical problems. This study revealed the possibility to differentiate between invasive and non-invasive benign meningioma based on the expression pattern of a few proteins.

Our approach, which includes sample analysis and data handling, is applicable to a wide range of screening studies. In this work we demonstrated that the combination of MS screening and multivariate analyses is a powerful tool in the search for patterns related to treatment effects and diagnostics in brain tumors.

3. Populärvetenskaplig sammanfattning på svenska

I denna avhandling har vi studerat proteiner och små molekyler i olika typer av hjärntumörer. Vi har visat att olika behandlingar leder till systematiska förändringar i protein- och metabolit sammansättningen i elakartade hjärntumörer. Vi har också visat att man med hjälp av skillnader i proteinsammansättningen kan skilja mellan olika typer av godartade hjärnhinnetumörer.

En cancercell skiljer sig från en vanlig cell genom att ett antal förändringar i dess gener inträffat. Dessa förändringar får konsekvenser då genens information överförs till proteiner, vilka är stora molekyler som reglerar cellens biologiska processer. De små molekyler som utgör mellan- eller slutprodukter i dessa reaktioner kallas metaboliter. Kollektivt brukar en cells samtliga proteiner benämnas dess proteom och tillika benämns samtliga metaboliter dess metabolom. Ytterst är det huvudsakligen proteinerna som reglerar diverse funktioner och är inblandade i istort sett alla biokemiska reaktioner som äger rum i en cell. Proteiner är uppbyggda av många mindre byggstenar, så kallade aminosyror. Ett protein kan bestå av allt från några tiotal aminosyror till flera tiotusentals. Hur ett protein monteras ihop bestäms av cellens arvsmassa (DNA) och har en avgörande betydelse för proteinets funktion. Om cellens DNA råkar ut för en förändring (mutation) kan detta leda till att proteinet monteras ihop felaktigt. Felaktigt monterade proteiner kan inte alltid utföra sina funktioner, åtminstone inte på ett kontrollerat och korrekt sätt, vilket i sin tur kan resultera i diverse sjukdomar, t. ex. cancer. Till metaboliterna räknas bl. a. aminosyrorna, men även många andra molekyler som är inblandade i cellens ämnesomsättning (metabolism) och restprodukter därav. Glukos, mjölksyra och aminosyran alanin är exempel på metaboliter i cellen.

Båda dessa klasser av molekyler har under årtionden studerats ingående med olika metoder. Många av molekylerna är noggrant beskrivna och i många fall har orsaken till vissa cancersjukdomar härletts till mutationer i specifika gener som leder till specifika funktionsförändringar i specifika proteiner. Eftersom ämnesomsättningen styrs av proteiner, påverkas i sin tur även metaboliterna av dessa funktionsförändringar vid olika sjukdomstillstånd. Historiskt sett har molekylärbiologisk forskning huvudsakligen bedrivits genom att studera ett protein eller en metabolit i taget för att på så vis kartlägga hur några av de olika molekylerna samverkar. Proteomet och metabolomet består dock av en så ofantlig mängd olika molekyler att det inte är rimligt att en sådan analysstrategi ska kunna ge en komplett karta över de cellulära processerna.

Med utveckling av nya analystekniker, kraftfulla datorer och dataanalysmetoder är det idag möjligt att studera proteomet och metabolomet i större skala. Dessa forskningsområden kallas proteomik respektive metabolomik. En speciell inriktning inom dessa områden går ut på att jämföra proteomet eller metabolomet mellan olika tillstånd i jakt på specifika skillnader. T. ex. kan man jämföra proteininnehållet mellan cancerceller och normala celler, där de proteiner som utgör skillnaden kan fungera som biomarkörer, d.v.s. proteiner som på något mätbart vis skiljer sig åt och därigenom indikerar sjukdom. I vissa fall kan enstaka proteiner eller metaboliter utgöra sådana biomarkörer och i andra fall kan mönster av proteiner eller metaboliter innehålla den information som eftersöks.

I denna avhandling har vi studerat proteomet och metabolomet i hjärntumörer. Cancer i allmänhet är komplexa sjukdomar som vanligtvis involverar förändringar i många cellulära processer och molekyler. Vi har därför valt att studera sammansatta mönster av proteiner och metaboliter, istället för en molekyl i taget. En typ av hjärntumör som vi undersökte var glioblastom, vilket är en av de vanligaste elakartade hjärntumörformerna hos vuxna. Idag är behandlingen vid glioblastom en kombination av operation och efterföljande strål- och cytostatikabehandling. Trots dessa insatser är prognosen fortfarande dålig vid denna allvarliga sjukdom. Det finns således ett stort behov av nya eller förbättrade behandlingsmetoder för patienter med glioblastom. För att nå dit krävs dock en ökad förståelse för den bakomliggande tumörbiologin samt metoder att snabbt och pålitligt kunna utvärdera resultatet av behandling. Syftet med denna avhandling har varit att undersöka proteomet och metabolomet hos glioblastom i samband med behandling för att utreda hur olika behandlingar påverkar tumören samt att undersöka möjligheten att hitta biomarkörer relaterade till behandlingseffekt. I en djurmodell visade vi att man kunde följa tumörutvecklingen genom att studera hur proteinuttrycksmönstret förändrades över tiden. Vi observerade även att proteinuttrycksmönstret på ett tidigt stadium påverkades av strålningsbehandling samt att denna påverkan tycks relaterad till en hämmad tumörutveckling. Vidare undersökte vi, i samma djurmodell, hur proteomet påverkades av en experimentell blodkärlshämmande behandling. Syftet med behandlingen var att strypa tumörcellernas tillgång till näring och därigenom döda tumörcellerna. Även denna behandlingsform gav upphov till förändrade proteinuttrycksmönster som tydligt kunde avläsas med hjälp av de analysmetoder vi använde oss av.

I en annan studie undersökte vi skillnader i metabolomet mellan själva tumören och hjärnan i området runt tumören hos en grupp patienter med glioblastom. Studien genomfördes under den första veckan patienterna genomgick strålbehandling. Vi kunde observa tydliga skillnader i metabolomet mellan de olika områdena i hjärnan samt specifika förändringar till följd av de de första dagarnas strålbehandling.

Slutligen studerade vi även patientmaterial från meningeom, en annan vanlig form av hjärntumör som utgår ifrån hjärnhinnorna. Meningeom kan oftast botas på kirurgisk

väg men i en del fall växer tumören in i hjärna eller skallben och är därmed svårare att bota. För att kunna fatta beslut rörande hur dessa patienter ska behandlas vore det bra om det redan vid den första operationen gick att förutsäga vilka patienter som löper risk att utveckla invasivt växande meningeom. Vi visade i denna studie att man baserat på proteomets sammansättning kunde särskilja invasivt växande tumörer från icke invasivt växande.

Studier av proteomet och metabolomet är komplexa och resulterar i mycket stora datamängder. För att effektivt hantera dessa datamängder samt för att tillåta jämförelser av sammansatta mönster har vi genomgående tillämpat multivariata analysmetoder speciellt framtagna för detta ändamål. Gemensamt för dessa metoder är att de kan hantera stora datamängder och ge en förenklad bild för tolkning av de komplexa samband som existerar i data som beskriver proteomet och metabolomet. Sammanfattningsvis kan vi konstatera att storskaliga studier av proteomet och metabolomet är lovande för att detektera behandlingseffekter på ett tidigt stadium. Sådana studier kan i framtiden underlätta arbetet med förbättring av befintliga behandlingar samt utveckling av nya behandlingsmetoder vid hjärntumörer.

4. Abbreviations and notations

4.1 Abbreviations

BAT	brain adjacent to tumor
CEC	circulating endothelial cells
CEP	circulating endothelial progenitor cells
CNS	central nervous system
CT	computer tomography
DA	discriminant analysis
DNA	deoxyribonucleic acid
ECM	extracellular matrix
EGF(R)	epidermal growth factor (receptor)
GBM	glioblastoma multiforme
GC	gas chromatography
H-MCR	hierarchical multivariate curve resolution
HIF-1α	hypoxia-inducible factor-1α
IMAC	immobilized metal affinity chromatography
ITOT	individual treatment over time
LOH	loss of heterozygosity
MALDI	matrix assisted laser desorption/ionization
MGMT	methylguanine-DNA methyltransferase
MR(I/S)	magnetic resonance (imaging / spectroscopy)
(m)RNA	(messenger) ribonucleic acid
MS	mass spectrometry
MVD	microvascular density
PCA	principal component analysis
PDGF(R)	platelet derived growth factor (receptor)
(O)PLS	(orthogonal) partial least squares / projection to latent structures
PRESS	prediction error sum of squares
PTEN	phosphatase and tensin homolog
ROI	region of interest

SC	subcutaneous
SELDI	surface enhanced laser desorption/ionization
TIC	total ion current
TMS	trimethylsilyl
TOF	time of flight
VEGF(R)	vascular endothelial growth factor (receptor)
WHO	World Health Organization

4.2 Notations

The following notations will be used throughout this thesis: (i) scalars are dented by italic letters, (ii) vectors are denoted by lower-case and bold font, (iii) matrices are denoted by upper-case and bold font, and (iv) vector or matrix inverses are denoted by a single quotation character, e.g. **p'**.

K	number of descriptor variables (columns in **X**)
M	number of response variables (columns in **Y**)
N	number of observations (rows in **X** and **Y**)
A	number of latent variables
X	matrix of descriptor variables [$N \times K$]
Y	matrix of response variables [$N \times M$]
y	response vector, i.e. a single response variable [$N \times 1$]
B	matrix of regression coefficients [$K \times M$]
T	matrix of score vectors for **X** in PCA, PLS and OPLS [$N \times A$]
t	vector of scores for **X** in PCA, PLS and OPLS [$N \times 1$]
E	matrix of residuals for **X** [$N \times K$]
F	matrix of residuals for **Y** [$N \times M$]
P	matrix of loading vectors for **X** in PCA, PLS and OPLS [$K \times A$]
p	loading vector for **X** in PCA, PLS and OPLS [$K \times 1$]
C	matrix of loading vectors for **Y** in PLS and OPLS [$M \times A$]
c	vector of loadings for **Y** in PLS and OPLS [$M \times 1$]
w	vector of weights for **X** in PLS and OPLS [$K \times 1$]
m	mass
z	number of charges

5. Introduction

Tumors inside the skull are commonly referred to as brain tumors. They may be divided into primary or secondary brain tumors, where primary brain tumors arise *de novo* in the brain, contrary to secondary ones that develop in some other part of the body and metastasize to the brain. Collectively, primary brain and central nervous system (CNS) tumors account for about 2 % of all cancers (ACS, 2008) and the incidence is about 18 cases per 100,000 individuals per year, where 11 of these are non-malignant and 7 are malignant (CBTRUS, 2009).

According to the World Health Organisation's (WHO) classification of primary brain tumors are typically divided into different histological subtypes based on the type of tissue the tumor is believed to originate from. Each subtype is then further graded into four malignancy grades, where I is considered benign and IV is the most malignant. The two most frequently diagnosed primary brain and CNS tumors are meningioma (~33%) and glioblastoma (~18%) (CBTRUS, 2009). These two types of tumors have been the subjects of study in this thesis.

The focus of this work is to screen the proteome and the metabolome for effects of treatment in glioblastoma (papers **I**-**III**) and to distinguish between groups of meningioma patients that may benefit from different treatment approaches (paper **IV**). With these efforts we hope to facilitate the improvement of treatment modalities and treatment decisions. However, to achieve this, basic understanding of brain tumor biology is fundamental. The following sections aim to briefly introduce some key elements of brain tumor biology, focusing on GBM, as well as the concept of non-targeted proteomic and metabolomic screening.

5.1 Glioblastoma

Glial cells are on average about equal in numbers to neural cells within the CNS, where they are also known as neuroglia. When they were first discovered, they were believed to function mainly as supportive tissue to the neurons, or in other words, as the glue that kept the neurons together ("glia" is Greek for "glue"). Nowadays, however, it is recognized that the neuroglia cells have many more important functions. The neuroglia cells are in turn divided into different groups, including astrocytes, oligodendrocytes, and ependymal cells, each with different primary functions. Of these, the astrocytes are the most abundant. Apart from their functioning as connective tissue, they are also involved in trauma repair, molecular transport, and scaffolding for the brain blood barrier (Raine, 1999).

Tumors originating from neuroglia cells are collectively called gliomas. In similarity with the neuroglia cells, gliomas are in turn further categorized based on what subtype of neuroglia cell is believed to be the originally transformed one, such as oligodendroglioma (originating from oligodendrocytes), ependymoma (originating from ependymal cells) and astrocytoma (originating from astrocytes).

As the astrocyte is the most abundant neuroglia cell, astrocytoma is the most common form of glioma, constituting 54% thereof (CBTRUS, 2009). Astrocytomas are normally graded according to the WHO's classification, which is based on morphological characteristics of the tumor and where the grades correspond to tumor malignancy and patient prognosis. The grades are: pilocytic astrocytoma (WHO grade I), diffuse astrocytoma (WHO grade II), anaplastic astrocytoma (WHO grade III) and glioblastoma, also known as glioblastoma multiforme (GBM) (WHO grade IV) (Kleihues et al., 2007; Louis et al., 2007). Of these, the last two are referred to as high-grade and are the most malignant, while pilocytic astrocytomas are considered benign (Wen & Kesari, 2008). GBM, i.e. the last and most malignant form (figure 1), has been the subject of study in this thesis (papers **I-III**).

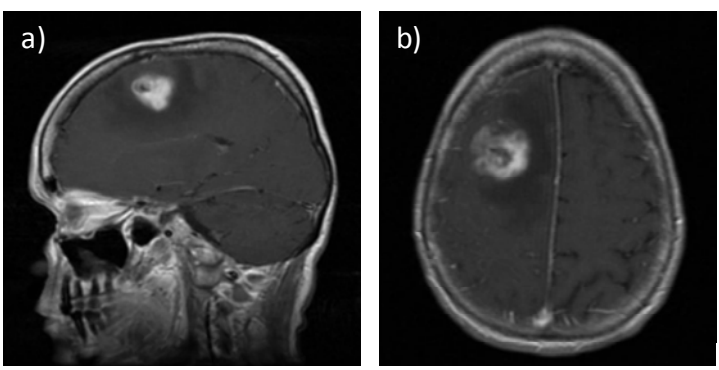

Figure 1. Magnetic resonance image of a GBM. **a)** Sagittal view. **b)** Transaxial view.

As implied by the name, glioblastoma multiforme consists of cells that are morphologically heterogeneous. They are characterized by dense cellularity, nuclear pleomorphism (i.e. cell nuclei with atypical, irregular shapes) and microvascular proliferation (Wen & Kesari, 2008). Other characteristic features of GBM commonly include microscopic necrotic areas and often also a large central necrotic area that is frequently identified by imaging techniques, such as magnetic resonance imaging (MRI) or computer tomography (CT), as well as a diffusely infiltrative growth pattern, where tumor cells invade the surrounding tissue (Ichimura et al., 2004).

The underlying cause (i.e. the etiology) of GBMs is largely unknown. About 5% of all cases are familiar, where some of these can be ascribed to a few rare hereditary

syndromes (Malmer et al., 1999), whereas the rest are believed to be caused by sporadic mutations (see section 5.1.1). The only environmental factor indisputably associated with an increased risk of developing the disease is exposure to high-dose irradiation (Ohgaki & Kleihues, 2005a).

The prognosis for patients diagnosed with GBM varies, but is generally very poor. It is mainly dependent on the available treatment options (see section 5.1.5), but also on factors such as age and performance status (Laws et al., 2003; Ohgaki & Kleihues, 2005b). The median survival for untreated patients is approximately 3 months from the time of diagnosis. Given current standard care, the median survival is about 14 months, where ~1-5% of the patients survive for 3 years or more (Henriksson et al., 2006; Krex et al., 2007; Ohgaki & Kleihues, 2005b; Stupp et al., 2005).

5.1.1 Primary and secondary glioblastoma

As previously mentioned, little is known about the GBM etiology. On a molecular level, however, there is an ever-increasing amount of knowledge continuously unraveling the underpinnings of the disease. Although it is uncertain exactly how GBM arises, the process is traditionally ascribed to one of two different pathways; the *de novo* or the progressive pathway (figure 2). The distinction is mainly based on clinical presentation, where GBM that arises *de novo* are called primary and GBM resulting from progression from previously diagnosed lower grade astrocytomas are referred to as secondary (Furnari et al., 2007; Kleihues & Ohgaki, 1997).

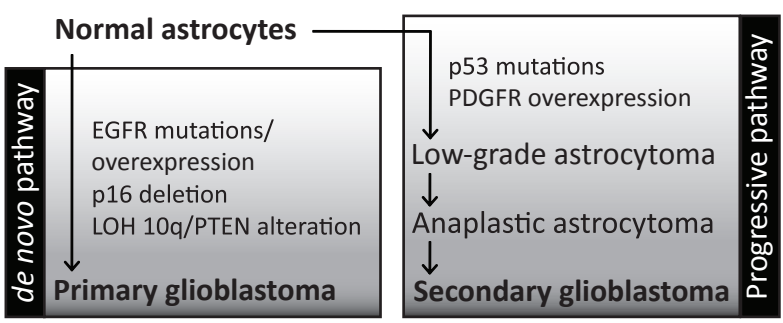

Figure 2. A simplified illustration of the suggested model for GBM progression.

Primary GBM constitute the vast majority of all cases (90-95%). The two entities are morphologically and clinically indistinguishable and share the same poor prognosis. They do, however, display notably different genetic characteristics. These characteristics include specific transformations of well-known oncogenes and tumor suppressor genes. Such genetic characteristics associated with primary glioblastoma include: amplifications and activating mutations of the epidermal growth factor receptor (EGFR) oncogene, loss of hetrozygosity (LOH) in the long arm of chromosome 10

(10q), which houses the phosphatase and tensin homolog (PTEN) tumor suppressor gene, as well as deletion of the tumor suppressor protein p16. Secondary GBM on the other hand typically exhibit mutations in the p53 tumor suppressor protein, over expression of the platelet derived growth factor receptor (PDGFR) oncogene, etc (Furnari et al., 2007).

5.1.2 Tumor stem cells

The concept of tumor stem cells is being investigated in several types of cancers. A tumor stem cell is usually defined as a pluri- or multipotent tumor cell with the ability to undergo asymmetric cell-division and thereby self-renew (i.e. generate additional tumor stem cells) and differentiate into various types of tumor cells (Pardal et al., 2003). Tumor stem cells in general are believed to be associated with various distinguishing features that can be used as stem cell markers, such as the surface protein CD133 for example, which is considered associated to brain tumor stem cells (Dirks, 2008). Tumor cells with stem-cell like characteristics were first isolated from GBM samples in 2003, and were shown to express CD133 to a large extent (Singh et al., 2003). These cells may be the result of malignant transformation of normal adult neural progenitor cells (Vescovi et al., 2006) and are thought to constitute a cell population that may ultimately be responsible for colonizing the tumors as they develop (Galli et al., 2004). The tumor stem cell concept is however currently being debated and no specific markers for these cells have been established as of yet (Bjerkvig et al., 2005; Dirks, 2008).

5.1.3 Neovascularization

All cells are dependent on a sufficient supply of nutrients and oxygen, which is provided by the vascular network. Tumor cells typically proliferate at a higher rate than normal cells and thus require a high oxygen and nutrient supply to sustain progression. During early tumor progression a growing GBM can satisfy this demand by growing along already existing vessels, a process that is referred to as vascular co-option (Leenders et al., 2002). During the course of progression, however, the ability to attract new blood vessels becomes important. As mentioned, one characteristic feature of GBM is its extensive microvascular proliferation, which it may achieve by any of the three following mechanisms: (i) vasculogenesis, i.e. *de novo* formation of new vessels through recruitment of circulating endothelial precursor cells (Asahara et al., 1997; Santarelli et al., 2006), (ii) intussusception, i.e. splitting an existing vessel in two through invagination of its endothelial cells (Kurz et al., 2003), and (iii) angiogenesis, the formation of new vessels from pre-existing ones by vascular sprouting. In growing GBM, angiogenesis is the most predominant cellular mechanism of neovascularization and it is further discussed below.

5.1.3.1 Angiogenesis

The process of angiogenesis is the main process by which GBM forms new blood vessels (Carmeliet & Jain, 2000) and it can be outlined in a series of consecutive steps: (i) degradation of an existing blood vessel's basement membrane and extracellular

matrix (ECM), (ii) proliferation and migration of blood vessel endothelial cells, (iii) formation of the lumen of the new blood vessel, through cell-cell interactions, and (iv) capillary formation and maturation of the new vessel (Alberts et al., 2002).

Angiogenesis is a complex progression that involves many different factors, where members of the vascular endothelial growth factor (VEGF) protein family are key components. In man, the VEGF family consists of five known members (VEGF-A, -B,-C,-D and PlGF), with VEGF-A (also often referred to as VEGF) being the most abundant. VEGF-A is also generally considered to be the most important positive regulator of angiogenesis (Ferrara et al., 2003) and it is extensively expressed in GBM, where it is believed to be the principle angiogenesis factor (Plate et al., 1993). Its importance in development is emphasized by the fact that single allele deletions are lethal in early embryonic development (Carmeliet et al., 1996). It is however not the only family member linked to GBM angiogenesis. Recently, expression of VEGF-C and -D was also reported in human malignant glioma, suggesting a functional role for these proteins as well in glioma angiogenesis (Grau et al., 2007; Jenny et al., 2006).

The members of the VEGF protein family operate through binding to different members of a receptor tyrosine kinase family, namely VEGF-R1, R2 and R3, which are located on the cell surface of endothelial cells. VEGF-R2 is considered to be the main receptor for VEGF-A, and phosphorylation thereof induces endothelial cell proliferation, migration and tube formation (Ferrara et al., 2003). The expression of the VEGF proteins is in turn affected by several factors, one of which is the state of hypoxia that is known to induce VEGF expression through the transcription factor hypoxia-inducible factor-1α (HIF-1α), as a mean to affect the cells' oxygen supply (Plate et al., 1992; Shweiki et al., 1992).

Considering the importance of the angiogenesis process in GBM biology, and the fact that it today is very well documented, it constitutes a prime target for the development of new GBM treatments. There are several experimental studies indicating that angiogenesis inhibition in general, and inhibition of the VEGF signaling pathway in particular, may be a successful treatment of experimental malignant glioma (Cheng et al., 1996; Kirsch et al., 1998; Lai et al., 2008; Peroulis et al., 2002). Furthermore, the combined VEGF-R/EGF-R tyrosine kinase inhibitor vandetanib, which was used for treatment in paper **III** of this thesis, has been shown to inhibit experimental glioma growth and potentiate the effects of chemotherapy and radiotherapy (Sandstrom et al., 2004; Sandström et al., 2008).

5.1.4 Invasiveness

The invasive phenotype is often found in both low and high-grade gliomas, indicating that it is acquired at an early stage of tumorigenesis. The tumor can thus grow more or less unrestrained in the brain before yielding any symptoms. Consequently, patients are typically not diagnosed until late in the course of the disease. Furthermore, the

infiltrative growth pattern also renders radical resection by means of surgery impossible, ultimately and invariably resulting in the disease recurring after treatment due to re-growth of invasive cells (Hoelzinger et al., 2007). Taken together, the infiltrative growth aspect can be considered one of the main reasons why treating GBM is so difficult.

Degradation of components of the ECM is thought to be a prerequisite for the widespread infiltration of glioma cells into the surrounding brain. To this end, the plasminogen activator system and the matrix metallo proteinase system are thought to be important contributors (Sandström et al., 1999). Both are also involved in angiogenesis, as well as several other important physiological processes where regulated degradation of ECM components is necessary. Components of both these systems are dysregulated in several malignant tumors, including GBM (Skrzydlewska et al., 2005).

Specific conditions of the tumor microenvironment are thought to trigger certain GBM cells to turn invasive, such as hypoxia for instance. Hypoxia is known to induce the transcription complex HIF-1, which in turn increases the expression of genes related to invasion (Ruan et al., 2009), which are to some extent the same as those involved in angiogenesis. As the invading cell migrates away from the tumor, it will travel through microenvironments of different characteristics. To sustain the invading phenotype it may rely on autocrine motility signaling and paracrine effects thereof (Hoelzinger et al., 2007). Invasion is however not a random event, instead GBM cells have been shown to primarily migrate along specific anatomic structures, such as myelinated axons. This is indicative of tumor cells interacting with specific, preferred constituents of the microenvironment, a process which may be related to the developmental process of neuronal guidance (Hoelzinger et al., 2007).

5.1.5 Glioblastoma treatment

GBM is normally treated by surgery, radiotherapy and chemotherapy. Due to its characteristic infiltrative growth, it is virtually impossible to surgically resect a GBM tumor completely. Nevertheless, surgical resection has been shown to have a significant impact on survival (Bergenheim et al., 2007; Mineo et al., 2007) and is always performed if possible, based on tumor location and patient performance status. In cases where surgical resection is not an option, patients are subjected to a biopsy procedure to obtain tissue samples for histopathological diagnosis. All patients (regardless of whether resection was performed or not) are then treated with fractionated radiotherapy. Each fraction is typically 2 Gy and is administered five days a week to a total dose of 60 Gy over six weeks. Several studies have shown that post-operative radiotherapy has a positive effect on patient survival (Filippini et al., 2008; Kristiansen et al., 1981; Mineo et al., 2007; Walker et al., 1979). Today, GBM patients are also routinely treated with chemotherapy during the radiotherapy course and as adjuvant cycles after completion of radiotherapy. Most commonly the alkylating agent temozolomide is administered per orally daily at 80 mg/m^2 during radiotherapy. After

completion of radiotherapy, adjuvant temozolomide 150-200 mg/m² is administered per orally for five days every four weeks, for up to six cycles. This multimodal regiment is based on a randomized controlled trial showing that the addition of temozolomide increases the median survival from 12.1 to 14.6 months, as compared to radiotherapy treatment alone (Stupp et al., 2005). Similarly, the 2 year survival increases from 10.4% to 26.5%. The recently published follow-up from this study demonstrate that also the five year survival is increased, from 1.9% to 9.8% (Stupp et al., 2009). Surgery followed by radiotherapy with concomitant temozolomide-based chemotherapy and then adjuvant temozolomide is therefore currently considered the standard of GBM treatment in most centers worldwide (Hart et al., 2008).

Despite the recent advances in therapy, the prognosis for GBM is still poor and the search for new treatments is intensive. Several therapies targeted for specific molecular pathways involved in GBM are currently under development. For instance, pathways involved in angiogenesis are considered possible future treatment targets for GBM. As previously mentioned, VEGF-A is a key positive angiogenesis regulator and an interesting target for new GBM treatments (Miletic et al., 2009). Bevacizumab is a humanized monoclonal antibody against VEGF-A, approved for use in combination with chemotherapy in the treatment of advanced colon cancer, non-small cell lung cancer and breast cancer and with interferon 2α in renal cell carcinoma (Escudier et al., 2007; Hurwitz et al., 2004; Miller et al., 2007; Sandler et al., 2006). Bevacizumab in combination with the topoisomerase I inhibitor irinotecan have recently been reported to result in promising response rates in the treatment of recurrent GBM (Vredenburgh et al., 2007) and is now entering the first line setting in combination with chemoradiotherapy and temozolomide (Lai et al., 2008).

5.2 Meningioma

The CNS is lined with a set of three membranes, namely the dura, the arachnoid and the pia mater. Collectively the three membranes are referred to as the meninges, where the dura is the outermost membrane, and also the thickest and most durable in the set. The arachnoid is the middle membrane. Lastly, the pia mater is the innermost as well as the most delicate membrane and adheres to the CNS surface. Tumors arising from the meninges are called meningioma, and they are thought to develop from a specific cell population called arachnoid cap cells that form the outer layer of the arachnoid membrane (Black, 1997).

As mentioned previously, meningioma is the most common primary brain or CNS tumor. The incidence is increasing with age and there is a considerable female preponderance, where the annual incidence rate is 3.6 for males and 8.1 for females per 100.000 individuals (CBTRUS, 2009). Regarding all primary intra-cranial tumors, meningiomas make up for approximately 20% and 38% in males and females, respectively (Claus et al., 2005; Lamszus, 2004). The WHO classification divides the tumors into three grades that reflect the risk of recurrence and aggressive behavior,

where each grade involve several histological subtypes (Kleihues et al., 2002; Kleihues et al., 2007). The grades may be categorized as: benign meningioma (WHO grade I), atypical meningioma (WHO grade II), and anaplastic/malignant meningioma (WHO grade III). Benign meningiomas are by far the most common and constitute 80-90% of all cases (Claus et al., 2005; Lamszus, 2004). These tumors are mostly non-invasive to their character but may sometimes, despite their benign histologic appearance, invade surrounding tissues (Bikmaz et al., 2007; Lamszus, 2004). Benign meningiomas are in turn divided into several histological subtypes, which act strictly descriptively and contain no prognostic information (Black, 1997). Two of the most common of these subtypes are meningothelial and fibrous meningioma, and they were the focus of study in paper **IV**.

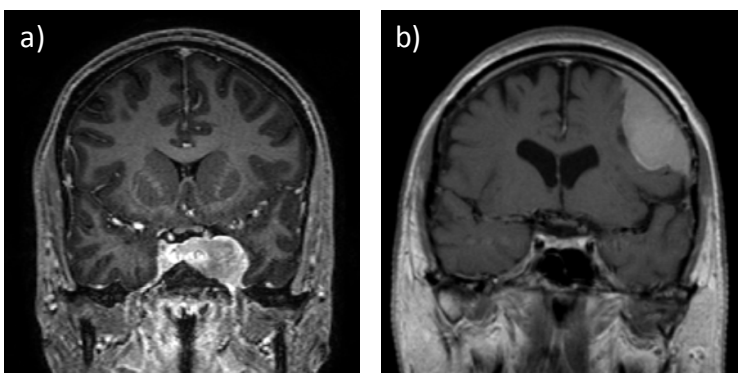

Figure 3. MR image of two patients diagnosed with benign meningioma (WHO grade I), coronal view. **a)** Skull base meningioma, growing invasively into the adjacent bone. **b)** Convexity meningioma.

There is a significant association between meningioma diagnosis and a parental history of meningioma (Claus et al., 2005). This is especially evident among patients suffering from the genetically inherited disease neurofibromatosis type 2 that have a strong predisposition to also develop meningioma. However, there appears to be additional genetic loci (unrelated to the NF2-gene on 22q) that may be involved in meningioma tumorigenesis (Louis et al., 1997). Among the sporadic cases, there are mainly two factors that have been associated to the etiology of meningioma, namely hormones and ionizing irradiation. Studies have suggested that even low irradiation doses increases the risk for meningioma (Claus et al., 2005). The latency period between radiation exposure and diagnosis is correlated to the radiation dose, but is on average 2-3 decades (Louis et al., 1997). Furthermore, the overrepresentation of female patients (the female:male ratio is approximately 2:1) implicates involvement of sex hormones in meningioma tumorigenesis. This is further addressed in section 5.2.3.

Benign meningiomas may be cured by surgical resection. However prognosis depends largely on tumor location (figure 3). Convexity tumors are often totally resected and thereby cured, whereas skull base tumors often have a poor outcome, due to difficulties to achieve a complete excision (Condra et al., 1997; Lamszus, 2004). The overall recurrence rate for benign meningiomas is reported to be 7-20%, and that of malignant meningiomas is 50-78% (Louis et al., 1997). Inoperable tumors, or tumor remnants, may remain histologically benign or, alternatively, they may progress towards a more malignant phenotype over the years.

5.2.1 NF2-gene

The most frequently observed gene alteration in meningiomas is mutations of the NF2-tumor suppressor gene, which has been reported to be mutated in up to 60% of all sporadic meningiomas (Ruttledge et al., 1994). It is located on the long arm of chromosome 22 (22q12.2), where chromosomal losses are often observed in sporadic meningiomas. These mutations of the NF2-gene or 22q-losses are frequently observed in benign meningiomas, and appear to be early events in meningioma tumorigenesis (Lamszus, 2004). They have also been shown to be significantly more frequent in fibrous meningioma (70%) than in meningothelial meningioma (25%) (Wellenreuther et al., 1995). Furthermore, most patients suffering from the dominantly hereditary genetic defect neurofibromatosis type 2 develop meningiomas. Commonly these patients develop multiple meningioma, and the onset of the disease normally occurs several decades earlier in life than it does in patients struck with sporadic disease. The frequency of atypical meningioma is however not increased in neurofibromatosis type 2 patients (Lamszus, 2004). The mechanism by which NF2 mutation leads to meningioma formation is unknown. However, based on the NF2 gene product's function, speculations have been presented. The protein encoded for by the NF2-gene, Neurofibromin 2 (also known as Merlin), is known to be involved in cytoskeletal stability and is predominantly located at the interface between the cell membrane the cytoplasm. Accordingly, it has been hypothesized that NF2 mutation or loss may lead to impairment of cell contact inhibition and ultimately to uncontrolled growth (Black, 1997).

5.2.2 Vascular endothelial growth factor

The previously described angiogenesis factor VEGF-A, which in GBM is associated with high micro vascular density is also expressed in meningioma. In meningiomas this expression is mainly associated with peritumoral edema (Goldman et al., 1997; Provias et al., 1997), whereas there appears to be no clear correlation between VEGF-A expression and micro vascular density (Lamszus et al., 2000). At the same time, in meningiomas micro vascular density does not correlate to malignancy grade, which contrasts with GBM. VEGF-A has, on the other hand, been found expressed at 10-fold higher levels in anaplastic meningiomas and at 2-fold higher levels in atypical meningiomas, as compared to benign tumors (Lamszus, 2004).

5.2.3 Sex hormones

As mentioned above, meningiomas are 2-fold more frequent in women than in men and are to some extent in association with breast carcinoma (Black, 1997). Meningiomas have also been observed to display an accelerated growth during pregnancy as well as during the menstrual cycle's luteal phase (Lamszus, 2004). Altogether, these observations indicate that sex hormones may play an etiological role in meningioma tumorigenesis. The estrogen receptor is sometimes detected in meningiomas but only in about 10% of all cases and only at low levels. Thus it appears not very important for meningioma development (Black, 1997). The progesterone receptor, on the other hand, is the most commonly found sex hormone receptor in meningiomas, and is expressed in 40-100% of the cases (Claus et al., 2005; Kyritsis, 1996). Although the biological function of the progesterone receptor in relation to meningioma tumorigenesis is unclear, it has been shown to be a positive prognostic factor that is strongly inversely correlated with mitotic index as well as tumor grade (found in 96% and 40% of benign and malignant meningiomas, respectively) (Claus et al., 2005; Hsu et al., 1997).

5.2.4 Meningioma treatment

Surgical resection is the basis for treatment of all types of meningiomas. Depending on the tumor location, complete resection is usually the goal (Black, 1997). As mentioned above, the extent of surgical resection is a key factor regarding the recurrence rate. The degree of resection is typically assessed in terms of a classification scheme presented by Simpson in 1957 (Simpson, 1957), see table 1.

Table 1. Simpson grading of resection versus recurrence

Simpson Grade	Description	Recurrence
I	Macroscopically complete removal, including dura and bone	9%
II	Macroscopically complete removal, with dural coagulation	19%
III	Macroscopically complete removal, without dural coagulation	29%
IV	Subtotal removal	44%
V	Decompression	-

Recurrence rates represent those reported by Simpson 1957. The follow-up time was 10 years (Simpson, 1957).

Radiotherapy-based treatment is considered the primary treatment option for critically located tumors (Kondziolka et al., 2008; Mendenhall et al., 2004; Mendenhall et al., 2003; Walsh & Couldwell, 2009). Also in tumors that are treated by surgical resection,

radiotherapy has become an important management tool that may be used both as an adjunct to surgery or as treatment for recurrence (Black, 1997). Regarding atypical and anaplastic meningiomas, postoperative radiotherapy is considered standard care. However, the timing of radiotherapy treatment of benign meningiomas is a matter of much debate, i.e. whether it should be given post-operatively or at the time of recurrence (Campbell et al., 2009). Several studies have reported that adjuvant radiotherapy treatment after subtotal resection lowers the recurrence frequency in patients with benign meningioma (Barbaro et al., 1987; Condra et al., 1997; Milker-Zabel et al., 2006), and thereby ultimately increases the cause specific survival. However, given the relative slow progression of the disease and the severe adverse side effects of radiotherapy, some argue that it is preferable to delay its use until signs of progressive disease (Miralbell et al., 1992; Newman, 1994; van Nieuwenhuizen et al., 2007).

Various agents for chemo-, immuno- and hormonal therapy have been suggested for patients suffering from recurrent, unresectable and previously irradiated meningiomas (Kyritsis, 1996). For example, treatments of benign meningiomas with the progesterone receptor inhibitor RU-486 as well as with interferone-α have been reported to have some beneficial effects (Kyritsis, 1996; Lamberts et al., 1992; Ragel & Jensen, 2003).

5.3 Proteomics & Metabolomics

The central dogma of molecular biology describes the transfer of sequential information from deoxyribonucleic acid (DNA) to proteins (Crick, 1970) (figure 4). The starting point is the genome, i.e. the complete set of DNA in an organism. In the DNA the organism's hereditary information is encoded and, as a fundamental part of cell division, it possesses the ability to duplicate itself through a process known as replication. In mankind the genome is thought to consist of 20,000-25,000 protein-coding genes (Collins et al., 2004) that remain static throughout life, save for various mutations. Through the process of transcription, this information is in turn transferred to messenger ribonucleic acid (mRNA). The transcription process is responsive to a variety of external factors, thus the transcription of genes coding for certain proteins may be up or down regulated as a result of cell-cell communication or exogenous stress factors acting upon the cell. Consequently, and in contrast to the genome, the set of mRNA molecules present within a cell or an organism (i.e. the transcriptome) is dynamic and depending on various environmental features. The information encoded in the mRNA molecules is eventually transmitted to proteins through a process known as translation. Proteins are in turn the principal modulators of most biological processes, including those of replication, transcription and translation. However, unlike DNA and RNA, proteins cannot reproduce the encoded information, meaning that the information flow is unidirectional. In similarity with the genome and the transcriptome, a system's complete set of proteins is referred to as the proteome. Moreover, these biological entities (i.e. genome, transcriptome and proteome) are usually complemented

by the metabolome, which represents a system's complete set of low molecular weight metabolites.

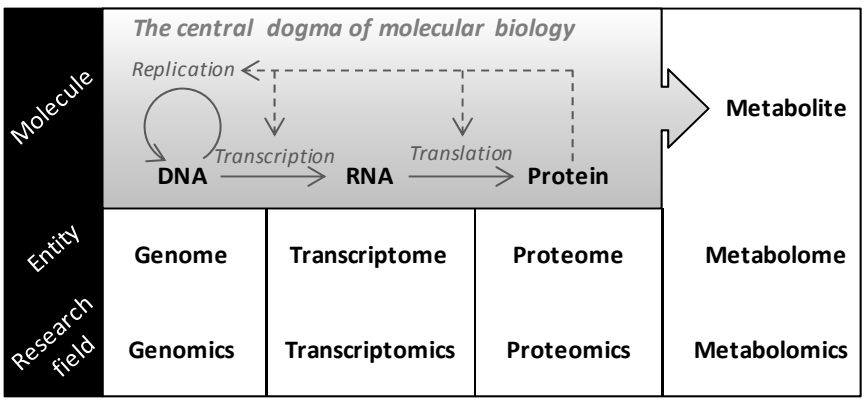

Figure 4. The central dogma of molecular biology and the associated fields of research, i.e. the "omics cascade".

These biological entities may be studied by two philosophically different approaches. Either the study is designed to investigate one or a few specific molecules in a set (genes, proteins etc.), based on a preconceived hypothesis that the specific molecules are carrying important information. Alternatively, the study is designed to globally screen these entities in an unbiased manner. The former approach has traditionally been used in molecular biology research, and has successfully identified most of the components in the various biological entities. However, it does not consider the vast complexity of the biological system (Sauer et al., 2007). With the advent of technologies that allow for global screening, the latter approach has become increasingly popular. It hopes to gain insight into the complexity of the system by a more holistic approach, where all constituents, or as large a portion as possible, of the whole system are quantified simultaneously.

For each of the mentioned biological entities there now is an associated research field concerned with investigating it on a global level. The research fields are: (i) genomics, (ii) transcriptomics, (iii) proteomics, and (iv) metabolomics. Collectively they have been referred to as the "omics cascade", and they constitute the foundation for systems biology, where metabolomics is considered to be the furthest downstream and as such reflecting the events of the upstream processes (Dettmer et al., 2007; Dunn & Ellis, 2005) (figure 4). The papers presented herein are based on proteomics (papers **I**, **III** and **IV**) and metabolomics (paper **II**) investigations.

Whereas the genome is static, in humans consisting of roughly 20,000-25,000 genes, the proteome is dynamic in response to environmental factors and consists of >200,000

proteins (Nishimura et al., 2005). The discrepancy is largely due to alternative splice variants and post translational modifications. For the purpose of investigating effects of treatment, as was the aim in papers **I**-**III**, it is obvious that one must study events occurring downstream of the static genome. Studying the proteome holds further advantages over studying the genome, as it is the proteins that predominantly are the biologically active components. At the same time as the relationship between the level of transcription of a specific gene (i.e. abundance of a specific mRNA molecule) and the abundance of the corresponding protein has been shown to be weak (Chen et al., 2006; Persson et al., 2009). However, this is a far less mature field of study and as of yet there exist no techniques that allow for a truly global analysis of the proteome, as it does for the transcriptome.

The size of the metabolome in plants is estimated to amount to 200,000 molecules, and that of man is speculated to be larger still (Dunn & Ellis, 2005). As mentioned above and illustrated in figure 4, the metabolome may be regarded as downstream to the processes of the central dogma of molecular biology. As such it constitutes a key link between genotype and phenotype. It is expected to reflect the status of a cell more directly than the transcriptome or proteome (Dunn & Ellis, 2005), and may be regarded the ultimate read out of a biological system (Fiehn, 2002). It is therefore an entity that is of great interest to investigate in respect to effects of treatment, as was done in paper **II**.

5.4 Biomarkers

The unbiased screening approaches mentioned above, utilizing various "omics" techniques, are often applied with the intent of distinguishing between samples collected from different populations (e.g. from cancer patients and healthy controls). Once the discriminating features have been characterized and validated, they may represent biomarkers that can be used prospectively to differentiate between the populations included in the study. In this sense, biomarkers are molecules (genes, transcripts, proteins or metabolites) that may be objectively quantified and that are indicative of a specific physiological state.

Depending on the characteristics of the biomarker it may then be used as an indicator for various physiological purposes, such as early detection of disease, susceptibility to treatment, prognostication, or as a measure of response to treatment. However, as can be anticipated from the intricacy of a biological system, in many cases there is no single marker that holds this indicative potential by itself. Rather, complex patterns of markers might be needed (Clarke et al., 2008; Wulfkuhle et al., 2003). The first such multivariate diagnostic assay approved by the Food and Drug Administration is a prognostic test called MammaPrint (Glas et al., 2006). MammaPrint is based on the expression profile of 70 genes that are well-known to be involved in breast cancer (van't Veer et al., 2002), and may be used to assess each individual patient's risk of recurrent disease, which in turn might facilitate treatment decisions.

5.4.1 Brain tumor biomarkers

Most brain tumors are diagnosed and classified based on histological methods and related to WHO's classification scheme, which is used as a measure of malignancy and provides some prognostic information. There is however a degree of inconsistency in the diagnostic procedure, largely due to tumor heterogeneity and limitations in the employed histological techniques. Thus it is difficult to predict how individual patients will respond to treatment, as this typically differs even for patients within the same pathological sub group. To this end, molecular markers with predictive capacity are warranted, in particular to aid the development and use of targeted treatments.

Although there are no validated molecular markers for treatment response in GBM that are universally used in the clinic, there are some that are promising in this regard and indicate that molecular prognostication may one day be feasible. One such marker is the methylation status of the O-6-methylguanine-DNA methyltransferase gene (MGMT), where methylation is positively associated with response to treatment with alkylating agents, e.g. temozolomide, as well as with radiotherapy (Hegi et al., 2005; Hegi et al., 2008). MGMT methylation status is currently evaluated in many centers.

As mentioned above, the receptor tyrosine kinase EFGR is typically over-expressed and mutated in GBM. Yet, treatments with EGFR tyrosine kinase inhibitors have showed little effect. However, it has been shown that the mutation status of the EGFR gene, specifically the EGFRvIII mutation which renders the EGFR constitutively active, as well as the PTEN status may hold prognostic information regarding this type of targeted treatment (Mellinghoff et al., 2005).

Expression of the angiogenesis promoter VEGF has been shown to correlate with malignancy grade (Schmidt et al., 1999). Concerning anti-angiogenesis treatments and targeted therapies against VEGF and its receptor, there are no reliable measures of treatment effect. Microvascular density (MVD), which in itself has shown to be a prognostic marker in GBM (Leon et al., 1996; Yao et al., 2005), is commonly used to assess effects of anti-angiogenesis treatment, however its use as such is debated (Hlatky et al., 2002). Furthermore, circulating endothelial cells (CEC) and circulating endothelial progenitor cells (CEP) may represent candidate predictive markers for this type of treatment (Beaudry et al., 2005; Willett et al., 2004). For example, changes in CEC levels have been correlated to changes in tumor volume during treatment, and low levels of viable CECs have been shown to be predictive for better treatment response in GBM treated with a pan-VEGFR tyrosine kinase inhibitor (Batchelor et al., 2007).

Use of magnetic resonance imaging (MRI) and spectrometry (MRS) may provide useful biomarkers of therapeutic response. For instance, MRI quantification of the Brownian motion of water within glioma tumor tissue after three weeks of treatment, is reported to successfully predict response to chemo- and/or radiotherapy, as assessed volumetrically by traditional imaging means almost two months later (Hamstra et al.,

2005; Moffat et al., 2005). Furthermore, an MRS study of glioma patients undergoing standard treatment indicates that spectral changes during treatment may hold prognostic information regarding tumor response to treatment and outcome (Alexander et al., 2006).

The possibility of utilizing complex biomarker patterns in GBM was recently demonstrated by Persson et al. (Persson, 2008). Based on the expression analysis of a panel of 60 proteins involved in immunregulation, they report that patients may be stratified according to beneficial effect of an experimental immunotherapy. The study also suggests that the treatment effects may be monitored over time.

6. Aims of the study

The overall aims of this study was to investigate the possibility to (i) detect molecular effects following glioblastoma treatment, and (ii) distinguish between different regions of GBM and subtypes of meningioma, by means of non-targeted screening methods. The specific aims to attain this were:

I. To use proteomic and metabolomic screening methods in the search for single biomarkers or patterns thereof, in primary brain tumors.

II. To apply multivariate projection methods for analyses of data acquired from primary brain tumor tissue and microdialysis samples.

III. To investigate molecular changes induced in malignant glioma by conventional radiotherapy on the proteome and in the metabolome.

IV. To investigate proteomic changes in malignant glioma and normal brain tissue induced by experimental anti-angiogenesis treatment.

V. To investigate protein expression patterns as markers of invasiveness in meningioma.

7. Methods

7.1 Samples and sampling

All work in this thesis was based on analysis of biological samples. In papers **I** and **III**, brain tissue samples from an experimental animal model were examined. In paper **II**, microdialysis samples from high-grade glioma patients were analyzed. The last paper (paper **IV**) focused on analyzing meningioma tumor tissue samples from patients with benign meningioma. The following subsections will briefly describe the samples and the sampling procedures.

7.1.1 Animal model

Papers **I** and **III** were based on animal studies, conducted in the BT4C rat glioma model (Bergenheim et al., 1994; Johansson et al., 1998). The model is a transplacental nitrosurea induced rat tumor that has been previously characterized as a gliosarcoma. The BT4C rat glioma cells were grown in fetal calf serum and harvested in log-growing phase, then diluted to a concentration of 4,000 cells/µl, in Dulbecco's modification of Eagle's minimal essential medium (DMEM). A 5 µl aliquot of the cell suspension was injected into the right caudate nucleus of anesthetized BDIX rats, by means of stereotactic equipment and a 22G micro syringe. Thereafter the animals were monitored by an experienced animal keeper for the duration of the experiment, and kept in controlled conditions with 12 h light/dark cycles, freely available with food and water.

The BT4C rat glioma model is both syngenic and orthotopic, meaning that tumor cells originate from the same species and are implanted in the same organ as the cells originated from. Ultimately, the animals were sacrificed and the tumor was cautiously dissected out and snap-frozen in liquid nitrogen. In addition, contralateral normal brain tissue was collected and frozen. The samples were stored at -80°C until analyzed.

7.1.2 Meningioma samples

In paper **IV,** human tissue samples from fibrous and meningothelial meningiomas of WHO grade I were investigated. The samples were collected during surgery. The tissue samples were snap-frozen in liquid nitrogen immediately after resection, and stored at -80°C until analyzed.

The study exclusively included WHO grade I meningiomas that were clearly characterized as invasive or non-invasive, as assessed by means of computed tomography/magnetic resonance (CT/MR) investigations and intra-operative findings.

For this purpose, only tumors that were found to invade into bony tissue by both assessment criteria were classified as invasive. Similarly, only tumors showing no evidence of invasiveness in relation to any other tissue were classified as non-invasive. All patients gave their informed consent to participate and both involved hospitals' ethics committees approved the study.

7.1.3 Microdialysis samples

Microdialysis is a method to sample low molecular weight compounds from within the living tissue of an organ (Ungerstedt, 1997). The method essentially consists of a microdialysis catheter, equipped with a semi-permeable membrane, which is connected to a syringe placed in a micro infusion pump (figure 5). To employ microdialysis for sampling, the catheter is initially positioned within the tissue of interest and slowly perfused with a physiological saline solution. At the location of the semi-permeable membrane, constituent molecules of the extra cellular space can diffuse through the membrane and mix with the perfusion solution. As such, the system mimics a capillary blood vessel, in that it exchanges constituents with the tissue extracellular space. The dialysate (i.e. perfusion medium) is then collected in micro vials, and its molecular composition subsequently analyzed.

The microdialysis technique of sampling has been adopted for use in human brains (Meyerson et al., 1990). Because the microdialysis approach allows for continuous sampling over time it is ideally suited for longitudinal studies of patients, especially in organs where biopsies can not readily be taken, such as the brain (Tabatabaei et al., 2008). Here, in paper **II**, it was utilized to collect extra cellular fluid from high-grade glioma patients before and during 5 days of conventional radiotherapy treatment. The catheters were implanted by means of a stereotactic technique, to allow for reliable sampling from different intra-cranial regions (Roslin et al., 2003).

The included patients all had a radiological suspicion of high-grade glioma, considered not suitable for surgical resection. They underwent a stereotactic biopsy followed by tissue diagnosis before the catheters were implanted along the biopsy trajectory. One catheter was placed in the contrast enhancing tumor region and one in the brain adjacent to tumor (BAT) region (figure 5). In addition, a reference catheter was implanted subcutaneously on the patient's abdomen. The semi-permeable membranes of the catheters had a 100 kDa cut-off, theoretically allowing molecules with a molecular weight smaller than that of the cut-off to freely diffuse across the membrane. The perfusion medium used was Ringer solution (Perfusion fluid T1; CMA Microdialysis, Stockholm, Sweden) mixed with Dextran (30 g Dextran 60/1000 ml) to prevent microfiltration (Hillman et al., 2005), and the flow rate was set to 0.3 µl/min. The radiotherapy began within two to five days after the microdialysis catheters were implanted, and the schedule was normally 2 Gy x 30 (i.e. 2 Gy fractions daily, for 30 days), and the patients were monitored by microdialysis for the first five days of treatment. During this time microdialysate samples were collected every second hour,

and stored at -80°C until analysis. The included patients all gave their informed consent and partook in the study voluntarily. The study was approved by the Regional Ethical Review Board in Umeå.

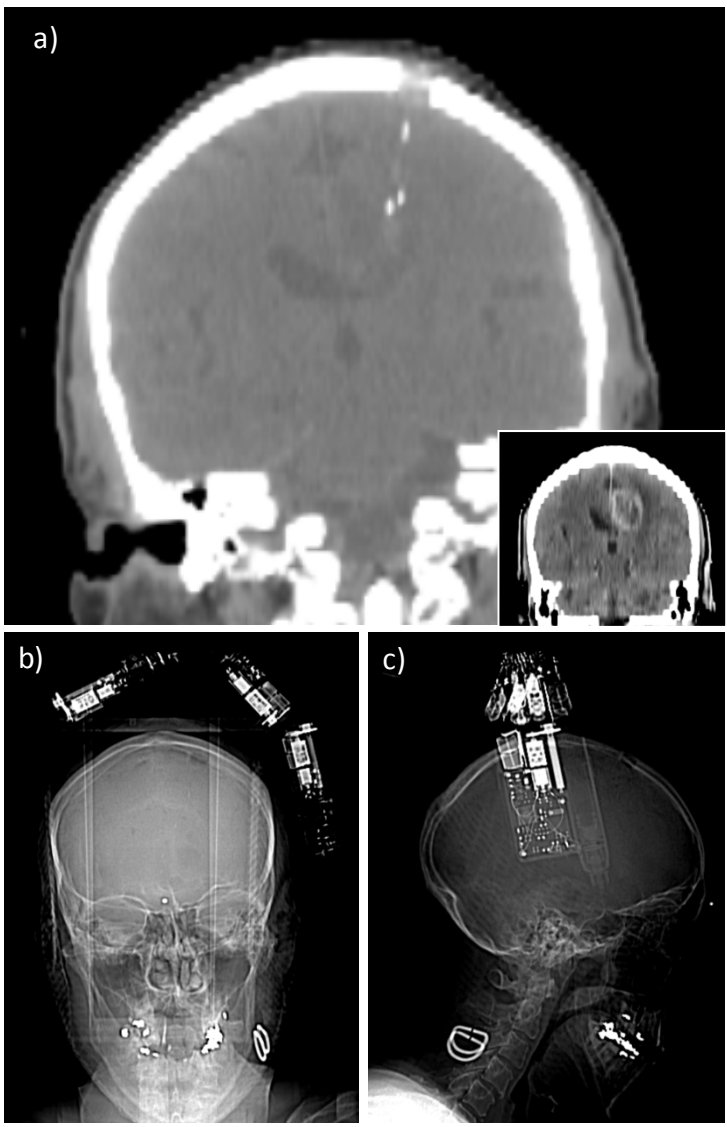

Figure 5. a) Coronar reconstruction of dose planning CT, showing implanted microdialysis catheters. Two catheters were placed within the contrast enhancing tumor and one in the BAT region (see inset panel for preoperative contrast enhanced coronar CT reconstruction). **b) and c)** Coronar and sagittal CT scouts of a patient with micro infusion pumps in place.

7.2 Instrumentation

In all studies presented herein, the samples were analyzed by either surface enhanced laser desorption/ionization - time of flight - mass spectrometry (SELDI-TOF-MS) (papers **I**, **III** and **IV**), or gas chromatography - time of flight - mass spectrometry (GC-TOF-MS) (paper **II**). In this work, both these methods represent screening approaches, in the sense that they aim to quantify and characterize as many molecules as possible in each sample, to generate a profile representative of that sample. The fundamental analysis techniques that are they based on (i.e. chromatography and mass spectrometry) are very broad, with a large range of application areas and many different technical solutions to each involved step. The following subsections aim to introduce the methods as they were applied in the studies presented here.

7.2.1 Mass spectrometry

Mass spectrometry (MS) is the collective term for techniques concerned with determining the masses of atoms or molecules. It is an immensely useful technique that is commonly applied in a variety of fields, including identification of chemical structures (Papac & Shahrokh, 2001) and determination of the composition of mixtures (e.g. detection of proteins in complex biological samples) (Issaq et al., 2002). A mass spectrometer typically consists of three integral parts: (i) an ion source, (ii) a mass analyzer and (iii) a detector, all connected to a computer that controls the system and enables registration and evaluation of obtained data.

The ion source is the part of the mass spectrometer where the molecules in the sample are ionized, producing charged molecular species. There are several different ionization methods available, including matrix assisted laser desorption/ionization (MALDI) and electron ionization (EI), which are the methods that were used here. The choice of ionization method depends primarily on the analyte (i.e. the sample to be analyzed) and the purpose of the MS analysis, a topic that will be further discussed in the following subsections.

After ionization, the charged species in the sample are directed into the mass analyzer, which acts to separate the species in the sample based on their mass to charge ratio (m/z) (El-Aneed et al., 2009). In similarity with the ionization step, this step too can be accomplished by a multitude of different techniques. In this work, however, only one type of mass analyzer was used, namely time of flight (TOF) (figure 6). A TOF instrument is defined by the International Union of Pure and Applied Chemistry (IUPAC) as one that makes use of the fact that ions with different mass to charge ratios require different times to traverse a given distance in a field-free region, provided that the ions have the same initial translational energy. The initial energy is gained by subjecting the analyte ions to an electric field of known strength and distance, where each ion gains electric potential energy (E_p) proportional to the number of charges it carries (z) (Merchant & Weinberger, 2000), according to:

(Eq. 1) $$E_p = zV$$

where V is the potential of the electric field (i.e. the voltage). As the ions accelerate through the electric field the potential energy is converted to kinetic energy (E_k), which in turn is directly proportional to the ions' mass (m) and velocity (v). This relationship is generally expressed as follows:

(Eq. 2) $$E_k = \frac{mv^2}{2}$$

From Eq. 2, given that two ions of different mass ($m1<m2$) are given the same kinetic energy (as they are when the number of charges are the same, $z1=z2$ in Eq. 1), it is easily deducted that the one with the smaller mass must have a greater velocity ($v1>v2$). After exiting the electric field, the accelerated ions enter into a drift tube of known length. The drift tube is a field-free region, hence no forces other than the ions' velocity act to determine the time (t) it takes for the ions to traverse the distance (d) to the end of the tube. The time is given by Eq. 3.

(Eq. 3) $$t = \frac{d}{v}$$

Consequently, ions with smaller masses, and thus greater velocities, reach the detector at the end of the flight tube before larger ions travelling at lower velocities. Taken together, by combining and rearranging Eq. 1–3, it is possible to express an ion's flight time (t) in terms of its mass, as follows:

(Eq. 4) $$t = \sqrt{\frac{d^2 m}{2Vz}} = k\sqrt{\frac{m}{z}}$$

where k denotes a constant that encompasses factors associated to the properties and settings of the instrument. k may be determined through calibration of the MS instrument, which is generally achieved by analyzing molecular species with known molecular masses.

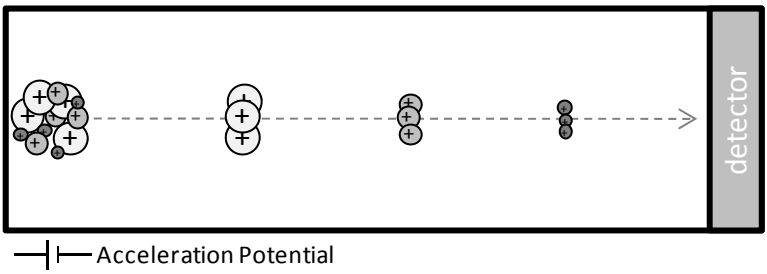

Figure 6. Schematic representation of a TOF mass analyzer. Gaseous analyte ions are accelerated in an electric field. They then drift through a vacuum tube with a detector located at the end. The time it takes for a specific ion to reach the detector is measured. Based on the time of flight the ion's m/z ratio can be calculated (Eq. 4).

Finally, after time separation, the analyte ions reach the detector at the end of the drift tube (figure 6). The detector measures the current produced as the ions reach it and converts it to a signal that is registered by a computer. The signal from the detector is recorded with a high frequency, and plotted in a graph (mass spectrogram), where the time point of the recorded signal, after its recalculation to mass units (Eq. 4), is represented by the horizontal axis and the amplitude of the signal is represented by the vertical axis (figure 7). The signal's amplitude is proportional to the number of ions that reach the detector at any given time point, hence when a particular ion species reaches the detector, the amplitude spikes and results in a peak in the spectrogram. The amplitude, or the area under the peak, may be used to quantify an ion species.

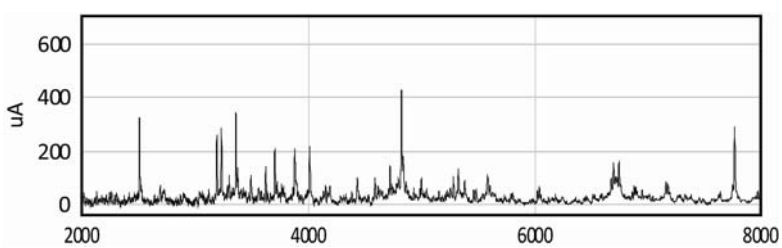

Figure 7. A typical mass spectrogram, produced by SELDI-TOF-MS. The horizontal axis represents time of flight, recalculated to mass to charge (m/z) (Eq. 4). The vertical axis corresponds to the detector signal output, which is proportional to the relative abundance of the separated ion species. *The figure is adapted from paper **IV**, with permission.*

7.2.2 Surface Enhanced Laser Desorption/Ionization

Although mass spectrometry was invented at the beginning of the 20th century, it was not until the invention of so called soft ionization techniques in the 1960's that it became possible to apply it for analyzes of larger molecules, such as peptides and

proteins (El-Aneed et al., 2009). The most important development in this field was the development of matrix assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) in the late 1980's (Zhou & Veenstra, 2008), which enabled researchers to readily analyze biological material. Prior to the advent of soft ionization, all available ionization procedures applied to large molecules either could not ionize them at all, or produced highly unstable ions that instantly fragmented in the ion source.

As mentioned, MALDI (Karas et al., 1987; Tanaka et al., 1988) is one example of a soft ionization method. The MALDI process can be illustrated in four consecutive steps: (i) mixing of the sample to be analyzed (i.e. the analyte) with an energy absorbing molecule (EAM), also referred to as a matrix, (ii) co-crystallization of the analyte matrix mixture on a surface, (iii) application of a laser pulse to the surface, and (iv) formation of gaseous analyte ions. The last step is a result of the laser pulse energy being absorbed by the matrix molecules and subsequently transferred to the analyte, which consequentially is desorbed and ionized in one step (This processed is thoroughly reviewed by *Zenobi and Knochenmuss* (Zenobi & Knochenmuss, 1998)). The gaseous analyte ions then enter the mass analyzer, typically a TOF, as described above.

Employing conventional MALDI in combination with a TOF mass analyzer (MALDI-TOF-MS) is an established method to analyze complex biological samples, such as blood plasma for instance (Hortin, 2006). However, to allow for meaningful analysis of such complex samples, they need to be pre-fractionated and purified prior to MS analysis.

In papers **I, III** and **IV**, SELDI-TOF-MS (Hutchens & Yip, 1993) was used to characterize the samples. SELDI is a variant of MALDI that also allows for analysis of large molecules. The "surface enhanced" part of the acronym refers to the surface that is used to introduce the sample to the MS system. In contrast to conventional MALDI, where the sample is presented to the MS system on a passive surface, SELDI utilizes an active surface that is used both for sample pre-fractionation and purification as well as for introduction to the MS system. The surface is designed to actively retain subsets of molecules in a sample based on their chemical properties, whereas molecules with different properties may be washed away together with various contaminants. This way both pre-fractionation and purification of the sample is obtained. The retained molecules are subsequently dissolved and co-crystallized with the matrix solution, whilst still on the active surface, before being introduced to the MS system.

In practical terms this is done by means of the commercially available ProteinChip® arrays on which the SELDI method is based. The arrays contain spots with chromatographically active surfaces, designed to retain molecules with specific chemical properties. There is a range of array types with different chromatographic properties to choose from, including anion exchange, cation exchange, reverse phase chromatography, and immobilized metal affinity chromatography (IMAC). The

SELDI-MS system is fitted with an interface suitable for the ProteinChip® arrays, allowing characterization of the retained molecules directly off the ProteinChip® spot surfaces.

7.2.3 Gas Chromatography

Gas chromatography (GC) is a widely used analytical technique that aims to separate molecules in a sample in order to quantify and/or identify them individually. Rather than separating the molecules based on their different masses, as is the case with the previously described TOF-MS, GC does so based on other physico-chemical properties of the analyte molecules (further explained below). A GC system generally contains five integrated components: (i) a carrier gas delivery system that continuously pumps an inert gas (e.g. helium) through the column, (ii) an injector, through which the sample is injected into the system, (iii) an analytical column, which essentially is a narrow tube coated on the inside with a specific material, (iv) an oven that contains the column and controls its temperature, and (v) a detector (figure 8).

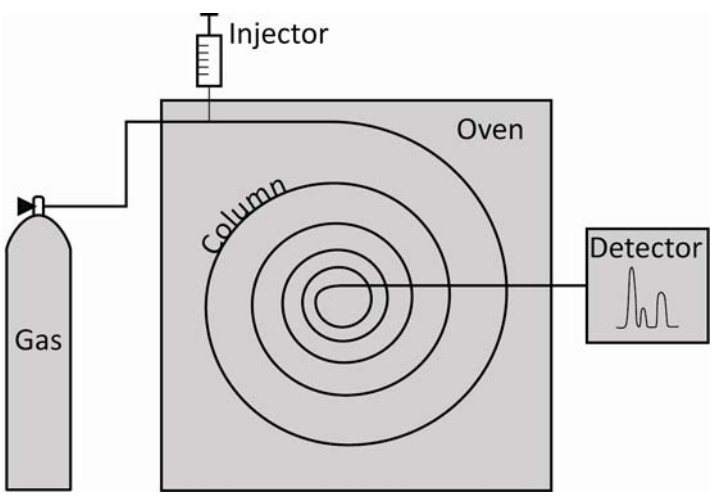

Figure 8. Schematic representation of a GC system, with the five most central parts indicated.

Once a sample is injected into the column, the different molecular species in the sample are vaporized and carried through the column by the constant flow of the carrier gas (i.e. the mobile phase). Whilst travelling through the column, the analyte molecules may interact with the stationary phase, i.e. the material making up the coating on the inside of the column walls. Eventually, the analyte molecules elute from the column and enter the detector, where the quantity of each molecular species is determined in relative terms. The time it takes a molecule to travel through the column is called retention time and is largely determined by the molecule's interaction with the stationary phase

(Pasikanti et al., 2008). Because these interactions differ between different molecules, it is possible to use GC to separate the molecules in a sample and generate a profile reflecting the molecular contents of that sample. The profile is called a chromatogram, and is basically a graph with the retention time on the horizontal axis and the detector response (i.e. relative quantity) on the vertical axis (figure 9a).

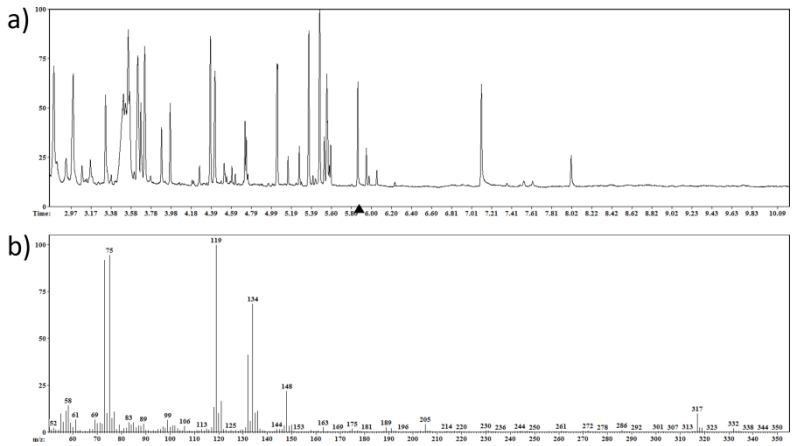

Figure 9. Graphical representation of the output from GC-MS analysis of one sample (a tumor microdialysis sample, included in paper **II**). **a)** A total ion current (TIC) chromatogram representative of the sample, with retention time plotted on the horizontal axis and signal response (i.e. relative quantity) on the vertical axis. **b)** A mass spectrum (plotted from 50 to 350 m/z), collected at the specific retention time marked by the triangle in panel **a** (retention time = 5.864 min).

For separation in GC systems, column temperatures up to 325°C are commonly used (Pasikanti et al., 2008). Therefore, only molecules that are thermally stable and have reasonably low boiling points lend themselves to separation by GC. In practice, this means molecules that are relatively small (~<500 Da) and non-polar (Halket et al., 2005). In a biological context this translates into the metabolites' family of molecules, and consequently makes GC one of the main analytical methods applied in the field of metabolomics (Dunn & Ellis, 2005). To increase the range of molecules in a sample that can be successfully analyzed by this approach, polar molecules are often derivatized to increase volatility and improve GC properties, prior to analysis. In paper **II**, two derivatization steps were used: oximation and silylation. Oximation involves transformation of carbonyl groups to corresponding oximes, in order to prevent enolization (i.e. the transformation of ketons to enols). This could otherwise result in the same molecule existing in several isomeric forms in the sample, and thereby complicate the chromatogram (Halket et al., 2005). Silylation is the process of substituting a proton on polar functional groups of the analyte, such as OH-, NH-, or SH-, by an alkylsilyl group (e.g. trimethylsilyl (TMS), which was used here). By these means the analyte's volatility and thermal stability is increased (Dettmer et al., 2007).

7.2.3.1 Gas Chromatography coupled to Mass Spectrometry

Several different techniques can be applied for detection of the eluted molecules. By employing an MS system as the GC detector (as was done in paper **II**) it is possible to both quantify and further characterize the eluting molecules structurally. To this end, a hard ionization method is warranted, such as electron ionization (EI). This involves bombardment of the analyte with electrons, which is thereby ionized and fragmented prior to introduction to the mass analyzer (Pasikanti et al., 2008). The MS system may be equipped with any of a range of mass analyzers. In the work presented here, the previously described linear TOF was applied, which separates the analyte fragments based on their individual m/z ratios, resulting in a mass spectrum that represents the fragmentation pattern of the analyte (figure 9b). This pattern is unique to each molecular species, as a consequence of the differences in chemical structure between them, and may as such be considered a molecule's fingerprint. It can in turn be used to search databases containing mass spectra obtained from molecules analyzed under the same experimental conditions. Provided that the analyzed molecule is represented in the database, it is possible to identify the analyte. In contrast to the fingerprint analogy *de novo* interpretation of fragmentation patterns can also be used for identification (Dunn & Ellis, 2005).

The MS system scans the GC effluent with a certain frequency (for instance 30 times per second, as in paper **II**). The total ion current (TIC) of each scan can be plotted against the retention time, ultimately resulting in a TIC chromatogram representative of the sample (figure 9a). The detected analyte molecules in the chromatograms are subsequently quantified by calculation of the area under the chromatogram peak that represents the analyte of interest (see section 7.5).

7.3 Data analysis using multivariate projection methods

The amount of data generated by chromatographic and/or mass spectrometric characterization of samples is typically very large. During the analytical process, hundreds or thousands of measured data points (variables) can be associated to each sample (observation). For practical reasons, this data is normally assembled and stored in data tables (also called data matrices), where each observation is represented by one row in the table and each of the descriptor variables is represented by one column. This data assembly is only possible if all samples in a study are characterized by the same, well defined, descriptor variables. The data matrices holding sample information are commonly denoted **X**, where the size of **X** is given by the number of samples (N) and the number of descriptive variables (K) used to characterize each observation in the study (figure 10a).

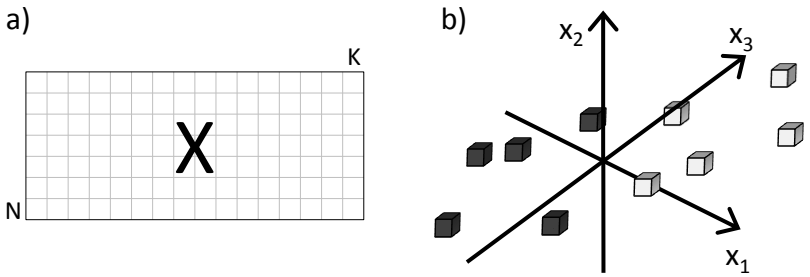

Figure 10. a) Illustration of a $N \times K$ sized matrix. **b)** If K=3, the data can easily be represented by a scatter plot in a 3-dimensional co-ordinate system, where interesting features in the data can be observed.

A data table with K columns can be represented by a point swarm in K-dimensional space (figure 10b). However, when K>3, this is not helpful in terms of data interpretation. To interpret the data in a large matrix, appropriate tools are needed.

In all papers presented herein, multivariate projection methods have been applied to analyze and interpret the acquired data. These methods all attempt to reduce the dimensionality of the original data by finding latent variables onto which the original data is projected to obtain a facilitated interpretation of the complex interactions often occurring in biological data of multivariate character. The latent variables can be viewed as hidden structures in the data, i.e. variables that are not directly measured but systematically affect the output data. For instance, consider standard blood samples that are screened with regard to metabolite or protein content. Although factors relating to the individual patients from whom blood was drawn (such as gender, age, disease state, etc.) are not directly measured, it is likely that they are systematically reflected in the samples. This underlying variation may cause structures in the data, which can be found and explained by projection to latent variables.

When the data analysis is applied with the specific purpose of distinguishing between predefined classes of samples, the analysis approach is termed classification. All studies presented here are solely concerned with multivariate data analysis for classification (e.g. distinguishing between tumor and normal samples, tumors of different growth characteristics, or between samples from treated and samples from untreated animals or patients). To this end, there are two conceptually different approaches to find the latent variables in a dataset, namely unsupervised and supervised methods. The supervised methods rely on additional input that defines the sample classes beforehand, and will then extract systematic variation that separates the predefined classes, if such variation exists. Unsupervised analysis methods do not use this type of additional input, but will instead extract the main systematic variation in the data, with no regard to possible sample classes.

7.3.1 Principal Component Analysis

Principal Component Analysis (PCA) (Hoskuldsson, 1995; Wold et al., 1987) is an unsupervised multivariate projection method, commonly used to obtain an overview of complex datasets. Ultimately, PCA reduces the dimensionality of the data by means of latent variables, which allows for the data to be represented by meaningful and easily interpreted plots in only two or three dimensions. Metaphorically, PCA can be described as opening up a window into the multidimensional space. Depending on the position and the angle of the window, the observer looking through it will get different views of the multidimensional reality.

Conceptually, the dimensionality reduction is achieved through construction of a new co-ordinate system, of fewer dimensions than the original one, to which each observation is assigned a new value. The new co-ordinate system is based on so-called principal components, which are calculated in a stepwise manner according to specific criteria: (i) the first principal component should explain as much as possible of the variation in the dataset, (ii) subsequent components should explain as much as possible of the variation left unexplained by the previous component(s), and with the added constraint that they must be orthogonal to the previous principal component(s). This set of principal components can be said to span a hyper plane in *K*-dimensional space. Each observation is then assigned values associated to the principal components, called score values, by means of projection onto the hyper plane (figure 11a). The euclidian length of a principal component, i.e. its eigenvalue, is a direct quantification of the amount of variation explained by it.

By plotting this new co-ordinate system and the associated score values, in a so-called PCA score plot (figure 11b), it is possible to obtain a good overview of the main variation in the original dataset and thereby observe possible groupings, trends, and outliers in the data. For the purpose of classification, the possible clustering of different sample classes in score space can be interpreted as the presence of systematic variation among the descriptor variables that can be used to discriminate between the sample classes.

When the principal components are fitted, it is also determined to which degree each of the original variables contributes to the direction of each principal component. The angle between a descriptor variable (i.e. an axis in the original co-ordinate system) and a principal component is a measure thereof (figure 11a). A small angle indicates a high degree of similarity between the descriptor variable and the principal component, and vice versa. The cosine transformation of this angle is called the loading value (p) and is stored separately. These loading values constitute the link between the principal components, thus also the score values, and the original data, and are important for interpretation of features observed in the score plot. For example, if the first score vector ($\mathbf{t}_1$) can be used to discriminate between the sample classes (as in figure 11b) the following two conclusions can be drawn: (i) the largest variation in the data is that

between the different sample classes and (ii) the descriptor variables with the largest loading values with respect to $\mathbf{t}_1$ (i.e. $\mathbf{p}_1$ values) are the original variables that contribute the most to the observed separation.

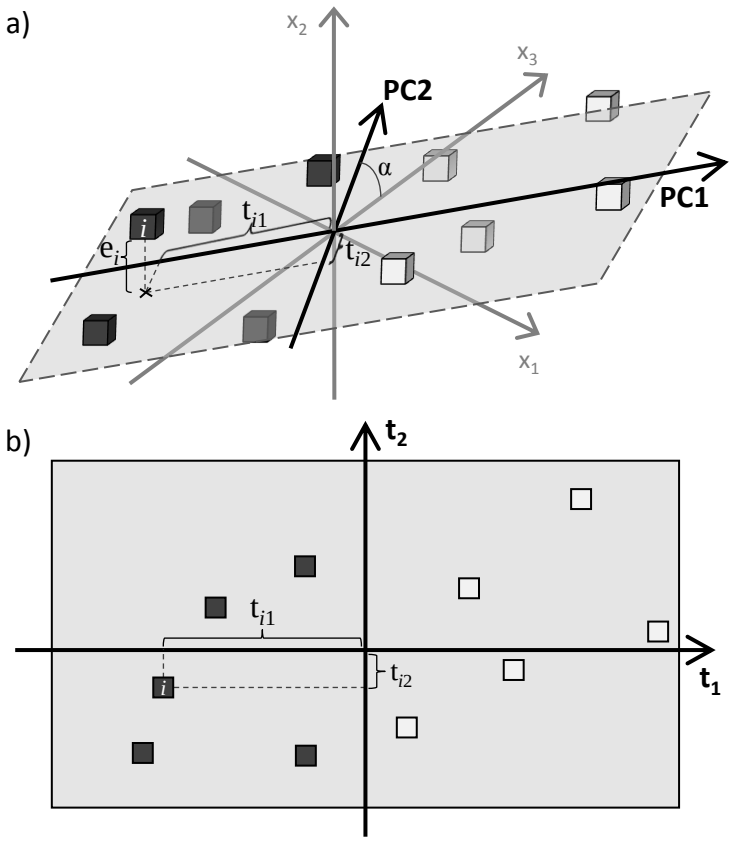

Figure 11. a) Two Principal Components (PC1 and PC2) have been fitted to data in a 3-dimensional co-ordinate system. Observation *i* is projected onto the hyper plane spanned by the principal components, and assigned score values t_{i1} and t_{i2}. The distance between observation *i* and the plane represents the residuals for this particular observation (e_i). The cosine transformation of the angle α represents the loading value for the original descriptor variable x_3 to the second principal component (PC2). Corresponding angles between descriptor variables and principal components (not marked in the figure) are transformed in the same way to calculate loading values for all combinations of descriptor variables and principal components. **b)** A schematic illustration of a PCA score plot, where the first and second score vectors ($\mathbf{t}_1$ and $\mathbf{t}_2$) are plotted against each other. The main variation of the original data is conserved, and the clustering of black and white observations is still obvious, but the dimensionality has been reduced from three to two dimensions.

The distance between each observation and the hyper plane represents the variation in the original data that is left unexplained by the PCA model (figure 11a). This variation is called the model residuals.

The principal components are determined either by Singular Value Decomposition (SVD) (Jolliffe, 2002) or by the NIPALS algorithm (Wold, 1966), where the latter was used in the work presented here. In terms of linear algebraic equations, the PCA model can be formulated as follows:

(Eq. 5) $$\mathbf{X} = \mathbf{T}\mathbf{P}' + \mathbf{E} = \sum_{a=1}^{A} \mathbf{t}_a \mathbf{p}'_a + \mathbf{E}$$

where **X** is the original data matrix, **T** the matrix of score vectors ($\mathbf{t}_a$), **P'** the transpose of the matrix with loading vectors ($\mathbf{p'}_a$) and **E** the residual matrix. The index *A* denotes the total number of principal components calculated.

7.3.2 Partial Least Squares

Partial least squares projection to latent structures (PLS) (Wold et al., 2001a; Wold et al., 2001b) is a multivariate regression method, useful to relate the descriptor variables in **X** to a set of response variables (**Y**). The response variables are particular properties of the observations in **X**. A typical regression equation formulating the linear relationship between **Y** and **X** is expressed in Eq. 6.

(Eq. 6) $$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{F}$$

where **F** represents the residuals and **B** the regression coefficients. This equation implies that (i) by measuring both the response properties (**Y**) and the descriptor variables (**X**) for a set of samples, it should be possible to solve the equation for **B** and (ii) once **B** is known it would suffice to measure the descriptor variables for future samples and use them in combination with **B** to predict **Y** for the new samples.

For example, let **Y** be a set of response variables describing the treatment response for a set of patients and **X** the metabolic or proteomic profile of samples collected from the same patients. If one could then solve Eq. 6 for **B**, and apply this solution to a new set of patients, it would be possible to estimate a patient's treatment response (**Y**) based on the metabolic or proteomic content of a sample from this patient (**X**).

The key to the reasoning above lies therefore in solving **B**. One way to do this is by means of Multiple Linear Regression (MLR). Through a set of algebraic operations to Eq. 6 the solution to **B** then becomes:

(Eq. 7) $$\mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

where **X'** is the transpose of **X**, and $(\mathbf{X'X})^{-1}$ denotes the inverse of the matrix **X'X**. However, there exists no inverse of matrices that contain co-linearity among its variables, or if it has more columns than rows (i.e. more variables than samples ($K>N$)). In most realistic scenarios, especially when the descriptor matrix consists of data from samples characterized with mass spectrometry or chromatography, **X** will contain many more variables than samples ($K>>N$). On top of that, the variables will display a large degree of co-linearity. Hence, $(\mathbf{X'X})^{-1}$ does not exist, and the MLR approach is not applicable to solve **B**.

There are a few different methods to handle this problem, where PLS is one of them. PLS first finds a set of latent variables (scores (**T**)), similar to the principal components described earlier. The scores are then employed to successfully solve the problem described above. Essentially, $(\mathbf{X'X})^{-1}$ is replaced by $(\mathbf{T'T})^{-1}$, which always exists as a result of the linearly independent (orthogonal) properties of the score vectors in **T**.

PLS focuses on describing the main variation in **X** that can be used to predict **Y**. Therefore, in contrast to PCA, where the latent variables (principal components in that case) are calculated to maximize the variation explained in **X**, the latent variables (scores) in PLS are calculated to maximize the co-variation between the latent structures of **X** and **Y**. The PLS scores are formed as linear combinations of the original variables in **X**, according to $\mathbf{t}_a = \mathbf{X}\mathbf{w}_a$, where $\mathbf{t}_a$ represents the score vector and $\mathbf{w}_a$ the corresponding weight vector (also known as the co-variance loading). For instances where **Y** consists of a single response (**y**), the scores are calculated to constitute good estimators of both **X** and **Y**, in the sense that the residuals from both matrices should be small (**E** and **F** in Eqs 8 and 9, respectively).

(Eq. 8) $$\mathbf{X} = \mathbf{T}\mathbf{P}' + \mathbf{E} = \sum_{a=1}^{A} \mathbf{t}_a \mathbf{p}'_a + \mathbf{E}$$

(Eq. 9) $$\mathbf{Y} = \mathbf{T}\mathbf{C}' + \mathbf{F} = \sum_{a=1}^{A} \mathbf{t}_a \mathbf{c}'_a + \mathbf{F}$$

where **X** denotes the original descriptor data matrix, **Y** the response data matrix, **T** the matrix of latent variables ($\mathbf{t}_a$), **P'** and **C'** represent the transpose of the matrix with loading vectors ($\mathbf{p'}_a$ and $\mathbf{c'}_a$) for **X** and **Y**, respectively, and **E** and **F** the residual matrices of **X** and **Y**, respectively. A is the total number of latent variables calculated.

The PLS score vectors are calculated in a stepwise manner. When score vectors subsequent to the first one are calculated, **E** and **F** replace **X** and **Y**, meaning that each new vector will explain the co-variation left unexplained by the previous vector(s). For

interpreting features observed in PLS score plots, the **X**-related PLS weights **w** are the most important (Johansson et al., 2003). The first **w** weight vector ($\mathbf{w}_1$) holds information regarding how the original descriptor variables combine to for the relation to **Y**, where descriptor variables with large absolute $\mathbf{w}_1$-values ($|\mathbf{w}|$) have a large impact on $\mathbf{t}_1$, and vice versa. Subsequent weight vectors ($\mathbf{w}_a$) are interpreted in a similar way, with the difference that they originate from the comparison of **E** and **F**, rather than that of **X** and **Y**. However, for score vectors subsequent to the first, a weight relating to the original variables (**X**) is also calculated, commonly denoted **w***.

When PLS is used for classification it is called PLS discriminant analysis (PLS-DA), and the **Y**-matrix is constructed by dummy variables used to divide the observations in **X** into predefined classes. Hence, PLS-DA is a supervised classification method.

7.3.3 Orthogonal Partial Least Squares

In all empirical settings, the measured data is likely to contain a large amount of variation that is uncorrelated to the responses of interest. For instance, if patient samples are collected at the clinic and subsequently characterized by means of metabolite profiling, the data output is likely to contain a lot of different variation from several possible sources. The sources of variation may be related to the heterogeneity of the patients (such as gender, age, etc.), or variation resulting from differences in sample handling (such as storage time) as well as random noise from the analysis. If the study is designed as a discriminant analysis with a single response (**y**), where the objective may be to discriminate between patients suffering from cancer and healthy controls, all other sources of variation could obscure the information of interest. Applying PLS-DA to the problem may then result in the y-correlated variation (i.e. the variation of interest, which may be used to predict **y**) being split up into several different latent variables, and mixed together with the other sources of y-orthogonal systematic variation (figure 12a). This will not inflict on the models predictive ability of the response variation, however, the interpretation of such a model can be misleading (Trygg & Wold, 2002).

Orthogonal Partial Least Squares (OPLS) is a further development of PLS, where an orthogonal signal correction filter (Wold et al., 1998) has been made an integral part of the PLS algorithm (Trygg & Wold, 2002). It aims at dividing the systematic variation in **X** into two parts; one that is linearly correlated to **Y** (and hence can be used to predict **Y**) and one that is uncorrelated (orthogonal) to **Y**. This enables interpretation of the two sorts of variation separately, which improves the interpretability of the model (Stenlund et al., 2009). The OPLS decomposition of **X** is expressed in Eq. 10:

(Eq. 10) $$\mathbf{X} = \mathbf{T}_p\mathbf{P}_p' + \mathbf{T}_o\mathbf{P}_o' + \mathbf{E}$$

where the indexes *p* and *o* denote the predictive and orthogonal part, respectively.

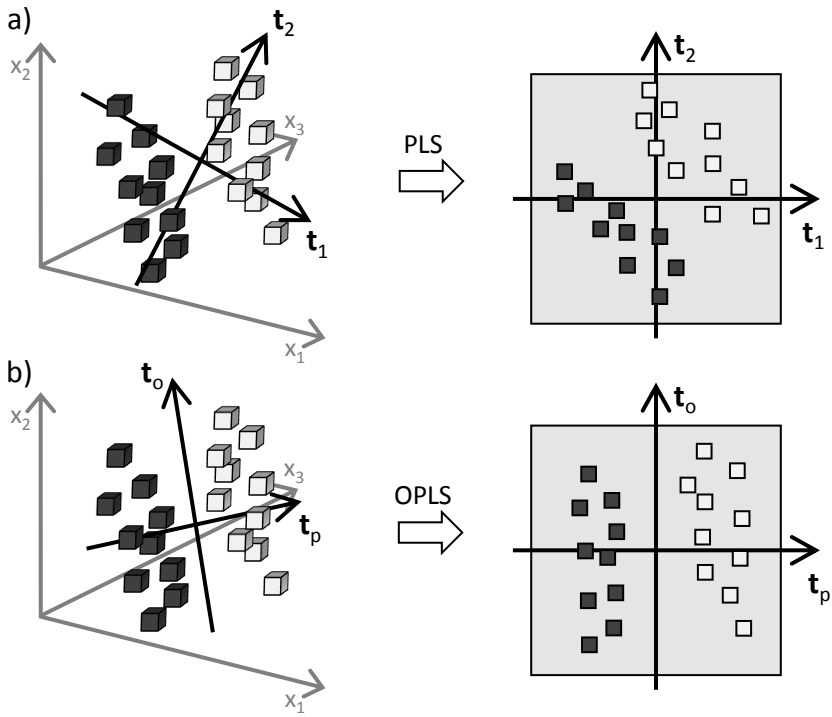

Figure 12. Schematic representation of the differences between PLS-DA and OPLS-DA for distinguishing between two classes of samples. **a)** When PLS is applied, the systematic variation within the different classes results in the y-predictive variation being divided between both latent variables ($\mathbf{t}_1$ and $\mathbf{t}_2$), making interpretation difficult. **b)** OPLS models the y-predictive and the y-orthogonal variation separately, resulting in a separation of the two classes solely based on one latent variable ($\mathbf{t}_p$), and the systematic variation uncorrelated to **y** is modeled by the second latent variable ($\mathbf{t}_o$). OPLS enables investigation of the two types of systematic variation separately, facilitating interpretation.

Similarly to PLS-DA, the classification implementation of OPLS is termed OPLS discriminant analysis (OPLS-DA) (Bylesjö et al., 2006) and is a supervised classification method. Applying OPLS-DA to the example above would result in all the systematic y-correlated variation in **X** being explained by a single score vector ($\mathbf{t}_p$). The remaining systematic variation incorporated into **X** would be modeled by a separate set of **y**-orthogonal score vectors ($\mathbf{T}_o$) (figure 12b), which facilitates model interpretation as well as detection of original variables contributing to the class separation. The latter can be achieved by studying the model loadings associated with the predictive components.

7.3.4 Validation

In any regression or classification situation, the ability to use the multivariate model to predict the response for new observations is of great importance. Considering that

systematic variation stems from some underlying cause, which is common between the observations and will be the same for future samples, and that random noise is unpredictable, the following reasoning is quite intuitive: In order to achieve the best possible predictive ability, a model should describe all systematic variation in the dataset that is correlated to the response, without including any of the random noise.

As previously described, the latent variables are calculated stepwise, where each new component explains a portion of the variation in the original data that has not been described by previous components. This process should stop when all systematic variation is accounted for, leaving the random noise in the model residuals. If too many latent variables are calculated, the model is "over fitted", meaning that random noise in **X** has been incorporated into **T**. This will affect the model's predictive ability negatively, since the noise will not be reproducible for future samples. In the classification situation, particularly when applying supervised methods, over fitting may also result in apparent discrimination between the observation classes due to chance correlations between the noise and the response. On the other hand, calculating too few latent variables ("under fitting") will also affect the model's predictive ability negatively. To determine the appropriate number of latent variables requires careful validation of the model.

Cross-validation (Stone, 1974; Wold, 1978) is an example of an internal validation procedure, commonly used to estimate the appropriate number of latent variables. The procedure involves excluding a subset of the observations in **X**, and calculating the latent variables based on the remaining ones. Typically a random selection corresponding to 1/7 of all observations in **X** is excluded in each CV round, however different sized subsets can be excluded. If one observation at a time is removed the procedure is referred to as "leave one out cross-validation". The responses of the excluded samples are then predicted by the fitted model. This procedure is then repeated in a stepwise manner, with a new subset of samples excluded each time, until every observation has been excluded exactly once. At each step, the predicted values for the excluded samples are compared to its observed values. The results of these comparisons are ultimately summarized in a term called prediction error sum of squares (*PRESS*) (Eq. 11).

(Eq. 11)
$$PRESS = \sum_{i=1}^{CVgrp} (y_i - \hat{y}_i)^2$$

where *CVgrp* represents the number of observations excluded in any given cross-validation round, y_i represents the observed response for observation i and $\hat{y}_i$ its corresponding predicted value. In essence, the *PRESS* value can be used as a direct measure of a model's predictive ability, and the number of latent components in a model can be adjusted to maximize this property, i.e. minimizing PRESS.

Ideally, a model's predictive ability should also be tested on an external test set (i.e. a set of observations that were excluded during the entire calculation of the model). How well the responses of observations in this set are predicted would obviously be the most relevant evaluation of the model. However, when samples in a study are few, and depending on the study objective, the study may benefit from including all observations in the calculation of the model (thereby achieving a more representative study population, at the expense of external validation). When this is the case, examining and presenting the cross-validated score values (i.e. the score values that is predicted for each observation during the cross-validation round when the observation in question is excluded), in so called cross-validated score plots, is as close to external testing as it is possible to get.

Furthermore, whenever it is applicable, the validation of any given model should also include a biological aspect. In other words, the results should be interpreted in terms of biological relevance, and when possible also functionally validated.

7.4 Analysis of SELDI-MS spectra

SELDI mass spectra are typically generated in search of features, i.e. protein levels, distinguishing between samples from different populations. Naturally, as this is a screening approach, the chance to detect features of interest increases with the number of features included in the investigation. Seeing as SELDI is designed to capture and analyze only a subset of the proteins in a sample, it is not uncommon that researchers wish to examine more than one such subsets. At the same time, SELDI is a high throughput method that requires fairly small sample volumes for each analysis, making it feasible to design experiments that include several molecular subsets. In practical terms, this can easily be achieved by varying the SELDI conditions, for instance through the use of more than one ProteinChip® array type, or by using buffers with different properties in the preparation of the arrays. Another possibility is pre-fractionation of the samples. Then, by the same logic, analyzing all fractions increases the chance to make relevant findings. Furthermore, the SELDI-MS system may be optimized for different mass regions. Therefore, once a sample is spotted to a ProteinChip® array, it is often analyzed with several different settings to obtain the best quality data for each region. Taken together, this frequently results in numerous mass spectra derived from the same sample.

Papers **I**, **III** and **IV** were all concerned with the analysis of SELDI data from several data subsets. The analysis approach was generally the same in the three papers, although somewhat refined for papers **III** and **IV** compared to paper **I**, and can be outlined as follows: (i) pre-treatment of data, (ii) multivariate characterization of each subset separately, (iii) merging of the most interesting features from the different subsets into one dataset, (iv) further multivariate analysis of the merged features dataset, (v) statistical validation of findings. Below follows a more detailed description of these steps.

7.4.1 Data pre-treatment

To allow for meaningful analysis of the generated mass spectra, each has to be constructed by the same well defined variables. As the values for the m/z variable were inconsistent between the numerous collected spectra, each spectrum was initially subjected to a binning procedure, where (i) small mass intervals (bins) were constructed, (ii) the mean intensity within each bin was calculated for every spectrum, and (iii) the data matrix was constructed by associating the mean values from the numerous spectra to its corresponding bin variable.

Prior to further analysis, the binned spectra were normalized individually, by dividing every bin variable by the total (paper **I**) or mean (papers **III**-**IV**) intensity of the whole spectrum. Lastly, data was centered by subtracting each variable by its mean, and scaled by pareto scaling (papers **III-IV**). Pareto scaling is the process of dividing each variable by the square root of its standard deviation.

7.4.2 Detecting features of interest

Once the SELDI mass spectra have been assembled into proper datasets ($\mathbf{X}_a$, where the index separates the different subsets), initial characterization thereof is generally warranted. This was done by applying PCA to each subset separately, where both the score values and residuals calculated for each observation were scrutinized. The next step included leave one out OPLS (i.e. OPLS where the cross-validation was designed to exclude one observation or one group of analytical replicates, in each round) where the cross-validated score values were examined. Observations not fitting the rest of the data would thereby become apparent. Potential outliers were investigated further, mainly through examination of their individual raw mass spectra, and removed from further analysis if necessary.

OPLS-DA was then applied to the pareto scaled data, consisting of complete spectral profiles of each sample ($\mathbf{X}_a$). Naturally, the response ($\mathbf{Y}$) was made up by dummy variables representing the sample classes of interest. For instance, in paper **III** the responses corresponded to type of tissue (tumor or normal), treatment, and the interaction term between the two. This model was in turn used to define spectral regions of interest (ROI) for distinguishing between the sample classes, essentially by utilizing the S-plot for each predictive component. The S-plot depicts both the modeled co-variation ($\mathbf{w}^*$ loadings) and the correlation ($\mathbf{p}_{corr}$ loadings) of each descriptor variable to a response (Wiklund et al., 2008). Spectral regions consisting of variables at either end of the S-plot (above or below a specific cut off value) were defined as interesting (figure 13).

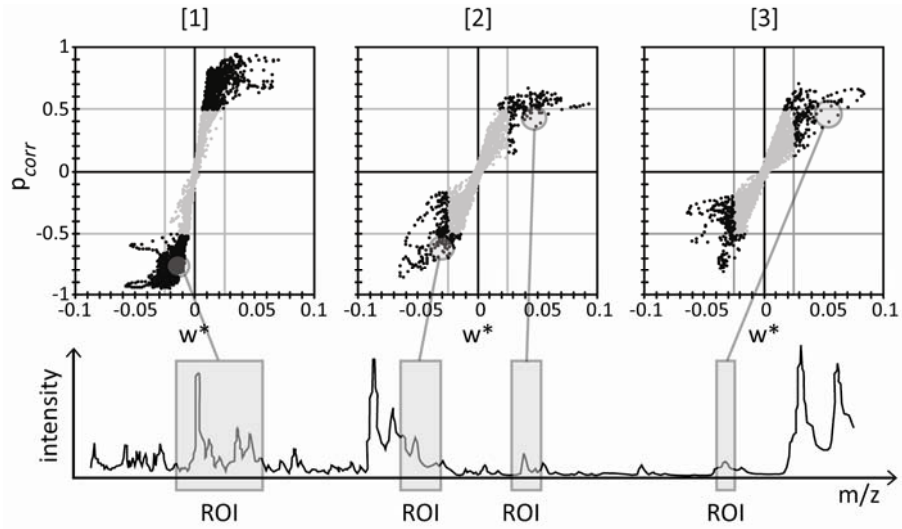

Figure 13. Schematic illustration of how regions of interest (ROI) are defined based on the weights (**w***) and correlation loadings (**p**$_{(corr)}$) from the predictive components of the OPLS-DA model (denoted by numbers in square brackets).

Lastly, variables within a defined ROI that formed local maxima (i.e. peaks) in the average spectral profile of at least one of the defined observation classes in **Y** were located. For each selected peak, corresponding intensity data was extracted from all observations in **X**. In order to compensate for possible mass shifts, this was achieved by defining a window around the peak of interest (i.e. the peak observed in the mean profile of at least one of the observation classes in **Y**), sized ±0.2% of that peak's m/z value. The window was then used to search for local maxima in each individual observation's mass spectrum (figure 14). The data was discarded if the extracted average normalized intensity was <1 in all observation classes, or if the peak resolution was <100 in all observation classes, as calculated by full width at half maximum on the average mass spectrum from each class.

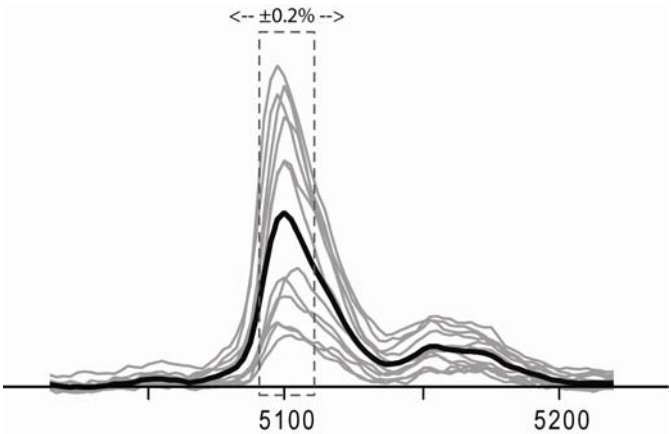

Figure 14. Schematic illustration of the window within which peak intensities from individual observations were extracted. The spectra represent brain tumor tissue samples, collected from 4 different animals within the same treatment group, that were analyzed in triplicates by SELDI-TOF-MS (paper **III**). The different colors distinguish between the individual spectra from each sample (gray) and the mean spectrum from the whole group (black). Here, a mean spectrum peak at *m/z* 5100 Da was found within a ROI. There was a slight mass shift between the individual spectra, and to compensate for this, the peak intensity values from each of the individual spectra were sought after within a window, sized ±0.2% of the mean peak's m/z (here ~±10 Da).

The above procedure was then repeated for all included data subsets, and the extracted peak intensity data from all subsets was assembled in a separate dataset ($\mathbf{X}_b$).

7.4.3 Analysis and validation of extracted features

$\mathbf{X}_b$, consisting of peak intensity data for all observations (including the ones that may have been excluded from the analysis of one or several subsets ($\mathbf{X}_a$)), was scaled to unit variance and initially screened for possibly deviating samples, by means of PCA and OPLS-DA using leave one out cross-validation. Following appropriate removal of outliers, $\mathbf{X}_b$ was further modeled by OPLS-DA, employing the same $\mathbf{Y}$ as previously, to unravel the most interesting peaks from all subsets.

This was realized by interpreting the scores and weights from the OPLS-DA model, and can be exemplified by the work presented in paper **III** for instance, where the responses ($\mathbf{Y}$) consisted of variables representing types of tissue (tumor/normal) and treatment (treated/untreated), as well as an interaction term between them. The aim of the study was to detect differences induced by treatment in either of the analyzed tissue types. The score plot from the OPLS-DA model of $\mathbf{X}_b$ clearly revealed that the first latent variable ($\mathbf{t}_{p1}$) explained the difference between normal and tumor tissue, whereas the effects of treatment were modeled by the second and third predictive components ($\mathbf{t}_{p2}$ and $\mathbf{t}_{p3}$) (figure 15). Thus, the correlation loadings from the latter two components

($\mathbf{p}_{(corr)2}$ and $\mathbf{p}_{(corr)3}$) were used to extract the most interesting peaks in $\mathbf{X}_b$, with regards to treatment effects.

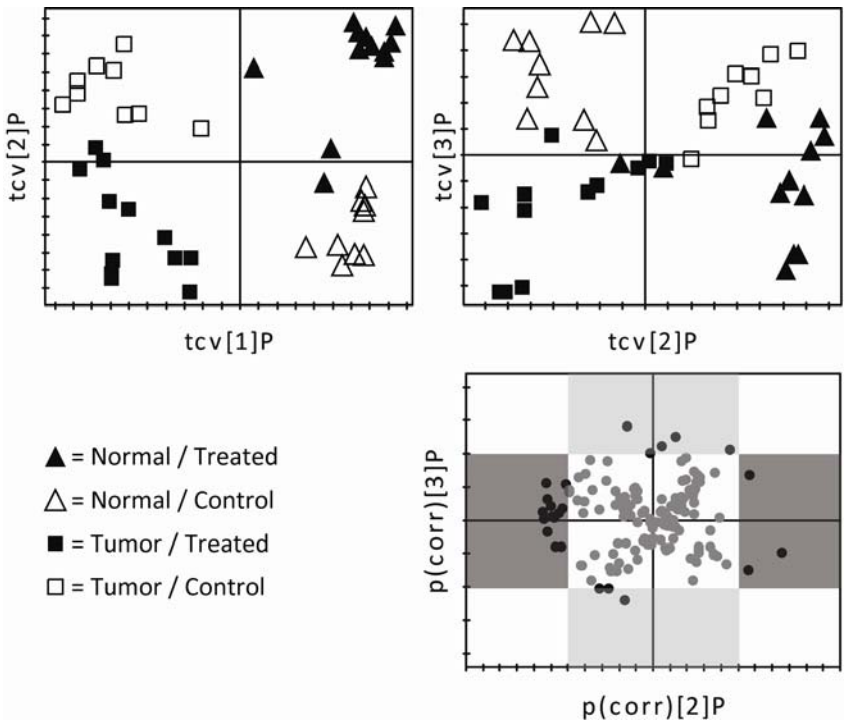

Figure 15. The upper two panels display cross-validated score plots from the OPLS-DA model of $\mathbf{X}_b$ (paper **III**). Here it was obvious that the effects of treatment were modeled by the second and third predictive component ($\mathbf{t}_{p2}$ and $\mathbf{t}_{p3}$), whereas the first predictive component modeled the difference between the tissue types. Consequently, the correlation loadings from $\mathbf{t}_{p2}$ and $\mathbf{t}_{p3}$ were used to locate the peaks strongest correlated to treatment in either of the two tissue types (lower panel). Variables in shaded areas were considered interesting, where those in the light shaded area displayed similar effects of treatment in both tissue types and those in the dark shaded area displayed opposite effects of treatment in the different tissue types.

Permutation analysis was also applied to $\mathbf{X}_b$. This approach can generally be described by the following steps: (i) calculation of the correlation between a descriptor variable ($\mathbf{X}_{b,k}$) and the response of interest ($\mathbf{y}$, i.e. treatment), (ii) random permutation of the response variable, (iii) calculation of the correlation between $\mathbf{X}_{b,k}$ and the randomly permuted $\mathbf{y}$, (iv) repetition of the last two steps a large number of times (typically 1,000 or 10,000 times), (v) comparison of the descriptor variable's correlation to the original response and that to each of the randomly permuted responses, and at last (vi) variables that display a stronger correlation to any of the permutated responses than to the original response are discarded. Taken together it is a very stringent approach to analyze correlations.

As the permutation analysis was applied in paper **III**, it acted as a complement to the multivariate modeling. Firstly, in cases where the same variables were picked up by both methods, the findings were obviously highly interesting and the permutation analysis could be viewed as a means to validate the findings made from the OPLS-DA analysis. Secondly, if the OPLS-DA model failed to divide the sought after effects into separate components, it would be possible to have missed important descriptor variables in $\mathbf{X}_b$ by looking exclusively at the model loadings. By looking at the permutation results in parallel, at least the variables that displayed a very strong correlation to the response was detected. Lastly, as mentioned, permutation is a very stringent validation criteria, as such it risks ruling out biologically interesting variables that do not display the strong correlation to the response needed to pass. By also interpreting the OPLS-DA loadings, variables that displayed weaker, yet interesting, correlations were picked up as well.

However, as was the case in paper **III**, when the dataset contains classes with large differences (i.e. tissue type) that are uncorrelated to the response (i.e. treatment), the different classes need to be analyzed separately by permutation.

7.5 Analysis of GC-MS data

In paper **II,** the samples were analyzed by GC-TOF-MS, which constitutes a hyphenation of chromatography and mass spectrometry. Thus, the data output from this method is two dimensional, where one dimension corresponds to the separation on the GC column and the other to MS fragmentation pattern of the GC column effluent. In other words, the data from a single sample can be stored in a simple table ($\mathbf{X}_{samp}$), where the number of rows (N) corresponds to the total number of MS scans performed during the analysis (for instance in paper **II**, the scan rate was set to 30 scans per second for about 10 minutes, i.e. $N \approx 18{,}000$) and the number of columns (K) corresponds to the number of m/z channels (the range in paper **II** was m/z 50 – 800, i.e. $K = 750$) (figure 16). The value in each cell of the table represents the recorded signal from the MS detector at a specific time point and m/z channel. Hence, plotting the values in one row yields a graphical representation of the mass spectrum obtained at a specific time point (see also figure 9b). Similarly, plotting the values of one column yields a graphical representation of the ion chromatogram from a specific mass channel. An overview of the sample is acquired by plotting the TIC chromatogram. The TIC is the sum of each row in the table, i.e. the total of all recorded responses for a specific time point (see also figure 9a).

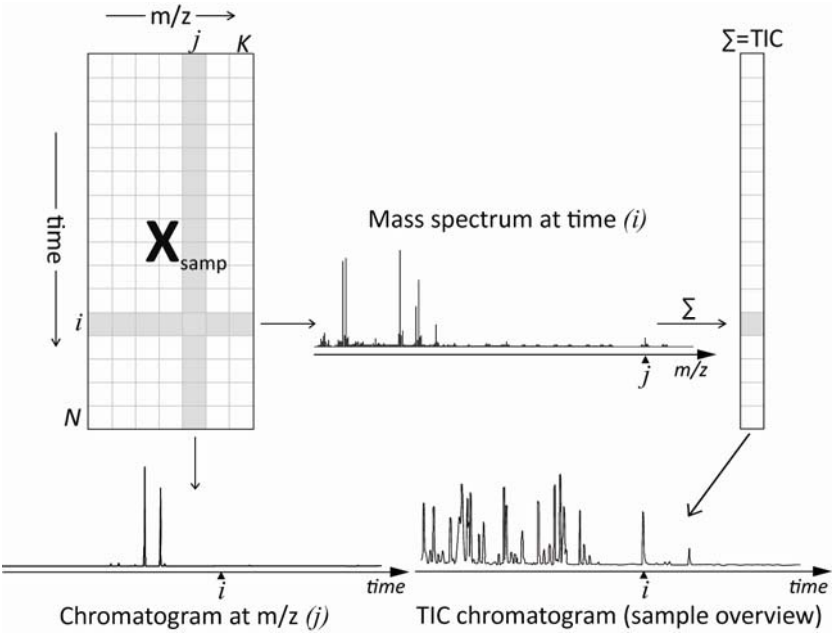

Figure 16. Schematic representation of the hyphenated data obtained from GC-MS analysis of one sample.

When complex samples are analyzed by GC, normally not all molecules can be separated on the column. Instead, similar molecules are likely to elute simultaneously from the column and enter the mass spectrometer at the same time. This makes it problematic to quantify and identify the molecules correctly. Therefore, prior to quantification, identification and further analysis, the collected data is first subjected to a process called deconvolution (or curve resolution) that aims to mathematically separate the signals from the molecules that did not separate on the column.

7.5.1 Deconvolution and data pre-treatment

Here, deconvolution (figure 17) was performed by means of hierarchical multivariate curve resolution (H-MCR) (Jonsson et al., 2005; Jonsson et al., 2006), which constitutes one of several methods to deal with the issues caused by co-eluting analyte molecules. H-MCR features the possibility to process multiple samples simultaneously, as well as the option to use the H-MCR solution predictively to rapidly process new samples. The H-MCR procedure ultimately produces a data table (**X**), where each sample is represented by a row and each resolved compound by a column. The value in each cell then corresponds to the area under the deconvoluted peak from a specific compound in a specific sample. Moreover, the resolved peaks are linked to pure mass spectral profiles (i.e. fragmentation patterns) that in turn can be used in efforts towards compound identification.

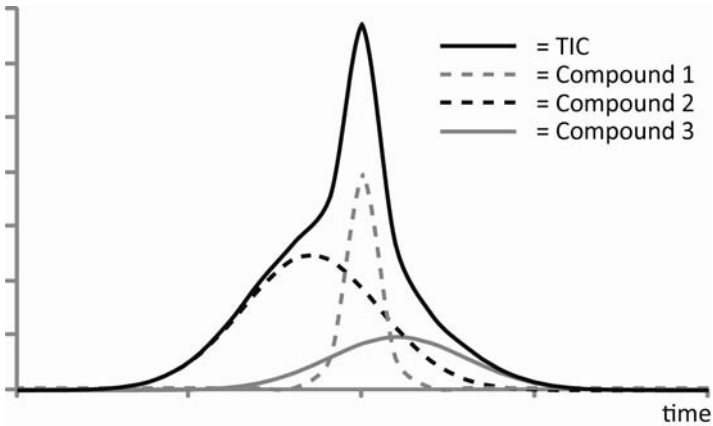

Figure 17. A schematic representation of curve resolution. The TIC peak, which represents the unprocessed data, is in fact the sum of three different compounds that did not resolve on the chromatographic column. By means of mathematical curve resolution, the different compound's pure profiles are calculated, and each compound can subsequently be quantified separately by calculating the area under the peak in its pure profile. Furthermore, each of the pure profiles is associated to a pure mass spectrum, which in turn can be used in an effort to identify the compound.

In paper **II**, the total number of samples (198) was too large to analyze all in one set (batch). The samples were therefore analyzed in two separate batches, where batch one included samples from the first five patients and batch two samples from the remaining six patients. The analysis order was randomized within each batch to avoid introducing systematic bias. H-MCR was applied to the obtained data from the first batch, and the calculated H-MCR parameters were subsequently used to predict the resolved profiles of the samples in the second batch. Prior to further analyses, all peak areas were normalized based on the data from 11 internal standards (added to each sample prior to GC-MS-analysis), eluting at various time points spanning the chromatographic time range. Combining the predictive feature of H-MCR with model predictions by means of multivariate projection methods is known as predictive metabolomics and has recently been shown to be useful in global metabolomics screening studies (Chorell et al., 2009; Jonsson et al., 2006; Pohjanen et al., 2007).

The aim of the study was to characterize differences between microdialysis samples collected within different tissue types (i.e. contrast enhancing tumor tissue, brain adjacent to tumor (BAT) tissue, and subcutaneous tissue from the abdomen) as well as potential differences in the metabolic profile induced by conventional radiotherapy treatment. To this end, samples from the various tissue types were collected both before and after treatment initiation and analyzed by GC-MS. The resolved data obtained from the samples collected before treatment onset was selected and used in a simple discriminant analyses directed towards distinguishing between the different

tissue types. However, prior to investigating possible metabolic differences induced by treatment, the data in **X** was subjected to further pre-treatment.

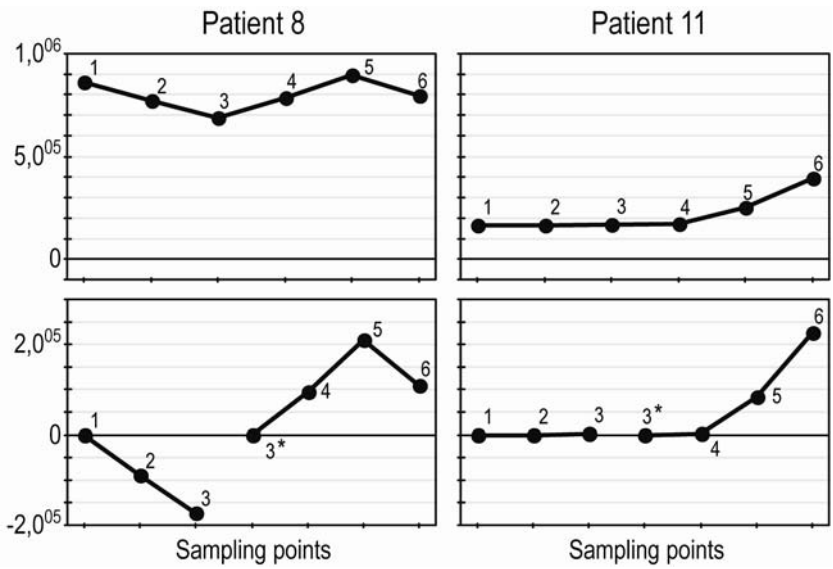

Figure 18. A graphical representation of individual treatment over time (ITOT) normalization, applied prior to investigation of potential effects induced by radiotherapy treatment. The upper two panels represent the relative abundance of the compound alanine (vertical axes), in tumor samples from two selected patients. The numbers by each observation denote the sequential sampling points, where samples 1-3 were collected prior to treatment initiation and sample 4-6 were collected thereafter. The asterisk (3*) is used to denote that the original data from time point 3 was used as the reference point in the group of treated samples (i.e. the value set to zero, by subtracting it from itself as well as all subsequently collected samples).

The additional data pre-treatment was done because the average discrepancy in abundance for any given compound could be quite large between patients, compared to the relative alteration induced by treatment within a patient. In addition, time related factors, such as the duration for which a specific catheter had been implanted, could potentially confound the effects of treatment. Therefore, to focus the analyses on systematic changes following treatment, irrespective of a compounds' initial level in different patients as well as irrelevant time related effects, the data obtained for all metabolites was set to zero for the first collected sample in both the treated and untreated group of samples (figure 18). In other words, the abundance level obtained from the first sample collected from a specific tissue type in a given patient was subtracted from all samples collected prior to treatment onset (including itself), from the same patient and tissue type. Likewise, the original data obtained from the last sample before treatment initiation was subtracted from all samples collected thereafter. We termed this data pre-treatment procedure individual treatment over time (ITOT)

normalization, and it was performed individually for each resolved metabolite (i.e. column in **X**). The result was stored in a separate data matrix ($\mathbf{X}_{\mathrm{ITOT}}$). $\mathbf{X}_{\mathrm{ITOT}}$ was in turn used for investigating effects of treatment in the included sample types.

8. Results and discussion

8.1 Sample characterization by means of non-targeted screening methods

In all papers presented herein the samples were studied by means of the non-targeted screening methods SELDI-MS and GC-MS to generate, respectively, protein and metabolite profiles representative of each sample. The profiles could in turn be used for further analysis and sample classification.

In papers **I** and **III** we showed that SELDI-MS derived protein profiles distinctly differed between tumor tissue samples and contralaterally collected normal tissue samples, in the BT4C rat glioma model, as made obvious by unsupervised PCA analysis of the complete protein profiles (figure 19a). This is not surprising considering the vast biological difference between the sample types, and comparable findings have been made by other groups studying GBM (Iwadate et al., 2004; Zheng et al., 2003). Similarly, in paper **II** we also detected a marked difference between the included sample types, this time based on the metabolic profiles in extracellular fluid from GBM patients. The most obvious distinction was observed between samples collected subcutaneously (SC) and intra-cranially, as illustrated here by a PCA score plot (figure 19b). Noticeably, here we could also quite clearly distinguish between the intra-cranially collected tumor and BAT samples (figure 2 in paper **II**).

Furthermore, both in papers **I** and **III**, we found that tumor progression in the BT4C model was associated to clear differences in protein profiles, as revealed by subjecting the full profiles to unsupervised PCA. The tumor tissue profiles from animals sacrificed shortly after implantation were more similar to those from normal brain tissue than the tumor profiles from animals that were sacrificed at a later stage (figure 2 in paper **III**). Moreover, a distinction could be made between the profiles of non-tumor tissue from tumor bearing rats and brain tissue from tumor-free rats. The BT4C glioma model has been shown to grow partly invasively, mimicking the clinical progression of the disease (Bergenheim et al., 1994), thus the latter finding might be a result of invading tumor cells. Alternatively, reflecting a systemic response to the disease.

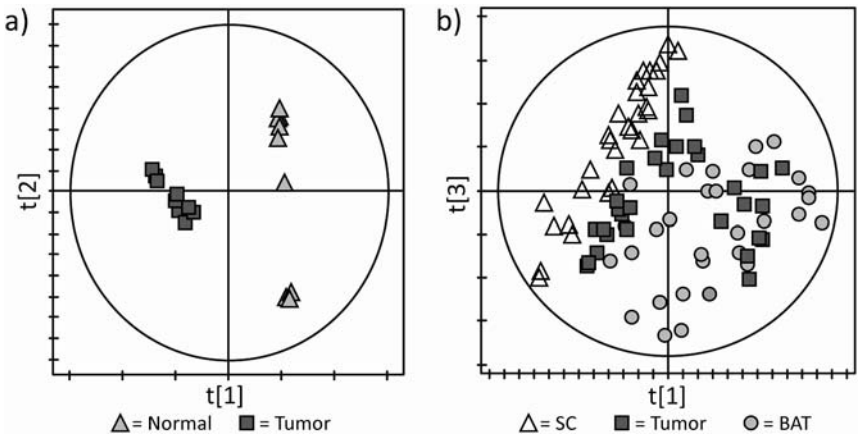

Figure 19. PCA score plots representing differences between samples from different types of tissue. **a**) Analysis of complete, SELDI-MS derived, protein profiles from GBM and contralaterally collected normal samples in the BT4C rat model. All samples are from untreated animals, sacrificed 24 days after tumor implantation, and included in the study presented in paper **I**. **b**) Analysis of metabolite profiles generated by GC-MS on microdialysate samples collected subcutaneously (SC) and intra-cranially, both in the contrast enhancing glial tumor and in the BAT region. All samples were collected prior to treatment and are included in the study presented in paper **II**.

A separate and important issue in all studies is avoidance of introducing bias to the data. This is especially true for label free MS-based proteomics studies where biological validation may be difficult to achieve (see below). It can be exemplified by the groundbreaking study presented by Petricoin and Liotta in 2002 (Petricoin et al., 2002), where they demonstrated how serum SELDI-MS patterns could almost perfectly discriminate between early stage ovarian cancer and healthy controls. This study was however shortly thereafter vigorously questioned, and it was suspected that the reported findings were in fact related to sample processing and therefore unrelated to biologically relevant discrepancies (Baggerly et al., 2004).

The above example also epitomizes the most obvious inherent drawback of SELDI-MS-based proteomics: the lack of protein identification. Because there is no easy way to achieve protein identification, there is no way to know whether the discriminating peaks are biologically relevant or not. In other words, the results cannot be biologically validated and the classifications may consequently be viewed as a kind of "black box" approach. The validity of simply relying on proteomic patterns, of unidentified proteins, to discriminate between sample classes has been the matter of a lively debate (Diamandis, 2003; Garber, 2004; Petricoin & Liotta, 2003), focused in particular on early cancer diagnosis based on blood sample analyses. On the plus side, the debate has brought attention to the importance of standardization of sample pre-treatment and analyses protocols (Mischak et al., 2007; Ransohoff, 2005a; Ransohoff, 2005b).

In our small and isolated studies presented here, rigorous care was taken to avoid introducing bias to our data, mainly through standardization and randomization of each step involving sample handling and analysis. Thus we have no reason to believe that the presented findings are reflective of anything but actual biological events. As such they demonstrate the feasibility of the approach.

While the SELDI-MS platform does not offer easily obtained biological validation, i.e. interpretation of the results in a biological context, at present there are no established methods that allow for protein profiling with the same sensitivity, intra-experimental analytical reproducibility and throughput capacity. Thus, it is a highly useful tool for biomarker discovery and for establishing potential proteomic differences between investigated samples classes. Thereafter there are methods to successfully identify the proteins of interest, and thereby achieve biological validation, although laborious (Orre et al., 2007).

GC-MS on the other hand offers opportunities for biological validation, as demonstrated in paper **II** where 67 metabolites (44%) were identified by using the obtained and deconvoluted mass spectra for each analyte in MS database searches. The limiting factor to identification is normally the completeness of the databases used for comparisons of mass spectra.

The concept of utilizing non-targeted screening methods and multivariate data analysis for classifications has been established for a long time. Almost 30 years ago Jellum et al reported that brain tumor tissue could be distinguished from normal brain tissue, by means of 105 GC-MS-derived metabolite peaks (Jellum et al., 1981). Today, the multi-parametric approach has gained widespread acceptance and is used in a multitude of cancer research fields, perhaps best exemplified by the FDA-approved multivariate prognostic test for breast cancer recurrence, MammaPrint (mentioned in section 5.4).

8.2 Analysis of proteomic and metabolomic patterns

8.2.1 Protein expression data generated by SELDI-MS

One of the initial steps in SELDI data analysis often involves peak identification and annotation. This implies that intensity data from peaks that by some criteria can be uniformly detected across a percentage of all samples is extracted, and the rest of the data is considered as noise and hence discarded. Peak detection may be achieved by a variety of methods, for instance by means of the algorithm implemented in the vendor's included software, ProteinChip® Software, which is commonly used (Qu et al., 2002; Rogers et al., 2003; Vanhoutte et al., 2007). There are however many more methods developed for this purpose (Cruz-Marcelo et al., 2008; Emanuele & Gurbaxani, 2009).

In paper **I** we showed that by subjecting complete MS profiles to unsupervised PCA it is possible to distinguish both between tumor and normal tissue as well as between radiotherapy treated and untreated tumor tissue, without initial peak selection. These observations could be made both in the low end (2.5-20 kDa) and in the high end (20-50 kDa) of the mass data, indicating that there was discriminating information across the full spectral range. The full MS profiles were further analyzed by PLS-DA, and the obtained classification model was successfully validated by prediction of external samples.

Furthermore, we demonstrated a novel method for peak selection, which was based on the initial analysis of complete MS profiles. It made use of the PLS weights (**w***) to define spectral regions of interest (ROI), within which peaks were defined as local maxima in the mean spectra profiles. This way, instead of including all detectable peaks for further examination, only peaks within regions holding probable discriminating information was further investigated, thereby decreasing the risk of spurious findings. This approach was refined for papers **III** and **IV** to be based on a combination of OPLS weights (**w***) and correlation loadings ($\mathbf{p}_{corr}$). This can be compared to a peak detection method referred to as annotated regions of significance (ARS) (Tan et al., 2006), which was developed for SELDI data and operates by initially defining spectral regions with significant intensity differences between spectra, irrespective of possible sample classes, within which peaks are detected. Thereafter, the selected peaks need to be subsequently assessed to avoid including noise and artifacts. In all papers presented here this step was done manually, whereas the ARS algorithm has an integrated automated peak quality control that discards low-quality peaks, as defined by a set of criteria including measurements of peak shape and location in relation to neighboring peaks.

The next step is peak annotation in the individual sample spectra and extraction of intensity data for each peak. This is typically complicated by minor misalignments of the m/z axis between the samples. Here, this problem was dealt with by fitting a window around the peak defined in the mean profile, sized ±0.2% of mean peak's m/z value. The window size was based on reports of SELDI mass accuracy (Grizzle et al., 2005).

In papers **III** and **IV**, we demonstrated that this approach may be used on individual SELDI data subsets, followed by merging the selected peak data to construct a joint dataset of all peaks of interest suitable for multivariate data analysis and classification of sample groups.

8.2.2 Metabolic data generated by GC-MS

Predictive metabolomics is a term for employing chemometric methods to all steps of a metabolomics study, and thereby assure generation of reliable data that allows for comparison of many samples simultaneously. This includes careful study design, data

deconvolution by predictive H-MCR (Jonsson et al., 2006) and data analysis by means of multivariate projection methods, which allows for prediction of external samples. Altogether, this approach makes screening of large sample sets possible and ensures interpretability of the results (Jonsson et al., 2006; Pohjanen et al., 2007). The predictive metabolomics approach was successfully employed in paper **II**. The involved issues of alignment, normalization and deconvolution were effectively handled by the established H-MCR method, which ultimately produces a data matrix containing concentration data for each detected metabolite in each sample, suitable for further multivariate data analysis and sample classification.

To allow for meaningful analysis of effects of treatment, we employed an additional data normalization procedure, referred to as ITOT (see section 7.5). This is a modified version of other normalization procedures that have been used previously to interpret treatment effects in metabolomic studies (Chorell et al., 2009). ITOT normalization constituted a key step in the successful investigation of the extracellular metabolic alterations induced by radiotherapy in GBM patients.

8.3 Treatment-induced changes in protein and metabolite composition

8.3.1 Radiotherapy treatment

Radiotherapy constitutes a cornerstone in GBM treatment, yet little is known about its molecular effects as well as the effects related to the often observed radioresistance. Moreover, there are no molecular markers for efficacy of radiotherapy treatment, which is typically assessed months after treatment initiation by radiological means. By that time it may be too late to significantly affect the tumor by second line treatment. Thus, identification of molecular markers of treatment efficacy could improve GBM management considerably. As an initial step in this direction papers **I** and **II** aimed at screening for molecular effects of radiotherapy.

In paper **I** tumor tissue samples from the experimental BT4C rat glioma model were investigated by means of SELDI-MS. The generated protein profiles were subsequently compared between animals subjected to a single dose of radiotherapy (12 Gy) and untreated animals, at different time points after treatment. Principal component analysis (PCA) of complete MS-profiles from animals sacrificed 12 days after treatment revealed distinct and systematic treatment induced proteomic alterations (figure 4 in paper **I**). To our knowledge there are no other reports where the proteome has been screened for alterations following radiation in glioma. Based on work performed in glioma cell lines it has however been reported that DNA-repair genes are up-regulated (Otomo et al., 2004), and the expression level of a few selected proteins have also been studied, including VEGF which has been found up-regulated by radiation (Steiner et al., 2004). Furthermore, there are screening studies of the proteome performed in relation to irradiation and other types of cancer, for instance colon and lung cancer (Orre et al.,

2007), indicating distinct alterations following radiotherapy. In addition, radiation-induced changes have been detected in unfractionated serum by means of SELDI-MS (Menard et al., 2006). This is in concordance with what might be expected, given the complex cellular machinery that is activated by radiation-induced DNA damage (Meek, 2004; Zhou & Elledge, 2000).

As mentioned earlier, by applying PCA to whole protein profiles of tumor tissue samples from the BT4C rat glioma model we found that the tumor progression was associated to distinct protein profiles. Interestingly, by comparing samples from treated and untreated animals in relation to the progression, it was evident that the untreated animals sacrificed at day 1 and 12 after irradiation had progressed further than their treated counterparts. Similar, but less evident, trends could be observed for animals sacrificed at day 5 and 7 after irradiation (figure 20). Altogether, this might imply that irradiation to some degree hampers tumor progression.

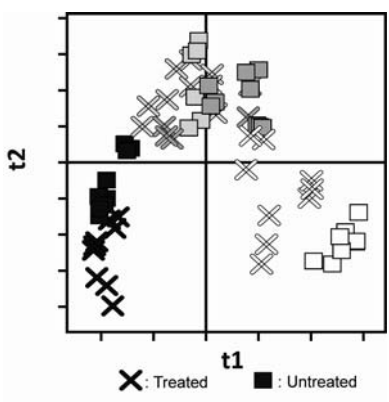

Figure 20. PCA score plot representing differences in SELDI-MS derived protein profiles between tumor tissue samples from animals sacrificed at different time points after irradiation. The time points are represented by various colors, as follows: black = day 1, light gray = day 5, dark gray = day 7, and white = day 12 after treatment. *The figure is adapted from paper **I** (no permission necessary).*

Furthermore, effects of radiotherapy on the extracellular metabolome were investigated in a series of 11 patients, diagnosed with GBM considered unsuitable for resection (paper **II**). OPLS-DA was applied to the deconvoluted and ITOT-normalized GC-MS data in order to detect the most interesting metabolites, i.e. metabolites that seemed affected by treatment in the intra-cranially collected microdialysis samples. From this, we detected 55 metabolites of interest in the samples collected within the contrast enhancing tumor and 72 in samples collected in the BAT region. Further analysis of the metabolite profiles of the selected compounds revealed clear and systematic differences between samples collected before and after treatment, both among tumor and BAT

samples. This was illustrated by the cross-validated score values from the OPLS-DA models based on the ITOT-normalized data of the selected metabolites (figure 21).

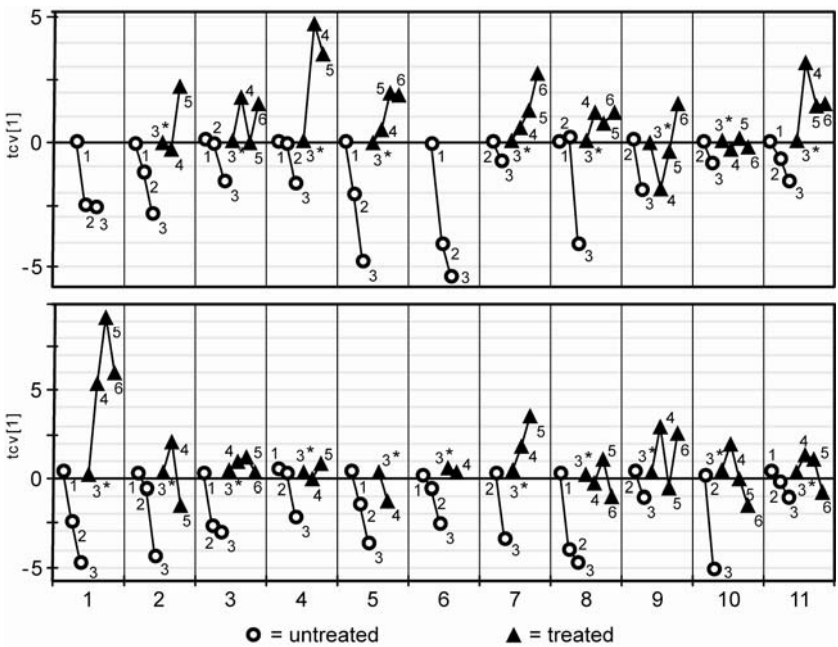

Figure 21. Cross-validated OPLS-DA score values (vertical axes) from analysis of the 55 and 72 most interesting metabolites affected by treatment in tumor (upper panel) and BAT (lower panel) samples respectively. The horizontal axes represent individual patient numbers. The numbers by each observation denote the sequential sampling points, where sample 1-3 were collected prior to treatment initiation and sample 4-6 were collected thereafter. The asterisk (3*) denotes that the original data from time point 3 was used as the reference point in the group of treated samples during ITOT normalization (see figure 18).

This study stretched over the first 5 fractions (á ~2 Gy) of conventional radiotherapy. Typically a complete radiotherapy course consists of about 30 fractions, received 5 days a week to a total of 60 Gy. The observed alterations therefore represented an early response to treatment, and as such the future implications are interesting. If it would be possible to correlate these early changes in metabolomic patterns to treatment outcome or radioresistance, it could have an impact on patient management. Considering that resistance to anti-cancer treatment is likely to be a multifactorial event (Smith et al., 2006), an approach such as the one used here might effectively lend itself to assessment thereof.

Although we screened the metabolome for alterations following treatment, the technique did not allow for a truly global analysis of all metabolites. To date, there is no technique that does. Thus, our observations may be thought of as representing traces

or fragments of the complex cellular processes that are activated upon irradiation induced cytotoxicity. The individual metabolites may be indicative of various parts of this response. Consider glutamate for instance, which we found up-regulated in tumor following treatment and that has also been reported up-regulated in ischemic and traumatic brain injury (Nordmark et al., 2009). Perhaps our result in this case was reflective of tumor cells damaged by radiation. In similar studies of various cancers, attempts have been made to put the individual metabolites into networks and biological pathways, such as performed by Denkert et al. in their study of colon carcinoma (Denkert et al., 2008). Although complicated by fragmented input data, pathway analysis or similar approaches hold great promise to facilitate the understanding of the biochemical and physiological processes under study.

Regarding the findings in both papers **I** and **II**, it should be mentioned that the differentially expressed proteins/metabolites do not necessarily represent markers of treatment efficacy, but rather effects thereof. In both cases there are no response variable to relate the effects to, thus no correlation to treatment efficacy can be made at this point. Nevertheless, we have shown that it is possible to detect early molecular effects of radiotherapy both in tissue and microdialysate samples, which in future studies may prove clinically important.

8.3.2 Experimental angiogenesis treatment

Vandetanib (ZD6474) is a small molecule tyrosine kinase inhibitor with demonstrated effects on both VEGFR and EGFR (Hennequin et al., 2002), and as such it may reduce both angiogenesis and tumor cell proliferation. These properties make it a highly interesting substance for treatment of GBM. In paper **III** we investigated the effects of vandetanib treatment on protein expression in the BD IX orthotopic rat glioma model, by means of non-targeted SELDI-MS proteome screening. Vandetanib has previously been shown to significantly affect tumor growth in the same model system when administered alone (Sandstrom et al., 2004) or in combination with chemotherapy or radiotherapy (Sandström et al., 2008). In the present study we found systematic alterations in protein expression patterns related to treatment both in tumor and contralateral brain tissue. This observation was apparent at all investigated treatment durations, i.e. following: 14 days, 6 days, and after three different time points shortly after a single dose of treatment.

The effects of a single dose were studied at 2, 8, and 24 hours post treatment, and there were clear treatment-induced differences in protein expression at all time points. However, there appeared to be a quite limited overlap among the individual peaks that most prominently contributed to the observed separations of treated and untreated tissues in the different comparisons. This was also the case when results from 6 and 14 days of treatment were compared, both between themselves and in relation to the single dose. All animals were implanted with tumors at the same time, but the animals included in the experiment group in which we investigated effects after 6 days of

treatment were sacrificed 7-8 days earlier than the rest, i.e. the tumors had not grown for as long in this group as in the rest. Given that the protein expression characteristics seems largely dependent on tumor progression, it is not surprising that there was a limited overlap among the individual proteins found to be differentially expressed in response to treatment between the experiment groups. Taken together, this points to a large degree of complexity in establishing proteomic markers that consistently reflect the cellular response to treatment.

Nevertheless, development of molecular markers that can be used to facilitate assessment of anti-angiogenesis treatment is of importance for continued progress in drug development (Jubb et al., 2006). Especially since the standard monitoring methods, such as radiological tumor size evaluation, is not an indicator of anti-angiogenesis treatment effects *per se.* As mentioned in section 5.4 there are a number of surrogate markers that have been suggested to this end, however, none that is clinically established. Given the complexity of the problem, it seems likely that no single marker alone will provide a solution, however little efforts have been made in this direction and our study (paper **III**) is the first that investigates the influence of anti-angiogenesis treatment on the protein expression pattern *in vivo* using a glioma model.

A potential limitation of our work was the lack of any response parameters, such as tumor volume or micro vascular density. However, these parameters have already been established in the same model system in response to vandetanib treatment, and little intra-individual variation was observed, as can be expected in an inbred animal model (Sandström et al., 2008). Thus our study was designed to be descriptive only. Furthermore, an inherent drawback of the study design is that it does not allow for distinction between tumor cellular response to treatment and effects of the normalization of the blood brain barrier. In other words, given that vandetanib is shown to close a leaky blood brain barrier and reduce vessel dilation, like other VEGFR tyrosine kinase inhibitors (Batchelor et al., 2007), the tissue samples from treated animals may contain less extravasated blood plasma proteins, which in turn might be reflected in our results. Obviously, such changes in the proteome will not translate into relevant markers for treatment response, as ultimately they need to be easily available, for instance through analyses of blood samples. A study investigating possible alterations in the blood serum proteome is underway.

Regarding EGFR tyrosine kinase inhibitor treatment, it was recently reported that a panel of MALDI-MS derived protein peaks may be used to predict the outcome of treatment within non-small cell lung cancer patients (Taguchi et al., 2007), demonstrating the usefulness of MS-based proteomics and a multi-parametric approach to this field of study.

8.4 Classification of benign meningioma growth behavior

In the study presented in paper **IV** we investigated protein profiles of invasive and non-invasive benign (WHO grade I) meningiomas, and the potential to use these for distinguishing between the two. At present, there are no established methods to prospectively achieve this distinction, although from a treatment decision point of view it would be highly desirable as it would facilitate early decisions regarding treatment. The tissue samples we examined were collected from patients suffering from fibrous or meningothelial benign meningioma, i.e. the two most common histological subgroups, and their growth behavior were characterized by surgical and radiological means. Although the protein profiles contained large variations between individuals, even within the same patient groups, our analysis approach (described in section 7.4) extracted peaks that systematically differed between invasive and non-invasive benign meningiomas. The numbers of peaks extracted were 22 and 6 among the fibrous and meningothelial meningiomas, respectively. Further analysis of the patterns that these panels of peaks represent revealed that they could be used to discriminate invasive tumors from non-invasive. For fibrous samples the area under the ROC-curve was calculated to 1.0, and the corresponding value calculated for meningothelial samples was 0.873.

One clinical problem associated with this disease is the inability to foresee which of the incompletely resected tumors that may develop an invasive growth characteristic. Knowledge thereabout could facilitate decisions regarding further treatment, such as the use of adjunct radiotherapy. To this end, prognostic markers are highly desirable (Commins et al., 2007). At present, only one molecular marker is commonly used to evaluate meningioma specimens in the clinic, the proliferation marker Ki 67, which typically can be used to aid in tumor grading and as such acts as a prognostic marker. However, Roser et al. have shown that Ki 67 levels do not correlate to recurrence among benign meningiomas, although the authors of this study did not address the issue of invasiveness (Roser et al., 2004). For this purpose a few molecular markers have been proposed, including the extracellular matrix-associated protein SPARC, which is involved in cell adhesion and migration (Rempel et al., 1999) as well as the protease cathepsin B (Strojnik et al., 2001).

Our study was of retrospective design, i.e. we investigated differences between tumors that were classified as non-invasive and those with manifested bone invasive growth, thus our findings do not necessarily hold prognostic information. However, while this was a small, retrospective study, our findings point to the viability of using proteomics screening studies to find markers for invasiveness in benign meningioma. The usefulness of applying this approach to classification studies of meningioma is further exemplified by a report by Okamoto et al., where they show that protein expression patterns may be used to discriminate between meningiomas of different WHO grades (Okamoto et al., 2006). Altogether, these findings warrant further investigations into the possibilities of finding discriminating information in the meningioma proteome.

9. Conclusions

We have demonstrated the usefulness of unbiased screening of the proteome and the metabolome to characterize different types of brain tumor tissue as well as how they were affected by treatment. The large amounts of data generated by these screening methods have been effectively analyzed by multivariate projection methods. The combined approach of high-throughput screening and multivariate statistical analysis proved useful in assessing early effects of treatment as well as for distinguishing between different types of brain tissues.

Unbiased screening analyses of the proteome by means of SELDI-MS revealed distinct differences in protein profiles associated to tumor progression, in the BT4C rat glioma model. Although expected, this observation highlights the importance of suitable control groups, thus it will have an impact on the design of experimental treatment studies in the future search for biomarker patterns. Furthermore, by comparing these protein patterns to those from animals subjected to a single dose of radiotherapy, it appeared as if the tumor progression was held back to some extent by the treatment. In addition, effects on protein expression patterns of an experimental angiogenesis inhibitor, vandetanib, were investigated in the same animal model and employing the same analysis approach. We found that treatment with vandetanib results in altered protein expression both in tumor and normal tissue.

The extracellular metabolome of contrast enhancing tumor and BAT was investigated in GBM patients. Stereotactic microdialysis allowed for longitudinal sampling, before and during the five first days of conventional radiotherapy treatment. The samples were screened with regards to metabolite content, by means of GC-MS. Subsequent multivariate data analysis disclosed noticeable baseline differences between samples collected within the tumor and those collected in the BAT region. Even though GBM is a highly heterogeneous disease, this approach allowed us to discover differences that were systematic across the whole patient cohort and fundamentally related to the molecular pathophysiology of the disease. Following radiotherapy treatment, we could detect an alteration in the metabolite composition, both in tumor and BAT samples. The patterns of affected metabolites were different between the compartments, though in both instances they contained highly discriminating information. Taken together, we demonstrated that stereotactic microdialysis sampling combined with a predictive metabolomics analysis approach could detect early molecular effects of radiotherapy. This opens for the possibility of finding clinically valuable predictive biomarkers or marker patterns in relation to GBM treatment.

Tumor tissue samples from patients with benign meningioma were analyzed for protein expression using SELDI-MS and subsequent multivariate data analysis. The samples were either of meningothelial or fibrous histology, and collected from patients with invasive or non-invasive disease. Although the samples protein expression was highly heterogeneous, our analyses strategy detected a few MS-peaks whose combined expression pattern could be used to largely distinguish between samples based on invasiveness. This study was designed to be a retrospective study and as such our findings does not necessarily hold prognostic information. If these results can be repeated in a prospective study, it may have an impact on clinical decision making.

Large scale screening studies of the proteome and the metabolome have the potential to detect molecular effects of treatment at an early stage. If these can be correlated to treatment outcome, they could facilitate improvement of conventional treatment strategies as well as the process of developing novel therapies.

10. References

ACS (American Cancer Society) (2008) **Cancer Facts & Figures 2008**. (Accessed July 13, 2009, at: http://www.cancer.org/downloads/STT/2008CAFFfinalsecured.pdf)

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2002) **Blood vessels and endothelial cells**. In *Molecular Biology of the Cell* pp. 1279-1283. Garland Science: New York.

Alexander A, Murtha A, Abdulkarim B, Mehta V, Wheatley M, Murray B, Riauka T, Hanson J, Fulton D, McEwan A, Roa W (2006) **Prognostic significance of serial magnetic resonance spectroscopies over the course of radiation therapy for patients with malignant glioma**. *Clinical and Investigative Medicine. Medecine Clinique et Experimentale*, 29, 301-11.

Asahara T, Murohara T, Sullivan A, Silver M, van der Zee R, Li T, Witzenbichler B, Schatteman G, Isner JM (1997) **Isolation of putative progenitor endothelial cells for angiogenesis**. *Science*, 275, 964-7.

Baggerly KA, Morris JS, Coombes KR (2004) **Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments**. *Bioinformatics*, 20, 777-85.

Barbaro NM, Gutin PH, Wilson CB, Sheline GE, Boldrey EB, Wara WM (1987) **Radiation therapy in the treatment of partially resected meningiomas**. *Neurosurgery*, 20, 525-8.

Batchelor TT, Sorensen AG, di Tomaso E, Zhang WT, Duda DG, Cohen KS, Kozak KR, Cahill DP, Chen PJ, Zhu M, Ancukiewicz M, Mrugala MM, Plotkin S, Drappatz J, Louis DN, Ivy P, Scadden DT, Benner T, Loeffler JS, Wen PY, Jain RK (2007) **AZD2171, a pan-VEGF receptor tyrosine kinase inhibitor, normalizes tumor vasculature and alleviates edema in glioblastoma patients**. *Cancer Cell*, 11, 83-95.

Beaudry P, Force J, Naumov GN, Wang A, Baker CH, Ryan A, Soker S, Johnson BE, Folkman J, Heymach JV (2005) **Differential effects of vascular endothelial growth factor receptor-2 inhibitor ZD6474 on circulating endothelial progenitors and mature circulating endothelial cells: implications for use as a surrogate marker of antiangiogenic activity**. *Clinical Cancer Research*, 11, 3514-22.

Bergenheim AT, Elfverson J, Gunnarsson PO, Edman K, Hartman M, Henriksson R (1994) **Cytotoxic Effect And Uptake Of Estramustine In A Rat Glioma Model**. *International Journal of Oncology*, 5, 293-299.

Bergenheim AT, Malmström A, Bolander H, Michanek A, Stragliotto G, Damber L, Bjor O, Henriksson R (2007) **[Registration on regional basis of patients with primary brain tumors. Regional differences disclosed]**. *Läkartidningen*, 104, 332-341.

Bikmaz K, Mrak R, Al-Mefty O (2007) **Management of bone-invasive, hyperostotic sphenoid wing meningiomas**. *Journal of Neurosurgery*, 107, 905-12.

Bjerkvig R, Tysnes BB, Aboody KS, Najbauer J, Terzis AJ (2005) **Opinion: the origin of the cancer stem cell: current controversies and new insights**. *Nature Reviews. Cancer*, 5, 899-904.

Black PM (1997) **Meningiomas**. In *Handbook of clinical neurology*, Vecht CJ (ed), Vol. 68. pp. 401-420. Elsevier Science B.V.: Amsterdam.

Bylesjö M, Rantalainen M, Cloarec O, Nicholson JK, Holmes E, Trygg J (2006) **OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification**. *Journal of Chemometrics*, 20, 341-351.

Campbell BA, Jhamb A, Maguire JA, Toyota B, Ma R (2009) **Meningiomas in 2009: controversies and future challenges**. *American Journal of Clinical Oncology*, 32, 73-85.

Carmeliet P, Ferreira V, Breier G, Pollefeyt S, Kieckens L, Gertsenstein M, Fahrig M, Vandenhoeck A, Harpal K, Eberhardt C, Declercq C, Pawling J, Moons L, Collen D, Risau W, Nagy A (1996) **Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele**. *Nature*, 380, 435-9.

Carmeliet P, Jain RK (2000) **Angiogenesis in cancer and other diseases**. *Nature*, 407, 249-57.

CBTRUS (Central Brain Tumor Registry of the United States) (2009) **CBTRUS 2009 Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2004-2005**. (Accessed July 13, 2009, at: http://www.cbtrus.org/reports/2009-NPCR-04-05/CBTRUS-NPCR2004-2005-Report-.pdf)

Chen YR, Juan HF, Huang HC, Huang HH, Lee YJ, Liao MY, Tseng CW, Lin LL, Chen JY, Wang MJ, Chen JH, Chen YJ (2006) **Quantitative proteomic and genomic profiling reveals metastasis-related protein expression patterns in gastric cancer cells**. *Journal of Proteome Research*, 5, 2727-42.

Cheng SY, Huang HJ, Nagane M, Ji XD, Wang D, Shih CC, Arap W, Huang CM, Cavenee WK (1996) **Suppression of glioblastoma angiogenicity and tumorigenicity by inhibition of endogenous expression of vascular endothelial growth factor**. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 8502-7.

Chorell E, Moritz T, Branth S, Antti H, Svensson MB (2009) **Predictive metabolomics evaluation of nutrition-modulated metabolic stress responses in human blood serum during the early recovery phase of strenuous physical exercise**. *Journal of Proteome Research*, 8, 2966-77.

Clarke R, Ressom HW, Wang A, Xuan J, Liu MC, Gehan EA, Wang Y (2008) **The properties of high-dimensional data spaces: implications for exploring gene and protein expression data**. *Nature Reviews. Cancer*, 8, 37-49.

Claus EB, Bondy ML, Schildkraut JM, Wiemels JL, Wrensch M, Black PM (2005) **Epidemiology of intracranial meningioma**. *Neurosurgery*, 57, 1088-95; discussion 1088-95.

Collins FS, Guyer M, Peterson J, Felsenfeld A, Wetterstrand K, Kamholz S (2004) **Finishing the euchromatic sequence of the human genome**. *Nature*, 431, 931-45.

Commins DL, Atkinson RD, Burnett ME (2007) **Review of meningioma histopathology**. *Neurosurgical Focus*, 23, E3.

Condra KS, Buatti JM, Mendenhall WM, Friedman WA, Marcus RB, Jr., Rhoton AL (1997) **Benign meningiomas: primary treatment selection affects survival**. *International Journal of Radiation Oncology, Biology, Physics*, 39, 427-36.

Crick F (1970) **Central dogma of molecular biology**. *Nature*, 227, 561-3.

Cruz-Marcelo A, Guerra R, Vannucci M, Li Y, Lau CC, Man TK (2008) **Comparison of algorithms for pre-processing of SELDI-TOF mass spectrometry data**. *Bioinformatics*, 24, 2129-36.

Denkert C, Budczies J, Weichert W, Wohlgemuth G, Scholz M, Kind T, Niesporek S, Noske A, Buckendahl A, Dietel M, Fiehn O (2008) **Metabolite profiling of human colon carcinoma--deregulation of TCA cycle and amino acid turnover**. *Molecular cancer*, 7, 72.

Dettmer K, Aronov PA, Hammock BD (2007) **Mass spectrometry-based metabolomics**. *Mass Spectrometry Reviews*, 26, 51-78.

Diamandis EP (2003) **Point: Proteomic patterns in biological fluids: do they represent the future of cancer diagnostics?** *Clinical Chemistry*, 49, 1272-5.

Dirks PB (2008) **Brain tumor stem cells: bringing order to the chaos of brain cancer**. *Journal of Clinical Oncology*, 26, 2916-24.

Dunn WB, Ellis DI (2005) **Metabolomics: Current analytical platforms and methodologies**. *Trends in Analytical Chemistry: TRAC*, 24, 285-294.

El-Aneed A, Cohen A, Banoub J (2009) **Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers**. *Applied Spectroscopy Reviews*, 44, 210-230.

Emanuele VA, 2nd, Gurbaxani BM (2009) **Benchmarking currently available SELDI-TOF MS preprocessing techniques**. *Proteomics*, 9, 1754-62.

Escudier B, Eisen T, Stadler WM, Szczylik C, Oudard S, Siebels M, Negrier S, Chevreau C, Solska E, Desai AA, Rolland F, Demkow T, Hutson TE, Gore M, Freeman S, Schwartz B, Shan M, Simantov R, Bukowski RM (2007) **Sorafenib in advanced clear-cell renal-cell carcinoma**. *The New England Journal of Medicine*, 356, 125-34.

Ferrara N, Gerber HP, LeCouter J (2003) **The biology of VEGF and its receptors**. *Nature Medicine*, 9, 669-76.

Fiehn O (2002) **Metabolomics--the link between genotypes and phenotypes**. *Plant Molecular Biology*, 48, 155-71.

Filippini G, Falcone C, Boiardi A, Broggi G, Bruzzone MG, Caldiroli D, Farina R, Farinotti M, Fariselli L, Finocchiaro G, Giombini S, Pollo B, Savoiardo M, Solero CL, Valsecchi MG (2008) **Prognostic factors for survival in 676 consecutive patients with newly diagnosed primary glioblastoma**. *Neuro-oncology*, 10, 79-87.

Furnari FB, Fenton T, Bachoo RM, Mukasa A, Stommel JM, Stegh A, Hahn WC, Ligon KL, Louis DN, Brennan C, Chin L, DePinho RA, Cavenee WK (2007) **Malignant astrocytic glioma: genetics, biology, and paths to treatment**. *Genes and Development*, 21, 2683-710.

Galli R, Binda E, Orfanelli U, Cipelletti B, Gritti A, De Vitis S, Fiocco R, Foroni C, Dimeco F, Vescovi A (2004) **Isolation and characterization of**

tumorigenic, stem-like neural precursors from human glioblastoma. *Cancer Research*, 64, 7011-21.

Garber K (2004) **Debate rages over proteomic patterns**. *Journal of the National Cancer Institute*, 96, 816-8.

Glas AM, Floore A, Delahaye LJ, Witteveen AT, Pover RC, Bakx N, Lahti-Domenici JS, Bruinsma TJ, Warmoes MO, Bernards R, Wessels LF, Van't Veer LJ (2006) **Converting a breast cancer microarray signature into a high-throughput diagnostic test**. *BMC Genomics*, 7, 278.

Goldman CK, Bharara S, Palmer CA, Vitek J, Tsai JC, Weiss HL, Gillespie GY (1997) **Brain edema in meningiomas is associated with increased vascular endothelial growth factor expression**. *Neurosurgery*, 40, 1269-77.

Grau SJ, Trillsch F, Herms J, Thon N, Nelson PJ, Tonn JC, Goldbrunner R (2007) **Expression of VEGFR3 in glioma endothelium correlates with tumor grade**. *Journal of Neuro-Oncology*, 82, 141-50.

Grizzle WE, Semmes OJ, Bigbee W, Zhu L, Malik G, Oelschlager DK, Manne B, Manne U (2005) **The Need for Review and Understanding of SELDI/MALDI Mass Spectroscopy Data Prior to Analysis**. *Cancer informatics*, 1, 86-97.

Halket JM, Waterman D, Przyborowska AM, Patel RK, Fraser PD, Bramley PM (2005) **Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS**. *Journal of Experimental Botany*, 56, 219-43.

Hamstra DA, Chenevert TL, Moffat BA, Johnson TD, Meyer CR, Mukherji SK, Quint DJ, Gebarski SS, Fan X, Tsien CI, Lawrence TS, Junck L, Rehemtulla A, Ross BD (2005) **Evaluation of the functional diffusion map as an early biomarker of time-to-progression and overall survival in high-grade glioma**. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 16759-64.

Hart MG, Grant R, Garside R, Rogers G, Somerville M, Stein K (2008) **Temozolomide for high grade glioma**. *Cochrane Database Syst Rev*, CD007415.

Hegi ME, Diserens AC, Gorlia T, Hamou MF, de Tribolet N, Weller M, Kros JM, Hainfellner JA, Mason W, Mariani L, Bromberg JE, Hau P, Mirimanoff RO, Cairncross JG, Janzer RC, Stupp R (2005) **MGMT gene silencing and benefit from temozolomide in glioblastoma**. *New England Journal of Medicine*, 352, 997-1003.

Hegi ME, Liu L, Herman JG, Stupp R, Wick W, Weller M, Mehta MP, Gilbert MR (2008) **Correlation of O6-methylguanine methyltransferase (MGMT) promoter methylation with clinical outcomes in glioblastoma and clinical strategies to modulate MGMT activity**. *Journal of Clinical Oncology*, 26, 4189-99.

Hennequin LF, Stokes ES, Thomas AP, Johnstone C, Ple PA, Ogilvie DJ, Dukes M, Wedge SR, Kendrew J, Curwen JO (2002) **Novel 4-anilinoquinazolines with C-7 basic side chains: design and structure activity relationship of a series of potent, orally active, VEGF receptor tyrosine kinase inhibitors**. *Journal of Medicinal Chemistry*, 45, 1300-12.

Henriksson R, Malmstrom A, Bergstrom P, Bergh G, Trojanowski T, Andreasson L, Blomquist E, Jonsborg S, Edekling T, Salander P, Brannstrom T, Bergenheim AT (2006) **High-grade astrocytoma treated concomitantly with estramustine and radiotherapy**. *Journal of Neuro-Oncology*, 78, 321-6.

Hillman J, Aneman O, Anderson C, Sjogren F, Saberg C, Mellergard P (2005) **A microdialysis technique for routine measurement of macromolecules in the injured human brain**. *Neurosurgery*, 56, 1264-8; discussion 1268-70.

Hlatky L, Hahnfeldt P, Folkman J (2002) **Clinical application of antiangiogenic therapy: microvessel density, what it does and doesn't tell us**. *Journal of the National Cancer Institute*, 94, 883-93.

Hoelzinger DB, Demuth T, Berens ME (2007) **Autocrine factors that sustain glioma invasion and paracrine biology in the brain microenvironment**. *Journal of the National Cancer Institute*, 99, 1583-93.

Hortin GL (2006) **The MALDI-TOF mass spectrometric view of the plasma proteome and peptidome**. *Clinical Chemistry*, 52, 1223-37.

Hoskuldsson A (1995) **A Combined Theory for PCA and PLS**. *Journal of Chemometrics*, 9, 91-123.

Hsu DW, Efird JT, Hedley-Whyte ET (1997) **Progesterone and estrogen receptors in meningiomas: prognostic considerations**. *Journal of Neurosurgery*, 86, 113-20.

Hurwitz H, Fehrenbacher L, Novotny W, Cartwright T, Hainsworth J, Heim W, Berlin J, Baron A, Griffing S, Holmgren E, Ferrara N, Fyfe G, Rogers B, Ross R, Kabbinavar F (2004) **Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer**. *The New England Journal of Medicine*, 350, 2335-42.

Hutchens TW, Yip TT (1993) **New Desorption Strategies For The Mass-Spectrometric Analysis Of Macromolecules**. *Rapid Communications in Mass Spectrometry*, 7, 576-580.

Ichimura K, Ohgaki H, Kleihues P, Collins VP (2004) **Molecular pathogenesis of astrocytic tumours**. *Journal of Neuro-Oncology*, 70, 137-60.

Issaq HJ, Veenstra TD, Conrads TP, Felschow D (2002) **The SELDI-TOF MS approach to proteomics: protein profiling and biomarker identification**. *Biochemical and Biophysical Research Communications*, 292, 587-92.

Iwadate Y, Sakaida T, Hiwasa T, Nagai Y, Ishikura H, Takiguchi M, Yamaura A (2004) **Molecular classification and survival prediction in human gliomas based on proteome analysis**. *Cancer Research*, 64, 2496-501.

Jellum E, Bjornson I, Nesbakken R, Johansson E, Wold S (1981) **Classification of human cancer cells by means of capillary gas chromatography and pattern recognition analysis**. *Journal of Chromatography*, 217, 231-7.

Jenny B, Harrison JA, Baetens D, Tille JC, Burkhardt K, Mottaz H, Kiss JZ, Dietrich PY, De Tribolet N, Pizzolato GP, Pepper MS (2006) **Expression and localization of VEGF-C and VEGFR-3 in glioblastomas and haemangioblastomas**. *Journal of Pathology*, 209, 34-43.

Johansson D, Lindgren P, Berglund A (2003) **A multivariate approach applied to microarray data for identification of genes with cell cycle-coupled transcription**. *Bioinformatics*, 19, 467-73.

Johansson M, Bergenheim AT, D'Argy R, Edman K, Gunnarsson PO, Widmark A, Henriksson R (1998) **Distribution of estramustine in the BT4C rat glioma model**. *Cancer Chemotherapy and Pharmacology*, 41, 317-25.

Jolliffe IT (2002) ***Principal component analysis***. Springer Series in Statistics. Springer-Verlag: New York. 2nd edition.

Jonsson P, Johansson AI, Gullberg J, Trygg J, A J, Grung B, Marklund S, Sjostrom M, Antti H, Moritz T (2005) **High-throughput data analysis for detecting and identifying differences between samples in GC/MS-based metabolomic analyses**. *Analytical Chemistry*, 77, 5635-42.

Jonsson P, Johansson ES, Wuolikainen A, Lindberg J, Schuppe-Koistinen I, Kusano M, Sjöstrom M, Trygg J, Moritz T, Antti H (2006) **Predictive Metabolite Profiling Applying Hierarchical Multivariate Curve Resolution to GC/MS Data: A Potential Tool for Multi-parametric Diagnosis**. *Journal of Proteome Research*, 5, 1407-1414.

Jubb AM, Oates AJ, Holden S, Koeppen H (2006) **Predicting benefit from anti-angiogenic agents in malignancy**. *Nature Reviews. Cancer*, 6, 626-35.

Karas M, Bachmann D, Bahr U, Hillenkamp F (1987) **Matrix-assisted ultraviolet laser desorption of non-volatile compounds**. *International Journal of Mass Spectrometry and Ion Processes*, 78, 53-68.

Kirsch M, Strasser J, Allende R, Bello L, Zhang J, Black PM (1998) **Angiostatin suppresses malignant glioma growth in vivo**. *Cancer Research*, 58, 4654-9.

Kleihues P, Louis DN, Scheithauer BW, Rorke LB, Reifenberger G, Burger PC, Cavenee WK (2002) **The WHO classification of tumors of the nervous system**. *Journal of Neuropathology and Experimental Neurology*, 61, 215-25; discussion 226-9.

Kleihues P, Louis DN, Wiestler OD, Burger PC, Scheithauer BW (2007) **WHO grading of tumours of the central nervous system**. In *WHO Classification of Tumours of the Central Nervous System*, Louis DN, Ohgaki H, Wiestler OD,Cavenee WK (eds) pp. 10-11. International Agency for Research on Cancer: Lyon.

Kleihues P, Ohgaki H (1997) **Genetics of glioma progression and the definition of primary and secondary glioblastoma**. *Brain Pathology*, 7, 1131-1136.

Kondziolka D, Lunsford LD, Flickinger JC (2008) **The application of stereotactic radiosurgery to disorders of the brain**. *Neurosurgery*, 62, Suppl. 2, 707-19; discussion 719-20.

Krex D, Klink B, Hartmann C, von Deimling A, Pietsch T, Simon M, Sabel M, Steinbach JP, Heese O, Reifenberger G, Weller M, Schackert G (2007) **Long-term survival with glioblastoma multiforme**. *Brain*, 130, 2596-606.

Kristiansen K, Hagen S, Kollevold T, Torvik A, Holme I, Nesbakken R, Hatlevoll R, Lindgren M, Brun A, Lindgren S, Notter G, Andersen AP, Elgen K (1981) **Combined modality therapy of operated astrocytomas grade III and IV. Confirmation of the value of postoperative irradiation and lack of potentiation of bleomycin on survival time: a prospective multicenter trial of the Scandinavian Glioblastoma Study Group**. *Cancer*, 47, 649-52.

Kurz H, Burri PH, Djonov VG (2003) **Angiogenesis and Vascular Remodeling by Intussusception: From Form to Function**. *News in Physiological Sciences*, 18, 65-70.

Kyritsis AP (1996) **Chemotherapy for meningiomas**. *Journal of Neuro-Oncology*, 29, 269-272.

Lai A, Filka E, McGibbon B, Nghiemphu PL, Graham C, Yong WH, Mischel P, Liau LM, Bergsneider M, Pope W, Selch M, Cloughesy T (2008) **Phase II pilot study of bevacizumab in combination with temozolomide and regional radiation therapy for up-front treatment of patients with newly diagnosed glioblastoma multiforme: interim analysis of safety and tolerability**. *International Journal of Radiation Oncology, Biology, Physics*, 71, 1372-80.

Lamberts SW, Tanghe HL, Avezaat CJ, Braakman R, Wijngaarde R, Koper JW, de Jong H (1992) **Mifepristone (RU 486) treatment of meningiomas**. *Journal of Neurology, Neurosurgery and Psychiatry*, 55, 486-90.

Lamszus K (2004) **Meningioma pathology, genetics, and biology**. *Journal of Neuropathology and Experimental Neurology*, 63, 275-86.

Lamszus K, Lengler U, Schmidt NO, Stavrou D, Ergun S, Westphal M (2000) **Vascular endothelial growth factor, hepatocyte growth factor/scatter factor, basic fibroblast growth factor, and placenta growth factor in human meningiomas and their relation to angiogenesis and malignancy**. *Neurosurgery*, 46, 938-47; discussion 947-8.

Laws ER, Parney IF, Huang W, Anderson F, Morris AM, Asher A, Lillehei KO, Bernstein M, Brem H, Sloan A, Berger MS, Chang S (2003) **Survival following surgery and prognostic factors for recently diagnosed malignant glioma: data from the Glioma Outcomes Project**. *Journal of Neurosurgery*, 99, 467-73.

Leenders WP, Kusters B, de Waal RM (2002) **Vessel co-option: how tumors obtain blood supply in the absence of sprouting angiogenesis**. *Endothelium*, 9, 83-7.

Leon SP, Folkerth RD, Black PM (1996) **Microvessel density is a prognostic indicator for patients with astroglial brain tumors**. *Cancer*, 77, 362-72.

Louis D, Ohgaki H, Wiestler O, Cavenee W, Burger P, Jouvet A, Scheithauer B, Kleihues P (2007) **The 2007 WHO Classification of Tumours of the Central Nervous System**. *Acta Neuropathologica*, 114, 97-109.

Louis DN, Budka H, Von Deimling A (1997) **Meningeal tumours**. In *Tumours of the Nervous System*, Kleihues P,Cavenee WK (eds) pp. 133-151. International Agency for Research on Cancer: Lyon.

Malmer B, Gronberg H, Bergenheim AT, Lenner P, Henriksson R (1999) **Familial aggregation of astrocytoma in northern Sweden: an epidemiological cohort study**. *International Journal of Cancer*, 81, 366-70.

Meek DW (2004) **The p53 response to DNA damage**. *DNA Repair (Amst)*, 3, 1049-56.

Mellinghoff IK, Wang MY, Vivanco I, Haas-Kogan DA, Zhu S, Dia EQ, Lu KV, Yoshimoto K, Huang JH, Chute DJ, Riggs BL, Horvath S, Liau LM, Cavenee WK, Rao PN, Beroukhim R, Peck TC, Lee JC, Sellers WR, Stokoe D, Prados

M, Cloughesy TF, Sawyers CL, Mischel PS (2005) **Molecular determinants of the response of glioblastomas to EGFR kinase inhibitors**. *New England Journal of Medicine*, 353, 2012-24.

Menard C, Johann D, Lowenthal M, Muanza T, Sproull M, Ross S, Gulley J, Petricoin E, Coleman CN, Whiteley G, Liotta L, Camphausen K (2006) **Discovering clinical biomarkers of ionizing radiation exposure with serum proteomic analysis**. *Cancer Research*, 66, 1844-50.

Mendenhall WM, Friedman WA, Amdur RJ, Foote KD (2004) **Management of benign skull base meningiomas: a review**. *Skull base: official journal of North American Skull Base Society ... [et al.]*, 14, 53-60; discussion 61.

Mendenhall WM, Morris CG, Amdur RJ, Foote KD, Friedman WA (2003) **Radiotherapy alone or after subtotal resection for benign skull base meningiomas**. *Cancer*, 98, 1473-82.

Merchant M, Weinberger SR (2000) **Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry**. *Electrophoresis*, 21, 1164-77.

Meyerson BA, Linderoth B, Karlsson H, Ungerstedt U (1990) **Microdialysis in the human brain: Extracellular measurements in the thalamus of parkinsonian patients**. *Life Sciences*, 46, 301-308.

Miletic H, Niclou SP, Johansson M, Bjerkvig R (2009) **Anti-VEGF therapies for malignant glioma: treatment effects and escape mechanisms**. *Expert Opinion on Therapeutic Targets*, 13, 455-68.

Milker-Zabel S, Zabel-du Bois A, Huber P, Schlegel W, Debus J (2006) **Fractionated stereotactic radiation therapy in the management of benign cavernous sinus meningiomas: long-term experience and review of the literature**. *Strahlentherapie und Onkologie*, 182, 635-40.

Miller K, Wang M, Gralow J, Dickler M, Cobleigh M, Perez EA, Shenkier T, Cella D, Davidson NE (2007) **Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer**. *The New England Journal of Medicine*, 357, 2666-76.

Mineo JF, Bordron A, Baroncini M, Ramirez C, Maurage CA, Blond S, Dam-Hieu P (2007) **Prognosis factors of survival time in patients with glioblastoma multiforme: a multivariate analysis of 340 patients**. *Acta neurochirurgica*, 149, 245-52; discussion 252-3.

Miralbell R, Linggood RM, de la Monte S, Convery K, Munzenrider JE, Mirimanoff RO (1992) **The role of radiotherapy in the treatment of subtotally resected benign meningiomas**. *Journal of Neuro-Oncology*, 13, 157-64.

Mischak H, Apweiler R, Banks RE, Conaway M, Coon J, Dominiczak A, Ehrich JHH, Fliser D, Girolami M, Hermjakob H, Hochstrasser D, Jankowski J, Julian BA, Kolch W, Massy ZA, Neususs C, Novak J, Peter K, Rossing K, Schanstra J, Semmes OJ, Theodorescu D, Thongboonkerd V, Weissinger EM, Van Eyk JE, Yamamoto T (2007) **Clinical proteomics: A need to define the field and to begin to set adequate standards**. *Proteomics Clinical Applications*, 1, 148-156.

Moffat BA, Chenevert TL, Lawrence TS, Meyer CR, Johnson TD, Dong Q, Tsien C, Mukherji S, Quint DJ, Gebarski SS, Robertson PL, Junck LR, Rehemtulla A,

Ross BD (2005) **Functional diffusion map: a noninvasive MRI biomarker for early stratification of clinical brain tumor response**. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 5524-9.

Newman SA (1994) **Meningiomas: a quest for the optimum therapy**. *Journal of Neurosurgery*, 80, 191-4.

Nishimura T, Ogiwara A, Fujii K, Kawakami T, Kawamura T, Anyouji H, Kato H (2005) **Disease proteomics toward bedside reality**. *Journal of Gastroenterology*, 40 Suppl 16, 7-13.

Nordmark J, Rubertsson S, Mortberg E, Nilsson P, Enblad P (2009) **Intracerebral monitoring in comatose patients treated with hypothermia after a cardiac arrest**. *Acta Anaesthesiologica Scandinavica*, 53, 289-98.

Ohgaki H, Kleihues P (2005a) **Epidemiology and etiology of gliomas**. *Acta Neuropathologica*, 109, 93-108.

Ohgaki H, Kleihues P (2005b) **Population-based studies on incidence, survival rates, and genetic alterations in astrocytic and oligodendroglial gliomas**. *Journal of Neuropathology and Experimental Neurology*, 64, 479-89.

Okamoto H, Li J, Vortmeyer AO, Jaffe H, Lee YS, Glasker S, Sohn TS, Zeng WF, Ikejiri B, Proescholdt MA, Mayer C, Weil RJ, Oldfield EH, Zhuang ZP (2006) **Comparative proteomic profiles of meningioma subtypes**. *Cancer Research*, 66, 10199-10204.

Orre LM, Pernemalm M, Lengqvist J, Lewensohn R, Lehtio J (2007) **Up-regulation, modification, and translocation of S100A6 induced by exposure to ionizing radiation revealed by proteomics profiling**. *Molecular and Cellular Proteomics*, 6, 2122-31.

Otomo T, Hishii M, Arai H, Sato K, Sasai K (2004) **Microarray analysis of temporal gene responses to ionizing radiation in two glioblastoma cell lines: up-regulation of DNA repair genes**. *Journal of Radiation Research*, 45, 53-60.

Papac DI, Shahrokh Z (2001) **Mass spectrometry innovations in drug discovery and development**. *Pharmaceutical Research*, 18, 131-145.

Pardal R, Clarke MF, Morrison SJ (2003) **Applying the principles of stem-cell biology to cancer**. *Nature Reviews. Cancer*, 3, 895-902.

Pasikanti KK, Ho PC, Chan ECY (2008) **Gas chromatography/mass spectrometry in metabolic profiling of biological fluids**. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences*, 871, 202-211.

Peroulis I, Jonas N, Saleh M (2002) **Antiangiogenic activity of endostatin inhibits C6 glioma growth**. *International Journal of Cancer*, 97, 839-45.

Persson O (2008) **Gene and Protein Expression in Human Gliomas**. *Doctoral Dissertation*. Department of Clinical Sciences. Lund University, Lund, Sweden

Persson O, Brynnel U, Levander F, Widegren B, Salford LG, Krogh M (2009) **Proteomic expression analysis and comparison of protein and mRNA expression profiles in human malignant gliomas**. *Proteomics. Clinical applications*, 3, 83-94.

Petricoin E, 3rd, Liotta LA (2003) **Counterpoint: The vision for a new diagnostic paradigm**. *Clinical Chemistry*, 49, 1276-8.

Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, Mills GB, Simone C, Fishman DA, Kohn EC, Liotta LA (2002) **Use of proteomic patterns in serum to identify ovarian cancer**. *Lancet*, 359, 572-7.

Plate KH, Breier G, Millauer B, Ullrich A, Risau W (1993) **Up-regulation of vascular endothelial growth factor and its cognate receptors in a rat glioma model of tumor angiogenesis**. *Cancer Research*, 53, 5822-7.

Plate KH, Breier G, Weich HA, Risau W (1992) **Vascular endothelial growth factor is a potential tumour angiogenesis factor in human gliomas in vivo**. *Nature*, 359, 845-8.

Pohjanen E, Thysell E, Jonsson P, Eklund C, Silfver A, Carlsson IB, Lundgren K, Moritz T, Svensson MB, Antti H (2007) **A multivariate screening strategy for investigating metabolic effects of strenuous physical exercise in human serum**. *Journal of Proteome Research*, 6, 2113-2120.

Provias J, Claffey K, delAguila L, Lau N, Feldkamp M, Guha A (1997) **Meningiomas: role of vascular endothelial growth factor/vascular permeability factor in angiogenesis and peritumoral edema**. *Neurosurgery*, 40, 1016-26.

Qu Y, Adam BL, Yasui Y, Ward MD, Cazares LH, Schellhammer PF, Feng Z, Semmes OJ, Wright GL, Jr. (2002) **Boosted decision tree analysis of surface-enhanced laser desorption/ionization mass spectral serum profiles discriminates prostate cancer from noncancer patients**. *Clinical Chemistry*, 48, 1835-43.

Ragel B, Jensen RL (2003) **New approaches for the treatment of refractory meningiomas**. *Cancer Control*, 10, 148-58.

Raine CS (1999) **Neurocellular Anatomy**. In *Basic Neurochemistry: Molecular, Cellular and Medical Aspects*, Siegel GJ (ed) pp. 1-30. Lippincott-Raven Publishers: Philadelphia.

Ransohoff DF (2005a) **Bias as a threat to the validity of cancer molecular-marker research**. *Nature Reviews. Cancer*, 5, 142-9.

Ransohoff DF (2005b) **Lessons from controversy: ovarian cancer screening and serum proteomics**. *Journal of the National Cancer Institute*, 97, 315-9.

Rempel SA, Ge S, Gutierrez JA (1999) **SPARC: a potential diagnostic marker of invasive meningiomas**. *Clinical Cancer Research*, 5, 237-41.

Rogers MA, Clarke P, Noble J, Munro NP, Paul A, Selby PJ, Banks RE (2003) **Proteomic profiling of urinary proteins in renal cancer by surface enhanced laser desorption ionization and neural-network analysis: identification of key issues affecting potential clinical utility**. *Cancer Research*, 63, 6971-83.

Roser F, Samii M, Ostertag H, Bellinzona M (2004) **The Ki-67 proliferation antigen in meningiomas. Experience in 600 cases**. *Acta neurochirurgica*, 146, 37-44; discussion 44.

Roslin M, Henriksson R, Bergstrom P, Ungerstedt U, Bergenheim AT (2003) **Baseline levels of glucose metabolites, glutamate and glycerol in malignant glioma assessed by stereotactic microdialysis**. *Journal of Neuro-Oncology*, 61, 151-60.

Ruan K, Song G, Ouyang G (2009) **Role of hypoxia in the hallmarks of human cancer**. *Journal of Cellular Biochemistry*, 107, 1053-1062.

Ruttledge MH, Sarrazin J, Rangaratnam S, Phelan CM, Twist E, Merel P, Delattre O, Thomas G, Nordenskjold M, Collins VP, et al. (1994) **Evidence for the complete inactivation of the NF2 gene in the majority of sporadic meningiomas**. *Nature Genetics*, 6, 180-4.

Sandler A, Gray R, Perry MC, Brahmer J, Schiller JH, Dowlati A, Lilenbaum R, Johnson DH (2006) **Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer**. *The New England Journal of Medicine*, 355, 2542-50.

Sandstrom M, Johansson M, Andersson U, Bergh A, Bergenheim AT, Henriksson R (2004) **The tyrosine kinase inhibitor ZD6474 inhibits tumour growth in an intracerebral rat glioma model**. *British Journal of Cancer*, 91, 1174-80.

Sandström M, Johansson M, Bergström P, Bergenheim AT, Henriksson R (2008) **Effects of the VEGFR inhibitor ZD6474 in combination with radiotherapy and temozolomide in an orthotopic glioma model**. *Journal of Neuro-Oncology*, 88, 1-9.

Sandström M, Johansson M, Sandström J, Bergenheim AT, Henriksson R (1999) **Expression of the proteolytic factors, tPA and uPA, PAI-1 and VEGF during malignant glioma progression**. *International Journal of Developmental Neuroscience*, 17, 473-81.

Santarelli JG, Udani V, Yung YC, Cheshier S, Wagers A, Brekken RA, Weissman I, Tse V (2006) **Incorporation of bone marrow-derived Flk-1-expressing CD34+ cells in the endothelium of tumor vessels in the mouse brain**. *Neurosurgery*, 59, 374-82; discussion 374-82.

Sauer U, Heinemann M, Zamboni N (2007) **Genetics. Getting closer to the whole picture**. *Science*, 316, 550-1.

Schmidt NO, Westphal M, Hagel C, Ergun S, Stavrou D, Rosen EM, Lamszus K (1999) **Levels of vascular endothelial growth factor, hepatocyte growth factor/scatter factor and basic fibroblast growth factor in human gliomas and their relation to angiogenesis**. *International Journal of Cancer*, 84, 10-8.

Shweiki D, Itin A, Soffer D, Keshet E (1992) **Vascular endothelial growth factor induced by hypoxia may mediate hypoxia-initiated angiogenesis**. *Nature*, 359, 843-5.

Simpson D (1957) **The recurrence of intracranial meningiomas after surgical treatment**. *Journal of Neurology, Neurosurgery and Psychiatry*, 20, 22-39.

Singh SK, Clarke ID, Terasaki M, Bonn VE, Hawkins C, Squire J, Dirks PB (2003) **Identification of a cancer stem cell in human brain tumors**. *Cancer Research*, 63, 5821-8.

Skrzydlewska E, Sulkowska M, Koda M, Sulkowski S (2005) **Proteolytic-antiproteolytic balance and its regulation in carcinogenesis**. *World Journal of Gastroenterology*, 11, 1251-66.

Smith L, Lind MJ, Welham KJ, Cawkwell L (2006) **Cancer proteomics and its application to discovery of therapy response markers in human cancer**. *Cancer*, 107, 232-41.

Steiner HH, Karcher S, Mueller MM, Nalbantis E, Kunze S, Herold-Mende C (2004) **Autocrine pathways of the vascular endothelial growth factor (VEGF) in**

glioblastoma multiforme: clinical relevance of radiation-induced increase of VEGF levels. *Journal of Neuro-Oncology*, 66, 129-38.

Stenlund H, Johansson E, Gottfries J, Trygg J (2009) **Unlocking interpretation in near infrared multivariate calibrations by orthogonal partial least squares**. *Anal Chem*, 81, 203-9.

Stone M (1974) **Cross-Validatory Choice and Assessment of Statistical Predictions**. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36, 111-147.

Strojnik T, Idanik B, Kos J, Lah TT (2001) **Cathepsins B and L Are Markers for Clinically Invasive Types of Meningiomas**. *Neurosurgery*, 48, 598-605.

Stupp R, Hegi ME, Mason WP, van den Bent MJ, Taphoorn MJ, Janzer RC, Ludwin SK, Allgeier A, Fisher B, Belanger K, Hau P, Brandes AA, Gijtenbeek J, Marosi C, Vecht CJ, Mokhtari K, Wesseling P, Villa S, Eisenhauer E, Gorlia T, Weller M, Lacombe D, Cairncross JG, Mirimanoff RO (2009) **Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial**. *Lancet Oncology*, 10, 459-66.

Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJ, Belanger K, Brandes AA, Marosi C, Bogdahn U, Curschmann J, Janzer RC, Ludwin SK, Gorlia T, Allgeier A, Lacombe D, Cairncross JG, Eisenhauer E, Mirimanoff RO (2005) **Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma**. *The New England Journal of Medicine*, 352, 987-96.

Tabatabaei P, Bergstrom P, Henriksson R, Bergenheim AT (2008) **Glucose metabolites, glutamate and glycerol in malignant glioma tumours during radiotherapy**. *Journal of Neuro-Oncology*, 90, 35-9.

Taguchi F, Solomon B, Gregorc V, Roder H, Gray R, Kasahara K, Nishio M, Brahmer J, Spreafico A, Ludovini V, Massion PP, Dziadziuszko R, Schiller J, Grigorieva J, Tsypin M, Hunsucker SW, Caprioli R, Duncan MW, Hirsch FR, Bunn PA, Jr., Carbone DP (2007) **Mass spectrometry to classify non-small-cell lung cancer patients for clinical outcome after treatment with epidermal growth factor receptor tyrosine kinase inhibitors: a multicohort cross-institutional study**. *Journal of the National Cancer Institute*, 99, 838-46.

Tan CS, Ploner A, Quandt A, Lehtio J, Pernemalm M, Lewensohn R, Pawitan Y (2006) **Annotated regions of significance of SELDI-TOF-MS spectra for detecting protein biomarkers**. *Proteomics*, 6, 6124-33.

Tanaka K, Waki H, Ido Y, Akita S, Yoshida Y, Yoshida T, Matsuo T (1988) **Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry**. *Rapid Communications in Mass Spectrometry*, 2, 151-153.

Trygg J, Wold S (2002) **Orthogonal projections to latent structures (O-PLS)**. *Journal of Chemometrics*, 16, 119-128.

Ungerstedt U (1997) **Microdialysis--a new technique for monitoring local tissue events in the clinic**. *Acta Anaesthesiologica Scandinavica. Supplementum*, 110, 123.

Walker MD, Strike TA, Sheline GE (1979) **An analysis of dose-effect relationship in the radiotherapy of malignant gliomas**. *International Journal of Radiation Oncology, Biology, Physics*, 5, 1725-31.

Walsh MT, Couldwell WT (2009) **Management options for cavernous sinus meningiomas**. *Journal of Neuro-Oncology*, 92, 307-16.

van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH (2002) **Gene expression profiling predicts clinical outcome of breast cancer**. *Nature*, 415, 530-6.

van Nieuwenhuizen D, Klein M, Stalpers LJ, Leenstra S, Heimans JJ, Reijneveld JC (2007) **Differential effect of surgery and radiotherapy on neurocognitive functioning and health-related quality of life in WHO grade I meningioma patients**. *Journal of Neuro-Oncology*, 84, 271-8.

Vanhoutte KJ, Laarakkers C, Marchiori E, Pickkers P, Wetzels JF, Willems JL, van den Heuvel LP, Russel FG, Masereeuw R (2007) **Biomarker discovery with SELDI-TOF MS in human urine associated with early renal injury: evaluation with computational analytical tools**. *Nephrology, Dialysis, Transplantation*, 22, 2932-43.

Wellenreuther R, Kraus JA, Lenartz D, Menon AG, Schramm J, Louis DN, Ramesh V, Gusella JF, Wiestler OD, von Deimling A (1995) **Analysis of the neurofibromatosis 2 gene reveals molecular variants of meningioma**. *American Journal of Pathology*, 146, 827-32.

Wen PY, Kesari S (2008) **Malignant gliomas in adults**. *The New England Journal of Medicine*, 359, 492-507.

Vescovi AL, Galli R, Reynolds BA (2006) **Brain tumour stem cells**. *Nature Reviews. Cancer*, 6, 425-436.

Wiklund S, Johansson E, Sjostrom L, Mellerowicz EJ, Edlund U, Shockcor JP, Gottfries J, Moritz T, Trygg J (2008) **Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models**. *Analytical Chemistry*, 80, 115-122.

Willett CG, Boucher Y, di Tomaso E, Duda DG, Munn LL, Tong RT, Chung DC, Sahani DV, Kalva SP, Kozin SV, Mino M, Cohen KS, Scadden DT, Hartford AC, Fischman AJ, Clark JW, Ryan DP, Zhu AX, Blaszkowsky LS, Chen HX, Shellito PC, Lauwers GY, Jain RK (2004) **Direct evidence that the VEGF-specific antibody bevacizumab has antivascular effects in human rectal cancer**. *Nature Medicine*, 10, 145-7.

Wold H (1966) **Estimation of principal components and related models by iterative least squares**. In *Multivariate Analysis*, Krishnaiah PR (ed) pp. 391-420. Academic press: New York.

Wold S (1978) **Cross-Validatory Estimation of Number of Components in Factor and Principal Components Models**. *Technometrics*, 20, 397-405.

Wold S, Antti H, Lindgren F, Öhman J (1998) **Orthogonal signal correction of near-infrared spectra**. *Chemometrics and Intelligent Laboratory Systems*, 44, 175-185.

Wold S, Esbensen K, Geladi P (1987) **Principal component analysis**. *Chemometrics and Intelligent Laboratory Systems*, 2, 37-52.

Wold S, Sjostrom M, Eriksson L (2001a) **PLS-regression: a basic tool of chemometrics**. *Chemometrics and Intelligent Laboratory Systems*, 58, 109-130.

Wold S, Trygg J, Berglund A, Antti H (2001b) **Some recent developments in PLS modeling**. *Chemometrics and Intelligent Laboratory Systems*, 58, 131-150.

Vredenburgh JJ, Desjardins A, Herndon JE, 2nd, Dowell JM, Reardon DA, Quinn JA, Rich JN, Sathornsumetee S, Gururangan S, Wagner M, Bigner DD, Friedman AH, Friedman HS (2007) **Phase II trial of bevacizumab and irinotecan in recurrent malignant glioma**. *Clinical Cancer Research*, 13, 1253-9.

Wulfkuhle JD, Liotta LA, Petricoin EF (2003) **Proteomic applications for the early detection of cancer**. *Nature Reviews. Cancer*, 3, 267-75.

Yao Y, Kubota T, Takeuchi H, Sato K (2005) **Prognostic significance of microvessel density determined by an anti-CD105/endoglin monoclonal antibody in astrocytic tumors: comparison with an anti-CD31 monoclonal antibody**. *Neuropathology*, 25, 201-6.

Zenobi R, Knochenmuss R (1998) **Ion Formation in MALDI Mass Spectrometry**. *Mass Spectrometry Reviews*, 17, 337-366.

Zheng Y, Xu Y, Ye B, Lei J, Weinstein MH, O'Leary MP, Richie JP, Mok SC, Liu BC (2003) **Prostate carcinoma tissue proteomics for biomarker discovery**. *Cancer*, 98, 2576-82.

Zhou BB, Elledge SJ (2000) **The DNA damage response: putting checkpoints in perspective**. *Nature*, 408, 433-9.

Zhou M, Veenstra TD (2008) **Mass spectrometry: m/z 1983-2008**. *Biotechniques*, 44, 667-670.

11. Acknowledgements

There are so many people who in various ways have contributed to this work. First and foremost, I want to express my sincerest gratitude to my team of supervisors, who have provided an excellent mix of analytical thinking from different points of view as well as solid support throughout this project.

Tommy – the driving force and principle conceiver of the project. Thank you for including me! Thanks also for continuous guidance and for encouragement when I have been overly critic. Your enthusiasm and optimism have been vital ingredients to this work. Thanks also for excellent travelling company on some unforgettable journeys, and for literally demonstrating that it is okay to drive on the wrong side of the road as well as on bicycle paths (so long as it is done carefully and only for short distances, in a foreign country).

Henrik – my knight in shining armor. Thank you for taking me on (in spite of our earlier encounters during my undergraduate studies). Thank you for educating me in multivariate analysis, for your invaluable input to all things data analysis related, and for always making time for my nonsense, even at the most inconvenient times. Thanks also for your many attempts at reviving early 90-ies music (I am looking forward to your future attempts at Wonderwall, I am sure there will be many of them). I have very much appreciated all of it.

Mikael – my beacon at the roaring sea of tumor biology. Thank you for introducing me to the project and for believing in me from start to finish. Your enormous expertise, guidance and critic evaluation of the work has been invaluable throughout this work. Thanks also for introducing me to the world of C64 gaming, scuba diving and for some classic parties.

Roger – although we have rarely interacted directly I have always felt your presence. It has been reassuring. Thank you for your contributions to the project and to the manuscripts. Let me know when you are up for that game of golf we have been talking about!

At this point I have spent an awfully long time at the very same desk, in what can only be described as one of the western world's most insufferably horrid offices. During this time people have come and gone at the remaining three desks in the office, and you have each contributed to make this a fun place to work. **My Office Buddies**, in (what

might be) chronological order: **Camilla** - for teaching me proper ethics and for educating me about the importance of breakfast (I now eat relatively decent ones). **Ulrika** - for many good laughs. **Anna** - for suffering though the initial SELDI troubles together with me. **Terese** - for many pleasant talks, for letting me pilfer teabags for several years and for attempting to improve on the office environment by keeping plants here. **Jonas** - for proving that all one needs is coffee and ice cream to survive and sustain a high analytical capacity as well as an impeccable athletic physique, for always being so amusingly negative towards everything (save science and Aston Villa, funnily enough), and for some mad nights of partying at various locations in Europe. **Carin** - my beloved sister, for taking very good care of me and for equipping me with my own mug (unfortunately I believe it has perished in the years since then). **Yie Wei** and **Linn** - for introducing me to Chinese culture and for taking very good care of us when we visited in China. **Annika** - for actively making a change to our working environment, by changing the orange/turquoise/salmon/beige, ripped (and taped) and poorly hung curtains to slightly less nasty ones, by constructing some spectacular make-shift holiday decorations, by having the fluorescent lights changed after months of malfunctioning, and by replacing the phone that did not work for many years with an equally complicated but seemingly operational one. Thanks also for co-inventing several more or less suitable in-office-games. **Fleur** - for the endless supply of coffee and nice long chats. **Lina** - for being the only one I have ever shared an office with whose eyes did not go completely blank when I spoke ill of western blots or mentioned anything multivariate. **Mahmood** - for conceiving a method to keep the office summer temperature at a cool 29.6 degrees (rather than the usual 33), by closing the window blinds. *It seems none of us shall ever find out the answer to the eternal question: what is the mystery hose, poking out of the never-working ventilation?*

Everyone at the oncology lab! Special thanks to **Håkan** - for great help with proof reading manuscripts, for contagious scientific enthusiasm, for suddenly turning psychic and (semi-successfully) guiding us out of Aberdeen and for an unforgettable interpretation of Clawfinger. **Lotta** and **Mikael** - for outstanding support and collaboration in preparation of samples. **Yvonne** and **Annika** - for your patience with my many questions whenever I strayed into the lab and found myself completely lost. And for instructing me on how to operate the dishwasher about 72 times. **Marcus** - for introducing me to the animal model. **Maria** - for being an easy going and excellent co-author.

Henke's Chemometrics Group - for always making me feel welcome, even though I am an outsider from all the way across campus. Special thanks to **Anna** - for taking time and showing me around both at the lab and in the intriguing world of GC-MS-data analysis, and for suggesting plenty of interesting music to listen to. **Elin C** - for great patience with my many questions and for inspiring me to attempt cardio workouts in the early mornings. Too bad that did not stick. **Elin T** - for very nice company as a tourist in Italy. I will work on improving my sneaky-tourist-photographer skills until

next time. **Izabella** - for great collaborations, for proofing parts of this thesis and for demonstrating that it is possible to properly do Rome in a day and a half. **Lina** - for being an excellent student, for putting in a lot of hard work, and for putting up with my (mild) abuse. Thanks also to **Michael Sjöström** for fruitful collaborations and to **Fredrik Pettersson** for re-introducing me to the world of chemometrics as well as for a productive and inspiring collaboration on the first paper.

Pia, **Carina, Monica** and **Anna** - for constantly providing all the right answers and for always lending a helping hand in big and small matters.

All the brilliant people who have put up with me outside of work! **David, Britta** and **Jasmin** - for high quality after work sessions at Rött. I hear they have contained arm wrestling, the invention of the "Dirty Dave", "after work-ångest", talk about strangulation games, etc... I only wish remembered some of it. **Johanna** - for your support during this work. **Former football team mates,** especially **Uffe, Butta, Julla, Pelle, Wörlén, Nicke, Wänman, Tommy, Karlsson, Forre, Östh, Luvan, ...** - for staying in touch. Although we watch more games than we play these days, it is still all good times to hang out. **All the participants of the highly prestigious, bi-annual golf tournament Touren** - for making it what it is, a definite highlight each spring and autumn. All hail the Champ! **Linda** - for very nice company whenever I have visited Stockholm over the years. Before you became all continental that is... Now, quit your nonsense and get back :) **Joanna** - for continually inviting me to tag along on exciting adventures (which I keep bailing out of in order to squeeze in a few more hours at the office). Hence forth I will make time to actually come along sometimes. Also, thanks for showing an interest in the progress of this thesis and for your encouragement during the writing up process. **Åsa** and **Jean-Francois** - for solid support in matters of all sorts. Special thanks for proof-reading the entire text and making valuable contributions to it! You are thereby, quite possibly, the only ones to have voluntarily read it. I am sorry it was not more entertaining. **Nitte** - for interesting lunch company over the years, for some awesome parties and adventures, and for having excellent taste in movies, music and single malts. There will be more time for all that from now on! **Johan** - for not making me sleep on a dodgy bench in the middle of foreign cities too often, for your friendship, and for everything else over the last decade and a half. Thanks also for proposing that I should find myself a supervisor at the opposite end of campus. That was good advice. Now, I reckon it is time you make yourself more geographically available again so you can tend to my whiskey stash on a monthly basis. It has been neglected for way too long. **Björn** and **Cissi** - for many rounds of golf, nice dinners, and for always just being there! **Erik** and **Mattias** - for being top blokes and great friends. You know I would turn to you if I would ever wake up in Vegas next to a lifeless lady of negotiable affection :)

Annika - for your tremendous efforts during the last months of this project, including your work as a layout artist, secretary, cook, proof-reader, escort, zombie slayer and everything else. As fantastic as Däcklasse is, he's got nothing on you!

My beloved family - for always being there for me no matter what. **Carin** and **Erik** - my extraordinarily amazing sister and half-decent brother in-law, thank you for always looking after me in every conceivable way. And thank you **William** for bringing so much joy and positive energy! **Mor** och **Far** – for 100% support in everything I have ever done! For this I am eternally grateful.