Integrative modeling of intratumoral heterogeneity, plasticity and regulation in nervous system cancers

IDA LARSSON
Abstract


The adult brain tumor glioblastoma (GBM) is characterized by short survival and a lack of efficient treatments. Median survival is 15 months from time of diagnosis and the 5-year survival rate is only 7 %. There is an urgent need for more efficient treatment against GBM, but there are many challenges, including the high extent of heterogeneity of GBM. The tumoral heterogeneity of GBM ranges from interpatient to intratumoral. The aim of this thesis has been to address unanswered questions relating to the intratumoral heterogeneity of GBM, with three specific focuses; (1) the organization of GBM cell state transitions (paper I and III), (2) the regulation of cell states and cell state transitions (paper II), and (3) targeted interventions against cell states (paper II and IV).

In paper I, we develop an experimental-computational method to measure and quantify cell state transitions. We find that GBM cell states organize hierarchically, with a clear “source state” feeding cells downwards in the hierarchy towards a “sink state” with negative growth rate, but with multi-directional transitions between intermediate states.

In paper II, we address the lack of computational methods to identify regulators of intratumoral heterogeneity by developing an algorithm called scRegClust that uses scRNA-seq data to estimate regulatory programs. Through an integrative study of the regulatory landscape of neuro-oncology we find two potential regulators of the macrophage-induced mesenchymal transition in GBM.

In paper III, we explore the energy-concept as a way of measuring differentiation potential of single cells, instead of relying on gene markers or gene signatures of stemness. We fit a model called the Ising model from statistical mechanics to scRNA-seq data and show both on synthetic and real data that the estimated Ising energy is a good measure of a cell’s differentiation potential, where high Ising energy indicate a high degree of stemness.

Finally, in paper IV, another experimental-computational method is developed to investigate drug-induced effects on both inter- and intratumoral heterogeneity.

In summary, the high extent of intratumoral heterogeneity in nervous system cancer is a major caveat for the development of more efficient treatments. In this thesis we have taken a systems biology approach to understand how this heterogeneity is structured and how we can exploit that knowledge for therapeutic purposes.

Keywords: Nervous system cancer, Glioblastoma, Heterogeneity, Plasticity, Mathematical modeling

Ida Larsson, Department of Immunology, Genetics and Pathology, Neurooncology and neurodegeneration, Dag Hammarskjölds väg 20, Uppsala University, SE-751 85 Uppsala, Sweden.

© Ida Larsson 2023

ISSN 1651-6206
URN urn:nbn:se:uu:diva-498239 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-498239)
To family and friends
The more I learn,
the more I realize how much I don’t know.

Albert Einstein
List of papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.


*indicates equal contribution.

Reprints were made with permission from the publishers.
Related work by the author


# Contents

1 Introduction ................................................................................................ 11

1.1 Cancer ............................................................................................. 11
  1.1.1 Hallmarks of cancer ......................................................... 11
  1.1.2 Unlocking phenotypic plasticity ..................................... 12

1.2 Glioblastoma .................................................................................. 14
  1.2.1 Glioblastoma genomics and heterogeneity .................... 14
  1.2.2 Glioblastoma cell states .................................................. 16
  1.2.3 Glioblastoma cell state dynamics ................................... 18

1.3 Pediatric nervous system cancers .................................................. 19
  1.3.1 Medulloblastoma ............................................................. 19
  1.3.2 Neuroblastoma ................................................................. 19
  1.3.3 Experimental models of nervous system cancers .......... 20

1.4 Data-driven study of tumor heterogeneity .................................... 21
  1.4.1 Single cell RNA sequencing ........................................... 21
  1.4.2 Plasticity ........................................................................... 23
  1.4.3 Entropy as a measure of stemness .................................. 25
  1.4.4 Drug-induced changes in intratumoral heterogeneity ... 26

1.5 Mathematical optimization ............................................................ 27

2 Present investigations ................................................................................ 29

2.1 Paper I: Modeling glioblastoma heterogeneity as a dynamic network of cell states ........................................ 29

2.2 Paper II: Reconstructing the regulatory programs underlying the phenotypic plasticity of neural cancers ......................... 31

2.3 Paper III: Estimating the differentiation potential and plasticity of cancer cells using statistical mechanics .................. 33

2.4 Paper IV: Using drug-induced cell states to build therapeutic combinations against nervous system cancers. ................... 35

3 Discussion and future perspectives ........................................................... 36

4 Populärvetenskaplig sammanfattning ....................................................... 40

5 Acknowledgements ................................................................................... 42

References ........................................................................................................ 44
Abbreviations

AC    Astrocyte
ADMM  Alternating direction method of multipliers
ADR N  Adrenergic
BBB   Blood-brain-barrier
CDK4  Cyclin-dependent kinase 4
CSC   Cancer stem cell
DCX   Doublecortin
DNA   Deoxyribonucleic acid
EGFR  Epidermal growth factor receptor
FACS  Fluorescence-activated cell sorter
GBM   Glioblastoma
GFAP  Glial fibrillary acidic protein
HOPX  Homeodomain-only protein
IDH   Isocitrate dehydrogenase 1
IRF8  Interferon regulatory factor 8
KLF4  Kruppel-like factor 4
MB    Medulloblastoma
MES   Mesenchymal
MGMT  Methylguanine methyltransferase
mRNA  Messenger ribonucleic acid
mTOR  Mammalian target of rapamycin
NB    Neuroblastoma
NF1   Neurofibromin 1
NPC   Neural progenitor cell
OCT3/4 Octamer-binding transcription factor 3/4
OLIG1 Oligodendrocyte transcription factor 1
OMG   Oligodendrocyte-myelin glycoprotein
OPC   Oligodendrocyte progenitor cell
OS    Overall survival
OSM   Oncostatin-M
OSMR  Oncostatin-M receptor
PBA   Population-based analysis
PDE   Partial differential equation
PDGFRA Platelet-derived growth factor receptor alpha
PI3K  Phosphoinositide 3-kinases
PPI   Protein-protein interaction
QC    Quality control
RB1   Retinoblastoma
RTK   Receptor tyrosine kinase
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHH</td>
<td>Sonic hedgehog</td>
</tr>
<tr>
<td>scRNA-seq</td>
<td>Single-cell RNA sequencing</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>SOX2</td>
<td>SRY-box transcription factor 2</td>
</tr>
<tr>
<td>STAG</td>
<td>State Transition and Growth</td>
</tr>
<tr>
<td>STAT1</td>
<td>Signal transducer and activator of transcription 1</td>
</tr>
<tr>
<td>TCGA</td>
<td>The Cancer Genome Atlas</td>
</tr>
<tr>
<td>TERT</td>
<td>Telomerase reverse transcriptase</td>
</tr>
<tr>
<td>TF</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>TME</td>
<td>Tumor microenvironment</td>
</tr>
<tr>
<td>TMZ</td>
<td>Temozolomide</td>
</tr>
<tr>
<td>TP53</td>
<td>Tumor protein P53</td>
</tr>
<tr>
<td>TTF</td>
<td>Tumor Treating Fields</td>
</tr>
<tr>
<td>UMI</td>
<td>Unique molecular identifier</td>
</tr>
<tr>
<td>WHO</td>
<td>World health organization</td>
</tr>
</tbody>
</table>
1. Introduction

1.1 Cancer

Cancer is one of the leading causes of death worldwide, accounting for around 10 million deaths annually. In Sweden, the cancer incidence is almost 300 new cases per 100,000 residents per year, with prostate and breast cancer being the two most common cancer incidences, while lung cancer is the most common cancer-related cause of death [1].

1.1.1 Hallmarks of cancer

Biologically, cancer is a heterogeneous group of disorders whose common feature is uncontrolled cell growth and spreading, either to surrounding tissue or other parts of the body [2]. In 2000, Hanahan and Weinberg made an effort to organize the collected knowledge from cancer research to define the "hallmarks of cancer" - a tool for describing the essential alterations to a normal cell to make it malignant [3]. Including the updates in 2011 and 2022, the hallmarks of cancer consists of eight acquired capabilities, two enabling characteristics and four emerging traits that are common to the majority, if not all, of cancers (Figure 1) [4, 5].

Four hallmarks directly concern the various strategies a cancer cell acquire to evade normal regulatory mechanisms of cell growth and division; sustaining proliferative signaling, evading growth suppressors, resisting cell death and enabling replicative mortality. As an example, cancer cells develop growth factor independence in a number of ways, e.g. by producing growth factor ligands themselves, stimulating normal cells to produce more growth factor ligands and increasing the number of growth factor receptors at the cancer cell surface, all strategies to maximize proliferation [6, 7]. Further, normal cells typically have an upper limit on the number of cell divisions they can go through, imposed by the length of the telomeres at the end of the chromosomes. At each division, the telomeres are shortened and finally crisis and apoptosis are triggered in the cell leading to cell death. Cancer cells on the other hand evade this mechanism by expressing telomerase, an enzyme capable of adding telomere sequences to the end of chromosomes [8, 9].

The four additional acquired capabilities concern other processes essential to tumorigenesis. To get access to nutrients and oxygen and to clear toxic waste tumors induce angiogenesis, the process in which new blood vessels are formed from existing ones. In normal vasculature, angiogenesis is only
active during embryogenesis and temporally during processes such as wound-healing, while it is continuously turned on in tumors. The ability of tumor cells to invade surrounding tissue and metastasize to locations distant from the primary tumor site is another hallmark, often indicative of a higher grade of malignancy. Finally, the deregulation of cellular metabolism and the ability to avoid immune destruction were added to the acquired capabilities in the 2011 edition of the hallmarks of cancer [4].

The acquisition of above mentioned capabilities is enabled by two specific characteristics of the cancer cells; genome instability and mutation, and tumor-promoting inflammation.

The hallmarks of cancer presents a summary of the collected knowledge gathered in the cancer research community and as the field have progressed, more hallmarks have been added to the list. Last year, four new emerging hallmarks and enabling characteristics were added, one of them being the unlocking of phenotypic plasticity [5].

1.1.2 Unlocking phenotypic plasticity

Cell differentiation during normal tissue development is considered a highly hierarchical process, where self-renewing stem cells give rise to progenitor cells and finally terminally differentiated cells that have lost their ability to proliferate. However, the terminally differentiated state is not necessarily set in stone. One famous example is the Yamanaka factors [10], the Nobel Prize-rewarded discovery that by inducing the expression of four transcription factors (TFs) (Oct3/4, Sox2, c-Myc, Klf4) a mature cell can be reprogrammed to a pluripotent stem cell, demonstrating that terminally differentiated cells still
carry the potential for plasticity. During tumorigenesis, cancer cells use this potential to disrupt the normal differentiation processes in order to evade a state of terminal differentiation [5]. The disrupted differentiation in cancer cells manifests either as dedifferentiation, blocked differentiation or transdifferentiation (Figure 1). An example of dedifferentiation has been observed in malignant melanoma. The cell-of-origin was through a lineage-tracing study identified as a mature pigmented melanocyte which during tumorigenesis looses the expression of MITF, a TF driving differentiation. Simultaneously, the cell starts expressing genes characteristic of neural crest cells, indicating that the cell has dedifferentiated towards a more progenitor-like state [11]. Through transdifferentiation, a cell committed to a specific developmental path can switch to follow a completely different path dependent on other processes and pathways than its cell-of-origin. The consequences of disrupted differentiation in cancer are many, e.g. the acquisition of drug resistance through transdifferentiation [12], which has been observed in basal-cell carcinoma (BCC) upon Hedgehog pathway inhibition treatment [13, 14]. The existence of phenotypic plasticity is extensive in the malignant brain tumor glioblastoma [15].
1.2 Glioblastoma

Glioblastoma (GBM) is a diffuse, astrocytic glioma diagnosed on the basis of \textit{IDH} mutational status (wildtype) and the presence of any of the following features: microvascular proliferation, necrosis, \textit{TERT} promoter mutation, \textit{EGFR} gene amplification and +7/-10 chromosome copy-number changes [16]. It’s the most common malignant brain tumor in the adult population, accounting for approximately 50\% of all primary malignant brain tumors. Patients are typically between the age of 55-85 at onset and average incidence rate is 3-6 cases per 100 000 people [16, 17]. The existence of any underlying high penetrance risk factor for GBM is largely unknown [18], but in recent years it has been reported that the incidence rate is increasing globally [19, 20]. The increase cannot be fully explained by better diagnostic tools and thus lifestyle choices and environmental factors, including non-ionizing radiation from mobile phones, have been investigated as contributing factors [21, 22, 23, 24]. However, conclusive evidence for a causal relationship are yet to be found.

The prognosis for affected patients is poor, median survival from time of diagnosis is around 15 months and 5-year survival rate is barely 7 \% [16], making GBM one of the most severe cancer diagnoses to get. The first-line treatment is maximal surgical resection followed by radiotherapy and concomitant chemotherapy with temozolomide (TMZ) [25]. Although the effect is modest, the addition of TMZ to the treatment regimen was considered the biggest breakthrough in GBM treatment for decades, increasing the median survival from 12.1 to 14.6 months [26]. Methylation of the O\textsuperscript{6}-methylguanine DNA methyltransferase (MGMT) promoter region is a prognostic biomarker for GBM, where patients with a methylated MGMT promoter respond better to TMZ treatment and have a longer overall survival (OS) [27, 28]. There is no standard treatment option for recurrent GBM, but eligible patients are recommended clinical trials [25]. Some novel treatment options for GBM, including Tumor Treating Fields (TTF) [29, 30] and cell vaccines [31], have shown survival benefit in clinical trials when added to the standard treatment regimen.

There is not a lack of ambitious efforts to find new treatments against GBM. There is, however, a lack of successful trials where a proposed treatment prolongs survival [32, 33]. Challenges include the blood-brain-barrier (BBB), which is a significant hurdle for efficient drug delivery, the immunosuppressive nature of the tumor, and the notorious invasive behaviour of GBM cells [34, 35, 36]. Another main challenge is the high extent of tumor heterogeneity.

1.2.1 Glioblastoma genomics and heterogeneity

The genomic and epigenetic landscape of GBM has been mapped through many efforts [37, 38, 39, 40, 41, 42]. GBM was the first cancer to be systematically studied by the Cancer Genome Atlas (TCGA) initiative through
Figure 2. The main altered pathways in GBM. Red indicate the occurrence of activating mutations, blue indicate the occurrence of inactivating mutations. Adapted from [37].

The molecular profiles of the tumors are used to stratify patients into subtypes. The number and exact definition of the subtypes have evolved over the years [38, 39, 40], but currently tumors are classified as either proneural, mesenchymal or classical [47]. Common gene alterations can be linked to the subtypes, with EGFR amplification being more frequent in the classical subtype while NF1 deletion occurs mostly in the mesenchymal subtype and amplification of PDGFRA in the proneural subtype. Although important for the understanding of GBM biology, the subtypes have limited clinical rele-
vance [38] and factors such as patient age and the amount of surgical resection are more relevant predictors of survival [48]. The subtype classification highlights the interpatient heterogeneity in GBM, but the tumors also display a high degree of intratumoral heterogeneity.

1.2.2 Glioblastoma cell states

The high extent of cell-to-cell heterogeneity in individual GBM tumors was described already in the 1930’s by the German neuropathologist Hans Joachim Scherer [36, 49]. Recently, this intratumoral heterogeneity has been studied in greater detail using modern approaches [50, 51, 52]. Sottoriva, et al. developed a multisampling scheme to collect tumor fragments from several regions of a patient tumor and found that regions from the same tumor had different subtype classifications [51], a finding that was corroborated by Patel, et al. in the first single cell study of GBM [52]. In the same study, the 430 cells from 5 patient tumors were shown to have distinct transcriptional profiles that clustered into 4 meta-modules representing processes such as the cell cycle or hypoxia. Many single cell studies of GBM have followed with slight variations in cell classification [47, 53, 54, 55, 56, 57, 58], e.g. whether cells are characterized using a discrete or continuous scale, but with resulting cell states significantly overlapping. The leading classification scheme today was introduced by Neftel, et al. in 2019 [16, 53] and comprises four main states, or meta-modules, that tumor cells can reside in, either driven by the recapitulation of normal neurodevelopmental cell types or cues from the microenvironment.

**Neural progenitor-like state**

Cells in the neural progenitor cell-like (NPC-like) state upregulate markers common to normal neural progenitors, such as $SOX4$, $SOX11$ and $DCX$ [59]. The state is further subdivided into NPC1 or NPC2, distinguished by markers of oligodendrocyte progenitor cells (OPCs) or neurons, respectively. In line with the resemblance of a progenitor cell type, a substantial portion of cells in the NPC-like states are classified as cycling (30 % and 22.5 %). Individual patient tumors in general contain cells in all states, but with a bias towards one or two states. To understand this, genetic drivers influencing the distribution of cell states were investigated, and $CDK4$ could be linked to the NPC-like state, both through computational analysis and validation in mouse models [53]. The NPC-like state was recently shown to be enriched in GBM cells driving diffuse tumor invasion [60].

**Oligodendrocyte progenitor-like state**

The oligodendrocyte progenitor cell-like (OPC-like) state is characterized by markers of the oligodendroglial lineage, such as $OLIG1$ and $OMG$ [59], and
contains the largest proportion of cycling cells (~32%) of all states. Amplification of *PDGFRA* was associated with the OPC-like state, suggesting *PDGFRA* as a genetic driver of this state. Both the NPC-like and OPC-like states have been associated with the proneural subtype [53]. OPC-like cells preferentially invade the brain through vessel co-option in an individual manner, as opposed to AC-like cells (below) who also prefer invasion through vessel co-option but in a collective manner [61, 62].

**Astrocyte-like state**
The astrocyte-like (AC-like) state is characterized by markers of normal astrocytes, such as *GFAP* and *S100B* [63, 59], but also by markers of radial glia cells, e.g. *HOPX* [64]. The AC-like state have the highest overlap with the classical subtype and as expected, high-level *EGFR* amplifications were detected in the AC-like cells, pointing towards *EGFR* as a genetic driver of this state [53].

**Mesenchymal-like state**
The mesenchymal-like (MES-like) state is characterized by markers, such as *VIM*, that are normally expressed by non-cancerous mesenchymal cells and cells undergoing mesenchymal processes. The state is further subdivided into hypoxia-independent MES1 and hypoxia-dependent MES2, the latter associated with hypoxia-response genes and glycolysis [53]. In contrast with the other states, the MES-like state have a limited overlap with normal neurodevelopmental cell types but has been observed to interact with cell types in the tumor microenvironment (TME), especially tumor-associated macrophages (TAMs) but also T-cells [47, 65, 66]. It’s been suggested that this interaction with the TME is driving the MES-like state, possibly through a mechanism where the macrophage-derived factor *OSM* binds to its receptor *OSMR* on GBM cells and initiates *STAT3*-signaling [67]. Another study observed upregulation of *IRF8* and initiation of *STAT1*-signaling in MES-like cells following a sustained immune attack. *IRF8* is a TF normally only active in hematopoietic cells (such as macrophages) but the authors speculated that GBM cells could hijack the expression of *IRF8* in a form of "myeloid mimicry" [68]. Recurrent therapy-resistant GBM tumors display a higher fraction of MES-like cells [65, 69], possibly due to a phenotypic shift in the cells from non-MES to MES upon treatment [70].

Evidently, GBM intratumoral heterogeneity is highly complex, consisting of distinct transcriptional states with diverse functional implications. To make the system even more complex, evidence of a high degree of phenotypic plasticity is amounting with cells being able to switch between states.
1.2.3 Glioblastoma cell state dynamics

The cancer stem cell (CSC) theory was originally presented as a model for explaining the observed heterogeneity between cancer cells in a tumor. The theory proposed that there exists a small population of CSCs at the top of a hierarchy, and that these differentiate into progenitors and finally fully differentiated cells [71]. A strict hierarchy imply that the most efficient therapeutic strategy would be to target the CSC population, since non-CSCs would not be able to sustain tumor growth [72]. In GBM, a putative CSC was identified on the basis of tumor initiation capacity in vivo and expression of the surface marker CD133 [73] and evidence of an invariant stem cell hierarchy was demonstrated using genetic barcoding [74]. However, accumulating evidence indicate that the CSC theory is not entirely applicable in GBM and point towards a less hierarchical, plastic organization of states. First, the experimental technique of identifying CSCs using sorting based on surface markers and tumor-initiating properties has limitations [15, 75] and several publications have shown that diverse GBM populations, including those not expressing stemness markers such as CD133, are able to form tumors in vivo [76, 77, 78]. Second, a more plastic organization of GBM phenotypic states has been demonstrated [53, 79]. Neftel, et al. used lineage tracing in combination with scRNA-seq and showed that cells from the same lineage, i.e. carrying the same barcode, recapitulated all four states. Dirkse, et al. used a different approached and sorted cells into 16 subpopulations based on the expression of four surface markers. By following the pure subpopulations in culture over time, they showed that the majority of populations could repopulate all other states, indicating a high degree of plasticity [79] (Figure 3).
1.3 Pediatric nervous system cancers

Cancer is typically a disease of age where mutations, oncogenic hits, are accumulated over time [80]. For that reason, pediatric cancers are fortunately rare disorders that are expected to affect only around 0.02 % of all Swedish children [81]. The OS rate has dramatically increased during the last 50 years to 80 % with modern treatment [82, 83]. However, even though these numbers are encouraging, not only overall survival but quality of life has to be considered for these small children receiving aggressive cancer treatment during years of critical development. Severe side effects of radiation and chemotherapy treatment include neurocognitive impairment, growth deficits, and development of secondary cancer. Therefore, the progress that has been seen in pediatric oncology treatment the last years has to do with the acknowledgement of the heterogeneic nature of these diseases and an increased risk stratification of patients, to identify low-risk patients and without survival loss de-escalate treatment for this group with improved quality of life as a result [82, 84]. The challenge that remains is to find efficient targeted treatments for high-risk patient groups, where the overall survival is significantly less than the average 80 % and severe side-effects of current treatment are experienced [85]. Two of the most common pediatric cancers originate in the nervous system.

1.3.1 Medulloblastoma

Medulloblastoma (MB) accounts for 20 % of all pediatric brain cancers, making it one of the primary reasons for cancer-related death among children under the age of 19 [86]. MB tumors are grouped into one of four main subgroups; WNT, SHH, group 3 or group 4, and each of these groups are further divided into subtypes [87, 88, 89]. The WNT-subgroup has the most favourable outcome (OS = 98 %) while the prognosis for patients with group 3 tumors is poor (OS < 50 %) [90]. Unlike GBM, individual MB tumors don’t appear to be a mixture of subgroups, but rather display subgroup-specific intratumoral heterogeneity, as demonstrated by recent single cell studies [91, 92]. One big challenge remaining for MB treatment is to find efficient therapy that targets the group 3 tumors, which are known to be driven by MYC amplification, an intractable drug target [93].

1.3.2 Neuroblastoma

Neuroblastoma (NB) arises in the sympathetic nervous system, most commonly in the adrenal medulla. It primarily affects small children, the median age of diagnosis is two years. The age of incidence together with other clinical and molecular markers, such as MYCN amplification status, stratifies the patients into low-, intermediate-, or high-risk, with the median survival ranging from almost 90 % for low-risk patients but decreasing dramatically
to below 40 % for high-risk patients [94]. A common scenario for high-risk patients is complete remission upon initial treatment, followed by tumor relapse and highly increased therapy resistance, which points toward a changed tumor cell composition after treatment. In a paper from 2017, van Groningen and colleagues [95] addresses this issue by defining two differentiation states of NB representing either an adrenergic (ADRN) or mesenchymal (MES) lineage. They show that the states are able to interconvert [96] and that treatment enriches for the MES state, indicating a higher degree of therapy resistance of this state compared to the ADRN state. Treatments targeting the MES state, and/or the conversion from ADRN to MES, could therefore be a promising strategy to battle therapy resistance in high-risk neuroblastoma. Recent scRNA-seq studies have revealed differences in cell type composition between high-risk and low-risk tumors, highlighting the heterogeneous nature of the tumor [97, 98].

1.3.3 Experimental models of nervous system cancers

The choice of disease model is an important consideration in the pre-clinical research setting. Many different model systems are used to study nervous system cancers, including in vitro, in vivo, ex vivo and hybrid systems [99] in between these. Traditionally, the most frequently used in vitro model is established cell lines from patient-derived cell cultures grown in serum-containing media [100]. The drawback with the serum-containing media is that cell profiles tend to drift away from the tumor of origin [101]. As a consequence, patient-derived cell cultures today are preferentially established in serum-free media to better reflect the original patient material [102, 103]. Commonly used in vivo models of neural cancers are rodents (mice and rat). Mice are either genetically engineered to develop tumors spontaneously or tumor cells are injected into immunodeficient mice [104, 105, 106, 107]. In vivo models have many advantages, including the ability to model the tumor’s interaction with a tumor microenvironment, but tumors often take long time to develop and they are demanding both in terms of cost and handling. Recently, zebrafish has been presented as a scalable in vivo alternative [108, 109].
1.4 Data-driven study of tumor heterogeneity

The technology to sequence DNA wasn’t developed until the later part of the 20th century, but has already revolutionized our understanding of biology and enabled ground-breaking discoveries in areas such as medicine, evolutionary biology and forensics, just to name a few (Figure 4) [110]. The continued advances within sequencing technology and the enormous amounts of data being generated enables the use of a data-driven perspective on biomedical research in general, and tumor heterogeneity studies in particular, but also challenges us to develop computational methods to accurately analyze and get the most out of these many terabytes of biological information.

Transcriptomics, the use of high-throughput methods to study the transcriptome of a sample, has during the last decade seen a shift towards using sequencing instead of microarrays to measure mRNA levels. RNA sequencing (RNA-seq) holds several advantages over microarrays, including sensitivity, dynamic range and no need for prior knowledge about the reference genome [111]. In the standard bulk RNA-seq approach, the average gene expression of all cells in a sample is measured. This is useful in many settings, e.g. when comparing the expression of genes between two conditions, but is insufficient when studying intratumoral heterogeneity where the main assumption is that the gene expression varies between cells in a single sample. A more appropriate technique is single cell RNA sequencing (scRNA-seq), a technology that offers unprecedented opportunities to study aspects of intratumoral heterogeneity such as cell state composition, plasticity, regulation and drug response [112].

1.4.1 Single cell RNA sequencing

The first paper on scRNA-seq was published already in 2009 [113], but the method’s popularity started to rise around 2014 when it was made more accessible as a result of better protocols and decreasing sequencing costs [114]. There are a multitude of available scRNA-seq protocols with slight differences [115, 116, 117], but the core of the method is the isolation of single cells before sequencing, e.g. in microfluidic droplets. The same procedures as during bulk RNA-seq are carried out, but confined to the isolated reaction chambers. Due to the low amount of starting material and the high cell-to-cell variability, the data generated from scRNA-seq experiments comes with its own set of unique computational and analytical challenges.

One of the first steps in pre-processing of the raw data is quality control (QC) of the cells. For scRNA-seq data, three important metrics to consider are number of counts per cell (count depth), number of genes per cell and number of mitochondrial gene counts per cell. In conjunction, these metrics can indicate whether a cell is viable or not and the possibility of it being a doublet (two cells sequenced in the same reaction chamber) [118]. However,
caution should be taken before throwing away cells in the QC step, e.g. high mitochondrial gene count could be an indication of a dying cell, but it could also be a viable cell with a high energy demand [119].

Accurate gene count quantification and normalization is an important step to be able to compare gene levels between cells. In a bulk RNA-seq experiment, a common assumption is that count depth (number of mRNA molecules) is the same in all samples and that any variation is due to technical reasons. The assumption is central for many of the normalization strategies developed for bulk RNA-seq data. On a single cell level, variation in count depth between cells are to be expected for biological reasons and therefore spike-ins and/or unique molecular identifiers (UMIs) are often part of the experimental protocol to facilitate gene or read quantification and comparison between cells [120, 121]. Further, many normalization strategies have been specifically designed for scRNA-seq data to account for factors such as dropout events leading to zero-inflated data. Important to consider when choosing normalization method is the protocol used, data generated by the 10X Chromium platform is different from data generated according to the Smart-seq2 protocol and must be treated accordingly [122, 123, 124]

Finally, additional pre-processing steps might include correcting for unwanted biological effects (most commonly cell cycle) [125], batch correction [126] and imputation to recover gene expression, but the need for these are dependent on the intended downstream analysis [127, 128].
1.4.2 Plasticity

To combine scRNA-seq with clustering algorithms has rapidly become the standard way of identifying known and novel cell states within a tumor [47, 52, 53, 129]. However, the study of intratumoral plasticity, how cells switch between phenotypic states, is complicated by the fact that a cell is destroyed during sequencing, making tracking of one cell’s transcriptional changes over time impossible. To overcome this limitation, one set of methods are concerned with developing algorithms to infer the underlying dynamical system from the snapshot of cell states that scRNA-seq provides [130, 131, 132], while another set of ideas develop experimental techniques to obtain time series data used to model the dynamical system [79, 133, 134] (Figure 5).

Trajectory inference and pseudotime

If cells haven’t been experimentally synchronized, a scRNA-seq measurement will contain cells distributed along a continuum of states, representing various stages of a biological process such as cell differentiation [130]. Mathematically, we can think of the sequenced cells and their inferred states as representing a sample from a probability distribution over cell states. Methods concerned with trajectory inference seeks to model the underlying stochastic process that has given rise to the observed probability distribution and through that computationally order cells along a trajectory representing the biological process. To define the state coordinates, the first steps in a trajectory inference method are feature selection and dimensionality reduction. If the underlying stochastic process is assumed to be a Markov process, single cell dynamics can be formulated as a set of partial differential equations (PDEs) known as drift-diffusion equations [135]. The Population Balance Analysis (PBA) algorithm solves this set of equations under steady-state conditions [130] and the Markov process is inferred as a Markov chain on a cell-cell graph from which lineage trajectories and pseudotime can be derived. PBA is not the only available algorithm for trajectory inference [132, 136, 137, 138]. Pseudotime as a concept was first introduced in the Monocle toolkit [132] and can be described as a value assigned to each cell quantifying it’s distance from the start of the trajectory (the root state) [139]. In Monocle, a minimum spanning tree (MST) is constructed and the longest trajectory as well as branching points are inferred. A more recent but already frequently used algorithm is STREAM [54, 138]. Following dimensionality reduction, STREAM infers the trajectory using ElPiGraph, an Elastic Principal Graph implementation [140], and in addition allows new cells to be projected onto the previous trajectory without the need to rerun the entire pipeline.

All the described trajectory inference models are examples of top-down modeling, where the data from many cells are analyzed together to infer the state of individual cells. An alternative approach, less used in the single cell context, is bottom-up modeling where an explicit dynamic model is assumed.
and used to model the data from each cell, independent of the other cells in the data set [135]. An example of a bottom-up modeling technique is RNA velocity.

**RNA velocity**

RNA velocity [131, 141] exploits a mistake in scRNA-seq protocols. These protocols should enrich for spliced mRNA, but around 15-20 % of the detected molecules are in fact unspliced pre-mRNA sequences. By quantifying the abundance of spliced versus unspliced RNA for each gene and using a model of transcriptional dynamics which takes transcription, splicing and degradation rates into account, the RNA velocity vector for each cell can be derived. RNA velocity is defined as the time-derivative of the gene expression state and predicts the future state of a cell. Since 2018 when the method was published it has gained widespread popularity, and has been employed in several GBM studies [142, 143].

The above described approaches have in common that they infer dynamical cell behaviour from a snapshot of the current state of the system. A limitation with this type of approach is that many different dynamical behaviours can give rise to the same state distribution, leading to an inherent uncertainty in the inferred dynamics [130]. A way to circumvent the problem is to make certain assumptions about the system, e.g. the principle of maximum parsimony which dictates that the preferred dynamical model is the one with a minimum number of transcriptional changes [112]. An alternative strategy is to combine computational modeling with experimental techniques to collect time series data, where the dynamical behaviour is observed and not inferred.

**Population sorting**

If the phenotypic states of interest are previously known, one well-established approach to study state transitions is to sort cells into pure subpopulations based on surface markers, e.g. using FACS, and measure how the state distribution over time change from being a pure population to reaching an equilibrium where potentially all states are present. This type of experimental data can then be modelled using a Markov chain model to derive transition rates [79, 133]. Limitations with this approach are the dependence on finding proper surface markers for each state and the lack of flexibility. States are defined prior to the start of the experiment and if later data show that these states are not representative, the time-series data generated is unusable. A second, perhaps less obvious drawback, is that Markov chains cannot represent individual growth rates for each state [144], meaning that if one state is rapidly proliferating, this could be misinterpreted as high transition rates to this state.

**Lineage tracing of single cells**

An alternative to the population-based approach is lineage tracing of single
1.4.3 Entropy as a measure of stemness

To infer lineage trajectories and pseudotime, a root state needs to be defined. Defining the root state is not trivial, especially not in the cancer context, and goes back to the problem of identifying CSCs. As already discussed, the traditional approach to identifying CSCs are based on the expression of certain markers, e.g. CD133, CD44 and SOX2 [73, 146, 147] in GBM. A related approach is to examine the enrichment of stemness gene signatures in individual cells. However, there is no consensus set of GBM CSC markers and stemness signatures derived from different GBM studies show surprisingly low overlap ([148]). Entropy-based models are emerging as an alternative measure of
stemness [149, 150, 151], or differentiation potential, where a stem-like cell can give rise to many different cell types and therefore are considered to carry a higher differentiation potential than an already differentiated cell.

In SCENT, differentiation potential is measured through signaling entropy. The transcriptomic profile of a cell is integrated with a protein-protein interaction network (PPI) to quantify the activity of the cell’s signaling pathways, it’s signaling entropy. The rationale behind a high signaling entropy as a proxy for stemness is that a stem-like cell are not committed to any specific lineage and thus the activity of all lineage-specific pathways should be equal, while a differentiated cell activates specific pathways and thereby decreases its signaling entropy [149, 152].

In the CytoTRACE algorithm, the number of expressed genes per cell is used as a measure of a cell’s differentiation potential, as this metric was found to be highly correlated with differentiation status [150].

1.4.4 Drug-induced changes in intratumoral heterogeneity

Large-scale cancer profiling projects have been, and are still, generating extensive knowledge and new insights about tumor biology. However, to reach a more complete understanding of the tumor biological system and learn the most effective interventions against it, we want to perturb the system and profile the downstream effects of these perturbations. In the context of tumor heterogeneity, it’s of particular interest to learn whether drugs induce subgroup-specific effects, which could inform the development of combination therapy to efficiently target the whole tumor. To this end, it’s of importance that drug screening efforts employ more informative readouts than viability.

Many efforts have been made to reconcile the need of a high-throughput, cost-effective assay for large-scale screening of the cellular response to various chemical perturbations [153, 154, 155, 156]. One of the more successful ones are the L1000 assay. Through a data-driven approach, the developers identified a subset of 978 so-called landmark genes. They showed that from the expression of these genes, 81 % of the non-profiled transcriptome could be inferred, but at a considerably reduced cost compared to sequencing the entire transcriptome [157]. Currently, the L1000 has profiled how 99 different cell lines respond to 32855 small molecule drugs, making it a highly valuable resource. A second assay worth mentioning is the Multiplexed Interrogation of gene eXpression through single cell RNA Sequencing (MIX-Seq) assay [158]. MIX-Seq allows for large-scale profiling of the response to perturbations on a single cell level in a highly cost-effective way. In short, this is achieved by pooling cellines and perturbing these using chemical or genetic perturbations followed by scRNA-seq. The original celline identity of each individual cell is inferred post-sequencing using SNP demultiplexing with high accuracy.
A strength with the MIX-Seq approach is the ability to profile the response of single cells to perturbations, making it straightforward to investigate the changes imposed by the drugs to population structure. In contrast, the majority of pharmaco-transcriptomics efforts are generating bulk RNA-seq data, hence lacking cell population resolution. Many computational methods have been developed to resolve this issue, commonly referred to as cell type deconvolution strategies. Traditionally, these methods are based on cell type reference profiles [159], but recently a spectrum of strategies has emerged that utilizes the concept of mutual linearity [160], scRNA-seq data [161, 162] or deep learning [163].

1.5 Mathematical optimization

Mathematical optimization is, in simple terms, the process of finding the best solution to a problem out of all possible solutions using mathematical techniques. It is used in a wide range of disciplines, from portfolio optimization in finance to metabolic engineering in computational systems biology. A general optimization problem can be expressed as

$$\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq b_i, \quad i = 1, \ldots, m
\end{align*}$$

where $x$ is the optimization variable, $f_0(x)$ is the objective function that is optimized with respect to (in the example above minimized), $f_i(x)$ are constraint functions and $b_i$ are the limits for the constraints. Depending on the form of the objective and constraint functions, the optimization problems are classified into different categories. One class of optimization problems are convex optimization problems.

Convex optimization

Convex optimization problems are those where the objective and constraint functions are convex, i.e. satisfy

$$f_i(\alpha x + \beta y) \leq \alpha f_i(x) + \beta f_i(x)$$

for all $\alpha, \beta \in \mathbb{R}$ with $\alpha + \beta = 1$, $\alpha \geq 0$, $\beta \geq 0$. The convex formulation holds several known optimization classes as special cases, including least-squares problems and linear programming. Generally, optimization problems are difficult to solve, with some exceptions where effective algorithms and accessible software exist to solve large problems with many variables and constraints. Convex optimization problems are an example of a problem class where effective solution methods are available. CVX is a Matlab-based modeling framework, but with extensions to e.g. R, developed for convex programming [164, 165].
One solution method for convex optimization problems that are especially well suited for the large data sets being analyzed in fields such as computational biology, applied statistics and machine learning is the Alternating Direction of Method Multipliers (ADMM) method. ADMM is an iterative algorithm that solves convex optimization problems by decomposing them into smaller subproblems that can be solved more easily [166].
2. Present investigations

2.1 Paper I: Modeling glioblastoma heterogeneity as a dynamic network of cell states

![Figure 6. Graphical abstract of paper I.](image)

In paper I, the overall aim was to address the lack of consensus on the organization of cell state transitions in GBM. Previous publications had reported either an invariant hierarchy [74] or complete plasticity between stem-like states [79], with the therapeutic implications differing between models.

Here, we develop an experimental-computational method to observe and quantify the cell state transitions. By combining lineage tracing, scRNA-seq and mathematical modeling we follow cells over time and map the phenotypic states they adopt, their state switching behavior and the state-specific growth rates. The model we develop to build these transition networks are called STAG (State Transitions And Growth). In STAG, we model the rate of change of cells in state $i$ as
\[\frac{dX_i}{dt} = (\alpha_i - \beta_i)X_i - \sum_{k \neq i} \gamma_{kj}X_i + \sum_{j \neq i} \gamma_{ij}X_j\]

and expressed on a matrix format as

\[\frac{d}{dt}X(t) = AX(t)\]

The analytical solution for \(X(t)\) is given by the matrix exponential and to fit the model we solve a constrained convex optimization problem. The A-matrix will contain transition rates between states, as well as individual growth rates for each state defined as the difference between proliferation and death rates.

Our most important findings are that GBM cell states organize hierarchically, with a clear "source state" feeding cells downwards in the hierarchy towards a "sink state" with negative growth rate, but with patient-specific multidirectional transitions between intermediate states. We show that external perturbations disrupt the transition network, indicating that GBM cell states have a preferred organization but with a large potential of adaptive plasticity. Finally, we use our model and estimated transition network to mathematically predict the minimal intervention needed to reduce tumor growth. The resulting prediction is that a combination of TMZ and an intervention that block the transitions to state 5 would successfully reduce tumor growth.

The continued work on paper II-IV is based on the findings of paper I.
2.2 Paper II: Reconstructing the regulatory programs underlying the phenotypic plasticity of neural cancers

In paper II, we address the lack of computational methods to identify regulators of intratumoral heterogeneity. We develop an algorithm called \textit{scRegClust} that uses scRNA-seq data to estimate regulatory programs. A regulatory program, as defined in this paper, consists of a small set of regulators (transcription factors, kinases etc) and a set of target genes with common biological function (driving specific cell states).

We make an in-depth analysis of model properties and robustness through a simulation study, and demonstrate the applicability of \textit{scRegClust} through three use cases on real data. In use case 1, we benchmark our algorithm against the most comparable algorithm we can find, SCENIC+. Despite the fact that SCENIC+ uses a second data layer, scATAC-seq data, there is a good level of agreement between methods. In use case 2, we run our algorithm on the data generated in paper I to identify regulators of above mentioned state 5. We find that state 5 is positively regulated by several RTKs and that a combination of TMZ and dasatinib (TK inhibitor) have synergistic effect on GBM patient-derived cells. Finally, in use case 3, we collect publically available scRNA-seq data from GBM, MB, NB and healthy brain and make an
integrative study of the regulatory landscape of neuro-oncology. From the integrative analysis, certain meta-modules emerge consisting of several modules with similar gene content from different studies that are regulated by the same regulators. Among other things, we identify a meta-module representing the macrophage-induced mesenchymal transition in GBM and propose two regulators of this. The developed algorithm is distributed as an R-package and its use is not restricted to neuro-oncology but can be applied to any scRNA-seq data set.
2.3 Paper III: Estimating the differentiation potential and plasticity of cancer cells using statistical mechanics

In paper III, we explore the energy-concept as a way of measuring differentiation potential of single cells, instead of relying on gene markers or gene signatures of stemness. We adapt a model from statistical mechanics called the Ising model, originally used to model magnetism, and fit the model to scRNA-seq data. The Ising energy of the system is described by the Hamiltonian

\[ H(x) = -\sum J_{i,j} x_i x_j - \mu \sum h_j x_j \]

where \( x \in \{-1, 1\} \), is the binarized gene expression vector for each cell, \( J \) is a gene-gene interaction matrix and \( h \) is an external field vector.

To demonstrate it’s applicability we fit the model to synthetic and real data and show that the estimated Ising energy is a good measure of a cell’s differentiation potential, where high Ising energy indicate a high degree of stemness. Further, we use the estimated Ising energy to order cells according to pseudo-time and create a barcode-free version of the STAG model from paper I. Fitted to the generated scRNA-seq data in paper I, we recreate the transition matrix with high accuracy (\( r = 0.92, p < 0.001 \)). This development opens for a broader use of the STAG model, since the experimental design using barcodes are both costly and time-consuming. Finally, by exploring model properties, we find that the external field vector \( h \), when ordered according to chromosomal location, can be used to identify chromosomal aberrations in the individual cells. In the paper we demonstrate this by identifying the well-known chromosome 7 gain in GBM cells.

In conclusion, the Ising model offers a robust mathematical framework to identify CSC-like states from scRNA-seq data. This framework opens for the possibility of not only estimating the degree of stemness of cells, but also sim-
ulate how external perturbations affect the cell distribution and predict what interventions could force cells towards a more differentiated state.
2.4 Paper IV: Using drug-induced cell states to build therapeutic combinations against nervous system cancers.

In paper IV, we describe a data-driven strategy to identify drugs that modulate the intratumoral heterogeneity of nervous system cancers. First, building on the proof-of-concept study in [167] we present a scalable method for transcript profiling of 120 drugs in 10 patient-derived cell lines from GBM, MB, NB and healthy brain cells. The library of 120 drugs were selected either based on previous drug screens [168], computational predictions or drugs under current investigation for neuro-oncology use, found through literature studies. Following some initial analysis of the drug profiling data, we integrate it with scRNA-seq data from the three diseases using an integrative non-negative matrix factorization (iNMF) strategy to deconvolute how drugs induce shifts in cell states. The original iNMF model [169] formulates the optimization problem

\[
\min \sum_{k=1}^{K} ||Z^k - (B + B^k)V^k||^2 + \lambda \sum_k ||B^kV^k||^2
\]

and we extend this framework to separate the drug-induced effects on cell states into common, disease-specific and cell line-specific. Finally, the drug-induced cell states are used to predict efficient combination treatments based on anticorrelation between vectors of drug-induced effects. To make the generated drug profiling data easily accessible to the neuro-oncology research community, we provide it as a web tool.

Figure 9. Graphical abstract of paper IV.
3. Discussion and future perspectives

The high extent of intratumoral heterogeneity in nervous system cancer is a major caveat for the development of more efficient treatments. With a particular focus on glioblastoma, this thesis have taken a systems biology approach to understand how the heterogeneity is structured and how we can exploit that knowledge for therapeutic purposes. The central themes of the thesis are

1. The organization of GBM cell state transitions (paper I and III).
2. The regulation of cell states and cell state transitions (paper II).
3. Targeted interventions against cell states (paper II and IV).

Computationally, the contribution of the thesis is the development of tools for analyzing scRNA-seq data, where each paper presents either a new model, a new algorithm or a novel application of an existing model. In paper I, the STAG model is introduced and the organization of cell state transitions in three GBM cell lines are investigated, either in an unperturbed setting or during treatment. We find that the organization of the state transitions is patient-specific with varying degrees of hierarchy and with a high potential of adaptive plasticity upon external perturbations. The work also raised the question on how specific cell states could be targeted, which motivated the development of the scRegClust-algorithm in paper II. In paper III, the Ising model from statistical mechanics was fitted to scRNA-seq data to calculate the differentiation potential of individual cells, and a barcode-free version of the STAG model was developed. Finally, in paper IV the effect of 120 drugs on 10 cell lines are investigated and the iNMF framework is extended to map drug-induced effects on cell state composition. For paper III and IV, I discuss the direction of future work below.

Continued work on paper III
In paper III, Ising models are introduced as an alternative way of measuring differentiation potential in individual cells from scRNA-seq data, instead of methods such as SCENT and CytoTRACE. The performance of the Ising model in terms of separating un-differentiated from differentiated cells and resolving differentiation stages is in parity with mentioned entropy-based methods, but its big advantage is the robust mathematical framework underlying the estimated cell energies. An unanswered question is how we can use the Ising model to simulate the effect of targeted perturbations on the cells. This is being explored in on-going work. The Ising energy is defined by the Hamiltonian

\[ H = \sum_{i,j} J_{ij} S_i S_j - \sum_i h_i S_i \]
\[ H(x) = -\sum J_{i,j}x_i x_j - \mu \sum h_j x_j \]

We are proposing to use the external field vector \( h \) in the Hamiltonian to simulate external perturbations to the system, e.g. gene knockouts or drug perturbations, as outlined in the discussion in paper III. Remaining questions to be solved are how to speed up the simulation to efficiently investigate targeted perturbations in thousands of genes in tens of thousands of cells. Once in place, interesting new data sets are emerging that combine genetic perturbations with single cell-level read outs [170], which would be suitable to validate our approach on.

Another future direction of study is to continue the work with the barcode-free STAG version. STAG, as introduced in paper I, is an experimental-computational pipeline that due to the experimental design are quite time-consuming and costly. In paper III, we show that by ordering the cells according to Ising energy and use the ordered cell matrix as input to a Markov model, the state transitions can be reconstructed with high accuracy - the correlation between the STAG and the barcode-free STAG network for the U3065MG cell line is 0.92. This opens for the possibility of using the barcode-free STAG model to infer state transitions from a large number of GBM cell lines, since already generated scRNA-seq data from single time point measurements can be used. This would allow for a systematic analysis of the similarities and differences between patients’ transition networks that we’ve started to appreciate from the three profiled cell lines in paper I.

**Continued work in paper IV**

For paper IV, the continued work will revolve around a couple of important questions. First, the reproducibility of the drug profiling platform. In ongoing work we are investigating if a pre-profiling step could predict what drugs would yield a strong enough response to be reproducibly profiled using our technique. The pre-profiling step currently under investigation is to measure drug-induced morphological changes to the cells with Cell Painting [171]. Second, a systematic analysis of the stability of the mathematical model, e.g. when changing the number of components, perturbing the data and varying input parameters (such as \( \alpha \) and \( \lambda \)), is under way. Third, in the paper we propose to search for synergistic drug pairs among the pairs that have anti-correlated drug effect vectors, since it has been shown that drugs that target independent processes are more likely to show synergistic effects [172]. Experimental follow-up of the nominated drug pairs should be done to validate this approach.
Pooled CRISPR screening based on scRegClust predictions

Through the collected efforts from many researchers in the GBM field we’ve come a long way in understanding GBM intratumoral heterogeneity; the cell states, their functional implications and to some extent also their dynamic behaviour. An aspect that is less explored is the regulation of these cell states and if, and how, we can force certain states.

In the cell, TFs act as interpreters of the genome and by regulating what genes are transcribed and the amount of transcription, they control processes such as cell type specification [173]. Studies aimed at understanding the regulatory actions of the > 1800 TFs in the human genome have been constrained by technological limitations to focus on either a small set of TFs with a detailed readout, or a large set of TFs with a simpler readout, e.g. proliferation. However, technological advancements are now enabling more systematic approaches. Joung, et al. used a barcoded ORF library to overexpress each of the human TFs in human embryonic stem cells (hESCs) with a scRNA-seq read out [174]. The result of this effort is an atlas of TFs driving directed differentiation. In the GBM context, a similar type of study could map how TFs and other regulatory proteins control cell states, which would be a valuable resource.

In paper II, we present the scRegClust-algorithm, which takes as input scRNA-seq data and performs clustering and regulatory inference, generating a prediction of the landscape of cell state programs and their regulators present in the data. As an extension to this project, we are now planning to use the predictions made by scRegClust to guide the design of a pooled CRISPR library. By combining the pooled CRISPR screening with scRNA-seq, as has been successfully done by other groups [175, 176], we hope to be able to assess the effect of inactivating a certain TF or kinase on the transcriptional profile of single cells. In addition to serving as a validation of the scRegClust predictions, this effort will be a starting point to effectively identify the regulatory networks underlying GBM cell states and plasticity.

Conclusions

Glioblastoma remains one of the worst cancers a person can be diagnosed with and the more knowledge we gather about the underlying biology, the more apparent is the need for personalized treatment of GBM patients. In a single patient tumor, cells can adopt functionally distinct states that are to some extent genetically driven, but also influenced by other intrinsic and extrinsic factors, such as the tumor microenvironment. The most important result in this thesis is in my opinion the fact that even in an unperturbed 2D culture environment, cells switch in a non-trivial fashion between states, demonstrating a high degree of intrinsic plasticity. It’s not hard to imagine that this switching pattern is more complex in the real tumor, where external factors such as interaction with other cell types and variation in nutrient- and oxygen supplies are added to the mix. In the same way as the cell type composition differs between patients,
we can expect that the cell state transition networks will be patient-specific, as has been touched upon in this thesis based on the results from paper I. An attempt to account for cell state plasticity when designing radiation schedules showed improved survival in mouse studies [177] and the approach has now been moved to phase I study with positive results [178]. The progress and outcome of this study will be of high interest. Moving forward, accounting for cell state plasticity when searching for future treatments against GBM will probably be an important factor for success.
Den här avhandlingen har haft som syfte att undersöka heterogeniteten mellan och inom tumörer som uppstår i nervsystemet, framförallt hjärntumören glioblastom. Avhandlingen består av fyra delarbeten som fokuserar på olika aspekter av tumörheterogenitet.

Glioblastom är den vanligaste typen av elakartad hjärntumör bland vuxna. Det är en förhållandevis ovanlig cancerform att drabbas av, ca 3 av 100 000 personer blir diagnostiserade med glioblastom varje år i Sverige, men den har en mycket dålig prognos. Medelöverlevnaden från tidpunkten då en person diagnostiseras är bara 15 månader och endast 7% lever längre än 5 år. Den behandling som finns att tillgå består av operation där så mycket som möjligt av tumören tas bort, följt av strålning i kombination med cellgiftsbehandling. Den typ av cellgift som används idag heter temozolomid och förlänger medelöverlevnaden med ca 3 månader. Det finns flera anledningar till att glioblastom är så pass svårt att behandla, t.ex. så är tumörcellerna extremt duktiga på att infiltrera resten av hjärnan, vilket betyder att det är omöjligt att få bort hela tumören vid operation. En annan viktig anledning är tumörheterogeniteten.

Tumörheterogenitet kan antingen syfta till att tumörer skiljer sig åt mellan patienter, inter-patient heterogenitet, eller att tumörceller inom samma tumör skiljer sig från varandra, intra-tumör heterogenitet. Det är lätt att förstå varför tumörheterogenitet försvårar behandling. En behandling som fungerar bra för en patient kan ha mycket lägre effektivitet på en annan patients tumör, eller så är behandlingen väldigt effektiv mot en viss typ av tumörceller, medan andra tumörceller i samma tumör är resistenta. I fallet med glioblastom är både inter-patient och intra-tumör heterogeniteten utbredd, men det är framförallt intra-tumör heterogeniteten som behandlas i den här avhandlingen.

Tidigare arbeten om glioblastom har visat att tumörceller inom en tumör kan befinner sig i ett av fyra huvudsakliga tillstånd. Varje tillstånd har sin specifika molekylära profil och funktion, t.ex. föredrar celler i ett tillstånd att invadera omkringliggande vävnad genom att klättra längs med blodkärlen i hjärnan, medan celler i ett annat tillstånd främst invaderar i hjärnans vita substans. För att komplicerar bilden ännu mer så har man sett att tumörceller som befinner sig i ett tillstånd kan hoppa över till ett annat tillstånd, intra-tumör heterogeniteten är med andra ord inte statisk utan föränderlig över tid (plastisk).

I delarbete I vill vi undersöka fenomenet att tumörceller kan röra sig mellan olika tillstånd och framförallt vill vi bestämma om de rör sig enligt en förutbestämd hierarki eller om de kan röra sig fritt mellan tillstånd. För att göra
detta utvecklade vi en experimentell metod för att följa många tusen tumör-
celler och deras avkomma över tid, och en matematisk modell som kartlägger
hur tumörcellerna har rört sig över tid. Det vi såg var att det fanns en viss
hierarki, men att denna varierade från patient till patient och att den kunde
ändras vid yttre påverkan, t.ex. när vi tillsatte läkemedel. I delarbete II har vi
tagit fram en algoritm där man matar in den molekylära profilen för enskilda
celler och får ut en förutsägelse om hur de olika tillstånden regleras i cellen,
t.ex. vilka proteiner som styr övergången från ett tillstånd till ett annat. I delar-
bete III har vi anpassat en modell som ursprungligen används inom statistisk
mekanik för att modellera magnetism till att modellera energi-tillståndet hos
enskilda tumörceller. Energi-tillståndet hos en cell är ett bra mått på hur högt
upp i hierarkin den befinner sig. Slutligen, i delarbete IV har vi utvecklat en
metod för att undersöka hur läkemedel ändrar sammansättningen av tumör-
cellstillstånd i en tumör.

För att sammanfatta så har syftet med den här avhandlingen varit att utveckla
metoder för att undersöka tumörheterogenitet hos glioblastom. Varje delarbete
bidrar med en ny matematisk modell, algoritm eller ett nytt sätt att använda en
befintlig modell. Avhandlingen bidrar till en ökad förståelse för hur tumör-
cellers inneboende förmåga att ordna sig i och röra sig mellan tillstånd är or-
organiserad, samt hur det kan angripas i behandlingssyfte.
This work was conducted with support from the Swedish Foundation for Strategic Research, the Swedish Research Council, the Swedish Cancer Society and the Swedish Childhood Cancer Foundation.

There are so many people that in different ways have contributed to making this work possible, but most importantly, have made these last 4.5 years amazingly fun. This is my attempt at thanking all of you.

To my supervisor, Sven Nelander. Thank you for welcoming me to your group and for being an excellent mentor. Thank you for believing in me and giving me the freedom to develop my own ideas, attend all the conferences and courses I’ve been interested in and including me in other aspects of scientific life, such as grant application writing. I’ve truly enjoyed doing my PhD in your group and even though it will be scary to move away from here, I feel like working with you really prepared me for the next step.

To my co-supervisor Patrik Johansson, thank you for your guidance during my first year and for teaching me about everything from databases to where to eat lunch. To my co-supervisor Fredrik Swartling, thank you for providing expertise in the medulloblastoma field and always giving an encouraging word when we meet in the corridor. There are three additional persons that I consider my informal co-supervisors and want to give an extra warm thank you to. Rebecka Jörnsten, you are truly a role model in science for me, thank you for guiding me in the world of mathematical statistics and always being so enthusiastic about new ideas. Bengt Westermark, thank you for all inspiring conversations, especially during the work with paper I. Cecilia Krona, thank you for always being generous with your time and for planning group-activities with me.

To my faculty opponent, Dr Itay Tirosh, thank you for taking the time to read and scrutinize my thesis and coming to Uppsala for the defense day. I’m looking forward to meeting you!

Thank you to all co-authors, Ramy, Mia, Gergana, Anders, and especially my amazing co-first authors on paper I and II, Erika and Felix. If there is something I’ve learnt during these years it is to choose your co-first authors with care. Erika, I’m so proud of the barcoding-project. We made a great
team and I really hope we get to work together again. Also, thank you for always being 1.5 months ahead of me, it has made putting this thesis together a lot easier. And most importantly, thank you for being such a good friend. **Felix**, I’m very happy Rebecka asked you to work on paper II with me, I feel like we complemented each other in a nice way and I’ve really enjoyed getting to know you. Good luck finishing your thesis, soon it’s your turn!

To the incredible **Nelander lab**, where should I start? I’ve had so much fun with all of you. **Rebecka**, I can’t believe that we’ve known each other for almost 10 years already, from the first days at KTH until today, when I consider you one of my closest friends. You’re such a brilliant person - brave, caring and smart - and I feel so incredibly grateful to have had the privilege to share all the ups and downs of PhD life (and commuting...) with you. **Adam**, my best master thesis student. Hold on to your curiosity, your eagerness to learn and your hard work. I’m confident you will do great. **Irem**, I’m so happy you started in the group and that I got the chance to get to know you, you are amazing and you have such a bright future ahead of you. **Souni** and **Ludmila** thank you for being there from the start, always so fun, knowledgeable and helpful. **Milena**, thank you for being a great friend and for entertaining us with your amazing song-skills. **Hitesh**, you really take it to the next level with your amazing videos. **Faidra**, thank you for always keeping my blood sugar up with chocolate. **Alper, Lisa, Madeleine, Mar, Paula, Trupthi** - I feel safe about the future of science with students like you. And thank you to all former members of the group that I’ve had the privilege to work with. **Emil**, I really miss my roommate and workout buddy. **Elin**, it didn’t take long to realise that we were gonna be very good friends. Thank you for moving just a few blocks from where I live, and all your pep talks during the years. **Josi**, looking forward to having you in the group again!

To all of Neuro-oncology - I think it will be hard to find a better workplace than this one. Thank you for all lunches, all delicious fika, and all boosts over the years. **Ines**, I really admire your ambition and drive, but mostly, thank you for always making me laugh. **Ananya**, thank you for being an excellent roommate. **Anna**, we made a great team at Elin’s dissertation. I’m really looking forward to being colleagues soon again, thank you for getting me my next job!

References


[33] Wen, P. Y. *et al.* Glioblastoma in adults: a Society for Neuro-Oncology (SNO) and European Society of Neuro-Oncology (EANO) consensus review on


[90] Juraschka, K. & Taylor, M. D. Medulloblastoma in the age of molecular


[129] Trapnell, C. Defining cell types and states with single-cell genomics. Genome...


[148] Lathia, J. D., Mack, S. C., Mulkearns-Hubert, E. E., Valentim, C. L. & Rich,


A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine”.)