UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1920

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

IDA LARSSON

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2023

ISSN 1651-6206
ISBN 978-91-513-1753-3
URN urn:nbn:se:uu:diva-498239

Dissertation presented at Uppsala University to be publicly examined in Rudbecksalen, Rudbecklaboratoriet, Dag Hammarskjölds väg 20, Uppsala, Friday, 5 May 2023 at 13:00 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Dr Itay Tirosh (Weizmann Institute of Science; Department of Molecular Cell Biology).

Abstract

Larsson, I. 2023. *Integrative modeling of intratumoral heterogeneity, plasticity and regulation in nervous system cancers*. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 1920. 53 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-1753-3.

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.

ISSN 1651-6206
ISBN 978-91-513-1753-3
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.

I **Larsson I***, Dalmo E*, Elgendy R, Niklasson M, Doroszko M, Segerman A, Jörnsten R, Westermark B, Nelander S. *Modeling glioblastoma heterogeneity as a dynamic network of cell states.* Molecular Systems Biology. 2021 17:e10105

II **Larsson I***, Held F*, Popova G, Koc A, Jörnsten R, Nelander S. *Reconstructing the regulatory programs underlying the phenotypic plasticity of neural cancers.* Preprint available on bioRxiv 2023.03.10.532041.

III Lång A, **Larsson I**, Skeppås M, Jörnsten R, Nelander S. *Estimating the differentiation potential and plasticity of cancer cells using statistical mechanics.* Manuscript

IV **Larsson I**, Popova G, Elgendy R, Sundström A, Krona C, Jörnsten R, Nelander S. *Using drug-induced cell states to build therapeutic combinations against nervous system cancers.* Manuscript

*indicates equal contribution.

Related work by the author

1. Almstedt E, Elgendy R, Hekmati E, Rosén E, Wärn C, Olsen T.K, Dyberg C, Doroszko M, **Larsson I**, Sundström A, Påhlman S, Bexell D, Vanlandewijck M, Kogner P, Jörnsten R, Krona C, Nelander S. *Integrative discovery of treatments for high-risk neuroblastoma.* Nature Communications 11.1 (Jan. 2020), p. 71

2. **Larsson I**, Uhlén M, Zhang C, Mardinoglu A. *Genome-Scale Metabolic Modeling of Glioblastoma Reveals Promising Targets for Drug Development.* Frontiers in Genetics 11 (Apr. 2020), p. 381

Contents

Abbreviations

AC	Astrocyte
ADMM	Alternating direction method of multipliers
ADRN	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

SHH	Sonic hedgehog
scRNA-seq	Single-cell RNA sequencing
SNP	Single nucleotide polymorphism
SOX2	SRY-box transcription factor 2
STAG	State Transition and Growth
STAT1	Signal transducer and activator of transcription 1
TCGA	The Cancer Genome Atlas
TERT	Telomerase reverse transcriptase
TF	Transcription factor
TME	Tumor microenvironment
TMZ	Temozolomide
TP53	Tumor protein P53
TTF	Tumor Treating Fields
UMI	Unique molecular identifier
WHO	World health organization

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 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

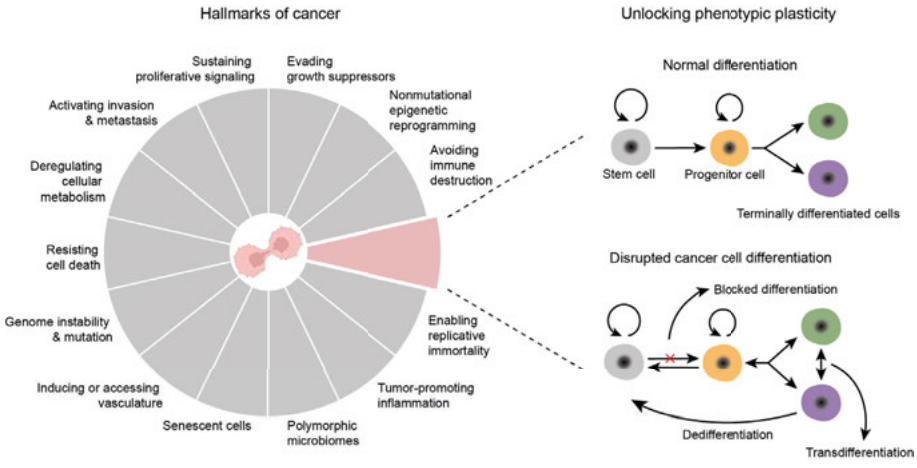

Figure 1. The hallmarks of cancer (left) and the emerging trait unlocking phenotypic plasticity (right). Adapted from [5].

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 *IDH* mutational status (wildtype) and the presence of any of the following features: microvascular proliferation, necrosis, *TERT* promoter mutation, *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^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

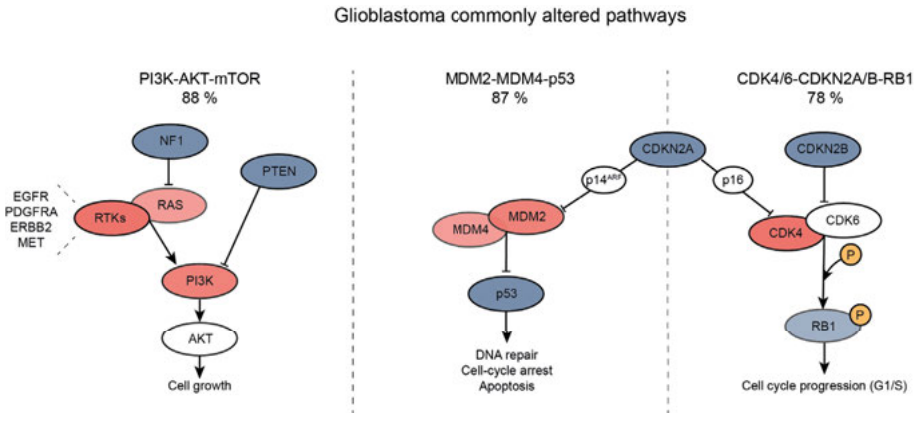

Figure 2. The main altered pathways in GBM. Red indicate the occurrence of activating mutations, blue indicate the occurrence of inactivating mutations. Adapted from [37].

analysis of copy number, gene expression and DNA methylation of 206 patient samples [37]. The most frequently altered genes coincide in three main pathways; PI3K-AKT-mTOR, p53 and RB1 signaling (Figure 2) [16]. Focal amplification and/or somatic mutation of the receptor tyrosine kinase (RTK) *EGFR* is found in 45 % of GBM patients and together with amplification of other RTKs, such as *PDGFRA*, and other gene alterations, this leads to activation of the PI3K-AKT-mTOR pathway and increased cell growth. The tumor suppressor gene *TP53* is mutated in 35 % of GBM patients, but as many as 87 % of GBM patients harbor mutations in any of the signaling components of the p53 pathway. *TP53* is sometimes called the "guardian of the genome", reflecting it's important role in preventing tumorigenesis by regulating the cell cycle, activating DNA repair mechanisms and initiating apoptosis [43, 44]. Mutation of *TP53* is common across different cancer forms and leads to inactivation [45]. *RB1* is a tumor-suppressor gene with a role in regulating a cell's progression through the cell cycle, more specifically by blocking the transition from G1 to S. Once the cell is ready to divide, *RB1* is inactivated through phosphorylation [46]. Alterations in the RB pathway occurs in approximately 78 % of GBM patients and typically leads to increased phosphorylation of *RB1* and increased cell growth.

The molecular profiles of the tumors are used to stratify patients into subtypes. The number and exakt 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.

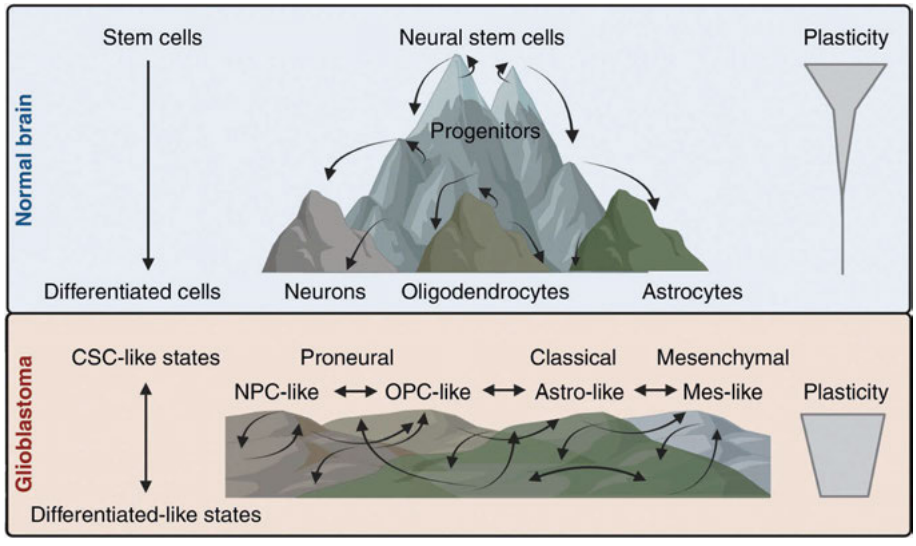

Figure 3. Landscape of phenotypic plasticity. Reprinted with permission [15].

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 heterogeneic 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,

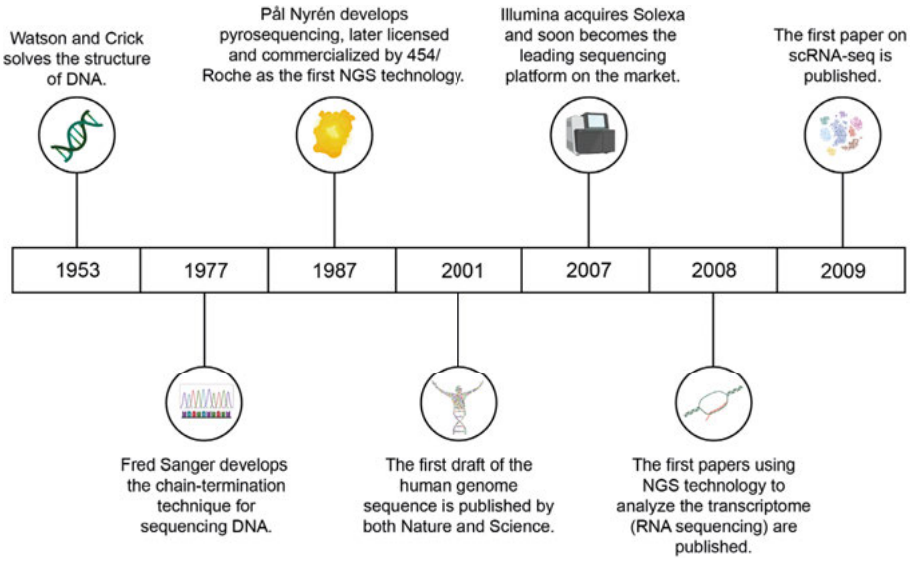

Figure 4. The history of DNA sequencing at a glance, adapted from [114].

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

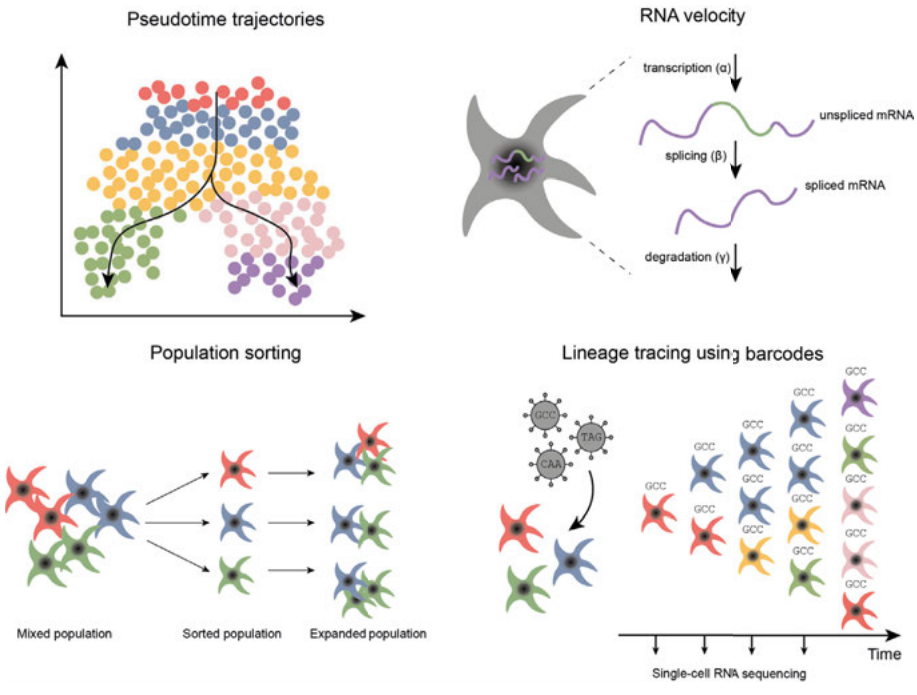

Figure 5. Different approaches to studying phenotypic plasticity.

cells to get at cell fate combined with other measurements to infer cell state, e.g. morphology or gene expression. The most high-throughput approach to lineage tracing is based on DNA sequence barcodes that are integrated in the genome of single cells and identified using sequencing. The descendants of the single cells inherit the barcode and clonal relationships between cells can be inferred [144]. Sampling cells at more than one time point allow for tracking of cell trajectories, as is done using CellTagging to reconstruct the trajectories of fibroblasts being reprogrammed to endoderm progenitors [134] or LARRY to study cell fate decisions in hematopoiesis [145].

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 straight-forward 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

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

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 α, $\beta \in R$ 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

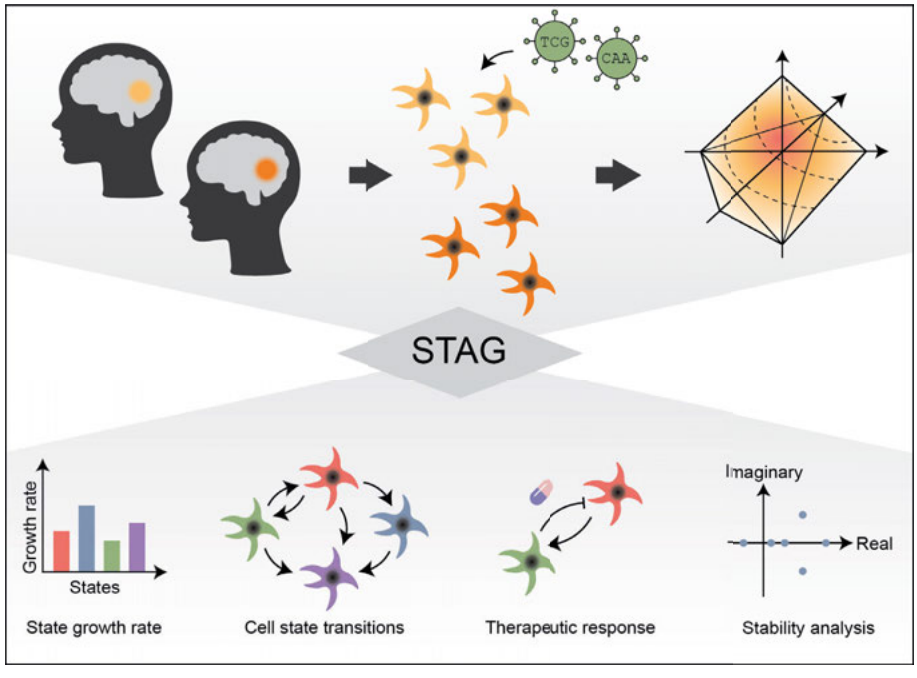

Figure 6. Graphical abstract of paper I.

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

$$dX_i/dt = (\alpha_i - \beta_i)X_i - \sum_{k \neq i} \gamma_{kj} X_i + \sum_{j \neq i} \gamma_{ij} X_j$$

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

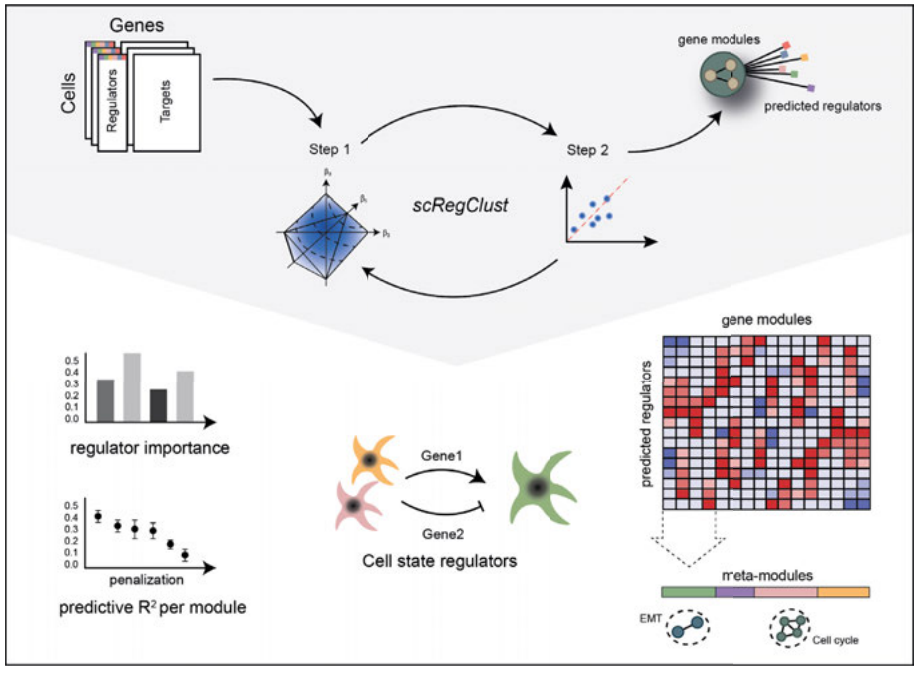

Figure 7. Graphical abstract of paper II.

In paper II, we address the lack of computational methods to identify regulators of intratumoral heterogeneity. We develop an algorithm called *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 *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

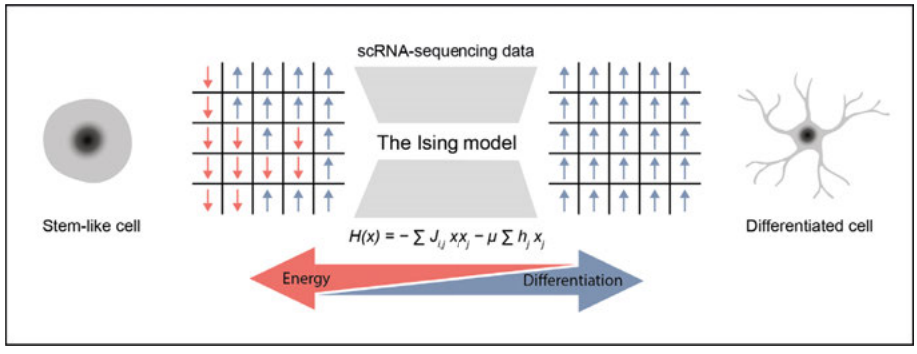

Figure 8. Graphical abstract of paper III.

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 pseudotime 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.

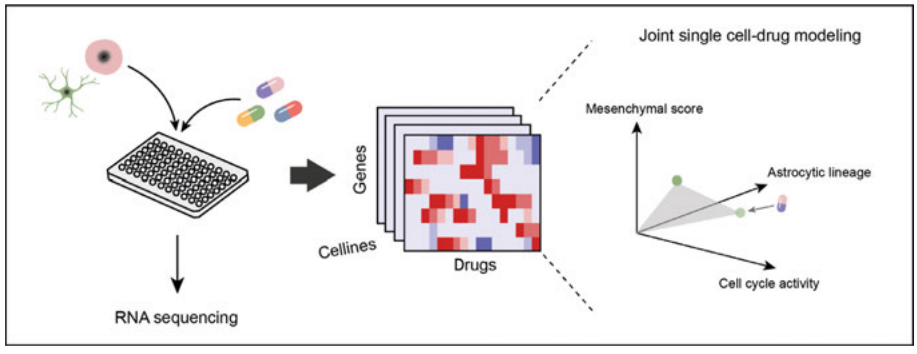

Figure 9. Graphical abstract of paper IV.

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

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

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.

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

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

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 α and λ), 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.

4. Populärvetenskaplig sammanfattning

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 befinna 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 komplicera 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 tumörceller 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 delarbete 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 tumörcellstillstånd 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 tumörcellers inneboende förmåga att ordna sig i och röra sig mellan tillstånd är organiserad, samt hur det kan angripas i behandlingssyfte.

5. Acknowledgements

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. **Soumi** 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!

Tack till min familj och alla vänner utanför jobbet, för all uppmuntran och allt stöd genom åren, ni är alla otroliga. **Mamma** och **pappa,** tack för att ni är, och alltid har varit, min trygga punkt. **Hanna**, du är den bästa syster man kan ha. Tack för att jag får ha ett andra hem hos dig i Uppsala och tack för att du planerat min fest. **Pelle**, vilken lyx det är att ha sin morbror precis runt hörnet, tack för alla kvällsfikor och pratstunder. Och **Filip**, tänk att jag hade turen att träffa dig. Tack för att du gör alla dagar bättre.

References

[1] Ferlay, J. *et al.* Global cancer observatory: Cancer today. Available at `https://gco.iarc.fr/today/home` (2023/01/04).

[2] What is cancer? (2022). URL `https://www.cancerresearchuk.org/about-cancer/what-is-cancer`.

[3] Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).

[4] Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).

[5] Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discov* **12**, 31–46 (2022).

[6] Witsch, E., Sela, M. & Yarden, Y. Roles for growth factors in cancer progression. *Physiology (Bethesda)* **25**, 85–101 (2010).

[7] Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141**, 1117–1134 (2010).

[8] Blasco, M. A. Telomeres and human disease: ageing, cancer and beyond. *Nat Rev Genet* **6**, 611–622 (2005).

[9] Killela, P. J. *et al.* TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. *Proc Natl Acad Sci U S A* **110**, 6021–6026 (2013).

[10] Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).

[11] Köhler, C. *et al.* Mouse cutaneous melanoma induced by mutant BRaf arises from expansion and dedifferentiation of mature pigmented melanocytes. *Cell Stem Cell* **21**, 679–693.e6 (2017).

[12] Boumahdi, S. & de Sauvage, F. J. The great escape: tumour cell plasticity in resistance to targeted therapy. *Nat Rev Drug Discov* **19**, 39–56 (2020).

[13] Von Hoff, D. D. *et al.* Inhibition of the hedgehog pathway in advanced basal-cell carcinoma. *N Engl J Med* **361**, 1164–1172 (2009).

[14] Biehs, B. *et al.* A cell identity switch allows residual BCC to survive Hedgehog pathway inhibition. *Nature* **562**, 429–433 (2018).

[15] Yabo, Y. A., Niclou, S. P. & Golebiewska, A. Cancer cell heterogeneity and plasticity: A paradigm shift in glioblastoma. *Neuro Oncol* **24**, 669–682 (2022).

[16] Osborn, A. G., Louis, D. N., Poussaint, T. Y., Linscott, L. L. & Salzman, K. L. The 2021 World Health Organization Classification of Tumors of the Central Nervous System: What Neuroradiologists Need to Know. *AJNR Am J Neuroradiol* **43**, 928–937 (2022).

[17] Ostrom, Q. T. *et al.* CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2015-2019. *Neuro Oncol* **24**, v1–v95 (2022).

[18] Schwartzbaum, J. A., Fisher, J. L., Aldape, K. D. & Wrensch, M. Epidemiology and molecular pathology of glioma. *Nat Clin Pract Neurol* **2**, 494–503 (2006).
[19] Philips, A., Henshaw, D. L., Lamburn, G. & O'Carroll, M. J. Brain Tumours: Rise in Glioblastoma Multiforme Incidence in England 1995-2015 Suggests an Adverse Environmental or Lifestyle Factor. *J Environ Public Health* **2018**, 7910754 (2018).
[20] Ostrom, Q. T. *et al.* CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2011-2015. *Neuro Oncol* **20**, iv1–iv86 (2018).
[21] Ostrom, Q. T. *et al.* The epidemiology of glioma in adults: a "state of the science" review. *Neuro Oncol* **16**, 896–913 (2014).
[22] Hardell, L. & Carlberg, M. Mobile phone and cordless phone use and the risk for glioma - Analysis of pooled case-control studies in Sweden, 1997-2003 and 2007-2009. *Pathophysiology* **22**, 1–13 (2015).
[23] Frei, P. *et al.* Use of mobile phones and risk of brain tumours: update of Danish cohort study. *BMJ* **343**, d6387 (2011).
[24] Cardis, E. *et al.* Brain tumour risk in relation to mobile telephone use: results of the INTERPHONE international case-control study. *Int J Epidemiol* **39**, 675–694 (2010).
[25] Wen, P. Y. *et al.* Glioblastoma in adults: a Society for Neuro-Oncology (SNO) and European Society of Neuro-Oncology (EANO) consensus review on current management and future directions. *Neuro Oncol* **22**, 1073–1113 (2020).
[26] Stupp, R. *et al.* Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med* **352**, 987–996 (2005).
[27] Dunn, J. *et al.* Extent of MGMT promoter methylation correlates with outcome in glioblastomas given temozolomide and radiotherapy. *Br J Cancer* **101**, 124–131 (2009).
[28] Hegi, M. E. *et al.* Clinical trial substantiates the predictive value of O-6-methylguanine-DNA methyltransferase promoter methylation in glioblastoma patients treated with temozolomide. *Clin Cancer Res* **10**, 1871–1874 (2004).
[29] Stupp, R. *et al.* Effect of Tumor-Treating Fields Plus Maintenance Temozolomide vs Maintenance Temozolomide Alone on Survival in Patients With Glioblastoma: A Randomized Clinical Trial. *JAMA* **318**, 2306–2316 (2017).
[30] Rominiyi, O. *et al.* Tumour treating fields therapy for glioblastoma: current advances and future directions. *Br J Cancer* **124**, 697–709 (2021).
[31] Liau, L. M. *et al.* Association of Autologous Tumor Lysate-Loaded Dendritic Cell Vaccination With Extension of Survival Among Patients With Newly Diagnosed and Recurrent Glioblastoma: A Phase 3 Prospective Externally Controlled Cohort Trial. *JAMA Oncol* **9**, 112–121 (2023).
[32] Gilbert, M. R. *et al.* A randomized trial of bevacizumab for newly diagnosed glioblastoma. *N Engl J Med* **370**, 699–708 (2014).
[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

current management and future directions. *Neuro Oncol* **22**, 1073–1113 (2020).
[34] Aldape, K. *et al.* Challenges to curing primary brain tumours. *Nat Rev Clin Oncol* **16**, 509–520 (2019).
[35] Frederico, S. C. *et al.* Making a cold tumor hot: The role of vaccines in the treatment of glioblastoma. *Front. Oncol.* **11**, 672508 (2021).
[36] Cuddapah, V. A., Robel, S., Watkins, S. & Sontheimer, H. A neurocentric perspective on glioma invasion. *Nat Rev Neurosci* **15**, 455–465 (2014).
[37] Network., C. G. A. R. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
[38] Brennan, C. W. *et al.* The somatic genomic landscape of glioblastoma. *Cell* **155**, 462–477 (2013).
[39] Phillips, H. S. *et al.* Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* **9**, 157–173 (2006).
[40] Verhaak, R. G. *et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17**, 98–110 (2010).
[41] Noushmehr, H. *et al.* Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* **17**, 510–522 (2010).
[42] Beroukhim, R. *et al.* Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc Natl Acad Sci U S A* **104**, 20007–20012 (2007).
[43] Lane, D. P. Cancer. p53, guardian of the genome. *Nature* **358**, 15–16 (1992).
[44] Toufektchan, E. & Toledo, F. The Guardian of the Genome Revisited: p53 Downregulates Genes Required for Telomere Maintenance, DNA Repair, and Centromere Structure. *Cancers (Basel)* **10** (2018).
[45] Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).
[46] Murphree, A. L. & Benedict, W. F. Retinoblastoma: clues to human oncogenesis. *Science* **223**, 1028–1033 (1984).
[47] Wang, Q. *et al.* Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. *Cancer Cell* **32**, 42–56 (2017).
[48] Li, Y. M., Suki, D., Hess, K. & Sawaya, R. The influence of maximum safe resection of glioblastoma on survival in 1229 patients: Can we do better than gross-total resection? *J Neurosurg* **124**, 977–988 (2016).
[49] Scherer, H. J. Structural development in gliomas. *The American Journal of Cancer* **34**, 333–351 (1938). URL `https://cancerres.aacrjournals.org/content/34/3/333`. `https://cancerres.aacrjournals.org/content/34/3/333.full.pdf`.
[50] Wenger, A. *et al.* Intratumor DNA methylation heterogeneity in glioblastoma: implications for DNA methylation-based classification. *Neuro Oncol* **21**, 616–627 (2019).
[51] Sottoriva, A. *et al.* Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc Natl Acad Sci U S A* **110**, 4009–4014 (2013).

[52] Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
[53] Neftel, C. *et al.* An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* **178**, 835–849 (2019).
[54] Garofano, L. *et al.* Pathway-based classification of glioblastoma uncovers a mitochondrial subtype with therapeutic vulnerabilities. *Nat Cancer* **2**, 141–156 (2021).
[55] Richards, L. M. *et al.* Gradient of Developmental and Injury Response transcriptional states defines functional vulnerabilities underpinning glioblastoma heterogeneity. *Nat Cancer* **2**, 157–173 (2021).
[56] Castellan, M. *et al.* Single-cell analyses reveal YAP/TAZ as regulators of stemness and cell plasticity in Glioblastoma. *Nat Cancer* **2**, 174–188 (2021).
[57] Couturier, C. P. *et al.* Single-cell RNA-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy. *Nat Commun* **11**, 3406 (2020).
[58] Darmanis, S. *et al.* Single-Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the Migrating Front of Human Glioblastoma. *Cell Rep* **21**, 1399–1410 (2017).
[59] Nowakowski, T. J. *et al.* Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318–1323 (2017).
[60] Venkataramani, V. *et al.* Glioblastoma hijacks neuronal mechanisms for brain invasion. *Cell* **185**, 2899–2917 (2022).
[61] Griveau, A. *et al.* A Glial Signature and Wnt7 Signaling Regulate Glioma-Vascular Interactions and Tumor Microenvironment. *Cancer Cell* **33**, 874–889 (2018).
[62] Seano, G. & Jain, R. K. Vessel co-option in glioblastoma: emerging insights and opportunities. *Angiogenesis* **23**, 9–16 (2020).
[63] Zhong, S. *et al.* A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524–528 (2018).
[64] Pollen, A. A. *et al.* Molecular identity of human outer radial glia during cortical development. *Cell* **163**, 55–67 (2015).
[65] Bhat, K. P. L. *et al.* Mesenchymal differentiation mediated by NF-κB promotes radiation resistance in glioblastoma. *Cancer Cell* **24**, 331–346 (2013).
[66] Hara, T. *et al.* Interactions between cancer cells and immune cells drive transitions to mesenchymal-like states in glioblastoma. *Cancer Cell* **39**, 779–792 (2021).
[67] Chanoch-Myers, R., Wider, A., Suva, M. L. & Tirosh, I. Elucidating the diversity of malignant mesenchymal states in glioblastoma by integrative analysis. *Genome Med* **14**, 106 (2022).
[68] Gangoso, E. *et al.* Glioblastomas acquire myeloid-affiliated transcriptional programs via epigenetic immunoediting to elicit immune evasion. *Cell* **184**, 2454–2470 (2021).
[69] Segerman, A. *et al.* Clonal Variation in Drug and Radiation Response among Glioma-Initiating Cells Is Linked to Proneural-Mesenchymal Transition. *Cell Rep* **17**, 2994–3009 (2016).

[70] Wang, L. *et al.* A single-cell atlas of glioblastoma evolution under therapy reveals cell-intrinsic and cell-extrinsic therapeutic targets. *Nat Cancer* **3**, 1534–1552 (2022).
[71] Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105â111 (2001).
[72] Dingli, D. & Michor, F. Successful therapy must eradicate cancer stem cells. *Stem Cells* **24**, 2603–2610 (2006).
[73] Singh, S. K. *et al.* Identification of human brain tumour initiating cells. *Nature* **432**, 396–401 (2004).
[74] Lan, X. *et al.* Fate mapping of human glioblastoma reveals an invariant stem cell hierarchy. *Nature* **549**, 227–232 (2017).
[75] Batlle, E. & Clevers, H. Cancer stem cells revisited. *Nat Med* **23**, 1124–1134 (2017).
[76] Beier, D. *et al.* CD133(+) and CD133(-) glioblastoma-derived cancer stem cells show differential growth characteristics and molecular profiles. *Cancer Res* **67**, 4010–4015 (2007).
[77] Wang, J. *et al.* CD133 negative glioma cells form tumors in nude rats and give rise to CD133 positive cells. *Int J Cancer* **122**, 761–768 (2008).
[78] Ogden, A. T. *et al.* Identification of A2B5+CD133- tumor-initiating cells in adult human gliomas. *Neurosurgery* **62**, 505–514 (2008).
[79] Dirkse, A. *et al.* Stem cell-associated heterogeneity in Glioblastoma results from intrinsic tumor plasticity shaped by the microenvironment. *Nat Commun* **10**, 1787 (2019).
[80] Gröbner, S. N. *et al.* The landscape of genomic alterations across childhood cancers. *Nature* **555**, 321–327 (2018).
[81] Gustafsson, G., Kogner, P. & Heyman, M. Childhood cancer incidence and survival in sweden 1984â2010. report 2013 from the swedish childhood cancer registry (2013). `http://www.forskasverige.se/wp-content/uploads/ChildhoodCancerIncidenceandSurvivalinSweden1984_2010.pdf`.
[82] Saletta, F., Seng, M. S. & Lau, L. M. Advances in paediatric cancer treatment. *Transl Pediatr* **3**, 156–182 (2014).
[83] Bhakta, N. *et al.* Childhood cancer burden: a review of global estimates. *Lancet Oncol* **20**, e42–e53 (2019).
[84] Hovestadt, V. *et al.* Medulloblastomics revisited: biological and clinical insights from thousands of patients. *Nat Rev Cancer* **20**, 42–56 (2020).
[85] Almstedt, E. *et al.* Integrative discovery of treatments for high-risk neuroblastoma. *Nat Commun* **11**, 71 (2020).
[86] Curtin, S. C., Minino, A. M. & Anderson, R. N. Declines in Cancer Death Rates Among Children and Adolescents in the United States, 1999-2014. *NCHS Data Brief* 1–8 (2016).
[87] Northcott, P. A. *et al.* The whole-genome landscape of medulloblastoma subtypes. *Nature* **547**, 311–317 (2017).
[88] Northcott, P. A. *et al.* Medulloblastoma comprises four distinct molecular variants. *J Clin Oncol* **29**, 1408–1414 (2011).
[89] Taylor, M. D. *et al.* Molecular subgroups of medulloblastoma: the current consensus. *Acta Neuropathol* **123**, 465–472 (2012).
[90] Juraschka, K. & Taylor, M. D. Medulloblastoma in the age of molecular

subgroups: a review. *J Neurosurg Pediatr* **24**, 353–363 (2019).
[91] Hovestadt, V. *et al.* Resolving medulloblastoma cellular architecture by single-cell genomics. *Nature* **572**, 74–79 (2019).
[92] Vladoiu, M. C. *et al.* Childhood cerebellar tumours mirror conserved fetal transcriptional programs. *Nature* **572**, 67–73 (2019).
[93] Llombart, V. & Mansour, M. R. Therapeutic targeting of "undruggable" MYC. *EBioMedicine* **75**, 103756 (2022).
[94] Maris, J. M., Hogarty, M. D., Bagatell, R. & Cohn, S. L. Neuroblastoma. *Lancet* **369**, 2106–2120 (2007).
[95] van Groningen, T. *et al.* Neuroblastoma is composed of two super-enhancer-associated differentiation states. *Nat Genet* **49**, 1261–1266 (2017).
[96] van Groningen, T. *et al.* A NOTCH feed-forward loop drives reprogramming from adrenergic to mesenchymal state in neuroblastoma. *Nat Commun* **10**, 1530 (2019).
[97] Kildisiute, G. *et al.* Tumor to normal single-cell mRNA comparisons reveal a pan-neuroblastoma cancer cell. *Sci Adv* **7** (2021).
[98] Bedoya-Reina, O. C. *et al.* Single-nuclei transcriptomes from human adrenal gland reveal distinct cellular identities of low and high-risk neuroblastoma tumors. *Nat Commun* **12**, 5309 (2021).
[99] Jacob, F. *et al.* A Patient-Derived Glioblastoma Organoid Model and Biobank Recapitulates Inter- and Intra-tumoral Heterogeneity. *Cell* **180**, 188–204 (2020).
[100] Ponten, J. & Macintyre, E. H. Long term culture of normal and neoplastic human glia. *Acta Pathol Microbiol Scand* **74**, 465–486 (1968).
[101] Lee, J. *et al.* Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines. *Cancer Cell* **9**, 391–403 (2006).
[102] Xie, Y. *et al.* The Human Glioblastoma Cell Culture Resource: Validated Cell Models Representing All Molecular Subtypes. *EBioMedicine* **2**, 1351–1363 (2015).
[103] Pollard, S. M. *et al.* Glioma stem cell lines expanded in adherent culture have tumor-specific phenotypes and are suitable for chemical and genetic screens. *Cell Stem Cell* **4**, 568–580 (2009).
[104] Uhrbom, L., Hesselager, G., r, M. & Westermark, B. Induction of brain tumors in mice using a recombinant platelet-derived growth factor B-chain retrovirus. *Cancer Res* **58**, 5275–5279 (1998).
[105] Holland, E. C. *et al.* Combined activation of Ras and Akt in neural progenitors induces glioblastoma formation in mice. *Nat Genet* **25**, 55–57 (2000).
[106] Weiss, W. A., Aldape, K., Mohapatra, G., Feuerstein, B. G. & Bishop, J. M. Targeted expression of MYCN causes neuroblastoma in transgenic mice. *EMBO J* **16**, 2985–2995 (1997).
[107] Swartling, F. J. *et al.* Pleiotropic role for MYCN in medulloblastoma. *Genes Dev* **24**, 1059–1072 (2010).
[108] Letrado, P., de Miguel, I., Lamberto, I., nez, R. & Oyarzabal, J. Zebrafish: Speeding Up the Cancer Drug Discovery Process. *Cancer Res* **78**, 6048–6058 (2018).

[109] Almstedt, E. *et al.* Real-time evaluation of glioblastoma growth in patient-specific zebrafish xenografts. *Neuro Oncol* **24**, 726–738 (2022).
[110] Heather, J. M. & Chain, B. The sequence of sequencers: The history of sequencing DNA. *Genomics* **107**, 1–8 (2016).
[111] Stark, R., Grzelak, M. & Hadfield, J. RNA sequencing: the teenage years. *Nat Rev Genet* **20**, 631–656 (2019).
[112] Tanay, A. & Regev, A. Scaling single-cell genomics from phenomenology to mechanism. *Nature* **541**, 331–338 (2017).
[113] Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* **6**, 377–382 (2009).
[114] Wu, X. *et al.* Research techniques made simple: Single-cell RNA sequencing and its applications in dermatology. *J. Invest. Dermatol.* **138**, 1004–1009 (2018).
[115] Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. Cel-seq: Single-cell rna-seq by multiplexed linear amplification. *Cell Reports* **2**, 666–673 (2012). URL `https://www.sciencedirect.com/science/article/pii/S2211124712002288`.
[116] Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
[117] Picelli, S. *et al.* Full-length RNA-seq from single cells using smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
[118] Ilicic, T. *et al.* Classification of low quality cells from single-cell RNA-seq data. *Genome Biol* **17**, 29 (2016).
[119] Galow, A. M. *et al.* Quality control in scRNA-Seq can discriminate pacemaker cells: the mtRNA bias. *Cell Mol Life Sci* **78**, 6585–6592 (2021).
[120] Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol* **15**, e8746 (2019).
[121] Stegle, O., Teichmann, S. A. & Marioni, J. C. Computational and analytical challenges in single-cell transcriptomics. *Nat Rev Genet* **16**, 133–145 (2015).
[122] Baran-Gale, J., Chandra, T. & Kirschner, K. Experimental design for single-cell RNA sequencing. *Brief Funct Genomics* **17**, 233–239 (2018).
[123] Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* **20**, 296 (2019).
[124] Tirosh, I. *et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* **539**, 309–313 (2016).
[125] Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411–420 (2018).
[126] Buttner, M., Miao, Z., Wolf, F. A., Teichmann, S. A. & Theis, F. J. A test metric for assessing single-cell RNA-seq batch correction. *Nat Methods* **16**, 43–49 (2019).
[127] van Dijk, D. *et al.* Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell* **174**, 716–729 (2018).
[128] Li, W. V. & Li, J. J. An accurate and robust imputation method scImpute for single-cell RNA-seq data. *Nat Commun* **9**, 997 (2018).
[129] Trapnell, C. Defining cell types and states with single-cell genomics. *Genome*

Res **25**, 1491–1498 (2015).
[130] Weinreb, C., Wolock, S., Tusi, B. K., Socolovsky, M. & Klein, A. M. Fundamental limits on dynamic inference from single-cell snapshots. *Proc Natl Acad Sci U S A* **115**, E2467–E2476 (2018).
[131] La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
[132] Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381–386 (2014).
[133] Gupta, P. B. *et al.* Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* **146**, 633–644 (2011).
[134] Biddy, B. A. *et al.* Single-cell mapping of lineage and identity in direct reprogramming. *Nature* **564**, 219–224 (2018).
[135] Teschendorff, A. E. & Feinberg, A. P. Statistical mechanics meets single-cell biology. *Nat Rev Genet* **22**, 459–476 (2021).
[136] Coifman, R. R. *et al.* Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps. *Proc Natl Acad Sci U S A* **102**, 7426–7431 (2005).
[137] Setty, M. *et al.* Characterization of cell fate probabilities in single-cell data with Palantir. *Nat Biotechnol* **37**, 451–460 (2019).
[138] Chen, H. *et al.* Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. *Nat Commun* **10**, 1903 (2019).
[139] Saelens, W., Cannoodt, R., Todorov, H. & Saeys, Y. A comparison of single-cell trajectory inference methods. *Nat Biotechnol* **37**, 547–554 (2019).
[140] Gorban, A. N. & Zinovyev, A. Principal manifolds and graphs in practice: from molecular biology to dynamical systems. *Int J Neural Syst* **20**, 219–232 (2010).
[141] Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol* **38**, 1408–1414 (2020).
[142] Wang, L. *et al.* The Phenotypes of Proliferating Glioblastoma Cells Reside on a Single Axis of Variation. *Cancer Discov* **9**, 1708–1719 (2019).
[143] Couturier, C. P. *et al.* Single-cell RNA-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy. *Nat Commun* **11**, 3406 (2020).
[144] Wagner, D. E. & Klein, A. M. Lineage tracing meets single-cell omics: opportunities and challenges. *Nat Rev Genet* **21**, 410–427 (2020).
[145] Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F. D. & Klein, A. M. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science* **367** (2020).
[146] Nishikawa, M. *et al.* Significance of Glioma Stem-Like Cells in the Tumor Periphery That Express High Levels of CD44 in Tumor Invasion, Early Progression, and Poor Prognosis in Glioblastoma. *Stem Cells Int* **2018**, 5387041 (2018).
[147] Gangemi, R. M. *et al.* SOX2 silencing in glioblastoma tumor-initiating cells causes stop of proliferation and loss of tumorigenicity. *Stem Cells* **27**, 40–48 (2009).
[148] Lathia, J. D., Mack, S. C., Mulkearns-Hubert, E. E., Valentim, C. L. & Rich,

J. N. Cancer stem cells in glioblastoma. *Genes Dev* **29**, 1203–1217 (2015).
[149] Teschendorff, A. E. & Enver, T. Single-cell entropy for accurate estimation of differentiation potency from a cell's transcriptome. *Nat Commun* **8**, 15599 (2017).
[150] Gulati, G. S. *et al.* Single-cell transcriptional diversity is a hallmark of developmental potential. *Science* **367**, 405–411 (2020).
[151] Grun, D. *et al.* De Novo Prediction of Stem Cell Identity using Single-Cell Transcriptome Data. *Cell Stem Cell* **19**, 266–277 (2016).
[152] Chen, W. *et al.* Single-cell landscape in mammary epithelium reveals bipotent-like cells associated with breast cancer risk and outcome. *Commun Biol* **2**, 306 (2019).
[153] Kutalik, Z., Beckmann, J. S. & Bergmann, S. A modular approach for integrative analysis of large-scale gene-expression and drug-response data. *Nat Biotechnol* **26**, 531–539 (2008).
[154] Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
[155] Yang, W. *et al.* Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* **41**, D955–961 (2013).
[156] Basu, A. *et al.* An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell* **154**, 1151–1161 (2013).
[157] Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437–1452 (2017).
[158] McFarland, J. M. *et al.* Multiplexed single-cell transcriptional response profiling to define cancer vulnerabilities and therapeutic mechanism of action. *Nat Commun* **11**, 4296 (2020).
[159] Avila Cobos, F., Alquicira-Hernandez, J., Powell, J. E., Mestdagh, P. & De Preter, K. Benchmarking of cell type deconvolution pipelines for transcriptomics data. *Nat Commun* **11**, 5650 (2020).
[160] Zaitsev, K., Bambouskova, M., Swain, A. & Artyomov, M. N. Complete deconvolution of cellular mixtures based on linearity of transcriptional signatures. *Nat Commun* **10**, 2209 (2019).
[161] Dong, M. *et al.* SCDC: bulk gene expression deconvolution by multiple single-cell RNA sequencing references. *Brief Bioinform* (2020).
[162] Newman, A. M. *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* **37**, 773–782 (2019).
[163] Menden, K. *et al.* Deep learning-based cell composition analysis from tissue expression profiles. *Sci Adv* **6**, eaba2619 (2020).
[164] Boyd, S. & Vandenberghe, L. *Convex Optimization* (Cambridge University Press, 2004).
[165] Grant, M. & Boyd, S. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx` (2014).
[166] Boyd, S., Parikh, N., Chu, E., Peleato, B. & Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning* **3**, 1–122 (2011).
[167] Almstedt, E. *et al.* Integrative discovery of treatments for high-risk neuroblastoma. *Nat Commun* **11**, 71 (2020).

[168] Johansson, P. *et al.* A Patient-Derived Cell Atlas Informs Precision Targeting of Glioblastoma. *Cell Rep* **32**, 107897 (2020).
[169] Yang, Z. & Michailidis, G. A non-negative matrix factorization method for detecting modules in heterogeneous omics multi-modal data. *Bioinformatics* **32**, 1–8 (2016).
[170] Replogle, J. M. *et al.* Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell* **185**, 2559–2575 (2022).
[171] Bray, M. A. *et al.* Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat Protoc* **11**, 1757–1774 (2016).
[172] Narayan, R. S. *et al.* A cancer drug atlas enables synergistic targeting of independent drug vulnerabilities. *Nat Commun* **11**, 2935 (2020).
[173] Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650–665 (2018).
[174] Joung, J. *et al.* A transcription factor atlas of directed differentiation. *Cell* **186**, 209–229 (2023).
[175] Datlinger, P. *et al.* Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods* **14**, 297–301 (2017).
[176] Dixit, A. *et al.* Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853–1866 (2016).
[177] Leder, K. *et al.* Mathematical modeling of PDGF-driven glioblastoma reveals optimized radiation dosing schedules. *Cell* **156**, 603–616 (2014).
[178] Dean, J. A. *et al.* Phase I study of a novel glioblastoma radiation therapy schedule exploiting cell-state plasticity. *Neuro Oncol* (2022).

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1920

Editor: The Dean of the Faculty of Medicine

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".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-498239

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2023