Dissertation

submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

presented by

M.Sc. Björn Benjamin Eismann

born in: Konstanz, Germany

oral examination: 09.09.2019

Comparative assessment of induced abnormal mitotic events by high-throughput light sheet imaging and image analysis

Referees: Prof. Dr. Roland Eils

Prof. Dr. Frank Lyko

Contributions

If not stated otherwise, all experiments and results presented in this thesis were conducted, processed, and analyzed by myself, under direct supervision of Dr. Christian Conrad and Prof. Dr. Roland Eils. Nevertheless, some steps of the experiments or of the data analysis where conducted in collaboration with colleagues. Hence, I will present the project as a group effort. I highly appreciate their significant help to this project and will specify their contributions as follows.

Carl Herrmann assisted with the detection of regulatory CpGs of mitotic genes. Jürgen Beneke and Ruben Bulkescher helped to prepare, test, and conduct siRNA coating in 96-well plates. Jürgen Beneke helped to develop the Hamilton liquid handling robot Matrigel cell spotting protocol. The "hSPIM" image processing software was developed in collaboration with mBits imaging GmbH represented by Markus Fangerau. Teresa Krieger helped with analysis of nuclei features and data visualization in R. Lukas Adam has designed and trained the convolutional neuronal network for image classification.

Contents

Abstract

In recent years, three-dimensional (3D) *in vitro* cell culture models such as spheroids and organoids have revolutionized life science research by providing a much more reliable context resembling the *in vivo* microenvironment. These systems yield important cell-to-cell interactions and induce cell differentiation. However, no conventional microscopy setup can provide sufficient imaging throughput as well as spatial and temporal resolution to enable full 3D live imaging and analysis down to subcellular processes.

In this project, we established state-of-the-art light sheet microscopy for live, long-term imaging of a short interfering ribonucleic acid (siRNA) treated 3D cell culture model. Due to the high temporal and special resolution of the light sheet microscope, we minimized imaging artifacts and achieved unprecedented visual representations of spheroids throughout development and upon gene knock-down by siRNAs. Furthermore, we deployed a high-throughput image analysis pipeline and machine learning classification to evaluate global, cellular and subcellular features for a precise, quantitative gene knock-down phenotype description. The RNA interference (RNAi) induced gene knock-down phenotypes were replicated and compared by a novel molecular, site-specific epigenome modifying method.

Throughout this project, we carefully evaluated every step of the workflow to improved its throughput and increased its reproducibility and usability. We addressed the key challenges in light sheet microscopy, such as sample preparation, data handling, image processing and analysis, thereby establishing quantitative light sheet microscopy screening of 3D cell culture models for many research applications. In total, we believe that our workflow can provide the basis for high-content analysis of 3D cell culture models for future research, enabling much more detailed functional experiments and basic research studies.

Zusammenfassung

Dreidimensionale (3D) *in vitro* Zellkulturmodelle wie Spheroide und Organoide haben in den letzten Jahren den Bereich der Lebenswissenschaften revolutioniert, da sie die zelluläre Mikroumgebung in einem deutlich verlässlicheren Kontext wiederspiegeln In diesen Zellkulturmodellen können wichtige Zell-Zell-Kontakte ausgebildet und die Zelldifferenzierung induziert werden. Dennoch können etablierte Mikroskopie-Methoden bisher keine ausreichende Aufnahmegeschwindigkeit erzielen und weder die räumliche noch zeitliche Auflösung liefern, um vollständige 3D Aufnahmen und Bildanalyse von sub-zellulären Prozessanalysen zu ermöglichen.

In dieser Arbeit haben wir die neuste Lichtblattmikroskopie für Langzeitaufnahmen von einem lebenden 3D Zellkulturmodellen nach einer Behandlung durch siRNA (kleine eingreifende Ribonukleinsäure) etabliert. Durch die hohe zeitliche und räumliche Auflösung des Lichtblattmikroskops konnten wir Mikroskopieartefakte minimieren und erzielten bisher unerreichte visuelle Darstellungen von Spheroiden sowohl während ihrer Entwicklung, als auch nach Verminderung der Expression spezifischer Gene durch siRNA Behandlung. Für die präzise und quantitative Beschreibung von Phänotypen, die durch die Herunterregulation bestimmter Gene hervorgerufen wurde, haben wir eine Hochdurchsatzbildanalyse-Pipeline sowie Klassifizierungen mittels künstlicher Intelligenz zur Bewertung von globalen, zellulären, sowie subzellulären Eigenschaften etabliert und verwendet. Außerdem wurden die interferierende RNA induzierten Phänotypen mittels einer neuen molekularen, zielgerichteten Methode zur Modifizierung des Epigenoms repliziert und verglichen.

Während der gesamten Projektes haben wir jeden Schritt des Arbeitsablaufs kritisch evaluiert um Durchsatz, Reproduzierbarkeit und Verwendbarkeit zu verbessern. Die Schlüsselprobleme der Lichtblattmikroskopie wie Probenvorbereitung, Datentransfer, Bildbearbeitung und Bildanalyse konnten wir durch einen hohen Grad an Automation von der Zellpertubation bis hin zur abschließenden Analyse beheben. Dadurch haben wir ein quantitatives Lichtblattmikroskopie-Screening von 3D Zellkulturmodellen für weitreichende wissenschaftliche Anwendungen etabliert. Insgesamt glauben wir, dass unser Arbeitsablauf die Basis für eine Hochdurchsatzanalyse von 3D Zellkulturmodellen für zukünftige Forschungsprojekte bereitstellen kann, da er detailliertere funktionale Analysen ermöglicht.

1 Introduction

Naturally arising oxidative damage to the human genome occurs around 10,000 times per day in every single cell [1]. Due to DNA repair mechanisms or induced cell death, these mutations do not have a detrimental effect on the tissue [2] or organism. Yet, if these mechanisms fail, mutations are acquired over time and can induce uncontrolled cellular expansion, resulting in cancer development. Growth and spreading of cancer tissue into essential organs can result in organ failure and patient death [3]. Today, cancer is still one of the leading causes of death despite medical improvements in recent years in developed countries [3]. To understand how cancer develops, how it spreads, and how it can be cured, cancer research has to study the underlying cellular mechanisms inducing cancer development, and how these changes can be reverted, or how cancer tissue can be specifically treated to inhibit cellular growth. Novel 3D human cell culture models provide a versatile system to study cancer development, tumor remission, as well as potential treatments [4].

1.1 Cell culture models in cancer research

Human and animal cell culture models are an essential, irreplaceable component in any life science research from basic research to translational cancer research [5, 6]. The term cell culture model refers to living, immortalized cells that live and proliferate under defined culture conditions (Figure 1). Cell lines can be established from tumor or somatic tissue [7]. Immortality can be acquired through naturally occurring mutations in tumor tissue or through introduction of a viral gene or modulated gene expression that deregulates the cell cycle [8]. Cell lines resemble properties of the tissue of origin, while xenografts in mice are a common method to improve tumor resemblance of the cell line as patient-derived xenografts accumulate copy number alterations due to selection of minor clones (Figure 1) [9].

Cell lines facilitate the study of cellular components, properties, processes, and functional responses to internal and external cues in a simple, robust, and easy-to-use setup with many established imaging and omics analysis methods. These omics methods allow evaluation and quantification of different aspects of RNA and protein biosynthesis, function, localization, homeostasis as well as their dynamics (Figure 1) [4, 10, 11].

Throughout history, these cost-effective *in vitro* cell cultures have led to significant insights into biological processes, circumventing elaborate work with model organisms as well as enabling studies on live human tissue [12]. One of the greatest advantages of cell culture research is the abundancy of different cell lines and the possibility to rapidly enlarge sample size and the number of cells. Additionally, many methods have been established to modify the

genome or transcriptome and to generate transgenic cell lines that express, for example, a recombinant protein or lack an endogenous gene of interest.
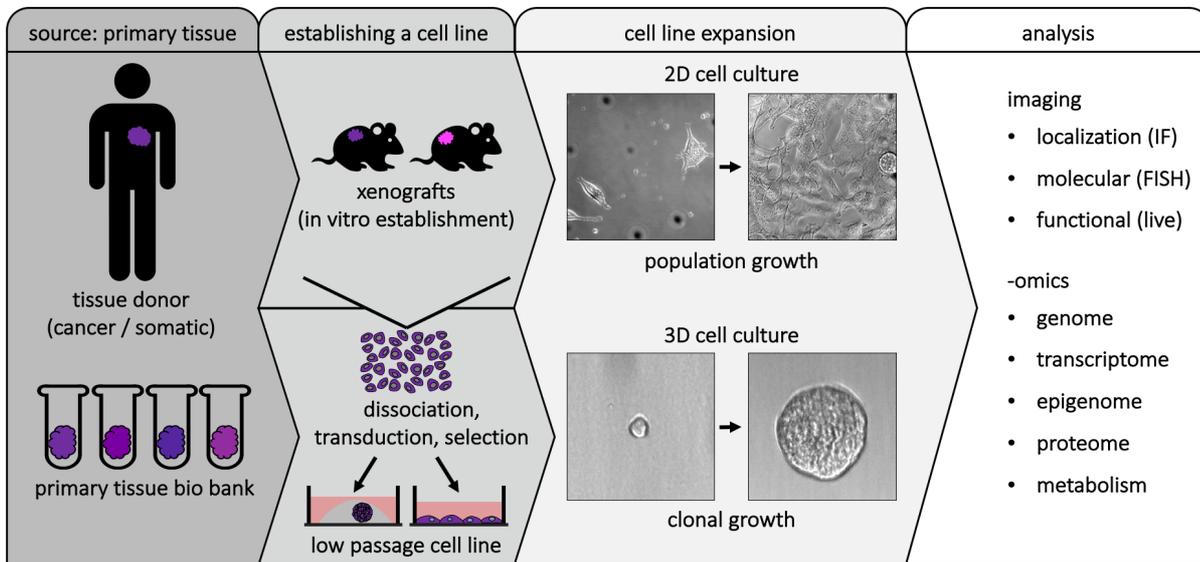
Figure 1 – Cell culture models for molecular and phenotypic analysis

purposely immortalized with viral genes and subsequently expanded in defined growth media, ensuring high proliferation and defined cell characteristics through culture media supplements. An additional step of mice xenografts improves viability and tumor properties of the cells. Established cell lines can subsequently be used in 2D or 3D environments for functional experiments analyzing structural features by immunofluorescence (IF), or molecular properties by fluorescent in-situ hybridization (FISH). Functional and omics methods characterize the genotype and phenotype of the tissue of origin.

1.1.1   2D cell culture – a basic *in vitro* model systems for research

For many decades, two-dimensional (2D) cell culture models have been the state-of-the-art system for any *in vitro* study, even though the physiologically relevant microenvironment is missing. In 2D, cells are usually grown on a stiff plastic surface, covered by their respective culture medium and passaged every few days, depending on their replication time. The hard surface of the culture dish modifies tissue-specific cell architecture and forces a defined polarity and cell shape onto the single cells, which alters mechanical and biochemical signals and subsequent cell-to-cell communication [13]. Modern alterations of 2D systems like transwell polycarbonate membrane cell culture inserts allow for more complex cultivation, as for example reconstituting the air-tissue interface of the lung [14].

Different implementations of 2D cell culture have helped extensively to understand cellular properties, though lacking the 3D context of the *in vivo* tissue [6, 12, 13]. Important cell-to-cell interactions between cells or the extracellular matrix (ECM) are also absent. Differentiation inducing or retaining physical microenvironment properties are lacking. Yet, these non-

physiological culture properties influence gene expression profiles and therefore the phenotype of the cells [15].

Due to these non-physiological culture conditions, cells grown in monolayer can lose distinct properties and features that are potential drug targets in medicine. For instance, primary hepatocytes cultured in 2D lose their differential state within a few days and monolayer cells silence drug component metabolizing enzymes, which makes 2D models much more sensitive to drugs [16]. Administering the same drug concentration that was effective in a 2D cell culture model to a patient, it would be highly inefficient and potentially even induce drug resistance in the cancer tissue [7].

Last but not least, a surrounding tissue can influence availability of nutrition and oxygen, limiting the cellular metabolism and replication time, introducing cell differentiation [13, 17] as well as influencing the drug permeability to the target cells [7].

1.1.2   3D spheroid cell culture in life science research

Most of these above mentioned major drawbacks of 2D cell culture models are addressed by the application of 3D cell culture models in research, first used in the late nineteen-fifties [18]. Today, there are two established nomenclatures for cultured 3D cell clusters – spheroids and organoids. While these two 3D cell culture systems differ in size and progress of differentiation of single cells within the cluster, but attributes that define the transition from a spheroid to an organoid are not clearly outlined and subject to ongoing discussions.

There is a great variety of tested and continuously improving methods to induce 3D spheroid development. Mechanical solutions like the "hanging drop", "ultra-low attachment plates", "spinner flask culture" or the "rotary cell culture system" [19] do not allow any adhesion of cells to a surface and therefore cause cells to aggregate. Other approaches reconstitute the basal membrane through a gelatinous protein mixture called Matrigel derived from Engelbreth-Holm-Swarm mouse sarcoma cells. Matrigel contains essential ECM proteins such as laminin, collagen IV, heparin sulfate proteoglycans, entactin or nidogen and a number of selected growth factors [12, 13].

Results obtained from research using 3D cell culture models like spheroids have shown that the additional dimension and loss of stiff culture surface changes gene expression dramatically, thus resembling more closely the respective (tumor-) tissue [7, 20]. Furthermore, key events of cellular development, such as proliferation, migration and induced apoptosis are defined by the extra cellular context [17, 21, 22], which means that essential factors and physical properties of the ECM need to be established throughout spheroid development, simulating the *in vivo* context. Another strong point for the usage of 3D cell culture models is

the high degree of heterogeneity between the cells due to differentiation and sub-type development [23], as it can be found in tumor tissue. It was shown by Ghosh and colleagues that genes upregulated in tumors match with their gene expression in their respective 3D cell culture model [20]. The close imitation of *in vivo* tissue improves reliability and reproducibility of research results, and therefore closes the gap between cell culture and the respective physiological *in vivo* tissue [24].

Especially in the field of oncology and the selection of promising drugs in pre-clinical screens, the discussed advantages of 3D cell culture models [5] provide the essential context for the characterization of cancer subtype and its therapeutic treatment [7, 22].

1.1.3   Tumor initiation

Cancer development describes a process of successive accumulation of (epi-) genome mutations that alter the gene expression profile of a cell and thereby cause this cell to undergo an abnormal phenotypic transformation with acquisition of biological capabilities [9, 25]. Post-mitotic somatic cells can reacquire stem cell like properties re-entering a highly proliferative state and invade into other tissues or organs, where the misallocated cells can cause organ failure through growth and nutrient deprivation of the physiological tissue [25, 26]. The general theory implies that a single mutation is most of the time not sufficient to induce cancer development, but it is rather the interplay of accumulated mutations, which change the cell's gene expression. Genes regularly activated in this process are called oncogenes, while genes which often repressed throughout oncogenesis are called tumor suppressor genes [27-29].

Tumor tissue can be separated into two states, benign and malignant [23]. Benign tumors remain at the original tumor site and have not (yet) acquired the capability to invade neighboring tissue, while malignant tumors are able to migrate and invade other body parts [29]. Both groups can show a high degree of cell proliferation, a key hallmark of cancer [25]. Further cancer hallmarks include genome instability, evasion of growth suppression and immune destruction including apoptosis resistance, replicative immortality, reprogrammed metabolism, induced angiogenesis in the tumor tissue, and invasive migration capabilities [5].

1.1.4   The role of mitotic defects in tumorigenesis

Cell division or mitosis is a highly regulated, cellular process describing the replication and separation of the full genome into two daughter cells, which is a prerequisite for life and growth of any organism. The cell cycle can be divided into different stages or phases that separate the replicated genome into the two daughter cells. During the **prophase**, the DNA is compacted into chromosomes with two identical sister chromosomes attached at the

centrosome and the nuclear envelope is being disassembled. The chromosomes are subsequently aligned at the mid plane of the cell during the **metaphase** with the help of the mitotic spindle apparatus. These microtubules are attached to the centrosomes and separate the sister chromosomes to one pole of the dividing cell during the **anaphase**. The cell division is completed during the **telophase**, when two daughter cells complete the cytokinesis, disassemble the mitotic spindles, decondense the DNA, and reform the nuclear membrane as well as acquire normal cellular shape.

Numerous mechanisms such as the spindle assembly check point and the no-cut checkpoint control define steps of the mitosis and ensure accurate distribution of the genome to the daughter cells [30]. Minor mistakes throughout mitosis can result in daughter cells with an incomplete genome or too many chromosomes, a feature called aneuploidy. Aneuploidy is regarded as common trait of cancer cells, arising in almost 70% of solid human tumors, as chromosome misdistribution provides cancer cells with a mechanism to lose tumor suppressor loci and gain extra copies of oncogenes [31]. Furthermore, features of cancer cells include structural alterations such as genome region translocation, amplification or deletion, resulting from high proliferation and loss of DNA repair mechanisms. These traits are referred to as chromosomal instability and are a leading cause for tumor development [32], from initiation, to growth, and acquisition of additional cancer hallmarks. In total, mitotic defects are considered the most influential step in tumor initiation and subtype formation, resulting in intratumor heterogeneity [23, 33].

1.1.5   Resolving intratumor and intertumor heterogeneity

Intratumor and intertumor heterogeneity in genotype, epigenotype, phenotype, and underlying gene expression are conceived as the most crucial aspects that needs to be analyzed to treat cancer patients [23]. The heterogeneity among patients makes it challenging to decide on a defined treatment solely by the tumor type. Small differences in tumor initiation and progression can cause distinct phenotypes within a single patient and a single tumor, with various resistances to one or the other chemotherapy [23, 33-35]. Furthermore, depending on the tissue of origin and the tumor microenvironment, a single tumor type could display distinct properties [35], drug resistance [36] and relapse properties [37]. For example, although almost indistinguishable by histo-cytochemistry, the tumor medulloblastoma displays a high chance of remission when originating from the dorsal brainstem, while medulloblastoma arising in the cerebellum causes a much shorter life expectancy [38]. To tackle tumor heterogeneity, it is not only essential to establish analysis methods that allow a focusing on the individual cell in the context of a microenvironment. Furthermore, it is essential to establish 3D cell culture

models representing defined tissue characteristics that can be used to analyze the initial steps of tumor development as well as tumor progression.

1.1.6 MCF10A – a benign breast epithelia model for tumor initiation studies

One of the most commonly occurring cancer types for females is breast cancer. In the United States in 2017, women in the United States had a 12.4% (1 in 8) lifetime risk to be diagnosed with breast cancer [breastcancer.org]. There are two main groups of breast cancer described, the ductal carcinoma *in situ* and lobular carcinoma *in situ* also known as lobular neoplasia [39]. In both breast cancer types, epithelial cells acquire in a stepwise transformation a highly proliferative phenotype starting from local hyperplasia to premalignant carcinoma *in situ* and developing into the highly motile, invasive carcinoma [28]. Breast cancer cells invade and expand into the ducts and lobules of the breast. Causing mutations are highly probable in the *BRCA1* and *BRCA2* genes [2]. These tumor suppressor genes encode for nuclear phosphoproteins that play a key role in maintaining genomic stability by regulating the homologous recombination pathway for double-stranded DNA repair [2, 31, 40]. A defect in this repair system causes a high likelihood to acquire genetic mutations that cause the development of breast cancer.

Michigan Cancer Foundation 10A cell line (MCF10A) is a benign, *in vitro*, human breast epithelium derived, 3D cell line that develops under 3D culture conditions in Matrigel into polarized micro tissue called spheroids [41]. These spheroids represent the acini of the human breast and display essential functions such as polarization and differentiation and exhibit the cell-to-cell variability found in the respective *in vivo* organ [12, 22]. These properties make the MCF10A cell line the ideal candidate for tumor initiation screens, in which the acquisition of (epi-) genetic mutations is replicated and reconstitutes the first steps of tumor development *in vitro*.

1.2 Epigenetics

Genetic and epigenetic mutations equally contribute to loss or gain of gene expression and oncogenic cell transformation [42, 43]. Epigenetics describe all non-genomic but heritable DNA modifications that are passed from the mother-cell to the daughter-cells [44]. The importance of aberrant DNA modifications and DNA associated protein modifications in tumorigenesis and tumor metastasis is highlighted by the increasing number of known tumor associated epigenetic alterations [45]. The two principal epigenetic modifications known for their significant role in cancer development are DNA methylation at 5-methylcytosine of the cytosine guanine dinucleotide (CpG) and covalent post-translational histone modifications

[46]. The latter include an addition or removal of a methyl group, acetyl group, phosphorylation, ubiquitination and sumoylation at the histone tails of the different subunits of the protein complex. And they are mediated by the enzyme group of histone modifiers and the chromatin remodeling complex [47-52]. The second main class of epigenetic modifications describes an addition of a methyl group to the cytosine of a CpG dinucleotide. Genomic regions that show distinct differential DNA methylation levels between tumor or non-tumor tissue at CpGs are called differentially methylated regions (DMR) and are regularly associated with CpG islands, DNA regions with high percentage (> 60%) of cytosine guanine dinucleotide and a length of at least 200 base pairs [53]. The DNA methyltransferase family (DNMT) catalyzes the addition of the methyl group to the C5 position of the cytosine ring (5-mC) in CpG dinucleotide. DNMT1 preferentially binds to replication sites in the S phase and methylates hemimethylated DNA, maintaining the methylation pattern on the newly synthesized strand [54] (Figure 2a). DNMT3a and DNMT3b catalyze *de novo* methylation at loci throughout development of an organism or a tissue [55] (Figure 2a). DNMT3l is an important stimulatory factor for DNMT3A, which regulates the duration of the attachment of methyltransferases to the DNA [56]. 5-mC is a reversible DNA modification that can either be lost through semi conservative DNA replication, or through a multistep enzymatic process. The initial step of DNA demethylation is catalyzed by Ten-eleven translocation methylcytosine dioxygenase 1 (TET1), oxidizing the modified DNA base 5-methylcytosine to 5-hydroxymethylcytosine (5-hmC) [57] (Figure 2a). 5-hmC is then further processed to 5-carboxylcytosine (5-caC) which is subsequently excised by a thymine DNA glycosylase (TDG) generating an abasic site as part of the base excision repair (BER) process [58] (Figure 2a).

1.2.1   Epigenetics in tumor development

Epigenetic modifications at DMRs are able to significantly impact gene expression through rearrangement of the nucleosomes, changing the chromatin structure as well as allowing or inhibiting the binding of transcription factor and regulating the recruitment of histone modifiers (Figure 2b) [59]. DMRs are found mainly in promoter regions, but also in gene bodies and intergenic regions. Although not exclusive, hypermethylated DMRs generally correlate with gene repression, while hypomethylated promoters are associated with higher gene expression [53, 60].

1.2.2   Epigenome modifying molecular tools

The fact that epigenetic states are in some genomic regions highly responsive to external influences or the cellular, developmental progress in a highly reproducible manner, make

epigenetic markers a more promising target for tumor therapy than genome modifications [61-64]. Genomic locus directed epigenome modifying enzymes in theory allow a complete remission of the tumor state, reverting the oncogenic phenotype of the epigenome misregulated tumor cell [64] (Figure 2b). Furthermore, epigenome modifying molecular tools are able to induce a defined gene expression pattern that drives a cell into apoptosis or inhibit certain cancer hallmark phenotypes [65-68]. Epigenome remodeling enzymes have been described already in the 1990s, but the key challenge is the accurate positioning of these enzymes to a genomic locus for site-specific epigenome remodeling [64, 69, 70].

Several molecular tools have been developed over the last years, including Transcription Activator-Like Effector Nuclease (TALEN) [71], and zinc finger DNA binding proteins (ZF) [65, 72] (Figure 2c). These systems have been successfully adapted to modify epigenetic traits at a specific genomic locus [49, 73, 74]. Most problematic for these methods were cost and effort required to generate a single functional epigenome modifier in relation to the epigenome modification success rate [75]. Each target site requires a redesign and protein synthesis of the ZF and TALEN with no guarantee of high effectivity [71].

The clustered regularly interspaced short palindromic repeats (CRISPR) system, described by three groups almost simultaneously in 2012 [76-78] was adapted from its function as microbial adaptive immune system to a recently broadly used genome editing tool [76]. In the following years, the original CRISPR system was adapted to a broad range of additional applications, using the deactivated Cas9 (dCas9) to target specific effector domains (dCas9-ED) to a specific genomic locus (Figure 2c) [79-84]. In the dCas9 protein, the two endonuclease domains RuvC and HNH are mutated in a single residue each (D10A / H840A) [24], forfeiting the endonuclease activity of the molecular tool. The greatest advantage of the different al CRISPR systems is the high adaptability, since site specificity is defined by the RNA component, the single guide RNA (sgRNA), which can be administered separately to the dCas9 protein component [85, 86]. This trait enables a fast exchange of genomic targets and versatile synthesis. Furthermore, multiple genome sites can be targeted simultaneously [87-89]. A drawback of the usage of the CRISPR-dCas system for locus specific epigenomic editing is the lower target specificity, which allows base pairing with up to 5 mismatches between sgRNA and the DNA [90] thus requiring cautious experimental design. Rapid development and constant improvement of online target selection and off-targeting prediction tools have made sgRNA synthesis and specificity evaluations a relatively easy task [87, 91, 92].
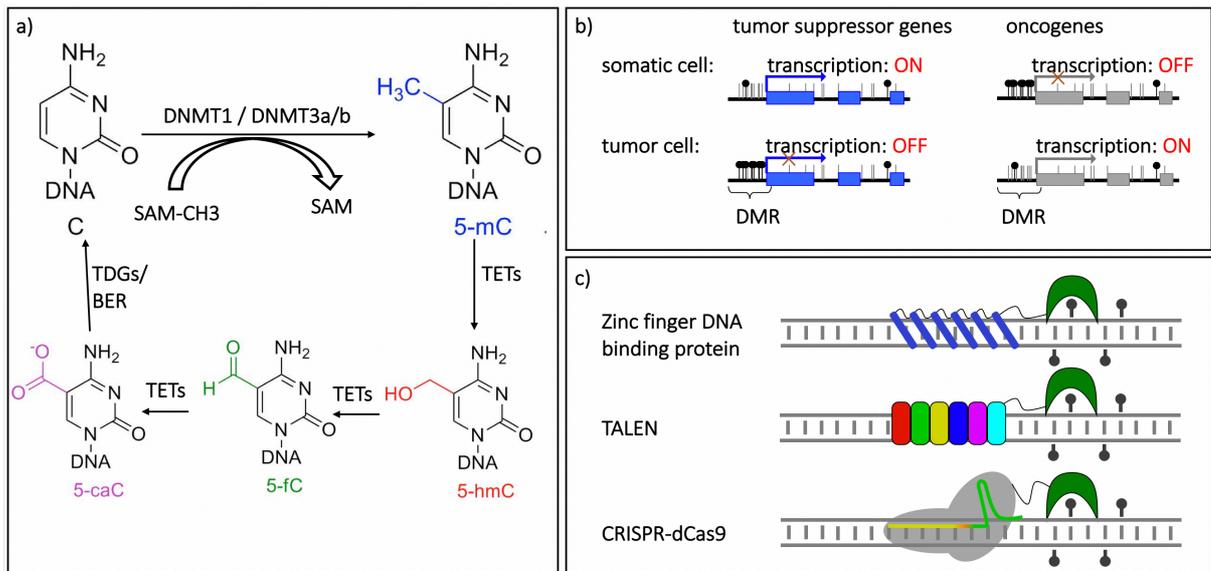
Figure 2 – Gene regulation by methylome modification in tumor tissue

a) The addition of a methyl group to 5C is catalyzed by DNMT1 (methylation pattern maintenance) or by DNMT3a/b (*de novo* methylation). DNA methylation is a reversible modification that can be actively removed by a multistep enzymatic reaction induced by TET1 catalyzing the initial step of demethylation b) Many tumor types are associated with aberrant DNA methylation on single CpGs or entire regions. Particular aberrant methylation pattern in DMRs of oncogenes and tumor suppressor genes result in tumorigenesis through abnormal gene expression. c) Zinc finger DNA binding proteins, TALEN, or CRISPR-dCas9 fused to methylome-modifyingifying enzymes have been shown to be functional in site directed methylome modification inducing selected transcription activation or repression.

In the last three years, a number of epigenome modifying applications utilizing the CRISPR-dCas9 system to site-specific epigenome editing have been published (Figure 2c). Hilton and colleagues achieved gene activation through histone H3 lysine 27 (H3K27) acetylation by fusing the acetyltransferase p300 core domain to the dCas9 [93], while Kearns and colleagues described a similar approach with gene repression through coupling of the LSD1 to the dCas and targeting H3K4Me2 and H3K27Ac histone modifications around the targeted Tbx3 enhancer region [94]. Other epigenetic enzymes like KRAB, DNMT3A, and DNM3L and TET1 also allowed for long-term epigenetic silencing or activation of the target endogenous genes through site directed methylation and demethylation of specified genomic regions [63, 95, 96] (Figure 2c). Although highly innovative, all these approaches display one common, crucial difficulty besides the methylome engineering efficiency. The alterations at the epigenome could be analyzed by methods such as pyrosequencing or ChiP-seq technology, but functional validations of the effect onto the cellular phenotype were not definite [52, 61, 93].

1.2.3 3D context and epigenetics – a new challenge in cancer research

Conventional methods like siRNA knock-down functional experiments and the described novel epigenome modifying tools can be used to study gene function, regulation, and individual

contribution to oncogenesis in 3D model systems. The contribution of the epigenome to tumorigenesis as well as the capability of the epigenome modifying molecular tools to precisely change these properties provide novel possibilities to revert cancerous characteristics of cells. Major challenges for this method are off-targeting effects, insufficient understanding of comprehensive regulatory mechanisms and effects of site directed epigenetic modifications [73]. The advantage of targeted epigenetic modification over traditional methods such as siRNA for a high-throughput knock-down screen needs to be determined, as CRISPR-Cas9 based targeted epigenetic modifications require a protein component as well as a RNA component delivered into each cell to induce genomic site directed epigenome modification. Many analysis methods such as (epi-) genome and transcriptome sequencing, metabolism analysis and imaging are currently adapted, improved or developed to suite the requirements of 3D cell culture models. Especially difficult is the visual representation of the spheroid or organoid. Due to light scattering, conventional microscopy techniques such as epifluorescent or confocal microscopes are insufficient to visualize and analyze the entirety of these 3D cell culture models. Novel imaging techniques such as light sheet microscopy for the first-time facilitate full 3D acquisition of spheroids and organoids.

1.3 Fluorescent light sheet microscopy

To study subcellular processes, development, or interactions between cells of a tissue, light microscopes are the most important and powerful analytic tool, as they can resolve subcellular components smaller than one micrometer [97].

1.3.1 Fluorescent microscopy

The development of the fluorescent microscope improved resolution and enabled a selected visualization of single structures and subcellular components. In a fluorescent microscope, the specimen is illuminated by a defined wavelength that is absorbed by the fluorophores [97]. These excited fluorophores emit the absorbed energy at a longer wavelength than used for excitation [98]. This shift between excitation and emitted wavelength is used to split the optical signal enabling fluorophore localization [97]. Fluorophores, such as dyes or fluorescent fusion proteins enable a visualization of cellular structures or components [99-102].

Key fluorescent microscopy technologies available to almost all life science researchers are conventional epifluorescence microscopy (Figure 3a), confocal laser scanning microscopy (CLSM) (Figure 3b) and Spinning-disk (Nipkow disk) confocal microscopy (SDM). Today, many different advances in microscopy development such as super resolution microscopy, have pushed the resolution limit further, making cellular processes or structures visible that

could not have been observed otherwise [103]. Additionally, constant improvement of molecular staining tools such as fluorescently labeled antibodies, fluorescent dyes, and fluorescent tags for proteins [99] exemplified by the green fluorescent protein (GFP) [102] and its deviates, facilitate novel applications and methods.

Although broadly established, conventional bright field and fluorescent microscopy have a substantial disadvantage as the excitation light illuminates the entire specimen, causing photo-bleaching and phototoxic effects in all planes of the entire sample, while the emitted light is detected only in a single plane [104]. With the development of the light sheet microscope, the issue of redundant light overexposure of the sample is addressed [10, 104-106].

1.3.2 Light sheet microscopy

In 1903, Siedentopf and Zsigmondy described the very first concepts of a light sheet microscope [107]. The so-called "ultramicroscope" employed a slid aperture to project sun light onto a sample at a 90° angle to the detection objective. This setup illuminated only a thin section of the sample, providing lateral sectioning. Almost a century later, Voie and colleagues reimplemented the orthogonal orientation of excitation and detection path to image the internal architecture of the cochlea with their orthogonal-plane fluorescence optical sectioning (OPFOS) microscope [108]. Their microscope already comprised many features of today's commercially available light sheet microscopes, such as laser excitation, beam expander for light sheet modulation, a cylindrical lens to generate the light sheet and Z-stages to move the specimen through the light sheet. Although acknowledged as innovative imaging technique, light sheet fluorescent microscopy (LSFM) remained largely unrecognized as method for life science research. In 2004, Jan Huisken and colleagues [109] published their single-plane illumination microscopy (SPIM) setup, which was capable to visualize *in vivo* muscles in a transgenic Medaka (O*ryzias latipes*) fish model organism as well as image development of a *Drosophila melanogaster* embryo for up to 17 hours at unmatched spatial and temporal resolution.

Throughout the subsequent years, SPIM imaging was accredited as surpassing imaging technique [105] with superior properties to conventional microscopy, which caused a rapid development of many different light sheet technology implementations. Many research groups developed their own implementation of the optical setup to study mainly embryogenesis of different model organisms, focusing on tissue differentiation, organ development and cellular positioning throughout the development of the embryo [104].

1.3.3 Opto-physical principles of light sheet microscopy

All light sheet microscopes are based on the same basic, opto-physical principles. But they highly differ in light sheet generation, in number and properties of the excitation and detection paths, as well as in the orientation of the setup [106, 110]. In comparison to any other fluorescent microscope, in which a single objective emits the excitation light along the same axis as the fluorescent signal is collected (Figure 3a-b), the light sheet microscope splits excitation and detection light path at a commonly 90° angle and illuminates the sample from the side (Figure 3c).
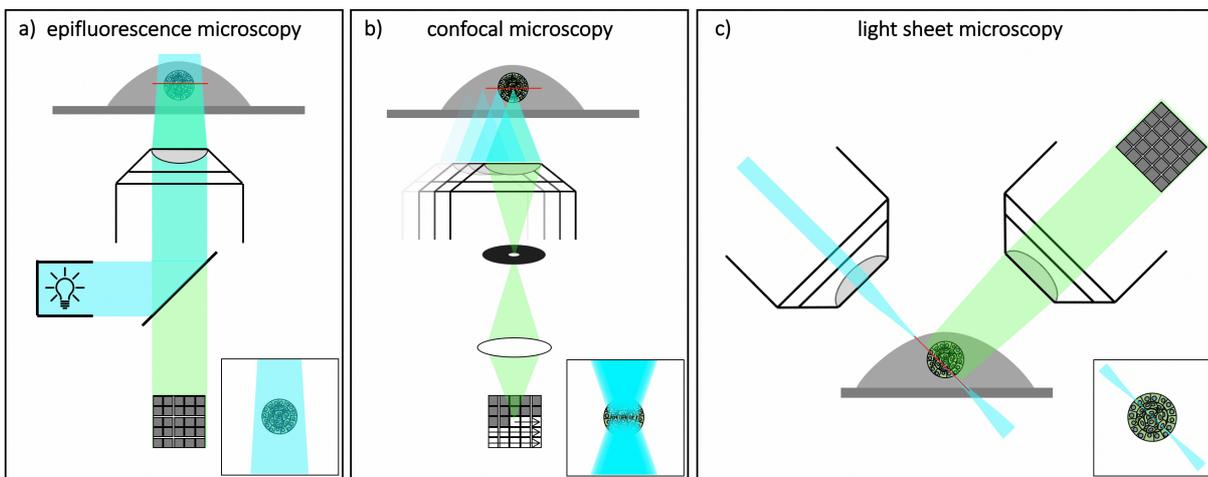
Figure 3 – Overview on fluorescent imaging techniques

Schematic (not drawn to scale) of microscopes acquiring a single image (red line) of a 3D spheroid sample. Insets describe illumination pattern needed for single slice acquisition. The specimen is illuminated with light at a fluorophore-specific excitation wavelength (blue). The fluorophores at ground state absorb the light energy and acquire an excitation singlet state. Immediately, fluorophores recover the stable ground state through energy and light emission at a longer emission wavelength detected through the detection path (green). Objectives, filters and dichroic beamsplitters direct and split the different excitation and emission wavelengths from the light source onto the sample and from the sample onto the detection apertures. a) In epifluorescent microscopes, the excitation light is transmitted through the sample, which allows for full frame detection, while in a confocal microscope b) the light is focused in a single point and scanned over time across the focus plane creating a scanned image. A pinhole (black disc) dismisses out-of-focus light. In both microscopes, excitation and detection optical path are passing through the same objective. c) In a light sheet microscope, excitation and detection paths are separated by 90°. Only a thin sheet of light or Gaussian beam is used to scan across the sample in synchronization with a full frame camera, illuminating only a thin layer of the sample per frame.

The light sheet can either be generated by a collimated laser passing through a slit aperture and a cylindrical lens forming a static light sheet or a virtual light sheet can be generated by a Gaussian beam that is scanned by a Galvo scanning mirror across the field of view [4, 104, 110-112]. Both technical implementations form a thin sheet of light that illuminates only a thin layer within the specimen, reducing the overall light exposure onto the sample drastically [104] and provide already a good sectioning of the sample (Figure 3c) [104]. The detection objective is placed at a perpendicular angle to the light sheet at focal distance, and transmits the

emission signal onto modern, full-frame scientific complementary metal-oxide-semiconductor (sCMOS) cameras [113].

From the most basic setup of a single excitation and secondary single detection objective, many novel implementations of the light sheet technology have evolved [110]. Some microscopes combine two detection and two excitation objectives orientated horizontally around a vertically placed sample, called a multi view setup [114]. This can be utilized to acquire four images of a specimen without rotation, each view acquiring a quarter of the sample at the highest resolution [114]. Another custom-built light sheet setup was developed to acquire mouse embryos at resolution limit [115]. Strnad and colleagues used an asymmetric setup with high numerical aperture (NA) detection objectives combined with a low NA excitation objective to acquire developing mouse embryos at the highest possible spatial resolution [115] while minimizing light exposure.

1.3.4 Sample mounting in light sheet microscopy

LSFM imaging requires a higher effort to prepare samples for imaging in comparison to standardized sample mounting for conventional microscopy with samples mounted on a simple glass slide or culture plate, making sample mounting one of the key challenges for SPIM imaging [10, 110, 116]. The optical geometry of a LSFM with two or more objectives at an extremely close proximity makes mounting and positioning the sample between the objectives an issue [114]. Furthermore, to obtain the optimal imaging conditions, it is essential that all components of the optical path and mounting media are adjusted to the refractive index (RI) of the sample [117]. Additionally, optimal environmental conditions like temperature, gas exchange, and nutrition must be provided without affecting light transition. Samples mounted in transparent fluorinated ethylene propylene (FEP) tubes and foils (Figure 4a), that have an RI of 1.33 identical to aqueous solutions ensure the continuity of the RI throughout the optical paths while also improving position and imaging stability [118, 119]. One major drawback of the usage of FEP foil is the reduced gas and nutrition exchange. Translucent gels like agarose and Matrigel are broadly used for aqueous imaging media and minimize light scattering (Figure 4b-c), while constraining the sample to a defined position [10].

Motorized positioning stages with extremely high precision and long travel range provide the capability to position the mounted sample exactly at the focal plane between the objectives for imaging. Larger samples with a size of 200 μm like developing embryos, plant roots or large organoids require additional imaging strategies like multi view acquisition to compensate for in depth light scattering [4, 109, 114].

1.3.5 Sample acquisition and acquisition modes

Jan Husiken *et al.* [109] were the first to implement a rotary stage that is capable of rotating the sample around an orthogonal axis and acquire the sample at multiple angles (Figure 4a). Only one excitation and one detection optical path, as exemplified in Figure 3c and Figure 4a, are needed for multi view imaging, but multiple excitation and detection paths can reduce imaging time.

Another approach for multi view acquisition, circumventing the rotation of the sample, is the usage of multiple, alike excitation and detection paths at either a 90° or 180° angel [112, 120]. For these light sheet microscopes, we can distinguish between two distinct imaging techniques to acquire a volume (Z-stack) of the specimen. With stage scan acquisition (SSA), the sample is translated through a static light sheet. On the one hand, SSA is advantageous for larger samples, because less optical components need to be adjusted which reduces vibrations and component cost, but on the other hand it is prone to produce excessive single view data for smaller samples (Figure 4b). Stage scan acquisition does not require an additional piezo motor moving the detection objective in sync with the light sheet. But it relies on a constant, uninterrupted and precise movement of the stage, translating the sample through the light sheet. SSA is furthermore challenging for image processing as the acquired data is skewed when the sample translation is not at an orthogonal angle to the excitation axis (Figure 4b) [121]. XY pixel positions do not directly correlate between slices in a Z-stack. And finally, to image a volume at multiple views, the Z-stack limits have to be set outside of the field of view (FOV), to ensure the acquisition of the FOV from each angle.

The second method is synchronous piezo/slice acquisition (SyncA) (Figure 4c). Here, the stage is stationary, while the light sheet is translated at focal distance in a synchronous movement together with the detection objective through the Z-stack [121, 122]. In comparison to SSA, SyncA requires more optical components to be adjusted like the position of the detection objective and the Z-position of the light sheet [112]. At these small intervals high frequency movements cause vibrations, and therefore elongate the minimal time between acquisitions of each slice of the Z-stack. However, only a defined volume is acquired from two views, restricted by the imaging properties of the objectives, which accelerates and simplifies multi view image reconstruction (MVR) (Figure 4c).

In conclusion, multi view imaging by sample rotation is optimal for large samples like organoids and model organism embryos with an extend of 200 μm and larger. Stage scan acquisition is highly effective for single sided, large area, thin sample acquisition, while synchronous piezo/slice acquisition minimizes acquired data and data processing effort for spheroid sized samples.
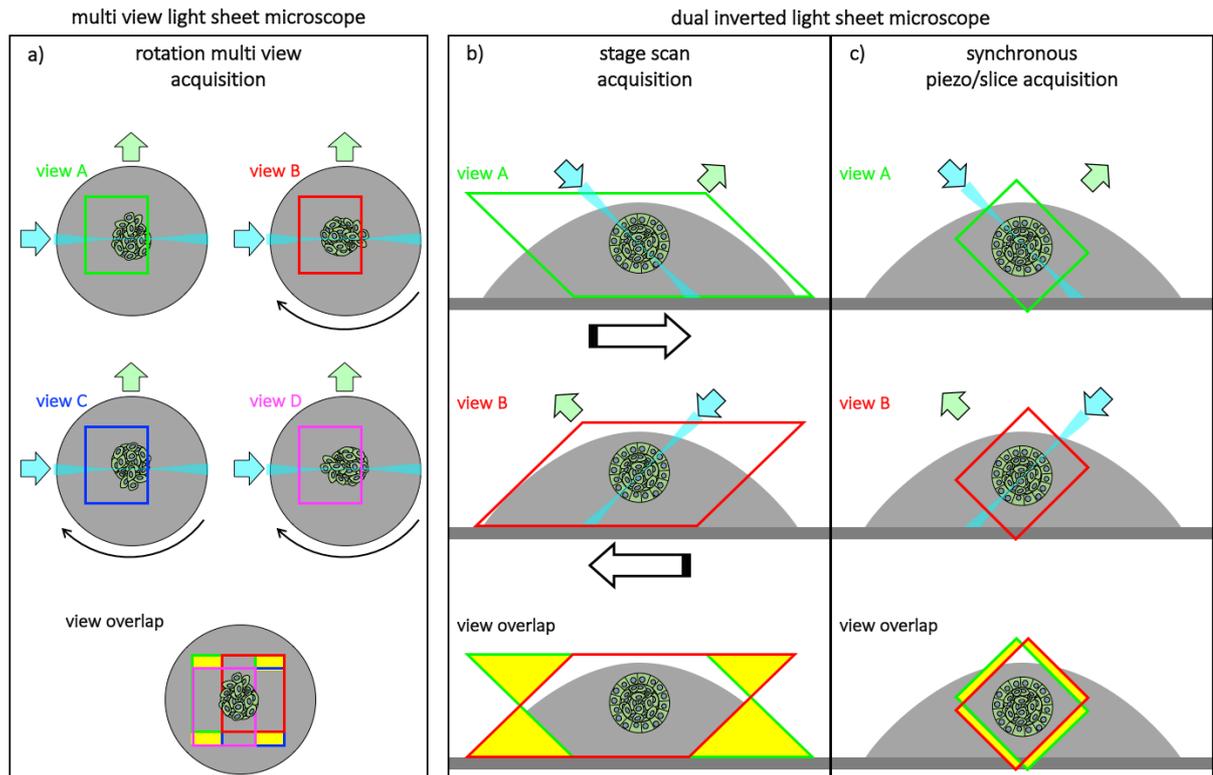
Figure 4 – Light sheet acquisition modes

a) A stage rotates the sample in an FEP tube to allow for multi view acquisition. One excitation path (light blue arrows and single slice light sheet) combined with rotation of the sample reduce light scattering by minimizing the distance the light sheet needs to travel through the tissue for each view. Fluorophores are detected at a 90° angle (light green arrow). The final sample image is calculated by fusion of the signal from the multiple views. b) SSA allows to quickly image a large area, but for smaller samples like a single spheroid mounted in a Matrigel spot (gray), a dual view acquisition (red and green bounding box) generates a high ratio of single view, excessive data (yellow). Additionally, acquired stacks are skewed, which requires additional image processing steps. c) SyncA acquires a defined FOV with fixed parameters, but in comparison to SSA the acquired data displays a higher ratio of volume imaged by both views (yellow volume).

1.3.6   Data generation, data handling, and multi view processing

Light sheet microscopes are equipped with sCMOS cameras that are capable to acquire large quantities of data with up to 1 GB/s [113, 123]. Therefore, data management, storage, and processing with high performance recourses are essential for SPIM imaging [124]. Restrictive acquisition by focusing imaging only on the field of interest (FOI) can reduce data size, but intelligent solutions such as intelligent imaging and adaptive optics are shaping the current light sheet microscopy development [125]. The term intelligent imaging describes a partly automated microscope that can actively change acquisition patterns upon a specific stimulus and selectively store or discard data depending on its content [126]. Adaptive optics describes microscopes that can evaluate acquired data and actively change the detection path to improve image quality and focus on an automatically defined FOI [125].

As a result of the large data sets acquired by LSFM, data science with focus on high-throughput image processing and analysis has become an integrated part in the field of light

sheet imaging. Image data of multiple views need to be fused and processed [114, 127]. The most commonly used processing pipeline is the multi view reconstruction (MVR) plugin developed in the group of S. Preibisch [128]. The processing tool detects points of interest within all views, like fluorescent beads embedded into the mounting gel and calculates from these positions a transformation matrix for each view. The combined image of the specimen is subsequently deconvolved and can be used for further analysis. The MVR pipeline is especially useful for multi view imaging with two or more acquired views, needed in particular for reconstruction of large embryos where the embryo surpasses the FOV multiple times.

1.3.7 Sample handling for light sheet based screens

The optimal imaging capabilities of the light sheet microscope, shift the key challenges of an imaging screen towards sample handling and data processing. Due to the physical positioning of the excitation and detection objective at close proximity, commercial high-content plates like the 94- or 384-well plates used in 2D cell culture screen are not suited for the multi-objective setup of some SPIM microscopes. Samples have to be positioned within a confined space. Recent development of single objective light sheet microscopes like the ssOPM or SCAPE [129, 130], enable the usage of the standard cell culture plates due to the single objective setup, but lack spatial resolution in comparison to the dual inverted single plane illumination microscope (diSPIM) system and require a positioning of the spheroid near the bottom of the well [129, 130]. In comparison, Strnad and colleagues developed a now commercially available setup, with a v-shaped FEP channel, carrying up to twenty mouse embryos [115]. This setup allowed for high spatial and temporal resolution, but was lacking sufficient throughput and suffers from elaborated sample mounting.

In general, only one commercially available geometry represented by the ASI diSPIM system or the Luxendo QuVi microscope allow sufficient sample number as well as spatial and temporal resolution to facilitate a light sheet-based, high-content imaging screen of 3D spheroids.

1.3.8 Dual inverted single-plane illumination microscopy

Though originating in the field of developmental biology, novel commercial light sheet microscopy provides promising capabilities for many different research applications. The development of the diSPIM by Kumar and colleges [112] has provided for the first time a commercially available setup that allows to image in a horizontal cell culture setup. Although the nomenclature suggests otherwise, the diSPIM was conceived as upright add-on to an inverted confocal microscope. The dual view acquisition by two high magnification objectives

positioned at an angle of 90°, with a 45° angle between the objective and the stage, facilitate upon image processing an isotropic, high spatial resolution in samples of up to 130 μm in diameter for subcellular phenotypic analysis. The combination of an upright stage for standardized cell culture plates and the strong advantages of a LSFM system, makes the diSPIM system the optimal tool for a 3D high-content, live, high temporal and spatial resolution, quantitative screen of 3D cell culture models.

1.4 Quantitative imaging screen

In the last three decades, digital microscopy has evolved with sophisticated acquisition automation [126, 131], high quantum efficiency cameras [113, 132], and reproducible specimen labeling [100, 101]. To allow an impartial, comparative analysis of different specimens, a quantitative image analysis is key and the prerequisite for any imaging screen [133]. The comparison can be conducted between treated and untreated specimen or between tumor and non-tumorigenic tissue to detect communalities and differences between the samples. This analysis allows to characterize protein function [134], potential therapeutic drug targets [135], and effective drug combinations [7] as well as describe potential side effects of the experimental procedures.

1.4.1 From a digital image to cellular phenotype description

In a quantitative image analysis, samples are described by numerical descriptors, characterizing differential properties, which define the phenotype [136-138]. These can be features such as signal intensity, textural features, shape descriptors, quantities or differential values analyzed over time describing different biological processes [139]. Traditional machine learning methods like random forest (RF) classifier [140] or the advancing convolutional neuronal network classification [141-143] allow a stochastic determination of a phenotype class to group samples [7, 144].

A parallelization of many functional experiments with different treatments applied, different cell lines, different conditions, concluded with quantitative image analysis is called a high-throughput imaging screen. The high degree of statistical significance resulting from the high-throughput of many samples, allows to detect strong target candidates that evoke a desired phenotype, but also provide means to detect subtle differences between control and treated sample (Figure 5) [7, 145]. Numbers of differential treatments can range from a few tens to hundreds of millions of different treatments, as for example applied in drug discovery in pre-clinical trials [135]. Academic research and the private sector, for example pharmaceutical companies, have invested extensively in establishing small molecule inhibitor libraries to test

for different effects and their potential use as therapeutic [143, 146]. Other molecular tools such as siRNA libraries, or the discussed dCas9 fused to different effector domains and active CRISPR-Cas9 are nowadays also evaluated as potential methods to treat cancer cells and are already used to study the underlying cellular mechanisms that induce tumor initiation or can revert the tumorigenic, cellular phenotype [147].

A high degree of automation, reduced experimental volume per treatment and sophisticated analysis pipelines are essential to an establishment of such a high-throughput screen [148]. Furthermore, a high classification accuracy, rapid processing, and strong statistical quantification of the phenotype are essential to a resilient evaluation [149, 150].
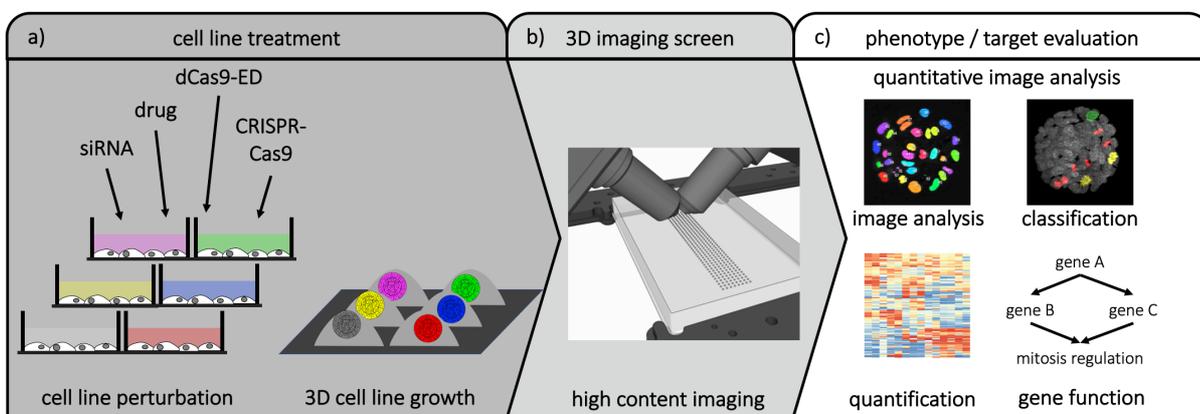
Figure 5 – Quantitative imaging screen of 3D cell lines for functional evaluation

a) Cell lines are treated with different molecular perturbation methods such as drugs, siRNA, CRISPR-Cas9 or dCas9-effector domains (dCas9-ED) to mediate specific modifications (color-coded). b) Spheroids are imaged in a high-content setup with multiple samples (different treatment / different cell line / different target) being acquired at the same time at adequate spatial and temporal resolution to follow the process or development of interest (e.g. mitosis). c) High-content quantitative image analysis pipeline with phenotype classification and comparative evaluation allows elucidating the function of the target protein.

1.4.2   Automated phenotype classification

Different machine learning (ML) algorithms have been deployed in life science research for many years to process, analyze, and evaluated biological data sets. For many years, random forest (RF) algorithms were the state-of-the art machine learning implementation to detect patterns in large data sets. Image classification by RF algorithms rely on quantitative values describing an image such as textural features (Haralicks features), sizes, pixel intensities, shapes or other measurable descriptors (Figure 6a), which require a preprocessing step of the images to calculate these quantitative values.

Deep neuronal networks (DNN) have gained in recent years an extensive rise in utilization in many everyday applications, such as voice recognition tasks or image processing [151]. A subgroup of DNNs is convolutional neuronal networks (CNN) which provide the basis for fast

and accurate image processing. In comparison, CNNs do not require extensive preprocessing of the images as compared to other ML classification algorithms as they analyze the entirety of the visual feature space (Figure 6b). These visual features can be for example the eyes, mouth and nose in face recognition software or in a biological related matter the state of the chromosomes throughout the cell cycle.

For research applications, the capability of CNN to process, segment, and classify large image data sets [150] at exceeding speeds displays highly promising capabilities [151, 152], especially in high-throughput imaging screens [149].

Inspired by stimuli processing of neurons in the brain [153], CNNs are complex mathematical models of many interconnected nodes, so called neurons ordered in layers (Figure 6b). Each neuron receives multiple excitation inputs from the previous layer, called activations. Activations are processed in the neuron through the weights and biases (Figure 6b). When the sum of all activations reaches a threshold, the neuron emits an activation signal to the next layer. The sum of all neurons, layers and their interconnections define the properties of the network and how an input is processed [154]. Weights and biases and therefore the properties of each neuron are adjusted by the training of the CNN by a large, labeled data set. Functions such as gradient descent or back propagation make iterative adjustments to the weights and biases to optimize a model. Underlying patterns of the input image, called features, are detected and define the output class of the input file.

Due to different intermediate calculations steps, of pooling multiple neuron layers and therefore subsampling the data, CNNs are highly robust and display low signal to noise interference [155]. Additional data argumentation of the training data by rescaling, rotation, introduction of random noise and signal variance enlarges the training data set [141]. This counteracts the problem of over fitting the CNN to the training data [156].

For classification of a small image in a DNN, already many computational matrix convolution calculations are needed [137, 154]. Therefore, CNNs require extensive computational performance power. Only in the last years, high performance graphic processing units (GPUs) have obtained sufficient calculation performance to do these calculation steps in parallel in a reasonable time frame, making DNN applicable to high-throughput image analysis steps in a high-content imaging screen for image classification [157, 158].

nuclei	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	F12	f13	image	class
1	0.08	2.36	0.82	6.60	0.67	16.63	12.64	3.64	4.88	1.62	1.77	-0.1	0.87		interphase
2	0.00	64.43	0.86	22.01	0.17	16.68	10.41	6.84	10.71	26.45	4.10	-0.2	0.94		prophase
3	0.00	35.20	0.73	64.04	0.19	75.26	51.84	5.90	9.44	13.34	3.69	-0.6	0.83		metaphase
4	0.00	65.58	0.74	15.03	0.15	87.80	72.61	6.39	10.32	25.29	4.11	-0.1	0.85		anaphase

b) random forest

- feature trees
- extensive pre-processing
- Input: quantitative value
- analytical
- user defined feature selection needed
- low computational demand

inter: 0.01
pro: 0.93
meta: 0.05
ana: 0.01

c) convolutional neuronal networks

- interconnected neurons
- low pre-processing
- Input: image / pixel value
- "black box"
- feature learning
- high computational demand

inter: 0.001
pro: 0.991
meta: 0.002
ana: 0.006

Figure 6 – Automated classification of mitotic phases by random forest or CNN

a) Today two main methods allow for automated image classification by analysis of quantitative features (f1 – f13) or by analysis of the image. Both require labeled training data to define the classes. b) Random forest consist of an ensemble of decision trees. These trees utilize defined quantitative properties of a nucleus such as textural features (Haralicks features), shapes, sizes, pixel intensities and others to define the image class. Appropriate feature selection for classification and training defines the analysis process. b) In comparison, CNN imports and processes full images and calculates the class from pixel values. Class defining features are self-taught during network training through weights and biases modulation. The CNN can be seen as a black box as the entirety of the network, the interconnected neurons (mathematical calculations) cannot be comprehended and requiring extensive computational performance.

1.4.3 Challenges of light sheet microscopy in high-content imaging screens

For years, the state-of-the art microscopy system to acquire large biological data sets at high repetition for screening purposes was the spinning-disk confocal microscope [159]. These systems are highly adapted for long-term imaging, high image quality, and culture stability as well as equipped with sophisticated adaptive optics-modules [126, 160]. As the field of light sheet microscopes develops, no or only early beta adaptations of adaptive optics technologies ensuring long-term image quality are available for light sheet microscopes [117, 125, 161] and require adjustment to the specific demands of the microscopy system. Furthermore, most of the light sheet microscope systems in research were custom made and tailored to a specific usage, making a general usage of light sheet microscopes challenging for high-content imaging.

2 Aim of study

Cell culture models and functional evaluation upon perturbation are fundamental tools to decipher gene function, cell function, development progresses, and tissue characteristics. 3D cell lines allow due to the established microenvironment a much more detailed analysis, resembling the *in vivo* physiology.

In this project, we wanted to establish a high-content imaging screen, with a high degree of process automation from cell culture treatment to the final analysis of the effect of the treatment. Through the usage of 3D cell culture models and fast, high-content, live light sheet microscopy as a screening method, we wanted to optimize treatment induced phenotype evaluation and reduce imaging induced phototoxic effects. Additionally, we wanted to harness the rapid imaging capabilities of light sheet microscopes to achieve unmatched sample numbers.

The workflow was intended to be applicable both in basic and medical oncology research. To display the capabilities in tumor research, we intended to apply the screen to study tumor initiation factors and induce by different means of gene regulation an abnormal, tumorigenic phenotype in a benign tumorigenic breast epithelia cell line. With the focus on mitotic gene knock-down, we aimed to induce abnormal cell divisions, incomplete genome separation, chromosomal instability or misdistribution of chromosomes into the daughter cells, which are strong hallmarks of cancer. We aimed to detect these phenotypes by high-throughput image analysis employing cutting edge deep neuronal network classification for high-throughput performance and exceeding classification accuracy. Additionally, we wanted to evaluate established and novel methods of gene regulation, such as the CRISPRi gene knock-down or dCas9-ED mediated gene regulation, and compare the cellular phenotypic effects of the different molecular gene regulation methods.

3 Results

The project was developed and conducted in two major discrete parts that together constitute a workflow for high-content perturbation analysis in 3D spheroids. First, we established a high-content imaging screen utilizing advanced light sheet 3D microscopy to acquire images of spheroids at high temporal and spatial resolution. In the second part, we applied this workflow to MCF10A and HEK293 3D spheroids and analyzed the effects of epigenomic or transcriptomic modifications on global and cellular phenotype. The quantitative image analysis of the induced cellular and spheroid characteristics allowed an evaluation of target gene function in spheroid development and clonal growth as well as epigenetic regulatory mechanisms.

3.1 3D cell culture imaging by diSPIM

3.1.1 Comparison of light sheet and spinning disc microscopy

With the development of the commercial dual inverted single plane illumination light sheet microscope licensed by Applied Scientific Instrumentation (ASI), the advantages of light sheet microscopy were for the first time available for high-content imaging of 3D cell cultures in standardized cell culture environments.

To evaluate the imaging parameters and improve image quality in comparison to high-content screen (HCS) microscopy systems, we acquired images of six day old MCF10A H2B-GFP spheroids with a diameter of about 80 μm with a spinning disc microscope and the diSPIM light sheet system. The stable labeling of histone subunit 2B in the MCF10A cell line with a GFP fluorophore enabled us to track the DNA throughout the cell cycle as well as image nuclear morphology.

Following the suggestion by Cole and colleagues [162], we decided to acquire with the diSPIM at a three-fold Abbe resolution limit oversampling in all axes, which would optimize image quality, but also increase data size, single file acquisition time, and light exposure. In total, a single position, single time point, dual view (90° rotated view) acquisition with a single channel (excitation wavelength), from now on termed "position scan", comprised 520 slices (260 slices per view) each slice with 1024 x 1024 pixels and a resolution of 0.1625 pixel / μm, resulting in a position scan data size of 1.1 gigabyte (GB). When the two orthogonal views per position scan were aligned, a total volume of 130 μm x 130 μm x 162.5 μm could be acquired from both views.

For comparability, the SDM system was set to acquire at the same spatial XYZ resolution and maximal signal intensity. We optimized the excitation path settings of the diSPIM to form a

Gaussian beam with a threefold larger YZ-diameter at the FOV border, in relation to the YZ extend at the image center giving us the highest optical sectioning at the X-axis center.

In comparison to the current spinning disk microscope system used for live, high-content, 3D imaging screens, the diSPIM setup outperforms the conventional SDM setup in almost every acquisition parameter such as laser power density, acquisition duration and signal-to-noise ratio (Table 1), though acquiring at a higher spatial resolution and shorter exposure time.

The biggest advantage of the light sheet microscope over the SDM became apparent when the region of interest is located deeper inside the tissue or spheroid (Figure 7). Due to signal scattering as the photons travel through the tissue, we were not able to acquire the strong signal of the MCF10A H2B-GFP nuclei deeper than 80 μm inside the sample with the SDM. In comparison, using the same specimen, we were able to image throughout the entire spheroid with the diSPIM system. By applying image processing and combining visual information from both optical paths, we were able to achieve exceptional isotropic image quality, essential for quantitative image analysis of the entire spheroid (Figure 7).

Table 1 – Comparison of spinning disc and diSPIM imaging parameter

Quantitative comparison of acquisition properties and resulting image quality between spinning disc microscopy and diSPIM systems. Acquisition parameters were set to generate Z-stacks with close to identical Z-stack acquisition properties. Imaging parameters listed describe acquisition parameter, light input (measured at the sample) and acquired image properties.

	spinning disc microscope	diSPIM
XYZ stack (px x px x n_slices)	1004 x 1002 x 233	dual (1024 x 1024 x 260)
XY pixel resolution	0.2 µm / px	0.1625 µm / px
measured laser power	1,320 µW / s	320 µW / s
exposure / slice	50 ms	1.75 ms
stack acquisition duration	53.2 s	4.5 s (+ 15 s processing)
signal-to-noise ratio	43.25	127
avg. background signal	18.4	0.941
power density / phototoxicity	168,000 W/cm2	40,700 W/cm2
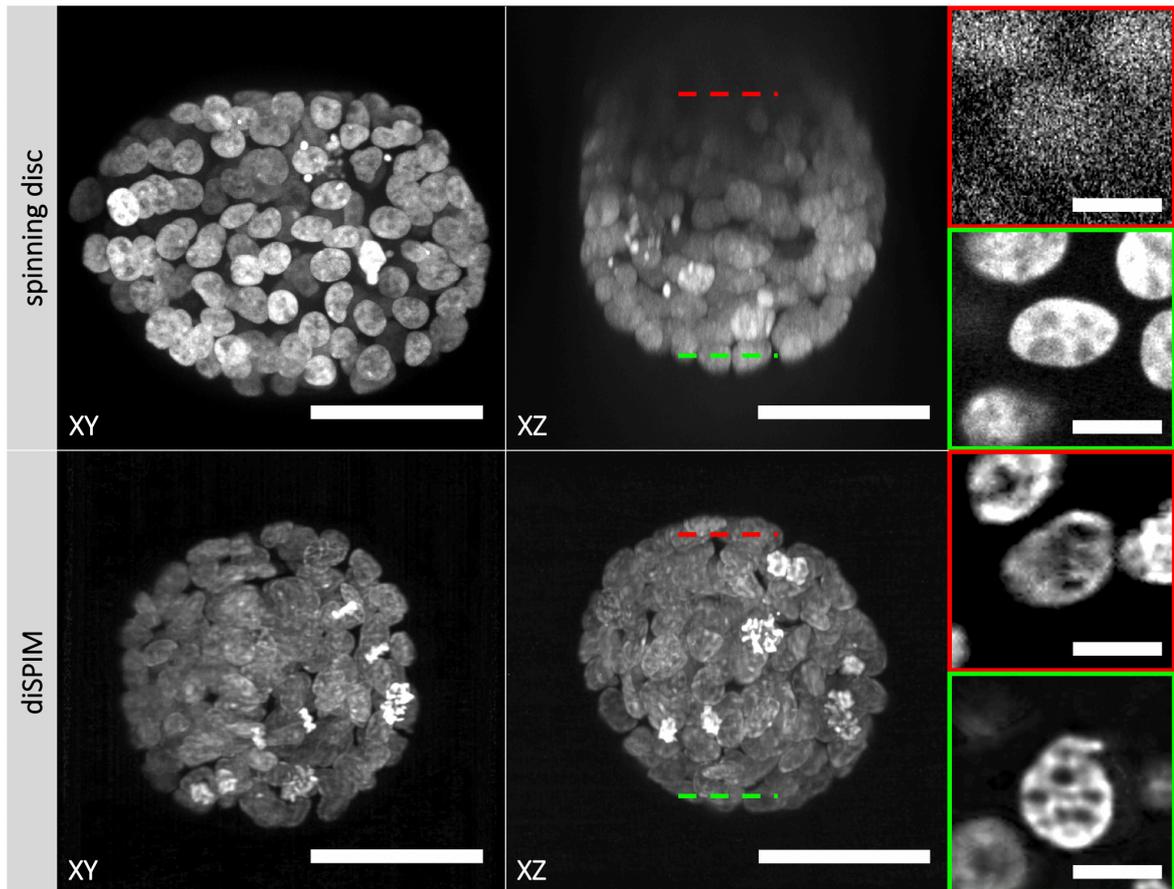
Figure 7 **– Comparison of spinning disc and light sheet imaging of 3D spheroids**

Direct comparison of spinning disc and light sheet imaging performance of MCF10A histone 2B – GFP spheroids acquired six days after seeding of single cells in Matrigel with a size of about 80 μm in diameter. XY and XZ maximum projection of full 3D stack (scale bar = 50 μm) represent the XYZ acquisition capabilities of SDM and the diSPIM. Inserts represent single nuclei close to the detection objective (green box) or imaged 80 μm inside the sample (red box) (scale bar (insert) = 10 μm), with insert Z-stack slice depicted by red and green line.

3.1.2   MCF10A – a valuable model for siRNA knock-down studies

For this project, we decided to use mainly the MCF10A benign cell line as a model to establish the high-content perturbation analysis workflow utilizing light sheet microscopy technology for phenotype evaluation. Single MCF10A cells grow clonally under 3D culture conditions to multicellular spheroids, following a defined developmental trajectory (Figure 8a), which allowed us to study the function of single genes in development and cellular growth by siRNA-mediated knock-down.

We used an MCF10A cell line endogenously expressing H2B-GFP proteins that label the DNA [163]. This cell line allowed us to follow and describe in detail single nuclei and DNA properties throughout the different stages of the cell cycle, as well as analyze the development of the full spheroid (Figure 8b). Furthermore, the MCF10A cell line is highly susceptible to the transfection of siRNA and therefore allowed an effective influence on cellular homeostasis. Initial experiments showed that siRNA solid phase reverse transfection of PLK1 targeting

siRNA evoked in 94% of all MCF10A cells a metaphase cell cycle arrest, while only 43 % of HEK293 cells displayed this phenotype (Figure 21a). PLK1 is a serine/threonine-protein kinase essential for the centrosome maturation and spindle assembly as well as the removal of cohesins from chromosome arms. A knock-down of PLK1is known to arrest the cell cycle in prometaphase.

The high transfection efficiency of siRNA into MCF10A cells allowed us to use this benign cell culture model in a HCS as only few cells would not be transfected. Therefore this approach provided us with a highly reliable readout of the knock-down phenotype. For the low transfection efficiency of HEK293 cell lines, many analyzed spheroids would have been untransfected and therefore alter the phenotype quantification by showing wildtype properties.

3.1.3   Long-term imaging of 3D spheroids by dual inverted light sheet microscopy

With the established imaging setup and acquisition parameters, we achieved high image quality of our 3D cell culture model (Figure 7), while administrating a low light dose onto the spheroids (Table 1). The non-invasiveness of the imaging can best be described by the fitness and viability of the imaged MCF10A spheroids throughout high repetition, long-term imaging (Figure 8a).

We were able to image a wild type MCF10A H2B-GFP spheroid, stained with a tubulin live dye (SiR-Tubulin) in dual color throughout the entire development from the two-cell stage (five hours post seeding) to a full-sized spheroid at an interval of ten minutes for up to six and a half days (155.6 hours) (Figure 8a). We could detect cell divisions and cell motility throughout the entire time lapse. With the ten-minute time interval, all stages of the mitosis are visible (Figure 8b).

Throughout the acquisition, we only needed to do minor position adjustments to some of the acquired spheroids after two days of constant imaging, highlighting the high position stability of the microscope, while traveling large distances between different positions. Spheroid movement was mostly caused through shrinkage of the Matrigel with an average displacement of 13.97 μm in X, Y and Z over 24 hours of imaging at a position approximately 200 μm above the plate (Supplementary Figure 1).

For the purpose of phenotype evaluation upon target gene knock down, we were not able to image the full development of six days of each spheroid due to data size constrains. We therefore selected day five to six for the 3D high-content screening approach. Spheroid "imaged" at a five-minute interval for 24 hours (avg. 62.25 nuclei count) during this defined time frame, show the same increase in nuclei total number as "non-imaged" spheroids (avg. 57.7 nuclei count), analyzed by start- and end-point evaluation (Figure 8c)
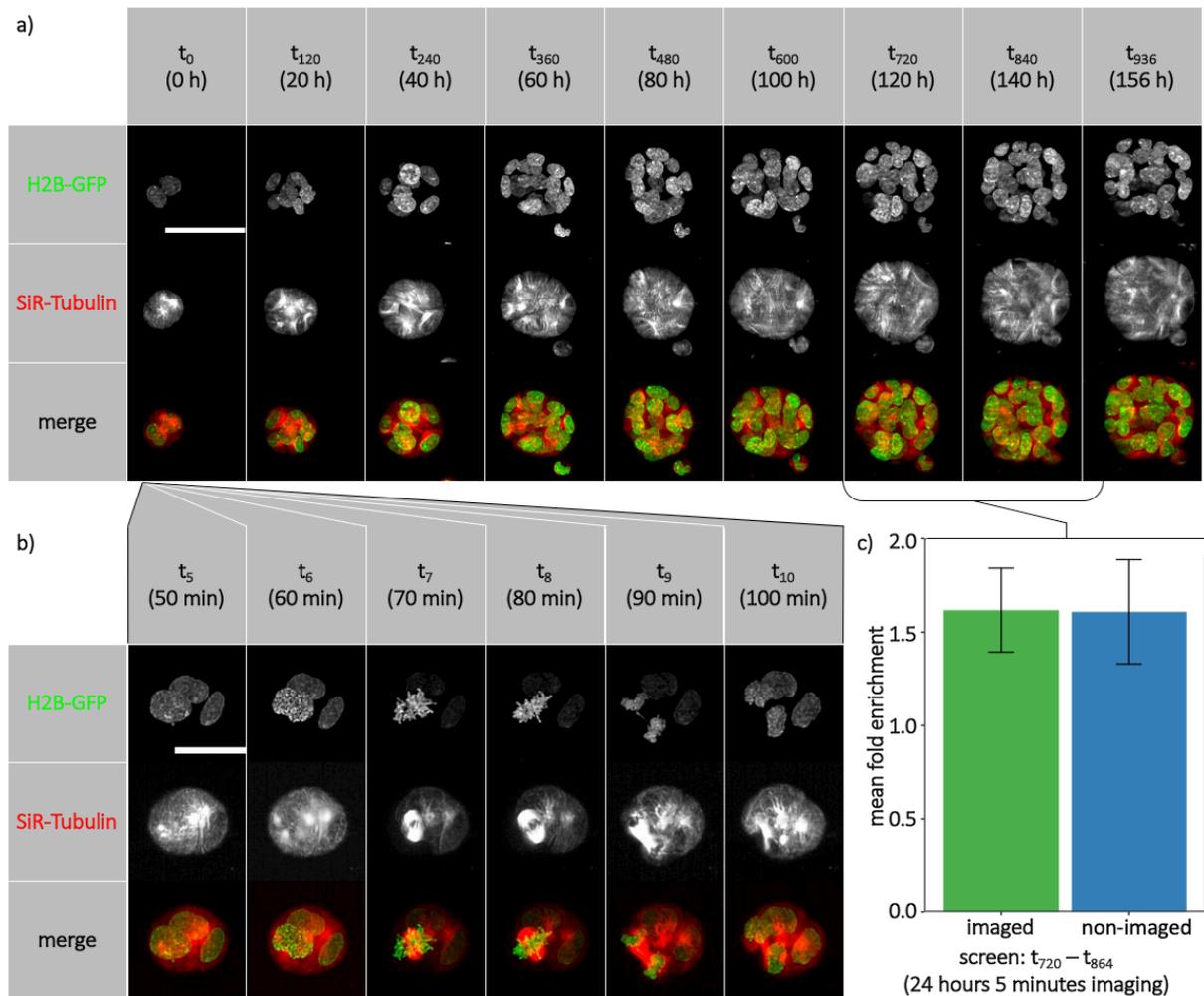
Figure 8 – Long-term imaging capabilities of the diSPIM

a) Example untreated MCF10A H2B-GFP spheroid imaged over 156 hours / 936 time points (t) every 10 minutes from the two-cell stadium to the fully developed spheroid in dual channel (H2B-GFP / SiR-Tubulin dye) (scale bar = 50 μm). b) High temporal and spatial resolution enable a detection of the different cell cycle stages (pro-, meta-, anaphase), while distinct features of the cytoskeleton could be detected (scale bar = 25 μm). c) Nuclei count mean fold enrichment of imaged and non-imaged spheroids in 24-hour acquisition cycle representing the developmental stage of the spheroids during the HCS acquisition time frame of MCF10A spheroid development (n_{imaged} = 31 / $n_{non-imaged}$ = 33, error bars represent standard deviation).

3.2 High-content light sheet imaging screen

In this project, we aimed to utilize the capabilities of the light sheet microscope to analyze and compare multiple different treatments that influence the cellular phenotype associated with tumor development.

3.2.1 Target selection

As imaging screens rely on visual features, we decided to focus on perturbing the cell cycle, since alteration in this key cellular process can be visualized by DNA labeling and as described in 1.1.4 are a regular cause for cancer initiation.

For phenotype validation and comparative analysis, we used different methylome-modifyingifying effector domains fused to deactivated CRISPR-Cas9 (dCas9-ED) for epigenome modulation. In comparison to RNAi, the dCas9-ED relies on a protein component (Cas9 fused to a methylome-modifyingifying effector domain) and an RNA component (single guide RNA) that need to be introduced into a single cell to induce epigenome alterations (1.2.2). Both of these methods can be applied to induce a specific transcriptome alteration, targeting only single genes, which allowed us to compare these methods by the phenotype.
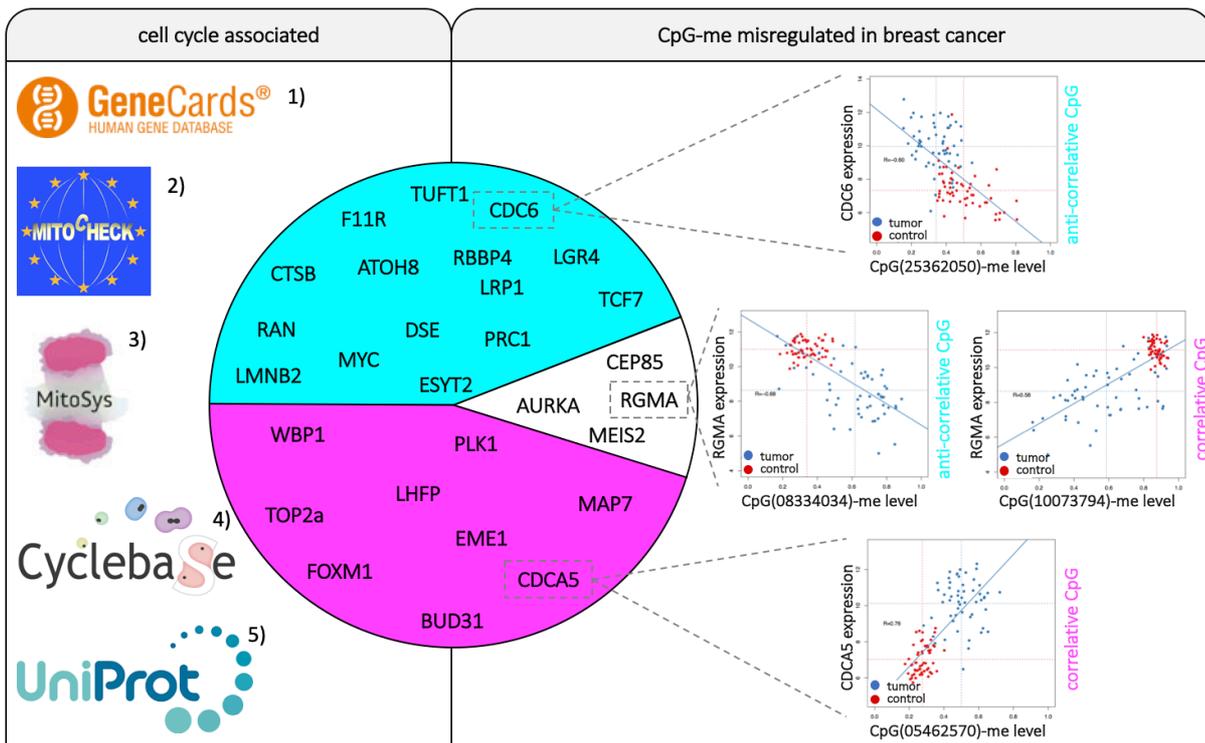
Figure 9 – Target gene selection

Target genes for the screen were selected by their role in cell cycle progression, as described by multiple online gene data bases (1. genecards.org / 2. mitocheck.org / 3. mitosys.org / 4. cyclebase.org / 5. uniprot.org) and a strong correlation of CpG-me level and gene misregulation. This allows a comparison of the knock-down phenotype induced at translational level in comparison to the knock-down induced at the transcriptomic level. Regulatory CpGs can be correlated (low CpG-me level = reduced gene expression (magenta)) or anti-correlated (high CpG-me level = reduced gene expression (cyan)) with expression of the target gene. Single target genes, could display both types of CpG-me relations - correlated and anti-correlated regulatory CpGs such as found in AURKA, CEP85, MEIS2 and RGMA (white area). Example of correlated and anti-correlated CpG displayed for CDC6 (one of two anti-correlated CpGs), RGMA (one of four anti-correlated CpGs and one of three correlated CpGs) and CDCA5 (one of seven correlated CpGs). Tumor samples are depicted by a blue dot, while control (non-tumorigenic) tissue measurement is marked by a red dot.

Due to the fact that siRNAs can be designed specifically for any target gene as well as the sgRNA for a site-specific dCas9 targeting of the methylome-modifying effector domain, it was essential to select for target genes involved in mitosis with distinct knock-down phenotypes and most importantly show a strong correlation of gene expression an alter methylation level

at defined regulatory, genomic regions. With the modulation of the methylome at these regulatory sites, we expect to induce in comparison to the siRNA regulating at translational level, a knock down through the modulation at the transcriptomic level (1.2.1).

In collaboration with Dr. Carl Herrmann, we bioinformatically selected 28 target genes that matched these criteria. These target genes displayed either in the promoter region or a distal regulatory region one or multiple CpGs that displayed a high correlation (> 0.5) or anti-correlation (< -0.5) between CpG methylation level and gene expression (Supplementary Tables 1 / Figure 9). The correlation was detected by comparison of tumor- and non-tumorigenic patient tissue CpG methylation level and gene expression data from a 450k Illumina breast cancer patient screening data set [164] available on UCSC genome browser [165].

3.2.2 High-content 3D light sheet screen workflow

The fundamental aim of this work was the establishment and biological evaluation of a high-content live imaging screen that can be used in 3D cell culture and efficient perturbation assays and tumorigenesis studies. Therefore, we optimized the workflow comprising all steps of a HCS from sample preparation, to high-content imaging, to image analysis described in detail in Figure 10.

To conduct an imaging screen using the diSPIM system, we planned to evaluate the phenotype of many different modified samples simultaneously, with each spheroid sample grown from a single treated cell. The manual preparation of each sample by hand would have required an intensive manual effort and led to small individual differences in transfection mix, spot size, cell seeding density and preparation duration, which in combination would have corrupted the outcome of the screen.

Due to these reasons, we established a solid-phase reverse transfection mix procedure, optimized for maximal siRNA and plasmid transfection efficiency (Figure 21a) of the different cell lines (Figure 10): Following transfection in 2D in a 96-well format, we mixed the cells with Matrigel and spotted the cell-Matrigel mix of all samples simultaneously with a liquid handling robot. The usage of a robot further reduced manual labor, reduced spot volume to a minimum of 0.2 μl, and positioned samples precisely in a defined pattern.

Furthermore, the usage of the Hamilton liquid handling robot for sample spotting facilitated defined pre-screen spheroid position detection as well as reduced the duration of the sample mounting in the imaging dish to 30 minutes. Subsequently, spotted samples were relocated to the incubator for spheroid development. Five days into 3D spheroid development under standard culture conditions, we detected in a low-resolution diSPIM pre-screen each

spheroid's position within each of the 320 Matrigel spots as well as general properties such as spheroid size and circularity. These features allowed us to automatically select the maximal number of 38 spheroids to be imaged at five minutes interval for the subsequent HCS by the optimal imaging position at the apex of the matrigel spot, while eluding 2D cell layers. Additionally, we defined manually two dual view stack positions of beads immersed in Matrigel essential for point spread function (PSF) detection and dual view registration, giving a total of 40 positions acquired per HCS at a five-minute interval reaching close to the maximum imaging performance of the diSPIM system.

3.2.3   3D spheroid diSPIM imaging screen of siRNA mediated knock down phenotypes

In total, we used for each of the 28 target gene two distinct siRNAs (siRNA set 1 and set 2) in three replicates, giving us a total of six perturbation samples per target gene. These six spheroids were subsequently imaged by high-content light sheet imaging. Together with the positive and negative transfection control we acquired a total of 228 spheroids in six individual experiments.

3.2.4   Image processing pipeline

The data size produced per HCS of ~10 terabyte (TB) (total of 60 TB in six experiments) raw data challenges standard IT infrastructures with regard to storage capacity and data processing. Therefore, we developed a data processing tool called "hSPIM", which ran directly on the workstation controlling the diSPIM system (Figure 10, Figure 11a). This tool avoided the need for a large data set transfer infrastructure or high performance cluster calculations. HSPIM combines the signal information of the two diSPIM views to calculate an estimate of the most probable image, given the point spread function provided by the bead reference image. As signal quality degrades deeper in the sample tissue due to light scattering (1.3.2), the second view was essential to the screen to compensate for signal loss in the large spheroids and to provide isotropic resolution to display the spheroid samples.
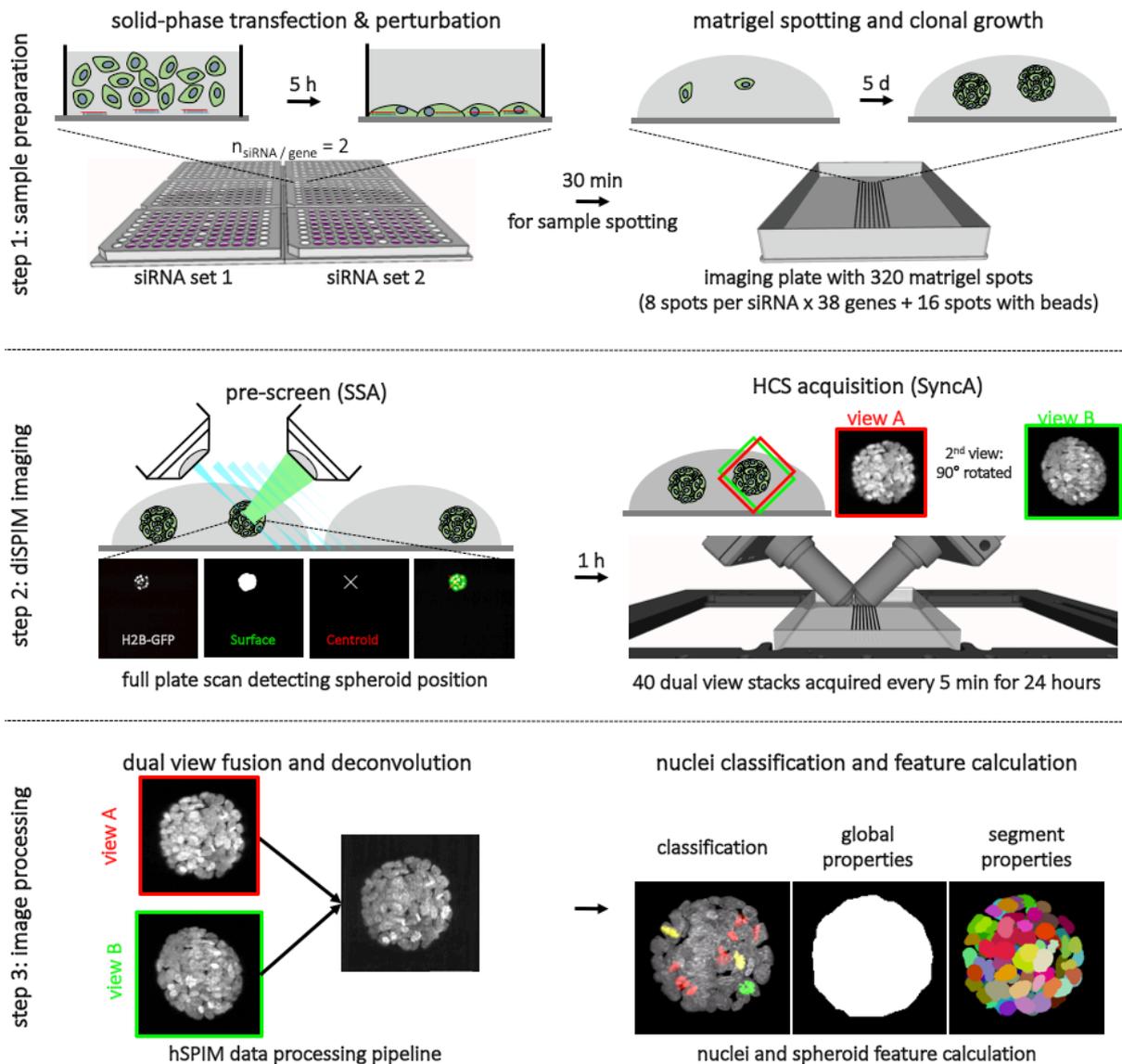
Figure 10 – siRNA konock-down high-content light sheet screen

There are three key steps of the high-throughout light sheet imaging screen:

Step 1 – sample preparation: solid-phase reverse transfection of cell line in 2D in a 96-well format with two different siRNAs (siRNA set 1 and siRNA set 2) per target gene transfected separately. A total of three complete HCS were conducted, resulting in six acquired samples per target gene. Treated cells were subsequently mixed with Matrigel and spotted onto a one-well imaging plate. Over five days spotted cells clonally expand into 3D spheroids growing in the Matrigel spots.

Step 2 – diSPIM imaging: In a low-resolution stage scan pre-screen, positions of all spheroids were detected. Samples for imaging were selected by position in Matrigel spots, while unsuited spheroid clusters of two or more fused spheroids as well as 2D cell layers were excluded. Subsequently, 38 individually treated samples were imaged for 24 hours every 5 minutes by dual view light sheet imaging acquiring two full stacks of view A (red) and view B (green) at a 90° angle.

Step 3 – data processing: Raw data was processed by image deconvolution and fusing visual information of view A and B together. Additionally, the image bit depth was reduced to 8-bit. This processed data was subsequently used to analyze the phenotype of each spheroid throughout the entire screen in regard to global spheroid properties as well as single segment (single nuclei) features and development.

With the available hardware and most calculations executed in parallel on a high-performance GPU, we were able to process the raw data of one position stack file within 8.23 seconds which reduced the data size from 1.1 GB to 80.1 megabyte (MB) per position scan with significantly improved XYZ isotropic resolution and signal-to-noise ratio (Figure 11a). Because the background signal from the surrounding Matrigel could be deducted and foreground signal of the nuclei were automatically detected, a reduction from a 16-bit to 8-bit TIF file format did not result in significant information loss, as well as a two by two-pixel binning, enabling the strong decrease of total data size from 1 GB to 80 MB per single time point position stack.

The hSPIM tool is in its principle based on the multi-view deconvolution developed by Stephan Preibisch and colleagues [128, 166]. Novel in our SPIM image processing approach was the separation of beads and sample, meaning that the registration matrix and PSF were detected in a dedicated spot of beads mixed in Matrigel (Figure 11a). PSF and registration were subsequently applied to all other processing calculations of sample images.

Even for large distances between individual samples of up to 10 centimeters, the transfer of the registration matrix and PSF was highly successful. When we applied the PSF and registration of bead position 0 (Pos0) onto a secondary bead position 39 (Pos39) at the end of a multi position scan cycle with long traveling ranges of the stage, we found a displacement between the two registration matrixes of less than 162 nm (Supplementary Figure 2). The transfer of registration matrix and PSF from one position onto others thus produces accurately registrated and deconvolved images throughout a full screen in time and space. Thus, the approach of separating bead and spheroid positions reduced the background signal in the images, avoided the need to subtract the beads from the output image and reduced the necessary steps to a single calibration calculation per time point, while retaining optimal quality for all other sample images (Figure 11a).

The hSPIM pipeline provided additionally to the fusion image the possibility to segment nuclei and calculate texture features for each nuclear segment, which subsequently can be used for segment classification and a quantitative analysis.

Computation duration in the given hardware framework (6.3.1.1) took 35.6 seconds per time point to detect PSF and registration matrix, while image fusion and deconvolution required 4.8 seconds and the segmentation and feature extraction an additional thirteen seconds of computation per position scan. Due to the parallel processing capabilities of the modern GPUs with five files being processed in parallel the total duration for the hSPIM pipeline was on average eleven hours and fifteen minutes.
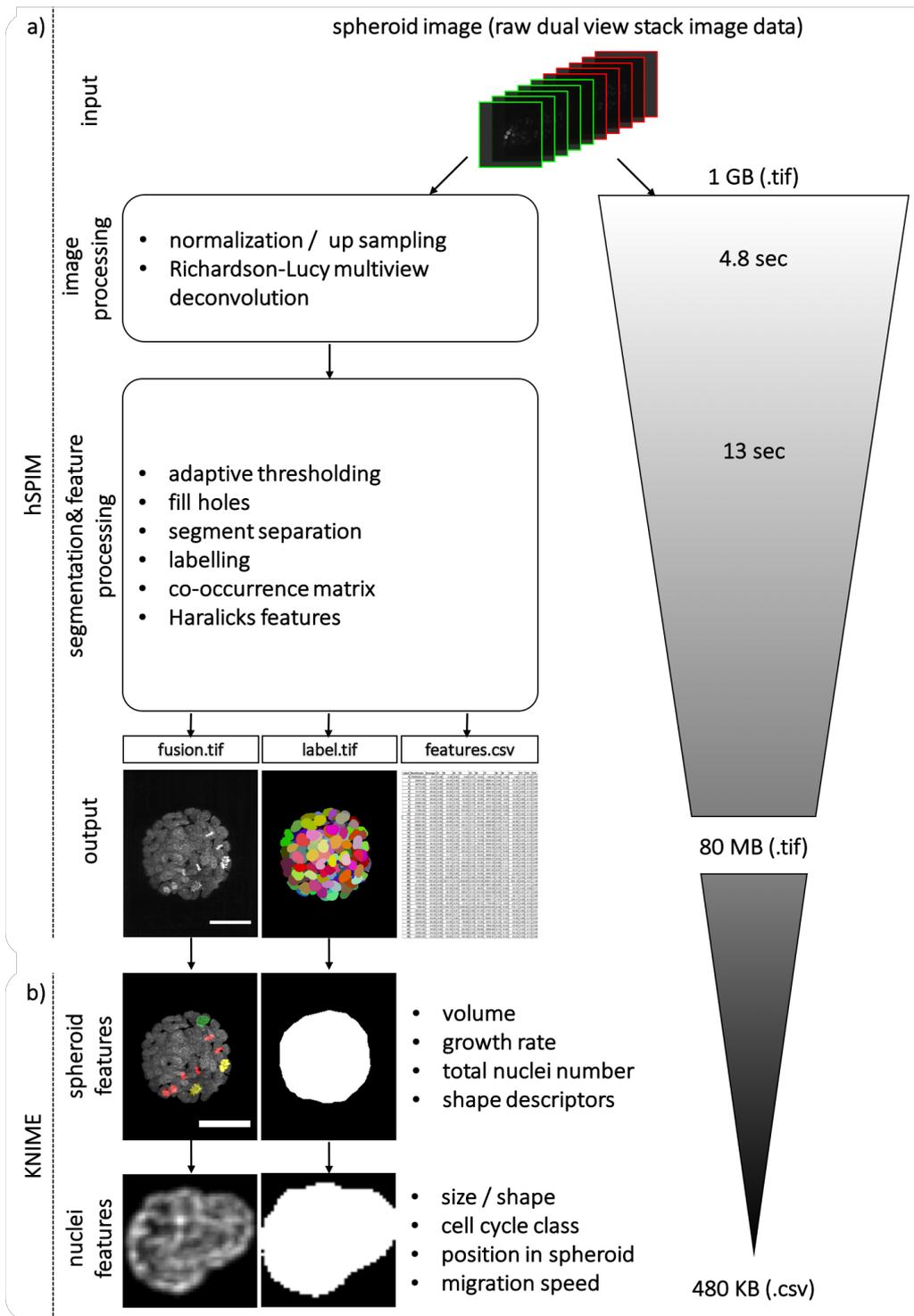
Figure 11 – diSPIM data processing and analysis

Raw diSPIM imaging data with 1 GB data size was processed and analyzed quantitatively. a) hSPIM is a processing pipeline developed specifically to diSPIM acquisition parameters and data structure. In a calibration step, the registration matrix and the PSF are detected in the dedicated bead position. These are subsequently applied to process all sample positions by Richardson-Lucy multiview deconvolution. From each processed image, nuclei are detected, segmented, and for each segment textural features (Haralicks features) are calculated. The processed image (fusion), the features and the label image are stored as output reducing the data to 80 MB per spheroid and time point. b) KNIME processing pipeline for high-content quantitative analysis uses the hSPIM output data to analyze cellular and global features of the spheroid to evaluate and compare the phenotype of all samples resulting in a quantitative description of different global spheroid and single nuclei features stored in a 480 KB sized CSV file (scale bar = 50 μm).

The output of the hSPIM software package was subsequently imported to a pipeline implemented in the Konstanz Information Miner (KNIME) software [167] to compute key quantitative properties of the development of the spheroid and each single nucleus (Supplementary Tables 2). Together, these properties described in detail the individual spheroid's phenotype and put cellular and subcellular processes into the context of the development of the entire spheroid. Differential phenotype analysis of non-targeted control siRNA and target gene knock-down spheroids allowed a functional evaluation of each gene of interest.

3.2.5 Cell cycle stage classification with a deep neuronal network

The key feature of the KNIME image analysis pipeline was the detection of the prominent four cell cycle stages (inter-, pro-, meta- and anaphase) of every single nucleus. To define the cell cycle stage of each nucleus throughout the entire screen, we tested two different means of automated image classification: a random forest classifier and deep learning image classification using a convolutional neuronal network (Figure 12).

For RF nuclei classification, we used the thirteen textural Haralick features calculated by the hSPIM light sheet data processing software (Figure 12a). Our RF model was trained with these features taken from 1,643 manually classified nuclei. With this training set and our computational hardware, we achieved a classification accuracy of 83.1% and 4.25 seconds computation time per nucleus (Figure 12c).

In comparison, DL classification performed much faster with only 3.7 ms classification calculation time per image with the applied VGG-based convolutional neuronal network [154] (Supplementary Figure 3). However, using the maximum intensity projection image of the same 1,643 nuclei as for our RF model training as input for the DNN training resulted here in a considerably lower prediction accuracy of only 74.75% as compared to the RF classification (Supplementary Figure 4). Nonetheless, the accuracy could be dramatically increased by increasing the training data set and most notably through the implementation of the "3D-2D plane classification" (Figure 12c). In developed 3D-2D plane classification, we preprocessed each 3D nucleus stack individually and used the 2D slices of the XY, XZ, and YZ orientation for training and later for segment classification (Figure 12a). This preprocessing increased the number of images representing each nucleus from a single maximum projection to about 60 images in three orthogonal view angles.

Predominantly, we established 3D-2D plane DL classification to avoid misclassifications caused by mitosis orientation or maximum projection artifacts. Cells in 3D cell culture divide in all 3 dimensions, unlike in 2D cell culture, where the culture dish defines the orientation of

division and therefore defines the visual traits of the different cell stages captured by the microscope. Only the inter- and prophase are in their respective 3D visual representation rotation invariant.

With this 3D-2D plane classification preprocessing step implemented, we achieved a much higher classification accuracy of 95.8% and computation duration of only 88 ms per nucleus classification (Figure 12b-c).
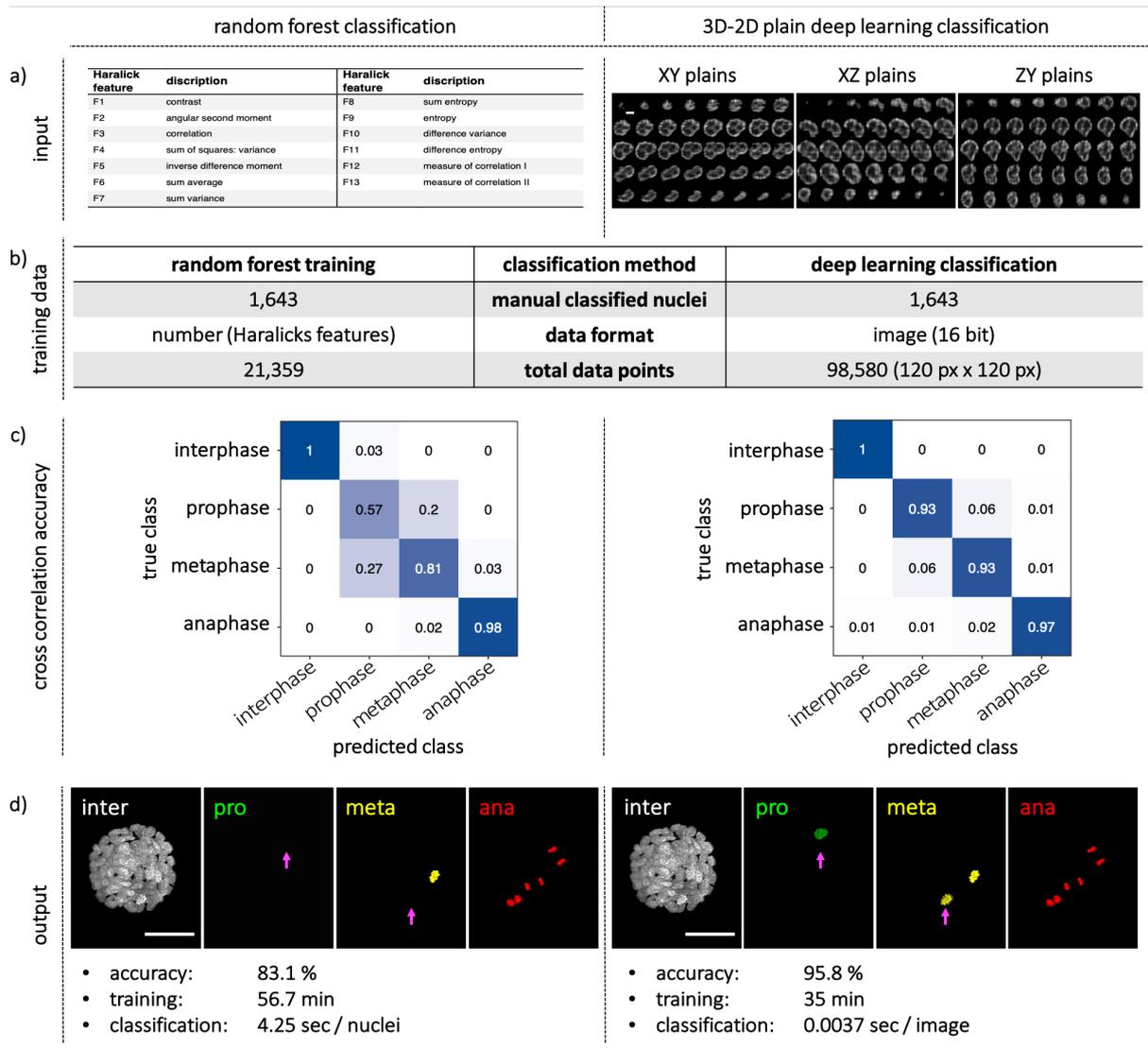
Figure 12 – Prediction accuracy of machine learning classifications

a) As input we used the 13 Haralicks features (F1-13) calculated by the hSPIM diSPIM data processing pipeline for random forest classification and 2D images of the XY, XZ, and YZ slices of a 3D nucleus for the deep learning classification, respectively (scale bar = 5 μm). b) The different input formats of the same nuclei (n=1,643) manually classified results in 21,359 data points used for the RF classifier training, while the 3D-2D plane deep learning classification training was conducted with a total of 98,580 16-bit images. c) Cross correlation measurement describes the classification accuracy and allows an evaluation of potential misclassification. d) In direct comparison, classification differences between RF and 3D-2D plane DL classification can be visually detected (magenta arrows), when applied to the same data (scale bar = 50 μm).

3.2.6 Control MCF10A spheroid cell cycle under screening conditions

The high accuracy of cell phase detection by DL classification paired with the high temporal resolution of the acquisition allowed us to follow a single cell through the different phases of the cell cycle (Figure 13a), while the high isotropic spatial resolution allowed us to calculate the ratio and size reliably of the different cell cycle stages (Figure 13b-c). In the control sample transfected with a non-targeted siRNA, 94.7% of all nuclei throughout the time lapse were classified as interphase, 0.9% as prophase, 1.2% as metaphase and respective 3.2% as anaphase (Figure 13b / Table 2). When we tracked the average duration of the different cell cycle phases, we detected a similar mitotic phase distribution with the prophase progressing for an average of 13.3 minutes, the metaphase for 11.25 minutes and the anaphase for 13.75 minutes (Figure 13a / Table 2) resulting in an average of 38.3 minutes for a cell to transition through the four cell cycle stages. Due to the time resolution of five minutes' acquisition intervals a more precise determination of the duration was not possible, but similar stage duration times with a total of ~ 50 minutes were also detected by Cai and colleagues [168] in HeLa Kyoto and U2OS cells. From the individual segment size, we detected that the prophase segments were on average the largest with 28,311 voxels, followed by inter- and metaphase nuclei with a size of 17,238 voxels and 16,865 voxels, respectively. Anaphase nuclei were the smallest on average with 9,660 voxels per segment (Figure 13c / Table 2).

Additionally, we analyzed the linear regression of the ratio of the cell phases throughout the screen time lapse. On average, none of the cell phases of the control spheroids gained or lost significant representations with differences in ratio between start and end of the acquisition of less than 0.01% (Table 2).

Furthermore, we evaluated the position of each segment in relation to the center of the spheroid, as it is often speculated that position within the tissue influences cell cycle localization [169]. In contrast, we could not find a clear predominant localization of the pro-, meta- and anaphase relative to the surface of the spheroid, as all classes are homogenously radially distributed in our MCF10A spheroids at day six of clonal development (Figure 13d).

Table 2 – Control sample properties throughout 24 hours development

developmental feature	interphase	prophase	metaphase	anaphase
average class percentage (%)	94.7	0.9	1.2	3.2
class duration (min)		13.3	11.25	13.75
segment size (voxel)	17,238	28,311	16,865	9,660
segment size (ratio to interphase)	1	1.64	0.98	0.56
gain / loss class percentage (%)	-0.0062	-0.0027	0.0003	0.0038

Throughout the 24-hour imaging time course, control spheroids gained an average of 16.5 additional nuclei and grew by an average of 46.4% in volume (Figure 13e-f / n = 12 control spheroids).
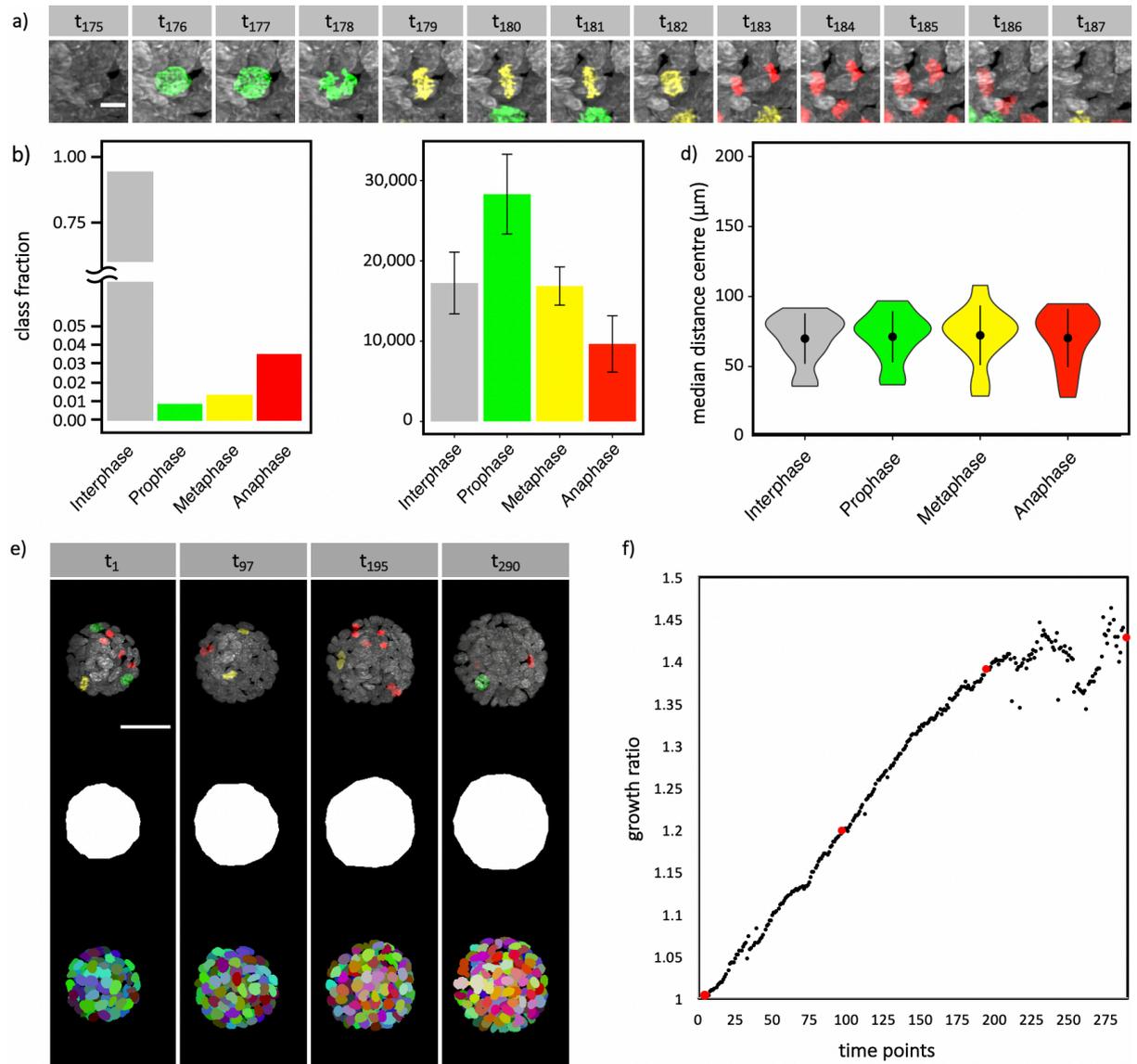
Figure 13 – Image analysis of processed diSPIM data of control spheroid

Global spheroid and single nuclei properties of the negative control samples throughout the HC diSPIM imaging screen. a) Example time lapse (t (time point) 175 – 187) of a cell undergoing mitosis with the main stages of the mitosis detected by deep learning image classification (scale bar = 5 μm) with inter- (white), pro- (green), meta- (yellow), anaphase (red). b) Bar plot displaying total class fraction of the different cell cycle stages throughout the HCS (n = 205,068). c) Bar plot showing average class segment size and standard deviation from voxel count (n = 205,068). d) Violin plot describing class, distance and number of segments in regard to their distance from center of the spheroid (n = 205,068). The black dot describes the median and the whiskers the 25%-75% interquantile range e) Four exemplary time points (t) of the spheroid development imaged over 24 hours (time point 1 – 290) with the classified cells, hull and segment maximum projection displayed (scale bar = 50 μm). f) Average growth ratio of all control samples (non-targeted siRNA; n = 12) displayed in dot plot with the time points displayed in e) highlighted in red.

When we analyzed all cell cycle transitions of each of the nuclei in the control samples, where we expect a biological cell to transition throughout the cell cycle from interphase to prophase to metaphase to anaphase, we detected a ratio of 99% cell cycle transitions that are following this normal transition pattern (normal = 146,101 / abnormal: 1,380 cell cycle transitions), again displaying the high classification accuracy of the deep learning (Figure 12) classifier exemplified in a mitotic cell in Figure 13a and the low effect on the cellular viability during light sheet imaging. We observed that single dying cells display a high degree of abnormal transition patters as the apoptotic condensed DNA (ACD) is regularly classified as pro- or anaphase (Supplementary Figure 5).

3.3 Detection of aberrant mitotic phenotype by automated visual feature evaluation

3.3.1 Screening for siRNA-mediated phenotype induction

In total, we acquired 3D time laps images of six spheroids per target gene (Supplementary Tables 1) of each of the 28 selected target genes. To detect abnormal mitotic phenotypes upon gene knock-down in each of the spheroids, we used the different visual features (Supplementary Tables 2 / Table 2) to characterize the individual spheroid phenotype. We evaluated each spheroid independently and combined all 23 features to rank and group phenotypes (Figure 19e-g). We used INCENP as positive transfection control as the siRNA-mediated knock-down of this gene results in a severe macronuclei phenotype that can be easily detected and distinguished (Figure 15b). Non-coding siRNA was used as negative control.

3.3.2 Aberrant mitosis duration

Cell cycle arrest or an elongated mitosis are severe phenotypes describing strong perturbations in cell growth and development. To detect an abnormal mitosis duration or a prevalent cell phase, describing a possible cell cycle arrest, we analyzed the relation of the different cell phases for each sample on average throughout the entire time lapse (Figure 14). The non-coding siRNA transfected negative control displayed on average 5.3% of all cells at a mitotic class of the cell cycle (Figure 14a). The INCENP positive control on the other hand displayed a strong increase of the prophase class to 46%, while the number of nuclei in interphase decreases to 44% (Figure 14a). Other gene knock-downs such as PRC1 and RGMA also induced an increase in the prophase class, although less pronounced than INCENP knock-down (Figure 14b). PLK1 and ESYT2 knock-down spheroids displayed an increase in all mitotic classes, indicating an elongated mitosis (Figure 14c). MYC and ATOH8

knock-down display an increase in anaphase classed segments, possibly describing an increase in apoptosis (Supplementary Figure 5) or cell cycle arrest during chromosome segregation (Figure 14d). An overview of all cell cycle class ratios can be found in Supplementary Figure 6.
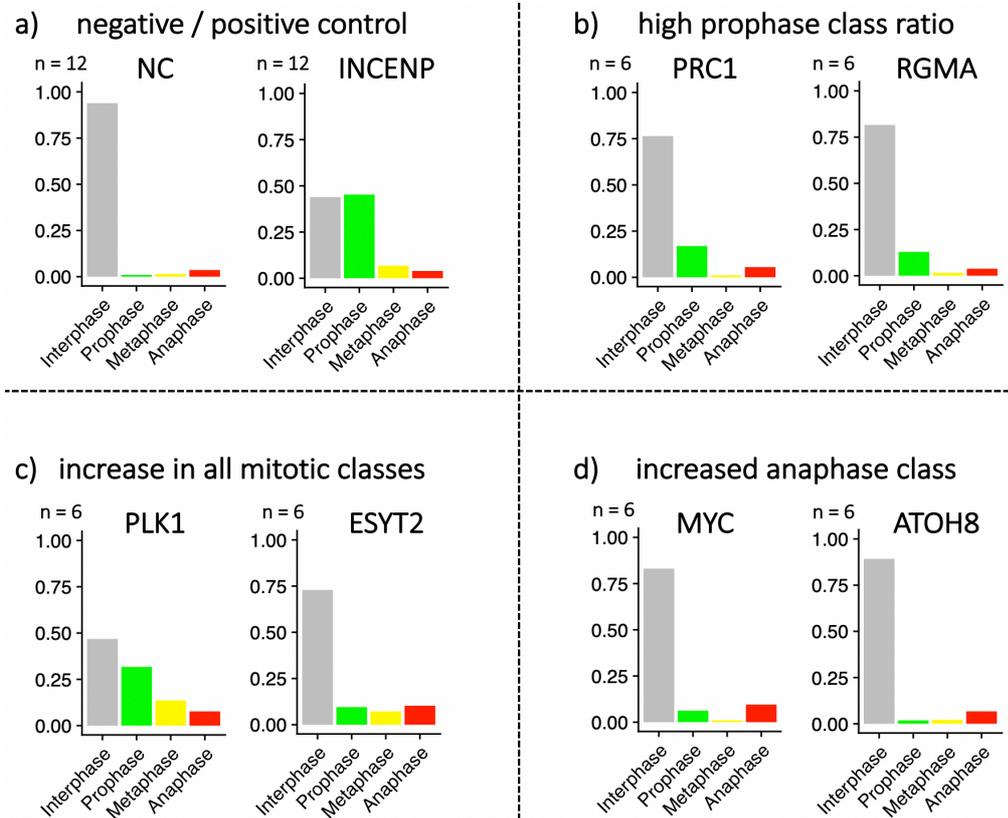
Figure 14 – Detection of abnormal mitosis duration induced by different siRNAs

Bar plot describing the total nuclei count as fraction of the different cell cycle classes. a) Negative control samples, transfected with non-coding siRNA (NC) and positive transfection control spheroids (INCENP) display distinct mitotic class distributions. b) PRC1 and RGMA showed an increase in prophase, c) PLK1 and ESYT2 show an overall increase of all mitotic phases, d) while MYC and ATOH8 displayed an increase of the anaphase class.

3.3.3 Abnormal positioning of cell cycle phases

Next, we evaluated the localization of the different mitotic phases within the spheroid under perturbation conditions. We used the distance to spheroid center measurement and displayed these in a violin plot (Figure 15). This analysis method allowed us to evaluate a possible biased localization of any class to the surface or inside of the spheroid. Additionally, this analysis technique provided us with the possibility to visualize size difference between the control siRNA and the target gene knock-down phenotype across each of the different samples.

The non-coding siRNA transfected spheroid had a mean distance of all segments of 95.8 pixel or 31.1 μm from the center of mass. None of the four cell phase classes had a preferred localization to the outside or inside of the spheroid (Figure 15a) which would describe an active

movement of the nuclei for the purpose of cell division induction, as known in neuronal progenitor cells [170]. Additionally, the density plot displays a homogenous distribution within the 25%-75% interpercentile range, describing a homogeneous phenotype throughout all six replicates.

The knock-down of INCENP in the positive control sample on the other hand induced reduced spheroid growth indicated by a decrease in the median with an average of ~26 px or 8.5 μm distance for all nuclei from the center of the spheroid (Figure 15b). Other samples, like DSE knock-down spheroids displayed also a reduced growth throughout all imaged samples and cell cycle stages (Figure 15c), depicted by the unimodal violin plot, with an average distance of ~48 px or 15.6 μm.
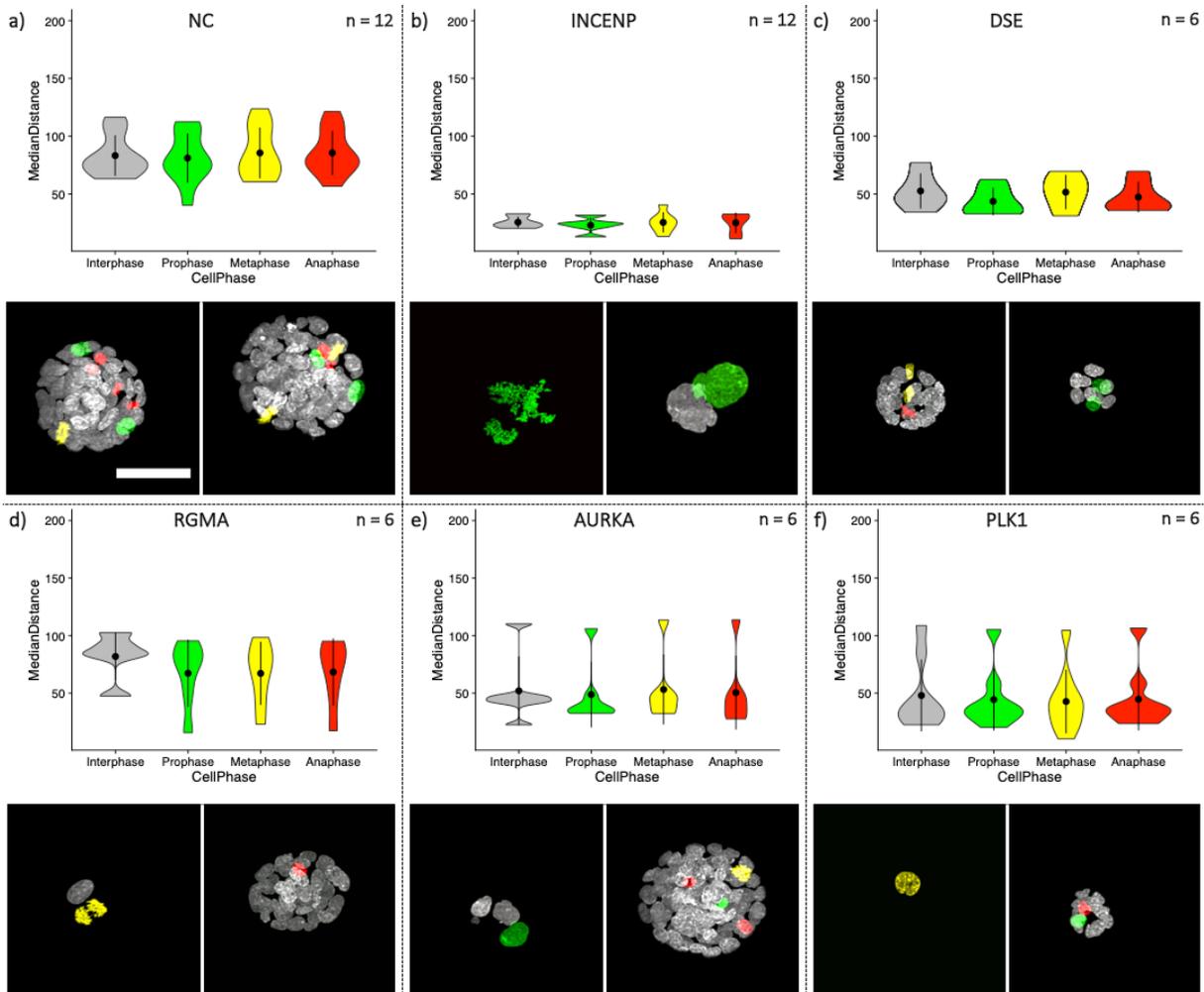
Figure 15 – Class positioning describes phenotypes across multiple siRNA knock-down samples

The class distribution displayed as violin plots of different selected samples with distinct abnormal properties appended by two representative images of maximum and minimum sized spheroids of the respective sample at the start of the time laps acquisition. a) Negative control (non-targeted siRNA) and b) positive control siRNA (INCENP) spheroids. Additional example specimens with abnormal class distribution are displayed for c) DSE, d) RGMA, e) AURKA f) PLK1 (scale bar = 50 μm). The black dot describes the median and the whiskers the 25%-75% interpercentile range.

The violin plot of PLK1, RGMA, and AURKA knock-down samples displayed multimodal distance-to-center distributions (Figure 15d-f). This distinct distribution described a reduced growth of single spheroids, while others develop similar to the negative control without any constraints in growth which could represent untransfected cells. A summary of all violin plots for each knock-down sample can be found in Supplementary Figure 7.

3.3.4 Cell cycle transitions

Due to the high temporal resolution of the acquisition, we were able to track single cell throughout the time laps and its individual cell cycle. This allowed us to determine whether a cell underwent in normal sequence the key phases of the cell cycle with the succession of interphase followed by prophase, followed by metaphase, followed by anaphase and back to interphase. This biologically "normal" progression can describe a healthy spheroid and no "abnormal" cell cycle progression would be expected. The 95% accuracy of the cell cycle classification (Figure 12) provided the essential prerequisite for this analysis.

High amounts of misclassification can occur due to abnormal nuclear shapes or textural features as found in abnormal nuclear phenotypes. Macronuclei are often misclassified as prophases or apoptotic cells or apoptotic condensed DNA phenotype can be misclassified as prophase or anaphase (Supplementary Figure 5). These misclassifications were detected as abnormal cell cycle progression. In these cases, a prophase was regularly followed by anaphase (ACD phenotype) or a cell transitioned back and forth between a pro- and interphase (macronuclei phenotype). The ratio of normal cell cycle progression to an abnormal cell cycle progression allowed us to evaluate viability of the sample upon target gene knock-down and detect cells with abnormal visual features or apoptotic cells (Figure 16).

In this analysis, beside the positive control of INCENP knock-down (p-value ≤0.001) with the described macronuclei phenotype, PLK1 (p-value ≤0.01), EME1 and CEP85 (p-value ≤ 0.05), displayed a high degree of false transitions due to cell cycle arrest and apoptosis. Conclusive count of normal and abnormal cell cycle transitions for all target genes can be found in Supplementary Tables 3.
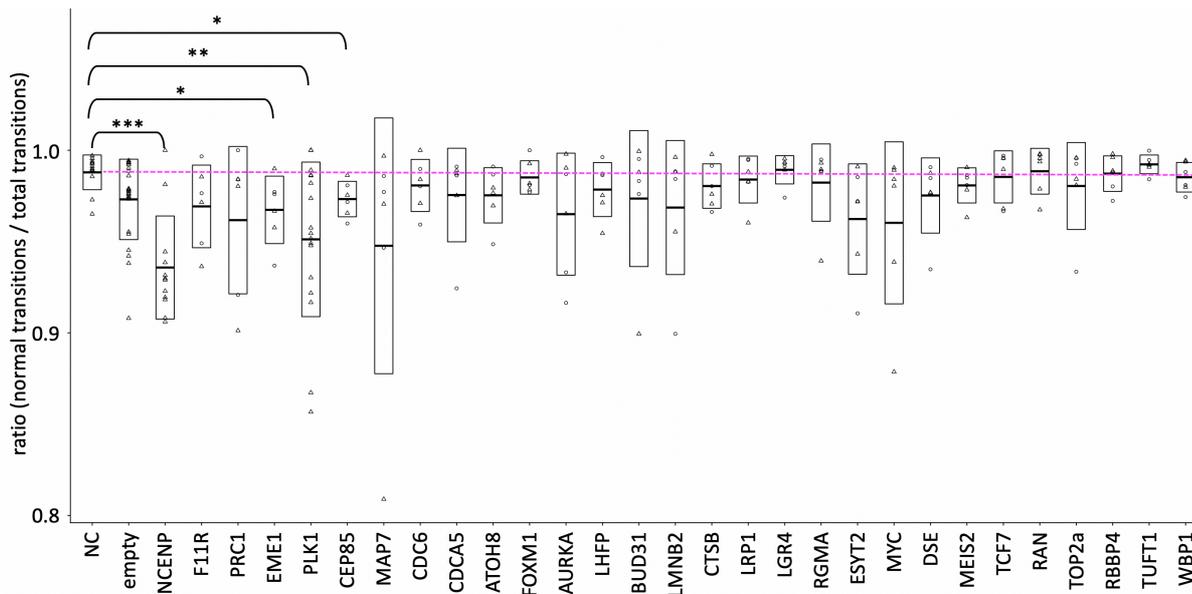
Figure 16 – Normal versus abnormal transitions describe sample viability

Box blot displaying the ratio of normal and total number of cell transitions of each of the target gene knock-down samples with the 95-confidence interval displayed as box, as well as the mean of all samples. Magenta line represents average ratio of the negative control sample (NC). t-test describing statistical significance with a p-value below 0.05 between the individual sample and the negative control are marked (p-values: *** ≤ 0.001, ** ≤0.01, * ≤0.05).

3.3.5 Cell cycle class size

As cells transition through the different stages of the mitosis, we detected a change in the segment size of the different mitotic stages described in Figure 13. We also compared on average the size in voxel of the different segment classes individually and clustered them by their size. This analysis allowed us to describe in comparison to the negative control sample, the cellular properties of the knock-down phenotype and detect outliers such as the positive control INCENP (cluster V - Figure 17) and BUD31 (cluster IV - Figure 17), which have such a distinct difference in size of the inter- and prophase that these phenotypes define a cluster each on their own. Increased prophase (cluster III - Figure 17) and respective increased interphase (cluster II - Figure 17) nuclei define the other clusters describing individual effects on the cell upon target gene knock-down.
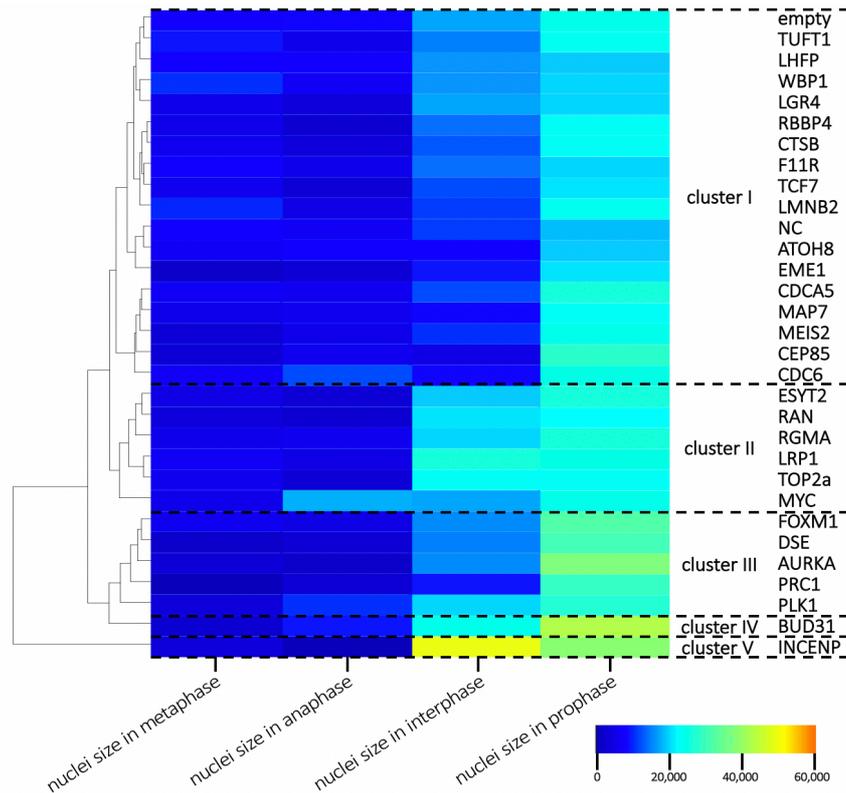
Figure 17 – Nuclei size throughout the cell cycle

Ranked based knock-down target gene clustering (cluster I-V) of the different target knock-down samples by the average nuclei size in voxel (color gradient) of the different mitotic classes (meta-, ana-, inter-, prophase).

3.3.6 Global spheroid structure evaluation

In addition to the analysis of single cells and their individual properties, we also analyzed the global spheroid appearance and its development over time. A combination of spheroid properties such as nuclei count, spheroid volume and the gain in volume as well as spheroid roundness allowed a further evaluation and comparison of the phenotypes induced upon target gene knock-down. A ranked based clustering of these properties facilitated furthermore the detection of samples with similar properties such as spheroid growth, nuclei density, abnormal spheroid shape and volume increase (Figure 18). Non-coding siRNA transfected negative control spheroids as well as all samples of cluster A displayed in comparison to severe knock-down phenotypes of the positive control (INCEMP) a strong growth in segment (nuclei) count, while retaining a roundish shape throughout the screen that resembles closely a sphere. In comparison, reduced growth and an abnormal shape was detected in samples of cluster E (Figure 18) such as CEP85 for example.
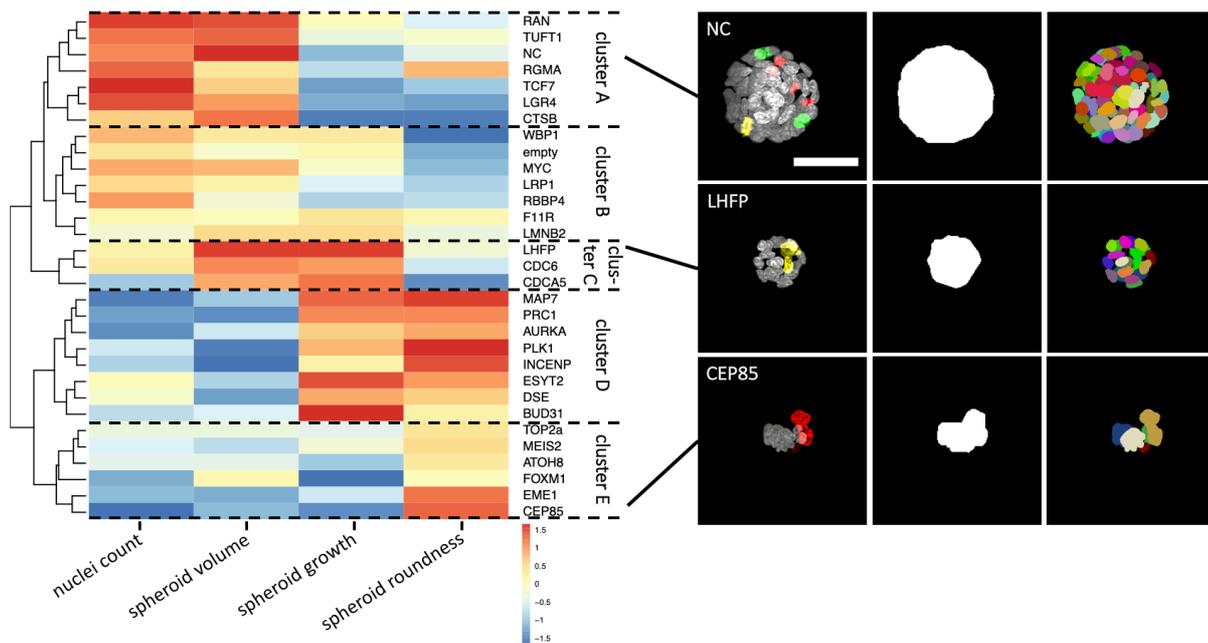
Figure 18 – Global spheroid properties

Global spheroid properties such as total nuclei count, spheroid volume at the acquisition start, spheroid growth throughout the screen and spheroid shape describe on average the global phenotypes of each target gene. By ranked based gene clustering (cluster A – E) of these properties, we achieved a clustering of similar phenotypes. Example images for cluster A, C and E display a non-coding negative control, LHFP and CEP85 siRNA transfected spheroid, the hull describing the spheroid volume and the individual segments maximum projected onto a single plane (scale bar = 50 μm).

3.3.7 Completing the picture – ranked based clustered phenotype analysis

To complete the picture of the individual effect of the gene knock-down onto spheroid development and the individual cell, we compared all 23 features (Supplementary Tables 2) from the entire screen of 228 individually evaluated spheroids, describing features of single nuclei and global spheroid properties. By ranked based clustering of the results, we obtained 5 clusters of knock-down phenotypes that correlate in their quantitative (Figure 19a) and visual features (Figure 19b-f). The negative control samples transfected with non-coding siRNA clusters together with LGR4, TCF7, TUFT1, WBP1 and CTSB (Figure 19 - cluster 1). This cluster comprises samples that were not or only slightly affected in their development and visual appearance. In comparison to the other clusters, these samples exhibited a high degree of growth represented in their growth rate by nuclei count, spheroid volume and high number of true transitions. These samples also displayed a high sphericity and migratory speed of the cells describing a highly vital phenotype.

Cluster 2 (Figure 19 – cluster 2) includes samples with minor phenotypes with a strong focus on increased prophase size (LRP1, RAN, TOP2A and RGMA) and reduced spheroid volume (TOP2A and RGMA) (Figure 19c).

Cluster 3 (Figure 19d – cluster 3) comprises knock-down phenotypes with spheroids depicting an increased volume growth rate (LHFP, CDC6, CDCA5) and shorter cell phase transition times (MYC, CDC6) while exhibiting high degree of false cell cycle transitions (MYC, LMNB2, F11R).

In cluster 4 (Figure 19e – cluster 4) strong phenotypes with reduced spheroid volume growth, reduced gain in nuclei count and low total number of cell cycle transitions (MEIS2, ATOH8, EME1 and CEP85) are grouped.

The most distant cluster to the NC phenotype (Figure 19 - cluster 5) includes the positive control INCENP, as well as PLK1, MAP7, DSE, PRC1, BUD31, ESYT2 and AURKA. These samples showed aberrant phenotypes in up to seven statistically significant features (p-values ≤ 0.05), describing grave development malfunctions as well as cellular and mitotic defects. Most notably were defects in spheroid growth, mitotic arrests, false cell cycle transitions, macronuclei formation and apoptotic cells. In comparison to these strong phenotypes, cluster 2-3 phenotypes comprise mainly singular or low statistical phenotypes (Figure 19a).
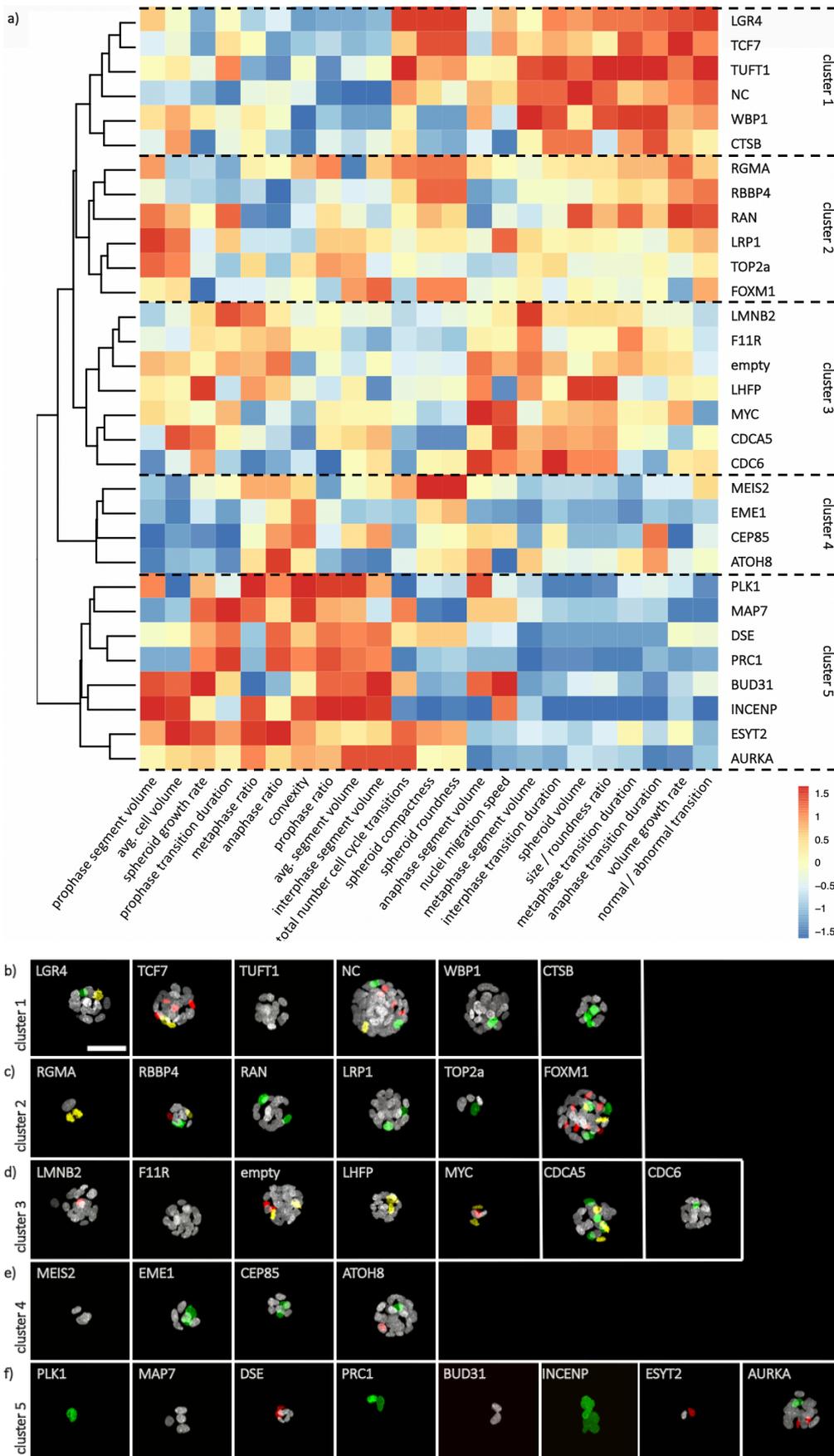
Figure 19 – Clustered phenotype analysis of all features detected in diSPIM HCS

Figure legend continued from previous page: a) Ranked based clustering of the phenotypes by their individual features describing the global and nuclei specific properties displays distinct clusters of siRNA target genes. b-f) Cluster 1-5 example images of each phenotype. Scale bar = 50 μm

3.3.8 Detected phenotype in comparison to the MitoCheck online data base

To evaluate the performance of our screening workflow and analysis pipeline, we compared the detected phenotypes to the online MitoCheck data base [148] (mitocheck.org), used during the planning phase of this project to select for the used target gene targets. In the MitoCheck screen only the first two to four cell divisions of HeLa H2B-GFP cells were imaged and analyzed for any abnormal phenotypes upon target gene knock-down. In comparison to the conducted high-content diSPIM screen, this results for the different MitoCheck screens in a focus on severe mitotic defects rather than on long-term development. Additionally, a heterogeneous population of cells was analyzed for the MitoCheck screen, whereas in this project we clonally expand a single cell and evaluate the single cell development. These differences in the two approaches on a high-content screen is reflected in the comparison of our results to the MitoCheck data base with a high degree of knock-down phenotype overlap. Notably, all of the strong and severe phenotypes identified in our study, such as complete cell cycle arrest, apoptosis, growth arrest and macronuclei formation induced by the knock-down of INCENP, PLK1, MAP7, DSE, PRC1, BUD31, ESYT2 and AURKA, described by Figure 14, Figure 15, Figure 16, Figure 17, Figure 18 and Figure 19 were also described in the MitoCheck screen even though a different cell line was used. Minor phenotypes such as the macronuclei formation in some spheroids induced by CEP85 and MEIS2 knock-down (Figure 19e) or interphase arrest induced by MYC siRNA transfection (Figure 19d) were not described. Additionally, TOP2a knock-down induced in our screen a higher ratio of prophase-classed segments, characterizing an elongated prophase. This phenotype was also identified by the MitoCheck screen, however in a validation experiment, a reduced prophase duration was described by the authors.

3.4 Specific methylome modification and analysis of induced aberrant mitosis

To confirm and evaluate the detected phenotypes of the siRNA-mediated knock-down diSPIM imaging screen, we modified the gene expression of our target genes by applying a molecular tool based on the recently described CRISPR-Cas9 system. This CRISPR-dCas9 methylome-modifying effector domain fusion protein has been shown to modify the epigenome at a targeted genomic location defined by the sgRNA as introduced in 1.2.2.

3.4.1 Epigenetic modifiers as gene regulatory method

We fused the catalytic C-terminal effector domains of epigenome modifying enzyme (DNMT3a, TET1) C-terminally to the dCas9 via a linker and added two nuclear localization sequences (NLS) for improved nuclear localization and a M2 flag to the N-terminus (Figure 20a). A dCas9 with no C-terminal addition of an ED was used as binding control (Figure 20a) and to physically block binding sites for regulatory factors. The EDs originate from the catalytically active C-terminal domains of the respective methylome-modifying enzyme (Figure 20b). We targeted the different epigenome modifying molecular tools to specific genomic sites by combining different sgRNAs with the different effector domains to modify distal or proximal regulatory CpGs with regulatory properties (Figure 20c). sgRNA target sites were selected to be approximately 33 base pairs away from the CpG to maximize epigenetic modifying efficiency following the results on methylome modifications by Vojta and colleagues [171].

Based on a 450k Illumina breast cancer patient screening data set comprising gene expression and CpG methylation data from human breast cancer patients, we selected CpGs that show high correlation (Pearson correlation > 0.5) between the expression of the selected target genes and the CpG methylation level (Figure 9 / Supplementary Tables 1). We detected up to seven regulatory CpGs per target gene (MEIS2 and RGMA) with correlative and anti-correlative regulatory properties. From the bioinformatics side, we expected that correlated CpGs (low CpG-me results in reduced expression) would show a gene knock-down dependent phenotype when these CpGs were targeted by the TET1 dCas9-ED and induce a phenotype resembling tumorigenesis in breast cancer, as these CpGs are demethylated in breast cancer. Correspondingly, we expected anti-correlated CpGs (high CpG-me results in reduced expression) to show the abnormal mitotic phenotype, when targeted by the methyl-group transferring DNMT3A.

Target genes that had only a single regulatory CpG with a correlation between expression level and CpG methylation below 0.6 were not further analyzed to focus our effort on most significant and promising target genes. Due to these restrictions, ATHOH8, AURKA, BUD31, CTSB, DSE, ESYT2, LGR4, RAN, and RBBP4 were not further investigated.
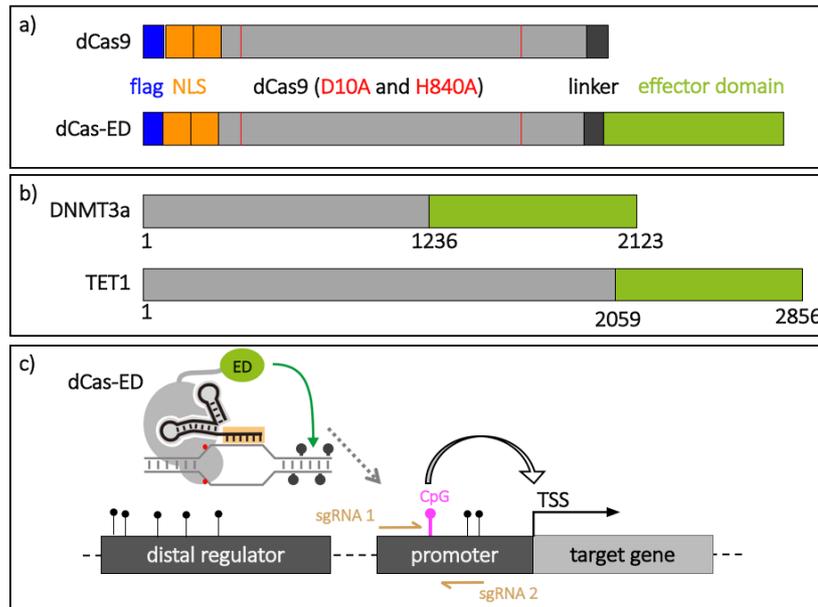
Figure 20 – dCas9-ED constructs targeting regulatory CpGs

a) CRISPR-Cas9 based epigenetic modifiers and binding control were composed of a M2-flag (blue), two nuclear localization sequences (orange), the dCas9 (light gray) mutated at residue 10 and 840 (red) deactivating the endonuclease function of the Cas9, and the effector protein (green). The effector domains were fused C-terminally via a linker (dark grey). b) Effector domains are the catalytically active, C-terminal domains of DNMT3A and TET1 from residue 1236 (DNMT3a) and 2059 (TET1) to the C-terminus of the protein. c) CRISPR-dCas9 fused with the effector domain was located to specific target sites defined by the sgRNA. dCas9-ED and sgRNA combination defined gene regulatory properties and targeted correlated or anti-correlated CpGs. Per CpG, we transfected two (upstream and downstream) sgRNAs with opposing orientation. CpGs were located in promotor and distal regulatory regions.

3.4.2 The selected molecular tool for gene knock-down

The remaining 18 target genes, had a total of 60 regulatory CpGs (mean Pearson correlation > 0.6), with 76% (n = 46) CpGs located in distal regulatory regions, while 23% (n = 14) CpGs are within 1 kb of the transcription start site. 42% (n = 25) CpGs show a positive correlation between methylation level and gene expression, while 58% (n = 35) are anti-correlated (Supplementary Tables 1). The genes CEP85, MEIS2 and RGMA were target genes that displayed correlated as well as anti-correlated CpGs making these targets highly interesting for quantitative analysis and epigenetic regulation comparison.

SgRNAs directing the dCas9 effector domain fusion protein to the specific genomic site were designed to locate the methylome-modifying enzymes around on average of 33 base pairs upstream from their corresponding target CpG, since it was described by Vjota and colleagues [171] that the dCas9-ED shows highest epigenome modifying effectivity 27 bp (+/-17 bp) from the PAM sequence of the sgRNA. Per regulatory CpG, we designed two opposing sgRNAs, one binding to the sense and one binding to the anti-sense strand of the DNA. Furthermore, sgRNA target sites were checked to display a minimum of two mismatches to the next off-target site, which should reduce off-targeting effects.

3.4.3 Control constructs to validate the methylome-mediated phenotype

To verify the induced phenotype and relate the observed effect solely to the modification of the methylome rather than the binding of a large protein to a regulatory region, we included in the pre-screen a third construct, which was comprised of only the dCas9 protein without any further enzyme added to the C-terminus. This construct allowed us, to evaluate the effect of the binding of the dCas9 onto the target site and distinguish the effects caused by plasmid transfection, site-specific protein binding and the effective methylome modification.

Furthermore, we wanted to induce and analyze a gene knock-down, through binding of the dCas9 to the transcription start site (TSS), a method in this thesis referred to as CRISPRi. We used the FANTOM5/CAGE online atlas (http://fantom.gsc.riken.jp/5/) to define the TSS of our target genes and selected a single sgRNA binding site at an average of 50 bp upstream of the TSS for optimal gene repression effect [70, 172].

3.4.4 Pre-screen in HEK293 cells

MCF10A cells are highly susceptible to siRNA transfection, but showed even with specialized transfection reagents a strong resistance against transfection with plasmids, displaying a transfection efficiency of ~2% detected by GFP expression (Figure 21a). This made it highly problematic to introduce both the sgRNA and the protein component of the functional dCas9-ED into this cell line by means of plasmid transfection for a statistically relevant quantitative analysis. We therefore decided to change to another cell culture model, the human embryonic kidney 293 cell line (HEK293). This cell line can be easily transfected with sgRNA expressing plasmids in comparison to MCF10A cells (Figure 21a).

When we compared the regulatory CpG methylation levels available from the ENCODE data base between MCF10A and HEK293 cells, we detected a good average correlation (0.69 Pearson correlation / 8.2% mean difference) of the target CpG methylation level between the two cell lines. Only individual CpGs displayed a larger than two-fold difference in the CpG-me (Figure 21b). Key differences were detected in the MEIS2 expression correlated CpGs, where four out of the seven regulatory CpGs showed a stronger decrease in methylation level in HEK293 cells than in the respective CpG in the MCF10A cell line. MEIS2 is a homeobox protein of the TALE (three amino acid loop extension) protein family involved in transcription regulation and shows a 47% higher transcription level in HEK293 cells in comparison to the MCF10A cell line (proteinatlas.org). This discrepancy of gene expression had to be considered for result evaluation of the MEIS2 target gene. In total, an increase in methylation level at a single CpG of more than two-fold was observed for CpGs regulating CDC6, F11R, FOXM1, MYC, TCF7 and PRC1 (Figure 21b).

To modulate the methylation level at the target CpGs in HEK293 cells, we introduced and selected for the stable expression of the different dCas9-ED constructs and evaluated the expression and localization of the recombinant protein by immunofluorescent staining (Figure 21c). The stable cell line expressing the dCas9-ED fusion protein allowed us to introduce in the pre-screen only the sgRNA delivering plasmid, to complete the methylome-modifying tool.
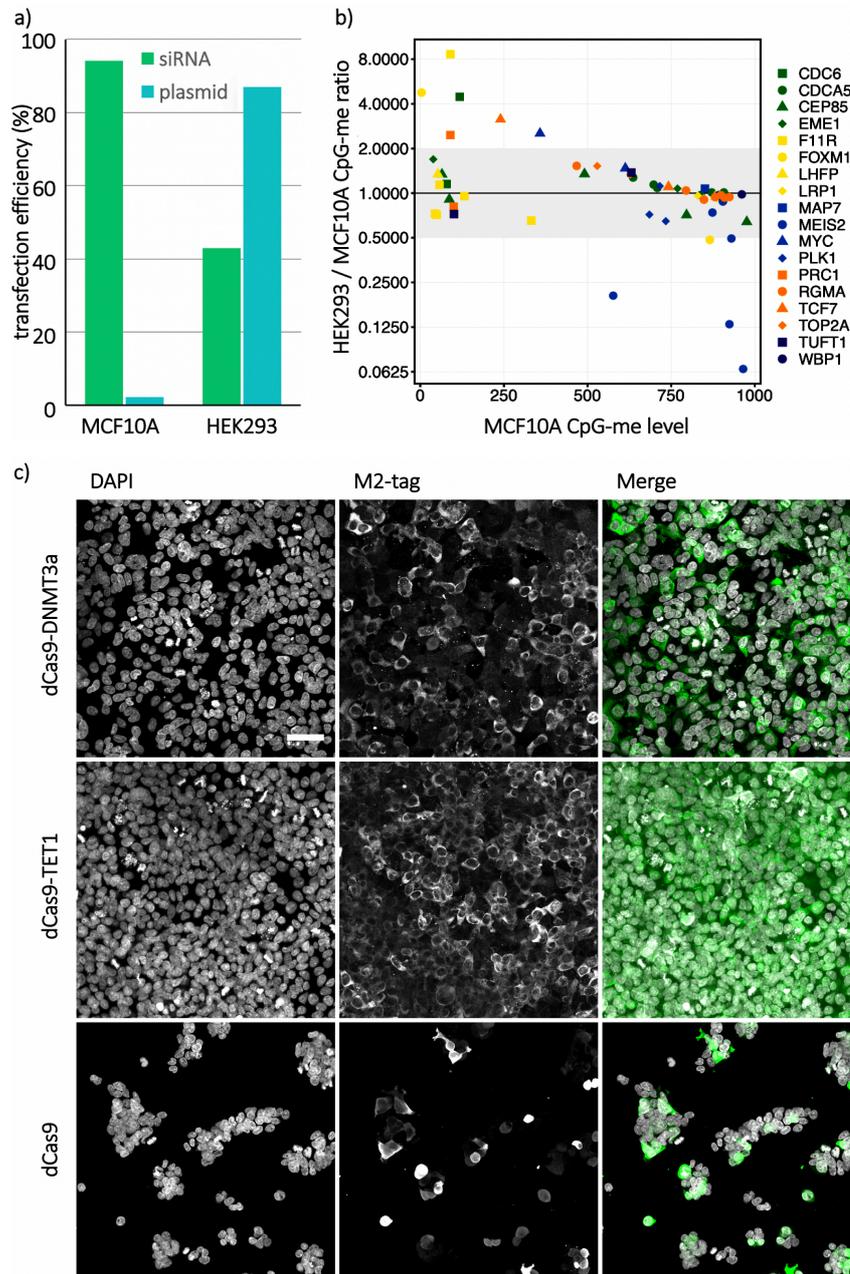
Figure 21 – HEK293 as model for epigenetic remodeling

a) Average transfection efficiency of MCF10A and HEK293 of siRNA targeting PLK1 (light green) and a sgRNA control plasmids (light blue) for a minimum of 1,000 cells. b) Comparison of MCF10A CpG methylation level (X-axis) to the ratio of the target CpG methylation level of HEK293 and MCF10A cell line (y-axis) for the different target genes and their respective regulatory CpG. The light grey area indicates changes in the methylation level less than 2-fold. c) Immunostaining of HEK293 cells expressing the different dCas-ED molecular tool constructs (scale bar = 50 μm).

3.4.5 Selection of effective molecular knock-down tools

Imaging by light sheet microscopy of the phenotype was not feasible, due to the extensive data that would have been generated with different combinations of the dCas-ED, dCas9 control construct, and sgRNA combinations. To circumvent the generation of an unmanageable data set, and difficulties arising from 3D immunofluorescent staining of samples mounted in Matrigel we decided to select first for significant methylome regulated target genes, that showed an abnormal phenotype already in 2D HEK293 cells acquired by 2D confocal microscopy. Subsequently, we would follow up with in depth analysis of the top target by means of high-content 3D live light sheet imaging.

In this pre-screen, we combined a total of 129 sgRNAs targeting regulatory CpGs or TSS or a non-targeting control sgRNA with the two dCas9-ED constructs and the inactive Cas9 construct stably expressed in HEK293 cells (Figure 21c). As before, we reduced transfection and handling artifacts by using solid-phase reverse transfection to deliver the sgRNA expressing plasmid into our dCas9-ED expressing cell lines. The sgRNA plasmid could be detected by GFP expression in transfected cells (Figure 21). SgRNAs that target correlative or respectively anti-correlated CpGs of a single target gene were mixed and transfected together, again to reduce the number of samples to 48 per dCas9-ED.

Although we achieved high transfection efficiencies and used HEK293 that stably express the dCas9-ED or the dCas9 constructs, we additionally used immunofluorescent staining to validate the presence of the sgRNA and fusion construct in each individual cell, due to varying dCas9-ED expression levels, as seen in Figure 21c. If only a single component (sgRNA or dCas-ED) was detected within a single cell, this cell was considered as negative and was not used for quantitative analysis.

We evaluated and classified the nuclear phenotype of double positive cells (++ / cells expressing dCas9-ED and the sgRNA) at different time points post transfection for up to 8 days by deep learning classification (Figure 22). Nuclei classes comprised the four main stages of the cell cycle (inter-, pro-, meta-, anaphase) as well as two additional phenotypes that describe nuclear aberrations. These abnormal nuclear phenotypes were macronuclei and apoptotic condensed DNA. A macronuclei phenotype describes a relatively large nucleus, displaying polyploidy through cell cycle progression without mitosis, while ACD describes the chromatin condensation induced by cell death. These two prominent phenotypes were not included in the CNN classification of the diSPIM scree in 3.3, as not sufficient training data could be generated prior to the light sheet screen. Additional features such as nuclei size were furthermore calculated from the immunofluorescence images. A phenotype was considered

strong when a nuclei class or class segment size was 1.5-fold increased over negative cells or the control samples transfected with non-targeted sgRNA.
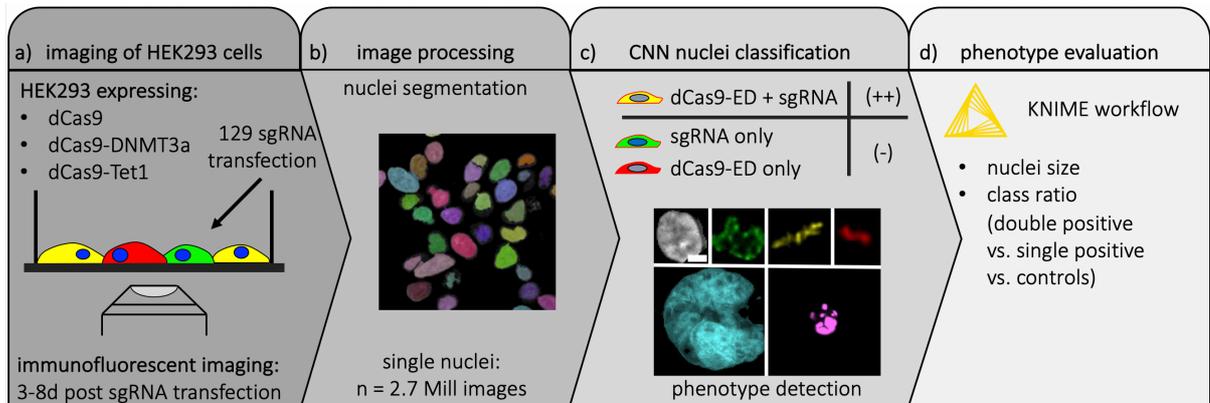
Figure 22 – dCas9-ED pre-screen for distinct phenotypes

a) dCas9-ED expressing HEK293 cells were transfected with sgRNA expression plasmids, completing the active fusion protein for methylome remodeling. Immunostaining for GFP (sgRNA) and M2-flag (dCas9-ED) allows to detect complete epigenome editing molecular tool present in the cell, while DNA was labeled by DAPI for b) segmentation and c) phenotype detection by CNN nuclei classification (scale bar = 5 *μ*m). d) Subsequently, analysis detects significantly occurring phenotypes and calculated nuclei features.

3.4.6   Correlation of phenotypes

A total of 129 possible dCas9-ED-sgRNA combinations were evaluated with an average of 6856 cells analyzed per combination, giving a total of 2.7 million nuclei that were evaluated in this pre-screen summarized in Table 3 for dCas9 genomic localization and in Table 4 for the different dCas-ED applied to target regions.

We detected in a total 18% of all sgRNA-dCas9 combinations strong phenotypes (n = 23) with a feature / class increase of more than 1.5-fold in comparison to untransfected (-) cells of the same class. 30% (n = 7) of the detected phenotypes were caused solely by localizing the deactivate Cas9 to the DNA (Table 3). 43% (n = 10) of 1.5-fold increased phenotypes were detected in dCas9-TET1 targeting and 26% (n = 6) in dCas-DNMT3a expressing cells.

Strong phenotypes included high ratio of apoptotic cells (ACD phenotype) (n = 4; 17%), macronuclei formation (n = 10; 43%), metaphase arrest (n = 2; 9%) and prophase arrest (n = 6; 26%). Additionally, significantly larger macronuclei (n = 2; 7%) and prophase (n = 1; 4%) segments were detected in direct comparison to the untransfected cells.

In conclusion, 88% of the CRISPRi and dCas9-ED evoked phenotypes (n = 22) matched in the described nuclei characteristics to the cells with the respective siRNA mediated gene knock-down described by the MitoCheck database or the Cyclebase3.0 [173] (Table 3, Table 4). An additional 85% (n = 17) of the evoked targeted CRISPRi and dCas-ED phenotypes

were comparable described as result of the same target gene knock-down in the high-content light sheet microscopy screen (3.3), highlighted by the check marks in Table 3 and Table 4.

Comparable phenotypes include an increase in macronuclei phenotype represented by an increase in nuclei size or high ratio of prophase. As discussed above (3.3.4), macronuclei phenotypes are often misclassified by the convolutional neuronal network as prophase in the high-content diSPIM knock-down screen. Severe phenotypes, such as ACD, cell cycle elongation or arrest and macronuclei formation correlate with reduced spheroid growth and reduced motility as these cells are strongly restricted in their capability to divide and overall viability. Furthermore, can a cell cycle arrest at a defined state also prevent proliferation, therefore result in a reduced spheroid growth or elongated cell cycle phase duration.

Although not conclusively analyzed in their mode of action, the high correlation of phenotypes between the different methods of gene modulation show a high similarity between the outcome on the cell upon three different target gene perturbation.

Table 3 – CRISPRi: evoking a phenotype by dCas localization to target site

Listing of target genes that show more than 1.5-fold increase phenotype, such as apoptotic condensed DNA (ACD) and increase in cell cycle class upon dCas9 positioning (CRISPRi) to anti-correlated (cyan), correlated (magenta) regulatory CpGs or transcription start site (orange). Detected CRISPRi phenotypes were compared with online databases (CB: Cyclebase.org / MC: MitoCheck.org) and siRNA diSPIM screen results described in Chapter 3.3. To the CRISPRi comparable induced spheroid and mitotic phenotypes are highlighted by green check mark, correlating the methods. incr. = increased

gene	dCas9 (CRISPRi)		phenotype described by CB/MC:	siRNA screen		
	sgRNA target	phenotype / class		spheroid phenotype		mitotic phenotype
MEIS2	(cyan up)	ACD	(green)	reduced growth ✔		---
PLK1	(cyan up)	prophase	(green)	reduced growth ✔	incr. ana-, pro- meta- phase / abnormal transitions	✔
TUFT1	(cyan up)	prophase	(green)	---		elongated prophase ✔
EME1	(magenta down)	prophase	(green)	reduced growth / abnormal shape ✔	abnormal cell transitions	✔
	(orange down)	prophase	(green)			
FOXM1	(orange down)	prophase	(green)	incr. growth		---
LHFP	(orange down)	ACD / prophase	(green)	reduced growth ✔	reduced nuclei motility	✔

Table 4 – Strong phenotype induced by dCas9-ED targeting regulatory CpGs

Listing of target genes that show more than 1.5-fold increase phenotype, such as apoptotic condensed DNA and increase in cell cycle class upon a) dCas9-DNMGT3a positioning (CpG methylation) or b) dCas9-TET1 positioning (CpG demethylation) to anti-correlated (cyan), correlated (magenta) regulatory CpGs or transcription start site (orange). Detected phenotypes were compared with online databases (CB: Cyclebase.org / MC: MitoCheck.org), CRISPRi phenotypes and siRNA diSPIM screen results described in Chapter 3.3. To the dCas9-ED comparable induced spheroid and mitotic phenotypes are highlighted by green check mark, correlating the methods. DCas9-DNMT3 is expected to target anti-correlated sgRNA CpGs targets (cyan arrow) while dCas9-TET is expected to induce knock-down phenotype in correlated sgRNA CpGs targets (magenta arrow). incr. = increased

a) dCas9-DNMT3a site directed epigenome modification

gene	dCas9 – DNMT3a sgRNA target	phenotype	phenotype correlates with CB/MC:	CRISPRi phenotype	siRNA screen spheroid phenotype	siRNA screen mitotic phenotype
PRC1	(cyan arrow)	macronuclei	(green)	---	reduced growth ✔	incr. ana-, pro- and meta-phase / abnormal transitions ✔
RGMA	(cyan arrow)	macronuclei	(green)	---	---	incr. nuclei size / elongated cycle phases ✔
RGMA	(orange arrow)	macronuclei	(green)			
CDC6	(orange arrow)	macronuclei	(green)	---	reduced growth ✔	larger prophase / higher pro-, metaphase ratio ✔
TOP2a	(magenta arrow)	macronuclei	(green)	---	reduced growth / incr. avg. nuclei size ✔	---
EME1	(magenta arrow)	macronuclei	(green)	prophase ✔	reduced growth / abnormal shape ✔	abnormal cell transitions ✔
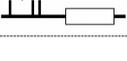
b) dCas9-TET1 site directed epigenome modification

gene	dCas9 – TET1 sgRNA target	class / phenotype	phenotype correlates with CB/MC:	CRISPRi phenotype	siRNA screen spheroid phenotype	siRNA screen mitotic phenotype
CDC6	(cyan arrow)	macronuclei	(green)	---	reduced growth ✔	larger prophase / higher pro-, metaphase ratio ✔
CEP85	(cyan arrow)	macronuclei	(white)	---	reduced growth / larger avg. nuclei ✔	abnormal cell transitions ✔
LHFP	(cyan arrow)	ACD	(green)	---	reduced growth / reduced nuclei motility ✔	---
PRC1	(cyan arrow)	ACD	(green)	---	reduced growth ✔	incr. ana-, pro- and meta-phase / abnormal transitions ✔
PRC1	(orange arrow)	macronuclei	(green)			
TUFT1	(orange arrow)	metaphase	(green)	---	---	elongated prophase
EME1	(magenta arrow)	macronuclei /metaphase	(white)	prophase ✔	reduced growth / abnormal shape ✔	abnormal transitions ✔
FOXM1	(magenta arrow)	prophase	(green)	---	incr. growth	---
RGMA	(magenta arrow)	macronuclei	(green)	---	incr. nuclei size ✔	elongated cycle phases ✔
TOP2a	(magenta arrow)	macronuclei	(green)	---	reduced growth / incr. avg. nuclei size ✔	---
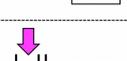
To evaluate the effectivity of methylome modification, we compared the anticipated effect of dCas9-ED targeting correlated CpG, anti-correlated CpG and TSS targeting with the measured influence on the cellular phenotype. 50% (n = 6) of the twelve dCas9-DNMT3a (Table 4a) and dCas-TET1 (Table 4b) samples with elevated levels of elicited phenotypes, targeted CpG properties matched with the bioinformatically predicted gene knock-down effect of the methylome modulation. In an additional 4 cases the localization of the dCas9 to the transcription start site was sufficient to induce a knock-down phenotype. In the other 27% (n = 6) sgRNA and dCas9-ED combinations, the localization in theory would increase the expression of the target gene, while still inducing an anticipated knock-down phenotype. The underlying cellular processes and reasons these phenotypes evolve cannot be defined through the conducted experiments at this point and would need further evaluation and differentiated analysis of the effectivity, regulatory mechanisms and capability of the different dCas9-ED.

3.4.7   In depth analysis of RGMA regulating dCas9-ED targets

Subsequently to the conducted pre-screen, we focused on RGMA for a in depth analysis, as this gene showed as only target after different methylome modifications a significant, predictable and reproducible phenotype (Table 4a, b) correlating in the evoked macronuclei phenotype with the online data bases and siRNA mediated segment size increase in the HC light sheet siRNA screen. MitoCheck.org describe the RGMA phenotype as strange nuclear shape, with nuclei segregation problems and multiple DNA masses, a phenotype that was detected as macronuclei class by the deep learning nuclei classifier. Additionally, RGMA is highly interesting to study the effectivity and possibilities of targeting different effector domains with different properties to defined target CpGs, as RGMA has both, correlated and anti-correlated CpGs (Supplementary Tables 1).

The RGMA (Repulsive Guidance Molecule Family Member A) gene encodes a glycosylphosphatidylinositol-anchored glycoprotein that has a known function as an axon guidance protein in the developing and adult central nervous system. For RGMA we detected a total of seven regulatory CpGs, three with correlated and four with anti-correlated regulatory properties.

These regulatory CpGs displayed a high correlation of the methylation level between HEK293 cells and MCF10A cell lines (Pearson correlation: 0.86) (Figure 21b), giving evidence that the different means of RGMA gene knock-down can be directly compared in the different cell lines.

When further evaluation the RGMA knock-down phenotype in the conducted pre-screen, we detected in three dCas9-ED-sgRNA combinations a high ratio of the cells the macronuclei phenotype targeting the TSS, correlative and anti-correlative RGMA regulatory CpGs (Table 5 / Figure 23b). On average the segment sizes increase by 2.2-fold in the evoked macronuclei from 1229 pixels to 2739 pixels (dCas9-DNMT3a targeting anti-correlated CpGs) or by 3.3-fold to 4111 pixels (dCas-TET1 targeting correlated CpGs) respectively, when compared to interphase classed nuclei. The increase in nuclei size and the macronuclei phenotype induced by the different sgRNA-dCas9-ED combinations are depicted exemplary in Figure 23.

Furthermore, the ratio of the macronuclei class increased significantly when anti-correlated CpG targeting sgRNA with dCas9-DNMT3a was present in the cells increasing the macronuclei count by 56%. dCas9-DNMT3a targeting the TSS mediating CRISPRi increased the ratio of macronuclei class segments to 1.94-fold (Table 5). dCas9-TET1 failed to significantly increase the ratio of macronuclei found in double positive and negative cells, resulting in a 1.36-fold increase of the occurrence of the macronuclei class.

Finally, to validate the obtained the results, we used any of the highly effective dCas9-ED constructs and combined these with a non-targeted sgRNA. In these samples, we could not detect a significant difference between double positive and negative cells in class occurrence or size increase (Table 5 – non-targeted control). Some sgRNA-dCas9-ED combinations even achieved a reduction of the macronuclei class ratio (Table 5 – dCas-TET1 and dCas9 combined with non-targeted sgRNA).

Table 5 – RGMA is an effective dCas9-ED target

In-depth comparison of the effect on the macronuclei class by the different active dCas9-ED targeted against regulatory elements of the RGMA gene. As control a non-targeted sgRNA was used in combination with all dCas9-ED. Asterisks labels ratio below 1.5-fold increase between double positive and untransfected cells. Red highlights a decrease of the macronuclei class ratio, Occurrences of the macronuclei class depicts ratio of samples comprising the phenotype class to total sample count.

	dCas9-ED	target site	macronuclei class ratio (++/-)	occurrence of maconuclei class
targeted dCas9-ED	dCas9-DNMT3a	TSS	1.94	0.24
		anti-correlated CpG	1.56	0.22
	dCas9-Tet1	correlated CpG	1.38*	0.17
non-targeted control	dCas9-DNMT3a	non-targeted	1.39*	0.06
	dCas9-Tet1	non-targeted	0.88*	0.07
	dCas9	non-targeted	0.76*	0.05

a) H2B-GFP

MitoCheck: RGMA knock down in HeLa

dCas9-DNMT3a x anti-correlative CpG

dCas9-TET1 x correlative

dCas9-TET1 x TSS
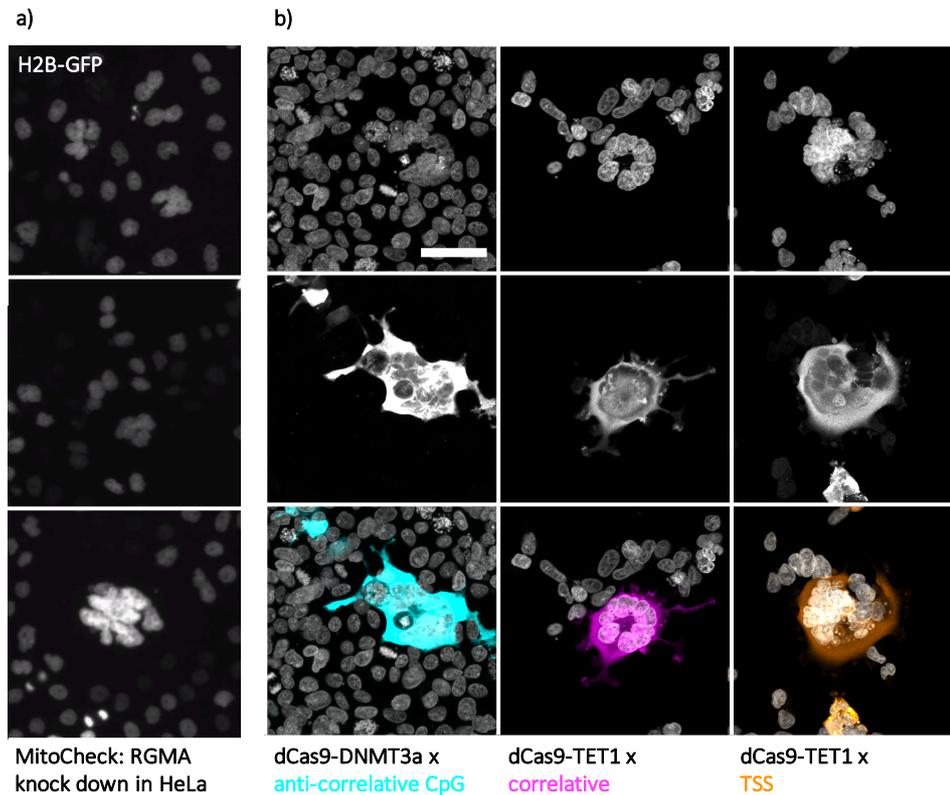
Figure 23 - RGMA knock-down phenotype in 2D

a) Screen images of RGMA knock-down in HeLa cells (MitoCheck.org). b) Example images of phenotypes evoked in 2D HEK293 cells upon dCas9-ED localization to anti-correlated (cyan), correlated (magenta) regulatory CpGs as well as the TSS (orange).

3.4.8 RGMA regulating CpGs targeted in 3D spheroids

HEK293 cells seeded in Matrigel can develop into multi cellular spheroids, just like MCF10A cells described in 1.1.6. This trait of the HEK293 cells allowed us to study the RGMA knock-down phenotype after methylome modification in 3D under the more physiological growth conditions. We therefore used the same sample preparation approach as described in section 3.2.3, transfecting into the different dCas9-ED cell lines the sgRNA by plasmid solid-phase reverse transfection and mounting the cells subsequently in matrigel onto diSPIM one-well imaging plates. Three and five days after sgRNA plasmid transfection, we imaged every GFP positive (sgRNA positive) spheroid at low temporal resolution. We evaluated the nuclear phenotype and spheroid development over time (Figure 24 / Table 6) through Hoechst staining. As the dCas-ED fusion protein were not fluorescently labeled, we were not able to detect expression levels of these recombinant proteins.

Similar to the prior conducted experiment, we combined the dCas9-DNMT3a with sgRNAs targeting anti-correlated CpGs and the TSS as well as, dCas9-TET1 with sgRNAs targeting correlated CpGs as these combinations resulted in a strong phenotype in the pre-screen

(Table 5). Furthermore, we combined the dCas9 without any effector domain with sgRNAs targeting either correlated, anti-correlated CpGs or the TSS as control.

A total of 273 spheroids were imaged by diSPIM light sheet microscopy and evaluated for all of the following properties, resembling the same phenotypes described in 3.4: macronuclei development, extended mitosis duration (mitotic arrest), reduced growth and ACD development.

The macronuclei phenotype was detected in ~40-50% of all spheroids, that either combine the dCas9-DNMT3a with sgRNA targeting anti-correlated CpGs (Figure 24a), dCas9-TET1 targeting correlated CpGs (Figure 24b) or through positioning the dCas9 and dCas9-DNMT3a to the TSS (Figure 24c,f). When the dCas9 control fusion protein was targeted to regulatory CpGs, we could not detect a high number of spheroids, incorporating macronuclei (Figure 24 d, e) with 5.6% and 7.4% of all spheroids showing macronuclei properties. Likewise, the phenotype of reduced growth and dying cells were distributed (Table 6), resembling a severe effect on the cellular homeostasis upon provisional RMGA knock-down. Elongated mitosis or mitotic arrest was detected in every third spheroid of dCas9-TET1 targeting correlated CpGs.

Table 6 – Phenotype analysis of dCas9-ED targeting RGMA regulatory CpGs in 3D HEK293 spheroids

For different sgRNA-dCas9-ED combinations targeting RGMA regulating CpGs and transcription start site (TSS), we replicated the pre-screen (3.4.4) in 3D spheroids and compared the effect of dCas9-ED targeting on different phenotypes. We evaluated the percentage of spheroids displaying abnormal cellular and global properties, such as macronuclei formation, extended mitosis duration, reduced spheroid growth and apoptotic cells (ACD). Example images for each combination are displayed in Figure 24.

dCas9-ED-sgRNA combination		detected phenotypes				
dCas-ED	target CpG properties	macronuclei	extended mitosis duration	reduced spheroid growth	ACD	n-spheroid
dCas9-DNMT3a	anti-correlated CpG	52%	12%	50%	73%	118
dCas9	anti-correlated CpG	6%	0%	11%	11%	18
dCas9-TET1	correlated CpG	41%	30%	32%	54%	37
dCas9	correlated CpG	7%	4%	26%	11%	27
dCas9-DNMT3a	TSS	45%	0%	45%	24%	38
dCas9	TSS	49%	3%	40%	31%	35

In total, these results constitute the bioinformatically predicted target gene knock-down upon site directed regulatory CpG methylation modulation and the resulting manifestation of the knock-down phenotype in macronuclei formation describing for the RGMA regulatory CpG an

effective possibility to regulate RGMA gene transcription. For RGMA, the conducted experiments strongly correlate the phenotypes to the suppression of the target gene independent of the applied method allowing a provisional correlation and comparison of the evoked phenotypes between the different methods and different cell lines.

For all methods modulating RGMA gene expression either on a transcriptional level by modulating regulatory CpGs with dCas-ED, or on the translational level through the siRNA, the evoked cellular phenotype represents a macronuclei formation resulting from incomplete cytokinesis while replication of the genome is continued, resulting multiple copies of the genome in a single cell.
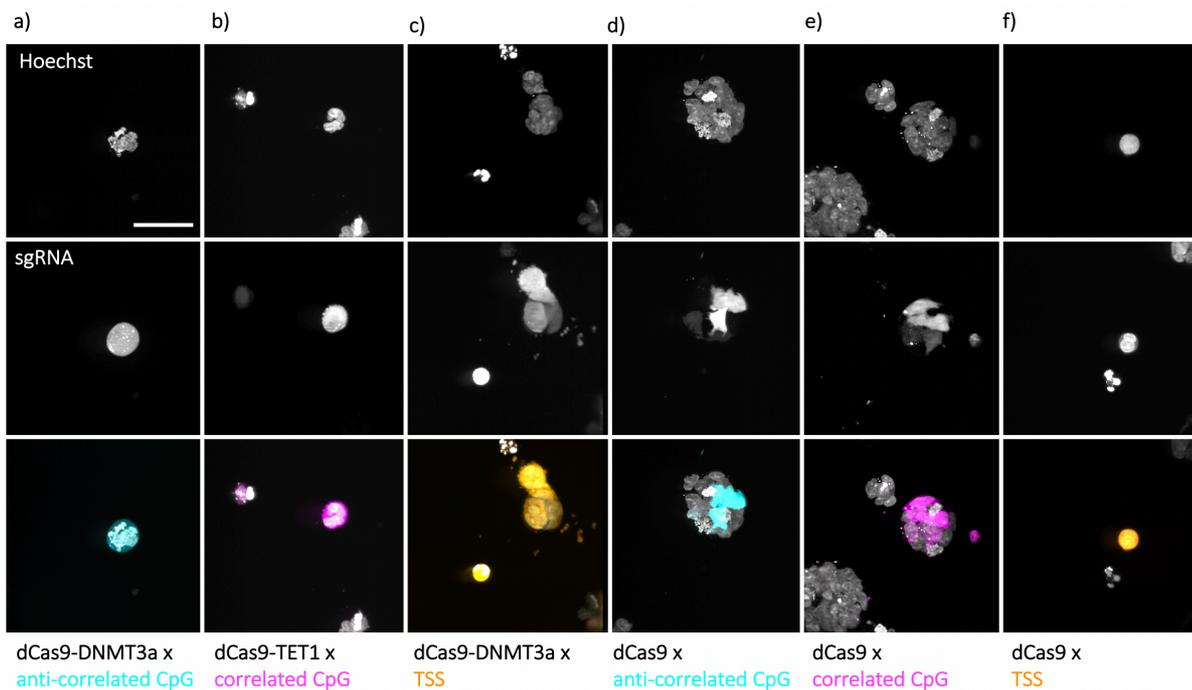
Figure 24 – Phenotypes evoked by dCas9-ED targeted against regulatory CpGs in 3D

Example images of HEK293 spheroids grown in Matrigel after dCas9-ED-sgRNA targeting regulatory CpG. a) Anti-correlated CpG targeting sgRNA (cyan) combined with dCas9-DNMT3a. b) Correlated CpG targeting sgRNA (magenta) combined with dCas9-TET1. c) TSS targeting sgRNA (orange) combined with dCas9-DNMT3a. d) anti-correlated CpG targeting sgRNA (cyan) combined with dCas9. e) correlated CpG targeting sgRNA (magenta) combined with dCas9. f) TSS targeting sgRNA (orange) combined with dCas9 (scale bar = 50 μm).

4 Discussion

In this project, we developed the methodological framework to study spheroid development and subcellular phenotypes in 3D cell culture systems in great depth and large number upon a perturbation. First, we established light sheet microscopy as tool to image 3D spheroids in a high-content screen at unmatched spatial and temporal resolution. Second, we used this method to analyze siRNA-mediated knock-down phenotypes in spheroids in a highly automated fashion. Third, we compared the siRNA and epigenetic methylation regulatory methods for target gene knock-down and the evoked phenotype. Thereby, we were able to compare usability, effectivity, and the induced phenotype between the two molecular methods.

4.1 Light sheet microscopy as a method for 3D cell line imaging

Image-based phenotype analysis is a very established method and has already been used extensively in imaging-based screens to detect abnormal phenotypes in cell culture systems upon molecular perturbations [126] or drug administration [7, 160]. Key applications of these screens can be found in pre-clinical trials in search for novel, effective inhibitor molecules and selection of potential treatment targets, as well as in basic research, analyzing individual gene functions or the functional interplay of associated genes.

For many years, spinning disc microscopes were the state-of-the-art system applied in these imaging screens due to their superior speed and low light dose in comparison to laser scanning confocal microscopes [7, 131, 174, 175], while retaining confocality and image resolution. For any imaging application in 2D cell culture models, the SDM or CLSM systems provide an effective, easy, usable, and established platform with sophisticated microscope controls and optimized incubation appliances. However, all single objective confocal microscopes display a major problem with larger samples such as 3D spheroids and organoids. Due to light scattering as photons pass through the tissue, these systems are not able to provide isotropic resolution deeper within larger tissues (Figure 5). SDM and CLSM systems are not able to image the entirety of the 3D cell culture model at high resolution and are restricted to cellular analysis on the surface or outline of the specimen. Additionally, acquisition times of seconds to minutes restrict high-throughput applications, short imaging intervals and large sample count, while intensive light exposure can induce phototoxic effects (Table 1).

With the commercial development of light sheet microscopes with an inverted geometry such as the ASI diSPIM system [112] or the Luxendo QuVi system, the advantages of light sheet imaging are now applicable to standardized 3D cell culture formats. Other light sheet microscopy geometries such as the multi view SPIM developed by Keller and colleagues [104,

176] allow high spatial resolution of large specimens, but lack capability to image more than a few samples. Asymmetric light sheet technology implementations such as developed by Strnad and colleagues [115] enable acquisition of up to 20 samples but are labor intensive and challenging to mount as well as restricted in their in-depth imaging capabilities, making the upright, dual-view geometry of the diSPIM the most suited system for high-content 3D cell culture imaging (Figure 4).

Dual view light sheet imaging is, as we have shown, an ideal light sheet geometry to acquire 3D spheroids with a size up to 100 μm at high temporal and isotropic spatial resolution (Figure 8). The extremely fast acquisition capabilities of the diSPIM allow a high sample number to be imaged at extremely short acquisition intervals (Figure 10). The low light intensity needed to provide sufficient signal to noise ratio allows imaging with an undetectable phototoxic effect on the sample. Dual view acquisition combined with image fusion and 3D deconvolution provides unmatched image quality as well as high, isotropic resolution, with the second view compensating for in depth light scattering (Figure 11). In summary, the light sheet microscope outperforms the spinning disc microscope in almost every measurable imaging parameter (Table 1) and cell viability experiment (Figure 8).

The ASI diSPIM system displays exceptional performance when handled correctly, but usability was restricted by user experience. Further challenges of the diSPIM such as imaging stability, sophisticated incubations, data size, and data processing were addressed by us and improved to a stage, where the entire development of a spheroid from a single cell could be imaged without influencing the development (Figure 8). Acquired data was instantly reduced by ~75% through image processing by the hSPIM diSPIM data processing tool, which uses state-of-the-art local, parallel GPU processing hardware for maximal performance and throughput (Figure 11). This allowed usage of standard IT infrastructure for data transfer in a reasonable time frame. Major limitations of the diSPIM system in 3D spheroid acquisition were only posed by samples larger than 100 to 120 μm in diameter. These samples outgrew the FOV quickly and loss in isotropic resolution was detected due to single view in depth signal loss.

4.1.1 Automation and reproducibility

Our key focus throughout the establishment of the high-content screen was on a high degree of reproducibility within the screen and for any possible future imaging application. This meant a high degree of automation while retaining flexibility to use different 3D cell culture systems and methods of cell culture perturbations. Additionally, any software used in this project was open source, with processing workflows quickly adaptable to the individual project focus.

With the usage of solid phase siRNA reverse transfection and the liquid handling robot we achieved high transfection rate, high precision sample mounting and strong reproducibility of the sample preparation minimizing individual human impact. This allows future research to apply the developed methods instantly on other cell lines and perturbation methods. Additionally, we implemented a pre-screen step to autonomously select samples for 24 hours imaging. Not only did this step reduce manual labor in sample selection, but it also provided an impartial selection of the spheroid with the only factors that define the sample selection being the position in the Matrigel, bypassing any 2D cell layers at the bottom of the dish and fused spheroids originating from two consolidated cell clusters.

4.1.2 Optimizing acquisition for a high-content screen

The fast imaging capabilities and low phototoxic effects of the light sheet microscopes entice for excess acquisition of unnecessary data. This can be due to unnecessarily short acquisition intervals or excess data covered by only a single view such as produced for small samples by the stage scan acquisition. Though possibly not affecting the sample viability, a key focus of light sheet microscopy should always be data reduction through data restricted acquisition. This will facilitate data handling and shorten data processing time and processing performance demand.

In this project, we therefore assessed every aspect of the screen in regard to data reduction. We selected piezo/slice acquisition due to the higher ratio of dual view acquired volume in comparison to stage scan acquisition, and for simplified post-acquisition data processing, as no deskewing of the data was necessary (Figure 4). Premises for the piezo/slice acquisition were the sphericity with a spheroid diameter of at most 100 μm and a confined single position, provided by five day old MCF10A spheroids.

For larger samples beyond 100 μm in width, the optimal acquisition mode is stage scan acquisition as it reduces movement of mechanical parts during acquisition and is not limited in the stack size by piezo motors, but by the size of the sample. Additionally, imaging speed is not limited by microscope mechanics but by camera acquisition capabilities, outperforming the SyncA image acquisition by five-fold (Figure 4). Due to skewing of the data, XY pixel positions cannot be correlated directly and dual view registration and deconvolution would require additional image processing. We utilized this acquisition mode in the low-resolution pre-screen, since the aim was to detect a spheroid and its position within the imaging plate and not to provide high resolution image quality, which would have required dual view acquisition and image registration.

The number of acquirable samples during the high-content screen was defined by different factors restricting the acquisition speed of the microscope. Stage velocity, data transfer rate, and set-imaging parameters defined the acquisition duration per position scan of 4-5 seconds. We decided to acquire at three-fold oversampling, since this would support the highest spatial resolution, but also increase dual stack acquisition duration. Furthermore, we were limited in the stage velocity, as rapid accelerations and decelerations would have resulted in detaching of Matrigel spots from the imaging plate. Thirdly, with 1 GB of data acquired every four to five seconds, a reliable fast data transfer was essential. Standard local area network performances of ~100 MB per second were not sufficient to directly transfer the data onto a server, which again highlights the hardware demands and needed infrastructure always required for light sheet imaging. With the set imaging parameters and maximal mechanical movements of the different stages, we were limited to 40 positions imaged every five minutes. The high temporal resolution was essential to track single cells throughout cell division (Figure 11).

4.1.3   Separating beads and samples for image processing

With multi view imaging, acquired data has to be conjoined through image registration and fusion of the different views. Two main methods have been used in the community for image registration. In contrast to signal or content based multi view registration detecting local intensity maxima within the sample, the registration matrix detection based on single fluorescent beads is independent of fast, dynamic intensity changes within the sample as seen during cell mitosis as well as intensity differences between the different views. Furthermore, bead based registration does not require a large sample with variations of signal intensities, but can be performed on single cells, the only requirement is the acquisition of beads in the sample or in a separate acquired position. Additionally, the PSF can be extracted from an image stack displaying beads for all acquired views. Bead based registration and deconvolution allows to applying the PSF directly to the processed image and compensates for minor optical aberrations in the detection path, as these are also affecting the PSF and therefore deducted from the output image.

A standard approach in multi view SPIM imaging with bead-based registration would be to mix fluorescent beads into the mounting gel of the sample. In this project, we physically separated beads from the sample and imaged stacks of beads immersed in Matrigel separately, which allowed us to detect registration matrix and PSF independently from any sample. We then applied these factors onto all other samples acquired at the same acquisition interval, without impairing image quality by high signal beads within the sample (Supplementary Figure 2). Furthermore, this meant that the registration matrix only needed to be calculated once per

time point, reducing computational demands drastically. Additionally, minor optical artifacts in the detection path induced by a small shift in refractive index between culture medium and Matrigel were compensated for, as the PSF was not theoretically calculated from the wavelength and objective properties but was extracted from imaging data. This improved overall image quality in long-term acquisitions, as minor shifts in the excitation and detection path originating from microscope de-alignment were compensated for. Additionally, registration matrix and PSF can be transferred between different color channels illuminating other fluorophores as used in Figure 8 with only minor increase in image blurring, again reducing required calculation steps in image processing. Additionally, we were not able to purchase fluorescent beads with sufficient signal and required bead size for the blue (405 nm) channel primarily used for DAPI / Hoechst (nuclei) staining. The transfer of PSF and registration matrix from the green channel (488 nm) to the blue (405 nm) channel allowed us to process the short wavelength (blue channel) data in Figure 24. The missing chromatic correction of the objective in the short wavelength range did not alter the image quality significantly.

4.1.4   High-throughput image processing

Besides simplifying sample preparation and imaging, we additionally focused on facilitating image processing (Figure 11). The established pipelines where developed for processing of any large diSPIM data while remaining open-source throughout. In the future, this developed workflow will allow any dual view light sheet user to easily process their own data without the need to script their own image processing pipeline, a major hindrance and concern for researchers to use the novel light sheet technology.

4.1.5   Image classification by deep learning

Grouping of individual images or subparts of an image, called image classification, is a key component of any quantitative image analysis [140, 177, 178]. We compared the effectivity of classical machine learning methods (random forest) to the novel convolutional neuronal network image classification methods (Figure 12). In terms of accuracy and processing speed, the used CNN classifier outperforms the RF classification significantly. On the other hand, CNNs provide the challenge to generate a sufficient sized training data set with appropriate but diverse sample images. These large data sets need to be generated *a priori* and with accurate selected data, in order to train the CNN to perform accurately. Additionally, CNN properties and setup have to match the application, as shallow networks tend to be inaccurate while deep networks require too much computational performance [151, 178]. Additionally,

empirical knowledge on CNN application in life science and microscopy data analysis is restricted, resulting in a trial-and-error approach to CNN implementation in the analysis workflow. On the other hand, demands of an imaging screen analysis of high-throughput und high accuracy with strong robustness towards minor imaging artifacts, signal loss, and image noise are met exceptionally by CNN image classification as seen in the results of this project (3.2.5). We believe that any future image segmentation, processing, classification, and analysis will be done using CNN as they provide strong performance improvements in comparison to classical image processing tools.

4.1.6   Phenotype assessment through feature evaluation

This project was conceived as a proof-of-concept and evaluation of different methods in gene regulation, phenotype imaging, and analysis. Therefore, we selected target genes on the basis of correlative knock-down methods (Figure 9). We aimed to induce gene knock-down on a transcriptional level through methylome modulation and on a translational level with siRNA to be able to compare effectivity and phenotype of these two molecular methods. Throughout, we focused on genes involved in the cell cycle, as gene regulation highly influences cell cycle progression, the role of mitosis in cancer initiation and cellular development [179, 180]. We set a high threshold on plausible effectivity of the knock-down, described by strong knock-down phenotype and high correlation of CpG-me and gene expression, resulting in 28 possible target genes (Supplementary Tables 1). Due to the fact, that we only selected for genes with a known knock-down phenotype allowed us to compare the detected results for all methods of gene expression reduction to known phenotypes provided by online databases such as MitoCheck and CycleBase.

Upon knock-down of the target genes by siRNA, we analyzed the phenotype by a total of 23 features that describe both global spheroid features, as well as features describing the single nucleus (Supplementary Tables 2). This combination of quantitative features allowed a comprehensive understanding of single nuclei and their underlying characteristics as well as the context of the spheroid as a global growing tissue by analyzing spheroid growth, development in reminiscence of the induced phenotype.

The negative, non-coding transfection control siRNA used as reference throughout this project showed even under the established high-content sample preparation and imaging conditions a close to normal development of the five-day old spheroid growing in Matrigel clonally from a single cell (Figure 8). The siRNA transfection as well as the described low phototoxicity of the diSPIM system did not interfere significantly with the development of the spheroid even when imaged at high temporal and spatial resolution.

The positive siRNA transfection control targeting INCENP allowed to evaluate the high efficiency of the sample preparation and HCS as we could detect and compare the knock-down phenotype of INCENP in all acquired samples. The macronuclei phenotype detected by the large segment size, reduced growth and abnormal cell cycle progression with elongated mitotic phases is highly reminiscent with the described gene function in the chromosomal passenger complex and knock-down phenotype of INCENP [134, 181-183]. The macronuclei formation originates as described and detected in the HCS images through the inhibited cytokinesis while the cell enters repeatedly the mitosis as the correct chromosome alignment and segregation is misregulated through the loss of function of the chromosomal passenger complex. In the INCENP samples, we always detected two separate macronuclei (Figure 19). We postulate that after knock-down sufficient functional INCENP protein remained in the cell to complete one additional mitosis with completed cytokinesis.

Other strong knock-down phenotypes of PLK1, MAP7, DSE, PRC1, BUD31, ESYT2 and AURKA such as found in cluster 5 (Figure 19) all resulted in significant growth reduction through mitotic arrest or prolongation that we were able to detect and quantify by the cell cycle analysis, nuclei and spheroid size and growth as well as global feature descriptors (Figure 14 - 19). These phenotypes strongly resemble in their key elements the described phenotypes of the used online databases (genecard.org, MitoCheck.org, cyclebase.org) as well as match with known gene functions as discussed below.

PLK1 is a serine/threonine-protein kinase, which plays an essential role in centrosome maturation and spindle assembly as well as the removal of cohesins from chromosome arms [184]. A knock-down of PLK1 is known to arrest the cell cycle in prometaphase [134, 148] which was described in the HCS by significant increased pro- and metaphase occurence, growth arrest and increased abnormal cell cycle transition (Figure 19).

MAP7, a microtubule-associated protein with a known function in microtubule stabilization, microtubule function regulation as well as cell polarization is key in the cell cycle as the correct distribution of the chromosomes to the daughter cells relies on the microtubule spindle apparatus [185, 186]. We postulate, that the loss of MAP7 proteins in the cell could result in a loss of cell polarization prior to the distribution of the chromosomes to the cell poles as well as uncontrolled chromosome segregation resulting in loss or gain of chromosome number subsequent to the cell division. This hypothesis is underlined by the strong increase of the prophase duration as the cell potentially struggles to polarize and form the spindle apparatus (Figure 14).

PRC1 or Protein Regulator of Cytokinesis 1 is a key regulatory protein of cytokinesis and spindle formation. PRC1 functions in cross-linking antiparallel microtubules and shows

essential functions of defining the spatiotemporal formation of the mid-body. Without the mid-body defining the region of final separation of the daughter cells, a completion of the mitosis is not possible [187]. These two functions have been detected by the siRNA HCS as we found a strongly increased ratio of nuclei segments classified as pro- and anaphase (Figure 14, Supplementary Figure 6) while samples also failed to grow describing an arrest of the cell either at prophase or in the final stage of cytokinesis.

BUD31 or Functional Spliceosome-Associated Protein 17 has a described function in mRNA splicing, an essential step of mRNA maturation and protein synthesis. A loss of all protein synthesis in a cell has a dramatic effect on the cellular homeostasis as all cellular processes rely on new synthesis of functional proteins due to constant protein turnover [188]. Loss of BUD31 results in fast loss of many mitotic proteins as these proteins display a higher turnover rate resulting from cell cycle stage specific expression in comparison to other housekeeping genes in frequently dividing cell culture model systems. Loss of growth and cell cycle progression detected in the HCS (Figure 18) can be the plausible result of mRNA maturation defects.

AURKA or Aurora Kinase A is a serine/threonine kinase with a well-known function in regulating the cell cycle. It establishes the mitotic spindle apparatus, induces centrosome duplication, centrosome separation as well as maturation. AURKA is essential for chromosomal alignment in metaphase, as well as the separation of the daughter cells [189]. The detected increased number of all mitotic stages (Figure 14) as well as reduced growth capabilities of the samples (Figure 18) allows us to postulate and correlate a complete delay of the different steps of mitosis to the function of AURKA in all mitotic phases.

ESYT2 or Extended Synaptotagmin 2 forms tethers to bind the endoplasmic reticulum to the cell membrane but has not been described directly to be associated with the cell cycle or mitosis [190]. Also, DSE or Dermatan Sulfate Epimerase has no close relation with the cell cycle as it functions as tumor rejection antigen with a specific catalytic function of converting D-glucuronic acid to L-iduronic acid during the biosynthesis of dermatan sulfate [191]. Downstream associations of DSE or ESYT2 with a specific step of the cell cycle cannot be excluded, that could explain the detected reduced growth and prophase arrest, but further functional experiments need to be conducted.

In total, we can correlate in almost all analyzed target genes of cluster 5 the known and published gene function with the detected knock-down phenotype, with the exception of ESYT2 and DSE. These results highlight the capabilities of the conducted HC screen. In comparison to published screens such as the MitoCheck consortium, the 3D cell culture models and the light sheet imaging methods provide a more detailed, sensitive and reliable

insight onto the target gene function and the knock-down effect onto the phenotype (3.3.8). Subtle heterogeneity among evaluated cells in terms of knock-down level and genomic heterogeneity in the cell culture model resulted in distinct phenotypes even within a single spheroid. This shows the strong and sensitive capability, this screening setup provides to study intercellular heterogeneity upon treatment, while also providing conclusive insight into the gene knock-down.

4.2 Knock-down using epigenome modifying molecular tools

A siRNA gene knock-down allows for efficient gene function evaluation, but it is usually unresolved how the different methods can influence the readout of gene function and whether novel methods can compete in effectivity and usability with established siRNA knock-down approaches. Additionally, the mode of action of the molecular tool, targeting the transcription or translation of the gene, can allow for compensating mechanisms. We therefore compared the detected phenotypes induced by siRNA-mediated gene knock-down to a novel, CRISPR-dCas9-ED based regulatory molecular tool, as well as compared their practical properties.

The deactivated CRISPR-Cas9 protein targeting enzymatic effector domains towards defined genomic regions allows the genomic site directed application of many different effector domains with their defined molecular functions. Site directed modification of the epigenome allows potentially lasting modifications of epigenetic properties of the cell, as the epigenome is like the genome highly conserved even throughout cell division to retain cellular properties and characteristics [82, 192].

Problematic in this experimental design is the lack of definite knowledge on regulatory mechanisms, feedback between genome and epigenome and how the dCas9-ED is capable of influencing the cellular homeostasis [44, 53]. Furthermore, additional or novel compensating mechanisms cannot be excluded. Bioinformatic analysis on one hand allow to detect the connection between CpG methylation level and gene expression, as well as description of other features of the target site such as regulatory factor binding sites or 3D genome interactions [193]. But on the other hand, complex regulatory mechanisms, cellular feedback cycles, low method effectivity or off-target effects of the CRISPR-dCas9 protein can render straightforward theoretic approaches ineffective [52, 194, 195].

To date, many different applications of the deactivated CRISPR-Cas9 targeting different effector domains such as DNMT3a [61, 171], TET1 [62, 63], Krüppel-associated box (KRAB) repressor [196-198] and p300 core protein [93, 196] have been used to modify the epigenome while the direct comparison to siRNA-mediated gene regulatory methods are excluded from these studies. In this project, we directly compared the effect of both molecular methods on

the cellular development focusing on the used set of mitotic target genes. This enabled us to evaluate siRNA and dCas-ED site directed targeting in their effectiveness and applicability.

To detect the methylome modifications and provide the highest success premises, we opted to provide the best prerequisites for the experiment. Therefore, we decided to change into another cell culture line and conducted a 2D pre-screen to define the most effective effector domain and target CpG combinations. The change of model system on the other hand also had an impact on the comparability of the evoked effect on the cell and made any correlation more questionable. Comparison of methylation level and target gene expression partly can compensate the unknown influence of the exchange of the used cell line.

With the different CRISPR-dCas9-ED stably expressed in HEK293 cells, we could combine the active methylome-modifying molecular tool (sgRNA and dCAs9-ED) by simple and effective sgRNA expressing plasmid transfection. Developing the same experimental setup in MCF10A cells was not easily feasible as well as labor intensive, as this cell line is resilient to plasmid transfection and gene integration requires lentiviral transduction of the MCF10A cells (Figure 21).

Through combining different sgRNAs targeting CpGs with identical regulatory properties (correlated and anti-correlated CpG methylation and target gene expression level), we diminished the individual effect of a CpG methylation modification, but on the other hand we also reduced the number of samples to be studied (Table 5 / 3.4.1). Effective sgRNA transfection into the established dCas9-ED cell lines, high-throughput image acquisition, and image analysis based on deep learning classification allowed for an effective evaluation of many different phenotypes in comparison to untransfected cells (3.4.4). Additionally, we compared the dCas-ED with the sole localization of the dCas to a target site perturbation (Table 3).

4.2.1   Epigenetic modulation detected by cellular phenotype

In total, we detected only in few (18%) sgRNA-dCas9-ED combinations a strong phenotype evoked by dCas9-ED targeting against the genome. All of these evoked, strong, abnormal nuclear shapes have been described by online databases or match with results obtained in the conducted HC siRNA screen, when the same gene was targeted by different means of modulation.

In 30 % of these cases, a binding of the dCas9 without any effector protein to the regulatory site or TSS was sufficient to induce a knock-down phenotype (Table 3) describing the potential of CRISPRi. We presume, that binding of the dCas to the genome is on its own capable to prevent regulatory proteins from DNA binding at this regulatory domain.

Furthermore, we were successful in applying the different dCAs9-ED to induce a significant mitotic aberration that correlate strongly with the known (online data base) and in this project described (siRNA HCS utilizing light sheet imaging) phenotypes with very high correlation.

These results provide the assumption that knock-down of the target gene was successful, especially for RGMA but also for other target genes.

On the other hand, 12% of the phenotypes were evoked with sgRNA-dCas9 combinations that theoretically should not be able to have an effect on the methylome and therefore should not induce abnormal phenotypes in these cells.

In conclusion, the methylome modification 2D pre-screen did not display adequate effectivity of the dCas-ED comparable to other established methods such as siRNA in inducing target gene knock-down phenotypes. CRISPRi, the localization of a large protein to a genomic domain deregulating this region shows high effectivity in comparison to the desired effect of methylome modulation. Underlying effects of the site-specific methylome alteration need further evaluation as well as regulatory mechanisms, that compensate the potential loss of gene expression. Potentially, each individual regulatory CpG needs to evaluated separately for dCas9-ED mediated deregulation. Furthermore, the obtained results reflect and highlight the key concern of the different described dCas9-ED molecular tools named in 4.2. The dCas9-ED have been first postulated in 2013 and established now for several years. The fusion of a large effector domain to a 160 kDa sized dCas protein results in a large protein that is questioned to be transported into the nucleus [199]. Furthermore, sophisticated delivery systems that avoid the transfection of two components into a single cell are being tested and slowly established but do not exhibit the efficiency of well-established molecular perturbations yet. Additionally, the biggest concern to date is the detection and evaluation of off-target effects. For gene knock-down studies the novel dCas9-ED tool as shown cannot compete in effectivity, usability and reliability with a siRNA knock-down, yet.

In total, we must conclude that we were able to evoke upon targeted gene deregulation with both used methods the same visual phenotype in almost all samples with strong knock-down phenotypes, but experimental difficulties and limited understanding of the molecular processes of the dCas9-ED described above, make epigenetic remodeling by dCas9-ED problematic for screening purposes.

On the other hand, the methylome modification allows for the first time through their targeted localization a functional, experimental evaluation of epigenetic regulatory elements and verifies bioinformatics predictions with a functional experiment such as we were able to conduct when targeting RGMA regulatory elements. The potentials for single gene

modulations are extensive, as the deeper level of transcriptomic modulation adds many possibilities to the research, allowing for much more complex genomic and regulatory functional evaluations.

4.2.2 dCas9-ED mediated RGMA knock-down

RGMA is a glycoprotein that exhibits a function as an axon guidance protein [200]. Knock-down of this protein in cell culture models induced a binucleated or macronuclei phenotype ([201]/ Mitosys ID: MCG_0007587), as detected in the different knock-down screens. In 2018, a study associated RGMA knock-down with high risk of luminal breast cancer and cancer progression [202, 203].

The methylome modification induced down regulation of RGMA did show the strongest correlation with the siRNA-mediated gene knock-down in the 2D pre-screen as well as in a subsequent in-depth analysis using long-term light sheet live imaging of 3D spheroids (Table 5). The evoked phenotype constitutes a reduced growth rate, elongated prophase and increased nuclei size resembling macronuclei formation described for this gene (Figure 24 / [201]).

For RGMA, the induced phenotype was only achieved in sgRNA-dCas9-ED combinations that align with the knock-down prediction by bioinformatics analysis of the target site and the appropriate effector domain function. This results in the assumption that we were successful in RGMA knock-down by methylome modification of predicted correlated and anti-correlated regulatory CpGs. The localization of a large protein to the TSS on the other hand was similarly successful.

The precise function and the effect of the loss of RGMA protein onto the cell cycle progression is not yet fully understood. We postulate that the RGMA knock-down phenotype was induced through the described altered utilization of the bone morphogenetic protein (BMP) type II receptors described by Xia and colleagues [204]. The authors describe a RGMA functions in the BMP pathway as co-receptor. A loss of the RGMA protein in the BMP signaling pathway can therefore have a dramatic effect on the intercellular interaction and consequently on the development of the spheroid model system through the BMP signaling.

4.3 Comparison of siRNA and dCas9-ED as molecular regulator

In total, we can conclude that in some cases the dCas9 fused with a methylome-modifying effector domain can provide the same effectivity as established knock-down methods and provides similar insight onto gene functionality as siRNA-mediated knock-down. But these results need to be carefully examined, replicated and the cause of the detected phenotype

needs to be validated. Comparison of the visual phenotype by light microscopy is insufficient to provide conclusive insight into gene function, mode of action and molecular pathways, but provides an exceptional assessment of the effect onto cellular processes and allows correlating gene function and cellular morphology. A combinatorial approach, such as combining RNAi with site directed positioning of epigenetic effector domains can validate and enrich knock-down experiments as well as insights into regulatory mechanisms.

For studies analyzing the individual effect of a single CpG or a regulatory region in the context of its epigenetic regulatory properties, the different effector domains provide an exceptional opportunity to conduct functional analysis and challenge bioinformatic predictions. Especially for these defined functional experiments, the additional effort necessary to introduce a functional dCas9-ED in the cell is valid, but for knock-down high-content screens the best and simplest method of choice will remain RNAi.

5 Conclusion and Outlook

In vitro cell culture systems are inevitably necessary models for human molecular biology research and translational medicine. 3D culture models, an innovative advancement from 2D cell culture, display more physiological *in vivo* characteristics, while posing significant challenges to novel and established analysis methods. These challenges have to be overcome to utilize the profound advantages of 3D cell culture models.

The application of light sheet microscopy in this project to study 3D spheroid or organoid systems is highly beneficial due to the improved 3D imaging capabilities of the light sheet microscopes and the rapid acquisition capabilities provide high-throughput screening possibilities. Drawbacks of light sheet microscopy, mostly in data handling and image processing were addressed. High-content image processing pipelines, combined with state-of-the art deep learning image classification are highly accurate and provide quantitative readout to evaluate treatment induced phenotype detection. The biggest advantage of the described screen was the individual analysis of single cells developing over time to spheroids and analyzing their properties.

Furthermore, we developed and test two highly innovative methods to modify cellular properties and study them by the named 3D microscopy and image analysis. Conclusive insight on target gene function using the CRISPR-dCas9 based epigenetic modulation were not possible. The mode of action for the dCas9-ED requires in-depth evaluation of the underlying cellular processes and therefore require target site-specific evaluation.

In the future, we expect to see a trend towards combining imaging and sequencing tools on a single cell level, where the phenotype and transcriptome are analyzed in a correlative manner. Additionally, deep learning image processing will be incorporated in many parts of the experimental process, from sample selection, to image acquisition, to microscopy control, to image analysis of raw data, and finally correlative analysis of sample properties and sequencing data, fusing visual and sequencing data into a single analysis step. A direct usage of deep learning classification of defined phenotypes such as macronuclei formation, ACE or tripolar cell division as well as other phenotypes was not feasible in this work as sufficient training data could not be provided, describing the difficulties arising from deep learning implementation into image analysis. Improvement in acquisition speed, imaging capabilities and integration of state-of-the-art illumination technologies such as structured illumination, will further increase throughput and image quality and resolution in HCS. With these technical improvements, a more detailed and extensive analysis will be possible by light sheet microscopy screening describing subcellular processes, structures and components in

unmatched detail. This will allow researchers to comprehensively study in a single experiment many different functional aspects in a highly quantitative fashion.

Furthermore, we addressed the effectivity and capabilities of the novel, site-directed methylome-modifying effector domains and compared siRNA and regulatory CpG methylome modulation based gene knock-down phenotypes. Though being capable of inducing a gene knock-down, the effectivity and usability of the dCas9-ED does not rival that of conventional knock-down methods such as siRNA, yet. The long-term capabilities to induce lasting gene expression modulations are appealing, but need to be validated. We expect in the future an improvement of the site directed methylome modification, by improved protein and RNA delivery into the cell through reduced protein size and optimized delivery systems. Furthermore, we expect improved understanding and prediction of the dCas9 fused to different effector domains and its properties such as effectivity, off-targeting, long-term capabilities and application capabilities. The capability of the different effector domains to selectively modify epigenetic traits at defined genomic loci will provide in the future significant insights into the regulatory networks of the epigenome and the transcriptome as well as detailed descriptions of regulatory regions.

6 Materials and Methods

6.1 High-content volumetric imaging screen utilizing advance light sheet 3D microscopy

6.1.1 Cell culture of MCF10A H2B-GFP

MCF10A H2B-GFP cells [41] (P25 to P31) were cultured and passaged in 25 cm^2 culture flasks (Greiner bio-one) in DMEM/F12 (ThermoFisher Scientific #11039) with supplements (5% horse serum, 10 μg/ml Insulin (Life Technologies), 20 ng/ml EGF, 0.5 mg/ml hydrocortisone and 100 ng/ml Cholera Toxin (Sigma)) under standard culture conditions (5% CO$_2$ / 37 °C) and passaged with 0.05% Trypsin (Life Technologies) in 2D at 80-90% confluency every three to four days.

6.1.2 Solid-phase reverse siRNA transfection

Solid-phase reverse transfection siRNA transfection mix was prepared as described by Erfle *et al.* [181]. In short: 4.00 μl OptiMEM (ThermoFisher Scientific #51985026), containing 1.82 M trehalose dehydrate (Merck, #T9531), was added in a single well of a 384-well plate. Then 2.50 μl Lipofectamine® RNAiMAX (ThermoFisher Scientific #13778075) transfection reagent and 5μl of a 5 μM siRNA solution were added and incubated for 5 min at room temperature to allow complex formation. Finally, 7.25 μl Collagen type IV solution (Merck #C5533) (1 mg/mL in ddH2O) was added and the transfection solution was gently mixed. The solution was diluted in 468.75 μl ddH2O of which 35 μl were transferred to a well of a 96-well plate. The plates were placed in a -20°C freezer for several hours, followed by a drying step (lyophilization or speed-vac), and were stored at room temperature until cell seeding.

Trypsinated MCF10A H2B-GFP were diluted in growth medium to a density of 5x10^5 cells/ml. 10'000 cells in 100 μl cell suspension were added to each well of the solid-phase reverse transfection mix. After five hours, cell medium was removed and cells were resuspended by directly adding 50 μl 0.25% Trypsin (Life Technologies #25200056) to each well.

6.1.3 High-content cell spotting in Matrigel

Mixing and spotting of cells into OneWell plate (Greiner bio-one CELLSTAR® OneWell Plate™ #670180) with Matrigel (Corning Matrigel Matrix) by an automated liquid handling robot from Hamilton Robotics following was conducted, following a customized protocol. In short: From each cell suspension transfected with individual siRNA, 60 isolated cells in 3 μl medium were mixed with 10 μl Matrigel. Subsequently each mixture was spotted eight times with a single spot volume of 0.2 μl in two columns by four rows array, giving a total of 320 spots in 40 columns and eight rows array centred on the imaging plate. One sub-array of spots was always

dedicated for beads (ThermoFisher Scientific #7220) mixed with Matrigel, used for registration of the two acquired views. Positioning of each spot is identical with the positions of a standard 1536-well plate. After 10 minutes at 37°C for Matrigel solidification culture/imaging medium was added and samples were incubated under standard culture conditions until imaging.

6.1.4 diSPIM imaging

Imaging was conducted with a dual-view inverted selective plane illumination microscope (diSPIM) as described by Kumar *et al.*, 2014 [112]. The microscope was equipped with LMM5 laser (Spectral Applied Research Laser Illumination Laser Merge Module 5) and AHF Quand Filterset (F59-405 / F73-410 / F57-406). Images were acquired by two water-cooled ORCA-Flash4.0 Hamamatsu sCMOS cameras. Cooling was provided with Julabo F250 cooling circuit. Standard culture conditions were generated by (3i ECS2) and direct airflow over the SPIM head was minimized to avoid unnecessary vibrations. All imaging time lapse acquisitions were conducted with 320 µW laser power for 488 nm excitation wavelength (measured at the sample).

6.1.5 Low resolution pre-screen

Readjusting the fine alignment of the microscope was conducted shortly before the start of the acquisition with special attention to retain temperature within the microscope chamber.

To detect each spheroid's positions and select the spheroids to be imaged, we conducted a low resolution, fast stage-scan pre-screen. A grid of imaging positions was spanned across the imaging plate, each position placed at the center of one column of spots. As the automated spotting process resulted in spots with defined positions and sizes, we were able to repeatedly use the same grid of stage scan acquisition positions for every pre-screen. Potentially due to small manufacturing differences of the imagining plate, we solely needed to adjust the general Z-position off-set, underlining the robustness of our sample preparation process. Each position acquisition resulted in a $X_{microscope}$-Stack of 1'200 slices with a step size of 5 µm and a pixel resolution of 0.648 µm/px and a field of view of 333 µm.

The acquisition of the pre-screen took 31 minutes and produced 96'000 images. This pre-screen data was subsequently analysed by a KNIME image processing workflow detecting the $XYZ_{microscope}$ position of each cell cluster, size and shape. Per spot we detected an average of 2.6 spheroids. From this data, a ranking of spheroids was conducted, based on their Z-position, discarding small, flat and elongated cell clusters selecting for the spheroids with the maximal Z-coordinates. The KNIME workflow file can be found for review and further development on the Isilon server (\diSPIM_prescreen_stagescan_Pos_analysis.knwf).

The high positioning of the selected spheroid in the Matrigel spot ensured maximal image quality, as no additional cells obstructed the illumination and detection path. These 38 defined spheroids with individual treatment plus two positions with fluorescent beads that were essential for later image processing steps were imaged with the diSPIM microscope for 24 hours at maximal temporal and spatial resolution for treatment evaluation. Incubation conditions in the microscope were improved to resemble as close as possible the standard culture conditions of an incubator.

6.1.6 Position scan acquisition

Imaging parameters for dual view synchronous piezo/slice scan (stack acquisition) were set, to acquire two stacks of 1024 px^2 in XY $_{image}$ with the maximal camera resolution of 0.1625 µm/px centered to the field of view of the camera and 260 slices in Z $_{image}$ with a slice interval of 0.5 µm starting with view A (right camera acquisition). Sample exposure was set to 1.5 ms and the option for "Minimized Slice Period" enabled. The option for "Autofocus during acquisition" was enabled with the autofocus running on the registration position imaging beads every acquisition cycle with 40 slices acquired every 0.5 µm. The off-set was detected by with "Volath" script.

Preselected positions copied from the KNIME analysis of the pre-screen were checked and if necessary manually corrected. An additional registration position was added as first and last position, imaging beads mixed in Matrigel.

Due to the acquisition limitations of the microscope of minimal four to five seconds per position scan and stage repositioning, we imaged at an interval of five minutes for 24 hours, resulting in a total data volume of 10.06 terabytes (TB), which was stored locally. The high temporal resolution was essential for tracking of nuclei as they progressed through the cell cycle. Throughout time lapse acquisition, we did not need to adjust for any position off-set introduced by deformation of the matrigel or external influences as samples remained almost exclusively in the FOV.

6.1.7 Image processing (hSPIM)

Raw data was processed by a custom software named "hSPIM", designed specifically to the acquisition geometry of the di-SPIM and the separate acquired registration position imaging beads in 3D. In short: The registration matrix of the two views is detected for each time point of the screen by registration of beads in 3D. Additionally, the PSF is extracted. This registration matrix and PSF are stored and used for registration and deconvolution of all other acquired positions of this time point. Furthermore, the software performs a segmentation of the nuclei, from which different geometrical and textural features are extracted for each segment.

Deconvolved fusion images and segment images as well as segment and feature table are stored and were used for further image analysis.

In addition, further functions are implemented into the hSPIM software. These can enable the user to directly visualize in 3D a registered and deconvolved snap image or to store a view angle and export a 3D movie of a single position.

6.1.8 High-content KNIME analysis workflow

Following the raw image processing, we have developed a workflow to analyze key cellular and global properties of each spheroid throughout the acquired time lapse by image analysis.

XYZ $_{microscope}$ displacement: By tracking the positions of single beats over time in the beat sample we could detect the global offset in all dimensions of the microscope introduced through fine displacements of the imaging plate or component expansion of the di-SPIM components.

Cell phase nuclei classification: For a precise nuclei classification, we used a CNN, trained with a selection of manually classified images by means of 3D-2D plane DL classification described in 3.2.5. The CNN calculated the probability for each of the four distinguishable cell cycle phases for each XY, XZ and YZ slices: Interphase, pro- and prometaphase, metaphase, anaphase. The class with the highest sum in likelihood for each segment was selected as defining class for this nucleus at this time point.

Clustering of segments into spheroids: To segregate segments from two spheroids acquired at a single imaging position into individual spheroid clusters, we analyzed the geometric distance of each segment to all others and clustered accordingly.

Spheroid size: The clustering enabled us to combine all segments of one spheroid for an image based size calculation.

Geometric nuclei class features: In the global context of the single spheroid we analyzed the single nuclei size, the intensity, the position of the segment from the center of the spheroid and the predicted class, over time.

Nuclei migration speed: By tracking the position of each segment over time, we also analyzed the median migration speed of all cells in each spheroid.

Time lapse movie: For individual evaluation, we exported the maximum projected time lapse movie of each position including cell cycle classification and spheroid hull.

The KNIME workflow file can be found for review and further development on the Isilon server (\diSPIM_phenotype_screen_analysis_3D_spheroids.knwf) as well as the template folder structure for hSPIM- and KNIME analysis workflow output

(\TLXXXX_S_hSPIM_Output_Template) with XXXX being place holders for the screen number.

6.1.9   Spheroid feature evaluation

Selected image features as quantified by the KNIME workflow (Supplementary Tables 2) were subsequently subjected to further quantitative analysis in R. Additional quantitative features such as cell volume (estimated as the ratio of spheroid volume to nuclei number) and average nuclei size in different cell phases were additionally computed from the quantified list of features. The fraction of cells detected in different cell cycle phases was averaged across all time points.

To calculate instantaneous spheroid growth rates from nuclei numbers, the number of nuclei over time was smoothed using the lowess function with parameters f=1/3, iter=3L, delta=0.01 * diff(range(NrCells[1:n.rows[[pos]],pos])), and differentiated using the diff function.

All feature measurements from all plates were combined into one matrix, centered by subtracting the column means from their corresponding columns, and scaled by dividing the centered columns their standard deviations. As mechanical plate drift resulted in spheroids lying partially outside the field of view in the case of plate 00941, affected features (nuclei and spheroid growth rate, compactness, convexity, sphericity, spheroid volume and cell volume) were excluded for this plate.

To identify clusters of siRNAs causing similar phenotypes, rank-based clustering and tSNE embedding were performed using the rank, dist, hclust and Rtsne functions. Heatmaps were created using the heatmap.2 function from the gplots package or the aheatmap function from the NMF package. t-tests for statistical comparisons were performed using the t.test function

6.2   dCas9-effector domains

6.2.1   Construct synthesis

All three used dCas-ED constructs (C49 – dCas9; C54 – DNMTA3-dCas9; C57 – TET1-dCas9) were assembled by Gibson Cloning (NEB #E2611) following manufacturer guidelines. Source constructs were obtained from AddGene: dCas9 [85], DNMT3a [171], TET1 [205]. Linker (GGGGS), NLS (PKKKRKV) and M2-Flag (DYKDHDG) DNA sequences as well as adapter primers were ordered from Eurofins Genomics. Successful cloning was assessed by sequencing by GATC Biotech AG following general guidelines, western blot (M2-flag) and expression in HEK293 cells detected by immunostaining (6.2.5).

Plasmid maps and construct components can be reviewed here (\EpiToolConstructs_synthesis)

construct	effector domain	description
C49	non	pcDNA_Tag_NLS_dCas_linker_STOP
C54	DNMT3a	C54_pcDNA_Tag_NLS_dCas_Linker_Dnmt3a
C57	TET1	C57_pcDNA_Tag_NLS_dCas_Linker_Tet1

6.2.2 Stable dCas9-ED expression in HEK293 cell line

5×10^5 HEK293 cells were transfected with 5 μg plasmid DNA of the different dCas9-ED constructs (C49, C54, C57) with Lipofectamin 2000 (Invitrogen) following manufacturer guidelines. 48 hours post dCas9-ED plasmid transfection, transfected cells were selected for by addition of G418 (Geneticin) selection antibiotic to the culture medium at a concentration of 500 μg / ml. Stock cultures were frozen after four passages under constant G418 selection resulting in stable dCas9-ED expressing cell lines.

6.2.3 Single guide RNA synthesis

SgRNA expression plasmids were designed and synthesized following the recommended SAM target sgRNA cloning protocol by Konermann and colleagues [194]. In short: the sgRNA(MS2) cloning backbone (AddGene: #61424) was digested with BbsI. Oligos representing the sgRNA target site with 20 bases in sense (Os) with an CACCG overhang and anti-sense (Oas) with an AAAC overhang were ordered from Eurofins and annealed. For genome reference, we used the UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19). Backbone and sgRNA defining insert were joined by a Golden Gate reaction. The resulting plasmid was expanded by bacterial transformation and assessed by sequencing.

6.2.4 Stable dCas9-ED cell lines sgRNA transfection

The HEK293 cells were transfected with the different sgRNA constructs by solid-phase reverse transfection as described in section 6.1.2, with Lipofectamine RNAiMAX being replaced by Lipofectamin 2000 (Invitrogen #11668027).

6.2.5 Immunostaining of HEK293 cells for DNA, sgRNA and dCas9-ED

HEK293 cells were fixed and stained at different time points between 3 and 9 days after transfection for the two components of the functional dCas9-ED by IF staining. Cells were

fixed by 4% PFA (Sigma-Aldrich #F8775) for 10 minutes in PBS with 0.5 % triton x-100 and blocked subsequently by 1% goat serum in PBS applied overnight. We used the mouse anti-Flag M2 monoclonal antibody from Sigma (#F1804) as primary and the goat anti-mouse Alexa 568 secondary antibody from Invitrogen (#A11004) to label the dCas9-ED in red. To detect the successful transfection with the sgRNA plasmid we applied rabbit anti-GFP monoclonal antibody from Cell Signaling (#2956) as primary and goat anti-rabbit Alexa 488 from Molecular Probes (#A11034) as secondary antibody. The DNA was stained by DAPI. Careful intermediate washes with PBS removed excess antibodies.

6.2.6 Confocal imaging of IF stained epigenome targeted HEK293 cells

Confocal imaging was conducted using the Zeiss LSM 780 under the control of AutofocusScreen macro (http://www.ellenberg.embl.de/apps/AFS/, 24.02.2016) acquiring 25 Z-stacks per well with each comprising five slices per dCas9-ED-sgRNA combination (one dCas9-ED / one target gene). Each stack was acquired with a bright field image additionally to the DAPI (405 nm), sgRNA (488 nm) and dCas9-ED (568 nm) channel.

6.2.7 Image analysis of dCas9-ED HEK293 cells and phenotype evaluation

Raw HEK293 images of each sgRNA-dCas9-ED combinations were smoothened by Gaussian convolution and single nuclei were segmented by Otsu thresholding. Single segments were further processed and spitted if necessary by segment erosion. Single nuclei were classified by a DNN into cell cycle stages (inter-, pro-, met-, anaphase) as well as significant phenotypes (macronuclei and apoptotic condensed DNA). Additionally, the transfection state of the cell was evaluated by the presence of an dCas9-ED (M2-flag) and the sgRNA (GFP). Cells with both dCas9-ED components present were labeled double positive. KNIME workflow can be found here: (\EpiTool_confocal_nuclei_classification.knwf)

Detected classes were further evaluated in comparison to non-targeted sgRNA transfected dCas9-ED cell lines as well as to non-transfected cells. All acquired time points (3, 5, 7 days post transfection) were combined during analysis. Cells with a significant increase (> 1.5 ratio) of a class between control cells were highlighted. KNIME workflow can be found here: (\EpiTool_Class_quantitative_analysis.knwf)

6.3 Materials

6.3.1 Hardware

6.3.1.1 Workstation

hardware	supplier	description
CPU	Intel	i9-7980XE
GPU	NVIDIA	Titan xp 12 GB
Hard drive (RAID0)	WD	WD-Red 8 TB
RAM	ECC	64 GByte DDR-4 PC2400
Motherboard	ASRock	X299 Taichi
Controller	Intel	SATA Controller, 10x 6 Gbit/s
Hard drive	Samsung	1 TB 960 Pro

6.3.1.2 ASI diSPIM hardware

hardware	supplier	description / number
camera cooling	Julabo	F250
quand filterset	AHF	F59-405 F73-410 F57-406
sCMOS cameras	Hamamatsu	ORCA-Flash4.0
laser	Spectral Applied Research	Laser Merge Module 5 (LMM5)

6.3.2 Software and workflows

6.3.2.1 Software

name	version	description
KNIME	3.5.5	Konstanz Information Miner
hSPIM	1.0	diSPIM raw image processing tool
MicroManager	1.4	microscope control software
diSPIM plugin	NB_20180116	nightly build MicroManager diSPIM controller plugin

6.3.2.2 KNIME workflows

name	folder
diSPIM_prescreen_stagescan_Pos_analysis	Group\Bjoern_Isilon\BjoernEismann_PhD_Thesis_Data
diSPIM_phenotype_screen_analysis_3D_spheroids	Group\Bjoern_Isilon\BjoernEismann_PhD_Thesis_Data
EpiTool_confocal_nuclei_classification	Group\Bjoern_Isilon\BjoernEismann_PhD_Thesis_Data
EpiTool_Class_quantitative_analysis	Group\Bjoern_Isilon\BjoernEismann_PhD_Thesis_Data

6.3.2.3   Haralick features

Haralick feature	discription	Haralick feature	discription
F1	contrast	F8	sum entropy
F2	angular second moment	F9	entropy
F3	correlation	F10	difference variance
F4	sum of squares: variance	F11	difference entropy
F5	inverse difference moment	F12	measure of correlation I
F6	sum average	F13	measure of correlation II
F7	sum variance		

6.3.3   Source constructs

construct	source	description / number
#46911	AddGene	Gilbert_pHR-SFFV-dCas [85]
#71666	AddGene	pdCas9-DNMT3A-EGFP [171]
#49792	AddGene	FH-TET1-pEF [205]
#61424	AddGene	sgRNA(MS2) cloning backbone

6.3.4   Antibodies

description	source	number	description
anti-Flag® M2	Sigma	F1804	primary mouse anti Flag M2 monoclonal antibody
anti-GFP	Cell signaling	2956	primary rabbit anti GFP monoclonal antibody
Anti-rabbit Alexa 488	Molecular Probes	A11034	fluorescent secondary goat anti rabbit antibody
Anti-mouse Alexa 568	Invitrogen	A11004	fluorescent secondary goat anti mouse antibody

6.3.5 Consumables and solutions

description	supplier	product number
beads: PS-Speck™ Microscope	ThermoFisher Scientific	P7220
Cell Culture Plate, 96-Well	Eppendorf	0030730119
CELLSTAR® OneWell Plate™	Greiner bio-one	670180
Cholera toxin	Sigma-Aldrich (Merck)	
Collagen type IV solution	Merck	C5533
culture flasks (25cm^2)	greiner bio-one	
DAPI	Sigma-Aldrich (Merck)	D9542
DMEM/F12	ThermoFisher Scientific	11039
G418 (Geneticin)	Sigma-Aldrich (Merck)	4727878001
Gibson Assembly Master Mix	NEB	E2611
Insulin	Life Technologies	
Lipofectamin 2000	Invitrogen	11668027
Lipofectamine® RNAiMAX	ThermoFisher Scientific	13778075
Matrigel	Corning	354248
OptiMEM	ThermoFisher Scientific	51985026
PCR plate, 96 well	Kisker	G060
trehalose dehydrate	Merck	T9531
trypsin	Life Technologies	25200056

6.3.6 siRNA – Ambion

gene name	siRNA ID 1	siRNA ID 2
ATOH8	s39645	s39643
AURKA	s196	s197
BUD31	s17010	s17009
CDC6	s2744	s2746
CDCA5	s41424	s41425
CEP85	s34959	s34961
CTSB	s3738	s3739
DSE	s26749	s26750
EME1	s44946	s44945
ESYT2	s33138	s33136
F11R	s27152	s27151
FOXM1	s5250	s5249
LGR4	s30840	s229314
LHFP	s19847	s19848
LMNB2	s39477	s39476
LRP1	s8278	s8280
MAP7	s17263	s17262
MEIS2	s8666	s8664
MYC	s9130	s9131
PLK1	s448	s450
PRC1	s17268	s17269
RAN	s11769	s11768
RBBP4	s55169	s56872
RGMA	s32498	s32500
TCF7	s13877	s13878
TOP2a	s14307	s14308
TUFT1	s14510	s14509
WBP1	s24095	s225969

7 Appendix

7.1 List of figures

7.2 List of tables

7.3 Units

	abbreviation	unit
weights	pg	picogram
	ng	nanogram
	μg	microgram
	mg	milligram
volume	nl	nanoliter
	μl	microliter
	ml	milliliter
distance	nm	nanometer
	μm	micrometer
	mm	millimeter
	cm	centimeter
time	sec	seconds
	min	minutes
	h	hours
rotation force	rpm	revolutions per minute
	g	relative centrifugal force
temperature	°C	degree celsius
	RT	room temperature (~21 °C)
optics	λ	wavelength
	NA	numerical aperature
data	GB	gigabyte
	TB	terabyte
	MB	megabyte
	KB	kilobyte

7.4 Abbreviations

2D	two-dimensional
3D	three-dimensional
ASI	Applied Scientific Instrumentation
bNGS	bulk next-generation sequencing
bp	base pairs
BSA	bovine serum albumine
CLSM	confocal laser scanning microscopy
CNNs	convolutional neuronal networks
CO_2	carbon dioxide
CpG	cytosine guanine dinucleotide
CRISPR	clustered regularly interspaced short palindromic repeats
CSC	cancer stem cell
diSPIM	dual inverted single plane illumination light sheet microscopy
DL	deep learning
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferases
DNN	deep neuronal networks
dNTP	deoxynucleotide triphosphate
dps	days post seeding
ECM	extracellular matrix
ED	effector domains
EGF	epidermal growth factor
FEP	fluorinated ethylene propylene
FOI	field of interest
FOV	field of view
GFP	green fluorescent protein
HC	high-content
HCS	high-content screen
HEK 293	human embryonic kidney cell line 293
HeLa	Henrietta Lacks cell line
HT	high-throughput
HTS	high-throughput screen
IF	immunofluorescence

KNIME	Konstanz Information Miner
LSFM	Light Sheet Fluorescence Microscopy
me	methylation
MVR	multi view reconstruction
NA	numerical aperture
NB	nightly built
NLS	nuclear localization sequence
OPFOS	orthogonal-plane fluorescence optical sectioning
OTP	one-touch-pipeline
P	Passage
PBS	phosphate buffered saline
PBS	phosphate-buffered saline
PDTXs	patient-derived tumor xenografts
RF	random forest
RI	refractive index
RNA	ribonucleic acid
RT	room temperature
sCMOS	scientific complementary metal-oxide semiconductor
SCS	single cell sequencing
SDM	Spinning-disk (Nipkow disk) confocal microscopes
sgRNA	single guide RNA
SPIM	selective plane illumination microscopy
ssOPM	stage-scanning OPM approach
SSA	stage scan acquisition
SyncA	synchronous piezp/slice acquisition
TALEN	Transcription Activator-Like Effector Nuclease
TET1	Ten-eleven translocation methylcytosine dioxygenase 1
TP	time point
tSNE	t-distributed stochastic neighbor embedding
TSS	transcription start site
ZF	zinc finger DNA binding proteins

8 Bibliography

1. Ames, B.N., M.K. Shigenaga, and T.M. Hagen, *Oxidants, antioxidants, and the degenerative diseases of aging.* Proc Natl Acad Sci U S A, 1993. **90**(17): p. 7915-22.

2. Yoshida, K. and Y. Miki, *Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage.* Cancer Sci, 2004. **95**(11): p. 866-71.

3. Zaorsky, N.G., et al., *Causes of death among cancer patients.* Ann Oncol, 2017. **28**(2): p. 400-407.

4. Pampaloni, F., E.G. Reynaud, and E.H. Stelzer, *The third dimension bridges the gap between cell culture and live tissue.* Nat Rev Mol Cell Biol, 2007. **8**(10): p. 839-45.

5. Thoma, C.R., et al., *3D cell culture systems modeling tumor growth determinants in cancer target discovery.* Adv Drug Deliv Rev, 2014. **69-70**: p. 29-41.

6. Benien, P. and A. Swami, *3D tumor models: history, advances and future perspectives.* Future Oncol, 2014. **10**(7): p. 1311-27.

7. Jabs, J., et al., *Screening drug effects in patient-derived cancer cells links organoid responses to genome alterations.* Mol Syst Biol, 2017. **13**(11): p. 955.

8. Kurita, R., et al., *Establishment of immortalized human erythroid progenitor cell lines able to produce enucleated red blood cells.* PLoS One, 2013. **8**(3): p. e59890.

9. Ben-David, U., et al., *Patient-derived xenografts undergo mouse-specific tumor evolution.* Nat Genet, 2017. **49**(11): p. 1567-1575.

10. Pampaloni, F., B.J. Chang, and E.H. Stelzer, *Light sheet-based fluorescence microscopy (LSFM) for the quantitative imaging of cells and tissues.* Cell Tissue Res, 2015. **360**(1): p. 129-41.

11. Rios, A.C. and H. Clevers, *Imaging organoids: a bright future ahead.* Nat Methods, 2018. **15**(1): p. 24-26.

12. Hoarau-Véchot, J., et al., *Halfway between 2D and Animal Models: Are 3D Cultures the Ideal Tool to Study Cancer-Microenvironment Interactions?* Int J Mol Sci, 2018. **19**(1).

13. Yamada, K.M. and E. Cukierman, *Modeling tissue morphogenesis and cancer in 3D.* Cell, 2007. **130**(4): p. 601-10.

14. Sip, C.G., N. Bhattacharjee, and A. Folch, *Microfluidic transwell inserts for generation of tissue culture-friendly gradients in well plates.* Lab Chip, 2014. **14**(2): p. 302-14.

15. Maitre, J.L., et al., *Asymmetric division of contractile domains couples cell positioning and fate specification.* Nature, 2016. **536**(7616): p. 344-348.

16. Gomez-Lechon, M.J., et al., *Long-term expression of differentiated functions in hepatocytes cultured in three-dimensional collagen matrix.* J Cell Physiol, 1998. **177**(4): p. 553-62.

17. Kleinman, H.K., D. Philp, and M.P. Hoffman, *Role of the extracellular matrix in morphogenesis.* Curr Opin Biotechnol, 2003. **14**(5): p. 526-32.

18. Lasfargues, E.Y., *Cultivation and behavior in vitro of the normal mammary epithelium of the adult mouse.* Anat Rec, 1957. **127**(1): p. 117-29.

19. Timmins, N.E., et al., *Method for the generation and cultivation of functional three-dimensional mammary constructs without exogenous extracellular matrix.* Cell Tissue Res, 2005. **320**(1): p. 207-10.

20. Ghosh, S., et al., *Three-dimensional culture of melanoma cells profoundly affects gene expression profile: a high density oligonucleotide array study.* J Cell Physiol, 2005. **204**(2): p. 522-31.

21. Petersen, O.W., et al., *Interaction with basement membrane serves to rapidly distinguish growth and differentiation pattern of normal and malignant human breast epithelial cells.* Proc Natl Acad Sci U S A, 1992. **89**(19): p. 9064-8.

22. Bissell, M.J., et al., *The organizing principle: microenvironmental influences in the normal and malignant breast.* Differentiation, 2002. **70**(9-10): p. 537-46.

23. Meacham, C.E. and S.J. Morrison, *Tumour heterogeneity and cancer cell plasticity.* Nature, 2013. **501**(7467): p. 328-37.

24. Kunz-Schughart, L.A., et al., *The use of 3-D cultures for high-throughput screening: the multicellular spheroid model.* J Biomol Screen, 2004. **9**(4): p. 273-85.

25. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.

26. Herrup, K., *Post-mitotic role of the cell cycle machinery.* Curr Opin Cell Biol, 2013. **25**(6): p. 711-6.

27. Leung, C.T. and J.S. Brugge, *Outgrowth of single oncogene-expressing cells from suppressive epithelial environments.* Nature, 2012. **482**(7385): p. 410-3.

28. Logan, G.J., et al., *Molecular drivers of lobular carcinoma in situ.* Breast Cancer Res, 2015. **17**: p. 76.

29. Pal, A. and C.G. Kleer, *Three dimensional cultures: a tool to study normal acinar architecture vs. malignant transformation of breast cells.* J Vis Exp, 2014(86).

30. Lawrence, K.S., T. Chau, and J. Engebrecht, *DNA damage response and spindle assembly checkpoint function throughout the cell cycle to ensure genomic integrity.* PLoS Genet, 2015. **11**(4): p. e1005150.

31. Duijf, P.H., N. Schultz, and R. Benezra, *Cancer cells preferentially lose small chromosomes.* Int J Cancer, 2013. **132**(10): p. 2316-26.

32. Vargas-Rondon, N., V.E. Villegas, and M. Rondon-Lagos, *The Role of Chromosomal Instability in Cancer and Therapeutic Responses.* Cancers (Basel), 2017. **10**(1).

33. Bedard, P.L., et al., *Tumour heterogeneity in the clinic.* Nature, 2013. **501**(7467): p. 355-64.

34. Beier, D., et al., *CD133(+) and CD133(-) glioblastoma-derived cancer stem cells show differential growth characteristics and molecular profiles.* Cancer Res, 2007. **67**(9): p. 4010-5.

35. Bruna, A., et al., *A Biobank of Breast Cancer Explants with Preserved Intra-tumor Heterogeneity to Screen Anticancer Compounds.* Cell, 2016. **167**(1): p. 260-274.e22.

36. Dexter, D.L. and J.T. Leith, *Tumor heterogeneity and drug resistance.* J Clin Oncol, 1986. **4**(2): p. 244-57.

37. Norum, J.H., K. Andersen, and T. Sorlie, *Lessons learned from the intrinsic subtypes of breast cancer in the quest for precision therapy.* Br J Surg, 2014. **101**(8): p. 925-38.

38. Gibson, P., et al., *Subtypes of medulloblastoma have distinct developmental origins.* Nature, 2010. **468**(7327): p. 1095-9.

39. Zhang, Y., et al., *Toward the precision breast cancer survival prediction utilizing combined whole genome-wide expression and somatic mutation analysis.* BMC Med Genomics, 2018. **11**(Suppl 5): p. 104.

40. Esteller, M., et al., *Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors.* J Natl Cancer Inst, 2000. **92**(7): p. 564-9.

41. Debnath, J., S.K. Muthuswamy, and J.S. Brugge, *Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures.* Methods, 2003. **30**(3): p. 256-68.

42. Keung, A.J., et al., *Chromatin regulation at the frontier of synthetic biology.* Nat Rev Genet, 2015. **16**(3): p. 159-71.

43. Cohen, A.J., et al., *Hotspots of aberrant enhancer activity punctuate the colorectal cancer epigenome.* Nat Commun, 2017. **8**: p. 14400.

44. Probst, A.V., E. Dunleavy, and G. Almouzni, *Epigenetic inheritance during the cell cycle.* Nat Rev Mol Cell Biol, 2009. **10**(3): p. 192-206.

45. Heyn, H. and M. Esteller, *DNA methylation profiling in the clinic: applications and challenges.* Nat Rev Genet, 2012. **13**(10): p. 679-92.

46. Farrell, W.E., *Epigenetic mechanisms of tumorigenesis.* Horm Metab Res, 2005. **37**(6): p. 361-8.

47. Jones, P.A. and D. Takai, *The role of DNA methylation in mammalian epigenetics.* Science, 2001. **293**(5532): p. 1068-70.

48. Jones, P.A. and P.W. Laird, *Cancer epigenetics comes of age.* Nat Genet, 1999. **21**(2): p. 163-7.

49. Jurkowski, T.P., M. Ravichandran, and P. Stepper, *Synthetic epigenetics-towards intelligent control of epigenetic states and cell identity.* Clin Epigenetics, 2015. **7**(1): p. 18.

50. Potaczek, D.P., et al., *Epigenetics and allergy: from basic mechanisms to clinical applications.* Epigenomics, 2017.

51. Wei, L., et al., *Histone methyltransferase G9a promotes liver cancer development by epigenetic silencing of tumor suppressor gene RARRES3.* J Hepatol, 2017.

52. O'Geen, H., et al., *dCas9-based epigenome editing suggests acquisition of histone methylation is not sufficient for target gene repression.* Nucleic Acids Res, 2017. **45**(17): p. 9901-9916.

53. Jones, P.A., *Functions of DNA methylation: islands, start sites, gene bodies and beyond.* Nat Rev Genet, 2012. **13**(7): p. 484-92.

54. Kinney, S.R. and S. Pradhan, *Regulation of expression and activity of DNA (cytosine-5) methyltransferases in mammalian cells.* Prog Mol Biol Transl Sci, 2011. **101**: p. 311-33.

55. Suetake, I., et al., *DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction.* J Biol Chem, 2004. **279**(26): p. 27816-23.

56. Bird, A.P. and A.P. Wolffe, *Methylation-induced repression--belts, braces, and chromatin.* Cell, 1999. **99**(5): p. 451-4.

57. Ito, S., et al., *Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine.* Science, 2011. **333**(6047): p. 1300-3.

58. Kohli, R.M. and Y. Zhang, *TET enzymes, TDG and the dynamics of DNA demethylation.* Nature, 2013. **502**(7472): p. 472-9.

59. Gros, C., et al., *DNA methylation inhibitors in cancer: recent and future approaches.* Biochimie, 2012. **94**(11): p. 2280-96.

60. Jurkowska, R.Z., T.P. Jurkowski, and A. Jeltsch, *Structure and function of mammalian DNA methyltransferases.* Chembiochem, 2011. **12**(2): p. 206-22.

61. McDonald, J.I., et al., *Reprogrammable CRISPR/Cas9-based system for inducing site-specific DNA methylation.* Biol Open, 2016.

62. Choudhury, S.R., et al., *CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter.* Oncotarget, 2016.

63. Xu, X., et al., *A CRISPR-based approach for targeted DNA demethylation.* Cell Discov, 2016. **2**: p. 16009.

64. Enriquez, P., *CRISPR-Mediated Epigenome Editing.* Yale J Biol Med, 2016. **89**(4): p. 471-486.

65. Garcia-Bloj, B., et al., *Waking up dormant tumor suppressor genes with zinc fingers, TALEs and the CRISPR/dCas9 system.* Oncotarget, 2016.

66.   Moses, C., et al., *Hallmarks of cancer: The CRISPR generation.* Eur J Cancer, 2018. **93**: p. 10-18.

67.   Kardooni, H., et al., *CRISPR-Mediated Reactivation of DKK3 Expression Attenuates TGF-beta Signaling in Prostate Cancer.* Cancers (Basel), 2018. **10**(6).

68.   Mabe, N.W., et al., *Epigenetic silencing of tumor suppressor Par-4 promotes chemoresistance in recurrent breast cancer.* J Clin Invest, 2018.

69.   Horlbeck, M.A., et al., *Nucleosomes impede Cas9 access to DNA and.* Elife, 2016. **5**.

70.   Lo, A. and L. Qi, *Genetic and epigenetic control of gene expression by CRISPR-Cas systems.* F1000Res, 2017. **6**.

71.   Pu, J., et al., *Utilization of TALEN and CRISPR/Cas9 technologies for gene targeting and modification.* Exp Biol Med (Maywood), 2015.

72.   Xirong, L., et al., *Hepatitis B virus can be inhibited by DNA methyltransferase 3a via specific zinc-finger-induced methylation of the X promoter.* Biochemistry (Mosc), 2014. **79**(2): p. 111-23.

73.   Falahi, F., A. Sgro, and P. Blancafort, *Epigenome engineering in cancer: fairytale or a realistic path to the clinic?* Front Oncol, 2015. **5**: p. 22.

74.   Stolzenburg, S., et al., *Stable oncogenic silencing in vivo by programmable and targeted de novo DNA methylation in breast cancer.* Oncogene, 2015.

75.   Brocken, D.J.W., M. Tark-Dame, and R.T. Dame, *dCas9: A Versatile Tool for Epigenome Editing.* Curr Issues Mol Biol, 2017. **26**: p. 15-32.

76.   Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-21.

77.   Cong, L., et al., *Multiplex genome engineering using CRISPR/Cas systems.* Science, 2013. **339**(6121): p. 819-23.

78.   Wiedenheft, B., et al., *Structures of the RNA-guided surveillance complex from a bacterial immune system.* Nature, 2011. **477**(7365): p. 486-9.

79.   Anton, T. and S. Bultmann, *Site-specific recruitment of epigenetic factors with a modular CRISPR/Cas system.* Nucleus, 2017: p. 1-8.

80.   Dong, X., et al., *CRISPR/dCas9-mediated inhibition of gene expression in Staphylococcus aureus.* J Microbiol Methods, 2017.

81.   Kao, P.H. and I.S. Ng, *CRISPRi mediated phosphoenolpyruvate carboxylase regulation to enhance the production of lipid in Chlamydomonas reinhardtii.* Bioresour Technol, 2017.

82.   Pulecio, J., et al., *CRISPR/Cas9-Based Engineering of the Epigenome.* Cell Stem Cell, 2017. **21**(4): p. 431-447.

83.   Martinez-Lage, M., R. Torres-Ruiz, and S. Rodriguez-Perales, *CRISPR/Cas9 Technology: Applications and Human Disease Modeling.* Prog Mol Biol Transl Sci, 2017. **152**: p. 23-48.

84.   Xie, N., et al., *Novel Epigenetic Techniques Provided by the CRISPR/Cas9 System.* Stem Cells Int, 2018. **2018**: p. 7834175.

85.   Gilbert, L.A., et al., *CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes.* Cell, 2013. **154**(2): p. 442-51.

86.   Hsu, P.D., et al., *DNA targeting specificity of RNA-guided Cas9 nucleases.* Nat Biotechnol, 2013. **31**(9): p. 827-32.

87.   Senis, E., et al., *CRISPR/Cas9-mediated genome engineering: an adeno-associated viral (AAV) vector toolbox.* Biotechnol J, 2014. **9**(11): p. 1402-12.

88.   Sachdeva, M., et al., *CRISPR/Cas9: molecular tool for gene therapy to target genome and epigenome in the treatment of lung cancer.* Cancer Gene Ther, 2015.

89.   Mei, Y., et al., *Recent Progress in CRISPR/Cas9 Technology.* J Genet Genomics, 2016. **43**(2): p. 63-75.

90.   Tsai, S.Q., et al., *Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing.* Nat Biotechnol, 2014. **32**(6): p. 569-76.

91.   Barrangou, R., et al., *Advances in CRISPR-Cas9 genome engineering: lessons learned from RNA interference.* Nucleic Acids Res, 2015.

92.   Xu, J., et al., *A Toolkit of CRISPR-Based Genome Editing Systems in Drosophila.* J Genet Genomics, 2015. **42**(4): p. 141-149.

93.   Hilton, I.B., et al., *Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers.* Nat Biotechnol, 2015.

94.   Kearns, N.A., et al., *Functional annotation of native enhancers with a Cas9-histone demethylase fusion.* Nat Methods, 2015. **12**(5): p. 401-403.

95.   Morita, S., et al., *Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions.* Nat Biotechnol, 2016. **34**(10): p. 1060-1065.

96.   Amabile, A., et al., *Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing.* Cell, 2016. **167**(1): p. 219-232 e14.

97.   Lichtman, J.W. and J.A. Conchello, *Fluorescence microscopy.* Nat Methods, 2005. **2**(12): p. 910-9.

98.   Jablonski, A., *Efficiency of Anti-Stokes Fluorescence in Dyes.* Nature, 1933. **131**: p. 839.

99.   Specht, E.A., E. Braselmann, and A.E. Palmer, *A Critical and Comparative Review of Fluorescent Tools for Live-Cell Imaging.* Annu Rev Physiol, 2017. **79**: p. 93-117.

100.   Lukinavicius, G., et al., *Fluorogenic probes for live-cell imaging of the cytoskeleton.* Nat Methods, 2014. **11**(7): p. 731-3.

101.   Lukinavicius, G., et al., *SiR-Hoechst is a far-red DNA stain for live-cell nanoscopy.* Nat Commun, 2015. **6**: p. 8497.

102.   Heim, R., A.B. Cubitt, and R.Y. Tsien, *Improved green fluorescence.* Nature, 1995. **373**(6516): p. 663-4.

103.   Hell, S.W. and J. Wichmann, *Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy.* Opt Lett, 1994. **19**(11): p. 780-2.

104.   Keller, P.J. and E.H. Stelzer, *Quantitative in vivo imaging of entire embryos with Digital Scanned Laser Light Sheet Fluorescence Microscopy.* Curr Opin Neurobiol, 2008. **18**(6): p. 624-32.

105.   *Method of the Year 2014.* Nature Methods, 2014. **12**: p. 1.

106.   Power, R.M. and J. Huisken, *A guide to light-sheet fluorescence microscopy for multiscale imaging.* Nat Methods, 2017. **14**(4): p. 360-373.

107.   Siedentopf H, a.Z.R., *Ueber Sichtbarmachung ultramikroskopischer Teilchen, mit besonderer Anwendung auf Goldrubingläser.* Drude's Annalen der Physik, 1903.

108.   Voie, A.H., D.H. Burns, and F.A. Spelman, *Orthogonal-plane fluorescence optical sectioning: three-dimensional imaging of macroscopic biological specimens.* J Microsc, 1993. **170**(Pt 3): p. 229-36.

109.   Huisken, J., et al., *Optical sectioning deep inside live embryos by selective plane illumination microscopy.* Science, 2004. **305**(5686): p. 1007-9.

110.   Huisken, J. and D.Y. Stainier, *Selective plane illumination microscopy techniques in developmental biology.* Development, 2009. **136**(12): p. 1963-75.

111.   Weber, M. and J. Huisken, *Light sheet microscopy for real-time developmental biology.* Curr Opin Genet Dev, 2011. **21**(5): p. 566-72.

112.   Kumar, A., et al., *Dual-view plane illumination microscopy for rapid and spatially isotropic imaging.* Nat Protoc, 2014. **9**(11): p. 2555-73.

113.   Kumar, A., et al., *Using Stage- and Slit-Scanning to Improve Contrast and Optical Sectioning in Dual-View Inverted Light Sheet Microscopy (diSPIM).* Biol Bull, 2016. **231**(1): p. 26-39.

114.   Swoger, J., et al., *Multi-view image fusion improves resolution in three-dimensional microscopy.* Opt Express, 2007. **15**(13): p. 8029-42.

115.   Strnad, P., et al., *Inverted light-sheet microscope for imaging mouse pre-implantation development.* Nat Methods, 2016. **13**(2): p. 139-42.

116.   Strobl, F., S. Klees, and E.H.K. Stelzer, *Light Sheet-based Fluorescence Microscopy of Living or Fixed and Stained Tribolium castaneum Embryos.* J Vis Exp, 2017(122).

117.   Ryan, D.P., et al., *Automatic and adaptive heterogeneous refractive index compensation for light-sheet microscopy.* Nat Commun, 2017. **8**(1): p. 612.

118. Bassi, A., B. Schmid, and J. Huisken, *Optical tomography complements light sheet microscopy for in toto imaging of zebrafish development.* Development, 2015. **142**(5): p. 1016-20.

119. Icha, J., et al., *Using Light Sheet Fluorescence Microscopy to Image Zebrafish Eye Development.* J Vis Exp, 2016(110).

120. Huisken, J. and D.Y. Stainier, *Even fluorescence excitation by multidirectional selective plane illumination microscopy (mSPIM).* Opt Lett, 2007. **32**(17): p. 2608-10.

121. Wu, Y., et al., *Reflective imaging improves spatiotemporal resolution and collection efficiency in light sheet microscopy.* Nat Commun, 2017. **8**(1): p. 1452.

122. Elisa, Z., et al., *Technical implementations of light sheet microscopy.* Microsc Res Tech, 2018.

123. Tomer, R., et al., *Quantitative high-speed imaging of entire developing embryos with simultaneous multiview light-sheet microscopy.* Nat Methods, 2012. **9**(7): p. 755-63.

124. Amat, F., et al., *Efficient processing and analysis of large-scale light-sheet microscopy data.* Nat Protoc, 2015. **10**(11): p. 1679-96.

125. Power, R.M. and J. Huisken, *Adaptable, illumination patterning light sheet microscopy.* Sci Rep, 2018. **8**(1): p. 9615.

126. Conrad, C., et al., *Micropilot: automation of fluorescence microscopy-based imaging for systems biology.* Nat Methods, 2011. **8**(3): p. 246-9.

127. Schmied, C., et al., *An automated workflow for parallel processing of large multiview SPIM recordings.* Bioinformatics, 2015.

128. Preibisch, S., et al., *Efficient Bayesian-based multiview deconvolution.* Nat Methods, 2014. **11**(6): p. 645-8.

129. Bouchard, M.B., et al., *Swept confocally-aligned planar excitation (SCAPE) microscopy for high speed volumetric imaging of behaving organisms.* Nat Photonics, 2015. **9**(2): p. 113-119.

130. Maioli, V., et al., *Time-lapse 3-D measurements of a glucose biosensor in multicellular spheroids by light sheet fluorescence microscopy in commercial 96-well plates.* Sci Rep, 2016. **6**: p. 37777.

131. Wachsmuth, M., et al., *High-throughput fluorescence correlation spectroscopy enables analysis of proteome dynamics in living cells.* Nat Biotechnol, 2015. **33**(4): p. 384-9.

132. Wang, Y., et al., *Quantitative performance evaluation of a back-illuminated sCMOS camera with 95% QE for super-resolution localization microscopy.* Cytometry A, 2017. **91**(12): p. 1175-1183.

133. Carpenter, A.E., et al., *CellProfiler: image analysis software for identifying and quantifying cell phenotypes.* Genome Biol, 2006. **7**(10): p. R100.

134. Kittler, R., et al., *Genome-scale RNAi profiling of cell division in human tissue culture cells.* Nat Cell Biol, 2007. **9**(12): p. 1401-12.

135.   Isherwood, B., et al., *Live cell in vitro and in vivo imaging applications: accelerating drug discovery.* Pharmaceutics, 2011. **3**(2): p. 141-70.

136.   Schmitz, A., et al., *Multiscale image analysis reveals structural heterogeneity of the cell microenvironment in homotypic spheroids.* Sci Rep, 2017. **7**: p. 43693.

137.   Shen, D., G. Wu, and H.I. Suk, *Deep Learning in Medical Image Analysis.* Annu Rev Biomed Eng, 2017.

138.   Strobl, F., A. Schmitz, and E.H.K. Stelzer, *Improving your four-dimensional image: traveling through a decade of light-sheet-based fluorescence microscopy research.* Nat Protoc, 2017. **12**(6): p. 1103-1109.

139.   Haralick, R.M., Shanmugam, K., and Dinstein, *Textural Features for Image Classification.* Systems, Man and Cybernetics, 1993.

140.   Misselwitz, B., et al., *Enhanced CellClassifier: a multi-class classification tool for microscopy images.* BMC Bioinformatics, 2010. **11**: p. 30.

141.   Zhang, L., et al., *DeepPap: Deep Convolutional Networks for Cervical Cell Classification.* IEEE J Biomed Health Inform, 2017. **21**(6): p. 1633-1643.

142.   Li, H., et al., *Cell dynamic morphology classification using deep convolutional neural networks.* Cytometry A, 2018.

143.   Scheeder, C., F. Heigwer, and M. Boutros, *Machine learning and image-based profiling in drug discovery.* Curr Opin Syst Biol, 2018. **10**: p. 43-52.

144.   Van Valen, D.A., et al., *Deep Learning Automates the Quantitative Analysis of Individual Cells in Live-Cell Imaging Experiments.* PLoS Comput Biol, 2016. **12**(11): p. e1005177.

145.   Gao, H., et al., *High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response.* Nat Med, 2015. **21**(11): p. 1318-25.

146.   Corsello, S.M., et al., *The Drug Repurposing Hub: a next-generation drug library and information resource.* Nat Med, 2017. **23**(4): p. 405-408.

147.   Miller, J.B. and D.J. Siegwart, *Design of synthetic materials for intracellular delivery of RNAs: From siRNA-mediated gene silencing to CRISPR/Cas gene editing.* Nano Research, 2018. **11**(10): p. 5310-5337.

148.   Erfle, H., et al., *Reverse transfection on cell arrays for high content screening microscopy.* Nat Protoc, 2007. **2**(2): p. 392-9.

149.   Sullivan, D.P., et al., *Deep learning is combined with massive-scale citizen science to improve large-scale image classification.* Nat Biotechnol, 2018.

150.   McQuin, C., et al., *CellProfiler 3.0: Next-generation image processing for biology.* PLoS Biol, 2018. **16**(7): p. e2005970.

151.   Schmidhuber, J., *Deep learning in neural networks: an overview.* Neural Netw, 2015. **61**: p. 85-117.

152. Faust, O., et al., *Deep learning for healthcare applications based on physiological signals: A review.* Comput Methods Programs Biomed, 2018. **161**: p. 1-13.

153. McCulloch, W.S. and W. Pitts, *A logical calculus of the ideas immanent in nervous activity. 1943.* Bull Math Biol, 1990. **52**(1-2): p. 99-115; discussion 73-97.

154. Angermueller, C., et al., *Deep learning for computational biology.* Mol Syst Biol, 2016. **12**(7): p. 878.

155. Dong, W., et al., *Denoising Prior Driven Deep Neural Network for Image Restoration.* IEEE Trans Pattern Anal Mach Intell, 2018.

156. Hussain, Z., et al., *Differential Data Augmentation Techniques for Medical Imaging Classification Tasks.* AMIA Annu Symp Proc, 2017. **2017**: p. 979-984.

157. Pastur-Romay, L.A., et al., *Deep Artificial Neural Networks and Neuromorphic Chips for Big Data Analysis: Pharmaceutical and Bioinformatics Applications.* Int J Mol Sci, 2016. **17**(8).

158. Majaj, N.J. and D.G. Pelli, *Deep learning-Using machine learning to study biological vision.* J Vis, 2018. **18**(13): p. 2.

159. Zemanova, L., et al., *Confocal optics microscopy for biochemical and cellular high-throughput screening.* Drug Discov Today, 2003. **8**(23): p. 1085-93.

160. Maier, L.J., et al., *Unraveling mitotic protein networks by 3D multiplexed epitope drug screening.* Mol Syst Biol, 2018. **14**(8): p. e8238.

161. Fu, Q., et al., *Imaging multicellular specimens with real-time optimized tiling light-sheet selective plane illumination microscopy.* Nat Commun, 2016. **7**: p. 11088.

162. Cole, R.W., T. Jinadasa, and C.M. Brown, *Measuring and interpreting point spread functions to determine confocal microscope resolution and ensure quality control.* Nat Protoc, 2011. **6**(12): p. 1929-41.

163. Liu, J.S., et al., *Programmed cell-to-cell variability in Ras activity triggers emergent behaviors during mammary epithelial morphogenesis.* Cell Rep, 2012. **2**(5): p. 1461-70.

164. Parashar, S., et al., *DNA methylation signatures of breast cancer in peripheral T-cells.* BMC Cancer, 2018. **18**(1): p. 574.

165. Kent, W.J., et al., *The human genome browser at UCSC.* Genome Res, 2002. **12**(6): p. 996-1006.

166. Schmied, C., et al., *An automated workflow for parallel processing of large multiview SPIM recordings.* Bioinformatics, 2016. **32**(7): p. 1112-4.

167. Dietz, C. and M.R. Berthold, *KNIME for Open-Source Bioimage Analysis: A Tutorial.* Adv Anat Embryol Cell Biol, 2016. **219**: p. 179-97.

168. Cai, Y., et al., *Experimental and computational framework for a dynamic protein atlas of human cell division.* Nature, 2018. **561**(7723): p. 411-415.

169. Desmaison, A., et al., *Impact of physical confinement on nuclei geometry and cell division dynamics in 3D spheroids.* Sci Rep, 2018. **8**(1): p. 8785.

170. Hu, D.J., et al., *Dynein recruitment to nuclear pores activates apical nuclear migration and mitotic entry in brain progenitor cells.* Cell, 2013. **154**(6): p. 1300-13.

171. Vojta, A., et al., *Repurposing the CRISPR-Cas9 system for targeted DNA methylation.* Nucleic Acids Res, 2016.

172. Radzisheuskaya, A., et al., *Optimizing sgRNA position markedly improves the efficiency of CRISPR/dCas9-mediated transcriptional repression.* Nucleic Acids Res, 2016.

173. Santos, A., R. Wernersson, and L.J. Jensen, *Cyclebase 3.0: a multi-organism database on cell-cycle regulation and phenotypes.* Nucleic Acids Res, 2015. **43**(Database issue): p. D1140-4.

174. Schmitz, M.H., et al., *Live-cell imaging RNAi screen identifies PP2A-B55alpha and importin-beta1 as key mitotic exit regulators in human cells.* Nat Cell Biol, 2010. **12**(9): p. 886-93.

175. Peng, J., et al., *High-throughput Screens in Mammalian Cells Using the CRISPR-Cas9 System.* FEBS J, 2015.

176. Cella Zanacchi, F., et al., *Live-cell 3D super-resolution imaging in thick biological samples.* Nat Methods, 2011. **8**(12): p. 1047-9.

177. Parnamaa, T. and L. Parts, *Accurate Classification of Protein Subcellular Localization from High Throughput Microscopy Images Using Deep Learning.* G3 (Bethesda), 2017.

178. Rawat, W. and Z. Wang, *Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review.* Neural Comput, 2017: p. 1-98.

179. Martinez, N.J., et al., *High-throughput fluorescence imaging approaches for drug discovery using in vitro and in vivo three-dimensional models.* Expert Opin Drug Discov, 2015. **10**(12): p. 1347-61.

180. Chittajallu, D.R., et al., *In vivo cell-cycle profiling in xenograft tumors by quantitative intravital microscopy.* Nat Methods, 2015. **12**(6): p. 577-85.

181. Erfle, H., et al., *Work flow for multiplexing siRNA assays by solid-phase reverse transfection in multiwell plates.* J Biomol Screen, 2008. **13**(7): p. 575-80.

182. Pau, G., et al., *Dynamical modelling of phenotypes in a genome-wide RNAi live-cell imaging assay.* BMC Bioinformatics, 2013. **14**: p. 308.

183. Billmann, M., et al., *A genetic interaction map of cell cycle regulators.* Mol Biol Cell, 2016. **27**(8): p. 1397-407.

184. Burkard, M.E., et al., *Plk1 self-organization and priming phosphorylation of HsCYK-4 at the spindle midzone regulate the onset of division in human cells.* PLoS Biol, 2009. **7**(5): p. e1000111.

185. Masson, D. and T.E. Kreis, *Identification and molecular characterization of E-MAP-115, a novel microtubule-associated protein predominantly expressed in epithelial cells.* J Cell Biol, 1993. **123**(2): p. 357-71.

186. Bulinski, J.C., et al., *Rapid dynamics of the microtubule binding of ensconsin in vivo.* J Cell Sci, 2001. **114**(Pt 21): p. 3885-97.

187. Jiang, W., et al., *PRC1: a human mitotic spindle-associated CDK substrate protein required for cytokinesis.* Mol Cell, 1998. **2**(6): p. 877-85.

188. Hla, T., et al., *Characterization of edg-2, a human homologue of the Xenopus maternal transcript G10 from endothelial cells.* Biochim Biophys Acta, 1995. **1260**(2): p. 227-9.

189. Kimura, M.T., et al., *Two functional coding single nucleotide polymorphisms in STK15 (Aurora-A) coordinately increase esophageal cancer risk.* Cancer Res, 2005. **65**(9): p. 3548-54.

190. Sandhu, J., et al., *Aster Proteins Facilitate Nonvesicular Plasma Membrane to ER Cholesterol Transport in Mammalian Cells.* Cell, 2018. **175**(2): p. 514-529.e20.

191. Maccarana, M., et al., *Biosynthesis of dermatan sulfate: chondroitin-glucuronate C5-epimerase is identical to SART2.* J Biol Chem, 2006. **281**(17): p. 11560-8.

192. Park, M., A.J. Keung, and A.S. Khalil, *The epigenome: the next substrate for engineering.* Genome Biol, 2016. **17**(1): p. 183.

193. Sloan, C.A., et al., *ENCODE data at the ENCODE portal.* Nucleic Acids Res, 2016. **44**(D1): p. D726-32.

194. Konermann, S., et al., *Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex.* Nature, 2015. **517**(7536): p. 583-8.

195. Mali, P., et al., *CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering.* Nat Biotechnol, 2013. **31**(9): p. 833-8.

196. Klann, T.S., et al., *CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome.* Nat Biotechnol, 2017.

197. Thakore, P.I., et al., *Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements.* Nat Methods, 2015.

198. Genga, R.M., N.A. Kearns, and R. Maehr, *Controlling transcription in human pluripotent stem cells using CRISPR-effectors.* Methods, 2015.

199. Lino, C.A., et al., *Delivering CRISPR: a review of the challenges and approaches.* Drug Deliv, 2018. **25**(1): p. 1234-1257.

200. Nohra, R., et al., *RGMA and IL21R show association with experimental inflammation and multiple sclerosis.* Genes Immun, 2010. **11**(4): p. 279-93.

201. Neumann, B., et al., *Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes.* Nature, 2010. **464**(7289): p. 721-7.

202. Xiao, B., et al., *Identification of methylation sites and signature genes with prognostic value for luminal breast cancer.* BMC Cancer, 2018. **18**(1): p. 405.

203. Zhang, N., et al., *Human breast cancer growth inhibited in vivo by a dominant negative pleiotrophin mutant.* J Biol Chem, 1997. **272**(27): p. 16733-6.

204. Xia, Y., et al., *Repulsive guidance molecule RGMa alters utilization of bone morphogenetic protein (BMP) type II receptors by BMP2 and BMP4.* J Biol Chem, 2007. **282**(25): p. 18129-40.

205. Tahiliani, M., et al., *Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1.* Science, 2009. **324**(5929): p. 930-5.

9 Supplementary Data

9.1 Supplementary Figures

TP	X (px)	Y (px)	Z (px)
0	0	0	0
...	...	...	...
298	-23	47	59
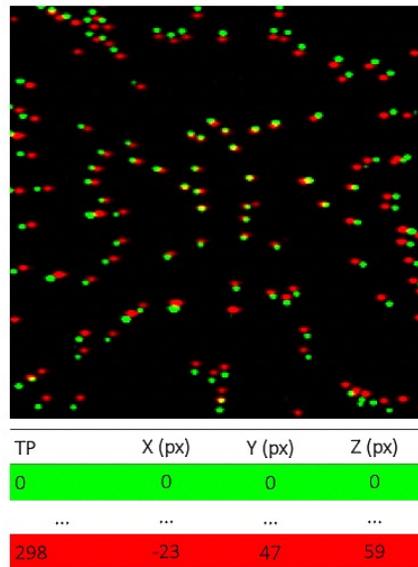
Supplementary Figure 1 – Matrigel displacement

Comparison of the positioning of beads immersed in matrigel at the beginning of the acquisition (TP0 - green) and at the last time point (TP298 - red) 24 hours later. Displacement was registered to the center of the image.
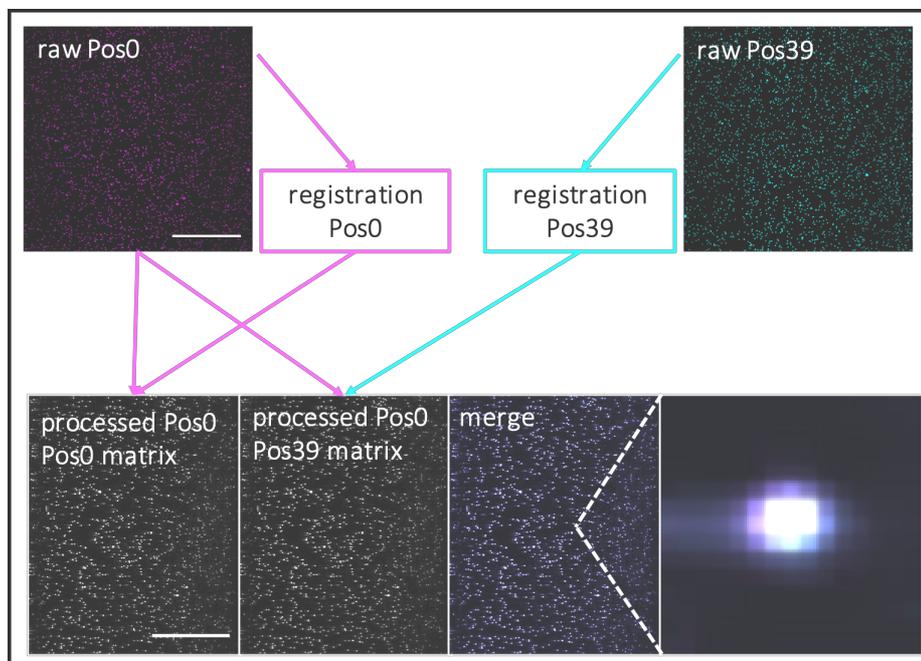
Supplementary Figure 2 – Transfer of registration matrix and PSF between positions

A key feature of the workflow is the separation of beads and sample. Position 0 and 39 of each acquisition were dedicated to beads mixed with Matrigel. When the registration matrix and PSF of both bead positions (Pos0 and Pos39) are applied to Pos0 for image processing, the shift in the resulting image is less than a pixel (< 0.325 μm) (scale bar = 50 μm).
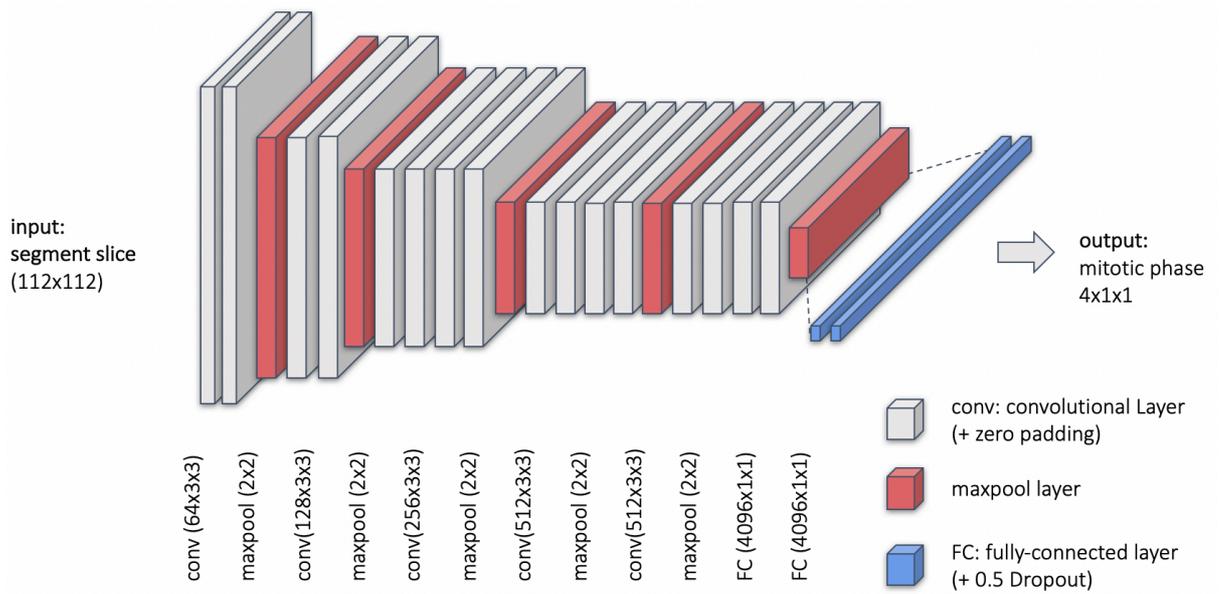
Supplementary Figure 3 – VGG-based convolutional neural network

Used convolutional neural network for nuclei cell cycle stage classification, comprised of convolutional layers and maxpooling layers. The output is combined by two fully-connected layers.
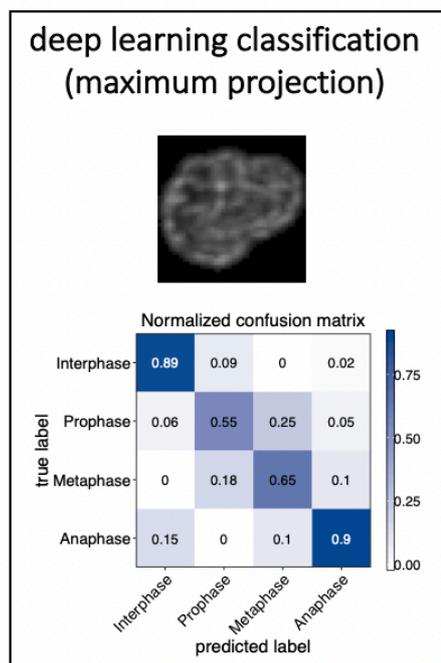
Supplementary Figure 4 – maximum projection DL classification

Utilizing the maximum projection of a nuclei segment for a more convenient and less computational costly cell phase classification results in a significant decreased accuracy of only 75%.
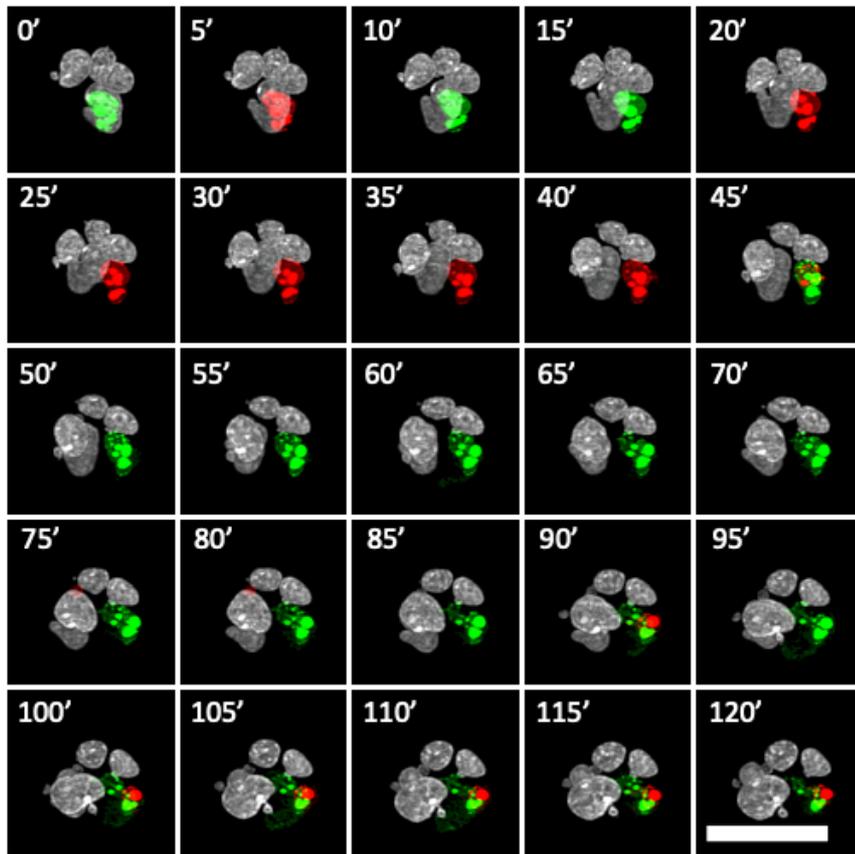
Supplementary Figure 5 – dying cell misclassification

Cell undergoing apoptosis are classified as prophase (green) and anaphase (red). Sample: Aurora kinase A knock-down spheroid at time point 1 to 26 of HCS at five minutes' interval.
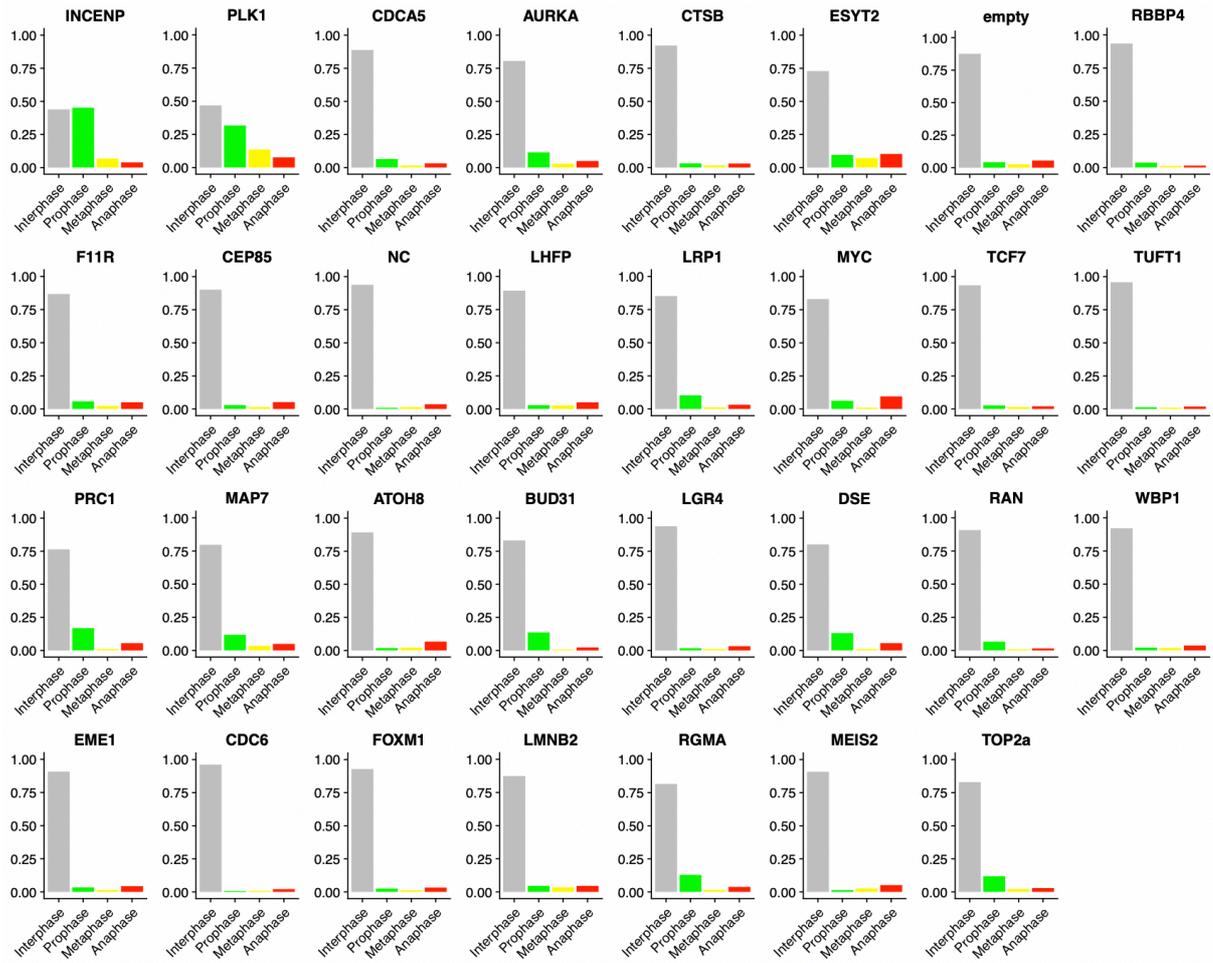
Supplementary Figure 6 – Ratio of cell phase upon knock-down

Bar plot for each all knock-down samples depicting the ratio of the four classified cell cycle stages.
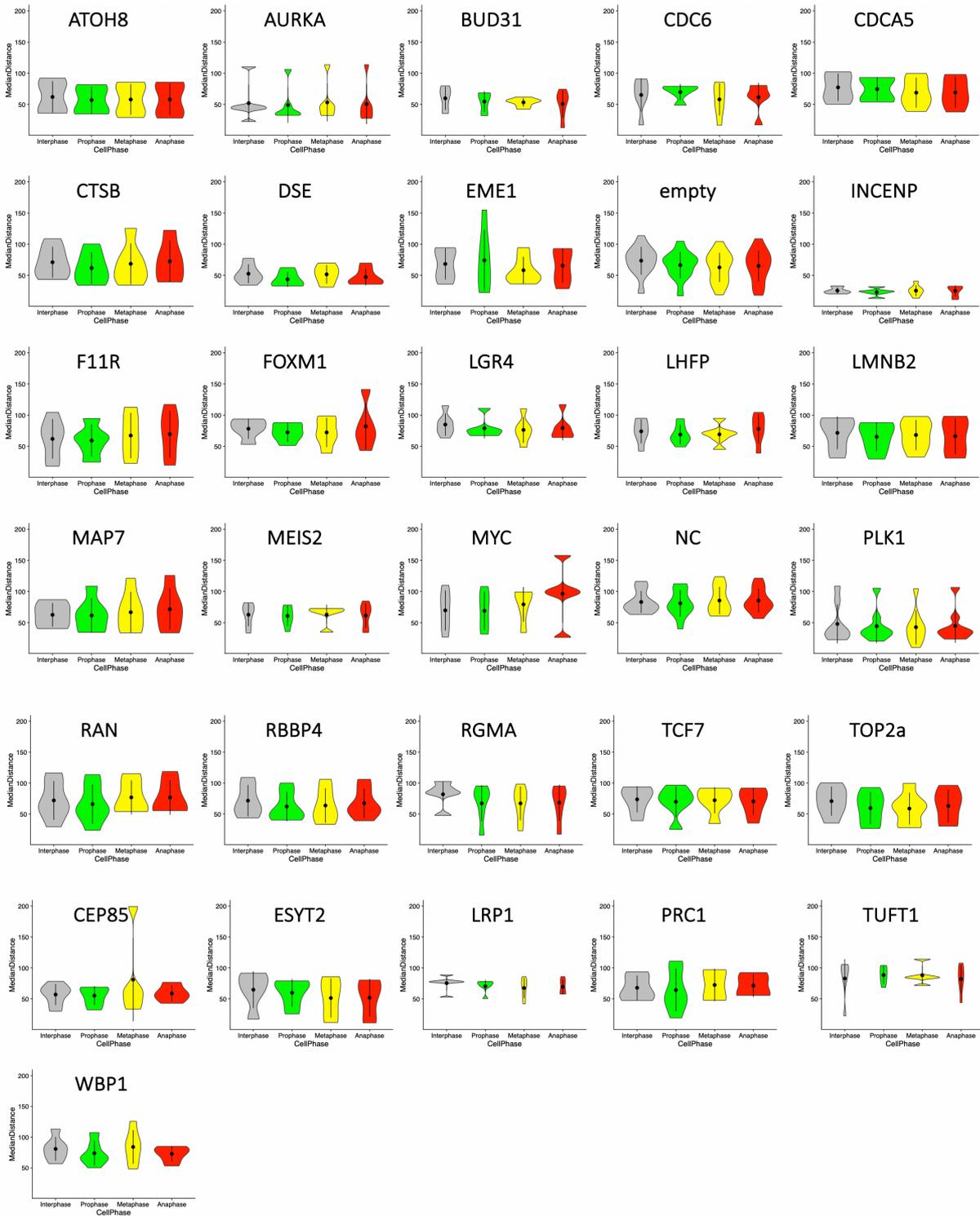
Supplementary Figure 7 – Violin plot of class distribution within the spheroid

Summary of the class distribution within the spheroid for all samples depicted by a violin plot.

9.2 Supplementary Tables

Supplementary Tables 1 - Selected target genes

Target genes were selected due to defined characteristics. They needed to be associated with the cell cyclce. The selected target genes also needed to display an expression regulation by either correlated or anti-correlated CpG methylation level with high Pearson correlation value (R).

target gene	name	Ambion siRNA#	regulatory CpG	CpG properties (expression/-me)	R_avg =
ATOH8	Protein atonal homolog 8	s39645 / s39643	1	anti-correlated	0.52
AURKA	Aurora kinase A	s196 / s197	2	correlated / anti-correlated	0.52
BUD31	Protein BUD31 homolog	s17010 / s17009	1	correlated	0.45
CDC6	Cell division control protein 6	s2744 / s2746	2	anti-correlated	0.65
CDCA5	Sororin	s41424 / s41425	6	correlated	0.70
CEP85	Centrosomal protein of 85 kDa	s34959 / s34961	5	correlated / anti-correlated	0.64
CTSB	Cathepsin B	s3738 / s3739	1	anti-correlated	0.53
DSE	Dermatan Sulfate Epimerase	s26749 / s26750	1	anti-correlated	0.54
EME1	Essential Meiotic Structure-Specific Endonuclease 1	s44946 / s44945	1	correlated	0.66
ESYT2	Extended synaptotagmin-2	s33138 / s33136	2	anti-correlated	0.75
F11R	F11 Receptor	s27152 / s27151	6	anti-correlated	0.55
FOXM1	Forkhead Box M1	s5250 / s5249	1	correlated	0.68
LGR4	Leucine-Rich Repeat G.Prot-Coupled Receptor 4	s30840 / s229314	1	anti-correlated	0.61
LHFP	Lipoma HMGIC Fusion Partner	s19847 / s19848	1	correlated	0.57
LMNB2	Lamin B2	s39477 / s39476	1	anti-correlated	0.60
LRP1	LDL Receptor Related Protein 1	s8278 / s8280	4	anti-correlated	0.70
MAP7	Ensconsin	s17263 / s17262	2	correlated	0.67
MEIS2	Meis Homeobox 2	s8666 / s8664	7	correlated / anti-correlated	0.57
MYC	Myc proto-oncogene protein	s9130 / s9131	2	anti-correlated	0.68
PLK1	Polo-like kinase 1	s448 / s450	4	correlated	0.64

PRC1	Protein regulator of cytokinesis 1	s17268 / s17269	1	anti-correlated	0.73
RAN	GTP-binding nuclear protein Ran	s11769 / s11768	1	anti-correlated	0.60
RBBP4	Histone-binding protein RBBP4	s55169 / s56872	1	anti-correlated	0.57
RGMA	Repulsive Guidance Molecule Family Member A	s32498 / s32500	7	correlated / anti-correlated	0.70
TCF7	Transcription factor 7	s13877 / s13878	2	anti-correlated	0.71
TOP2A	Topoisomerase II Alpha	s14307 / s14308	2	correlated	0.66
TUFT1	Tuftelin	s14510 / s14509	2	anti-correlated	0.59
WBP1	WW Domain Binding Protein 1	s24095 / s225969	1	correlated	0.52

Supplementary Tables 2 – 23 factors describing spheroid phenotype

A total of 23 factors were calculated to define the phenotype of the spheroid throughout the time lapse. These factors describe properties of the whole spheroid, spheroid development and single nuclei features throughout the twenty-four hour time laps.

factor	name	Description	global or nuclei feature
1	**spheroid growth rate (nuclei)**	increase of nuclei count over the course of the time lapse	global
2	**prophase ratio**	fraction of nuclei classified as "prophase" from all nuclei	nuclei
3	**metaphase ratio**	fraction of nuclei classified as "metaphase" from all nuclei	nuclei
4	**anaphase ratio**	fraction of nuclei classified as "anaphase" from all nuclei	nuclei
5	**avg. cell volume**	the nuclei size in voxel number of all nuclei in all cell phases	global
6	**prophase segment volume**	the nuclei size in voxel number of all nuclei classified as "prophase"	nuclei
7	**metaphase segment volume**	the nuclei size in voxel number of all nuclei classified as "metaphase"	nuclei
8	**anaphase segment volume**	the nuclei size in voxel number of all nuclei classified as "anaphase"	nuclei
9	**interphase segment volume**	the nuclei size in voxel number of all nuclei classified as "interphase"	nuclei
10	**spheroid volume**	the spheroids volume in voxel throughout the time lapse	global
11	**avg. segment volume**	the nuclei size in voxel number of all nuclei in all cell phases	global

#			
12	**spheroid growth rate (volume)**	increase in volume of the spheroid hull throughout the time lapse	global
13	**spheroid compactness**	factor describing the volume in relation to the largest extend	global
14	**convexity**	factor describing the volume in relation to the surface	global
15	**nuclei migration speed**	average movement of all nuclei in 3D space in pixel per time point	global
16	**interphase transition duration**	average duration of a nuclei in "interphase"	nuclei
17	**prophase transition duration**	average duration of a nuclei in "prophase"	nuclei
18	**metaphase transition duration**	average duration of a nuclei in "metaphase"	nuclei
19	**anaphase transition duration**	average duration of a nuclei in "anaphase"	nuclei
20	**total number cell cycle transitions**	total number of biological possible cell cycle phase transitions	global
21	**normal / abnormal transition**	faction of cell cycle phase transitions, that are biological impossible	global
22	**spheroid roundness**	factor describing shape of spheroid	global
23	**size / spheroid roundness ratio**	ratio of size and roundness	global

Supplementary Tables 3 – Number of normal and abnormal transition detected

target gene	normal transition	abnormal transition	ratio	n = spheroids
NC	146,101	1,380	0.99	12
EMPTY	192,450	1637	0.99	12
INCENP	4,193	317	0.92	12
ATOH8	39,421	514	0.99	6
AURKA	35,123	454	0.99	6
BUD31	33,567	368	0.99	6
CDC6	56,024	1,476	0.97	6
CDCA5	31,556	494	0.98	6
CEP85	23,590	520	0.98	6
CTSB	49,919	1,317	0.97	6
DSE	24,199	578	0.98	6
EME1	21,214	362	0.98	6
ESYT2	39,147	873	0.98	6

F11R	63,331	701	0.99	6
FOXM1	67,996	820	0.99	6
LGR4	47,385	577	0.99	6
LHFP	42,949	630	0.99	6
LMNB2	67,367	752	0.99	6
LRP1	38,655	571	0.99	6
MAP7	28,768	525	0.98	6
MEIS2	31,524	1,326	0.96	6
MYC	42,240	703	0.98	6
PLK1	21,475	450	0.98	12
PRC1	12,712	205	0.98	6
RAN	35,225	251	0.99	6
RBBP4	62,204	465	0.99	6
RGMA	56,631	792	0.99	6
TCF7	53,754	573	0.99	6
TOP2A	43,639	484	0.99	6
TUFT1	73,423	782	0.99	6
WBP1	51,438	784	0.98	6

10 Acknowledgements

I thank Dr. Christian Conrad for his support and guidance throughout this project starting at the DKFZ selection round, aiding my advancement to a light sheet microscopy expert and for keeping me curious on cutting-edge technologies, while allowing me to openly direct the project. I thank Prof. Dr. Roland Eils for his support, mentoring, guidance and funding during the last years and for giving me the opportunity to work with highly advanced technologies in a fruitful and supportive environment.

Furthermore, I thank Dr. Jan Felix Evers and Prof. Dr. Frank Lyko for their careful consideration and evaluation of the PhD project and their valuable comments during my thesis advisory committee meetings, as well as Prof. Dr. Gislene Pereira and Prof. Dr. Stefan Wiemann for their valuable time as part of my examination committee.

Additionally, I want to thank all members of the Intelligent Imaging group, the members of the Synthetic Biology group and the members of the RNAi Screening Facility at BioQuant as well as the whole Eilslabs department for their contributions to this project, their friendship and cooperate support throughout the years. Furthermore, Dr. Jan Eufinger, Corinna Sprengart, Manuela Schäfer were highly supportive regarding any administrative issues or travel planning.

I want to thank Jon Daniels (ASI) for his essential and quick technical support regarding any issues with the diSPIM systems, as well as many valuable discussion and early skype calls.

An extensive thank you goes to Lena Eismann, Teresa Krieger, Dominik Niopek and Timo Trefzer, Sarah Weiß, as well as Monika and Karlheinz Hülser for "carful" reading and highly valued critical discussions and feedback on the thesis manuscript.

Finally, but most notably I want to thank my wife Lena for her love and support in good and difficult moments of the last four years.

Lastly, I thank my family Monika, Karlheinz, Sarah, Mathias, Doris, Norbert, Raphaela, Nora, Erik and our dog Foot for their support throughout all these years. I couldn't have finished this project without your help.