Computational Approaches to Examine Cellular Signalling Networks

vom Fachbereich Biologie der Technischen Universität Darmstadt

zur Erlangung des Grades
Doctor rerum naturalium
(Dr. rer. nat.)

Dissertation
von Marcel Jentsch

Erstgutachter: Prof. Dr. Alexander Löwer
Zweitgutachter: Prof. Dr. Kay Hamacher

Darmstadt 2019

Jentsch, Marcel : Computational Approaches to Examine Cellular Signalling

Darmstadt, Technische Universität Darmstadt,

Jahr der Veröffentlichung der Dissertation auf TUprints: 2020

Tag der mündlichen Prüfung: 24.10.2019

March 5, 2020

Just remember that you're standing on a planet that's evolving
And revolving at 900 miles an hour.
It's orbiting at 19 miles a second, so it's reckoned,
The sun that is the source of all our power.
– Eric Idle

Contents

Part I

Abstract - Zusammenfassung

Abstract

Complex biological systems can only be analysed by utilizing computational and mathematical methods. They are essential for studying the interactions between the components of biological systems and generating an understanding how these interactions give rise to biological functions and mechanisms of cellular signalling networks.
In this work, I provide three examples on how the analysis of single cell data derived from live-cell time-lapse microscopy of fluorescent reporter systems benefits from the use of several approaches that originate from computer sciences.
The analysis of single cell data faces several challenges ranging from extracting single cell time series from the raw imaging data, identifying signalling classes to the identification of distinct patterns in single cell trajectories. Several methods were introduced in the course of this work to handle the experimental data appropriately.
For the extraction of single cell time series, a novel method to track cells in the imaging data based on Coherent Point Drift was introduced. The tracking benefits from the features of the Coherent Point Drift method that correspond to cellular motility patterns. The motion coherence constrain of the Coherent Point Drift mimics the observation that cells do not move independently; they are embedded in a neighbourhood that constrains their freedom of movement.
The Dynamic Time Warping framework was established as a useful approach to tackle several issues in the analysis of biological data. The opportunity to quantify the similarity of dynamics granted a new view on single cell data. While calculating the optimal alignment between two single cell trajectories their similarity can be quantified. Dynamic Time Warping was modified so that it constrains the flexibility of the alignment making the alignment more biological relevant. Based on Dynamic Time Warping estimated similarities among individual signalling dynamics distinct signalling classes could be identified in the datasets analysed. Dynamic Time Warping was as well utilized for multivariate time series. This allowed the comparison of single cells while taking the dependence of dynamics of several signalling components within a pathway into account. This gives a new way on the comparison of pathway activity among individual cells.
Different signalling pathways exhibit different signalling dynamics. Therefore, two feature detection methods were proposed that aim to quantify signalling dynamics from different angles. The Dynamic Time Warping framework was used to develop a feature detection method that identifies patterns in the time series flexible in the time domain and independent of the scaling. If dynamics lack repetitive patterns dynamics have to be quantified in a different way. Therefore, to identify global dynamics a supervised learning method was developed that reduces the dimensionality of the time series data and identifies fundamental dynamics that compose the observed individual dynamics.
To understand how cells, encode the extracellular input and transmit its information to elicit appropriate responses, quantitative time-resolved measurements of

pathway activation at the single-cell level was acquired for all three scenarios.
The application of the introduced set of tools provided new insight into fundamental biological questions. On the level of the raw imaging data the cell tracking step does only differ slightly between the three biological examples. On the single cell level the three signalling pathways studied exhibit different dynamics and demand therefore different requirements on the analysis.
The TGFβ pathway is a multi-functional signalling system regulating cellular processes ranging from proliferation and migration to differentiation and cell death. Alterations in the cellular response to TGFβ are involved in severe human diseases. It was revealed that the response to a given dose of TGFβ is determined cell specifically by the levels of defined signalling proteins and that the observed heterogeneity in signalling protein expression leads to decomposition of cells into classes with qualitatively distinct signalling dynamic. How the dynamics differ among the signalling classes could be quantified using the supervised learning approach.
Studies have shown the beneficial effects of hyperthermic treatment during radiation- or chemotherapy of cancers. I aimed to understand how p53 dynamics after genotoxic stress are modulated by temperature across a physiological relevant range. In the range of 33°C to 39°C, pulsatile p53 accumulation dynamics are modulated in frequency. Above 40°C, a temperature that corresponds to mild hyperthermia in the clinical setting, a reversible phase transition towards sustained hyperaccumulation was observed. This disrupts the canonical p53 response to DNA double strand breaks. Above 40°C mild hyperthermia alone was sufficient to induce a p53 response.
The view onto the p53 signalling was extended by simultaneously measurement of an additional pathway component. p21 as an inhibitor of cyclin-dependent kinases is the mediator of p53 in growth suppression and a marker of cellular senescence. p53 signalling encodes information about signal intensity, duration and identity in complex dynamics. I studied how these p53 dynamics are related to p21 dynamics in the same cell. It could be shown that p53 and p21 dynamics were not independent and that distinct signalling dynamic shape population response dynamics after application of genotoxic stress. This was shown by using clustering based on multivariate Dynamic Time Warping similarity estimates. Signalling classes and dynamics were connected to cell cycle state.

Zusammenfassung

Komplexe biologische Systeme können nur mit rechnerischen und mathematischen Methoden analysiert werden. Sie sind unerlässlich, um die Wechselwirkungen zwischen den Komponenten biologischer Systeme zu untersuchen. Sie erzeugen ein Verständnis dafür, wie die Wechselwirkungen zwischen zellulären Signalnetzwerken bestimmte biologische Funktionen und Verhaltensweisen hervorrufen und steuern.
In dieser Arbeit stelle ich drei Beispiele vor, wie die Analyse von Einzelzelldaten aus der Zeitraffermikroskopie fluoreszierender Reporterzelllinien, von der Verwendung verschiedener Ansätze aus den Computerwissenschaften profitiert. Die Analyse von Einzelzelldaten ist eine Herausforderung, die von der Extraktion von Einzelzell-Zeitreihen aus den rohen Bilddaten, über die Identifizierung von Signalklassen, bis hin zur Identifizierung unterschiedlicher Muster in Einzelzell-Trajektorien reicht.
Es wurden verschiedene Methoden eingeführt, um die entsprechenden experimentellen Daten zu verarbeiten. Zur Extraktion von Einzelzell-Zeitreihen wurde ein neuartiges Tracking, basierend auf der Coherent Point Drift Methode, entwickelt. Die Eigenschaften der Coherent Point Drift Methode entsprechen den Charakteristiken der Zellbewegung. Der Motion Coherence Constrain der Coherent Point Drift Methode spiegelt die Beobachtung das Zellen sich nicht unabhängig voneinander bewegen; sie sind eingebunden in einen Zellverbund welcher ihre Bewegungsfreiheit beschneidet.
Das Dynamic Time Warping Framework ist ein nützlicher Ansatz, um mehrere biologische Prozesse zu untersuchen. Durch die Möglichkeit die Ähnlichkeit von Dynamiken zu quantifizieren wird ein neuer Blick auf Einzelzelldaten gewährt. Mittels der Berechnung eines optimalen Alignements zwischen zwei Einzelzelltrajektorien kann deren Ähnlichkeit quantifiziert werden. Dynamic Time Warping wurde so verändert, dass es die Flexibilität des Alignments eingeschränkt wird, wodurch dieses biologisch relevanter wird.
Basierend auf den berechneten Ähnlichkeiten zwischen einzelnen Signaldynamiken konnten Signalklassen in den untersuchten Daten identifiziert werden. Dynamic Time Warping wurde des Weiteren für multivariate Zeitreihen verwendet, was es erlaubt, die Ähnlichkeit zweier Zellen in Abhängigkeit gleichzeitiger Dynamiken verschiedener Signalwegkomponenten zu quantifizieren. Dies erlaubt den Vergleich der Signalwegenaktivität zwischen Zellen unter Berücksichtigung der gleichzeitigen Dynamiken verschiedener Komponenten des untersuchten Signalweges.
Unterschiedliche Signalwege weisen eine unterschiedliche Signaldynamik auf. Daher wurden zwei Methoden zur Erkennung vorgeschlagen, die darauf abzielen, die Signaldynamik aus verschiedenen Blickwinkeln zu quantifizieren. Das Dynamic Time Warping Framework wurde genutzt, um eine Feature Detection Methode zu entwickeln, welches Muster in der Zeitreihe flexibel im Zeitbereich und unabhängig von der Skalierung identifiziert. Um globale Dynamiken zu identifizieren, wird ein überwachtes Lernverfahren eingeführt, dass die Dimensionalität der Zeitreihen-

daten reduzieren und grundlegende Dynamiken identifizieren kann, aus denen sich die beobachtete individuelle Dynamik zusammensetzt.
Um zu verstehen, wie Zellen den extrazellulären Input kodieren und Informationen übertragen, um angemessene Antworten zu erhalten, wurden für alle drei Szenarien quantitative, zeitaufgelöste Messungen der Signalwegaktivität auf Einzelzellebene erfasst. Die Anwendung der vorgestellten Tools bot einen neuen Einblick in grundlegende biologische Fragen. Auf der Ebene der rohen Bilddaten unterscheidet sich das Zelltracking zwischen den drei biologischen Beispielen kaum. Auf Einzelzellebene weisen die drei untersuchten Signalwege eine unterschiedliche Dynamik auf und stellen daher unterschiedliche Anforderungen an die Analyse.
Der TGFβ-Signalweg ist ein multifunktionales Signalsystem, das zelluläre Prozesse im Bereich von Proliferation und Migration bis hin zu Differenzierung und Zelltod reguliert. Störungen der Zellantwort auf TGFβ sind an schweren Erkrankungen des Menschen beteiligt. Es wurde gezeigt, dass die Reaktion auf eine gegebene Dosis von TGFβ zellspezifisch durch die Konzentration der relevanten Signalproteine bestimmt wird und dass die beobachtete Heterogenität bei der Signalproteinexpression zur Aufspaltung der Zellen in Klassen mit qualitativ unterschiedlicher Signaldynamik führt. Wie sich die Dynamik zwischen den Signalklassen unterscheidet, ließ sich mithilfe des Ansatzes des überwachten Lernverfahrens quantifizieren.
Studien haben die positive Wirkung einer hyperthermischen Behandlung in Kombination mit Bestrahlung oder Chemotherapie bei Krebs gezeigt. Ich wollte verstehen, wie die Dynamik von p53 nach genotoxischem Stress durch die Temperatur über einen physiologisch relevanten Bereich moduliert wird. Im Bereich von 33°C bis 39°C wird die pulsierende p53-Akkumulationsdynamik in der Frequenz moduliert. Die in einem klinischen Umfeld eingesetzte leichte Hyperthermie oberhalb von 40°C wurde simuliert und dabei konnte ein reversibler Phasenübergang in Richtung einer anhaltenden Hyperakkumulation beobachtet werden. Dies stört die kanonische p53-Antwort auf DNA-Doppelstrangbrüche.
Der Blick auf p53 wurde erweiterte um die gleichzeitige Betrachtung einer weiteren Komponente des Signalweges. p21, ein Inhibitor von Cyclin-abhängigen Kinasen, ist der Mediator von p53 in der Wachstumssuppression und ein Marker der zellulären Seneszenz. Das p53 Signal kodiert Informationen über Signalintensität, -dauer und -identität in komplexer Dynamik. Ich habe untersucht, wie diese Dynamik mit der Dynamik von p21 in derselben Zelle zusammenhängt. Es konnte gezeigt werden, dass die p53- und p21-Dynamik nicht unabhängig voneinander sind und dass bestimmte dynamische Muster die dynamische Antwort auf genotoxischen Stress die Populationsdynamik bestimmen. Dies wurde durch die Klassifikation der Dynamiken basierend auf multivariaten Dynamic Time Warping gezeigt. Signalklassen und -dynamik waren mit dem Zellzyklusstatus verbunden.

Part II

Introduction

1

Signalling in Cells

Cells permanently respond to changes in their immediate environment. They sense these changes by receiving multiple simultaneous signals that originate from the extrinsic physical environment and internal biological processes. To respond adequately cells integrate and process incoming signals by multiple signalling pathways that control the flow of information within the cell [7]. These pathways form a complex molecular network that controls the dynamics of the signalling proteins. Thus requires investigators to consider multiple time points when analysing pathway activity [324].
The dynamics of the signalling proteins induce and regulate gene expression programs that modulate the cellular response to altering environmental conditions [275]. Analysing signalling pathway components with high temporal and spatial resolution is essential for understanding of the dynamic and specific regulation of cellular processes. Roughly, cellular signalling can be classified into different biological mechanisms, the processing of extrinsic signals e.g. cell-to-cell signalling as well as sensing and responding to changes in the extracellular environment and intrinsic signals e.g. cellular DNA damage or metabolic state.
To enable the processing of extrinsic signals from their surrounding environment, cells bear specific membrane receptors on their surface that can sense different input types. These receptors initiate a quantitative cellular response by transmitting the information through signalling pathways to the nucleus [326], for example, during development, when morphogens precisely determine cell fates according to spatial localization [122].
Additionally, intracellular receptors are proteins found inside of the cell, typically in the cytoplasm or nucleus. In most cases, the ligands of intracellular receptors are small, hydrophobic molecules, since they must be able to cross the plasma membrane in order to reach their receptors. For example, estrogen and androgen receptors are located in the cytosol and upon activation translocate into the nucleus there they bind to DNA and regulate the activity of a specific set of target genes [76].
Both, the extra- and intracellular receptors trigger the intracellular machinery that initiate the cellular response. They are capable of and flexible as well as adaptive in binding one or more ligand. The receptors pass the information pro-

vided by the ligand thought the membrane by a conformational change that can be detected by intracellular proteins. Many of the membrane bound receptors oligomerize through lateral diffusion during signalling. These multi-pass receptors control cellular metabolism or cell migration by promoting for example immediate and reversible changes in pH, membrane polarity, calcium flux [44]. In general kinase domains of the intracellular machinery that responds to the activated receptor complex become activated upon ligand binding. Cellular signalling enables the cells to adjust their fate and function in accordance to the variable environmental context, steering diverse processes that range from guiding embryonic development to maintenance of adult tissue homoeostasis.

Sensing and processing of internal signals initiate a quantitative cellular response in a similar way. For example, DNA damage triggers recruitment of multiprotein complexes (sensors) that then activate the transducers ataxia telangiectesia mutated (ATM) and ATR (ATM and Rad3 related), which belong to the phosphoinositide 3-kinase–like kinase family [15]. Once activated, ATM and ATR phosphorylate a bunch of substrates, initiating a cascade that results in cell cycle arrest and DNA repair [384, 25].

Cellular signalling is characterized by gradual on-turning/off-turning mechanisms that regulate ongoing signal stability and strength on a spatial and temporal scale [202]. Multiple processes e.g. molecular interactions, chemical modifications, conformational changes, stability alterations, confined localization and time-dependent availability of co-factors [202] regulate this on-turning/off-turning mechanisms.

Transcriptional networks are essential for complex biological systems. The gene expression is primarily controlled by interaction of the transcriptional machinery with regulatory modules and induced changes within the chromatin structure [124, 191, 273]. To maintain cellular integrity cytoplasmic intracellular signalling is highly organized in space and time.

In the last years, system biology has provided profound insight into cellular signalling processes [44]. The growing understanding of cellular signalling networks could be achieved due to application of continued advances in biological, technical and computational methods for data acquisition and analysis.

In the course of the work, the focus will be set on the computational analysis of imaging data acquired using high-resolution time-lapse live-cell imaging of fluorescent reporter systems. The combination of fluorescent reporter systems and time-lapse microscopy enables quantitative time-resolved measurements of pathway activity at the single-cell level and may contribute as a cornerstone to complete the picture of cellular signalling mechanisms and to the elucidation of the spatial and temporal aspects that dynamically shape signalling at the single cell level.

1.1 Molecular dynamics encode cellular information flow

Multiple signalling pathways in a diverse set of organisms use dynamics to dictate distinct outcomes [262]. The cellular response to extrinsic or intrinsic signals is intrinsically encoded by the single-cell expression of the pathway components. A controlled equilibrium in which degradation, production and stability of the signalling pathways components is highly regulated is essential for responding appropriately to altering environmental conditions. Shifts in the equilibrium of the signalling chain, caused by e.g. mutations, can lead to serious malfunctions of living cells.

The study of signalling pathways on the single cell level has revealed important role of dynamics in regulating the fate of the cell [27, 274]. The information of the signal is encoded in the dynamics of quantity, localization and activity of signalling proteins. These encoding modulates several features of temporal signal characteristics like frequency, amplitude or duration. Complex negative and positive molecular regulatory mechanisms act on all levels of signalling pathways and provide instruments for a tightly temporal and spatial orchestrated flow of information intra and inter cells. Undisturbed procession and integration of inter- and intracellular information guarantees appropriate context dependent cell fate decisions. Inter- and intracellular information includes among others such diverse features like cell cycle status, activity of the multiple signalling pathways, cellular neighbourhood, electrolyte concentration or temperature [314]. As a consequence, genetic identical cells will not react homogeneous to a fix stimulus due to the heterogeneous inter- and intracellular context [326] they are embedded in.

Many chemotherapeutic drugs kill only a fraction of cancer cells even in a population of essentially identical cells. Studding the variation in the signalling dynamics of p53 in single cells to explain the heterogeneity in cellular response to drugs and other external stimuli [262] has revealed that several factors contribute to the heterogeneous response. Pro- and anti-apoptotic pathways are upregulated by the same stimulus and balanced out so that an isogenic populations exhibit a mixture of different outcomes.

Sequence specific transcription factors (TFs) steer the expression of target genes dynamically through the interaction with *cis*-regulatory TFs [155]. It has been shown that for severe pathways TGFβ, p53, TNF-α, and NF-κB signalling, that non-genetic heterogeneity shapes the signalling dynamics of individual cell and determine the prototypic response to extracellular stimulation [112, 13, 319, 332, 274, 188, 326]. Non-genetic heterogeneity in a genetically identical population of cells emerges from several sources that range from cell cycle state, the local environment to the state of the components of the signalling pathways.

The dynamic and heterogeneous cellular responses to external stimuli is in essence a description of the life process itself [44].

2

The central hub of the stress response p53

p53 was discovered in the late 70s by five independent groups studying SV40-derived tumour antigens. In 1979 five independent groups published similar findings, a complex between the SV40 large T-antigen and a non-viral 53 kDa protein [193, 184, 198, 173, 222, 312]. It was identified as an oncogene because it was detected in transformed cancer cells but not in healthy cells [72, 286]. Supported was the role of p53 as an oncogene by transducing p53 cDNA into normal cells inducing transformation into cancerous cells [90, 156]. The name p53 for this non-viral 53 kDa protein was established in 1983 during a Workshop in Oxted, UK. The name p53 is based on the estimated mass of 53 kDa. However, the actual weight of the human p53 protein is only 43.7 kDa.

In the late 80s evidences emerged that in contrast to previous indications, p53 is a tumour suppressor. For example, it was found that in the human leukaemia-derived cell line HL60 the coding sequence of p53 was virtually deleted [365]. At the same time a p53 cDNA clone was generated that was unable to reproduce the transforming effects observed with the earlier clones [193]. When comparing sequences of p53 (encoded by TP53) cDNA from different sources, published at this time, it became obvious that none of the published p53 sequences which originated from cancer cells were identical and carried mutations in the coding sequence.

In the early 90s it was shown that wtp53 effectively repressed the transformation [89, 97] while p53 mutants indeed possess attributes of oncogenes [19, 18]. Furthermore, humans carrying germ line mutations have an increased cancer predisposition e.g. Li-Fraumeni syndrome [213] and p53 knockouts in animal models develop cancer with a very high penetrance [82]. p53 become undoubtedly the most extensively studied gene and protein.

In this forty years that have passed since p53 was discovered it has made its way from just another interesting protein to one of the most extensively studied genes and proteins, from an oncogene to a tumour suppressor, unveiling its standing as the central hub of the cellular stress response [193].

By which means p53 regulates the mechanisms that allow p53 the suppression of tumours was enlightened in the following years.

Hence, p53 is a role model of a tumour suppressor with the central property of p53 to act as a transcription factor (TF). It controls among other biological processes

the fate of a cell in response to stress, steers cell cycle arrest [228], apoptosis [376] and senescence [371].
p53 plays a central role in the development of cancer. A great fraction of all tumours harbour a loss of p53 function either directly through mutation or indirectly through disturbances in several mechanisms that control the p53 response [44]. The fundamental function of p53 as tumour suppressor can be emphasised by the fact that it is the most commonly mutated gene in cancer with over 50 % of tumours expressing a mutant variant [318, 115]. 74% of the mutations occur in the central DNA binding domain of p53 with a significant accumulation at six hot spot codons [318]. If the function of p53 is not directly disturbed through mutation many cancer types bear indirect disturbances in other mechanisms that control the p53 response [44]. In this context, key components of the p53 network like Mouse double minute 2 homolog (MDM2), and the upstream kinase regulators Checkpoint kinase 2 (Chk2) and ATM are among the most frequently mutated genes in cancer [44].
In recent years, the interplay with metabolic pathways, cytokines required for embryo implantation [193] and the cross play with other signalling pathways has become of emerging interest. The focus on the tumour suppressor role 'guardian of the genome' has shifted towards the role of p53 as the central hub of the cellular stress response (Fig. 2.1). The p53 network is activated by several different distinct mechanisms [303] and orchestrates the response to various sources of stress ranging from DNA damage, hypoxia, nitric oxide to depletion of glucose and other nutrients [349].
The p53 protein is steady expressed and normally directly degraded so that it is present at low levels in unstressed cells. In response to stress it accumulates in the nucleus and binds to DNA sequences specific. The DNA damage response (DDR) is one of most intense studied activators of the tumour suppressor p53. DDR activation can induce transient cell cycle arrest in G1 or G2 or terminal cell fates, such as senescence and apoptosis [164].
Upstream kinases of the damage response activate p53 and induce specific expression patterns of p53 target genes[280]. The p53 response is remarkably flexible and depends on the cell type, its differentiation state, stress conditions, and collaborating environmental signals [164, 262].
Due to alternative splicing and alternative transcriptional initiation the TP53 gene encodes nine known different isoforms observed in different tissues and during different stages of development. The precise roles of each of the reported p53 isoforms remain largely unknown [193, 42].

2.1 The master regulator of the cellular stress response

The tumour suppressor gene p53, 'the guardian of the genome', has an outstanding role in preventing malignant transformation by maintaining the genetic integrity of proliferating cells [348, 183, 193]. The TP53 gene encodes among other isoforms

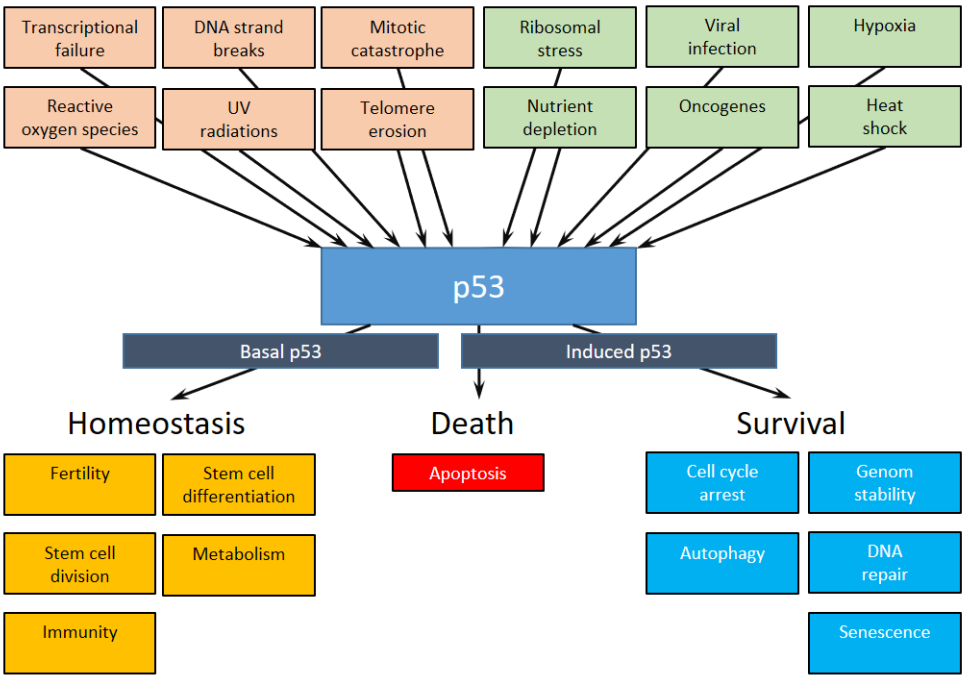

Fig. 2.1. Several sources of stress induce a p53 response

Multiple different stress stimuli trigger the p53 response. The biological outcomes of the response, that alters the gene expression of dozens of genes, depends on the cell type and the type, intensity and duration of the activating stress. Stress events that induce the p53 DDR pathways are arranged on the top left. The figure is inspired by a review from Meek [221].

a full length polypeptide with a length of 393 amino acids. The p53 sequence is divided in three functional domains: An N-terminal transactivation domain that interacts with several transcription factors and MDM2, the DNA binding core domain and the C-terminal domain that contains the tetramerization and regulatory function as well as nuclear localization and export signals [40]. p53 functions primarily as a tightly regulated homotetrameric transcription factor (TF) that encompasses both activation and transrepression activities.

p53 is part of a protective system that prevents cells from acquiring cancerous properties. The potency of p53 as a tumour suppressor is partially grounded on its ability to arrest or eliminate cells after DNA damage [200] and to halt the proliferation in response to aberrant oncogene expression [205]. Activation of p53 reduces the risk of propagating DNA damage by elimination or repair of damaged cells [364]. Elimination is driven by DDR kinases phosphorylating p53, driving cell-cycle arrest, senescence or apoptosis. p53 stimulates DNA repair by activating target genes that encode components of the DNA repair machinery. The tumour suppressor function of p53 is crucial for maintaining genomic integrity and controlled cell growth[192, 23, 104]. In response to a wide range of different stress

stimuli (Fig. 2.1), p53 is induced essentially by blocking its degradation, leading to an increase in its cellular levels [221]. Changes in the p53 levels alter, in a context-dependent manner, the outcome of several biological processes. Independent from its role as 'guardian of the genome', p53 regulates a range of genes essential for a variety of cellular processes, often when present at basal non-induced levels [221]. So for example, p53 regulates the proliferation and differentiation of stem cells [9, 316] or is able to restrict cellular longevity by promoting the ageing process [212, 340].

Mutated p53 can promote tumorigenesis due to a loss off, gain off or altered function [348, 115, 260] and changed regulation of hundreds of genes that are direct targets for p53-mediated trans-activation [225]. p53 activity has broadly diverse consequences and its activation is regulated in a context dependent manner [44]. Thus, activity of p53 must be tightly orchestrated.

p53 possesses two functional trans-activation domains [387] and it can bind to specific DNA sequences [88, 168]. This target response element (RE) sequences are central for the function of p53. The responsiveness of REs to p53 is dependent on several factors like sequence specificity [155], p53 level and cofactors[40, 269], thus yields a large variation in p53-dependent expression profiles [93, 134, 329]. p53 and other TFs can be in competition for neighbouring binding sites [139, 380] but also can cooperate having negative or positive effects on transactivation. Interactions of p53 and other TFs are reported e.g. for SMADs[67, 363] and NFκB [147, 356]. Depending on the interaction with TFs and cofactors, p53 either activates or represses target gene expression [177].

The set of cofactors that interact with p53 can dependent on the level of stress, for example cofactors that stimulate apoptotic genes (e.g. PIN1, ASPP1, ASPP2 [214, 34, 292]), cooperate with p53 in the case of high stress levels as irreparable damage while low levels of damage induces interaction with co-factors that influence cell cycle arrest (e.g. iASPP, HZF, MUC1 [214, 33, 70, 142]).

p53 is not limited to its function as a TF. p53 interacts with apoptosis related proteins of the BCL-2-like proteins that regulate the mitochondrial release of cytochrom c [256, 94]. Cytochrom c contributes to the apoptotic dismantling of the cell [111].

The human p53 gene is located at chromosome 17p13.1. It is composed of 19.198 nucleotides arranged in 11 exons[31]. Splicing isoforms of p53 can act as transcription factors, repressors of transcription and differ in stability of the Protein and DNA binding [269, 373, 223, 233, 43]. The N-terminal transactivation domain of p53 has 3 possible phosphorylation sites[2]. Phosphorylation patterns influence the affinity for binding of proteins like MDM2[297], CBP[334] or p300[186] that regulate the activity of p53. CBP and p300 are histone acetyltransferases that acetylate lysine residues of histones and thereby relax the chromatin [269].

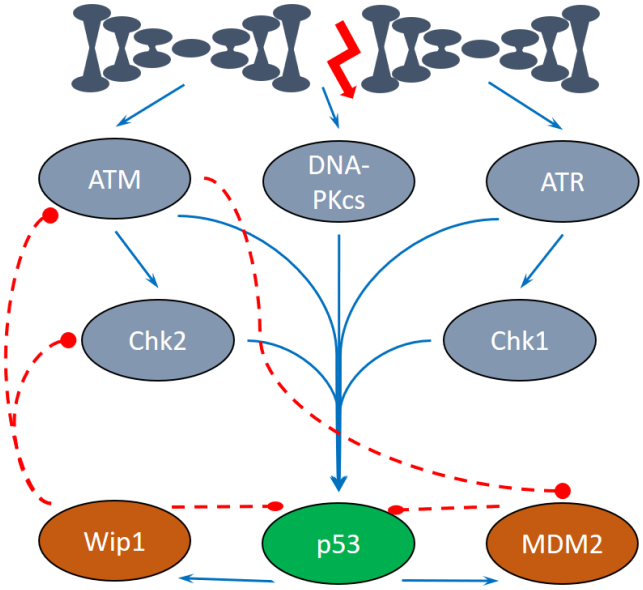

Fig. 2.2. DDR p53 network

After double strand breaks upstream kinases activate p53. P53 activation prevents p53 from proteasomal degradation promoted by MDM2, hence p53 can accumulate. The two negative feedback loops MDM2 and Wip1 shape the p53 response dynamics. The figure is based on a figure from [98].

2.2 p53 is the central hub in a broad and flexible network

To understand the biological properties of p53 and its surrounding regulatory network, it is crucial to explore its interaction partners.
In 1992 the MDM2 protein, probably the major negative regulator and gatekeeper of p53, was found to bind p53 and inhibit its activity [234, 193]. In the canonical model of the network p53 is kept at low levels in healthy cells by its counterpart MDM2 [265]. In absence of stress the p53-specific E3 ubiquitin ligase MDM2 ubiquitinates p53, thus initiating promote the proteasomal degradation of p53 [133, 178, 143]. Also, MDM2 binding blocks the trans-activation domain of p53 inhibiting its function. Upon exposure to stress p53, gets phosphorylated by different kinases within the N-terminal domain. These phosphorylation disrupt the interaction between p53 and MDM2 so that the proteasomal degradation is stopped and p53 is activated and can accumulate within the nucleus. Higher levels of p53 in the nucleus can unfold the central property to act as a TF by binding to target gene promotors. Homotetrameric complexes of p53 encompass both transactivation and transrepression activities by binding to sequence-specific gene promoters [177].
Most of the regulation of the p53 activity goes through the MDM2 protein interaction with other p53 network components like AKT-1, AMP kinase, ARF and cyclin D-PP2A [31, 129]. Activity of MDM2 is dependent on its cellular level,

p53 binding affinity and E3 ligase efficiency[193]. Exemplary, ARF activates p53 by binding MDM2 and promoting its rapid degradation [383]. The stress dependent switch-like activation of p53 is guaranteed substantially by MDM2. p53 stabilization happens immediately after exposure to various cellular stresses[227], including DNA damage and replication stress produced by deregulated oncogenes. Stress can have several faces, especially risk for the genome integrity.

This activation is accompanied by an increase in a variety of post-translational modifications of p53 and MDM2 ranging from acetylation, phosphorylation and methylation to poly(ADP-ribosyl)ation, neddylation, sumoylation, and non-proteolytic monoubiquitylation [336, 40, 258]. Mechanisms leading to p53 activation can be stimulus dependent. These post-translational modifications that can alter p53 affinity for different target genes and influence target gene bias [179]. These modifications are catalysed by a large number of enzymes; e.g. at least a dozen different kinases ensure p53 phosphorylation on at least 18 different residues[258]. Several of these modifications prevent the p53 and MDM2 interaction and promote nuclear p53 accumulation. Hence, DNA damage promotes p53 phosphorylation, blocking its MDM2-mediated degradation [307], whereas oncogenic signalling induces the tumour suppressor ARF to inhibit MDM2 [271, 276, 383].

The interaction of the p53 trans-activation domain and the MDM2 binding domain is essential for functional tumour suppression [2]. Furthermore, MDM2 is a direct transcriptional target of p53 forming a negative feedback loop. p53 induces MDM2 promoting consequently its own deactivation and degradation [22, 368, 267].

The kinases ATM, ATR and DNA-PK have been shown to phosphorylate p53 and MDM2, leading to stabilization of the p53 protein [220, 56]. Also the checkpoint kinases Chk1 and Chk2 are activated by ATM and ATR [177]. Chk1 and Chk2 phosphorylate p53 at multiple DNA damage-inducible sites regulating stability and function of the protein [306]. The variation in DNA repair efficiencies and ATM activities can explain the dynamical space of p53 across cell lines[324].

The MDM2 feedback is not sufficient to explain the complex p53 dynamics observed. An additional feedback loop was found between wild-type p53-induced phosphatase 1 (Wip1) and p53. The expression Wip1 is controlled by p53. Wip1 modulates several components of the p53 network as it is dephosphorylating p53 by PI3K-like kinases [207], inhibits p53 phosphorylation by MEK [125], dephosphorylates ATM [310] and Chk2 [106, 257]. The schema of the p53 network is shown in Fig. 2.2.

The cross play of p53 and its negative regulators induces several context-dependent p53 dynamics. For example, ionizing radiation (IR) induces repeated p53 accumulation in pulses of uniform amplitude [26, 203, 274] and duration and ultraviolet radiation (UV) triggers single sustained pulses with dose-dependent amplitude and duration [27].

Recent studies indicate that p53 itself is not only responsible for the regulation of transcription, while p21 is required for the indirect down regulation of genes induced by p53 [102].

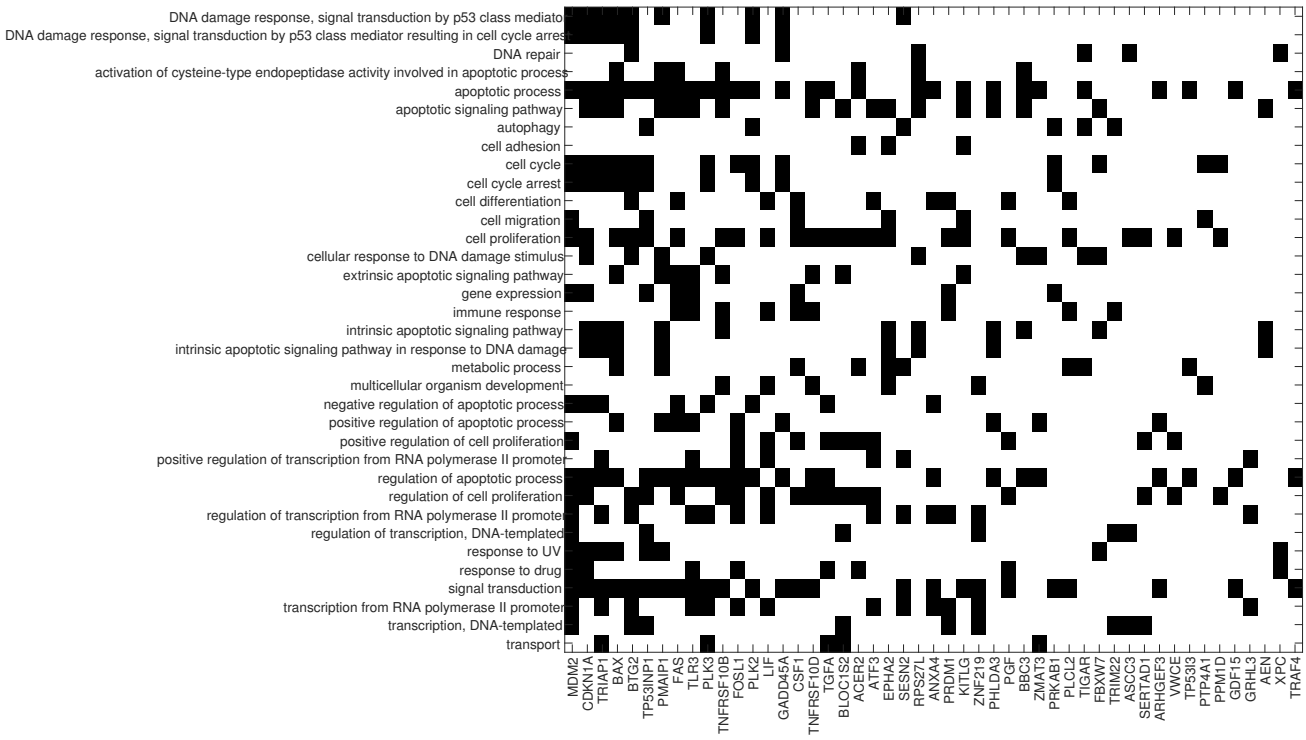

Fig. 2.3. Top p53 target genes proposed by Fischer and their biological function

From the top 116 genes identified as p53 targets in at least 6 out of 16 genome-wide data sets listed in Fischer [100] the functional annotation according to Gene Ontology [14] is shown for target genes that have at least 3 known biological functions.

2.3 p53 target genes

How the p53 tumour suppressor function, primary mediated by its primary role as a TF, is guaranteed by steering the expression profile of its target genes by dynamic rewiring of the p53 network is essential for the understanding of cancer development. Its various functions are based on controlling the expression of distinct sets of its target genes. In his 2017 review Fischer identifies 3661 (319 from individual gene studies and 3509 from high-throughput data sets) proposed p53 target genes involved in multiple cellular responses, including cell cycle arrest, DNA repair, apoptosis, metabolism, autophagy, mRNA translation, feedback mechanisms [100] repression of pluripotency, cellular plasticity and ferroptosis [16]. It is quite likely that there are many false positive findings among these 3661 potential p53 target genes. Likely the real number of target genes does not exceed a few hundred [101] (Fig. 2.3). Fischer found limited reproducibility and numerous contradictions in the different data sets. p53 target genes can be divided in genes that are activated and genes that are repressed by p53. The sequence specific TF p53 controls various distinct transcriptional programs.

p53 can induce or repress the expression of its target genes[138, 370]. Frequently encountered p53 target genes encode proteins that are involved in apoptosis, cell cycle control [225] and growth arrest. Cell cycle progression is controlled by p53 among others by the trans-activation of the cyclin-dependent kinase inhibitor p21 [87]. Target genes of special importance for the proapoptotic function of p53 are PUMA, NOXA (BH-3 only members of the BCL-2 family) and BAX [269, 231].
Cell stress induces p53 activation, for example, DNA damage initialises p53 target gene activation through a series of p53 pulses that encode the specific response [28]. p53 activates target genes by binding to p53 specific response elements (REs). There are canonical p53 REs, that are formed by two separated decameric half-sites with the consensus sequence RRRCWWGYYY (R = Purine, W = A—T and Y = Pyrimidine) [88], and non-canonical REs [333]. These p53 binding sides are independent of cell type and treatment [342].
In recently proposed models down regulation, of target genes in response to p53 is indirect, mainly directed through the cyclin-dependent kinase inhibitor 1A (CDKN1A) [102]. The gene CDKN1A encodes one of the main effectors of the p53 DDR p21. Best known in this context is the down regulation of numerous cell cycle genes by the p53-p21 pathway[201, 330, 337, 299, 169, 385]. Cell cycle transition from G1 to S and G2 to mitosis is regulated by cyclin-dependent kinase (CDK) family of proteins that are activated by cyclins [343] and inhibited by p21. Induction of p21 by p53 after exposure to stress prevents cell cycle progression. p21 is essential for the cell cycle arrest at transition from G1 to the S phase by inhibition of CDK2/4 [74] and contributes to G2 arrest by inhibition of CDK1 [48]. p21 was also reported to inhibit DNA synthesis and DNA repair and to promote apoptosis [303].
As mentioned before among the different targets genes of p53 is a group of p53 pathway regulators involved in negative feedback loops in particular MDM2 and Wip1.

2.4 Dynamic rewiring of the p53 network

Recent studies have reported that pulsatile p53 dynamics trigger the expression of genes involved in DNA repair and cell-cycle arrest, while sustained p53 accumulation initiates activation of senescence and apoptotic genes [27, 274]. The dynamic patters exhibited by p53 in response to radiation, UV, hypoxia or oncogenes induces different biological effects. p53 is the central hub of the molecular network which is dynamically steering the cellular response to stress. A fundamental question is what mechanisms contribute to the cell-to-cell variability in the stress dependent p53 activation.
Like several signalling pathways (nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB), nuclear factor of activated T cells (NFAT), and extracellular signal–regulated kinase (ERK) or Transforming growth factor beta (TGFβ) [64, 375, 21, 326]) the p53 signalling uses complex dynamics to encode information about signal intensity, duration and identity. The different dynamics observed

for p53 after exposure to stress are associated with the crosstalk between p53 and separate inputs and outputs of other cellular signalling cascades. P53 modulates stress driven cell fate by controlling and balancing out different biological responses. Its embedded in a densely populated and interconnected network of regulators and effectors. The crossplay between the different biological processes triggers context dependent the biological outcome. For example, alterations in p53 control of metabolism undoubtedly contribute to apoptosis, autophagy, and ferroptosis [110]. Autophagy delays apoptosis by p53 induced reduction of PUMA [335].
In response to double-strand breaks, feedback loops cause p53 to oscillate [180, 274] triggering a cascade of transcription enforcing cell cycle arrest and, if the damage is severe enough, committing the cell to a terminal fate. Pulsatile dynamics either result in resumption of proliferation or in permanent arrest if oscillations persist [324]. Sustained p53 accumulation in the nucleus is associated with permanent cell cycle arrest [57].
Stewart-Ornstein and Lahav [324] have shown that small-molecule inhibition of ATM converted sustained p53 signalling to the classical pulsatile behaviour, whereas inhibitors of DNA repair, or compounds that increase ATM activity, promoted sustained signalling. They could also show that small-molecule inhibitors of mTOR, CDK, and PARP also altered p53 response dynamic.
Biological information is not only encoded in the structural components of a cell, but also in the dynamics of signalling molecules [29]. P53 is a well embedded in the regulation of genes and pathways that contribute to the cellular homeostasis. For example, the uptake of Glucose, a key nutrient, into the cytosol is decreased by p53 directly by downregulating the gene expression of glucose transporters GLUT1 and GLUT4 [298]

2.5 Temperature effects on the p53 network

It is known that p53 plays often a central role in the development of cancer and that a combinational treatment of hyperthermia together with radiation and anticancer agents has been used clinically where it has shown to be beneficial to the treatment outcome [6]. It is known that several elements of the p53 pathway as well as other proteins that interact with p53 are modulated by temperature [120]. However, the underlying mechanism of signal transduction and set of genes involved in this process are still poorly understood. How are the single cell dynamics of the guardian of the genome p53 upon genotoxic stress modulated by the temperature of the surrounding environment?
Cancer treatment concepts aim to destroy neoplastic cells. The straightforward approach, surgical operations, aims to directly remove the cancerous tissue. Other more sophisticated methods focus on putting to death cancer cells in situ while avoiding so much harm as possible to healthy cells. Non-invasive methods range from the widely used radiation therapy and chemotherapy to novel approaches, such as immunotherapy, monoclonal antibodies or stem cell transplants. In addi-

tion, adjuvant therapies like hormone therapy or hyperthermia are used to improve the effectiveness of radiation therapy or chemotherapy. Hyperthermia is a procedure which elevates the temperature of tumour-loaded tissue to 39.5–43°C and is generally applied as an additive treatment to established non-surgical cancer treatments [60]. The elevation of the local or whole body temperature is achieved by different methods, as superficial hyperthermia, regional hyperthermia, interstitial hyperthermia, whole body hyperthermia and hyperthermic isolated limb perfusion [369].
In several countries authorities accept adjuvant hyperthermia for radiotherapy and chemotherapy as regular treatment for distinct cancer indications[295]. For example, the Charité - Universitätsmedizin Berlin is treating patients with radiotherapy and chemotherapy and adjuvant hyperthermia for more than two decades. An increasing number of phase II and III clinical trials have shown the effectiveness of a combined cancer therapy approach [71, 208, 54]. It has been hypothesized that cytotoxicity induced by chemo- and radiotherapy is enhanced by hyperthermia ('thermal chemosensitization', 'thermal radiosensitization') [137, 54].
Temperature has a direct influence on the tumour microenvironment and changes the blood flow as well as oxygen and nutrient distribution in the tumour tissue [215]. Direct cell killing effects have been shown to be markedly enhanced at temperatures above 43°C [79, 137].
The underlying molecular mechanisms of hyperthermia are investigated since decades. Temperature affects biological processes directly by changing the speed of enzyme-catalysed reactions. The Arrhenius equation states that the speed of a biochemical reaction doubles when the temperature rises by 10°C. This holds until the proteins intimately involved in the reaction denaturize [140]. Furthermore, increased temperature affects several molecular functions of proteins not just by denaturing also by inducting protein misfolding. The environmental temperature affects fluidity and stability of cellular membranes and induces changes of cytoskeletal organization (cell shape, mitotic apparatus, intracytoplasmic membranes such as endoplasmic reticulum and lysosomes) [137].
Hyperthermia modulates the activity of several pathways with different function. For example, hyperthermia can induce the expression of heat shock proteins, activate the immune response, activate apoptosis pathways or alter cell cycle regulatory signalling pathways [215].
Radiation therapy and chemotherapeutics targets rapidly dividing cancer cells by directly or indirectly inducing DNA damage [366]. Hyperthermia interferes with DNA repair pathways [162] thus DNA damage-based treatment modalities benefit from adjuvant heat application. It has been proposed that hyperthermia sensitizes cells to agents that interfere with cell cycle checkpoints [255].
Several processes involved in the cellular DDR or connected to the p53 signalling cascade are known to be heat sensitive. Direct effects of hyperthermia, that affect DNA integrity, are the induction of DNA lesions (SSBs and DSBs), promoting the formation of γH2AX foci and the autophosphorylation and activation of ATM [215]. DNA replication and repair is as well decelerated by reduced DNA poly-

merase and DNA topoisomerase activity. The pathway orchestrated by ATM and ATR is the central regulator of the DDR network. Cell cycle checkpoint activation in response to DNA damage is managed by ATM and ATR [390, 187]. ATM and ATR are two kinases that directly interact with p53 and steer p53 dynamics. ATM directly phosphorylates Chk2 in response to heat [230, 339]. ATR and Chk1 are as well heat activated [339, 108]. The degree of thermosensitivity is dependent on the cell cycle phase [360, 137]. Mre11, Rad50 and Nbs1 (MRN) form tight complexes which after irradiation rapidly form foci at sides of DSBs [300]. In general, the MRN complex is required in the initial processing of DSBs by the activation of ATM [341]. After cellular exposure to heat the MRN complex does not bind any longer to due to the activity of HSP70 [300]. The expression of members of the family of heat shock proteins (HSPs)is elevated after heat exposure and controlled by members of the protein family heat shock factors (HSF1-4). HSPs are the major molecular chaperones. During synthesis HSPs assist in proper protein folding, and during proteotoxic stress HSPs protect stress-labile proteins and contribute to proteolysis of damaged proteins [361].

The most frequently mutated gene in human cancer is p53. A clear appreciation, how spatial and temporal cellular context drives p53 diverse effects remains missing [164]. The kinases ATM and ATR trigger activation and accumulation of p53 [12, 308]. It had been shown that implicated activation of ATM and a subset of its downstream targets, including p53, is independent of the MRN complex [149, 230]. p53 is as well activated by the thioredoxin-dependent redox state and modulation of checkpoint regulators Gadd45a and Cdc2 at 41°C [160]. The proper functioning of p53 is affected by HSPs by balancing p53 synthesis, degradation and its nuclear translocation [361]. Binding of HSP90 to p53 makes p53 resistant to ubiquitination and degradation mediated by MDM2 and enhances the binding of p53 to the p21/WAF1 promoter. It has also been shown that Hsp70 and Hsp40 are involved in stabilization of p53-DNA complexes [351, 391].

Hyperthermia is not only an artificial issue the human body is dealing with. As homoeothermic animals, humans exhibit a stable core temperature of 37°C. At the periphery, especially at the extremities the temperature can be much lower and is depending on the environmental temperature[328, 114]. Not only the ambient temperature is a cause of deviating body temperature. To name a few age, time of day, sex and reproductive status have influence as well [250, 49]. Especially, the activity level has substantial impact. During physical exertion, the body core temperatures can reach above 40°C [197]. Altered temperature is as well observed in pathological context, associated with several diseases ranging from vascular issues where the blood flow associated heat dissipation is perturbed to fever, which is induced by multiple sources (infections, autoimmune issues) [121] or cancer [211]. Studies indicate that longer deviations in human core temperature of more than 4°C can result in physiological impairments and fatality [236].

3

Pathways in TGFβ signalling network

The transforming growth factor β (TGFβ) superfamily is comprised of over 30 different secreted factors [358]. Ligands of the TGFβ superfamily bind to serine/threonine kinase receptors that activate SMAD proteins. Activated SMAD proteins accumulate in the nucleus and regulate the gene expression. Beside the canonical signal transduction utilizing SMAD proteins other pathways are also employed in a cell-specific manner.
The TGFβ superfamily can be classified in two functional groups: the TGFβ like family members consisting of TGFβ 1-3, Activin and Nodal and the Bone Morphogenetic Proteins (BMP) like family members consisting BMP, Growth and Differentiation Factors (GDF) and anti-Müllerianhormone (AMH). The groups differ in their ligand structure, downstream SMADS, activation and induced signalling cascade [145].
TGFβ induced signalling is one of the most investigated signalling cascades [58]. Its key function is the maintenance of the tissue homoeostasis by regulating the transcriptions of genes responsible for proliferation, survival and cytostasis [135, 358]. TGFβ is a pleiotropic cytokine responsible for regulating a myriad diverse cellular processes, including differentiation, apoptosis, adhesion, motility, tissue regeneration, immune responses, tumorigenesis, embryonic development, adult tissue homeostasis and cellular microenvironment in nearly every human cell type [92, 218, 78, 91, 150, 305, 311, 36, 135, 289, 296, 135].
A disturbed information flow in the TGFβ pathway can change the cellular response to TGFβ ligand and cause severe human health defects like cancer, fibrosis, vascular disorders and autoimmune diseases [37, 218, 117]. In the canonical TGFβ signalling the essential step is the translocation of the active SMAD transcription factor complex from the cytosol to the nucleus, where it can control the rate of transcription of its target genes.
TGFβ has a complex dual role in cancer and can act as a tumour suppressor by inhibiting the uncontrolled growth of tissue especially epithelial cells by inducing cell cycle arrest and promoting apoptosis or act as a tumour promotor by inducing growth, survival, motility, invasion, and metastasis [150, 36, 218]. Multiple alterations in components of the TGFβ signalling pathways are likely to be found in human cancer cells [218, 150, 194]. In early stages of tumourige-

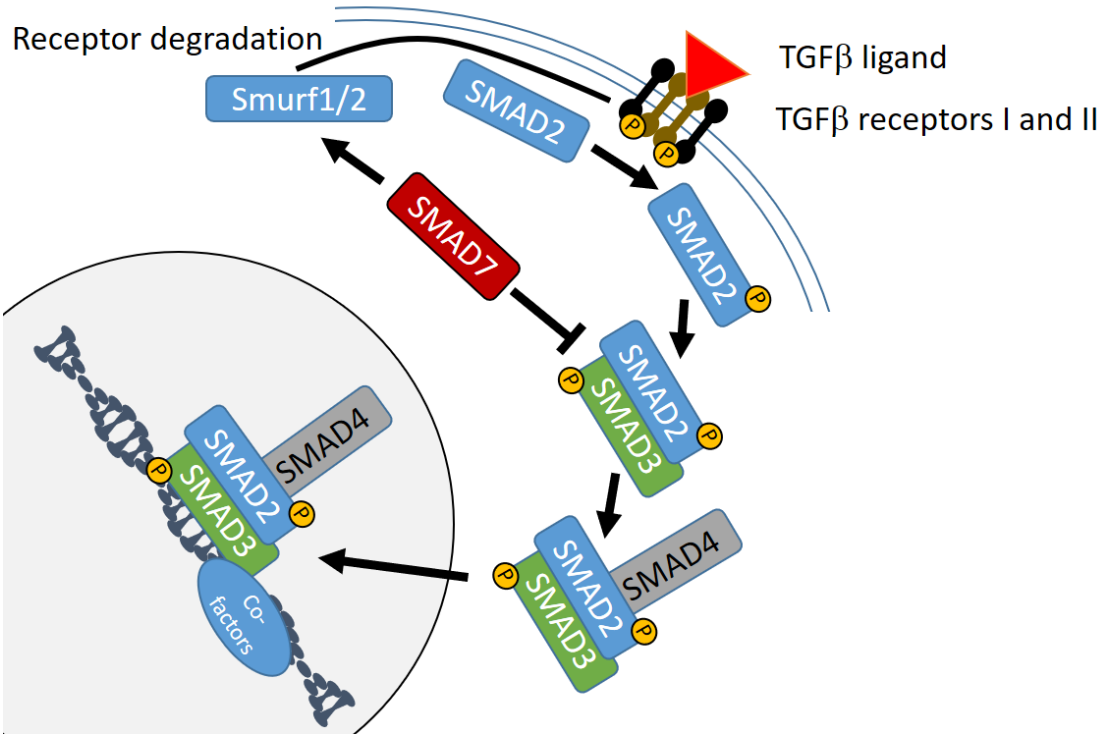

Fig. 3.1. Canonical pathways in TGFβ signalling network

After ligand binding to TGFβ Type II receptor the active peptide recruits a TGFβ Type I receptor and forms a complex. The active complex then phosphorylates receptor-associated SMAD proteins (R-SMADs), including SMAD2/3. Activation of SMAD2/3 leads to formation of complexes with SMAD4. This complex shuttles into the nucleus and regulates the expression of specific target genes. Among the target genes is SMAD7 that has the role of a negative feedback loop in the pathway. SMAD7 inhibits TGFβ prevents the formation of SMAD2/SMAD4 complexes, the activation of SMAD2 by TGFβ type I receptor by enhanced SMURF2 activity. SMAD7 inhibits TGFβ signalling by preventing the formation of SMAD2/4 complexes and the interaction of SMAD2 with the type I receptor. Additionally, SMAD7 can activate SMURF1/2, which induces degradation of receptors and SMAD7 via proteasomal and lysosomal pathways. Figure inspired by schema published in [291, 325]

nesis by evading the tumor-suppressive, cell cycle inhibitory role of TGFβ and in late stages [218, 91, 150, 283, 226, 151] by the induction of transcriptional transition that induces epithelial-to-mesenchymal transition (EMT)[116, 77, 218]. EMT is a physiological process, thus acquire cells to lose epithelial properties and gain characteristics of mesenchymal cells[378], a conversion from adherent cellular characteristic to migratory cellular behaviour. How the molecular switch between adherent cellular characteristic and EMT is triggered by $TGF\beta$ is under debate. Changes in the expression of different genes [243, 268] or pathway dynamics may cause the transition [253].

TGFβ alters gene expression of target cells mainly by propagating the signalling

downstream through canonical signal transduction utilizing SMAD proteins [358]. TGFβ furthermore regulates non-canonical SMAD independent pathways such as JNK/p38, Ras-ERK, PI3K/AKT and small GTPases like RhoA [78, 378, 240]. Crosstalk between SMAD dependent and SMAD independent TGFβ regulated pathways and other physiological pathways shape TGFβ signalling context specific [78, 92]. The combined interplay of the canonical and the non-canonical signalling pathways with other signalling cascades triggers the conclusive outcome of the cellular response to TGFβ. The TGFβ signalling is dynamically shaped by the activation state of the receptor complex[378]. TGFβ signalling is tightly regulated and conducted by post-translational modifications of the pathway components, that guide the spatial and temporal activity the of signal [240]. Mutation of TGFβR or SMADs are responsible TGFβ signalling perturbation [218, 117].
TGF-β signal transduction is controlled by multiple mechanisms and proteins, not only at the receptor level. Remarkable is the key negative regulator SMAD7 [264, 202, 153]. It was also proposed that SMAD7 plays an outstanding role as a cross-talk mediator between TGFβ signalling and other signal transduction processes[372].

3.1 The TGFβ network - SMAD-mediated canonical pathways

Three TGFβ (TGFβ1, TGFβ2, TGFβ3) isoforms have been identified in mammals [282, 321]. They are encoded from individual genes located on different chromosomes. TGFβ isoforms display similar, although not identical, biological activity and differential tissue expression [224]. TGFβ1 is the prototypical ligand of the TGFβ superfamily. TGFβ is secreted from cells in a large complex that is cleaved into mature TGFβ and latency-associated protein (LAP) [11, 39].
Canonical TGFβ signal transduction is ensured by transmembrane type I and type II serine/threonine kinase receptors (TGFβRI and TGFβRII), downstream SMAD proteins and other signalling mediators.
Upon ligand binding to the TGFβRII, two TGFβRII form a heterotetrameric complex with two TGFβRI. Within the active complex the TGFβRII autophosphorylates and catalyses the transphosphorylation of TGFβRI [305], thus activating TGFβRI kinase [96]. The activated TGFβRI then propagates signals to members of the SMAD family by phosphorylation of their C-terminal tails [75].
SMADs are transcription factors that constantly shuttle between the cytoplasm and the nucleus [219, 96, 296]. In most vertebrates eight SMADs are found [148]. SMADs are classified in three groups. Receptor regulated (R-) SMADs (SMAD1, SMAD2, SMAD3, SMAD5, SMAD9) that interact directly with the ligand-receptor complex. (Co-)SMADs (SMAD4) that partnering R-SMADs. Inhibiting (I-)SMADs (SMAD6, SMAD7) that suppress the activity of R-SMADs and Co-SMADs. The BMP-like subfamily activates SMAD1, SMAD5, SMAD9 and TGFβ-like subfamily activates SMAD2, SMAD3. The TGFβRI phosphorylated R-SMADs SMAD2 and SMAD3, heterotrimerize with (Co-)SMAD4 and

translocate to the nucleus, where they bind to specific promoter sites and regulate the expression of target genes in collaboration with other co-activators or co-repressor [96, 284].
The Ub-proteasome machinery regulates stability and amounts of R-SMADs and Co-SMADs at steady state and during the active TGFβ signalling cascade [202]. Lysosomal depletion of extracellular TGFβ due to internalization of receptor–ligand complexes [62, 388] and the associated temporal removal of the heterotetrameric receptor complex from the membrane [347] shapes the TGFβ response.
The expression of SMAD7, one of the negative feedback loops of the canonical TGFβ pathway, is increased after TGFβ exposure. SMAD7 is an important crosstalk mediator of TGFβ signalling with other signalling pathways. Stability and function of SMAD7 is regulated by a balance of acetylation (p300), deacetylation (SIRT1, HDAC1), methylation (PRMT1), phosphorylation (MPK38) and ubiquitination (SMURF1, Arkadia together with Axin) [372]. Also other transcription targets can effectively inhibit the signalling progresses, for example SMURF1 or SMURF2 [166, 246, 84] that are directly induced by activated Smad2/3/4 complexes. TGFβ dependent SMAD7 and SMURF1 or SMURF2 induction sustains and increases their cellular levels so that they can inhibit effectively the signalling progresses [202]. The (I-) SMAD7 binds to the activated TGFβR complex and competes with SMAD2 and SMAD3, thus prevents their phosphorylation [386]. The TGFβ induced targets SIK1 and SMAD7 target together TGFβRI for facilitate proteasomal degradation[172]. SMURF2 induces degradation of receptors and SMAD7 via proteasomal and lysosomal pathways [166]. SMAD7 activity is also tightly regulated for example by adaptor proteins like STRAP that stabilizes SMAD7/TGFβR complexes formation or SMURF1 that ubiquitinates SMAD7 [84, 166, 118]. SMAD7 competes with SMAD2 and SMAD3 for binding to activated type I TGFβR [135, 136, 218] and induces type I TGFβR degradation by interacting with SMURF1 and SMURF2 [84, 386]. SMAD7 also recruits E3 ubiquitin (Ub) ligases together with corresponding co-factors to receptor complexes and to the other SMADS [153, 154] inducing degradation by the proteasomal machinery. SMAD7 can also act as disruptor of R-SMADs and Co-SMAD or between R-SMADs and type I receptors interactions[153]. Altered SMAD7 expression can have serious consequences like cancer, tissue fibrosis and inflammatory diseases [372].
SMAD anchor for receptor activation (SARA) is known as a SMAD cofactor that interacts directly with SMAD2/3 and functions to recruit SMAD2/3 to the TGFβ receptor. SARA plays an essential role in TGFβ-induced SMAD2 activation and it may modulate TGFβ signalling through regulating the balance between SMAD2 and SMAD3. SARA also functions as an anchor for catalytic subunit of protein phosphatase 1 (PP1c) and maybe involved in the dephosphorylation of TGFβ type I receptor mediated by SMAD7. The expression of SARA changes as the development of epithelial to mesenchymal transition (EMT) and fibrosis and it plays a critical role in the maintenance of epithelial cell phenotype. Modulation

of SARA may provide a new therapeutic approach to TGFβ-mediated EMT and fibrosis [331].
Among the targets that are under transcriptional regulation of the active SMAD2 and SMAD3 complex are epithelial and cell cycle genes that are downregulated and mesenchymal marker and negative feedback regulators of the pathway that are upregulated [219]. TGFβ pathway activity can be modulated by different means as posttranslational modification of the SMAD proteins or the receptors[96, 367, 202]. Modification in conjunction with other transcription factors to coordinate the transcription induced by TGFβ. The canonical pathway is schematically shown in Fig. 3.1.

3.2 The TGFβ network - Non-SMAD pathways

Besides to the here presented protagonists in the TGFβ pathway many other signalling molecules and signalling adaptors have been identified that interact with the TGFβ receptor complex [24]. The diversity of TGFβ signalling responses is triggered by interaction of the 'core' (SMAD-dependent) with other pathways and the ability of TGFβ receptors to activate other signalling cascades [78, 238].
Non-canonical TGFβ pathways, activated by the TGF-β receptors through either phosphorylation or direct interaction, include several limbs of MAP kinase (MAPK) pathways, GTPase signalling pathways (Rho, Ras, Cdc42), phosphatases PP2A and Shc and PI3K/AKT pathway [378, 202].
Phosphorylation of cytoplasmic tyrosine residues of type I, type II serine/threonine kinase receptors and Shc is induced by TGFβ [185, 188]. This enables the binding of Grb2 [109] and the recruitment of SOS [188]. SOS then activates membrane bound RAS and in the following RAS activates RAF [359]. RAF phosphorylates MAP kinase kinases (MKKs) MEK1 and MEK2 leading to an activation of ERK1 and ERK2 [50]. ERKs regulate many transcription factors, e.g. NFκB, c-Myc, ELK1, SMAD1, SMAD2 and SMAD3 [377]. By phosphorylation of activated SMADs, ERK is able to inhibit SMAD activity[174, 175, 107]. Transcription factors regulated by TGFβ-stimulated ERK control several genes involved in cell-matrix interaction, cell motility, and endocytosis.
Another branch of MAPK pathways activated by TGFβ is the JNK and p38 MAPK signalling cascades that activates the transcription factors c-Jun and c-Fos of the AP1 complex. JNK and p38 get activated by the secondary layer, the MKKs MKK4 and MKK3/6 of the MAPK pathway. MKK4 and MKK3/6 are activated by the MKKK TAK1. MKKs get activated by MKKKs. In the JNK and p38 MAPK branch one MKKK responsible for the activation of MKK4 and MKK3/6 is TAK1. The activation of TAK1 by TGFβ is not well understood. Experiments have shown that TAK1 is activated through the catalytic activation of TRAF6 by TGFβ [182]. TRAF6 gets triggered by the TGFβ induced oligomerization of the type I/II receptor complex. Differentiation, cell survival and inflammatory responses are regulated by TAK1 through its activity on p38 and JNK by the induction of pro-survival gene transcription by components of the NFκB pathway

[4, 113, 281, 350]. It has been proposed that the MKKKs MEKK1 and MLK3 can as well activate JNK and p38 in a TGFβ dependent manner [382, 170].
SMAD7 associates with TAK1, MKK3, and p38 [85, 135, 182] modulating the activity of the JNK and p38 pathway. SMAD7 functions as an adaptor protein, underpinning non-SMAD signalling pathways activation involving TAK1, p38 and JNK MAPK pathways [240]. SMAD7 can also bind on key player of the Wnt signalling pathway, β-catenin [86]. SMAD7 is an important crosstalk mediator between canonical and non-canonical TGFβ signalling and connects TGFβ signalling with other signalling pathways.
TGFβ induces the activation PI3K/mTOR/AKT pathway in a non-SMAD signalling manner. The activation of PI3K leads to a phosphorylation of AKT by PIP3 and PDPK1[20, 309, 345, 362, 181, 86]. Type I/II TGFβR are required for the activation of the (PI3K)-AKT signalling pathway. The interaction is mediated by receptor associated p85 [374], a regulatory subunit of PI3K. TGFβ mediated EMT is also supported by the PI3K/Akt pathway by contributing to actin filament organisation and cell migration. AKT induces activation of mTor[378] and its downstream targets S6k and 4E-BP1 [181]. TGFβ induced cellular response by PI3k/AKT activation plays often a counterpart to the canonical SMAD effects [309, 55, 317]. This antagonistic effects are for example borne by interaction of SMAD3 and AKT [279, 65] there AKT prevents Type I TGFβR driven phosphorylation and subsequent cytosolic-nuclear translocation of Smad3.

4

Experimental techniques and data acquisition

The acquisition of large amounts of data to understand biological relationships and systems has become in the recent years state of the art in modern biology. In the course of this work several biological datasets that aim to illuminate different biological aspects will be examined. By combining fluorescent reporter cell lines with live-cell time-lapse microscopy data could be acquired, that allows the simultaneous quantification of signalling dynamics with high temporal and spatial resolution in thousands of individual cells. This enables the characterization of long-term dynamics of signalling in individual cells over several hours up to days.

4.1 Fluorescent reporter cell lines

To measure signalling dynamics and follow pathway activity in single cells the location and abundance of signalling proteins was monitored with high spatial and temporal resolution. In the chapters of the result section in the experiments carried different cellular fluorescent reporter systems were used.
The reporter system are shown schematic in Fig. 4.1.

Fluorescent reporter system to measure SMAD signalling dynamics

To monitor SMAD2 translocation data of a live-cell reporter system was used, that was based on the breast epithelial cell line MCF10A. MCF10A cells are an established model for TGFβ signalling [381]. The used MCF10A SMAD2-YFP reporter cell line has been described before [326, 325].
As cell reporter system a clonal cell line expressing a YFP-SMAD2 fusion protein under the control of a constitutive promoter (UbCp) and histone H2B-CFP as a nuclear marker was established. That the SMAD2-YFP fusion protein expression corresponds to approximately 50% of the endogenous SMAD2 protein was shown with western blot analysis. It was as well shown that this over expression does not disturb the SMAD2 signalling dynamics and that the expression of TGFβ target genes is unaffected [326].
To study the feedback in TGFβ signalling, a SMAD7 knock out MCF10A cell line

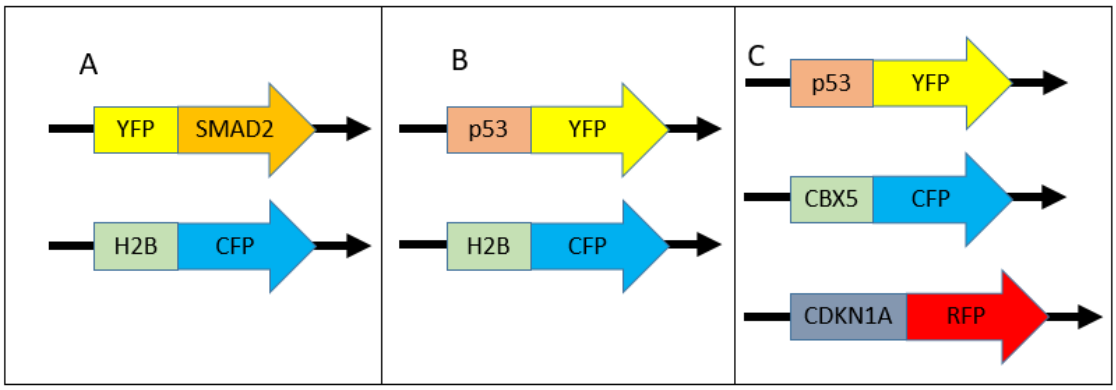

Fig. 4.1. Schematic fluorescent endogenous reporter systems

A - Fluorescent reporter system to measure SMAD signalling dynamics in individual cells. SMAD2 was fused to the yellow fluorescent protein mVenus (YFP). As nuclear marker, the constitutively expressed histone 2B (H2B) was fused to the cyan fluorescent protein mCerulean (CFP).
B - Fluorescent reporter system to measure p53 signalling dynamics in individual cells. p53 was fused to the yellow fluorescent protein mVenus (YFP). As nuclear marker, histone 2B (H2B) was fused to the cyan fluorescent protein mCerulean (CFP).
C - Fluorescent reporter system to measure simultaneous p53 and p21 signalling dynamics in individual cells. p53 was fused to the yellow fluorescent protein mVenus (YFP). CDKN1A was fused to the yellow fluorescent protein mCherry (RFP). CDKN1A encodes p21. As nuclear marker, CBX5 was fused to the cyan fluorescent protein mCerulean (CFP).

was created utilizing the CRISPR/Cas9 system. In the clonal SMAD2-YFP Cas9 reporter cell line knockout of SMAD7 affects both alleles of the gene to ensure a complete loss of the corresponding protein.
The cell lines were generated by Jette Strasen at the MDC Berlin. A detailed picture of the cell lines is given in her PhD thesis [325]. An example of images acquired with this cell line is given in Fig. 4.2.

Fluorescent reporter system to measure p53 signalling dynamics

To monitor p53 dynamics, a clonal A549 lung carcinoma cell line stably expressing p53 fused to the yellow fluorescent protein mVenus and the nuclear marker H2B fused to a cyan fluorescent protein (mCerulean) was employed. The A549 p53-Venus reporter cell line has been described before [98, 99].
In brief, it expresses a p53 cDNA fused to the mVenus coding sequence under the control of the human EF1A promoter as well as a histone H2B cDNA fused to the mCerulean coding sequence under the control of the human Ubiquitin C promoter.

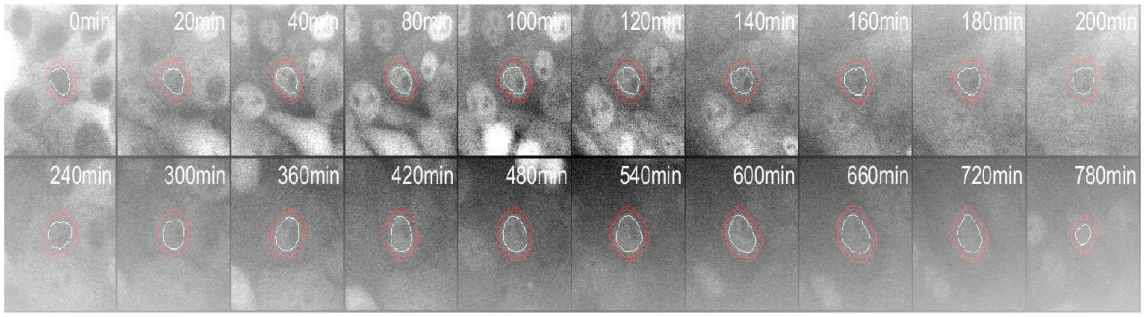

Fig. 4.2. Fluorescent reporter system to measure SMAD signalling

Live-cell time-lapse microscopy images of MCF10A cells expressing SMAD2-YFP following treatment with 100 pM TGFβ1. White circles indicate the segmented nucleus, and the estimated cytoplasmic area is represented by red annuli. Image taken from [326].

An example of images acquired with this cell line is given in Fig. 4.3.

Fluorescent reporter system to measure simultaneous p53 and p21 signalling dynamics
To simultaneously monitor p53 and p21 dynamics, a clonal MCF10A breast epithelial cell line stably expressing p53 fused to the yellow fluorescent protein mVenus, p21 fused to the red fluorescent protein mCherry and the nuclear marker CBX5 fused to a cyan fluorescent protein (mCerulean) was employed. CBX5 is a highly conserved non-histone protein, which is a member of the heterochromatin protein family. The MCF10A p53/p21-Venus reporter cell line has been described before [304, 303]. The cell lines were generated by Caibin Sheng at the TU Darmstadt. A detailed picture of the cell line is given in his PhD thesis [303].

4.2 Live-cell time-lapse microscopy

Fluorescence time-lapse microscopy opened new frontiers in biology research by accessing spatial and temporal dynamics of fluorescence signals. The fluorescence signals are identified with the proteins of interest labelled in the reporter cell lines. To study biological processes *in-vivo* high-throughput time-lapse fluorescence microscopy is used to image cells. Long-term imaging studies of cellular dynamics increase the understanding of molecular functions in developmental and cell biology [17].
With molecular florescence labelling of specific components signalling pathways dynamics were monitored. The molecular dynamics can be investigated over several hours up to days. Cells were imaged at regular intervals of several minutes. Looking at the sequence of acquired images makes time appear to be moving faster and lapsing compared to the sequence played at real time. Therefore, biological

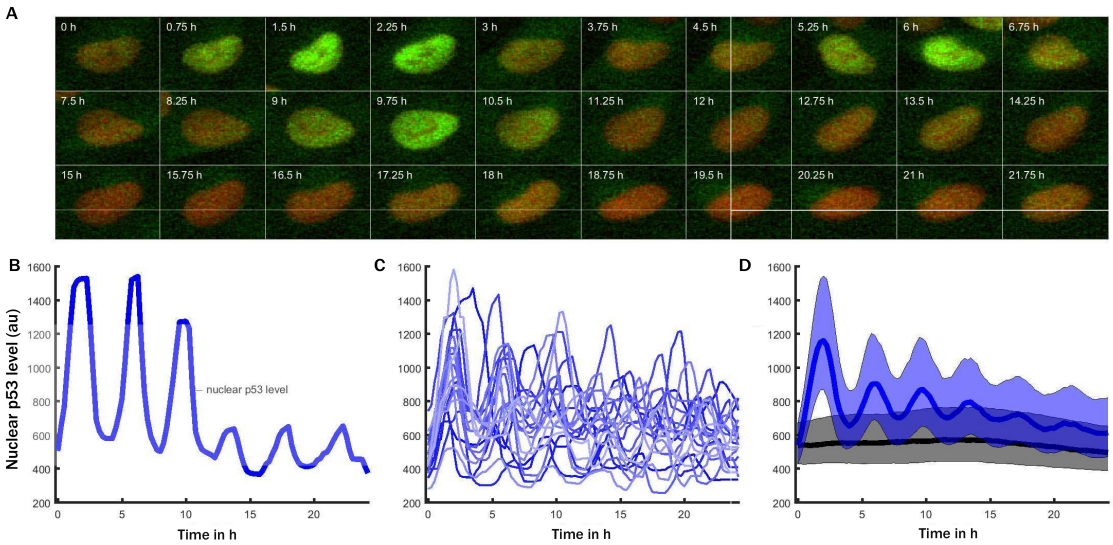

Fig. 4.3. Combination of time-lapse microscopy and fluorescence reporter system

A - Time-lapse microscopy images of an individual cell irradiated with 10Gy to demonstrate the p53 fluorescence reporter system. P53 is fused to the yellow fluorescence protein mVenus (YFP) and stably expressed under the control of the EF1A promoter. To mark the nucleus H2B is fused with mCerulean (CFP). P53 is shown in red and the H2B in green.
B - Time resolved median nuclear p53 level for the cell shown above.
C - With the combination of time-lapse microscopy and fluorescence reporter system one could quantify the response of several thousand cells simultaneously. To demonstrate the heterogeneity of the cellular response 20 randomly selected trajectories of the median nuclear p53 level after stimulation with 10 Gy ionizing radiation were plotted.
D - The p53 response on the population level (black: non-irradiated control cells; blue: cells irradiated with 10 Gy). Bold lines indicate median p53 levels, shaded areas reflect the inter-quantile range between the 0.25 and 0.75 quantile.

processes that take severalty minutes or hours can be monitored. Time-lapse microscopy generates large amounts of image data files.

4.3 From live-cell time-lapse microscopy to time series

From the raw imaging data, the signalling dynamics hundreds of individual cells were extracted using automated image analysis. The steps necessary to extract single cell time series that keep track of cellular features over time are described in the sections 6.1, 6.2 and 6.3. Time series is a common type of dynamic data that naturally arises in many different scenarios. Time series data is in general connected with characteristics of large size and high dimensionality. The dimensionality of a time series data object can be commonly associated with its sampling rate and length [5] and in the case of multivariate time-series with the set

of features measured over the course of time. The sampling rate determines the temporal resolution of features that can be detected [254].

5

Specific Aims

Cellular information processing from exterior sources involves sensing mechanisms like cell-surface receptors and signalling pathways that steer the stream of the information into the nucleus. Cellular signalling pathways are made up of a highly connected network that regulates multiple functions in a context dependent manner [159]. This enables cells to quantitatively respond sensitive and sensitive to fine-grained inputs. How these signalling pathways transmit the information through the different components of the network is roughly understood and under debate. Recently, dynamics of pathway activity could be identified as a link between external stimuli and cellular responses[274].

Imaging of fluorescent reporters in living cells to quantify activity of cellular signalling pathway has proven to be a powerful tool due to high temporal and spatial resolution on the level isogenetic individual cell [320]. Monitoring of thousands of genetically identical cells cancels out heterogeneity in pathway activity due to genetic differences.

Heterogeneity in pathway responsiveness of isogenetic cells has multiple molecular sources. This sources range from internal sources like the stage of the cell cycle or stochastic intracellular processes like the degradation of molecules to external sources that form the environment the cell faces [204, 314]. Internal stochastic processes lead to cell-to-cell variability in the concentrations of signalling proteins [95].

Different signalling pathways were investigated in the course of this work to quantify the distinct dynamics they exhibit. To study the interactions between the components of biological systems and generate an understanding how these interactions give rise to biological functions and mechanisms within cellular signalling network it is necessary to employ computational methods. To this end, a set of novel computational tools was developed that provided new insights and perspectives onto fundamental biological questions regarding the studied pathways.

5.1 Computational methods to examine cellular signalling networks

Dealing with live-cell time-lapse microscopy data requires a considerably amount of computational effort. First the raw imaging data must be processed to extract data at the single cell level. Followed by the analysis of the single cell data so that relevant biological questions can be illuminated to answer questions relevant for the data studied.
A complete framework was established that allows the extraction of single cell data from the raw imaging data and as well grants a set of methods that can be used to quantify single cell time series. A non-rigid method point cloud registration, Coherent Point Drift [249], was introduced into the field of cell tracking. The method exhibits several features that fit to the problem of tracking cells over time. The framework enables the extraction of single cell time series of features like the nuclear level of one of the labelled proteins.
Another crucial question tackled was how can similarity between the measured signalling dynamics be quantified. To this end, the concept of Dynamic Time Warping (DTW) was introduced to quantify similarity. The opportunity to quantify the similarity of dynamics grand a new view on single cell data. The flexibility in time domain made DTW the method of choice.
Based on the quantification of similarity it was studied how single cell data can be classified. Clustering identified distinct signalling dynamics that contribute to the shape average dynamic of a population of cells.
The DTW concept is additionally used to quantify similarity among multivariate time series. Hence, similarity among cells can be quantified among cells where simultaneous measurements of several pathway components are given.
To describe time series distinct features are taken into account. This features can be dynamical patterns in the time series like pulses or long lasting effects like a constant change in the level of the investigated protein. Therefore, the problem of feature detection in single cell data was tackled from two different sides. An approach is proposed that is based on DTW that is able to detect repetitive patterns in the data. The repetitive features will be found independent to disturbances in the time domain or the scaling. The other approach proposed tackles the feature detection problem from the global point of view. Using the concept of Principal Components Analysis (PCA), a supervised learning approach was established that identifies a set global patterns that can describe the underlying dynamics in a set time series.
Time-lapse microscopy data suffers to some extend by the strong variability measured while repeating the same experiments. The dynamics are reproducible but the measured florescence intensities vary due to technical noise. To reduce the technical noise a novel normalization method is proposed, that overcomes this problem.
To demonstrate the potential of the proposed methods their performance was studies in three different real world scenarios which require different analytic

strategies.

5.2 Temporal dynamics of SMAD signalling

It has been proposed that TGFβ controls cell fate by transmitting signal duration and magnitude faithfully into the nucleus in a dose dependent manner [296].
It was examined, how the cytokine TGFβ induces individual cell responses. Will the response be homogeneous or exhibit a wide range of dynamics that might be based on distinct signalling classes? Hence to some extent, how the signalling network underneath shapes the dynamic responses observed so that external signals are processed in such a way that the appropriate cellular responses is initiated.
Monitoring pathway activity with quantitative time-resolved single cell measurements enables the comparison of signalling dynamics of thousands of individual cells. To quantify the similarity of the pathway activity dynamics among different cells the Dynamic Time Warping framework, introduced in chapter (7), provided the necessary set of methods. The combinational approach of experimental and computational methods, granted the capability to decompose the population into distinct classes with qualitatively different signalling dynamics. If such signalling classes could be identified can they be linked to the phenotypic outcome induced by the TGFβ stimulation?
Most studies that quantitatively describe TGFβ induced SMAD signalling investigate this process on the cell population level by quantifying the average behaviour at fixed time points. This simplification will be overcome by combining live-cell imaging with automated image analyses, that provides the spatial and temporal resolution to generate a much more precise picture. Combination of live-cell imaging of fluorescent SMAD2 fusion proteins with automated image analyses provides the tools that permit monitoring of TGFβ long-term signalling dynamics on the single cell level. To systematically characterize TGFβ signalling dynamics on the single-cell level it is essential to integrate experimental measurements with computational approaches, that connect response dynamics either in a cell-specific or concentration-dependent manner.
The methods introduced chapter 7 was used to characterize the SMAD2 signalling dynamics in response to vary doses of TGFβ. It will be studied what causes the individual response of a cell. The findings were validated, in *silico* and experimentally, with a knock-out of SMAD7. SMAD7 is considered to be the main feedback regulators of TGFβ-induced signalling and acts at the level of TGFβ receptors [239].

5.3 Effect of temperature on p53

In relation to the scientific effort put on the understanding of p53 and its role as nature's safeguard against cancer the amount of studies that investigate the effect

of hyperthermia on the p53 response is underrepresented. This is even more surprising having in mind the beneficial effects of hyperthermia on cancer treatment. Especially, on the single cell level the dependence of p53 dynamics on the temperature has not been studied yet. This work will aim to fill this gap by studying how temperature influences p53 dynamics, particular in the context of hyperthermia. This will provide new insights and maybe give a new perspective on the biological processes that a combinational treatment of hyperthermia together with radiation and anticancer agents modulates.
Understanding the diversity and dependences of dynamics is crucial to predict potential toxicities in the body and which tumours may be sensitive to certain time scales of treatments [324]. Is the transmission of information about DNA damage to the main human tumour suppressor buffered against the environmental temperature changes or are p53 signalling dynamics temperature sensitive? To this end, p53 signalling dynamics were monitored in thousands of cells in a range from 33°C to 41°C with or without radiation induced DNA damage.
A common problem while working with time-lapse microscopy data is bringing together different experiments that were acquired on different microscopes, by different experimentalists, on different days. To overcome this problem a normalization method was developed that reduces the inter-experimental deviation by fitting a set of experiments to one reference experiment.

5.4 p53/p21 pathway

Studying only a single protein of interest at the time is a very simplified perspective on how information flows within cells. Cellular signalling pathways are composed of several network components, that change dynamically. This flexibility enables the cell to make appropriate context dependent cell fate decisions. Having an extended view at more than one of the network component, will provide a better understanding how information flows and is processed in a signalling pathway.
The view onto the p53 signalling will be extended by investigating an additional pathway component. To analyse the interaction of two network components, an endogenous reporter system will be utilized that simultaneously provides measurements of the nuclear levels of p53 and p21 within the same cells.
Mutations in the p53 gene have a prominent role in the formation of cancer cells. p53 is embedded as central hub in a molecular network that steers several biological functionalities ranging from cell division, cell death, senescence, angiogenesis, differentiation and DNA metabolism. One prominent member of the network components is p21.
Cell cycle transition from G1 to S and G2 to mitosis is regulated by cyclin-dependent kinase (CDK) family of proteins that are activated by cyclins [343] and inhibited by p21. p21 is essential among others for the cell cycle arrest at transition from G1 to the S phase by inhibition of CDK2/4 [74] and contributes to G2 arrest

by inhibition of CDK1 [48]. Also, it is required for proper cell cycle progression, contributes to cell death program, DNA repair, senescence, aging and the induction of pluripotent stem cell reprogramming [161]. The central stress signalling hub p53 directly regulates gene expression of multiple target genes by binding to p53 REs [87, 100]. The p53 target p21 promoter contains two highly conserved p53-RE. The cellular sensor p53 is induced by various extrinsic stresses, including DNA damage and oxidative stress, triggers the transcriptional upregulation of p21 [161] in dependence of the cell cycle position [323, 206, 53]. Cyclin-dependent kinase inhibitor p21 is both necessary and sufficient for the downregulation of known p53-repression targets, including survivin, CDC25C, and CDC25B in response to p53 induction [32].

It will be examined how the DDR, induced by DNA double strand breaks (DSBs) as a consequence of exposure to ionized radiation, steers the fate of cells by triggering transient cell cycle arrest in the G_1 or G_2 phase or pushes cells into senescence or apoptosis. To this end, the dynamic nuclear accumulation of p53 and its target gene p21 is studied while exposed to genotoxic stress. This relation is crucial for arresting the cells at G_1 to S phase transition through CDK2/4 [74] inhibition and likely promotes cell cycle arrest in G_2 by inhibiting CDK1 [48] as a response to genotoxic stress.

Part III

Computational Techniques

6

Time-lapse microscopy image analysis

Time-lapse microscopy enables to address questions on the single cell level *in situ*, that are otherwise unattainable, then studying a bulk population of cells. Due to heterogeneity in the behaviour observed in eukaryotic cell populations, the analysis of thousands of simultaneous monitored individual cell is necessary to study biological processes and reveal new mechanisms [10]. The simultaneous monitoring and analysis of genetically identical cells enables the study of the individual dynamic response to intrinsic and extrinsic changes that shape the composition of the cell population. To measure these dynamics within cellular signalling, live-cell time-lapse microscopy of fluorescent reporters emerged as a powerful approach [320]. Analysis and quantification of imaging data depends on the given computational methods that extract the single cell data of interest and provide the tools to study them subsequent. The processing and analysis requires a not negligible technological and computational effort.

The raw data of a time-lapse microscopy experiment is made up of several thousand images. For each time point cells studied in this work were monitor at several field of views at the emission wavelengths of the excited fluorescent proteins used for labelling separately. Exemplary, if given a 24 h experiment with a temporal resolution of 5 min and 60 different field of views (each with 1608x1608 pxl), yields 17280 images Each of the images has a size of 4.9 MB, which makes 84672 MB in total for each of the fluorescent labelled proteins separately. This amounts of data cannot require automated imaging analysis.

In Fig. 6.1, one complete field of view is shown for the nuclear marker labels (A) and for one simultaneous acquired protein of interest (B) that is predominately located in cytoplasm in this example. To uniformly stain the nuclei a stably express protein fused to a fluorophore only present in the nucleus can be used as nuclear marker e.g. histones. In general, in the course of a study several dozens of experiments are carried out. While having this large amount of data it is obvious that methods, algorithms and computational frameworks are required that automatize the repetitive and computational intense steps of the raw data processing. The aim is to extract a simplification of the experimentally acquired data that is more meaningful and easier to analyse, while reducing complexity and maintaining information. The requirements range from the identification of the

nuclei in the images, the tracking of these nuclei overtime to the extraction of time series that record a distinct cellular characteristic or feature as well as the computational and statistical analysis of the extracted data from the time-lapse imaging.
In this chapter, an overview is given on how raw microscopy data is processed within the in this thesis proposed workflow, so that the study of single cell signalling dynamics is enabled. Initially, the workflow changes the representation of raw data into a set of trajectories that characterize individual phenotypic features of a cell over time. While the workflow is presented, a new cell tracking method based on the non-rigid point set registration algorithm proposed by Myronenko and Song [248, 249] is introduced. The workflow described enables, together with the time-lapse fluorescence microscopy for the data acquisition, the opportunity to analyse and study thousands of cells with high temporal and spatial resolution and precession for up to several dozens of hours.
Raw time-lapse image series were processed using custom-written Matlab (Mathworks) framework based initially on code developed by the Alon lab [63] and the free, open-source software for quantitative analysis of biological images CellProfiler [51]. The framework provides several functionalities to analyse and manipulate image data, ranging from simple filters to object measuring tools and segmentation methods.

6.1 Processing of time-lapse microscopy images

Before working with the data the raw images acquired experimentally need to be normalized. To this end flat field correction and background subtraction is used. This pre-processing step removes errors due to technical illumination artefacts from the camera and normalize local differences of fluorescence measurement within the field of view. Flat field correction removes artefacts from microscopy images that are caused by variations in the pxl-to-pxl sensitivity within recorded focal plane. Calibration images were acquired by imaging resuspension medium with increasing exposure times, starting with 0 ms, until the signal saturated. The auto-fluorescent cell culture medium provides a strong auto-fluorescence signal at all wavelength due to the high concentration of horse serum [41]. The offset and gain for every pixel were calculated on the basis of the calibration images. This information is then used to correct the images, which were taken during time lapse experiments. The flat field correction assures that the background is removed by taking into account that the background noise is not evenly distributed over focal plane.

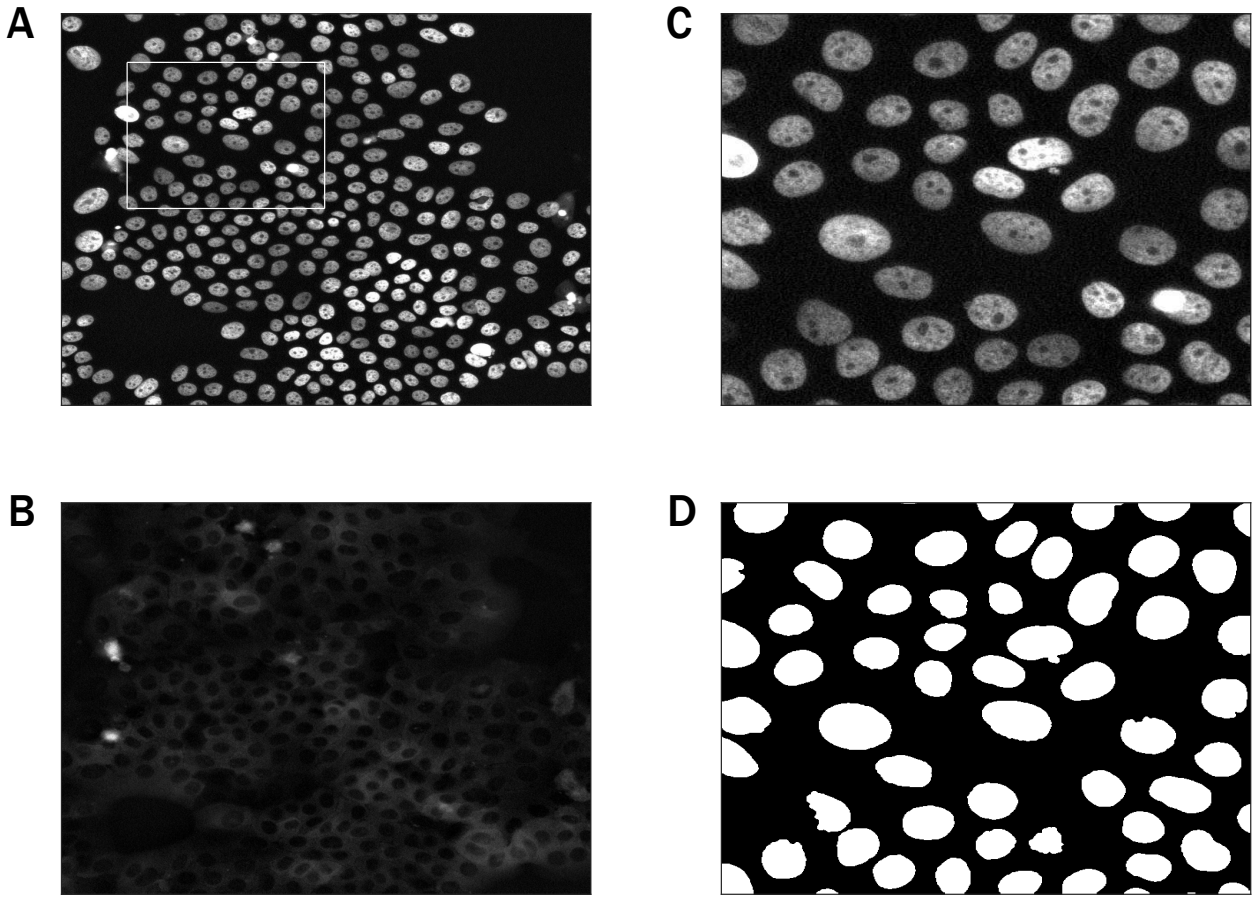

Fig. 6.1. Example of data acquired by fluorescence microscopy (SMAD reporter system introduced in 4.1)

A – Example of a complete field of view. The channel intensities acquire for the labelled nuclear marker are shown.
B - The same field of view as shown in (A). The fluorescence measures for one protein of interest are shown. In this example the protein of interest is SMAD2. SMAD2 is predominantly distributed in the cytoplasm.
C - A snippet marked in the frame shown in (A) as a white rectangle.
D - The segmentation of the snippet shown in (C). The segmentation automatically identifies the nuclei that persist in an image. The segmentation computes a partitioning of the image into disjoint regions by automatic thresholding and seeded watershed. The regions correspond to nuclei(white) or the background(black).

6.2 Segmentation of individual nuclei and the surrounding cytoplasm

To analyse time-lapse microscopy data, the initial corner stone is the identification of nuclei of the cells present in each field of view (Fig. 6.1 (A)), that are labelled

with a specific florescence marker (4.1). The segmentation problem aims at partitioning an image into disjoint regions. These regions label areas in the image that can be linked to distinct objects. The segmentation task is the essential initial step that grants the spatial single cell resolution by changing the representation of an image into a set of objects (the nuclei) easier to study.
The segmentation procedure applied to identify the individual nuclei in the images where the nuclear marker is monitored uses Otsu's method that performs automatic image thresholding [302] and seeded watershed algorithms [63, 98]. The watershed framework provides a very powerful tool for image segmentation. Seeded watersheds have shown to be useful for object detection in images of cells [270]. These methods assume the image as a topographic surface of hills, valleys, mountains and plains. The inverted pxl-intensities correspond to the altitude. Pxls with the highest gradient magnitude intensities represent the region boundaries that enclose a 'basins' (which represent the disjoint segments) with a central common local intensity minimum. In Fig. 6.1 (C, D), a snippet of a microscopy image and the corresponding segmentation calculated using automatic thresholding and seeded watershed is shown.
As shown in Fig. 6.1 (B) the protein of interest is not always located in the nucleus. Therefore, an identification of the cytoplasm around the nuclei is necessary. As were is no fluorescence marker for the cytoplasm a direct segmentation cytosol is not possible. The identified nuclei can be used to heuristically determine the cytoplasm of each individual cell. This can be achieved by generating a ring around each nucleus while taking care that the individual rings do not overlap. In Fig. 4.2 several examples of such ring around each nucleus are shown. By applying the morphological image processing operations erosion and dilation these rings can be efficiently computed. To minimize the error for the cytoplasmic measurements by accidentally covering nuclear areas, some space between the rings and the segmented nuclei is left.
Based on the segmentations of the images several features of the measured fluorescence labelled proteins can be quantified. Within the area of each segmented object, nucleus or cytoplasm, the different fluorescence intensities of the proteins correspond to their local abundance. Several statistical readouts are stored, like median, mean, variance or quantiles, for each cell. In addition, also spatial features, unrelated to the labelled protein, like the size of the nucleus or the position are extracted.

6.3 Tracking of cells in time-lapse microscopy data

Time-lapse microscopy data exhibits complex spatial and temporal information of individual cells on features as the abundance of the fluorescent labelled proteins, morphology and motility. To extract trajectories of the phenotypic features of interest from the imaging data two challenging tasks have to be solved beforehand. The previous described segmentation problem that allows the automatic identification of the objects of interest in the images and the tracking problem that enables to follow these distinct object over the course of the experiment in the series of microscopy images. The objects of interest in the imaging data are the cell nuclei in the channel that monitors the emission wavelength emitted by the excited nuclear fluorescence marker. The two problems can be treated separately or synergistic. The synergistic approach connects both steps so that the tracking improves the segmentation and vice versa iteratively.

The presented approach starts with an initial segmentation generated as described in the previous section for all images in a time-lapse series. After the segmentation step, a greedy match approach tries to find the corresponding nuclei in the temporal adjacent images primarily on their spatial properties. Greedy algorithms are simple heuristic top down algorithms, that implement take the optimal option and repeat.

The registration of the corresponding segmented nuclei in following images is determined either in a rigid or a non-rigid way. The difference between the two tracking approached are elaborate and discussed in the next section.

The non-rigid tracking method, proposed represents a novel way to solve the cell tracking problem. The non-rigid approach fit to several assumptions as observations that cells do not move independently but rather in clusters of neighbouring cells. Exemplary this is shown in Fig. 6.2 (C).

The described tracking method uses a synergistic approach where inconsistent merging or splitting of nuclei emerging in tracks due to segmentation errors are identified and used to refine the segmentation. In Fig. 6.2 (A,B) the common observed inconsistencies are shown. For example, if two tracks merge into one and at a later time point split again into two it is likely that this originates from an under-segmentation error in the initial segmentation where two neighbouring nuclei are accidently merged. If an inconsistency is detected the segmentation is redone for the images at the time points and at the location where the inconsistency was detected with altered parameters and variations of watershed algorithms [326] according to the identified problem in the tracking. If the mis-segmentation is resolved the tracks are rebuild using the same greedy match algorithm as before, else the track is removed from further analysis.

6.3.1 Rigid vs. non-rigid point cloud registration

In computer vision, image analysis and pattern recognition the point cloud registration problem is one essential task [46, 210, 209]. The problem tackled is the

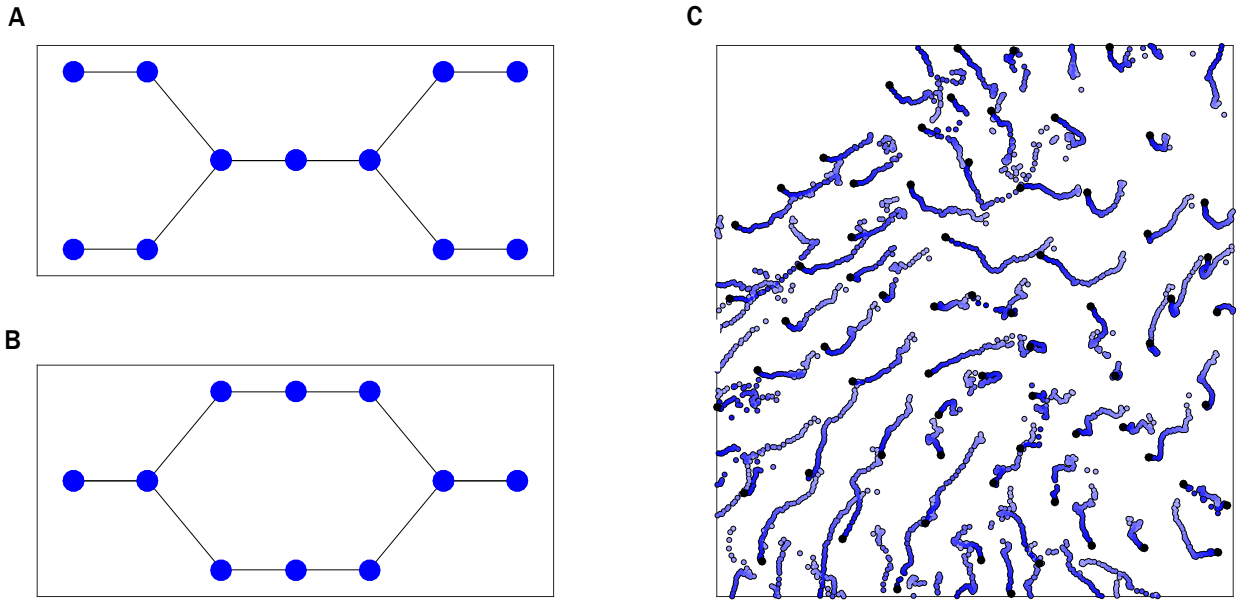

Fig. 6.2. Tracking errors may be linked to errors in the segmentation and cells do no move independently

A - Merging error, where two tracks merge into one and split again into two at later time points, indicates a potential under-segmentation.
B - Splitting error, where one track splits into two and merge again into one at later time points, indicates a potential over-segmentation.
C - Cells do not move independently. The neighbouring cells form clusters that move together due to physical restrictions. In this figure the locations of the cells are shown over fifty time points. The progress in time is colour coded. The more time has passed the lighter the blue. One can observe that neighbouring cells exhibit similar temporal and spatial motility patterns.

finding of the optimal spatial alignment of two set of points. The spatial alignment is based on estimating an underlying spatial transformation that optimally maps two point sets onto each other. Optimality is estimated with respect to set off allowed transformations. In our case these points are in particular the coordinates of the cells. The dimensionality can be extended apart from spatial information by adding additional phenotypic features like the mean nuclear fluorescence level of one of the labelled proteins or the size of the nucleus. If additional features are taken into account, they have to be scaled properly. For simplification only the spatial location of the nuclei is taken into account while introducing the approach in this section.
Given two finite point sets {P} and {Q} in a d dimensional space $\mathbb{R}^d$ of size p and q. It is reasonable and desired in many applications to find a mapping K and the corresponding transformation {P'} of {P} onto {Q}, which optimally aligns

the two point sets. Optimality is defined as the minimum of some distance, dissimilarity or in short cost function $dist(P', Q)$ between two point sets. A simple approach is find a mapping that minimizes the sum of all pairwise Euclidean distances between the points sets, $dist(P', Q) = \sum_{\forall a \in P'} \sum_{\forall b \in Q} \sqrt{\sum_{k=1}^{d} (a(k) - b(k))^2}$. In the practice more robust cost functions like an $M - estimator$ are used [242].
There are two distinct classes of registration algorithms based on the allowed transformations. The set of computational less complex rigid and the more sophisticated non-rigid algorithms. In the rigid case the transformation does not change the proportion of all the pairwise-distances between the points. This is achieved by applying only linear transformations like scaling, rotations and translations [103]. In real world applications this holds for registration problems where the points have a fix spatial relation to another and cannot move freely. A simple example would be a set of points that outlines a bone. This is necessary for many navigated surgical procedures, where precise registration of preoperative data sets with bones of the patient is an important requisite [45]. This is for the most application a much to simple assumption. In the most scenarios, non-rigid transformations are experienced in nature. Non-rigid registration algorithms exhibit higher degrees of freedom and make use of non-rigid, non-linear transformations. That makes non-rigid registration more difficult because the true underlying non-rigid transformations are in general unknown and modelling them is a challenging task [59]. The increase in complexity requires greater computational effort. Rigid transformations may be sufficient for one bone, but if the two bones are connected by a joint or the bone breaks, non-rigid registration methods are necessary.
There exists a bunch of different algorithms for the rigid and non-rigid point set registration. Non-rigid registration methods typically reckon the alignment of two point clouds as a probability density estimation problem and use Gaussian Mixture Models [248, 144]. An overview and a discussion of these methods is beyond the scope of this work.
The tracking method presented in the following section, makes use of the Coherent Point Drift method [248, 249], a state of the art method for either rigid or non-rigid point cloud registration that makes use of a probabilistic framework. Coherent Point Drift carries out several features that fit the requirements to find an optimal mapping of cell correspondences among adjacent microscopy images.

6.3.2 Novel cell tracking method based on probabilistic point set registration

Coherent Point Drift proposed by Myronenko and Song in 2007 [248, 249] is a probabilistic method for either rigid or non-rigid point cloud registration. The method is robust to noise, outliers and missing points and one of the current state-of-the-art point set registration algorithms. The algorithm makes the reasonable assumption that points from one cloud are normally distributed around points belonging to the other cloud. As a consequence, the alignment problem can

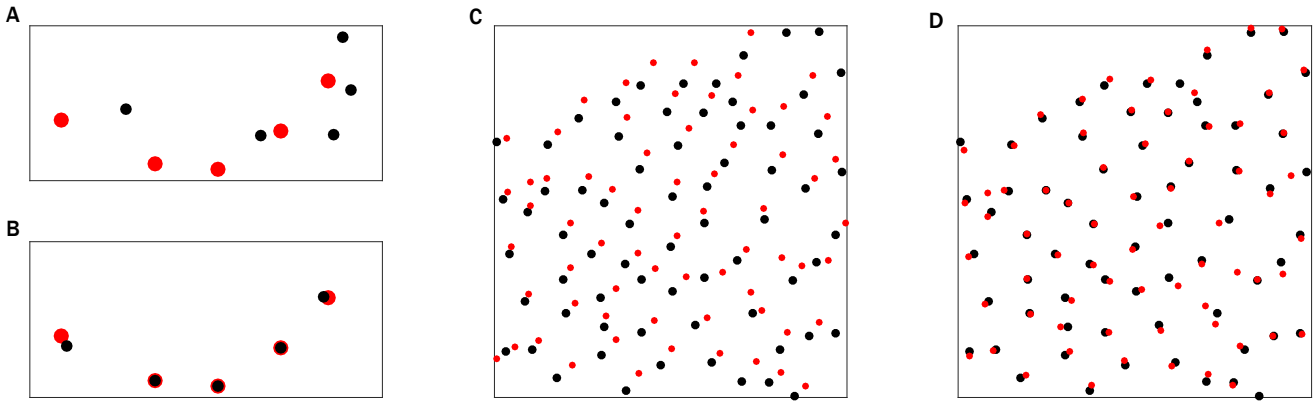

Fig. 6.3. Cell tracking benefits from the application of non-rigid point cloud registration

A - An example of two small points clouds. The black point set is a slightly rotated, shifted version of the red point cloud and as well randomly disturbed in the location of the individual points.
B - By applying Coherent Point Drift one can find a non-rigid registration of the disturbed black point cloud onto the red shown in (A). The rotation, shifting and random displacement are nearly completely removed.
C - Location of cells from cells segmented at two adjacent time points in one field of view. To mimic a technical error caused by a shift in the position of the field of view the red cell point set is also shifted north and east by 50 pxl.
D - By applying Coherent Point Drift one can find a non-rigid registration of black set of cell locations onto the red set shown in (C).

be reformulated into estimating the parameters of a mixture distribution [209].
One of the central contributions of the Coherent Point Drift algorithm is the formulation of the motion coherence constrain in the non-rigid scenario. This constrain preserves the topological structure by forcing points to move as a group. To find the point-to-point correspondence it assumes that the approximate arrangement of a point cloud is preserved.
Cells do not move independently; they are embedded in a neighbourhood that constrains their freedom of movement. If the rough topological structure is not preserved and the points are placed random it is very unlikely to find the correct correspondence. Like raisins in a growing yeast dough that are not moving independently due to physical constraints. Similar motility patterns are observed in microscopy imaging data (Fig. 6.2 (C)). This gives the impression of a coordinated motion of neighbouring cells grouped in clusters. For example, cells from highly crowded areas tend to move to less densely populated areas.
An additional advantage of Coherent Point Drift that fits our cell tracking problem is the robust procession of unequal number of objects in the two different point set clouds that have to be aligned. In general, there will not be the same number of cells in two temporal and spatial adjacent microscopy images. This is

either due to biological reasons like cell death or divisions or has non-biological origin such as cells that evade or invade into the area observed under the microscope or due to erroneous segmentation.
Coherent Point Drift treats the alignment of two point sets as a probability density estimation problem, where one point cloud represents the Gaussian Mixture Model centroids, and the other one represents the data points [248, 249]. By maximizing the likelihood, the Gaussian Mixture Model centroids are fitted to the data while preserving the topological structure of the point sets by forcing the Gaussian Mixture Model centroids to move coherently as a group. The Gaussian Mixture Model centroids were fitted iteratively by maximizing the likelihood and finding the posterior probabilities of centroids, which provide the correspondence probability. Myronenko and Song provide a Matlab implementation of the Coherent Point Drift available at *https : //sites.google.com/site/myronenko/research/cpd*. In the following, a tracking method that makes use of Coherent Point Drift and its advantages is proposed.
In Fig. 6.3 (A,B) a simple example is presented. Coherent Point Drift can compensate rotations, shifts and random deviations. In Fig. 6.3 (C,D) an additional real world example is shown. The method aligns the cell populations monitored at two different time points in the presence of dividing cells and is able to compensate technical errors like the shift in the position of the field of view.
The tracking procedure starts with the calculation of the initial segmentation of the nuclei using Otsu's automatic threshold method and seeded watershed algorithms for all images in the microscopy image stack that contains all images from one point of view. The segmentation is done on the image data from the channel that monitors the labelled marker protein in the nucleus as shown in Fig. 6.1 (A). For each time-lapse microscopy image a point cloud is generated based on the coordinates of the centre of mass of each segmented nucleus and optional as an additional dimension the mean fluorescence of the marker in the segmented area. If the mean fluorescence of the marker is used the values are scaled so that they contribute 20% of spatial information. In the next step the temporal adjacent point clouds are registered onto another using Coherent Point Drift. For each point cloud, except the first and the last, two alignments were generated that contain the correspondence between the nuclei in the current frame with either the nuclei from the previous or the next. To this end, the transformed coordinates of the nuclei from the previous or next frame onto the current frame, generated by Coherent Point Drift registration, were used to map the nuclei from the current frame onto the nuclei in the temporal adjacent frames in a heuristic manner. The heuristic greedy match algorithm used always aligns the closest points with each other so that for the all nuclei from the previous, current and next frame a correspondence in the temporally adjacent frame is calculated. A nucleus can have therefore afterwards multiple correspondences. The correspondences were filtered by restricting the moved distance of the nuclei from one frame to the other. To this end, the distances from the untransformed positions of the nuclei are used.
Once, all the correspondences are generated, the tracks were checked for incon-

sistencies (see section above). If an inconsistency is detected the segmentation is redone for the images at the time points and at the location where the inconsistency was detected with altered parameters and variations of watershed algorithms [326] according to the identified problem in the tracking (Fig. 6.2(A,B)). If the inconsistencies could be resolved by re-segmenting, the correspondences are recomputed and the tracks are rebuilt.
Finally, the tracks were filtered based on criteria, that depend on the actual requirements. For most analyses presented in the results sections, tracks that will reach from the first till the last frame and use only one daughter after a cell division are extracted. For the analysis of the relation among sister cells, after division both offspring were tracked under the assumption that the initial mother cell is present in the first frame and all tracked offspring are present in the last. As a consequence, trajectories of sister cells built upon the tracking information are identical at time points before division. In the end, the tracks are used to generate single cell trajectories of phenotypic features that are used for the further analysis of the feature of interest over time.
To speed up the computation time the framework can easily be parallelized. Commonly in a time-lapse microscopy experiment several field of views are monitored simultaneously. The data of each field of view can be processed separately without interfering with other processing of raw data.

6.3.3 Generation of single cell time series

After computing the correspondences of cells in the time-lapse image series and having extracted the related tracks the next task is to read out the time series interest. In the common scenario, one fluorescence channel where the labelled nuclear marker is measured and one or two additional channels of different wavelength where the labelled proteins of interest there monitored (4.1) are given. The segmentation of the nuclei and the temporal correspondence calculated in the tracking is used to generate the single cell trajectories of different phenotypic cellular features. In general, the extracted trajectories monitor the dynamical changes of distinct statistical characteristics of cellular features over time. The statistical characteristics are computed for all the given channels and for each tracked cell using the measured intensities within the segmented area of the nucleus. Also trajectories of spatial and geometrical characteristics like the nuclear area or the position of the nucleus are extracted from the cell tracks. This is of importance as for example that grants access to studies of the motility of a cell population.
In the addition to keeping track of nuclear measurements also cytoplasmic changes in the abundance in the proteins of interest are of importance for further analysis. To this end, the heuristic estimated non-overlapping rings around the segmented nuclei, that mimic the cytoplasm, allow the read-out of statistical characteristics of the protein of interest within the cytoplasm. Because lacking a cytosolic marker, it is not guarantee that the ring only covers cytoplasmic areas.
Features linked to motility over time can be accessed directly from the change in

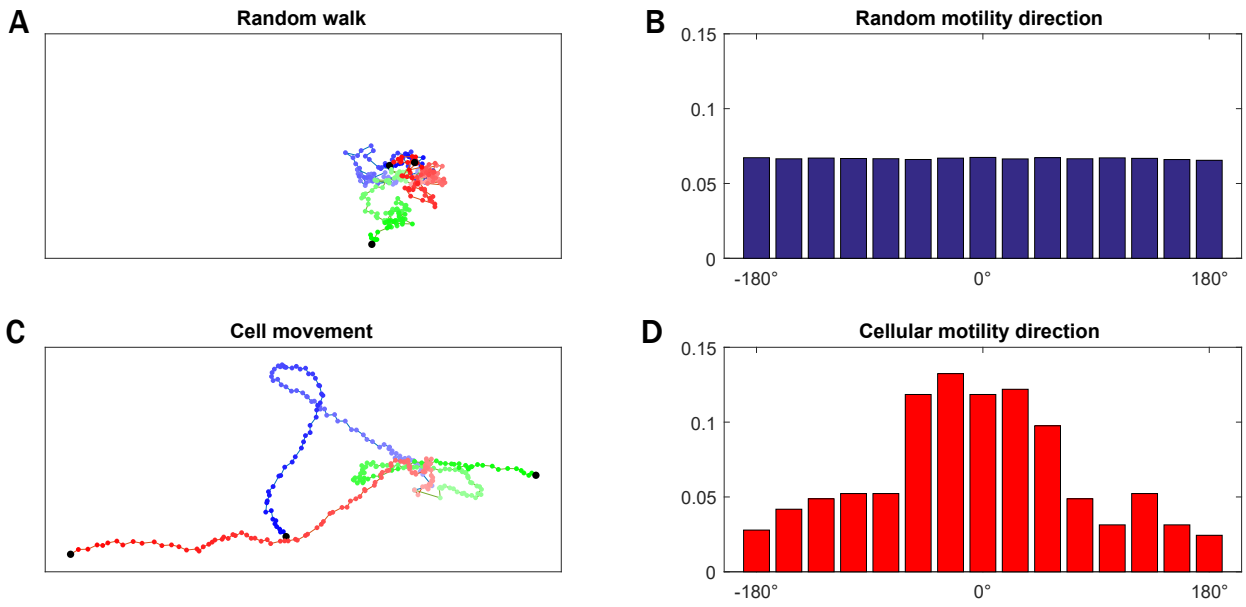

Fig. 6.4. Monitoring cellular motility over time

A -Three randomly moving objects. The average speed of objects is the same as the average speed in the plot below with the same colour code. All have the same starting coordinate.
B - Change of direction of randomly moving cells. Random moving objects change their direction uniformly distributed.
C - Motility over time of three randomly selected tracked cells. The tracks are normalized to the starting position so that all three start at the same spot.
D - Change of direction observed in tracked cells. Cells move far beyond from random. In general, they keep their direction $\pm 30°$.

the position of a cell in adjacent time frames. The positions of a cell over time encodes several features. In Fig. 6.4 the motility between randomly moving objects and cells monitored is compared. Roughly, locomotion can be largely summarized by the distribution of speed and direction change. Especially the distribution of direction change will give valuable information on how ordered cells move.
The cellular density was estimated with respect to all segmented cells in each frame of the time series of microscopy images. Around the position of each segmented cell all neighbouring cells in a radius of 640 μm were identified. To get an estimate for the local cell density each of the neighbours were weighted according to the distance to the analysed cell and these weights were summed up (Fig. 6.5). Cells that are closer then 320 μm to the rim of microscopy image are excluded because they will disturb the density measure. In Fig. 6.5 the density estimate is shown for one set of identified cells in a microscopy image.

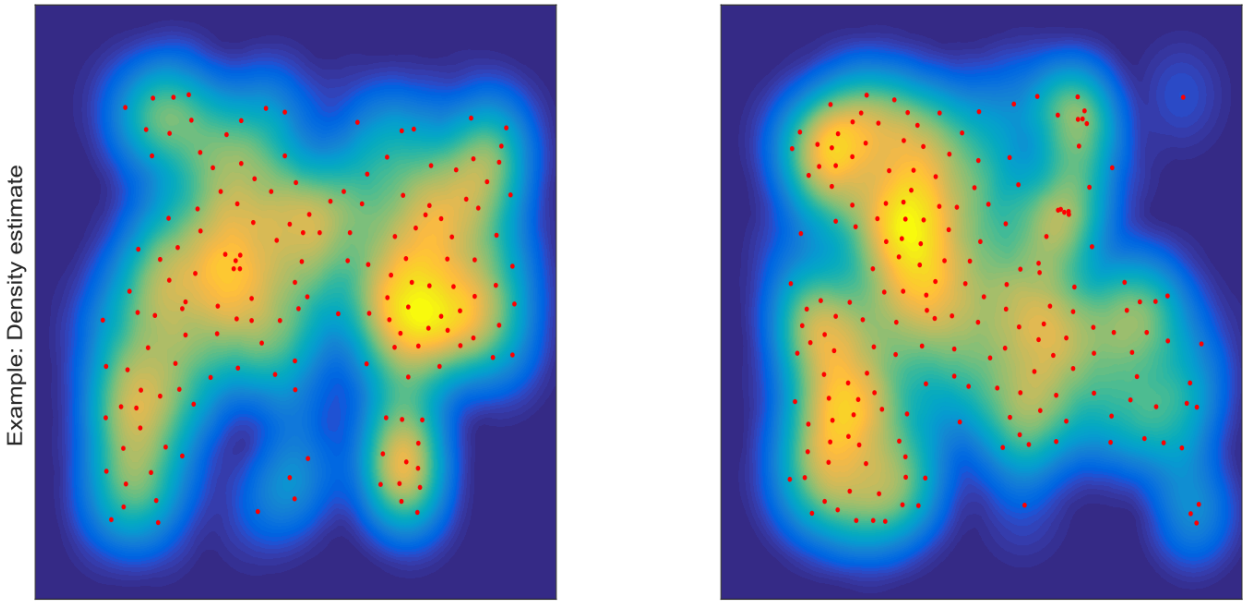

Fig. 6.5. Example of the density estimate for two field of views at a fixed time point.

Measuring local cell density by live-cell imaging. Local cell density is measured in a 200px radius around each cell for each time point by applying a bell shaped kernel to obtain a weighted sum of all neighbouring cells. The resulting density scores are demonstrated using a randomly chosen time point. Red circles indicate the centroid of cells identified. Warmer colours indicate higher density scores.

7

Quantification of similarity among single cell signalling dynamics

Based on the tracking of individual cells, dynamic cellular features (e. g. intensity, size, speed) can be successively measured over a time interval in an unbiased manner. These calculated time series enable deep insights into cell-to-cell variability that otherwise cannot be investigated by other experimental approaches. The temporally and spatially resolved data can be a potential cornerstone to illuminate and characterize dynamic changes of biological processes in different genetically identical individuals and populations across different conditions.
To reveal how dynamic changes in biological processes depend on, for example, conditional perturbations or individual phenotypic differences, time series need to be compared in a precise and robust, quantitative and qualitative way. Observations that unfold over time represent valuable information subject to analysis, classification, indexing, prediction, or interpretation of biological processes [163, 126, 196, 105]. One possible way to answer, if cell-to-cell variability is related or grounded on distinct dynamics, is by applying clustering techniques. To this end, measurements that estimate pairwise distance or dissimilarity among two trajectories, quantifying the degree to which a given time series resembles another are required. Even, if there is an increasing interest in time series clustering [196] in various application domains, little effort is carried out on transferring existing methods into the context of single cell data [146]. In general, single cell data is classified and analysed mainly on the level of extracted features and not on basis of the complete measured time series [235, 288].
To apply techniques like clustering on the level of single cell time series a pairwise dissimilarity or distance measure that incorporates the whole trajectories must be employed. Dissimilarity and distance measures differ essentially in their characteristics. Dissimilarity measures are less restricted. Distance measures have to fulfil the triangle inequality, while this is not true for dissimilarity measurements. The triangle inequality states: given three point A, B and C, the distance from A to C cannot be longer than the distance from A to B plus the distance from B to C. Four commonly used methods in the field of single cell data analysis are the Euclidean distance, Fourier coefficients, auto-regressive models and correlation [301]. They suffer from several weaknesses and limitations that make them an inappropriate choice as measurement for the similarity among single cell dy-

namics of biological processes. The weaknesses will be discussed in the following. To overcome the limitations an alternative methodology is introduced, to quantify the similarity between single cell trajectories independent of non-linear variations in the time domain.
In this chapter, the concept of Dynamic Time Warping (DTW) [290] will be introduced into the field of single cell time series analysis. DTW will be used to quantify dynamical differences among single cell trajectories. It has been shown that DTW is found to be superior to other methods estimating similarity among time series in various scientific domains [353].
DTW has several advantages compared to other methods like the Euclidean distance, especially because it is able to compensate temporal shifts, variations and deviations in the observed cellular dynamics. After the introduction of DTW a novel constrained version of DTW approach (cDTW) will be proposed is later used for the comparison of a huge dataset of single cell time series based on the dynamics observed. The modifications of the classic DTW improves its applicability to biological systems, by introducing an elastic constrain that mimics the elasticity of a spring. cDTW constrains the flexibility in the time domain by the introduction of an additional parameter. To illustrate the performance of cDTW and analyse its precision and robustness, the method was tested in different artificial scenarios.
Based on cDTW, in the course of this work a clustering framework for single cell trajectories is established that enables new mechanistic insights into the underlying individual molecular dynamical patterns that shape the cell composition of the population in the given data set.

7.1 Measuring cell-to-cell distance among trajectories

The fundamental part in many analysis of large time series datasets is providing a quantification of the similarity between two-time series. The simplest way to estimate the dissimilarity between two different time series is to use any L_n norm. The l_1 and the l_2 vector norms, also known as Manhattan and Euclidean distance, are used in general to measure distances between objects and are the appropriate choice in many scenarios. The Manhattan distance between two points is given as the sum of the absolute differences of their Cartesian coordinates and the Euclidean distance is length of the direct connection between two points in Euclidean space. All l_p vector norms are metrics. Hence, they satisfy 4 properties: non-negativity, identity of indiscernibles, symmetry and triangle inequality [52]. The Pythagorean formula for the Euclidean distance for two points p and q is in l dimensions is given by $dist(p,q) = (\sum_{i=1}^{l}(p(i) - q(i))^n)^{1/n}$ with $n = 2$.
In the context of two different time series t and s of length n the Euclidean distance is the area between the two trajectories.
It is obvious that the temporal order of the measurements in the trajectories does

not play any role. Therefore, the order of the measurements can be shuffled and the result remains the same. The same is also true for correlation, Fourier coefficients and auto-regressive.
So, the employment of such a method to compare time series is accompanied by losing the 'essential' - the temporal information that structures and shapes the dynamics present in the data. For example, that means that given two periodic signals of the same frequency and amplitude already a small temporal shift changes the result of the Euclidean distance drastically. Euclidean distance does not measure more than the sum the total difference in the amplitude of two trajectories no matter of their temporal order. Also, any l_n norm is sensitive to noise, scale and time shifts [294]. Methods like correlation suffer from the same weaknesses but are more robust to scaling. From this point of view, a method that takes into account the temporal aspect while comparing time series would be desirable.
Several shape-based similarity measures have been proposed in the past decades in order to overcome the limitations of the vector norms. Methods available are for example: global alignment kernel distance [68], Soft DTW [69] or shape-based distance used in k-Shape clustering [263] to name a few. Reviewing them all was beyond the scope of this work as the focus was on establishing of the Dynamic Time Warping framework into the field of single cell analysis.

7.2 Application of Dynamic Time Warping on single cell time series

DTW is a class of dynamic programming based algorithms for comparing time series with each other and it is the standard method to address temporal alignment problems [290, 35, 167]. DTW overcomes the known weakness of the Euclidean distance and other rigid methods like correlation, especially due to the sensitivity to distortion in time axis and by respecting the temporal order of the measurements. The method was originally developed for application in the field of speech recognition [277]. The DTW method is based on the dynamic programming approach proposed by Bellman [30] in the middle of the last century. This approach has proven to be of great benefit to several fields of application. Just to name two of the most popular application in life science, the Needleman-Wunsch algorithm [252] for comparing protein or nucleotide sequences and the Dijkstra's algorithm [81] to solve the shortest path problem are based on dynamic programming. The idea is to solve a complex problem by breaking it apart into simpler sub-problems. By solving the simple sub-problems in a recursive manner one gets access to the solution of the initial complex problem.
DTW calculates an optimal alignment between two time-dependent sequences $X := (x_1, x_2, \dots x_n)$ of length $N \in \mathbb{N}$ and $Y := (y_1, y_2, \dots y_m)$ of length $M \in \mathbb{N}$, while stretching or compressing them locally in the time domain to compensate time deformations and different speeds [241]. A pairwise alignment of two sequences tries to find the most likely correspondence of the measurements given

in X and Y. The correspondence reflects the dynamic relation of the two trajectories to another. To this end, the trajectories are warped to match each other. DTW takes into account that distinct patterns and features within trajectories may differ in scaling, frequency or length and this affects their shape in the time domain. The non-linear mapping between the two trajectories is found by warping the time axes.
In the following, the notation from Mueller will be used [241]. Let F denote a feature space so that $x_n, y_m \in F$ for $n \in [1:N]$ and $m \in [1:M]$ To compare instances within the feature space F a distance or cost function is needed, which is defined as a function $c : F \times F \to \mathbb{R}_{\geq 0}$. In general, $c(x, y)$ is small if the two feature characteristics are similar and increases with fading similarity. A set of such functions, as discussed in the previous section, are the L_n vector norms like the Euclidean distance. By computing each pair of elements of the sequences X and Y, one obtains the cost matrix $C \in \mathbb{R}^{N \times M}$ defined by $C(n, m) := c(x_n, y_m)$. The alignment desired runs in a warping path along a "valley" of low cost in the cost matrix C from the lower right to the upper left corner. Such a warping path is a sequence $p = (p_1, \ldots, p_l)$ with $p_l = (n_l, m_l) \in [1:N] \times [1:M]$ for $l \in [1:L]$ that can be defined using 3 simple rules:

(i) *Boundary condition:* $p_1 = (1, 1)$ and $p_l = (N, M)$ so that the first elements and last element of X and Y are aligned to each other.
(ii) *Monotonicity condition:* $n_1 \leq n_2 \leq \ldots n_L$ and $m_1 \leq m_2 \leq \ldots m_L$ guarantees that the order of the elements in of X and Y is conserved in the alignment.
(iii) *Step size condition:* $p_{l+1} - p_l \in \{(1, 0), (0, 1), (1, 1)\}$ for $l \in [1 : L - 1]$ works as a continuity condition that prevents omitting of elements of X or Y and repeating elements in the alignment.

The total cost of a warping path p between the two sequences X and Y is defined as the sum of all cost along the path: $c_p(X, Y) := \sum_{l=1}^{L} c(x_{p(l,1)}, y_{p(l,2)})$. A specific warping path is of special interested, the optimal warping path p^* between X and Y with the lowest cost among all possible warping paths. DTW is the method, that based on dynamic programming can compute the cost of the optimal path p^*, $DTW(X, Y) := c_{p^*}(X, Y)$. If needed the warping path p^* itself is easily accessible. The optimal path is not necessary unique, there may be several paths of equal cost.
The cost of optimal warping path p^* can be interpreted as an estimate of the similarity of two different time series. All pairwise costs of the optimal warping paths calculated by DTW for a set of trajectories does not exhibit a metric. The method satisfies only 3 of the properties that frame a metric: non-negativity, identity of

indiscernibles and symmetry. The 4th feature of an metric d, the triangle inequality assuming that a metric fulfils $d(x,z) \leq d(x,z) + d(y,z)$, cannot be guaranteed by DTW.
As stated before the algorithm to compute the optimal warping path is based on dynamic programming. The heart of dynamic programming is the idea to disassemble the big problem into simple sub problems. Let D be a $N \times M$ matrix defined by $D(n,m) := DTW(X(1:n), Y(1:m))$ for trajectories $X(1:n) := (x_1, x_2, \ldots x_n)$ and $Y(1:m) := (y_1, y_2, \ldots y_m)$. From the step size condition it is known that $p_{l+1} - p_l \in \{(1,0),(0,1),(1,1)\}$ for $l \in [1:L-1]$. This indicates that the entries of the matrix D can be calculated recursively by $D(n,m) = min\{D(n-1,m), D(n,m-1), D(n-1,m-1)\} + c(x_n, y_m)$.

Pseudo code: constrained DTW using Sakoe-Chiba band

double cDTW(X:array[0 ... n-1], Y:array[0 ... n-1], malus:double, bandWidth:int)

// **Init**

D = array[0 ... n, 0 ... n] % init zero matrix of size N x N

D[1:n,0] = Inf

D[0,1:n] = Inf

B = array[0 ... n, 0 ... n] % init zero matrix of size N x N

B[0:n,0] = Inf

B[0,1:n] = Inf

C = $c(X,Y)$ % distance between each pair of observations in X and Y

sC = quantile(C(:) ,malus)

// **Recurrence**

for i = 0:n-1

for j = max(i-bandWidth,1):min(i+bandwidth,n)

h = $\begin{pmatrix} D[i,j] \\ D[i+1,j] + max(0, sC * B(i+1,j)) \\ D[i,j+1] - min(0, sC * B(i,j+1)) \end{pmatrix}$

D[i+1,j+1] = min(h)+C(i,j)

idx = arg min (h)

B[i+1,j+1] = $\begin{cases} 0 & idx == 0 \\ max(B[i+1,j]+1, 1) & idx == 1 \\ min(B[i,j+1]-1, 1) & \text{else} \end{cases}$

end

return $D(n,n)$

end

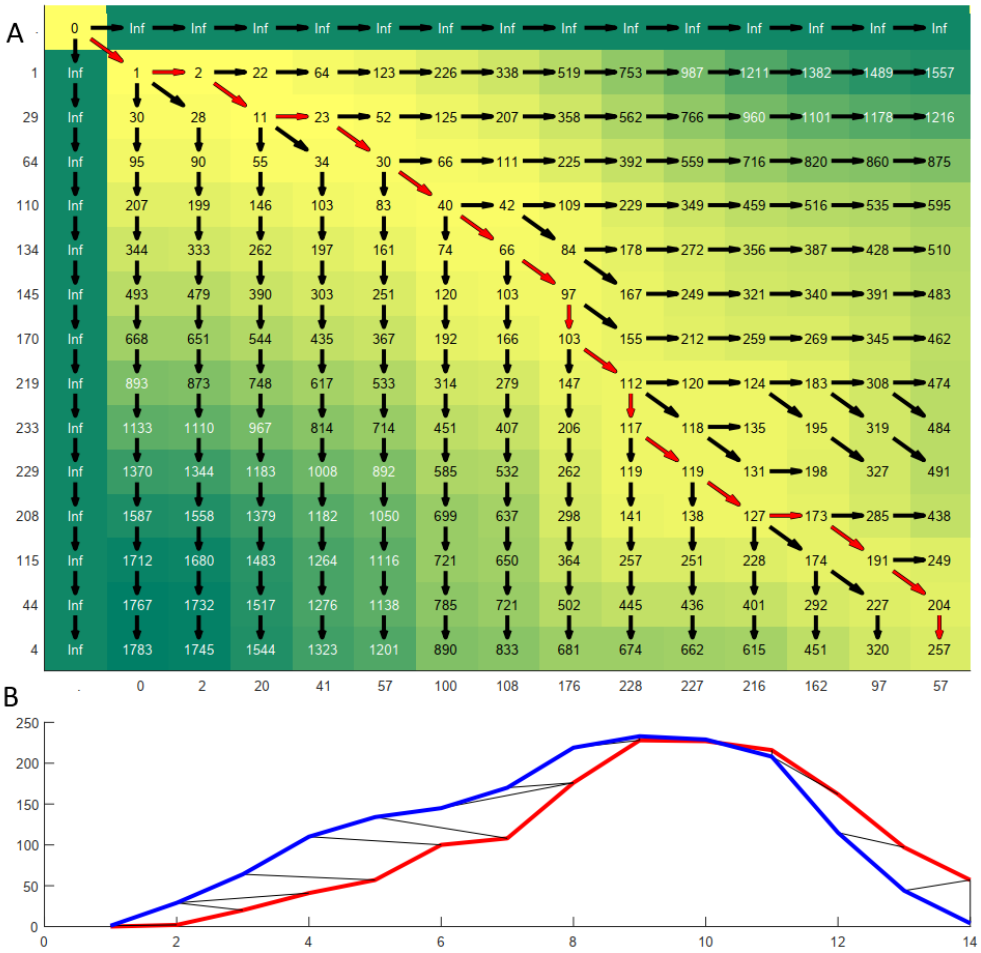

Fig. 7.1. Recursive filling of the dynamic programming matrix by Dynamic Time Warping

A - DTW calculates an optimal match between two sequences with certain constrains by calculating the distance between each point of the two sequences and finding the optimal match between them. The two sequences are shown left and at the bottom of the matrix. After initialisation by assigning infinity to the top row and leftmost column and setting the top left entry of the matrix to zero the matrix is filled recursive. The arrows indicate if a stretch or directs match was obtained. The DTW similarity score can be found in the lower right corner of the dynamic programming matrix. Backtracking (red arrows) all the way back to the top upper corner allows to extract the alignment between the two sequences. These backtrack corresponds to the alignment.
B - The trajectories used in the illustration above and corresponding alignment obtained by DTW. The mapping calculated in the alignment is shown in black.

At each step, the algorithm finds the direction in which the cost increases the least under the chosen constraints. In the end, one can read off the dissimilarity measurement, the cost of the optimal path p^*, between the two trajectories X and Y from the matrix entry $D(n, m)$ The computational effort of the central

recursion of the DTW algorithm has a quadratic complexity of $O(NM)$.
If the optimal path itself is of interest, not just its cost, an additional step to the computation is necessary. There are two distinct ways to estimate the optimal alignment path. Either, the $arg\ min\{D(n-1,m), D(n,m-1), D(n-1,m-1)\}$ are stored during the execution of the recursion in an additional matrix B and tracing back from $B(n,m)$ returns the optimal alignment or the optimal alignment is extracted by starting at $D(N,M)$ by computing $p_{l-1} := arg\ min$ $\{D(n-1,m-1), D(n-1,m), D(n,m-1)\}$.
One drawback of the standard DTW is that it does not constrain directly the flexibility of the alignment. That means for example, that a pattern that occurs in both time series can be aligned onto another no matter how much time has passed between their occurrence or that a pattern that appears in both trajectories differently stretched is aligned onto another no matter how the position or the length of the patterns differs. While working with real world data this is an inappropriate feature that contradicts the biological reasoning. The advantages of an elastic constrain for DTW is that, due to the penalization of longer stretching, it avoids inappropriate time shifts in the alignment. So that the method avoids to align substantial different temporal occurrences of patterns. The pseudo code of the adjusted DTW method is given in the box above. The parameter *malus* controls the 'stiffness' of the alignment and is explained later in this section.
In Fig. 7.1 (A) an example is given how the dynamic programming matrix computed by DTW is filled for a pair of short trajectories (B). After the initialisation of the left column and the top row the matrix is filled starting from the upper left corner. The entry $D(i,j)$ is introducing a stretch if it originates from one of the entries $D(i-1,j)$ $D(i,j-1)$ or a match if it originates from the entry $D(i-1,j-1)$ into the alignment in recursive fashion. From the rightmost corner of the matrix one can read of the similarity score. Backtracking from the rightmost corner yields the optimal warping path.

In Fig. 7.2 the advance made by the use of DTW is shown in comparison to the Euclidean distance. In this example DTW outperforms Euclidean distance. While the Euclidean distance cannot compensate temporal deviations, one cannot prevent that dynamics that are obviously similar are in the framework of the Euclidean distance less similar than completely unrelated trajectories. This limitation of the Euclidean distance will make it inappropriate for the detailed analysis of dynamics and promotes the employment of the DTW framework.
To set the parameter *malus* that controls the 'stiffness' of the cDTW approach two different experiments with artificial time series were carried out. In both experiments, it was studied how the parameter *malus* controls the behaviour of cDTW with respect to deviations in the time domain. Initially, a time series was generated were a pulse of length 4 h is embedded at the begin with a equidistant temporal resolution of 5 min (Fig. 7.3). To determine a valid value the pulse was either shifted (7.3)(A)) or stretched (7.3)(A)) to generate a set of artificial time series.
Using a *malus* of zero corresponds to classic DTW scenario, even temporally dis-

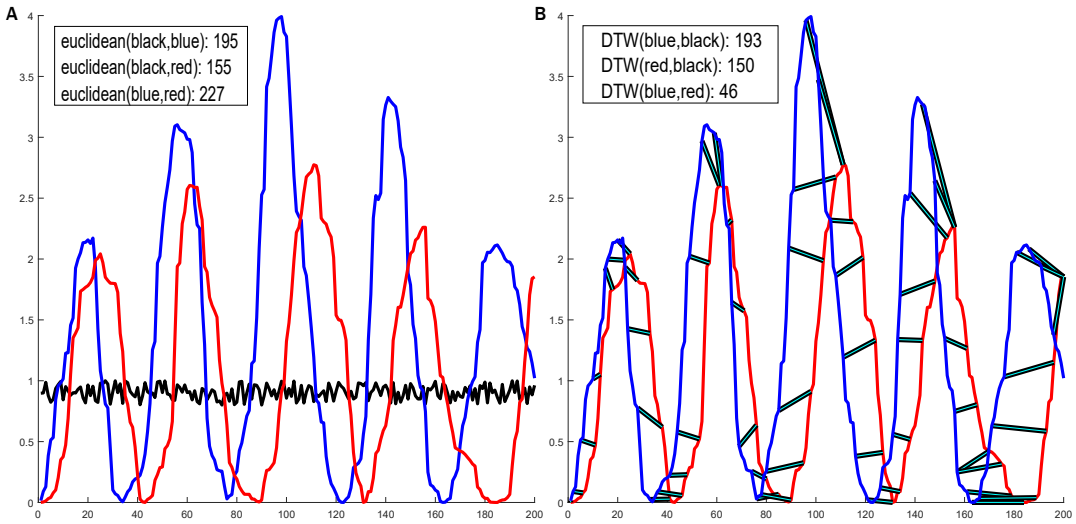

Fig. 7.2. Measuring dissimilarity between time series with the temporal rigid Euclidean distance vs. non-rigid DTW

A - Example of three artificial trajectories and the Euclidean distances between them. The red trajectory is a warped and disturbed progeny of the blue trajectory. Due to the stiffness in the temporal domain the distance of the black trajectory to the red and the blue is smaller than the distance between the red and the blue even if the red and the blue share a similar temporal dynamic.
B - The DTW dissimilarity score of the same three artificial trajectories and the alignment of the red and the blue trajectory. DTW can compensate shifts and warps. That enables DTW to compare trajectories with respect to their dynamical patterns.

tant or highly distorted features were aligned, leading to similar low cDTW scores as for only slightly changed time series. The ability of aligning features depends on the value of *malus*. If *malus* is set to 0.25, patterns shifted by more than 6h were no longer aligned and the cDTW scores was close to the simple Euclidean distance. In the second experiment were stretching is investigated this time frame is slightly longer. The shift/stretch induced in the time series had a direct impact on the cDTW score. cDTW scores increased with increasing shift/stretch. The analysis of how the parameter *malus* influences the capability of cDTW to compensate temporal deviations yields a reasoning for the choice for the value of *malus* used in chapter 11. In the following applications of cDTW the *malus* was set to 0.25 [326].
Next, the question of how robust the cDTW method is to experimental noise will be answered. For this end, a noise titration on a real data set was employed and the impact of the noise onto the cDTW measurements was analysed (Fig. 7.4 (A)). It could be shown that cDTW is very robust to noise. Even at high signal-to-noise ratios (SNR) the cDTW scores do not differ substantially from the undisturbed result.

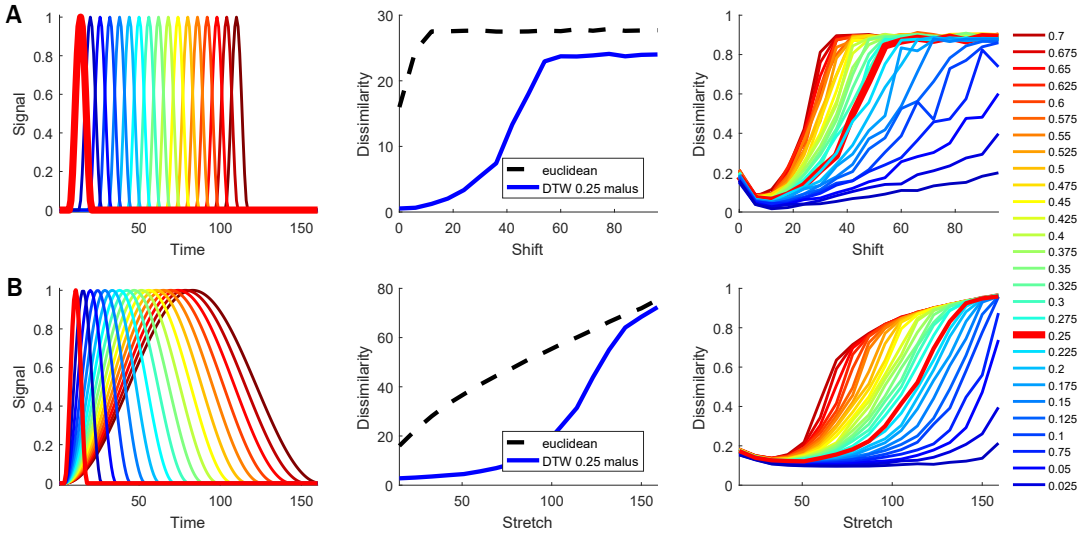

Fig. 7.3. cDTW can adjust the strength of the compensation of disturbances in the time domain.

A - The parameter *malus* gives control to which extend cDTW can compensate temporal shifts. An artificial trajectory with one single pulse was generated and this pulse was shifted in his timing (left panel). The impact of this shift was calculated for the Euclidean (middle panel, dashed line) and the cDTW measure (middle panel, solid line, $malus = .25$) With increasing shifts the cDTW score grows gradual until it reaches values similar to the Euclidean distance at around 50 time points (this corresponds to around 4 h if the temporal resolution is 5 minutes).
B - The parameter *malus* gives control to which extend cDTW can compensate temporal stretches. An artificial trajectory with one single pulse was generated and this pulse was stretched in time (left panel). The impact of this stretching was calculated for the Euclidean (middle panel, dashed line) and the cDTW measure (middle panel, solid line, $malus = 0.25$) With larger stretches the Euclidean distance grows gradually in contrast to shifts without reaching a maximum. The DTW score grows much slower for a certain lag time. After the lag the value reached values similar to the Euclidean distance fast. This lag phase is for a *malus* of 0.25 approximately double the time observed for shifts. The parameter *malus* allows to adjust the stiffness of cDTW between the performance of the standard DTW method ($malus = 0$) and the Euclidean distance ($malus = 1$).

SNR is a measure that compares the level of an original signal to the level of background noise. For SNR the definition $SNR = mean(Signal^2)/\sigma_N^2$ with σ_N^2 the variance of the background noise was used. In section 9.2.3 a detailed view on the SNR range one can expect to face in real world imaging data is given.
DTW has a quadratic running time of $O(NM)$. If each pairwise DTW for set a of K time series is computed, a total running time of $O(K^2NM)$ is necessary. This demands a huge computational effort for big data sets. One simple way to reduce the effort is to constrain the DTW calculation. A common DTW variant

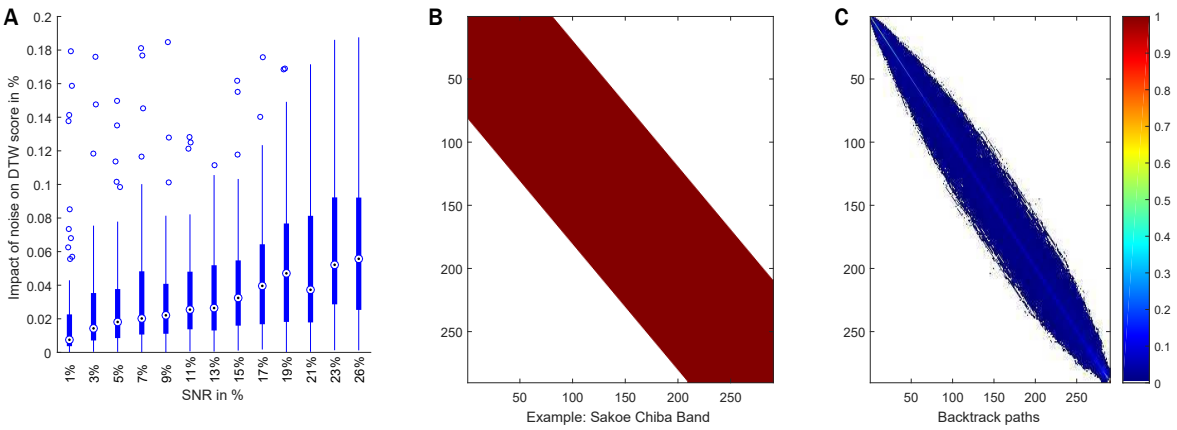

Fig. 7.4. cDTW is robust to noise and its computation can be speeded up using a simple band constrain.

A - cDTW scores are robust to noise. From a set of single cell measurement two cells were randomly picked and white noise was added according to the indicated SNR. cDTW scores between undisturbed and disturbed trajectories were calculated and normalized to the cDTW score of the two undisturbed trajectories. This was done for 2500 distinct pairs of trajectories for each SNR level.
B - Simple example of a Sakoe-Chiba band to constrain the computational effort of DTW. The entries of the dynamic programming matrix are only calculated within the red area.
C - Using backtracking and the same SMAD dataset as in (A) it was measured which entries of the dynamic programming matrix are used for the alignment based on the optimal warping path. The observations assure what the use of the Sakoe-Chiba band has no real effect on the cDTW score.

that constrains the admissible warping paths uses a Sakoe-Chiba band[290] (Fig. 7.4 (B)). The Sakoe-Chiba band restricts how much the warping path can deviate from the main diagonal of the matrix D. This global constrain added to DTW does not only reduce the computational effort. It also gives control to prevent unreasonable alignments by preventing extreme temporal warping. To find a biological reasonable width of the Sakoe-Chiba band, a real data set was used, to measure how much deviation from the main diagonal, one can expect from the warping paths (Fig. 7.4 (C)). In general, a “time warp” of less than 75 minutes (mean: 46 min) was observed. To avoid bias because only one data set was used a band width of around 4 h was used. For trajectories of length 24 h this reduces the computational effort to 1/6.

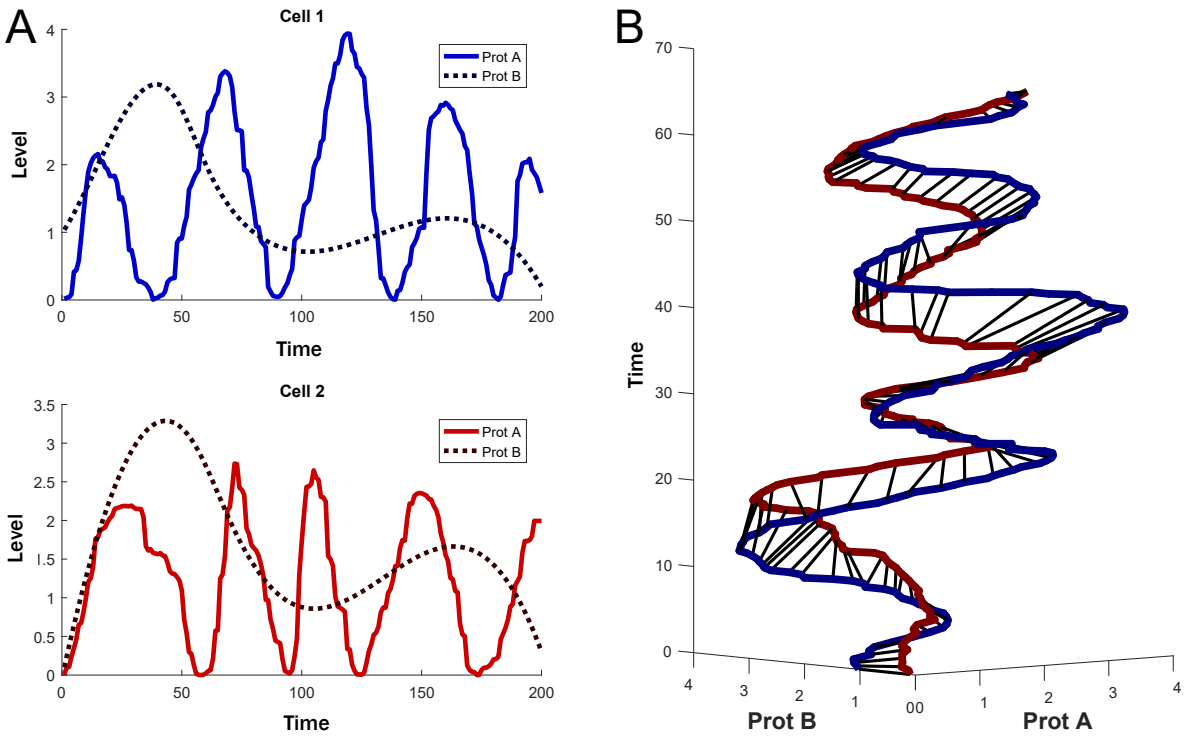

Fig. 7.5. DTW can easily be extended to multidimensional measures.

A - Two artificial multidimensional measures of two 'proteins' (solid and dashed line) over time and two 'cells' (blue and red).
B - Alignment of the two 'cells' shown on the right.

7.3 Extensions of Dynamic Time Warping for multivariate time series

Single cell measurements are not restricted to only one single features single feature as multivariate time series can have several time-dependent variables. It is likely that the variables depend not only on their past values but also have some dependencies among each other. The single cell time series data sets are composed of several measurements over time. For each cell tracked there are for each time point a collection of measurements monitoring several phenotypic characteristics. Specifically, if more than one protein of interest was labelled, it would be desirable to quantify the similarity between cellular dynamics while looking at the labelled proteins simultaneously.

The DTW framework can be uncomplicated extended for the application in the multivariate time series scenario by adjusting the cost function $c(x, y)$ accordingly. The cost function used so far in the DTW was the Euclidean distance, $c(x_i, y_j) = ((x_i - y_j)^2)^{1/2}$. There is no need to change the cost function. It must only be incorporate that a multivariate measurement d is given, so that

$X := (x_{1,1}, \ldots, x_{1,d}, \ldots, x_{n,1}, \ldots, x_{n,d})$ and $Y := (y_{1,1}, \ldots, y_{1,d}, \ldots, y_{m,1}, \ldots, y_{m,d})$. Therefore, the Pythagorean formula of the Euclidean cost function over all variables $c(x_i, y_j) = c(x_{i,:}, y_{j,:}) = (\sum_{k=1}^{d}(x_{i,k} - y_{j,k})^2)^{1/2}$ with $i \in 1, .., n$ and $j \in 1, .., m$ is applied. In Fig. 7.5 an example for a set of multivariate artificial trajectories is given.

If multivariate DTW is applied the different variables need to be scaled using appropriate normalization methods such as min-max normalization to weight each variable equally.

8

Using similarity measures to classify single cell trajectories

Cluster analysis methods are widely used exploratory data mining techniques that can provide starting points for other analyses. Cluster analysis is a technique based on unsupervised learning which characterizes objects by their properties and features and classifies them into groups. These clusters and their characteristics are created by the clustering analysis itself. The whole set of the data objects is divided into multiple clusters, where the assignment relies on their distance or similarity. The initial essence of clustering, was describing spatial characteristics of the objects of concern. However, from social to biological studies, researcher's interests in temporal characteristics often rival their interests in spatial features [146].
Clustering relies on functions that measure the pairwise distance/dissimilarity among data objects. To cluster single cell time series a shape-based time series clustering based on DTW dissimilarity measures was employed. The in the following proposed single cell time series clustering approach identifies different groups of distinct behaviours that exist within the given single cell data based on the previously introduced dissimilarity measure. The framework assures that each single cell trajectory is allocated to the appropriate cluster. Clustering can identify patterns in the data that cannot investigated or identified manually.

8.1 Clustering methodology

Various clustering algorithms have been proposed that differ on how they handle the data and how the groups are created [294]. The prevalent clustering methods are connectivity models (e.g. hierarchical clustering) and centroid models (e.g. k-means) that will be discussed in the next sections. Less common are based on distribution models (based on the probability the objects in the same cluster came from the same distribution e.g. expectation maximization algorithm [73]) or density model (assumes that objects from the same cluster came from areas of similar density e.g. DBSCAN [94]).
The connectivity model assumes that to some extend all data objects within a data set pertain to the same cluster [247]. Most common is the agglomerative bottom up clustering approach, where in the beginning each object is its own cluster.

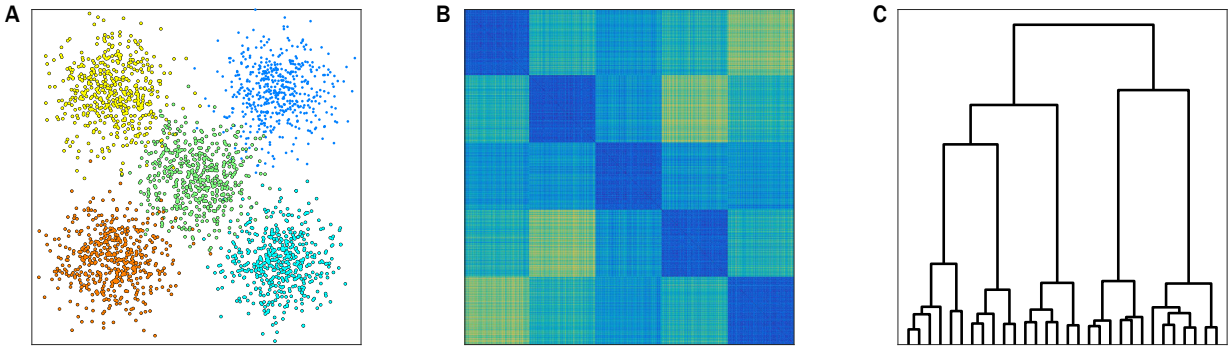

Fig. 8.1. Example of the clustering methodology

A - A test dataset was generated with points that there drawn from five different normal distributions.
B - Based on the coordinates of the points from (A) a distance matrix was calculated using the Euclidean distance.
C - Based on the distance matrix shown in (B) the Ward method for hierarchical clustering was applied to identify the random generators that produced the points in (A). The dendrogram shown indicates that the points can be grouped in five clusters.

In Fig. 8.1 an example of hierarchical clustering is given. The pairwise distance between objects is used to merge the most similar together. The technique creates a hierarchy of the objects in which, with increasing level in the hierarchy, clusters are created by merging the clusters from the next lower level, such that an ordered sequence of groupings is obtained [130]. This merging is done in agglomerative clustering until all objects are linked into one huge cluster. Different approaches have been proposed varying in the linkage criterion. The linkage criterion decides how merging is performed based on a specified inter-cluster distance measure. Several algorithms [322] for computing the distance between clusters have been proposed. As an example, **WPGMA** (Weighted Pair Group Method with Arithmetic Mean) linkage uses the average distance between all pairs of objects in any two clusters. The alternative opposite top down divisive technique is rarely used. The divisive method starts with all objects in one cluster and splits until each objects is his own cluster.
A hierarchical binary cluster tree called dendrogram (Fig. 8.1 (C)) is used to illustrate the linkage of the objects. In the dendrogram U-shaped lines connect the leafs. Each of leafs corresponds to one data points[322]. The height of each node is proportional to the value of the intergroup dissimilarity between its two daughter nodes[130]. It can be a useful tool for identifying the distinct number of clusters that may exist within a dataset by implying cuts in the dendrogram that push the result in the desired clustering. The cuts are placed ideally at the height of the dendrogram where the largest dissimilarity increase occurs.
The idea of the clustering methods based on a centroid model employs a differ-

ent strategy. It is based on an optimization problem that tries to maximize the intra-cluster similarity and the inter-cluster dissimilarity. In general, finding the global optimum requires the enumeration of all permutations of object to cluster mappings. Therefore, iterative methods are used that converge to a local optimum in a subset of mappings.

Commonly the centroid model driven clustering starts with choosing an appropriate number of clusters k. Next, k centroids are randomly assigned, usually by choosing k objects from the dataset. These centroids are the initial centre of each of the k clusters. Based on these centroids each object is assigned to the cluster that corresponds to the centroid closest to the object. After assigning each object the centroids get newly calculated based on the objects that are included in a certain cluster.

The centroid update function differs among the different centroid based methods. For example, the $k-means$ algorithm uses as new centroid of a cluster the object within a cluster that has on average the lowest inter-cluster dissimilarity. The steps of assigning objects to the clusters using the current centroids and updating the centroids are repeated iteratively until a certain number of iterations is reached or the method converges and the object assignment to clusters does not change.

Due to the random start the results of centroid model cluster methods are stochastic. Therefore, several local optima are estimated using different initial centroids and the best is chosen. Centroid model driven methods are heuristic algorithms that only guarantee to find local optima. Finding the global optima remains a computational difficult NP-hard problem [8].

The two discussed clustering approaches differ in their computational complexity. The hierarchical approach has time and memory complexity of $O(N^2)$ for a dataset of size N, because all pairwise similarities between the data objects are necessary for the sequential grouping or splitting of the data objects. Partitional clustering has a lower linear complexity of $O(N)$ because at each iteration the similarities of each objects to the k centroids have to be estimated.

Both discussed clustering models have their advantages and disadvantages. The choice of the algorithm that fits best mainly depends on the particular application. Partitional clustering can deal with big data but assumes that the shape of the clusters is hyper spherical and produces several different results depending on the randomly selected initial centroids. Hierarchical clustering is deterministic and reproducible. In addition, one requires prior knowledge about the number of clusters you expect the data is grounded on in the centroid model, while you can create any number of clusters by interpreting the dendrogram in hierarchical clustering.

In the course of this work a two staged clustering approach is used. The methodology is based on an initial hierarchical clustering step followed by a centroid model based refinement step. For hierarchical clustering the agglomerative Ward's method[354] is used. Ward's method minimizes the within cluster variance based on the sum of squares between two clusters. The advantage of Ward's method

is that it yields very homogeneous clusters. Next, the initial clustering is refined with a centroid model based algorithm utilizing the Hausdorff distance [119]. The Hausdorff distance measures how far two subsets of a metric space are from each other. Iteratively, the Hausdorff distance is calculated for each trajectory to the different distinct clusters and the trajectory is assigned to its closest cluster. This was repeated until an optimal assignments or the maximal number of allowed iterations was reached. The second refinement step is introduced as a sanity check based on the assumption, that the number of iterations is low if the Ward's method classified the data well. For the data presented in the following chapters 11 and 13, a single iteration was sufficient, after which more than 95% of the trajectories remained assigned to their initial cluster.

8.2 Cluster Validation

A validation criterion for number of assumed clusters and object allocation is necessary. Several methods that address these issues have been proposed.
All validation techniques suffer from specific weaknesses due to the complex interaction of clustering algorithms, validation measures and underlying data. Results should be double-checked using alternative validation approaches [127].
The jump method [327] was used to identify the reasonable unbiased choice for the number of clusters and the silhouette method [287] to access allocation quality. Both methods use intrinsic information within the data to assess the quality of the clustering.
The Jump method uses the sum of square errors as a measure of intra-cluster dispersion. This value is computed by $d_k = \sum_{j=1}^{k} \sum_{i=1}^{o \in C_j} dist(o_i, C_j)$ for different number of clusters k. The largest jump $J_k = d_k - d_(k-1)$ is an indicator for the number of clusters that exist in the underlying dataset.
The silhouette method provides a measure how well each objects lies within its own cluster and a simple graphical representation to assess the allocation quality. Given the matrix $D \in \mathbb{R}^{N \times N}$ defined by $D(i,j) := c(x_i, x_j)$ for $i = 1, ..., n$ and $j = 1, ..., n$, the function c measures the distance/dissimilarity among two objects. Let $a(i) = \frac{1}{|C_i|-1} \sum_{j \in C_i, i \neq j} D(i,j)$ be an estimator for the assignment quality of each $i \in C_i$ to the cluster C_i. The smaller $a(i)$ the better the assignment. Let $b(i) = min_{k \neq i} \frac{1}{|C_k|} \sum_{j \in C_k} D(i,j)$ be an estimate for each $i \in C_i$ best alternative assignment to one of the other clusters. The silhouette coefficient is defined as $s(i) = \frac{b(i)-a(i)}{max(a(i),b(i))}$. The silhouette coefficients range from -1 to 1. The closer the coefficient is to 1 the better is the assignment and vice versa.

9

Non-rigid feature detection based on Dynamic Time Warping

Feature detection or pattern recognition aims to find certain features or patterns within the given data. The field of searching for patterns is concerned with the automatic discovery of regularities in the data through the use of computer algorithms. With the use of these regularities actions such as classifying the data into different categories [38] are taken.
Feature detection problems can be found in nearly all scientific realms. And as wide as its application domains is the variety of methods and concepts used. The detection of nuclei in the imaging data is an example already discussed in a previous chapter. To go deeper into this matter is beyond the scope of this work. For a general overview the book of Bishop [38] can recommend.
What constitutes a feature depends on the actual problem and the type of data given. Therefore, the feature detection method used also depend on the scenario one faces when analysing data. The patterns of interested in the course of this work will be temporally connected measurements within the time series that differ from the background like e.g. pulses or sustained changes in the amplitude. Even such a simple feature as a pulse may have many different characteristics and features in the time domain, so that two different pulses may have very different lengths, amplitudes, slopes or min and max values. The feature detection methods used must be therefore selected in a way that fits the characteristic of the pattern one aims to detect in the given time series. In the course of this work, I was interested in features associated with the dynamics of the protein of interest (p53 or SMADs), the detection of cell division events or features in the data linked to errors in the tracking.
This section, starts with the simple problem of detecting cell division events in the trajectories. Afterwards, a novel feature detection method is proposed grounded on the application of DTW. The DTW framework provides the features necessary for dealing more complex problems.
The DTW framework, will equip us with the amount of flexibility necessary in the time domain to find disturbed instances of a specific patterns. This means, it is sufficient to define a pattern primarily by its shape without paying much attention on other characteristics of the feature like amplitude and length. By using DTW one can detect locally connected areas in the time series that exhibit a

certain shape. The different characteristics of this areas can be used to access how a certain stimulus can alter the temporal dynamics of the protein of interest on the single cell level. In this chapter a benchmark is performed on how the novel feature detection approach performs in different scenarios and the potential in real world applications demonstrated with examples dealing with measured time series of nuclear p53 level.

9.1 Detection of cell division events

One potential source of cell-to-cell variability is the cellular state. The cellular state can be directly linked to cell division events. In the imaging data it can be observed that as cells divide the nuclear envelope breaks down during mitosis and two new daughter cells emerge with much smaller nuclei [326]. This drop in

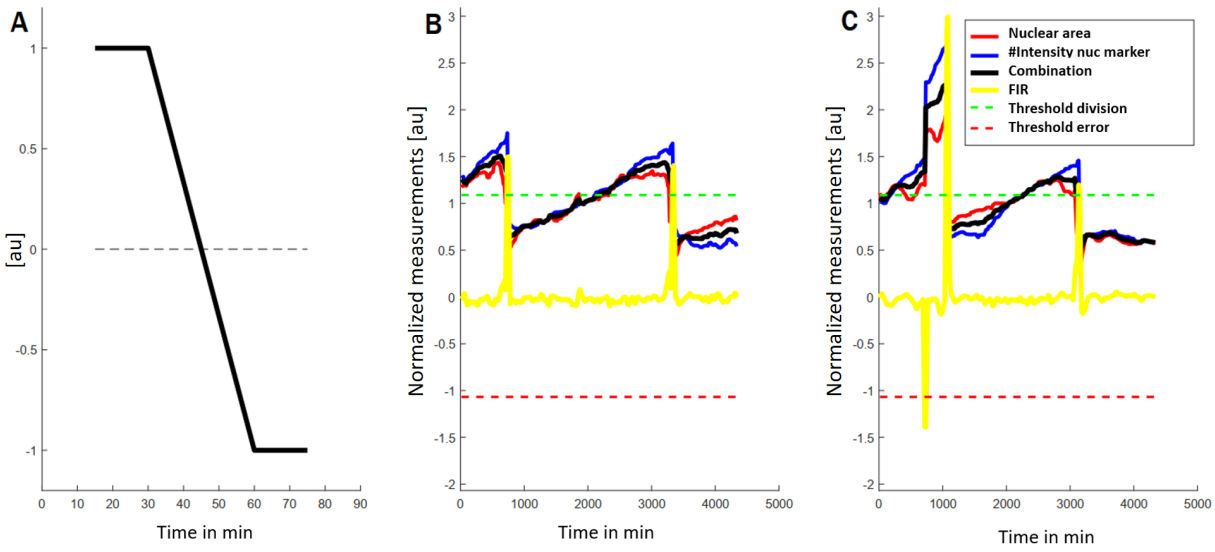

Fig. 9.1. Use of FIR to detect cell division events

A - FIR kernel used to amplify rapid changes in the time series.
B - Example of two detected divisions. The red line displays the size of the normalized nuclear area and the blue line the integrated nuclear measures of the marker over time. The black line is a combination of the normalized nuclear size and integrated marker intensity and is filtered by the kernel on the left. The yellow line shows the result of the FIR. The two dashed lines correspond to the automatically determined thresholds for cell division events (green) and segmentation or tracking errors (red). The cell shown undergoes two divisions during the time of the tracking.
C - Example of an erroneous track. This cell gets excluded from the further analyses, because an unreasonable increase in the size of the nucleus is detectable.

the size of the nucleus can be directly observed in the data as a sudden decrease in trajectories that keep track of the nuclear size and the integrated fluorescence intensity of the nuclear marker. This observation is used to identify the cell division events in the data (Fig. 9.1). To identify cell divisions in the time series the measurements of the integrated intensities of the nuclear marker and the size of the nucleus is utilized. In a first step both time series are normalized by their respective mean and are combined by averaging. A finite impulse response (FIR) filer that mimics a 1D Prewitt filter was applied [199] onto the combined time series to amplify discontinuities, a drop in the size of the nucleus, that correspond to cell division. The window length of the filter was set to 75 minutes. A threshold based on the standard deviation of the values generated by the FIR was used to detect the events of interest. The threshold was set to 4 times the distance of the observed standard deviation from the observed mean. The threshold was confirmed on a set of several hundred established cell divisions identified visually in a set of real data. For the cell divisions identified visually both sensitive and specific there found to be beyond 99%. The pattern that encodes a cell division event in the data is of very low complexity so this simple approach presented is already sensitive and specific enough to produce results of high quality. This is caused by the fact that the event the approach aims to detect is very short and homogeneous in the time domain.
In addition to the detection of the cell division events this method also can be used to improve the quality of the data. This approach can also identify errors caused by faulty segmentation and tracking. A massive increase, the opposite of what is expected in the case of a cell division, in the combined time series, cannot be linked to biological events. Such events can occur in the data if a faulty merging of two nuclei in the segmentation procedure happened. This kind of event can also be easily accessed by applying the FIR. Time series where such kind of error was observed were excluded from further analyses. In Fig. 9.1 all these different steps are shown. The method has proven to be robust and valuable tool within our lab. Also the detected cell division event itself can help to improve data quality. As the nuclear envelop breaks down during mitosis and morphological changes lead to increased auto-fluorescence, signals measured during cell division were disregarded and removed from the trajectories by interpolation.

9.2 Feature detection based on local Dynamic Time Warping

The type of filtering and feature detection presented in the previous section based on signal amplification using FIR is simple and straightforward but suffers from lack of flexibility the time domain. The pattern aimed to detect in the data was short and could not differ in its shape. For the most applications the features of interest in the data are much more complex.

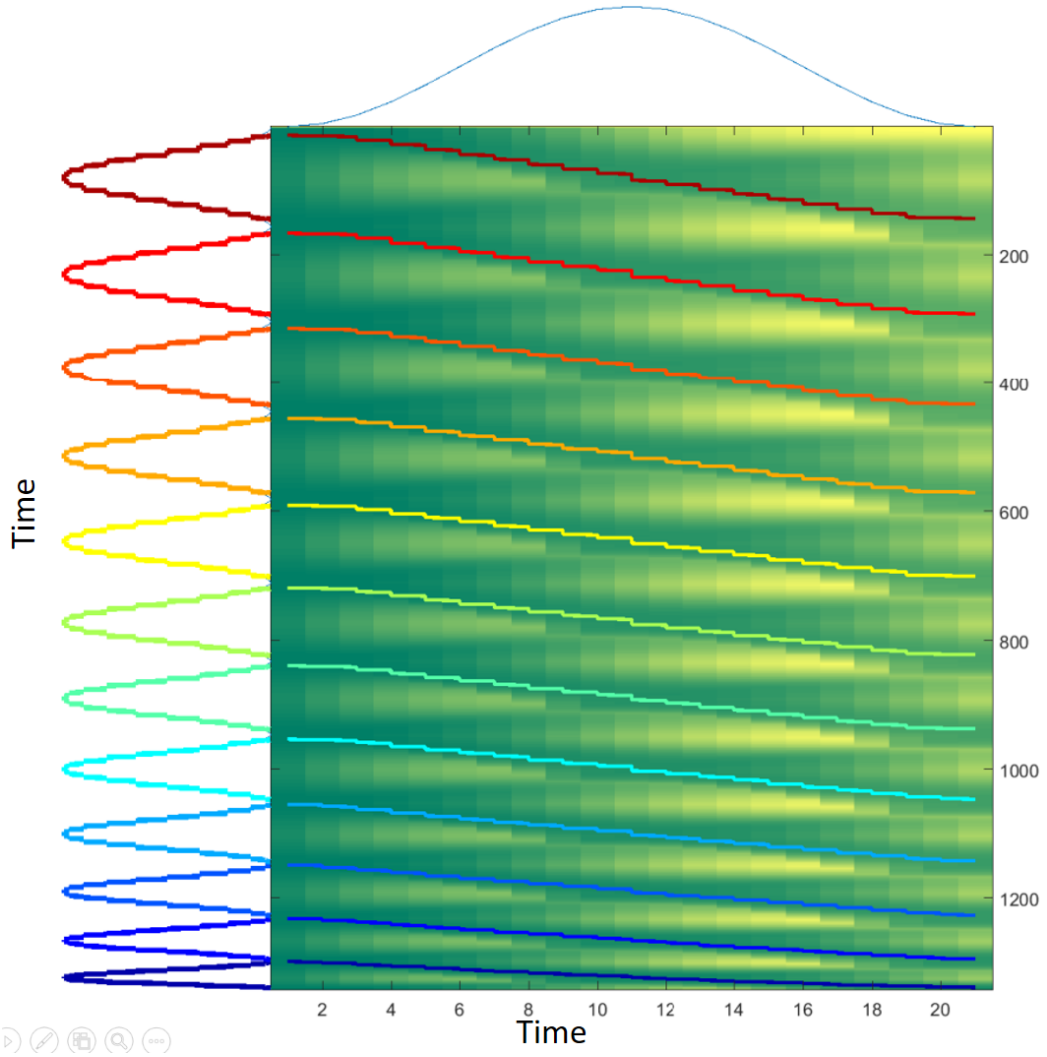

Fig. 9.2. Local DTW is able to detect patterns independent of the time domain

To detect the simple pattern of a pulse shown at the top in the trajectory displayed on the left a matrix is filled as described at the next page of this section using a dynamic programming approach. The trajectory in which the pulse has to be detected is composed of several pulses of decreasing length. The backtracks that correspond to detected features are shown in the dynamic programming matrix. The backtracks start at local minima that corresponds with the endpoint of a feature in the rightmost column. The detected features are displayed in different colour code. Feature detection based on local DTW has the ability of stretching and squeezing the pattern. This allows the detection of variable temporal disturbed occurrences of the pattern in the trajectory of interest.

A method that is based on the concept of FIR but has flexibility in the time and frequency domain is the wavelet transformation, a general tool for pattern recognition. A peak detection framework based on wavelets for the analysis of p53 dynamics from single cell measurements with respect of inevitable noise was

previously proposed [235]. Similar approaches exist for mass spectra [83] or EEG analysis [272]. In the course of my work on the topic of feature detection in single cell time series also wavelets had been taken into account. But the results had not convincing me. Therefore, the perspective was changed towards a novel way to detect features in time series.

The idea is that the strength of the Dynamic Time Warping approach might transfer the freedom and flexibility on the time domain to develop a novel feature detection approach. The idea is inspired by the dynamic programming based Smith-Waterman algorithm [313] that performs local sequence alignments. The Smith-Waterman algorithm is a variation of the Needleman-Wunsch algorithm [252] for global sequence alignments. The Needleman-Wunsch algorithm and DTW differ only slightly in their implementation. Historically DTW was derived from the Needleman-Wunsch algorithm. A basic assumption of DTW is, that one sequence is a time-warped version of the other, so that one of the sequences is either stretched (one-to-many), condensed (many-to-one), or fix (one-to-one) onto the time points of the other sequence. The Needleman-Wunsch uses gaps, where one or more elements in one sequence are not matched by any elements in the other sequence (one-to-none or none-to-one alignment). So the Needleman-Wunsch is appropriate for discrete sequences like DNA sequences and DTW has its application if you can assume contiguous signal that are discretized like our single cell time series. A local version of the DTW algorithm will inherit the characteristics of the Smith-Waterman algorithm.

Pseudo code: local Dynamic Time Warping

double localDTW(X:array[0 ... n-1], pattern:array[0 ... m-1])
  // Init
  D = array[0 ... n, 0 ... m] % init zero matrix of size N x M
  D[0,1:m] = Inf
  B = array[0 ... n-1, 0 ... m-1] % init zero matrix of size N-1 x M-1
  C = $c(X, pattern)$ % pairwise distances between observations and the pattern
  // Recurrence
  for i = 0:n-1
    for j = 0:m-1
      h = $\begin{pmatrix} D[i,j] \\ D[i+1,j] \\ D[i,j+1] \end{pmatrix}$
      D[i+1,j+1] = min(h)+C(i,j)
      B(i,j) = arg min (h)
    end
  end
  return D B

Based on this assumptions DTW will be adjusted so that it performs a local alignment similar to the Smith-Waterman algorithm. The algorithm aims to find

the occurrences of short given pattern in a longer time series. Therefore, the DTW algorithm is altered in a way that the pattern sequence can start and end anywhere in the time series in those the occurrences of a pattern should be detected. The pseudocode of the method is given at the bottom of the previous page.
The pseudo code of the local DTW differs mainly from the pseudocode presented in section 7.2 in the initialisation of the dynamic programming matrix D. After the programming matrix is filled the occurrences of the pattern could be found by backtracking from the local minima in the rightmost column of the dynamic programming matrix. The local minima refer to the ends of local alignments between the pattern and subsequences of given trajectory. So the backtracking to the start of the local alignment gives the potential occurrence of the pattern. In Fig. 9.2 an example of a filled dynamic programming matrix is shown together with the backtracks that refer to the occurrences of the pattern hidden in the artificial trajectory composed of patterns of different length. Next, the behaviour of the local DTW approach for the feature detection is studied given different lengths and amplitudes of the pattern and with respect to noise.

9.2.1 Local Dynamic Time Warping - Robustness to disturbances in the time domain

The expectation that reasoned the transfer of the DTW framework to the field of pattern recognition was, that the novel method will inherit the advantage of flexibility in the time domain. To study if the characteristic is preserved and if the local DTW is capable of detecting a given pattern p in a trajectory independent of the length a time series t was generated that contained n distinct occurrences $q_1, ..., q_n$ of the pattern. Each of the occurrences $q_1, ..., q_n$ has a different lengths that is increasing with the position in t. The pattern had a fixed length of 21 time points while the length of the pattern in the trajectory ranges from 7 for the first occurrence to 2323 time points for the last occurrence in t. In this experiment effects of noise or scaling are not taken into account.
In Fig. 9.2 the results partially are shown for occurrences of patterns q in the range of 39 to 157 time points. Local DTW detects all the occurrences q of a pattern p within the artificial time series t no matter the length, if the occurrences q have the same scaling as the pattern p and no noise is applied. As expected the characteristic of the Smith-Waterman algorithm could be transferred to the DTW framework.

9.2.2 Local Dynamic Time Warping - Robustness to different scaling of the amplitude

While it has been shown that local DTW gives us the flexibility in the time domain, the next thing to investigate was the scaling problem. Can the given pattern

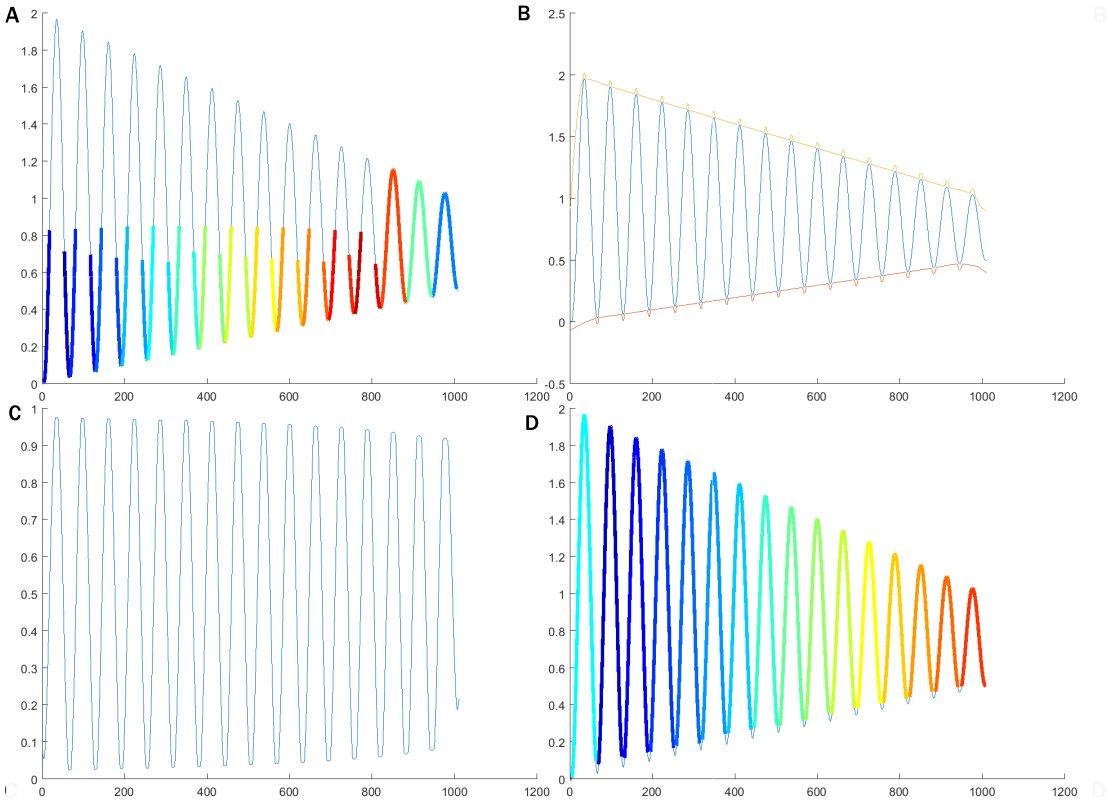

Fig. 9.3. Detection of features by local DTW is sensitive to differences in the amplitude of the pattern and the feature of interest. A normalization step that uses a band around the trajectory can overcome this issue.

A - A pulsatile trajectory with pulses of decreasing strength of amplitude. If the feature detection is applied as shown in the previous figure local DTW fails to identify the pulses correctly. Detected features are shown with different colour code.
B - Estimated upper and lower bound computed mainly on the basis of the local maxima and minima. This bound form a band around the trajectory.
C - With the band from (B) the trajectory can be normalized and eliminate the local differences in the amplitude. The normalized trajectory can be used for the estimation of the occurrence of a pattern in the trajectory.
D - The pulses are correctly detected using the band filter normalized trajectory.

be detected in a trajectory no matter the differences in the amplitude of the occurrences q and the pattern p in the trajectory.
To study this issue, an artificial trajectory t was generated that mimics this question. After applying local DTW on this artificial data (9.3 (A)), it was shown that local DTW itself cannot detect the occurrences q in the trajectory independent of the scaling.
To overcome the scaling issue the local DTW approach was extended by an additional step to bring the pattern and the occurrences of the pattern in the trajec-

tory onto the same scale. While the occurrences of the pattern may have different amplitudes within the trajectory a global scaling of the trajectory could not be appropriate choice. The idea is to generate a band which encloses the trajectory by using a lower and an upper band that incorporates the local minima and maxima of the trajectory. The 'band normalization' computes at the begin constants based on the time series that are used for generating a band around each trajectory. These constants define different attributes of the bands like the width, a maximum value for the lower bound and a minimal level of the upper bound. Using these constants and anchors for the past and the future, for each trajectory an individual band was generated. The anchors for the past and the future are extensions of the given time series based on the average start and end value of the time series 9.5(B). The anchors smooth the band while having equal starts for all trajectories. The upper and the lower band are used to scale the trajectory locally.
Examples of such normalization bands are shown in Fig. 9.3 (B) for the artificial trajectory and in Fig. 9.5(B) for a real world example.
The band is used to perform a local min-max normalization and transform the original trajectory into a range between 0 and 1. Therefore, while the normalization transforms the trajectory into a range between 0 and 1 the pattern itself must fall in the same range. These normalized transformed trajectory is shown in 9.3 (C) for the artificial trajectory and in 9.5(C) for a real world example.
Applying local DTW onto the transformed trajectory returns all the occurrences q of the pattern p. After the positions of the pattern occurrences have been found the positions can be mapped onto the untransformed trajectory as shown in Fig. 9.5(D).
Local DTW alone cannot detect pattern completely independent of the amplitudes, but with the introduction of the additional normalization step the scaling issue can be solved.

9.2.3 Local Dynamic Time Warping - Robustness to different levels of noise

After having studied the behaviour of the local DTW pattern recognition approach in the context of temporal disturbances and scaling and also having extended the framework by an additional 'band normalization' step, it was analysed how robust the method performs while different levels of noise are applied.
To study the influence of noise, a simple time series with 100 pulses was generated and titrated with different levels of noise. The SNR used ranged from the signal 10-fold the noise to the signalling and noise of equal strength. For each of the SNR levels the positions of the 100 occurrences of the pattern was estimated using local DTW. This was done for each of the SNR levels a thousand times with random noise of fixed strength (Fig. 9.4 (A,B,C)).
SNR was defined by the mean of the trajectory x^2 divided by σ^2 of the added white noise. The results are presented in Fig. 9.4 (D). The method started to produce errors if the SNR is below 1.5. But the error rate is very low at in this noise range

so that in less than 5% of the trajectories one extra occurrence is detected, so 99% of the features are still correctly detected. The error rate increases with the decrease in the SNR level. If the signal has the same level as the noise it is not likely that all features are correctly detected.
While studying the imaging data a SNR of 0.59 ± 0.05 can be estimated for the nuclear marker. Therefore, it can be assumed that the SNR in the extracted time series is much lower because the time series are a simplification of the original data. This could be shown for the p53 data in Fig. 9.5 (E). The average SNR measured was beyond 30. In this SNR range the impact of noise on the performance of local DTW is negligible.

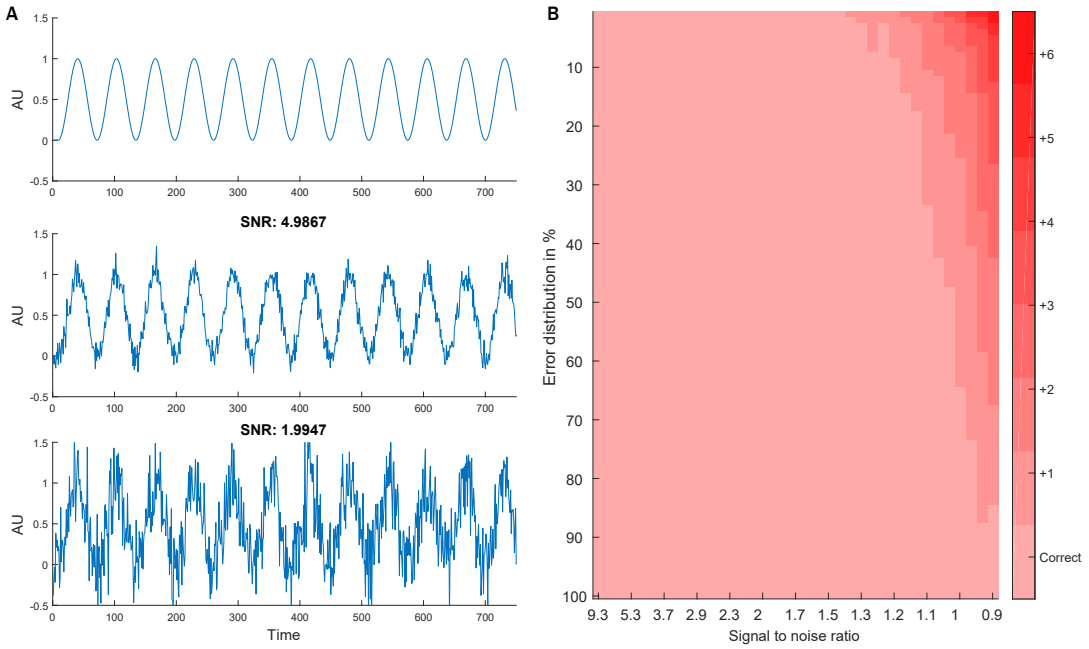

Fig. 9.4. Local DTW is robust to noise

A - An artificial time series with 100 pulses was generated and different levels of noise there added. At the top a subsequence of this trajectory which contains 12 pulses is shown. The same subsequence is shown in the middle with a signal-to-noise ratio of 5 and at the bottom of 2. The SNR was defined by the mean of the clean trajectory x^2 divided by σ^2 of the added white noise.
B - The undisturbed trajectory with 100 pulse shown in (A) was titrated with different levels of noise in the range from 9.35 till 0.93 and aimed to detected the 100 pulses. In different colour code the deviation in the number of detected pulses is shown. As long as the noise is not greater that the signal the local DTW approach for the feature detection identifies all artificial pulses.

9.3 Use case: detecting p53 pulses with local dynamic time warping

To demonstrate the potential of the workflow a real world example of a p53 trajectory is revised. In p53 trajectories the main interest is the detection of pulses. To quantify the characteristics of the identified pulses different features can be extracted (Fig. 9.5 (A)).
That enables the comparison of dynamics among different single cells and different experimental conditions.
As described before for each trajectory a band was calculated that normalizes each trajectory individually while incorporating the temporal order of the measurements (Fig. 9.5 (B,C)). One important aspect in the generation of the 'band' is the use of anchors into the past and the future of each trajectory based on the starting and end values of all trajectories in the data set. The anchors are used to elongate each trajectory and therefore increase the quality of the estimated band that encloses a trajectory. The 'band normalization' step ensures that local DTW detects disturbed occurrences of a pattern that differ in length and amplitude as shown in Fig. 9.5 (D).
Local DTW in combination with local 'band' normalization gives a novel, robust and flexible tool for the pattern recognition problem. Especially the flexibility in the time domain makes it a promising approach for the analysis of single cell data.

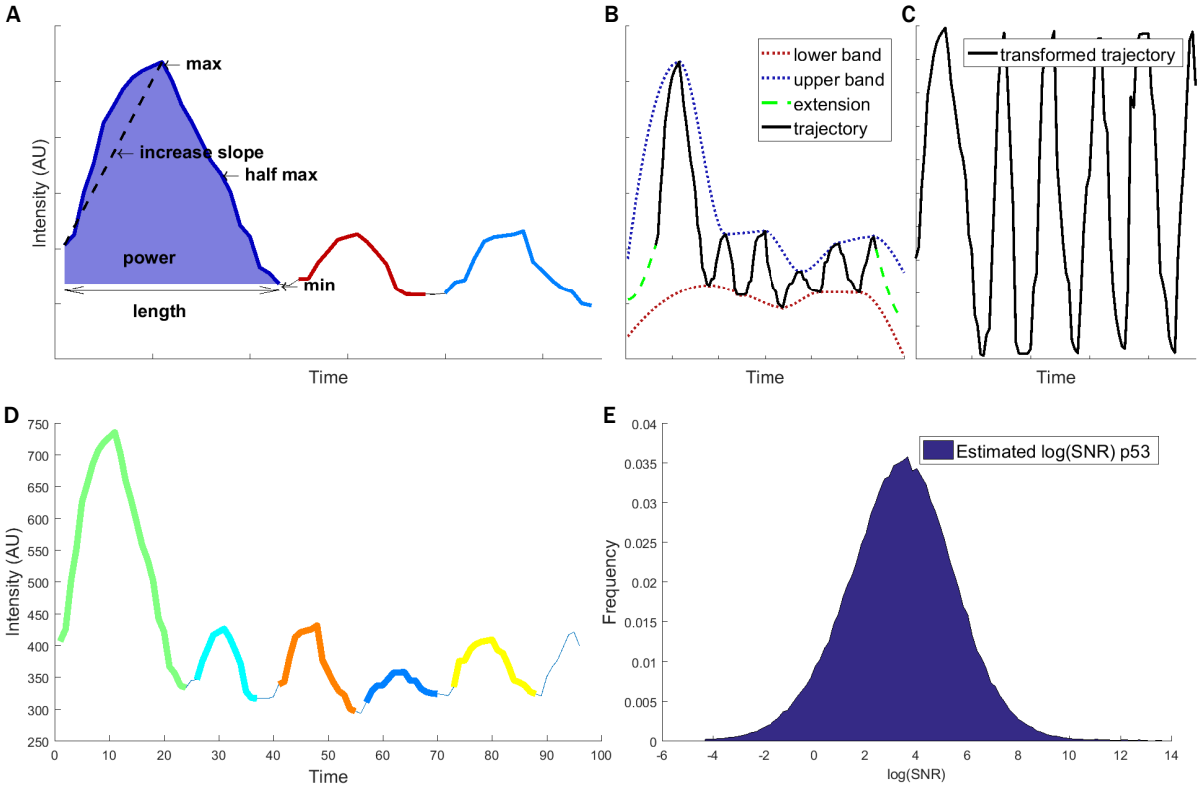

Fig. 9.5. Feature detection based on local DTW works on non-artificial data

A - Example of a trajectory with detected features and a schematic representation of feature characteristic readouts.
B - The feature detection method based on local DTW generates initially a band around the input trajectory to level local amplitude differences. In addition to local minima and maxima anchors into the past and future are added based on the average initial and final intensities of the population the cell was drawn from.
C - The band normalized trajectory shown in (B). The upper and lower bound of the band are used to normalize the original trajectory into a fix range between 0 and 1.
D - The detected features in the trajectory already used in (B) and (C). Pulses are detected independent of their dynamic, length or amplitude.
E - SNR approximation estimated from non-artificial trajectories of p53 labelled cells. The standard deviation of non-irradiated cells is used as a model for the noise and the mean of the quadric minimum normalized intensity of irradiated cells as a model for the signal $(mean(irradiated^2)/std(non-irradiated))$. The SNR levels estimated fall in a range there the local DTW feature detection approach should work proper.

10

Detecting global patterns in single cell trajectory datasets

In the previous section a method was proposed that is able to detect local patterns in the single cell trajectories. The local DTW method is appropriate for p53 time series data, where the detection of pulses is of main interest. For SMAD2 trajectories this is not that simple. SMAD2 dynamics are composed out of a great variety of single patters ranging from simple pulses, sustained plateaus to complex shifts in the SMAD2 nuc/cyt ratio. Therefore, the problem of quantifying SMAD dynamics will be tackled from a global point of view.
In this chapter, it is demonstrated how principal component analysis (PCA) might contribute to change the perspective from distinct features within a trajectory to a global view where the individual dynamic of a single cell is composed of fixed proportions of underlying dynamics. Based on the observations in the context of PCA the focus will be shifted from the unsupervised learning procedure to a supervised that incorporates the knowledge gathered while working with the data. The framework presented will exhibit several new ways to gather insights into the structure of the single cell data.

10.1 Principal component analysis for single cell time series

PCA is a commonly used statistical algorithm that reduces the dimensionality of the data, while retaining most of its variation [1]. This is achieved by finding principal components which contribute the most to the variation in the data. PCA uses orthogonally transformation to change a set of observations into a new data set of linearly uncorrelated vectors of variables called principal components. The variables that constitute an observation or trajectory in our single cell data are the measurements at the different time points. The transformations are defined in the way that the first component has maximal variance and the other components have the maximal variance under the constrain that they are orthogonal to the previously determined component.
In the actual scenario the principal component are fundamental trajectories, which can describe underlying dynamics present in the single cell data. PCA will be used to identify some set trajectories, the principal components, whose weighted linear

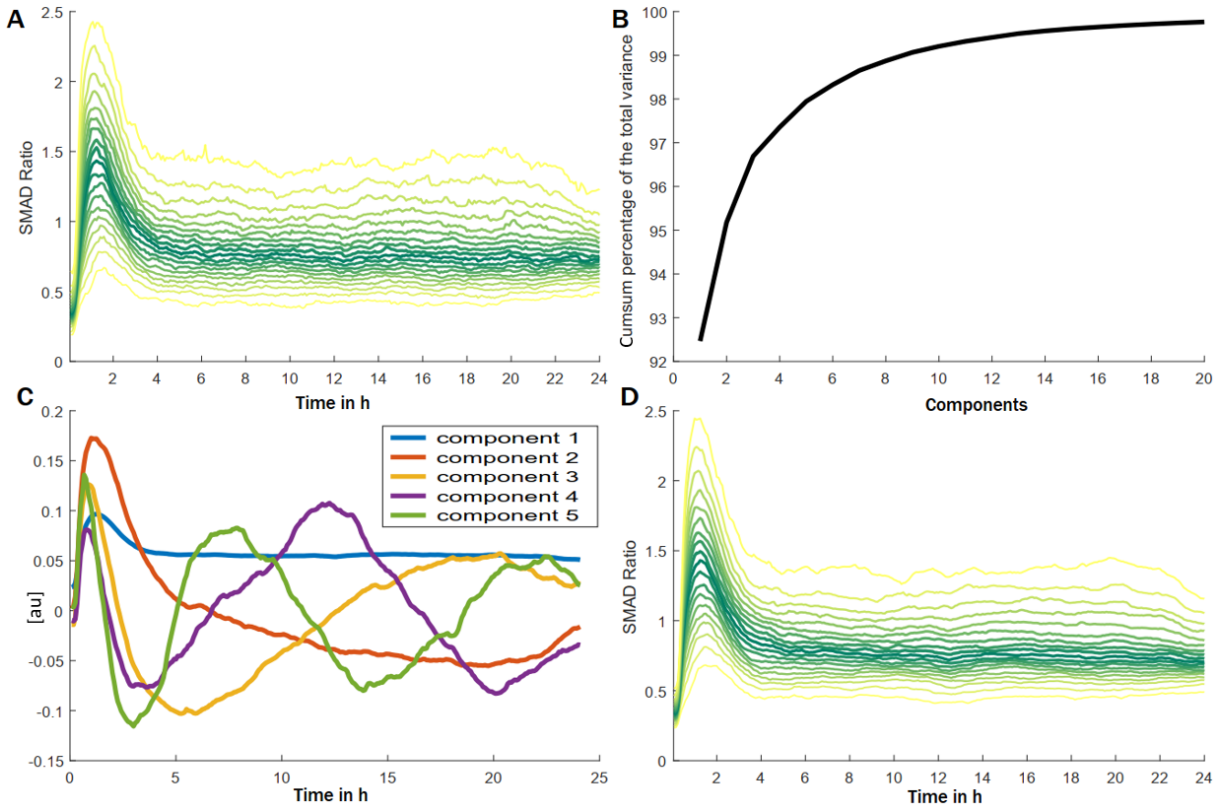

Fig. 10.1. Unsupervised learning for time series data: Principal component analysis (PCA) provides a dimensional approximation that can be used to reduce the dimensionality of the data

A – The dataset of SMAD2 labelled cells stimulated with 100 pM TGFβ (352 cells) used to illustrate the PCA. The different colour coded lines represent the quantiles p in the interval from 0 to 1 with steps of size 0.05. Hence the most yellowish outer lines are identified as the min and max observed at the certain time points.
B - Cumulative sum of the percentage of variance explained by the first 20 temporal principal components. PCA enables us to reduce the dimensionality of the original data set shown in (A) Nine principal components are able reproduce already more than 99% of the original dataset.
C - Patters of the first 5 temporal principal components.
D - Data reconstructed with the first 5 principal components. The reconstructed data is smoother than the original data.

combination will reconstruct the original trajectories. The PCA implementation provided by Matlabs Statistics and Machine Learning Toolbox [322] was used.
Given a single cell time series data matrix D for N cells (the observations) and T time points (the variables). The rows correspond to the cells and the columns to the time points. The in general applied first step of PCA of subtracting the mean of each variable from the dataset to centre the data around the origin is skipped. The function $[C, L] = pca(D)$ returns two matrices C and L. The T by

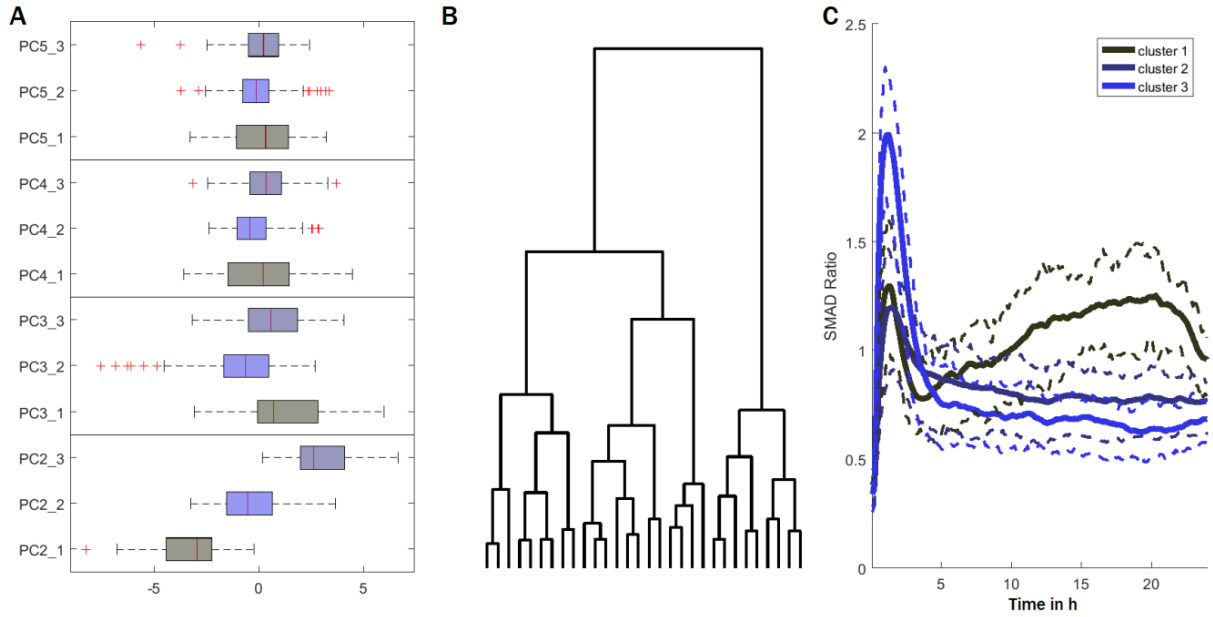

Fig. 10.2. The loadings estimated for temporal principal component can be used to quantify the different behavioural subgroups in data sets.

A - Based on the estimated loadings of the temporal principal components that allow the reconstruction of the original data set (see previous figure) signalling classes can be identified in the data. Hierarchical clustering was used on the loadings of the first five components except the first to identify three distinct behavioural patterns in the original dataset. Show are the weights of the loadings with respect to the clusters identified.
B - Dendrogram generated using 'Wards' method for hierarchical clustering on basis of the loadings of the first 5 components except the first estimated for the reconstruction of the original data set.
C - Median trajectories of the three identified signalling classes (cluster 1 - 60 cells, cluster 2 - 185 cells, cluster 3 - 107 cells) in the data set of cells stimulated with 100 pM TGFβ. The dashed lines indicate the 25% and 75% quantile.

T matrix C where each column contains coefficients for one principal component with columns in descending order in terms of the explained variance and the N times T principal component loadings matrix L, which is the representation of D in the principal component space. Because the first step of centring the data around the origin is skipped D can be reconstructed by LC'. For deeper insight into this topic the book from Bishop [38] is recommended.
In Fig. 10.1 an example is given on how PCA works on real dataset of single cell trajectories. The data contributes the nuclear to cytoplasmic SMAD2 ratio of 352 cells stimulated with 100 pM TGFβ. To reduce the dimensionality of the data one can make use of the order of the principal component in the matrix C. In Fig. 10.1 (B) it is shown how many principal components are necessary to explain a certain amount of the variance in the data and vice versa and to allow the recon-

struction of the original data with respective precision. In the example presented a dimensionality of five is sufficient to reconstruct nearly 98% of the data. So, $L(:, 1:5)C(1:5)'$ will give us a very a good approximation of the original data D. Consequently, each original trajectory can be approximated by a weighted linear combination of the top five principal components where the weights differ among the different trajectories. In Fig. 10.1 (C) the shape of the first 5 principal components is presented. One can interpret these principal components as 'fundamental' dynamics that shape the dynamic of the population of cells studied.
As has been shown, one does not only access the shape of the 'fundamental' components C, but also an estimate L to which extend each component contributes to the single cell trajectories is available. These loadings can be used to identify classes of signalling dynamics that contribute to the composition of the cell population. In Fig. 10.2 the loadings of the top five principal components except the first of each cell are used to cluster the population into three distinct groups. The first principal component can be associated with the mean and the approach aims to enhance less obvious dynamics. With the loadings one can embed each single cell trajectory into a five dimensional space where each dimension corresponds to the loading of the top five principal components. As the principal components are sorted by the variance they explain the impact of the first principal components for the clustering is much larger than for the later. This can be seen in the boxplots that show the contribution of each component to the cluster (Fig. 10.2(A)). Hence, the 5th component does only have minor impact on the clustering.
PCA can provide a useful toolbox while working with single cell time series data. While working with this framework two drawbacks were identified that restricts the usability. First, the method is unsupervised and therefore the interpretation of the principal components and the corresponding loadings is not always clear. Second, there is no fixed range for the values the principal components and the loadings can attain. This makes the interpretation as well complicated. To overcome these problems but keep the advantages of the PCA, in the next section a supervised learning method is introduced into the analysis of single cell time series.

10.2 Supervised identification of fundamental patterns in the single cell time series data

In this section, an alternative approach is introduced to quantify single cell dynamics that incorporates experience and knowledge gathered while studying the data and biological background. Based on artificially constructed trajectories, that are expected to encode the underlying dynamics of the data it aims to find a dimensionality reduction that allows the reconstruction of the original data to a certain degree. The knowledge driven constructed trajectories will be named in the following as fundamental patterns. The fundamental patterns will all have only values in the range from 0 to 1.

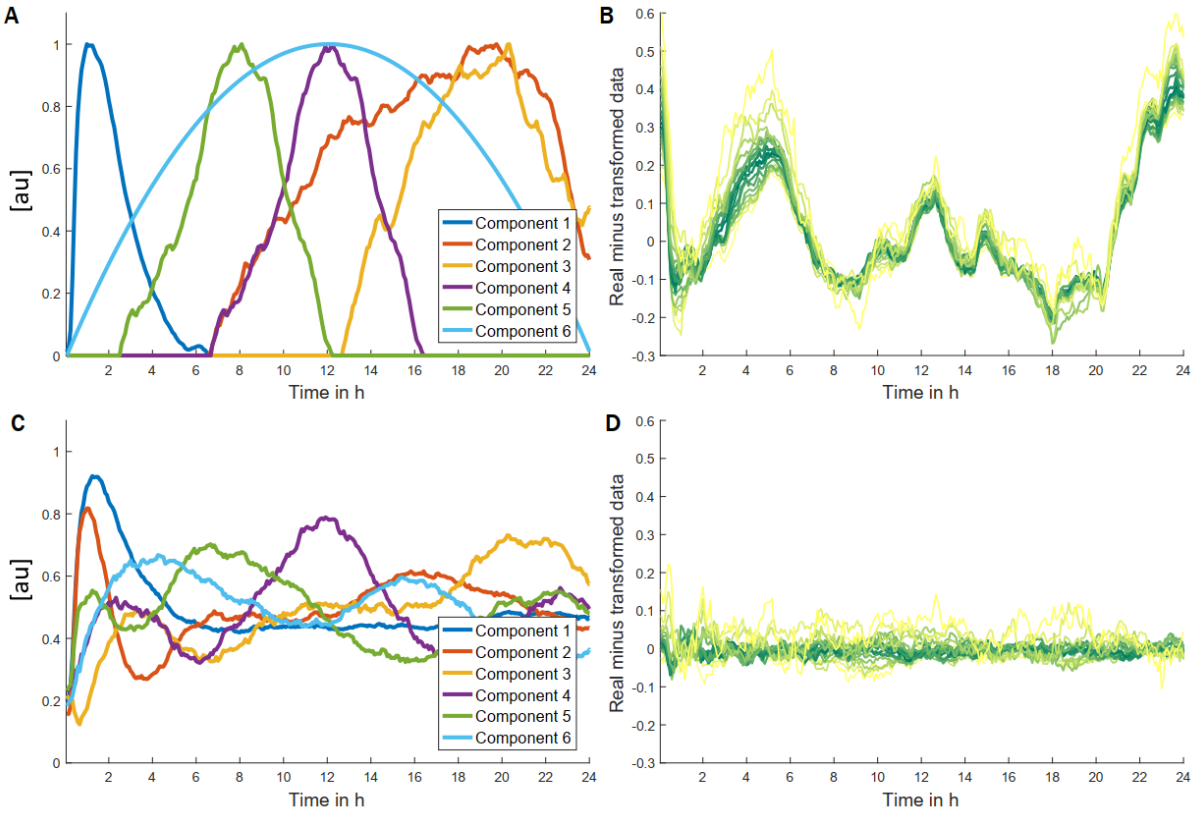

Fig. 10.3. Supervised learning for time series data: Identifying fundamental temporal components based on knowledge

A - Starting with the construction of the fundamental patterns one expect to find in the data. Each of the component describes a distinct feature in the data like the initial response or a late plateau.
B - While using the initial knowledge driven components, only around 34% of the original variance in the data could be explained. Shown is the distribution of the distance of the reconstructed data set to the original data set.
C - Using a constrained non-linear programming solver one can fit the fundamental patterns based on our initial guess so that the explained variance in the data is increased. After the fit the estimated temporal fundamental components explain more than 99% of the variance of the original data set.
D - The distribution of the distance of the reconstructed data using the fitted components set to the original data set.

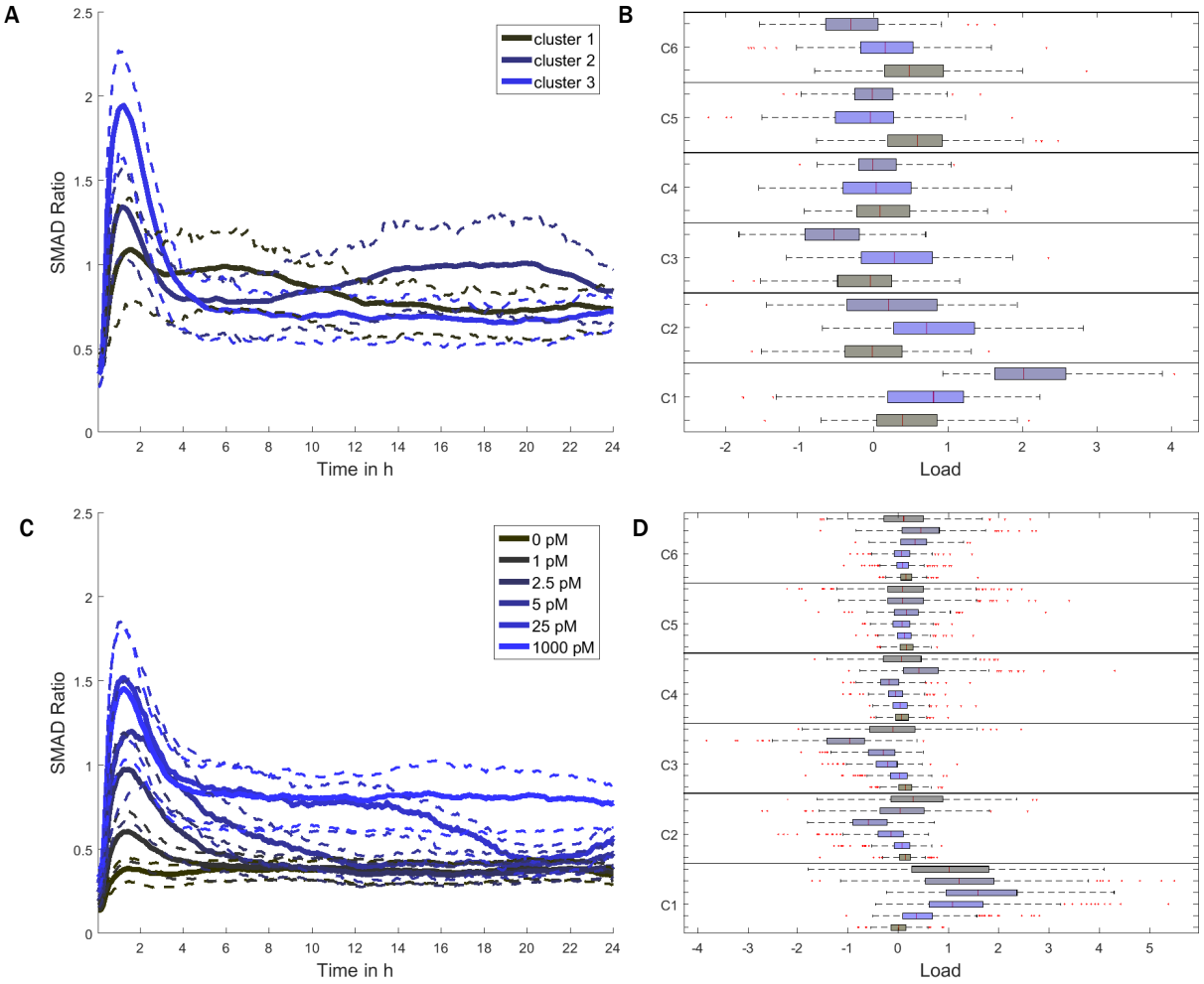

Fig. 10.4. Supervised learning for time series data: Identifying fundamental temporal components based on knowledge

A - As in the previous figure the loadings can be used to identify distinct behavioural classes in the 100 pM dataset (cluster 1 - 60 cells, cluster 2 - 185 cells, cluster 3 - 107 cells).
B - The weights of the fundamental patterns describe the dynamics of the clusters shown in (E).
C - The 100 pM data set is a taken from a larger experiment with six different condition (control - 358 cells, 1 pM TGFβ - 395 cells, 2.5 pM TGFβ - 314 cells, 5 pM TGFβ - 295 cells, 25 pM TGFβ - 351 cells, 100 pM TGFβ - 352 cells). Using the fitted components more than 97% of the variance of the complete data set is reproduce, while the initial components only reproduce 22% of the variance. D - The loadings of the fundamental patterns with respect to the condition reconstructed.

Given as before a single cell time series data matrix D for N cells (the observations) and T time points (the variables). Based on experience gained while working with the data one constructs M different fundamental patterns stored in a T by M matrix C where each column of contains coefficients for one fundamental pattern. It is of interest to identify a N times M loading matrix L so that the origin D can be reconstructed by LC'. L can be estimated by solving the linear equation $CL = D'$ for L. Shown in Fig. 10.3 (A) are six fundamental patterns constructed for data that contributes the dynamics of the nuclear cytoplasmic SMAD2 ratio after stimulation with 100 pM TGFβ over a period of 24 h. The different distinct fundamental patterns mimic several aspects of the dynamics that were observed in the data. With the estimated weight of each of the corresponding loadings L one can try to reconstruct the original data matrix D with the fundamental patterns. As shown in Fig. 10.3 (B) this simple approach can only explain a little less than 34% of the original data. This is far beyond the precision required.

To improve the precision constrained non-linear multivariable function is minimized and the fundamental patterns are fitted so that the variance explained is maximized. The function to minimize is $\sum(D - (C\backslash D') * D')$ where $L = C\backslash D'$ solves the system of linear equations $CL = D'$. The system is constrained so that C can only utilize values in the range from 0 to 1. To minimize and fit the fundamental patterns onto the original data the function $fmincon$ from Matlabs Optimization Toolbox [259] was used. After the fitting of the fundamental patterns 10.3 (C,D) 99% of the original data could be reconstructed. This is slightly better than the same number of principal components for the test data (Fig. 10.1 (B)). The fitted fundamental patterns inherit basal characteristic from the initial trajectories that where constructed using existing knowledge.

This alternative supervised learning framework exhibits the advantages of the PCA analysis. As in the section before, one can use the loadings L to identify signalling classes that contribute to the composition of the cell population (Fig. 10.4 (A, B)). The supervised approach has the advance that in contrast to the PCA the components are not ordered by the variance they explain. Therefore, the fundamental components do not naturally differ in the amount of variance they explain. While comparing the loadings with respect to the clustering between PCA and the supervised approach it is clear that the several fundamental patterns contribute much more equally to the classification in contrast to PCA where the differences among the loadings of the principal components came mainly from the first few. In the PCA the 5th used component only explains less than 0.5% of the variance, therefore the influence on the classification is only minor. The fundamental patterns are also easier to interpret than the principal components while they fall all in the range between 0 and 1 and they already incorporate expectations generated by experience.

To test the potential of new approach for the analysis and quantification of single cell time series it was studied to which extend the estimated fundamental patterns can explain additional data. The data D used in the example to fit the fundamental patterns ((Fig. 10.1 (A)) and 10.1 (C))) is only a subset of a bigger data

set that incorporates six different conditions in total. The conditions differ in the TGFβ stimulus strength that ranges from 0 to 100 pM. How much of the variance of the total dataset can be explained with the fundamental pattern fitted only onto the 100 pM conditions? With the initial unfitted only 22% can explain, with the principal components extracted from the 100 pM conditions close to 94% and with the fitted fundamental patterns one can reconstruct more than 97% of the complete data (Fig. 10.4 (C, D)).

The concept that allows the use of the fitted fundamental components to compute loadings and reconstruct data that was not used in the fitting process can also be used to map new trajectories onto already existing classifications. This classification must not originate from the loadings. New trajectories can also be mapped on clusters estimated using other methods like DTW.

Part IV

Heterogeneity in canonical TGFβ signalling is grounded on distinct signalling classes

11

Cell-specific TGFβ signalling

With a combination of time-resolved monitoring of the nuclear to cytoplasmic SMAD2 translocation at the single cell level and quantitative mathematical modelling, cell-specific long-term dynamics of SMAD mediated signalling will be analysed and it will be studied how phenotypic responses of epithelial cells are modulated by the strength of TGFβ stimulation.
The following chapter is based on previously published data [326, 325]. The experiments used in the following sections were carry out Henriette Strasen at the Max Delbrück Center for Molecular Medicine and Stefan Bohn at the TU Darmstadt.

11.1 Heterogeneity in SMAD dynamics

Canonical TGFβ signalling processes the external information by ligand binding to transmembrane TGFβ type II receptor (TGFβR2) followed by bridging of TGFβR2 and TGFβ type II receptor (TGFβR1) dimers [217]. TGFβR2 undergoes autophosphorylation and catalyses transphosphorylation TGFβR1 within the activated heterotetrameric receptor complex [96]. The activated receptor complex phosphorylate SMAD2/3, which then heterotrimerize with SMAD4 and translocate to the nucleus and where is binds to target gene promoters and regulates transcription. Recent studies revealed that nuclear translocation of SMAD2-SMAD4 occur with pronounced cell-to-cell variability [355, 389]. To analyse TGFβ pathway activity the translocation of the SMAD complexes was monitored in individual cells using time-lapse imaging and the live-cell reporter system, introduced in section 4.1. Imaging data was acquired over a time period of 24 h with temporal resolution of 5 min.
As an estimate for the shuttling of SMADs between the cytoplasm and the nucleus the ration between the average cytoplasmic and nuclear level of SMAD2 was used. The cytoplasmic measurements were accessed heuristically as described in section 6.2. In brief, for all identified nuclei non-overlapping rings of fixed size around the segmented nucleus were generated. In the area of the rings the cytoplasmic abundance of SMAD2 was estimated.

Initially, the SMAD translocation after stimulating the cells with a high dose of

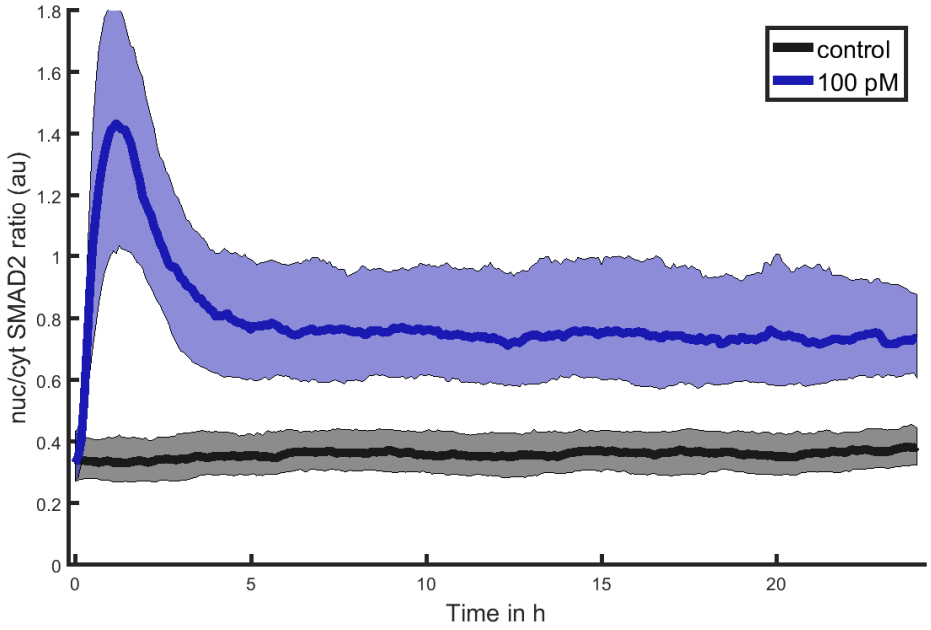

Fig. 11.1. Time-resolved analysis of TGFβ induced SMAD2 nuclear to cytoplasmic translocalization for stimulus levels in the range from 0 pM (black, 358 cells) and 100 pM (blue, 352 cells). Nuc/cyt SMAD2 ratios for six individual cells are shown. The median is shown a bold line and the 25% and 75% is represented by the coloured area.

100pM TGFβ was studied in relation to unstimulated cells (Fig. 11.1). In unstimulated cells the SMAD2 equilibrium is shifted primarily to the cytoplasmic side and only very low levels of SMAD can be found in the nucleus. The population dynamic observed, after stimulation with 100 pM TGFβ, is that the equilibrium turns massively towards the nuclear side, so that SMAD2 accumulates in the nucleus within the first two hours after stimulation until it reaches its maximum. The fast increase is followed by a slower re-shuttling of SMAD2 to the cytosol for the next three to four hours. After around five hours the population average level of SMAD2 stays elevated for the period measured during the experiment.
While investigating the single cell dynamics of several hundred of cells, stimulated with 100 pM TGFβ, a substantial heterogeneity in the translocation dynamics of SMAD2 could be observed (Fig. 11.2). Cells differ in their amplitudes of the direct response of nuclear SMAD2 accumulation and in their late phase dynamics. In the later elevated plateau phase, cells differ substantially in their nuc/cyt SMAD2 ratio or exhibit additional periods of SMAD2 accumulation and asynchronous fluctuations could be observed while in the state of the equilibrium.
On the average level the results fit to previous studies done with populations of cells [152, 62, 388, 347], but they deliver a more fine grained picture of signalling dynamics due to the single cell level analysis. The simplified view on the cell population masks the majority of the dynamics and gives only a poor representation of the pathway activity.

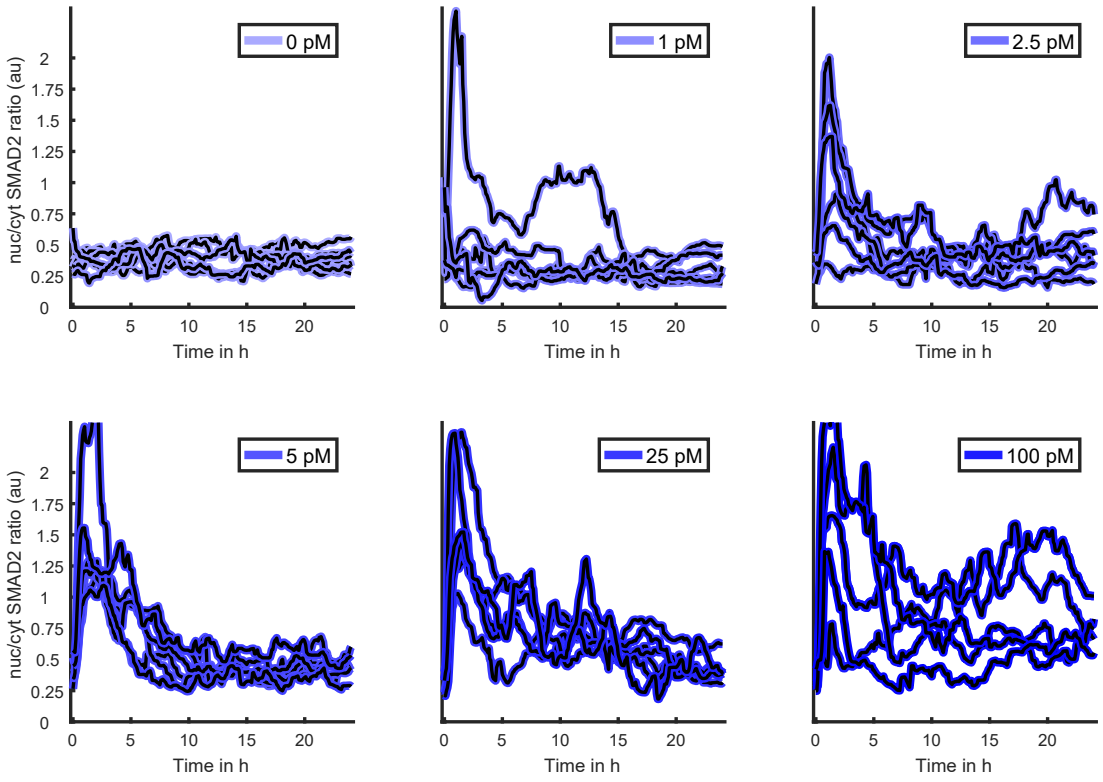

Fig. 11.2. Heterogeneity of TGFβ induced SMAD2 nuclear to cytoplasmic translocalization for stimulus levels in the range from 0 pM and 100 pM. For each indicated doses of TGFβ six randomly chosen single cell trajectories of the nuc/cyt SMAD2 are shown.

To validate our reporter system several experiments were carried out [326, 325]. Western blot analysis revealed that the amount of SMAD2-YFP fusion protein corresponds to approximately 50% of the endogenous SMAD2 protein. By monitoring TGFβ-induced dynamics of phosphorylation of endogenous SMAD2 in the parental and reporter cell lines it could be shown that the induced overexpression did not disturb SMAD translocation dynamics. The expression of SMAD target genes SMAD7, SnoN and PAI-1 upon stimulation with 100pM TGFβ in parental and the reporter cell lines remained unchanged. It also was shown, that the translocation of the SMAD complex into the nucleus is fully dependent on the activity of the TGFβ type I receptor [326]. Canonical TGFβ was prevented by or directly terminated after the application of the inhibitor SB431542 [152] independent of the time passed since stimulation. In an ongoing response SMAD2 was imminently removed from the nucleus and shuttled back to the cytoplasm.
After it was ensured that the reporter system is a suitable tool to quantify TGFβ signalling dynamics, it was studied if and how the strength of the stimulus is reflected by SMAD translocation dynamics. To this end, the initial experiment was

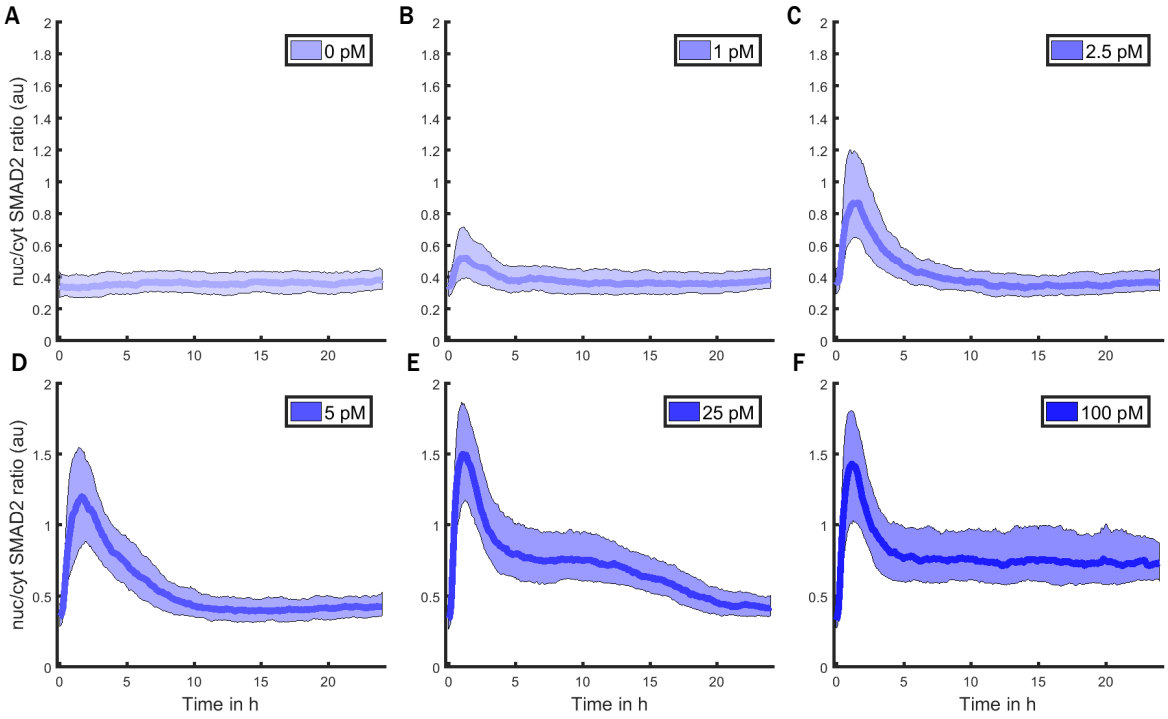

Fig. 11.3. Time-resolved analysis of TGFβ induced SMAD2 nuclear to cytoplasmic trans-localization for varying TGFβ stimulus levels.

A-F - Median nuc/cyt SMAD2 ratio of cells stimulated with varying concentrations of TGFβ over 24 h. The shaded area represents the 25% to 75% quantile range. (total 2065 cells, control - 358 cells, 1 pM TGFβ - 395 cells, 2.5 pM TGFβ - 314 cells, 5 pM TGFβ - 295 cells, 25 pM TGFβ - 351 cells, 100 pM TGFβ - 352 cells)

repeated with varying doses of TGFβ. In Fig. 11.3 (A-E) the response on the level of the population is shown by the median and the 25% and 75% inter-quantile range of the nuc/cyt. The spread of the inter-quantile range is an indicator for the heterogeneity of the pathway activity. Heterogeneity increases with the stimulus strength. The increase in heterogeneity starts at low doses in the temporal range of the direct SMAD response and with increasing doses spreads out to later time points. The time dependent inter-quantile ranges of the obtained ratios overlap among the different conditions. From the population perspective, the strength of the initial response as well as the strength and length of the secondary response are directly determined by the strength of the extracellular stimulus.

To quantify the direct response to the TGFβ stimulation, the average of the greater half of ratio measurements within the first 4 h of the experimental course (Fig. 11.4 (A)) is plotted. This measurement was chosen to compensate lags in the timing of the direct response. A steady increase could be observed until higher doses are reached. The strength of the direct response becomes independent at doses greater than 25 pM TGFβ. With increasing stimulus strength, the variability of the first response is increased but at higher doses also the strength of the

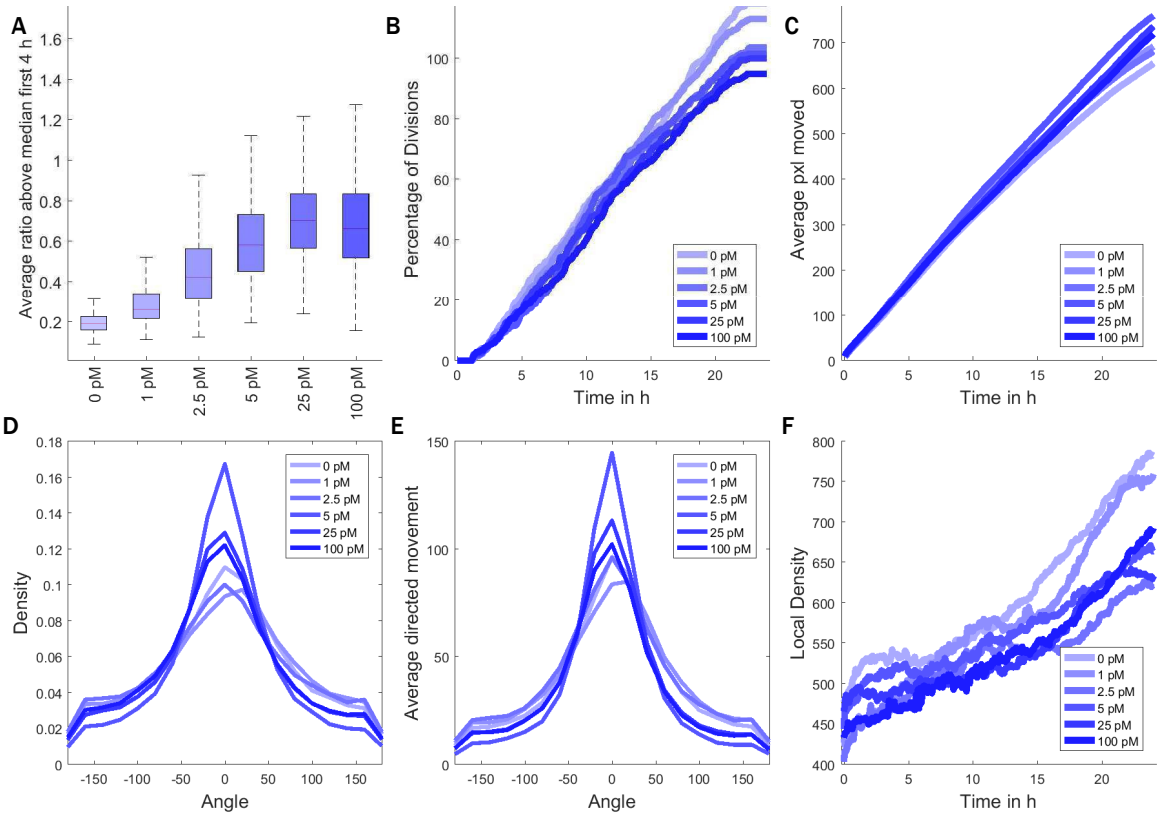

Fig. 11.4. Time-resolved analysis of TGFβ induced phenotypic responses for varying TGFβ stimulus levels.

A - Average nuc/cyt ratio of values greater than the median within the first 4 h with respect to varying concentrations of TGFβ.
B - Cumulative probability of the occurrences of a cell division event with respect to varying concentrations of TGFβ.
C - Cumulative average distance moved (pxl) over the course of the experiment with respect to varying concentrations of TGFβ. A pxl has the size of 1.6 μm.
D - Change of direction with respect to varying concentrations of TGFβ.
E - Moved distance in relation to direction of the movement with respect to varying concentrations of TGFβ.
F - Density of cells over the course of the 24 h of the experiment with respect to varying concentrations of TGFβ.

first response obtained to low doses could be detected.
At the single cell level, the picture is much more delicate and complex. Again variability in the signalling dynamics induced by a fixed stimuli (Fig. 11.2) could be observed. The strength and the variability of the initial response is positively correlated with the stimulus strength. Also the responsiveness itself is connected to the stimulus strength, as at low doses, only a small fraction of cells shows a direct response. The number of the responding cells increases with the strength of the stimulation. Additional later translocation events occur as well more often

at higher TGFβ doses and are mostly absent at low stimulus concentrations. At all doses one can observe outliers, that behave beyond the expectations as shown in Fig. 11.2 for the 1 pM condition.
How is the extracellular concentration of the ligand TGFβ connected to the phenotypic responses? From the microscopy data several phenotypic features like cell division events, quantity and quality of the cellular motility or the local cell density could be quantified. While looking at the motility, special attention was paid onto how directed cells move by investigating the distribution of speed and direction change [141]. The cellular density was estimated with respect to all segmented cells in each frame of the time series of microscopy images. In Fig. 6.5 the density estimates are shown for cells identified in a microscopy image. The methods to identify cell division events and estimate the local density and the motility are described in detail in chapter 6.
Only a minor correlation between the TGFβ dose and the phenotypic readouts could be observed. Roughly, unstimulated and cells exposed to low extracellular doses of ligand divide a little more frequent, move a little slower and less directed and grow to a denser population (Fig. 11.4 (B-F)).
The central observation is that signalling dynamics upon different stimulation doses overlap among different conditions. So that trajectories with similar dynamical patterns could be found in populations stimulated with different doses of the ligand. If dynamical similarities could be quantified it should be possible to sort cells according to their dynamic response independent from the condition. This might give an alternative view to quantify and understand the observed heterogeneity in the response dynamics given a fixed TGFβ stimulation dose. Based on these findings the conclusion is that the strength of the TGFβ stimulus determines only partially the dynamical response and the phenotypic outcome.

11.2 Signalling classes decompose the heterogeneous SMAD pathway activity into distinct dynamics

Signalling dynamics upon different stimulation doses overlap among different conditions so that the strength of the TGFβ stimulus determines only partially the phenotypic response. This raises the question, whether the observed dynamics of SMAD2 translocation predict the phenotypic responses more precise than the extracellular concentration of the ligand. To test this hypothesis, the similarity among SMAD dynamics needs to be quantified. To describe the whole population of cells quantitatively the similarity estimate could be used to identify classes of signalling dynamics. This gives the opportunity to link signalling dynamics independent of the stimulus strength to phenotypic characteristics.
To shed light on this question, the constrained dynamic time warping (cDTW) approach, introduced in section 7.2 is employed. With cDTW, it is possible to compare SMAD dynamics across multiple stimulation regimes. cDTW is adjusted

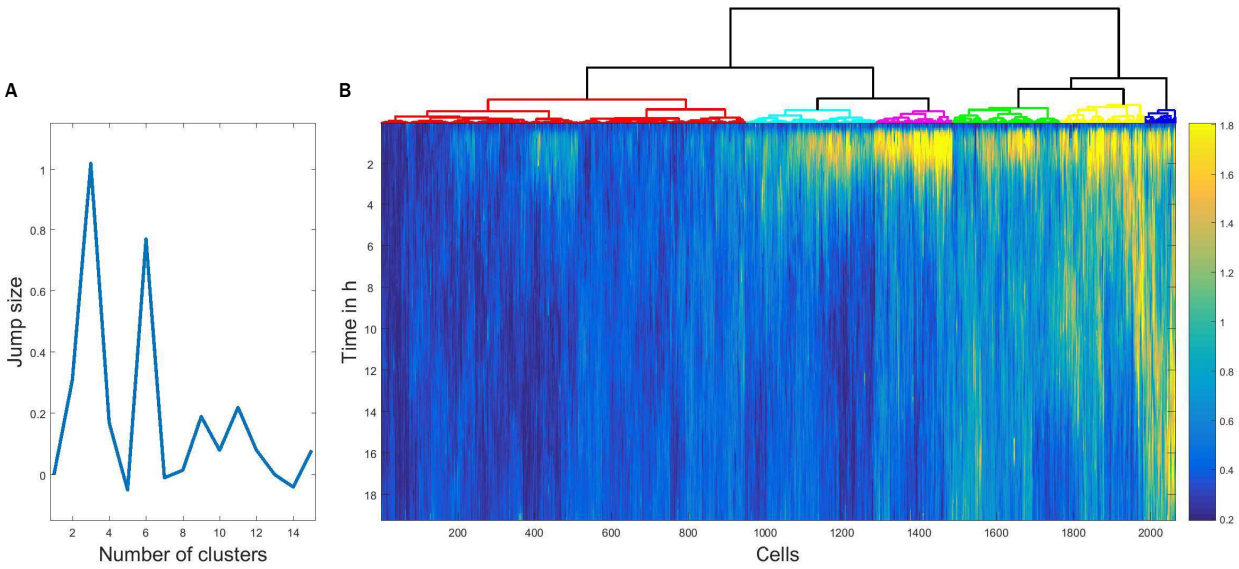

Fig. 11.5. SMAD2 dynamics induced by varying TGFβ can be classified based on pairwise cDTW dissimilarity estimates

A – To determine a good choice for the number of clusters the Jump method can be used. For different cluster numbers, the jump size is calculated using sum of square errors as a measure of intra cluster dispersion. Jump size reaches maxima at 3 and 6 clusters.
B - Heatmap of single cell time courses sorted according to hierarchical clustering. The corresponding dendrogram is shown on top. The colour code is linked to the individual clusters (red - cluster 1, red - cluster 2, cyan - cluster 3, purple - cluster 3, green - cluster 4, yellow - cluster 5, blue - cluster 6).

to biologically relevant boundaries. The concept of comparing signalling dynamics utilizing cDTW assumes that temporal shifts and stretches of dynamical signalling patterns found in the trajectories have to be compensated. The classification of single cell trajectories relies on the calculation of a dissimilarity matrix (Fig. 11.6 (C)) that contains all the pairwise cDTW dissimilarity measurements among nuc/cyt ratio trajectories for the cells (Fig. 11.3 (A-F)) stimulated with different extracellular concentration of TGFβ.
Using the pairwise cDTW similarity measurements hierarchical clustering based on the 'Ward' method [354] was applied as described in section 8.1. The resulting dendrogram, together with the accordingly sorted heatmap of the data, is shown in Fig. 11.5. Six signalling response clusters could be identified. The six clusters can be characterizing roughly by three features. Cluster 1 contains cells that show no or only a minimal response, Cluster 2 and 3 cells are characterized by a transient response and cells from cluster 3, 4 and 5 exhibit sustained signalling dynamics. The number of the clusters is assigned according to the sorted value of the average cumulative nuc/cyt SMAD2 ratio. Clusters that show similar characteristics differ in length and level of the initial and late response phase.

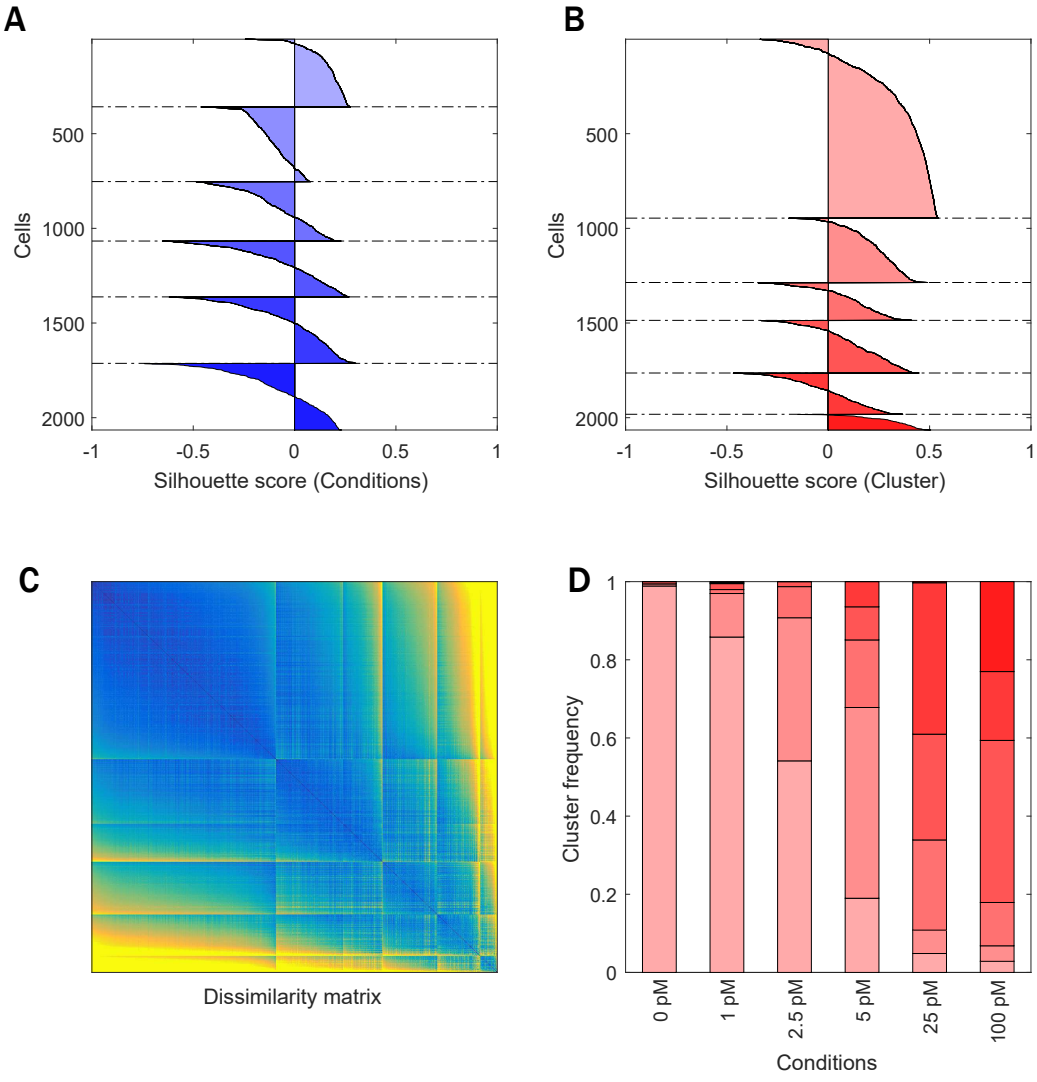

Fig. 11.6. Cells that share the same signalling classes are more similar than cells stimulated with the same concentration of TGFβ

A and B - Silhouette plots of cells sorted according to TGFβ concentration (B) or cluster(C) provide a graphical representation of the classification based on the cDTW similarity score. Positive silhouette scores indicate that SMAD2 responses are more similar to the own group, while negative scores signify that the corresponding trajectory is closer to any of the other groups. In general, signalling classes provide better classification than sorting to the level of stimulus.
C - Dissimilarity matrix calculated pairwise by cDTW of single cell trajectories treated with varying TGFβ doses. The matrix is sorted with according to dendrogram shown in Fig. 11.5 (B).
D - Distributions of signalling classes depending on TGFβ dose.

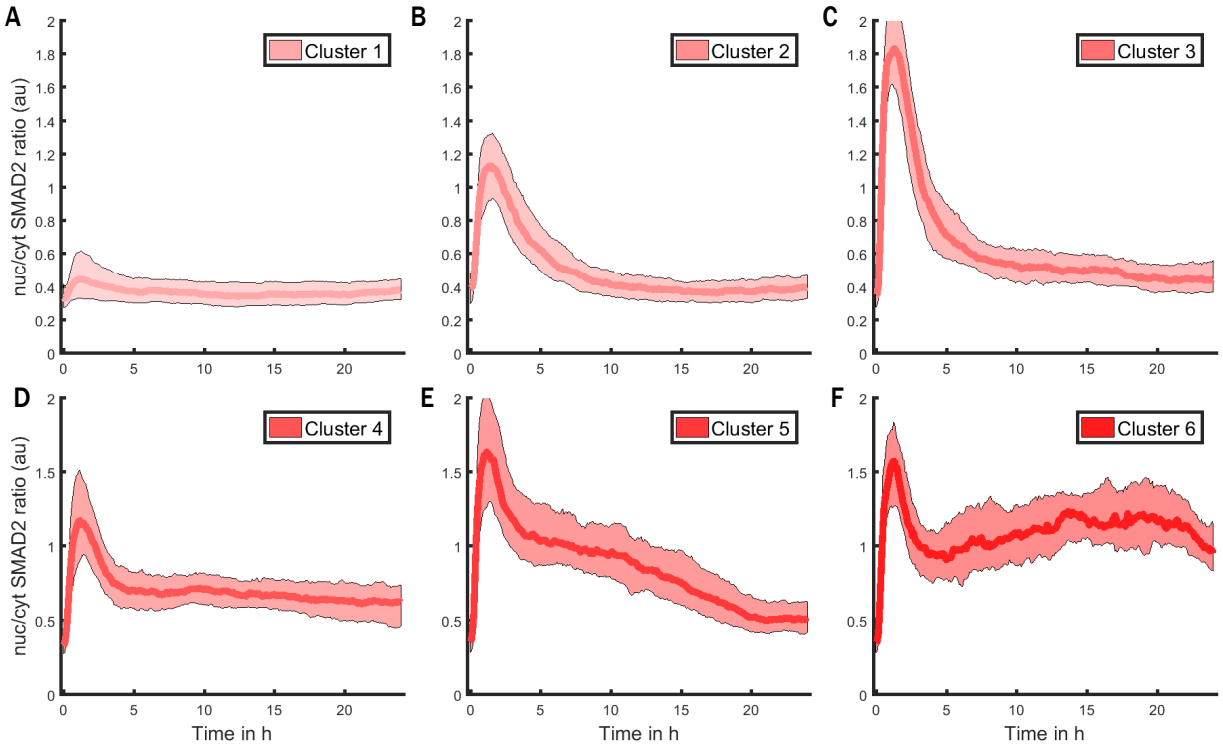

Fig. 11.7. SMAD dynamics observed after stimulation with varying doses of TGFβ decompose into distinct signalling classes.

A-F - Individual cells were clustered into six signalling classes according to their time-resolved nuc/cyt SMAD2 ratio using cDTW. Each line represents the median over all cells of the indicated cluster. Cells stimulated with varying TGFβ concentrations as indicated in the previous figure were included in the analysis. The shaded are represents the 25% to 75% quantile range. (Total – 2065 cells, Cluster 1 - 946 cells, Cluster 2 - 340 cells, Cluster 3 - 200 cells, Cluster 4 - 278 cells, Cluster 5 - 218 cells, Cluster 6 - 83 cells).

It could be shown, with respect to the cDTW scores, that the partitioning of the data into the six signalling classes groups the observed dynamics better than the strength of the applied TGFβ stimulus by the calculation of the silhouette coefficients (Fig. 11.6 A,B). The silhouette coefficient [287] indicates how similar an object is to objects of its own cluster compared to objects from other clusters. The value of the coefficient ranges from minus 1 to 1. The closer the value is to 1 the more similar the element is to the other elements of its own cluster, while lower values indicate the opposite. The silhouette coefficients provide a visual opportunity to inspect the quality of the clustering. The silhouette coefficients were estimated on the pairwise cDTW similarity measurements.
The decision for six clusters was based on the results of the 'jump method' (8.2), that is based on measuring the average distance, per dimension, between each observation and its closest cluster center [327]. The results (Fig. 11.5 (A)) of

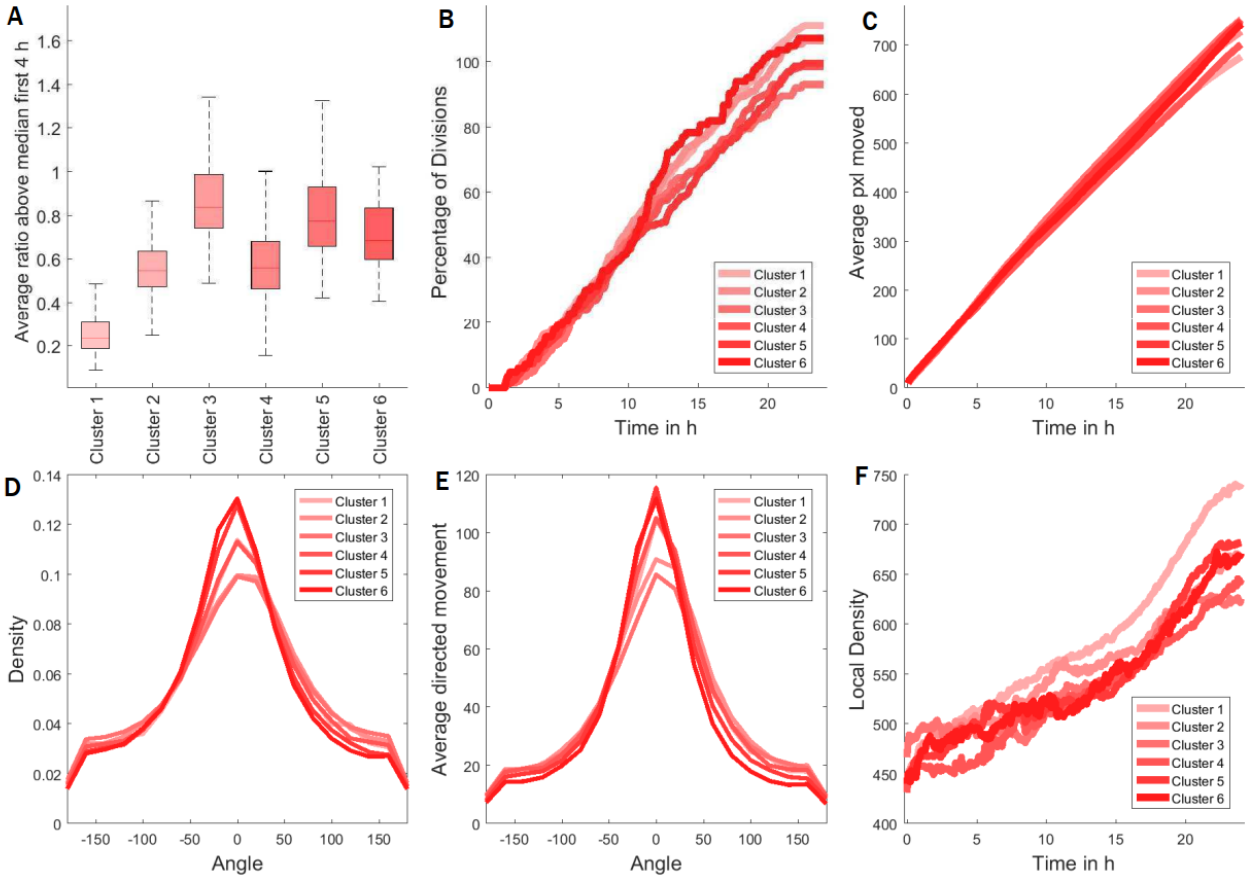

Fig. 11.8. Relation of the identified classes of signalling dynamics and phenotypic responses.

A - Average nuc/cyt ratio of values greater than the median within the first 4 h with respect to the six signalling classes.
B - Cumulative probability of the occurrence of a cell division event with respect to the six signalling classes.
C - Cumulative average distance moved (pxl) over the course of the experiment with respect to the six signalling classes.
D - Change of direction with respect to the six signalling classes.
E - Movement distance in relation to direction of the movements with respect to the six signalling classes.
F - Density of cells over the course of the 24 h of the experiment with respect to the six signalling classes.

the 'jump' method suggested a classification either into three or six clusters. Six clusters were chosen, to gain a more precise picture of the underlying signalling dynamics of interest. Three clusters would have split the dynamics into the signalling classes non-, transient- and sustained-responders.
In Fig. 11.7 (A-F) the median time courses, together with the inter-quantile range, of the identified response classes of signalling dynamics are shown. The

estimated signalling classes exhibit distinct dynamic patterns. With increasing level of the extracellular ligand the distribution of the signalling classes shifts from non-responders to more transient and at higher doses primary towards sustained dynamics.
The clusters represent signalling classes that are not evenly distributed over the different conditions (Fig. 11.6 (D)). Hence, observed pathway activity dynamics form a continuum of patterns, where the transition from one cluster to another cluster is smooth. For example, cluster 2 and 4 differ mainly in the level of the plateau after the initial response. While the nuc/cyt SMAD2 ratio of cluster 4 cells remains elevated it returns to basal levels in cluster 2 cells. The overlap of inter-quantile ranges obtained for the clusters is much lower than obtained for the conditions. Also the signalling dynamics are much more distinct compared to the conditions. For example, the cluster 3 allows the conclusion that a very strong first response leads to no sustained later dynamics.
When comparing the phenotypic readouts (Fig. 11.8) according to the signalling classes with results for the conditions (Fig. 11.4), the relation of the phenotypic features and the signalling classes does not produce a clearer result. Minor trends in motility could be observed so that cells from lower cluster id move a little less and a little less directed, but no direct effects on proliferation (Fig. 11.3 (B-F)) could be observed. That confirms the finding that the stimulation dose of TGFβ does not have a strong impact on the studied phenotypic responses. Neither the strength of the stimulus nor the observed dynamics trigger the outcome.
It's not surprising that greatest difference is observed on the level of the dynamics. For example the strength of the direct response (Fig. 11.8 (A)). The distribution for non-responders from cluster 1 does overlap less than the same readout for untreated cells and cells treated with low doses of TGFβ.
The initial question, if dynamics dictate the phenotypic response, remains unclear. Maybe because the time window of 24 h studied was too short to observe a change in the phenotypic characteristics or other factors like the state of the cell play a role. The fraction of cells that were assigned to signalling classes is correlated to the dose of TGFβ applied. The distribution of the signalling classes according to a fixed stimulus mirrors the observed heterogeneity. The signalling classes describe the spectrum of dynamics within the studied population.

11.3 SMAD dynamics can be decomposed into fundamental sub-patterns

While analysing single cell trajectories, one is in general interested in the occurrence and characteristics of certain temporal dynamic features like pulses. These patterns can differ in their length and amplitude but do exhibit a similarity in the shape. What kind of features are expected to occur in nuc/cyt SMAD2 ratio time series, that are connected to TGFβ signalling dynamics?

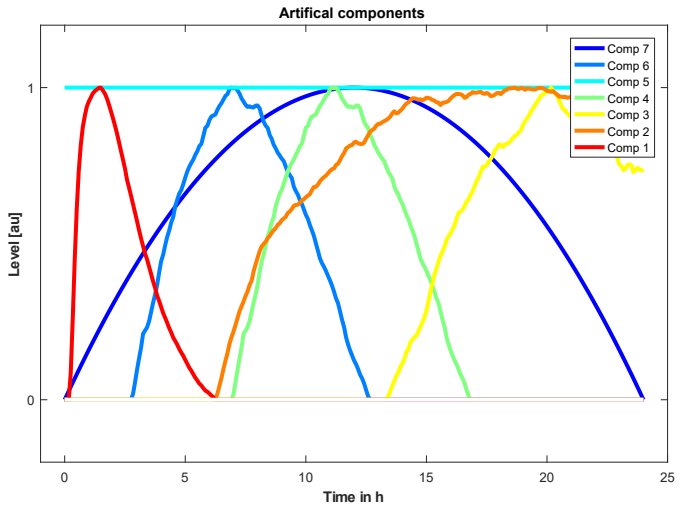

Fig. 11.9. Knowledge based initial components Seven constructed global dynamics that aim to describe the data partially.

The SMAD2 dynamics are mostly composed of an initial first response and later responses that are characterized either by stochastic fluctuations or an elevated sustained plateaus of vary length. As SMAD2 dynamics do not exhibit uniquely shaped repetitive patterns, the point of view is changed to a global perspective, that incorporates the whole length of the trajectories. The resulting of features will be called fundamental components in the following. A feature or fundamental component is from this point of view a trajectory, that contributes to a certain degree to a distinct time series. A linear combination of these fundamental components will be able to reproduce the whole original time series with only a small error. To this end, the framework introduced in section 10.2 will be applied. To study the signalling dynamics of a whole population of cells fundamental components will be identified in the time series in a supervised fashion. The fundamental components give a reduction in the data dimensionality and grand an alternative view on the observed signalling dynamics.

The supervised fundamental component framework will be applied on the same TGFβ titration dataset D of SMAD2 dynamics used in the previous section. Initially seven patterns or components C were constructed, based on the experience gained while working with single cell SMAD2 signalling dynamics (Fig. 11.9). The initial artificial components characterize different distinct features, that are expected to describe the observed dynamics in the data well. For example, the first initial artificial component characterizes the direct response, the second a long sustained response after the initial pulse, the third an influx of SMAD2 into the nucleus in the last 10 h, the forth an elevated level of SMAD in the middle of the experimental course and so on. With the initial hand-made components not even 30 % of the titration dataset could be reproduced. The initial components C were fitted iteratively so that, the difference $D-((C \setminus D')'*C')$ is minimized, where the operator $\setminus$ used solves the linear equation $Cx = D$. After the fitting the yielded fundamental components finally reproduce more than 99 % of the experimental data. The fitted fundamental components are shown in Fig. 11.10 (A).

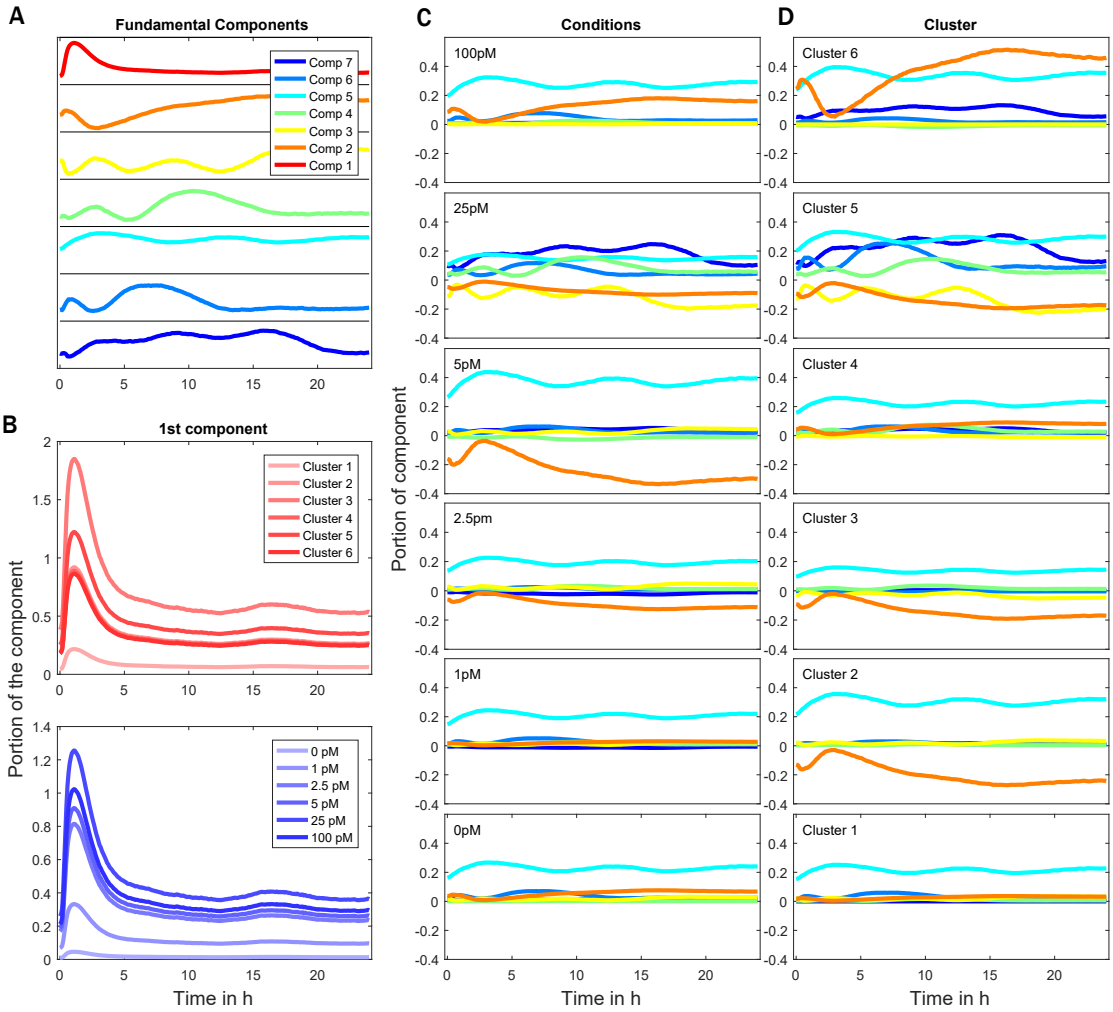

Fig. 11.10. Fundamental temporal components quantify the observed SMAD2 dynamics

A - Fitted fundamental temporal components of the titration data set.
B - Proportion of the first component to the total observed ratio within each of the distinct sub-population of cells. At the bottom, with respect to varying concentrations of TGFβ. At the top, with respect to the classification. The proportion is determined much more if one looks at the signalling classes.
C - Portion of the other components to the total observed ration within the sub-population of cells with respect to varying concentrations of TGFβ.
D - Portion of the other components to the total observed ration within the sub-population of cells with respect to the classification. The different components have a more distinct influence on the clusters than on the conditions.

In Fig. 11.10 the portion each of the fundamental components contributes on average to the cells treated with a certain level of ligand or to cells that belong to one of the distinct signalling classes is shown. How much each fundamental

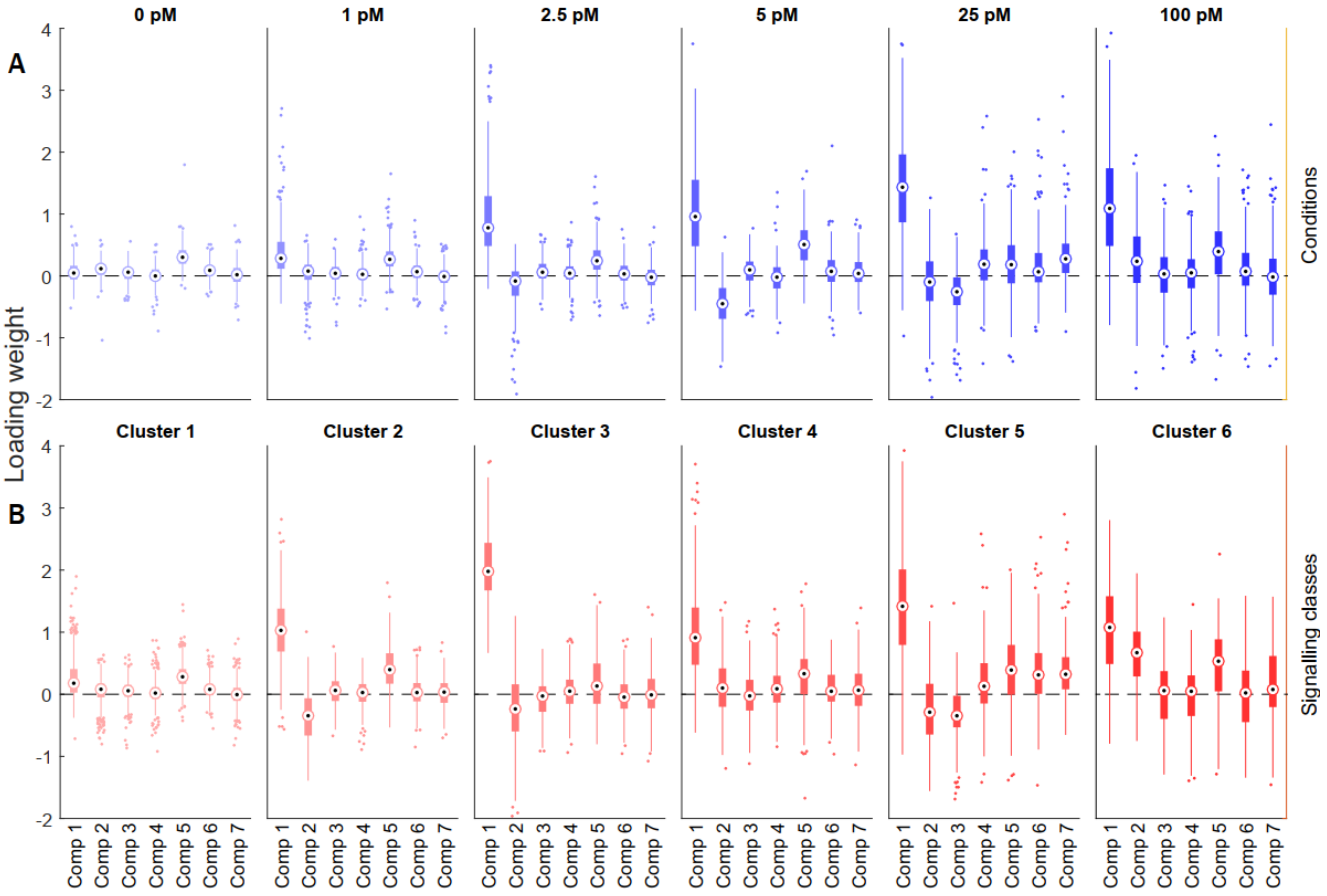

Fig. 11.11. Distribution of loadings of the fundamental components obtained for single cell trajectories with respect to condition or signalling class.

A - The distribution of the component loading with respect to the condition. The heterogeneity rises with increasing stimulation strength.
B - The loading of the components with respect to the signalling classes. The range of the weights is greater within the components of a cluster than compared to the conditions.

component contributes to a trajectory is encoded in the loadings of each of the components. Not surprisingly, the distributions of loadings of the different fundamental components are more heterogeneous among the signalling classes than among cells that came from different conditions. This could be expected because the classification of signalling dynamics as well as the fundamental components aim to emphasize distinct dynamics within the time series.
For the first component we see the clearest trend. In the range from 0 pM to 25 pM the higher the dose the more the first component contributes to the observed dynamics (Fig. 11.10 (B)). The trend observed for the conditions, is due to the changing mixture of cells from signalling classes in the different conditions. At the level of signalling classes four average levels of weights with a greater spread (Fig. 11.10 (C)) could be found for the first component. In cluster two, three and

six the first component contributes equally to the dynamics. The second components, that is responsible for the late sustained response, contributes mostly to cells from the sixth cluster. Else the second components act as counterpart to the third component that plays mainly the role in the late plateau level (prominent in the 5 pM TGFβ condition and in cells from cluster two) or the two components work synergistically as for 100pM or cluster six. The other components contribute mainly to the length and shape of the late response phase.
While investigating the loadings with respect to the conditions or clusters clear trends in their weights could be quantified (Fig. 11.11). The clusters differ in general to a greater extend in their setup of the loadings associated to the fundamental component that shape to the signalling dynamics.
The fundamental components give a new tool to access how dynamics are composed and generated by simpler dynamics underneath. As a prove of concept it was shown in section 10.2, that the subset of the cells stimulated with 100 pM TGFβ was sufficient to create a fit that explains around 97 % of the complete titration dataset. The loadings that determine this contribution can be used to map new data sets onto the six identified signalling classes. This feature will be used in the following sections. This mapping can be done with the freedom of allocating new data with different temporal resolution or different length onto an existing classification. It is only necessary to have an overlapping window of measurements within the data acquisition.

11.4 Dynamics of SMAD signalling drive phenotypic cellular features with temporal delay

Are the effects of TGFβ signalling on proliferation and motility more prominent at later time points compared to the previously studied 24 h dataset? Does the cell cycle state or the local cell density influence the dynamics observed? The second question also addresses two potential sources of heterogeneity in signalling dynamics: the cell cycle and state local density [204, 314]. To this end, two additional microscopy experiments are included in this study. A 60 h long experiment with a resolution of 15 minutes and the stimulus set at the beginning of the experiment and an experiment of 36 h length with a temporal resolution of 10 minutes where the stimulus was set after the first 24 h have passed.
The results of the supervised learning method from the previous section are used to map the six signalling classes identified so far onto the new data. This approach uses the dynamics that made up a response to map a given classification onto new data based on overlapping observation times.
In the previous section, the concept of fundamental components was applied to reduce the dimensionality of the data. Now the previously identified signalling classes are mapped onto the new datasets with the identified fundamental components based on estimates of the corresponding loadings for each of the trajectories

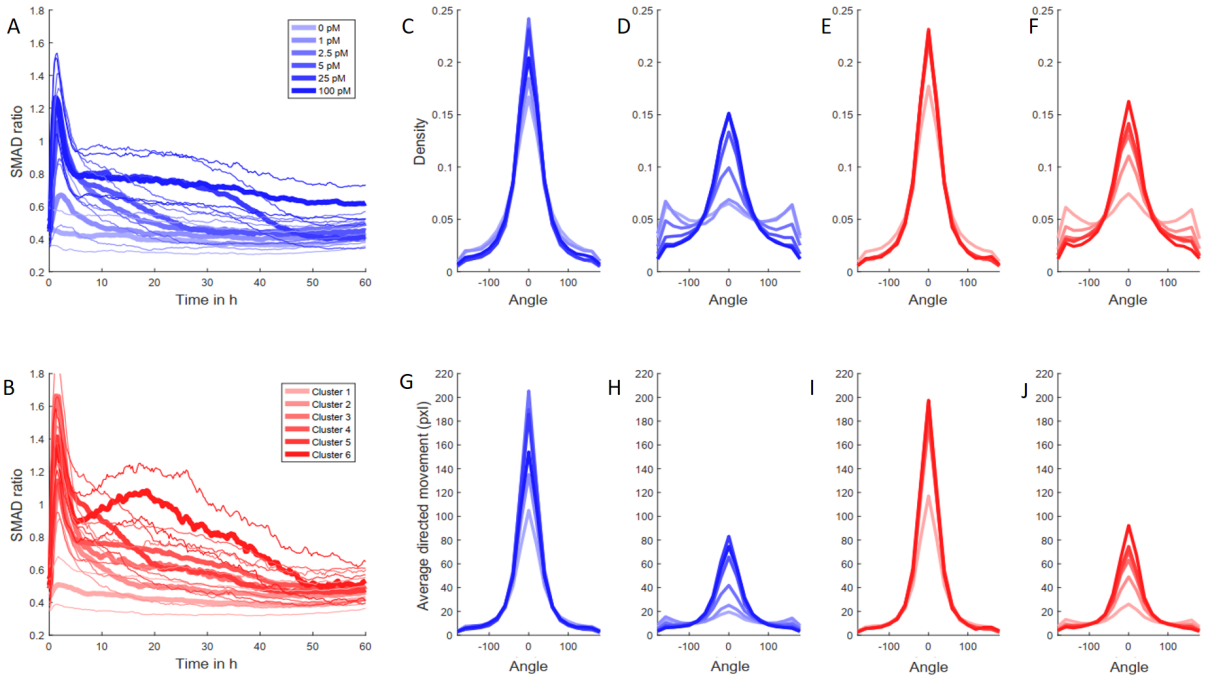

Fig. 11.12. Effect of TGFβ dynamics on the cellular motility enfolds at later time points

A - Cells were imaged for 60 h with varying TGFβ concentrations (5935 cells, control - 1442 cells, 1pM TGFβ 948 cells, 2.5pM TGFβ 776 cells, 5pM TGFβ 813 cells, 25pM TGFβ 1024 cells, 100pM TGFβ 932 cells). The median and the 25% and 75% quantile nuc/cyt SMAD2 ratios for the conditions are shown.
B - Mapping of SMAD2 trans-location dynamics of the cells in individual cells shown in (A) to previously identified signalling classes (see Fig. 13.3). For each trajectory, the most similar signalling class was determined using the loadings of the fundamental temporal components. The median and the 25% and 75% quantile nuc/cyt SMAD2 ratios for resulting mapped sub-populations (Cluster 1 - 2493 cells, Cluster 2 - 875 cells, Cluster 3 - 327 cells, Cluster 4 - 1579 cells, Cluster 5 - 460 cells, Cluster 6 - 201 cells) are shown.
C, D - Change of direction with respect to different TGFβ doses for the first 30 hours (C) and the hours 31 to 60 (D) of the experiment.
E, F - Change of direction with respect to the six signalling classes for the first 30 hours (E) and the hours 31 to 60 (F) of the experiment.
G, H - Average movement in pxl according to the change of direction with respect to different TGFβ doses for the first 30 hours (G) and the hours 31 to 60 (H) of the experiment.
I, J - Average movement in pxl according to the change of direction with respect to the six signalling classes for the first 30 hours (I) and the hours 31 to 60 (J) of the experiment.

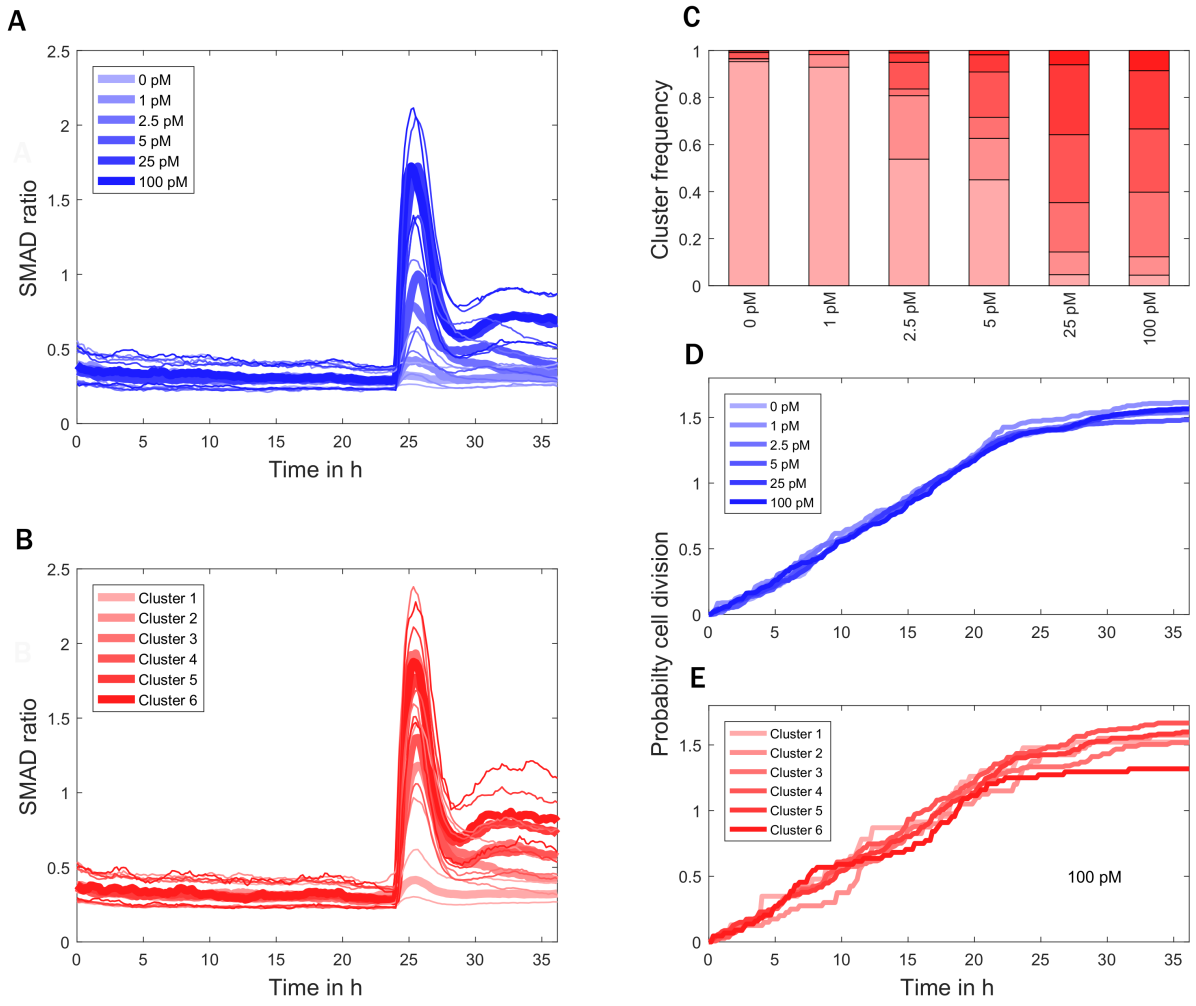

Fig. 11.13. Connection of the SMAD2 dynamics to the cell cycle

A - Cells were imaged 24h before stimulation with varying TGFβ concentrations. The effect on the SMAD2 dynamics was monitored for additional 12 h (3575 cells, control - 671 cells, 1 pM TGFβ1 - 282 cells, 2.5 pM TGFβ1 - 697 cells, 5 pM TGFβ1 - 602 cells, 25 pM TGFβ1 - 810 cells, 100 pM TGFβ1 - 513 cells). The median and the 25% and 75% quantile nuc/cyt SMAD2 ratios for the conditions are shown.
B - Mapping of SMAD2 trans-location dynamics of the cells in individual cells shown in (A) to previously identified signalling classes (see Fig. 13.3). For each trajectory, the most similar signalling class was determined using the loadings of the fundamental components. The median and the 25% and 75% quantile nuc/cyt SMAD2 ratios for resulting mapped sub-populations (cluster 1 - 1608 cells, cluster 2 - 435 cells, cluster 3 - 386 cells, cluster 4 - 589 cells, cluster 5 - 445 cells, cluster 6 - 112 cells) are shown.
C - Distributions of signalling classes shown in (B) in dependence to the applied TGFβ dose for the cells shown in (A).
D - Cumulative probability of the occurrence of a cell division event with respect to varying concentrations of TGFβ.
E - Cumulative probability of the occurrence of a cell division event with respect to the six signalling classes.

in the new datasets. The trajectories of the new dataset are mapped onto the signalling class with the setup of loadings that reflects the loadings of the new trajectory the most.
Special attention was paid to the different temporal resolutions among the experiments. The length of the fundamental components identified in section 11.3 must also be kept in mind. To adjust different temporal resolution levels spline interpolation was applied. For the 60 h dataset, only the first 24 h of the dataset were used. For the 36 h dataset only the first 12 h from the fundamental components were aligned to the last 12 h of the dataset.
The distribution of the signalling classes estimated for the new datasets with respect to the applied stimulus does reproduce the distribution found in the initial 24 h titration dataset. This is shown for the 36 h dataset in Fig. 11.13 (C). To avoid redundancy this is not shown for the 60 h experiment but the numbers are given in the caption of Fig. 11.12.
In the 60 h long experiment the main interest was, if a stronger effect on the motility at later time points could be detected. The results are shown in Fig. 11.12. As expected the effects on the motility were lower for the first 24 h af-

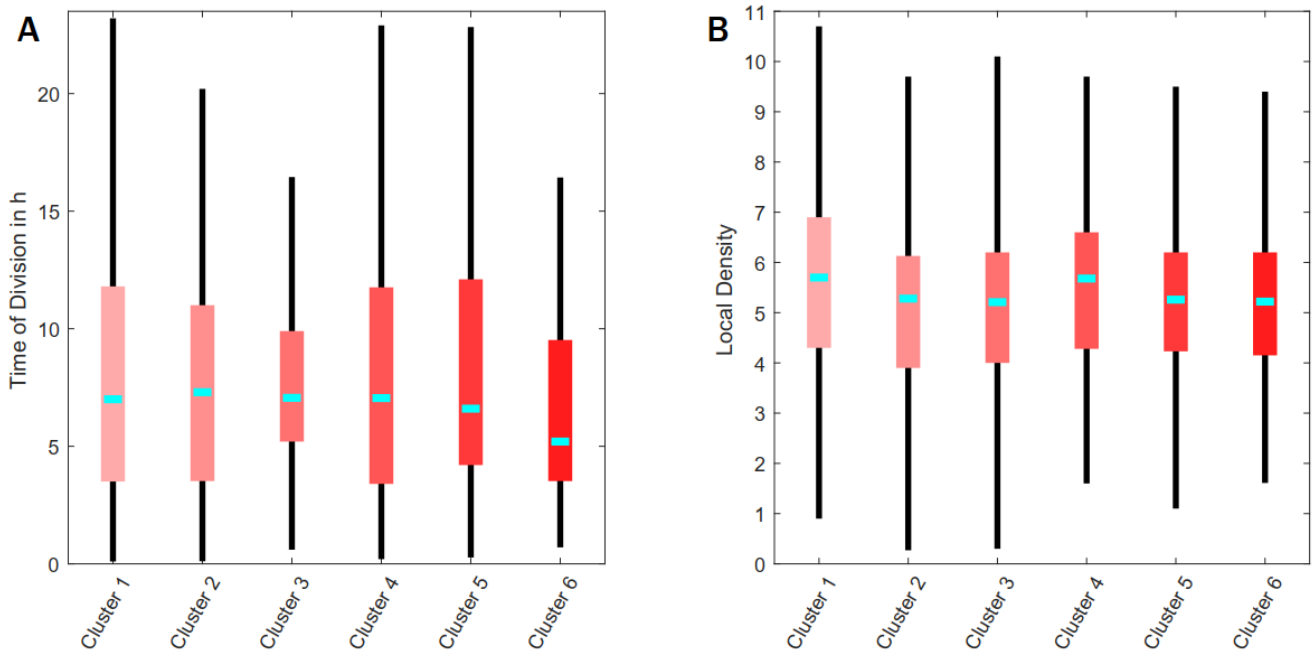

Fig. 11.14. Effect of the cell cycle and the local density before the stimulation

A - Time passed since the last cell division before stimulus for each signalling class. Distributions are overlapping. There is no significant effect of cell division time is observable.
B - Cell density before stimulus for the signalling classes. Density scores represent a weighted sum of all neighbouring cells within 640 μm distance. Distributions are overlapping. There is no significant effect of cell division time is observable.

ter stimulation, than for the second half of the experiment. Over the whole time course the motility drops, likely due to an increase in the local cell density. But, a clear difference among cells treated with different doses or even stronger among cells that were assigned to different signalling classes could be detected at later time points. While investigating the dose effect on the motility a clear correlation between how fast and how directed cells move could be detected. Cells stimulated with low doses move less and more undirected at later time points compared to cells from higher TGFβ doses. The picture is similar but simpler for the signalling classes. Non-responders move slow and undirected while responders move faster and directed. This changes the perspective on the dose dependence. Changes in the motility correlated to the ligand dose are most likely connected to the decreasing number of non-responders with increasing stimulus strength. The difference among the signalling classes is less strong than the difference among the cells with respect to the applied dose of TGFβ. Both transient and sustained SMAD signalling classes have similar effects on the motility. For the late motility it is sufficient to distinguish between responders and non-responders and the detailed shape of the dynamics plays a less important role.

In the 36 hour dataset (11.13(A)), cells were monitored before and after the stimulus to answer two questions. The impact of the cell division and the local cell density on the observed signalling dynamics is studied in this dataset. The time passed since the last division before the stimulus was used as estimate for the cell cycle state. Cells that have divided shortly before the stimulus was set are likely in G1. The cell cycle stage does not have any detectable influence on the signalling dynamics 11.14 (A). No relation could be found between the time of division before the stimulation and the mapping of the new trajectories onto the six signalling classes (11.13(B,C)) . Neither the responsiveness nor the strength of the response could be connected to the local cell density or the cell cycle state before the stimulus is applied. The same is true for the local cell density. The local cell density measures how crowded the neighbourhood of a cell is with other cells. While investigating the impact of the local cell density onto the formation of the signalling classes 11.14 (B) there could also no relation be detected. At least for the signalling classes an effect on the proliferation could be found (11.13(D,E)). Cells from the sixth cluster nearly stopped proliferating after the TGFβ stimulation.

In general, the signalling classes give a better correlation to some of the phenotypic readouts compared to the TGFβ doses. Cellular behaviour is partially more driven by the dynamics of the signalling than from the strength of the applied stimulus. The signalling classes explain to some extend the heterogeneity in the phenotypic responses to a fixed stimulus.

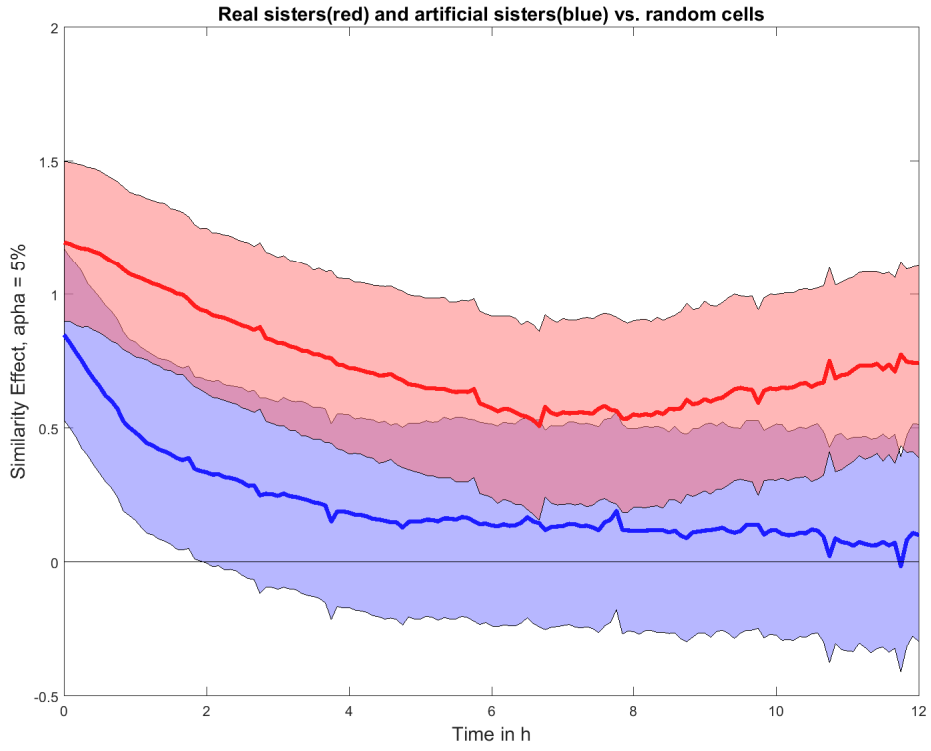

Fig. 11.15. Sister cells analysis

Analysis of SMAD2 trans-location dynamics in sister cells. SMAD2 trans-location dynamics in sister cells after division and unrelated cell pairs with the same nuc/cyt SMAD2 ratio were compared using cDTW. Resulting similarity scores were aligned in time and compared to those from randomly selected cell pairs. Effect size (solid lines) and 95% confidence intervals (shaded areas) were estimated by bootstrapping. The analysis shows that recently divided cells are more similar than control cell pairs and remain correlated over time, indicating that heterogeneity arise from differences in cellular state.

11.5 Dynamics of SMAD signalling are determined by the state of the individual cell

How is the information encoded in the TGFβ dose processed by a cell, so that a population of cells build-up by different fractions of entities from the distinct signalling classes is generated? As potential sources of heterogeneity the cell cycle state, local density and variable cellular protein composition were identified in previous studies[204, 314] . The analysis of the 36 h dataset showed that the cell cycle state before the stimulus as well as the local density could not be correlated with the assignment of trajectories to the signalling classes (Fig. 11.14 and 11.13). Neither the cell cycle state nor the local density can predict the signalling class or fate of a cell.

One remaining origin of the heterogeneity observed is the internal protein composition of the cell. This aspect cannot be studied directly from the phenotypic read-outs of the data. The use of sister cell analysis has proven to be a helpful tool to analyse characteristics of signalling pathways [112, 319, 293]. The assumption is that sister cells are more similar in their molecular setup than a randomly picked pair of cells. Hence, to investigate the influence of the variable protein composition of cells, recently divided cells were compared with randomly picked cells. The approach will show if the observed heterogeneity is due to stochastic fluctuations or based on the internal state of the cell, defined by the cellular molecular setup. Therefore, it could be assumed that heterogeneity is linked to the cellular state, if it could be shown that the signalling dynamics of sister cells are more similar over a longer period of time compared to dynamics of randomly selected cells.
To make the analysis more sustained in addition 'artificial sister cells' are introduced. As 'artificial sister cell' cells, cells were labelled that coincidentally shared the same nuc/cyt SMAD2 ratio at a certain time point ('artificial divisions'). These set was used to exclude that the observed effect for sister cells is not just due to the same ratio at the moment of division.
To measure how sister cells, differ from 'artificial sister cells' and randomly picked cells, again the DTW framework is applied. The similarity of SMAD2 dynamics was estimated for a longer period of time after the cell division or a 'artificial divisions' was obtained in the pair of cells analysed.
To compare the dynamics of sisters with other cells, data from 11 different experimental replicates of cells stimulated with 100 pM TGFβ was combined. In all the experiments the cells were imaged for 24 h with a temporal resolution of 5 minutes. In total, an initial population of about 6000 cells was used, giving rise to 13.000 cells after 24 h as the result of about 7000 recorded cell division events. DTW was applied within a sliding window of 4 h length. By sliding the window along the time axis a series of similarity measurements s was generated staring from the time of division i by $s(j-i+1) = dtw(t_1(j:j+48), t_2(j:j+48))\ for\ j = i:i+n$. By t_1 and t_2 a pair of sister cell trajectories is identified.
To answer the initial question, sisters and 'artificial sisters' were compared with randomly selected cells from the non-sister sub population. Bootstrapping was used to estimate the effect size and the corresponding confidence intervals. The results are shown in Fig. 11.15. This method yields an estimate on the divergence of sister cells in a time-dependent manner.
The results show much more correlated and similar response dynamics among sister cells, while the similarity between the control 'artificial sister cells' drop very fast within 4 h to levels undistinguishable from randomly picked cells. The similarity between sisters decreases much slower for the first 6 h and remains significantly elevated over the time period of 12 h that has been investigated by shifting the sliding window. Similar temporal patterns were proposed in previous studies [112, 319].
The results indicate that a primary source of heterogeneity of the signalling response dynamic is linked to the cellular state and is not intrinsically unpredictable

and stochastic. The cellular state drives the signalling dynamics to a certain degree.

11.6 Negative feedback regulator SMAD7 knock-out explains heterogenity partially

Sister cell analysis has shown that TGFβ signalling dynamics are driven by the molecular setup of the cell. Hence, heterogeneity in SMAD signalling arises from varying levels of signalling proteins. In cellular signalling pathways it is assumed that dynamics are as well driven by the availability of the extracellular ligand that undergoes steady decay and the activity of transcriptional negative feedback loops [326]. To investigate how the negative feedback loop modulates the signalling dynamics, the effect of SMAD7 gene knock-out is studied. SMAD7 is considered to be one of the main feedback regulators of TGFβ-induced signalling and acts at the level of TGFβ receptors [239]. The analysis of the impact of the SMAD7 knock-out can confirm the hypothesis that variable cellular protein composition is the main source of heterogeneity detected in the analysed data of TGFβ signalling dynamics. Will there be a shift in the signalling dynamics, if the negative feedback in the reporter cell system is turned off?
The hypothesis was tested in SMAD2 reporter cells using Cas9-mediated gene knock-out of the prominent feedback regulator of the TGFβ pathway SMAD7. With these cell line, the initial titration experiment was repeated for the parental and the SMAD7 knock-out. The results are shown in Fig. 11.16 and Fig. 11.17. The dynamics of the parental cell line (Fig. 11.16 (A)) differ to some extend from the previously used titration dataset. Using the fundamental components, the signalling classes have been mapped onto the datasets introduced in this section as in section 11.4. The distribution off the signalling classes estimated in the dataset of the parental cell line (Fig. 11.17 (A)) does not to a full extent reproduce what was initially found (Fig. 11.6 (D)). But the trend of the fraction of the distinct signalling classes with respect to the stimulation strength could be repeated. The titration data of the knock-out cell line exhibits on the population level a stronger signalling (Fig. 11.16 (B)). Especially if the level of the sustained late response is compared with what was obtained in the cells there SMAD7 is normally expressed. If the mapping of the signalling classes onto the parental and the SMAD7 knock-out is compared the distribution of the signalling classes, with respect to the dose of the ligand, was shifted. With the decreasing strength of the negative feedback the signalling class composition within the population shifts toward those classes that exhibit a higher responsiveness, stronger signalling and towards more sustained dynamics line (Fig. 11.17 (B)). While transient signalling could still be observed, the existence of additional negative feedback loops can be assumed due to the presence of redundant transcriptional feedback regulators in TGFβ signalling [357]. Hence, negative feedback within the TGFβ signalling

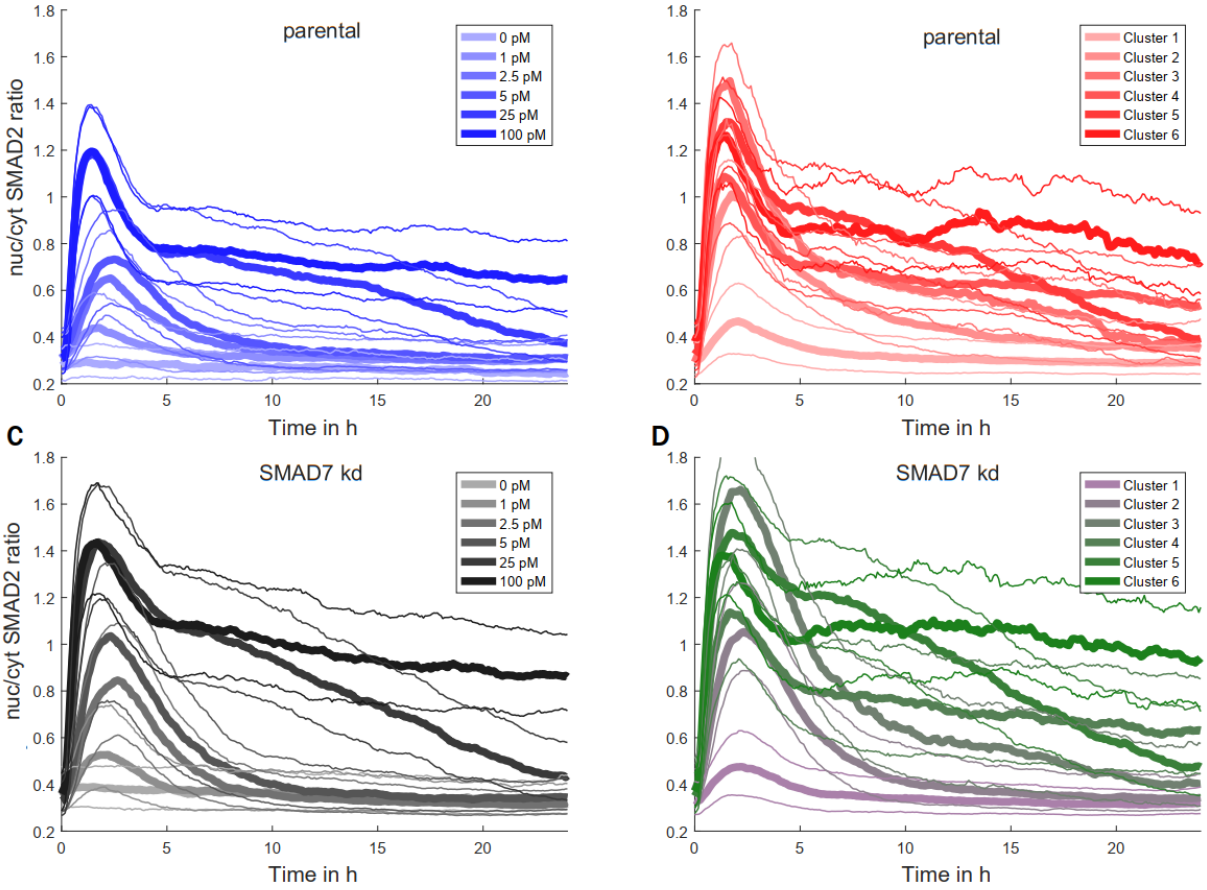

Fig. 11.16. Negative feedback triggered by SMAD7 influence the signalling dynamics.

A - Median nuc/cyt SMAD2 ratio of parental cells of the SMAD7 knock-out cell line stimulated with the indicated concentrations of TGFβ (3571 cells, control - 266 cells, 1 pM - 743 cells, 2.5 pM - 655 cells, 5 pM - 568 cells, 25 pM - 686 cells, 100 pM - 653 cells).
B - The parental cells of the SMAD7 knock-out cell shown in (A) are mapped to the signalling classes using the fundamental temporal components. (cluster 1 - 1803 cells, cluster 2 - 524 cells, cluster 3 - 133 cells, cluster 4 - 676 cells, cluster 5 - 300 cells, cluster 6 - 135 cells).
C - Median nuc/cyt SMAD2 ratio of the SMAD7 knock-out cell line stimulated with the indicated concentrations of TGFβ (3998 cells, control - 475 cells, 1 pM - 670 cells, 2.5 pM - 821 cells, 5 pM - 614 cells, 25 pM - 771 cells, 100 pM - 647 cells).
D - Distribution of signalling classes in SMAD7 knock-out cells (cluster 1 - 1520 cells, cluster 2 - 796 cells, cluster 3 - 357 cells, cluster 4 - 505 cells, cluster 5 - 630 cells, cluster 6 - 190 cells).

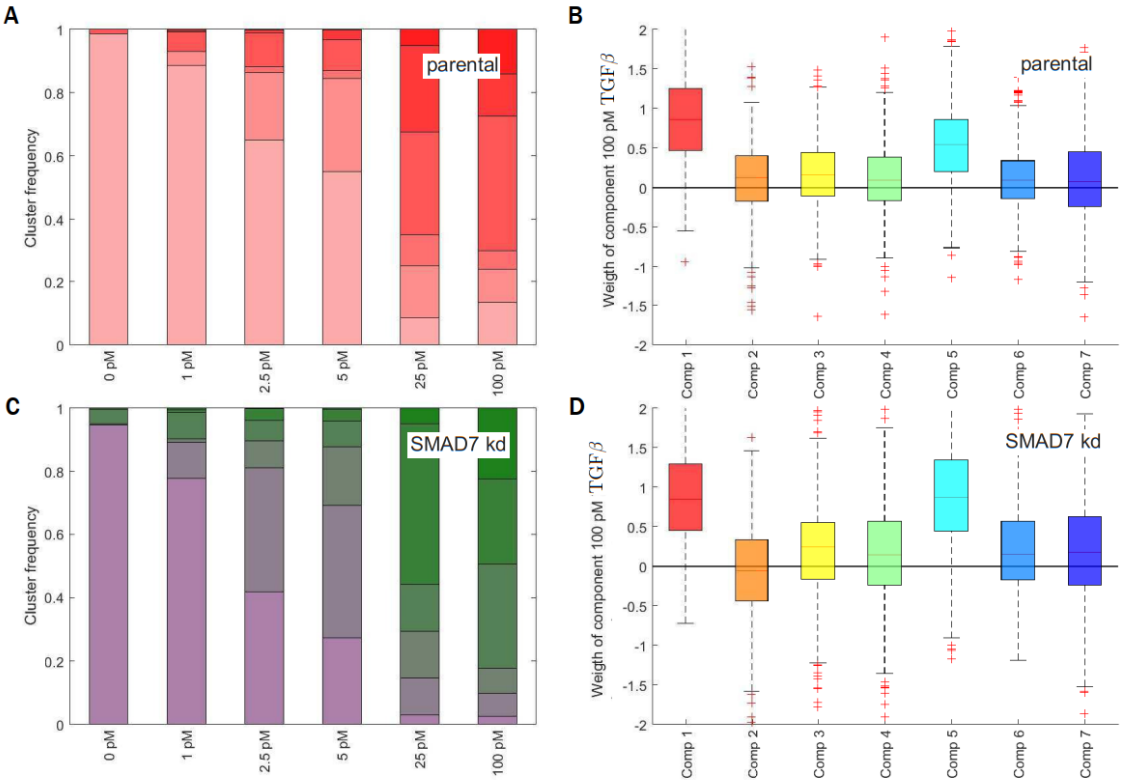

Fig. 11.17. Negative feedback triggered by SMAD7 shifts composition of signalling classes and is a cause for the heterogeneity observed.

A - Proportion of signalling classes among the different conditions in parental cell line of the SMAD7 knock-out cells.
B - Distribution of the loadings of the fundamental temporal components for cells from the parental cell line for 100 pM.
C - Proportion of signalling classes among the different conditions of the SMAD7 knock-out cells.
D - Distribution of the loadings of the fundamental temporal components for cells from the SMAD7 knock out for 100 pM. The SMAD7 knock-out cells differ significantly in the proportion that the 5th component. The 5th component contributes to the late sustained late response.

pathway is not only carried out by SMAD7.
Also it was compared how the loadings of the fundamental components within a population of cells stimulated with the dose changes due to the SMAD7 knock out. The dynamics observed in SMAD7 wild type and the SMAD7 knock out cells are altered predominantly in the 5th component. The 5th component contributes notably to the sustained response. Surprisingly, the loadings of the other fundamental components are only affected weakly. This is shown in Fig. 11.17 (C,D) for the stimulation with a dose of 100 pM TGFβ.

The CRISPR/Cas9-mediated knock-out of SMAD7 indicates that a major part of the observed heterogeneity, can be attributed to fluctuations in feedback proteins. The SMAD7 knock-out specifically affected the signal duration and shifts the observed dynamics towards more sustained signalling. The findings confirm our hypothesis that cell-to-cell variability in signalling protein concentrations is the main source of the observed heterogeneity in the signalling dynamics.

Part V

P53 dynamics are sensitive to changes in temperature

12

Temperature modulates p53 dynamics upon genotoxic stress

Time-lapse microscopy is used to monitor p53 and p21 in a live-cell reporter system introduced in section 4.1. The live-cell reporter system expresses fusion proteins of p53 and p21 with fluorescent proteins from endogenous gene loci [304], over time in individual cells.
All the single cell data presented in this chapter was acquired by Petra Snyder at the TU Darmstadt.

12.1 Reduction of variance among different experimental replicates

A common problem when working with time-lapse microscopy data is bringing together different experiments that were acquired on different microscopes, by different experimentalists, on different days. Combining such diverse experiments to analyse them quantitatively and qualitatively is a non-trivial task. The distributions of the fluorescence intensity measured differ even among identical experimental settings. The sources of the observed variability have multiple origins ranging from biological sources such as different passage numbers of the cells monitored to technical sources such as different light sources. As it is aimed to investigate the effects of varying environmental conditions, it is necessary to remove the unwanted variability caused by other biological and technical sources. In this section, a normalization method is proposed that reduces the inter-experimental deviation by fitting a set of experiments to one reference experiment. The method works similar to quantile normalization, a technique for making two distributions identical in statistical properties. But, in contrast to quantile normalization the measured distributions are mapped onto one reference distribution. The central assumption of the normalization algorithm is trivial. If a subset of measurements is given for all experiments with identical experimental settings, it can be assumed that they all have a similar distribution of the measurement. In this scenario, the observed deviations of the measurements can be directly linked to sources not related to the biological heterogeneity. Therefore, the heterogeneity observed is directly linked to non-biological variation in the data acquisition procedure.

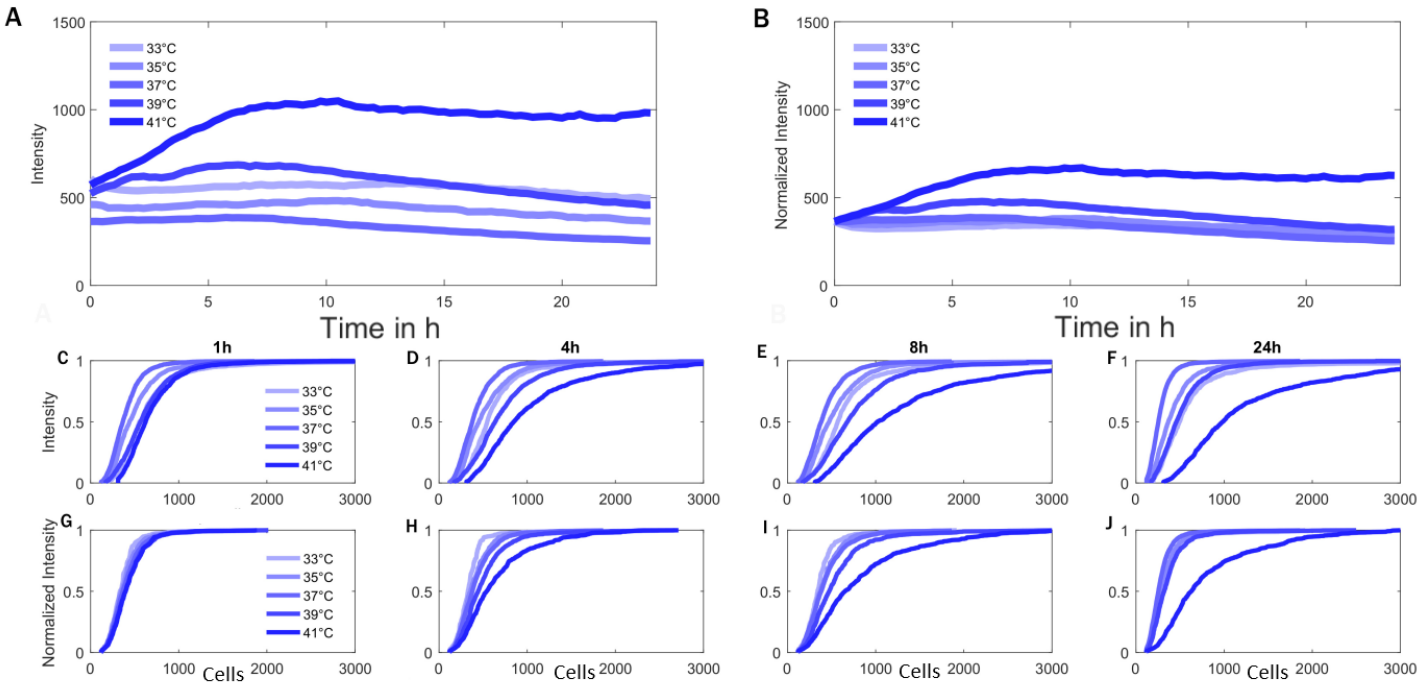

Fig. 12.1. Normalization with respect to identical experimental settings at a certain time point

A - Median trajectories of the nuclear p53 level for cells growing at different temperatures. The temperature was changed at the beginning of the experiment. The change in the temperature was roughly 1°C per hour. Each temperature setting was performed in a distinct experiment. From this setting one could expect that all measures at the beginning of the experiment should be equally distributed. Hence, the differences among the measures at the first time point are experimental noise. It can be observed, especially for 33°C, 35°C and 37°C, that the initial error propagates thought the whole time course of the experiment.
B - Median trajectories of the normalized cells shown in panel A. The normalization is done with respect to the first time point of the experiments.
C-F - Distribution of the p53 fluorescence measures at different time points (1h, 4h 8h and 24h). We cannot exclude that the source of the heterogeneity, observed among the different experiments, is connected to the experimental effects. Therefore, it cannot be concluded that the different p53 levels originate from biological sources modulated by temperature.
G-J - After normalization with respect to the first time point of measurement it can be expected that the experimental technical error effects are removed from the observations. The normalization removes a large proportion of the detected differences in the level of p53 among the different experiments.

In the setup of this study, at the beginning of each experiment the nuclear p53 level is measured in identically treated cells. At the first time point there will always be non-irradiated cells incubated at 37°C. The differences detected among ex-

periments should therefore only be to a minor extent related to biological sources. In the example, presented in this section, an equal number cells in different experiments is assumed. This simplification will not hold in the scenarios faced in the real world. Hence, usually the procedure requires additional interpolation or extrapolation steps. Interpolation or extrapolation is done using splines.

Example of normalization procedure:
Let E_{ijk} be a measurement in a set of experiments, for the ith experiment, at the jth time point and for the kth cell. All experiments have the same experimental biological setup at first time point $j = 1$. The xth experiment is set as the reference. The aim is to fit another experiment y onto x so that the effect of non-biological error sources is reduced. The integral part of the method is the initial estimation of coefficients of different weight. The calculation starts with computing the coefficient by $weight = sort(E_{x,1,:})./sort(E_{y,1,:})$. These weights are used in in the further steps to fit the experiment y onto x by applying the coefficient to each distinct time point.
The pseudocode of the method is shown in the text field below.

```
weight = sort(E_{x,1,:})./sort(E_{y,1,:})
for k = 1 : NrOftimePoints
    [iFjF] = sort(E(y, k, :))
      % iF sorted values and jF the corresponding order so that
      % E_{y,k,:} equals iF_{jF}
    tmp = weight. * iF
    fittedExp(k, :) = tmp(jF)
      % the order jF is used to reorder tmp so that the mapping of
      % remains identical between the original and the normalized data
end
```

The idea is that the non-biological error is constant over the whole time curse of the experiment. Therefore, this error can be estimated at the first time point where all cells from the different experiments have received the same treatment. The differences observed do not originate from biological sources. In Fig. 12.1 results computed by the proposed method are shown. In the example shown five different experiments are given. At the beginning of all different experiments the cells face the same experimental condition. After the start the temperature was set to different temperatures in the range from 35°C to 41°C. While studding the raw data (In Fig. 12.1 (A,C-F)) a huge difference among the different experiments could be detected. The difference is already strong at the first time point (Fig. 12.1(C)). Hence the differences in the measurements at the first time point cannot be related to the any biological difference in the experimental setup. Therefore, it is not clear to which extent the later differences are related to the different treatment. After we applied the proposed normalization method the heterogeneity among different experiments is reduced drastically (In Fig. 12.1 (A,C-F)).

After the non-biological effects were removed only cells treated with 41°C differ significant from the other conditions. The temporal dynamics and differences encoded in the strength of the response are conserved after normalization among the different experimental. This method ensures that, one can directly compare normalized measures of the p53 level within the analysed cell populations among diverse experimental settings monitored in different experiments.

12.2 P53 dynamics upon genotoxic stress are temperature sensitive

To investigate, if temperature has an impact on p53 signalling dynamics, we equilibrated cells to temperature in the range from 33°C to 41°C and monitored the p53 dynamics on the single cell level for 24 h at for non-irradiated and irradiated cells by time-resolved live-cell microscopy (Fig.12.2 (A-E)). The irradiation dose was set to 10Gy X-ray (250kV). The data acquired in the different experiment where normalized by the method presented in the previous section.
In untreated cells no differences were detected in p53 dynamics among cells equilibrated at temperatures below and at normotherm 37°C. With increasing temperature, one could observe an effect on the p53 dynamics, even in the absence of DNA damage induced by radiation. At 39°C a minor increase in the p53 level could be detected. Surprisingly at 41°C p53 accumulates significantly. Within the first 6 h of the experiments the nuclear p53 level increases constantly and remains elevated for the complete time course of the experiment (Fig. 12.2 (E)). On average the p53 level accumulates in non-irradiated cells at 41°C to around 60 % of the first observed canonical pulse, detected as response to radiation at 37°C.
Cells irradiated with 10Gy exhibit a much more complex picture. Pulse duration and pulse frequency are correlated with temperature in the range from 33°C to 39°C. The lower the temperature the slower pulses appear (Fig. 12.2 (A-D)). Pulses were asynchronous and heterogeneous across the population equilibrated at different temperatures. Therefore, pulses appear damped in the median p53 level in the different cell populations irradiated and equilibrated between 33°C to 39°C. At 41°C a surprising non-canonical response as detected for the non-irradiated cells was observed. Instead of pulsatile dynamics, a strong accumulation within the first 6 hours was detected ensued by a slow decrease over the following hours monitored in the experiment. The p53 accumulates much stronger than what had been observed at other temperatures and reaches levels around 1.5 fold than the average amplitude of the first pulse at 37°C. After 24h the p53 level is similar in irradiated and non-irradiated cells grown at 41°C.
The result so far is challenging, especially at 41°C, because even if the cells were equilibrated to the temperature before the experiment was started, a direct temperature response at the beginning of the experiments could be observed. This could as well be observed in a strong increase in the heterogeneity, directly after

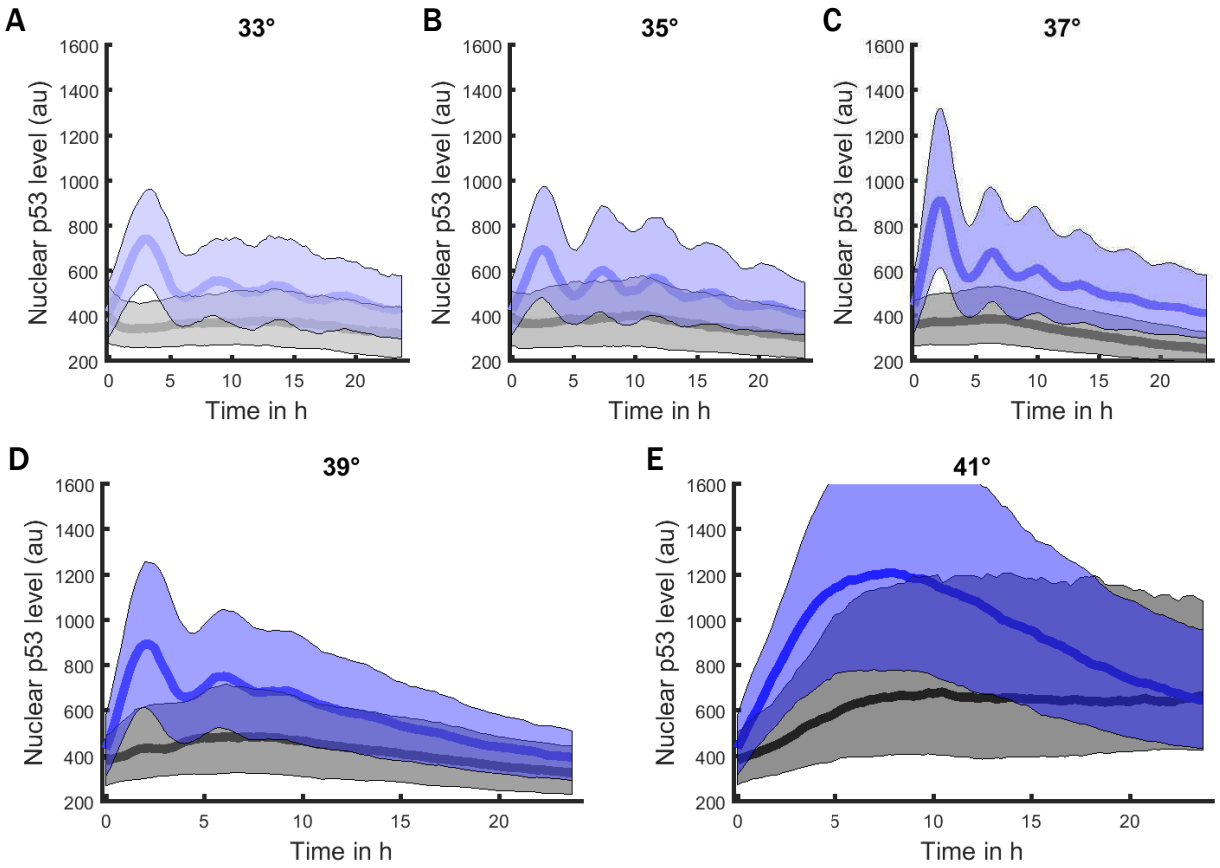

Fig. 12.2. p53 dynamics are modulated by temperature (all cells there normalized to the 37°C condition)

A-E - Cells pre-equilibrated were monitored for 24 h in a temperature range from 33°C (A) to 41°C (E). The incubation chamber of the microscope had a temperature of 37°C. The temperature switch was initiated at the begin of the experiment inducing a gradual temperature change over time. The cell population shown in blue was irradiated with 10 Gy at the begin of the experiment. Non-irradiated cells are shown in grey. Bold lines indicate median p53 levels, shaded areas reflect the inter-quantile ranges from 0.25 to 0.75. To our surprise at 41°C the canonical pulsatile p53 dynamics are replaced by a sustained accumulation response. High temperature alone is sufficient to induce a p53 response even in the absence of radiation. The number of shown cells in the order of the temperature was for non-irradiated 1214, 1759, 3820, 2436, 1139 and for irradiated 782, 1098, 2857, 1630, 1007.

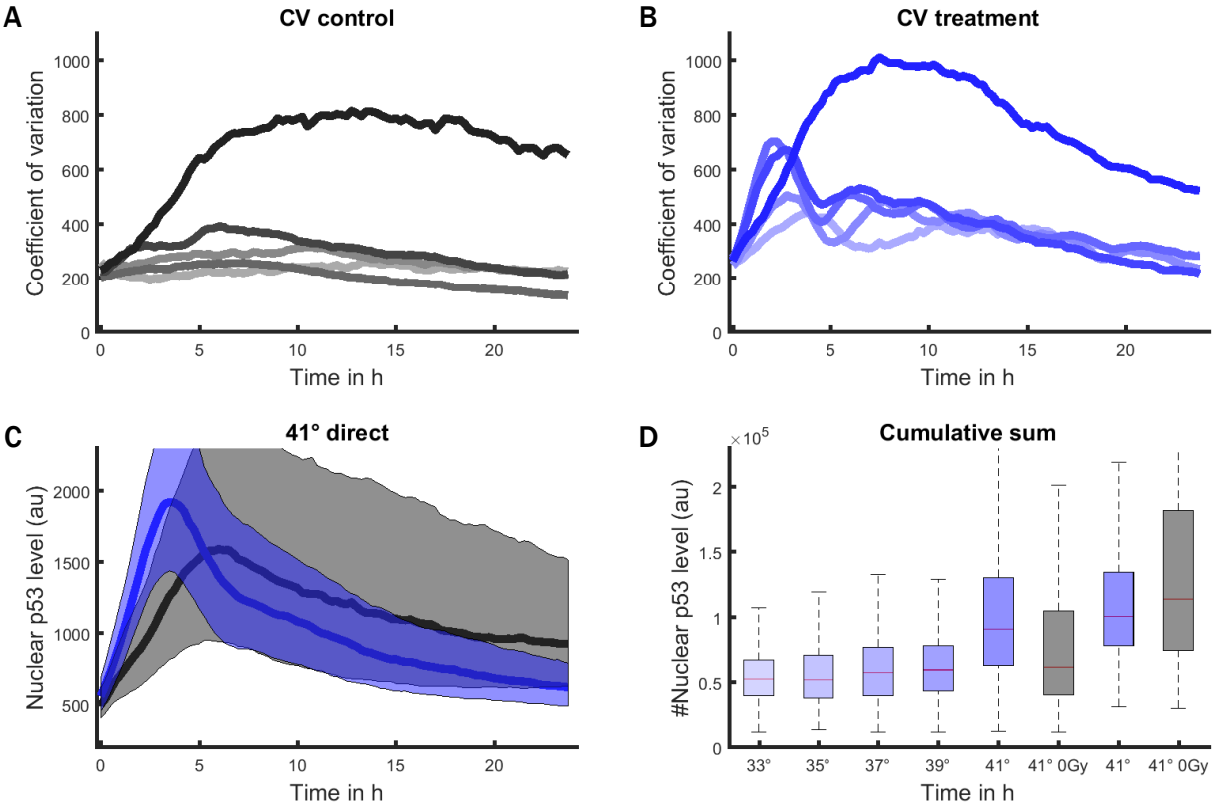

Fig. 12.3. p53 dynamics are modulated by temperature (all cells there normalized to the 37°C condition)

A - Coefficients of variation over time for control cell populations shown in 12.2(A-E). At 41°C a strong increase in the heterogeneity of the measurements could be detected.
B - Coefficients of variation over time for cell populations irradiated with 10Gy x-rays shown in 12.2(A-E). At 41°C a strong increase in the heterogeneity of the measurements could be detected. At the other temperatures monitored the heterogeneity measured can be linked to the asynchrony in radiation induced p53 pulsing.
C - p53 response of control and irradiated cell populations acutely shifted to 41°C without prior pre-equilibration. A stronger and in the dynamics slightly shifted p53 response compared to cells pre-equilibrated at 41°C 12.2(E) could be observed. Data shown for 721 non-irradiated and 994 irradiated cells.
D - Boxplot indicating the cumulative p53 level for the single cell trajectories. A minor positive correlation in the temperature range from 33°C to 39°C could be detected. The strongest effect is estimated for 41°C, where a sustained response was obtained for control and irradiated cells. Surprisingly the most heterogeneous and on average strongest response was observed for non-irradiated cells pre-equilibrated at 41°C 12.2(C).

the start of the experiment shown in Fig. 12.3 (A,B). This raises the question, why does this direct response after the start of the experiment, even in pre-equilibrated cells, occur.

Therefore, it was investigated how cells react to an acute temperature change from 37°C to 41°C (Fig. 12.3 (C)). The p53 response observed while exposed to a direct acute shift in temperature differs substantially from what was monitored for cells that were pre-equilibrated. In general, the response was much stronger compared to cells pre-equilibrated. In non-irradiated cells p53 accumulates during the first 8 h and decreases afterwards slowly. The p53 level reached is on average much higher than what had been observed in pre-equilibrated cells. Radiation in combination with acute temperature shift induced an even stronger immediate response. p53 accumulates fast, within the first 4 h, to levels twice of what have been quantified in cells pre-equilibrated. Interestingly, this accumulation is less long-lasting then in the non-irradiated 41°C experiment. After 10 h the p53 level is on average only half of the average amplitude of the direct response and after 24 h a fraction of cells has again reached the initial p53 level. Another interesting observation made in the acute temperature shift experiment is, that the average response of non-irradiated cells is stronger than the response of irradiated after around 8 h where the average amplitude of the non-irradiated cells reaches its maximum. While the p53 decrease slower in non-irradiated cells than in irradiated cells, the average p53 level of non-irradiated cells remains higher than in irradiated cells for the rest of the experiment. Due to this observations, it can be assumed that the pre-equilibration reduces the sensitivity of the p53 dynamics to temperature changes.

To access the variability over time within each of the datasets the coefficient of variation (CV) (Fig. 12.3 (A,B)) was estimated. In non-irradiated cells a similar low level in the temperature range from 33°C to 39°C could be observed. The CV increases drastically at 41°C were the fast increase in the first 6 h to levels 4-fold of the other conditions was obtained and remained elevated for the whole course of the experiment. For irradiated cells in the range from 33°C to 39°C differences in the CV related to the induced p53 pulsing where detected. At higher temperatures the pulsing is faster and stronger and at lower slower and weaker as quantified in the temporal dynamics of the CV measurement, especially for the first pulse. At 41°C a faster and longer increase as for non-irradiated cells at the same temperature can be found. In contrast to the non-irradiated cells the CV starts to decrease after around 10 h. The induced sustained p53 response at 41°C has a drastic impact on the observed variability.

To complete the initial analysis, the integrated nuclear p53 level over the whole time course of the experiment (Fig. 12.3 (D)) was compared. In the temperature range from 33°C to 39°C the cumulative p53 level remains constant. The median integrated p53 levels of non-irradiated 41°C pre-equilibrated cells were similar to what we observed for damaged cells at lower temperature, but with wider distribution across the population. Irradiated cells at 41°C exhibit a substantially higher accumulated p53 level. Between pre-equilibrated and acute shift to 41°C

the accumulated response is comparable even if the dynamics are different. Surprisingly the accumulated level of non-irradiated cells is on average the strongest and exhibits the strongest heterogeneity.
The differences between cells pre-equilibrated at or directly exposed to 41°C point to a decreased sensitivity to hyperthermia of pre-equilibrated cells. The question, why a response in pre-equilibrated cells could be detected remains open. It might be that the short time frame necessary for the transfer from the incubator to the microscope cools down the cells and that the short temperature reduction stops ongoing p53 responses. This will be investigated this in the following sections in detail.

12.3 Hyperthermia has a direct effect on the p53 response

In this section, p53 dynamics are analysed in more detail to get a better understanding of the temperature dependence of p53 signalling dynamics induced by DNA damage at the single cell level. To change perspective, the auto-correlation between p53 levels in the population at two different time points (Fig. 12.4) was determined. Repetitive patterns in the temperature range from 33°C or 39°C could be detected as lines, areas of increased auto-correlation in (Fig. 12.4), parallel to the main diagonal. At 41°C the picture changed completely. Any kind of repetitiveness vanishes completely and is replaced by a sustained accumulation pf p53. At 41°C the level of p53 accumulation is correlated over a longer time period. This indicates that a strong immediate response lead to a strong p53 level at later time points and vice versa. Hence, the p53 level remains relatively stable after the initial increase at 41°C. In the range between 33°C and 39°C the pulsing frequency is positively correlated with temperature. In addition, it could be observed that

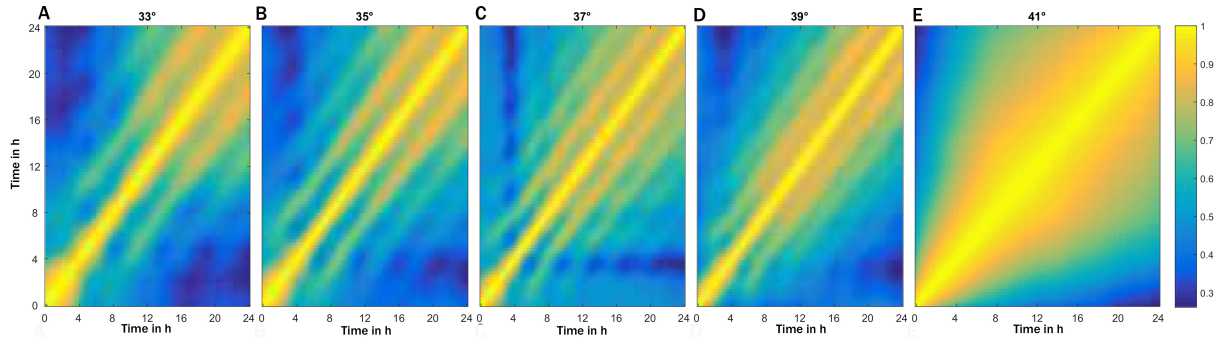

Fig. 12.4. Auto-correlation

A-E - Auto-correlation among cells irradiated with 10 Gy at the indicated temperatures. Each point in this graph represents the correlation coefficient between p53 levels in the population at indicated time points. Repetitive pulses could be detected in the temperature range from 33°C to 39°C. The pulsatile dynamical patterns vanish at 41°C and are replaced by a long monotone response.

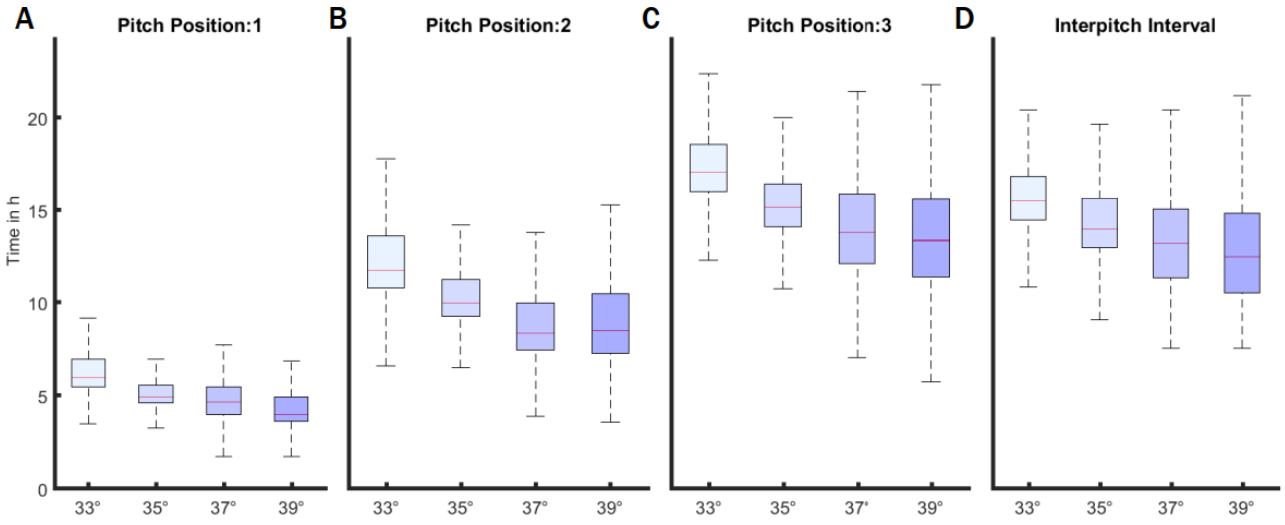

Fig. 12.5. Estimation of the pitch detection using the average magnitude difference function

A-C - The distribution of the first three peak positions estimated using the average magnitude difference function for cells irradiated with 10 Gy over a temperature range 33°C to 39°C. Red lines indicate medians of distributions, boxes include data between the 25th and 75th percentiles, whiskers extend to maximum values within 1.5 times the interquartile range. Cells pulse slower the lower the temperature at 33°C to 35°C compared to cells at 37°C or 39°C. The pulsing frequency at 37°C and 39°C is similar. D - The inter-peak interval estimated for cells irradiated with 10 Gy over a temperature range 33°C to 39°C based on the first five detected pitches. The inter-peak interval gets longer with reduced temperature while the range of the distribution gets wider with increasing temperature.

the parallel lines get much more distinct at later time points of the experimental course. At 39°C the lines parallel to the main diagonal spread out and the valleys between the lines become less clear shaped. This indicates that the pulsing frequency is more heterogeneous at 39°C.

To estimate the pitch of p53 pulses the Average Magnitude Difference Function (AMDF) proposed by Ross at al. in 1974 [285] was applied. AMDF is the commonly used method for pitch detection. Window lengths between 4.5h - 7h were used and a pitch period lower bound of 2h was assumed. For robustness the different pitch positions for the different window sizes were integrated. The mean over all window sizes was used as estimate for a certain pitch position.

The pitch positions for the first 3 pulses differed among the cell populations from 33°C to 39°C (Fig. 12.5 (A-C)). 41°C data was excluded due to the fact that no pulsing is expect. The difference of the normotherm population to cells equilibrated at lower temperatures is much stronger than the difference to the 39°C cell population. The colder the temperature the slower the pulsing. Pulsing at 39°C

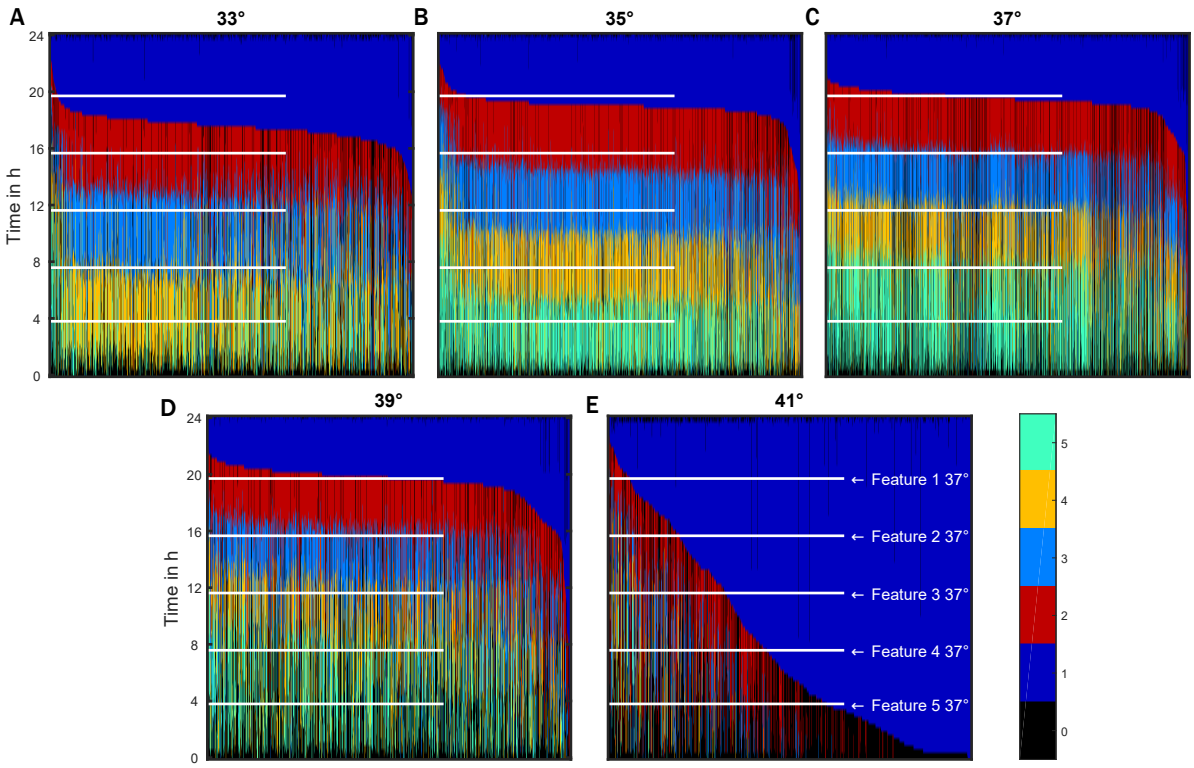

Fig. 12.6. Dynamic of p53 pulses is modulated by temperature

A-E - Visualization of the detected features upon damage induction for the temperature range from 33°C to 41°C. The different pulses are shown in different colour as indicated next to (E). The legend at the right side refers to which colour corresponds to the temporal order of the detected features. The white lines indicate the average end of the pulses at 37°C.

is only slightly faster than at 37°C. From a broader perspective the inter-pitch intervals (Fig. 12.5 (D)) were calculated. The inter-pitch intervals confirm the previous findings. Pulsing frequency is positively correlated with the temperature in the temperatures below 41°C and accompanied by an increase in the variability of the pitch position with increasing temperature.

The two methods applied at the begin of this section aim to find repetitive patterns by using the concept of auto-correlation. In the rest of this section will focus on the well described DNA damage induced canonical p53 pulses. The p53 pulses are essential in the encoding and processing of the DNA damage information in the DDR signalling network [180, 275]. The framework based on local DTW introduced in section 9 will be applied to detect and characterize the pulsing with respect to the temperature applied. This approach will gain deeper insight on the p53 dynamics and how temperature modulates the p53 DDR. Does temperature alter characteristics of the p53 pulses?

The local DTW feature detection method was applied to the dataset and the

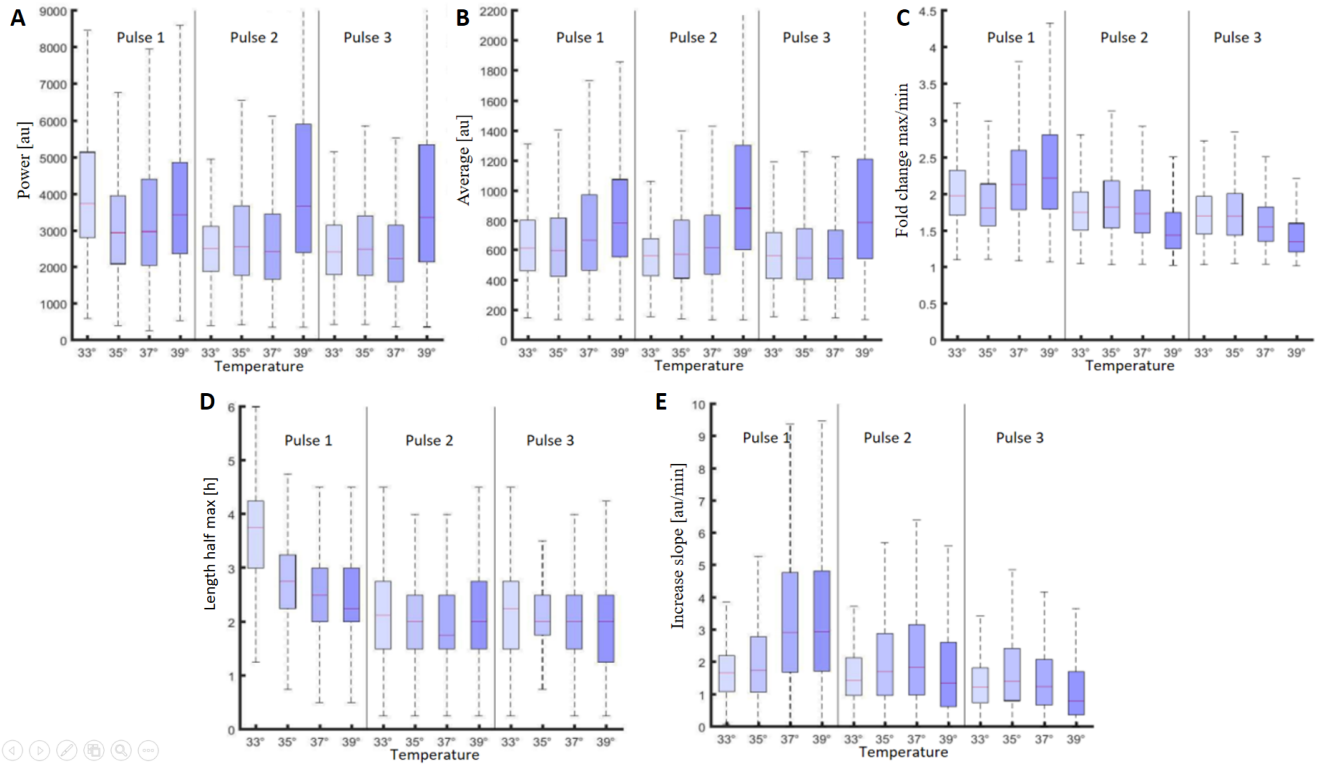

Fig. 12.7. Dynamic of p53 pulses is modulated by temperature

A - Power (integrated p53 level) for the first 3 pulses for the temperature range from 33°C to 39°C. The features of a pulse are shown in (Fig. 9.5(A)). Pulses are detected using a local dynamic time warping based approach.
B - Average p53 level for the first 3 pulses for the temperature range from 33°C to 39°C.
C - Fold change (maximum divided by minimum level of a pulse) for the first 3 features for the temperature range from 33°C to 39°C.
D - Length in hours that a pulse stays above the half-maximal amplitude of the first 3 pulses for the temperature range from 33°C to 39°C.
E - Increase slope for the first 3 pulses for the temperature range from 33°C to 39°C. The increase is measured by the change in the fluorescence intensity [au] per minute.

pulses shown in (Fig. 12.6 (A-E)) were identified. In this figure white lines as a reference to the 37°C cell population were added and the cells were sorted with respect to the length of the first detected feature. At 41°C the spread of the length of the first pulse ranges over the whole experimental course. At the other temperatures p53 pulsing was obtained in response to induced DNA damage. It could be observed that the bands formed by the colour coded pulses get broader with lowering the temperature. The opposite was detected at 39°C where only the first two pulses form a clear band in the plot and later pulses appear less homogeneous in their timing within the investigated cell population. The heterogeneity in this timing is positively correlated in the range from 33°C to 39°C.
The pulses are characterized by several features. These features are modulated by

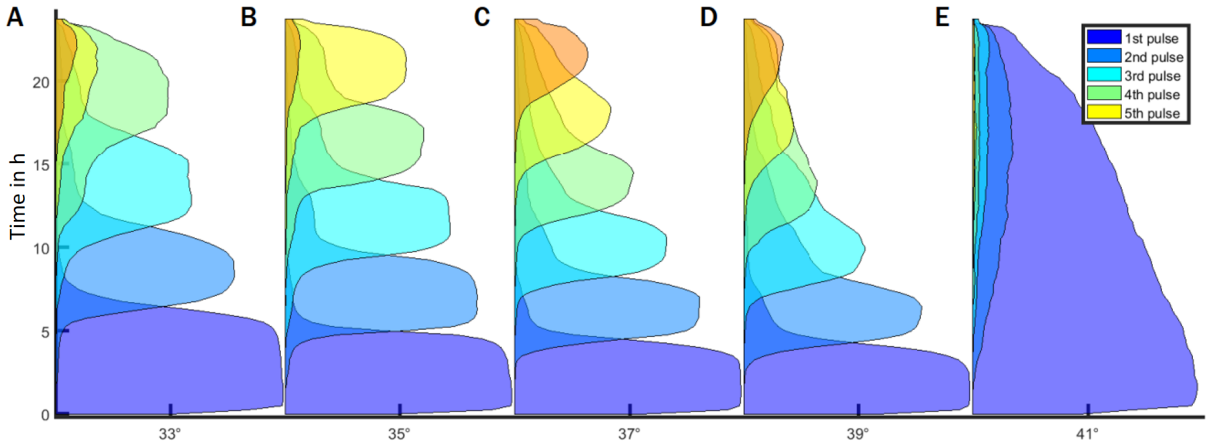

Fig. 12.8. Probability of a cell of being in a certain pulse

A-E - The probability of being in a certain pulse at a given time point for the temperature range from 33°C to 41°C. Each of the x-axes of the sub-plots has a range from 0 to 1. The pulsing frequency is positively correlated with the temperature within the range between 33°C to 39°C. At 41°C the pulsing dynamics is completely lost and replaced by a sustained response.

the temperature in the range from 33°C to 39°C. 41°C was excluded due to the exceptional different dynamic, that is not directly comparable with the pulsing observe at temperatures below. In Fig. 12.7(A-E) for the first tree detected pulses in a trajectory, the power, the average, the max-mix fold-change, the time a pulse is above half the amplitude and the slope of the increase are presented. For all the read-outs one could quantify substantial differences.
For the power, the integrated intensities of a pulse, the 2nd and the 3rd pulse at 39°C are stronger than for the other temperatures (Fig. 12.7(A)). For the average intensity of a pulse, a similar trend (Fig. 12.7(B)) was observed, with the exception of the first pulse at 33°C which has the hugest power due to the increased length. The average can be interpreted as the length adjusted power. The strength of the first pulse likely depends on the temperature, while later pulses at have similar strength for 33°C to 37°C. Later pulses at 39°C exhibit a higher average intensity compared to lower temperatures. This is accompanied by a reduced fold change (Fig. 12.7(C)) indicating a higher basal level. For the max-min fold-change higher values could be observed for the first pulse at higher temperature. The opposite was observed for the 2nd and 3rd pulse. This is due to the fact that the min value of later pulses is higher at higher temperatures. Next, the time a feature is above half the maximum amplitude of the pulse (Fig. 12.7(D)) is quantified. The differences in the length of pulses a mainly found in the first detected pulse, while the 2nd and 3rd pulse exhibit a similar length no matter the temperature. Finally, the increase slope of the pulses (Fig. 12.7(E)) is compared. The slope differs especially in the first pulse. The increase is much slower for cells from 33°C and 35°C compared to cells equilibrated to 37°C to 39°C.

Based on the estimated pulses, the probabilities for being in a certain feature at a given time point were calculated. The probabilities were estimated with respect to the populations monitored at different experimental conditions. Again temperature-dependent changes in timing from 33°C to 39°C (Fig. 12.8) were observed. This analysis confirmed changes in synchrony of features in a temperature dependent manner. The probability distributions overlap stronger with increasing temperature until the pulsing dynamics break down and are replaced by a strong sustained response at 41°C were canonical p53 pulses are not detectable.

12.4 P53 response is modulated directly by temperature

So far, it has been observed that temperature modulates p53 dynamics and hyperthermia alone is sufficient to induce a p53 response. The question why a response in pre-equilibrated cells could be detected remains still an open question. Therefore, it will be investigated how dynamic patterns of p53 might change if temperature is altered during an already ongoing response to stress induced by DNA damage.

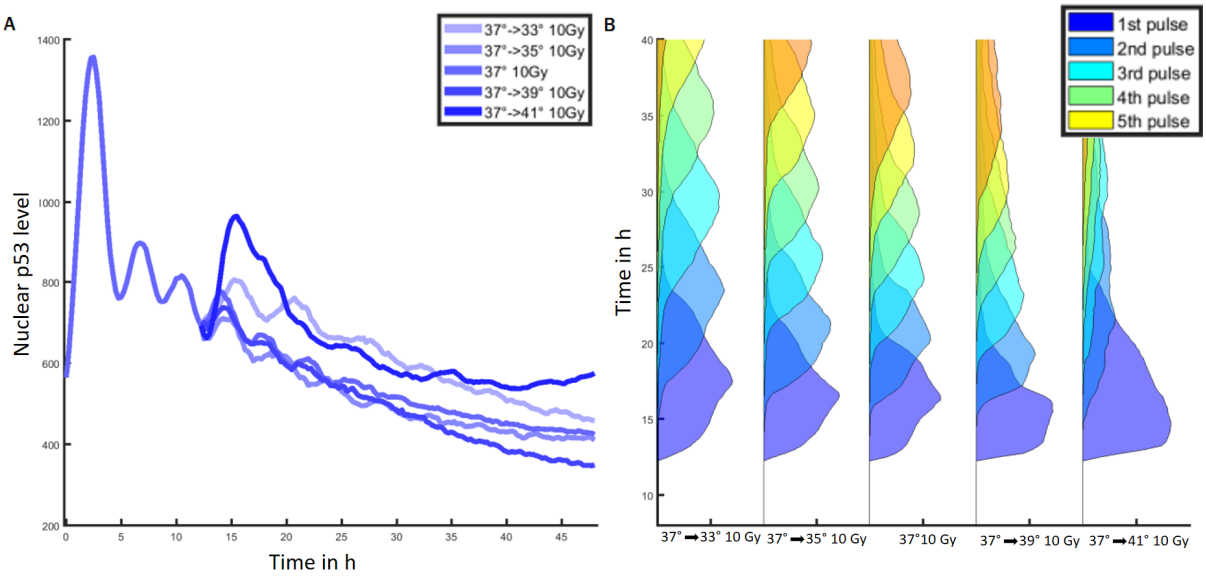

Fig. 12.9. Temperature change alters the dynamics of an ongoing p53 response

A - Cells at 37°C were irradiated and monitored for 12 h. After the 12 h the temperature was altered in a range from 33°C to 41°C. Time point 12 h were chosen to induce the temperature effect approximately after the 3rd pulse. After the temperature switch the cells were tracked for the following 36 h. The data for 1407, 1488, 1420, 935 and 909 cells are shown. The order of the number of cells corresponds to the order of the temperature switch.
B - Probability of being in a certain pulse after the temperature was switched at 12 h. Pulses that had started before the temperature change are removed.

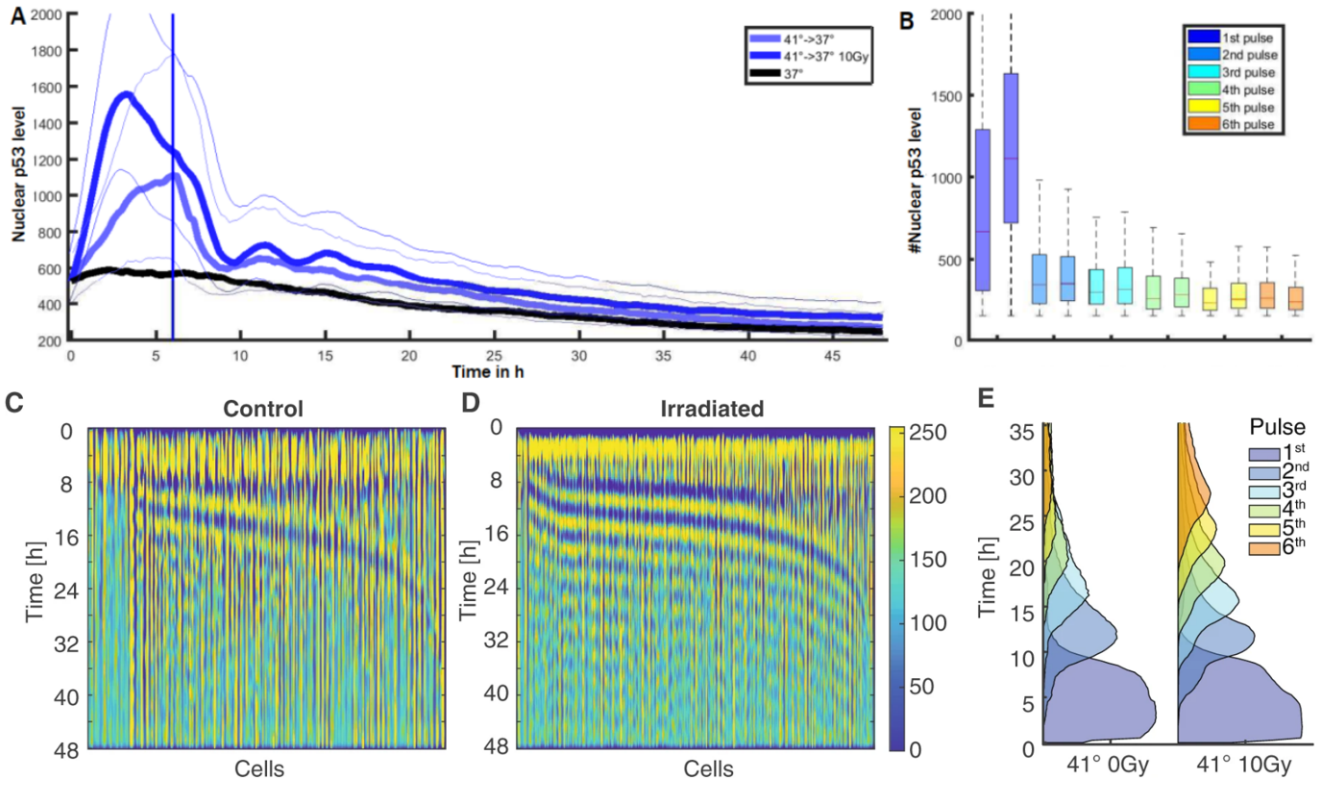

Fig. 12.10. p53 response to hyperthermia is reversible

A - Cells pre-incubated at 41°C were monitored for 48 h. After the 6th hour of the experiment the temperature was switched to 37°C normothermia. The sustained response to 41°C stops abruptly and one can detect pulses of low amplitude. Data shown for 449 non-irradiated and 2213 irradiated cells. As a reference data for 519 non-irradiated cells incubated at 37°C is shown in black.

B - Power of the detected pulses. The first pulse corresponds to 1st acute p53 response to 41°C. The strength of the detected pulses after the acute temperature response does not differ substantially among radiated cells and irradiated cells. Red lines indicate medians of distributions, boxes include data between the 25th and 75th percentiles and whiskers extend to maximum values within 1.5 times the interquartile range.

C - Heatmap of p53 levels for the non-irradiated cell population sorted by the timing of the first pulse that appeared after the temperature switch to 37°C after 6h. To amplify the pulses in the heatmap, especially in the non-irradiated scenario, image processing techniques were applied. First, the cells were sorted according to the end of the first pulse detected after the temperature was reduced. Then a Gaussian filter (11x11) was applied on the sorted data matrix followed by an application of an additional median filter (31x31). The difference between the two results after the filtering steps were plotted. The pulses induced by the transmission from the acute 41°C to 37°C are not present in all cells and are very heterogeneous in their timing.

D - Heatmap of p53 levels for the irradiated cell population sorted by the timing of the first pulse that appeared after the temperature switch to 37°C after 6h. To enhance visualization of pulsatile dynamics the data was filtered using images processing techniques as in C.

E - Probabilities of being in a certain feature at a given time point for irradiated and non-irradiated cells. The induced pulsing is less heterogeneous for irradiated cells. Non-irradiated cells have on average less pulses.

First, cells were equilibrated at 37°C. To induce genotoxic stress, the cells were irradiated with 10Gy and time-lapse microscopy was carried out to monitor the p53 dynamics. After 12 h, the temperature was changed in the range from 37°C to 41°C and the cells were monitored for additional 36 h. 12 h were chosen to induce the expected change in the ongoing p53 dynamics approximately after the end of the 3rd pulse of the canonical p53 response to radiation (Fig. 12.9(A)). Before analysing the new dataset, the normalization step presented in section 12.1 was carried out. The results reproduce to a certain extent the observations made in the previous two sections. At the first 12 h cells response to genotoxic in a regular pulsatile manner. Compared to 37°C, after the temperature switch the pulsing frequency was reduced while the temperature was lowered and at 39°C the pulsing frequency increased together with an increase in the heterogeneity of the timing of the pulses. At 41°C cells lose their pulsatile dynamical patterns and exhibit a strong immediate transient response. The temperature change disrupts the ongoing canonical dynamics of p53 DDR and a new dynamic regime is exhibited. This could be shown by calculating the pulses, with the novel local DTW based feature detection method (Fig. 12.9(B)). Only pulses that appeared after the 12th h of the experiment passed were investigated. The temporal pattern of the appearance of pulses was studied according to the different temperatures.
Finally, the acute p53 response by changing the temperature from 41°C to 37°C is studied. It was already hypothesized before that short-term temperature change in pre-equilibrated cells during the transfer from the incubator to the microscope may be sufficient to reset the p53 system. It had been observed so far that a change to 41°C directly induces a dynamic p53 response that overrides the dynamics exhibited by the cell before the temperature change. For this end, it was studied how a change in temperature alters an ongoing p53 response to hyperthermia in irradiated and non-irradiated. The experiment from section 12.2 was repeated with cells pre-equilibrated at 41°C but the temperature was not kept constant over the whole time course. After 6 h the temperature was decreased to 37°C normothermia (Fig. 12.10(A)). This temperature drop took around 30 minutes due to the technical capabilities of the incubation system incorporated in the microscope.
For the first 6 h the initial findings for hyperthermia conditions could be reproduced. A strong accumulation of p53 in the nucleus was measured. This accumulation is faster for irradiated cells and slower for non-irradiated cells. Irradiated cells reach their maximum amplitude already before the temperature was changed. For non-irradiated cells it still seems as if p53 level is still increasing while the temperature reduction to 37°C was induced. In irradiated and non-irradiated cells, the p53 level dropped immediately the moment the temperature was change. The p53 level decreased and irradiated cells started to show pulsatile dynamical patterns around 3h post the temperature shift (Fig. 12.10(E,F)). Surprisingly, p53 pulses could be detected in non-irradiated cells after the direct p53 response to hyperthermia was terminated (Fig. 12.10(C,D)).
To quantify the pulses, feature detection based on local DTW was applied. In

Fig. 12.10(B)) the power of the different detected pulses is shown. The first pulse shown corresponds to the direct response to hyperthermia. For the first response significant differences in the accumulated p53 level could be quantified. This was not the case for the pulses that followed after the temperature was set to normotherm. Also it was analysed when the pulses appear (Fig. 12.10(D,E)). It was found that the timing and synchrony of p53 pulses differs notably (Fig. 12.10(D,E)). Irradiated cells exhibit, after the direct response to hyperthermia was terminated, homogeneous dynamical patterns comparable to what had been observed in irradiated cells equilibrated at 37°C after the first 2 or 3 pulses have passed. In non-irradiated cells p53 pulses vary to a much greater extent. They occur asynchronous over a wider time period. The timing of first pulse after the temperature was changed has a much greater interquartile ratio than the same pulse in irradiated cells.

While nearly all irradiated cells show a pulsatile behaviour after the temperature shift 11% of non-irradiated cells did not exhibit any secondary response after the direct response to hyperthermia. In addition, irradiated cells produced several pulses, while most non-irradiated cells have only 2 distinct pulses. The dynamical differences point to an alternative molecular mechanism that steers pulsatile p53 accumulation in the absence of radiation induced genotoxic stress related kinase activity.

Part VI

Simultaneous monitoring of P53 and P21 upon genotoxic stress

13

P53 modulates P21 in dependence of the cell cycle state

The following sections are based on data; we have published in 2019 [304]. The experiments shown in this chapter were carry out by Caibin Sheng at the TU Darmstadt.

13.1 p53 and p21 DNA damage response dynamics are related to the cell cycle

To investigate how p53 and p21 dynamics in the DDR are connected, live-cell microscopy was used to monitor the nuclear level of the proteins over time in single cells, using the fluorescence reporter system introduced in section 4.1. All experiments presented and analysed in this chapter consist of two phases. For the first 24 h cells were grown under normal conditions without any treatment. After 24 h the cells were exposed to ionizing radiation. Afterwards, the dynamics of p53 and p21 were monitored for additional 24 h. The first phase was carried out to monitor the protein levels of p53 and p21 and cell division events while cells were undamaged and therefore able to proliferate. Hence the DDR could be analysed in context of previous monitored molecular dynamics of the p53-p21 pathway and the cell cycle state. The exposure to irradiation induced DSBs in the cell and stopped detectable cell proliferation immediately. In this section all cells in the experiment were exposed to the same 5 Gy dose of genotoxic stress after the proliferation phase of 24 h passed (Fig. 13.1 (A,C) and Fig. 13.2 (A,B)).
For the first 24 h stochastic p53 pulsing was detected in some cells, that was in general not accompanied by a change in the nuclear p21 level. Upon DSB induction by irradiation p53 starts to pulse in almost all cell. The first pulse directly starts upon damage induction and reaches a maximum after around 4 h and is homogeneous in the whole population of the monitored cells. Later pulses appear damped at the population level due to the heterogeneity in their timing (Fig. 13.1 (A, B)). On the population level, p21 accumulates with a small time lag of 1-2 h within the first 6 h after the irradiation and remains elevated for the rest of the course of the experiment (Fig. 13.1 (C)). However, at the level of single cell trajectories, the observed p21 dynamics exhibit heterogeneity to a large scale

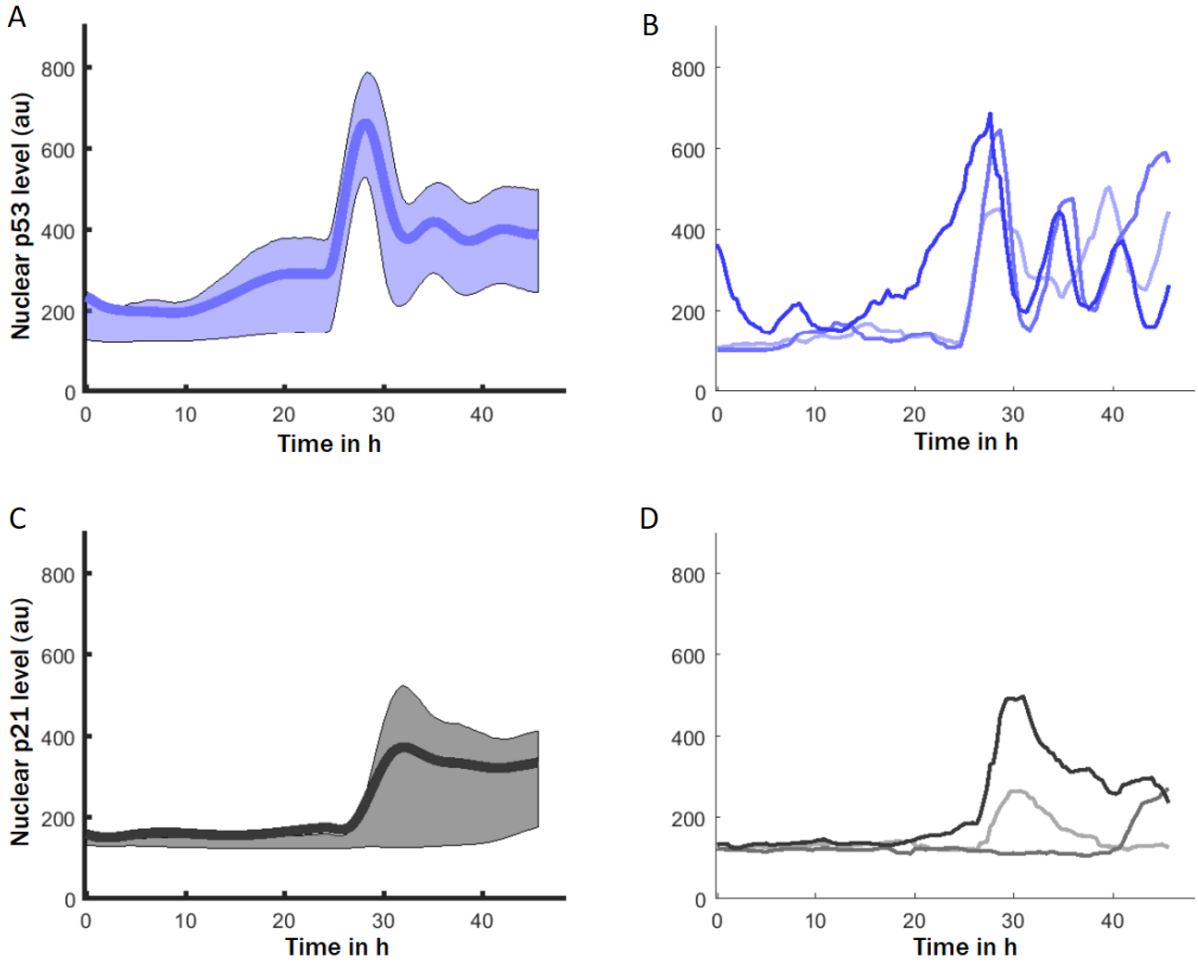

Fig. 13.1. The dynamics of p53 and p21 in response to genotoxic stress can be monitored simultaneous at the single cell level

A - Median nuclear p53 level of cells tracked for 48 h. After 24 h, cells were stimulated with 5 Gy and the change in the p53 dynamics was monitored. The shaded area represents the 25% and 75% quantile range. In total 9868 cells were tracked.
B - To visualize heterogeneity in the data three p53 trajectories were randomly selected and shown. The first direct p53 pulse of the DDR is homogeneous, while the later pulses vary in the timing and overlap. These overlap let the pulse appear damped in the median plot shown in (A). In the first 24 h cells exhibit some stochastic pulsing even in absence of induced DNA damage.
C - Median nuclear p21 level of cells shown in (A). After 24 h cells were stimulated with 5 Gy and the induced change in the p21 dynamics was monitored. The shaded area represents the 25% and 75% quantile range.
D - To visualize heterogeneity in the data the p21 level of the randomly chosen trajectories from (B) were shown. The p21 response varies in strength and timing. The lightness of the colour identifies the corresponding cells in the plots.

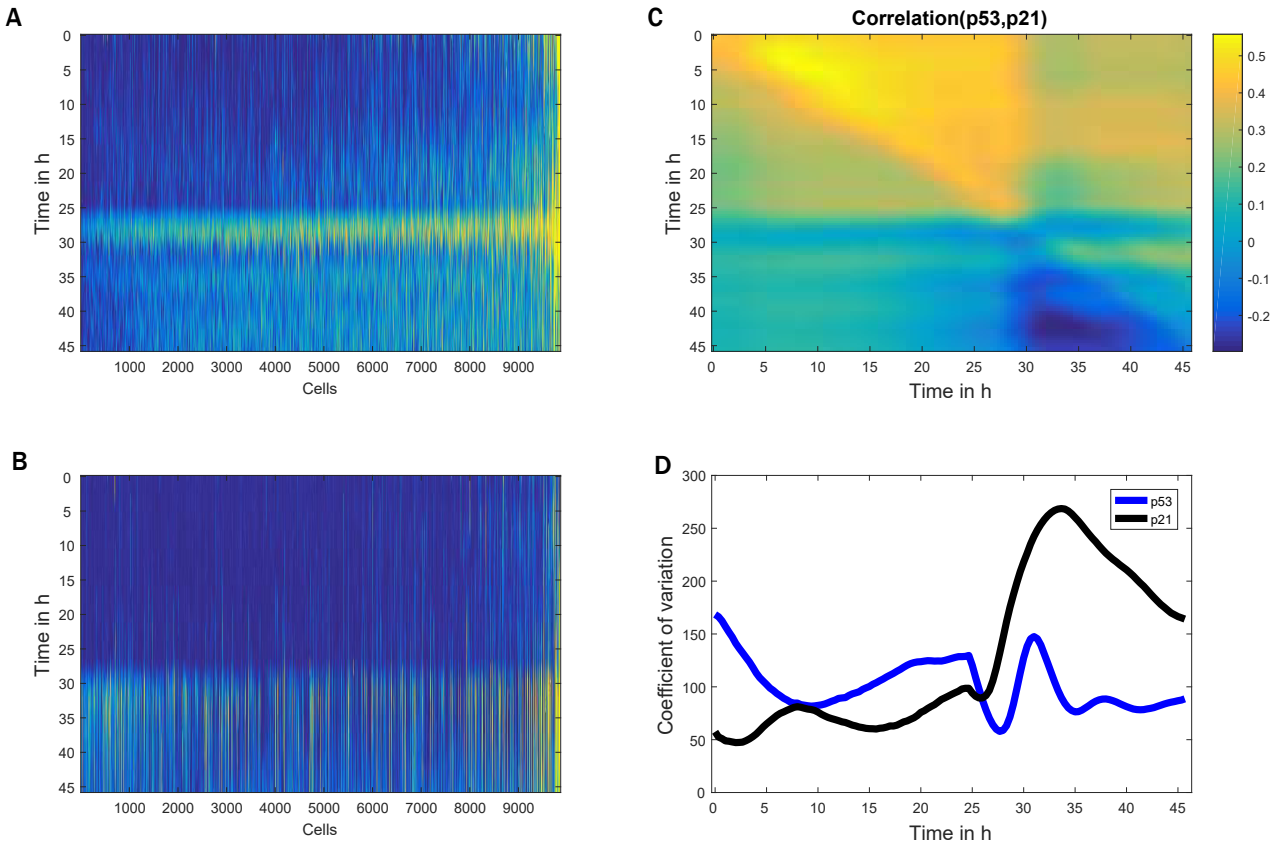

Fig. 13.2. The dynamics of p53 and p21 in response to genotoxic stress can be monitored simultaneous at the single cell level

A - Heatmap of the p53 single cell trajectories. The cells were sorted with respect to the accumulated p53 level over time. The more bluish the colour the lesser is the measured nuclear level of p53 in the nucleus.
B - Heatmap of the p21 single cell trajectories. The cells were sorted with respect to the accumulated p53 level as used in (C). We could not identify a direct correlation between the strength of the p53 response and arrangement p21 trajectories. The more bluish the colour the lesser is the measured nuclear level of p53 in the nucleus.
C – Cross-correlation between p53 relative to p21. Before the DNA damage induction, the p53 measurements distribution is positively correlated with later p21 levels. P21 exhibits no specific correlation with later p53 measurements. After the damage induction any p53 correlation to p21 vanishes.
D - Coefficients of variation over time for control cell populations shown in (A, B). After the DNA damage is induced the heterogeneity in the p21 measurements increases drastically.

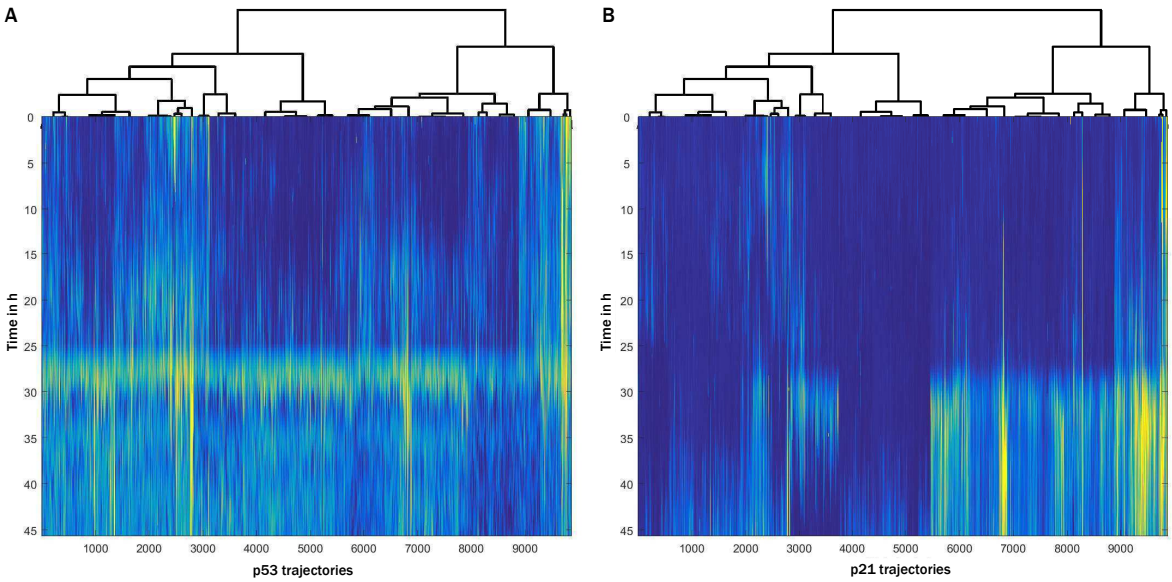

Fig. 13.3. Structure of the classification of single cell p53-p21 dynamics based on DTW similarity estimates

A - At the top the dendrogram calculated for the classification of single cell trajectories. The heatmap of p53 trajectories is sorted according to the leafs of the dendrogram.
B - At the top the same dendrogram as shown in (A). The heatmap of p21 trajectories is sorted according to the dendrogram.

(Fig. 13.2 (D)). For example, cells were observed that accumulate p21 directly after the treatment in one big pulse, cells that start to accumulate p21 hours after the irradiation or cells that do not show any kind of p21 response. At first glance, any correlation between p53 and p21 dynamics is hidden from direct observation. While sorting p21 trajectories according to the integrated p53 intensity no simple linear relation or pattern (Fig. 13.2 (A,B)) could be detected.
In Fig. 13.2 (C) the correlation of the p53 and p21 measurements at different time points is shown. Before DNA damage induction p53 is positively correlated with current and future p21 levels. This correlation is reduced for p21 levels measured after the irradiation. P21 measurements are not correlated with future p53 measurements. Remarkable the relation of p53 and p21 seems to vanish after the treatment.
While no direct linear relation is obvious between p53 and p21 measurements after DNA damage induction - how else are obtained dynamics connected to each other? To shed light on the issue, a DTW based approach to classify single cell dynamics was applied. This could allow the identification and quantification of non-linear correlations between p21 and p53 dynamics. In chapter 11, the single cell trajectories were clustered with respect to the observed SMAD2 dynamics. The analysis of single cell dynamics in this chapter goes a step further, while incorporating that more than one readout of a protein of interest is given. The approach pro-

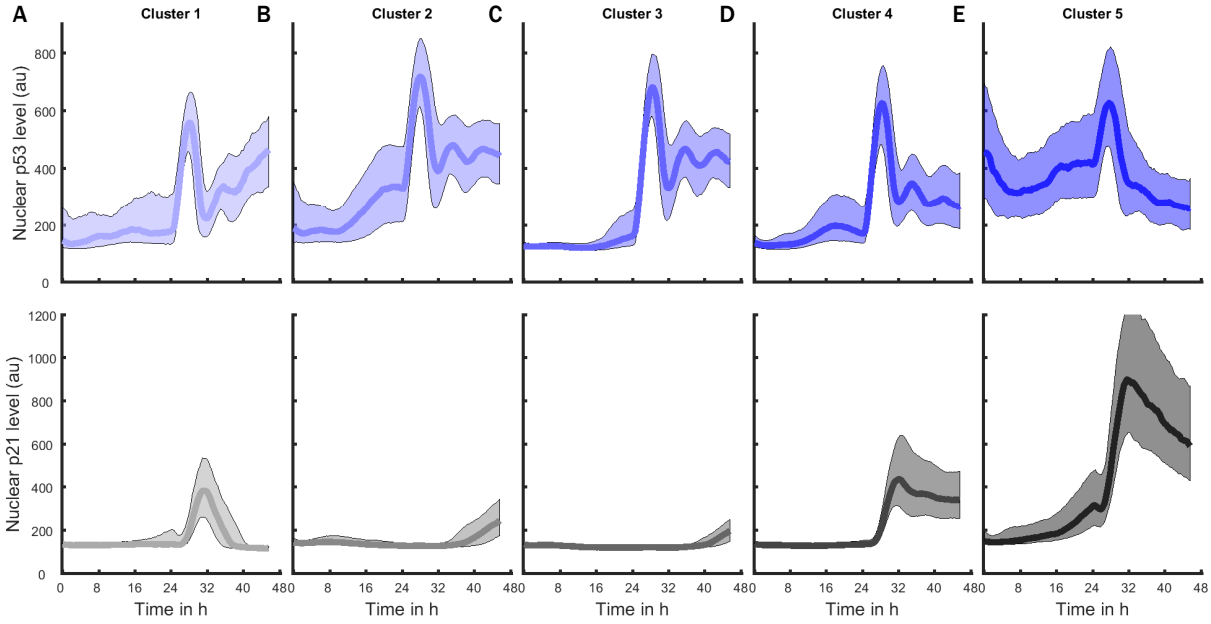

Fig. 13.4. The p21-p53 signalling can be classified in distinct dynamics (At the top in blue the p53 dynamics and below in black the corresponding p21 response)

A - The signalling class is 1 formed by 913 cells (p53 in blue and p21 in grey). The shaded area represents the 25% and 75% quantile range. These cells exhibit a direct p53 pulse followed by a steady increase of p53 accompanied by a shortly delayed pulse like p21 response.
B - The signalling class is 2 formed by 2830 cells. One could observe a direct p53 pulse followed by asynchronous pulses. p21 starts very late in the course of the experiment to accumulate.
C - The signalling class is 3 formed by 1721 cells. Surprisingly, this cluster does not show any p53 activity before the irradiation. One could observe a direct p53 pulse followed by asynchronous pulses. p21 activity is negligible.
D - The signalling class is 4 formed by 3416 cells. One could observe a direct p53 pulse followed by asynchronous pulses of low amplitude in contrast to cluster 2 and 3. p21 accumulates fast with a short delay and remains elevated for the rest of the course of the experiment.
E - The signalling class is 5 formed by 988 cells. p53 already shows a relatively high activity before the irradiation. Irradiation induces on direct pulse followed by a decrease in the p53 level. These cluster exhibits a strong p21 activity already before irradiation.

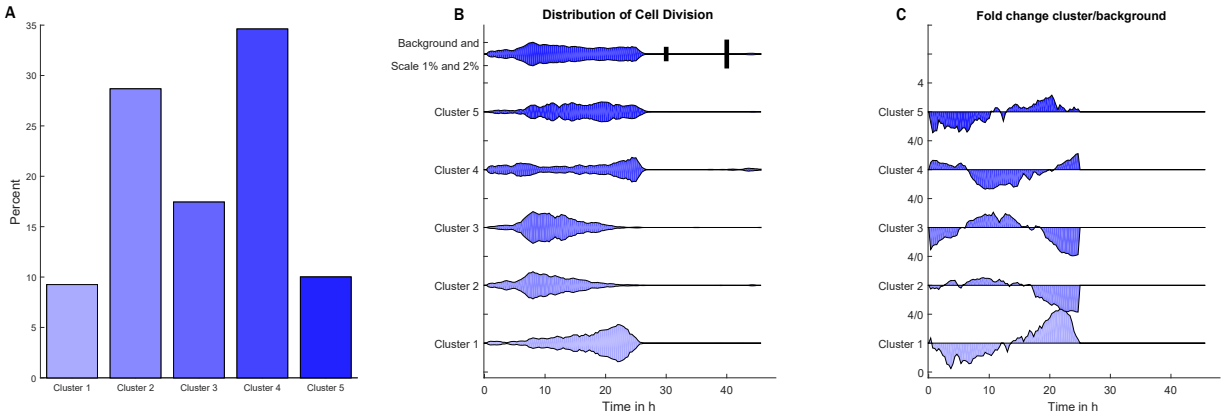

Fig. 13.5. The clusters are not even distributed over the population. One of the main drivers behind the formation of signalling classes is the state of the cell cycle before the irradiation.

A - Distribution of the clusters among the cell population. The number of p21 none responders (Clusters 2,3) and the number of cells that exhibit a 'normal' p21 response (Clusters 1,4) have the similar proportion of the population.
B - Temporal distribution of the cell division events. At the top the background distribution of the whole cell population is shown. At the same top line 2 vertical black lines on the right serve as a scale for 1% and 2%. The other 5 shown distributions illustrate the relation of clusters and cell cycle state. All clusters show a specific shape of the observe cell division distribution.
C - To enhance the difference in the cell division distribution of the different signalling classes the difference to the background (distribution at the top of (B)) is shown.

posed in section 7.3, that enables quantification of dissimilarity among different multidimensional single cell time series, will be utilized. Multidimensional in this context means 2D single cell trajectories of simultaneously measured nuclear p21 and p53 intensity.
The trajectories were aligned as shown in Fig. 7.5. Based on these alignments qualitative estimates on how much p53-p21 signalling dynamics differ among cells were made. To guarantee that none of the two proteins of interest gets overestimated in the 2D DTW dissimilarity measurement, the p21 and p53 measures were normalized, using the min-max method. Min-max normalization scales the intensity measurements into a range between zero and one. For all the intensity measurements x of each protein, the normalized values z by $z_i = (x_i - min(x))/(max(x) - min(x))$ were calculated. The normalization step was executed globally on all p53 or p21 measurements on the whole cell population and not separately for each trajectory. After having computed all pairwise dissimilarities the clustering procedure as described in 8.1 was used. The resulting dendrograms are shown in figure 13.3. They illustrate the arrangement of the classification of the signalling dynamics produced by the clustering approach. By investigating the heatmaps sorted according to the hierarchical clustering, one observed again that the p53 dynamics

are relatively homogeneous while the p21 dynamics exhibit a strong heterogeneity. However, several patterns that shape the relation among p53 and p21 dynamics can be detected. The number of signalling classes was set, as in chapter 11 based on the results of the Jump method [327]. Following this procedure, five distinct clusters could be identified, that incorporate the p53 as well as the p21 dynamics with equal proportion.

The dynamics from the population perceptive are shown in figure 13.4. The clusters differ in their p53 dynamics as well as in their p21 dynamics. p53 differs among clusters in the activity before the irradiation, the strength of the pulse that directly follows the treatment and the shape of the late response that follows. p21 exhibits as expected a more heterogeneous picture of different dynamics with respect to the clusters. The analysis included cells that do not show any p21 activity (cluster 3), cells that show late accumulation of p21 (cluster 2), cells that exhibit a short pulse like response (cluster 1), cells that accumulate p21 fast and have an increased p21 level for the rest of the course of the experiment (cluster 4) to cells that accumulate fast vast amounts of p21 followed by a steady decrease in the p21 level (cluster 5).

While looking at the median trajectories of the clusters there is not always a direct visible link between the p53 and p21 dynamics observed (Fig. 13.4). For cluster 5 cells that that have already a high basal p53 activity before the treatment exhibit the strongest, around 2-fold of what was observe in the other clusters, p21 response. This strong p21 response is accompanied by losing the pulsatile p53 dynamics and an increase in the p53 level on average to levels lower than that was observed before the stimulus. In cluster 5, already a significantly enhanced p21 activity before irradiation compared to the other signalling classes could be detected. Cells from cluster 1 show an increase in the late p53 level, while the p21 is removed from the nucleus. In contrast, cells from cluster 4 have a low median p53 level accompanied by a relatively high p21 level in the nucleus. The late p21 responders, cluster 2 and 3, differ mainly in the p53 activity before the treatment. Cells from cluster 3 show, as the only cluster, no p53 activity before the irradiation. This increased p53 activity in cluster 2 leads to an earlier accumulation of p21 compared to cluster 3, where nearly no p21 activity is detectable over the whole course of the experiment.

The frequency of the signalling classes is not evenly distributed in the cell population studied(fig.13.5 (A)). A little less than 2/3 of the cells come from the clusters 2 and 4 and cluster 1 and 5 provide around 10%. Remarkable the number of p21 none responders (Clusters 2,3) and the number of cells that exhibit a ‘normal‘ p21 response (Clusters 1,4) have the same share of the population.

While p21 as a major target of p53, the p21-p53 pathway integrates exposure to stress with the cell cycle. To this end, it is of special interested if the estimated signalling classes and their underlying signalling dynamics have a connection to the cell cycle state. As estimate for the cell cycle state, the cell was in before the genotoxic stress was applied, the time passed since the last detected division before the treatment was used (Fig. 13.5(B,C)) . A clear relation between the

classification of the p53-p21 dynamics and the cell cycle state could be detected. Cells from cluster 2 and 3 that show only a minor late p21 response originate from cells that are in S or G_2. Cells from cluster 1 that exhibit a short pulse-like direct p21 response are mainly cells from G_1 that divided near to the time of irradiation. For cells that form cluster 4 and 5 the connection to the cell cycle is less clear. Cluster 4 is composed of cells that have divided recently and cells that have divided at the begin of the experiment, so cells either in G_1 or G_2. Cluster 5 cells show no clear trend of the time passed since the last division, with a small under representation of cells from S and G_2 and small over representation of cells from G_1.

The clusters differ as well in the number of expect divisions before the irradiation. In cluster 1 around 90% of the cells have divided, while in cluster 3 and 5 this number drops to around 75% and in cluster 2 and 4 around 67%. Surprisingly, the increased p53 activity before irradiation in cluster 5 had no impact on the proliferation.

The p53 response within the population was quite homogeneous while the p21 response is variable. One main driver for this might be the cell cycle state, a cell is in while exposed to genotoxic stress. It could be also reveal that p21 and p53 signalling dynamics are connected. The p21 accumulation is commonly accompanied with low nuclear level of p53 in the nucleus at later time points. Especially, for cluster 1 and 5 this is obvious. The direct p53 pulse induced by genotoxic stress has no influence on the behaviour of p21 within the experiment.

13.2 Dynamics are linked to the strength of the genotoxic stress

After it has been shown, that one can identify several distinct signalling classes, in a population genetically identical of cells, treated with the same level of genotoxic stress, it will be studied if and how the strength of the applied ionizing radiation has an impact on the recorded p53-p21 dynamics. Will where be distinct signalling classes for each level of applied genotoxic stress or will the same signalling classes be detected in each of the conditions? To this end, it will be studied how stronger genotoxic stress as applied in the previous chapter shape and modulate the response dynamics of the p53-p21 signalling pathway. In addition to the previous applied 5 Gy also higher doses of 10 Gy and 20 Gy were investigated.
As in the previous section, the experiment analysed in this section consisted of two distinct phases. A first phase where the cell proliferation and pathway activity in absence of a stimulus and a second there the p53 and p21 dynamics in response to different levels genotoxic stress were monitored.
On the population level similar dynamics were observed as already described for 5 Gy (Fig. 13.6. The nuclear p53 level of the DDR is positively correlate with the strength of the stimulus. The amplitude of the pulse directly induced by genotoxic stress as well as the level of the late phase increased with the dose. The dynamics of p21 appear less sensitive to the irradiation dose. The median p21 response is a little stronger and the width of the distribution of the measured p21 intensity increases with the dose.
As in the previous section, the DTW framework was used to estimate all pairwise similarities among the single cell trajectories of the new dataset. To avoid a bias towards one of the proteins of interest, min-max normalization was applied as described in the previous section. After calculation of all cell to cell pairwise similarities, the clustering procedure as described in 8.1 was applied again. To estimate the number of clusters the Jump method [327] was used as well. The Jump method indicated to use either 3 or 5 distinct signalling classes. As five clusters were identified in the previous section, the same amount of clusters was chosen for the dataset analysed in this section.

The computed classes of p53-p21 signalling dynamics are shown in figure 13.7. The signalling classes estimated exhibit remarkable similarities to the clusters identified in the previous section. The identified cluster 1* and 5* exhibit similar dynamics (fig. 13.7 (A,D)). A direct p53 pulse after irradiation is detected, followed by a steady increase for the rest of the experimental course on the level of the population. The two cluster differ primary in the amplitude of p53. Cells from cluster 5* exhibit a higher nuclear p53 level. The p21 response is in both clusters nearly negligible. Cells from signalling class 1*, show a minor late response by staring to accumulate p21 slowly around 16 h post irradiation. Also the clusters 2* and 4* are similar in their observed dynamics and differ primary in the strength of the p53 response, which is higher in cluster 4* (fig. 13.7 (B,D)). A direct p53 pulse is observed followed by increased relatively stable level of p53. This late p53

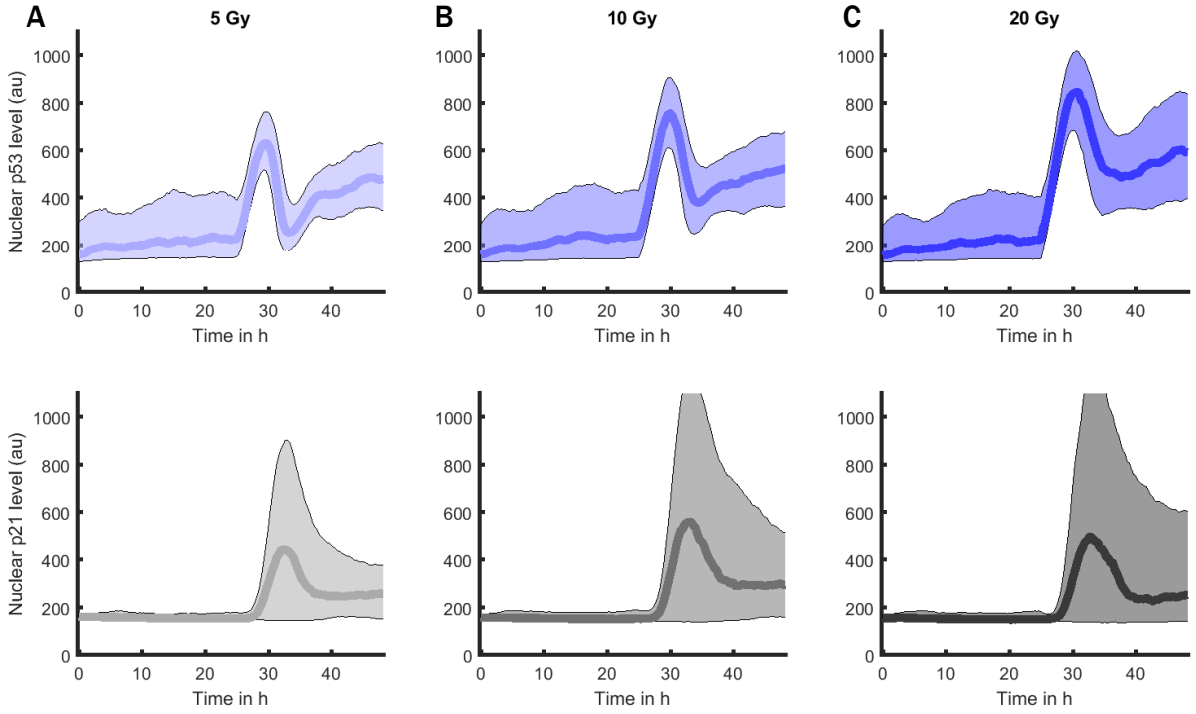

Fig. 13.6. The p53-p21 dynamics induced by different strength of the genotoxic stress.

A - Dynamics of p53 and p21 on the population level. Cells were treated with 5 Gy after 24 h have passed. In total 982 cells were tracked for this condition. p53 in blue and p21 in grey. The shaded area represents the 25% and 75% quantile range.
B - The same experimental setting as in (A) with 10 Gy. 1263 cells were tracked.
C - The same experimental setting as in (A) with 20 Gy. 1186 cells were tracked. The increased dose enhances the p53 response. The p21 dynamics are relatively unperturbed by the dose.

level is lower than what had been monitored for cells from cluster 1* and 5*. The nuclear level of p21 increases fast after irradiation with a short temporal delay and reached a maximum amplitude around 8 h after the exposure to genotoxic stress. Cluster 3* cells already show increased p53 activity before irradiation accompanied minor p21 activity (fig. 13.7 (C)). After irradiation only one direct strong p53 pulse could be monitored followed by a decreased nuclear level of p53. Surprisingly, this decreased late p53 level in cluster 3* cells is even lower than the level detected before irradiation. Additionally, the p21 response after irradiation is very strong. p21 accumulates fast and reaches values of more than two fold of the p21 amplitude, that has been measured in the other clusters. After a maximum is reached, around 5 h after irradiation, the p21 level is decreasing, but is still around the level of maximum amplitude at the end of the experimental course, measured for cluster 2* and 4*.
The signalling classes identified in this dataset do not reproduce directly the clus-

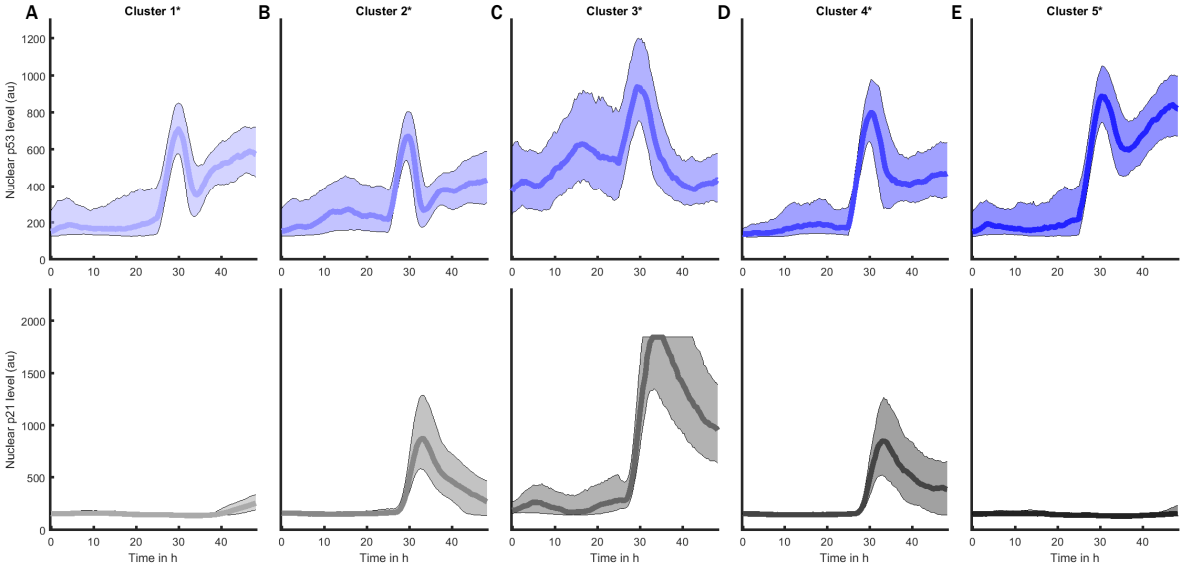

Fig. 13.7. The p21-p53 signalling induced by different levels of genotoxic stress can be classified in distinct dynamics

A - The signalling class 1* is formed by 934 cells (p53 in blue and p21 in grey). The shaded area represents the 25% and 75% quantile range. These cells exhibit a direct p53 pulse followed by a steady increase of p53. A minor p21 response is detectable at the end of the course of the experiment.
B - The signalling class 2* is formed by 1206 cells. One could observe a direct p53 pulse followed by asynchronous pulses that are shaded by their heterogeneity in the timing. p21 responses accumulates fast after the irradiation with a short delay. The p21 level reaches a maximum around 33 h and decreases slowly.
C - The signalling class 3* is formed by 346 cells. This cluster exhibits the strongest p53 activity before irradiation. After irradiation p53 one can observe one big pulse. After the first pulse has ended p53 activity is the lowest among all clusters.
D - Signalling class 4* formed by 476 cells. The p53 and p21 dynamics equal the dynamics of cluster 2* with a higher p53 amplitude.
E - The signalling class 5* formed by 469 cells. The p53 and p21 dynamics equal the dynamics of cluster 1* with a higher p53 amplitude and negligible p21 activity.

ters found in the previous section. As an exception cluster 5 from the previous section could be directly linked to cells from cluster 3*. The p21 non-responding classes 1* and cluster 5* are likely to have connection to the clusters 2 and 3 from the previous section as well as that a connection between cluster 2* and 4* and the cluster 1 and 4 from the previous section is not negligible. The connection between the clusters identified in the different sections was primary based on the observed p21 dynamics.
While having a glance at the distribution of signalling classes with respect to the applied level of genotoxic stress, the precision of the mapping of the two different

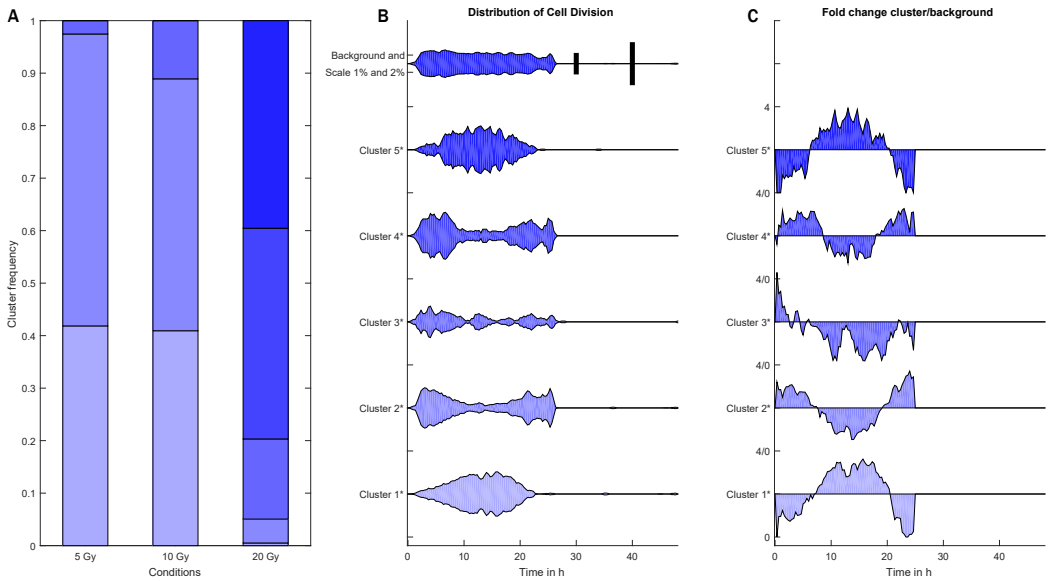

Fig. 13.8. Clusters are differently distributed with respect to the condition and the signalling classes are dependent on the cell cycle state.

A - Distribution of the clusters among the different conditions. The number of p21 none responders (Clusters 1,5) and the number of cells that exhibit a 'normal' p21 response (Clusters 2,4) have the similar proportion of each differently treated population. At 20 Gy the population shifts from cluster 1 and 2 cells to mainly cluster 4 and 5 cells.
B - Temporal distribution of the cell division events. At the top the background distribution of the whole cell population is shown. At the same top line 2 vertical black lines on the right serve as a scale for 1% and 2%. The other 5 shown distributions illustrate the relation of clusters and cell cycle state. All clusters show a specific shape of the observe cell division distribution.
C - To enhance the difference in the cell division distribution of the different signalling classes the difference to the background (distribution at the top of (B)) is shown.

sets of signalling classes (fig. 13.8(A)) onto another could be increased. No matter the stimulus cells from cluster 3* at all irradiation levels could be found. They were present in all three conditions in only a small fraction, slowly growing with the strength of the applied ionizing radiation. At 5 Gy and 10 Gy majority of

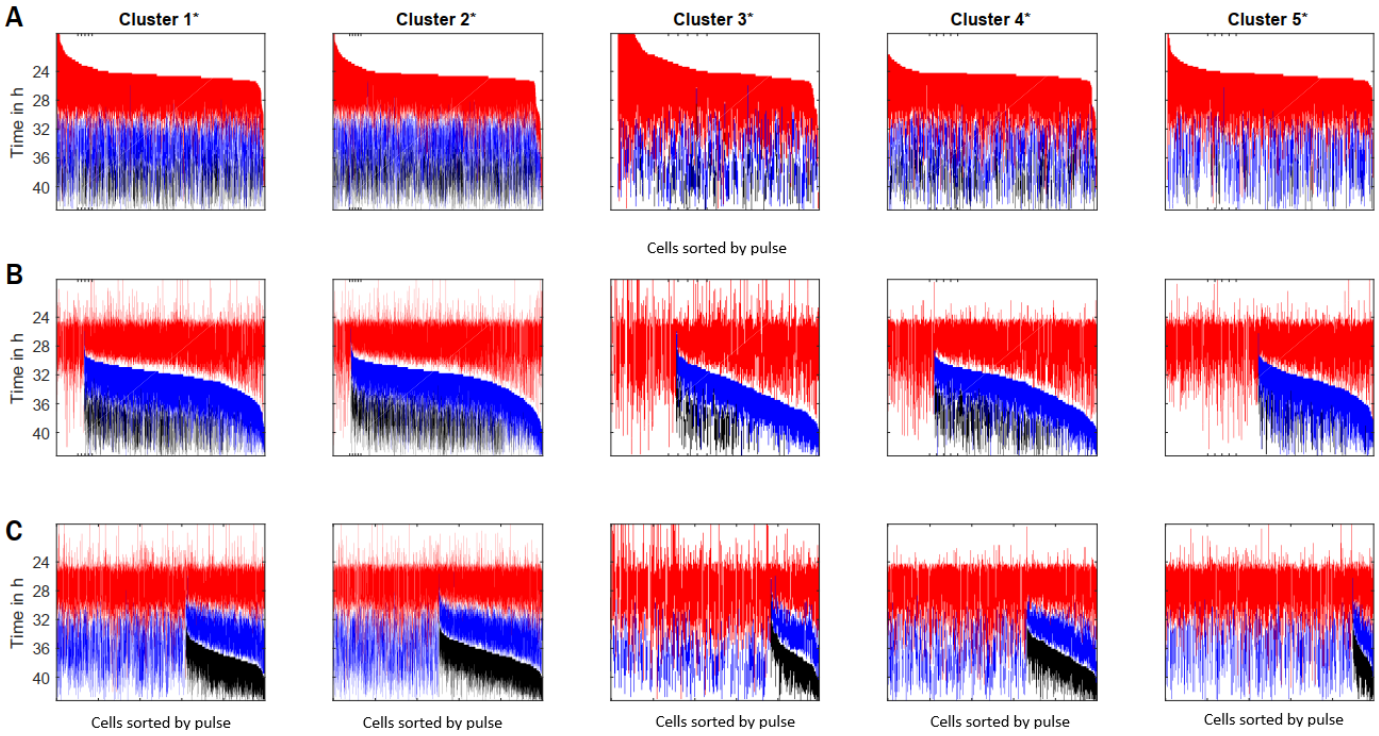

Fig. 13.9. Temporal distribution of the first three features detected, that were already ongoing while genotoxic stress was applied or were induced by the genotoxic stimulus.

A - Detected features sorted by the first pulse (red). The first was detected in nearly all cells. A not small proportion of cells from cluster 3 were due to the enhanced p53 activity before the treatment already in a pulse before the stimulus was set.
B - Detected features sorted by the second pulse (blue). The temporal distribution of the second pulse is more homogeneous in cluster 1 and 2.
C - Detected features sorted by the third pulse (black).

the cell population is contrived by cells that originate from 1* or 2*, both with similar proportion, while none of the cell could be assigned to cluster 4* and 5*. At 20 Gy the population shifts towards cluster 4* and 5* cells. Cluster 1* cells are completely lost and cluster 2* cell are a minority of less than 5%.
With these additional observations, it could be assumed that cluster 1* is likely composed of cells from initial cluster 2 and 3, that cluster 2* cells are aggregated from initial cluster 1 and 4 cells and that the remaining cells cluster 3* cells were directly connected to the previous cluster 5. As it is obvious that cluster 1* and 5* show similar dynamics. Cluster 1* are likely composed of cells from the previous identified clusters 2 and 3. I could be claimed that cluster 5* was likely composed of cells with dynamics similar to the previous identified clusters 2 and 3 but with a stronger p53 response due to the increased level of genotoxic stress. The same can be stated for cluster 2* and 4*.
To support these assumption, the state of cell cycle (fig. 13.8(B,C)) cells were

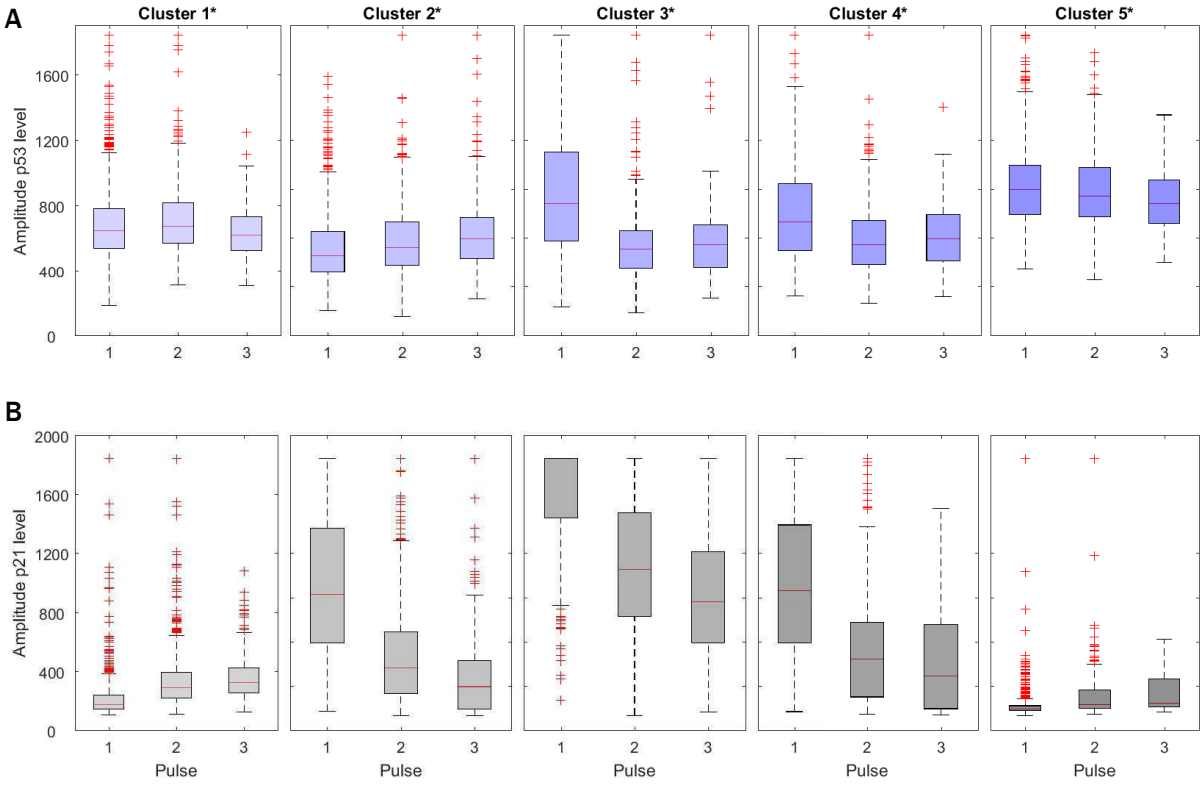

Fig. 13.10. Amplitude of the first three p53 pulses and the simultaneous measured amplitude of p21 while in a pulse.

A - The amplitude of the first three p53 pulses in dependence of the classification.
B - The amplitude of p21 while being in one of the first three p53 pulses.

in before irradiation was analysed in relation to the assigned signalling classes. Again a clear connection between the cell cycle state and the classification could be observed. Cluster 2* and 3* as well as cluster 1* and 5* exhibit a similar distributions of detected cell division events in the first 24 h of the experiment. Cluster 1* and 5* originate from cells that are in S or G_2, while cells from Cluster 2* and 3* are in G_1 or G_2.

The analysis of the p53-p21 dynamics in dependence to different doses of genotoxic stress was completed with the study of p53 pulses induced by the irradiation. The quantification of the pulsing was set into relation with the nuclear p21 dynamics. To this end, as in the previous chapter, the method proposed in section 9.2 to detect the p53 pulses in the dataset is utilized. From the detected pulses all pulses were removed that had ended before the cells were irradiated and the first 3 pulses that remain for each time series were analysed. In Fig. 13.9 the temporal order and frequency of the first three pulses is shown. It is not that unlikely that a cell from cluster 3* was already in a p53 pulse while the cells were irradiated. In Fig.

13.9, the corresponding amplitude of the p53 pulses in relation to simultaneous measured amplitude of p21 is shown. In close to all cells, an in the timing homogeneous first pulse could be detected, that differs in the p53 amplitude among the clusters. Cells from cluster 5* have a homogeneous strong response. The average response in cluster 3* cells is similar but with a wider distribution. The first p53 pulse in cells from cluster 2* has the lowest amplitude. The second and the third pulse have either a similar amplitude as the first pulse (cluster 1*,2* and 5*) or a reduced p53 pulse amplitude (cluster 3* and 4*). The clusters 1* and 5* that exhibit the highest amplitude in the second and third pulse show the lowest p21 activity. Interestingly, one cannot find a relation of the strength of the first pulse and the observed p21 activity. Cells treated with 5 Gy and 10 Gy, identified as cells that belong to cluster 1* and 2*, show as well a more homogeneous pulsing behaviour compared to cells treated with 20 Gy.

From the cluster analysis, it can be concluded that p21 dynamics are not directly influenced by the strength of the genotoxic stress and mainly driven by the cell cycle state and that the strength of the p53 is steered by the level of applied irradiation. The strength of late p53 nuclear level is negatively correlated with the level of p21 in the nucleus. Cells that show already p53 activity before radiation, likely struggling already with their genome integrity, are most p21 responsive due to genotoxic stress. These cells show in addition a damped late p53 response while huge amount p21 can be found in the nucleus.

Part VII

Discussion

14

Discussion

14.1 Classification of SMAD signalling

A disturbed information flow in the TGFβ pathway can change the cellular response to TGFβ ligand and cause severe human health defects like cancer, fibrosis, vascular disorders and autoimmune diseases [37, 218, 117]. There is a precise understanding of the signalling TGFβ pathway components, but it is rather unclear how the pathways can transmit context-dependent and concentration-dependent signals through the TGFβ signalling pathway [131]. Illuminating the logic that integrate these various inputs to explain the multifunctional nature of the TGFβ pathway is a central challenge [131].
A combination of experimental and computational methods was used to investigate how information is processed within the TGFβ signalling pathway. Experimentally, the shuttling of SMAD between the cytoplasm and the nucleus and phenotypic characteristics at the single cell level could be measured for up to 60 hours. It could be shown that SMAD translocation dynamics induced by TGFβ encode phenotypic responses of epithelial cells.
In general, SMAD signalling dynamics can be split in two phases. The first direct transient response of fast nuclear SMAD accumulation followed by a slower decrease in the nuclear SMAD level of more or less dose dependent amplitude. The strength of the nuclear SMAD2 accumulation is to some extent positively correlated with the strength of the TGFβ stimulus. The later second phase is more heterogeneous characterized by multiple features ranging from sustained plateaus to fluctuation in the SMAD nuclear cytoplasm equilibrium. The average dynamics observed are consistent with previously published studies based on population measurements [152, 62, 388, 347].
Especially, three processes shape the signalling dynamics: the availability of the extracellular ligand that is limited mainly by degradation due to endocytosis, the cellular state defined by the internal molecular setup and redundant transcriptional feedback [80, 378, 372, 165]. If the transcriptional feedback is reduced, the responsiveness of the cells is increased and the observed signalling dynamics are shifted towards more sustained characteristics.
Dynamic Time Warping, as a method for non-linear alignment of time series and

a supervised learning method called fundamental components were established as new tools to quantify single cell trajectories. The standard DTW method does not constrain temporal shifts, so that features can be aligned no matter how distant in time they are. Hence, the framework was modified to a stiffer version that ensures that the results remain biological relevant by constraining the extent of allowed squeezing and stretching. A modified DTW method (cDTW), that constrains the flexibility in the time domain to generate biological more relevant results, was used to analyse the translocation dynamics of thousands of cells. With cDTW the similarity among thousands of individual single cell trajectories can be quantified. DTW is less sensitive to asynchronies than simpler similarity measures such as the Euclidian distance and more important respects the order of the measurements acquired at different time points. The DTW framework provides a toolbox that can quantify similarity among time series while being flexible in the time domain. Based on the estimated pairwise cell similarity obtained by cDTW, hierarchical clustering was applied and different signalling classes of translocation dynamics could be identified. By clustering the time series, a pronounced cell-to-cell variability could be quantified even within cells stimulated with a fixed dose of TGFβ. The clustering yields a more homogeneous grouping of the dynamics compared to other sorting such as ligand concentration applied or the state of the cell cycle. The transitions from signalling class to signalling class were not sharp, rather they form a continuum. The decomposition of the cell population into signalling classes is dose depended and allows a quantification of the heterogeneity of the signalling dynamics. Heterogeneity in signalling dynamics of the response increases with the strength of the stimulus.

With the introduction of the concept of the fundamental components distinct underlying dynamics in the time series data were identified. How much the identified distinct fundamental contribute to the observed dynamics depends on the strength of the stimulus or on the presence of the negative feedback regulator SMAD7.

It could be shown, that heterogeneity in the response of the TGFβ arises mainly from the cellular state. Cell-to-cell variability in signalling protein concentrations is the main source of the observed heterogeneity. It was shown that local cell density or cell cycle state play only a minor role. Cell cycle state has for example impact on signalling through the ATM and ATR kinases[3]

Several processes have been reported to have an impact on the cellular heterogeneity [204, 314]. Cell cycle state or local cell density cannot explain the observed heterogeneity. It has been shown that the cell density sensing YAP/TAZ pathway does not influence SMAD signalling [251]. YAP and TAZ are the major downstream effectors of the Hippo pathway, which regulates tissue homeostasis, organ size, regeneration and tumorigenesis [237].

Using sister cell analysis, it could be shown that the molecular setup of a cell can explain cell-to-cell variability. As sister cells are more similar than other cells for a longer period of time after division, it can be assumed that long-lasting components dominates cellular heterogeneity. The use of sister cell analysis has proven to

be a helpful tool to analyse characteristics of signalling pathways [112, 319, 293]. The assumption is that sister cells are more similar in their molecular setup than a randomly picked pair of cells.
Negative feedback is essential for most cellular signalling pathways [189]. The primary negative transcriptional feedback in the TGFβ pathway is SMAD7. It is expected that the feedback comes into play at high ligand doses, where it is limiting the sustained secondary response. This was confirmed by studying the TGFβ response in a SMAD7 knock out cell line. By suppressing the negative transcriptional feedback, by the SMAD7 knock out, it could be observed that signalling shifts to stronger more sustained responses and to a more heterogeneous distribution of the signalling classes with respect to the applied TFGβ dose. While not only sustained dynamics could be observed additional redundant transcriptional feedbacks can be assumed that need to be investigated further. The negative regulators Ski and SnoN of TGFβ signalling are potential components worth an additional look. How the inhibition of SNoN or Ski changes the dynamics while SMAD7 is present or knocked-out can give new insight on the topology of the TGFβ signalling.
During later signalling phases especially at higher stimulation doses several stochastic fluctuations in the nuclear SMAD level could be observed. These fluctuations cannot be explained by current models of the TGFβ pathway topology [326]. It can be speculated that vesicle-mediated recycling of receptors to the cell surface leads to stochastic increases in the cellular sensitivity to the ligand. In the context of EGF signalling this has been observed [344]. In additional studies, combined live-cell reporters may shed light on the SMAD dynamics, by monitoring other known players in the TGFβ signalling like the receptors, SMURF1/2, SARA or p38.
Signalling dynamics trigger the phenotypic features at the longer time scales. Phenotypic detectable changes occur several hours post ligand application. TGFβ pathway stimulation promotes motility in general while the anti-proliferation effect is connected to strong and sustained signalling responses.
It could be shown, that the sensitivity of the TGFβ pathway within an individual cell can be controlled by the concentrations of the signalling protein. This might give cells a regulatory ability to adjust the TGFβ response according to the demands, by adjusting the signalling protein composition by other signalling pathways. This may give cells or on a larger scale the epithelium they form the potential to incorporate additional information and provide the variability of responses to function proper in different scenarios.

14.2 p53 dynamics are modulated by temperature

P53 as guardian of the genome and the central hub of the cellular stress response is crucial to prevent of tumorigenesis. In this context the regulation of cell cycle arrest, senescence and apoptosis upon stress are the best understood functions of

p53, but p53 has a much broader role, controlling many aspects of cellular homeostasis [176]. An increasing number of studies highlight the role of mutant p53 proteins in cancer cell growth and in the worsening of cancer patients' clinical outcome [66]. The p53 status in cancer predicts patient survival [229]. Disturbances in the information processing by the p53 pathway leads to transformation of individual cells and subsequent tumorigenesis. 40 years of research on p53 accompanied by more than 80000 publications still has not explained the full complexity of the p53 signalling network [158].

A combinatorial cancer treatment of hyperthermia together with radiation and anticancer agents has been used clinically where it has shown positive results to a certain degree [6]. Several elements of the p53 pathway as well as other proteins that interact with p53 are modulated by temperature [120]. However, the underlying mechanism of signal transduction and the set of genes involved in this process are still poorly understood. Progress in screening methods has paved the way for tackling the many ways in which dynamics are modulated. To understand how varying temperature modulates single cell dynamics of p53 upon genotoxic stress the nuclear p53 protein levels of thousands of genetically identical cells exposed to the physiologically relevant temperature range of 33°C to 41°C were monitored.

It could be shown that below 39°C the frequency of p53 accumulation pulses that encodes the DNA damage is positively correlated with the temperature. Also, it could be observed that the heterogeneity of the pulsing in the populations of cell exposed to a fixed temperature increases with the temperature. Surprisingly, in spite of the outstanding importance of p53 dynamics for the maintenance of genomic integrity the network p53 is embedded in as central hub cannot compensate environmental temperature. Interestingly suchlike was quantified for the NF-κB pathway in 2018 [128]. Harper et al. have shown that the frequency of cytoplasmic to nuclear translocation of the transcription factor RelA is as well positively correlated with the temperature and that this is caused by differential timing in the feedback expression. This only affects a subset of the RelA target genes. If a similar change in the timing of the negative feedback loops of the p53 pathway cause the change in p53 accumulation frequency or other factors contribute to this behaviour has to be investigated in further studies.

The most surprising finding was that at 41°C the canonical p53 pulsing induced by irradiation vanishes and is replaced by a strong initial accumulation of p53 that is followed by a sustained increased p53 nuclear level. The nuclear p53 level could be decreased to a normal level by changing the temperature back to normothermal 37°C. Also astonishing was that a change in temperature to 41°C without DNA damage induction was sufficient to induce a heterogeneous p53 response. The return to 37°C induced in irradiated and non-irradiated cell p53 pulses. The response induced by mild hyperthermia leads to an increased nuclear p53 level that stayed constant after around 6 h have passed since the temperature was altered. The molecular mechanisms underneath are unclear. Potential sources may be the integration of p53, MDM2 and HSPs [351] or structural properties of p53. Unfolding of p53 at 41°C could inhibit the interaction with the negative feedback MDM2

preventing p53 proteasomal degradation. Unfolded p53 is as well transcriptional inactive and is unable to induce its own negative transcriptional feedback loops. It is known that cancer related p53 mutants exhibit not unlikely temperature sensitive phenotypes [158, 157]. Furthermore, the binding of the chaperon HSP90 induced by elevated temperatures may stabilize p53 conformation and prevent MDM2 binding and promote MDM2 degradation [244, 266, 195].

The effect of mild-hyperthermia discussed so far dealt with cells that were pre-incubated at 41°C. Compared to cell that were pre-incubated a direct acute shift from normotherm to 41°C induced a much stronger response. As expected the initial response is much stronger for irradiated cells but surprisingly p53 accumulates drastically in untreated cells while the p53 level in irradiated returns on average to basal levels observed before the temperature was changed and the DNA damage was induced. This indicates that the p53 network acquires tolerance to heat shocks. The molecular mechanisms that contribute to this observation and the potential role of the HSPs system are unclear and require further investigations. The acquired heat shock tolerance may play an important role in described thermal chemosensitization and thermal radiosensitization [137, 54] that may contribute to the positive effect of combinational cancer treatment .

Hyperthermia induced p53 accumulation may ensure sufficient expression of target genes, such as p21 while p53s function as transcription factor is reduced. The low complexity in the sustained dynamics may not encode the DNA damage and steer a flexible response but ensure cell cycle arrest. Studies have shown that hyperthermia facilitates genome instability [346] and attenuated damage response [338], so that sustained nuclear p53 accumulation may guarantee cells enter cycle arrest and go through mitosis only if DNA integrity is ensured.

The combination of hyperthermia with radio- or chemotherapy has proven to be beneficial for the outcome of cancer treatment with the additional advantage of neglectable side effects [315, 278]. The molecular mechanisms of hyperthermia that contribute to the positive effects on cancer therapy by modulating molecular pathways is not well studied and remains an open question [137]. The p53 level may be an indicator for the forecast of therapeutic success of cancer treatment in combination with temperature [261]. The molecular dynamics observed may indicate that sustained p53 accumulation induces cell cycle arrest and that this prevents cell killing by mitotic catastrophe. Mitotic catastrophe is a failure to undergo complete mitosis after DNA damage. Hyperthermia may recover functions of p53 for example where p53 is disabled by a amplification of its negative feedback MDM2 as it is found in sarcomas [256, 190].

The complex interactions of HSPs with the p53 network may also contribute to the positive effect of hyperthermia on the cancer treatment. Heat shock induced conformational change of wild type p53 and binding to Hsp90 [352]. Hsp90 may promote refolding of destabilized mutant p53 upon application of hyperthermia is ended and thus compensating functional deficiencies in cancer cells [244, 245]. Hsp70 family members participate in the cytoplasmic sequestration of wild-type p53 in cancer cells [391]. This leads to p53 resistance to Mdm2-mediated degra-

dation [379] and to a high degree of aggregation that impairs the G1 checkpoint following DNA damage [232].
The observed dynamics and the observation that the p53 network acquires tolerance to heat shocks provide molecular evidence that the precise timing of radiation and hyperthermic treatments is a critical parameter for successful combination therapies [171, 132].
In further studies it should be determined if the expression of p53 target genes upon damage induction is temperature-dependent as well. To this end, high confidence target genes that contribute to distinct cellular response pathways, including CDKN1A/p21, GADD45 (both cell cycle arrest), XPC (DNA damage repair) and BAX (apoptosis) should be studied by measuring corresponding RNA levels at different time points and temperatures in untreated and irradiated cells by qPCR [100]. Also the use of the same cell line as in chapter VI may enlighten how the dynamics of p21 are influenced by temperature.

14.3 p53 and p21 dynamics in response to DNA damage are cell cycle dependent

p53 is embedded as central hub in a molecular network that steers several biological functionalities ranging from cell division, cell death, senescence, angiogenesis, differentiation and DNA metabolism. Among the target genes in response to p53 is the cyclin-dependent kinase inhibitor 1A (CDK1A) [102]. The gene CDK1A encodes one of the main effectors of the p53 DDR p21. Best known is its function as a CDK2 inhibitor that prevents hyperphosphorylation of RB and thereby arrest cells at the restriction point. The downregulation of cell cycle regulated genes in this context by the p53-p21 pathway [201, 330, 337, 299, 169, 385] hence regulating the cell cycle progression. Cell cycle transition from G1 to S and G2 to mitosis is regulated by cyclin-dependent kinase (CDK) family of proteins that are activated by cyclins [343] and inhibited by p21. To understand the dynamic relations between p53 and p21, after exposure to irradiation, both signalling components were monitored within thousands of individual cells.
p53 signalling encodes information about signal intensity, duration and identity in complex dynamics. After exposure to irradiation of 5 Gy homogeneous p53 dynamics were observed at the population level in contrast to p21 where a substantial heterogeneity in the accumulation was detected. The observed heterogeneity of the p21 response observed is in accordance with the current literature [61, 323]. Consistent with previous studies it was observed that not all p53 dynamics induce CDKN1A expression [203]. It could be shown that before irradiation the nuclear level of p53 is positively correlated with future nuclear p21 levels. The correlation is disrupted if irradiation is applied. How characteristics of the p53 dynamics result in CDKN1A expression could be answered using classification of signalling dynamics.
To quantify similarity among cells the DTW framework for multivariate time se-

ries, introduced in section 7.3, was used. While having labelled p53 and p21 simultaneous it would be desirable to quantify the similarity between cellular dynamics while looking at the labelled proteins simultaneously. Based on the estimated similarities hierarchical clustering was applied to identify distinct signalling classes that incorporate information of p53 and p21. In a recent study that used the same reporter system the data was clustered only by taking the p21 measures into account [304] using the k-shape algorithm [263]. As its distance measure k-Shape uses a normalized version of the cross-correlation measure in order to consider the shapes of time series while comparing them [263]. Sheng observed with this approach that p21 dynamics after DNA damage diverge from p53 dynamics in Single Cells [304], while it could be shown utilizing the DTW approach that p21 and p53 are connected. The identified signalling classes connect p53 and p21 dynamics and split the heterogeneous p21 response. The p21 dynamics in the signalling classes range from a strong direct response to no accumulation of p21 at all. The stronger the p21 accumulation the lower is the p53 at later time points. If a direct response in p21 cannot be detected the p53 level is higher. In cells that do not show any p21 response the p53 pulsing is much more prominent.
As in previous studies [304] a relation was observed between p21 dynamics and the cell cycle position at the moment of damage was observed. In addition, also the p53 dynamics could be linked to the cell cycle, as it could be shown that only cells that were in S or G1 before the irradiation exhibit a clearly pronounced pulsatile behaviour. These cells normally lack a p21 accumulation.
Cells that exhibit already p53 and p21 activity before irradiation is applied likely struggle already with their genomic integrity and exhibit the strongest response to genotoxic stress. These group shows as well no connection to the cell cycle.
Cells need p21 degradation for faithful repair and replication of the genome, as inappropriate p21 accumulation during S-phase led to increased genomic instability [304]. At can be assumed that a main driver behind the observed p21 heterogeneity is explained by post-translational regulation as heterogeneity is largely explained by cell cycle position at the moment of damage. It was shown by Sheng [304] that heterogeneous p21 dynamics post damage are caused by S phase-specific degradation through PCNA-CRL4^{cdt2}, which is necessary for maintaining genome stability.
It was also studied how signalling is modulated by different intensities of applied genotoxic stress. The dynamics similar but the strength of the p53 response is dependent of the applied irradiation.

14.4 Application of computational methods on single cell data

Complex biological signalling pathways can be analysed by utilizing computational and mathematical methods. These methods are essential for the study of the interactions between the pathway components and illumination of how these

interactions give rise to biological functions and dynamical behaviour of cellular signalling networks. These pathways form a complex molecular network that controls the dynamics of the signalling proteins, requiring investigators to consider multiple time points when analysing pathway activity [324].
In the chapter III several methods were introduced that were used in the flowing chapters IV, V and VI to handle and analyse the raw live-cell time-lapse imaging data and the extracted single-cells time series.
Coherent Point Drift, a non-rigid point set registration algorithm, was proposed as a method of choice for the cell tracking problem. To find the point-to-point correspondence it assumes that the approximate arrangement of a point cloud is preserved. Points in corresponds to the center of mass of segmented nuclei in the imaging data. The motion coherence constrain of the Coherent Point Drift mimics the observation that cells do not move independently; they are embedded in a neighbourhood that constrains their freedom of movement. This gives the impression of a coordinated motion of neighbouring cells grouped in clusters. Non-rigid registration methods outperform rigid methods while generating cell tracks. Visual inspection of the generated tracks in the imaging data ensured the quality of the tracks. But to validate the methodology further work is necessary. Using the datasets provided by the Biomedical Imaging 2013 Cell Tracking Challenge may be an appropriate way to benchmark and compare the method to other state of the art methods [216].
To analyse signalling dynamics in thousands of genetically identical reporter cells, I established constrained dynamic time warping as a tool for non-linear alignment of time series data. Dynamic time warping both emphasizes similarities in dynamic patterns of the time courses, and allows quantifying differences in signal amplitude, thereby improving the grouping of noisy single-cell trajectories. By allowing for stretching and squeezing of time courses, DTW is less sensitive to asynchronicity than simpler similarity measures such as the Euclidean distance. However, constraining the extent of temporal alignments in DTW is critical to ensure that results remain biologically significant. To this end the DTW was modified by introducing an additional parameter that reduces the flexibility of DTW by an elastic constrain.
Based on the non-linear alignments signalling classes can be identified in the data. Hierarchical clustering was used. In the course of this work, the identification of signalling classes within a population of cells has proven to be a precious tool to link dynamics to other phenotypic features.
The DTW approach of quantifying similarity and subsequent clustering was also utilized for multivariate time series. This enables the comparison of cell on the level of several pathway components simultaneous while the dynamics of the components are not treated independent. This revealed in chapter VI that p53 and p21 dynamics are connected.
Clustering of single cell data was applied before for example in Sheng [304] and Ryu [288]. Sheng used k-Shape [263] to identify the data only on p21 even if simultaneous p53 measurements were given and therefore missed the interplay between

p21 and p53 dynamics. Rhyu used the Euclidean distance and k-means to identify signalling classes in ERK activity trajectories. I would discourage the use of the Euclidean distance because of its high sensitivity to asynchronicity.
I used the Dynamic Time Warp framework to develop a feature detection method that is characterized by its flexibility in the time domain and its independence to scaling.
To identify underlying dynamics that contribute to the observed dynamics a supervised learning method was introduced. The method produces a set of fundamental components taking account existing knowledge. Linear combinations of the fundamental components can reproduce the original dataset to some extent. The concept reduces the dimensionality of the data and can be used to identify signalling classes in the data or map new data on existing classifications. This perspective onto single cell dynamics is new. Similar approaches like PCA or diffusion maps have been used for single-cell sequencing data [47, 123].
To understand complex biological systems requires the integration of experimental and computational research.

Part VIII

Appendix

15

Appendix

15.1 List of Experiments

Datasets are listed in the order of their appearance.

15.1.1 SMAD single cell data

Data used in chapter IV
SMAD2 titration 24 h
Number of cells: control - 358, 1pM tgfβ - 395, 2.5pM tgfβ - 314, 5pM tgfβ - 295, 25pM tgfβ - 351, 100pM tgfβ - 352
Carried out by Henriette Strasen

SMAD2 titration 60 h
Number of cells: control - 1442, 1pM tgfβ - 948, 2.5pM tgfβ - 776, 5pM tgfβ - 813, 25pM tgfβ - 1024, 100pM tgfβ - 932
Carried out by Stefan Bohn

SMAD2 titration 36 h, stimulus after 24 h
Number of cells: control - 671, 1pM tgfβ - 282, 2.5pM tgfβ - 697, 5pM tgfβ - 602, 25pM tgfβ - 810, 100pM tgfβ - 513
Carried out by Henriette Strasen

SMAD2 titration 24 h, parental cell line SMAD7 kock down
Number of cells: control - 266, 1pM tgfβ - 743, 2.5pM tgfβ - 655, 5pM tgfβ - 568, 25pM tgfβ - 686, 100pM tgfβ - 653
Carried out by Henriette Strasen

SMAD2 titration 24 h, SMAD7 knock down
Number of cells: control - 475, 1pM tgfβ - 670, 2.5pM tgfβ - 821, 5pM tgfβ - 614, 25pM tgfβ - 771, 100pM tgfβ - 647
Carried out by Henriette Strasen

Sister cell analysis
Number of cells: 11 replicates 100pM tgfβ - 989, 938, 1550, 1440, 1417, 796, 647, 1208, 2050, 115, 745
Carried out by Henriette Strasen

15.1.2 p53 single cell data

Data used in chapter V
All experiments carried out by Petra Snyder
Irradiation 10Gy

Non-irradiated cells equilibrated to 37°C - 2857 cells
Irradiated cells equilibrated to 37°C - 1182 cells
Non-irradiated cells equilibrated to 33°C - 1214 cells
Irradiated cells equilibrated to 33°C - 782 cells
Non-irradiated cells equilibrated to 35°C - 1759 cells
Irradiated cells equilibrated to 35°C - 1098 cells
Non-irradiated cells equilibrated to 37°C - 3830 cells
Irradiated cells equilibrated to 37°C - 2857 cells
Non-irradiated cells equilibrated to 39°C - 2436 cells
Irradiated cells equilibrated to 39°C - 1630 cells
Non-irradiated cells equilibrated to 41°C - 1139 cells
Irradiated cells equilibrated to 41°C - 1007 cells
Non-irradiated cells acutely shifted to 41°C - 721 cells
Irradiated cells acutely shifted to 41°C - 994 cells
Irradiated cells damaged at 37°C - 1420 cells
Irradiated cells damaged at 37°C and shifted to 33°C after 6h - 1407 cells
Irradiated cells damaged at 37°C and shifted to 35°C after 6h - 1488 cells
Irradiated cells damaged at 37°C and shifted to 39°C after 6h - 935 cells
Irradiated cells damaged at 37°C and shifted to 41°C after 6h - 909 cells
Non-irradiated cells incubated at 41°C and shifted to 37°C after 6h - 449 cells
Irradiated cells damaged at 41°C and shifted to 37°C after 6h - 2213 cells
Non-irradiated cells incubated at 37°C - 519 cells

15.1.3 Simultaneous p53-p21 single cell data

Data used in chapter VI
All experiments carried out by Caibin Sheng

5Gy - 9968 cells, 5Gy - 982 cells, 10Gy - 1263 cells, 20Gy - 1186 cells

15.2 List of Abbreviations

ASPP1	Apoptosis-stimulating of p53 protein 1
AKT1	RAC-alpha serine/threonine-protein kinase
AMH	Anti-Müllerian protein
AMDF	Average Magnitude Difference Function
AMP	Adenosine monophosphate
ARF	ADP ribosylation factor
ASPP1	Apoptosis-stimulating of p53 protein 1
ASPP2	Apoptosis-stimulating of p53 protein 2
ATM	Ataxia telangiectasia mutated
ATR	ATM and Rad3-related
BAX	Bcl-2-associated X protein
BMP	Bone morphogentic protein
CBP	CREB-Binding Protein
cDNA	copy DNA
cDTW	constrained Dynamic Time Warping
CFP	Cyan fluorescence protein
Cdc2	Cell division cycle protein 2 homolog
Cdc42	Cell division control protein 42 homolog
CDKN1A	Cyclin Dependent Kinase Inhibitor 1A
CDK	Cyclin Dependent Kinase
Chk1	Checkpoint kinase 1
Chk2	Checkpoint kinase 2
CPD	Coherent point drift
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DDR	DNA damage response
DNA	Deoxyribonucleic acid
DSB	Double strand breaks
DTW	Dynamic Time Warping
ERK	Extracellular signal-regulated kinases
EMT	Epithelial-to-mesenchymal transition
FIR	Finite impulse response
Gadd45	Growth arrest and DNA-damage-inducible protein GADD45 alpha
GDF	Growth and differentiation factor
GLUT1	Glucose transporter 1
GLUT4	Glucose transporter 4
GTP	Guanosine-5'-triphosphate
H2B	Histone H2B
HDAC1	Histone deacetylase 1
HSF	Heat shock factor
HSP	Heat shock protein
HZF	Hematopoietic zinc finger protein
iASPP	Inhibitor Of Apoptosis Stimulating Protein Of P53

JNK	c-Jun N-terminal kinase
LAP	Latency-associated protein
MDC	Max-Delbrück Center
MDM2	Mouse double minute 2 homolog
MAPK	Mitogen-activated protein kinase
MKK	Mitogen-activated protein kinase kinase
MKKK	Mitogen-activated protein kinase kinase kinase
mRNA	Messenger ribonucleic acid
mTOR	mechanistic Target of Rapamycin
Mre11	Double-strand break repair protein MRE11A
MUC1	Mucin-1
Nbs1	Nibrin
NFAT	Nuclear factor of activated T-cells
NF-κB	Nuclear factor 'kappa-light-chain-enhancer' of activated B-cells
NOXA	Phorbol-12-myristate-13-acetate-induced protein 1
p21	CDK-Inhibitor 1
p300	Histone acetyltransferase p300
p38	P38 mitogen-activated protein kinases
PI3K	Phosphoinositid-3-Kinase
PARP	Poly [ADP-ribose] polymerase
PCA	Principal components analysis
PIN1	Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1
PK	Protein kinase
PP2A	Protein phosphatase 2
PRMT1	Protein arginine N-methyltransferase 1
PUMA	p53 upregulated modulator of apoptosis
qPCR	quantitative polymerase chain reaction
Rad50	DNA repair protein RAD50
RE	Response element
RelA	Nuclear factor NF-kappa-B p65 subunit
RFP	Red fluorescence protein
Rho	Ras homolog family member
RhoA	Ras homolog family member A
SARA	SMAD anchor for receptor activation
Shc	SHC-transforming protein
SIK1	Serine/threonine-protein kinase SIK1
SIRT1	Sirtuin-1
SMURF1	E3 ubiquitin-protein ligase SMURF1
SMURF2	E3 ubiquitin-protein ligase SMURF2
SNR	Signal to noise ratio
SSB	Single strand breaks
SOS	Son of Sevenless
TAK1	Mitogen-activated protein kinase kinase kinase 7
TAZ	Transcriptional coactivator with PDZ-binding motif

TF	Transcription factor
TRAF6	TNF receptor associated factor 6
TGFβ	Transforming growth factor beta
TGFβRI	Transforming growth factor beta receptor 1
TGFβRII	Transforming growth factor beta receptor 2
TNFα	Tumor necrosis factor alpha
TU	Technische Universität
UbCp	Ubiquitin-like domain-containing CTD phosphatase
UV	Ultraviolet radiation
Wip1	P53-Induced Protein Phosphatase 1
WPGMA	Unweighted Pair Group Method with Arithmetic mean
Wnt	P53-Induced Protein Phosphatase 1
XPC	Xeroderma pigmentosum, complementation group C
YAP	Yes-associated protein
YFP	Yellow fluorescence protein

15.3 List of Figures

List of Figures

15.4 Acknowledgements

First of all I want to thank Prof. Dr. Alexander Loewer not only for giving me the opportunity to work in his lab, but also for his support and understanding during the challenging times, for his trust, and for his patience with my questions and inestimable help during the writing of this thesis.

I also would like to thank Prof. Dr. Kay Hamacher for his efforts on my doctoral study and reviewing this thesis.

Very special thank go to my valued colleagues at the Loewer lab: Caibin Sheng, Petra Snyder, Anna Mittermeier, Laura Friedel, Stefan Bohn, Jette Strasen, Jennifer Mahr, Ana Finzel, Sabine Henke, Dhana Friedrich, Ilias Nolis and Ulrike Burk. It was a great time to having fun and working together at the MDC Berlin and the TU Darmstadt. We'll meet again.

I would also like to express my thanks to Prof. Dr. Stefan Legewie, Uddipan Sarma and Manuela Benary.

Finally, my deep and sincere gratitude to my family for their continuous and unparalleled love, help and support. This journey would not have been possible without the support of my three beloved girls Sandra, Emma and Elza. Without their enthusiasm, encouragement, support and continuous optimism this thesis would hardly have been completed.
I am grateful to my sister for always being there for me as a friend. I am forever indebted to my parents for giving me the opportunities and experiences that have made me who I am. They selflessly encouraged me to explore new directions in life and seek my own destiny. This journey would not have been possible if not for them, and I dedicate this milestone to them.

Curriculum Vitae

Marcel Jentsch
07.04.79

Education

Doctoral candidate
Technische Universität Darmstadt 2016-2019
Thesis: **Using Computational Classification Techniques to Examine Dynamic Rewiring of the Cellular Signalling**
Supervisors: Prof. Dr. A. Löwer and Prof. Dr. K. Hamacher

M.Sc. Bioinformatics
Freie Universität Berlin 2007-2011
Thesis: **Alternative splicing detection algorithms for Affymetrix exon array data – comprehensive evaluation and new methods based on Kullback Leibler divergence and nonparametric statistics**
Supervisors: Prof. Dr. U. Leser and Prof. Dr. M. Vingron

B.Sc. Bioinformatics
Freie Universität Berlin 2002-2007
Thesis: **Rekonstruktion regulativer zellulärer Hierarchien**
Supervisors: Prof. Dr. R. Spang and Prof. Dr. K. Reinert

Academic Employment

Research Assistant
Technische Universität Darmstadt 2016-2019
Systems Biology of the Stress Response

Max-Delbrueck-Center Berlin 2013-2015
Medical Systems Biology

Charité Berlin 2012-2013
Charité Comprehensive Cancer Center

Humboldt Universität Berlin 2010-2011
Knowledge Management in Bioinformatics

Student Assistant
Max-Planck-Institute for Human Development Center 2008-2010
Adaptive Behaviour and Cognition

Publications

Jacob, J. Jentsch, M. Kostka, D. Bentik, S. Spang, R.
Detecting hierarchical structure in molecular characteristics of disease using transitive approximations of directed graphs
Bioinformatics 2008, 24 (7): 995-1001

Eulenstein, S. Jentsch, M. Schlag, P.M.
Präklinische Validierung für das US-gestützte Navigationssystem LiverTack
11. CURAC, 2012

Zimmermann, K. Jentsch, M. Rasche, A. Hummel, M. Leser, U.
Algorithms for differential splicing detection using exon arrays: a comparative assessment
BMC Genomics 2015, 16(1):136

Strasen, J. Sarma, U. Jentsch, M. Bohn, S. Sheng, C. Horbelt, D. Knaus, P. Legewie, S. Löwer, A.
Cell-specific responses to the cytokine TGFβ are determined by variability in protein levels.
Molecular Systems Biology, 2018, 14(1)

Sheng, C. Mendler, I.H. Rieke, S. Snyder, P. Jentsch, M. Friedrich, D. Drossel, B. Löwer, A
PCNA-Mediated Degradation of p21 Coordinates the DNA Damage Response and Cell Cycle Regulation in Individual Cells
Cell Reports, 2019, 27 (1)

Jentsch, M. Snyder, P. Sheng, C. Cristiano, E. Löwer, A.
p53 dynamics in single cells are temperature-sensitive
Scientific Reports, 2020, 10, 1481

Ehrenwörtliche Erklärung:

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit entsprechend den Regeln guter wissenschaftlicher Praxis selbstständig und ohne unzulässige Hilfe Dritter angefertigt habe.

Sämtliche aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie sämtliche von Anderen direkt oder indirekt übernommenen Daten, Techniken und Materialien sind als solche kenntlich gemacht. Die Arbeit wurde bisher bei keiner anderen Hochschule zu Prüfungszwecken eingereicht.

Darmstadt, den 06.09.2019

References

1. *Principal Component Analysis*. Springer, New York, 2002.

2. Aberg, E., F. Saccoccia, M. Grabherr, W. Ore, P. Jemth and G. Hultqvist: *Evolution of the p53-MDM2 pathway*. BMC Evolutionary Biology, 17, 2017.

3. Abraham, R.: *Cell cycle checkpoint signaling through the ATM and ATR kinases*. Genes and Development, 15:2177–2196, 2001.

4. Adhikari, A., M. Xu and Z. J. Chen: *Ubiquitin-mediated activation of TAK1 and IKK*. Oncogene, 26:3214–3226, 2007.

5. Aghabozorgi, S., A. Seyed Shirkhorshidi and T. Ying Wah: *Time-series clustering – A decade review*. Information Systems, 53:16–38, 2015.

6. Ahmed, K., T. Yoshiaki and K. Takashi: *Hyperthermia: an effective strategy to induce apoptosis in cancer cells*. Apoptosis, 20:1411–1419, 2015.

7. Alberts, B., D. Bray, K. Hopkin, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter: *Essential Cell Biology, 4th Edition*. Garland Science, New York, 2014.

8. Aloise, D., A. Deshpande, P. Hansen and P. Popat: *NP-hardness of Euclidean sum-of-squares clustering*. Machine Learning, 75:245–248, 2009.

9. Aloni-Grinstein, R., Y. Shetzer, T. Kaufman and V. Rotter: *p53: the barrier to cancer stem cell formation*. FEBS Lett., 588:2580—2589, 2014.

10. Altschuler, S. J. and L. F. Wu: *Cellular Heterogeneity: Do Differences Make a Difference?*. Cell, 114, 2010.

11. Annes, J., J. Munger and D. Rifkin: *Making sense of latent TGFbeta activation*. J Cell Sci, 116:217–224, 2003.

12. Appella, E. and C. Anderson: *N Signaling to p53: breaking the posttranslational modifications code*. Pathol Biol, 48:227–245, 2000.

13. Ashall, L., C. Horton, D. Nelson, P. Paszek, C. Harper, K. Sillitoe, S. Ryan, D. Spiller, J. Unitt, D. Broomhead, D. Kell, D. Rand, V. See and M. White: *Pulsatile stimulation determines timing and specificity of NF-B-dependent transcription*. Cell, 324:242–246, 2009.

14. Ashburner, M., C. Ball, J. Blake, D. Botstein, H. Butler, J. Cherry, A. Davis, K. Dolinski, S. Dwight, J. Eppig, M. Harris, D. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. Matese, J. Richardson, M. Ringwald, G. Rubin and G. Sherlock: *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium*. Nat Genet, 1:25–29, 2000.

15. Ashwell, S. and S. Zabludoff: *DNA Damage Detection and Repair Pathways—Recent Advances with Inhibitors of Checkpoint Kinases in Cancer Therapy*. Clinical Cancer Research, 14, 2008.

16. Aylon, Y. and M. Oren: *The paradox of p53: what, how, and why?*. Cold Spring Harb. Perspect. Med., 6:6, 2016.

17. Baker, M.: *Cellular imaging: Taking a long, hard look*. Nature, 466:1137—-1140, 2010.

18. Baker, S.: *Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas*. Science, 244:217–221, 1989.

19. Baker, S.: *p53 gene mutations occur in combination with 17p allelic deletions as late events in colorectal tumorigenesis*. Cancer Res, 50:7717–7722, 1990.

20. Bakin, A., A. Tomlinson, N. Bhowmick, H. Moses and C. Arteaga: *Phosphatidylinositol 3-kinase function is required for transforming growth factor beta-mediated epithelial to mesenchymal transition and cell migration*. J Biol Chem, 275:36803–36810, 2000.

21. Bar-Or, R., R. Maya, L. Segel, U. Alon, A. Levine and M. Oren: *Generation of oscillation by the p53-Mdm2 feedback loop: A theoretical and experimental study*. Proc. Natl. Acad. Sci. U.S.A., 97:11250–11255, 2000.

22. Barak, Y., T. Juven, R. Haffner and M. Oren: *mdm2 expression is induced by wild type p53 activity*. Embo J, 12:461–468, 1993.

23. Bargonetti, J. and J. Manfredi: *Multiple roles of the tumor suppressor p53*. Perspectives in Biology, 14:86–91, 2002.

24. Barrios-Rodiles, M., K. Brown and B. Ozdamar: *High-throughput mapping of a dynamic signaling network in mam-malian cells*. Science, 307:1621–1625, 2005.

25. Bartek, J. and L. J.: *Mammalian G1- and S-phase checkpoints in response to DNA damage*. Curr Opin Cell Biol, 13:738–747, 2001.

26. Batchelor, E., A. Loewer and G. Lahav: *The ups and downs of p53: understanding protein dynamics in single cells*. Nat. Rev. Cancer, 9:371—-377, 2009.

27. Batchelor, E., A. Loewer, C. Mock and G. Lahav: *Stimulus-dependent dynamics of p53 in single cells*. Molecular Systems Biology, 7:488–488, 2011.

28. Batchelor, E., C. Mock, I. Bhan, A. Loewer and G. Lahav: *Recurrent initiation: a mechanism for triggering p53 pulses in response to DNA damage*. Mol Cell, 30:277–289, 2008.

29. Behar, M. and A. Hoffmann: *Understanding the Temporal Codes of Intracellular Signals*. Curr Opin Genet Dev, 20:684—693, 2010.

30. Bellman, R.: *The theory of dynamic programming*. Bulletin of the American Mathematical Society, 60:503—-516, 1954.

31. Belyi, V., P. Ak, E. Markert, H. Wang, W. Hu, A. Puzio-Kuter and L. A.J.: *The Origins and Evolution of the p53 Family of Genes*. Perspectives in Biology, 2 no. 6, 2010.

32. Benson, E., S. Mungamuri, O. Attie, M. Kracikova, R. Sachidanandam, J. Manfredi and S. Aaronson: *p53-dependent gene repression through p21 is mediated by recruitment of E2F4 repression complexes*. Oncogene, 33:3959—-3969, 2014.

33. Bergamaschi, D.: *iASPP oncoprotein is a key inhibitor of p53 conserved from worm to human*. Nature Genet, 33:162–167, 2003.

34. Bergamaschi, D.: *iASPP preferentially binds p53 proline-rich region and modulates apoptotic function of codon 72-polymorphic p53*. Nature Genet, 38:1133–1141, 2006.

35. Berndt, D. and J. Clifford: *Using dynamic time warping to find patterns in time series.*. Proc. 3rd Int. Conf. Knowl. Discov. Data Min., pp. 359–370, 1994.

36. Bierie, B. and H. Moses: *Tumour microenvironment: TGFβ: the molecular Jeckyll and Hyde of cancer*. Nature Reviews Cancer, 6 no. 7:506–520, 2006.

37. Biernacka, A., M. Dobaczewski and N. Frangogiannis: *TGF-beta signaling in fibrosis*. Growth Factors, 29:196–202, 2011.

38. Bishop, C. M.: *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006.

39. Blobe, G., W. Schiemann and H. Lodish: *Role of transforming growth factor beta in human disease*. N Engl J Med., 342:1350–1358, 2000.

40. Bode, A. and Z. Dong: *Post-translational modification of p53 in tumorigenesis*. Nat Rev Cancer, 4:793–805, 2004.

41. Bohn, S.: *Investigating molecular mechanisms mediating stimulus-dependent SMAD signaling in single cells*. Master's thesis, TU Darmstadt, 2016.

42. Bourdon, J.: *p53 family isoforms*. Curr. Pharm. Biotechnol, 8:332—-336, 2007.

43. Bourdon, J., K. Fernandes, F. Murray-Zmijewski, G. Liu, A. Diot, D. Xirodimas, M. Saville and D. Lane: *p53 isoforms can regulate p53 transcriptional activity*. Genes Dev, 19:2122–2137, 2005.

44. Bradshaw, R. and E. Dennis: *Handbook of Cell Signaling, 2nd Edition*. Science Direct, 2009.

45. Brendel, B., S. Winter, A. Rick, M. Stockheim and H. Ermert: *Bone registration with 3D CT and ultrasound datasets*. CARS 2003. Computer Assisted Radiology and Surgery. Proceedings of the 17th International Congress and Exhibition, 1256:426–432, 2003.

46. Brown, L.: *A survey of image registration techniques*. ACM Comput. Surv., 24:325–376, 1992.

47. Buettner, F., K. Natarajan, F. Casale, A. Scialdone, F. Theis, S. Teichmann, J. Marioni and O. Stegle: *Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells*. Nature Biotechnology, 33:155–160, 2015.

48. Bunz, F.: *Requirement for p53 and p21 to sustain G2 arrest after DNA damage*. Science, 282:1497—1501, 1998.

49. Byrne, C. and C. Lim: *The ingestible telemetric body core temperature sensor: a review of validity and exercise applications*. Br. J. Sports Med., 41:126–133, 2007.

50. Cantwell-Dorris, E., J. OLeary and O. Sheils: *BRAFV600E: implications for carcinogenesis and molecular therapy*. Mol Cancer Ther, 10:385–94, 2011.

51. Carpenter, A., T. Jones, M. Lamprecht, C. Clarke, I. Kang, O. Friman, D. Guertin, J. Chang, R. Lindquist, J. Moffat, P. Golland and D. Sabatini: *CellProfiler: image analysis software for identifying and quantifying cell phenotypes*. Genome Biology, 7:R100, 2006.

52. Cech, E.: *Point Sets*. New York: Academic Press, p. 42, 1969.

53. Chai, Y., J. Cui, N. Shao, E. Reddy and V. Rao: *The second BRCT domain of BRCA1 proteins interacts with p53 and stimulates transcription from the p21(WAF1/CIP1) promoter*. Oncogene, 18:263–268, 1999.

54. *Information Brochure for Patients Hyperthermia*, 2016. Department of Radiation Oncology and Radiotherapy.

55. Chen, R., Y. Su, R. Chuang and T. Chang: *Suppression of transforming growth factor-beta-induced apoptosis through a phosphatidylinositol 3-kinase/Akt-dependent pathway*. Oncogene, 17:1959–1968, 1998.

56. Chen, Z. andf Yoo, S., Y. Park, K. Kim, S. Wei, E. Buhrd, Z. Ye, H. Pan and J. Takahashi: *Identification of diverse modulators of central and peripheral circadian clocks by high-throughput chemical screening*. Proc. Natl. Acad. Sci., 109:101–106, 2011.

57. Chen, X., J. Chen, S. Gan, H. Guan, Y. Zhou, Q. Ouyang and J. Shi: *DNA damage strength modulates a bimodal switch of p53 dynamics for cell-fate control*. BMC Biol., 11:1440—-1444, 2013.

58. Chen, Y. and X. Wang: *A special issue on TGFβ signaling.* Cell Research, 19:1–2, 2009.

59. Chui, H. and A. Rangarajan: *A new point matching algorithm for non-rigid registration.* Comput. Vis. Image Understand., 89:114–141, 2003.

60. Cihoric, N., A. Tsikkinis, G. van Rhoon, H. Crezee, D. Aebersold, S. Bodis, M. Beck, J. Nadobny, V. Budach, P. Wust and P. Ghadjar: *Hyperthermia-related clinical trials on cancer treatment within the ClinicalTrials.gov registry.* Int J of Hyperthermia, 31:609 – 614, 2015.

61. Ciznadija, D., X. Zhu, and A. Koff: *Hdm2- and proteasome-dependent turnover limits p21 accumulation during S phase.* Cell Cycle, 10:2714—-2723, 2011.

62. Clarke, D., M. Brown, R. Erickson, Y. Shi and X. Liu: *Transforming growth factor beta depletion is the primary determinant of Smad signaling kinetics.* Mol Cell Biol, 29:2443 —- 2455, 2009.

63. Cohen, A., N. Geva-Zatorsky, E. Eden, M. Frenkel-Morgenstern, I. Issaeva, A. Sigal, R. Milo, C. Cohen-Saidon, Y. Liron, Z. Kam, L. Cohen, T. Danon, N. Perzov and U. Alon: *Dynamic proteomics of individual cancer cells in response to a drug.* Science, 322:1511–1515, 2008.

64. Cohen-Saidon, C., A. Cohen, A. Sigal, Y. Liron and U. Alon: *Dynamics and variability of ERK2 response to EGF in individual living cells.* Mol. Cell, 36:885–893, 2009.

65. Conery, A., Y. Cao and E. Thompson: *Akt interacts directly with Smad3 to regulate the sensitivity to TGF-beta induced apoptosis.* Nat Cell Biol, 6:366–372, 2004.

66. Cordani, M. Butera, G., R. Pacchiana and M. Donadelli: *Molecular interplay between mutant p53 proteins and autophagy in cancer cells.* Biochimica et Biophysica Acta (BBA) - Reviews on Cancer, 1:19–28, 2017.

67. Cordenonsi, M.: *Links between tumor suppressors: p53 is required for TGF-β gene responses by cooperating with Smads.* Cell, 113:301–314, 2003.

68. Cuturi, M.: *Links between tumor suppressors: p53 is required for TGF-β gene responses by cooperating with Smads.* In Proceedings of the 28th international conference on machine learning (ICML-11), 113:926–936, 2011.

69. Cuturi, M. and M. Blondel: *Soft-DTW: a Differentiable Loss Function for Time-Series.* Proceedings of the 34th International Conference on Machine Learning, 70:894–903, 2017.

70. Das, S.: *Hzf determines cell survival upon genotoxic stress by modulating p53 transactivation.* Cell, 130:624–637, 2007.

71. De Haas-Kock, D., J. Buijsen, M. Pijls-Johannesma, L. Lutgens, G. Lammering and G. van Mastrigt: *Concomitant hyperthermia and radiation therapy for treating locally advanced rectal cancer.* Cochrane Database Syst Rev, 3, 2009.

72. DeLeo, A. e. a.: *Detection of a transformation-related antigen in chemically induced sarcomas and other transformed cells of the mouse*. Proceedings of the National Academy of Sciences of the United States of America, 76:2420–2424, 1979.

73. Dempster, A., N. Laird and D. Rubin: *Maximum likelihood from incomplete data via the EM algorithm*. JOURNAL OF THE ROYAL STATISTICAL SOCIETY, pp. 1–38, 1977.

74. Deng, C., P. Zhang, J. Harper, S. Elledge and P. Leder: *Mice lacking p21CIP1/WAF1 undergo normal development, but are defective in G1 checkpoint control*. Cell, 82:675—-684, 1995.

75. Dennler, S. and P. Dijke: *Smad Proteins in TGF-Beta Signaling*. Encyclopedia of Cancer, pp. 294–345, 2011.

76. Deroo, S. and K. Korach: *SEstrogen receptors and human disease*. Journal of Clinical Investigation, pp. 561–570, 2006.

77. Derynck, R. and R. Akhurst: *Differentiation plasticity regulated by TGF-beta family proteins in development and disease*. Nat Cell Biol, 9:1000–1004, 2007.

78. Derynck, R. and Y. Zhang: *Smad-dependent and SMAD-independent pathways in TGFβ familiy signalling*. Nature, 425 no. 6958:2078–2093, 2003.

79. Dewey, W.: *rrhenius relationships from the molecule and cell to the clinic*. A. Int J Hyperthermia, 10:457—-483, 1994.

80. DiGuglielmo, G., C. LeRoy, A. Goodfellow and J. Wrana: *Distinct endocytic pathways regulate TGFbeta eceptor signalling and turnover*. Nature Cell Biology, 5:410–417, 2003.

81. Dijkstra, E. W.: *A note on two problems in connexion with graphs*. Numerische Mathematik, 1:269—-271, 1959.

82. Donehower, L.: *Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours*. Nature, 356:215–221, 1992.

83. Du, P., S. M. Lin and W. A. Kibbe: *Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching*. Bioinformatics, 22(17):2059–2065, 07 2006.

84. Ebisawa, T., M. Fukuchi, M. G., T. Chiba, K. Tanaka, T. Imamura and K. Miyazono: *Smurf1 interacts with transforming growth factor-beta type I receptor through Smad7 and induces receptor degradation*. J Biol Chem, 276:12477–12480, 2001.

85. Edlund, S., S. Bu, N. Schuster, P. Aspenstrom, R. Heuchel, N. Heldin, P. Dijke, C. Heldin and M. Landstrom: *Transforming growth factor-beta1 (TGF-beta)-induced apoptosis of prostate cancer cells involves Smad7-dependent activation of p38 by TGF-beta-activated kinase 1 and mitogen-activated protein kinase kinase 3*. Mol Biol Cell, 14:529–544, 2003.

86. Edlund, S., S. Lee, S. Grimsby, S. Zhang, P. Aspenstrom, C. Heldin and M. Landstrom: *Interaction between Smad7 and beta-catenin: importance for transforming growth factor beta-induced apoptosis.* Mol Cell Biol, 25:1475–1488, 2005.

87. El-Deiry, W.: *WAF1, a potential mediator of p53 tumor suppression.* Cell, 75:817–825, 1993.

88. El-Deiry, W., S. Kern, J. Pietenpol, K. Kinzler and B. Vogelstein: *Definition of a consensus binding site for p53.* Nat Genet, 1:45–49, 1992.

89. Eliyahu, D., D. Michalovitz, S. Eliyahu, O. Pinhasi-Kimhi and M. Oren: *Wild-type p53 can inhibit oncogene-mediated focus formation.* Proc Natl Acad Sci U S A, 86:8763–8767, 1989.

90. Eliyahu, D. e. a.: *Participation of p53 cellular tumour antigen in transformation of normal embryonic cells.* Nature, 312:646—649, 1984.

91. Elliot, R. and G. Blobe: *Role of transforming groth factor factor β in human cancer.* Journal of Clinical Oncology, 23 no. 9:2078–2093, 2005.

92. Elston, R. and G. Inman: *Crosstalk between p53 and TGF-β Signalling.* Journal of Signal Transduction, 2012, 2012.

93. Espinosa, J. M.: *Mechanisms of regulatory diversity within the p53 transcriptional network.* Oncogene, 27:4013–4023, 2008.

94. Ester, M., H. Kriegel, J. Sander and X. Xu: *A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise.* Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, pp. 226–231, 1996.

95. Feinerman, O., J. Veiga, J. Dorfman, R. Germain and G. Altan-Bonnet: *Variability and robustness in T cell activation from regulated heterogeneity in protein levels.* Science, 321:1081 — 1084, 2008.

96. Feng, X. and R. Derynck: *Specificity and versatility in tgf-beta signaling through Smads.* Annu Rev Cell Dev Biol, 21:659–693, 2005.

97. Finlay, C., P. Hinds and A. Levine: *The p53 proto-oncogene can act as a suppressor of transformation.* Cell, 57:1083–1093, 1989.

98. Finzel, A.: *Upstream control and downstream responses of p53 are involved in its tumor suppression functions upon genotoxic stress.* PhD thesis, Freie Univerity Berlin, 2016.

99. Finzel, A., A. Grybowski, J. Strasen, E. Cristiano and A. Loewer: *Hyperactivation of ATM upon DNA-PKcs inhibition modulates p53 dynamics and cell fate in response to DNA damage.* Mol. Biol. Cell, 27:2360–2367, 2016.

100. Fischer, M.: *Census and evaluation of p53 target genes.* Oncogene, 36:3943–3956, 2017.

101. Fischer, M., P. Grossmann, M. Padi and J. DeCaprio: *Integration of TP53, DREAM, MMB-FOXM1 and RB-E2F target gene analyses identifies cell cycle gene regulatory networks*. Nucleic Acids Res, 44:6070–6086, 2016.

102. Fischer, M., L. Steiner and K. Engeland: *The transcription factor p53: not a repressor, solely an activator*. Cell Cycle, 13:3037–3058, 2014.

103. Fitzgibbon, A.: *Robust registration of 2D and 3D point sets*. Image and Vision Computing, 21:1145–1153, 2003.

104. Fridman, J. and S. Lowe: *Control of apoptosis by p53*. Journal of Signal Transduction, 22:9030–9040, 2003.

105. Fu, T.: *A review on time series data mining*. Engineering Applications of Artificial Intelligence, 24:164–181, 2011.

106. Fujimoto, H., N. Onishi, N. Kato, M. Takekawa, X. Xu, A. Kosugi, T. Kondo, M. Imamura, I. Oishi, A. Yoda and Y. Minami: *Regulation of the antioncogenic Chk2 kinase by the oncogenic Wip1 phosphatase*. Cell Death Differ, 13:1170—-1180, 2006.

107. Funaba, M., C. Zimmerman and L. Mathews: *Modulation of Smad2-mediated signaling by extracellular signal-regulated kinase*. J Biol Chem, 277:41361–41368, 2002.

108. Furusawa, Y., T. Iizumi, Y. Fujiwara, Q. Zhao, Y. Tabuchi and T. Nomura: *Inhibition of checkpoint kinase 1 abrogates G2/M checkpoint activation and promotes apoptosis under heat stress*. Apoptosis, 17:102—-112, 2012.

109. Galliher, A. and W. Schiemann: *Src phosphorylates Tyr284 in TGF-beta type II receptor and regulates TGF-beta stimulation of p38 MAPK during breast cancer cell proliferation and invasion*. Cancer Res, 67:3752–3758, 2007.

110. Gao, M., P. Monian, Q. Pan, W. Zhang, J. Xiang and X. Jiang: *Ferroptosis is an autophagic cell death process*. Cell Res., 26:1021–1032, 2016.

111. Garrido, C., L. Galluzzi, M. Brunet, P. Puig, C. Didelot and G. Kroemer: *Mechanisms of cytochrome c release from mitochondria*. Cell Death and Differentiation, 13:1423–1433, 2006.

112. Geva-Zatorsky, N., N. Rosenfeld, S. Itzkovitz, R. Milo, A. Sigal, E. Dekel, T. Yarnitzky, Y. Liron, P. Polak, G. Lahav and U. Alon: *Oscillations and variability in the p53 system*. Mol Syst Biol, 2:2006–0033, 2006.

113. Gingery, A., B. E.W., L. Pederson, M. Ruan, N. Horwood and M. Oursler: *TGF-beta coordinately activates TAK1/MEK/AKT/NFkB and SMAD pathways to promote osteoclast survival*. Exp Cell Res, 314:2725–2738, 2008.

114. Gisolfi, C. and F. Mora: *The hot brain survival temperature and the human body*. MIT Press, 2000.

115. Goh, A., C. Coffill and D. Lane: *The role of mutant p53 in human cancer*. Journal of Pathology, 223 no. 2:116–126, 2011.

116. Gonzalez, D. and D. Medici: *Signaling mechanisms of the epithelial-mesenchymal transition*. Sci Signal, 7, 2014.

117. Gordon, K. and G. Blobe: *Role of transforming growth factor-beta superfamily signaling pathways in human disease*. Biochim Biophys Acta, 1782:197–228, 2008.

118. Groenroos, E., U. Hellman, C. Heldin and J. Ericsson: *Control of Smad7 stability by competition between acetylation and ubiquitination*. Biochim Biophys Acta, 10:483–493, 2002.

119. Gross, W.: *Grundzüge der Mengenlehre*. Monatshefte für Math. und Phys., 26:A34–A35, 1915.

120. Guan, J., E. Stavridi, D. Leeper and G. Iliakis: *Effects of hyperthermia on p53 protein expression and activity*. Journal of cellular physiology, 190:365–374, 2002.

121. Guilherme, L. and J. Kalil: *Rheumatic Fever and Rheumatic Heart Disease: Cellular Mechanisms Leading Autoimmune Reactivity and Disease*. J. Clin. Immunol., 30:17–23, 2010.

122. Gurdon, J., S. Dyson and D. StJohnston: *Cells' perception of position in a concentration gradient*. Cell, 95:159–162, 1998.

123. Haghverdi, L., F. Buettner and F. Theis: *Diffusion maps for high-dimensional single-cell analysis of differentiation data*. Bioinformatics, 31:2989–2998, 2015.

124. Hallikas, O., K. Palin, N. Sinjushina and et al.: *Genome-wide Prediction of Mammalian Enhancers Based on Analysis of Transcription-Factor Binding Affinity*. Cell, 124 no. 1:47–59, 2006.

125. Han, H.: *The estrogen receptor alpha pathway induces oncogenic Wip1 phosphatase gene expression*. Mol Cancer Res, 7:713–723, 2009.

126. Han, J. and M. Kamber: *Data mining: concepts and techniques*. Morgan Kaufmann, Waltham, USA, 2005.

127. Handl, J., J. Knowles and D. Kell: *Computational cluster validation in post-genomic data analysis*. Bioinformatics, 21:3201–3212, 2005.

128. Harper, C., D. Woodcock, C. Lam, M. Garcia-Albornoz, A. Adamson, L. Ashall, W. Rowe, P. Downton, L. Schmidt, S. West, D. Spiller, D. Rand and M. White: *Temperature regulates NF-kappaB dynamics and function through timing of A20 transcription*. Proc. Natl. Acad. Sci. U. S. A., 115:5243—-5249, 2018.

129. Harris, S. and A. Levine: *The p53 pathway: Positive and negative feedback loops*. Oncogene, 24:2899–2908, 2005.

130. Hastie, T., R. Tibshirani and J. Friedman: *In The Elements of Statistical Learning*. Springer-Verlag, 2009.

131. Hata, A.: *TGF-beta signalling from receptors to Smads*. Cold Spring Harb Perspect Biol., 8, 2016.

132. Hauck, M. and M. Zalutsky: *Enhanced tumour uptake of radiolabelled antibodies by hyperthermia: Part I: Timing of injection relative to hyperthermia*. Int. J. Hyperth., 21:1–11, 2005.

133. Haupt, Y., R. Maya, A. Kazaz and O. M.: *Mdm2 promotes the rapid degradation of p53*. Nature, 387:296–299, 1997.

134. Hearnes, J. M.: *Chromatin immunoprecipitation-based screen to identify functional genomic binding sites for sequence-specific transactivators*. Mol. Cell, 25:10148–10158, 2005.

135. Heldin, C., M. Landstrom and A. Moustakas: *Mechanism of TGF-β signaling to growth arrest, apoptosis and epithelial-mesenchymal transition*. Current Option in Cell Biology, 21 no. 2:166–176, 2009.

136. Heldin, C., K. Miyazono and P. Dijke: *GF-beta signalling from cell membrane to nucleus through SMAD proteins*. Nature, 390:465–471, 1997.

137. Hildebrandt, B., P. Wust, O. Ahlers, A. Dieing, G. Sreenivasa, T. Kerner, R. Felix and H. Riess: *The cellular and molecular basis of hyperthermia*. Critical Reviews in Oncology Hematology, 43:33–46, 2002.

138. Ho, J. and S. Benchimol: *Transcriptional repression mediated by the p53 tumour suppressor*. Cell Death Differ, 10:404–408, 2003.

139. Hoffman, W. H., S. Biade, J. T. Zilfou, J. Chen and M. Murphy: *Transcriptional repression of the anti-apoptotic survivin gene by wild type p53*. J. Biol. Chem., 277:3247–3257, 2002.

140. Holleman, A., E. Wiberg and N. Wiberg: *Anorganische ChemienBand 1, Grundlagen und Hauptgruppenelemente*.

141. Holzmann, H., A. Munk, M. Suster and W. Zucchini: *Hidden Markov models for circular and linear-circular time series*. Environmental and Ecological Statistics, 13:325—347, 2006.

142. Homer, C.: *Y-box factor YB1 controls p53 apoptotic function*. Oncogene, 24:8314–8325, 2005.

143. Honda, R., H. Tanaka and H. Yasuda: *Oncoprotein MDM2 is a ubiquitin ligase E3 for tumor suppressor p53*. FEBS Lett, 420:25–27, 1997.

144. Horaud, R., F. Forbes, M. Yguel, G. Dewaele and J. Zhang: *Rigid and articulated point registration with expectation conditional maximization*. IEEE Trans. Pattern Anal. Mach. Intell., 33:587–602, 2011.

145. Horbelt, D., A. Denkis and P. Knaus: *A portrait of Transforming Growth Factor β superfamily signalling: Background matters*. Int. J. Biochem. Cell Biol., 44:469–74, 2012.

146. Huang, W., X. Cao, F. H. Biase, P. Yu and S. Zhong: *Time-variant clustering model for understanding cell fate decisions*. Proceedings of the National Academy of Sciences, 111(44):E4797–E4806, 2014.

147. Huang da, W., B. T. Sherman and R. A. Lempicki: *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources*. Nature Protoc, 4:44–57, 2009.

148. Huminiecki, L., L. Goldovsky, S. Freilich, A. Moustakas, C. Ouzounis and C. Heldin: *Emergence, development and diversification of the TGF-beta signalling pathway within the animal kingdom.*. BMC Evol Biol, 9, 2009.

149. Hunt, C., R. Pandita, A. Laszlo, R. Higashikubo, M. Agarwal and T. Kitamura: *Hyperthermia activates a subset of ataxia-telangiectasia mutated effectors independent of DNA strand breaks and heat shock protein 70 status.s*. Cancer Res., 67:3010–3017, 2007.

150. Ikushima, H. and K. Miyazono: *TGFβ signalling: a complex web in cancer progression*. Nature Reviews Cancer, 10 no. 6:415–424, 2010.

151. Inman, G.: *Switching TGFβ from a tumor suppressor to a tumor promoter*. Current Opinion in Genetics and Development, 21 no. 1:93–99, 2011.

152. Inman, G., F. Nicolás and C. Hill: *Nucleocytoplasmic shuttling of Smads 2, 3, and 4 permits sensing of TGF-beta receptor activity*. Mol Cell, 10:283 —- 294, 2002.

153. Itoh, S. and P. ten Dijke: *Negative regulation of TGFb receptor/Smad signal transduction*. Curr Opin Cell Biol, 19:176–184, 2007.

154. Izzi L, A. L.: *Regulation of the TGFb signalling pathway by ubiquitin-mediated degradation*. Oncogene, 23:2071–2078, 2004.

155. Jegga, A., A. Inga, D. Menendez, B. Aronow and A. Resnick: *Functional evolution of the p53 regulatory network through its target response elements*. Proceedings of the national academy of science, 103 no. 3:944–949, 2008.

156. Jenkins, J., K. Rudge and G. Currie: *Cellular immortalization by a cDNA clone encoding the transformation- associated phosphoprotein p53*. Nature, 312:651–654, 1984.

157. Joerger, A. and A. Fersht: *Structural Biology of the Tumor Suppressor p53 and Cancer-Associated Mutants*. Adv. Cancer Res, 97:1–23, 2007.

158. Joerger, A. and A. Fersht: *The p53 Pathway: Origins, Inactivation in Cancer, and Emerging Therapeutic Approaches*. Annual Review of Biochemistry, 85:375–404, 2016.

159. Jordan, J., E. Landau and R. Iyengar: *Signaling networks: the origins of cellular multitasking*. Cell, 103:103–200, 2000.

160. Jung, H. and Y. Seo: *Protective effects of thioredoxin-mediated p53 activation in response to mild hyperthermia*. Oncol Rep., 27:650—-656, 2012.

161. Jung, Y., Y. Qian and X. Chen: *Examination of the expanding pathways for the regulation of p21 expression and activity*. Cellular Signalling, 22:1003–1012, 2010.

162. Kampinga, H. and A. Konings: *Inhibition of repair of X-ray-induced DNA damage by heat: the role of hyperthermic inhibition of DNA polymerase alpha activity*. Radiat. Res., 112 no. 1:86—-98, 1987.

163. Kantz, H. and T. Schreiber: *Nonlinear time series analysis*. University Press, Cambridge, UK, 2004.

164. Kastenhuber, E. and S. Lowe: *Putting p53 in Context*. Cell, 170:1062—-1078, 2017.

165. Katsuno, Y., S. Lamouille and R. Derynck: *TGF-beta signaling and epithelial–mesenchymal transition in cancer progression*. Current Opinion in Oncology, 25:76–84, 2013.

166. Kavsak, P., R. Rasmussen and C. Causing: *Smad7 binds to Smurf2 to form an E3 ubiquitin ligase that targets the TGF β receptor for degradation*. Mol Cell, 6:1365–1375, 2000.

167. Keogh, E. and M. Pazzani: *Scaling up dynamic time warping for datamining applications*. In Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '00, pp. 285–289, 2000.

168. Kern, S.: *Identification of p53 as a sequence-specific DNA-binding protein*. Science, 252:1708–1711, 1991.

169. Kidokoro, T., C. Tanikawa, Y. Furukawa, T. Katagiri, Y. Nakamura and K. Matsuda: *CDC20, a potential cancer therapeutic target, is negatively regulated by p53*. Oncogene, 27:1562–1571, 2008.

170. Kim, K., B. Kim, Z. Xu and S. Kim: *Mixed lineage kinase 3 (MLK3)-activated p38 MAP kinase mediates transforming growth factor-beta-induced apoptosis in hepatoma cells*. J Biol Chem, 279:29478–29484, 2004.

171. Kinuya, S., K. Yokoyama, T. Hiramatsu and N. Tonami: *Optimal Timing of Administration of Hyperthermia in Combined Radioimmunotherapy*. Cancer Biother. Radiopharm., 15:373—-379, 2000.

172. Kowanetz, M., P. Lonn, M. Vanlandewijck, K. Kowanetz, C. Heldin and A. Moustakas: *TGFbeta induces SIK to negatively regulate type I receptor kinase signaling*. J Cell Biol, 182:655–662, 2008.

173. Kress, M., E. May, R. Cassingena and P. May: *Simian virus 40-transformed cells express new species of proteins precipitable by anti-simian virus 40 tumor serum*. J Virol, 31:472–483, 1979.

174. Kretzschmar, M., J. Doody and J. Massague: *Opposing BMP and EGF signalling pathways converge on the TGF-beta family mediator Smad1*. Nature, 389:618–622, 1997.

175. Kretzschmar, M., J. Doody and J. Timokhina, I. Massague: *A mechanism of repression of TGFbeta/ Smad signaling by oncogenic Ras*. Genes Dev, 13:804–816, 1999.

176. Kruiswijk, F., C. Labuschagne and K. Vousden: *P53 in Survival, Death and Metabolic Health: a Lifeguard With a Licente To Kill*. Nat Rev Mol Cell Biol, 16:393—-405, 2015.

177. Kruse, J. and W. Gu: *Modes of p53 Regulation*. Cell, 137:609—-622, 2009.

178. Kubbutat, M., S. Jones and K. Vousden: *Regulation of p53 stability by Mdm2*. Genes Dev, 387:299–303, 1997.

179. Kumari, R., S. Kohli and S. Das: *p53 regulation upon genotoxic stress: intricacies and complexities*. Mol. Cell. Oncol., p. e969653, 2014.

180. Lahav, G., A. Rosenfeld, N. Sigal, N. Geva-Zatorsky, A. Levine, M. Elowitz and U. Alon: *Dynamics of the p53-Mdm2 feedback loop in individual cells*. Nat. Genet., 36:147—-150, 2004.

181. Lamouille, S. and R. Derynck: *Cell size and invasion in TGF-beta-induced epithelial to mesenchymal transition is regulated by activation of the mTOR pathway*. J Cell Biol, 178:437–451, 2007.

182. Landstrom, M.: *he TAK1-TRAF6 signalling pathway*. Int J Biochem Cell Biol, 42:585–589, 2010.

183. Lane, D.: *p53, guardian of the genome*. Nature, 358 no. 6381:15–16, 1992.

184. Lane, D. and L. Crawford: *T antigen is bound to a host protein in SV40-transformed cells*. Nature, 278:261–263, 1979.

185. Lawler, S., X. Feng and R. Chen: *The type II transforming growth factor-beta receptor autophosphorylates not only on serine and threonine but also on tyrosine residues*. J Biol Chem, 272:14850–14859, 1997.

186. Lee, C., J. Ferreon, A. Ferreon, M. Arai and P. Wright: *Graded enhancement of p53 binding to CREB-binding protein (CBP) by multisite phosphorylation*. Proc Natl Acad Sci, 107:19290–19295, 2010.

187. Lee, J. and T. Paull: *Direct activation of the ATM protein kinase by the Mre11/ Rad50/Nbs1 complex*. Science, 304:93—-96, 2004.

188. Lee, M., C. Pardoux and M. Hall: *TGF-beta activates Erk MAP kinase signalling through direct phosphorylation of ShcA*. Proc Natl Acad Sci, 26:3957–3967, 2007.

189. Legewie, S., H. Herzel, H. Westerhoff and N. Blüthgen: *Recurrent design patterns in the feedback regulation of the mammalian signalling network.*. Mol Syst Biol, 4:190, 2008.

190. LeGuellec, S., F. Chibon, M. Ouali and A. Neuville: *Are Peripheral Purely Undifferentiated Pleomorphic Sarcomas With MDM2 Amplification Dedifferentiated Liposarcomas?*. Am. J. Surg. Pathol., 38:293—304, 2014.

191. Lemon, B. and R. Tjian: *Orchestrated response: a symphony of transcription factors for gene control*. Genes and Developmentz, 14:2551–2569, 2000.

192. Levine, A.: *Tp53, the cellular gatekeeper for growth and division*. Cell, 88:323–331, 1997.

193. Levine, A. and M. Oren: *The first 30 years of p53: growing ever more complex*. Nature Reviews Cancer, 9 no. 10:749–758, 2009.

194. Levy, L. and C. Hill: *Alterations in components of the TGF-β superfamily signaling pathways in human cancer cells*. Cytokine and Groth Factor Reviews, 17 no. 1-2:41–58, 2006.

195. Li, D., N. Marchenko, R. Schulz, V. Fischer, V. VelascoHernandez, F. Talos and U. Moll: *Functional Inactivation of Endogenous MDM2 and CHIP by HSP90 Causes Aberrant Stabilization of Mutant p53 in Human Cancer Cells*. Mol. Cancer Res., 9:577—-588, 2011.

196. Liao, T.: *Clustering of time series data—a survey*. Pattern Recognition, 38:1857–1874, 2005.

197. Lim, C., C. Byrne and J. Lee: *Human thermoregulation and measurement of body temperature in exercise and clinical settings*. Ann. Acad. Med. Singapore, 37:347–353, 2008.

198. Linzer, D. and A. Levine: *Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells*. Cell, 17:43–52, 1979.

199. Lipkin, B.: *Picture Processing and Psychopictorics*. Elsevier Science, 1970.

200. Livingstone, L., A. White, J. Sprouse, E. Livanos, T. Jacks and T. Tlsty: *Altered cell cycle arrest and gene amplification potential accompany loss of wild-type p53*. Cell, 70:923–935, 1992.

201. Loehr, K., C. Mo, A. Contente and M. Dobbelstein: *p21/CDKN1A mediates negative regulation of transcription by p53*. J Biol Chem, 278:32507–32516, 2003.

202. Loenn, P., A. Moren, E. Raja, M. Dahl and A. Moustakas: *Regulating the stability of TGFβ receptors and Smads*. Cell Research, 19:21–35, 2009.

203. Loewer, A., E. Batchelor, G. Gaglia and G. Lahav: *Basal dynamics of p53 reveal transcriptionally attenuated pulses in cycling cells*. Cell, 142:89–100, 2010.

204. Loewer, A. and G. Lahav: *We are all individuals: causes and consequences of nongenetic heterogeneity in mammalian cells*. Curr Opin Genet Dev, 21:753 — 758, 2011.

205. Lowe, S., T. Jacks, D. Housman and H. Ruley: *Abrogation of oncogene-associated apoptosis allows transformation of p53-deficient cells*. Proc. Natl. Acad. Sci. USA, 91:2026–2030, 1994.

206. Lu, M. and B. Arrick: *Transactivation of the p21 promoter by BRCA1 splice variants in mammary epithelial cells: Evidence for both common and distinct activities of wildtype and mutant forms*. Oncogene, 19:6351–6360, 2000.

207. Lu, X.: *The Wip1 Phosphatase acts as a gatekeeper in the p53-Mdm2 autoregulatory loop*. Cancer Cell, 12:342–354, 2007.

208. Lutgens, L., J. van der Zee, M. Pijls-Johannesma, D. De Haas-Kock, J. Buijsen and G. Mastrigt: *Combined use of hyperthermia and radiation therapy for treating locally advanced cervix carcinoma*. Cochrane Database Syst Rev, 3, 2010.

209. Ma, J., J. Zhao and A. Yuille: *Non-Rigid Point Set Registration by Preserving Global and Local Structures*. IEEE Transactions on Image Processing, 25:53–64, 2015.

210. Ma, J., H. Zhou, J. Zhao, Y. Gao, J. Jiang and J. Tian: *Robust feature matching for remote sensing image registration via locally linear transforming*. IEEE Trans. Geosci. Remote Sens., 53:6469–6481, 2015.

211. Mahadevan, S. and G. Garmel: *An introduction to clinical emergency medicine*. Cambridge University Press, 2005.

212. Maier, B., W. Gluba, B. Bernier, T. Turner, M. K., T. Guise, A. Sutherland, M. Thorner and H. Scrable: *Modulation of mammalian life span by the short isoform of p53*. Genes Dev., 18:306—-319, 2004.

213. Malkin, D.: *Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms*. Cell Research, 250:1233–1238, 1990.

214. Mantovani, F.: *The prolyl isomerase Pin1 orchestrates p53 acetylation and dissociation from the apoptosis inhibitor iASPP*. Nature Struct. Mol. Biol, 14:912–920, 2007.

215. Mantso, T., G. Goussetis, R. Franco, S. Botaitis, A. Pappa and M. Panayiotidis: *Effects of hyperthermia as a mitigation strategy in DNA damage-based cancer therapies*. Seminars in Cancer Biology, 37:96–105, 2016.

216. Maska, M. and at. al.: *A benchmark for comparison of cell tracking algorithms*. Bioinformatics, 30:1609–1617, 2014.

217. Massague, J.: *TGF-beta signal transduction*. Annu Rev Biochem, 67:753–791, 1998.

218. Massague, J.: *TGFβ in cancer*. Cell, 134 no. 2:215–230, 2008.

219. Massague, J., J. Seoane and D. Wotton: *Smad transcription factors*. Genes Dev, 19:2783–2810, 2005.

220. Maya, R., M. Balass, S. Kim, D. Shkedy, J. Martinez-Leal, O. Shifman, M. Moas, T. Buschmann, Z. Ronai, Y. Shiloh, M. Kastan, E. Katzir and M. Oren: *ATM-dependent phosphorylation of Mdm2 on serine 395: Role in p53 activation by DNA damage*. Genes Dev, 15:1067—-1077, 2001.

221. Meek, D.: *Regulation of the p53 response and its relationship to cancer*. Biochemical Journal, 469:325–346, 2015.

222. Melero, J., D. Stitt, W. Mangel and R. Carroll: *Identification of new polypeptide species (48-55K) immunoprecipitable by antiserum to purified large T antigen and present in SV40-infected and -transformed cells*. Virology, 93:466–480, 1979.

223. Melino, G., X. Lu, M. Gasco, T. Crook and R. Knight: *Functional regulation of p73 and p63: Development and cancer*. Trends Biochem Sci, 28:663–670, 2003.

224. Memon, M., M. Anway, T. Covert, M. Uzumcu and M. Skinner: *Transforming Growth Factor Beta (TGFbeta1, TGFbeta2 and TGFbeta3) Null-Mutant Phenotypes in Embryonic Gonadal Development*. Mol Cell Endocrinol, 294:70–80, 2008.

225. Menendez, D., A. Inga and M. Resnick: *The expanding universe of p53 targets*. Nat Rev Cancer, 9 no. 10:724–737, 2009.

226. Meulmeester, E. and P. Dijke: *The dynamic roles of TGF-β in cancer*. Journal of Pathology, 223 no. 2:205–218, 2011.

227. Michael, D. and M. Oren: *The p53-Mdm2 module and the ubiquitin system*. Semin Cancer Biol, 13:49–58, 2003.

228. Michalovitz, D., O. Halevy and M. Oren: *Conditional inhibition of transformation and of cell proliferation by a temperature-sensitive mutant of p53*. Cell, 62:671–680, 1990.

229. Miller, L., J. Smeds, J. George, V. Vega, L. Vergara, A. Ploner, Y. Pawitan, P. Hall, S. Klaar, E. Liu and J. Bergh: *An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival*. PNAS, 102:13550—-13555, 2005.

230. Miyakoda, M., K. Suzuki, S. Kodama and M. Watanabe: *Activation of ATM and phosphorylation of p53 by heat shock*. Oncogene, 21:1090—1096, 2002.

231. Miyashita, T. and J. Reed: *Tumor suppressor p53 is a direct transcriptional activator of the human bax gene*. Cell, 80:293–299, 1995.

232. Moll, U., A. Ostermeyer, R. Haladay, B. Winkfield, M. Frazier and G. Zambetti: *ytoplasmic sequestration of wild-type p53 protein impairs the G1 checkpoint after DNA damage*. C. Mol Cell Biol, 162, 1996.

233. Moll, U. and N. Slade: *p63 and p73: Roles in development and tumor formation*. Mol Cancer Res, 2:371–386, 2004.

234. Momand, J., G. Zambetti, D. Olson, D. George and A. Levine: *The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation*. Cell, 69:1237–45, 1992.

235. Mönke, G., E. Cristiano, A. Finzel, D. Friedrich, H. Herzel, M. Falcke and A. Loewer: *Excitability in the p53 network mediates robust signaling with tunable activation thresholds in single cells*. Scientific Reports, 7, 2017.

236. Moran, DS.and Mendal, L.: *Core Temperature Measurement*. Sport. Med., 32:879—-885, 2002.

237. Moroishi, T., C. Hansen and K. Guan: *The emerging roles of YAP and TAZ in cancer*. Nature Reviews Cancer volume, 3215:73—-79, 2015.

238. Moustakas, A. and C. Heldin: *Non-Smad TGF-beta signals*. J Cell Sci, 118:3573–3584, 2005.

239. Moustakas, A. and C. Heldin: *The regulation of TGF signal transduction*. Development, 136:3699 — 3714, 2009.

240. Mu, Y., S. Gudey and M. Landstrom: *Non-Smad signaling pathways*. Cell and Tissue Research, 347 no. 1:11–20, 2009.

241. Mueller, M.: *Information Retrieval for Music and Motion*. Springer, Berlin, Heidelberg, 2007.

242. Mukhopadhyay, P.: *An Introduction to Estimating Functions*. Alpha Science International, Harrow, UK, 2004.

243. Mullen, A., D. Orlando, J. Newman, J. Loven, R. Kumar, S. Bilodeau, J. Reddy, M. Guenther, R. DeKoter and R. Young: *Master transcription factors determine cell-type-specific responses to TGF-b signaling*. Cell, 147:565 — 576, 2011.

244. Müller, L., A. Schaupp, D. Walerych, H. Wegele and J. Buchner: *Hsp90 Regulates the Activity of Wild Type p53 under Physiological and Elevated Temperatures*. J. Biol. Chem., 279:48846—48854, 2004.

245. Müller, P., P. Ceskova and B. Vojtesek: *Hsp90 Is Essential for Restoring Cellular Functions of Temperature-sensitive p53 Mutant Protein but Not for Stabilization and Activation of Wild-type p53 IMPLICATIONS FOR CANCER THERAPY*. J. Biol. Chem., 280:6682–6691, 2004.

246. Murakami, G., T. Watabe, K. Takaoka, K. Miyazono and T. Imamura: *Cooperative inhibition of bone morphogenetic protein signaling by Smurf1 and inhibitory Smads*. Mol Biol Cell, 14:2809–2817, 2003.

247. Murtagh, F.: *A Survey of Recent Advances in Hierarchical Clustering Algorithms*. The Computer Journal, 26(4):354–359, 1983.

248. Myronenko, A. and X. Song: *Point set registration: Coherent Point drift*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32:2262–2275, 2010.

249. Myronenko, A., X. Song and M. Carreira-Perpinan: *Non-rigid point set registration: Coherent Point Drift*. NIPS, pp. 1009–1016, 2007.

250. Nakamura, K.: *Central circuitries for body temperature regulation and fever*. Am. J. Physiol. Integr. Comp. Physiol., pp. 1207–1228, 2011.

251. Nallet-Staub, F., X. Yin, C. Gilbert, V. Marsaud, S. BenMimoun, D. Javelaud, E. Leof and A. Mauviel: *Cell density sensing alters TGF-b signaling in a celltype-specific manner, independent from Hippo pathway activation*. Dev Cell, 32:640–651, 2015.

252. Needleman, S. and C. Wunsch: *A general method applicable to the search for similarities in the amino acid sequence of two proteins*. Journal of Molecular Biology, 48 no. 3:443–453, 1970.

253. Nicolas, F. and C. Hill: *Attenuation of the TGF-beta-Smad signaling pathway in pancreatic tumor cells confers resistance to TGF-beta-induced growth arrest*. Oncogene, 22:3698 —- 3711, 2003.

254. Nyquist, H.: *Certain Topics in Telegraph Transmission Theory*. Transactions of the American Institute of Electrical Engineers, 47:617—644, 1926.

255. Oei, A., L. Vriend, J. Crezee, F. NAP. and P. Krawczyk: *Effects of hyperthermia on DNA repair pathways: one treatment to inhibit them all*. Radiation oncology, 10, 2015.

256. Ohnstad, H., R. Castro, J. Sun, K. Heintz, L. Vassilev, B. Bjerkehagen, S. Kresse, L. Meza-Zepeda and O. Myklebost: *Correlation of TP53 and MDM2 genotypes with response to therapy in sarcoma*. Cancer, 119:1013—1022, 2013.

257. Oliva-Trastoy, M., V. Berthonaud, A. Chevalier, C. Ducrot, M.-C. Marsolier-Kergoat, C. Mann and F. Leteurtre: *The Wip1 phosphatase (PPM1D) antagonizes activation of the Chk2 tumour suppressor kinase*. Oncogene, pp. 1449—-1458, 2007.

258. Olsson, A., C. Manzl, A. Strasser and A. Villunger: *How important are post-translational modifications in p53 for selectivity in target-gene transcription and tumour suppression?*. Cell Death Differ, 14:1561–1575, 2007.

259. *Statistics and Machine Learning Toolbox*, 2019. The MathWorks, Natick, MA, USA.

260. Oren, M. and V. Rotter: *Mutant p53 gain-of-function in cancer*. Cold Spring Harbor Perspectives in Biology, 2 no. 2, 2010.

261. Ota, I., N. Okamoto, K. Yane, A. Takahashi, T. Masui, H. Hosoi and T. Ohnishi: *Therapeutic strategies for head and neck cancer based on p53 status*. Exp. Ther. Med., 3:585—-591, 2012.

262. Paek, A., J. Liu, A. Loewer, W. Forrester and G. Lahav: *Cell-to-Cell Variation in p53 Dynamics Leads to Fractional Killing*. Cell, 165:631–642, 2016.

263. Paparrizos, J. and L. Gravano: *k-Shape: Efficient and Accurate Clustering of Time Series*. In Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, pp. 1855–1870, 2015.

264. Park, S. H.: *Fine tuning and cross-talking of TGF-β signal by inhibitory Smads*. J. Biochem. Mol. Biol., 38:9–16, 2005.

265. Pei, D., Y. Zhang and J. Zheng: *Regulation of p53: a collaboration between Mdm2 and MdmX*. Oncotarget, 3:228–235, 2012.

266. Peng, Y., L. Chen, C. Li, W. Lu and J. Chen: *Inhibition of MDM2 by hsp90 contributes to mutant p53 stabilization*. J. Biol. Chem., 276:40583—-40590, 2001.

267. Picksley, S. and D. Lane: *The p53-mdm2 autoregulatory feedback loop: a paradigm for the regulation of growth control by p53?*. Bioessays, 15:689–90, 1993.

268. Piek, E., W. Ju, J. Heyer, D. Escalante-Alcalde, C. Stewart, M. Weinstein, C. Deng, R. Kucherlapati, E. Böttinger and A. Roberts: *Functional characterization of transforming growth factor beta signaling in Smad2-and Smad3-deficient fibroblasts*. J Biol Chem, 276:19945 — 19953, 2001.

269. Piesch, E., S. Sykes, S. McMahon and M. Murphy: *The p53 family and programmed cell death*. Nature, 27:6507–6521, 2008.

270. Pinidiyaarachchi, A. and C. Wahlby: *Seeded Watersheds for Combined Segmentation and Tracking of Cells*. mage Analysis and Processing, pp. 336–343, 2005.

271. Pomerantz, J., N. Schreiber-Agus, N. Liégeois, A. Silverman, L. Alland, L. Chin, J. Potes, K. Chen, I. Orlow and H. Lee: *The Ink4a tumor suppressor gene product, p19Arf, interacts with MDM2 and neutralizes MDM2's inhibition of p53*. Cell, 92:713–723, 1998.

272. Prochazka, A., J. Kukal and O. Vysata: *Wavelet transform use for feature extraction and EEG signal segments classification*. pp. 719 – 722, 04 2008.

273. Ptashne, M. and A. Grann: *Transcriptional activation by recruitment*. Nature, 386:569–577, 1997.

274. Purvis, J., K. Karhohs, C. Mock, E. Batchelor, A. Loewer and G. Lahav: *p53 dynamics control cell fate*. Science, 336:945–956, 2012.

275. Purvis, J. and G. Lahav: *Encoding and decoding cellular information through signaling dynamics*. Cell, 152:945–956, 2013.

276. Quelle, D., F. Zindy, R. Ashmun and C. Sherr: *Alternative reading frames of the INK4a tumor suppressor gene encode two unrelated proteins capable of inducing cell cycle arrest*. Cell, 83:993—1000, 1005.

277. Ramanan, D. and D. Forsyth: *Automatic annotation of everyday movements*. Advances in Neural Information Processing Systems, 16, 2003.

278. Rao, W., Z. Deng and J. Liu: *A Review of Hyperthermia Combined With Radiotherapy/Chemotherapy on Malignant Tumors*. Crit. Rev. Biomed. Eng., 38:101—-116, 2010.

279. Remy, I., A. Montmarquette and S. Michnick: *PKB/Akt modulates TGF-beta signalling through a direct interaction with Smad3*. Nat Cell Biol, 6:358–365, 2004.

280. Riley, T., E. Sontag, P. Chen and A. Levine: *Transcriptional control of human p53-regulated genes*. Nat. Rev. Mol. Cell Biol., 9:402—412, 2008.

281. Rincon, M. and R. Davis: *Regulation of the immune response by stress-activated protein kinases*. Immunol Rev, 228:212–224, 2009.

282. Roberts, A. and M. Sporn: *Transforming growth factor beta*. Adv Cancer Res, 51:107–145, 1988.

283. Roberts, A. and L. Wakefield: *The two faces of transforming growth factor β in carcinogenesis*. Proceedings of the National Academy of Sciences of the United States of America, 100 no. 15:8621–8623, 2003.

284. Ross, S. Hill, C.: *How the Smads regulate transcription*. Intern J Biochem Cell Biol, 40:383–408, 2008.

285. Ross, M., H. Shaffer, A. Cohen, R. Freudberg and H. Manley: *Average magnitude difference function pitch extractor*. IEEE Trans. Acoust., 22:353–362, 1974.

286. Rotter, V.: *p53, a transformation-related cellular-encoded protein, can be used as a biochemical marker for the detection of primary mouse tumor cells*. Proceedings of the National Academy of Sciences of the United States of America, 80:2613–2617, 1983.

287. Rousseeuw, P. J.: *Silhouettes: A graphical aid to the interpretation and validation of cluster analysis*. Journal of Computational and Applied Mathematics, 20:53 – 65, 1987.

288. Ryu, H., M. Chung, M. Dobrzynski, D. Fey, S. Blum, Y. Lee, M. Peter, B. Kholodenko, N. Jeon and O. Pertz: *Frequency modulation of ERK activation dynamics rewires cell fate*. Mol Syst Biol, 11:838, 2015.

289. Saito, A., H. Masafumi and N. Takahide: *TGFb Signaling in Lung Health and Disease*. Int J Mol Sci., 19(8), 2018.

290. Sakoe, H. and S. Chiba: *Dynamic programming algorithm optimization for spoken word recognition*. IEEE Trans. Acoust., 26:43–49, 1978.

291. Samanta, D. and P. Datta: *Alterations in the Smad pathway in human cancers*. Front Biosci, 17:1281–1293, 2012.

292. Samuels-Lev, Y.: *ASPP proteins specifically stimulate the apoptotic function of p53*. Mol. Cell, 8:781–794, 2001.

293. Sandler, O., S. Mizrahi, N. Weiss, O. Agam, I. Simon and N. Balaban: *Lineage correlations of single cell division time as a probe of cell-cycle dynamics.*. Nature, 519:468–471, 2015.

294. Sarda-Espinosa, A.: *Comparing Time-Series Clustering Algorithms in R Using the dtwclust Package*. Techn. Rep., 2019.

295. Sauer, R., M. Crezee, H. Hulshof, O. Issels, R. Ott and G. van Rhoon: *The letter of the Ludwig Boltzmann Institute is unnecessarily polarizing to the discussion on whether or not hyperthermia is evidence based*. Strahlenther Onko, 189:84–86, 2013.

296. Schmierer, B. and C. Hill: *TGFbeta-SMAD signal transduction: molecular specificity and functional flexibility*. Nat Rev Mol Cell Biol, 8:970–982, 2007.

297. Schon, O., A. Friedler, M. Bycroft, S. Freund and A. Fersht: *Molecular mechanism of the interaction between MDM2 and p53*. J Mol Biol, 323:491–501, 2002.

298. Schwartzenberg-Bar-Yoseph, F., M. Armoni and E. Karnieli: *The tumor suppressor p53 down-regulates glucose transporters GLUT1 and GLUT4 gene expression*. Cancer Res, 64:2627—-2633, 2004.

299. Scian, M., E. Carchman, L. Mohanraj, K. Stagliano, M. Anderson, D. Deb and et al.: *Wild-type p53 and p73 negatively regulate expression of proliferation related genes*. Oncogene, 27:2583–2593, 2008.

300. Seno, J. and J. Dynlacht: *Intracellular Redistribution and Modification of Proteins of the Mre11/Rad50/Nbs1 DNA Repair Complex Following Irradiation and Heat-Shock*. Journal of cellular physiology, 199:157–170, 2004.

301. Serra, J. and J. Arcos: *An Empirical Evaluation of Similarity Measures for Time Series Classification*. Knowledge-Based Systems, 67:305–312, 2014.

302. Sezgin, M. and B. Sankur: *Survey over image thresholding techniques and quantitative performance evaluation*. JJournal of Electronic Imaging, 13:146—-165, 2004.

303. Sheng, C.: *Cellular heterogeneity in the DNA damage response is determined by cell cycle specific p21 degradation*. PhD thesis, Humbolt Univerity Berlin, 2017.

304. Sheng, C., I. Mendler, S. Rieke, S. Snyder, M. Jentsch, D. Friedrich, B. Drossel and A. Loewer: *PCNA-Mediated Degradation of p21 Coordinates the DNA Damage Response and Cell Cycle Regulation in Individual Cells*. Cell Reports, 27:48–58, 2019.

305. Shi, Y. and J. Massague: *Mechanisms of TGF-β signalling from cell membarn to the nucleus*. Cell, 113 no. 6:685–700, 2003.

306. Shieh, S., J. Ahn, K. Tamai, J. Taya and C. Prives: *The human homologs of checkpoint kinases Chk1 and Cds1 (Chk2) phosphorylate p53 at multiple DNA damage-inducible sites.* Genes and Development, 14:289–300, 2000.

307. Shieh, S., M. Ikeda, Y. Taya and C. Prives: *DNA damage-induced phosphorylation of p53 alleviates inhibition by MDM2.* Cell, 91:325–334, 1997.

308. Shiloh, Y.: *ATM and ATR: networking cellular responses to DNA damage.* Current option in genetics and development, 11:71–77, 2001.

309. Shin, I., A. Bakin, U. Rodeck, A. Brunet and C. Arteaga: *Transforming growth factor beta enhances epithelial cell survival via Akt-dependent regulation of FKHRL1.* J Biol Chem, 12:3328–3339, 2001.

310. Shreeram, S.: *Wip1 phosphatase modulates ATM-dependent signaling pathways.* Mol Cell, 23:757–764, 2006.

311. Siegel, P. and J. Massague: *Cytostatic and apoptotic actions of TGF-β in homeostasis and cancer.* Nature Reviews Cancer, 3 no. 11:807–821, 2003.

312. Smith, A., R. Smith and E. Paucha: *Characterization of different tumor antigens present in cells transformed by simian virus 40.* Cell, 18:335–346, 1979.

313. Smith, T. F. and M. S. Waterman: *Identification of common molecular subsequences.* In *Journal of Molecular Biology*, vol. 147(1), pp. 195–197, 1981.

314. Snijder, B. and L. Pelkmans: *Origins of regulated cell-to-cell variability.* Nat. Rev. Mol. Cell Biol., 12:119–125, 2011.

315. Soares, I., I. Ferreira, R. Igreja, C. Novo and J. Borges: *Application of Hyperthermia for Cancer Treatment: Recent Patents Review.* Recent Pat. Anticancer. Drug Discov., 7:64—-73, 2012.

316. Solozobova, V. and C. Blattner: *p53 in stem cells.* World J. Biol. Chem., 2:202—214, 2011.

317. Song, K., H. Wang, T. Krebs and D. Danielpour: *Novel roles of Akt and mTOR in suppressing TGF-beta/ALK5-mediated Smad3 activation.* EMBO J, 25:58–69, 2006.

318. Soussi, T.: *TP53 mutations in human cancer: database reassessment and prospects for the next decade.* Advances in Cancer Research, 10:107–139, 2011.

319. Spencer, S., S. Gaudet, J. Albeck, J. Burke and P. Sorger: *Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis.* Nature, 459:428–432, 2009.

320. Spiller, D., C. Wood, D. Rand and M. White: *Measurement of single-cell dynamics.* Nature, 465:736–745, 2010.

321. Sporn, M. and A. Roberts: *Peptide growth factors: current status and therapeutic opportunities.* Important Adv Oncol, pp. 75–86, 1987.

322. *Statistics and Machine Learning Toolbox*, 2019. The MathWorks, Natick, MA, USA.

323. Stewart-Ornstein, J., and G. Lahav: *Dynamics of CDKN1A in single cells defined by an endogenous fluorescent tagging toolkit*. Cell Rep., 14:1800–1811, 2016.

324. Stewart-Ornstein, J., and G. Lahav: *p53 dynamics in response to DNA damage vary across cell lines and are shaped by efficiency of DNA repair and activity of the kinase ATM*. Sci. Signal., 10:10, 2017.

325. Strasen, J.: *Dynamics and variability of SMAD signaling in single cells - The activity of MAP kinases determines long-term dynamics of SMAD signaling*. PhD thesis, Humbolt Univerity Berlin, 2019.

326. Strasen, J., U. Sarma, M. Jentsch, S. Bohn, C. Sheng, D. Horbelt, P. Knaus, S. Legewie and A. Loewer: *Cell-specific responses to the cytokine TGFβ are determined by variability in protein levels*. Molecular systems biology, 14:1744–4292, 2018.

327. Sugar, C. A. and G. M. James: *Finding the Number of Clusters in a Dataset: An Information-Theoretic Approach*. Journal of the American Statistical Association, 98(463):750–763, 2003.

328. Sund-Levander, M., C. Forsberg and L. Wahren: *Finding the Number of Clusters in a Dataset: An Information-Theoretic Approach*. Scand. J. Caring Sci., 16:122—128, 2002.

329. Szak, S., D. Mays and J. Pietenpol: *Kinetics of p53 binding to promoter sites in vivo*. Mol Cell, 21:3375–3386, 2001.

330. Tabach, Y., M. Milyavsky, I. Shats, R. Brosh, O. Zuk, A. Yitzhaky and et al.: *The promoters of human cell cycle genes integrate signals from two tumor suppressive pathways during cellular transformation*. Mol Syst Biol, 1:2005–2022, 2005.

331. Tang, W., G. Ling, L. Sun and F. Liu: *Smad anchor for receptor activation (SARA) in TGF-beta signaling*. Front Biosci (Elite Ed), 2:857–860, 2010.

332. Tay, S., J. Hughey, T. Lee, T. Lipniacki, S. Quake and M. Covert: *Single-cell NF-jB dynamics reveal digital activation and analogue information processing*. Nature, 466:267–271, 2005.

333. Tebaldi, T., S. Zaccara, F. Alessandri, A. Bisio, Y. Ciribilli and A. Inga: *Whole-genome cartography of p53 response elements ranked on transactivation potential*. BMC Genomics, 16, 2014.

334. Teufel, D., M. Bycroft and F. A.R.: *Regulation by phosphorylation of the relative affinities of the N-terminal transactivation domains of p53 for p300 domains and Mdm2*. Oncogene, 28:2112–2118, 2009.

335. Thorburn, J., Z. Andrysik, L. Staskiewicz, J. Gump, P. Maycotte, A. Oberst, D. Green, J. Espinosa and A. Thorburn: *Autophagy controls the kinetics and extent of mitochondrial apoptosis by regulating PUMA levels*. Cell Rep, 7:45–52, 2014.

336. Toledo, F. and G. Wahl: *Regulating the p53 pathway: in vitro hypotheses, in vivo veritas*. Nat Rev Cancer, 6:909–923, 2006.

337. Toledo, S. de, E. Azzam, P. Keng, S. Laffrenier and J. Little: *Regulation by ionizing radiation of CDC2, cyclin A, cyclin B, thymidine kinase, topoisomerase IIalpha, and RAD51 expression in normal human diploid fibroblasts is dependent on p53/p21Waf1*. Cell Growth Differ, 9:887–896, 1998.

338. Turner, T. and T. Caspari: *When heat casts a spell on the DNA damage checkpoints*. Open Biol, 4:140008—140008, 2014.

339. Tuul, M., H. Kitao, M. Iimori, K. Matsuoka, S. Kiyonari and H. Saeki: *Rad9, Rad17, TopBP1 and claspin play essential roles in heat-induced activation of ATR kinase and heat tolerance*. PLoS One, 8, 2013.

340. Tyner, S., S. Venkatachalam, J. Choi, S. Jones, N. Ghebranious, H. Igelmann, X. Lu, G. Soron, B. Cooper and C. Brayton: *p53 mutant mice that display early ageing-associated phenotypes*. Nature, 415:45—53, 2002.

341. Uziel, T., Y. Lerenthal, L. Moyal, Y. Andegeko, L. Mittelman and Y. Shiloh: *Requirement of the MRN complex for ATM activation by DNA damage*. The EMBO Journal, 22:5612–5621, 2003.

342. Verfaillie, A., D. Svetlichnyy, H. Imrichova, K. Davie, M. Fiers, Z. Atak and et al.: *Multiplex enhancer-reporter assays uncover unsophisticated TP53 enhancer logic*. Genome Res, 26:882–895, 2016.

343. Vermeulen, K., D. Van Bockstaele and Z. Berneman: *The cell cycle:a review of regulation, deregulation and therapeutic targets in cancer*. Cell Proliferation, 36:131—149, 2003.

344. Villasenor, R., H. Nonaka, P. DelConteZerial, Y. Kalaidzidis, M. Zerial and S. Pfeffer: *Regulation of EGFR signal transduction by analogue-to-digital conversion in endosomes*. eLife, 4, 2015.

345. Vinals, F. and J. Pouyssegur: *Transforming growth factor beta1 (TGF-beta1) promotes endothelial cell survival during in vitro angiogenesis via an autocrine mechanism implicating TGF-alpha signaling*. Mol Cell Biol, 21:7218–7230, 2001.

346. Vitale, I., L. Galluzzi, M. Castedo and G. Kroemer: *Mitotic catastrophe: a mechanism for avoiding genomic instability.*. Nat. Rev. Mol. Cell Biol., 12:385—-392, 2011.

347. Vizan, P., D. Miller, I. Gori, D. Das, B. Schmierer and C. Hill: *Controlling longterm signaling: receptor dynamics determine attenuation and refractory behavior of the TGF- pathway*. Sci Signal, 6, 2013.

348. Vousden, K. and C. Prives: *Blinded by the light: the growing complexity of p53*. Cell, 137 no. 3:413–431, 2009.

349. Vousden, k. and K. Ryan: *p53 and metabolism*. Nature Rev. Cancer, 9, 2009.

350. Wagner, E. and A. Nebreda: *Signal integration by JNK and p38 MAPK pathways in cancer development*. Nat Rev Cancer, 9:537–549, 2009.

351. Walerych, D., M. Olszewski, M. Gutkowska, A. Helwak, M. Zylicz and A. Zylicz: *Hsp70 molecular chaperones are required to support p53 tumor suppressor activity under stress conditions*. Oncogene, 28:4284–4294, 2009.

352. Wang, C. and C. Jiandong: *Phosphorylation and hsp90 Binding Mediate Heat Shock Stabilization of p53*. Journal of Biological Chemistry, 278:2066–2071, 2002.

353. Wang, X., A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann and E. Keogh: *Experimental comparison of representation methods and distance measures for time series data*. Data Mining and Knowledge Discovery, 26:275—-309, 2013.

354. Ward, J.: *Hierarchical Grouping to Optimize an Objective Function*. J. Am. Stat., 58:236—-244, 1963.

355. Warmflash, A., Q. Zhang, B. Sorre, A. Vonica, E. Siggia and A. Brivanlou: *Dynamics of TGFb signaling reveal adaptive and pulsatile behaviors reflected in the nuclear localization of transcription factor Smad4*. Proc Natl Acad Sci, 109, 2012.

356. Webster, G. A. and N. D. Perkins: *Transcriptional cross talk between NF-κB and p53*. Mol. Cell. Biol, 19:3485–3495, 1999.

357. Wegner, K., A. Bachmann, J. Schad, P. Lucarelli, S. Sahle, P. Nickel, C. Meyer, U. Klingmüller, S. Dooley and U. Kummer: *Dynamics and feedback loops in the transforming growth factor b signaling pathway*. Biophys Chem, 162:22–34, 2012.

358. Weiss, A. and L. Attisano: *The TGFbeta Superfamily Signaling Pathway*. WIREs Dev Biol, 2:47—-63, 2013.

359. Wellbrock, C., M. Karasarides and R. Marais: *The RAF proteins take centre stage*. Nat Rev Mol Cell Biol, 5 no. 11:875–85, 2004.

360. Westra, A. and W. Dewey: *Variation in sensitivity to heat shock during the cell-cycle of Chinese hamster cells in vitro*. Int. J. Radiat. Biol. Relat. Stud. Phys. Chem Med., 19:467—-477, 1971.

361. Widlak, P., M. Gramatyka and M. Kimmel: *Crosstalk between stress-induced NF-kB, p53 and HSF1 signaling pathways – review*. Proceedings of the 19th World Congress The International Federation of Automatic Contro, 2014.

362. Wilkes, M., H. Mitchell and S. Penheiter: *Transforming growth factor-beta activation of phosphatidylinositol 3-kinase is independent of Smad2 and Smad3 and regulates fibroblast responses via p21-activated kinase-2*. Cancer Res, 65:10431–10440, 2005.

363. Wilkinson, D. S., W. W. Tsai, M. A. Schumacher and M. C. Barton: *Chromatin-bound p53 anchors activated Smads and the mSin3A corepressor to confer transforming growth factor β-mediated transcription repression*. Mol. Cell. Biol., 28:1988–1998, 2008.

364. Williams, A. and B. Schumacher: *p53 in the DNA-damage-repair process*. Cold Spring Harb. Perspect. Med., 6, 2016.

365. Wolf, D. and V. Rotter: *Major deletions in the gene encoding the p53 tumor antigen cause lack of p53 expression in HL-60 cells*. Proc. Natl Acad. Sci. USA, 82:790—-794, 1985.

366. Woods, D. and J. Turchi: *Chemotherapy induced DNA damage response*. Cancer Biol Ther., 14:379–389, 2013.

367. Wrighton, K., X. Lin and X. Feng: *Phospho-control of TGF-beta superfamily signaling*. Cell Research, 19:8–20, 2009.

368. Wu, X., J. Bayle, D. Olson and A. Levine: *The p53-mdm-2 autoregulatory feedback loop*. Genes Dev, 7:1126–1132, 1993.

369. Wust, P., B. Hildebrandt, G. Sreenivasa, B. Rau, J. Gellermann and H. Riess: *Hyperthermia in combined treatment of cancer*. Lancet Oncol, 3:487—-497, 2002.

370. Xu, X.: *Genetic interactions between tumor suppressors Brca1 and p53 in apoptosis, cell cycle and tumorigenesis*. Proc Natl Acad Sci USA, 28:266–271, 2001.

371. Xue, W.: *Senescence and tumour clearance is triggered by p53 restoration in murine liver carcinomas*. Nature, 445:656–660, 2007.

372. Yan, X. and Y. Chen: *Smad7: not only a regulator, but also a cross-talk mediatorof TGF-β signalling*. Biochem J, 434:1–10, 2011.

373. Yang, A., M. Kaghad, Y. Wang, E. Gillett, M. Fleming, V. Dotsch, N. Andrews, D. Caput and F. McKeon: *p63, a p53 homolog at 3q27-29, encodes multiple products with transactivating, death-inducing, and dominant-negative activities*. Cell Research, 2:305–316, 1998.

374. Yi, J., I. Shin and C. Arteaga: *Type I transforming growth factor beta receptor binds to and activates phosphatidylinositol 3-kinase*. J Biol Chem, 288:10870–10876, 2005.

375. Yissachar, N., T. Sharar Fischler, A. Cohen, S. Reich-Zeliger, D. Russ, E. Shifrut, Z. Porat and N. Friedman: *Dynamic response diversity of NFAT isoforms in individual living cells*. Mol. Cell, 49:1322—1330, 2013.

376. Yonish-Rouach, E.: *Wild-type p53 induces apoptosis of myeloid leukaemic cells that is inhibited by interleukin-6*. Nature, 352:345–347, 1991.

377. Zahng, X., L. Ma, H. Qi, J. Shan, W. Yu and Y. Gu: *MAPK/ERK signaling pathway-induced hyper-O-GlcNAcylation enhances cancer malignancy*. Cell Research, 410 no. 1-2:101–110, 2015.

378. Zahng, Y.: *Non-Smad pathways in TGF-β signaling*. Cell Research, 19 no. 1:128–139, 2009.

379. Zaika, A., N. Marchenko, and U. Moll: *Cytoplasmically 'sequestered' wild type p53 protein is resistant to Mdm2-mediated degradation.*. J Biol Chem, 274:27474—-27480, 1999.

380. Zaky, A.: *Regulation of the human AP-endonuclease (APE1/Ref-1) expression by the tumor suppressor p53 in response to DNA damage*. Nucleic Acids Res, 36:1555–1566, 2008.

381. Zhang, J., X.-J. Tian, H. Zhang, Y. Teng, R. Li, F. Bai and J. Elankumaran, S. Xing: *GF-b–induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops*. Sci Signal, 7, 2014.

382. Zhang, L., W. Wang, Y. Hayashi and et al.: *A role for MEK kinase 1 in TGF-beta/activin-induced epithelium movement and embryonic eyelid closure*. EMBO J, 22:4443–4454, 2003.

383. Zhang, Y., Y. Xiong and W. Yarbrough: *ARF promotes MDM2 degradation and stabilizes p53: ARF-INK4a locus deletion impairs both the Rb and p53 tumor suppression pathways*. Cell, 92:725–734, 1998.

384. Zhou, B. and E. Sausville: *Drug discovery targeting Chk1 and Chk2 kinases*. Prog Cell Cycle Res, 5:413–421, 2003.

385. Zhu, H., B. Chang, T. Uchiumi and I. Roninson: *Identification of promoter elements responsible for transcriptional inhibition of polo-like kinase 1 and topoisomerase IIalpha genes by p21(WAF1/CIP1/SDI1)*. Cell Cycle, 1:59–66, 2002.

386. Zhu, H., J. Iaria and A. Sizeland: *Smad7 differentially regulates transforming growth factor beta-mediated signaling pathways*. J Biol Chem, 274:32258–32264, 1999.

387. Zhu, J., W. Zhou, J. Jiang and X. Chen: *Identification of a novel p53 functional domain that is necessary for mediating apoptosis*. J Biol Chem, 273:13030–13036, 1998.

388. Zi, Z., Z. Feng, D. Chapnick, M. Dahl, D. Deng, E. Klipp, A. Moustakas and X. Liu: *Quantitative analysis of transient and sustained transforming growth factor-b signaling dynamics*. Mol Syst Biol, 7:492, 2011.

389. Zieba, A., K. Pardali, O. Söderberg, L. Lindbom, E. Nystrom, A. Moustakas, C. Heldin and U. Landegren: *Intercellular variation in signaling through the TGF-b pathway and its relation to cell density and cell cycle phase*. Mol Cell Proteomics, 11, 2012.

390. Zou, L. and S. Elledge: *Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes*. Science., 300:1542—-1548, 2003.

391. Zylicz, M., F. King and A. Wawrzynow: *Hsp70 interactions with the p53 tumor suppressor protein*. EMBO J, 17:4634–4638, 2001.