Doctoral Thesis in Biotechnology

Finding order in chaos

Dissecting single-cell heterogeneity in space and time

CHRISTIAN GNANN
Finding order in chaos
Dissecting single-cell heterogeneity in space and time
CHRISTIAN GNANN

Academic Dissertation which, with due permission of the KTH Royal Institute of Technology, is submitted for public defence for the Degree of Doctor of Philosophy on Friday the 3rd May 2024, at 10:00 a.m. in Air&Fire, Tomtebodavägen 23, Solna.

Doctoral Thesis in Biotechnology
KTH Royal Institute of Technology
Stockholm, Sweden 2024
Thesis Defense

The public defense of this thesis will take place on May 3\textsuperscript{rd} 2024 at 10:00 in the room Air&Fire at Science for Life Laboratory, Tomtebodavägen 23A, Solna, Sweden. The attempted degree is Doctor of Philosophy (PhD) in Biotechnology.

**Respondent:** M.Sc. Christian Gnann, Department of Protein Science, KTH Royal Institute of Technology, Science for Life Laboratory, Sweden

**Faculty Opponent:** Prof. Lucas Pelkmans, Department of Molecular Life Sciences, University of Zurich, Switzerland

**Host:** Dr. Burcu Ayoglu, Department of Protein Science, KTH Royal Institute of Technology, Science for Life Laboratory, Sweden

**Evaluation committee**

**Prof. Anna Wredenberg,** Department of Medical Biochemistry and Biophysics, Karolinska Institute, Sweden

**Docent Arne Lindqvist,** Department of Cell and Molecular Biology, Karolinska Institute, Sweden

**Prof. Ylva Engström,** Department of Molecular Biosciences, Stockholm University, Sweden

**Respondent's supervisor:** Prof. Emma Lundberg, Department of Protein Science, KTH Royal Institute of Technology, Science for Life Laboratory, Sweden

**Respondent's co-supervisor:** Dr. Manuel Leonetti, Chan Zuckerberg Biohub, San Francisco, USA
Abstract

The cell is the smallest unit of life and contains DNA, RNA, proteins and a variety of other macromolecules. In recent years, technological advances in the field of single cell biology have revealed a staggering amount of phenotypic heterogeneity between cells in a population, which were previously considered homogenous. Previous work has largely been focused on studies of RNA. As proteins however are the ultimate effectors of genetic information, this thesis aims to provide a protein-centered view on cellular heterogeneity, particularly focusing on cell cycle and cellular metabolism.

Most of my work has been performed within the framework of the Human Protein Atlas project. In the context of this project, we mapped the spatial distribution of more than 13,000 human proteins with subcellular resolution and found that around a quarter of all human proteins exhibit protein expression heterogeneity.

In Paper I, we hypothesized that a majority of the observed cellular heterogeneity can be explained by differences in cell cycle progression. Therefore, we generated a map of proteomic and transcriptomic heterogeneity at subcellular resolution, which we precisely aligned to the cell cycle position of individual cells. This approach allowed us to identify hundreds of previously unknown cell cycle-related proteins. With sustained proliferative signaling representing a hallmark of cancer, novel cell-cycle proteins could serve as potential new drug targets against cancer. We further show that a large part of cell cycle dependent proteome variability is not established by transcriptomic cycling. This suggests that post-translational modifications are a major contributor to the regulation of cell cycle dependent protein level changes. Therefore, in Paper II, we carried out a deep phosphoproteome mass spectrometry profiling of the same cellular model as in Paper I and identified almost 5,000 cell cycle dependent phosphosites on over 2,000 proteins. The unprecedented scale of our phosphoproteomic data allows us to link cell cycle dependent protein expression dynamics to phosphorylation events. Furthermore, we identify a large set of proteins with stable expression levels and fluctuating phosphorylation patterns along cell cycle progression that likely alters protein function.

Despite identifying hundreds of novel cell cycle dependent proteins in paper I, we observed that the majority of heterogeneously expressed proteins display variable expression independent of cell cycle progression, among them a large number of metabolic enzymes. Thus, we sought to describe the extent of subcellular metabolic complexity in human cells and tissues in Paper III. While we confirm metabolic compartmentalization in our dataset, we show that around 50% of metabolic enzymes localize to multiple cellular compartments. By integrating public protein-protein interaction data with our subcellular location information, we identify several enzymes with novel compartment-specific functions. Additionally, we observe a strongly elevated number of heterogeneously expressed enzymes compared to the background of the human proteome that is largely independent of cell cycle progression. We show that this heterogeneity can be manifested in the lineage of a single cell and is conserved in situ. To conclude, we suggest that
the extensive metabolic heterogeneity can establish functional metabolic states in a population of human cells. Finally, in Paper IV, we assessed the heterogeneity of the mitochondrial proteome as they are metabolic powerhouses containing an elevated number of cell cycle independent variably expressed proteins. In this study, we correlated the variable expression of over 400 mitochondrial proteins to the expression of rate limiting enzymes in important mitochondrial pathways; such as the TCA cycle and ROS metabolism. We show that enzymes in the same pathways often correlate in their expression, indicating that their expression variability may contribute to the establishment of metabolic states. Altogether, the thesis illuminates the spatiotemporal complexity of the human proteome established by protein multilocalization and expression heterogeneity as fundamental non-genetic means of functional cell regulation.

**Keywords:** non-genetic single-cell heterogeneity, cell cycle, metabolism, imaging-based subcellular proteomics, phosphoproteomics
Sammanfattning

Cellen är den minsta enheten av liv, och varje cell innehåller en komplex uppsättning av DNA, RNA, proteiner och andra makromolekyler. Under de senaste åren har teknologiska framsteg inom cellbiologiska studier av enskilda celler avslöjat en överväldigande mängd fenotypisk heterogenitet mellan cellpopulationer som tidigare betraktades som homogena. Tidigare kartläggning av denna heterogenitet har främst fokuserat på studier av RNA. Denna avhandling syftar dock till att ge en proteincentrerad syn på cellulär variabilitet, med särskilt fokus på cellcykeln och cellulär metabolism.

Det mesta av mitt arbete innefattar data från Human Protein Atlas-projektet. I detta projekt har vi kartlagt den spatiala fördelningen av över 13 000 mänskliga proteiner med subcellulär upplösning. Vi finner att ungefär en fjärdedel av alla mänskliga proteiner uppvisar heterogenitet i proteinuttryck mellan enskilda celler.

I Artikel I undersökte vi hypotesen att majoriteten av den observerade cellulära heterogeniteten kan förklaras av skillnader i cellcykelprogression. Därför genererade vi en stor karta över proteomisk och transkriptomisk uttrycksdata i relation till den mänskliga cellcykeln i enskilda celler. Vi identifierade hundratals nya proteiner relaterade till cellcykeln, vilka kan komma att fungera som potentiella läkemedelsmål vid sjukdomar som cancer. Vi visar vidare att en stor del av den cellcykelberoende variabiliteten på proteinnivå inte etableras genom cykliska ändringar av transkriptomet. Detta antyder att posttranslationella modifieringar i betydande utsträckning bidrar till regleringen av dynamiska förändringar i proteinuttryck under cellcykeln.

Därför utförde vi i Artikel II en djup masspektrometriprofiling av fosfoproteomet längs samma cellulära modell som i Artikel I och identifierade över 2000 cellcykelberoende fosforyleringsplatser hos människans proteiner. Den oöverträffade omfattningen av denna fosfoproteomiska data gör att vi kan koppla cellcykelberoende dynamik i proteinuttryck till fosforyleringshändelser. Dessutom identifierar vi en stor uppsättning proteiner med stabila uttrycksnivåer och varierande fosforyleringsmönster längs cellcykelns progression, vilket sannolikt reglerar deras funktion.

I Artikel I observerade vi också att majoriteten av heterogen uttryckta proteiner visar variation som är oberoende av cellcykelprogression, och bland dessa proteiner finns ett stort antal metaboliska enzymer. Därför strävade vi efter att beskriva omfattningen av subcellulär metabolisk komplexitet hos mänskliga celler och vävnader i Artikel III. Samtidigt som vi bekräftar att det finns en spatiell uppdelning av cellulära metaboliska processer visar vi att ungefär 50% av de metaboliska enzymerna lokaliserar till flera subcellulära avdelningar. Genom att integrera offentliga protein-protein-interaktionsdata med vår information om subcellulär lokalisering identifierar vi flera enzymer som utför olika funktioner i olika cellulära avdelningar. Dessutom observerar vi en starkt ökad mängd heterogen uttryckta enzymer. [UNA3] Vi föreslår att denna omfattande metaboliska heterogenitet kan etablera olika funktionella metabola tillstånd i en population av mänskliga celler.
Sammanfattning

Slutligen utvärderade vi i Artikel IV heterogeniteten hos det mitokondriella proteomet eftersom de är metabola kraftverk som innehåller en stor andel proteiner vars uttryck varierar beroende av cellcykeln. I denna studie korrelerade vi det variabla uttrycket av över 400 mitokondriella proteiner med uttrycket av hastighetsbegränsande enzymer i viktiga mitokondriella processer, såsom citrondsyracykeln och metabolism av reaktiva syreföreningar. Vi visar att enzymer i samma reaktionsvägar ofta korrelerar i sitt uttryck, vilket indikerar att deras uttrycksvariabilitet kan bidra till etableringen av metabola tillstånd.

Sammanfattningsvis kunde vi genom att korrelera proteinuttryck med cellcykeln i Artikel I förklara en betydande mängd heterogenitet mellan enskilda celler i vår data. Artikel I-II ger insikter om dessa proteiner reglering medan Artikel III + IV vidareutforskar en del av den cellcykelberoende variabilitet som observerats i Artikel I. Som helhet belyser avhandlingen den rumsliga och temporala komplexiteten hos det mänskliga proteomet, etablerad genom multilokalisering och heterogenitet i uttryck hos proteiner som grundläggande icke-genetiska medel för funktionell cellreglering.
“One in 30 trillion!” Those are not lyrics from a popular love song you heard on the radio and cannot get out of your head, but an accurate description of a cell in your body. Cells are the building blocks of life and similar to cities, each cell is further organized into spatially separated compartments called organelles. Organelles serve as the infrastructure in the city; microtubules (MT) provide the roads, the Golgi apparatus is the cellular post office and the mitochondria provide energy as the cellular power plant. But who is running the cell? Proteins, akin to the diverse roles of humans in a city, serve as the workforce within cells. A human cell can contain up to 20,000 different types of proteins that all perform different tasks in the cellular cities. By adjusting the amount and the location of each of those proteins, a cell can regulate its function.

Populations of cells of the same same kind were previously considered identical. However, should you consider two identical twins as the same person or should they be considered as unique individuals, shaped by their environment and experiences? Since we consider identical twins as unique, we should also apply this idea to homogenous cell populations - regarding each cell as 1 in 30 trillions. Importantly, recent research has revealed that there are striking differences between cells, for example the types of proteins present in single cells as well as their amount. However, the full extent of those differences remains elusive. Additionally, knowledge of the biological consequence of this variability is sparse, although it has been linked to clinically relevant phenomena such as drug resistance in cancer treatment for example.

In this thesis, I studied the differences between supposedly identical cells by looking at their protein composition. Our initial hypothesis was that most of the cellular differences are caused by the cell’s progression through the cell cycle. The cell cycle is a series of carefully orchestrated events leading to growth of the cell, the duplication of its genetic information and finally the division of one cell into two daughter cells. Dysregulation of the events of the cell cycle is one of the major drivers for uncontrolled cell growth which in aggressive forms can lead to one of the most dangerous diseases of the western world: cancer. We identified hundreds of novel cell cycle related proteins, many of which could serve as novel drug targets to treat cancer. To better understand how the levels of these proteins were regulated we aimed to characterize the origins of the differences, by looking into the production of the proteins. Each cell contains DNA, which acts as the genetic blueprint of life and contains all the information required to assemble each of the 20,000 proteins. Copies of the DNA, called mRNA, can be obtained and translated into proteins, the molecular machines of the cell. By simultaneously measuring the number of mRNA copies and protein copies, we discovered that only 15% of the cell cycle related proteins were regulated through the levels of mRNA.
Popular scientific summary

How are the cell cycle dependent levels of the remaining 85% of proteins regulated? That was a big question that led us to set up a follow-up study. Proteins can be modified by the addition of molecules, similar to a christmas trees decorated with diverse ornaments. However, these modifications are not just ornaments, they possess the capability to significantly alter the turnover and function of the proteins they decorate. A prominent and well studied modification are protein phosphorylations. The addition of a phospho-group is a reversible modification of a protein and often acts as a molecular switch. In a follow-up study, we characterized the landscape of all protein phosphorylations across the cell cycle using a method called mass spectrometry. Thereby, we identified thousands of cell cycle dependent phosphorylations capable of altering a protein’s function, turnover or location independent of the amount of mRNA copies present in a cell.

Another major finding in the initial study was that the cell cycle could not explain the differences in protein levels for three quarters of the tested proteins. Therefore, we explored cellular metabolism as an alternative driver for variable protein composition between cells. Metabolism is a collective term for all cellular processes that produce new building blocks, break down molecules and produce energy. As such, it is an essential cellular process and therefore differences in metabolism are likely to be responsible for functional differences between individual cells. We identified that a large number of metabolic proteins show variable expression levels between individual cells. We show that differences in metabolic protein compositions are decoupled from the cell cycle. In conclusion, we suggest that metabolic variations can establish functional differences in a seemingly identical population of cells.

To test this hypothesis we wanted to better understand the role of metabolic differences in the cellular city. The mitochondria is the powerplant of the cellular city and a central organelle for metabolism. By mapping the protein content of the mitochondria in single cells, we hoped to gain novel insights into the role of metabolic differences. We find over 1000 proteins in the mitochondria, almost half were previously not known to localize to the mitochondria. A third of all mitochondrial proteins exhibit differences in their levels between individual cells. By correlating their abundances with important metabolic proteins in the mitochondria, we were able to show that the variability of metabolic protein compositions establishes metabolic states.

Overall, this thesis maps out the differences between individual cells with a birds eye view through a microscope. We characterize the cell cycle and metabolism as drivers of cellular variability. We provide new insights about the regulation of said heterogeneity and propose an extent of cellular heterogeneity far exceeding previous expectations. We show that each cell is one in 30 trillion. This thesis paves the way towards a deeper understanding of cellular dynamics, offering potential implications for targeted drug treatments.


Proteine herzustellen. Dafür werden Kopien der DNA erstellt, mRNA genannt, die danach in Proteine übersetzt werden. Indem wir zeitgleich die Menge an Protein und zugehöriger mRNA Kopie gemessen haben, konnten wir feststellen, dass lediglich 15% der Proteine, die zum Ablauf des Zellzyklus beitragen, durch die Menge an mRNA Kopien reguliert werden.


List of Publications

Publications and manuscripts included in this thesis

I. **Spatiotemporal dissection of the cell cycle with single-cell proteogenomics.**

II. **Deciphering hierarchical cell cycle controls by near-saturation phosphoproteomics**
Cesnik A.J., Gnann, C., McCarthy, F., Itzhak, D.N., Lundberg, E.
Manuscript

III. **Widespread enzyme expression variations underlie diverse metabolic capacities within cell types.**
Manuscript

IV. **An image-based map of the mitochondrial proteome reveals widespread metabolic heterogeneity**
Manuscript
List of Publications

Respondent’s contribution to the included work

**Paper I:**
Contributed to the experimental work for the imaging-based subcellular proteomics screen, responsible for generation of mNG-tagged cell lines for ab validation and time lapse experiments as well as data analysis and interpretation thereof. Joint responsible for data visualization. Participated in manuscript writing.

**Paper II:**
Joint responsible for experimental design and experimental work, contributed to data analysis concerning location overrepresentation, joint responsible for data visualization and manuscript writing.

**Paper III:**
Main responsible for study design, experimental work, data analysis, interpretation, visualization and manuscript writing.

**Paper IV:**
Main responsible for study design, shared main responsible for experimental work, data analysis and interpretation, main responsible for visualization and manuscript writing.
Extended list of publications not included in this thesis

**Deep Visual Proteomics defines single-cell identity and heterogeneity.**

**OpenCell: proteome-scale endogenous tagging enables the cartography of human cellular organization.**

**Illuminating Non-genetic Cellular Heterogeneity with Imaging-Based Spatial Proteomics.**

**Mapping the nucleolar proteome reveals a spatiotemporal organization related to intrinsic protein disorder.**

**A Sample Preparation Protocol for High Throughput Immunofluorescence of Suspension Cells on an Adherent Surface.**
Acknowledgements

ting section almost exactly 9 years after I first moved to Stockholm. At that time, I had not even wasted a single thought about a career in academia or doing a PhD. And now this thesis marks the end of my time in Stockholm (for now). Those 9 years and especially the last 5 years during which I did my PhD took me on a rollercoaster ride with many ups and many downs, they shaped me to become a better version of myself and most importantly they allowed me to meet many incredible people along the way. And while this short section cannot possibly do justice to the impact that all of you had on my work and more importantly, my life, here is an attempt to put my gratitude to paper.

Emma - if the PhD was a rollercoaster you were the rails guiding me safely through the entire ride! Thank you for being with me from the start and welcoming me into Cell Profiling all the way back in 2015. I have looked up to you since day one since your scientific vision and capability to connect the dots is unmatched. After every single one of our interactions I was in a better place than before. Thank you for believing in me when I did not myself, for pushing me when I needed it and for always listening. Thanks for letting me stay in your apartment, letting me borrow your car and feeding me the best Swedish julbord that I have ever had - ironically in the US. It has been a privilege working for and with such a great scientist and even better person.

Manu - thank you for welcoming me into the beautiful world of OpenCell and the ML group. Your enthusiasm for and knowledge of cell biology is infectious and inspires me to this day and your humor and your jokes turned working with you into a real blast. I made countless valuable memories and experiences and your karaoke skills are unmatched.

To all my colleagues and co-workers who became friends and made me enjoy coming into work. Diana, you left - but then you came back and you can’t imagine how happy I was when I heard the news. You are one of the calmest and most caring people I know and have been a rock for me to lean on in the last few months. Anthony, you are a magician of the English language. I learned so much from you working together on all these papers - one could say you “boosted” my scientific development? Thanks for exploring SF with me and taking me to random restaurants in the Tenderloin that I would have never found without you. Xiuqi, you were the best Master student I could have imagined. Thank you for all the help over the months and good luck with your future endeavors - you are absolutely fantastic! Alina, give me the email of your biochemistry professors; I will let them know that you are a metabolism genius! Thanks for all the valuable input and the conversations about everything and nothing. The backseat boys - Jan, Casper, Emmie and Frida along with our number one supporter Nicolai. I miss the Baywatch evenings in our Airbnb but “I’ll be ready” to tune in again! Casper, I think of you at every red light. Your baby animal smile can get me through a day when things are not great and...
your computational expertise got me through countless coding and statistics problems. Jan, I still believe that controlled randomness is the best strategy at the Kicker table. Thank you for taking my mind off work during all those afternoon sessions in the gamma lunchroom and as a result even inspiring the title of my thesis! Emmie, I will not forgive you for winning at Bingo and not screaming out loud! Thank you for all the laughs around the office, your positive vibe is exhilarating. Frida, how’s life? You are the best desk neighbor and the true baKING! Thank you for all the great conversations and for sharing your wisdom with me - “don’t judge a cake until it’s cut” could turn into my life motto. Nicolai, I’m glad we survived the Shining hotel in the Swiss alps and all our other “sinful” experiences - but I think it’s a generational thing. Anna MC and Anna B - my Fisherman’s friends. Anna B you are probably the nicest person I know. I will miss your smile when coming to the office in the morning. Anna MC, you tried so hard but I still have not seen Totoro. Thank you for always being so incredibly helpful and supportive and for making my days so many times. Mariya, your drive is inspiring. Thank you for all the late evening conversations when the office was dark already and all the fikas we shared. Sanem, we suffered together in the last few months but we got it done. Your support meant so much to me! Thank you for giving me access to a private zoo. Anna M and Johan my Strava buddies. I’m looking forward to many more kudos in the future! Trang, we started at the same time and will end it on different continents! Thank you for all the nice memories of picnics and Christmas cookie baking. Charlotte, you left a lasting mark when I first met you and you started rearranging the desks in the inner office. I really appreciate your get it done mentality. Burcu, thank you for listening and helping with important life decisions about moving to the US. Ulrika, you are the glue that holds this group together! Thank you for showing an incredible level of empathy and for always being there. Jenny, one of the last dinosaurs. The most beautiful voice in science and an amazing person on top of it. Iva, don’t let anything bug you! Tony and Pranauthi, you might sit on the dark side but you shine bright like a diamond. Eleanor, thank you for all the laughs - the inner office would not be as fun without you! Frede, “naturally if you collect such a large amount of data” you will need a genius providing you with annotation tools! Carolina, you keep this lab running! Alex, I aspire to be appreciated by other people the way you appreciate your leverpastej sandwich. Kalliopi thank you for keeping a big smile on your face even when the robot is failing you. Katrin, danke für die Langlauftrips. Ankit, Ishan, Xikun, Will, Ernst, Maddie, Irene, Thanks for all the valuable scientific discussions. Maria, Bruno, Rafael, you only joined the lab recently but you make it a better place.

Thank you to all former CellProfiling members. Over the years, you have not only left the group but also a mark in my heart. Andreas, we had to pivot a bit due to the pandemic but I still had an amazing time supervising you. Rutger for dancing to “All the single ladies” in the cell lab and teaching me everything I needed to know about immunofluorescence as well as some survival skills. Lovisa, thank you for teaching me a random collection of Swedish words and terms. Wei, thank you for all the discussions around cell segmentation and introducing me into the world of artificial intelligence and Chinese cuisine! Peter, “CONTAMINATION” - you were my supervisor, then my boss and a major reason why I even started this PhD journey. Oana and Devin, in you I see the scientist that I aspire to be. Devin, I could mention project unicorn or project phoenix
but much more importantly - you introduced me to sour beers! Thank you for super bowl watch parties, exchange rate misunderstandings during our culinary adventures in Tijuana and for discovering a bridge under a bridge under a bridge. Oana, I have learned so much during our biology talks at the Biohub. THANK YOU for pointing me to the shrimp burrito in Pescadero, showing me around Pittsburgh and all our phone calls.

Thank you to everyone at alfa 2 and Scilifelab for making work feel like a second home all those years. Jennie, if you “come to work to fika” then I come to work to fika with you. Thank you for all the afterworks and the Vasaloppet training! And I am looking forward to open the BathPub, even though I still prefer toilet seats. Sara, Max, David, thank you for turning Diskus- into Discovägen on a regular basis. There is nothing like screaming Tokio Hotel lyrics at 2 am. Hasta la vista, Ceke, thank you for teaching me everything I know about the Tecans. Lovisa, thank you for helping out so much with the Tecans and for all the beers outside work! Eni, I am still a research engineer at heart. August, your halloween costumes are unmatched. Sofia, your plant is still thriving. Annika, Leo, Matilda, Maryam, it’s 11:30 - the race for the microwaves is on. André, at some point I’ll actually join Advent of Code. Andreas and Fredrik, your encouragement was a big reason for me to do Vasaloppet and get outside my comfort zone. Jochen, Danke für die regelmäßigen, aufmunternden Worte! Sami, thank you for all the locker room talk. Banu, thank you for our Indian-German cook off. I learned so much from you and I am still amazed by your enthusiasm for Schupfnudeln. Angelo, keep those puns coming! Nils, for driving me mad with your blocking skills at the kicker table. Phil, Dörte, Rozina, Maria, Chus, Lea, Shuai, Amena, Julia, thank you for all the small but meaningful conversations in the hallways, it is so nice to bump into a smiling face.

A big thank you to Mathias for starting the HPA project and enabling the starting point for this thesis. HPA IT and Fredric in particular for helping out with so many small things along the way. And of course, thank you to the entire HPA family. None of this would have been possible without you.

To everyone at the CZ Biohub, my home for the first few months of my PhD. Preethi, Nathan, Joana, Laura, James, Edna, Rachel, Hera, Jason, I do not miss culturing 1,000 cell lines at a time but I do miss our Monday sorting sessions. All for the pipeline! Thank you for making the ML group so special, for meme of the week, quote of the week, all the laughs in and out of the lab and for feeding my love of burritos. Preethi, thank you for all the good memories and introducing me to dosas. Baby goats and deers and screw pelicans! Joana, you are the reason why I was able to genuinely enjoy a few haircuts in my life. James, I think we might have lost the battle of the clones but at least we had Champions league watch parties during lunch breaks. Hera, the sequencing guru! Thank you for teaching me about the evolutionary advantages of being hangry. Dan, you are a mass spec wizard. The three weeks in the Stanford basement with you and Frank, and all the jokes and football trashtalk, were some of the best in this PhD.
Acknowledgements

Thank you to all my collaborators over the years. Yun Chien, thank you for showing me around your lab in Munich. Fabian and Andreas, I’m proud to have worked with you - thank you for all the chats at conferences. Malene and Eric, skeletal muscle is stunning! It has been fun working with tissue. Hanna, thanks for running the opera for us and generating a huge amount of data in very little time. Olivia, it didn’t work out but at least we had a good reason to drink beers at conferences.

To all the people that have kept me sane outside work. Zaneta, it’s crazy how we went from sightseeing in SF to living together and surviving the pandemic together! And would you look at those joshua trees now! Mike and Margaux, I am so glad I met you. You are some of the funniest people I know. Thanks for always having my back, for all the cabin weekends and for revolutionizing the hydration drink industry - Ölectrolytes will be huge. Miren and Marcel, I was incredibly happy when you got the apartment below us. Our weekly food ordering tradition got me through the last few weeks. Sandra, only an absolute badass runs a marathon without training. Thank you for letting me (or rather Kalle) convert you into a dog person! Alma, you are an inspiration - both in science and on the trails. Everyone from the Master’s for all the Theodoras beers over the years. We might be in different places now but we are still close! Okan, you are always there for me and you give the best hugs! Let’s get this German castle, a food truck and a couple doves. Marcus, how fast does it go? You are the human version of a Gambler shirt - no bullshit and I can always rely on you. Matina, you are amazing and I trust you with everything - except for your egg soup recommendation. Rico, Rebeca, Milo thank you for all the barbeques, the beers, the football games, the chats and so much more in the Kungshamra forest. Schorsch, danke für die Kayakingtrips und ein Gefühl von Heimat durch deine Knauzen! Lena, dein Midsommarkuchen war der Beste, den ich hatte! Elena, wir sind zusammen in Stockholm angekommen und haben dadurch erst so richtig zueinander gefunden. Danke, dass es so leicht ist, mit dir zu reden. Dein Drive ist inspirierend! Fenja, wo ist das Mammut?! Krass, wie schnell du nach nur ein paar Tagen auf einer Konferenz, so schnell so wichtig für mich geworden bist. Tanja und Oscar Danke für die Pizzaabende und die Spaziergänge! Felix, danke für die (Lang)Lauftouren - Wenn ich in den nächsten Jahren meinen Ironman anpeile, dann auch wegen dir. Paw Patrol, Axa, Oscar, Ärla, Bea, Pontus, Aska. Thank you for all the tips, the training sessions and helping out with Kalle. Huchi, ich habe überlegt, ein “h” im Titel wegzulassen. CAOS forever! Wir sollten mal wieder die Delfindecke auspacken...

Jutta, Gerd und Moritz, danke, dass ihr mich so schnell in eure Familie aufgenommen habt. Jutta, ich bin wahnsinnig stolz, deine Kunst auf dem Cover zu haben.

# Table of Contents

**Thesis Defense** ........................................................................................................................ i
**Abstract** ..................................................................................................................................... iii
**Sammanfattning** ......................................................................................................................... v
**Popular scientific summary** ........................................................................................................ vii
**Populärwissenschaftliche Zusammenfassung** ................................................................. ix

**List of Publications** ................................................................................................................. xiii
  - Publications and manuscripts included in this thesis ............................................................ xiii
  - Respondent’s contribution to the included work ................................................................. xiv
  - Extended list of publications not included in this thesis ...................................................... xv
**Acknowledgements** ....................................................................................................................... xvii

**Table of Contents** ................................................................................................................... xxiii

## Chapter 1 - Cells: the building blocks of life .................................................................................. 1
  1.1 The central dogma of biology ............................................................................................... 1
    - DNA - the genetic blueprint .............................................................................................. 1
    - RNA - the messenger......................................................................................................... 2
    - Proteins - the machinery ................................................................................................. 3
  1.2 Beyond the central dogma - a complex proteomic landscape .............................................. 4
    - Morenaments - Post-translational modifications ................................................................ 4
      - Switchcraft - Phosphorylations as regulatory protein power buttons ......................... 4
    - Quantity matters - Regulation of protein levels ................................................................ 5
  1.3 Biological Cities - Subcellular Compartmentalization .......................................................... 5
    - Fire up the sauna - Mitochondria, the powerhouse of the cell ....................................... 6
  1.4 The ever-changing subcellular landscape ............................................................................ 7
    - A phoenix without ashes - The eukaryotic cell cycle....................................................... 7
    - An intricate maze - subcellular metabolism ...................................................................... 9

## Chapter 2 - The complexity of the cellular landscape ................................................................. 11
  2.1 Swiss pocket knives - Multilocalizing proteins .................................................................. 11
  2.2 Same same but different - Cell to cell Heterogeneity ....................................................... 12
    - “The differences that make a difference” ........................................................................ 13
    - Random noise or organized chaos? - Studying cellular heterogeneity ......................... 15
Table of Contents

Chapter 3 - The technological toolbox.................................................................17
  3.1 It’s all about reads - Single-cell RNA Sequencing .....................................18
  3.2 Peak performance - A new era of mass spectrometry...............................18
  3.3 Location is everything - Subcellular Proteomics.......................................19
    3.3.1 Everything everywhere all at once - MS based subcellular proteomics.....19
    3.3.2 Seeing is believing - Imaging based Spatial proteomics ........................20
      3.3.2.1 Green lantern - Fluorescent protein tagging .................................21
      3.3.2.2 Antibody sandwiches - Immunofluorescence .................................21
        3.3.2.2.1 Human Protein Atlas - Mapping the human proteome one protein a time..23
    3.3.3 “Hello world” - Converting images into numbers .................................23
Chapter 4 - Present investigations and thesis aims .............................................25
  Paper I: Spatiotemporal dissection of the cell cycle with single-cell proteogenomics......25
  Paper II: Deciphering hierarchical cell cycle controls by near-saturation phosphoproteomics .....27
  Paper III: Widespread enzyme expression variations underlie diverse metabolic capacities within cell types...........................................................................................................28
  Paper IV: An image-based map of the mitochondrial proteome reveals widespread metabolic heterogeneity ........................................................................................................29
Chapter 5 - Concluding remarks.........................................................................31
References .............................................................................................................35
List of abbreviations ...........................................................................................48
At the core of biology lies the fundamental unit of life - the cell. Through billions of years, life has evolved from unicellular organisms to increasingly complex multicellular species such as humans. A human individual contains around 30 trillion cells\(^1\), each with an intricate molecular machinery. Cells work together to form tissues, organs and entire organisms. Each tissue carries out specialized tasks. Muscles are needed for movement, while the gastrointestinal tract is required for digestion of food and the extraction of nutrients. Similarly, even the building blocks of life - cells - are divided into a variety of functional subcellular compartments called organelles. In this chapter, I will cover different areas of the multidisciplinary field of cell biology to reveal how cells are able to translate their biological information into function, how their intricate subcellular organization allows them to run many biological processes simultaneously, and as a consequence why the subcellular location of cellular parts matters. Finally, we will dive into a novel area of biological research: Dissection of how heterogeneity between supposedly identical cells can fine-tune cellular function and drive dynamic cellular processes, ultimately uncovering its roles in the complex symphony of life.

### 1.1 The central dogma of biology

A cell is composed of approximately 70% water, with the remaining dry mass containing largely organic molecules as well as a small number of inorganic ions. However, a simple list of ingredients does not provide a lot of information about the final product. For example, would you be able to confidently state that carbon and oxygen along with trace amounts of zinc, sulfur and other elements constitute a car tire? Similarly, in a cell, these chemicals form larger functional molecules, e.g. rubber, that collectively constitute a cell (or car tire). Examples for such macromolecules are nucleic acids like DNA and RNA as well as proteins and lipids.

But how do we get from macromolecules to biological function, and how do the different macromolecules interact with each other? At the center of all cellular activities lies the central dogma of biology.\(^2\) In short, this principle describes the flow of biological information from DNA to RNA and subsequently to proteins. The DNA can be seen as a genetic blueprint, a copy of which is sent to the factory in the form of messenger RNA (mRNA), where the actual functional machine - the proteins - are produced (Fig 1).

#### 1.1.1 DNA - the genetic blueprint

The DNA stores all information about the possible functions an organism can exert. Essentially, DNA is a chain made up of four different nucleotides; adenine (A), cytosine (C), guanine (G) and thymine (T). This nucleic acid chain is organized into a double helix by establishing hydrogen...
bonds between the complementary base pairs A and T as well as C and G. The total genetic material of one human cell is estimated to a length of around 2 meters. With human cell sizes ranging from a few to hundreds of micrometers depending on the cell type, those helices have to be strongly compressed which is achieved by compacting the DNA into chromatin. A sequence of nucleotides of which a copy (=RNA) can be obtained is defined as a gene. The number of genes in a human cell is currently estimated to be ~46,000, among them around 20,000 protein coding genes.

Figure 1 | The central dogma of biology (and beyond).
The central dogma describes the flow of genetic information in a cell. A functional segment of DNA (gene) is transcribed into mRNA which is translated into amino acid chains (proteins) at the ribosomes. The protein then folds into a 3D structure allowing it to carry out its specific cellular function. While the central dogma suggests that one gene codes for one protein, alternative splicing and post translational modifications (PTMs) establish a multitude of different functional proteoforms from just one genetic sequence.

1.1.2 RNA - the messenger

Each cell in the human body contains the same genetic information (DNA) or a complete list of parts. However, as cells in different tissues are functionally different, they require a different set of parts and therefore not every gene is produced at all times in every cell. Instead, each cell only produces copies of the required parts. Those copies are called RNA. While the central dogma specifies the flow of information from DNA to RNA to protein, there are genes that do not lead to the production of proteins. Those genes encode non-coding RNAs such as lncRNA or miRNA, that have extended functions in many biological processes and diseases. However in the context of this thesis, the focus lies on the around 20,000 protein coding genes that can be transcribed into mRNA. Similar to DNA, mRNA molecules are made up of 4 nucleotides; however, instead of T, mRNA contains the base uracil (U). Furthermore, mRNA molecules do not form a double stranded helix but are instead single-stranded. Once a copy of the DNA has been obtained by transcription, the genetic information has to be translated into function. This is achieved by translating mRNA into a protein. Importantly, a cell can obtain multiple copies of a gene depending on its current needs. The abundance of mRNA transcripts for a particular need is therefore a way to regulate cellular function by adjusting protein levels.
1.1.3 Proteins - the machinery

Proteins are the ultimate effectors of the genetic information and the workhorses of the cell. They consist of long chains composed from a combination of 20 different amino acids. Evolution has established a mechanism, where each sequence of three unique nucleotides on the transcribed mRNA encodes for one particular amino acid. This genetic code ensures the translation from each mRNA to the correct incorporation of amino acids into the protein chain. Each of the 20 different amino acids has different chemical properties. This enables a multitude of non-covalent chemical interactions between amino acid residues across the entire protein molecule eventually leading to the folding of the amino acid chain into a complex three dimensional (3D) structure. Ultimately, the wide variety of naturally evolved 3D structures enabled the astonishing diversity of molecular functions performed by proteins. However, some regions can also remain unfolded and are referred to as intrinsically disordered regions (IDRs). Notably, the lack of structure can in some cases be as important as a completely fixed structures and protein disorder has important implications across many biological functions, such as signaling or subcellular organization.

Therefore, proteins are the centerpieces of the cellular machinery as well as the link between genotype (the genetic information) and phenotype (what is happening in the body). As an example, the central dogma would fail without proteins at the first essential step of transcribing DNA into mRNA. During the process of transcription, the mRNA copy is synthesized by RNA polymerases. In those protein complexes, multiple independent proteins come together as subunits to form a larger structure. Another example are ribosomes, protein complexes responsible for translating the RNA template into protein. Comprising two ribosome subunits and ribosomal RNA (rRNA), the small subunit identifies and reads the mRNA, recruiting complementary transfer RNA (tRNA) anticodons (each representing a codon of three nucleotides) carrying specific amino acids. The large ribosomal subunit then incorporates the amino acids into the growing polypeptide chain. This process repeats as the ribosome moves three nucleotides along the mRNA until the entire protein synthesis is completed. Those examples underscore the fundamental role of protein structure for protein function.

While the most straightforward definition would simply label proteins as amino acid chains, their significance in the biological realm transcends their molecular composition. They can solve complex cellular problems with astonishingly high speed and accuracy.

Borrowing the voice of Queen Beyoncé: Who runs the cell? Proteins!
1.2 Beyond the central dogma - a complex proteomic landscape

The central dogma proposes that one gene will establish one protein. The human cell contains approximately 20,000 protein coding genes implying a similar number of proteins. However, recent estimates suggest that the number of unique molecular forms of proteins, also termed proteoforms, may be 100,000 or even millions. How is an over 5-fold increase from the number of genes to the number of proteins possible? Firstly, post-transcriptional modifications, such as alternative splicing, can generate distinct proteins or isoforms from an identical genetic sequence by cutting the RNA. Secondly, amino acids can be modified via many different post-translational modifications (PTMs). In summary, each amino acid sequence with modifications at specific positions defines a new protein molecule, i.e. a new proteoform, each with a potentially distinct structure and function.

1.2.1 Morenaments - Post-translational modifications

PTMs are a collective term for protein processing events through proteolytic cleavage or by addition of a modifying group to one or more amino acids. With over 400 types of PTMs, proteins can be conceptually compared to christmas trees decorated with diverse ornaments. However, PTMs are not just ornaments, they possess the capability to significantly alter protein properties, influence their activity state, interactions with other proteins, protein turnover, and more. The most abundant PTMs include phosphorylation, acetylation and ubiquitination which can be found on over 90% of the proteins.

1.2.1.1 Switchcraft - Phosphorylations as regulatory protein power buttons

The most extensively studied PTM and one of the focal points of this thesis is protein phosphorylation. Phosphorylation is a modification of typically a serine, threonine, or tyrosine residue by replacing the neutral hydroxyl group of the amino acid residue with a negatively charged phosphate group. This modification is performed by protein kinase enzymes. Due to the resulting changes in molecular charge, the protein may undergo conformational and thus functional alterations. Notably, phosphorylation is a reversible process, facilitated by phosphatase enzymes capable of removing the phosphate group. Therefore, protein phosphorylation functions as an important “regulatory switch” that is crucial for myriad cellular processes. They allow for rapid transition of information in cellular signaling, play important roles in cellular metabolism and are integral to progression through the cell cycle. Consequently, aberrant protein phosphorylation states can lead to severe pathologies, making protein kinases an important drug target. A more in-depth discussion on cell cycle and metabolism, and to an extent the role of protein phosphorylation in those processes, will be provided in subsequent sections of this thesis.
1.2.2 Quantity matters - Regulation of protein levels

The cell carefully maintains homeostasis of its proteome landscape, or the entirety of all proteoforms in a cell, as it is integral for cellular identity. Yet cellular processes can be dynamic, which may require an identity change. To allow for those changes, a cell can adjust its proteome by regulating for example the abundance of relevant proteins. But what factors can affect the abundance of proteins in a cell and how can differences be established? According to the central dogma, the most obvious answer is the abundance of RNA. RNA is often used as a proxy for protein levels but a linear correlation between protein and RNA is heavily discussed. Actually, protein expression regulation goes far beyond transcriptional regulation. In general, protein levels are in an equilibrium of production and degradation. Increased transcription can lead to increased protein production but proteins can also undergo targeted degradation. Additionally, PTMs have been shown to affect protein stability and might thus make an impact on protein degradation rate. However, a cell might not even be required to adjust the abundance of proteins to alter its performance but can introduce functional changes through PTMs. For instance, metabolic state switches are often regulated independent of RNA for example through structural changes of the associated proteins.

1.3 Biological Cities - Subcellular Compartmentalization

In the last chapter we have dissected the chemical composition of cells and established the importance of DNA, RNA as well as proteins and their modifications. However, the cell represents much more than a bag of macromolecules. The eukaryotic cell contains several subcompartments, termed organelles, which are separated from the remaining cell either by membranes or through their physicochemical properties (Fig 2). Organelles provide a variety of environments each suited for certain biological functions. For example, lysosomes are responsible for the digestion of macromolecules. This is enabled by their highly acidic environment (pH 4.5-5.0) compared to the rest of the cell (pH 7.0-7.4). If not for the segregation of those processes, the digestive enzymes from the lysosome would leak into the cell and the cell would digest itself. Furthermore, a change of intracellular pH would lead to protein misfolding, affecting the function of many proteins as well as causing a large amount of stress on the entire cell. However, subcellular compartmentalization does not only allow for the establishment of different chemical environments. Each organelle is populated by a unique set of proteins. These proteins together with the unique chemical environment of each organelle establishes their highly specialized function and turns the smallest unit of life – the cell – into a complex subcellular city. On a regular day in this city, microtubules (MT) provide the roads, the Endoplasmic reticulum (ER) produces proteins which are shipped to their various destinations through the Golgi apparatus, the cellular post office. Lysosomes handle waste and the nucleus controls life in the city by storing all of the cell’s DNA. None of these processes would be possible without energy, which is provided by the mitochondria – the cell’s powerplant (Fig 3).
1.3.1 Fire up the sauna - Mitochondria, the powerhouse of the cell

The mitochondria are widely known as the powerhouse of the cell.\(^47\) They produce the majority of the cellular energy in the form of adenosine triphosphate (ATP) through respiration, which is then transported to areas with cellular energy needs. But would it not be easier to produce the energy directly where it is required? Theoretically yes, however this reaction generates a large amount of reactive oxygen species (ROS) which can for example cause DNA damage in the nucleus. Furthermore, energy production in the mitochondria establishes an up to 15 °C warmer environment compared to the remaining cell.\(^48\) Collectively, these aspects illustrate the importance of subcellular compartmentalization for efficiency in cell biology. However, that is not the only difference between mitochondria and other specialized organelles. For instance, in contrast to other compartments, they are engulfed by a double membrane and are the only organelle in mammalian cells containing their own DNA.\(^49\) However, only 13 mitochondrial proteins are actually transcribed and translated in the mitochondria, while the rest of the mitochondrial proteome is transcribed in the nucleus and the proteins are then imported from the cytoplasm. Furthermore, the number of mitochondria per cell is maintained independently of the human cell through fission and fusion.\(^50\) Those unique characteristics all stem from a symbiotic relationship of an archaeon which engulfed alpha-proteobacteria. Those bacteria later evolved into mitochondria.\(^51,52\) Ultimately, this symbiotic relationship kickstarted eukaryotic evolution as it provided the energy necessary for the transcription and expression of complex...
Chapter 1 - Cells: the building blocks of life

Mitochondria are essential in the regulation of cell death (apoptosis) and are increasingly recognized as cellular signaling hubs due to their physical interactions with other organelles such as the ER and the nucleus. Interestingly, recent studies have discovered cell-free intact mitochondria in blood which could indicate not only intra- but intercellular signaling functions for this fascinating organelle.

1.4 The ever-changing subcellular landscape

Similar to a city, the spatial organization of the cell enables the performance of multiple tasks at the same time without interference. In a city, spatial separation is not the sole factor; processes can also be temporally segregated — consider street construction work, which is typically scheduled during early or late hours to minimize disruption to traffic. Similarly in a cell, several processes are highly dynamic in space and time; for example the cell cycle and metabolism, which will be covered in more detail in the subsequent sections. Consequently, we can describe a cell as both spatially and temporally compartmentalized.

1.4.1 A phoenix without ashes - The eukaryotic cell cycle

As mentioned in the previous sections, the human body consists of trillions of cells. The lifetime of some of these cells, such as most neurons in the brain and cells in the eye lens, is only limited by the lifetime of the human. However, the majority of cells in the human body experience wear and tear and need to be replaced. As a result, an average human replaces >300 billion cells every day. Thus, the process of producing fresh cells is essential to life and is achieved through cell division. During the course of the cell cycle, a cell grows, duplicates its DNA and then divides into two daughter cells with the same genetic information.

This delicate process is highly conserved across cell types and species, carefully controlled and takes place in 4 distinct phases; G0/G1 (gap 1), S (DNA synthesis), G2 (gap2) and M (mitosis).
The G1 and G2 phases are growth and checkpoint phases preparing the cell for the next stage of the division for example by increasing the cell size and production of proteins and genes required in the next steps of the cell cycle. During S phase, the DNA gets replicated to ensure that both daughter cells receive the same genetic information. Finally, after completing some final checks during the G2 (gap 2) phase, the cell enters mitosis where it actually separates and forms two new cells. Cells can temporarily exit the cell cycle and enter a reversible quiescent state (G0) after mitosis due to external signals like nutrient deprivation. However, upon stimulation quiescent cells are capable of re-entering the cell cycle which differentiates them from other non-proliferative cell states like terminally differentiated or senescent cells.

Figure 4 | The human cell cycle - a piece in 5 acts.
A) The human cell cycle takes place in 4 phases. During G1 the cell prepares for the division process. The cell may exit the cell cycle (G0) or duplicate its DNA during S phase. G2 is an important control phase before the cell enters the final act - mitosis where the cell divides into 2 daughter cells. G0 phase is a dormant stage in the cell cycle. B) The transition through the cell cycle is carefully controlled by CDKs and cyclins. When forming a complex, the CDKs get activated and induce the expression or phosphorylation of many targeted proteins necessary for the next cell cycle phase.

The transition into a new phase is carefully controlled by several cell cycle checkpoint proteins, many of which are cyclins and cyclin dependent kinases (CDKs) (Fig 4B). They are expressed in a cyclic manner and form a complex allowing the CDKs to activate their target genes by phosphorylation. This is one of the reasons why phosphorylation is considered one of the most important PTMs in cell biology. Ultimately, the activation of the checkpoint protein then leads to the expression of genes and proteins required for the next phase of the cell cycle. Importantly, their activity is inhibited by cell cycle checkpoint signals. There are three major cell cycle checkpoints. During G1 and G2 cell size checkpoints ensure that the cell is large enough for further progression through the cell cycle. Next, a DNA damage checkpoint in late G2 ensures correct replication of the DNA; i.e. if DNA damage pathways are activated, the cell will not be able to progress from G2 into mitosis. Lastly, during mitosis the mitotic spindle checkpoint ensures correct chromosome segregation.

Importantly, if cellular control fails and the cells keep dividing; in spite of for example DNA damage, this can lead to uncontrolled cell division which is one of the hallmarks of cancer.
Ironically, this can make the cell cycle not only an essential for life but also a possible driver to the end of life. In summary, the cell cycle is the birth of two new cells from the mother cell, which in theory is a simple concept. In reality, over the course of the cell cycle the entire cellular landscape gets remodeled and hundreds of genes, proteins and PTMs change their abundances.\textsuperscript{74,75} Therefore, the cell cycle showcases the dynamic and versatile nature of each cell in our body, underscoring that cell biology is never in a steady state.

1.4.2 An intricate maze - subcellular metabolism

Over the course of the cell cycle each cell switches between various functional states. For example, a cell’s energy needs differ between different cell cycle phases; the cell requires different building blocks to produce new DNA or new membranes to form the daughter cell.\textsuperscript{76,77} The process of breaking down nutrients (anabolism) and producing new building blocks or provide energy for the cell (catabolism) along with the handling of cellular waste is called metabolism.

While we can think of the cell cycle as a circular process (with some exit points), cellular metabolism is a complex maze. The different paths and areas of this network correspond to different metabolic pathways. Each metabolic pathway consists of a chain of chemical reactions converting molecules to better suit the functional needs of the cell. For example, the cholesterol biosynthesis pathway is responsible for the production of cholesterol, an important building block of cell membranes.\textsuperscript{78,79} Pathways in amino acid metabolism produce amino acids which can then be integrated into newly produced proteins. The pyrimidine and purine metabolism pathways provide nucleotides, the building blocks of RNA and DNA. The tricarboxylic acid cycle (TCA) cycle in the mitochondria produces most of the cell’s energy in the form of ATP. Other pathways such as xenobiotics metabolism get rid of exogenous toxic substances.

When cells undergo functional changes and the nutrient or energy needs fluctuate, the metabolic network can get rewired dramatically, e.g. over the course of the cell cycle\textsuperscript{80,81} or the circadian rhythm.\textsuperscript{82,83} Additionally, cells undergo metabolic shifts as a result of signaling or activation events. For example, macrophages adjust their metabolism after activation\textsuperscript{84,85} and the metabolism of stem cells is different from differentiated cells.\textsuperscript{86-88} However, quite often the relationship between cellular processes and metabolism is bidirectional. Metabolism is thus not only a product of the current cellular state but it can also be a driver of cellular fate. For instance, a lack of certain nutrients during a particular point in the G1 phase of the cell cycle, termed the restriction point, can stall cell cycle progression.\textsuperscript{89,90} Another example are metabolites that can directly alter transcriptional and epigenetic programs in pluripotent stem cells thus affecting their differentiation trajectory.\textsuperscript{91} Additionally, metabolism is integral to the identity of a cell as it allows cells to perform their specific functions. For example collagen synthesis pathways are much more active in fibroblasts,\textsuperscript{92} acinar cells in the pancreas produce many digestive enzymes\textsuperscript{93} and hepatocytes express higher levels of enzymes involved in drug metabolism.\textsuperscript{94} In summary, metabolism is a complex cellular network that can be finely tuned in response to changes in the cellular environment or the trajectory of the cell.
Chapter 2 - The complexity of the cellular landscape

Chapter I established that cellular processes are compartmentalized and that organelles have distinct proteomes allowing them to fulfill their specific tasks. I also covered several dynamic processes capable of changing protein and gene expression over time. In this chapter, I will introduce additional levels of complexity in cell biology that have previously often been overlooked.

2.1 Swiss pocket knives - Multilocalizing proteins

Proteins compose the cellular machinery; and akin to humans, diverse tasks unfold in different localizations: A PhD student spends long days in the lab while a lumberjack can be found in a forest. Consequently, mapping a protein’s subcellular location could provide insights about its biological role. A protein in the mitochondria for example is likely involved in energy metabolism, whereas a protein in the Golgi apparatus is likely involved in cellular transport. Importantly, protein location is independent of the expression level of the protein, and provides a level of cellular regulation on a similar scale compared to changes in protein abundance. Interestingly, around 50% of all proteins localize to multiple compartments, establishing a striking intricacy of the proteomic distribution in the cell (Fig 6). In other words, both the PhD student and the lumberjack might enjoy a cold beer in a bar while watching football together. Generally, there are multiple reasons for protein multilocalization: i) A protein can localize to different compartments if they are involved in similar processes, e.g. ribosomal proteins often localize to the cytosol and ER, where translation takes place, and to the nucleoli, the site of ribosome biogenesis and assembly. ii) Proteins can carry out independent functions in multiple locations turning multilocalizing proteins into the swiss pocket knives of the cell. A prominent example for such a multifunctional protein is GAPDH, a glycolytic enzyme, which is often considered a cytoplasmic marker protein. However, research from the last decade has revealed a role in apoptosis and transcriptional gene regulation in the nucleus for GAPDH. Interestingly, there is a discrepancy between the number of multilocalizing proteins (~50%) and the number of predicted multifunctional proteins (25%). This prediction could indicate that only half of the multilocalizing proteins are indeed multifunctional. However, it could also highlight a knowledge gap warranting the need for additional studies to decipher the potential non-primary functions of multilocalizing proteins.

Proteins can diffuse passively between different compartments, however protein localization may also actively change in response to cellular events. For instance, the cell cycle checkpoint protein CCNB1 increases in cytoplasmic expression during the G2 phase but its subsequent phosphorylation enables its translocation into the nucleus. This relocation ultimately results in the initiation of mitosis and emphasizes protein multilocalization as an essential tool for the
regulation of cellular processes. While the transcription and translation of a protein can take hours, protein modification as well as (re)location provides a rapid regulatory mechanism for the cell. This results in an incredibly dynamic subcellular proteomic landscape in response to changes in cellular state. Hundreds of proteins change their subcellular localization profile in response to signal pathway activation, progression through the cell cycle or pathogen infection. In a cell population, these processes are regulated in an asynchronous manner, i.e. each cell is in a different phase of the cell cycle or a different subset of cells have different signaling pathways activated. Therefore, protein localization differences, along with changes in their abundance and PTMs, can be sources for variability in a homogenous cell population.

2.2 Same same but different - Cell to cell Heterogeneity

In the previous sections, we established that the cell cycle and metabolism are dynamic processes with an enormous impact on the presence of proteins, transcripts and metabolites in each cell over time. This dynamic enables the cell to adapt its cellular function or phenotype over time. Ultimately, this raises the question whether we should consider a homogenous population of cells, i.e., genetically identical cells of the same cell type, as truly identical or whether we should appreciate that there is cellular diversity and complexity even among supposedly homogenous cell populations.

The term “heterogeneity” is defined differently across varying research fields; in this thesis, it relates to differences of clonal, genetically identical cell populations (Fig 5). “Different” and “identical” are commonly not used in the same sentence. However, should you consider two identical twins as the same person or should they be considered as unique - or heterogeneous - shaped by their environment and experiences? Since we consider identical twins as unique, we should apply this idea to homogenous cell populations as well. This chapter provides an overview about the “differences [that] make a difference” and their implications in many biological areas. Finally, I will highlight important aspects and considerations for the study of cellular heterogeneity.
Figure 5 | Definition of Heterogeneity across varying research fields

The term “heterogeneity” is defined differently depending on the field. i) In personalized medicine, it can be used to describe differences between individuals in a population (e.g., male or female sex). ii) In the cancer field, tumors are often described as heterogeneous depending on their cellular composition.114 iii) However, heterogeneity in tumors might also refer to genetic differences between cells in the same tumor due to a subset of the cells acquiring unique mutations.115-118 In the context of this thesis, heterogeneity is defined as the differences between genetically identical cells.

2.2.1 “The differences that make a difference”

In this section, I will provide a short overview about the myriad layers of cellular heterogeneity. A central piece to my work is based on microscopy following the dogma of “seeing is believing”. Observing cells under a microscope, one quickly realizes that cells, similar to snowflakes, are unique and appear in many different shapes and sizes.

“Looks don’t matter” does not apply in cell biology. Differences in cell size are linked to metabolic capacity or nutrient uptake.119,120 For example energy metabolism in the mitochondria is optimal
Chapter 2 - The complexity of the cellular landscape

at intermediate cell sizes and decreases with increasing cell sizes. As a consequence, cell size can have severe implications for cellular fate, for example by driving cells into senescence. Additionally, changes in cell size can coincide with differences in cellular shape, which in turn can affect cellular motility; for example the speed of macrophages chasing a bacterium. Furthermore, cells rarely exist by themselves and are heavily influenced by the cells around them, their microenvironment. Cells in the center of a colony are in contact with many other cells. This enables efficient cell-cell communication resulting in variability in signaling pathway activation across cells in a colony. Simultaneously, local cellular crowding also results in changes in the morphology of organelles. Furthermore, it has been shown that the cellular microenvironment can influence a cell’s susceptibility to viral infection.

Figure 6 | Cell-cell heterogeneity and protein multilocalization establish remarkable complexity in cell populations. ENO1 and GAPDH are known multifunctional proteins. They both serve as glycolytic enzymes in the cytosol and perform non-canonical functions in the nucleus. GAPDH is involved in apoptotic pathways and transcriptional gene regulation while ENO1 has RNA/DNA regulatory functions. Additionally, the two enzymes exhibit cell-cell heterogeneity in protein location and expression levels which establishes diverse cellular phenotypes in a cell population. (green=protein of interest, blue = DAPI, red = MT); scalebar corresponds to 10 µm)

Upon looking through a microscope, cell size and shape as well as the microenvironment are examples for obvious variability between individual cells. However, these obvious differences are only the tip of the iceberg. Under the surface, the Atlantis of heterogeneity remains to be discovered. I already touched upon heterogeneity in organelle morphology, which is common among many organelles including the endoplasmic reticulum, the cytoskeleton or the mitochondria. The organellar heterogeneity of the mitochondria, whether in size or morphology of the network, the amount of mitochondria per cell, the positioning of the mitochondria in the cell or mitochondrial DNA (mtDNA) mutations is linked to several diseases such as cancer, metabolic or neurodegenerative disorders.
In my thesis work however, I did not focus on the heterogeneity in size, microenvironment or organelle morphology. Instead, my research focussed on heterogeneity in protein or gene expression caused by dynamic cellular processes (Fig 6). For example, hundreds of genes and proteins change their expression - and/or location in the case of proteins - over the course of the cell cycle\textsuperscript{74,75,147}, the circadian rhythm\textsuperscript{83,148}, during cellular stress\textsuperscript{149} or after a signaling event\textsuperscript{108}. If not synchronized, cells in a population will differ in their respective states and thus exhibit differences in their gene and protein expression profiles. The timeline of those dynamics can range from a few minutes in a signaling response to multiple hours or days to complete a cell cycle.\textsuperscript{150} Cellular metabolism often changes as a result of the previously mentioned processes. For example nucleotide metabolism is more active during the S phase of the cell cycle and the levels of TCA cycle metabolites oscillate along with cell cycle progression.\textsuperscript{78,151} Additionally, recent work provided examples of metabolic cross-feeding in yeast. Here, a subset of cells produces metabolites that are then used by other cells. Not only were the cells phenotypically different, they also responded differently to drug treatment.\textsuperscript{152} Those examples highlight the importance of heterogeneity in cell biology. Furthermore, even cells in the same cellular state may exhibit variable gene expression as a result of stochasticity in transcription, overall transcriptional or translational rates or transcriptional bursting.\textsuperscript{153-155} Finally, recent studies revealed that the expression levels of certain genes can be heritable, establishing heterogeneity over generations.\textsuperscript{156,157}

In conclusion - there is not one type of cellular heterogeneity, but rather diverse and interconnected modes of variability. The constant crosstalk between many biological processes as well as morphological features and cell-cell interactions that themselves exhibit heterogeneity can trigger dynamic cell state switches. As a result, cellular heterogeneity is a central mechanism of cell biology with important implications for disease and its treatment. But how do we make sense of a system that is so intertwined? How do we bring order into the chaos? How do we make sure that what we observe is not just random noise but true biological meaning?

2.2.2 Random noise or organized chaos? - Studying cellular heterogeneity

Between all the layers of heterogeneity described in the previous section, the easy way out is to consider differences within cell populations as random. In fact, single cell heterogeneity in gene or protein expression has in the past been referred to as noise.\textsuperscript{155,158-162} While biological processes can be stochastic, the term “noise” does not resonate with me as it indicates a lack of biological meaning for cellular heterogeneity. For instance, stochastic heterogeneity serves as a tool to allow for state changes in cancer cells.\textsuperscript{163} Secondly, cell-to-cell differences could serve as a bet-hedging strategy, where (microbial) cell populations rely on phenotypic variability to survive in rapidly changing environments, for example during drug treatment.\textsuperscript{164,165} To illustrate, variability in the expression of antibiotic resistance genes aids a subset of bacterial cells to survive antibiotic treatment.\textsuperscript{166} While most bacteria would die in this setting, the survival of a few individuals with high expression of those genes ensures the survival of the population. Similar discoveries have been made with drug resistance in cancer cells\textsuperscript{167}, highlighting cellular heterogeneity as an inherent mechanism in cell biology and evolution. In fact, several
mechanisms have been identified in bacteria that can regulate the degree of stochasticity in gene expression.\textsuperscript{168,169} As a result, recent studies pointed out that gene expression exhibits chaotic patterns rather than random noise.\textsuperscript{170-172} In other words, while a chaotic system might appear noisy on first sight, there is order or a determinism to the cellular heterogeneity underneath. In order to understand whether cellular variability is deterministic, it is thus crucial to dissect its origin.\textsuperscript{153} Studying heterogeneity requires the measurement of the variability in relation to a biological phenotype; e.g. the cell cycle stage of a cell, its size, shape as well as its microenvironment or past and future trajectory.\textsuperscript{173} In fact, Foreman et al. recently challenged the view of stochastic heterogeneity and claimed that the “majority of expression variability results from cell state differences”.\textsuperscript{174} Therefore, alignment of protein expression measurements to an axis of heterogeneity, such as cell cycle progression, allows for disentanglement of seemingly random biological noise into organized chaos of deterministic heterogeneity (Fig 7).

\textbf{Figure 7 | Revealing order in chaos
When studying heterogeneity, the measurements have to be aligned to a potential origin of heterogeneity, e.g. cell cycle progression. Using this approach reveals determinism in seemingly random distributions of cells.}
In the previous chapters, I discussed how work of the last two decades, including the work of this thesis, challenges the traditional view of cell biology. We are now acknowledging the currently unfathomable complexity of a cell as the number of possible functional proteoforms far exceeds the number of DNA encoded genes. One mechanism for the cell to potentially increase the amount of functional differences, which is also discussed in this thesis, is protein multilocalization. Simultaneously, we have come to realize that the traditional view of “homogenous” cell populations is flawed and instead shifted our focus on understanding the functions and origins of cellular heterogeneity.113

**Figure 8 | Single-cell resolution is required for the study of single-cell heterogeneity.**
Two different samples yield the same results when analyzed in bulk. Only single-cell analysis can reveal the phenotypic variability in the cell population.

The goal of this thesis is to study the origins and consequences of cellular heterogeneity and protein multilocalization by assessing the entire protein or gene content, i.e. the (phospho)proteome and transcriptome. This chapter aims to provide insights into the content of the technological toolbox that we have at our disposal to achieve those aims.
Chapter 3 - The technological toolbox

Traditionally, omics studies have been performed in bulk. Importantly, to study cellular heterogeneity on the systems level, those methods had to be improved to single-cell resolution, as cellular plasticity might be masked in bulk measurements (Fig 8). With the conceptual invention of single-cell RNA Sequencing (scRNA-Seq) in 2009\textsuperscript{175} and its commercialization only a few years later\textsuperscript{176} we have entered the age of single-cell biology in omics.\textsuperscript{177, 178}

3.1 It’s all about reads - Single-cell RNA Sequencing

scRNA-Seq can be seen as the door opener to study cellular heterogeneity on the systems level. Before its development, it was possible to look into specific genes or proteins or cellular phenotypes, such as size or cell cycle duration\textsuperscript{179}, but never the cell’s entire transcriptome. A scRNA-Seq experiment relies on the dissociation of the cells or the tissues and separation into wells or individual droplets. Afterwards the unstable RNA is converted into DNA by reverse transcription cell specific barcodes are added to the transcripts from each cell. This step allows pooling of the sample material into a so-called library before the sample is sequenced.\textsuperscript{180} Single-cell transcriptomics has revolutionized the field of investigating cellular heterogeneity. ScRNA-Seq technologies have been used to characterise the role of heterogeneity in drug resistance,\textsuperscript{167, 181–184} stem cell differentiation and development,\textsuperscript{185, 186} immune cell activation\textsuperscript{187} or cancer metastasis.\textsuperscript{188} There are however certain challenges with scRNA-Seq methods. These include scRNA-Seq protocols only reliably capturing 10-40\% of all cellular transcripts, due the low abundance of many transcripts.\textsuperscript{189} This can lead to sparsity in the sampling introducing significant technical noise in the data; i.e. many genes are not detected in a cell - even if they are present.\textsuperscript{190} This technical variability needs to be distinguished from biological variability, complicating studies of cell-cell heterogeneity.

3.2 Peak performance - A new era of mass spectrometry

As established previously, proteins are the ultimate effectors of cell biology. RNA is often used as a proxy for protein levels but the lack of a linear correlation between protein and RNA is well established.\textsuperscript{34–38} Therefore, measuring proteins is crucial to get close(r) to the discovery of functional cellular phenotypes.

Mass spectrometry (MS) allows the identification of molecules by measuring their mass in relation to their charge.\textsuperscript{191} Notably, MS is not restricted to the measurement of proteins. In fact, any kind of molecule can be detected by MS, including metabolites or lipids. As I studied proteins in this thesis, I will focus on MS in the context of proteomics. Generally, two approaches towards MS proteomics exist: bottom up and top down. Top down proteomics relies on the characterization of intact, native proteins and is an essential tool for the study of entire protein complexes and distinct proteoforms.\textsuperscript{192} On the contrary, bottom up proteomics, which was used in this thesis, detects peptides, which are smaller parts of the full length proteins.\textsuperscript{193} To perform a bottom up proteomics experiment, the proteins of a biological sample are extracted by cell...
lysis. Then they are denatured prior to digestion with a proteolytic enzyme, typically trypsin. Finally, the samples are injected into the mass spectrometer. In short, the molecules are ionized before their deflection in a magnetic field is recorded. This deflection relates to a molecule’s mass-to-charge (m/z) ratio which serves as a molecular fingerprint. The observed spectra can then be analyzed and assigned to the correct protein fragment using peptide search engines. Importantly, any modification on an amino acid residue will result in a mass shift of the peptide thereby enabling the identification of phosphorylations and other PTMs using MS. In conclusion, MS is a versatile method allowing for the accurate quantification of proteins and other molecules in complex biological samples.

However, only in the last 5-6 years, MS proteomics has entered the single cell era due to “breathtaking” developments. While the method was lacking depth in the past, it is now possible to accurately quantify hundreds to thousands of proteins from a single cell using several approaches. Single cell MS has already been applied to profile cells in different stages of the cell cycle, reveal metabolic heterogeneity in cancer cells, reveal variability among immune cells and study the heterogeneity to drug response, or during cell differentiation. Those developments establish MS proteomics as a valuable tool which could possibly revolutionize the study of non-genetic heterogeneity.

3.3 Location is everything - Subcellular Proteomics

The expression of genes and proteins contributes significantly to cell type, state and fate. However, protein location is a regulatory mechanism independent of the expression level of a protein. Mislocalized proteins may have severe physiological effects and can even represent a hallmark of disease. The study of the subcellular localization of proteins is referred to as “Spatial Proteomics” or “Subcellular Proteomics”. In this chapter, I will provide a short overview about the possible technologies, which are divided into MS and imaging-based methods.

3.3.1 Everything everywhere all at once - MS based subcellular proteomics

There are many different approaches towards MS based spatial proteomics. Fractionation-based approaches rely on the purification of organelles from a cell lysate. This approach has been successfully applied for drafting the proteome of mitochondria or lipid droplets. In a novel approach, marker proteins for an organelle were genetically modified with a fluorescent protein. This tag was used as an affinity handle and allowed for the pulldown and subsequent MS analysis of intact organelles. Focusing on individual organelles allows the mapping of organelle specific proteomes, however it is laborious to perform separate experiments for each organelle in a cell to map the entire cellular landscape. An alternative approach relies on partial separation of the organelles in a centrifugation gradient. Quantitative MS of each fraction then reveals the abundance profile for each protein across all fractions. Proteins with a similar location will exhibit similar distribution profiles allowing the definition of organellar specific profiles. At this
Chapter 3 - The technological toolbox

.point, there are a myriad of protocols available for fractionation-based spatial proteomics.⁹⁷,⁹⁸,¹⁰⁷,¹⁰⁸,²¹³ Notably, their robustness and relatively simple experimental set-up allows for the study of protein translocations, for example in response to signaling events or during cellular stress caused by starvation or unfolded protein response.¹⁴⁹ It is worth noting that approximately 25-50% of all proteins present ambiguous profiles in organellar mapping methods, likely due to protein multilocalization, complicating the correct assessment of their subcellular localization.

An alternative approach to fractionation-based methods is the mapping of physical interactions, i.e. interactomics or proximity labeling. In interactomics approaches, an affinity tag, e.g. green fluorescent protein (GFP) or Human influenza hemagglutinin (HA), is genetically attached to a bait protein. Afterwards, affinity purification enables the capture of the engineered bait protein along with its interaction partners.⁹⁶,²¹⁷–²¹⁹ Alternatively, antibodies can be used to pull down the bait protein without prior genetic manipulation. Since proteins are only able to interact when they are physically close, profiling the interactions between a sufficient number of proteins enables the mapping of the subcellular organization of the proteome. Interactomics profiling has revealed substantial differences in the interactomes of different cell lines and across different cell states.²²⁰ However, this method relies on the direct interaction between proteins and might fail to capture transiently or weakly interacting proteins. To exemplify, if you consider proteins in a cell similar to humans on a train you would only capture the people, which a person of interest is physically interacting with, for instance by holding hands. However, you would miss all other people on the same train that are not holding hands with the person of interest, for example the conductor.

Proximity labeling methods overcome that limitation.²²² Here, a bait protein with a known subcellular localization gets tagged with an enzyme, e.g. engineered ascorbate peroxidase (APEX) or biotin ligase (proximity-dependent biotin identification (BioID)).²²³,²²⁴ Once activated, the enzyme will biotinylate all proteins in close proximity, i.e. 1-10 nm.²²⁵ Biotin thus serves as a marker for proteins in the same cellular location and can be used as an affinity tag. Using a streptavidin pulldown, all biotinylated proteins can be recovered and assigned the same subcellular location as the bait protein. The biggest strength of proximity labeling comes from the low distance of the labeling chemistry; enabling for example the distinction of mitochondrial proteins localizing to the outer and inner mitochondrial matrix, the ER-mitochondria contact sites, protein localization to the mtDNA or the study of organellar dynamics across different cell states and perturbations.²³² However and despite the recent improvements in MS sensitivity, they still require a substantial amount of input material. Accordingly, studies of the spatial distribution of proteins in single cells are not yet possible using these methods and require alternative approaches.

3.3.2 Seeing is believing - Imaging based Spatial proteomics

Mass spectrometry produces highly quantitative data in the form of graphs and peaks. Therefore, the most obvious advantage of using large scale imaging approaches to map the spatial
proteome of the cell is that the starting data are pretty images instead. But that is by far not the only advantage! Spatial Proteomics by imaging allows the study of the spatial location of proteins in intact cells in their native environment with single-cell resolution through fluorescence microscopy. For this reason, it took center stage in my thesis work. The fluorescent visualization of a protein can be achieved by multiple approaches, either by tagging the proteins of interest with a fluorescent protein or by using antibodies or other affinity reagents.

3.3.2.1 Green lantern - Fluorescent protein tagging

The birth of imaging-based Spatial Proteomics occurred almost 20 years ago, when a large portion of the proteome was localized in yeast using high throughput tagging with GFP. Yeast studies have continued to provide valuable information over the years, largely due to the high efficiency and relative ease for endogenous genetic manipulation in yeast compared to human cells. Endogenous tagging is crucial for performing spatial proteomics experiments, as protein overexpression can lead to aggregation and mislocalization. With the introduction of the clustered regulatory interspersed short palindromic repeat associated protein 9 (CRISPR/Cas9) technologies, endogenous tagging of proteins became feasible on large scale for human cells. Fluorescent protein tagging allows for the study of protein location in live cells and thus enables large-scale studies of protein translocation events. However, any modification of the target protein may affect its localization and therefore its function. The largest database of endogenously tagged fluorescent proteins in human cells is “OpenCell”, currently containing location information for 1,310 mNeonGreen (mNG)-tagged proteins in Hek293T cells. One of the biggest strengths of fluorescent proteins is their versatility. Not only do they allow for protein visualization under a microscope but they also serve as excellent affinity handles for protein pulldowns. This versatility allows the combination of MS-based interactomics with imaging-based subcellular proteomics.

Libraries of genetically tagged cell lines have been used successfully to study single-cell heterogeneity. However, since their generation is very time consuming, optical pooled CRISPR screens could serve as an alternative. Here, a barcoded library of guide RNAs can be used to modify hundreds of proteins with a fluorescent protein tag in the same sample. The entire population of cells is then imaged and the identity of the protein determined by in situ sequencing, ultimately allowing the connection between protein identity and phenotype. However, a major limitation of fluorescent protein tagging based approaches is that they are currently far from full proteome coverage in human cells.

3.3.2.2 Antibody sandwiches - Immunofluorescence

Another alternative for visualizing proteins is by performing an immunofluorescence experiment. This method relies heavily on the use of antibodies, which are more than a biotechnological tool but function in vivo as part of the adaptive immune system. The large, y-shaped glycoproteins consist of two identical heterodimers, with a heavy and light chain (Fig 9A). Each dimer comprises constant and variable chains. The constant chain is conserved for antibodies of the
same subtype; i.e. IgG, IgM, IgA, IgD, and IgE. The variable chain on the other hand is the reason for the widespread use of antibodies in biological research. *In vivo* antibodies are produced by B cells with the purpose to recognize foreign antigens through the structure of the variable region.\(^{247}\) As the name indicates, the variable region is variable; i.e. each B cell clone produces a unique antibody with a different variable region and thus an affinity to a different molecular structure. This is achieved through somatic hypermutation\(^{248}\) establishing a theoretical human antibody repertoire over \(10^{15}\) unique antibody clones.\(^{249}\) The capacity of an animal to produce antibodies targeting foreign antigens can be leveraged to produce antibodies targeting a specific human protein. Hereby, an animal can for example be immunized with the human protein and the freshly produced antibodies can be purified from its blood.\(^{250}\)

![Imaging-based Subcellular Proteomics using indirect immunofluorescence.](image)

**Figure 9** | Imaging-based Subcellular Proteomics using indirect immunofluorescence.  
**A)** Schematic illustration of an IgG antibody, which is most commonly used for immunofluorescence experiments. Heavy chain is shown in red, light chain in orange. The constant region is conserved for all IgG antibodies, the variable region is unique and recognizes a specific antigen.  
**B)** Workflow of an immunofluorescence experiment. Cells are seeded and fixed, before incubation with protein specific antibodies. Afterwards the sample is incubated with secondary antibodies that are coupled to a fluorophore which allows for the visualization of the protein using a fluorescence microscope. Protein location can be assigned according to the pattern of the protein signal in the images (green: protein of interest, blue: DAPI; scalebar corresponds to 10 \(\mu\)m).

Antibodies can then be used for performing immunofluorescence experiments. In indirect immunofluorescence (Fig 9B), which was used in this thesis, antibodies specifically recognizing
the protein of interest are incubated on the sample. Subsequently, a secondary antibody, which is coupled to a fluorophore, recognizes the primary antibody creating an antibody sandwich. The fluorescent signal can then be revealed by fluorescence microscopy (Fig 9B). In contrast to tagging methods, immunofluorescence experiments require sample preparation including fixation and permeabilization. Therefore, antibody experiments are limited to endpoint measurements. Furthermore, protein localization as well as cellular morphology can be affected by the sample preparation.  

In recent years concerns about the validity of antibodies have emerged, as some antibodies can yield unspecific and non-reproducible results. Therefore - when working with antibodies - seeing is only believing when the antibodies are thoroughly validated.

### 3.3.2.2.1 Human Protein Atlas - Mapping the human proteome one protein a time

Due to the enormous effort and time required to produce a proteome wide library of well validated antibodies, the number of global subcellular proteomics studies is limited to the Human Protein Atlas (HPA), which provided the framework for all studies included in this thesis project. The HPA is a Swedish research initiative initiated in 2003. Its aim is to map the proteome organization in cells and tissues using its in-house-generated proteome wide antibody library. All antibodies are rabbit polyclonal antibodies derived after immunization with recombinant protein epitope signature tags (PrEST) as antigens and purification using the antigen as an affinity ligand. Antibodies are then quality controlled for sensitivity and crossreactivity using western blotting and protein arrays before used in immunocytochemistry experiments. The HPA Subcellular Section (v23) contains subcellular localization data for 13,147 proteins. The HPA antibodies targeting the protein of interest along with antibodies targeting the ER and the microtubules as well as the nuclear dye 4',6-Diamidino-2-phenylindol (DAPI) are used to systematically stain up to three cell lines, one protein at a time. The DAPI, ER and microtubules stainings serve as reference channels to outline cellular morphology. After image acquisition on a confocal microscope, the images are annotated and assigned to over 30 different subcellular structures. This database has enabled the in–depth characterization of organellar proteomes, e.g. by revealing the crucial role of nucleolar proteins during mitosis. Additionally, it also showcases the complexity of the subcellular protein landscape as >50% of all proteins localized to multiple compartments and 22% of all proteins display single-cell variability in their expression.

### 3.3.3 “Hello world” - Converting images into numbers

Images are not just pretty and colorful, images contain an almost endless amount of information. For example, the brightness of the fluorescence signal correlates with the amount of protein present in the sample. Furthermore, the distribution of bright pixels within a cell provides information about the subcellular localization of a protein. Importantly, fluorescence intensity obtained from immunofluorescence experiments is only a semiquantitative measurement of protein abundance as the signal depends on other factors, such as the affinity of the antibodies.
Therefore, differences in the fluorescence signal of two separate proteins do not provide information about differences in protein abundance. However, it is possible to compare the relative abundance of the same protein across different cells thus enabling the study of cellular heterogeneity. But how can we extract this information?

Images are an array of pixels and each pixel has a value depending on its brightness. By analyzing the local neighborhood of pixels it is possible to identify borders in the images. Those borders can then be used for cell segmentation. For example, the nuclear dye DAPI is predominantly present in the nucleus, which will therefore appear bright in the image while the cytoplasm will be dark. Detecting the border between dark and bright pixels allows for nuclear segmentation. Similarly a cytoplasmic or membrane marker can be used for cell segmentation. Segmentation of cells in an image is crucial to obtain single-cell resolution and allow for the study of heterogeneity with fluorescence microscopy. The pixel information inside the cell objects can be obtained and converted to a number which can be used in statistical analysis.

The rise of artificial intelligence (AI) has revolutionized the field of biological image analysis, by providing accurate cell segmentation classifiers\textsuperscript{260}, predicting protein localization\textsuperscript{261,262}, extracting information about the subcellular location of organelles\textsuperscript{127,263,264} and more. Images are thus much more than just images, fluorescence microscopy data holds intricate information about the subcellular organization of proteins, beyond protein localization. Similarities in the images can be extracted and reveal proteins that act in the same complexes. Additionally, protein localization data can be integrated with protein-protein interaction data to reveal additional information about the proteomic organization of a cell.\textsuperscript{96,265} AI in image analysis is incredibly powerful and will be necessary in future studies of cellular heterogeneity, however, a detailed discussion is outside the scope of this thesis. In conclusion, the future of microscopy is bright (pun intended) and it will continue to be at the forefront of cell biological research.
Chapter 4 - Present investigations and thesis aims

The overarching aim of this thesis is to study non-genetic protein expression heterogeneity in cell populations over time. In paper I, we characterized the single-cell expression heterogeneity of >1200 proteins in the context of the cell cycle through a combination of single-cell RNA Sequencing and imaging-based subcellular proteomics. We identified hundreds of novel cell cycle proteins but showed that the majority of those proteins are not regulated at the level of transcript expression. To assess whether those proteins are instead regulated through PTMs, we conducted a deep phosphoproteomic study with unprecedented depth in Paper II using the same cell line model. This dataset allowed us to identify over 2000 proteins with cell cycle dependent (CCD) phosphorylation changes. Thus, we were able to link the phosphorylation changes to dynamics in protein expression and localization changes. Paper I also established widespread cell cycle independent protein expression heterogeneity; particularly for metabolic enzymes. In Paper III, we leveraged imaging-based single-cell spatial proteomic data from the HPA database to characterize the spatiotemporal partitioning of the metabolic proteome in human cells. We show that around 40% of metabolic enzymes exhibit cell-to-cell expression variability, largely independent of cell cycle progression. This heterogeneity is conserved in situ and can be manifested from a single cell. Furthermore, we integrated public protein-protein interaction data to reveal novel, compartment-specific functions for multilocalizing enzymes. Taken together, we suggest that the expression heterogeneity of enzymes can establish diverse cellular phenotypes. To test this hypothesis, we devised a large imaging-based spatial proteomics screen in Paper IV. Here, we leveraged imaging-based single-cell spatial proteomic data to reveal the extent of non-genetic partitioning of the metabolic proteome in the mitochondria.

Paper I: Spatiotemporal dissection of the cell cycle with single-cell proteogenomics

The HPA Subcellular Section reveals that ~20% of all human proteins exhibit cell-cell heterogeneity in protein expression in a supposedly homogenous cell line population. We hypothesized that a large portion of this variability could be explained by cell cycle progression. Protein dynamics across the cell cycle have previously been studied in bulk, oftentimes by chemically synchronizing cell line populations. Here, we set up a targeted proteogenomics screen, where we integrated scRNA-Seq with immunofluorescence imaging of 1180 proteins in U-2 OS Fluorescence Ubiquitin Cell Cycle Indicator (FUCCI) cells to study cell cycle related expression dynamics with single-cell resolution. U-2 OS FUCCI cells express fluorescent protein tagged cell cycle marker proteins in the nucleus; mKO2-CDT1 is expressed during G1 and S phase and mAG-GMNN is expressed during the S and G2 phases. The FUCCI markers allows
for the precise cell cycle mapping of each cell and therefore an alignment of protein and transcript measurement to the cell cycle. (Fig 10) This approach allowed us to identify 320 proteins (27%) with CCD expression during interphase along with 230 proteins that localize to mitotic structures and are therefore classified as CCD. 301 proteins did not have previous associations to the cell cycle. Notably, the majority of protein expression variability could not be explained by cell cycle progression. By integrating scRNA-Seq data with our protein imaging data, we were able to investigate whether the expression of CCD proteins is regulated transcriptionally. We then validated cell cycle related functions for novel CCD proteins by performing functional short interfering RNA (siRNA)-mediated gene silencing of several proteins, assessing their effect on proliferation.

Figure 10 | Workflow for Paper I - a targeted single-cell proteogenomics screen. The expression of mKO2-CDT1 (red) and mAG-GMNN (green) in the nuclei of FUCCI cells can be quantified by FACS or fluorescence microscopy to pinpoint the exact cell cycle position of each cell. Protein expression can be measured by fluorescence microscopy and transcript levels are obtained by scRNA-Seq. Protein and RNA expression can then be correlated to cell cycle progression.

Altogether, this work established a spatiotemporal map of the human cell cycle, which is now integrated into the HPA database. We identify many novel CCD proteins that could potentially serve as new clinical markers for proliferation or even novel drug targets. Furthermore, we show that the majority of protein expression heterogeneity is independent from the cell cycle, warranting further investigation. Finally, the integration of scRNA-Seq data with imaging-based proteomics information reveals that only a small portion of CCD protein is regulated transcriptionally. This highlights the need for multi-omic studies to truly understand the causes and origins of cellular heterogeneity.
Paper II: Deciphering hierarchical cell cycle controls by near-saturation phosphoproteomics

In Paper I, we show that the majority of CCD proteins are regulated post transcriptionally. As phosphorylation plays a major part in cell cycle regulation, for example through the regulation of cell cycle checkpoints proteins, we hypothesized that many of the observed protein expression variations are explained by changes in their phosphorylation profile. To study the dynamics of the phosphoproteome during the cell cycle we combined fluorescence activated cell sorting with MS proteomics. We used the same cell line as in Paper I, which allowed us to connect the phosphoproteomics data with the imaging-based spatial proteomics data from Paper I.

Figure 11 | Workflow for Paper II - Cell cycle resolved phosphoproteomics screen
FUCCI cells are sorted into different cell cycle phases based on GMNN and CDT1 expression. Following TMT labeling mass spectrometry allows for the identification of phosphosites. Their abundance changes can then be correlated to cell cycle progression.

In a single experiment, we identified over 77,000 phosphosites, nearly as many as the 90,000 observed in a metaanalysis of many different human sample types over >500 days of instrument time.268 Notably the unprecedented depth of our data indicates a >10-fold increase compared to previous phosphoproteomics studies of the cell cycle.74,75,269,270 We quantified phosphopeptides during G1, S and G2 and identified 4860 CCD phosphosites corresponding to 2019 proteins, among them many unknown cell cycle proteins. We show that phosphoproteins form an extended network around the CCD proteins identified in Paper I. Furthermore, we provide evidence for PTM-mediated regulation of protein expression and explain CCD protein translocations observed in Paper I. For instance, UGDH translocates from the cytosol to the nucleus in G1 following a dephosphorylation of S88. Interestingly, the vast majority of phosphosites are present in disordered regions, indicating a high probability of conformational changes associated with the phosphorylation. Notably, this could explain the translocation of...
highly disordered nucleolar proteins to the mitotic chromosomes following hyperphosphorylation during G2. In conclusion, this modified phosphoproteomics approach enabled an unprecedented depth of analysis of the CCD phosphoproteome. Due to the depth of our dataset, its applications extend far beyond the cell cycle but will enable the systematic identification of functional phosphosites.

Paper III: Widespread enzyme expression variations underlie diverse metabolic capacities within cell types

In Paper I we revealed that the majority of protein expression variability between single cells is independent of the cell cycle. Several discoveries in that study pointed towards metabolism as a driver for non CCD heterogeneity. Therefore, we wanted to systematically characterize the intra- and intercellular organization of the metabolic proteome. The HPA database contains subcellular localization data for 2,126 out of 3,069 genes included in the Human1 database\textsuperscript{271}, which we used to define the metabolic proteome. First, we characterized the subcellular organization of the metabolic proteome and revealed distinct spatial partitioning of metabolic pathways, which we linked to metabolic compartmentalization. Next, we aimed to profile the heterogeneity of the metabolic proteome. Notably, 37.4\% of enzymes display cell-cell heterogeneity opposed to 22.4\% for the entire proteome (p=4.7e-47, binomial test) (Fig 12A). To validate that this widespread heterogeneity is not a result of potential genetic differences in the cancer cell lines used in the HPA, we performed a clonal expansion experiment of mNG-tagged Hek293T cells (Fig 12B). Finally, to emphasize the possibilities of studying metabolic heterogeneity in cell lines, we compared our data to single-cell MS proteomics data from liver hepatocytes\textsuperscript{272} which are known to exhibit a variety of metabolic states along the lobular axis between portal and central veins, commonly referred to as liver zonation. Notably, a significant increase of enzymes with variable expression in hepatocytes display variable expression in cancer cells as well (n=137, 47.9\%, p=0.00034, binomial test) highlighting cell line systems as a suitable model to study metabolic heterogeneity. Finally, we confirm that the majority of enzymes display cellular heterogeneity found in paper I is independent of cell cycle progression and primarily regulated on the post-transcriptional level. Therefore, we conclude that enzyme expression variability can establish functional metabolic states. We highlight that different layers of heterogeneity can act in parallel. For example; HMGCS1 is an enzyme in cholesterol biosynthesis that displays non-CCD expression heterogeneity along with a CCD translocation from the cytosol to the nucleus in early G1. This translocation turns HMGCS1 into a multilocalizing protein. In fact, over half of all metabolic enzymes localize to multiple compartments introducing an additional layer of spatial complexity along with the temporal complexity established by single-cell heterogeneity. We define multilocalizing enzymes as potentially multifunctional. By integrating public protein-protein interaction data with subcellular localization information from the HPA we revealed potential novel nuclear functions for multilocalizing proteins (Fig 12C).
In conclusion, this map of cellular metabolism highlights the complexity of the metabolic landscape both within and between cells. It is an important step towards understanding the implications of heterogeneity in health and disease. Ultimately, this could allow the targeted manipulation of metabolic states as a novel treatment strategy to overcome drug resistance in diseases such as cancer.

Figure 12 | Key findings from Paper III

A) Metabolic enzymes exhibit a higher degree of heterogeneity compared to the background of the human proteome. B) Clonal expansion experiment. A single cell can give rise to heterogeneous populations (blue = nucleus, grey = ENO1-mNG; scalebar corresponds to 10 µm) C) P4HA is a novel multifunctional protein. It interacts with proteins involved in collagen synthesis in the cytoplasm, and proteins involved in mRNA processing in the nucleus.

Paper IV: An image-based map of the mitochondrial proteome reveals widespread metabolic heterogeneity

Mitochondria are important organelles with a variety of important cellular functions such as the generation of ATP through oxidative phosphorylation, signaling or apoptosis. Notably, mitochondrial dysregulation and heterogeneity has implications in a myriad of diseases. However, information about proteomic heterogeneity is still lacking. The mitochondrial proteome has been mapped extensively using MS-based subcellular proteomics methods. Here, we provide the first mitochondrial map with single-cell resolution allowing for systematic study of mitochondrial protein expression heterogeneity. We identify over 500 novel mitochondrial proteins compared to MitoCarta3.0 which is referred to as the most comprehensive database for mitochondrial proteins.
By analyzing bulk RNA Sequencing data across human tissues, we confirm that mitochondria-associated gene expression generally scales with the energy demand of a tissue. However, various tissues differ in their pathway activity indicating that mitochondrial expression differences are integral for tissue identity. Thus, we hypothesize that mitochondrial expression heterogeneity in homogeneous cell populations can also establish multiple functional metabolic states. Additionally, mitochondrial enzymes displayed a particularly high degree of cellular variability in Paper III, and displayed disproportionately high non-CCD expression heterogeneity in Paper I. Therefore, we set up a large imaging-based spatial proteomics screen to analyze the co-expression of hundreds of variably expressed mitochondrial proteins with IDH3A and SOD2. IDH3A is a rate limiting enzyme in the TCA cycle while SOD2 is integral for mitochondrial ROS metabolism. We identify unique subsets of proteins correlating with SOD2 and IDH3A levels and demonstrate that mitochondrial heterogeneity indeed establishes metabolic states.
Chapter 5 - Concluding remarks

When I embarked on this PhD, I had an interest in the visual aspect of cell biology. During my work experience prior to the PhD I learned to appreciate how fluorescence microscopy allowed me to “see” textbook biology that I learned to love during my highschool days. I remember being amazed by the striking beauty of a mitotic cell with its condensed chromosomes being pulled apart by the spindle. I remember the beauty of a Golgi stack and the net-like structure of the ER. I remember the amazement of being able to see mitochondrial DNA after staining with DAPI and increasing the image signal. At the same time, I remember the confusion when I realized that mitochondria don’t exist as “blobs” in the cell but form complex networks of many individual mitochondria. 5 years later, this fascination has not changed. And why would it? Microscopy has been a cornerstone of cell biology since Leeuwenhook described the first cellular structures using the first ever microscope and it has been a cornerstone of my thesis project. At the start of my PhD, I was confident that we would solve some of the riddles of life, 5 years later I am looking at a few answers and a plethora of new questions. My interpretation of cell biology is very far from the simplified textbook idea. Instead, I regard a cell as a tiny universe, full of wonders and unanswered questions. And I believe we are only at the tip of an iceberg of understanding cell biology. To say it with the words of my supervisors: “Cells are not done surprising us”\(^{277}\) (Manuel Leonetti) and “we won’t be able to fully understand, predict, repair or control cell function until we have determined the multi-scale proteome architecture of cells and its dynamic rewiring”\(^{278}\) (Emma Lundberg).

At the same time, there has probably never been a more exciting time to be a PhD student studying cells with single-cell resolution due to the astonishing speed of method development and technological advancements in the field. During the last 5 years, I have studied cellular heterogeneity on the protein level with a focus on cell cycle and metabolism mostly through the objective of a microscope. Yet, I learned very quickly that cellular heterogeneity spans multiple levels of regulation, so studying heterogeneity from only one angle is incredibly difficult. In my studies, we integrated (phospho)proteomics and transcriptomics over time along the cell cycle axis. We integrated public interactomics data with subcellular protein location information to reveal potentially novel, context-specific functions of metabolic enzymes. Finally, we correlated the expression of hundreds of variably expressed mitochondrial proteins to rate limiting enzymes in important mitochondrial pathways. Altogether, this thesis provides the first proteomic dissection of the mammalian cell cycle with single cell resolution with novel insights about the transcriptional regulation of the cell cycle. Furthermore, we produced a phosphoproteomics dataset with unprecedented resolution to characterize the CCD proteome in more detail. Finally, we provide a proteome-wide map of metabolic heterogeneity which can establish metabolic states.
Nevertheless, after those 5 years studying cell cycle and metabolism on the proteome and transcriptome level, a lot of open questions remain. In the final chapter of this thesis I want to provide thoughts on two of them:

i) How is proteomic heterogeneity manifested and how is it regulated? Does it arise stochastically or is it linked to differences in cell states or morphology?

ii) What are the functional consequences of protein expression heterogeneity?

As shown in Paper I, to answer the question of the regulation of cellular heterogeneity, we need single cell measurements across different omics modalities. Single-cell sequencing is now a well established method and can even be applied across several omics layers simultaneously, for example for profiling chromatin accessibility together with the transcriptome. As mentioned earlier, MS based proteomics has entered the single-cell era as well and is already used for systematic analysis of cellular heterogeneity. Only a few years ago, Matthias Mann stated that it might become possible to handle and profile single-cells for proteomics analysis at some point in time but “not in my career”. Now he is at the forefront of technology development for single-cell MS. As a result, over the next few years, with further progress in sample handling and MS sensitivity, it might become feasible to perform top-down proteomics or MS-based subcellular proteomics with single-cell resolution. In fact, a proof of concept study has already revealed proteoform variability for dozens of proteins in skeletal muscle cells and around 2000 PTMs were identified in single cells in a recent study by Orsburn et al. Lastly, recent developments in the metabolomics field now allow for the quantification of metabolite and lipid levels in single cells using matrix-assisted laser desorption ionization (MALDI) or MS and have revealed substantial cellular variability in metabolite composition. While the metabolome has not been a part of this thesis, it will however be integral for truly understanding and interpreting metabolic states. Altogether, non-genetic heterogeneity is established across all omics layers, from the transcriptome to the proteome and finally the metabolome.

Now that we have single-cell technologies across all omics layers available, we can combine the pieces to solve the puzzle of cellular heterogeneity. Data modalities can be integrated by collecting them from the same cells or by aligning all omics measurement along the same axis of heterogeneity; e.g. cell cycle progression or the expression of rate limiting enzymes, as done in this thesis. In my opinion imaging provides the perfect modality to combine multiple omics approaches. Not only are we able to study cells in their native environment and therefore account for heterogeneity in a cell’s size, shape or microenvironment as well as their transcriptome, proteome and metabolome, but imaging-based methods are available for all mentioned omics modalities.

i) Transcript abundances can for example be measured using in situ Sequencing or Fluorescence In Situ Hybridization (FISH) methods. Due to the spectral overlap of the fluorophores it is not possible to visualize more than 4-8 proteins in the same sample using fluorescence microscopy. To overcome this limitation, a myriad of multiplexing imaging technologies have been developed enabling the staining of up to 100 proteins in the same sample. So far multiplexing microscopy has mostly been applied to the profiling of cell type composition in tissue environments. However, some studies have now focussed on
studying diverse cellular states in a cell line population making multiplexed microscopy a valuable tool for the study of cellular heterogeneity.\textsuperscript{131,264,299} Furthermore, the rapid improvement in single-cell MS methods can be combined with (multiplexed) imaging using Deep Visual Proteomics.\textsuperscript{300} Here, a sample is imaged and cell states of interest can be identified in the analysis. Afterwards, individual cells can be collected from the tissue or cell line population by laser microdissection and their proteome correlated to the features collected in the analysis. This approach has been successfully used to profile the proteomes of individual hepatocytes revealing metabolic zonation in the liver.\textsuperscript{272} iii) Lastly, metabolites can be studied in their native environment using MALDI-MS.

The integration of different omics modalities has started, for example with the integration of MALDI and transcriptomics\textsuperscript{301} and I am beyond excited for further developments in the field. Ultimately, the integration of various omics measurements combined with deep learning approaches for whole-cell modeling could provide us with answers towards the origin of heterogeneity and how much of it is regulated on the transcriptional, translational, post-translational or metabolite level. Many findings in this thesis point towards post-transcriptional regulation of protein expression and to protein location as an independent regulator of protein function. I am excited about the possibilities of compartment resolved phosphoproteomics to identify whether protein relocalization is caused by modifications to the peptide chain. The technical adjustments applied to the phosphoproteomics experiment in Paper II enabled a depth of analysis that was previously only possible through metaanalysis of dozens of phosphoproteomic studies across different human sample types. The applications of our dataset therefore extend far beyond the analysis of the CCD phosphoproteome. Instead, it can be used as a resource to identify functional phosphosites. Notably, most phosphosites are found in disordered regions of proteins. Intriguingly, phosphorylations in disordered regions are likely to induce conformational and thus functional changes of the proteins.\textsuperscript{302} Large scale structural modeling using Alphafold\textsuperscript{303} could reveal functionally relevant sites in an unperturbed cell culture system and therefore open up new possibilities for drug development.

However, as highlighted in earlier parts in the thesis, it is crucial to align heterogeneity measurements to an axis of heterogeneity. I am personally very excited about the role of metabolic states in drug resistance and infection. By testing how metabolic states change over time and during or after perturbations, we will shed light on the functional impacts of heterogeneity. I am particularly intrigued by following the cellular phenotype over time using timelapse microscopy of fluorescently tagged cell lines. This could be particularly powerful in CRISPR pooled libraries where the location dynamics of hundreds of enzymes can be studied simultaneously.\textsuperscript{245} Following a drug challenge the heterogeneous response of the cells can be analyzed and coupled to its history. This approach could be combined with live-Seq, a scRNA-Sequencing method that allows for sampling of a cell’s transcriptome without cell lysis.\textsuperscript{304} I wrote this paragraph specifically with clinical applications of infection and drug resistance in mind, however similar experiments could be designed to pinpoint the role of heterogeneity in aging, development or immune cell activation.
Finally, the work in this thesis has focused on cell lines. Heterogeneity in cell lines is already very complex but the complexity only increases in tissues due to the interaction with other cell types or the differences in nutrient availability. Nevertheless, plenty of examples for non-genetic heterogeneity have been already described in situ, from metabolic zonation in the liver\textsuperscript{305,306} to differences in the activation potential of immune cells.\textsuperscript{37} Those examples highlight cell lines as a suitable model for studying cellular heterogeneity.

Altogether, the integrated application of single-cell omics technologies over time and during perturbations will provide a better understanding of how cellular heterogeneity is established and how it fluctuates. This understanding will also unravel its roles in a myriad of diseases such as cancer and infectious diseases. Ultimately, the manipulation of cellular heterogeneity and forcing cells into a specific cellular state has the potential to serve as a novel treatment strategy in those diseases.
References

24. Huang KY, Lee TY, Kao HJ, et al. dbPTM in 2019: exploring disease association and cross-talk of post-


54. Lane N, Martin WF. Eukaryotes really are special, and mitochondria are why. *Proc Natl Acad Sci*. 2015;112(35):E4823-E4823. doi:10.1073/pnas.1509237112


63. Seok SH. Structural Insights into Protein Regulation by Phosphorylation and Substrate Recognize of Protein Kinases/Phosphatases. *Life*. 2021;11(9):957. doi:10.3390/life11090957


69. Ly T, Whigham A, Clarke R, et al. Proteomic analysis of cell cycle progression in asynchronous cultures,
90. Müller L, Gutschner T, Hatzfeld M. Going only half the way: cell cycle exit after the G1 restriction point. Signal Transduct Target Ther. 2023;8(1):1-2. doi:10.1038/s41392-023-01692-1


References


References


179. Zhu


183. Franca GS, Baron M, Pour M, et al. Drug-induced adaptation along a resistance continuum in cancer cells. Published online June 25, 2022;2022.06.21.496830. doi:10.1101/2022.06.21.496830


References


227. Bra...


263. Chandrasekaran SN, Ackerman J, Alix E, et al. JUMP Cell Painting dataset: morphological impact of 136,000 chemical and genetic perturbations. Published online March 24, 2023;2023.03.23.534023. doi:10.1101/2023.03.23.534023


References

doi:10.1126/scisignal.aaz1482
286. Sliuzak G. Subcellular quantitative imaging of metabolites at the organelle level. Nat Metab. 2023;5(9):1446-1448. doi:10.1038/s44155-022-00882-z
References


List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D</td>
<td>Three-Dimensional</td>
</tr>
<tr>
<td>A</td>
<td>Adenine</td>
</tr>
<tr>
<td>AI</td>
<td>Artificial intelligence</td>
</tr>
<tr>
<td>APEX</td>
<td>Ascorbate Peroxidase</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine Triphosphate</td>
</tr>
<tr>
<td>BiolID</td>
<td>proximity-dependent biotin identification</td>
</tr>
<tr>
<td>C</td>
<td>Cytosine</td>
</tr>
<tr>
<td>CCD</td>
<td>cell cycle dependent</td>
</tr>
<tr>
<td>CCNB1</td>
<td>Cyclin B1</td>
</tr>
<tr>
<td>CD2BP2</td>
<td>CD2 antigen cytoplasmic tail-binding protein 2</td>
</tr>
<tr>
<td>CDK</td>
<td>Cyclin-Dependent Kinase</td>
</tr>
<tr>
<td>CDT1</td>
<td>Chromatin Licensing and DNA Replication Factor 1</td>
</tr>
<tr>
<td>CRISPR/Cas9</td>
<td>Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated protein 9</td>
</tr>
<tr>
<td>DAPI</td>
<td>4',6-Diamidino-2-Phenylindole</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>DUSP18</td>
<td>Dual specificity protein phosphatase 18</td>
</tr>
<tr>
<td>ER</td>
<td>Endoplastic Reticulum</td>
</tr>
<tr>
<td>FISH</td>
<td>Fluorescence In Situ Hybridization</td>
</tr>
<tr>
<td>G</td>
<td>Guanine</td>
</tr>
<tr>
<td>G0</td>
<td>Gap 0 (Cell cycle phase)</td>
</tr>
<tr>
<td>G1</td>
<td>Gap 1 (Cell cycle phase)</td>
</tr>
<tr>
<td>GAPDH</td>
<td>Glyceraldehyde-3-Phosphate Dehydrogenase</td>
</tr>
<tr>
<td>GFP</td>
<td>Green Fluorescent Protein</td>
</tr>
<tr>
<td>GMNN</td>
<td>Geminin</td>
</tr>
<tr>
<td>HMGCS1</td>
<td>3-Hydroxy-3-Methylglutaryl-CoA Synthase 1</td>
</tr>
<tr>
<td>HPA</td>
<td>Human Protein Atlas</td>
</tr>
<tr>
<td>IDH3A</td>
<td>Isocitrate Dehydrogenase (NAD(+)) 3 Catalytic Subunit Alpha</td>
</tr>
<tr>
<td>IDR</td>
<td>intrinsically disordered regions</td>
</tr>
<tr>
<td>Ig</td>
<td>Immunoglobulin</td>
</tr>
</tbody>
</table>
List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>JPH3</td>
<td>Junctophilin 3</td>
</tr>
<tr>
<td>KLHL38</td>
<td>Kelch Like Family Member 38</td>
</tr>
<tr>
<td>lncRNA</td>
<td>Long non-coding RNA</td>
</tr>
<tr>
<td>M</td>
<td>Mitosis (Cell cycle phase)</td>
</tr>
<tr>
<td>m/z</td>
<td>Mass-to-Charge Ratio</td>
</tr>
<tr>
<td>mAG</td>
<td>Monomeric Azami-Green</td>
</tr>
<tr>
<td>MALDI</td>
<td>Matrix Assisted Laser Desorption/Ionization</td>
</tr>
<tr>
<td>miRNA</td>
<td>MicroRNA</td>
</tr>
<tr>
<td>mKO2</td>
<td>Monomeric Kusabira-Orange 2</td>
</tr>
<tr>
<td>mNG</td>
<td>mNeonGreen</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>MS</td>
<td>Mass Spectrometry</td>
</tr>
<tr>
<td>mtDNA</td>
<td>Mitochondrial DNA</td>
</tr>
<tr>
<td>PrEST</td>
<td>Protein Epitope Signature Tags</td>
</tr>
<tr>
<td>PTM</td>
<td>Post-translational Modification</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic Acid</td>
</tr>
<tr>
<td>ROS</td>
<td>Reactive Oxygen Species</td>
</tr>
<tr>
<td>rRNA</td>
<td>Ribosomal RNA</td>
</tr>
<tr>
<td>S</td>
<td>Synthesis (Cell cycle phase)</td>
</tr>
<tr>
<td>scRNA Seq</td>
<td>Single-Cell RNA Sequencing</td>
</tr>
<tr>
<td>SOD2</td>
<td>Superoxid Dismutase 2</td>
</tr>
<tr>
<td>SOX12</td>
<td>SRY-Box Transcription Factor 12</td>
</tr>
<tr>
<td>T</td>
<td>Thymine</td>
</tr>
<tr>
<td>TCA</td>
<td>Tricarboxylic Acid Cycle</td>
</tr>
<tr>
<td>tRNA</td>
<td>Transfer RNA</td>
</tr>
<tr>
<td>U</td>
<td>Uracile</td>
</tr>
<tr>
<td>UGDH</td>
<td>UDP-Glucose 6-Dehydrogenase</td>
</tr>
<tr>
<td>UMAP</td>
<td>Uniform Manifold Approximation and Projection</td>
</tr>
</tbody>
</table>