



UNIVERSITÀ DI PARMA

UNIVERSITA' DEGLI STUDI DI PARMA

DOTTORATO DI RICERCA IN
“Biotecnologie e Bioscienze”

CICLO XXXVIII

Overcoming the challenges in sample multiplexing for next generation single cell proteomics

Coordinatore:

Chiar.ma Prof.ssa Elena Maestri

Tutore:

Chiar.ma Prof.ssa Barbara Montanini

Dottoranda: Giulia Barotti

Anni Accademici 2022/2023-2024/2025

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ABSTRACT

CHAPTER 1: INTRODUCTION

- 1.1 Background and significance
- 1.2 Established single-cell technologies
- 1.3 Proteomics and mass spectrometry
- 1.4 Quantitative methods in proteomics
- 1.5 Single-cell proteomics
- 1.6 Single-cell heterogeneity in aging
- 1.7 Histone proteomics

CHAPTER 2: A NEW METHOD TO MULTIPLEX SINGLE CELLS FOR THE ANALYSIS OF HISTONE POST-TRANSLATIONAL MODIFICATIONS

CHAPTER 3: HYPERPLEXING: ACHIEVEMENTS AND CHALLENGES

- 3.1 Introduction: multiplexing labeling strategies background and aim of the project
- 3.2 Feasibility tests to combine two labeling strategies with different chemistry
- 3.3 Key parameters for developing a dual-labeling protocol
- 3.4 First single-cell experiment to test the protocol in yielding meaningful biological results

CHAPTER 4: SINGLE-CELL PROTEOMICS IN A MODEL OF EXTENDED HEALTHSPAN

- 4.1 Introduction
- 4.2 Methods
- 4.3 Results
- 4.4 Conclusions

CHAPTER 5: CONCLUSIONS AND PERSPECTIVES

- 5.1 Methodological lessons from single-cell proteomics
- 5.2 Current limitations and unresolved challenges
- 5.3 Future perspectives, overall contributions and outlook

REFERENCES

LIST OF ABBREVIATIONS

MS – Mass Spectrometry

ESI – Electrospray Ionization

MALDI – Matrix-Assisted Laser Desorption/Ionization

TOF – Time-of-Flight

FTICR – Fourier Transform Ion Cyclotron Resonance

IMS – Ion Mobility Spectrometry

DTIMS – Drift Tube Ion Mobility Spectrometry

TWIMS – Traveling Wave Ion Mobility Spectrometry

TIMS – Trapped Ion Mobility Spectrometry

FAIMS – High-Field Asymmetric Waveform Ion Mobility Spectrometry

PASEF – Parallel Accumulation–Serial Fragmentation

DDA – Data-Dependent Acquisition

DIA – Data-Independent Acquisition

LC – Liquid Chromatography

LFQ – Label-Free Quantification

hPTMs – Histone Post-Translational Modifications

MPA / MPB – Mobile Phase A / Mobile Phase B

HCD – Higher-Energy Collisional Dissociation

FT-ICR – Fourier Transform Ion Cyclotron Resonance (reappears in top-down MS)

ECD – Electron Capture Dissociation

ETD – Electron Transfer Dissociation

SCP – Single-Cell Proteomics

FDR – False Discovery Rate

TCA – Tricarboxylic Acid cycle

AGC – Automatic Gain Control

CV – Compensation Voltage

FAIMS–MS/MS – High-Field Asymmetric Ion Mobility Spectrometry coupled to Tandem MS

ABSTRACT

Single-cell proteomics has emerged in the past five years as a powerful approach to investigate the heterogeneity of proteomes within cell populations. Understanding how protein expression and modification vary from cell to cell is essential, as population averages can mask critical regulatory mechanisms and rare cell states. This question is particularly relevant in the context of aging, where tissues progressively lose cellular identity and individual cells accumulate distinct DNA mutations and proteomic profiles, potentially contributing to functional decline.

Among the molecular mechanisms that shape proteome regulation, chromatin plays a central role by modulating gene accessibility and expression through histone post-translational modifications (hPTMs). Therefore, studying chromatin-associated proteins at single-cell resolution can provide key insights into how epigenetic diversity influences cellular function and identity.

Despite recent advances, single-cell proteomics still faces important challenges in sensitivity and throughput. Unlike sequencing-based methods, proteins and peptides cannot be amplified or easily barcoded, making efficient sample processing and precise quantification.

During my PhD, I addressed these limitations through three complementary projects. First, I developed and optimized a workflow for single-cell analysis of histone post-translational modifications, enabling the study of chromatin regulation at unprecedented depth. Second, I refined a series of chemical labeling strategies to tag peptides site-specifically and multiplex samples, increasing the throughput of single-cell proteomics. Finally, I applied the optimized workflow to a mouse model of extended healthspan, using single-cell proteomics to quantify how the liver proteome and its heterogeneity are regulated during aging by metabolic intervention.

Together, these studies contribute methodological and conceptual advances toward a more comprehensive understanding of proteome regulation at the single-cell level and open new directions for applying proteomics to complex biological questions such as aging and epigenetic control.

CHAPTER 1: INTRODUCTION

1.1 Background and significance

Heterogeneity is a basic feature of life. It is now well known, both theoretically and experimentally, that almost all cellular systems are heterogeneous. Heterogeneity in nature manifests itself at different levels and is fundamental to the survival and functionality of living organisms. One example is the bet-hedging survival strategy in which a population produces offspring with different phenotypes, rather than just one optimal phenotype, in order to reduce temporal variation in adaptation. If environmental conditions change drastically, organisms that were originally less developed can adapt to the new environment, ensuring the survival of the population. Therefore, we can say that cellular systems with greater diversity have greater fitness (1).

Heterogeneity begins from the moment of conception: a single fertilized egg gives rise to all the cell types present in the human body. Despite containing the same genetic information, each cell in adult bodies is unique and exhibits considerable variability in phenotype compared to other cells (2). Currently, a fundamental challenge in biology is to understand how such cellular diversity is generated from a single cell, how it is regulated, and how it is exploited to implement appropriate responses to external perturbations in normal and diseased tissues (2). There are many ways to quantify heterogeneity, including genomic, epigenomic, transcriptomic, proteomic and metabolomics approaches (Figure 1). Importantly, variation seen at one layer may not align with the variation observed at another, reflecting the complexity of cellular regulation (1).

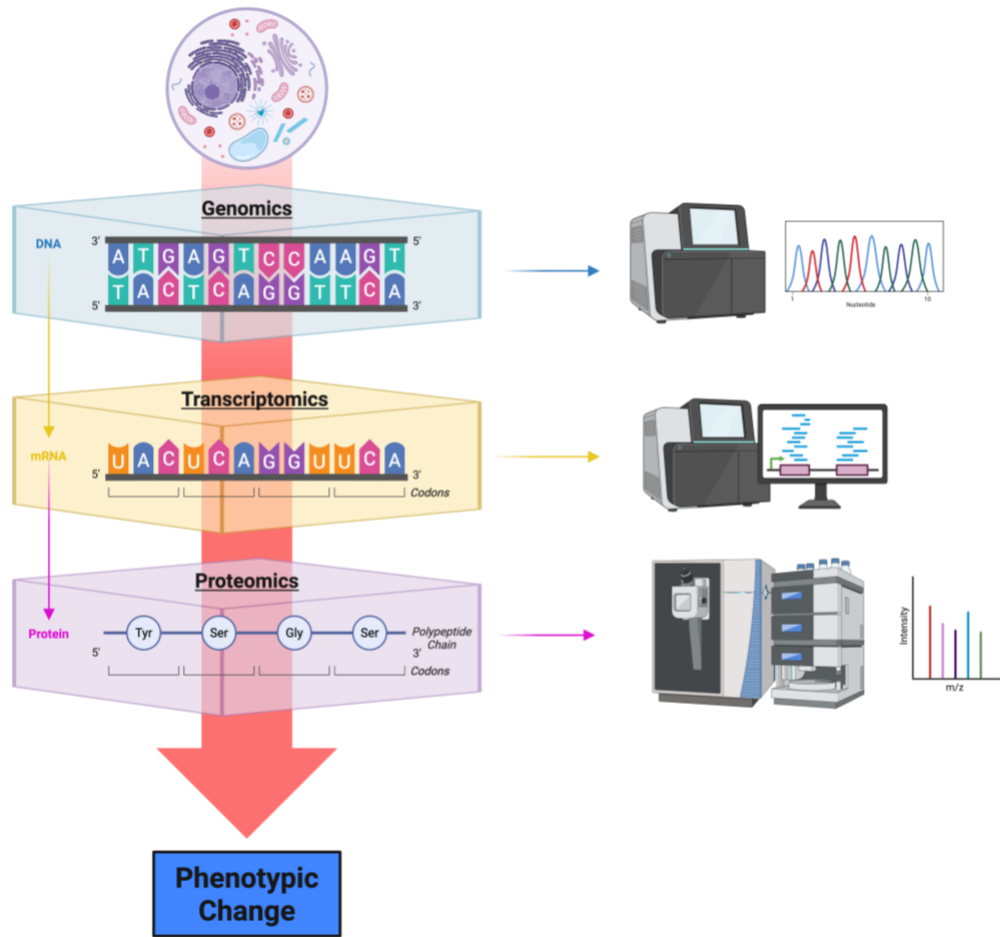


Figure 1: An overview of omics layers showing genomics, transcriptomics, and proteomics.

There are various causes of cellular heterogeneity. First, some cells within a population naturally contain mutations. If germline mutations provide a selective advantage, the cells that carry them can proliferate and become a significant portion of the population (1). Somatic mutations usually remain confined in selected cells, but they can also confer proliferative advantages. However, genetic differences explain only part of cellular heterogeneity. Many differences are not coded in the DNA sequence, and they are manifested as cell-to-cell variability in e.g. transcriptomes. These can derive from both intrinsic processes and extrinsic influences. Extrinsic signals, in particular, can trigger phenotypic plasticity in response to environmental changes, but they act only on the subset of cells that are exposed to those conditions (3). In the extrinsic influence, we can also include the different phases of the cell-cycle or the size of the cell (4). Even in a simple Petri dish culture, where cells are genetically identical and grown under the same conditions, populations

will contain a mixture of cells at different stages of the cycle or with different sizes. These differences influence transcriptional activity, metabolic demands, and protein content, and are therefore an immediate and visible source of heterogeneity (5). Intrinsic heterogeneity can be the result of stochastic events such as gene expression noise, and for that reason can be more a subtle phenomenon (3). Stochastic gene expression accounts for part of heterogeneity. Epigenetic regulation, alternative splicing and post-translational modifications can also impact the state of the genome or a cell, showing an effect on the heterogeneity more directed and less stochastic (6).

Also, the process of development is driven by non-genetic heterogeneity. When a small population of cell beginning to differentiate, the cells are considered identical, actually in this phase the cells display a non-genetic heterogeneity (1). Nonetheless, differences in transcriptional states can be heritable through cell divisions. Transcriptional diversity is due to signaling factors, developmental regulators and chromatin regulators (7). Development is driven by mechanisms that are involved in directed heterogeneity.

Following development and differentiation comes another level of heterogeneity: the tissues. A very interesting example is that of the endothelial tissue. Starting from a homogeneous population, endothelial cells differentiate into diverse subtypes, creating heterogeneity both within individual organs (intra-organ heterogeneity) and across different organs (inter-organ heterogeneity). Capillary endothelial cells were once considered a uniform vascular population, but recent evidence shows their phenotypic and structural features vary widely depending on the organ or tissue microenvironment (8). The immune system is another clear example of tissue heterogeneity. Not only there are different cell types such as monocytes, macrophages, B cells and T cells, but there is also fundamental diversity within the T cell subpopulation at the level of T cell receptors. This diversity allows for an effective immune response (9).

Heterogeneity in healthy cellular systems is a strength that promotes survival, adaptability and proper function. But in disease, the same variability often drives to dysfunction and pathology. One of the most straightforward examples of heterogeneity in disease is certainly cancer. Tumor diversity originates from the accumulation of mutations in cells in clonal expansion. These cells encounter the microenvironment surrounding them, known as the tumor microenvironment, and the combination of these two components, genetics and epigenetics, leads to the development of

cancer (10). Cancer cell heterogeneity develops and accumulates at different levels: epigenetic, transcriptional, protein and environmental. This results in intratumoral and intertumoral diversity. The first one includes copy number variation and epigenetic alterations. The second one manifests itself more at the phenotypic level, with differences even between patients, and is more dependent on the environment (11). Furthermore, tumor heterogeneity is known to facilitate adaptation to the environment, conferring resistance to therapeutic treatments. Therefore, tumor cell diversity confers plasticity, increasing tumor survival and making it difficult to eliminate the tumor both by the immune system and by drugs (10).

Heterogeneity is a feature that is also present in infections. Studies using single-cell RNA sequencing on Influenza infection in vivo showed that infected cells within the same tissue displayed variability in gene expression as well as the activation of host immune responses. After infection, some cells produced high levels of viral transcripts, others expressed very little, but only a small portion presented strong interferon responses. Therefore, also the infection consists of different cellular states. Some cells contribute significantly to viral replication, while others amplify the signal or produce a response (12). HIV infection provides another example of heterogeneity in infections. Single cell transcriptional profiling has shown that even in a population of clonal cells, they do not behave in the same way following infection. Some cells have been identified that support productive infection by presenting high levels of viral transcripts, while others remain latent, with the viral genome remaining in a silenced transcriptional state (13). This variability in the viral infections is critical because explains how the infections spread, how the immune system reacts and how the disease progresses and consequently the challenge of eradicating these diseases.

Senescence is another important process in which cellular heterogeneity is a big component and has a major impact on aging. It was erroneously believed that senescent cells were in a unique state of growth arrest, but recent single-cell studies reveal that they exhibit a wide range of phenotypes, as cells show different transcriptional states, secretory activities, and metabolic states. This diversity means that senescent cells do not contribute equally to tissue dysfunction: some remain relatively inert, while others, for example, may contribute more to the secretion of inflammatory factors that can damage local homeostasis. Therefore, it is important to define senescence as a state

of heterogeneity in order to understand how aging contributes to the decline of tissues and organs, and to enable targeted therapies (14).

The heterogeneity in aging is not only due to senescence, but a large contribution must also be attributed to transcriptional noise and variability between cells of the same type. By comparing immune system cells from older individuals with those from younger individuals, it has been shown that cells from older individuals exhibit greater cell-to-cell variability in terms of gene expression. This greater diversity leads to a less consistent and therefore dysfunctional cellular response to external stimuli (15).

Heterogeneity in aging is also present at the proteomic level. During aging, the proteome is not more balanced and functional but gradually declines. The proteostasis decline and the accumulation of misfolded and aggregated proteins does not occur at the same rate and in the same way, but manifests heterogeneously across cells and tissues (16). These multiple levels of cellular heterogeneity in aging contribute to tissue decline and increase the risk of degenerative diseases (Figure 2).

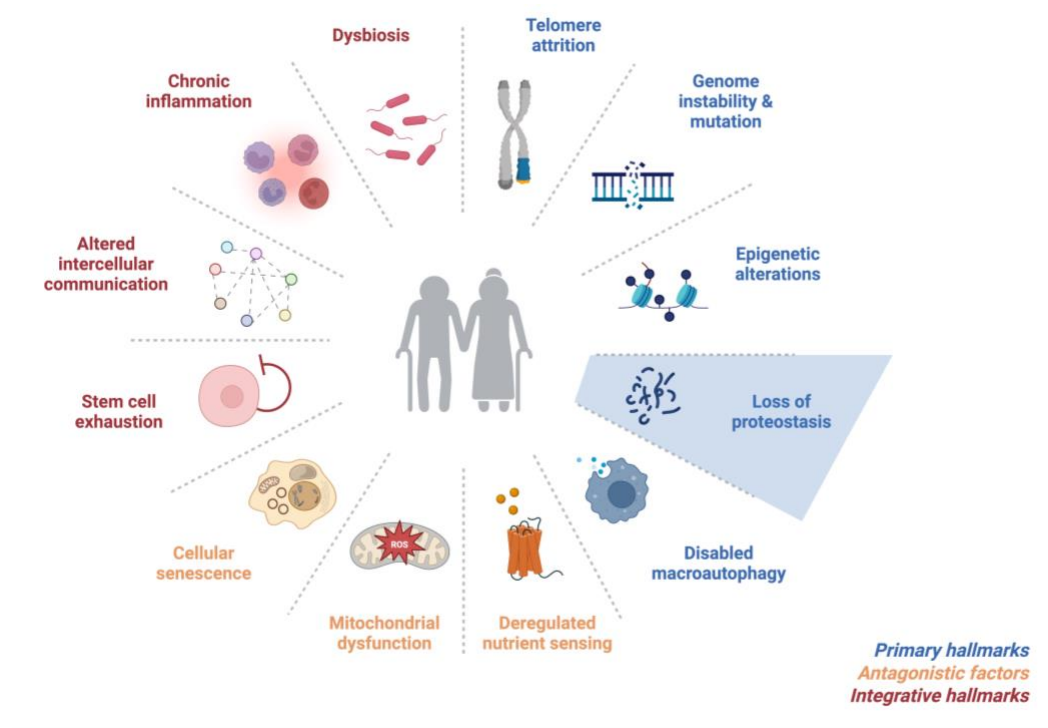


Figure 2: Loss of proteostasis, a central integrative hallmark influencing multiple aging mechanisms (adapted from (17)).

In conclusion, heterogeneity is part of the biology of cellular systems. In healthy contexts, it provides survival, adaptability, and functionality, while in disease it often drives progression, resistance, and decline. From stochastic gene expression to directed changes in chromatin, from tissue-specific diversity to diseases such as tumors, infections, and aging, variability is present at every level of life. Most of what we know about diversity comes from genomic and transcriptomic approaches. But protein abundance, modifications, and interactions ultimately shape phenotypes, is therefore fundamental to measure heterogeneity at the proteomic level with the resolution of the single cell.

1.2 Established single-cell technologies

Cells have always been considered a structural element of life and, for many years now, have also been redefined as functional units. Traditionally, cells are classified according to their morphology or the proteins they express (18,19). Attempts have always been made to capture the complexity of tissues and organs, but the techniques used until two decades ago were only capable of bulk measurements, which results in population averages (18,19). However, this masks the diversity, rare cell states, and cell-to-cell variability that is known to be fundamental. Over the last 20 years, technological advances have led to the development of single-cell technologies that allow us to capture heterogeneity. From high-multiplex cytometric and proteomic assays to advances in single-cell genomics and transcriptomics, these methods allow to resolve cellular subpopulations, detect rare cell states, define individual cells molecular profile. Each technology presents advantages and disadvantages, the challenge are the trade-offs in throughput, sensitivity, cost, and modality.

Starting with the isolation of single cells, several techniques are available, including Fluorescence-Activated Cell Sorting (FACS), Magnetic-Activated Cell Sorting (MACS), Laser Capture Microdissection (LCM), manual cell picking, and various microfluidic approaches (Table 1). Some of these methods, such as FACS, are not only used for isolation but also play a key role in measuring protein expression at the single-cell resolution. In addition, other approaches are dedicated to this purpose, such as Single-cell Western blotting (scWB), DNA reporter-based assays, and chemiluminescence or colorimetric methods (Table 2). Beyond these, mass

spectrometry-based platforms like CyTOF and Imaging Mass Cytometry further expand the possibilities for high-dimensional single-cell protein analysis (20) (Table 2).

Technique	Description	Advantages	Disadvantages	Applications
Density gradient centrifugation / filtration	Separates cells based on density or size via centrifugation or porous membranes.	Simple, label-free, inexpensive.	Low specificity; poor recovery for rare cells.	Basic enrichment of major populations.
Fluorescence-activated cell sorting (FACS)	Flow cytometry-based method using fluorescent antibodies to sort cells.	High throughput; high purity; multiparametric; live cell recovery.	Requires large input; can stress fragile cells; costly instrumentation.	Immunophenotyping, rare population isolation, downstream functional studies.
Magnetic-activated cell sorting (MACS)	Uses magnetic beads conjugated to antibodies to isolate labeled cells.	Gentle on cells; simple; cost-effective; good recovery rates.	Limited multiplexing; usually one marker per run.	Bulk enrichment or depletion of target populations.
Laser capture microdissection (LCM)	Laser-guided cutting of cells from tissue sections.	High precision; preserves spatial context.	Low throughput; expensive; can damage nucleic acids.	Tumor heterogeneity, developmental biology, neurodegeneration studies.
Manual cell picking / micromanipulation	Manual selection of single cells with micropipettes under a microscope.	Precise; versatile; direct live cell recovery.	Labor-intensive; very low throughput; variable reproducibility.	Targeted studies; clonal expansion; special cell-type isolation.
Microfluidics (Microwells)	Arrays of microwells trap single cells in parallel compartments.	Simple design; scalable; high recovery and purity.	Uneven reagent distribution; limited dynamics measurement.	Assays of secreted proteins, e.g., B cell studies.
Microfluidics (Droplets)	Encapsulates cells in nanoliter droplets within an oil phase for high-throughput isolation.	Very high throughput; minimizes cross-contamination; scalable.	Requires specialized equipment; challenging protein detection in droplets.	High-throughput sequencing, large-scale single-cell assays.

Table 1: List of techniques available for isolation of single cells.

Technique	Description	Advantages	Disadvantages	Applications
Fluorescence / image-based (Flow cytometry, FACS, imaging cytometry)	Fluorescent antibody labeling or genetic tags measured by flow or imaging cytometry.	High throughput; live cell sorting; multiparametric; mature technology.	Limited multiplexing due to spectral overlap; requires large input.	Immunophenotyping, rare population analysis, functional assays.
Single-cell Western blotting (scWB)	Miniaturized Western blot in microwells with antibody probing per cell.	Multiplex detection; resolves isoforms and PTMs; improved specificity.	Low throughput; limited sensitivity for secreted or rare proteins.	Protein isoform/PTM analysis at single-cell level.
DNA reporter-based (SPARC, CITE-seq)	Antibodies tagged with oligos measured with sequencing (SPARC, CITE-seq).	Unlimited multiplexing; integrates proteins and RNA; digital quantification.	Targeted panels only; CITE-seq mostly for surface proteins.	Multimodal single-cell omics (RNA + protein).
Chemiluminescence / colorimetric (ELISPOT, FluoroSpot)	Immunospot assays detect secreted proteins with enzyme-linked reactions.	Very sensitive; cost-effective; ideal for secreted proteins.	Low multiplexing; mainly endpoint; limited intracellular info.	Cytokine detection, immune monitoring, secretory assays.
Mass cytometry (CyTOF, IMC)	Antibodies tagged with rare metals quantified by mass spectrometry.	Very high multiplexing (>40); spatial resolution with IMC.	Lower throughput vs FACS; less sensitive for low-abundance proteins; costly.	Deep immunophenotyping, tumor microenvironment, tissue imaging.
Single-cell mass spectrometry (scMS / LC-MS proteomics)	Unbiased proteome-wide analysis using miniaturized LC-MS workflows.	Unbiased proteome coverage; discovery potential; detects PTMs.	Low throughput; sensitivity challenges; complex data analysis.	Unbiased proteome profiling, biomarker discovery, heterogeneity studies.

Table 2: List of techniques available for measuring protein expression at the single-cell resolution.

These approaches are still widely used for single-cell studies because they contribute to unraveling gene regulatory networks, cellular states, and disease mechanisms. On the other hand, they are limited by the availability of antibodies or predefined marker panels. To gain a greater understanding of cellular heterogeneity, researchers have focused on developing single-cell genomics and single-cell transcriptomics techniques.

Single-cell genomics embraces all techniques that profile the DNA of individual cells, allowing direct measurement of gene variations, copy-number alterations, structural variants, and mutational signatures at cellular resolution. Single-cell whole genome sequencing (scWGS) studies have demonstrated the ability of this technique to detect somatic mutations and clonal architecture in cancer, as well as lineage relationships in development (21).

Despite the successes, these technologies have technical limitations. A single cell contains only picograms of DNA, so uniform amplification of genetic information is necessary, but this process can introduce amplification bias, allele dropout, or uneven coverage. For this reason, strategies such as multiple displacement amplification (MDA), multiple annealing and looping-based amplification cycles (MALBAC), or hybrid methods have been developed which, although

imperfect, are able to meet experimental needs (22) (Table 3). On the computational side, calling somatic single-nucleotide variants (SNVs), copy-number variations (CNVs), and structural rearrangements from noisy amplification artifacts requires specialized algorithms and careful filtering (23).

Technique	Description	Advantages	Disadvantages	Applications
Multiple Displacement Amplification (MDA)	Isothermal whole-genome amplification using phi29 polymerase for high fidelity and long fragments.	High fidelity, long read lengths; widely used.	Amplification bias; uneven coverage; allele dropout.	Single-cell WGS; detection of SNVs and structural variants.
Multiple Annealing and Looping-Based Amplification Cycles (MALBAC)	Quasi-linear preamplification followed by exponential PCR; reduces bias compared to MDA.	Improved uniformity; lower amplification bias.	Shorter fragments; still prone to bias in GC-rich regions.	Clonal evolution studies; CNV detection.
Linear Amplification via Transposon Insertion (LIANTI)	Linear amplification using transposon-based insertion, minimizing errors and improving uniformity.	High accuracy; minimizes amplification errors.	Technically complex; lower throughput; requires specialized protocols.	High-quality SNV and CNV detection; mosaicism studies.
Microwell-based scWGS (e.g., DLP+)	High-throughput microwell-based workflow (e.g., Direct Library Preparation Plus) for scalable scWGS.	Scalable; cost-effective; suitable for large numbers of cells.	Lower coverage compared to full scWGS; specialized setup required.	Cancer genomics; large-scale clonal architecture studies.
Single-cell exome sequencing (scWES)	Captures coding regions of the genome from single cells for variant detection at exome scale.	Focus on coding variants; less sequencing required.	Misses regulatory variants and noncoding regions.	Rare disease studies; mutational landscape in coding genes.
Single-cell targeted sequencing (amplicon-based)	Targets predefined genomic regions via multiplex PCR or hybrid capture in single cells.	High sensitivity for specific variants of interest; cost-efficient.	Restricted to targeted loci; not genome-wide.	Tracking known mutations in cancer or microbial studies.
Single-cell copy number profiling (low-pass WGS)	Low-coverage sequencing of single cells to detect copy number alterations at genome-wide scale.	Robust detection of CNVs and aneuploidy; low sequencing cost.	Does not resolve SNVs; limited resolution for small variants.	Tumor heterogeneity; aneuploidy analysis; developmental genomics.

Table 3: List of techniques available for single-cell genomics.

Single-cell RNA sequencing (scRNA-seq) enables profiling the transcriptome of individual cells and allows the capture of cell-to-cell differences in gene expression, which is essential for discovering rare cell types, mapping developmental trajectories, and linking transcriptional programs to phenotypic states (18,24).

In practical terms, the scRNA-seq workflows involve several steps: 1. isolation and capture of single cells (or nuclei), 2. cell lysis and reverse transcription (RT) of RNA into complementary

DNA (cDNA), 3. barcoding and cDNA amplification, 4. library preparation and sequencing, and 5. computational processing (alignment, quantification, normalization, clustering, trajectory, etc.). In each of these steps, technical noise and biases may be introduced, as in the case of inefficient reverse transcription or amplification, which can usually lead to difficulties in identifying transcripts that are truly expressed and a distortion in the calculation of actual relative abundances. For this reason, data thus tend to be sparse and noisy, requiring specialized normalization, imputation, and batch correction methods (25).

Over the years, several protocols have been developed for scRNA-seq with the aim of optimizing certain trade-offs: between sensitivity, throughput, coverage, cost per cell, and transcript quantification accuracy. For example, plate-based protocols (like Smart-seq3) offer full-length transcript coverage and high sensitivity but lower throughput and higher cost per cell, whereas droplet-based methods (e.g. Drop-seq, 10x Genomics Chromium) scale to thousands or tens of thousands of cells but capture only the 3' or 5' ends of transcripts and miss lowly expressed genes (24) (Table 4).

In parallel, transcriptomic analyses have been expanded beyond a single focus on transcript expression, allowing the identification of secondary RNA structures and RNA-protein interactions through a sequencing-based method such as protein interaction profile sequencing (PIP-seq) using RNase footprinting followed by high-throughput sequencing (26). Although PIP-seq is not performed at a single-cell resolution, it shows how sequencing technologies can be exploited to obtain functional and regulatory information from RNA. These ideas together with the great work done by researchers in developing and optimizing protocols for scRNA-seq are inspiring in the development of single-cell proteomics and other multi-omic methods.

Technique	Description	Advantages	Disadvantages	Applications
Smart-seq / Smart-seq2 (plate based, full-length)	Cells are captured in wells; full-length cDNA generated via template-switching RT, amplified by PCR.	High sensitivity; detects splice isoforms and allelic expression; even coverage across transcripts.	Lower throughput; higher cost per cell; laborious; bias in amplification.	Deep profiling of rare cells; isoform and allelic expression studies.
Smart-seq3 (plate based, full-length)	Single-cell RNA-seq method combining template-switching reverse transcription with 5' UMIs and whole-transcript coverage.	High sensitivity; full-length transcript coverage; improved quantification accuracy with UMIs; good isoform detection.	Lower throughput; relatively high cost and hands-on time; higher mitochondrial RNA in some datasets; requires automation for best performance.	Detailed transcriptome profiling; isoform analysis; studies requiring high gene detection sensitivity rather than massive cell numbers.
Droplet-based methods (Drop-seq, inDrop, 10x Chromium)	Cells and barcoded beads encapsulated in droplets; mRNA captured with barcodes/UMIs.	Very high throughput; scalable; lower cost per cell; useful for heterogeneous tissues.	Only 3, $\bar{A} \leq 5$, $\bar{A} \leq$ end reads; lower sensitivity; high dropout rates; misses low-abundance genes.	Atlas-scale profiling; cell type discovery; clustering large populations.
Microwell / microfluidics plate-based methods	Cells partitioned into microwells; barcoded beads/oligos capture transcripts in small volumes.	Moderate throughput; low reagent use; integrates with other modalities.	Variability across wells; lower sensitivity vs plate-based full-length methods.	Medium-scale profiling; multiomic assays; targeted studies.
Single nucleus RNA-seq (snRNA-seq)	Isolates nuclei instead of whole cells; nuclear RNA sequenced using scRNA-seq protocols.	Works on frozen/fixed tissues; avoids dissociation bias; suitable for fragile tissues.	Only nuclear transcripts; lower RNA yield; may miss cytosolic RNAs and isoforms.	Brain research; archived/frozen samples; tissues difficult to dissociate.
Combinatorial indexing / split-pool methods (sci-RNA, SPLiT-seq)	Cells/nuclei barcoded in successive rounds (split-pool combinatorial indexing).	Ultra-high throughput; inexpensive for massive numbers of cells; no need for microfluidics.	Lower sensitivity; complex barcoding; risk of barcode collisions.	Large-scale mapping of heterogeneous tissues; ultra-high throughput studies.
PIP-seq (Protein Interaction Profile Sequencing)	Uses structure-specific RNases to simultaneously map RNA-protein interaction sites and infer RNA secondary structure genome-wide.	Enables parallel detection of RNA secondary structure and protein-binding regions; transcriptome-wide; works on tissues and cells; provides high-resolution interaction maps.	Laborious and time-consuming; may have biases from cross-linking and nuclease efficiency; data analysis can be computationally demanding.	Studying global RNA-protein interaction landscapes; investigating RNA structure-function relationships; analyzing post-transcriptional regulation; identifying protein-protected RNA regions across the transcriptome.
Spatial transcriptomics & in situ barcoding (Slide-seq, MERFISH, seqFISH, STARmap)	Spatially organized barcodes or in situ hybridization preserve spatial localization of transcripts.	Preserves spatial context; links gene expression to tissue architecture.	Lower transcript coverage; specialized protocols; computationally complex; lower throughput.	Tissue-level gene expression mapping; tumor microenvironment; brain architecture.

Table 4: List of techniques available for single-cell RNA sequencing.

1.3 Proteomics and mass spectrometry

Mass spectrometry (MS)-based proteomics includes a wide range of experimental approaches that answer questions about proteins, starting with the sequence or abundance levels of a protein, function, three-dimensional structure, chemical reactions, protein-protein interactions, and many others. The main technology used for these studies is the mass spectrometer which detects ions based on their mass-to-charge (m/z) ratios and signal intensities. Mass spectrometry can be applied to different scales, from the analysis of a single protein to profiling thousands simultaneously, and for samples of varying complexity (27).

Over the past 30 years, MS-based proteomics has advanced rapidly in terms of both instrumentation technology and computational analysis. Since the introduction of soft ionization

methods such as electrospray ionization (ESI) and MALDI, it has been possible to analyze peptides and proteins, leading the way for modern proteomics workflows. Subsequently, improvements have been made in all components of mass spectrometers, in particular in mass analyzers, where higher resolution, broader dynamic range, and faster scan speed allow thousands of proteins to be identified and quantified from complex mixtures with good sensitivity and reproducibility. In parallel, refinements in acquisition strategies such as data-dependent and data-independent approaches together with appropriate bioinformatics pipelines, allow more consistent identification of peptides across a wide dynamic range, including those present at low abundance(28).

Mass analyzers are the main component of a mass spectrometer. As mentioned above, separate ions according to their mass-to-charge ratio (m/z) and determine the resolution, accuracy, and speed of measurement. The first to be widely used were quadrupole, ion traps, and time-of-flight (TOF) instruments, thanks to their mass filtering, ion accumulation, and fast acquisition capabilities. More recently, Fourier transform-based instruments such as Orbitraps and FTICR have begun to be adopted, which have significantly increased resolution up to 480,000 FWHM at m/z 200 and sub-ppm accuracy but at the cost of acquisition speed. Today, hybrid technologies such as Q-TOF, Q-Orbitrap, or Tribrid instruments are being commercialized in order to optimize both sensitivity and resolution as well as acquisition speed (29).

In parallel with these developments, ion mobility spectrometry (IMS) was introduced, which involves multi-dimensional gas-phase separation, allowing ions to be distinguished not only by their m/z but also by their shape, size, and collisional cross section, which significantly improves separation capacity and dynamic range. Several IMS technologies have been developed, such as drift tube (DTIMS), traveling wave (TWIMS), cyclic IMS, trapped IMS (TIMS), and high-field asymmetric waveform ion mobility (FAIMS) (29). In proteomics, TIMS in combination with parallel accumulation–serial fragmentation (PASEF) has been particularly impactful, coupling ion mobility separation to MS/MS achieving scan rates of 120-130 Hz. FAIMS improved selectivity by filtering out chemical noise and reducing co-isolation, which is especially advantageous for low-input or single-cell applications (29).

Another important innovation is the proportion of ions that are measured compared to those entering the instrument. Before the mid-2010s, only a small fraction of ions contributed to usable spectra, limiting sensitivity. The first improvements led to instrumental designs that approached to near-complete usage of accumulated ions. Commercial mass spectrometers developed within the past 3-4 years such as the Orbitrap Astral (Thermo Scientific) or the timsTOF Ultra (Bruker) further optimize ion usage using parallelized acquisition and ion storage, minimizing idle time, ideal for single cell work. By increasing the effective duty cycle, modern instruments achieve both higher sequencing speed and sensitivity, which enables proteome profiling from small amounts of material (29).

Recently, few studies showed the progress made in analyzing sample with very low input with mass spectrometry. It was demonstrated that single-cell proteomics can now quantify 2,000 proteins from an individual HeLa cell with good reproducibility (30). Moreover, the Orbitrap-Astral system has taken this further, identifying over 7,000 proteins from 250 pg of peptides and more than 5,000 proteins from single cells running 50 samples per day (31). These examples show that now instruments combine deep coverage with high throughput in analysis from a limited amount of materials, enabling single-cell proteomics studies.

Traditionally, the acquisition strategy that are used in proteomics are data-dependent acquisition (DDA) and data-independent acquisition (DIA). The first one implies a scan that leads to the selection of the most intense precursor ions for fragmentation, producing high-quality spectra but with the risk of missing lower-abundant peptides. DIA, on the other hand, fragments all precursor ions within a defined m/z window, generating spectra that cover peptides more systematically but with a more complex data analysis. For low-input and single-cell proteomics, both approaches have been implemented to overcome their limitations. DDA when combined with multiplexing methods such as tandem mass tag (TMT) is widely used because the pooling of signals boosts peptide identification and allows the analysis of up to 32 single cells in one run (32). All the improvements made in recent years such as shorter gradients with liquid chromatography (LC), optimized ion accumulation, and multiplexed designs make both DDA and DIA suitable for single-cell proteomics, where DDA–TMT offers multiplexing efficiency, and DIA provides reproducibility and depth from very small inputs (32).

1.4 Quantitative methods in proteomics

Proteomics has become a widespread and important platform for studying both individual proteins and complex samples. Technological advances in mass spectrometry over the years have made it possible to answer more and more biological questions where the interest is not only qualitative but also quantitative. To answer the vast array of questions, there are two main quantification strategies: (a) untargeted, which we can define as a global quantification of hundreds or thousands of proteins for protein profile comparison, and (b) targeted, for one or more components, where the proteins of interest are isolated from a sample and quantified. This distinction mainly defines the purpose of the study. However, when we want to emphasize the methodology, we have another subgroup of fundamental importance, which is divided into: (a) label-free quantification, where the sample is kept in its native composition and the molecules are quantified based on the comparison of their features across independent LC-MS analysis, and (b) label-based quantification, where artificial labels are added to peptides and proteins, introducing an expected mass difference between the experimental conditions within the sample. Depending on the information obtained from these quantification methods, the abundance of a protein or peptide can be relative or absolute. In the former, a protein or proteome is compared between different samples, while in the latter, information is obtained regarding the absolute amount or the concentration of a protein within the sample (33) (Figure 3).

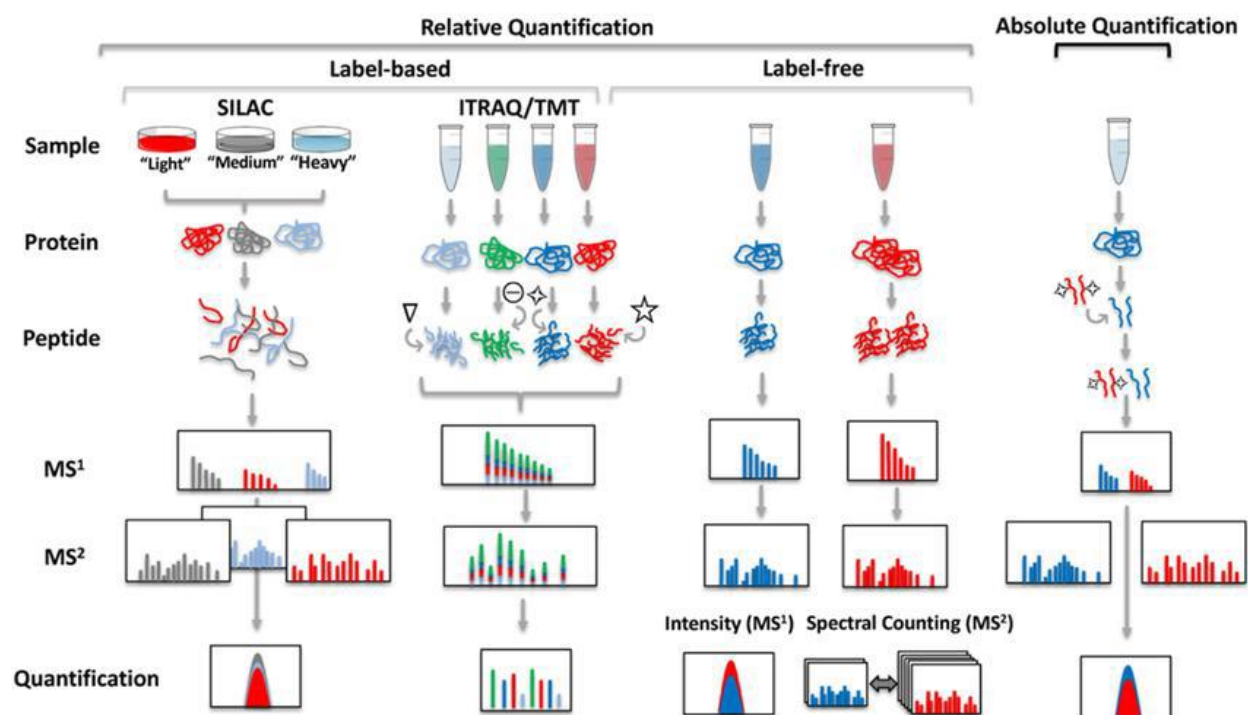


Figure 3: Quantitative proteomic strategies. Label-based methods include metabolic labeling (e.g., SILAC), where isotopic variants of amino acids are incorporated into proteins during cell growth, and isobaric tagging (e.g., TMT or iTRAQ), where peptides from different samples are chemically labeled after digestion. In SILAC experiments, quantification is based on comparing the MS¹ signal intensities of isotope-labeled peptide pairs, whereas in iTRAQ or TMT approaches, relative abundances are determined from reporter ion intensities measured in the MS² spectra. In label-free quantification, samples are analyzed separately and proteins are quantified based on ion intensity in MS¹ or spectral counting in MS². Finally, absolute quantification is achieved by adding known amounts of reference peptides or proteins, known as spike-ins, which serve as internal standards to determine the absolute concentration of the target proteins (34).

Label-free quantification (LFQ) is a widely used quantification method because it is less time-consuming during sample preparation, there is no limit to the number of experiments that can be compared, mass spectral complexity is not increased turning in more analytical depth, and there is evidence that can provide higher dynamic range of quantification than stable isotopic labeling. Traditionally, the LFQ strategies can be distinguished in: measuring and comparing the signal intensity of peptide precursor ions that belong to protein, or, counting and comparing the fragment

spectra that identify the peptides of a given protein. Label-free quantification is advantageous when there are large and global changes between conditions or for the discovery of new biomarkers but has also several limitations. LFQ methods are susceptible to run-to-run variability in sample preparation, ionization efficiency and LC-MS performance because every sample is measured separately. A robust retention time alignment and peak matching is essential for the comparison across runs but missing values are common especially in data-dependent acquisition due to the stochastic peptide selection. Moreover, the approach of spectral counting is affected by a non-linear response and early saturation, where the abundant proteins accumulate spectra disproportionately and low-abundant proteins are hardly detected. For these reasons, the label-free quantification methods are less accurate when small and meaningful differences need to be detected (35).

To achieve a good relative quantification with improved sensitivity, accuracy and reproducibility is used the samples multiplexing which involves label-based methods that comprise in vivo (metabolic) or in vitro (post-harvested) peptide or proteins labeling. Depending on the experimental needs, the tagging can be chemical, metabolic and enzymatic. Metabolic labeling introduces stable isotopes during protein synthesis in cells or organisms by supplying media enriched with labeled amino acids or nutrients such as ^{15}N or heavy-lysine/arginine in SILAC. This labeling methodology improves quantification accuracy because labels are incorporated at the early stage and samples can be mixed immediately after harvesting, but at the cost of a poor experimental variability. Enzymatic labeling introduces isotopes post-harvest during or after proteolysis, typically by digesting proteins in heavy (H_2^{18}O) or light (H_2^{16}O) water. This results in incorporation of two ^{18}O or ^{16}O atoms at the C-terminus of the generated peptides, resulting in a mass shift of 4 Da between heavy- and light-labeled peptides. This method is simple and can be applied to any biological sample, but incomplete or partial incorporation of the labels can complicate data analysis (33).

Chemical labeling methods introduce tags in vitro, after proteins or peptides are extracted from the biological sample. These reagents react with specific functional groups and create predictable mass differences that allow relative quantification between samples in a single MS run. Among these are ICAT (isotope-coded affinity tags), which targets cysteine residues but limited to proteins containing this amino acid, ICPL (isotope-coded protein labeling), which labels free amines and

thus covers nearly all peptides, though it can interfere with trypsin digestion and peptide fragmentation. Another step is the introduction of isobaric chemical tags, such as iTRAQ and TMT (tandem mass tags) which for their chemistry enable multiplexing (up to 8- or 10-plex in classical iTRAQ, and even higher with newer TMT35-plex reagents), making them particularly powerful for large comparative studies and single-cell proteomics. The main inconvenience of isobaric labeling is ratio compression, caused by co-isolated peptides that lead to the underestimation of actual protein abundance differences. This challenge can be overcome by additional fragmentation strategies such as MS3. Other reagents, such as DiLeu or IPTL, have been developed as cost-effective or alternative isobaric systems, offering similar multiplexing capacity with different chemistries (33).

Among these chemical methods, TMT have become the most widely adopted isobaric labeling approach because they provide a higher degree of multiplexing than other strategies. Isobaric tags are designed with three functional groups: a reporter ion group that enables relative quantification, a mass balancer group that ensures the overall tag mass remains identical across all channels, and a reactive group that couples the tag to peptides or proteins which is usually an amine-reactive NHS-ester group, thereby reacting with amines and modifying lysine residues and peptide N-termini. The interplay between the reporter and mass balancer groups is central: labeled peptides have the same precursor mass and chemical properties during LC-MS, showing the same retention time and ionization efficiency, and thus appear as a single precursor peak in the MS1 spectrum. Upon fragmentation, the reporter ions are cleaved, and their intensities provide relative quantification across multiple samples. While low-resolution MS can be sufficient for some applications, high-resolution ($R > 35,000$) is required to distinguish closely spaced reporter ion channels, particularly in the latest tag designs that rely on small mass differences between ^{13}C - and ^{15}N -based isotopologs (36). TMTpro now can multiplex up to 35-plex and this makes it particularly suitable for large-scale and single-cell proteomics studies.

In parallel, dimethyl labeling has emerged as a distinct isotopic labeling method that ensures high reaction yield, high accuracy and high reproducibility, as well as robustness, simplicity and low cost (37). The reaction involved in this strategy is the reductive amination where formaldehyde forms a Schiff base with the N-terminus or ϵ -amino group of a Lys residue of a peptide/protein. The aldimine intermediate is reduced by sodium cyanoborohydride (NaBH_3CN) to form a secondary

amine which is more nucleophilic than the primary amine and immediately reacts with another formaldehyde molecule to form dimethylamino group. By combining different isotopic forms of formaldehyde and sodium cyanoborohydride, it is possible to multiplex up to 5-plex, but the risk is the overlapping of neighboring isotopic clusters. To avoid that and increasing the accuracy of the quantification, is suggested to ensure at least 4Da of the mass shift between the nearest labelled forms. Therefore, the triplex combination of $\text{CH}_2\text{O}/\text{NaBH}_3\text{CN}$ (light), $\text{CD}_2\text{O}/\text{NaBH}_3\text{CN}$ (intermediate) and $^{13}\text{CD}_2\text{O}/\text{NaBD}_3\text{CN}$ (heavy) that respectively weight +28.0313, +32.0564, +36.0757, is considered the winning strategy as a trade-off between multiplexing and accuracy (37).

Over the last 10 years, with the increase in the size of experiments in terms of the number of samples and, more recently, the development of single-cell proteomics, there has been a need to develop more and more high-throughput applications. To increase multiplexing or hyperplexing, there is a tendency to combine isotopic labeling with isobaric labeling. One of the first was cPILOT (combined precursor isotopic labeling and isobaric tagging), which exploits the difference in pKa between peptide N-termini and lysine residues to introduce dimethylation at the N-terminus, followed by TMT labeling of lysines to get a site-specific labeling. By combining light and heavy dimethyl groups with TMT reagents, cPILOT doubled the multiplexing capacity from six to twelve channels, later reaching up to 24-plex using DiLeu tags, and has been applied in studies of aging, oxidative stress, and disease models. Similar combinations have been achieved by coupling SILAC with TMT or iTRAQ, where isotopic labels are incorporated metabolically before digestion, further expanding quantitative depth. Other developments include NETLOP (NHS-ester tandem labeling in one pot), which merges 3-plex mTRAQ with 16-plex TMTpro for a 48-plex setup in a simplified one-step workflow, and TAG-TMTpro, a highly multiplexed system that modifies N-termini and lysine residues with Boc-protected alanine or glycine prior to 18-plex TMTpro labeling, achieving up to 54 simultaneous samples per analysis. Hyperplexing increases the number of samples that can be analyzed in a single experiment, but this often comes at the cost of reduced protein identifications and, in some cases, lower sensitivity due to charge effects and co-isolation interferences. Moreover, effective quantification requires that all labeled precursors co-elute during chromatography so that they are selected together for MS/MS, ensuring that reporter ions from every sample are consistently detected (36). Hyperplexing represents a powerful strategy

to increase analytical throughput in the context of single-cell proteomics, where the amount of material is extremely limited and maximizing data acquisition efficiency is crucial. These advanced approaches reduce instrument time per sample and minimize technical variability between runs. Some trade-offs in sensitivity and identification depth remain, but the integration of hyperplexing with optimized acquisition methods and low-input workflows offers a way to increase throughput and make single-cell proteomics suitable for larger studies.

1.5 Single-cell proteomics

In the past, single-cell protein analysis relied almost entirely on antibody-based methods. Techniques such as fluorescence microscopy, flow cytometry, CyTOF, and single-cell Western blotting enabled the quantification of specific proteins or epitopes within individual cells, but they were limited by antibody availability, cross-reactivity, and the small number of targets that could be measured simultaneously. At the same time, one of the best techniques for identifying and quantifying hundreds and thousands of proteins is mass spectrometry, but until then it had only been applied to bulk samples. This gave rise to the idea of developing an approach to study the proteome of individual cells that would integrate the potential of mass spectrometry (38).

To do this, it started by trying to adapt established mass spectrometry-based proteomics protocols on to the requirements of a single-cell experiment. The first limitation was sample loss, traditional lysis and cleanup methods relied on detergents or urea, which require desalting and result in major peptide loss when starting from picogram quantities. The second was insufficient ion statistics: peptide ions derived from a single cell are so few that stochastic sampling and counting noise dominate the MS signal. The third was peptide identification, since reliable fragmentation and sequencing normally depend on thousands of ions per precursor. To overcome the first limitation, the cells were isolated manually under a microscope, lysed mechanically by sonication to avoid detergents and digested with trypsin. To address the second and third limitation, was introduced the use of isobaric tandem-mass-tag (TMT) labeling to multiplex single cells together with a “carrier” sample made of about 200 cells. The carrier channel provided abundant peptide ions for sequencing and stabilized peptide identification, while the reporter ions allowed quantification of each single cell within the same MS/MS scan. This approach was named Single Cell ProtEomics by Mass Spectrometry (SCoPE-MS) and successfully quantified hundreds to over a thousand

proteins per single mammalian cell and could distinguish distinct cancer cell types and differentiation states purely based on their proteomes (38).

SCoPE-MS was the first methodology developed and despite being extremely innovative and revolutionary, remained technically demanding and low throughput, with high cost per cell and considerable hands-on time. These limitations were addressed in SCoPE2, which replaced the manual lysis with Minimal ProteOmic sample Preparation (mPOP), a simple freeze–heat cycle performed in multiwell plates with inexpensive thermocyclers. It also introduced data-driven optimization of MS (DO-MS) to systematically tune instrument parameters for single-cell analysis, and DART-ID, a Bayesian algorithm that increases peptide identification confidence by integrating retention time information. By miniaturizing and automating preparation and shortening chromatographic gradients, SCoPE2 reduced cost by over ten-fold and improved quantitative reproducibility. The study quantified more than 3,000 proteins across 1,490 single monocytes and macrophages, uncovering a continuous gradient of macrophage polarization states even in the absence of external cytokines (39).

A step further toward automation and reproducibility came with the introduction of nPOP (nano-ProteOmic sample Preparation). In nPOP, thousands of single cells are individually lysed, digested, and labeled in nanoliter droplets directly on a coated glass slide using piezo-acoustic dispensing. This design eliminates the need for multiwell plates and reduces reagent volumes and handling losses. The use of dimethyl sulfoxide (DMSO) for lysis enables efficient protein extraction without cleanup steps, maintaining near-bulk digestion efficiency and minimizing contamination. In a single batch, nPOP can prepare more than 2,000 cells, reducing batch effects while maintaining quantitative consistency across cells. Importantly, nPOP is compatible with both workflows with isobaric labels and label-free designs. The large datasets generated using this platform allowed the analysis of protein covariation across thousands of single cells, revealing regulatory patterns such as cell cycle–dependent changes and metabolic heterogeneity within melanoma subpopulations (40).

At the same time, analytical tools have also been optimized to keep pace with the new direction of research. The introduction of Trapped Ion Mobility Spectrometry coupled to Parallel Accumulation–Serial Fragmentation (TIMS–PASEF) improved sensitivity and data completeness.

The integration of ion mobility separation within the mass spectrometer improves ion utilization and reduces chemical noise. In this system, peptide ions are trapped and released as concentrated ion packets, which are then fragmented in rapid succession. The diaPASEF (data-independent acquisition PASEF) mode further increases the duty cycle, stacking precursor ions in the ion mobility dimension to maximize sampling efficiency. This technology enables label-free single-cell proteomics, capable of quantifying over 2,000 proteins per cell with high quantitative precision and allows the characterization of functional perturbations. Single-cell proteomics could now be performed without the need for boosting channels, providing an unbiased and linear quantitative response across individual cells (41). By integrating nPOP-based preparation with modern data processing method such as DIA-NN and FragPipe, peptide identification and quantification from low-input DIA and DDA datasets are optimized improving sensitivity and reproducibility.

In only a few years, single-cell proteomics (SCP) has evolved rapidly. Recent developments in automated sample preparation and high-sensitivity mass spectrometers have enabled analyses of hundreds of single cells per day, with a dramatic increase in throughput compared to early studies. Label-free methods, such as those based on low-flow LC–MS coupled to DIA acquisition, now provide straightforward and low-cost workflows that can be implemented in almost any proteomics laboratory, while multiplexed strategies using TMT, plexDIA, or coupled isotopic–isobaric labeling schemes offer higher throughput for large-scale studies. Together, these approaches have expanded the applicability of SCP to a range of biological questions, from developmental and cancer biology or aging to drug response profiling and biomarker discovery (42).

An important step forward has been the application of single-cell proteomics to specific biological mechanism, such as chromatin regulation. In this context, was recently introduced single-cell histone proteomics, demonstrating for the first time the feasibility of mapping post-translational modifications (PTMs) of histones at single-cell resolution (43). It has been shown that histone proteoforms can now be identified and quantified from individual mammalian cells by integrating optimized chemical derivatization strategies with customized LC–MS workflow and data analysis. The approach revealed heterogeneous combinations of acetylation and methylation marks across single cells, reflecting distinct chromatin states that could not be resolved in population-level analyses (43). This work demonstrates how single-cell proteomics can address functional

proteome regulation and epigenetic heterogeneity, bridging single-cell biology and chromatin research.

Despite these advances, several challenges are still limiting. An important one is the trade-off between proteome depth and throughput: current workflows can quantify thousands of proteins from single cells, but not at that same depth for each cell. Similarly, the robustness of ultra-low-flow chromatography continues to be a practical bottleneck, as narrower column diameters improve sensitivity but can reduce system stability and increase maintenance demands. Another key issue is quantitative accuracy and missing data. Both isobaric and label-free SCP datasets suffer from limited ion statistics and stochastic precursor sampling (42).

In single-cell proteomics the quantification of the proteome of individual cell is obtained dissociating tissues or isolating cells, thereby losing spatial information about the original tissue structure. In contrast, spatial proteomics focuses on preserving the localization context of the proteins within tissues, allowing the study of cell-cell interactions, microenvironmental niche and tissue organization. This approach is based on antibody-based imaging methods such as immunofluorescence or imaging mass cytometry, obtaining a high spatial resolution limited by antibody availability, specificity and multiplexing capacity (44). Recent advances in this field have begun to overcome these limitations allowing unbiased protein identification directly from tissue section. For example, with deep visual proteomics (DVP) thousands of proteins can be quantified from spatially defined single cells, allowing a direct correlation of proteomic states with tissue organization and cellular morphology (45). Despite these advances, spatial proteomics is limited in proteome depth and throughput. Consequently, single-cell proteomics and spatial proteomics should be considered as complementary rather than competing approaches, where SCP provides molecular depth and statistical power, while spatial proteomics adds contextual information.

Overall, single-cell proteomics has rapidly evolved from a new technological challenge into an analytical approach capable of capturing the molecular heterogeneity of individual cells. The continuous development of ultra-sensitive instrumentation, automated sample processing, and multiplexed acquisition strategies is expanding its accessibility and depth. With ongoing improvements in throughput, reproducibility, and data analysis, single-cell proteomics may become a routine tool for studying cell states, signaling, and epigenetic.

1.6 Single-cell heterogeneity in aging

Healthspan has become the main objective of geroscience research and is defined as the portion of life spent in good health, free from chronic diseases and functional decline that come with aging. Compared to lifespan which is easier to define as the total duration of life, healthspan is more elusive because the concept of good health is not static (46). Health can vary overtime, reflecting the variety of multiple physiological systems that collectively determine a functional organism. For this reason, it would be wrong to consider healthspan as a simple chronological interval, but rather as a trajectory where the health index can decrease, stabilize, or temporarily improve over time (46). Therefore, studies on aging and healthspan should be conducted to develop approaches and interventions that not only extend lifespan but also preserve function and delay systemic decline (46). Another important consideration is that the biology of aging must be viewed as a complex, multidimensional system. Organisms are networks of molecules, cells, and physiological components that interact with each other and are evolutionarily programmed to maintain a dynamic equilibrium (47). Aging, therefore, represents the progressive failure of this balance, leading to a decrease in resilience, feedback loops becoming dysregulated, and a gradual decline in the ability of the entire system to recover from perturbations (47). Since aging is a process that is not defined by a single pathway but rather by a combination of processes ranging from mitochondrial function and cellular metabolism to the coordination of all organs, it is important to use healthspan models that reflect this complex system and are capable of measuring and capturing the physiological diversity and variability that define it (47).

Aging does not affect the tissues uniformly but changes the proportions and internal states of specific cell types, and it is precisely at a cellular level that variability begins to manifest. Bulk assays average across the mixtures and can miss the minority populations that drive dysfunction (48). Single-cell and single-nucleus RNA-seq overcome this by deconvoluting cell identity and state, showing how aging affect cell compositions and cell programs within a tissue. Single-cell studies detect an increased transcriptional heterogeneity, called “noise”, with age in defined cell types, an expansion of inflammatory subpopulations, and pathway shifts that differ by cell type e.g. proteostasis and lipid metabolism (48). One example of this picture is observed in inflammaging which is the chronic inflammation that rises with age. Single-cells analyses showed that myeloid populations with stromal and endothelial cells are the main sources of pro-

inflammatory cytokines and senescence-associated secretory phenotype (SASP) factors (49). Single-cell atlases provide a map of how the inflammation and senescence manifest within a tissue, demonstrating that only a portion of cells express genes related to those pathways with a strong upregulation of SASP and immune response, the others remain relatively stable or activate compensatory stress-response programs (49).

Variability in senescence is also evident in the search for specific biomarkers. In fact, the molecular signatures of senescence vary greatly depending on factors such as cell type, triggering stressors (replicative exhaustion, DNA damage, oncogenic signaling, oxidative or metabolic stress) and tissue environment, therefore no single marker reliably identifies all senescent cells (49). Classical markers such as p16 (Cdkn2a) and p21 (Cdkn1a) remain useful, but they capture only subsets of the senescent population. Some p16^{high} cells are not truly senescent, while other senescent cells lack p16 expression altogether (49).

Until now, to capture the complexity of aging single-cell RNA-seq have been done across multiple tissues such as brain, lung, pancreas, muscle, immune system, and liver, to dissect these changes. These studies revealed two main patterns: first, an increase in transcriptional heterogeneity within individual cell types, reflecting deregulated gene expression and loss of identity; and second, the emergence or expansion of rare subpopulations, such as senescent, inflammatory, or metabolically reprogrammed cells, that become more prevalent with age (50). In this field, the Tabula Muris Senis project generated over 500,000 single-cell transcriptomes from 23 tissues of young and old mice. This atlas enabled systematic mapping of how specific cell types change during aging. For instance, immune infiltration and activation increase across tissues, while stem and progenitor cells often show exhaustion or differentiation bias (50). Other focused studies have applied single-cell analyses to identify tissue-specific hallmarks of aging: in the pancreas, β -cell subtypes lose insulin granule genes and gain stress-response signatures; in muscle and bone marrow, satellite and hematopoietic stem cells exhibit transcriptional drift and reduced regenerative potential; in lung and kidney, epithelial and endothelial cells acquire pro-inflammatory, SASP-like phenotypes (50). Collectively, these datasets demonstrate that aging affects different cell types through different pathways, many of which converge on inflammation, metabolic imbalance, and loss of proteostasis.

Single-cell transcriptomics has increased our understanding of how aging changes cellular composition and transcriptional states, but gene expression at the RNA level does not always reflect functional protein output, particularly in older tissues. Multiple studies demonstrate that the correlation between mRNA and protein decreases during aging, leading to a post-transcriptional noise that cannot be captured by transcriptomics alone (51). A study comparing the prefrontal cortex in humans and macaques, showed that the strong mRNA-protein correlation during development collapsed in the brain during aging. It was observed that protein levels often change independently of transcript levels, with many genes showing opposite trends at RNA and protein stages (51). This phenomenon was enriched in pathways central to longevity such as mTOR signaling, mitochondrial function, and neurodegeneration, and could be mechanistically linked to RNA-binding proteins (RBPs) and microRNAs that control RNA stability and translation efficiency (51). Aging introduces dysregulation at translation and degradation that reshapes the proteome even when transcription remains stable. A more recent study on mouse and human midbrain, using matched transcriptomic and proteomic data, demonstrated that protein levels become more variable and less predictable from RNA, particularly in neurons, glia, and mitochondrial proteins (52). The study identified age-related shifts in translation efficiency and proteostasis, suggesting that impaired ribosomal fidelity, altered RNA-binding activity, and reduced protein turnover amplify stochastic variation. In older tissues, this loss of coordination means that transcript levels no longer faithfully reflect functional protein output (52).

These studies show that aging increases post-transcriptional and post-translational noise, driven by dysregulated translation, RNA processing, and protein degradation but transcriptomics alone cannot accurately describe the molecular state of aging cells. This highlights the need for single-cell proteomics, which directly measures proteins at the level where aging-related regulation actually occurs. Single-cell proteomics can detect cell-to-cell variability in protein abundance, modification, and turnover, providing the missing link between transcriptome predictions and the biochemical reality of aged cells. In tissues such as the liver, where metabolism and proteostasis are tightly regulated and highly heterogeneous, these methods are crucial to understanding how individual hepatocytes respond to aging and metabolic interventions, and to identify functional cell states that cannot be detected from RNA data alone.

1.7 Histone proteomics

Chromatin is the physiological template of all eukaryotic genetic information, is a dynamic polymer where the genomic DNA is highly folded and compacted, presents different states of organization that depend on the structuring order of the nucleosomes which are the basic repeating unit of chromatin. The nucleosomes are formed by an octamer of histone proteins composed of four histone partners: the H3-H4 tetramer and two H2A-H2B dimers, while the histone H1 sits on the outside and helps stabilize the structure. Histones, in turn, are proteins that have a globular domain and a flexible, charged “tail” at the NH₂-terminal that protrudes outward from the nucleosome. It is precisely at the level of histone tails that are located the vast majority of post-translational modifications that play a fundamental role in DNA regulation processes. In fact, together with DNA methylation, histone modifications are one of the mechanisms involved in epigenetic regulation because they are able not only to regulate chromatin structure but also to recruit proteins or complexes with specific enzymatic activities that contribute to chromatin remodeling (53,54). Among the more of 20 known modifications there are methylation, acetylation, phosphorylation, ubiquitination, SUMOylation, ribosylation, citrullination (deamination) and lactylation (55). Depending on the modifications on the histone tail, the chromatin results in a more open chromatin structure (euchromatin), which makes DNA more accessible to transcription factors and the transcriptional machinery or a close chromatin structure (heterochromatin) which, on the contrary, makes DNA inaccessible (53,54). Two of the most studied modifications that regulate chromatin structure and gene expression are acetylation and methylation. Histone acetylation generally correlates with gene activation. It occurs mainly on lysine residues of histone tails and is catalyzed by histone acetyltransferases (HATs). Acetylation neutralizes the positive charge of lysine, weakening the interaction between histones and negatively charged DNA. Histone deacetylases (HDACs) can remove these groups, leading to chromatin condensation and gene repression (54). Histone methylation can be associated with either activation or repression, depending on the specific residue and number of methyl groups added. Histone methyltransferases (HMTs) add one, two, or three methyl groups to lysine or arginine residues, creating signals that are read by specific binding proteins. Methylation does not change the charge of histones, but it provides a binding platform for effector proteins that control chromatin state and transcription (54). Histone tails often display combinatorial patterns of PTMs, where multiple modifications coexist on the same tail or even on the same amino acid residue.

These combinations can work together in a synergistic or antagonistic manner depending on the type of the modification and on the location, changing the chromatin state and gene activity. The set of all modifications and this combinatorial complexity determine the so-called histone code (53). Changes in chromatin structure are associated with cell cycle progression, DNA replication, DNA damage and its repair, recombination, and overall chromosome stability, and it is precisely for this reason that the study of histone modifications has become increasingly important over time, to the point of becoming one of the main focuses in understanding complex and relevant disease states such as cancer and aging (53).

Traditionally, the histones have been detected by antibody-labeling, however, in recent years, mass spectrometry has become one of the main techniques because it allows unbiased identification, localization, and quantification of multiple modifications within the same protein or peptide. With mass spectrometry, it is possible to identify PTMs with very similar masses, such as trimethylation and acetylation. Furthermore, combinatorial marks coexisting in the same histone can be identified (56).

The analysis of histones using mass spectrometry presents unique challenges compared to other proteomics experiments, making this type of study more complex. Histones are highly basic, small proteins (10–20 kDa), enriched in lysine and arginine residues which, following digestion with trypsin, form very short, hydrophilic peptides that are difficult to retain using reverse-phase chromatography and can easily be lost during sample preparation. Furthermore, lysine is a modification site that is often exploited as a target for modifications such as acetylation, methylation, propionylation, crotonylation, and many others, for which protease accessibility is reduced, thus creating a very heterogeneous mix of peptides that presents overlapping isobaric forms. Not only that, but many of these modifications are very low in terms of stoichiometry, requiring a very high instrument sensitivity. To overcome these issues, specialized workflows have been developed. A key innovation was chemical derivatization, typically with propionic anhydride in which amine groups on unmodified lysines are acylated, blocking further cleavage by trypsin producing Arg-C-like peptides that are well suited for reversed-phase nano-HPLC and increasing peptide hydrophobicity for improved LC separation (56).

However, because the mass of chemically added propionyl groups is identical to that of endogenous lysine propionylation, this strategy complicates the direct study of naturally occurring acylations. To overcome this limitation, deuterated propionic anhydride (D_5 -propionylation) was introduced, providing a distinct mass shift (+5 Da) that allows the discrimination between endogenous and derivatized propionyl groups maintaining identical chemical and chromatographic properties (57,58). Beyond that, deuterated propionylation also serves as a quantitative and multiplexing tool. By labeling different samples with isotopically light or heavy reagents, relative quantification of histone peptides can be achieved through differential isotopic encoding (57). These developments not only correct a long-standing analytical artifact in histone proteomics, but the combined use of light and deuterated propionic anhydride enables multiplexed quantitative analysis. By derivatizing different samples with isotopically distinct reagents and analyzing them together, relative abundance of histone peptides can be accurately compared within a single LC–MS run.

The main approaches that are employed are both bottom-up and top-down MS. Bottom-up MS, which digests histones into short peptides, offers high sensitivity and is compatible with chemical derivatization and multiplexed quantification strategies, enabling identification of thousands of modified peptides in a single run. In contrast, top-down and middle-down methods analyze large protein fragments or intact histones, preserving information on combinations of coexisting PTMs within the same molecule. These approaches are uniquely used to investigate combinatorial modification patterns, although they require high-resolution instrumentation (e.g., Orbitrap or FT-ICR) and advanced fragmentation methods such as electron capture dissociation (ECD) or electron transfer dissociation (ETD) to preserve labile PTMs during fragmentation (56). Quantitative strategies such as SILAC, TMT, iTRAQ, or label-free analysis complement these workflows, enabling relative or absolute quantification of modification states across conditions (56).

Further advances also rely on the development of dedicated software and database-search algorithms capable of handling the high number of possible modification states. Programs such as Mascot, Sequest, Andromeda, Skyline and PTMap allow mapping of known PTMs, while non-restricted search tools (e.g., InsPecT, MODa, or PILOT_PTMs) can discover unexpected mass shifts indicative of novel histone modifications (56).

One of the most widely used is EpiProfile, which accurately quantify histone peptides containing multiple and often isobaric post-translational modifications. EpiProfile integrates prior knowledge of peptide retention times and characteristic fragment ions to discriminate between co-eluting or isobaric species and to extract chromatographic peak areas with high precision. By combining MS1 and MS2 information, it reliably resolves complex mixtures. Compared to manual quantification or programs like Skyline, EpiProfile provides faster, more reproducible, and histone-specific quantification, enabling the large-scale and accurate profiling of histone modification patterns in biological samples (59).

Together, these technical innovations that combine optimized digestion protocols, chemical derivatization, high-resolution MS with appropriate fragmentation modes, and customized bioinformatic pipelines, have made it possible to map and quantify hundreds of distinct histone modifications with high accuracy and sensitivity.

CHAPTER 2: A NEW METHOD TO MULTIPLEX SINGLE CELLS FOR THE ANALYSIS OF HISTONE POST-TRANSLATIONAL MODIFICATIONS

Paper accepted by JoVE Journal.

TITLE

Automated Sample Preparation for the Multiplexed Analysis of Single-Cell Histone Post-Translational Modifications (sc-hPTM2)

AUTHORS AND AFFILIATIONS

Giulia Barotti, Ronald Cutler, Joshua Cantlon, Barbara Montanini and Simone Sidoli

KEYWORDS: histone post-translational modifications, single-cell proteomics, mass spectrometry, epigenetics, chromatin, isotopic labeling, sodium butyrate, hepatocellular carcinoma spheroids

ABSTRACT

Histone post-translational modifications (hPTMs) are central regulators of chromatin organization and gene expression. Their dysregulation is implicated in development, cancer, and aging. While mass spectrometry is the method of choice to study hPTMs, most protocols are designed for bulk samples and cannot resolve cell-to-cell variability. Extending histone PTM analysis to the single-cell level is therefore essential but technically challenging, as histones are low in abundance, lysine-rich, and heavily modified.

Here, we describe an automated single-cell proteomics workflow for histone PTM analysis using nano liquid handling. The system enables nanoliter-scale processing of individual cells with minimal handling and high reproducibility. The workflow includes digestion with ArgC Ultra

protease, which cleaves at arginine residues and avoids the need for lysine-blocking derivatization steps typically required in histone proteomics. To enable quantitative comparison between conditions, we apply a two-plex labeling strategy with propionic anhydride and its deuterated analogue, propionic anhydride-d10. This combination of automated sample preparation, isotopic multiplexing, and ArgC Ultra digestion results in a streamlined and sensitive protocol for single-cell histone PTM analysis.

The method allows to investigate chromatin heterogeneity and the effects of epigenetic perturbations at single-cell resolution. To demonstrate the workflow, we generated spheroids from HepG2/C3A hepatocellular carcinoma cells and treated them with sodium butyrate, a histone deacetylase inhibitor that induces hyperacetylation.

SUMMARY

Here, we present a protocol to automate single-cell histone post-translational modification analysis using an isotopic two-plex labeling strategy and ArgC digestion. This workflow enables quantitative, reproducible, and high-sensitivity profiling of chromatin heterogeneity and epigenetic responses at single-cell resolution.

INTRODUCTION

Histones are the protein components of nucleosomes, the repeating units of chromatin. Their N-terminal tails are rich in lysine and arginine residues and are highly regulated by post-translational modifications (PTMs), such as acetylation, methylation, and phosphorylation. These modifications influence DNA accessibility and thereby control processes including transcription, replication, and repair. Altered histone PTM patterns are strongly associated with disease states, from cancer to neurodegeneration and aging(54,60). For this reason, histone PTM analysis has become of great interest in epigenetics and proteomics research. Mass spectrometry (MS) provides site-specific and quantitative analysis of histone modifications, but histones present unique analytical challenges. Their amino acid composition, multiple modification sites, and complex combinatorial patterns

make them more difficult to analyze than typical proteins. Established protocols rely on chemical derivatization to block lysine residues before digestion, improving peptide recovery and chromatographic performance(61). While effective in bulk analyses, these multi-step derivatization protocols can introduce losses which are problematic for single-cell inputs. Single-cell analysis is increasingly important because biological systems are inherently heterogeneous. Bulk histone analyses average signals across many cells and thus fail to capture this variability. Subpopulations of cells may differ in their chromatin states, and these differences can drive development, cellular reprogramming, or drug responses(38,62).

Here, we present our automated sample preparation method for the multiplexed single-cell histone PTM analysis (sc-hPTM2). This method relies on the CellenONE system, which allows precise dispensing of individual cells and nanoliter-scale reagents in an automated fashion. This method is a variation on the nano proteomic sample preparation (nPOP) method, which first introduced the use of droplet preparation on a glass slide(63). The nPOP platform has advanced automated single-cell proteomics, yet its standard tryptic digestion protocol is suboptimal for histones due to their high lysine content and extensive modification density. Here, we adapted the CellenONE system to incorporate the gold-standard histone derivatization with propionic anhydride and introduced isotopic labeling, enabling accurate and comprehensive quantification of histone post-translational modifications at single-cell resolution. Specifically, we simplify the protocol by using ArgC Ultra protease for protein digestion. Unlike trypsin, ArgC cleaves only at arginine residues and does not require lysine blocking. This eliminates the need for a two-step derivatization strategy, streamlining the protocol. In addition, the method incorporates a two-plex isotopic labeling scheme using propionic anhydride and propionic anhydride-d10, allowing for 2 cells to be analyzed simultaneously(61,64). Following digestion and labeling, single-cell samples are recovered and analyzed by LC-MS/MS using data-independent acquisition (DIA). In our hands, this method has shown to provide an increase in both sensitivity and reproducibility for quantification of histone PTMs as compared to the previous version of this method(65).

To demonstrate our method, we profiled single-cell histone PTM dynamics in spheroids derived from HepG2/C3A hepatocellular carcinoma cells. We treated these spheroids with sodium butyrate, a well-characterized histone deacetylase inhibitor that increases histone acetylation and serves as a benchmark for probing chromatin regulation(66,67). By comparing control and sodium

butyrate-treated spheroids at single-cell resolution, we demonstrate a proof of principle of representative results, highlighting the potential for broader applications in chromatin biology and cancer epigenetics. Notably, for the protocol we refer to a prior publication describing growth and treatment of 3D spheroids(67), acquisition of histone peptides in mass spectrometry¹², and data analysis using the software EpiProfile(68).

PROTOCOL

Buffers and reagents:

1. Cell growth media for HepG2/C3A cells: Dulbecco's Modified Eagle's Medium (DMEM, 4.5g/L glucose) containing 10% fetal bovine serum (FBS), non-essential amino acids (1% v/v), L-glutamine (1% v/v) and penicillin/streptomycin (0.5% v/v). Growth media is stored at 4 °C.
2. Master mix: ArgC Ultra 60 ng/ul, HEPES 6 mM, DDM 0.03%, DTT 10 mM in LC-MS grade water.
3. HA solution: 1% (v/v) hydroxylamine in LC-MS grade water.
4. Wash solution: 10 mL of a 50 : 50 acetonitrile / water mixture containing 0.1 % (w/w) formic acid (all LC-MS grade)
5. Mobile Phase A (MPA) - 0.1% formic acid: Add 1 mL of concentrated formic acid to 999 mL of HPLC-grade water and mix well.
6. Mobile Phase B (MPB) - 80% HPLC-grade acetonitrile + 0.1% formic acid: Add 800 mL of HPLC-grade acetonitrile to 199 mL of HPLC-grade water. Next, add 1 mL of concentrated formic acid to this solution and mix.

1. Spheroids preparation

For the spheroids preparation we refer to Stransky S, Young D, Mikkelsen K, Thulesen AP, Frandsen HS, Sidoli S. Semi-Automated Phenotypic Analysis of Functional 3D Spheroid Cell Cultures. J Vis Exp. 2023 Aug 18;(198). doi: 10.3791/65086. PMID: 37607073.

Culture HepG2/C3A hepatocellular carcinoma cells in a rotating 3D bioreactor for 5–7 days to promote uniform spheroid formation until reaching ~400–600 μm diameter. Use DMEM supplemented with 10% FBS, 1% non-essential amino acids, 1% L-glutamine, and 0.5% penicillin/streptomycin. Maintain spheroids at 37 °C and 5% CO₂ until treatment.

NOTE: Figure 4A represents the equipment used to prepare the spheroids in the reference paper.

2. Cell culture and treatment

- 2.1 Put ~10 spheroids in a tissue-culture 6-wells containing 2 mL of growth media.
- 2.2 Incubate the cells at 37 °C in a humidified 5 % CO₂ incubator.
- 2.3 Treat the spheroids for 24 h with 5mM sodium butyrate or add the same volume of water for the control (Figure 4B).

3. Cell isolation and single cell suspension preparation

- 3.1 After 24h of treatment, incubate the cells with 1 mL of tryPLE at 37°C for 3-5 min.
- 3.2 Gently pipette up and down until the spheroid breaks apart and single cells are obtained.
- 3.3 Collect the cells in a 15 mL tube and wash diluting up to 10 mL with ice-cold 1X PBS, centrifuge at 500 x g for 5 min at 4°C and aspirate the supernatant.
- 3.4 Resuspend the cell in 1 mL ice-cold 1X PBS and add Sytox green dead cell stain with a final concentration of 1 μM .

- 3.5 Incubate the cell with Sytox green for 20 min on ice in a dark environment.
- 3.6 Wash the cell with 10 mL of ice-cold 1X PBS, centrifuge at 500 x g for 5 min at 4°C and aspirate the supernatant.
- 3.7 Resuspend the cell in 1 mL ice-cold 1X PBS and add paraformaldehyde (PFA) with a final concentration of 0.5%.
- 3.8 Incubate the cell with PFA for 15 min on ice.
- 3.9 Wash the cell with 10 mL of ice-cold 1X PBS, centrifuge at 500 x g for 5 min at 4°C and aspirate the supernatant.
- 3.10 Reconstitute the cells at final concentration of 200–500 cells/μL.

NOTE: For small cells that do not tend to deposit, a concentration of 500 cells/ul is recommended, while for large cells that tend to deposit, a concentration of 200 cells/ul is recommended. If the cell suspension is prone to aggregation, run through a 40-μm strainer before aspiration to avoid nozzle clogging. If cells larger than 40 μm are of interest and probably present in a sample, a 70-μm strainer can be used instead to remove very large aggregates.

- 3.11 Store the cells on ice until loaded into the instrument for sorting.

4. Software connection and system initialization

- 4.1 Turn on the cellenONE system (Figure 5A), boot the control computer, launch the software, and open the “**v2.0 -nPOP2**” user workspace; all required field files and run templates for nPOP are pre-loaded there.
- 4.2 Make sure the instrument’s chiller is switched on, then set the dew-point-chase value according to the internal temperature currently reported by the instrument.
- 4.3 Prime the system with PDC-70, coating-type 2 dispensing nozzles (Figure 5B) installed in slots 1 and 3, following the on-screen instructions.

- 4.4 Set each nozzle's pulse width and drive voltage to the values specified on the manufacturer's packaging.

NOTE: Using two separate nozzles preserves dispensing consistency: one delivers the cell suspension, the other the reagents, so any debris adhering to a nozzle will not compromise subsequent drops. Employing both nozzles during the final collection step also accelerates sample recovery.

CRITICAL STEP: For maximal pick-up efficiency, the z-offset between the two nozzles should differ by $< 50\ \mu\text{m}$. Determine this by defining the optimal contact position for each nozzle and comparing the Z-height read-outs in the Nozzle Setup tab.

- 4.5 Fill the humidifier reservoir with de-ionised water up to the indicated fill line.

5. Pre-run configuration on the liquid handler

- 5.1 Select the quantity of glass slides (Figure 5C) to be prepared. This choice directly defines the total number of cells you plan to analyze.
- 5.2 For a 2-plex multiplexing scheme (Figure 6A), a single slide is loaded with 320 cells.

6. Dispensing a DMSO “moat” around slide edges

NOTE: Figure 6B represents the steps for preparing the samples in the cellenONE starting from the moat of DMSO.

NOTE: As an additional prevention to droplet evaporation, DMSO is dispensed along the edges of the slides.

- 6.1 In the Main tab, select the run to “**0_Dispense_DMSO_Moat**”.
- 6.2 Go to Target Setup tab, load the **DMSO_Moat.fld** file from the appropriate scheme folder, and set No. of Fields to Y, matching the number of slides you plan to process.

- 6.3 Aliquot 150 μ L of LC-MS–grade DMSO into a new PCR tube and place it into position 2 of the CellenWASH station with the tube cap facing the instrument door.
- 6.4 Open the Run tab and click **Start Run**. The system will aspirate 10 μ L of DMSO and pause for user confirmation of the droplet stability. Fire three autodrop test droplets (Figure 7A); if all three are perfectly formed, click **Continue Run**.
- 6.5 After dispensing, the nozzles flush automatically and the instrument captures images across all slides for quality control.
- 6.6 Review the images to verify a uniform DMSO droplet on every slide and to check for missing or off-target droplets (Figure 7B).

7. DMSO dispense

- 7.1 In the Main tab, select the run to “**1_Dispense_DMSO**”
- 7.2 Open Target Setup tab, import the field file DMSO.fld (Figure 7C) from the appropriate scheme directory, and set Number of Fields to Y, matching the slide count you intend to process.
- 7.3 Aliquot 150 μ L of LC-MS–grade DMSO into a new PCR tube and place it into position 2 of the CellenWASH station with the tube cap facing the instrument door.
- 7.4 Open the Run tab and click **Start Run**. The system will aspirate 10 μ L of DMSO and pause for user confirmation of the droplet stability. Fire three autodrop test droplets; if all three are perfectly formed, click **Continue Run**.
- 7.5 When dispensing finishes, the software automatically flushes the nozzles and captures images across every slide for quality assurance.
- 7.6 Review the images to verify a uniform DMSO droplet on every slide and to check for missing or off-target droplets.

8. Cell dispense

- 8.1 In the 'Main' tab, select the run to “**2_Dispense_Cells**”
- 8.2 In Target Setup tab, open the Cells Field Files sub-folder of your chosen multiplexing-scheme directory.
- 8.3 Because the experiment involves two conditions, pick the field file that matches the sample you are about to sort—either CellType_A.fld or CellType_B.fld.
- 8.4 After the file loads, set Number of Fields to Y, so it equals the total slide count for this run.
- 8.5 Degas the cell suspension from the point 3.11 with a vacuum pump for 10 min on ice.
- 8.6 Transfer 150 µL of the prepared cell suspension into a clean CellenVIAL, then place the vial in slot 1 of the CellenWASH carousel with its cap pointing toward the instrument door.
- 8.7 Go to Nozzle Setup tab → Do Task and execute “Take10ul_CellenVIAL_nozzle1” to aspirate the sample from the vial.
- 8.8 Staying in Nozzle Setup tab, launch the cell-dispensing viewer (Figure 8A) and run the mapping routine to define the ejection zone and gate objects based on size and elongation. The single-cell isolation parameters shown in Figure 8A correspond to the exact settings used for dispensing HepG2/C3A cells dissociated from spheroids.
- 8.9 Then activate the fluorescence mode by selecting ‘T > F’. Set the selection mode to negative. Increase the isolation parameter range for size and intensity from the minimum to the maximum values to ensure that no dead cells are isolated.
- 8.10 Once you are satisfied with the mapping parameters, open the **Run** tab and click **Start Run**, then follow the software prompts. When dispensing ends, the instrument will automatically flush the nozzles and capture slide images so you can review droplet placement and uniformity.

NOTE: During cell sorting, images are taken of each individual cell that is isolated during the run (Figure 5B).

- 8.11 After cell sorting finishes, close the dispensing window.
- 8.12 Keep any leftover cell suspension on ice until Part 8 is over; the surplus can be used later for bulk samples, empirical-library construction, or additional validation studies.

9. Digestion master mix dispense

- 9.1 In the 'Main' tab, select the run to "**2_Dispense_Digest_10ulPickup**".
- 9.2 Open Target Setup tab, load Digest.fld from the appropriate multiplexing-scheme folder, and set Number of Fields to Y so it equals the slide count for this run.
- 9.3 Prepare a fresh digestion master mix and degas the solution with a vacuum pump for 10 min on ice. Master mix: ArgC Ultra 60 ng/ul, HEPES 6 mM, DDM 0.03%, DTT 10 mM in LC-MS grade water.
- 9.4 Pipette 20 μ L of the master mix into a 384-well plate and place the plate inside the X1 system.
- 9.5 In the **Run** tab, click **Start Run** and follow the prompts. After the system aspirates the reagent, confirm droplet stability when prompted.
- 9.6 When dispensing finishes, the instrument automatically flushes the nozzles and photographs every slide; review the images to verify that each droplet received the digestion mix uniformly.
- 9.7 Incubate the slides overnight for digestion, maintaining the specified humidity and dew-point settings. Each droplet contains 13.5 nL of digestion mix. Digestion should be carried out overnight at ~75 % relative humidity, with the dew point maintained 1–2 °C below the chamber temperature to prevent droplet evaporation.

10. Incubation overnight

10.1 In the 'Main' tab, choose the run to "4_Incubation_16h".

10.2 Switch to the **Run** tab, click **Start Run**.

NOTE: The instruments will take pictures of the slides every 30 min to monitor the condition of the droplets. If they dry out or swell during incubation, adjust the dew-point setting as needed.

10.3 When the incubation is done, a message will pop up saying that incubation is completed.

Note: Rapid digests (≤ 4 h) may be achievable with higher effective enzyme-to-substrate conditions; this was not evaluated here, so the validated condition remains 16 h.

11. Evaporation

NOTE: Before moving on to the labeling step, it is recommended to check the condition of the droplets to ensure that they are not too swollen with water, as water could interfere with the efficiency of labeling. For this purpose, the evaporation phase is recommended.

11.1 In the 'Main' tab, choose the run to "5_Evaporation".

11.2 Go to File, Cooling Unit Control and under the Dew Point Control set Fixed Control Temperature at 23°C.

11.3 Switch to the **Run** tab, click **Start Run**.

11.4 The dry will last for 5 min, and the instruments will automatically capture pictures of the slides after the drying.

11.5 Inspect the pictures, if the droplets are sufficiently dry stop the run. If not, click continue and dry for another 5 min (Figure 9A).

11.6 At the end of the run, switch back to the previous Dew Point Control.

12. Labels dispense

- 12.1 In the 'Main' tab, choose the run to "6_Dispense_PA_Labels".
- 12.2 Go to Target Setup tab, load the Labels.fld file from the appropriate scheme folder, and set No. of Fields to Y, matching the number of slides you plan to process.
- 12.3 Aliquot 150 μ L of LC-MS–grade DMSO into slot 2 of the CellenWASH station.
- 12.4 Prepare a stock solution of the two labelling stocks for multiplexed histone derivatisation: 25 % (v/v) propionic anhydride in DMSO and 25 % (v/v) propionic anhydride- d_{10} in DMSO.

NOTE: Degas both solutions before transferring them to the plate to minimize bubble formation.

- 12.5 Dispense 20 μ L of each labelling reagent into designated wells of a 384-well plate and place the plate inside the X1 system. Each droplet receives 10 nL of labeling reagent (25% propionic anhydride or 25% propionic anhydride- d_{10} in DMSO).
- 12.6 Switch to the Run tab, click Start Run, and follow the prompts.
- 12.7 Allow the labels to react on-slide for 1 h.
- 12.8 After dispensing, the instrument automatically flushes the nozzles and photographs every slide. Review the images to confirm even distribution and correct droplet placement across all slides.

13. HA dispense

- 13.1 Prepare the HA solution and degas it before use.
- 13.2 Load the vial. Pipette 150 μ L of the HA solution into a clean PCR tube, position it in CellenWASH slot 2, and orient the cap toward the instrument door.
- 13.3 Select the run profile. In the Main tab choose "7_Dispense_Quench."

- 13.4 Configure the field file. Open Target Setup tab, load HA.fld from the appropriate multiplexing-scheme folder, and set Number of Fields to Y so it equals the total slide count.
- 13.5 Start the run. Navigate to **Run**, click **Start Run**, and follow the on-screen instructions.
- 13.6 Incubate. Let the HA react on the slides for 30 min.
- 13.7 Verify dispensing. After completion, the system flushes the nozzles and captures images of each slide—review these images to confirm uniform droplet placement and volume.

14. Sample pickup

- 14.1 Prepare the wash solution into the WashTray XL.
- 14.2 Set up the collection plate. Dispense 2 μ L of 0.01 % (w/w) DDM in LC-MS–grade water into every well of a fresh 384-well plate and place the plate inside the liquid handler.
- 14.3 Choose the pickup run. In the Main tab, select the profile “8_Pickup_2plex.”
- 14.4 Load the field file. Open Target Setup tab, import Pickup.fld from the correct multiplexing-scheme directory, and set Number of Fields to Y so it matches your slide count.
- 14.5 Execute the run. Switch to the **Run** tab, click **Start Run**, and follow the prompts. The instrument will collect each droplet, flush the nozzles, and image the slides so you can confirm that every sample has been retrieved.

15. Post-pickup processing

- 15.1 Seal the 384-well plate with adhesive foil and centrifuge briefly to bring droplets to the bottoms of the wells.
- 15.2 Remove the foil and dry the samples at low heat in a SpeedVac concentrator.

- 15.3 Store the dried plate at -80 °C until you are ready to reconstitute the samples for LC-MS/MS analysis.

16. Bulk samples preparation

- 16.1 Use leftover suspension from part 8.
- 16.2 Thermal lysis: snap-freeze the remaining suspension at -80 °C for 10 min, then heat-shock at 90 °C for 10 min.
- 16.3 Allow the tube to cool back to room temperature.
- 16.4 Transfer 25 µL of the lysate to a PCR tube and add 5 µL of the previously prepared digestion mix.
- 16.5 Incubate overnight ($\approx$ 16 h) at 37 °C in a thermal cycler.
- 16.6 Light-label derivatisation: add 5 µL of a fresh 25 % (v/v) light label in acetonitrile and incubate for 1 h.
- 16.7 Reaction quench: add 5 µL of 1 % (w/w) hydroxylamine (HA) solution and incubate for 30 min.
- 16.8 Dry the sample to completeness in a SpeedVac set to medium heat, then store or reconstitute as required for LC-MS/MS.

17. LC/MS setup and data acquisition

NOTE: to collect the data for this manuscript, a nLC-MS/MS system setup with an Ion Optics 75 µm ID x 25 cm x 1.7 µm analytical column.

- 17.1 Prepare the mobile phases to run on the HPLC:

- a. Mobile Phase A (MPA): 0.1% formic acid in HPLC-grade water
- b. Mobile Phase B (MPB): 0.1% formic acid in HPLC-grade acetonitrile

17.2 Program the HPLC method:

- a. 4-30% MPB over 37.5 min, 30-99% MPB over 7.5 min, isocratic 99% MPB over 6 min
- b. Flow rate of 400 nl/min at low % of MPB and 200 nl/min at high % of MPB

17.3 Set up the MS acquisition method to perform data-independent acquisition (DIA):

- a. Set full MS scan to 300-1100 m/z in the orbitrap, with a resolution of 120,000 and an automatic gain control (AGC) target of 300.
- b. Set MS/MS in the orbitrap with sequential isolation windows of 50 m/z with an AGC target of 1000 and a Higher-energy collisional dissociation (HCD) energy of 27.

17.4 Cover the plate with a rubber seal mat and place it inside the autosampler.

17.5 Queue up runs: run first the bulk samples and then the sets of single cells.

18. Data analysis

- 18.1 Run the raw data with EpiProfile2.1 to get the output with the peptidoforms abundances for the cells labeled with propionic anhydride. EpiProfile 2.1 extracts approximately 250 peptidoforms with its canonical version, but it is editable depending on organism, additional modifications of interest, mutations, etc.¹³
- 18.2 Run the raw data with EpiProfile2.1_D5 to get the output with the peptidoforms abundances for the cells labeled with propionic anhydride-d10.
- 18.3 Perform the mapping linking the metadata with the EpiProfile outputs with the quantQC program.

- 18.4 Filter out failed samples by comparing with the blanks.
- 18.5 Filter out peptidoforms which are not in at least 50% of the samples.
- 18.6 Filter out peptidoforms whose signal of the single cell intensity is not greater than the 150% signal in blanks.
- 18.7 Obtain the results table (Figure 10A) and process the data further as needed.

RESULTS

To demonstrate the application of our single-cell histone proteomics workflow using our system, we analyzed spheroids of HepG2/C3A cells treated with 5 mM sodium butyrate for 24 h. After LC–MS/MS acquisition, raw data were processed with EpiProfile for both light (propionic anhydride) and heavy (propionic anhydride-d10) labeling, obtaining normalized data representing the abundances of the peptides and histone marks. The QuantQC R package was then used to map the acquisition metadata of the liquid handler to the EpiProfile outputs in order to get the information of the peptides content and related abundance for each cell, then were filter out peptidoforms not consistently detected in at least 50% of the single cells, and excluded peptidoforms whose signal of the single cell intensity is not greater than the 150% signal in blanks. Data were subsequently normalized using a $\log(1+x)$ transformation to obtain the results table that is used for downstream analysis, which consists of a column with the various peptidoforms and columns for each individual cell with the abundances of the various identified peptidoforms.

Figure 10A shows a representative subset of the results table containing the first 20 histone H3 peptidoforms quantified. The values are expressed as relative abundances in control and sodium butyrate–treated conditions. This overview highlights how the workflow allows the detection of diverse PTM states at single-cell resolution, including methylation, acetylation, and combinatorial marks. For example, H3K9ac and H3K14ac both show increased relative abundance upon sodium butyrate treatment, consistent with the expected inhibition of histone deacetylases (HDACs). Similarly, histones co-modified with H3K9acK14ac appear even more enriched in the treated cells, reflecting the cooperative nature of acetylation marks in regulating chromatin accessibility.

Notably, results can be dissected into multiple layers. First, a good quality control is to ensure that an overall increase in total histone acetylation upon sodium butyrate exposure is observed (Figure 10B). Importantly, the treated cells display a broader error distribution compared to controls, reflecting the higher heterogeneity captured at the single-cell level. This illustrates how single-cell proteomics can reveal not only average shifts in PTM levels but also the variability across individual cells within a given population. Figure 7C aims to demonstrate that, when histones are cleaved into relatively long peptides, it is possible to quantify not only the relative abundance of single modifications, but also their co-occurrence frequency. In the example, the effect of sodium butyrate is mostly evident in the combinatorial acetylation state H3K9acK14ac, which is often associated with transcriptional activation. These patterns align with the effect of sodium butyrate and show the ability of our approach to detect biologically meaningful changes in histone PTMs at the single-cell scale.

To explore global differences, we performed PCA across all quantified peptidoforms (Figure 10D). Control and treated cells separated along the first principal component, which captured ~40% of the total variance. The wider dispersion of the sodium butyrate-treated cells indicates increased biological heterogeneity, consistent with variable acetylation responses among individual cells within the spheroids following HDAC inhibition.

These representative results demonstrate the efficiency and robustness of the workflow in capturing histone post-translational modifications at single-cell resolution and preserves information on cell-to-cell variability. By integrating sample preparation, acquisition, and computational analysis, this pipeline provides a practical and reproducible approach for profiling histone PTMs in single cells, making it suitable for broader applications in chromatin biology and drug-response studies.

DISCUSSION

The ability to measure histone post-translational modifications at the single-cell level represents a critical step toward understanding the true heterogeneity of chromatin regulation. Traditional bulk methods have provided invaluable insights into the average distribution of hPTMs across tissues,

yet they mask cell-to-cell variability that is increasingly recognized as biologically meaningful. Single-cell proteomic approaches now allow us to resolve this hidden layer of complexity, revealing distinct chromatin states that may underlie differential gene regulation, cellular plasticity, and disease vulnerability. Histone modifications are particularly well suited for single-cell interrogation because they encode combinatorial information that directs transcriptional outcomes. For example, simultaneous presence of H3K9ac and H3K14ac marks is a stronger indicator of transcriptional activation than a single one of these marks. Capturing such co-occurrence patterns at single-cell resolution is a new perspective in characterizing chromatin landscapes and their dynamics.

Despite these advantages, single-cell histone PTM analysis remains technically challenging. Limitations include extremely low protein input, the need for precise sample handling, and the difficulty of distinguishing technical from biological noise. Furthermore, while histone purification would reduce background interference, it is technically incompatible with single-cell sample amounts because acid extraction requires centrifugation. Our workflow therefore processes the entire cell lysate, which maximizes peptide recovery and minimizes loss during handling. Histones remain the dominant signal due to their high abundance and favorable behavior under propionylation, which selectively enhances detection of lysine-rich peptides. Non-histone proteins contribute minimally and, when detected, can provide additional internal references for normalization. Our workflow aims to simplify some of these challenges by providing a labeling technique to combine multiple conditions and an optimized protocol specifically dedicated to single-cell histone preparation. Isotopic 2-plex labeling was preferred compared to e.g. isotopic labeling such as TMT, because it ensures accurate quantification of histone peptides that often exist as isobaric positional isomers. TMT labeling would require complex MS³ acquisition of unique fragment ions to distinguish isobaric species, which is not applicable in a straightforward manner in regular mass spectrometers. Because desalting causes disproportionate peptide loss at single-cell input, we omit cleanup and rely on SpeedVac concentration, direct low-volume injection, and DIA-targeted extraction to preserve sensitivity while minimizing matrix interference.

Throughput, computational analysis of combinatorial PTMs, and integration with other single-cell omics remain areas for continued development. The push behind the development of these

methods is the fact that studying histone modifications at the single-cell level has important biological implications. Cell-to-cell variability in hPTMs can reveal transient cell states, lineage trajectories, or subpopulations that respond differently to stimuli or treatment. In the context of development, aging, or disease, this resolution may uncover epigenetic mechanisms that drive heterogeneity in neuronal and glial function, explain why certain cells are more vulnerable to degeneration, or identify epigenetic subtypes predictive of therapeutic response. Ultimately, single-cell histone proteomics provides a framework to bridge chromatin biology with systems-level understanding of complex tissues.

FIGURES

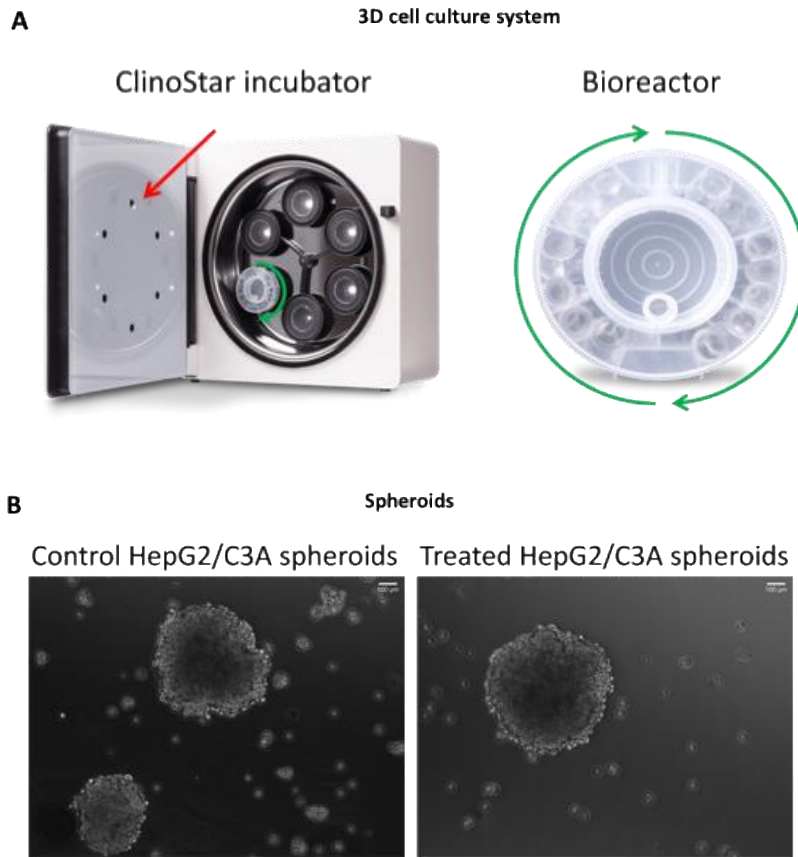


Figure 4: 3D culture system and spheroids. (A) 3D cell culture system. The bioreactor on the right is the device in which the cells are cultured. On the left is the incubator, which can hold up to 6 bioreactors and maintains a specific rotation speed, CO₂ level, and temperature. This figure has been modified from <https://celvivo.com/>. (B) HepG2/C3A spheroids during sodium butyrate treatment. The spheroids are cultured in 6-wells plate during the treatment with 5 mM of sodium butyrate for 24 h.

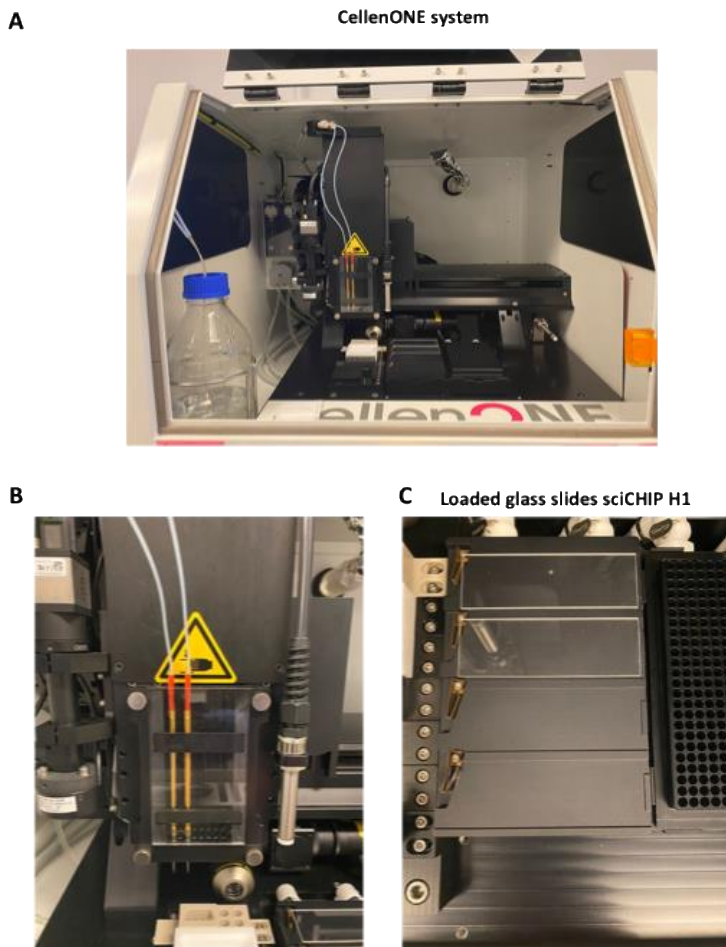


Figure 5: CellenONE instrument system. (A) Internal view of the instrument where sample preparation takes place. The figure shows the bottle containing milliQ water that feeds the instrument's liquid system, the nozzles involved in dispensing reagents and cells connected to the pumps, the glass slides loaded into the slide holder, the holder for the 384-well plate, and the cellenWASH station where the PCR tubes with reagents are placed. (B) PDC-70, coating-type 2 dispensing nozzles. The nozzles used during sample preparation, in positions 1 and 3. The nozzle in position 1 dispenses the cells, while the one in position 3 dispenses all the reagents used. The coated nozzles reduce cell and peptides sticking, useful for cell dispensing and optimal pickup. Moreover, give more reproducible droplet size. Compatibles with all the reagents used in this protocol. (C) Glass slides sciCHIP H1 loaded. In the picture there are two slides loaded in the slides holder that can hold up to 4 slides. The slides need to be very clean, flat and aligned with the top right corner.

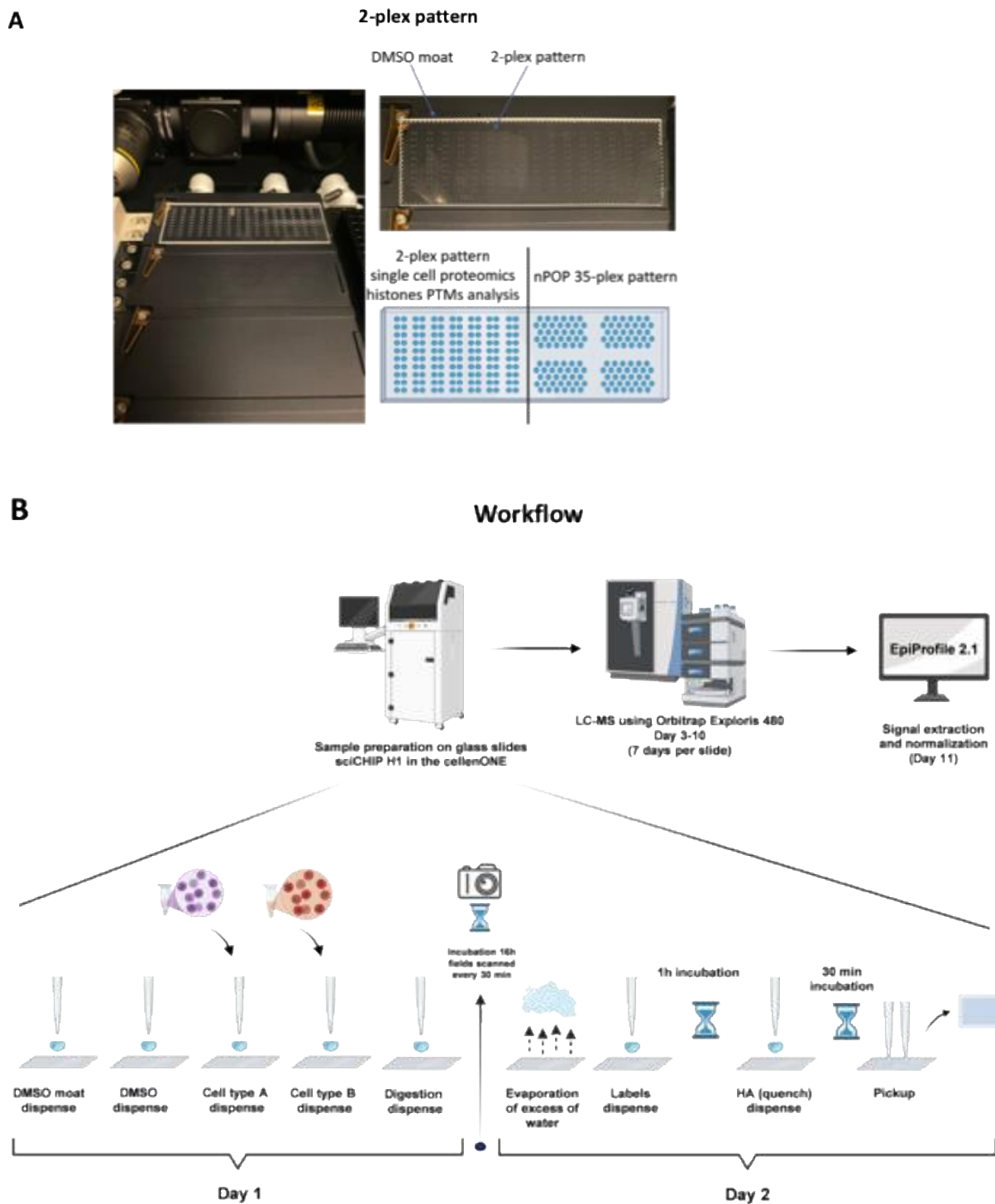
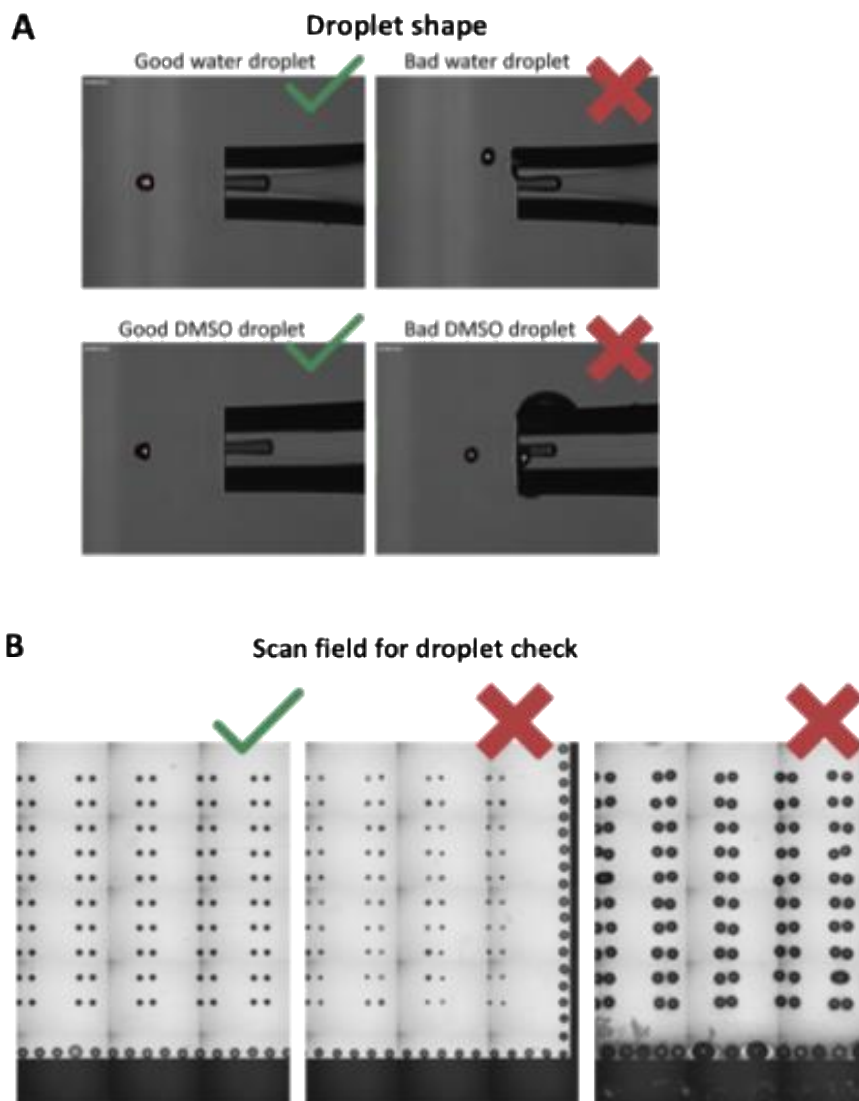


Figure 6: 2-plex pattern and workflow for sample preparation, data collection and data analysis. (A) Representation of the 2-plex pattern and DMSO moat. The figure shows the DMSO moat, which prevents excessive drying of the droplets during sample preparation, and the 2-plex

pattern following DMSO dispensing. It also shows the difference between the 2-plex pattern described in this protocol and what is expected in an nPOP experiment with TMT 35-plex labeling.

(B) Schematic overview of the experimental pipeline. Sample preparation takes place over two days: in the first part, DMSO, single cells, and the master mix for digestion are dispensed. The digestion is incubated overnight, during which photos of the droplets on the slides are taken every 30 min. In the second phase, after evaporation of any excess water that may have formed, labeling, quenching, and finally the pickup of the cell sets in a 384-well plate are performed. Processed samples are analyzed by mass spectrometry, and the resulting data are subjected to computational analysis to obtain quantitative profile of histones PTMs.



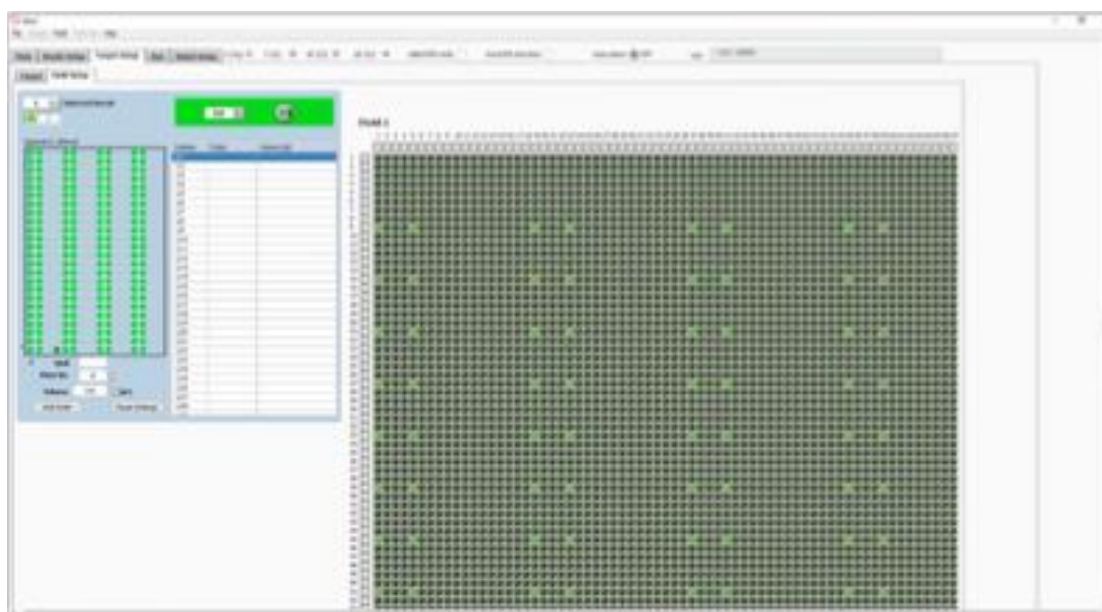
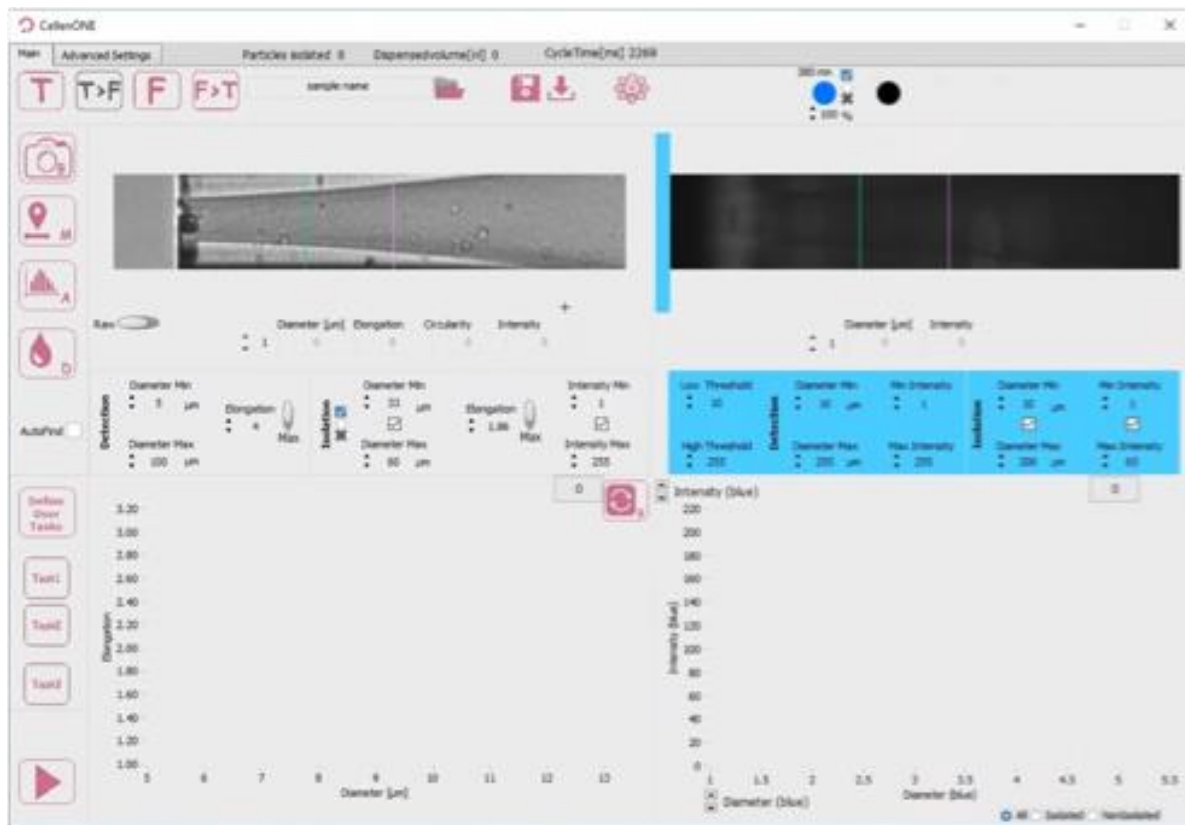
C**DMSO field file**

Figure 7: Accuracy of dispensed drops. (A) Images of autodrop test of the droplets. The first two images above show two autodrop test images in which the reagent is mainly composed of water. The first shows a correctly formed drop, while the second shows an incorrect drop that will need to be fixed before continuing with the preparation. The first two images below show two autodrop test images in which the reagent is mainly composed of DMSO. First, note how the DMSO appears transparent compared to water, and as in the previous images, you can see a correctly formed drop and an incorrect drop that will need to be corrected before continuing with the preparation. **(B)** Examples of fields scanned after dispensing. In the figure, the first image shows a successful dispensing in which the drops have the correct shape. The second shows drops that are too dry and therefore reduced in size, while the last shows drops that are too swollen and increased in size, where some have even merged. **(C)** Image of a field file used for DMSO dispensing. Representation of a file in which the position of the drops, the volume, the probe from which the reagent is dispensed, and the nozzle used are determined.

A

CellenONE cell dispenser viewer



B

Cell sorting nozzle view

HepG2 cell during the sorting



Figure 8: Types of visualizations during cell sorting. (A) CellenONE dispensing viewer. The figure shows the window that opens during cell sorting. Cells can be isolated based on transmission, which mainly determines size and elongation, while fluorescence is used to isolate live cells based on intensity following the use of dyes such as SytoxGreen, which tag dead cells.

The bottom part shows the cells plotted during analysis and isolation, depending on whether they are detected, isolated, or not. **(B)** Visualization of cell isolation based on transmission. The single cell is about to be isolated based on characteristics determined after mapping and analysis. Cells sticking to the glass tip outside the nozzle following sample aspiration are noted; these do not interfere with isolation as they are indicated as background.

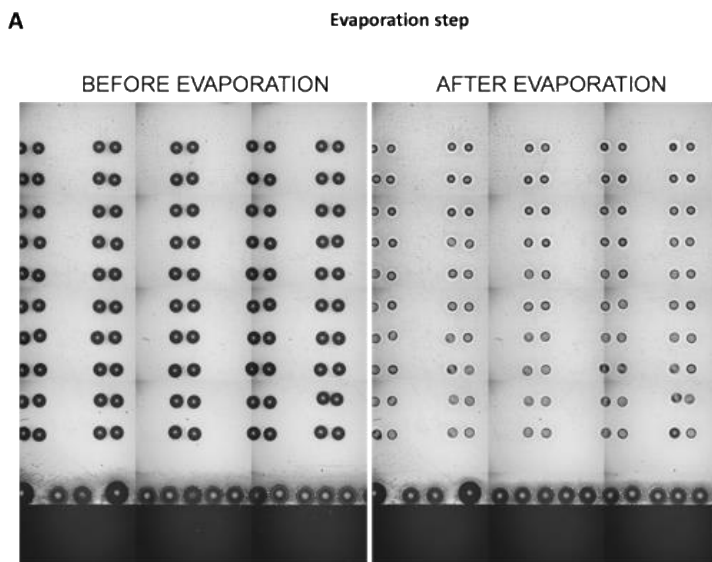


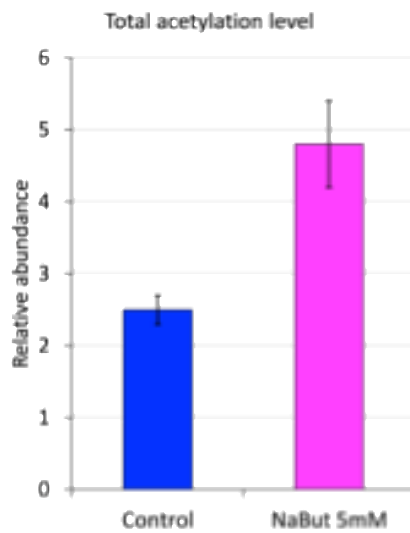
Figure 9: Evaporation step. The first image shows a scan field after overnight incubation of the digestion, in which the droplets are slightly swollen due to water collected from the humidity present in the system. The excess water is evaporated so that it does not interfere with the efficiency of labeling. The second image shows a scan field of droplets after the evaporation step, ready for the next labeling step.

A

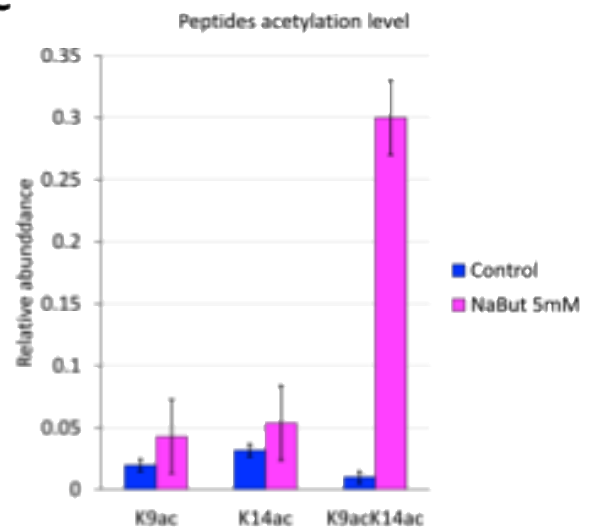
Results table

Peptide	Control	NaBut 5mM
H3_3_8 unmod	96.90%	96.20%
H3_3_8 K4me1	2.97%	3.51%
H3_3_8 K4me2	0.05%	0.18%
H3_3_8 K4me3	0.01%	0.05%
H3_3_8 K4ac	0.07%	0.06%
H3_9_17 unmod	14.38%	17.28%
H3_9_17 K9me1	5.60%	6.97%
H3_9_17 K9me2	23.02%	21.68%
H3_9_17 K9me3	19.65%	19.62%
H3_9_17 K9ac	2.21%	4.32%
H3_9_17 K14ac	3.23%	5.41%
H3_9_17 K9me1K14ac	4.27%	4.84%
H3_9_17 K9me2K14ac	15.28%	13.47%
H3_9_17 K9me3K14ac	7.33%	4.85%
H3_9_17 K9acK14ac	1.01%	30%
H3_18_26 unmod	71.09%	73.77%
H3_18_26 K23me1	0.09%	0.09%
H3_18_26 K18me1	0.35%	0.42%
H3_18_26 K18me1K23me1	0.00%	0.03%
H3_18_26 K18ac	2.26%	2.35%

B



C



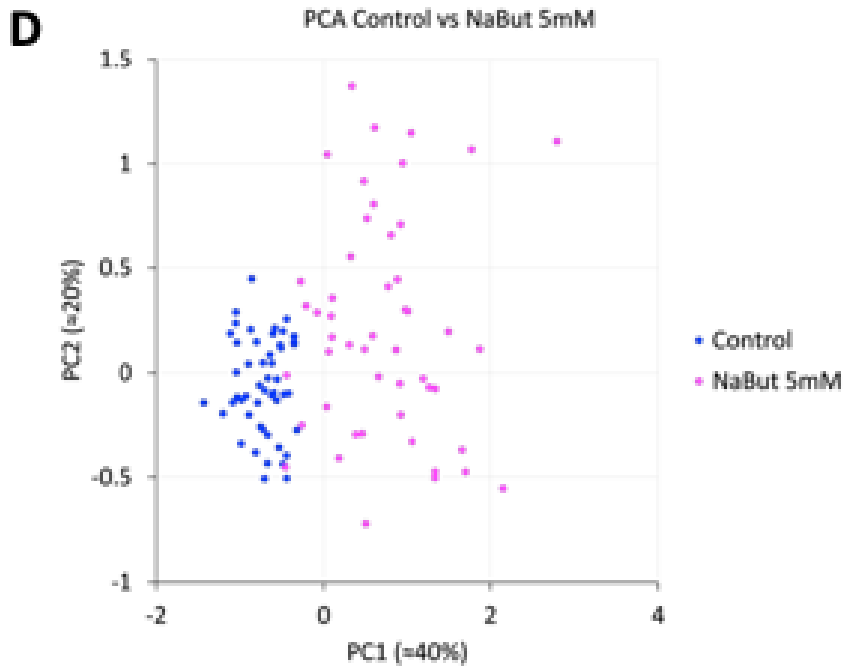


Figure 10: Data analysis of histones PTMs at a single cell level. (A) Table list of identified peptides with post-transcriptional modification sites and relative abundance of each of them in both conditions control and treated. In this results table every column corresponds to a cell. (B) Relative abundance of acetylation in the total of cells. (C) Relative abundance of acetylation for representative peptides of single and combinatorial modifications. (D) Principal component analysis (PCA) was performed to show the separation of the two cell groups and the heterogeneity of the treated cells.

CHAPTER 3: HYPERPLEXING: ACHIEVEMENTS AND CHALLENGES

While sensitivity and sample preparation represent two of the major challenges in single-cell proteomics, this research field is still constrained by throughput. Each individual cell represents an independent sample, and biological questions require the analysis of thousands of cells. Label free or limited multiplexed single-cell experiments demand LC-MS/MS acquisition, experimental time, instrument availability and cost that become prohibitive. In this context, hyperplexing is not just an improvement in but a necessity. This approach increases the number of samples that can be analyzed simultaneously within a single MS run, addressing the mismatch between the biological scale of single-cell studies and the limits of mass spectrometry acquisition. Moreover, combining multiple samples in a single run can enhance signal intensity and quantitative robustness. For these reasons, the development of hyperplex strategies represents a main requirement for scalable and biologically comprehensive single-cell proteomics studies.

3.1 Introduction: multiplexing labeling strategies background and aim of the project

Quantitative MS-based proteomics measures peptides derived from protein samples to determine how protein abundances change across different experimental conditions such as disease states, time points, drug perturbations, or biological replicates (36). The identification of proteins that vary in relative concentration provides valuable understanding of biological mechanisms, biomarker discovery, and systems regulation. However, achieving accurate and comprehensive quantification across large experimental cohorts remains the main challenge. For this reason, quantitative proteomics must be both high-throughput and robust, capable of handling hundreds of samples in parallel while maintaining quantitative accuracy (36).

This need is even more critical in single-cell proteomics, where each sample represents an individual cell, and experimental designs can involve thousands of injections. In single-cell proteomics throughput is one of the main bottlenecks, because the analysis speed of current mass spectrometers limits the number of samples that can be processed within a given time frame. Sample multiplexing provides a solution to address both quantification and throughput challenges. In the context of MS-based proteomics, multiplexing allows peptides from multiple samples to be combined and analyzed in a single LC-MS/MS run. This strategy not only enables relative quantitation across samples but also reduces the number of required injections, directly improving

experimental throughput. Moreover, by acquiring data from multiple samples in one analysis, multiplexing can enhance quantitative precision, lower technical variability, and reduce experimental costs (36).

Multiplexing two samples in one run halves the number of injections, expanding to dozens of samples can further accelerate data collection and facilitate large-scale cohort studies or time-course experiments (36) (Figure 11). These advantages make multiplexing not only a tool for quantitative accuracy but also a strategic solution to the throughput barrier in emerging single-cell proteomics workflows. Recent advances in multiplexing or hyperplexing have pushed the number of simultaneously analyzed samples from the early 6–10 channels to more than 50 in MS-based workflows (36).

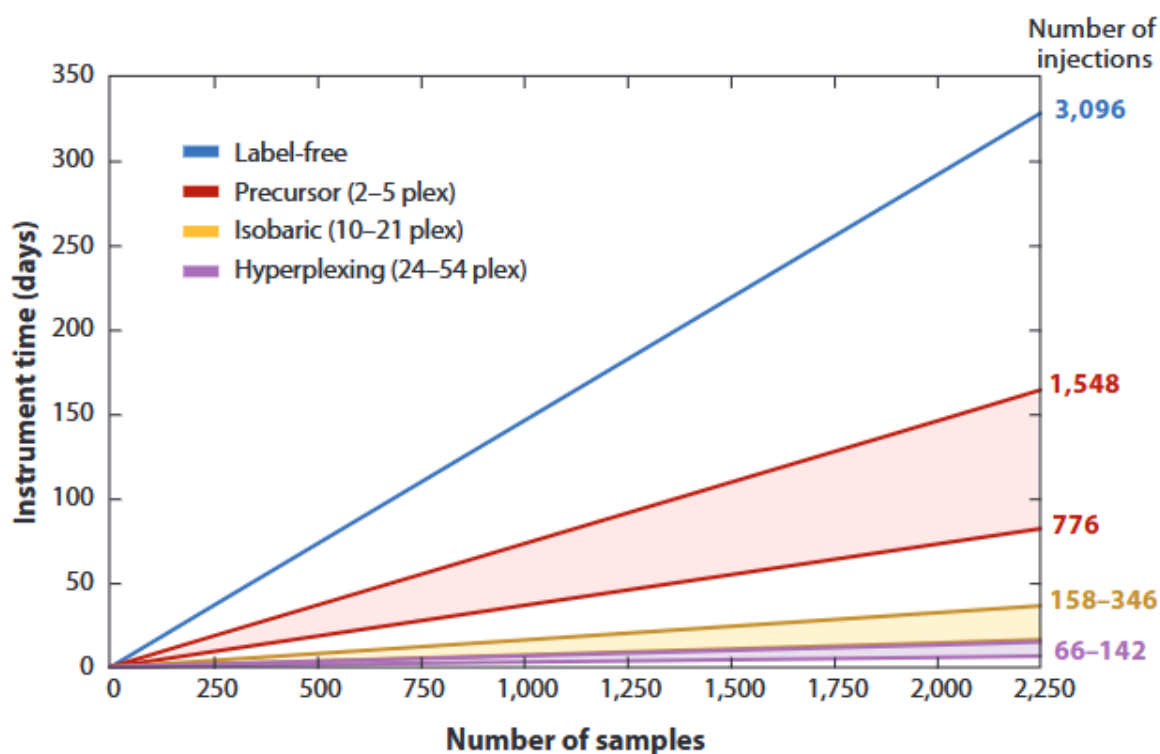


Figure 11: Amount of instrument time per number of samples (36).

Nevertheless, multiplexing introduces specific analytical and technical challenges, including ratio compression, co-isolation interference, and variability in sample preparation and acquisition

strategies (36). Addressing these limitations requires innovative labeling chemistry and optimized workflows that balance throughput with data quality. Among the available labeling approaches, isobaric tags such as Tandem Mass Tags (TMT) and isotopic labeling strategies such as dimethyl labeling have emerged as powerful and complementary techniques. In this work, I sought to combine the strengths of these two labeling strategies to develop an improved hyperplex workflow capable of addressing the throughput demands of single-cell proteomics.

TMTs are isobaric chemical labels used to enable multiplexed quantitative proteomics. Each tag has the same intact mass (isobaric) but releases reporter ions of different masses upon fragmentation, allowing multiple samples to be quantified in a single MS run. A TMT molecule is composed of three main parts: a reporter group, a balancer and a reactive group (Figure 12). At the MS1 level all tagged peptides appear as a single peak, but after fragmentation the TMT tags break at a specific site, releasing reporter ions of different masses corresponding to each sample. The intensity of these reporter ions reflects the relative abundance of the peptide across the multiplexed samples, enabling accurate quantitative comparison. Sample preparation is very simple and straightforward. After digestion, the peptides from different samples are labeled with different TMT tags, incubated for one hour, and the reaction is quenched with hydroxylamine to prevent cross-labeling. They are then pooled together into a single sample that will be analyzed by mass spectrometry.

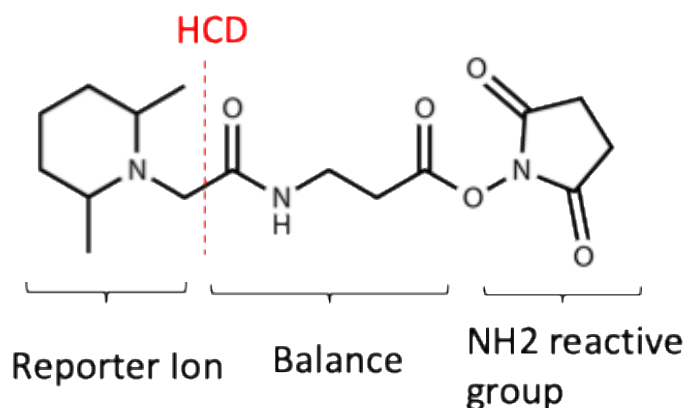


Figure 12: TMTs chemical structure.

Dimethyl labeling is a chemical isotopic labeling technique used for relative quantification in proteomics. It is based on the reductive amination of peptide primary amines which is a simple, low-cost, and efficient approach that introduces stable isotopic labels. Dimethylation modifies peptides by adding two methyl groups ($-\text{CH}_3$) to each primary amine through a Schiff base reaction between formaldehyde and the amine, followed by reduction with a cyanoborohydride. The modification converts each amine into a dimethylated tertiary amine, increasing its mass depending on the isotopic composition of the reagents used. Like TMT, dimethylation targets primary amines, labeling both the peptide N-terminus and the ϵ -amino group of lysine residues. Dimethyl labeling can generate from three isotopically distinct forms to five, but the one that I used in this work is the 3-plex one which allows up to triplex quantification in one LC-MS run (Figure 13): (a) light label uses regular formaldehyde (CH_2O) and sodium cyanoborohydride (NaBH_3CN); (b) intermediate label uses deuterated formaldehyde (CD_2O) and NaBH_3CN and (c) heavy label uses deuterated formaldehyde (CD_2O) and sodium cyanoborodeuteride (NaBD_3CN). Each version adds a distinct mass shift (+28.0313 Da for light, +32.0564 Da for intermediate, and +36.0757 Da for heavy per labeled site), which can be resolved by high-resolution mass spectrometry. At the MS1 level the labeled peptides appear as distinct peaks separated by the isotopic mass differences and the quantification is achieved by comparing the intensity of these isotopically resolved precursor ions. Sample preparation is very similar to that of TMT. After digestion, the peptides from different samples are labeled with different tags, incubated for one hour, and the reaction is quenched with ammonia solution to prevent cross-labeling. They are then combined into a single sample that will be analyzed by mass spectrometry.

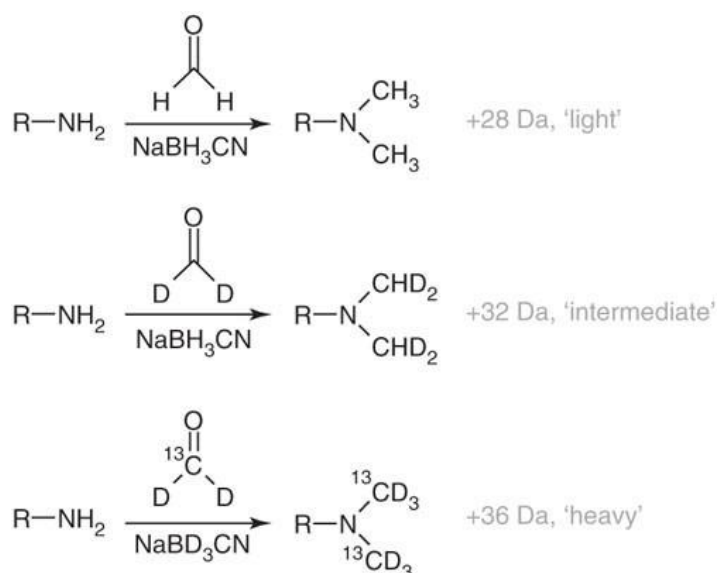


Figure 13: Combinations of isotope-labeled formaldehyde and cyanoborohydride to create dimethyl labels.

To overcome the current throughput limitations in single-cell proteomics, I aimed to develop a new hyperplex workflow that combines the advantages of both isobaric and isotopic labeling strategies. The rationale behind this approach was to expand multiplexing capacity beyond what can be achieved with either technique alone, while maintaining quantitative accuracy, reproducibility, and scalability. In this design, I employed TMTpro 35-plex reagents to enable high-level isobaric multiplexing within a single LC-MS/MS run, and to further enhance throughput, each TMT channel was dimethyl-labeled in three isotopic variants (light, intermediate, and heavy), providing an additional 3-fold multiplexing layer. This combination results in a theoretical 96-plex capacity (32×3) per experiment that not only increases the analytical throughput but also boosts the overall peptide signal intensity that is beneficial for single-cell experiments, where the total of protein content is very low. All the development and optimization steps for this approach were carried out using bulk proteomic samples, i.e. HeLa cells lysate. Working in bulk allowed me to systematically test labeling efficiency, mixing accuracy, quantitative performance, and data analysis strategies under controlled conditions. This step was essential to identify the most robust and reproducible workflow before adapting and scaling it for single-cell proteomics applications, where sample amounts and sensitivity requirements impose additional constraints.

Combining two labeling strategies that rely on different chemistries but share the same molecular targets presents significant experimental and analytical challenges. Both dimethylation and TMT labeling react with primary amines, the peptide N-terminus and the ϵ -amino group of lysine residues. When applied sequentially or in combination, this overlap can lead to competition between reagents, incomplete labeling, or undesired cross-reactions, ultimately affecting quantification accuracy and peptide identification. To address these limitations, the strategy was designed to achieve site-specific labeling, assigning each chemistry to a distinct reactive site. In this approach, dimethylation is directed exclusively to peptide N-termini, while TMT labeling is restricted to lysine side chains (Figure 14).



Figure 14: Representation of an ideal labeled peptide after site-specific labeling.

3.2 Feasibility tests to combine two labeling strategies with different chemistry

As an initial approach, we tested a workflow based on the standard dimethylation protocol, followed by digestion with trypsin and subsequent TMT labeling. This sequence was chosen because dimethylation is typically performed efficiently at the peptide level under mild conditions, and mainly we could have specifically labeled N-termini with dimethyl labels, and after digestion the lysine residues would have been available for reaction with TMT. However, when combined with TMT chemistry, this strategy proved unsuccessful for several reasons. First, we observed undesired cross-reactions between the TMT NHS-ester and the residual reducing agents or catalysts used in the dimethylation step. These chemical interferences not only compromised the efficiency of TMT labeling but also introduced background signals that masked peptide peaks, making it more difficult to detect and identify peptides during MS analysis. In addition, performing dimethylation first resulted in blocking of lysine ϵ -amino groups, which are the main reactive sites

for TMT tagging. Consequently, the overall TMT labeling efficiency decreased, leading to reduced proteome coverage and weaker quantitative performance. This first attempt clearly demonstrated that reaction compatibility and order are critical when combining two chemistries that target the same functional groups. The experiment revealed that simply applying existing labeling protocols in sequence is not sufficient, as chemical cross-reactivity and site blocking can significantly affect both labeling completeness and signal quality. Therefore, a revised, site-specific strategy was needed to spatially and temporally separate the two chemistries, ensuring that each reagent reacts exclusively with its intended target.

To verify whether a selective dual-labeling strategy was in principle feasible, we explored existing protocols that combined different labeling chemistries. Among these, we identified the cPILOT (combined Precursor Isotopic Labeling and Isobaric Tagging) workflow as a relevant example (69). In this method, light and heavy dimethylation and TMT labeling are achieved by carefully controlling reaction conditions and, importantly, by including a desalting step between the two labeling strategies and enzymatic digestion prior to labeling (69). This approach effectively separates the two labeling steps, preventing cross-reactions and allowing efficient tagging at both the N-terminus and lysine residues. We adapted and tested the cPILOT protocol under bulk conditions, and the results confirmed that the dual labeling concept works in practice. Labeling efficiency was around 70% at the N-termini for dimethylation and a 13% off-target labeling on lysine residues (Figure 15), and with 7,000 peptides identified. In this specific experiment, labeling with TMT showed 1% efficiency at both binding sites, because the reagents we were using had accidentally undergone hydrolysis and were no longer functional. However, the fact that we were able to identify and quantify labeled peptides and calculate labeling efficiency demonstrates that a site-specific combination of dimethylation and TMT is chemically possible.

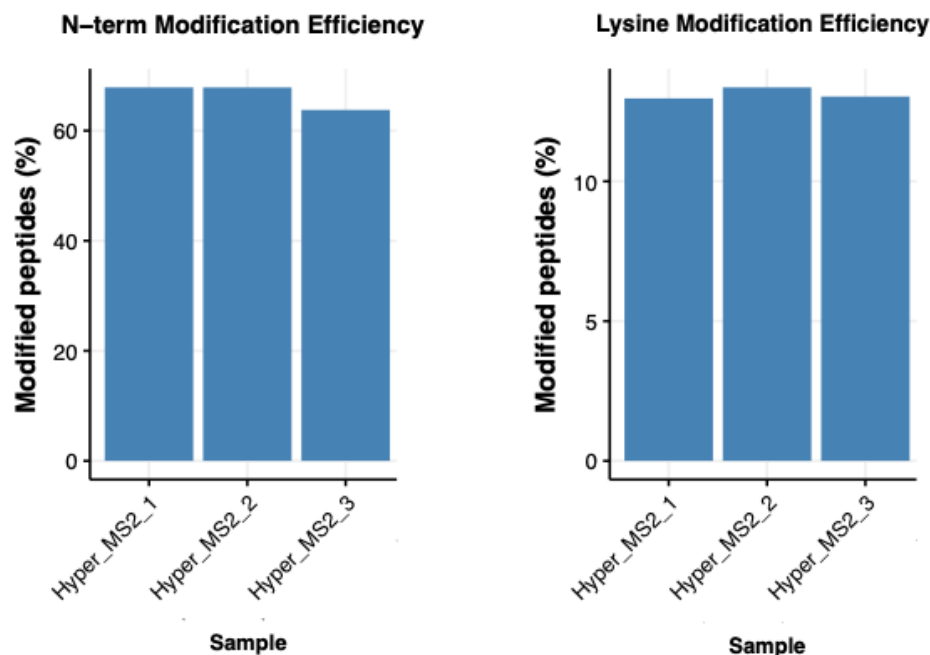


Figure 15: Dimethylation labeling efficiency.

While this validation was essential to prove the feasibility of the concept, the cPILOT workflow is not directly compatible with single-cell proteomics, especially in our setup based on the cellenONE instrument. The cellenONE system performs fully automated single-cell isolation and sample preparation at nanoliter volumes, using glass slides as the working surface for reagent dispensing, incubation, and transfer. All steps from lysis to digestion and labeling occur in confined droplets on these slides, with no intermediate handling possible. Because of that, off-line cleanup or desalting steps cannot be performed, as even minimal sample loss or droplet disruption would compromise recovery from single cells. Therefore, while the cPILOT approach confirmed that the chemistry itself can work, we needed to design a new workflow compatible with the physical and chemical constraints of the cellenONE, maintaining reaction specificity while eliminating any steps that require bulk handling or sample transfer.

Since desalting or cleanup steps were not an option, we began to reconsider the chemistry of the workflow itself. To overcome this, we searched for an alternative dimethylation chemistry that could achieve the same modification of primary amines but without relying on a reducing catalyst. During this search, we identified the Eschweiler–Clarke reaction as a promising substitute (Figure

16). This reaction performs the methylation of primary and secondary amines using formaldehyde and formic acid, where the formic acid acts both as the reducing agent and proton source, eliminating the need for cyanoborohydride or other additives (70). Mechanistically, formaldehyde first reacts with the amine to form an iminium intermediate, which is then reduced in situ by formic acid, yielding the dimethylated amine and releasing carbon dioxide (70). By replacing the traditional catalyst with formic acid, the Eschweiler–Clarke reaction aims to provide a cleaner and more MS-compatible chemistry, minimizing interference with TMT reagents. An additional advantage of this reaction is the low pH provided by formic acid, which creates a chemical environment favorable for selective labeling of the N-terminus over lysine side chains. This property aligned well with our goal of achieving site-specific dimethylation, as the lower pH would help exploit the natural difference in pKa between the two amine groups.

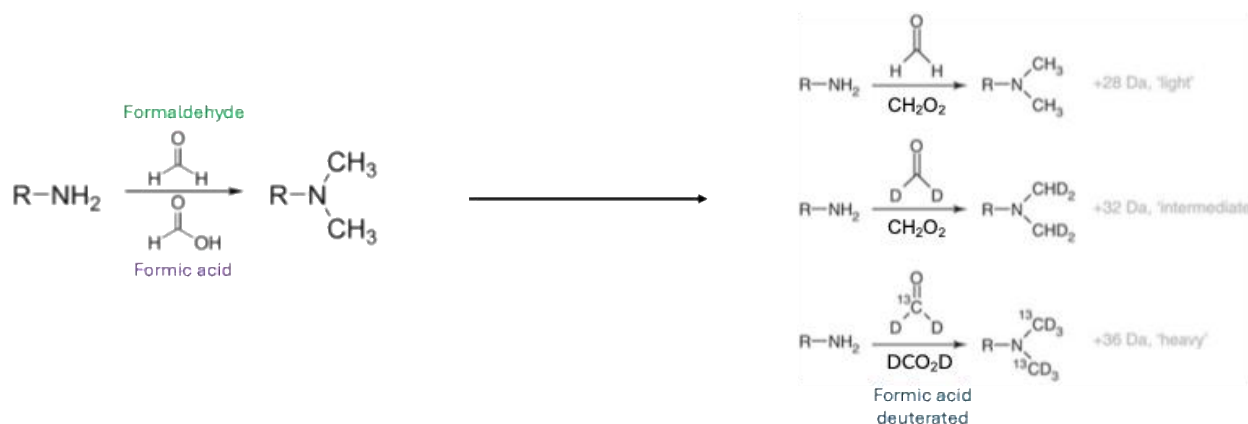


Figure 16: Eschweiler–Clarke reaction and labels resulting from the reaction between formic acid and formaldehyde.

However, the main limitation of the Eschweiler–Clarke reaction is that it is performed at near-boiling temperatures (around 95–100 °C) to drive the reduction step. Such conditions are incompatible with the cellenONE system, where it is not possible to heat a reaction and therefore work must be carried out at room temperature. Therefore, our next step was to adapt and optimize the Eschweiler–Clarke chemistry to function under milder conditions, testing whether efficient dimethylation could be achieved without heating, while preserving site selectivity and maintaining compatibility with downstream TMT labeling and single-cell sample handling.

As a first test, we evaluated whether the Eschweiler–Clarke reaction could effectively label peptides under different temperature conditions. We selected four incubation temperatures 25°C, 37°C, 56°C, and 95°C representing a range commonly used in peptide processing workflows. The rationale behind this design was to determine whether temperature had a significant influence on labeling efficiency, as the original Eschweiler–Clarke reaction requires near-boiling conditions, and to assess whether longer incubation (4 hours) could compensate for the lower reaction rates expected at milder temperatures. For comparison, we included a control reaction using the conventional dimethylation protocol with formaldehyde and sodium cyanoborohydride as the catalyst. Each condition was tested in triplicate using HeLa cell lysate to ensure reproducibility and provide a robust assessment of labeling performance.

The results show that the control dimethylation achieved approximately 80% labeling efficiency, consistent with expected performance for this reaction. In contrast, all formic acid–based conditions resulted in significantly lower labeling efficiencies, reaching only 10–15% modification at both the N-termini and lysine residues, regardless of temperature (Figure 17).

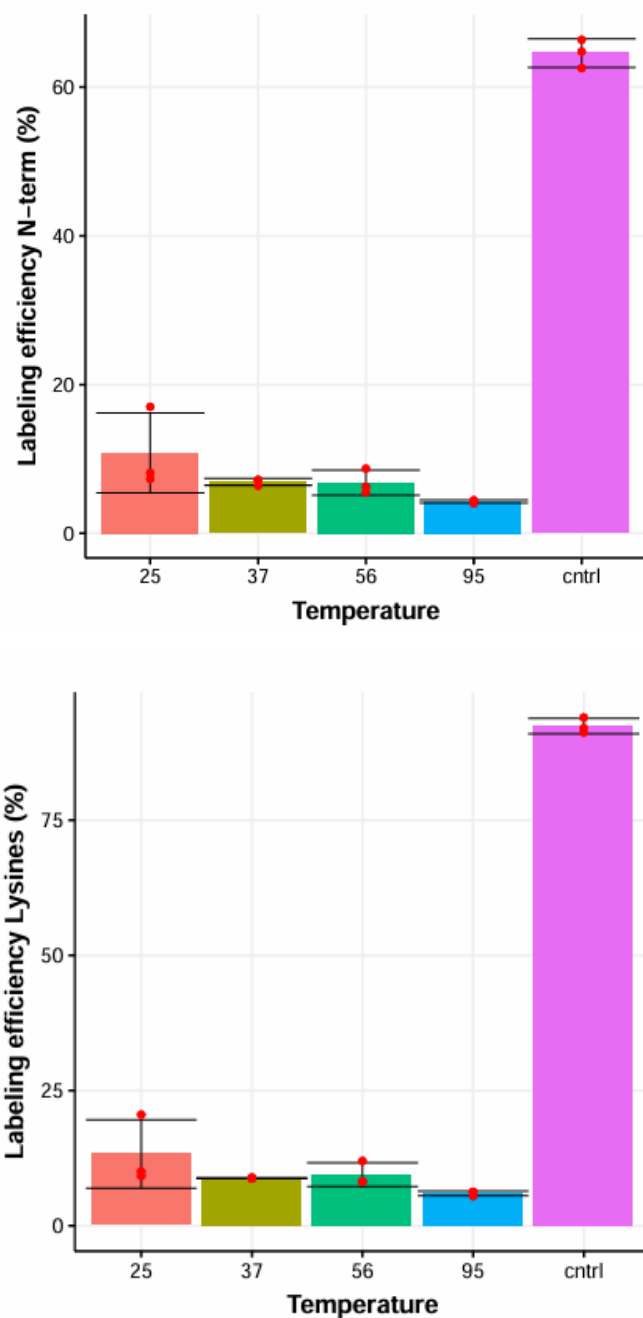


Figure 17: Labeling efficiency on N-termini and lysines after dimethylation with formic acid and formaldehyde at different temperatures 25°C, 37°C, 56°C, and 95°C and 4h incubation.

Based on these initial results, we hypothesized that the extended 4-hour incubation under strongly acidic conditions might have negatively affected the reaction efficiency or peptide stability. To test this possibility, we performed a second set of experiments using shorter incubation times of

15 minutes and 1 hour at room temperature (25°C). The 15-minute time point was chosen as the minimum duration required for the Eschweiler–Clarke reaction to occur, while the 1-hour condition was included to evaluate whether a slightly longer reaction could improve labeling yield. We focused on room temperature since the previous tests had shown no clear temperature-dependent improvement in efficiency, and because it is the only temperature we can use once the protocol is transferred for use in single cells with cellenONE system. In addition, we evaluated the effect of quenching by comparing reactions stopped with TEAB buffer to those without quenching, to determine whether residual acidity or reaction carryover could be influencing labeling outcomes. Each condition was again tested in triplicate using HeLa cell lysate.

The results showed no substantial differences between the tested incubation times or between quenched and unquenched samples. In all cases, the labeling efficiency remained around 20% for both N-terminal and lysine modifications, confirming that the Eschweiler–Clarke reaction does not proceed efficiently at low temperatures under our experimental conditions (Figure 18).

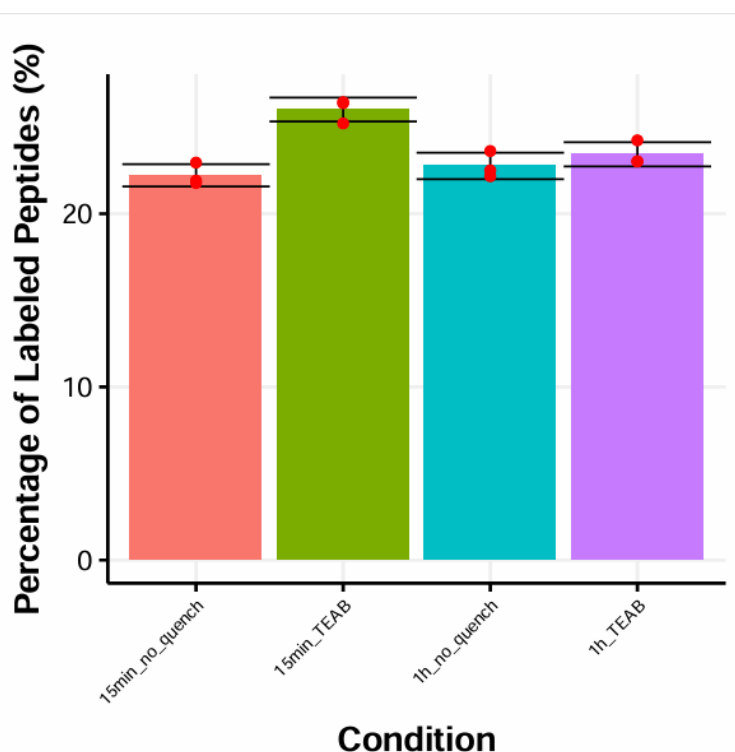


Figure 18: Second experiment dimethylation with formic acid and formaldehyde at room temperature, testing 15 minutes and 1 hour incubation time, with and without quenching.

From these observations, we concluded that although the Eschweiler–Clarke chemistry offers theoretical advantages in terms of catalyst-free and low-pH selectivity, it cannot be effectively adapted to mild or room-temperature conditions, as required for compatibility with the cellenONE platform. Therefore, it became clear that we needed to reconsider our strategy and explore alternative chemical or procedural solutions to achieve selective and efficient labeling suitable for single-cell proteomics.

After confirming that the Eschweiler–Clarke approach could not be adapted to low-temperature conditions, we decided to try a more unconventional strategy. The idea was to invert the labeling order, performing TMT labeling first under modified conditions, followed by dimethylation. Specifically, we hypothesized that if the TMT reaction could be carried out under acidic conditions, it might preferentially target the N-terminal amines, while keeping the lysine side chains less reactive. The TMT reagent could be therefore resuspended in formic acid to maintain a low-pH environment, and labeling attempted directly under these conditions. After this first reaction, we planned to quench the residual TMT reagent using a basic buffer to raise the pH and then proceed with dimethylation to target the lysine residues. The rationale behind this approach was that quenching the TMT reagent first would deactivate any remaining NHS-ester groups, thereby preventing interactions or cross-reactions during the subsequent dimethylation step.

First, we tested TMT labeling in acidic conditions in duplicate using HeLa cell lysate. The results showed that the reaction was ineffective. Labeling efficiency remained below 1% at both the N-termini and lysine residues, indicating that the NHS-ester chemistry of TMT is not compatible with low-pH environments. Under such conditions, the reactive ester is rapidly hydrolyzed, and the amine groups are protonated, preventing nucleophilic attack and formation of the covalent bond (Figure 19).

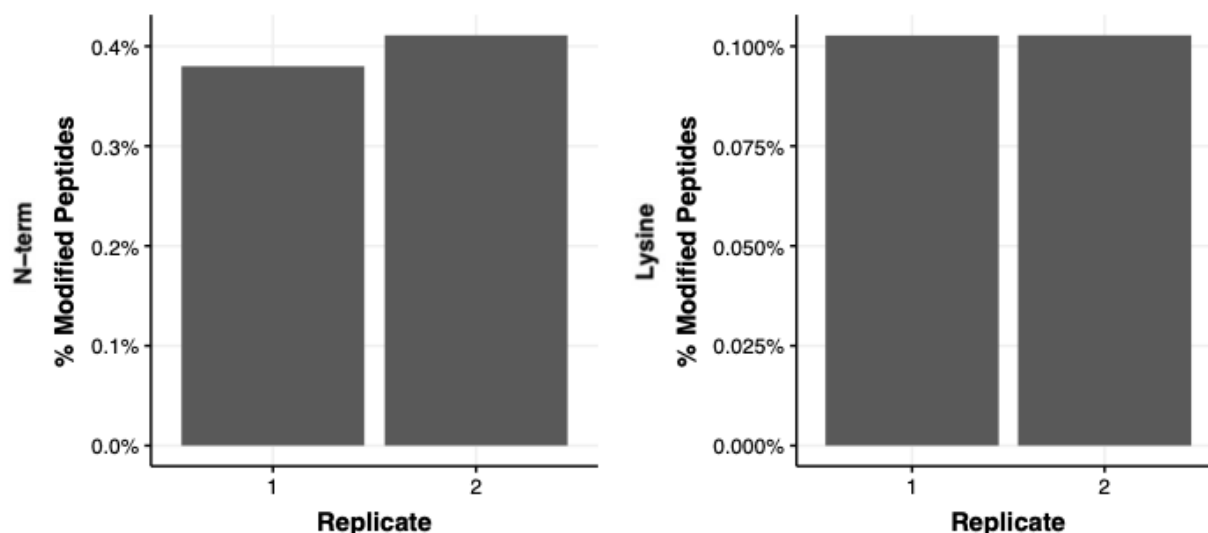


Figure 19: TMT labeling in acidic conditions. Labeling efficiency on N-termini and lysines residues.

These findings demonstrated that the TMT labeling step requires a neutral to slightly basic environment to proceed efficiently and therefore cannot be adapted for selective N-terminal labeling under acidic conditions.

At this stage, it became clear that the chemistry itself could not be fundamentally modified. Both TMT and dimethylation rely on well-established amine-reactive mechanisms that require specific pH and chemical environments. Therefore, instead of altering the reagents, we shifted our focus toward finding a way to neutralize either the TMT reagent or the dimethylation catalyst to prevent their mutual interference while maintaining reaction efficiency.

3.3 Key parameters for developing a dual-labeling protocol

After an extensive literature search, we identified the NETLOP (NHS-ester tandem labeling in one pot) workflow as a potentially suitable strategy (71). In this method, the TMT labeling is performed first, and the reaction mixture is incubated for several hours (around 4 hours) to promote the complete hydrolysis of any remaining TMT reagent, thereby eliminating its reactivity before proceeding with a secondary isotopic labeling step (71). This concept was particularly relevant to our work because it provided a way to inactivate the TMT reagent without altering its labeling

performance, reducing the risk of cross-reactions during subsequent labeling steps. We therefore adapted the NETLOP approach to our own experimental conditions, testing whether a similar controlled hydrolysis of TMT could allow a subsequent dimethylation reaction without interference. By this point, we were fully aware of the complexities and limitations involved in developing such a dual-labeling protocol, and we approached the optimization systematically, focusing on the key chemical parameters of both reactions:

(a) pKa of the targets and pH of the reactions:

The interplay between pKa and pH determines whether an amino group is protonated or deprotonated, which directly affects its nucleophilicity and its ability to react with labeling reagents. In peptides, the N-terminal amine typically has a pKa of $\sim 7-8$, while the ϵ -amino group of lysine is more basic, with a pKa of ~ 10.5 . This means that at a slightly acidic pH ($\sim 6-7$), most N-terminal amines are partially deprotonated and therefore reactive, whereas lysine side chains remain largely protonated and less nucleophilic. Conversely, at basic pH ($\sim 8.5-9.5$), lysine side chains become progressively deprotonated, allowing the TMT to react efficiently. This difference provides a way to control labeling selectivity. For example, performing the dimethylation reaction at slightly acidic pH favors N-terminal modification, as lysines are mostly protonated and less accessible. In contrast, maintaining a basic environment is essential for TMT labeling, since its NHS-ester reacts only with unprotonated amines capable of performing nucleophilic attack on the carbonyl carbon of the ester group. Furthermore, the protonation state also influences the stability of the labeling reagents themselves. At low pH, the NHS-ester group of TMT undergoes hydrolysis, which inactivates the reagent, while at high pH, amine deprotonation enhances reactivity but also increases the risk of side reactions or peptide backbone instability. Therefore, identifying the narrow pH window that ensures both reactivity and chemical stability was essential for combining the two chemistries.

(b) Protein-to-label ratio:

The stoichiometry between protein and labeling reagent determines both completeness and reproducibility. Insufficient reagent results in partial labeling, whereas excessive reagent increases

the risk of side reactions and background noise. Finding the optimal balance ensures efficient labeling while minimizing reagent waste and downstream variability.

(c) Incubation time:

Reaction duration is important to balance completeness and specificity. Short incubations lead to incomplete reactions, while prolonged exposure may cause hydrolysis or over labeling. In the NETLOP-like approach, the incubation time also determines the degree of TMT hydrolysis, making it a critical factor for compatibility between the two labeling steps (71).

(d) Quenching of the reaction:

Finally, appropriate quenching is crucial to stop residual reactivity before the next step. For dimethylation, removal or neutralization of residual formaldehyde and reducing agent prevents cross-labeling and interference. While TMT, in this specific case, is self-inactivated through hydrolysis during incubation.

Another important factor to consider in developing this hyperplex workflow compatible with single-cell proteomics is the presence of DMSO (dimethyl sulfoxide). In our setup, the presence of DMSO is essential, as it plays a crucial role in single-cell sample preparation on the cellenONE platform. DMSO is used as a carrier solvent during cell dispensing and lysis because it helps maintain droplet stability, ensures accurate cell transfer, and facilitates efficient protein solubilization at nanoliter volumes on glass slides.

However, the use of DMSO introduces an additional chemical constraint for labeling reactions. DMSO is a polar aprotic solvent that can significantly influence both reagent solubility and reaction kinetics (72–74). In the case of amine-reactive labeling, DMSO may affect the protonation equilibrium of amine groups and alter the reactivity of electrophilic reagents such as the NHS-ester of TMT or the formaldehyde used in dimethylation. Furthermore, DMSO can compete with water as a solvent, modifying the microenvironmental pH and potentially stabilizing reactive intermediates or delaying hydrolysis processes (72–74). These effects can be positive or negative depending on the reaction. For instance, DMSO can enhance peptide solubility and reagent

dispersion, improving local reaction efficiency, but it can also slow down hydrolysis or reduce nucleophilicity of amines by changing solvent polarity (72–74).

In our case, this meant that every optimization step, particularly those involving pH, incubation time, and reagent concentration, had to be carefully re-evaluated in the presence of DMSO to ensure that both TMT and dimethylation chemistries remained effective under the exact solvent conditions required for single-cell experiments. Therefore, the impact of DMSO added another layer of complexity to the workflow design, as we needed to balance solvent compatibility for cellenONE handling with chemical reactivity and stability for the labeling reactions.

To evaluate whether the presence of DMSO could interfere with amine-labeling reactions, we tested its impact on the standard dimethylation protocol under different solvent percentages. Peptide samples from HeLa cell lysate were resuspended in increasing proportions of DMSO (0%, 25%, 50%, 75%, and 100%) and subjected to the conventional dimethylation reaction using formaldehyde and sodium cyanoborohydride. Each condition was performed in triplicate to ensure reproducibility and to quantify the effect of solvent composition on labeling efficiency.

The results show that dimethylation efficiency remained consistently high across all DMSO concentrations, reaching nearly 100% labeling at both the N-termini and lysine residues in every tested condition. These findings indicated that DMSO does not significantly affect the performance of the dimethylation reaction, even when used as the primary solvent (Figure 20). This robustness is likely due to the fact that reductive amination in dimethylation relies on the formation of iminium intermediates and subsequent reduction, processes that are not substantially altered by the polar aprotic nature of DMSO.

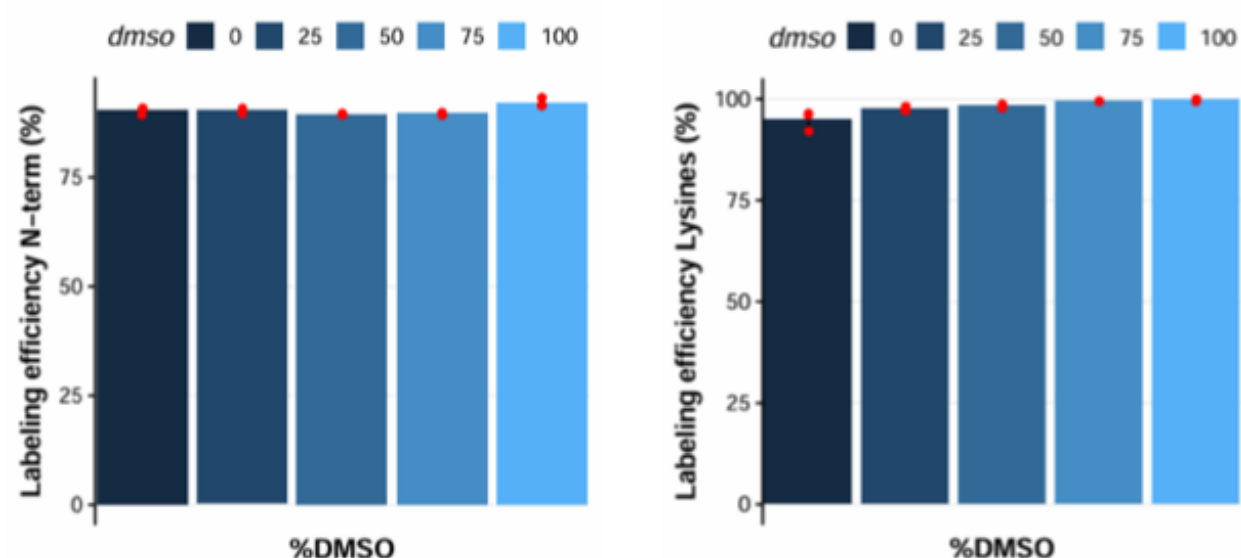


Figure 20: Labeling efficiency at N-termini and lysine residues of dimethylation at different percentages of DMSO.

From this experiment, we concluded that DMSO is fully compatible with dimethylation chemistry, and that its presence, required for single-cell sample preparation on the cellenONE platform, would not represent a limiting factor for implementing this reaction in our final workflow. Since we already had extensive experience with TMT 35-plex labeling in single-cell proteomics and knew from prior work that TMT reagents perform efficiently in the presence of DMSO, we did not conduct additional tests for the TMT reaction under these conditions. Therefore, we concluded that both labeling chemistries are compatible with DMSO.

After establishing that DMSO did not negatively affect dimethylation and was compatible with our workflow requirements, we proceeded to perform a scaled experiment in bulk under the exact conditions planned for the single-cell setup on the cellenONE platform. The goal was to validate the chemical compatibility of the final workflow before miniaturizing it to the single-cell format.

For this test, we used HeLa cell lysate prepared in triplicate, following a protocol that reproduced all the relevant parameters of the single-cell conditions. The workflow included the use of DMSO at the same concentration expected in the cellenONE droplets, a TMT labeling reaction at pH 9.5, and a dimethylation reaction at pH 6.8. The protein-to-TMT ratio was set at 1:5, while the

dimethylation was performed using its standard stoichiometry. Following the NETLOP-inspired approach, the TMT labeling was incubated for 4 hours to allow sufficient time for both labeling and hydrolysis of unreacted TMT, after which the dimethylation step was carried out under the classical conditions.

The results were unexpected. The digestion control confirmed that the samples were intact but showed no detectable labeling, indicating that the labeling steps, rather than digestion or sample handling, were responsible for the low performance. When analyzed site-specifically, lysine residues showed approximately 90% TMT labeling efficiency with less than 4% dimethylation, suggesting that the TMT reaction was dominant at this site. Conversely, at the N-termini, only about 6% dimethylation was detected, while TMT labeling reached roughly 80%, demonstrating that TMT was also reacting where dimethylation was expected to occur (Figure 21).

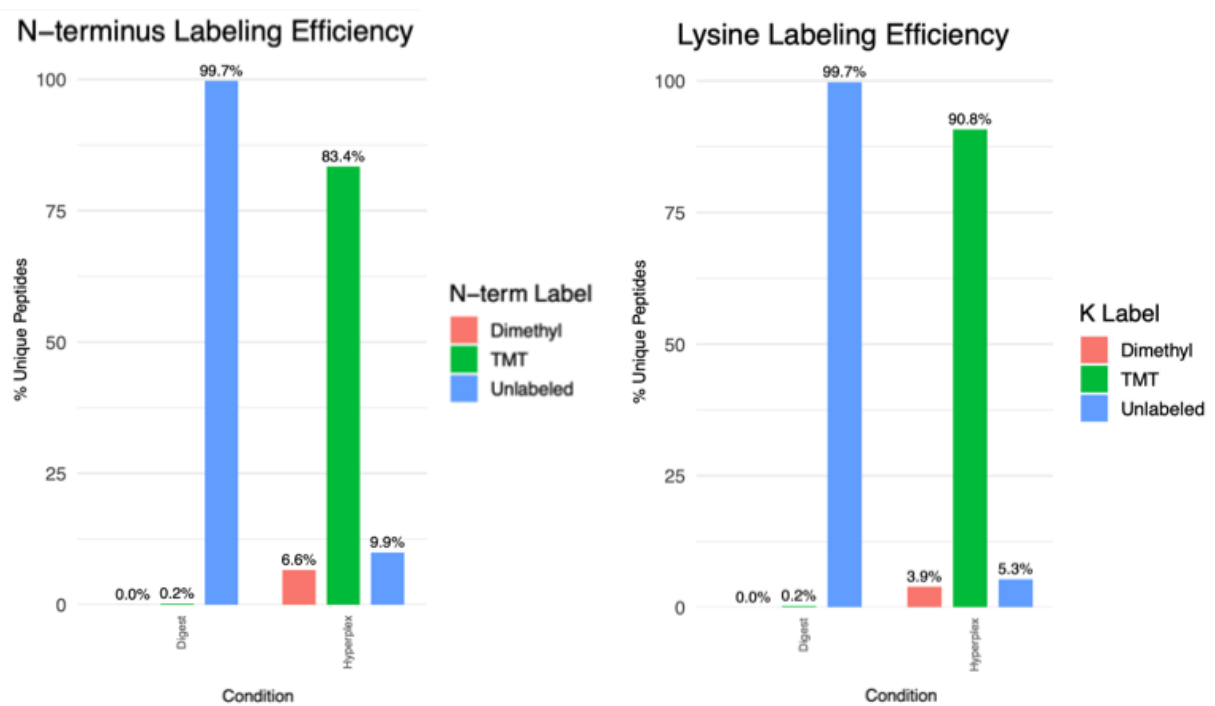


Figure 21: Labeling efficiency at N-termini and lysines of scaled Hyperplex protocol.

These results indicated that the intended site-specific labeling was not achieved. Several factors could explain this outcome. It is possible that DMSO influenced the chemical equilibrium, affecting reagent reactivity or hydrolysis rates. Alternatively, the excess of TMT reagent (1:5 ratio)

may have led to over labeling and saturation of both reactive sites. Another possibility is that the TMT reagent did not hydrolyze completely during the 4-hour incubation, leaving residual active ester groups capable of reacting during or after the dimethylation step. At this stage, while the labeling reactions were clearly active, the specificity and selectivity of the dual-labeling process were not achieved, suggesting that further optimization would be required.

3.4 First single-cell experiment to test the protocol in yielding meaningful biological results

In parallel with the bulk optimization, we performed a single-cell experiment using the cellenONE platform under the same reaction conditions, to evaluate whether the workflow could already yield biologically meaningful results, even at low labeling efficiency. We isolated approximately 2,000 single cells in total, comprising HeLa and HepG2 cells, and processed them using the same hyperplex protocol applied in bulk: TMT labeling at pH 9.5, dimethylation at pH 6.8, and DMSO as solvent. Our goal was to assess whether, despite incomplete labeling, the method could provide sufficient proteomic information to distinguish between the two cell types.

The results mirrored those obtained in bulk. At the lysine residues, we observed about 96% TMT labeling efficiency, while at the N-termini, the distribution was 8–16% dimethyl, 40–60% TMT and 30–40% unlabeled (Figure 22,23). These values confirmed that the intended site specificity was not achieved, and there are several reasons that could be led to this result:

1. Residual TMT activity: the 40–60% TMT at N-termini indicates incomplete hydrolysis of unreacted TMT after 4 h, allowing TMT to keep competing at the N-terminus.
2. Competition + site leakage: 8–16% dimethyl alongside high TMT and 30–40% unlabeled shows a three-way split (TMT vs dimethyl vs no reaction), a sign that the timing/pH window isn't tight enough to enforce site specificity.
3. pH window not restrictive enough for N-term selectivity: the conditions possibly left a meaningful fraction of N-termini deprotonated while TMT was still active, enabling TMT to capture those sites before (or instead of) dimethylation.
4. Possible solvent effects: DMSO may be slowing TMT hydrolysis and/or shifting local protonation equilibria, contributing to the mixed outcome.

5. TMT excess: the 1:5 protein:TMT ratio likely drives to TMT over labeling and reduces the effective opportunity for dimethylation at the N-terminus.

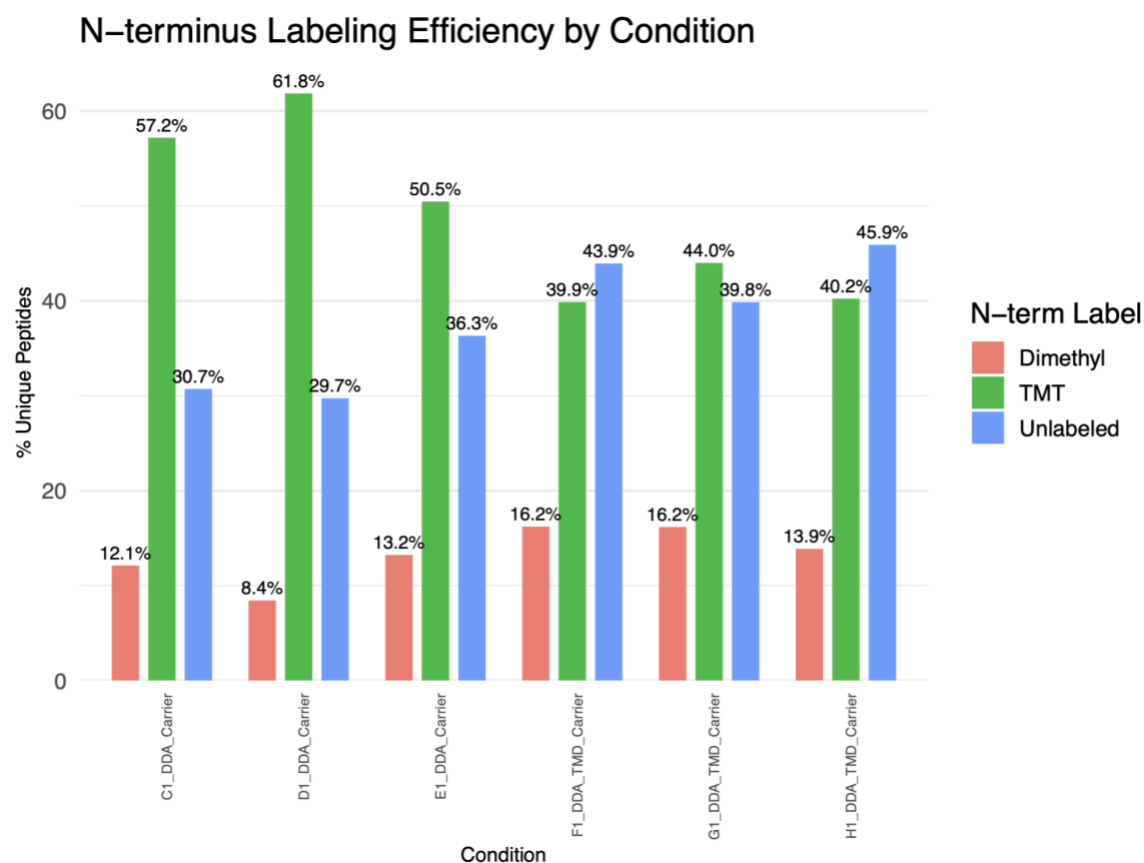


Figure 22: N-terminus labeling efficiency of the hyperplex protocol in single-cell experiment.

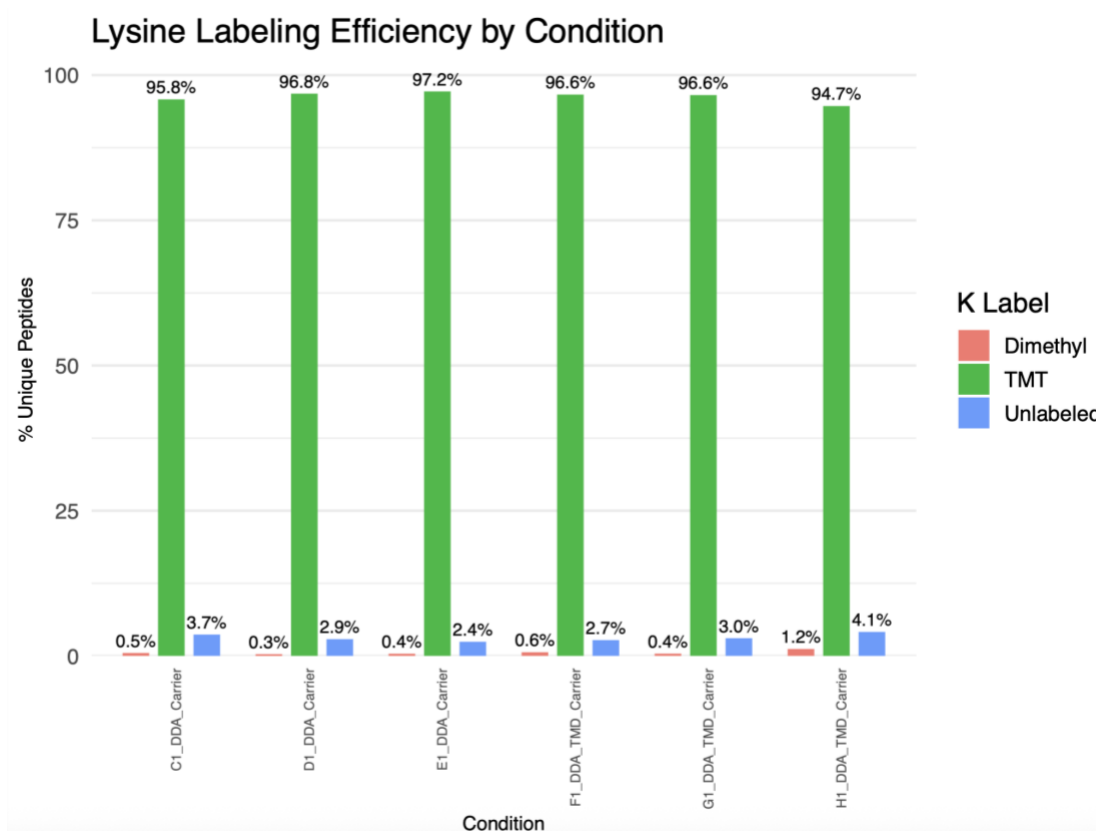


Figure 23: Lysine residues labeling efficiency of the hyperplex protocol in single-cell experiment.

Nonetheless, even with this incomplete labeling, enough peptides were successfully identified to allow biological interpretation. As shown in Figure 24, the average peptide identifications per cell were around 100–120 for HeLa and slightly lower for HepG2. While this depth is below the expected range for optimized single-cell proteomics, it was sufficient for downstream data analysis. The principal component analysis (Figure 25) clearly separated the two cell populations into distinct clusters, demonstrating that the workflow was able to preserve cell-type-specific proteomic signatures. Furthermore, the volcano plot (Figure 26) revealed several significantly differentially abundant proteins between HeLa and HepG2 cells, including canonical markers associated with their respective biological phenotypes.

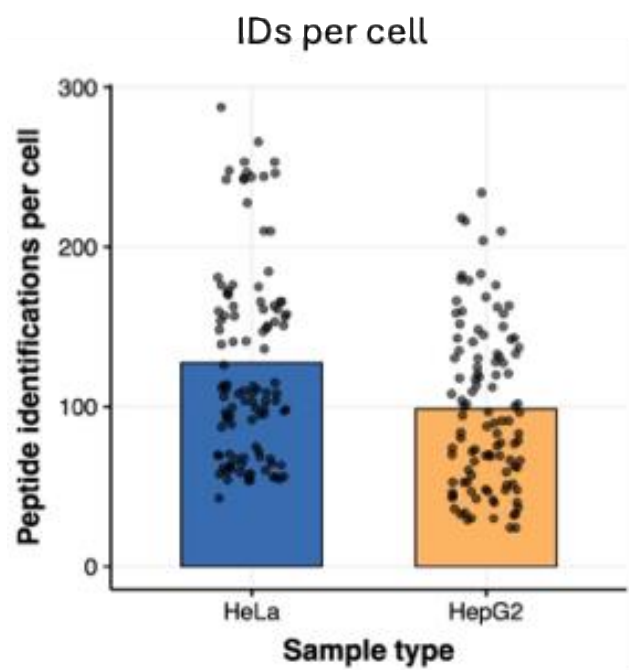


Figure 24: Peptide identifications per cell after hyperplex protocol in single-cell experiment.

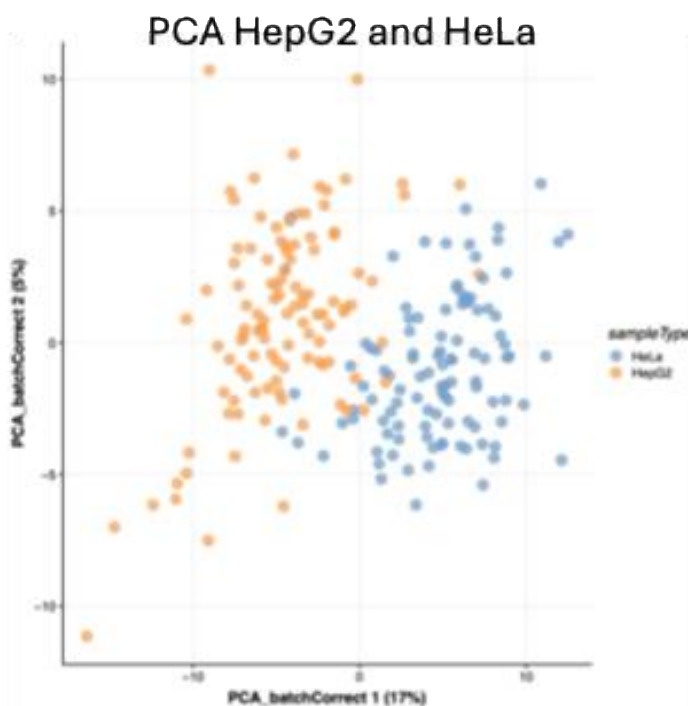


Figure 25: Principal component analysis after hyperplex protocol in single-cell experiment.

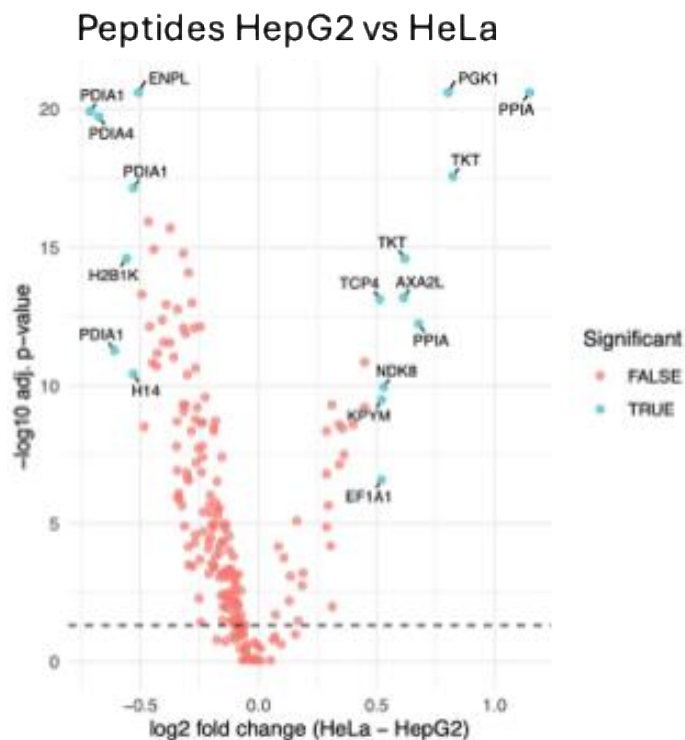


Figure 26: Volcano plot after hyperplex protocol in single-cell experiment.

Overall, these results are highly encouraging. Although the labeling efficiency remains low, the ability to discriminate between two distinct cell lines at the single-cell level indicates that the workflow, even in its current form, captures meaningful biological variation.

Developing a hyperplex workflow that combines isobaric (TMT), and isotopic (dimethyl) labeling strategies represents a complex chemical and methodological challenge. Throughout this work, we systematically explored the feasibility of integrating these two complementary chemistries to overcome one of the main bottlenecks in single-cell proteomics which is the limited throughput. The goal was to create a protocol that could simultaneously expand multiplexing capacity and boost the analytical signal, providing a scalable solution for large single-cell cohorts.

Through systematic testing, we identified the key parameters that must be finely controlled to achieve reproducible, site-specific labeling: the pKa of the reactive sites, pH of each reaction, protein-to-label stoichiometry, incubation time, and quenching efficiency. Our bulk experiments under cellenONE-like conditions provided critical insights into how these factors interact. These

findings underscore that the main challenge is not the lack of reactivity but the competition and kinetic imbalance between the two chemistries.

Encouragingly, the single-cell experiments conducted in parallel confirmed that, despite low overall labeling efficiency, the workflow can already generate biologically meaningful data. The successful discrimination between HeLa and HepG2 cells, supported by clear clustering in PCA and statistically significant protein differences, demonstrates that the workflow is chemically capable of capturing true biological variability. This outcome validates the conceptual foundation of the hyperplex approach and indicates that further refinements will likely yield a functional and scalable protocol.

Moving forward, the next steps will focus on refining the timing and extent of TMT hydrolysis, optimizing reagent stoichiometry to balance reactivity, and improving pH control to enforce site selectivity. Additional efforts will aim to better understand the impact of DMSO on hydrolysis kinetics and to adapt the workflow to the spatial and temporal constraints of cellenONE automation.

In summary, while many chemical and technical challenges remain, the progress achieved so far provides a solid foundation. The results demonstrate that the combination of TMT and dimethylation can be made compatible within a single workflow and that the chemistry, even if not yet optimized, is functionally effective at the single-cell level. These findings are a strong indication that we are on the right path toward the development of a robust, high-throughput hyperplex protocol, capable of transforming single-cell proteomics by drastically increasing throughput, sensitivity, and quantitative precision.

CHAPTER 4: SINGLE-CELL PROTEOMICS IN A MODEL OF EXTENDED HEALTHSPAN

4.1 Introduction

Aging is a complex and heterogeneous process that affects all levels of biological organization, from molecules to tissues and entire organisms. Rather than following a single linear trajectory, aging is the result of numerous molecular pathways and stochastic cellular events, leading to a progressive loss of homeostasis. Each cell within a tissue accumulates different types of damage and activates unique adaptive responses. As a result, the aging process is not uniform, but rather highly heterogeneous, both between individuals and among cells of the same tissue (75). This intrinsic heterogeneity represents one of the major challenges in aging research. Traditional bulk analyses provide averaged measurements across thousands or millions of cells, masking the underlying variability that may hold the key to understanding mechanisms of resilience or decline. Age-related dysfunctions often derive from the presence of rare or divergent cell states such as senescent, metabolically reprogrammed, or stressed cells, which are masked when analyzing cells in bulk (76).

Single-cell proteomics provides a powerful approach to address one of the key limitations in aging biology: the inability to resolve heterogeneity within complex tissues. Single-cell transcriptomics has been widely used in mapping cellular diversity across tissues and life stages, mRNA abundance does not always reflect protein levels or functional state, especially in aged cells where post-transcriptional regulation becomes prominent (77). Proteins are the direct effectors of cellular physiology, represent a more faithful readout of functional state. The advent of high-sensitivity mass spectrometry now allows the quantification of hundreds to thousands of proteins from individual cells, capturing cell-specific differences that are otherwise averaged out in bulk analyses (78). The application of single-cell proteomics to aging offers the opportunity to uncover how molecular heterogeneity affects the aging phenotype, revealing not only how different cells age within a tissue but also how they differentially respond to metabolic or pharmacological interventions designed to extend healthspan.

To understand the molecular basis of this cellular heterogeneity requires focusing on processes that directly couple metabolism to protein function. Among the different molecular mechanisms

that change with age, post-translational modifications (PTMs) play a central role in modulating protein activity, localization, and stability. These modifications act as regulatory marks that help to integrate cellular metabolic states with signaling and structural networks. In this context, lysine succinylation has emerged as a relevant PTM due to its tight connection with mitochondrial metabolism and energy production, both of which are deeply affected during aging. Lysine succinylation is a key metabolic regulator that connects mitochondrial activity to protein function. Succinylation involves the covalent addition of a succinyl group derived from succinyl-CoA, an intermediate of the tricarboxylic acid (TCA) cycle, onto lysine residues. This modification introduces a negative charge and causes a substantial change in protein structure, often altering enzymatic activity, protein–protein interactions, or subcellular localization(79).

Succinic acid, a natural intermediate of the tricarboxylic acid (TCA) cycle, plays a central role in sustaining mitochondrial respiration and metabolic signaling. Experimental evidence shows that succinic acid supplementation can improve mitochondrial function and enhance oxidative metabolism under stress. For instance, succinate treatment has been reported to restore mitochondrial respiration and ATP production in models of oxidative injury (80) and to ameliorate mitochondrial dysfunction in high-fat diet–induced metabolic stress by improving lipid metabolism and mitochondrial gene expression (81). Additionally, succinate metabolism has been proposed as a target for metabolic reprogramming in aging and disease, as it directly influences both redox balance and post-translational modifications such as succinylation(82).

Because succinylation is tightly related to the availability of succinyl-CoA, the cellular levels of its metabolic precursor, succinic acid (succinate), can directly influence the extent and distribution of this modification. By restore succinate pools, succinic acid enhances TCA cycle flux and supports succinyl-CoA synthesis, thereby influencing the balance between protein succinylation and desuccinylation. In this way, succinic acid acts not only as a substrate for energy production but also as a potential metabolic modulator of protein regulation. Restoring this metabolic balance has been proposed as a strategy to counteract mitochondrial decline and metabolic inflexibility during aging (82,83).

To explore these mechanisms experimentally, we sought to investigate how metabolic modulation through succinic acid treatment influences the aging liver at the molecular level. To this end, we

employed a mouse model of physiological aging, in which 17-month-old mice (corresponding to late adulthood) were administered succinic acid as a dietary supplement. The liver was selected as the target organ for analysis due to its central role in systemic metabolism and its pronounced vulnerability to age-related dysfunction. Hepatic aging is characterized by mitochondrial decline, increased oxidative stress, and impaired proteostasis, making it an ideal system to evaluate the effects of metabolic interventions. Based on unpublished data from our laboratory, succinic acid treatment mitigates several molecular alterations associated with aging, suggesting a potential rejuvenating effect at the proteomic level, supporting its potential as a candidate gerotherapeutic compound.

To achieve a comprehensive and quantitative assessment of proteome remodeling, we applied our single-cell proteomics workflow combined with TMTpro 35plex labeling, enabling simultaneous analysis of hundreds of individual liver cells from both control and treated animals. This high-throughput approach provides detailed insight into cell-to-cell variability, allowing us to resolve how succinic acid treatment influences proteomic composition, functional diversity, and molecular heterogeneity within the aged liver.

4.2 Methods

Mice treatment

Seventeen-month-old male mice were used as a model of physiological aging. Animals were randomly divided into two groups: a control group receiving standard chow and water, and a treatment group receiving succinic acid supplementation for 10 weeks. Succinic acid was administered in base chow diet at a concentration of 2%. All mice were maintained under controlled environmental conditions (12-hour light/dark cycle, 22 ± 2 °C, ad libitum access to food and water).

Liver collection and Hepatocytes isolation

Following the 20 weeks treatment, mice were anesthetized with isoflurane and sacrificed by portal vein perfusion for hepatocyte isolation. After induction of anesthesia, the abdomen was opened and the portal vein was cannulated using a 24 G catheter. The liver was first perfused in situ with

pre-warmed phosphate-buffered saline (PBS) to remove circulating blood, followed by perfusion with Earle's Balanced Salt (EBS) buffer containing Liberase (0.26 U/mL) to enzymatically digest hepatic tissue. The perfusion was performed using a peristaltic pump at a controlled flow rate until complete blanching of the liver was achieved. After perfusion, the liver was excised and transferred to ice-cold Dulbecco's Modified Eagle Medium (DMEM), and hepatocytes were released by gently tearing the Glisson's capsule and shaking the tissue to dissociate cells. The resulting cell suspension was filtered through a 100 μ m sterile cell strainer and centrifuged at 40 x g for 5 min at 4 °C to pellet viable hepatocytes. The supernatant containing dead cells and non-parenchymal cells was discarded, and hepatocytes were further purified by Percoll gradient centrifugation (80 x g for 10 min at 4 °C) to remove remaining debris and non-viable cells. The final hepatocyte pellet was resuspended in DMEM and cell viability was assessed by Trypan Blue assay. Freshly isolated hepatocytes were then processed immediately for single-cell proteomics sample preparation.

Single-cell sample preparation

The preparation of hepatocytes for single-cell proteomic analysis was performed following the same workflow described in Chapter 2, starting from cell washing and staining (steps 3.3–3.10). Briefly, after isolation, cells were collected in a 15 mL conical tube, washed with ice-cold 1 $\times$ PBS, and centrifuged at 50 $\times$ g for 5 min at 4 °C. The cell pellet was resuspended in 1 mL of PBS and incubated with 1 μ M Sytox™ Green to stain dead cells for 20 min on ice in the dark. After washing, cells were fixed in 0.5% paraformaldehyde (PFA) for 15 min on ice, followed by an additional PBS wash and resuspension at a final concentration of 200–500 cells/ μ L. Single-cell isolation and sample handling were carried out as previously described, using the cellenONE® system for automated single-cell dispensing. For this experiment, one slide was prepared containing 24 sets of 32 single cells, with control and succinic acid-treated cells distributed across all sets. In addition to sorted single cells, the sets contained negative one control sample each to be used for downstream quality-control analysis. Proteins were digested with trypsin overnight and labeled using the TMTpro 35plex reagents (1:8 proteins-to-label ratio) according to the optimized workflow detailed in Chapter 2. Each set of 32 single cells represents a sample that is collected and placed in a well of a 384-well plate that will be used for LC-MS/MS analysis.

LC-MS/MS

The workflow we used to perform the data acquisition was adapted from Huffman et al., *Nature Methods*, 2023. To establish a reference proteome and evaluate labeling efficiency, a carrier/reference mix was prepared from the leftover cells after hepatocyte isolation. Approximately the equivalent of 100 cells was labeled with TMTpro 126C (carrier) and 5 cells with TMTpro 127N (reference). The two samples were combined, dried down, and resuspended in 0.1% formic acid and 0.01% n-dodecyl maltoside (DDM) prior to LC-MS/MS analysis. This mix was first analyzed in DIA mode to verify TMT labeling efficiency and to generate a spectral library for precursor identification and retention time calibration, thereby facilitating peptide selection for prioritized acquisition. DIA analysis was performed on a Thermo Scientific Orbitrap Exploris mass spectrometer coupled to a nanoflow liquid chromatography Dionex Ultimate 3000. The instrument was operated in positive ion mode using a nano-electrospray ionization (NSI) source at 2.0 kV with an ion transfer tube temperature of 275 °C. FAIMS was used at standard resolution with a compensation voltage (CV) of -50 V. Peptide separation was achieved using an 83-minute gradient from 4% to 99% solvent B (0.1% formic acid in acetonitrile) at a flow rate of 0.5-0.2 μ L/min. Solvent A consisted of 0.1% formic acid in water. 1 μ L of sample was loaded onto a 25-cm x 75- μ m IonOpticks Aurora Series UHPLC column. Full MS scans were acquired at a resolution of 120,000 (m/z 450-1600), followed by DIA windows of 20 m/z across the same range, with HCD collision energy set to 33%, AGC target of 5×10^5 , and automatic maximum injection time. The resulting data were processed in FragPipe (v23) to assess labeling efficiency and to build experiment-specific spectral libraries. For FragPipe database searches, trypsin was specified as the digestion enzyme, allowing up to two missed cleavages. TMTpro (+304.2071 Da) was set as a variable modification on lysine residues and peptide N-termini for the analysis of labeling efficiency. To build the spectral libraries, TMTpro was set as a fixed modification on lysines and variable on N-termini, ensuring consistent quantification across datasets. Using a custom R script, peptide precursor information (m/z and retention time) from the DIA run was filtered and converted into an inclusion list for prioritized acquisition. This list was imported into MaxQuant.Live, where it was used to generate a DDA method. The DDA method employed the same LC gradient and FAIMS settings, with full scans acquired at a resolution of 120,000, an AGC target of 1×10^6 , and 36% normalized collision energy for MS2. After the initial DDA method was

created, we performed three additional DDA runs of independently prepared carrier/reference mixtures under identical gradient and instrument conditions. After each run, the data were processed with FragPipe, and the inclusion list was refined iteratively to maximize peptide identification consistency and protein coverage across runs. These iterative optimization experiments are referred to as “scout runs”. Once the DDA method had been refined across the three scout runs, it was used for the acquisition of single-cell proteomics samples. Each run included the carrier and reference mix, which was dried and resuspended in 0.1% formic acid and 0.01% DDM immediately before injection. The first single-cell set was analyzed to verify TMT labeling efficiency and to confirm the carrier-to-reference-to-cell ratio. Upon validation of these parameters, the remaining 23 sets of single-cell samples were acquired using the same DDA method and chromatographic gradient.

Raw data processing

Raw files from all LC-MS/MS acquisitions were processed using FragPipe (v23) with trypsin as the proteolytic enzyme, TMTpro (+304.2071 Da) defined as a fixed modification on lysines and a variable modification on peptide N-termini. The resulting quantitative tables were exported for downstream analysis in R (v4.5.0). Data integration and curation were performed using a custom R pipeline adapted from existing single-cell proteomics frameworks. The analysis started by importing the FragPipe output folder, together with a run-to-well annotation table linking LC–MS runs to well positions and plate layout, and the CellenONE cell isolation files containing metadata for each dispensed cell. These datasets were combined using the QuantQC package, generating a unified matrix that links quantitative values, acquisition metadata, and cell-level information. The combined dataset was then converted into an SCP object, allowing structured storage of all metadata and quantitative features for subsequent filtering and normalization.

Quality Filtering and PSM-Level Processing:

Data filtering was initially applied at the peptide–spectrum match (PSM) level.

- Runs with fewer than 1,500 identified PSMs were excluded;
- A dynamic median absolute deviation (MAD) filter was used to remove outlier runs based on total identifications;

- PSMs annotated as contaminants or matched to the decoy database were removed;
- Only PSMs with a precursor ion fraction (PIF) > 0.8 were retained, ensuring high spectral purity;
- A single-cell-to-carrier ratio (SCR) was calculated, and PSMs with $SCR > 10x$ were excluded to minimize interference from carrier channels;
- PSMs with an estimated false discovery rate (FDR) $> 1\%$ were filtered out.

After quality filtering, the reference channel (TMT 127N) was used to account for between-run variability by scaling intensities across acquisitions. The normalized PSM intensities were then aggregated to the peptide level, excluding the carrier, reference, and unused channels.

Cell- and Peptide-Level Filtering:

At the peptide level, we filtered out cells with low total intensity or high peptide coefficient of variation (CV) and removed blank samples. Peptides with a high proportion of missing values were also excluded. To correct for technical variation, we applied a two-step normalization:

- Cell normalization by dividing each cell's intensities by its median total signal, ensuring comparable overall intensity across cells;
- Peptide normalization by dividing each peptide's intensities by its mean value across cells, followed by $\log_2$ transformation and imputation of missing values.

Batch Correction and Dimensional Reduction:

To correct for batch effects, we used the function `removeBatchEffect` of `limma` package to perform linear modeling and batch correction by run, addressing technical variability between MS acquisitions. To control for biological heterogeneity, additional correction was performed using cell diameter metadata obtained during CellenONE isolation.

Principal component analysis (PCA) was then used to evaluate the data structure and to estimate cumulative variance for the selection of the number of principal components to retain for downstream analyses.

Protein-Level Quantification:

Peptide intensities were aggregated to the protein level by summing normalized peptide abundances per protein group. The resulting protein matrix was used for PCA, clustering, and exploratory data analysis to evaluate treatment-dependent proteomic differences and cellular heterogeneity within and across conditions.

Exploratory analysis:

To assess differences in cellular and proteomic characteristics between experimental conditions, we generated violin plots comparing cell diameter, median reporter ion intensity (MedianRI), and the number of identified proteins per cell between control and succinic acid-treated groups. Statistical differences in cell diameter and MedianRI distributions were evaluated using the Wilcoxon rank-sum test (two-sided), with $p < 0.05$ considered significant.

Correlations between cell diameter and MedianRI were assessed using Spearman's rank correlation coefficient, and results were visualized as scatterplots with fitted regression lines. Principal component analysis (PCA) was performed on normalized protein intensities to explore global structure and potential separation of proteomic profiles between treated and control cells.

For differential protein abundance analysis, we used the findMarkers function, applying the Wilcoxon rank-sum test to detect proteins significantly regulated between conditions. The results were corrected for multiple testing using the Benjamini–Hochberg false discovery rate (FDR) method. Log₂ fold changes were calculated as treated versus control means. Proteins with $FDR < 0.05$ and $|\log_2FC| > 0.25$ were considered significantly differentially abundant.

Enrichment analysis:

Single-cell proteomes were first compared with liver transcriptomic reference datasets to classify individual cells into known hepatic subtypes based on marker protein expression. Each cell was assigned to the most closely matching reference subtype using similarity scoring across marker sets. The resulting classifications were visualized using Uniform Manifold Approximation and Projection (UMAP), separating cells by condition.

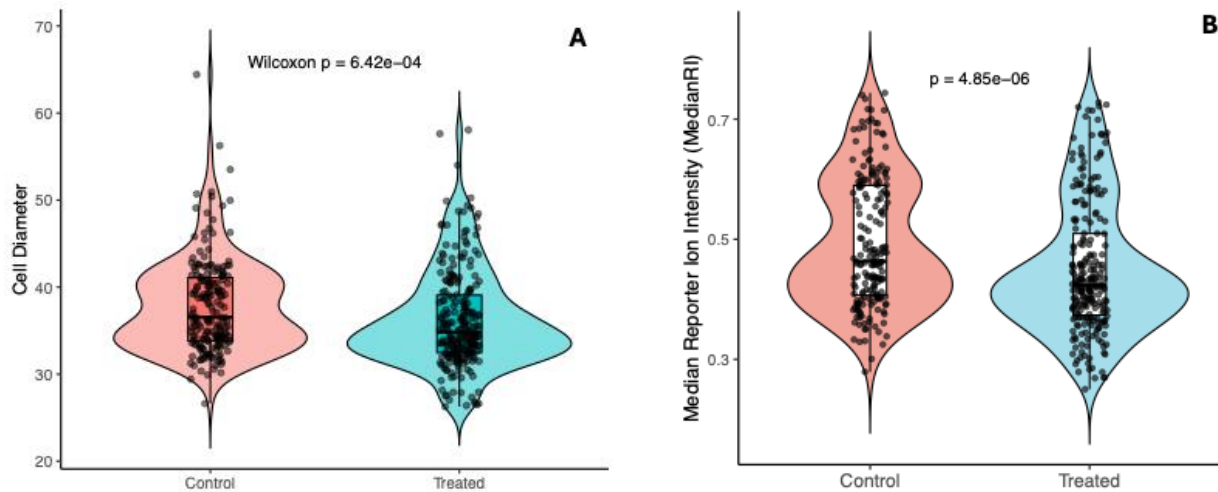
To explore functional heterogeneity independently of cell type annotation, we performed unsupervised clustering based on Gene Ontology (GO) biological process terms. For each cell, enrichment scores were computed by summarizing the abundance of proteins associated with each GO term. Clustering was visualized by hierarchical heatmap representation and UMAP projection, with clusters assigned numerical identifiers for reference

To dissect the contribution of specific biological pathways, enrichment scores for three functional categories such as senescence, metabolism, and chromatin organization were computed for each cell. These scores were converted in z-scores and visualized across UMAP coordinates to reveal the distribution of functional programs between control and treated conditions.

4.3 Results

To investigate the biological effects of succinic acid treatment on the liver of aged mice, we applied our optimized single-cell proteomics workflow using TMTpro 35plex reagents. In total, we analyzed 24 sets of cells, which after data processing resulted in the quantification of 431 individual cells distributed between the control and treated conditions. Across all cells, we detected 634 proteins, providing sufficient proteome depth to characterize major cellular processes while maintaining single-cell resolution. Each single-cell dataset contained the reporter ion intensities for the TMT channels, providing a quantitative measure of relative protein abundance at the cellular level. Single-cell proteomics not only enables the comparison of average protein levels between conditions but also provides access to cell-to-cell variability, which can reveal subtle population structures or treatment-dependent responses. When comparing cell diameter between control and treated cells (Figure 27A), we observed a significant difference. Cells derived from succinic acid-treated mouse showed a slightly reduced average diameter and a broader distribution, indicating increased morphological variability across the population. These data suggest that the treatment might modulate hepatocyte morphology or influence cellular state heterogeneity within the aged liver. This morphological variation was detected at the proteomic level. The median reporter ion intensity was also significantly different between the two conditions (Figure 27B). Treated cells tended to show lower median intensities overall, which is consistent with the smaller average cell size observed in the same samples. By directly comparing cell diameter and reporter ion intensity (Figure 27C), we confirmed a strong positive correlation

between these two parameters in both control and treated groups. The linear relationship indicates that protein content scales with cell size. The distribution of the points indicates that treated cells not only show a lower overall intensity but also a broader scatter around the regression line, suggesting increased heterogeneity in their protein content relative to size. This could reflect variable responses of individual hepatocytes to the treatment, revealing subtle phenotypic diversification that would remain undetected in bulk measurements. When considering proteome depth, measured as the number of identified proteins per cell, we found a similar pattern (Figure 27D). Control cells showed a slightly higher median number of identified proteins compared to treated cells, again consistent with the differences in reporter ion intensity. Importantly, this observation is not only relevant for biological interpretation but also for experimental design and data normalization in single-cell proteomics, as cell size introduces a natural source of variation in protein detectability.



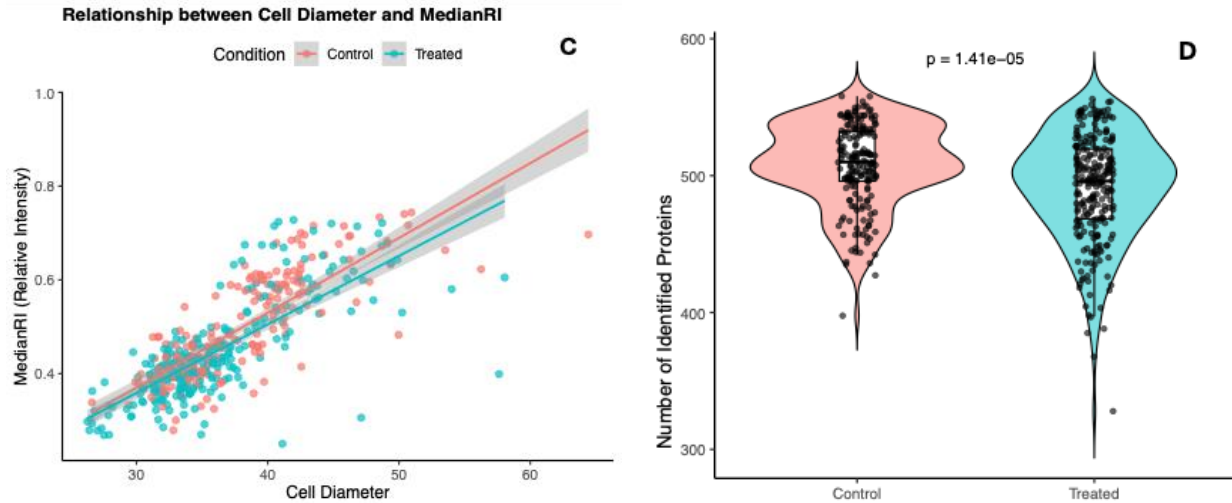


Figure 27: Relationship between cell diameter, protein content, and proteome depth in single hepatocytes from aged mice. (A) Comparison of cell diameter between control and treated groups; (B) The median reporter ion intensity (MedianRI), comparison by condition; (C) Correlation between cell diameter and MedianRI across individual cells; (D) Number of identified proteins per condition.

To further explore the effect of succinic acid treatment at the proteome level, we compared the single-cell protein expression profiles between control and treated liver cells. The principal component analysis (PCA) revealed a clear separation between the two conditions (Figure 28A). Cells from succinic acid-treated mice tended to cluster apart from those of untreated animals, indicating that the treatment induces a shift in proteomic composition across individual cells. This separation suggests that the succinic acid diet drives a measurable proteomic response at the single-cell level. The treated population also appeared more dispersed along the principal components, reflecting increased variability in the proteome consistent with the higher heterogeneity already observed in cell size and protein content distributions. To identify the proteins contributing to this separation, we performed a differential abundance analysis between control and treated groups (Figure 28B). The resulting volcano plot highlights several significantly regulated proteins, with the most evident differences observed in ribosomal proteins, which were enriched in the untreated group. This suggests that untreated aged liver cells retain higher ribosomal protein abundance. In contrast, succinic acid-treated cells exhibited reduced expression of these components, possibly indicating more efficient or restrained protein synthesis.

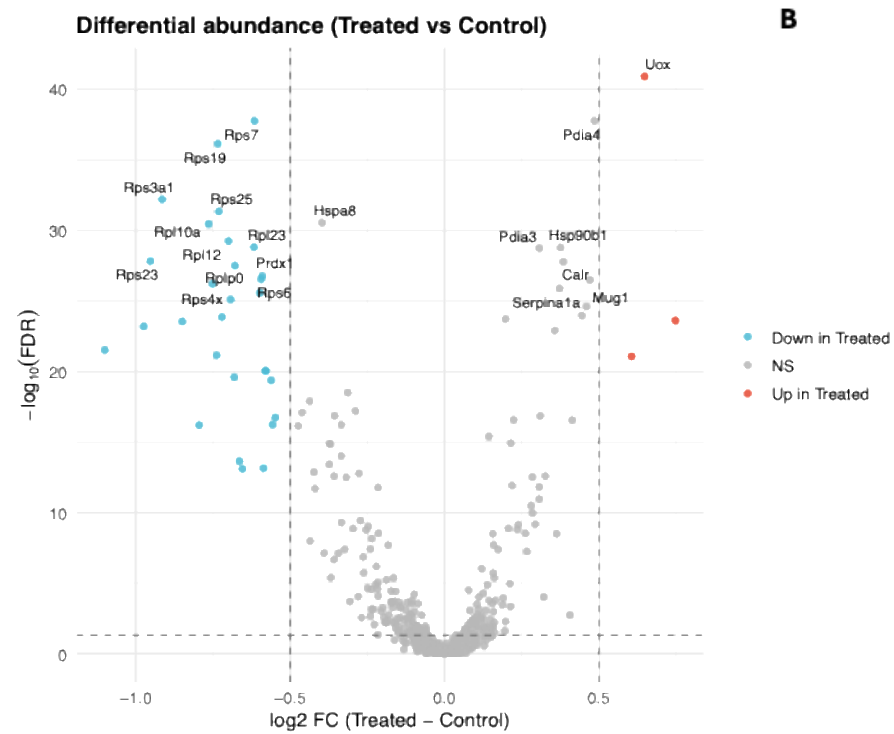
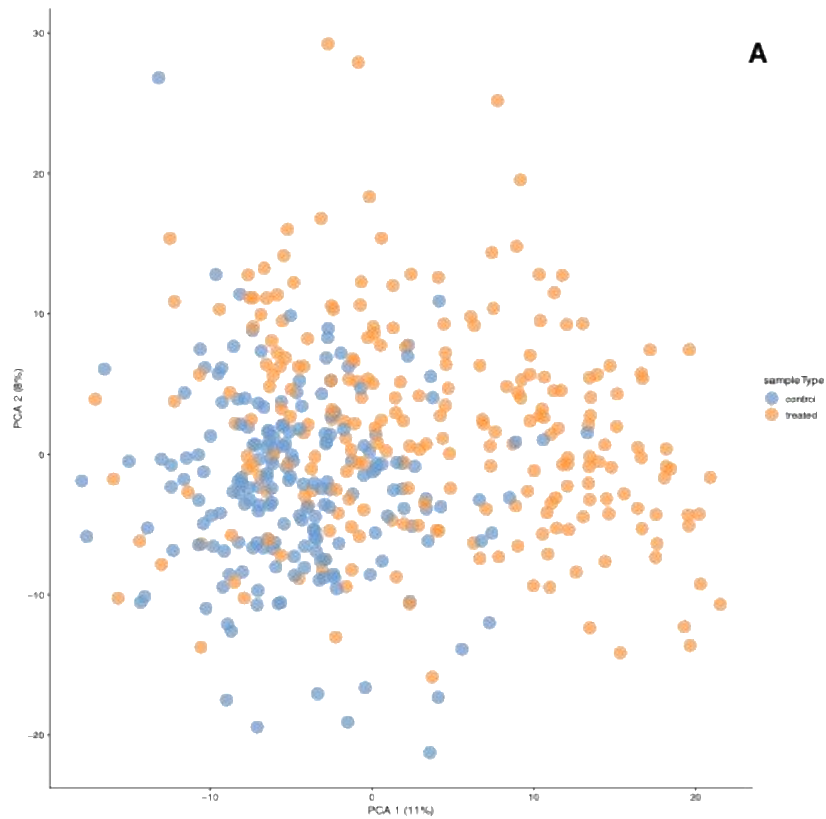


Figure 28: Proteomic differences between control and succinic acid–treated liver cells at single-cell resolution. (A) Principal component analysis (PCA) of single-cell proteomes shows distinct clustering of control (orange) and treated (blue) cells; (B) Volcano plot of differential protein abundance between control and treated conditions.

Based on the detected protein profiles, we classified individual liver cells into known subtypes according to marker proteins characteristic of each cell class (Figure 29). This classification approach allows us to interpret the single-cell proteomic data in a biologically meaningful way, associating each quantified cell with a specific liver cell population. The analysis revealed that hepatocyte-related clusters represented the predominant cell type across both control and succinic acid–treated conditions, as expected for liver tissue. Interestingly, the treated group displayed a higher proportion of cells belonging to the subclass C11, a hepatocyte subtype typically associated with periportal localization and enriched in oxidative, mitochondrial, and biosynthetic pathways. These cells are known to play a central role in energy metabolism and detoxification, suggesting that succinic acid treatment may promote or preserve hepatocyte states with enhanced metabolic capacity. This classification provides the basis for to determine whether certain functional subclasses are enriched in one condition compared to the other, and whether particular proteomic functions are associated with greater or reduced variability across single cells.

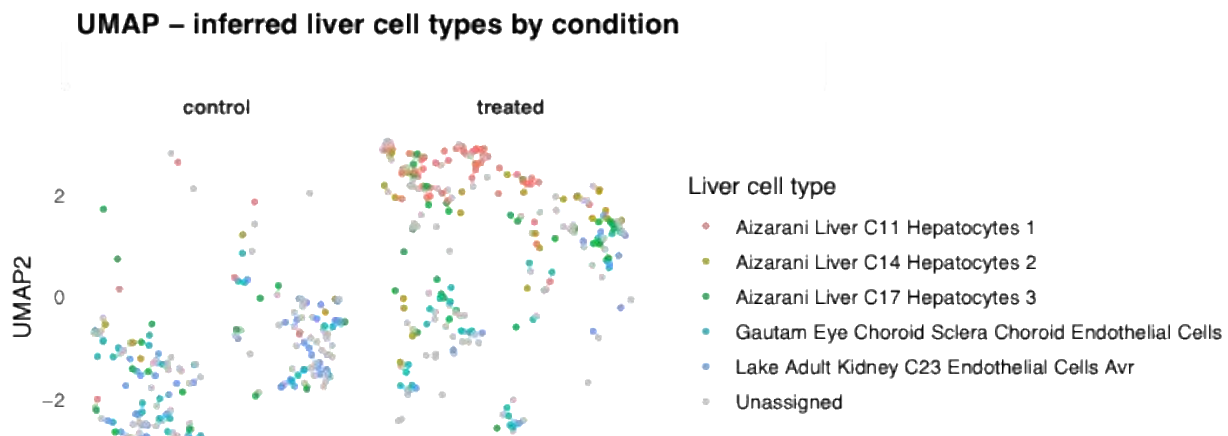


Figure 29: Classification of single liver cells into known cellular subtypes based on proteomic signatures. Single-cell proteomes were assigned to liver cell subclasses according to characteristic protein expression patterns.

Having established that single-cell proteomics can distinguish different liver cell types, we next aimed to identify how many functional groups could be resolved within our dataset and which molecular features define them. Since each cell is represented by its individual proteome, the main challenge was to determine how to classify these cells based on their global protein composition. To address this, we performed unsupervised clustering based on Gene Ontology (GO) terms, grouping cells according to the biological processes most represented in their proteomes. This approach allowed us to classify the entire dataset into four major clusters, each corresponding to a distinct functional profile (Figure 30A). The GO enrichment patterns indicated that these clusters reflect different proteomic programs, ranging from lipid and fatty acid metabolism to ribosome biogenesis and RNA processing. The separation observed in the heatmap was consistent with the UMAP projection, confirming the coherence and stability of the clustering approach (Figure 30B). When the same clusters were visualized separately for the two conditions (Figure 30C), both control and treated populations showed the same global organization, indicating that the same functional proteomic states are present in both conditions. However, the distribution and density of cells within each cluster differed substantially. In particular, the treated population showed a broader spread within clusters, indicating higher intra-cluster heterogeneity. This suggests that succinic acid treatment not only generate entirely new functional states but increases the variability of existing ones, possibly reflecting differential metabolic adaptation among hepatocytes in the aged liver. These analyses demonstrate that single-cell proteomics can capture coherent functional groupings across cells and conditions, while capturing differences in population variability.

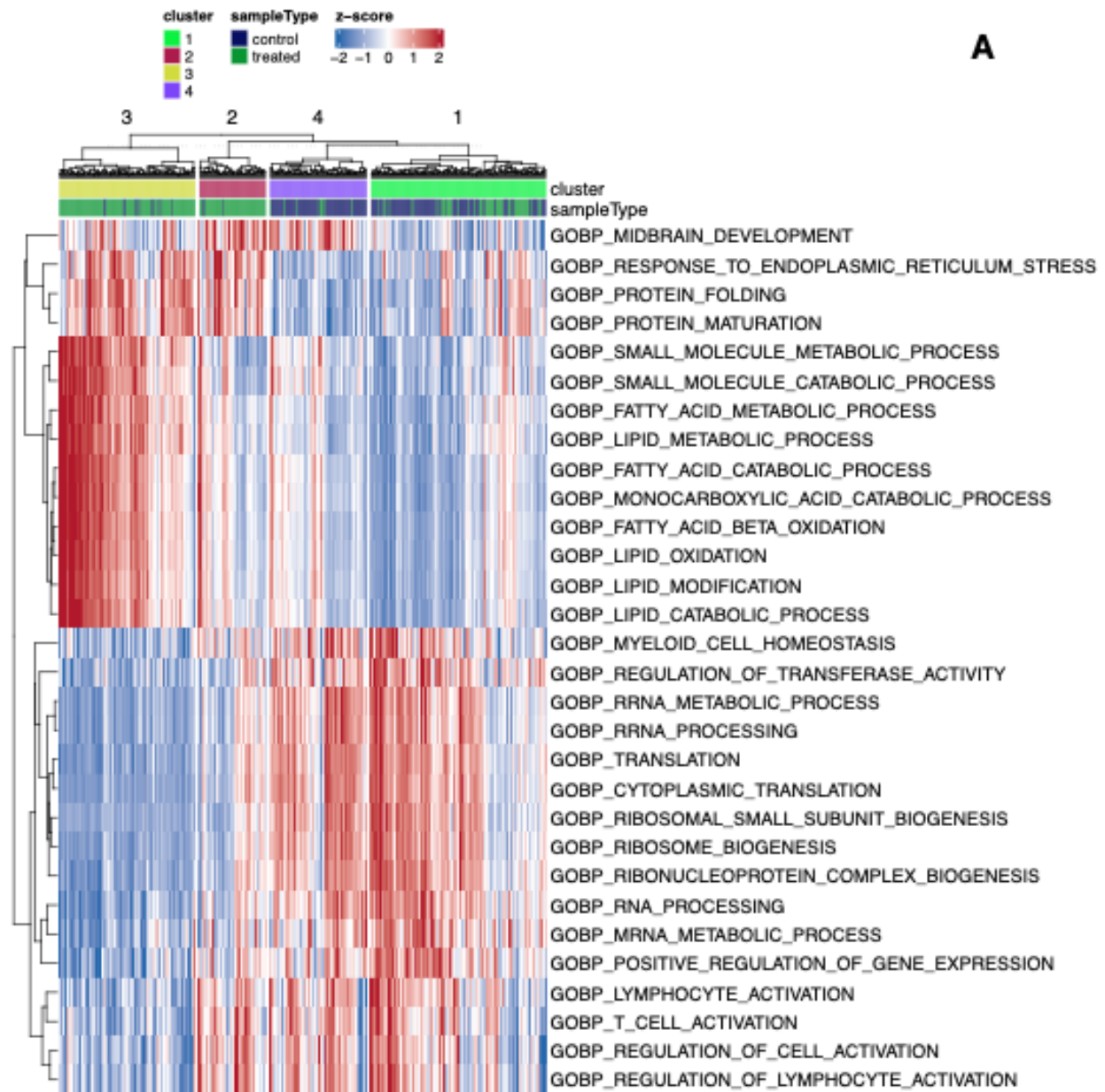
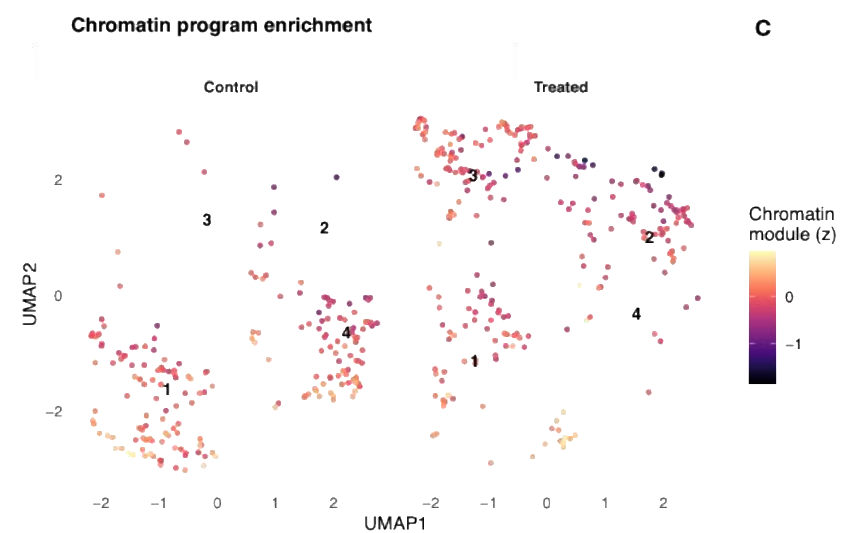
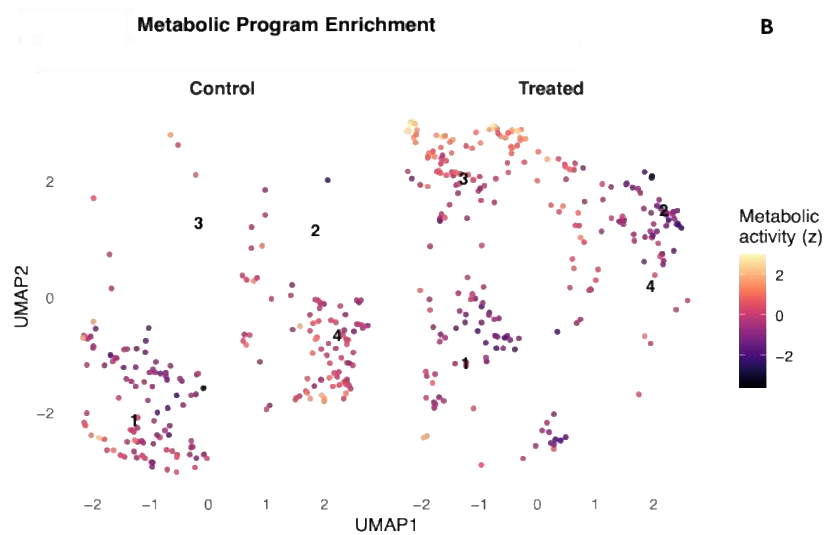
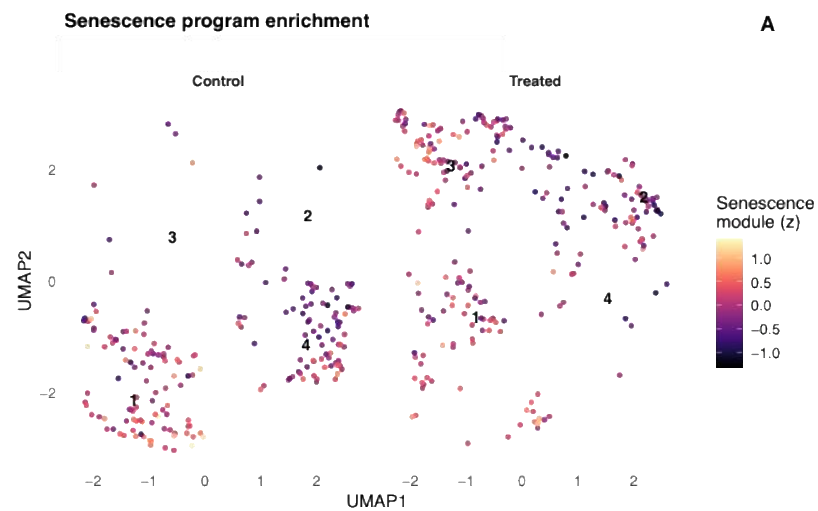




Figure 30: Unsupervised functional clustering of single-cell proteomes. (A) Heatmap of Gene Ontology-based clustering showing four main clusters (1–4) characterized by distinct functional enrichments. (B) UMAP representation of the same four clusters in the combined dataset. (C) Split UMAPs for control and treated cells.

To better understand the biological meaning of the observed heterogeneity, we next examined how specific functional pathways varied across individual cells in each condition. We focused on three key biological domains such as senescence, metabolism, and chromatin organization, that are central to the aging process and are expected to respond to succinic acid treatment. For each category, we mapped the enrichment of the corresponding Gene Ontology (GO) terms across all single-cell proteomes. These representations revealed striking differences between control and treated populations. Cells showing enrichment for senescence-related proteins (Figure 31A) were more heterogeneous in the treated group. This suggests that cells undergoing or maintaining senescence in the treated population exhibit more diverse proteomic compositions, potentially reflecting distinct senescence-associated states. This variability may contribute to differential responses to interventions, making senescent cells in treated tissues less uniform and potentially harder to target with single gerotherapeutic strategies. A similar observation was evident for metabolic processes (Figure 31B), where cells from the treated condition exhibited broader enrichment patterns. This variability indicates that succinic acid does not lead to a uniform metabolic reprogramming but instead increases the range of metabolic phenotypes within the liver cell population. This heterogeneity may reflect an adaptive response in which different cells adopt distinct metabolic strategies to maintain homeostasis under treatment. In contrast, control cells displayed a more compact distribution, consistent with a more homogeneous metabolic state typical of aged, untreated tissue. The pattern was also evident in chromatin organization (Figure 31C), where control cells showed consistent enrichment of epigenetic proteins, while treated cells presented a more fragmented and diverse proteomic signature. This variability could be indicative of differential chromatin remodeling responses to treatment, possibly linked changes in transcriptional activity or the activation of distinct gene expression patterns. These findings were confirmed by the heatmap analysis (Figure 31D), which revealed that enrichment of functional GO terms was more homogeneous in control cells, whereas in treated cells, the same functions were expressed in a more variable and cell-specific manner. This pattern demonstrates that succinic acid treatment increases functional heterogeneity rather than causing uniform changes in average protein expression.



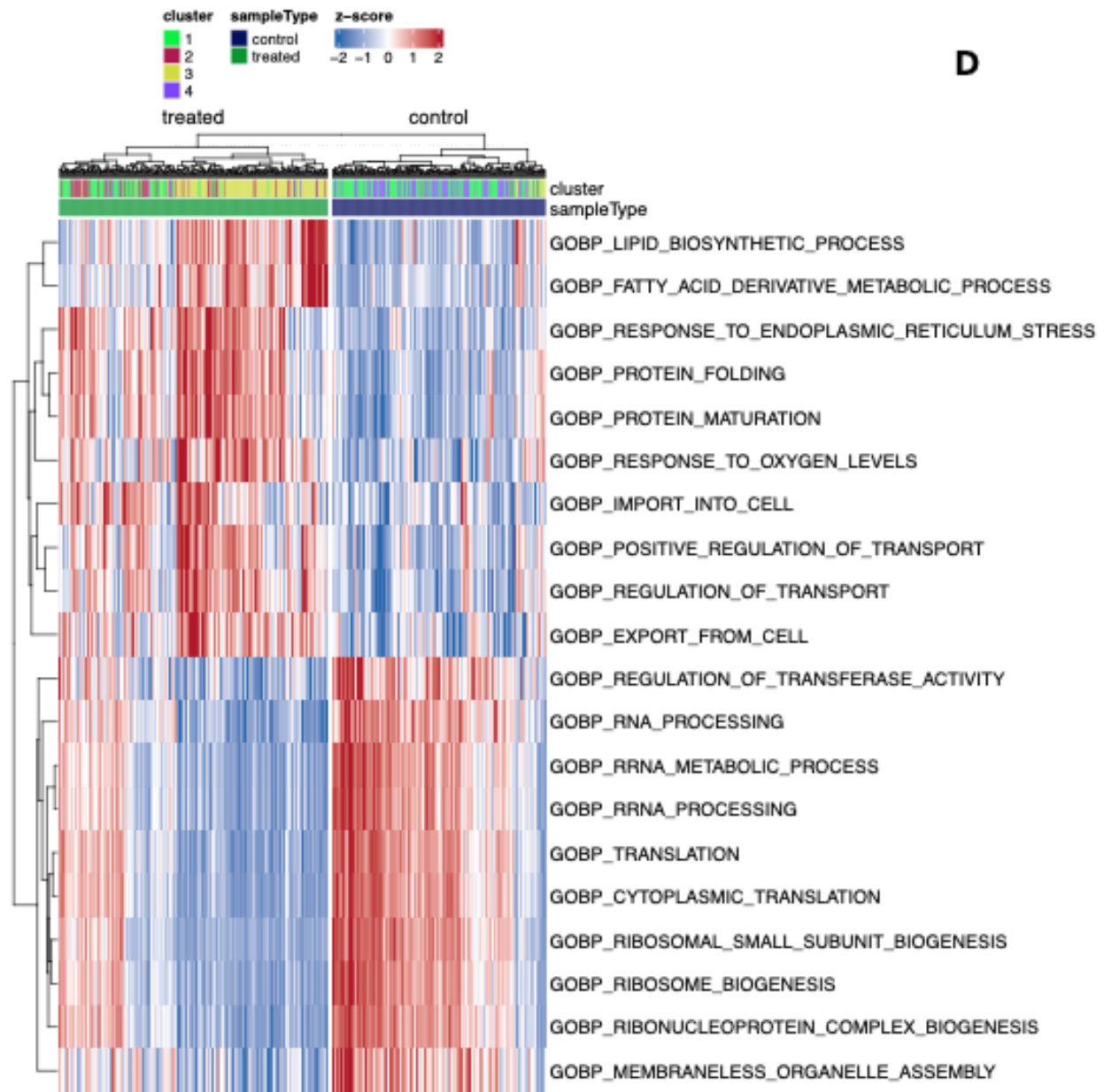


Figure 31: Functional pathways across individual cells in control and treated cells. (A–C) UMAP projections showing the enrichment of Gene Ontology terms related to (A) senescence, (B) metabolism, and (C) chromatin organization across individual cells in control and succinic acid–treated conditions; (D) Heatmap representation of the same GO terms.

4.4 Conclusions

This study demonstrates the potential of single-cell proteomics to uncover molecular complexity in aging tissues and to evaluate the effects of interventions at a high level of resolution. Aging is intrinsically heterogeneous, with cells within the same tissue exhibiting variable molecular profiles, functional states, and responses to external stimuli. By applying an optimized single-cell proteomic workflow to the livers of aged mice treated with succinic acid, we were able to dissect this heterogeneity quantitatively and to reveal cell-type and specific function adaptations that would remain masked in bulk analyses.

Our single-cell approach allowed us to link cell morphology, protein content, and functional diversity, showing that proteome depth correlates strongly with cell size and that succinic acid treatment modulates both cellular composition and molecular functions. The single-cell resolution of this analysis revealed that treatment with succinic acid not only affects average protein expression levels but also increases intra-population diversity, suggesting a more plastic and adaptive proteomes in treated cells. This insight into cell-to-cell variability represents one of the main strengths of single-cell proteomics.

Beyond the biological findings, this study introduced several technical innovations in data acquisition and analysis that enhance the accuracy, reproducibility, and interpretability of SCP data. The development of a custom R pipeline allowed rigorous quality filtering, adaptive normalization and batch correction that represent an analytical innovation that enhances both sensitivity and reproducibility in single-cell proteomics studies. We were able to map functional states and biological processes at the single-cell level, these analyses demonstrated that the treated liver displays a higher degree of functional heterogeneity, suggesting that succinic acid treatment modulates aging not by uniform rejuvenation but by expanding the spectrum of cellular states.

In conclusion, this work highlights the power of single-cell proteomics to characterize biological complexity, resolve subtle treatment-response effects, and provide mechanistic insight into how molecular heterogeneity contributes to tissue function and adaptation during aging. It also shows the importance of advanced computational approaches in translating high-dimensional single-cell data into biologically meaningful information.

CHAPTER 5: CONCLUSIONS AND PERSPECTIVES

This work describes the development, optimization, and application of advanced single-cell proteomics workflows, aimed at resolving biological heterogeneity and investigating the molecular basis of aging. The project has generated both technical and conceptual insights into the current state and future potential of proteomics at single-cell resolution. Working on these projects has been a gradual learning process on how to make single-cell proteomics reliable and biologically meaningful. Each part of the work (optimizing multiplexed labeling, adapting workflows for specific protein classes, and applying them to aging models) contributed to understanding what determines the success of this emergent technique and where its limitations currently lie.

5.1 Methodological lessons from single-cell proteomics

One of the main aspects learned during this experimental journey was how small differences in reaction conditions, peptide handling, or normalization strategies can result into substantial variation at the data level (62). Through the optimization of labeling strategies, it became clear that increasing throughput or multiplexing capacity is not just adding more reagents but controlling the chemistry with precision to preserve quantitative consistency (84). The experience with combined labeling approaches, showed that the complexity of the system increases with each additional layer of chemistry. Achieving balance between sensitivity, labeling completeness, and quantitative reproducibility became a technical and conceptual challenge (85). Similar reasoning guided the development of the workflow for the analysis of the histones: here, the challenges shifted from multiplexing to molecular diversity. Working with proteins that are highly charged and extensively modified showed that standard proteomics protocols cannot be applied uniformly across all protein classes (86). Instead, they must be tailored to the biochemical nature of the target proteome. This concept underscored the importance of method specificity, designing the workflow around the molecules of interest, rather than expecting universal applicability. Applying these optimized approaches to the study of aging provided a broader perspective on the role of single-cell proteomics as a discovery tool. The ability to distinguish old and treated cells, and to detect heterogeneity even within apparently homogeneous populations, demonstrated that proteomics can reach the resolution needed to describe functional diversity at the cellular level (87). This

experience also highlighted the current interpretational gap: while we can now measure proteomes of individual cells with increasing accuracy, linking these measurements to defined cell types or biological functions remains challenging because proteomic reference sets are still missing. This revealed that technical precision alone is insufficient but biological context and data integration are equally essential components (88).

5.2 Current limitations and unresolved challenges

Despite the progress achieved, several technical and analytical challenges remain as limitations to overcome. The first one concerns chemical reproducibility in multiplexed labeling. The challenge ahead is to design labeling chemistries that maintain high multiplexing depth and that are fully compatible with single-cell proteomics workflows (84). At the single-cell level, sample recovery represents another obstacle. Even with optimized carrier-based strategies and automated workflows, peptide loss during transfer is still critical, particularly for hydrophobic and low-abundant proteins (89). These losses not only limit proteome coverage but introduce variability that complicates downstream normalization and clustering. Improving the steps involved in collecting and transferring the sample are therefore immediate priorities for next-generation workflows. On the computational side, data analysis is still constrained by missing values, batch effects, and incomplete sampling. Even when technical noise is minimized, the stochastic nature of ion sampling leads to inconsistent peptides detection across cells (62). This limits the resolution of clustering and the reliability of differential expression analyses. There is a need for algorithms specifically designed for proteomics data that can model missingness rather than relying on computational frameworks borrowed from transcriptomics (90). A further limitation is cell-type classification. Nowadays, proteomic datasets rely almost entirely on transcriptomic-based reference annotations. Because mRNA and protein abundances often diverge, this dependency introduces systematic uncertainty into biological interpretation. The creation of proteome-based cell atlases, even for a limited set of tissues, will be essential to define cell identities directly at the protein level and to allow consistent cross-study comparisons (88). Finally, applying these methods to complex biological systems such as aging exposes their interpretational boundaries. While single-cell proteomics can detect differences in protein abundance across cells and conditions, it provides limited information on structural states, interactions, or subcellular localization, factors that are likely to change substantially during aging. Integrating proteomics

with complementary approaches that capture protein conformation and spatial organization will therefore be necessary to translate these molecular measurements into mechanistic understanding (45).

The question of which biological models best exploit the power of this technology remains central. Different tissues exhibit different forms of complexity: the cell types composition of the liver is relatively heterogenous and suitable for quantitative comparison, while the brain presents extreme heterogeneity with hundreds of interacting cell types (91). Skeletal muscle adds another layer of complexity, with multinucleated fibers that may hide intracellular heterogeneity among nuclei. The workflows developed here could be particularly informative in such systems, where variability across or within cells carries functional meaning. Extending these methods to brain or muscle tissue could reveal new aspects of proteomic regulation during aging that are invisible to bulk or transcriptomic analyses.

5.3 Future perspectives, overall contributions and outlook

In the next decade, single-cell proteomics is expected to evolve from a specialized experimental niche into an integrated platform for molecular biology. Advances in sample miniaturization, labeling chemistry, and data acquisition will likely enable near-complete proteome coverage for thousands of cells per experiment, while maintaining quantitative precision (62). Automation of sample preparation will reduce losses and increase throughput, bringing the field closer to the scale currently achieved in single-cell transcriptomics (89). Another transformative step will be the emergence of multi-omic and spatially approaches, linking proteomic information with transcriptomic, metabolomic, and imaging data within the same biological context (92). Such integration will make it possible to not only measure protein abundances but also interpret them in relation to cell state, environment, and spatial organization.

Overall, this thesis provides both methodological and conceptual advances in the field of single-cell proteomics. From a methodological point of view, the work introduced and validated a novel hyperplex strategy that combines two labeling chemistries for a site-specific labeling on lysine residues and peptide N-termini. This approach demonstrated the feasibility of expanding multiplexing capacity beyond standard configurations while preserving quantitative precision. The systematic evaluation of reaction parameters defined an operating window for improvements. In

parallel, a dedicated workflow for histone proteins was developed, enabling the detection and quantification of histone variants and post-translational modifications at single-cell level. By adapting digestion and derivatization chemistry to the specific properties of chromatin-associated proteins, this method extended the scope of single-cell proteomics beyond soluble or cytosolic proteins, providing a foundation for future studies on protein regulation and epigenetic remodeling. These methodological advances were integrated into an optimized single-cell workflow applied to the study of aging and metabolic intervention. Using a 35-plex TMT design, the analysis of liver cells from aged and succinic-acid-treated mice demonstrated the capacity of the approach to resolve cellular heterogeneity, separate biological conditions, and identify enriched functional pathways linked to treatment response. This represents an example where single-cell proteomics was successfully used to explore complex physiological processes such as aging.

Taken together, these achievements demonstrate that it is possible to move from proof-of-concept experiments toward quantitatively reliable and biologically informative single-cell proteomics. As the field advances, the boundary between method development and biological discovery will continue to blur. The future of proteomics will depend on how well we can link molecular measurement to biological mechanism, bridging chemistry and biology. In this context, the approaches and lessons from this thesis contribute to building the foundation for a next generation of proteomic analyses, ones capable of describing not only how proteins change, but how cells age, adapt, and maintain their identity at the molecular scale.

REFERENCES

1. Goldman SL, MacKay M, Afshinnkoo E, Melnick AM, Wu S, Mason CE. The Impact of Heterogeneity on Single-Cell Sequencing. *Front Genet*. 2019 Mar 1;10.
2. Choi YH, Kim JK. Dissecting Cellular Heterogeneity Using Single-Cell RNA Sequencing. *Mol Cells*. 2019 Mar 31;42(3):189–99.
3. Huang S. Non-genetic heterogeneity of cells in development: more than just noise. *Development*. 2009 Dec 1;136(23):3853–62.
4. Singh A, Soltani M. Quantifying Intrinsic and Extrinsic Variability in Stochastic Gene Expression Models. *PLOS ONE*. 2013 Dec 31;8(12):e84301.
5. Hsiao CJ, Tung P, Blischak JD, Burnett JE, Barr KA, Dey KK, et al. Characterizing and inferring quantitative cell cycle phase in single-cell RNA-seq data analysis. *Genome Res*. 2020 Apr 1;30(4):611–21.
6. Chang AY, Marshall WF. Organelles – understanding noise and heterogeneity in cell biology at an intermediate scale. *J Cell Sci*. 2017 Mar 1;130(5):819–26.
7. Kumar RM, Cahan P, Shalek AK, Satija R, DaleyKeyser Aj, Li H, et al. Deconstructing transcriptional heterogeneity in pluripotent stem cells. *Nature*. 2014 Dec 4;516(7529):56–61.
8. Zhong J, Gao R rong, Zhang X, Yang J xin, Liu Y, Ma J, et al. Dissecting endothelial cell heterogeneity with new tools. *Cell Regen*. 2025 Mar 23;14(1):10.
9. Durlanik S, Thiel A. Requirement of immune system heterogeneity for protective immunity. *Vaccine*. 2015 Sept 29;33(40):5308–12.
10. Proietto M, Crippa M, Damiani C, Pasquale V, Sacco E, Vanoni M, et al. Tumor heterogeneity: preclinical models, emerging technologies, and future applications. *Front Oncol*. 2023 Apr 28;13.
11. Gascoigne KE, Taylor SS. Cancer Cells Display Profound Intra- and Interline Variation following Prolonged Exposure to Antimitotic Drugs. *Cancer Cell*. 2008 Aug 12;14(2):111–22.
12. Steurman Y, Cohen M, Peshes-Yaloz N, Valadarsky L, Cohn O, David E, et al. Dissection of Influenza Infection In Vivo by Single-Cell RNA Sequencing. *Cell Syst*. 2018 June 27;6(6):679-691.e4.
13. Telwatte S, Morón-López S, Aran D, Kim P, Hsieh C, Joshi S, et al. Heterogeneity in HIV and cellular transcription profiles in cell line models of latent and productive infection: implications for HIV latency. *Retrovirology*. 2019 Nov 11;16(1):32.
14. Cohn RL, Gasek NS, Kuchel GA, Xu M. The heterogeneity of cellular senescence: insights at the single-cell level. *Trends Cell Biol*. 2023 Jan;33(1):9–17.

15. Martinez-Jimenez CP, Eling N, Chen HC, Vallejos CA, Kolodziejczyk AA, Connor F, et al. Aging increases cell-to-cell transcriptional variability upon immune stimulation. *Science*. 2017 Mar 31;355(6332):1433–6.
16. Klaips CL, Jayaraj GG, Hartl FU. Pathways of cellular proteostasis in aging and disease. *J Cell Biol*. 2018 Jan 2;217(1):51–63.
17. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. Hallmarks of aging: An expanding universe. *Cell*. 2023 Jan 19;186(2):243–78.
18. Jovic D, Liang X, Zeng H, Lin L, Xu F, Luo Y. Single-cell RNA sequencing technologies and applications: A brief overview. *Clin Transl Med*. 2022 Mar 29;12(3):e694.
19. Carangelo G, Magi A, Semeraro R. From multitude to singularity: An up-to-date overview of scRNA-seq data generation and analysis. *Front Genet*. 2022 Oct 3
20. Zhao Q, Li S, Krall L, Li Q, Sun R, Yin Y, et al. Deciphering cellular complexity: advances and future directions in single-cell protein analysis. *Front Bioeng Biotechnol*. 2025 Jan 14;12:1507460.
21. Evrony GD, Hinch AG, Luo C. Applications of Single-Cell DNA Sequencing. *Annu Rev Genomics Hum Genet*. 2021 Aug 31;22(Volume 22, 2021):171–97.
22. Kashima Y, Sakamoto Y, Kaneko K, Seki M, Suzuki Y, Suzuki A. Single-cell sequencing techniques from individual to multiomics analyses. *Exp Mol Med*. 2020 Sept;52(9):1419–27.
23. Valecha M, Posada D. Somatic variant calling from single-cell DNA sequencing data. *Comput Struct Biotechnol J*. 2022 Jan 1;20:2978–85.
24. Chen G, Ning B, Shi T. Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. *Front Genet*. 2019 Apr 5;10:317.
25. Haque A, Engel J, Teichmann SA, Lönnberg T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med*. 2017 Aug 18;9(1):75.
26. Foley SW, Gregory BD. Protein Interaction Profile Sequencing (PIP-seq). *Curr Protoc Mol Biol*. 2016 Oct 10;116:27.5.1-27.5.15.
27. Shuken SR. An Introduction to Mass Spectrometry-Based Proteomics. *J Proteome Res*. 2023 July 7;22(7):2151–71.
28. Keane RE, Tidy RJ, Parker GJ, Gummer JPA, Priddis C. Mass spectrometry based proteomics: Changing the impact of protein analysis in forensic science. *WIREs Forensic Sci*. 2024;6(4):e1516.
29. Peters-Clarke TM, Coon JJ, Riley NM. Instrumentation at the Leading Edge of Proteomics. *Anal Chem*. 2024 May 21;96(20):7976–8010.

30. Brunner A, Thielert M, Vasilopoulou C, Ammar C, Coscia F, Mund A, et al. Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Mol Syst Biol*. 2022 Mar;18(3):e10798.
31. Bubis JA, Arrey TN, Damoc E, Delanghe B, Slovakova J, Sommer TM, et al. Challenging the Astral mass analyzer to quantify up to 5,300 proteins per single cell at unseen accuracy to uncover cellular heterogeneity. *Nat Methods*. 2025 Mar;22(3):510–9.
32. Ghosh G, Shannon AE, Searle BC. Data acquisition approaches for single cell proteomics. *Proteomics*. 2025 Jan;25(1–2):2400022.
33. Rozanova S, Barkovits K, Nikolov M, Schmidt C, Urlaub H, Marcus K. Quantitative Mass Spectrometry-Based Proteomics: An Overview. In: Marcus K, Eisenacher M, Sitek B, editors. *Quantitative Methods in Proteomics*. New York, NY: Springer US; 2021. p. 85–116.
34. Calvete JJ, Lomonte B, Saviola AJ, Calderón Celis F, Ruiz Encinar J. Quantification of snake venom proteomes by mass spectrometry-considerations and perspectives. *Mass Spectrom Rev*. 2024;43(5):977–97.
35. Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem*. 2007 Oct;389(4):1017–31.
36. Bowser BL, Robinson RAS. Enhanced Multiplexing Technology for Proteomics. *Annu Rev Anal Chem*. 2023 June 14;16(1):379–400.
37. Hsu JL, Chen SH. Stable isotope dimethyl labelling for quantitative proteomics and beyond. *Philos Transact A Math Phys Eng Sci*. 2016 Oct 28;374(2079):20150364.
38. Budnik B, Levy E, Harmange G, Slavov N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol*. 2018 Oct 22;19(1):161.
39. Specht H, Emmott E, Petelski AA, Huffman RG, Perlman DH, Serra M, et al. Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. *Genome Biol*. 2021 Jan 27;22(1):50.
40. Leduc A, Huffman RG, Cantlon J, Khan S, Slavov N. Exploring functional protein covariation across single cells using nPOP. *Genome Biol*. 2022 Dec 16;23(1):261.
41. Brunner A, Thielert M, Vasilopoulou C, Ammar C, Coscia F, Mund A, et al. Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Mol Syst Biol*. 2022 Feb 28;18(3):e10798.
42. Truong T, Kelly R. What's New in Single-Cell Proteomics. *Curr Opin Biotechnol*. 2024 Apr;86:103077.

43. Cutler R, Corveleyn L, Ctortekca C, Cantlon J, Vaca SAJ, Deforce D, et al. Mass Spectrometry-based Profiling of Single-cell Histone Post-translational Modifications to Dissect Chromatin Heterogeneity. *bioRxiv*; 2024. p. 2024.07.05.602213.
44. Lundberg E, Borner GHH. Spatial proteomics: a powerful discovery tool for cell biology. *Nat Rev Mol Cell Biol*. 2019 May;20(5):285–302.
45. Mund A, Brunner AD, Mann M. Unbiased spatial proteomics with single-cell resolution in tissues. *Mol Cell*. 2022 June 16;82(12):2335–49.
46. Kaeberlein M. How healthy is the healthspan concept? *GeroScience*. 2018 Aug 6;40(4):361–4.
47. Cohen AA, Ferrucci L, Fülöp T, Gravel D, Hao N, Kriete A, et al. A complex systems approach to aging biology. *Nat Aging*. 2022 July;2(7):580–91.
48. Uyar B, Palmer D, Kowald A, Murua Escobar H, Barrantes I, Möller S, et al. Single-cell analyses of aging, inflammation and senescence. *Ageing Res Rev*. 2020 Dec;64:101156.
49. Cohn RL, Gasek NS, Kuchel GA, Xu M. The heterogeneity of cellular senescence: insights at the single-cell level. *Trends Cell Biol*. 2023 Jan;33(1):9–17.
50. Uyar B, Palmer D, Kowald A, Murua Escobar H, Barrantes I, Möller S, et al. Single-cell analyses of aging, inflammation and senescence. *Ageing Res Rev*. 2020 Dec;64:101156.
51. Wei YN, Hu HY, Xie GC, Fu N, Ning ZB, Zeng R, et al. Transcript and protein expression decoupling reveals RNA binding proteins and miRNAs as potential modulators of human aging. *Genome Biol*. 2015 Feb 22;16(1):41.
52. Buck SA, Mabry SJ, Glausier JR, Banks-Tibbs T, Ward C, Kozel JG, et al. Aging disrupts the coordination between mRNA and protein expression in mouse and human midbrain. *bioRxiv*. 2024 June 1;2024.06.01.596950.
53. Jenuwein T, Allis CD. Translating the Histone Code. *Science*. 2001 Aug 10;293(5532):1074–80.
54. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Res*. 2011 Mar;21(3):381–95.
55. Bozdemir N, Kablan T, Biyikli E, Cinar O, Uysal F. A comprehensive review of histone modifications during mammalian oogenesis and early embryo development. *Histochem Cell Biol*. 2025;163(1):70.
56. Huang Y, Lin S, Garcia B, Zhao Y. Quantitative proteomic analysis of histone modifications. *Chem Rev*. 2015 Feb 17;115(6):2376–418.
57. Zheng Y, Huang X, Kelleher NL. Epiproteomics: quantitative analysis of histone marks and codes by mass spectrometry. *Curr Opin Chem Biol*. 2016 Aug;33:142–50.

58. Vai A, Noberini R, Ghirardi C, Rodrigues de Paula D, Carminati M, Pallavi R, et al. Improved Mass Spectrometry–Based Methods Reveal Abundant Propionylation and Tissue-Specific Histone Propionylation Profiles. *Mol Cell Proteomics*. 2024 July 1;23(7):100799.
59. Zuo-Fei Yuan, Shu Lin, Rosalynn C Molden, Xing-Jun Cao, Natarajan V Bhanu, Xiaoshi Wang, Simone Sidoli, Shichong Liu, Benjamin A Garcia. EpiProfile Quantifies Histone Peptides With Modifications by Extracting Retention Time and Intensity in High-resolution Mass Spectra. *Mol Cell Proteomics MCP*. 2015 June;14(6):1696–707.
60. Jenuwein T, Allis CD. Translating the histone code. *Science*. 2001 Aug 10;293(5532):1074–80.
61. Sidoli S, Bhanu NV, Karch KR, Wang X, Garcia BA. Complete Workflow for Analysis of Histone Post-translational Modifications Using Bottom-up Mass Spectrometry: From Histone Extraction to Data Analysis. *J Vis Exp JoVE*. 2016 May 17;(111):54112.
62. Slavov N. Single-cell protein analysis by mass spectrometry. *Curr Opin Chem Biol*. 2021 Feb;60:1–9.
63. Leduc A, Koury L, Cantlon J, Slavov N. Massively parallel sample preparation for multiplexed single-cell proteomics using nPOP. *BioRxiv Prepr Serv Biol*. 2023 Nov 28;2023.11.27.568927.
64. Lund PJ, Kori Y, Zhao X, Sidoli S, Yuan ZF, Garcia BA. Isotopic Labeling and Quantitative Proteomics of Acetylation on Histones and Beyond. *Methods Mol Biol Clifton NJ*. 2019;1977:43–70.
65. Cutler R, Corveleyn L, Ctortocka C, Cantlon J, Jacome Vaca SA, Deforce D, et al. Mass Spectrometry-based Profiling of Single-cell Histone Post-translational Modifications to Dissect Chromatin Heterogeneity. *BioRxiv Prepr Serv Biol*. 2024 Aug 12;2024.07.05.602213.
66. Davie JR. Inhibition of histone deacetylase activity by butyrate. *J Nutr*. 2003 July;133(7 Suppl):2485S-2493S.
67. Stransky S, Cutler R, Aguilan J, Nieves E, Sidoli S. Investigation of reversible histone acetylation and dynamics in gene expression regulation using 3D liver spheroid model. *Epigenetics Chromatin*. 2022 Nov 21;15(1):35.
68. Yuan ZF, Sidoli S, Marchione DM, Simithy J, Janssen KA, Szurgot MR, et al. EpiProfile 2.0: A Computational Platform for Processing Epi-Proteomics Mass Spectrometry Data. *J Proteome Res*. 2018 July 6;17(7):2533–41.
69. Arul AB, Robinson RAS. Automated Sample Multiplexing by using Combined Precursor Isotopic Labeling and Isobaric Tagging (cPILOT). *J Vis Exp JoVE*. 2020 Dec 18;(166):10.3791/61342.

70. Zhou X, Ni X, Wu X, Yin L. Review of Modern Eschweiler–Clarke Methylation Reaction. *Molecules*. 2025 Aug 27;30(17):3504.
71. Xing S, Pai A, Wu R, Lu Y. NHS-Ester Tandem Labeling in One Pot Enables 48-Plex Quantitative Proteomics. *Anal Chem*. 2021 Sept 28;93(38):12827–32.
72. Klein ATJ, Holschbach M. Labelling of the solvent DMSO as side reaction of methylations with n.c.a. [11C]CH₃I. *Appl Radiat Isot*. 2001 Sept 1;55(3):309–13.
73. Di Mino C, Clancy AJ, Sella A, Howard CA, Headen TF, Seel AG, et al. Weak Interactions in Dimethyl Sulfoxide (DMSO)–Tertiary Amide Solutions: The Versatility of DMSO as a Solvent. *J Phys Chem B*. 2023 Feb 8;127(6):1357–66.
74. Creary X, Burtch EA, Jiang Z. Carbocation-Forming Reactions in Dimethyl Sulfoxide. *J Org Chem*. 2003 Feb 1;68(3):1117–27.
75. Zhu T, Li S, Liu D, Zhang X, Zhou L, Zhou R, et al. Single-worm quantitative proteomics reveals aging heterogeneity in isogenic *Caenorhabditis elegans*. *Aging Cell*. 2023 Dec 3;23(3):e14055.
76. Keele GR, Zhang JG, Szpyt J, Korstanje R, Gygi SP, Churchill GA, et al. Global and tissue-specific aging effects on murine proteomes. *Cell Rep*. 2023 July 25;42(7).
77. Li Y, Wang Q, Xuan Y, Zhao J, Li J, Tian Y, et al. Investigation of human aging at the single-cell level. *Ageing Res Rev*. 2024 Nov 1;101:102530.
78. Yang L, George J, Wang J. Deep Profiling of Cellular Heterogeneity by Emerging Single-Cell Proteomic Technologies. *Proteomics*. 2020 July;20(13):e1900226.
79. Zhao G, Zhen J, Liu X, Guo J, Li D, Xie J, et al. Protein post-translational modification by lysine succinylation: Biochemistry, biological implications, and therapeutic opportunities. *Genes Dis*. 2023 July 1;10(4):1242–62.
80. Nowak G, Clifton GL, Bakajsova D. Succinate ameliorates energy deficits and prevents dysfunction of complex I in injured renal proximal tubular cells. *J Pharmacol Exp Ther*. 2008 Mar;324(3):1155–62.
81. Yang Y, Luo L, Li Y, Shi X, Li C, Chai J, et al. Succinic Acid Improves the Metabolism of High-Fat Diet-Induced Mice and Promotes White Adipose Browning. *Nutrients*. 2024 Nov 8;16(22):3828.
82. Mills E, O'Neill LAJ. Succinate: a metabolic signal in inflammation. *Trends Cell Biol*. 2014 May;24(5):313–20.
83. Hou X, Zhu L, Xu H, Shi J, Ji S. Dysregulation of protein succinylation and disease development. *Front Mol Biosci*. 2024 May;11.

84. Petelski AA, Emmott E, Leduc A, Huffman RG, Specht H, Perlman DH, et al. Multiplexed single-cell proteomics using SCoPE2. *Nat Protoc.* 2021 Dec;16(12):5398–425.
85. Specht H, Slavov N. Transformative Opportunities for Single-Cell Proteomics. *J Proteome Res.* 2018 Aug 3;17(8):2565–71.
86. Sidoli S, Garcia BA. Middle-down proteomics: a still unexploited resource for chromatin biology. *Expert Rev Proteomics.* 2017 July;14(7):617–26.
87. Tabula Muris Consortium. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature.* 2020 July;583(7817):590–5.
88. Mahdessian D, Cesnik AJ, Gnann C, Danielsson F, Stenström L, Arif M, et al. Spatiotemporal dissection of the cell cycle with single-cell proteogenomics. *Nature.* 2021 Feb;590(7847):649–54.
89. Specht H, Slavov N. Transformative Opportunities for Single-Cell Proteomics. *J Proteome Res.* 2018 Aug 3;17(8):2565–71.
90. Vanderaa C, Gatto L. Revisiting the Thorny Issue of Missing Values in Single-Cell Proteomics. *J Proteome Res.* 2023 Sept 1;22(9):2775–84.
91. Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection and processing, Library preparation and sequencing, Computational data analysis, et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature.* 2018 Oct;562(7727):367–72.
92. Stuart T, Satija R. Integrative single-cell analysis. *Nat Rev Genet.* 2019 May;20(5):257–72.

