Dissecting autonomous enzyme variability in single cells

Christian Gnann[1,2], Anthony Cesnik[1,2,3], Alina Sigaeva[1], Trang Le[1,3], Sanem Sariyar[1], Diana Mahdessian[1], Rutger Schutten[4], Preethi Raghavan[2], Manuel D. Leonetti[2], Cecilia Lindskog[4], Mathias Uhlén[1,5], Ulrika Axelsson[1], Emma Lundberg[1,2,3,6*]

1 KTH Royal Institute of Technology, Science for Life Laboratory, Department of Protein Science, Stockholm, Sweden

2 Chan Zuckerberg Biohub, San Francisco, CA, USA

3 Stanford University, Department of Bioengineering, Stanford, CA, USA

4 Uppsala University, Rudbeck Laboratory, Department of Immunology, Genetics and Pathology, Uppsala, Sweden

5 Karolinska Institutet, Department of Neuroscience, Stockholm, Sweden

6 Stanford University, Department of Pathology, Stanford, CA, USA

* Corresponding author

Abstract

Metabolic enzymes perform life-sustaining functions in various cellular compartments. Anecdotally, metabolic activity is observed to vary between genetically identical cells, which impacts drug resistance, differentiation, and immune cell activation. However, no large-scale resource systematically reporting metabolic cellular heterogeneity exists. Here, we leverage imaging-based single-cell spatial proteomics to reveal the extent of non-genetic variability of the human enzymatic proteome, as a proxy for metabolic states. Nearly two fifths of enzymes exhibit cell-to-cell variable expression, and half localize to multiple cellular compartments. Metabolic heterogeneity arises largely autonomously of cell cycling, and individual cells reestablish these myriad metabolic phenotypes over several cell divisions. These results establish that cell-to-cell enzymatic heterogeneity is an organizing principle of cell biology and one that may rewire our understanding of drug resistance, treatment design, and other aspects of medicine.

Main text

Cellular metabolism is an essential process that orchestrates the production of energy and building blocks for the cell, while also disposing of cellular waste. This is coordinated by a plethora of metabolic enzymes (including transporters), cofactors, and metabolites that form a complex dynamic network of chemical reactions. The reactions are tunable in response to intracellular or extracellular signaling, nutrient availability, cell cycle progression or circadian rhythm. Recent advances in single-cell sequencing technologies have revealed cell-to-cell heterogeneity in genetically identical cell populations.[1–3] This heterogeneity results in phenotypic differences, including variable expression of metabolic enzymes that has been linked to cancer drug resistance[4,5], metastasis[6], differentiation[7,8], and immune cell activation[9,10].

Despite these insights, studies of cellular metabolism are mostly based on bulk proteomic or metabolomic measurements[1], thereby limiting the ability to study metabolism with single-cell or subcellular resolution. Single-cell transcriptomics suffers from systematic noise and dropouts that make it difficult to study cellular heterogeneity.[11] Additionally, although gene expression roughly predicts the expression of the corresponding protein, it fails to capture cellular variability that is established post-transcriptionally.[12] Spatially resolved measurements of metabolites by matrix-assisted laser desorption ionization mass spectrometry imaging (MALDI-MSI) have revealed heterogeneity in the lipid composition in single cells.[12,13] However, these studies observe a relatively small number of molecules and fail to capture the full complexity of metabolism on the single-cell level including its subcellular organization. Mass spectrometry (MS)-based subcellular proteomics can classically assess protein localization but does not provide single-cell resolution, rendering studies of single-cell heterogeneity impossible.[14] Imaging-based subcellular proteomics using antibodies[15] or fluorescently tagged proteins[16–18] overcomes those limitations by simultaneously characterizing protein levels and distribution in individual cells with high resolution. Such subcellular measurements can reveal spatial or temporal protein expression heterogeneity,[14,19] thus providing insights into the spatiotemporal partitioning of metabolic processes.

Here, we present the first global map of the metabolic proteome with single-cell and subcellular resolution. We show that more than half of all enzymes are multilocalizing (*i.e.,* localize to multiple cellular compartments). By integrating protein interactomic data with our imaging dataset, we dissect compartment specific interactions and identify novel multifunctional enzymes. Lastly, we show that metabolism exhibits a higher amount of heterogeneity than most other biological processes, with a comparable amplitude of variability to that of the cell cycle. Autonomous cell processes have been found in organelle biogenesis and cytoskeletal organization, with oscillations that are independent but influenced by the cell cycle[20]. We observe autonomy of metabolic heterogeneity from the cell cycle in this work and evidence that it is regulated mainly post-transcriptionally, resulting in a range of distinct metabolic states within one cell population. We believe that understanding these metabolic states is important to understanding human health and disease at a molecular level in fundamental, preclinical and clinical contexts. All data are publicly available as part of the Human Protein Atlas (HPA) database (www.proteinatlas.org)[15,21].

A global subcellular map of the metabolic proteome

To generate a global map of the subcellular distribution of the human metabolic proteome with single-cell resolution, we analyzed data that we previously generated for the subcellular section of the Human Protein Atlas (HPA).[15] This resource contains the subcellular distribution of the human proteome in a wide variety of human cell lines, observed using high-throughput immunofluorescence (IF) and confocal microscopy. Target proteins were detected with the highly validated[22–25] proteome-wide HPA antibody library.

The Human1 database[26], a human genome-scale metabolic model, estimates 3,069 genes to be involved in 13,070 biochemical reactions, involving 8,369 metabolites (Supplementary Table 1). Of these genes, we have measured the subcellular distribution for 2,126 proteins, (HPAv23 Subcellular Section, 17,180 IF images, Supplementary Table 2), representing an 8-fold increase compared to other imaging-based subcellular proteomics databases[16,27] (Fig. 1a, Extended Data Fig. 1a). Additionally it contains twice the number of enzymes compared to recent single-cell MS-based proteomics datasets[28], which are furthermore unable to also resolve protein subcellular localization[29]. Finally, the number of proteins mapped experimentally to cellular structures is much higher than for UniProt (Extended Data Fig. 1b,c), making the HPA subcellular section an unprecedented resource for studying the heterogeneity of the metabolic protein expression at the single cell level.[15] To exemplify the type of data and the wide variety of cellular compartments covered, IF images for enzymes from each pathway group are shown in Fig. 1b. The monocarboxylate transporter SLC16A1 can be found in the plasma membrane, while CS, a key enzyme in the citric acid cycle (energy metabolism) and ACAT1 (amino acid metabolism) are detected in the mitochondria. DUT is involved in nucleotide metabolism in the nucleus and CYP51A1 is an enzyme in cholesterol (lipid) metabolism in the endoplasmic reticulum (Fig. 1b). Overall, we found a well-known enrichment of enzymes in compartments of cellular metabolism such as the mitochondria, cytosol and secretory organelles and lower representation of enzymes in the nuclear compartment and cytoskeletal structures (Extended Data Fig. 1d).

Compartmentalization of the metabolic proteome

A pathway-specific analysis against the background of the metabolic proteome revealed distinct spatial partitioning of metabolic pathways (Fig. 1c). For example, mitochondria harbor many enzymes involved in energy and amino acid metabolism, while transporters are overrepresented in the membrane compartments, where they play a role in shuttling metabolites between different organelles. To quantify the similarity of subcellular distribution for different sets of metabolic enzymes, we embedded the spatial information from the images into a compressed representation using a machine learning approach (**Methods**).[30] After obtaining these embeddings in the latent space (Fig 1d, Extended Data Fig. 1e, Supplementary Table 3), we measured the average pairwise cosine distance of the image embedding vectors for different proteins. Enzymes within the same metabolic pathway were significantly more similar in their subcellular distribution than random similar-size sets of proteins (Mann-Whitney U; p=1.52e-41; Fig 1e). Therefore, our proteome data confirm metabolic compartmentalization, a well-established biological phenomenon[31], where similar metabolic reactions are spatially concentrated in similar cellular regions. We identified a correlation between the number of proteins in a pathway and the degree of their compartmentalization (Extended Data Fig. 1f, g; R2 = 0.363, p=8.23e-14; no correlation was observed for

random sets of proteins), indicating that pathways with fewer proteins are spatially more constricted whereas pathways with a greater number of proteins have a greater reach across the cell. Interestingly, we identified several outliers from this trend (Extended Data Fig. 1f). For example, oxidative phosphorylation in the mitochondria as well as aminoacyl-tRNA biosynthesis in the cytosol exhibit a significantly elevated level of compartmentalization (Fig. 1f-h). Oxidative phosphorylation is catalyzed by large supramolecular complexes, such as ATP synthase which consists of 17 distinct proteins encoded by 19 different genes.[32] tRNA charging requires 20 separate enzymes, one for each amino acid. Thus, in both cases, multiple proteins are essentially carrying out a single step in a metabolic pathway. Conversely, cholesterol biosynthesis follows the general trend and is less compartmentalized as it is tightly regulated in the endoplasmic reticulum but occurs in additional organelles such as the peroxisomes or the nuclear membrane (Fig. 1f,g,i).[33]

In conclusion, the subcellular resolution of our dataset allowed us to precisely chart the spatial compartmentalization of the metabolic proteome, revealing that metabolic networks can span several organelles, thereby establishing a complex subcellular landscape. Although partitioning of metabolic reactions has been long known, this is the first time it has been demonstrated at a global metabolic proteome level in single cells.

Non-genetic enzymatic heterogeneity establishes diverse metabolic phenotypes in cell populations

The single-cell resolution of imaging-based proteomics makes it possible to assess the dynamic properties of the metabolic proteome. Enzymes that are temporally or otherwise regulated in terms of abundance or localization will appear as heterogeneous cell populations in our static images of asynchronous cell populations.

We previously established that a large portion of the human proteome (22.5%, 2,960 of 13,147 proteins) shows significant cell-to-cell variability in protein expression levels or spatial distribution (Fig. 2a).[15,34] Strikingly, metabolic enzymes across all pathway groups exhibit a significantly higher degree of heterogeneous expression compared to the entire HPA mapped proteome (805 of 2,126 enzymes, 37.9%, *i.e.,* 68.4% higher than whole proteome, p=4.7e-47, binomial test; Fig. 2b). Interestingly, mitochondrial enzymes displayed a particularly high degree of cellular variability (p=7.0e-6, binomial test, BH corrected; Fig. 2c, Extended Data Fig. 2a). To further investigate the biological effects of metabolic heterogeneity on cellular function, we performed a gene ontology (GO)-term enrichment analysis of variably and stably expressed enzymes against the background of the entire metabolic proteome (Supplementary Table 4). For stably expressed enzymes, this analysis revealed a significant enrichment for terms related to protein ubiquitination, DNA damage response and Wnt signaling. For variably expressed ones, we found an enrichment of genes involved in central metabolism (*e.g.,* ATP generation and glycolysis), as well as nucleotide metabolism (Fig. 2d). Altogether, our data suggest a biological role for metabolic heterogeneity, although its origins and exact biological functions remain to be determined.

Metabolic heterogeneity is conserved *in situ* and manifested in the lineage of a single cell

More than half of these enzymes (396 of 733, 54%) showed similar cell-to-cell variation in more than one human cell line (Fig. 2e, Extended Data Fig. 2b), suggesting that they are controlled by conserved regulatory mechanisms. To confirm that the phenotypic heterogeneity is not driven by genetic differences caused by increased mutation rates in cancer cells[35], we

performed a clonal expansion experiment of mNG-tagged Hek293T cells. This key experiment demonstrated that individual cells have the capacity to reproduce highly heterogeneous cell populations, thus establishing myriad functional phenotypes over just a few cell divisions. Heterogeneity observed in the previous dataset was confirmed for several metabolic enzymes belonging to different pathways, including ENO1 and GAPDH (glycolysis), ASS1 (arginine metabolism), PHGDH (serine metabolism), NANS (nucleotide sugar metabolism) and HMGCS1 (cholesterol biosynthesis) (Fig 2f). To further validate metabolic heterogeneity in clonal lineages, we annotated images corresponding to 246 metabolic enzymes from a CRISPR tagging screen in HEK293T and HAP cells by Reicher et al (Supplementary Table 5).[27] Their results show heterogeneity for over half of the enzymes included in the dataset (53%, 131 of 246). Out of the enzymes showing similar localization to the HPA, a majority show conserved heterogeneity across cell lines (59%, 101 out of 171), similar to our previous analysis across different cell lines in the HPA dataset (Fig 2e). Taken together, these results highlight metabolic heterogeneity as an inherent mechanism of cell biology.

Metabolic heterogeneity *in situ* is impacted by different cell lineages and interactions between multiple cell types. To confirm that human metabolic enzymes exhibit robust phenotypic heterogeneity beyond cell line populations, we analyzed data from a MS-based study of murine liver zonation with single cell resolution.[36] Here, Rosenberger, *et al.,* identified over 700 proteins with varied expression related to the distance from central vein and portal triads, and over half were enzymes (384 of 742, 51.7%). Out of those enzymes, we observe a significantly higher proportion with variable expression in our dataset (137 of 286, 47.9%, p=0.00034, binomial test) compared to the entire metabolic proteome (805 of 2126 enzymes, 37.9%). For example, the well known zonation markers CYP2E1 and ASS1 display heterogeneity in cell lines (Extended Data Fig. 2c, Fig. 2a,e,f) and increase in expression towards the central vein (CYP2E1) or the portal triad (ASS1) (Fig. 2g). The variable enzymes HMGCS1 (Fig. 2a,e,f) and NDUFA1 (Fig. 2a), an enzyme in oxidative phosphorylation, display similar zonation characteristics to ASS1, with elevated expression towards the portal vein (Fig. 2g). Those findings are in line with previous studies confirming that oxidative phosphorylation is more active in the periportal zone.[37,38] Furthermore, they exemplify how metabolic enzyme expression heterogeneity can give rise to significant differences in the metabolic capacity of genetically identical cells and thus diversify cellular phenotypes. Altogether, we show that cell line systems are a suitable model to study non-genetic heterogeneity and can serve as a valuable tool to understand cellular plasticity *in situ*.

Autonomous metabolic heterogeneity is established post-transcriptionally

The cell cycle is a major contributor to cellular heterogeneity, and we interrogated whether it could possibly explain the staggering amount of metabolic heterogeneity in our data using a targeted single-cell proteogenomic screen of the U-2 OS cell line with precise cell cycle mapping over interphase.[34] Strikingly, as much as 76.6% of enzymes (188 of 245) included in the study displayed non-CCD protein expression heterogeneity. Interestingly, this is the case for so-called "housekeeping genes", such as enzymes involved in glycolysis or the citric acid cycle, including GAPDH, PDHB and LDHB. They are stably expressed on the transcriptional level, but exhibit significant variability of protein expression[39] that is independent of cell cycle progression (Fig. 3a). As a result, and in line with previous studies,[12,40,41] our results show that single-cell transcriptomic data alone is a poor predictor of phenotypic metabolic states. In fact, while the variability in protein expression for non-CCD enzymes is comparable to that for CCD proteins, the variability observed for their corresponding

RNAs is lower (p=0.0030, respectively, Kruskal-Wallis test, Fig. 3b) and independent of their overall expression levels (Fig S2D). Furthermore, the physical properties of variable metabolic enzymes differ from CCD proteins, with lower protein disorder fractions and elevated melting temperatures (Extended Data Fig. 2e,f, Supplementary Table 6). Those findings point to fundamental differences in how enzyme variability and protein cycling are regulated, such as through upstream kinase families[42].

We investigated differences in PTM regulation of metabolic enzymes compared to cycling proteins by comparing the kinases upstream of phosphosites on these proteins (Supplementary Table 7). While cycling proteins are enriched for CMGC kinases upstream of phosphosites (p=5.0e-15), which was reported previously[34] and is tied in function to regulating cell fate[43], metabolic enzymes are enriched for AGC and TK kinase families (p=0.0016 and 0.00078), which have respective functions in metabolism[44] and cell signaling[45,46]. Enzymes that exhibit variability between individual cells display a strong enrichment for upstream TKs when compared to non-variable enzymes (p=6.9e-7). This indicates that metabolic enzymes have distinct PTM regulation and that the regulation of enzymes displaying variable expression between individual cells is regulated by machinery involved in cell signaling transduction and cell-to-cell communication.

Importantly, the cell cycle and metabolic heterogeneity can still influence each other. HMGCS1 is an enzyme involved in cholesterol biosynthesis that has variable, non-CCD expression levels of both RNA and protein (Fig. 3a). However, time lapse microscopy reveals a coordinated CCD translocation of cytoplasmic HMGCS1 to the nucleus specifically during interphase (Fig. 3c), showing how different layers of heterogeneity can come together in the same cell. In this case, the differences in expression levels establish differences in metabolic capacity between individual cells, while the location of the protein is linked to cell cycle position. This result also highlights the fact that gene expression heterogeneity is neither able to capture the subcellular localization of a protein nor the spatial heterogeneity established by protein translocation. In conclusion, we find that metabolic heterogeneity is largely established independently of the cell cycle, mostly regulated post-transcriptionally, for example through translocation[47], targeted degradation[48], and post-translational modification[48] of enzymes.

In conclusion, we observe cellular heterogeneity across all metabolic pathways (Fig. 4) potentially establishing a plethora of metabolic states. This widespread variation in enzyme expression creates a need to reinspect the wiring of metabolic flux diagrams and may change our understanding of the metabolic regulation of cell types and single cells (Fig. 4).

The intricate subcellular landscape of the metabolic proteome

Multilocalizing enzymes, including HMGCS1, open up intriguing questions about the interconnectedness of enzyme subcellular distributions and corresponding metabolic processes. With the subcellular resolution of imaging proteomics, we found that more than half of all metabolic proteins are multi localizing (53.8%, 1,144 of 2,126; Fig. 5a,b), which is comparable to the entire proteome (56.5%, 7,434 of 13,147, p=0.011, binomial test; Extended Data Fig. 3a)[15,16,49,50]. Many enzymes localize to both the cytoplasmic and nuclear compartments, such as ALAS1 (mitochondria and nucleoplasm), NDUFB4 (mitochondria and nuclear membrane) and GCH1 (cytosol and nucleoplasm), while other enzymes localize to multiple structures in one meta-compartment, such as the transporter SLC7A5 (cytoplasm and vesicles; Fig. 5a). As

mentioned previously, enzymes are overrepresented in the cytoplasm (p = 5.0e-31, Fisher exact) and are less likely to be present in the nucleus compared to the entire human proteome (p = 1.3e-28, Fisher exact; Extended Data Fig. 1d). However, multilocalizing enzymes are more likely to be present in the nucleus compared to unilocalizing enzymes, *i.e.,* those found in a single location (Extended Data Fig. 3b).

Multilocalizing enzymes are functionally understudied

Our data indicate that 54% of proteins localize to multiple compartments, which is likely an underestimation, since spatiotemporal protein expression varies across cell types and contexts.[51] We define multilocalizing proteins as potentially multifunctional proteins. Text-mining studies predict that over a quarter of all proteins can carry out multiple functions[52,53] establishing a great discrepancy between the number of multilocalizing proteins and the number of multifunctional proteins with a strong overrepresentation of enzymes among known moonlighting proteins (39 of 86, 45%, p = 2.1e-10, binomial test)[54,55]. We compared the location data from the HPA subcellular section to experimentally verified location data extracted from the curated database UniProt (after subtracting locations derived from HPA; Supplementary Table 8). We found a significantly higher proportion of multilocalizing proteins in the HPA dataset than in UniProt across all metabolic pathways (p = 9.8e-32, Fisher exact, Benjamini-Hochberg (BH) corrected; Extended Data Fig. 3c) with an overrepresentation of nuclear localizations (Fig. 5c) across many metabolic pathways (Fig. 5 d,e, Extended Data Fig. 4a-i). For example, IF imaging revealed a localization to both the cytosol and nucleus for ALDH3A1 and ACLY, as well as to the Golgi apparatus and nucleus for CERT1 validated with multiple independent antibodies, while there is currently no experimental evidence for a nuclear location in the UniProt database for any of those proteins (Extended Data Fig. 3d). Those findings could point to uncharacterized non-metabolic nuclear functions for metabolic enzymes,[56] similar to the recently described involvement of Golgi proteins in DNA damage.[57] Interestingly, multilocalizing enzymes are more likely to exhibit variable expression than enzymes localizing to a single cellular compartment (35.8.% of unilocalizing proteins are variably expressed, 39.8% of multilocalizing; p = 7e-4 Fisher's exact test). The combination of cellular heterogeneity and multilocalization can greatly expand the functional diversity of cells in a population.

The variable expression patterns of ENO1, GAPDH and PDHB in the nucleus and cytoplasm (Extended Data Fig. 5a) serve as great examples of how the spatiotemporal regulation of enzymes establishes diverse functional capacities among cells. PDH is a complex that generates acetyl-CoA in the mitochondria and the nucleus. Mitochondrial acetyl-CoA is used in the TCA cycle while nuclear acetyl-CoA is crucial for histone acetylation.[58] Our data show stable expression of PDH subunit beta (PHDB) in the nucleus, while mitochondrial PDHB levels vary between individual cells. These findings indicate a decoupling of the two processes (Extended Data Fig. 5a) and compartment-specific regulation of PDHB levels. In other cases, multilocalizing proteins can perform distinct functions in different compartments. The glycolytic enzyme GAPDH is one such multifunctional enzyme; it carries out its canonical glycolytic function in the cytosol, while it is also involved in apoptotic pathways and transcriptional gene regulation in the nucleus (Extended Data Fig. 5a).[59] ENO1 is another glycolytic enzyme in the cytosol, but it also functions as a plasminogen receptor on the cell surface and has RNA/DNA regulatory functions in the nucleus[60]. Altogether, our analysis reveals new insights into multilocalizing characteristics of metabolic enzymes and highlights the need for additional studies to decipher their non-canonical functions.

Identification of multifunctional enzymes

Moonlighting enzymes perform multiple functions in different parts of the cell. To find such multifunctional proteins[61], we combined localization information from our global spatial proteomic data with protein-protein interaction (PPI) measurements that capture the molecular-scale local environment of proteins. We assembled enzyme interaction partners from various studies[16,62,63], as well as from the STRING database[64] (filtered for physical interactions; experimental evidence and confidence > 0.7), and constructed interaction networks for each multilocalizing metabolic enzyme (Supplementary Table 9). This allowed us to assess non-metabolic functions using a guilty-by-association approach. For example, P4HA2, an enzyme involved in collagen synthesis in the ER (Fig. 5f), interacts with other members of the collagen synthesis pathway in the cytoplasm. However, in the nucleus, it interacts with members of the CPSF complex and may be involved in mRNA polyadenylation as well as alternative splicing (Fig 5g, Extended Data Fig. 5b). Interestingly, recent studies have revealed that many metabolic proteins are capable of RNA binding, indicating that a large number of enzymes could have non-canonical functions in cell biology.[65,66] One example is ATP6V1D, which is a subunit of the vacuolar ATPase, a proton pump responsible for controlling the intra- and extracellular pH of cells (Fig. 5h). It interacts with NOTCH1 and USB7 in the nucleus and could thus play an additional role in the regulation of NOTCH signaling (Fig. 5i, Extended Data Fig. 5c). Altogether, combining spatial proteomic and interactomic analyses revealed potential non-canonical protein functions, including for well characterized enzymes such as P4HA2 and ATP6V1D, and provides a promising avenue for future systematic studies of multifunctional enzymes.

Exploration of metabolic proteome phenotypes in the HPA database

This image-based map of the metabolic proteome is freely available as part of the Human Protein Atlas database (www.proteinatlas.org/humanproteome/subcellular/metabolic+proteome v 24). The subtle variations in protein expression and subcellular localization between cells from undisturbed, well-characterized cell lines can hint at novel protein functions and cell-specific regulatory mechanisms. These findings offer a robust baseline for future research into genetic, environmental, and pharmacological perturbations of metabolic cellular phenotypes. Moreover, as antibody-based staining of histological samples is a routine procedure in clinical contexts, our resource allows for direct comparison between cell lines and tissues. Altogether, these data will facilitate constructing spatially resolved proteome-wide metabolic flux models to better understand heterogeneous drug responses in cell populations and ultimately provide new opportunities to drug specific cell states.

Conclusion

In this study, we describe the heterogeneity of the human metabolic proteome on single-cell and subcellular levels. We show pervasive non-genetic heterogeneity of metabolic enzyme expression. Over half of all enzymes localize to multiple locations in human cells, including several newly identified multifunctional proteins, and nearly two fifths exhibit cell-to-cell variable expression that is largely autonomous of the cell cycle. This heterogeneity is recapitulated in many cell lines, native tissue contexts, and clonal expansions of individual cells, and appears to be established at a post-transcriptional level.

This imaging-based subcellular proteomic dataset will be a valuable resource across many areas of research, including cell biology, metabolic modeling or pathology.

The compartmentalization of metabolism is a well established biological concept, yet global maps of metabolic compartmentalization are still lacking. In this study, we provide experimental evidence for the subcellular localization of over 2,100 metabolic enzymes, demonstrating global metabolic compartmentalization on the proteomic level. While the enzyme activity is certainly linked to subcellular localization and colocalization with other enzymes, it also depends on the presence of respective precursors, products and inhibitors. Recent developments in MALDI-MSI allow for the subcellular profiling of lipids and other metabolites[2,13], but future work is required to dissect the role of metabolite compartmentalization[67]. Additionally, enzyme activity can be regulated by post-transcriptional modifications, such as phosphorylation, which remains insufficiently characterized[42], especially with regards to spatial information. We showed that enzymes displaying cell-to-cell variable expression are particularly enriched for specific upstream tyrosine kinases involved in cell signaling when compared to all enzymes.

Protein localization is an important layer of cellular regulation that cannot be inferred from transcript or protein expression data alone[14]. We find that over half of all metabolic enzymes localize to multiple compartments, greatly expanding their context-dependent functionalities. Even so, the secondary functions of enzymes are rarely described. We show that compartment-resolved PPI data have the potential to unravel novel functions for multilocalizing proteins and provide a suitable starting point for in-depth functional studies of protein multifunctionality. Importantly, our imaging-based spatial proteomics approach does not distinguish between proteoforms[68]. As such, protein multilocalization and multifunctionality may be driven by post-translational modifications[69] or structural changes.[70]

Cellular metabolism is an extremely dynamic process that can change over the cell cycle, circadian rhythm, or following signaling events, yet global maps of metabolic variability are missing. In this study, we provide experimental evidence that 805 metabolic enzymes demonstrate variability in expression in terms of abundance or localization. This widespread variation in enzyme expression creates a need to reinspect the wiring of metabolic flux diagrams (*e.g.,* for pathways illustrated in Fig. 4). Metabolic flux is often assessed using bulk transcriptomic data, which is unable to resolve cell-to-cell enzymatic heterogeneity[71]. However, single-cell transcriptomic data alone is unlikely to accurately capture metabolic states, since it is a poor predictor of single-cell protein expression[34,72–75] and metabolite levels[12]. Consequently, our single-cell proteome-wide metabolic map provides a valuable step towards improved metabolic modeling[71] and may change our understanding of the metabolic regulation of cell types.

Inherent heterogeneity of metabolic protein expression may serve the purpose of a bet-hedging evolutionary strategy that allows entire cell populations to survive in rapidly changing environments, such as during drug treatments[76,77], as has been demonstrated in yeast,[78] bacteria[79] and cancer cells.[80] Sufficient heterogeneity may contribute to the resilience of the cell population as a whole. Reduced heterogeneity (*e.g.,* in aging, stress, or disease) may deplete the potential of the cell population to restore cell function and survive under a wider range of conditions. At the same time, the excessive

heterogeneity (e.g. observed in aging yeast cells[81,82] might indicate that the cells are no longer able to maintain the tight regulation of their metabolism that is necessary for effective and efficient cell functioning.

An immense number of possible metabolic states are possible, given the 800 metabolic enzymes that we show to have variable expression. Even so, established biochemical frameworks describe a relatively small number of high-level, often mutually exclusive metabolic states (switches), such as glycolysis-dominated or respiration-dominated energy production. Co-regulated enzymatic activity within a metabolic pathway can lead to co-variation of enzyme expression. In this case, it may be possible to pinpoint highly variable metabolic enzymes that are markers for distinct high-level metabolic states within the cell population. However, this co-variation is not absolute; one enzyme can show relatively higher expression than other enzymes in the same pathway if its product is required for another metabolic pathway or if it is performing several different functions.[83] Mapping such variations may reveal the complex landscape of more granular metabolic phenotypes, including the connections, dependencies and compensatory interactions between different metabolic pathways at the single-cell level. These diverse enzymatic expression patterns may also give rise to convergent metabolic phenotypes. This phenomenon was recently described for the protein components of the circadian clock, each of which can have widely varying turnover rates in individual cells yet show robust circadian rhythm periods as a whole population.[84] Converged high-level phenotypes like the circadian rhythm are still underpinned by diverse expression of their individual components, which notably opens up diverse response potential for cells when those individual components are targeted by drug treatments. Better understanding cellular metabolic states facilitates manipulating them and developing novel treatment strategies for diseases, such as cancer, to improve the efficacy of existing drugs and overcome drug resistance.

Measuring metabolic states is a difficult task, and the ways in which metabolic states can converge based on variable enzyme expression, such as was described above for circadian rhythms, point to the potential of enzyme (co)expression and (co)localization to be useful proxies for studying metabolic states. Additionally, the simultaneous spatial profiling of multiple omics layers may be able to measure these proxies and provide novel insights into the regulation of cellular metabolism.[85]

In conclusion, the single-cell map of autonomous enzymatic variability presented in this paper underlines the importance of studying metabolic heterogeneity in many areas of cell biology. This new perspective could revolutionize many areas of medicine, including the development of new treatment designs to avoid drug resistance.

Figures

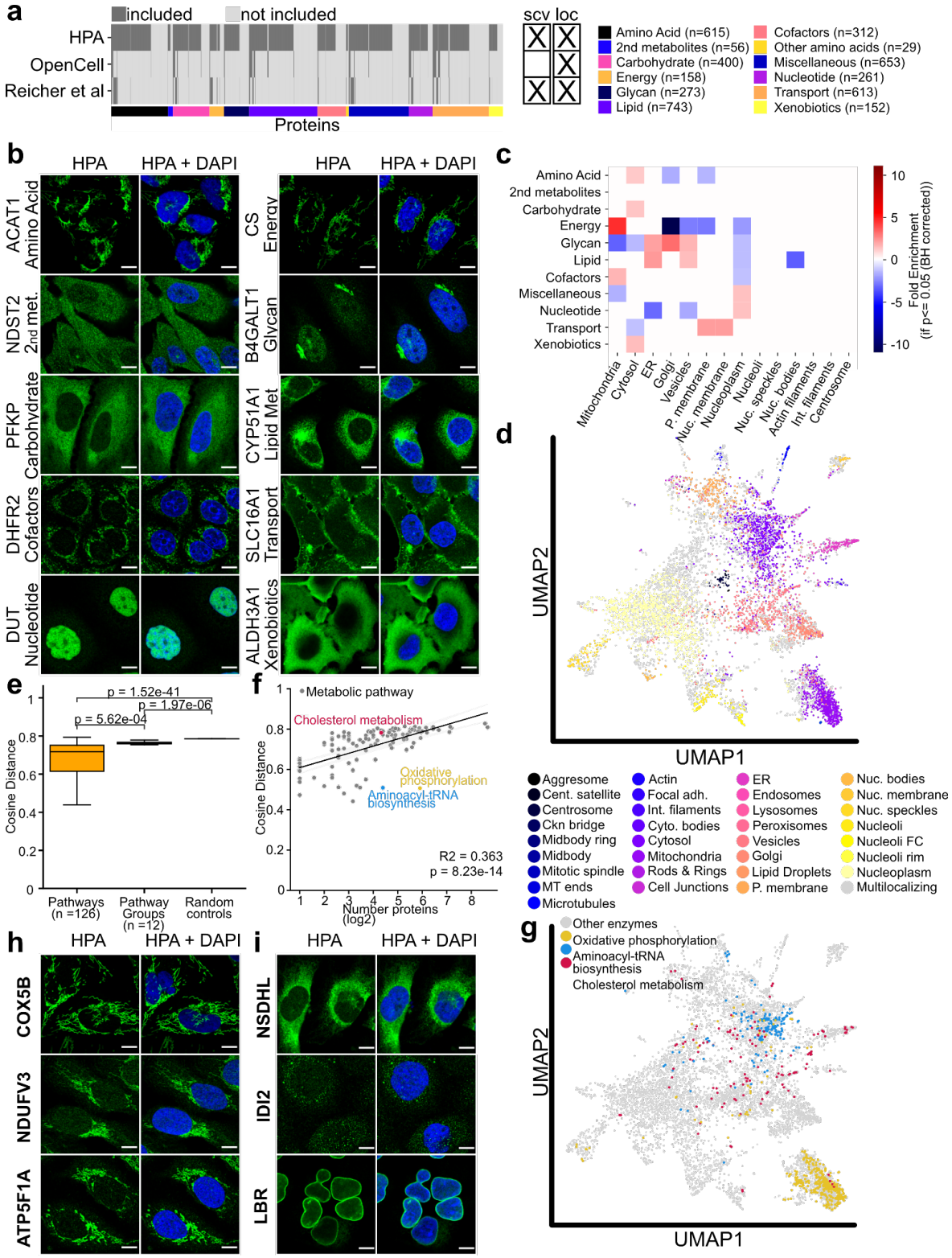

Fig. 1: The metabolic proteome.

a, Coverage of metabolic enzymes across pathway groups in different imaging-based subcellular proteomics dataset. The HPA dataset uses a proteome-wide antibody library and allows for the simultaneous study of single-cell heterogeneity and protein location. The OpenCell[16] and Reicher[27] datasets rely on endogenous tagging for protein localization. OpenCell

profiles a polyclonal pool of CRISPR/Cas9 edited cells while the Reicher study generated clonal cell populations. As a result, OpenCell is unable to resolve phenotypic cell-to-cell enzymatic heterogeneity from genetic heterogeneity. (scv = single-cell variability; loc = location) **b,** Example images for metabolic enzymes across all metabolic pathway groups, blue = DAPI, green = protein of interest; scale bar corresponds to 10 µm. **c,** Overrepresentation analysis for enzyme localisation in the HPA dataset across different metabolic pathways. Enrichment is shown for significantly overrepresented protein location (P >= 0.05; binomial test). **d,** UMAP visualization of the image features of the entire HPA image dataset.[30] Metabolic enzymes are visualized. Unilocalizing protein images are colored according to location, while gray data points correspond to multilocalizing proteins. **e,** Pairwise cosine distances calculated for proteins within the same pathway, pathway group, and across random controls in the latent space from image-classification model. Centre line, median; box, first (Q1) and third (Q3) quartiles; whiskers, 1.5× interquartile range (IQR) below Q1 and above Q3. Enzymes within the same pathways are significantly more compartmentalized compared to random protein sets (Mann-Whitney U test; p = 1.52e-41 **f,** Logarithmic correlation between the cosine distance from D and the number of proteins in a pathway. Cholesterol biosynthesis (red), oxidative phosphorylation (blue) and aminoacyl-tRNA biosynthesis (yellow) are highlighted. **g,** UMAP from c, proteins in cholesterol biosynthesis (red), oxidative phosphorylation (blue) and aminoacyl-tRNA biosynthesis (yellow) are highlighted. **h,** Example images for enzymes in oxidative phosphorylation (blue = DAPI, green = protein of interest; scale bar corresponds to 10 µm). **i,** Example images for enzymes in cholesterol biosynthesis (blue = DAPI, green = protein of interest; scale bar corresponds to 10 µm).

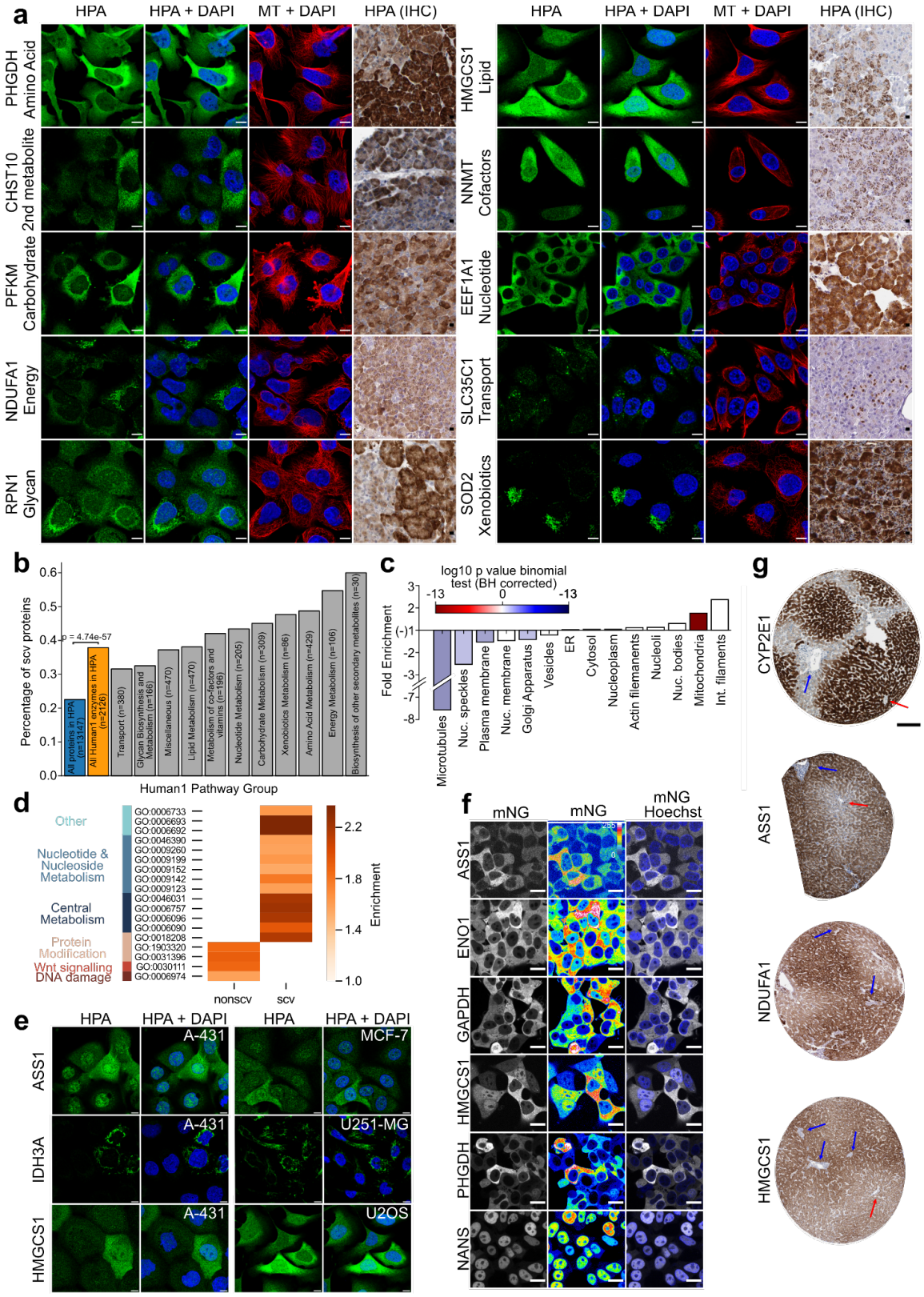

Fig. 2: Metabolic heterogeneity establishes phenotypic states.

a, Example images in cell lines and pancreatic acinar cells for variable metabolic enzymes across all metabolic pathway groups, blue = DAPI, green = protein of interest, red = microtubules; scale bar corresponds to 10 µm. **b,** Enzymes exhibit a higher degree of heterogeneity compared to the human proteome. **c,** GO-term enrichment analysis of variably and non-variably expressed enzymes against the background of the metabolic proteome. Fold-change >1.5 and P <= 0.05 were used

as thresholds. **d,** Immunofluorescence images of ASS1 (top), IDH3A (middle) and HMGCS1 (bottom) reveal conserved single-cell heterogeneity across multiple cell lines. **e,** Overrepresentation analysis for variable enzyme localisation in the HPA dataset compared to the background of the metabolic proteome. **f,** A single cell can reproduce heterogeneous populations, as shown in clonal expansion experiments of mNG-tagged cell lines, grey = mNG tagged protein of interest, blue = Hoechst. **g,** Cellular heterogeneity is conserved in tissue. Immunohistochemistry images for metabolic zonation in liver tissue. Red arrows indicate central vein, blue arrows display portal triads. Scale bar corresponds to 200 µm.

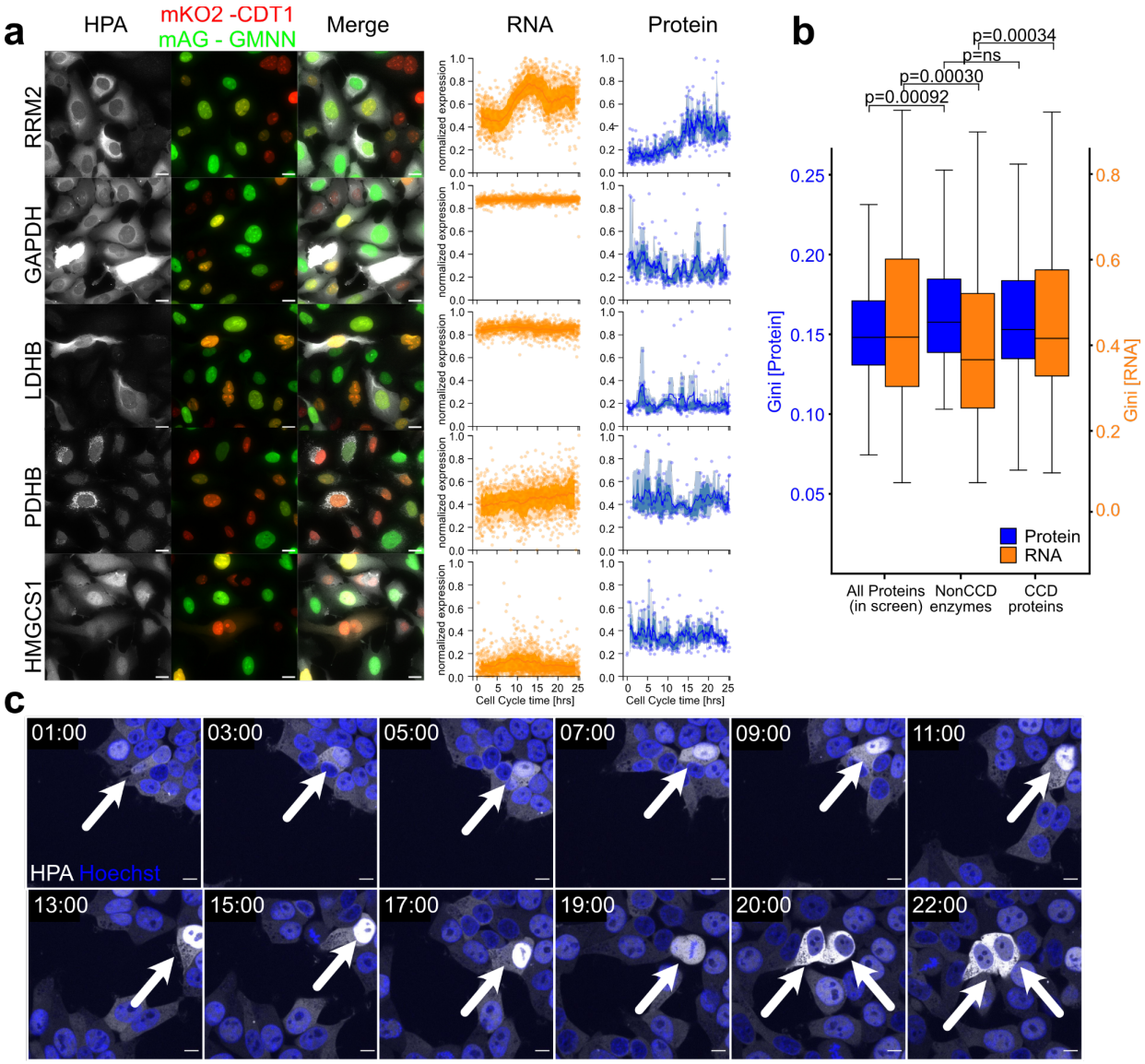

Fig. 3: Metabolic heterogeneity is exhibited independently of cell cycle progression and established post-transcriptionally.

a, Left to middle: Example immunofluorescence images for CCD (RRM2) and non-CCD enzymes (GAPDH, ENO1, PDHB, HMGCS1) in U2OS-FUCCI cells from Mahdessian *et al*[34], green = mAG-GMNN, red = mKO2-CDT2, grey = protein of interest; scale bar corresponds to 10 µm. Middle and right, single-cell temporal expression profiles for RNA (orange) and protein (blue). Line, moving mean; darker shade, 25th to 75th percentile range; lighter shade, 10th to 90th percentile range; points, individual cell data. **b,** Transcript profiles (orange) do not reflect the variability observed at the protein level (blue), and in fact while variability of non-CCD enzymes is higher than all proteins, the variability of their corresponding RNAs is lower than all proteins in the screen. For box plots: center line, median; box, first (Q1) and third (Q3) quartiles; whiskers, 1.5× interquartile range (IQR) below Q1 and above Q3 **c,** Time-lapse imaging data for mNG-tagged HMGCS1 protein (white) demonstrates cytoplasmic localization in G1, and then translocation to the nucleus (blue); scale bar corresponds to 10 µm. However, the overall expression levels of HMGCS1 remain stable, corroborating the findings of cell cycle independent expression heterogeneity from A.

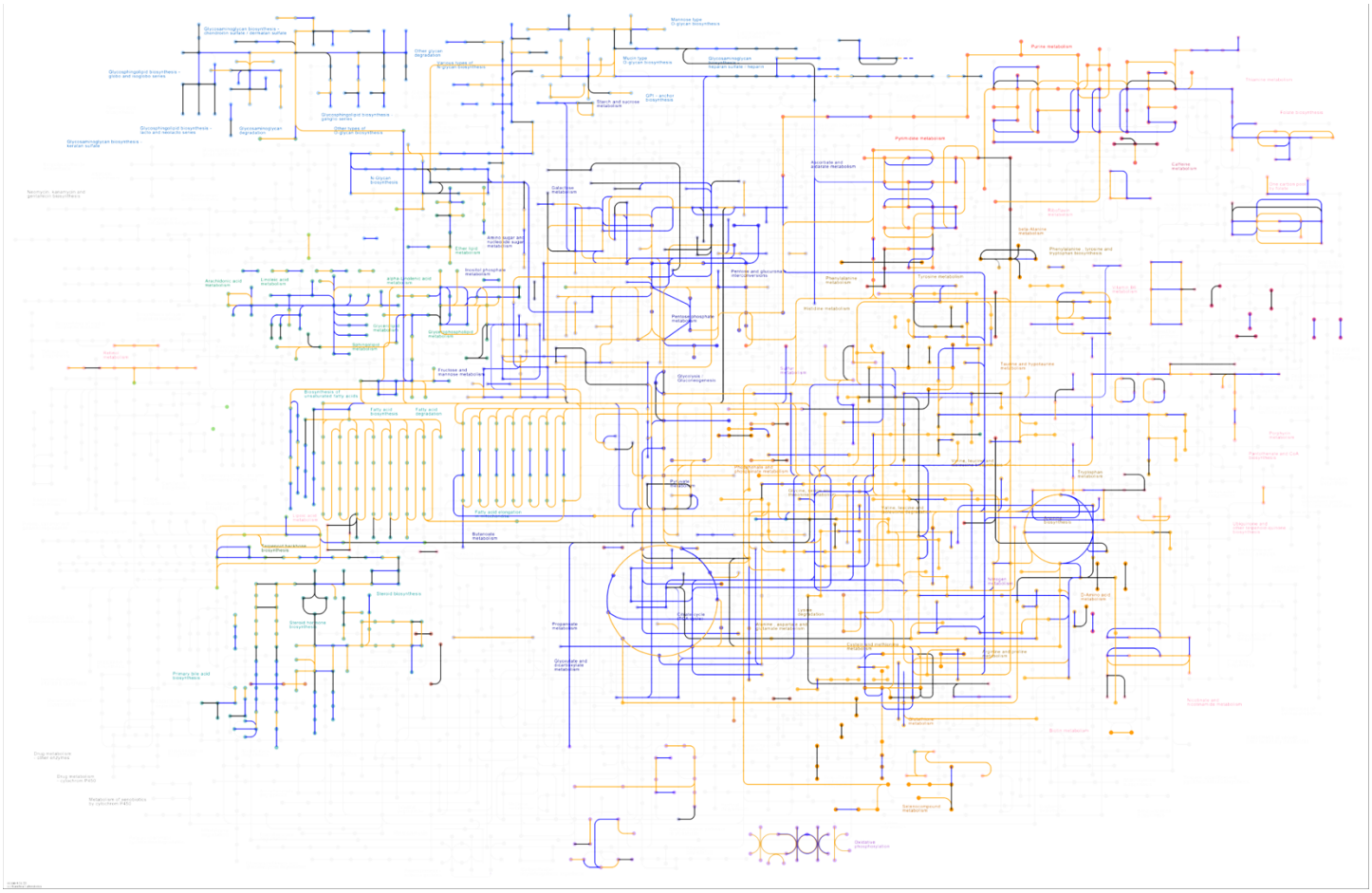

Fig. 4: Global reach of metabolic variability

KEGG metabolic pathway diagram highlighting human metabolic reactions (orange, blue, and black lines). The variability of enzymes is widespread: the reactions containing at least one protein displaying such variability are highlighted with orange lines, and the other reactions included in this study are highlighted with blue lines.

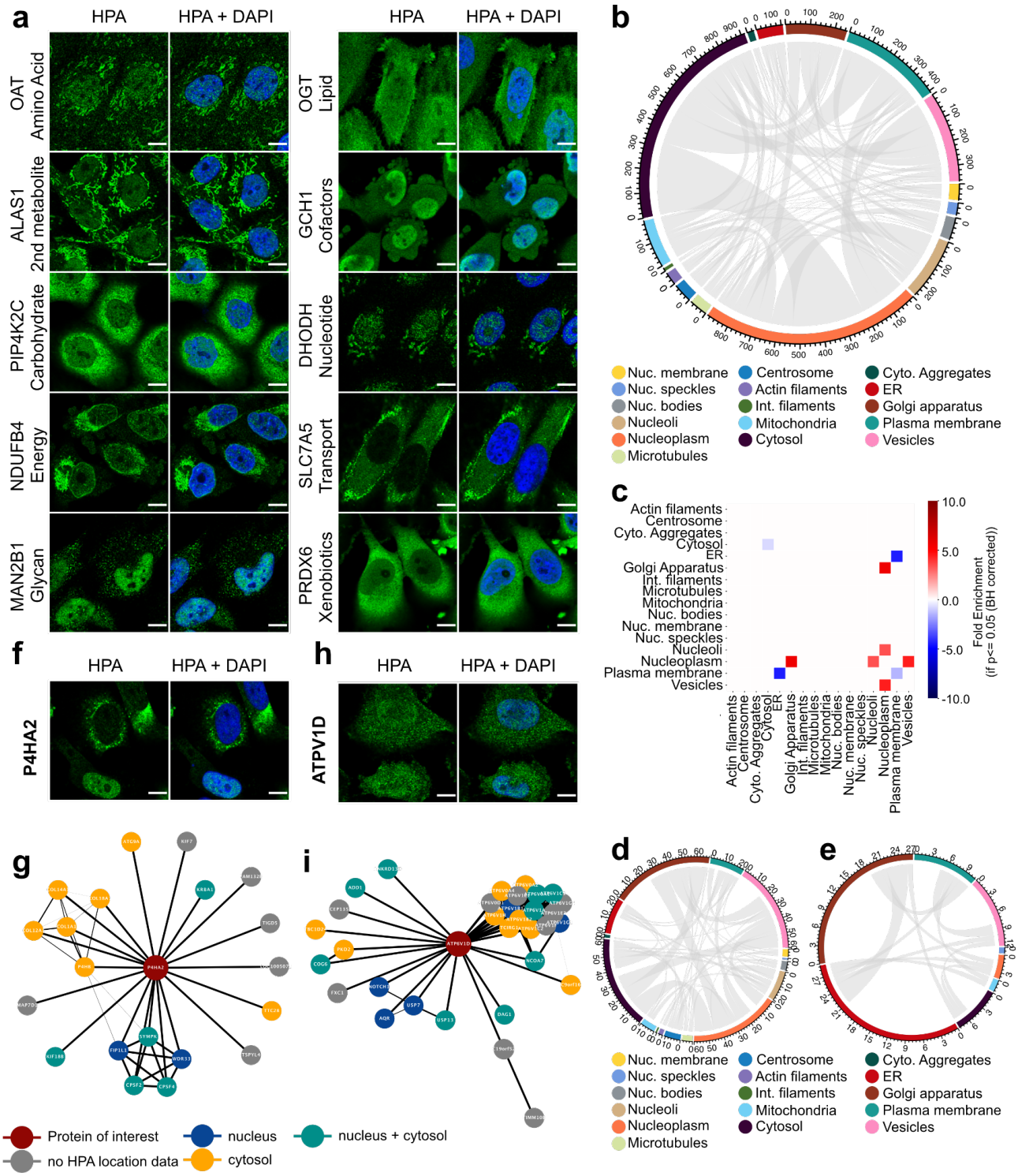

Fig. 5: The intricate subcellular landscape of the metabolic proteome.

a, Example images for multilocalizing metabolic enzymes across all metabolic pathway groups (blue = DAPI, green = protein of interest; scale bar corresponds to 10 µm). **b,** Circos plots with the identified proteins of each compartment presented and sorted by meta-compartments (cytoplasm = red, secretory pathway = green, nucleus = blue). Multilocalizing proteins are connected by a line (secondary colors represent multilocalization across metacompartments). The plots show connections among metabolic proteins in the HPA dataset. **c,** Overrepresentation analysis for enzyme co-localisation in the HPA dataset compared to UniProt localisation information. Enrichment is shown for significantly

overrepresented location pairs (P >= 0.05; binomial test, BH corrected). **d, e,** Circos plots with the identified proteins of each compartment presented and sorted by meta-compartments (cytoplasm = red, secretory pathway = green, nucleus = blue). Multilocalizing proteins are connected by a line (secondary colors represent multilocalization across metacompartments). The plots show connections among **(d)** enzymes involved in glycan biosynthesis in the HPA dataset and **(e)** enzymes involved in glycan biosynthesis with experimental evidence on UniProt. **f,** P4HA2 localizes to the ER and the nucleus (blue = DAPI, green = protein of interest, scale bar corresponds to 10 µm). **g,** Protein-protein interaction network of interactors of P4HA2, colored by protein localization. P4HA2 interacts with a cluster of proteins in the cytoplasm and the nucleus. **h,** ATP6V1D localizes to the cytosol and the nucleus (blue = DAPI, green = protein of interest, scale bar corresponds to 10 µm). **i,** Protein-protein interaction network of ATP6V1D, colored by protein localization. P4HA2 interacts with a cluster of proteins in the cytoplasm and the nucleus.

Materials and Methods

HPA antibodies and validation

All antibodies within the HPA project are rabbit polyclonal antibodies derived after immunization with recombinant protein epitope signature tags (PrEST) as antigens and purification using the antigen as an affinity ligand.[22] Antibodies were then quality-controlled for sensitivity and crossreactivity using western blotting and protein arrays before used in immunocytochemistry experiments for the Subcellular Section of the Atlas. The subcellular protein localization results are assigned a gene-specific confidence score. If there were no independent experimental data on uniProt contradicting the results obtained in the immunofluorescence experiment, the antibodies were labeled "approved", if the independent supported the results, the antibodies were labeled "supported". Antibodies validated according to the strategies outlined by the International Working Group for Antibody Validation (IWGAV)[86] were labeled "enhanced". These strategies include i) validation by co-staining with fluorescently tagged protein; ii) validation by gene silencing or knock out; iii) validation by immunocapture and mass spectrometry and iv) validation by independent antibody.[86]

Subcellular localization of proteins using immunofluorescence

The distinct localization patterns for each protein assigned on the basis of annotations of the target protein in the HPA Subcellular Section (Supplementary Table 2). In short, a standardized protocol for immunofluorescence was applied[87] and fluorescent images with confocal resolution were acquired as described in Thul et al.[15] Finally, the protein localizations were manually determined based on the signal pattern of the protein of interest in relation to cellular markers (DAPI, microtubules and the ER).

Identification of proteins with cell-to-cell heterogeneity as well as multilocalizing proteins

Protein cell-to-cell heterogeneity and multilocalization were manually identified in images from the HPA Cell Atlas (HPA v23).[15] Cell-to-cell heterogeneity was identified either as variation in staining intensity (= protein abundance) or variation in subcellular distribution between cells in the same microscopic image. Multilocalizing proteins were defined as proteins being present in multiple locations within or across all cell lines and/or images in the Subcellular Section of the HPA.

Simplifying the HPA protein location data

We extracted the protein location data from HPA v23 and collapsed the annotations for all 35 subcellular structures into 14 major organelles (Mitochondria, Cytosol, Cytoplasmic Aggregates, Endoplasmic Reticulum (ER), Golgi Apparatus, Vesicles, Plasma membrane, nuclear membrane, Nucleoplasm, Nucleoli, Nuclear speckles, nuclear bodies, Actin filaments, Intermediate filaments, Centrosome) that were used for the analysis according to Extended Data Table 1.

Definition of the metabolic proteome, pathways and pathway groups

The metabolic proteome was defined according to the Human1 database, a human genome-scale metabolic model containing 13,070 biochemical reactions with 8,369 metabolites and over 3,069 genes.[26] The Human1 database provides evidence about

the metabolic reactions and pathways for each enzyme and thus allows for grouping the enzymes into different metabolic pathways and pathway groups (Supplementary Table 1). 2,125 genes have corresponding protein data with available protein localization data in the HPA Subcellular Section v23 (Fig. 1a). The good coverage for enzymes across all pathway groups in the HPA Subcellular Section v23 enables a detailed analysis of the metabolic proteome using imaging-based Spatial Proteomics.

Subcellular location overrepresentation analysis

For the location overrepresentation analysis we extracted the number of proteins for each subcellular location across our test and background datasets. We calculated the overrepresentation of the location in the test vs the background protein set and performed a two-sided binomial test to test for statistical significance. The assumptions of this test were satisfied as follows: the data were large and nominal, the sample size was less than the population size, the samples were independent, and the probability of a given outcome did not affect the probability of the other. A p-value < 0.05 after Benjamini Hochberg correction was used as a cutoff. The significant enrichment for all locations in the test set versus the background set were visualized in a heatmap or barplot.

Cosine distance analysis

We extracted features from the entire HPA Cell Atlas image dataset using a machine learning model[30] and visualized them as a 2D projection of a UMAP (Extended Data Fig. 1e). We computed the average pairwise cosine distance (1 - cosine similarity) of all images belonging to genes from the same metabolic pathway or pathway group (Supplementary Table 3). We performed a random control by picking the equivalent number of random images and computing the average cosine distance across 1000 permutation experiments. Analysis was performed with *numpy* and *cupy* for GPU acceleration.

GO term enrichment analysis

We performed a GO term enrichment analysis for biological processes using GOrilla[88] and filtered for terms with a fold enrichment >1.5 and a p value < 0.05 after Bonferroni correction (Supplementary Table 4). The significant terms were manually grouped into categories (*e.g.,* cell cycle, metabolism) and visualized in a heatmap.

Comparison of metabolic heterogeneity with cell cycle heterogeneity

We obtained a list of cell cycle dependent (CCD) and independent (non-CCD) variable proteins[34]. By extracting the single-cell protein and RNA measurements from this study, we were able to compare the degree of variability on the proteomic and transcriptomic level across different non-exclusive groups of human proteins as defined by our data (all proteins in the HPV v23, enzymes in the HPA, enzymes exhibiting cell-cell heterogeneity in HPA v23, and CCD proteins) using Kruskal-Wallis. Furthermore, we obtained the physical properties of human proteins (fraction of disordered residues, melting temperature protein length) as described before[34] and compared them across the same groups as above using a Mann-Whitney U test with Benjamini Hochberg correction (Supplementary Table 6).

Upstream kinase investigation

The upstream kinases of phosphosites on various groups of proteins were evaluated using data from PhosphoSitePlus[89] (version May 13, 2024), grouped into the kinase families reported in KinMap[90] (accessed May 20, 2024). The counts of phosphosites for CCD proteins and enzymes were compared to those from the 13,147 proteins in the HPA Subcellular Atlas v23 to establish significance using the one-sided Fisher exact test, corrected by the Bonferroni method, and variable enzymes were similarly compared to all other enzymes (Supplementary Table 7). These results should be taken as qualitative assessments of PTM regulation, owing to the complexity of cellular regulation.

Generation of clonal mNG-tagged cell lines

CRISPR/Cas9 mediated mNG tagging was performed to establish a polyclonal mNG tagged Hek293T cell lines as described previously.[16,91] Afterwards, single cells were sorted into the wells of a 96 well plate and cultured in DMEM supplemented with 20% fetal bovine serum (FBS), 1 mM glutamine and 100 µg/ml penicillin/streptomycin (Gibco). to establish clonal populations. Cells were maintained in 96 well plates at 37 °C in a 5.0 % CO2 humidified environment and the medium was changed every three days until confluence was achieved. After 2-3 passages in complete DMEM with 10% FBS gDNA was extracted to confirm the genotype by Illumina amplicon sequencing and live cell microscopy was performed.

Evaluating the genotype of mNG-tagged cell lines

DNA repair outcomes were characterized by Illumina amplicon sequencing. Clonally expanded Hek293T cells were grown in a 96-well plate at around 80% confluency, washed with PBS and resuspended in 50 µl QuickExtract (Lucigen) at 65 °C for 20 min and 98 °C for 5 min. Next, 2 µl gDNA, 20 µl 2× KAPA, 1.6 µl of 50 µM forward and reverse primer, 8 µl 5 M betaine and 8.4 µl H2O were run on a thermocycler for 3 min at 95 °C followed by 3 cycles of 20 s at 98 °C, 15 s at 63 °C, 20 s at 72 °C, 3 cycles of 20 s at 98 °C, 15 s at 65 °C, 20 s at 72 °C, 3 cycles of 20 s at 98 °C, 15 s at 67 °C, 20 s at 72 °C and 17 cycles of 20 s at 98 °C, 15 s at 69 °C, 20 s at 72 °C. Next, 1 µl of 2 nM PCR product, 4 µl of forward and reverse indexed barcoding primer, 20 µl 2× Kapa and 11 µl H2O were run with the following thermocycler settings: 3 min at 95 °C, 10 cycles of 20 s at 98 °C, 15 s at 68 °C, 12 s at 72 °C. Product concentrations in the 200–600 bp range were quantified using a fragment analyzer (Agilent) and pooled at 500 nM. The sequencing library was purified using dual solid phase reversible immobilization at a 0.6 and 1.1× bead/sample ratio. Sequencing was performed using an Illumina MiSeq system at the CZ Biohub Sequencing facility. Sequencing outcomes were characterized using CRISPResso[92] (version 1.0.13). For clonal expansion and time lapse microscopy only clones with a homozygous homology directed repair (HDR) mediated knock in were considered.

Spinning-disk confocal live cell microscopy

Approximately 20,000 endogenously tagged HEK293T cells were grown on a fibronectin (Roche)-coated 96 well glass bottom plate (Cellvis) for 24 hours. Cells were counterstained in 0.5 µg/ml Hoechst 33342 (Thermo) for 30 min at 37 °C and imaged in complete DMEM without phenol-red. Live-cell imaging was performed at 37 °C and 5% CO_2 on a Dragonfly

spinning-disk confocal microscope (Andor) equipped with a 1.45 N/A 63x oil objective and an iXon Ultra 888 EMCCD camera (Andor) using the Micromanager software v1.4.[93] For time-lapse acquisition of HMGCS1-mNG cells approximately 10,000 endogenously tagged HEK293T cells were prepared for live cell imaging, as described earlier. Images were acquired in hourly intervals for 40 hours.

Circos plots for visualization of patterns of multilocalization within cells

The locations for each protein were extracted from the HPA subcellular section and grouped into meta structures and compartments as described previously. Multilocalizing proteins were visualized by a connection between the respective locations in the Circos plot, with proteins localizing to three or more locations displaying multiple connections in the plot. Unilocalizing proteins do not display any connections in the plot. To analyze the degree of experimentally determined protein multilocalization (excluding information derived from the HPA, we extracted subcellular location data for the proteins included on HPA subcellular section v23 from UniProt[94] (20230221). We filtered for locations with experimental evidence (ECO:0000269) and converted the UniProt locations to matching GO terms as described previously (Supplementary Table 8).[15] Finally, the degree of multilocalization in UniProt data were visualized as described before.

Interactomics analysis to identify potentially multifunctional enzymes

We identified a protein's interaction partners from various resources[16,62,63] as well as from the STRING database[64] (experimental evidence and confidence > 0.7) (Supplementary Table 9). Subsequently, a cellular component (nucleus, cytosol, nucleus+cytosol) was assigned to each interactor according to its localization data in the HPA Subcellular Section. Finally, an interaction network among all interactors was constructed with the color of the nodes indicating the subcellular protein location and the thickness of the edges showing the confidence of an interaction between 1 (experimentally verified location from the interactomics datasets) and 0.7 (lowest confidence extracted from theString database) using the matplotlib and *networkx* packages in python 3.10. This approach allowed us to evaluate the spatially-resolved interaction network for each enzyme and manually assess its non-metabolic functions using a guilty-by-association approach.

References

1. Evers, T. M. J. *et al.* Deciphering Metabolic Heterogeneity by Single-Cell Analysis. *Anal. Chem.* **91**, 13314–13323 (2019).
2. Rappez, L. *et al.* SpaceM reveals metabolic states of single cells. *Nat. Methods* **18**, 799–805 (2021).
3. Leduc, A., Huffman, R. G., Cantlon, J., Khan, S. & Slavov, N. Exploring functional protein covariation across single cells using nPOP. *Genome Biol.* **23**, 261 (2022).
4. Aissa, A. F. *et al.* Single-cell transcriptional changes associated with drug tolerance and response to combination therapies in cancer. *Nat. Commun.* **12**, 1628 (2021).
5. Torre, E. A. *et al.* Genetic screening for single-cell variability modulators driving therapy resistance. *Nat. Genet.* **53**, 76–85 (2021).
6. Tasdogan, A. *et al.* Metabolic heterogeneity confers differences in melanoma metastatic potential. *Nature* **577**, 115–120 (2020).
7. Fernández-García, J. *et al.* CD8+ T cell metabolic rewiring defined by scRNA-seq identifies a critical role of ASNS expression dynamics in T cell differentiation. *Cell Rep.* **41**, 111639 (2022).
8. Wang, G. *et al.* Spatial dynamic metabolomics identifies metabolic cell fate trajectories in human kidney differentiation. *Cell Stem Cell* **29**, 1580-1593.e7 (2022).
9. Hartmann, F. J. *et al.* Single-cell metabolic profiling of human cytotoxic T cells. *Nat. Biotechnol.* **39**, 186–197 (2021).
10. Adamik, J. *et al.* Distinct metabolic states guide maturation of inflammatory and tolerogenic dendritic cells. *Nat. Commun.* **13**, 5184 (2022).
11. Heumos, L. *et al.* Best practices for single-cell analysis across modalities. *Nat. Rev. Genet.* **24**, 550–572 (2023).
12. Capolupo, L. *et al.* Sphingolipids control dermal fibroblast heterogeneity. *Science* **376**, eabh1623 (2022).
13. Bien, T., Koerfer, K., Schwenzfeier, J., Dreisewerd, K. & Soltwisch, J. Mass spectrometry imaging to explore molecular heterogeneity in cell culture. *Proc. Natl. Acad. Sci.* **119**, e2114365119 (2022).
14. Lundberg, E. & Borner, G. H. H. Spatial proteomics: a powerful discovery tool for cell biology. *Nat. Rev. Mol. Cell Biol.* **20**, 285–302 (2019).

15. Thul, P. J. *et al.* A subcellular map of the human proteome. *Science* **356**, (2017).

16. Cho, N. H. *et al.* OpenCell: Endogenous tagging for the cartography of human cellular organization. *Science* **375**, eabi6983.

17. Reicher, A., Koren, A. & Kubicek, S. Pooled protein tagging, cellular imaging, and in situ sequencing for monitoring drug action in real time. *Genome Res.* **30**, 1846–1855 (2020).

18. Stadler, C. *et al.* Immunofluorescence and fluorescent-protein tagging show high correlation for protein localization in mammalian cells. *Nat. Methods* **10**, 315–323 (2013).

19. Gnann, C., Cesnik, A. J. & Lundberg, E. Illuminating Non-genetic Cellular Heterogeneity with Imaging-Based Spatial Proteomics. *Trends Cancer* **7**, 278–282 (2021).

20. Mofatteh, M. *et al.* Autonomous clocks that regulate organelle biogenesis, cytoskeletal organization, and intracellular dynamics. *eLife* **10**, e72104 (2021).

21. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).

22. Nilsson, P. *et al.* Towards a human proteome atlas: high-throughput generation of mono-specific antibodies for tissue profiling. *Proteomics* **5**, 4327–4337 (2005).

23. Skogs, M. *et al.* Antibody Validation in Bioimaging Applications Based on Endogenous Expression of Tagged Proteins. *J. Proteome Res.* **16**, 147–155 (2017).

24. Stadler, C. *et al.* Systematic validation of antibody binding and protein subcellular localization using siRNA and confocal microscopy. *J. Proteomics* **75**, 2236–2251 (2012).

25. Edfors, F. *et al.* Enhanced validation of antibodies for research applications. *Nat. Commun.* **9**, 4130 (2018).

26. Robinson, J. L. *et al.* An atlas of human metabolism. *Sci. Signal.* **13**, eaaz1482 (2020).

27. Reicher, A. *et al.* Pooled multicolour tagging for visualizing subcellular protein dynamics. *Nat. Cell Biol.* **26**, 745–756 (2024).

28. Wang, Y. *et al.* Pick-up single-cell proteomic analysis for quantifying up to 3000 proteins in a Mammalian cell. *Nat. Commun.* **15**, 1279 (2024).

29. Christopher, J. A., Geladaki, A., Dawson, C. S., Vennard, O. L. & Lilley, K. S. Subcellular Transcriptomics and Proteomics: A Comparative Methods Review. *Mol. Cell. Proteomics MCP* **21**, 100186 (2022).

30. Ouyang, W. *et al.* Analysis of the Human Protein Atlas Image Classification competition. *Nat. Methods* **16**,

1254–1261 (2019).

31. Bar-Peled, L. & Kory, N. Principles and functions of metabolic compartmentalization. *Nat. Metab.* **4**, 1232–1244 (2022).
32. Song, J., Pfanner, N. & Becker, T. Assembling the mitochondrial ATP synthase. *Proc. Natl. Acad. Sci.* **115**, 2850–2852 (2018).
33. Maxfield, F. R. & van Meer, G. Cholesterol, the central lipid of mammalian cells. *Curr. Opin. Cell Biol.* **22**, 422–429 (2010).
34. Mahdessian, D. *et al.* Spatiotemporal dissection of the cell cycle with single-cell proteogenomics. *Nature* **590**, 649–654 (2021).
35. Tomlinson, I. P. M., Novelli, M. R. & Bodmer, W. F. The mutation rate and cancer. *Proc. Natl. Acad. Sci.* **93**, 14800–14803 (1996).
36. Rosenberger, F. A. *et al.* Spatial single-cell mass spectrometry defines zonation of the hepatocyte proteome. *Nat. Methods* **20**, 1530–1536 (2023).
37. Hildebrandt, F. *et al.* Spatial Transcriptomics to define transcriptional patterns of zonation and structural components in the mouse liver. *Nat. Commun.* **12**, 7046 (2021).
38. Ben-Moshe, S. *et al.* Spatial sorting enables comprehensive characterization of liver zonation. *Nat. Metab.* **1**, 899–911 (2019).
39. Petrosius, V. *et al.* Exploration of cell state heterogeneity using single-cell proteomics through sensitivity-tailored data-independent acquisition. *Nat. Commun.* **14**, 5910 (2023).
40. Hrovatin, K., Fischer, D. S. & Theis, F. J. Toward modeling metabolic state from single-cell transcriptomics. *Mol. Metab.* **57**, 101396 (2022).
41. Alghamdi, N. *et al.* A graph neural network model to estimate cell-wise metabolic flux using single-cell RNA-seq data. *Genome Res.* **31**, 1867–1884 (2021).
42. Needham, E. J., Parker, B. L., Burykin, T., James, D. E. & Humphrey, S. J. Illuminating the dark phosphoproteome. *Sci. Signal.* **12**, eaau8645 (2019).
43. Varjosalo, M. *et al.* The protein interaction landscape of the human CMGC kinase group. *Cell Rep.* **3**, 1306–1320 (2013).

44. Pearce, L. R., Komander, D. & Alessi, D. R. The nuts and bolts of AGC protein kinases. *Nat. Rev. Mol. Cell Biol.* **11**, 9–22 (2010).

45. Thomas, S. M. & Brugge, J. S. Cellular functions regulated by Src family kinases. *Annu. Rev. Cell Dev. Biol.* **13**, 513–609 (1997).

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

47. Li, W. *et al.* Nuclear localization of mitochondrial TCA cycle enzymes modulates pluripotency via histone acetylation. *Nat. Commun.* **13**, 7414 (2022).

48. Lee, J. M., Hammarén, H. M., Savitski, M. M. & Baek, S. H. Control of protein stability by post-translational modifications. *Nat. Commun.* **14**, 201 (2023).

49. Foster, L. J. *et al.* A mammalian organelle map by protein correlation profiling. *Cell* **125**, 187–199 (2006).

50. Geladaki, A. *et al.* Combining LOPIT with differential ultracentrifugation for high-resolution spatial proteomics. *Nat. Commun.* **10**, 331 (2019).

51. Crook, O. M. *et al.* Inferring differential subcellular localisation in comparative spatial proteomics using BANDLE. *Nat. Commun.* **13**, 5948 (2022).

52. Shirafkan, F., Gharaghani, S., Rahimian, K., Sajedi, R. H. & Zahiri, J. Moonlighting protein prediction using physico-chemical and evolutional properties via machine learning methods. *BMC Bioinformatics* **22**, 261 (2021).

53. Li, Y. *et al.* De novo Prediction of Moonlighting Proteins Using Multimodal Deep Ensemble Learning. *Front. Genet.* **12**, (2021).

54. Chen, C., Zabad, S., Liu, H., Wang, W. & Jeffery, C. MoonProt 2.0: an expansion and update of the moonlighting proteins database. *Nucleic Acids Res.* **46**, D640–D644 (2018).

55. Gupta, M. N. & Uversky, V. N. Moonlighting enzymes: when cellular context defines specificity. *Cell. Mol. Life Sci. CMLS* **80**, 130 (2023).

56. Ruiz-Rodado, V., Lita, A. & Larion, M. Advances in measuring cancer cell metabolism with subcellular resolution. *Nat. Methods* **19**, 1048–1063 (2022).

57. Galea, G. *et al.* The Golgi complex is a regulatory hub for homologous recombination-mediated DNA repair. 2022.10.17.512236 Preprint at https://doi.org/10.1101/2022.10.17.512236 (2022).

58. Sutendra, G. *et al.* A Nuclear Pyruvate Dehydrogenase Complex Is Important for the Generation of Acetyl-CoA and Histone Acetylation. *Cell* **158**, 84–97 (2014).

59. Sirover, M. A. Pleiotropic effects of moonlighting glyceraldehyde-3-phosphate dehydrogenase (GAPDH) in cancer progression, invasiveness, and metastases. *Cancer Metastasis Rev.* **37**, 665–676 (2018).

60. Qiao, G., Wu, A., Chen, X., Tian, Y. & Lin, X. Enolase 1, a Moonlighting Protein, as a Potential Target for Cancer Treatment. *Int. J. Biol. Sci.* **17**, 3981–3992 (2021).

61. Chapple, C. E. *et al.* Extreme multifunctional proteins identified from a human protein interaction network. *Nat. Commun.* **6**, 7412 (2015).

62. Hein, M. Y. *et al.* A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances. *Cell* **163**, 712–723 (2015).

63. Huttlin, E. L. *et al.* Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. *Cell* **184**, 3022-3040.e28 (2021).

64. Szklarczyk, D. *et al.* The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646 (2023).

65. Curtis, N. J. & Jeffery, C. J. The expanding world of metabolic enzymes moonlighting as RNA binding proteins. *Biochem. Soc. Trans.* **49**, 1099–1108 (2021).

66. Queiroz, R. M. L. *et al.* Comprehensive quantitation of RNA-protein interaction dynamics by orthogonal organic phase separation (OOPS). *Nat. Biotechnol.* **37**, 169–178 (2019).

67. Wellen, K. E. & Snyder, N. W. Should we consider subcellular compartmentalization of metabolites, and if so, how do we measure them? *Curr. Opin. Clin. Nutr. Metab. Care* **22**, 347–354 (2019).

68. Smith, L. M. & Kelleher, N. L. Proteoform: a single term describing protein complexity. *Nat. Methods* **10**, 186–187 (2013).

69. Chang, Y.-C. *et al.* Decrypting lysine deacetylase inhibitor action and protein modifications by dose-resolved proteomics. *Cell Rep.* **43**, 114272 (2024).

70. Cappelletti, V. *et al.* Dynamic 3D proteomes reveal protein functional alterations at high resolution in situ. *Cell* **184**, 545-559.e22 (2021).

71. Chen, Y., Gustafsson, J., Yang, J., Nielsen, J. & Kerkhoven, E. J. Single-cell omics analysis with genome-

scale metabolic modeling. *Curr. Opin. Biotechnol.* **86**, 103078 (2024).

72. Specht, H. *et al.* Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. *Genome Biol.* **22**, 50 (2021).
73. Reimegård, J. *et al.* A combined approach for single-cell mRNA and intracellular protein expression analysis. *Commun. Biol.* **4**, 1–11 (2021).
74. Brunner, A. *et al.* Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Mol. Syst. Biol.* **18**, e10798 (2022).
75. Popovic, D., Koch, B., Kueblbeck, M., Ellenberg, J. & Pelkmans, L. Multivariate Control of Transcript to Protein Variability in Single Mammalian Cells. *Cell Syst.* **7**, 398-411.e6 (2018).
76. Grimbergen, A. J., Siebring, J., Solopova, A. & Kuipers, O. P. Microbial bet-hedging: the power of being different. *Curr. Opin. Microbiol.* **25**, 67–72 (2015).
77. Carja, O. & Plotkin, J. B. Evolutionary Rescue Through Partly Heritable Phenotypic Variability. *Genetics* **211**, 977–988 (2019).
78. Kamrad, S. *et al.* Metabolic heterogeneity and cross-feeding within isogenic yeast populations captured by DILAC. *Nat. Microbiol.* **8**, 441–454 (2023).
79. El Meouche, I., Siu, Y. & Dunlop, M. J. Stochastic expression of a multiple antibiotic resistance activator confers transient resistance in single cells. *Sci. Rep.* **6**, 19538 (2016).
80. Emert, B. L. *et al.* Variability within rare cell states enables multiple paths toward drug resistance. *Nat. Biotechnol.* **39**, 865–876 (2021).
81. Levy, S. F., Ziv, N. & Siegal, M. L. Bet hedging in yeast by heterogeneous, age-correlated expression of a stress protectant. *PLoS Biol.* **10**, e1001325 (2012).
82. Wang, J. *et al.* Single-cell RNA-seq reveals early heterogeneity during aging in yeast. *Aging Cell* **21**, e13712 (2022).
83. Arnold, P. K. & Finley, L. W. S. Regulation and function of the mammalian tricarboxylic acid cycle. *J. Biol. Chem.* **299**, 102838 (2022).
84. Gabriel, C. H. *et al.* Circadian period is compensated for repressor protein turnover rates in single cells. *Proc. Natl. Acad. Sci. U. S. A.* **121**, e2404738121 (2024).

85. Vicari, M. *et al.* Spatial multimodal analysis of transcriptomes and metabolomes in tissues. *Nat. Biotechnol.* 1–5 (2023) doi:10.1038/s41587-023-01937-y.

86. Uhlen, M. *et al.* A proposal for validation of antibodies. *Nat. Methods* **13**, 823–827 (2016).

87. Stadler, C., Skogs, M., Brismar, H., Uhlén, M. & Lundberg, E. A single fixation protocol for proteome-wide immunofluorescence localization studies. *J. Proteomics* **73**, 1067–1078 (2010).

88. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).

89. Hornbeck, P. V. *et al.* PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* **43**, D512-520 (2015).

90. Eid, S., Turk, S., Volkamer, A., Rippmann, F. & Fulle, S. KinMap: a web-based tool for interactive navigation through human kinome data. *BMC Bioinformatics* **18**, 16 (2017).

91. Leonetti, M. D., Sekine, S., Kamiyama, D., Weissman, J. S. & Huang, B. A scalable strategy for high-throughput GFP tagging of endogenous human proteins. *Proc. Natl. Acad. Sci.* **113**, (2016).

92. Pinello, L. *et al.* Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat. Biotechnol.* **34**, 695–697 (2016).

93. Edelstein, A. D. *et al.* Advanced methods of microscope control using μManager software. *J. Biol. Methods* **1**, e10 (2014).

94. UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2023).

Acknowledgements

We acknowledge the entire staff of the HPA. Funding was provided by the Knut and Alice Wallenberg Foundation to the Human Protein Atlas and E.L. (KAW 2021.0346 & KAW 2021.0189), Göran Gustafsson Stiftelse, Vetenskapsrådet to E.L (VR 2017-05327) and Schmidt Futures to E.L. A.S. was supported by a Rubicon grant from the Dutch Research Council (NWO, 019.231EN.013).

Author contributions

C.G. participated in the development of the idea, conducted the experiments, conducted the analysis, generated figures and co-wrote the manuscript. A.C. and A.S. co-wrote the manuscript. A.C. conducted upstream kinase analysis. T.L. performed cosine distance analysis. D.M. participated in manuscript writing and contributed to discussions. S.S. and R.S. performed tissue image annotation. P.R. performed live-cell time lapse experiments. M.L. provided resources for, aided and supervised the generation of clonal mNG-tagged cell lines. C.L. provided resources for tissue image annotation and

contributed to discussions. M.U. provided antibody resources. U.A. and A.S. were responsible for data integration into the proteinatlas.org database. E.L. developed the idea, designed and supervised the project and co-wrote the manuscript. All authors have read and approved the manuscript.

Competing interests

The authors declare no competing interests to this work.

Data and materials availability

All cell types are authenticated and listed in the Materials and Methods. Images are publicly available on the HPA website (www.proteinatlas.org), and single-cell transcriptomic data are available in GEO SRA project GSE146773. Uncompressed images for the cell cycle-resolved imaging proteomic dataset were annotated using IDR metadata templates and deposited in the BioImage Archive (accession S-BIAD34, https://www.ebi.ac.uk/biostudies/BioImages/studies/S-BIAD34). All other imaging data will be made publicly available upon publication.

Code availability

Code will be made publicly available upon publication.