Proteomics Reveals the Importance of the Dynamic Redistribution of the Subcellular Location of Proteins in Breast Cancer Cells

Gabriela Pinto‡, Abdulrab Ahmed M. Alhaiek and

Jasminka Godovac-Zimmermann*

Division of Medicine, University College London, Centre for Nephrology, Royal Free Campus, Rowland Hill Street, London NW3 2PF, UK, EU

‡Present address: Department of Pharmacy, University of Naples Federico II, Via D. Montesano 49, 80131 Napoli, Italy

*Corresponding author: Prof Jasminka Godovac-Zimmermann,

email:j.godovac-zimmermann@ucl.ac.uk

Abstract

At the molecular level living cells are enormously complicated complex adaptive systems in which intertwined genomic, transcriptomic, proteomic and metabolic networks all play a crucial role. At the same time, cells are spatially heterogeneous systems in which subcellular compartmentalization of different functions is ubiquitous and requires efficient cross-compartmental communication. Dynamic redistribution of multitudinous proteins to different subcellular locations in response to cellular functional state is increasingly recognized as a crucial characteristic of cellular function that seems to be at least as important as overall changes in protein abundance. Characterization of the subcellular spatial dynamics of protein distribution is a major challenge for proteomics and recent results with MCF7 breast cancer cells suggest that this may be of particular importance for cancer cells.

Cancer and complexity

Our oldest description of cancer was discovered in Egypt and dates back to 3000 BC. Breast cancer was described in the Edwyn Smith Papyrus with a written verdict: "There is no treatment" [1]. Much of the anatomical and clinical knowledge obtained in the 16th and 17th centuries was followed by progress in pathology and experimental research in cancer treatment and scientific surgery in the 18th and 19th centuries. Today cancer still constitutes a major scientific endeavour occupying chemists, biologist, geneticists and physicists [101]. Cancers have been found in everything from coral to budgerigars and dinosaurs [102]. Despite at least 3 013 022 publications on cancer (1818-2014) and large-scale "omics" studies for humanity, cancer is still the "Emperor of all maladies"[2]. Although mortality levels are lower than for infectious/parasitic and heart disease [103], cancer remains a fascinating scientific challenge, fundamentally due to its fearsome complexity.

A major complexity problem in cancer is occasioned by the fact that heterogeneity is a ubiquitous feature of tumors. Carcinogenesis is not a single step process and initiation through proliferation and metastasis stages are thought to be dynamically influenced by accumulation of a large number of genetic mutations [3]. Each round of cell division may cause increased instability and complexity of the genome that enhance tumor progress [3]. The enigmatic heterogeneity within the same kind of tumor makes manifestation of the disease unpredictable and greatly complicates clinical diagnosis via characteristics such as morphological features, cellular biomarkers, hormonal receptors, gene expression, propensity to metastasize and recurrence potentials [4,5]. It has been suggested that cancer should be regarded as a chaotic process [6] and there is some recent evidence for stochastic changes in breast cancer cell lines [7]. Although progress is being made, advances in effective disease diagnosis, management and prognosis would be achieved by further understanding of

the basis of this heterogeneity.

In the age of large-scale "omics" that followed on the completion of human genome sequencing, the development of ever more powerful DNA sequencing methods, of DNA-microarray technology and of proteomics led to two major large-scale methods for investigation of breast cancer and other diseases: (1) large-scale genome-wide-association-strategies (GWAS) that involve screening of large number of patients to identify genetic mutations correlated with disease [8,9], and (2) extensive investigation of differential gene expression via transcriptomics and proteomics [10]. The large-scale GWAS strategies have been found to have limitations, e.g. in complex diseases many common disease-correlated mutations are neither necessary nor sufficient for the disease and often seem only to be deleterious in the context of the genome/proteome of an individual [11]. This has led to proposals that such data needs to be integrated with transcriptomic, proteomic, metabolomic and clinical information to develop 'functional-network-based' conceptual models of normal function, disease, diagnostics and therapy [12,13]. Large-scale screening of mRNA levels showed that multiple and extensive changes in mRNA levels are commonly seen in breast cancer [14-17]. More recently such studies have been extended to involvement of micro-RNAs [18] and epigenetic modulation of chromatin [19].

There were important reasons [20] to complement this work with high throughput proteomics methods, including that protein abundance may be different than transcript abundance and that for genetic variation, translation and protein stability may be more determinant for protein abundance than transcript levels [21]. Recent work on tumor samples suggests that transcriptomics and proteomics measurements of total abundance monitor different aspects of cancer cell function and provide highly complementary information [22].

These methods are accompanied by very large numbers of more conventional but

increasingly powerful molecular and cell biology studies that focus on specific local features within highly complex cellular networks and which are collected and collated in data bases such as GO, REACTOME, etc. For example, cancer-related eukaryotic cell proliferation is known to involve complex molecular choreography of mitogens that stimulate cell growth, membrane receptors, their signaling pathways, and downstream effectors of cell division and cellular metabolic state [15,23,24]. Very complex cellular signalling systems modulate cancer cell function through post-translational modifications such as phosphorylation [25] and methylation/acetylation [26]. Large-scale proteomics detection of PTMs like phosphorylation or acetylation has been crucial in this area [27].

A key feature long known from molecular and cell biology, but so far often ignored in many large-scale studies, is that cellular function is highly dependent on the spatial distribution of many cellular components, ranging over metabolites, low molecular weight signaling molecules (e.g. GTP, Ca^{2+}, NADH, ROS), proteins, lipids, tRNA [28], etc. An emerging theme, that is the focus of this report, is that the subcellular distribution of proteins is dynamic and context-dependent, that proteins may have different functions at different subcellular locations, that the dynamic distributions are a crucial feature of cellular function and that perturbation of spatial control may be an important feature of cancer. Indeed, we suggest that dynamic alterations of subcellular spatial distribution of proteins is at least equally important to changes in total protein abundance in cellular function.

Indications for the importance of dynamic subcellular distribution of proteins.

There are innumerable conventional cell biology studies that demonstrate functionally relevant subcellular translocation of specific individual proteins, even for proteins that were once regarded as "housekeeping" proteins [29-31]. The different subcellular translocation of the isoforms of hexokinases, HKI and HKII, is

known to be a mechanism of cellular regulation addressing the cell to catabolic or anabolic glucose utilization both in proliferating cells or in cancer cells [32]. Two isoforms of pyruvate kinases, PKM1 and PKM2, have diverse involvement in metabolic pathways. These include the shuttling of pyruvate preferentially to lactate dehydrogenase instead of to mitochondria that underlines the major role of the cytoplasmic PKM2 isoform in tumor progression [33], translocation into the nucleus that occurs in response to different apoptotic stimuli [34], and participation in nuclear transcription complexes in response to hypoxia [35]. Another example of crucial subcellular redistribution is BRCA1, well known for its nuclear-cytoplasmic trafficking in breast cancer [36]; recently, its redistribution to the cytoplasm in malignant breast cancer tissues has been supposed to be a defence mechanism of the cell probably associated with a more intense cellular apoptotic activity [37]. However, study of expression and subcellular localization for single proteins or small groups of proteins may often limit understanding of cellular mechanisms because of the complex networking of biological systems. Moreover the high number of moonlighting proteins with different functions in various subcellular compartments as well as massive spatio-temporal and condition-dependent redistribution of proteins makes understanding even more complicated.

Methods for Global Determination of Protein Distributions.

There are a very large number of proposed methods for monitoring protein location within cells. These have been extensively reviewed recently [38]. The vast majority of these methods involve the tagging of proteins with some sort of detectable marker. This can limit their applicability to global monitoring of dynamic protein distribution, but they may be highly attractive as techniques for confirming results for specific proteins obtained with global methods. We do not consider these methods further here and refer readers to the above review.

At present there seem to be three technologies that are capable of global monitoring. For lower eukaryotes such as yeast, it is possible to use molecular biology methods to introduce markers on a genome-wide scale. Such methods have been used to show that about 4-5% of yeast proteins are subject to subcellular translocation under hypoxia [39] and to show that at least a third of yeast mitochondrial proteins have additional subcellular locations outside of mitochondria [40]. For higher eukaryotes and study of issues such as cancer, there currently seem to be only two appropriate technologies available, each of which has strengths and disadvantages.

Antibody-based proteomics includes techniques that are often used to detect/evaluate one or a few proteins such as immunohistochemistry (IHC), enzyme-linked immune-sorbent assay (ELISA), western blot, and immunoprecipitation (IP). The large-scale platforms include tissue microarrays (TMAs) and reverse phase protein arrays (RPPAs) that can provide the tools and the strategies to generate systematic analysis using specific antibodies. However, a common limitation is that antibodies must be available and are greatly variable in terms of sensitivity and specificity [41]. Furthermore, a quantitative approach cannot be easily robust; the order of magnitudes in the proteome is widely variable [42].

The other global method is MS-based analysis of subcellular fractions, which is the main focus in the following text. This has been used for some time to investigate protein content of selected subcellular organelles [38], but often without consideration of dynamic changes in their protein content. This method seems to have greater dynamic range and greater ability to accurately quantitate moderate changes in abundance at a given subcellular location compared to the antibody methods. These features can be crucial, e.g., in the detection of trace amounts of cytosolic metabolic enzymes involved in nuclear functions or in the measurement of changes in cellular function by coordinated, moderate changes of many proteins in

complex networks.

Workflows for MS-based quantitative shotgun subcellular proteomics using differential isotope labelling by metabolic incorporation or chemical labelling are now well established (Fig.1). The huge raw data files obtained from high-resolution mass spectrometers requires a powerful processing platform with quantitative proteomics software and subsequently rigorous procedures for result validation to avoid erroneous protein identification and to obtain statistically reliable quantitative information. An important consideration with this method is the reliability of protein distributions based on the breakage and fractionation of cells and this is considered in further detail below.

Static Studies of the Global Subcellular Distribution of Proteins.

The development of methods for subcellular fractionation to gain more insight into protein function has been emphasized since at least 1946 on the basis of the idea that "The physiology of the cell cannot be fully understood unless we succeed in determining the constitution of its parts..." [43]. The by now large body of work on the proteomes of specific subcellular organelles has often been dominated by the concept that highly pure organelle preparations are required [38]. This focus has carried over to a number of studies of the overall global distribution of proteins to multiple subcellular locations under static (non-stimulated) conditions, where emphasis has often been placed on determining *the* location of a protein. While highly purified organelles allow determination of what might be termed "permanent resident" proteins, we believe that this limited focus constrains elucidation of cellular function. Much as the original central dogma of molecular biology (one gene implies one mRNA implies one protein) has been well and truly superseded, so there has long been strong evidence that the often unconsciously assumed corollary (one protein implies one cellular location implies one function) is equally superseded. The

cell biology literature is full of thousands of examples of proteins with multiple subcellular locations and multiple functions. The GO database contains annotations to multiple subcellular locations for 36% of human proteins and this is certainly an underestimate. We consider that the most important proteins for coordination of cellular function over spatially and functionally inhomogeneous cells are likely to usually be those that are *not* "permanent residents", but rather translocate to different subcellular locations/functions in response to cellular state. These may be "peripheral" proteins in any given subcellular organelle and may often be lost during stringent preparation of highly purified organelles. To give a specific example, nuclear respiratory factor 2 (NRF2) is a critical transcription factor for response to oxidative stress [44]. It is normally present in the cytoplasm complexed with its inhibitor KEAP1 and tethered to the outside of mitochondria by interactions of this complex with PGAM5 [45], which places NRF2 in proximity to sources of reactive oxygen species (ROS) from the mitochondrial respiratory chain. A crucial aspect of NRF2 function is its translocation from this peripheral mitochondrial location to the nucleus under conditions of oxidative stress. Hence, in looking at cancer and other cells there hence seem to be three crucial questions. (1) how many such multiply located/potentially translocating proteins are there? (2) How many proteins are shared between functionally essential organelles such as the nucleus and mitochondria (this section)? (3) how can we reliably identify and verify the functional importance of the dynamic redistribution of such proteins (next section)?

In a paper published by Qattan et al [46], three crucial features of the subcellular distribution of proteins in MCF7 breast cancer cells were established: (a) large numbers of proteins are distributed over multiple subcellular locations, (b) there is substantial variation between different proteins in their relative amounts in different locations, but many proteins have appreciable quantities in various locations, and (c)

those proteins widely distributed over different locations are also those with currently annotated participation in the widest variety of functional processes. In this work, a quantitative proteomic method was used to study the static distribution of subcellular proteins in MCF7 breast cancer cells by combination of sucrose density gradient subcellular fractionation and tandem-MS-based shotgun proteomics analysis. Four separated portions of the sucrose gradient that roughly corresponded to four major organelle compartments (cytosol, plasma membrane, endoplasmic reticulum and mitochondrion) were subjected to detailed analysis. Stringent controls of the degree of purification/contamination/reproducibility of the gradient fractions were applied and the relative amounts of proteins localized in one or the other cellular compartment measured. A strong linear correlation was found between the relative abundance of each protein and the spectral counts of tryptic peptides, as already demonstrated by others [47,48]. Thus, spectral counts, calculated by the integration of the normalization by Scaffold software and the use of normalized spectral abundance factors (NSAF), the latter for counterbalancing the spectral contribution of larger proteins, were used as directly corresponding to relative abundance of characteristic proteins in the label-free quantification experiments [46,49]. Lower abundance proteins with only one assigned peptide or small numbers of counts were excluded from the analysis. Different proteins showed different abundance patterns over the four compartment (Fig. 2C) and these patterns provided evidence that the presence of proteins in multiple locations could not be due to artefacts of the sucrose gradient fractionation (Fig. 2D). Each protein was characterized by one or more (protein, location, abundance) data points. For 2184 proteins, the 4638 data points (an average of 2.1 locations per protein) were used to calculate apparent mole fraction distributions of the proteins over 1-4 of the cellular compartments. The allowed space for mole fraction distribution was widely occupied (Fig. 2E),

indicating good sampling and systematic dispersion of many proteins over multiple locations with appreciable abundance in the different locations. As an overall result, the data was consistent with detection of 46.7% of the 2184 proteins in multiple subcellular locations, this being a lower limit because of the sampling properties of spectral counting [46]. Examination of the biological processes in which these proteins have been implicated suggested correlation between multiple subcellular locations and multiple functional roles [46].

Similar methods were subsequently applied to the partitioning of proteins between the nucleus and mitochondria of MCF7 cells [49]. Western blot analysis of proteins constitutively present in mitochondria, the nucleus or other cytoplasmic organelles was used to validate the reproducibility of the fractionation procedures prior to MS analysis (Figure 3A). For instance, the detection of marker proteins in the expected N, M or C subcellular preparations {mitochondria: SDHB, MT-ND1; nucleus: ORC2; Golgi apparatus: KDEL; endoplasmic reticulum: ERN1, nucleus and cytoplasm: HDAC} confirmed the recovery, reproducibility and purity level of fractions used for MS analysis (Figure 3A).

Following the MS analysis, the subcellular localization attributed to functionally interesting proteins can be further confirmed by Western blot analysis (Figure 3B). Many functionally interesting proteins were found to have multiple locations, *e.g.* "mitochondrial" proteins (MT-CO2, CYC1, ATP5B, PCK2, SDHA) were observed in both the nucleus and mitochondria, in agreement with the MS results. Overall, 985 proteins were found to be common to mitochondria and the nucleus in MCF7 cells [49]. This is a large number compared to previous studies of highly purified mitochondria/nuclei, which we believe reflects the stripping of important proteins by stringent organelle purification methods, e.g. NRF2 or the hexose kinases (HK1, HK2) from the glycolysis enzymatic cascade that are known to shuttle between

mitochondria and nuclei in functional roles that determine the metabolic fate of glucose and are thought to be crucial to cancer metabolism [50].

As noted above collation of results in databases such as GO, LOCATE and iLOC-EUK [38] of large numbers of more conventional cell biology measurements are also consistent with multi-site location of high percentages of cellular proteins. Monitoring by antibody methods is also converging on this conclusion [51].

This brings us to the third critical question noted above: (3) how can we reliably identify and verify the functional importance of the dynamic redistribution of such proteins?

Global Dynamics of Subcellular Protein Distribution.

There are by now a number of studies of dynamic changes in the proteome of subcellular organelles using MS-based methods. These were reviewed recently [38]. Particularly notable have been several studies of expression levels and subcellular localization of endogenous proteins in HeLa cells [52,53] and viral infections [54]. In the following we limit discussion to several recent SILAC differential labelling studies using a "subcellular spatial razor" approach that are consistent with the conceptual framework outlined above: nucleo-ctyoplasmic trafficking of proteins following engagement of cell cycle checkpoints for DNA replication [55] or exposure to oxidative stress [56] (in human fibroblasts) as well as the response of MCF7 breast cancer cells to estrogen stimulation [57].

Tandem-mass spectrometry-based shotgun proteomics now allows the rapid, accurate and highly sensitive identification and quantification of several thousand proteins from complex biological samples. The variability caused by pre-MS steps has meant that the development of a reliable method for quantitative analysis is still considered challenging [58]. However, careful attention to the sources of variability can produce

data with a high degree of accuracy and reproducibility [46,49,55-57]. Software capable of dealing flexibly and efficiently with quantitative analysis of massive data sets, such as MaxQuant [58,59] and Perseus, are an essential requirement. In recent studies of the subcellular nucleo-cytoplasmic protein abundance changes for human IMR90 fibroblasts following mild *tert*-butyl peroxide (TBP) oxidative stress [56] or following cell cycle arrest at the origin activation checkpoint for DNA replication [55], as well as for MCF7 breast cancer cell following exposure to estrogen (E2) [57], 12-16 large data sets were processed in parallel with MaxQuant and Perseus. The processed data for the unfractionated total cell lysate (T), the nucleus-enriched samples (N), and the corresponding nucleus-depleted samples (C) in three replicates monitored large numbers of proteins. For example, this included 4386 different proteins for estrogen-stimulated MCF7 cells, (3604 reliably quantified according to stringent selection criteria [57]) and 3589 proteins for oxidatively stressed human IMR90 fibroblasts [56]. For each protein, the corresponding SILAC ratios provided measures of the changes in the total cellular abundance (S_t), or in the nuclear (S_n) or cytoplasmic (S_c) compartmental abundance. For the two studies of IMR90 fibroblasts, the distribution profiles of the SILAC ratios (S_n, S_c and S_t) showed typical Gaussian scatter and outlier analysis such as the Significance B score used in MaxQuant could be used to select proteins showing significant changes in total or compartmental abundance. Good reproducibility over the replicates for the SILAC ratios indicated that with careful sample preparation, reliable, global, dynamic protein quantification of different subcellular compartments is possible [56,57].

The subcellular spatial razor formulation can distinguish between changes in total protein abundance (S_t) and redistribution (S_n/S_c) to/from a target organelle (e.g. the nucleus). For all three of the systems so far analyzed with this framework, correlation between changes in total abundance and in subcellular nuclear/cytoplasmic

redistribution has been found to be low, i.e. changes in compartmental abundance do *not* simply mirror changes in total protein abundance [56,57]. The orthogonal basis set $\{S_t, S_c/S_t, S_n/S_t\}$ for the three measured SILAC ratios separates changes in total protein abundance (S_t) from a distribution plane (S_c/S_t, S_n/S_t) that reflects the redistribution of a protein. Conservation of mass (highly reproducible subcellular fractionation and no differential protein losses between stimulated/unstimulated samples during MS sample preparation) requires that the data points lie in 2 quadrants corresponding to N → C or C → N redistribution of the protein (Fig. 4A) and provides a convenient formulation for visualizing/evaluating the reliability of the compartmental changes [56,57]. The 3D spatial razor model allows characterization of both total and compartmental changes in protein abundance for each individual protein (Fig. 4B,C). For instance, the analysis of datasets for estradiol-stimulated MCF7 cells showed N → C redistribution of NHP2L1, AGR3, and ERP29 proteins accompanied by different changes in their total abundance, as well as marked N → C nuclear redistribution of EZR and SUB1 (PC4) with little or no change in total abundance as a result of cellular information transfer between different subcellular locations upon estradiol stimulation (Fig. 4B) [57]. Similarly, for oxidative stress of IMR90 cells, eight subunits of the CCT protein folding complex showed a substantial C → N redistribution while proliferating cell nuclear antigen (PCNA) showed an opposite N → C redistribution (Fig. 4C), both with a little or no change in total abundance and suggestive of "catalytic" transfer of information between different subcellular compartments [56].

An important feature of the subcellular spatial razor formulation is that it is overdetermined and explicitly includes conservation of mass in the quantitative evaluations [56,57]. This offers new perspectives for checking the selectivity of the detected changes and for evaluation of any effects of "impurity" of the subcellular

fractions used in the MS analyses. For example, consistent with the idea that many proteins are dispersed over multiple subcellular locations, the nuclear preparation analysed in the oxidative stress experiments on IMR90 cells contained 371 proteins annotated by GO to mitochondria, only 147 of which were currently also annotated to the nucleus by GO. However, only a small proportion of these "mitochondrial" proteins showed any change in their fraction in the nucleus in response to oxidative stress (Fig. 5) and similar changes in a small proportion of the nuclear proteins was observed for those annotated to endoplasmic reticulum, plasma, membrane, the Golgi apparatus, lysosomes, endosomes and peroxisomes [56]. Such data patterns are very strong evidence that those proteins showing changes in nuclear abundance represent selective trafficking for a specific subset of cellular proteins and are *not* a consequence of contamination of the nuclear preparations with other cellular components. How the inclusion of a conservation of mass test can be used with outlier analysis to select the most reliable changes in compartmental redistribution has been described [57]. Importantly, comparison with the results obtained for cell cycle arrest at the origin activation checkpoint showed that different proteins show nucleo-cytoplasmic trafficking, with different changes in the nucleus for specific proteins from cellular functional networks such as the TCA cycle, glycolysis and the proline regulatory axis [55,56]. That is, high specificity in the nuclear response to different cellular stimulations was demonstrated. In short, as predicted some years ago [46], evidence for the functional importance of the subcellular distribution of a protein can be detected by differential changes in its location under functional stimulation in highly enriched although not exhaustively purified subcellular fractions. This is crucial for detecting the functional involvement of translocating, but "non-permanent-resident" proteins in different subcellular locations/organelles.

For the response of IMR90 cells to oxidative stress or activation of DNA replication checkpoints, 3-4% of the monitored proteins were found to show appreciable changes in compartmental abundance, with changes in total abundance (S_t) and in compartmental redistribution (S_n/S_c) contributing roughly equally. Similarly, subcellular translocation of 4-5% of cellular proteins was observed in a genome-wide, fluorescence based study of the response of yeast cells to hypoxia [39].

For MCF7 breast cancer cells, a very different pattern was observed following exposure to estradiol. Many more of the monitored proteins (about 20%) showed substantial changes. However, of 331 proteins with >2-fold changes in at least one of the SILAC ratios, only 5 proteins showed >2-fold changes in total abundance (St). For the other proteins the >2-fold change corresponded to the changes in partitioning of the protein between the nuclear/cytoplasmic compartments, with a strong preponderance of proteins showing N $\rightarrow$ C redistribution (Fig. 6). A crucial feature was that >2-fold changes in compartmental abundance were *much* more prominent than changes in overall abundance. Furthermore, the identities of many of these proteins were consistent with previous studies of estrogen receptors. For example, in the nucleus the ERα and ERβ receptors have been shown to interact with 498 other proteins, only 70 of which are common to both [60]. Of these proteins, 357 were detected in the MCF7 estrogen stimulation experiments: 58 proteins showed >2-fold decrease in nuclear abundance, and a further 76 showed appreciable nucleo-cytoplasmic redistribution, but none of them showed >2-fold change in total abundance. Of the roughly 1000 proteins that are partitioned between the nucleus and mitochondria, 249 showed evidence of nucleus $\rightarrow$ cytoplasm redistribution upon exposure to estrogen. For the 134 proteins showing the most significant changes in nucleo-cytoplasmic distribution, investigation of the biological processes in which they are implicated using REACTOME [61,62] showed association with core cellular

processes (gene expression, the cell cycle, protein metabolism, mRNA metabolism) and potential changes in more specific biological processes distributed spatially over the cytoplasm, plasma membrane and nucleus [57]. In short, this work revealed that the dominant response of MCF7 breast cancer cells to estradiol is *not* changes in the total cellular abundance of proteins, but rather a massive change in their spatial distribution between the nucleus and cytoplasm, with indications that this influences functional processes at many other subcellular spatial locations.

Dynamic Subcellular Distribution of Proteins in Normal and Pathological Cellular Function.

Evidence of dynamic redistribution of proteins to and from the nucleus/mitochondria/cytoplasm and other subcellular locations following external perturbations such as environmental cues [63], cell cycle arrest [55,56], hormonal stimulation [64], oxidative stress [65-67], or viral infections [68-71] is increasing rapidly. In a model eukaryotic system, about 50% of a partially characterized yeast proteome (60%) showed dynamic redistribution between the cytosol and different organelles in response to the environmental switches [63]. There is no longer any doubt that dynamic, coordinated, context-dependent redistribution of multitudinous proteins over many subcellular locations is a central mechanism in cellular function.

These recent studies are beginning to characterize the general nature of the involvement of subcellular translocation of proteins in normal cellular function. For example, limited perturbations of healthy cells, including yeast subjected to hypoxia [39] and human fibroblasts subjected to mild oxidative stress [56] or to cell cycle arrest at the origin activation checkpoint for DNA replication [55], have typically found appreciable changes for 3-5% of proteins. Changes in total cellular abundance and in subcellular spatial redistribution were found to be of about equal importance in producing the spatial partitioning of proteins over compartments that are the basis

of cellular response to such stimulations. At the present time, there are indications from studies of viral infection [12,54,68-72] and from the studies of MCF7 breast cancer cells [46,49,57] that strong perturbations in basal protein distributions and massive protein redistribution following stimulations may be characteristic of disease.

Attempts to define general features of cancer [23,24,73] include uncontrolled proliferation and modified metabolism, especially of metabolism connected with the oxidative state of cells. The latter has been a focus of cancer research ever since Warburg [74]. The series of experiments on cell cycle arrest at the origin activation checkpoint [55] and response to oxidative [56] in fibroblasts as well as the response of MCF7 cells to estradiol stimulation thus provide a beginning to comparing protein redistribution in important core processes in healthy and cancerous cells. In fact, the series of investigations of cell cycle arrest in IMR90 human fibroblasts [55,56,75] were undertaken to investigate the possible use of pharmaceutical modulators of the CDC7-kinase in cancer therapeutics [55,56,75,76]. So far the overall indications are that healthy cells show strongly regulated, coordinated changes in both total abundance and subcellular spatial distribution of relatively small sets of proteins that are highly characteristic of the cellular response to specific stimulations. In contrast, massive changes in subcellular spatial distribution dominated much more limited changes in total abundance for proteins of MCF7 cells subjected to estradiol stimulation. In fact, estrogen receptors themselves undergo subcellular spatial redistribution in many functional contexts and this is known to be connected to core cellular processes, e.g. efflux of ERα from the nucleus is associated with repression of cell cycle progression and S-phase proliferation in MCF7 cells [77]. Estrogen receptors are targeted for proteasomal degradation through a transcription-coupled pathway requiring new protein synthesis [78,79] and the proteomics experiments also

detected major changes in proteasomes that could be coupled to estrogen receptor turnover as well as many other functional processes in both the nucleus and cytoplasm [80]. This led to suggestions that the modified metabolic properties of cancerous MCF7 cells may be mainly based on perturbed spatial distribution of proteins, that this also opens the possibility that transformation and tumorigenicity may also be strongly influenced by perturbed spatial distribution of proteins, and that refocusing on the dominant mechanism for response of breast cancer cells to estradiol may have important consequences for development of therapeutics [57].

New Challenges for Proteomics and Bioinformatics and Progress Towards Medical Applications.

The high degree of spatial/functional inhomogeneity of cells and the need for efficient subcellular communication means that the study of dynamic spatial distribution of subcellular proteins is of fundamental importance to improving understanding of cellular processes underlying human diseases. Indeed, this remarkable phenomenon occurs in any kind of cell: roughly half of all subcellular proteins are known to translocate into another compartment to reach their functional location [81], while most of the ~1,000 different "resident" proteins contained in mitochondria are imported from the cytosol to exert their functional role [82]. Moreover, numerous mislocalized proteins have been associated with human diseases as diverse as Alzheimer's disease or various types of cancer and aberrant localization of proteins has been shown to contribute to the pathogenesis of many human diseases [83].

The correct assignment/quantitation of proteins in diverse subcellular compartments based on quantitative high-throughput proteomics data promises to shed new light on many biological mechanisms. Continuing advances in proteomics MS technology will be important [58]. However, since each protein may have multiple subcellular

locations and its distribution over different locations may change as a function of cell cycle stage, metabolic, environmental and culture conditions, measuring context-dependent subcellular trafficking of proteins is a challenging task. For many subcellular organelles numerous examples of functionally important "peripheral" proteins are already known. This suggests that fractionation methods should be as simple and mild as possible and that compensatory analyses of the type exemplified by the subcellular spatial razor formulation are likely to be the most informative approach. There is also a need for improved databases that explicitly tie together function/location as pair relationships for individual proteins. At present databases such as GO obscure crucial information by neglecting this. Furthermore, the highly complex data that is now emerging from studies of context-dependent subcellular protein trafficking indicates that we should begin thinking about cellular function in terms of higher level cellular processes that are distributed over diverse cellular spatial locations and begin building models of cellular function that explicitly take such spatial dispersion into account.

The development of reliable quantitative proteomics has many potential applications to clinical settings. In this context, great challenges for oncologists are to ensure the availability of early detection/diagnostic tools and to develop methods for dealing with the enigmatic heterogeneity within the same kind of tumour (intra-tumour heterogeneity). Detection and therapeutic interventions for most cancer types have evolved over the past century, but are still limited to localized forms of cancer and/or advanced stages [84]. Moreover, better detection, targeted therapeutics and monitoring of cancer must be linked to the underlying biological processes associated with the initiation, proliferation, and metastasis stages [84,85]. Finding ways of rational diagnosis and prognosis is not an easy target and so far often does not take into consideration all the advances in the research platforms, in particular,

genomics and proteomics.

It is becoming more apparent that relying on global genomics information is a valuable asset to formulate a hypothesis, identify recognizable tissue signatures, and seek correlations to medical conditions, but the functional states of cells are beyond the scope of genomics alone. The more focussed approach of proteomics is complementary to genomics and can define the functional products (proteins) of the deregulated genes in space and time, which increases our understanding of the underlying biological processes [86,87]. One of the challenges but also opportunities faced by the comprehensive study of proteomes is that the estimated 20-30,000 human genes can give rise to over one million different proteins with very wide range in orders of magnitude of abundance [88]. Moreover, transformation of cancer cells is associated with major changes in the functional units (proteins), thereby altering the protein locations, concentrations, signalling pathways, and ultimately the function of the cells. Protein expression, function, location and structural mapping are all proteomics approaches that are integral to systems biology. They can enhance discovery of candidate cancer biomarkers with a high degree of specificity, sensitivity, classification and staging, thereby aiding in standardization of targeted approaches to therapy [89].

Expert commentary: "vision of the author"

Recent advances in the characterization of the subcellular location of proteins now indicate that dynamic trafficking of multitudinous proteins over many subcellular locations is a central mechanism in cellular function. This appears to be a key feature in coordination of the spatially heterogeneous distribution of different cellular functions. In MCF7 breast cancer cells massive nuclear-cytoplasmic redistribution of proteins is the dominant response to stimulation with estrogen, far outweighing changes in total protein abundance. Many examples of proteins with different

functional roles in different subcellular compartments are already known and the number of proteins so identified is poised to grow enormously. Proteins can only exert functions in locations where they exist and the recent results indicate that subcellular redistribution of proteins is of at least equal importance to changes in their total abundance in achieving the changes in compartmental abundances that are the basis of cellular response. There is now an urgent need for global characterization of the dynamic subcellular partitioning of proteins for many cell types and functional contexts.

Five year view:

We expect that in the next several years there will be major improvements in methodology and collection of massive amounts of global data on the dynamics of protein location in cells. We expect that such global analysis will be essential to reliable integration/interpretation of the massive numbers of more conventional investigations of localized features of complex cellular networks. For example, recent global analyses of nuclear cytoplasmic trafficking in the context of oxidative stress [56] and of transcriptional targets of NRF2 [90] suggest new interpretations of the crucial role(s) of NRF2 in oxidative stress responses that are linked to heme/iron homeostasis or to proteins that contain heme/iron as cofactors and that were largely missed by inference from large numbers of conventional experiments [56]. Similarly, results on trafficking in the context of response of MCF7 cells to estrogen have suggested that nuclear hormone receptors may be master integrators of spatial coordination in cells [57]. It is increasingly apparent that spatial coordination is a key aspect of cellular function in both healthy and disease states. We expect that differential network analysis [91] of global features of spatial control will begin to define high level functional processes/networks that are spatially dispersed across many cellular locations. Integration of such information with genomic,

transciptomic, and metabolomic information will be crucial to unravelling the enormously complicated complex adaptive systems that cells represent and to exploiting the information in medical and other applications.

Key issues:

1. Analytical methods that can follow dynamic subcellular redistribution of proteins have become available, with a prominent role for MS-based proteomics.

2. Large proportions of cellular proteins are distributed to multiple subcellular locations and may have different functions in those locations.

3. The subcellular spatial distribution of proteins is dynamic, changes with cellular state and is a central mechanism of cellular function. Nuclear-cytoplasmic trafficking of proteins is the dominant response for MCF7 cells exposed to estrogen.

4. Present evidence suggests that changes in subcellular location of proteins are of at least equal importance to changes in total protein abundance in cellular response to environmental cues, cell cycle stages, hormonal stimulation, oxidative stress, etc.

5. Dynamic protein redistribution provides a means to coordinate function over the spatial inhomogeneity of cells.

6. Coordinated changes in abundance/distribution of relatively small sets of proteins seem to be characteristic of cellular response to specific perturbations in healthy cells.

7. Early evidence from studies of viral infection and breast cancer cells suggests that disregulated cellular spatial control may be a feature of disease states.

8. New high level models of cellular function that include dispersion of functional systems over multiple subcellular locations and include dynamic redistribution of proteins need to be constructed.

9. The information needs to be integrated with genomics, transcriptomics and metabolomics to obtain integrated, comprehensive models of cellular function.

10. These new concepts of cellular function will offer many opportunities in medical diagnostics and therapeutics as well as other applications.

References:

Papers of special note have been highlighted as:
*   of interest
**   of considerable interest

1.   Allen JP, Art MMo. *The Art of Medicine in Ancient Egypt* (Metropolitan Museum of Art, 2005).
2.   Mukherjee S. *Emperor of All Maladies: A Biography of Cancer* (Scribner, 2010).
3.   Marusyk A, Polyak K. Tumor heterogeneity: causes and consequences. *Biochimica et biophysica acta*, 1805(1), 105-117 (2010).
4.   Dick JE. Stem cell concepts renew cancer research. *Blood*, 112(13), 4793-4807 (2008).
5.   Fidler IJ, Hart IR. Biological diversity in metastatic neoplasms: origins and implications. *Science*, 217(4564), 998-1003 (1982).
6.   Calin GA, Vasilescu C, Negrini M, Barbanti-Brodano G. Genetic chaos and antichaos in human cancers. *Medical hypotheses*, 60(2), 258-262 (2003).
7.   Casale FP, Giurato G, Nassa G *et al.* Single-Cell States in the Estrogen Response of Breast Cancer Cell Lines. *PLoS ONE*, 9(2), e88485 (2014).
8.   Bush WS, Moore JH. Chapter 11: Genome-Wide Association Studies. *PLoS Comput Biol*, 8(12), e1002822 (2012).
9.   Medland SE, Jahanshad N, Neale BM, Thompson PM. Whole-genome analyses of whole-brain data: working within an expanded search space. *Nat Neurosci*, 17(6), 791-800 (2014).
10.   Imielinski M, Cha S, Rejtar T, Richardson EA, Karger BL, Sgroi DC. Integrated proteomic, transcriptomic, and biological network analysis of breast carcinoma reveals molecular features of tumorigenesis and clinical relapse. *Molecular & cellular proteomics : MCP*, 11(6), M111.014910 (2012).
11.   Chakravarti A, Clark AG, Mootha VK. Distilling Pathophysiology from Complex Disease Genetics. *Cell*, 155(1), 21-26 (2013).
12.   Perez de Diego R, Mulvey C, Casanova JL, Godovac-Zimmermann J. Proteomics in immunity and herpes simplex encephalitis. *Expert review of proteomics*, 11(1), 21-29 (2014).
13.   Scholz SW, Mhyre T, Ressom H, Shah S, Federoff HJ. Genomics and bioinformatics of Parkinson's disease. *Cold Spring Harbor perspectives in medicine*, 2(7), a009449 (2012).
14.   Ince TA, Weinberg RA. Functional genomics and the breast cancer problem. *Cancer cell*, 1(1), 15-17 (2002).
15.   Sebastian T, Johnson PF. Stop and go: anti-proliferative and mitogenic functions of the transcription factor C/EBPbeta. *Cell cycle (Georgetown, Tex.)*, 5(9), 953-957 (2006).
16.   Sorlie T, Perou CM, Tibshirani R *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences of the United States of America*, 98(19), 10869-10874 (2001).
17.   van de Vijver MJ, He YD, van't Veer LJ *et al.* A gene-expression signature as a predictor of survival in breast cancer. *The New England journal of medicine*, 347(25), 1999-2009 (2002).
18.   Klinge CM. miRNAs and estrogen action. *Trends in endocrinology and metabolism: TEM*, 23(5), 223-233 (2012).
19.   Magnani L, Stoeck A, Zhang X *et al.* Genome-wide reprogramming of the chromatin landscape underlies endocrine therapy resistance in breast cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 110(16), E1490-1499 (2013).
20.   Chen EI, Yates JR, 3rd. Cancer proteomics by quantitative shotgun proteomics. *Molecular oncology*, 1(2), 144-159 (2007).

21. Foss EJ, Radulovic D, Shaffer SA, Goodlett DR, Kruglyak L, Bedalov A. Genetic Variation Shapes Protein Networks Mainly through Non-transcriptional Mechanisms. *PLoS Biol*, 9(9), e1001144 (2011).

22. Foss EJ, Radulovic D, Stirewalt DL *et al.* Proteomic classification of acute leukemias by alignment-based quantitation of LC-MS/MS data sets. *Journal of proteome research*, 11(10), 5005-5010 (2012).

23. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*, 100(1), 57-70 (2000).

24. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*, 144(5), 646-674 (2011).

25. Shaw RJ, Cantley LC. Decoding key nodes in the metabolism of cancer cells: sugar & spice and all things nice. *F1000 biology reports*, 4, 2 (2012).

26. Cedar H, Bergman Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet*, 10(5), 295-304 (2009).

27. Olsen JV, Mann M. Status of large-scale analysis of post-translational modifications by mass spectrometry. *Molecular & cellular proteomics : MCP*, 12(12), 3444-3452 (2013).

28. Chu HY, Hopper AK. Genome-wide investigation of the role of the tRNA nuclear-cytoplasmic trafficking pathway in regulation of the yeast Saccharomyces cerevisiae transcriptome and proteome. *Molecular and cellular biology*, 33(21), 4241-4254 (2013).

29. He H, Lee MC, Zheng LL, Zheng L, Luo Y. Integration of the metabolic/redox state, histone gene switching, DNA replication and S-phase progression by moonlighting metabolic enzymes. *Bioscience reports*, 33(2), e00018 (2013).

30. Sirover MA. Subcellular dynamics of multifunctional protein regulation: mechanisms of GAPDH intracellular translocation. *Journal of cellular biochemistry*, 113(7), 2193-2200 (2012).

31. Tristan C, Shahani N, Sedlak TW, Sawa A. The diverse functions of GAPDH: views from different subcellular compartments. *Cellular signalling*, 23(2), 317-323 (2011).

32. Neary CL, Pastorino JG. Nucleocytoplasmic shuttling of hexokinase II in a cancer cell. *Biochemical and biophysical research communications*, 394(4), 1075-1081 (2010).

33. Christofk HR, Vander Heiden MG, Harris MH *et al.* The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature*, 452(7184), 230-233 (2008).

34. Stetak A, Veress R, Ovadi J, Csermely P, Keri G, Ullrich A. Nuclear translocation of the tumor marker pyruvate kinase M2 induces programmed cell death. *Cancer research*, 67(4), 1602-1608 (2007).

35. Luo W, Semenza GL. Pyruvate kinase M2 regulates glucose metabolism by functioning as a coactivator for hypoxia-inducible factor 1 in cancer cells. *Oncotarget*, 2(7), 551-556 (2011).

36. Thompson ME. BRCA1 16 years later: nuclear import and export processes. *The FEBS journal*, 277(15), 3072-3078 (2010).

37. Mylona E, Melissaris S, Nomikos A *et al.* Effect of BRCA1 immunohistochemical localizations on prognosis of patients with sporadic breast carcinomas. *Pathology, research and practice*, 210(8), 533-540 (2014).

38. Satori CP, Henderson MM, Krautkramer EA, Kostal V, Distefano MM, Arriaga EA. Bioanalysis of Eukaryotic Organelles. *Chemical Reviews*, 113(4), 2733-2811 (2013).
**A comprensive review on organelle analysis and function by using multiple approaches.

39. Henke RM, Dastidar RG, Shah A *et al.* Hypoxia elicits broad and systematic changes in protein subcellular localization. *American journal of physiology. Cell physiology*, 301(4), C913-928 (2011).
** Describes the redistribution of yeast proteins in response to hypoxia/reoxygenation as regulatory mechanism underlying oxygen signaling.

40.   Ben-Menachem R, Tal M, Shadur T, Pines O. A third of the yeast mitochondrial proteome is dual localized: a question of evolution. *Proteomics*, 11(23), 4468-4476 (2011).

** Describes the multiple localization and function of mitochondrial yeast proteins outside the mitochondria as an evolutionary advantage of eukaryotic cells.

41.   Brennan DJ, O'Connor DP, Rexhepaj E, Ponten F, Gallagher WM. Antibody-based proteomics: fast-tracking molecular diagnostics in oncology. *Nat Rev Cancer*, 10(9), 605-617 (2010).

42.   Kim JW, You J. Protein Target Quantification Decision Tree. *International Journal of Proteomics*, 2013, 8 (2013).

43.   Claude A. FRACTIONATION OF MAMMALIAN LIVER CELLS BY DIFFERENTIAL CENTRIFUGATION : II. EXPERIMENTAL PROCEDURES AND RESULTS. *The Journal of experimental medicine*, 84(1), 61-89 (1946).

44.   Hayes JD, Dinkova-Kostova AT. The Nrf2 regulatory network provides an interface between redox and intermediary metabolism. *Trends in biochemical sciences*, 39(4), 199-218 (2014).

45.   Lo SC, Hannink M. PGAM5 tethers a ternary complex containing Keap1 and Nrf2 to mitochondria. *Experimental cell research*, 314(8), 1789-1803 (2008).

46.   Qattan AT, Mulvey C, Crawford M, Natale DA, Godovac-Zimmermann J. Quantitative organelle proteomics of MCF-7 breast cancer cells reveals multiple subcellular locations for proteins in cellular functional processes. *Journal of proteome research*, 9(1), 495-508 (2010).

* Describes the methods for analysis of multi-subcellular locations

47.   Liu H, Sadygov RG, Yates JR, 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Analytical chemistry*, 76(14), 4193-4201 (2004).

48.   McIlwain S, Mathews M, Bereman M, Rubel E, MacCoss M, Noble WS. Estimating relative abundances of proteins from shotgun proteomics data. *BMC Bioinformatics*, 13(1), 308 (2012).

49.   Qattan AT, Radulovic M, Crawford M, Godovac-Zimmermann J. Spatial distribution of cellular function: the partitioning of proteins between mitochondria and the nucleus in MCF7 breast cancer cells. *Journal of proteome research*, 11(12), 6080-6101 (2012).

** Provides a wide look at distribution of proteins envolved in oxidative phosphorylation, the tricarboxylic acid cycle, RNA processing/translation, glycolysis and Ras-related signaling traffiking between the nucleus and mitochondria as evidence for spatio-functional integration of these processes over the two different subcellular organelles.

50.   John S, Weiss JN, Ribalet B. Subcellular localization of hexokinases I and II directs the metabolic fate of glucose. *PLoS One*, 6(3), e17674 (2011).

51.   Lundberg E, Uhlen M. Creation of an antibody-based subcellular protein atlas. *Proteomics*, 10(22), 3984-3996 (2010).

52.   Boisvert FM, Ahmad Y, Gierlinski M *et al.* A quantitative spatial proteomics analysis of proteome turnover in human cells. *Molecular & cellular proteomics : MCP*, 11(3), M111.011429 (2012).

53.   Boisvert FM, Lamond AI. p53-Dependent subcellular proteome localization following DNA damage. *Proteomics*, 10(22), 4087-4097 (2010).

54.   Emmott E, Rodgers MA, Macdonald A, McCrory S, Ajuh P, Hiscox JA. Quantitative proteomics using stable isotope labeling with amino acids in cell culture reveals changes in the cytoplasmic, nuclear, and nucleolar proteomes in Vero cells infected with the coronavirus infectious bronchitis virus. *Molecular & cellular proteomics : MCP*, 9(9), 1920-1936 (2010).

55.   Mulvey CM, Tudzarova S, Crawford M, Williams GH, Stoeber K, Godovac-Zimmermann J. Subcellular proteomics reveals a role for nucleo-cytoplasmic trafficking at the DNA replication origin activation checkpoint. *Journal of proteome research*, 12(3), 1436-1453 (2013).

56.   Baqader NO, Radulovic M, Crawford M, Stoeber K, Godovac-Zimmermann J. Nuclear Cytoplasmic Trafficking of Proteins is a Major Response of Human Fibroblasts to Oxidative Stress. *Journal of proteome research*, 13(10), 4398-4423 (2014).

** Determines changes in total and compartmental protein abundance and protein redistribution between the nucleus and cytoplasm of human IMR90 fibroblasts subjected to mild oxidative stress by using the subcellular special razor method.

57.   Pinto G, Alhaiek AA, Amadi S *et al.* Systematic nucleo-cytoplasmic trafficking of proteins following exposure of MCF7 breast cancer cells to estradiol. *Journal of proteome research*, 13(2), 1112-1127 (2014).

** Describes the method subcellular special razor approach to look at changes in total protein abundance and in protein distribution between the nucleus and cytoplasm following exposure of MCF7 breast cancer cells to estradiol.

58.   Cox J, Mann M. Quantitative, high-resolution proteomics for data-driven systems biology. *Annual review of biochemistry*, 80, 273-299 (2011).

59.   Cox J, Matic I, Hilger M *et al.* A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nature protocols*, 4(5), 698-705 (2009).

60.   Nassa G, Tarallo R, Guzzi PH *et al.* Comparative analysis of nuclear estrogen receptor alpha and beta interactomes in breast cancer cells. *Molecular bioSystems*, 7(3), 667-676 (2011).

61.   Croft D, O'Kelly G, Wu G *et al.* Reactome: a database of reactions, pathways and biological processes. *Nucleic acids research*, 39(Database issue), D691-697 (2011).

62.   Milacic M, Haw R, Rothfels K *et al.* Annotating Cancer Variants and Anti-Cancer Therapeutics in Reactome. *Cancers*, 4(4), 1180-1211 (2012).

63.   Jung S, Smith JJ, von Haller PD *et al.* Global analysis of condition-specific subcellular protein distribution and abundance. *Molecular & cellular proteomics : MCP*, 12(5), 1421-1435 (2013).

* Develops a proteomics method for studying abundance and traffiking of subcellular yeast protein in response to the culture conditions switch.

64.   Mavinakere MS, Powers JM, Subramanian KS, Roggero VR, Allison LA. Multiple novel signals mediate thyroid hormone receptor nuclear import and export. *The Journal of biological chemistry*, 287(37), 31280-31297 (2012).

65.   Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotech*, 26(12), 1367-1372 (2008).

66.   Kodiha M, Stochaj U. Nuclear Transport: A Switch for the Oxidative Stress—Signaling Circuit? *Journal of Signal Transduction*, 2012, 18 (2012).

67.   Patel VP, Chu CT. Nuclear transport, oxidative stress, and neurodegeneration. *International journal of clinical and experimental pathology*, 4(3), 215-229 (2011).

68.   Lietzen N, Ohman T, Rintahaka J *et al.* Quantitative subcellular proteome and secretome profiling of influenza A virus-infected human primary macrophages. *PLoS pathogens*, 7(5), e1001340 (2011).

69.   Salsman J, Jagannathan M, Paladino P *et al.* Proteomic profiling of the human cytomegalovirus UL35 gene products reveals a role for UL35 in the DNA repair response. *Journal of virology*, 86(2), 806-820 (2012).

70. Skiba M, Glowinski F, Koczan D, Mettenleiter TC, Karger A. Gene expression profiling of Pseudorabies virus (PrV) infected bovine cells by combination of transcript analysis and quantitative proteomic techniques. *Veterinary Microbiology*, 143(1), 14-20 (2010).

71. Zhang H, Zhao C, Li X *et al.* Study of monocyte membrane proteome perturbation during lipopolysaccharide-induced tolerance using iTRAQ-based quantitative proteomic approach. *Proteomics*, 10(15), 2780-2789 (2010).

72. Perez de Diego R, Mulvey C, Crawford M *et al.* The proteome of Toll-like receptor 3-stimulated human immortalized fibroblasts: implications for susceptibility to herpes simplex virus encephalitis. *The Journal of allergy and clinical immunology*, 131(4), 1157-1166 (2013).

73. Hanahan D. Rethinking the war on cancer. *The Lancet*, 383(9916), 558-563 (2014).

74. Ngo H, Tortorella SM, Ververis K, Karagiannis TC. The Warburg effect: molecular aspects and therapeutic possibilities. *Molecular biology reports*, (2014).

75. Tudzarova S, Trotter MW, Wollenschlaeger A *et al.* Molecular architecture of the DNA replication origin activation checkpoint. *The EMBO journal*, 29(19), 3381-3394 (2010).
** The first report of a new DNA replication origin activation checkpoint.

76. Kulkarni AA, Kingsbury SR, Tudzarova S *et al.* Cdc7 kinase is a predictor of survival and a novel therapeutic target in epithelial ovarian carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 15(7), 2417-2425 (2009).

77. Lombardi M, Castoria G, Migliaccio A *et al.* Hormone-dependent nuclear export of estradiol receptor and DNA synthesis in breast cancer cells. *The Journal of cell biology*, 182(2), 327-340 (2008).

78. Alarid ET. Lives and times of nuclear receptors. *Molecular endocrinology (Baltimore, Md.)*, 20(9), 1972-1981 (2006).

79. Alarid ET, Bakopoulos N, Solodin N. Proteasome-mediated proteolysis of estrogen receptor: a novel component in autologous down-regulation. *Molecular endocrinology (Baltimore, Md.)*, 13(9), 1522-1534 (1999).

80. Komili S, Silver PA. Coupling and coordination in gene expression processes: a systems biology view. *Nat Rev Genet*, 9(1), 38-48 (2008).

81. Chacinska A, Koehler CM, Milenkovic D, Lithgow T, Pfanner N. Importing mitochondrial proteins: machineries and mechanisms. *Cell*, 138(4), 628-644 (2009).

82. Schmidt O, Pfanner N, Meisinger C. Mitochondrial protein import: from proteomics to functional mechanisms. *Nat Rev Mol Cell Biol*, 11(9), 655-667 (2010).

83. Hung MC, Link W. Protein localization in disease and therapy. *Journal of cell science*, 124(Pt 20), 3381-3392 (2011).

84. Brooks JD. Translational genomics: the challenge of developing cancer biomarkers. *Genome research*, 22(2), 183-187 (2012).

85. Chaffer CL, Weinberg RA. A Perspective on Cancer Cell Metastasis. *Science*, 331(6024), 1559-1564 (2011).

86. Lam SW, Jimenez CR, Boven E. Breast cancer classification by proteomic technologies: Current state of knowledge. *Cancer Treatment Reviews*, 40(1), 129-138 (2014).

87. Scott JD, Pawson T. Cell Signaling in Space and Time: Where Proteins Come Together and When They're Apart. *Science (New York, N.Y.)*, 326(5957), 1220-1224 (2009).

88. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Molecular & cellular proteomics : MCP*, 1(11), 845-867 (2002).

89. Breuer EK, Murph MM. The Role of Proteomics in the Diagnosis and Treatment of Women's Cancers: Current Trends in Technology and Future Opportunities. *Int J Proteomics*, 2011 (2011).

90. Campbell MR, Karaca M, Adamski KN, Chorley BN, Wang X, Bell DA. Novel Hematopoietic Target Genes in the NRF2-Mediated Transcriptional Pathway. *Oxidative Medicine and Cellular Longevity*, 2013, 12 (2013).

91. Ideker T, Krogan NJ. *Differential network biology* (2012).

101. The early History of cancer. American Cancer Society. Available at:

http://www.cancer.org/cancer/cancerbasics/thehistoryofcancer/the-history-of-cancer-

what-is-cancer (Last accessed 30 October 2014)

102. Dinosaurs got cancer. Nature, Macmillan Publishers Ltd. Available

at:http://www.nature.com/news/2003/031021/full/news031020-2.html (Last accessed

30 October 2014)

103. The WHO health report 2004. Available at:

http://www.who.int/whr/2004/annex/topic/en/annex_2_en.pdf) (Last accessed 30
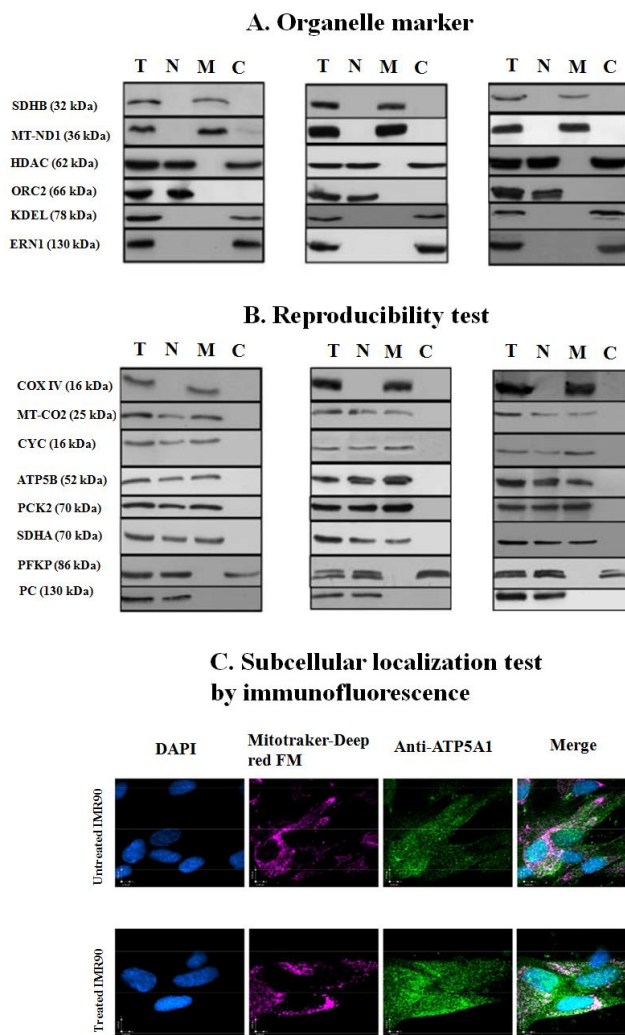
October 2014)
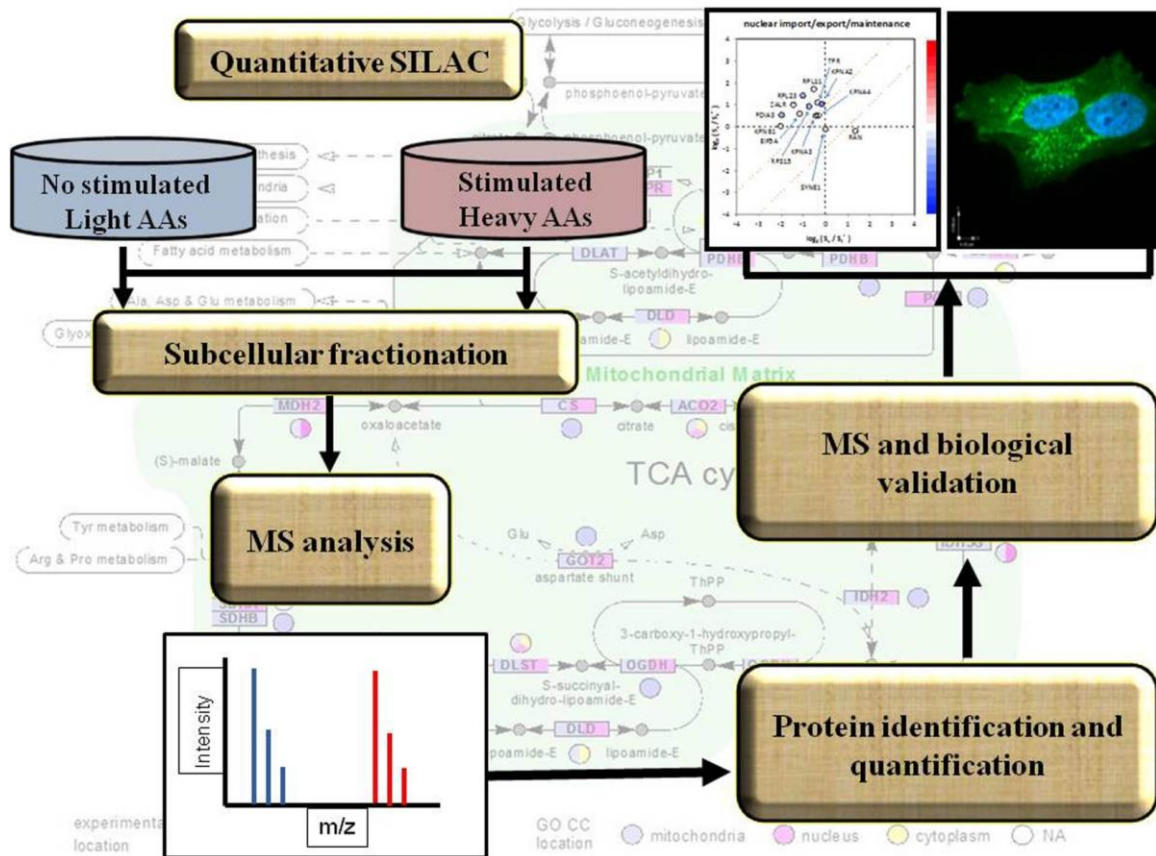
Figure 1. Workflow for subcellular analysis with quantitative shotgun proteomics.
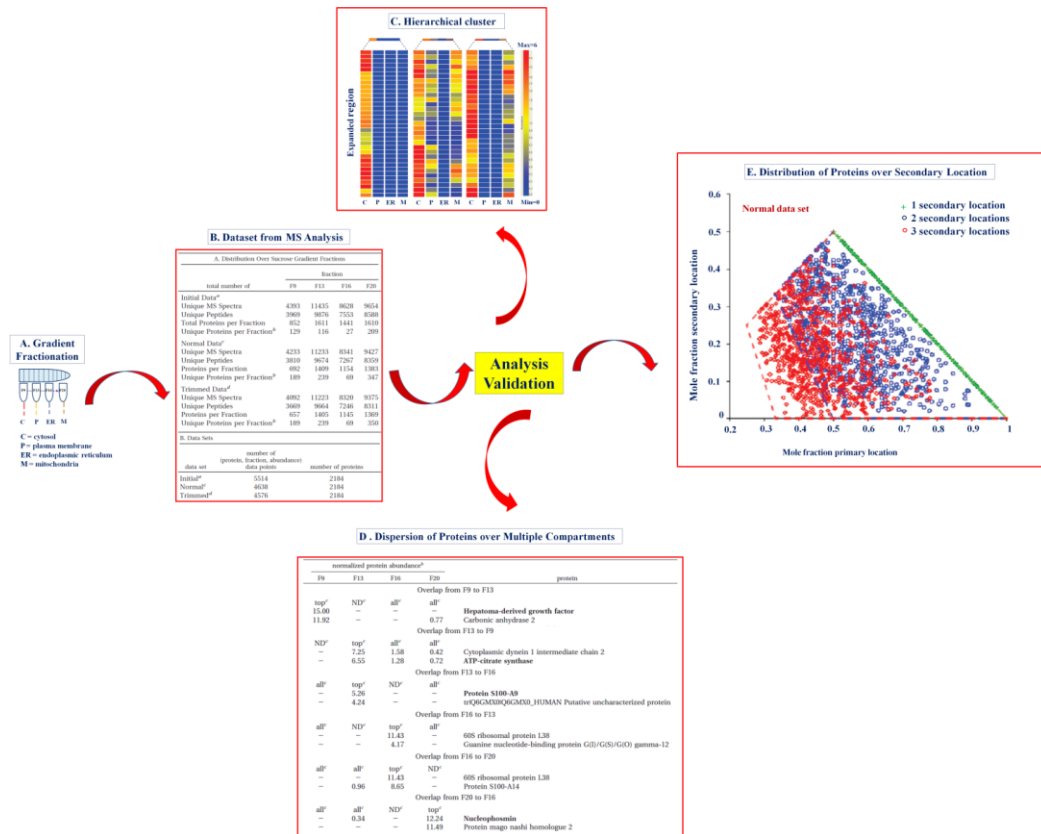
Figure 2. Proteomics approach to assess global static subcellular distribution of proteins. (A) Sucrose gradient fractionation for different organellar compartments (cytoplasm - C, plasma membrane - P, endoplasmic reticulum – ER, mitochondrion – M). (B) Proteins from each compartment are analysed by MS and a set of {protein, fraction, spectral counts} data points is assembled. Proteins identified by only one peptide or with few spectral counts are filtered out. (C) Hierarchical clustering and heat map for the amount of each protein in the four compartments. Individual proteins are represented by a single row, each fraction is represented by a single column and each cell represents the abundance of a protein in a compartment. The color scale is for normalized relative abundance from 6.0 (red) to 1.0 (yellow) to 0.0 (blue, not detected) (D) Quantitative abundances (relative) for the distribution of different proteins over the four compartments indicate the distributions are not due to artefacts in the sucrose gradient fractionation. The data is used to calculate the mole fraction of each protein in the four compartments. (E) Primary mole fractions vs secondary mole fractions for proteins with a primary location and 1 (green), 2 (blue), or 3 (red) secondary locations

A. Organelle marker

	T N M C	T N M C	T N M C
SDHB (32 kDa)			
MT-ND1 (36 kDa)			
HDAC (62 kDa)			
ORC2 (66 kDa)			
KDEL (78 kDa)			
ERN1 (130 kDa)			

B. Reproducibility test

	T N M C	T N M C	T N M C
COX IV (16 kDa)			
MT-CO2 (25 kDa)			
CYC (16 kDa)			
ATP5B (52 kDa)			
PCK2 (70 kDa)			
SDHA (70 kDa)			
PFKP (86 kDa)			
PC (130 kDa)			

C. Subcellular localization test by immunofluorescence

	DAPI	Mitotraker-Deep red FM	Anti-ATP5A1	Merge
Untreated IMR90				
Treated IMR90				

Figure 3. Alternative approaches to validation of results: (A) Demonstration of reproducibility of fractionation in three replicates by Western blotting of proteins contained in a cellular total lysate (T) or in nuclear (N), mitochondrial (M) and cytoplasmic (C) subcellular preparations. (B) Western blot [49], and (C) immunofluorescence [56] analysis to confirm subcellular localization obtained by MS analysis.
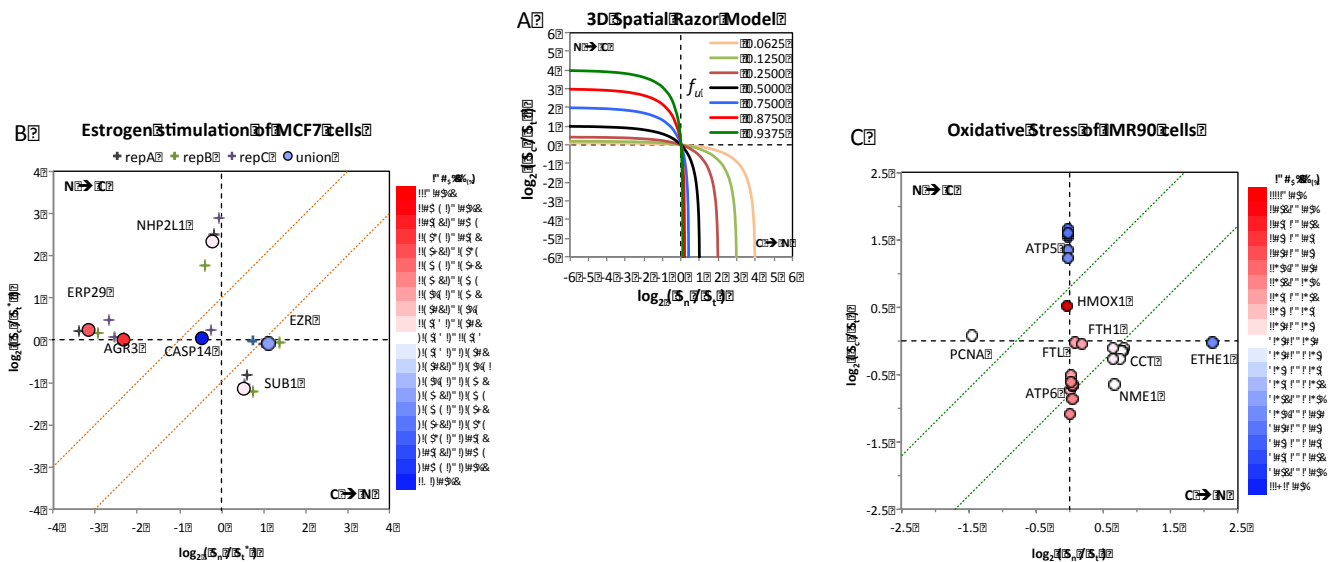
Figure 4. Three-dimensional data representation for the subcellular spatial razor model. (A) For the orthogonal 3D space $\{S_n/S_t, S_c/S_t, S_t\}$, the theoretical distribution plane $\{S_n/S_t, S_c/S_t\}$ for different values of f_u (the fraction of protein in the nucleus in the unstimulated cells) as the fraction of the protein in the nucleus in the stimulated cells (f_s) varies over $0 < f_s < 1$. (B) Six proteins plotted in the 3D space $\{S_n/S_t, S_c/S_t, S_t\}$ for estrogen stimulated/unstimulated MCF-7 cells showing scatter over the replicates [57]. (C) Proteins or protein complexes plotted in the 3D space $\{S_n/S_t, S_c/S_t, S_t\}$ for TBP-induced oxidative stress of human fibroblasts [56]. In panels B and C the axis perpendicular to the page is color coded for changes in total abundance (S_t). The orange bounding lines show 2-fold changes in (S_n/S_c).
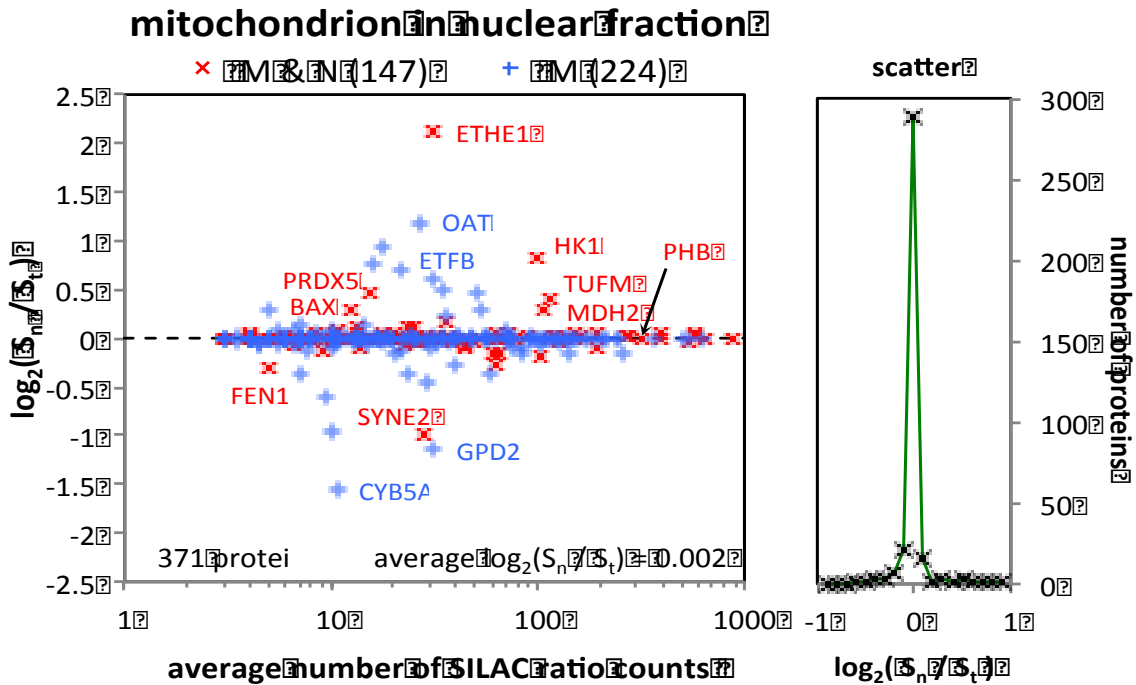
Figure 5. Analysis of the enrichment/purity of the nuclear fraction. Left: $\log_2(S_n/S_t) =$ $\log_2(f_s/f_u)$ as a function of the average number of ratio counts over the nucleus and total data sets for proteins with GO annotation to mitochondria and nucleus (red, 143 proteins) or to mitochondria but not nucleus (blue, 218 proteins). Right: number of proteins versus $\log_2(f_s/f_u)$.
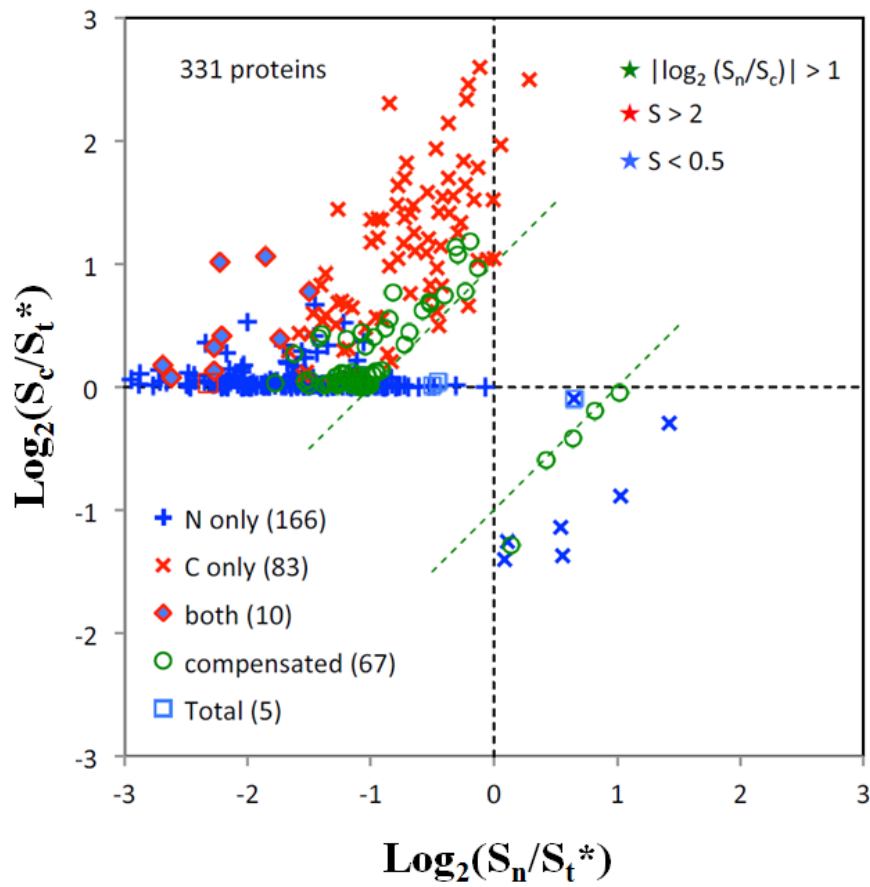
Figure 6. Three-dimensional plot for 331 proteins with significant changes upon estrogen stimulation of MCF7 cells. The legend (lower left) shows the symbol coding for >2-fold changes only in S_n, only in S_c, in both S_n and S_c, for S_n/S_c only, and for S_t. Increases/decreases are color coded (upper right legend).