HHS Public Access

Author manuscript

Curr Biol. Author manuscript; available in PMC 2016 October 19.

Published in final edited form as:

Curr Biol. 2015 October 19; 25(20): 2663–2671. doi:10.1016/j.cub.2015.08.047.

The Nuclear Proteome of a Vertebrate

Martin Wühr[1,2], **Thomas Güttler**[1], **Leonid Peshkin**[2], **Graeme C. McAlister**[1], **Matthew Sonnett**[1,2], **Keisuke Ishihara**[2], **Aaron C. Groen**[2], **Marc Presler**[2], **Brian K. Erickson**[1], **Timothy J. Mitchison**[2], **Marc W. Kirschner**[2,*], and **Steven P. Gygi**[1,*]

[1]Department of Cell Biology, Harvard Medical School, 02115 Boston, MA, USA

[2]Department of Systems Biology, Harvard Medical School, 02115 Boston, MA, USA

Summary

The composition of the nucleoplasm determines the behavior of key processes such as transcription, yet there is still no reliable and quantitative resource of nuclear proteins. Furthermore, it is still unclear how the distinct nuclear and cytoplasmic compositions are maintained. To describe the nuclear proteome quantitatively, we isolated the large nuclei of frog oocytes via microdissection and measured the nucleocytoplasmic partitioning of ~9000 proteins by mass spectrometry. Most proteins localize entirely to either nucleus or cytoplasm, only ~17% partition equally. A protein's native size in a complex, but not polypeptide-molecular-weight is predictive of localization: partitioned proteins exhibit native sizes larger than 100 kDa, while natively smaller proteins are equidistributed. To evaluate the role of nuclear export in maintaining localization, we inhibited Exportin 1. This resulted in the expected re-localization of proteins towards the nucleus, but only 3% of the proteome was affected. Thus, complex assembly and passive retention, rather than continuous active transport, is the dominant mechanism for the maintenance of nuclear and cytoplasmic proteomes.

Introduction

The organization of cells into membrane-enclosed compartments (i.e. organelles), each housing a characteristic set of macromolecules, is one of the foundations of complex, eukaryotic life [1]. Access of proteins to the nucleus is often highly regulated and controls critical steps in development, stress response, and general cell signaling [2].

Molecular traffic between nucleus and cytoplasm is routed through nuclear pore complexes (NPCs) embedded in the nuclear envelope [3]. These pores are permeable to ions, metabolites and small proteins (reported to be up to ~40 kDa in molecular weight) but do not allow larger macromolecules to pass efficiently unless they are bound by nuclear transport receptors (also called karyopherins) that include importins and exportins [4–6].

*Correspondence: marc@hms.harvard.edu, steven_gygi@hms.harvard.edu.

Their activity is rendered directional and energy-dependent by the coupling of transport to the RanGTPase system [7].

Despite the central role of the nucleus in multicellular biology, its protein content has never been satisfactorily catalogued, nor has the proteome's nucleocytoplasmic partitioning been quantified systematically. This is at least partly due to the fact that efficient separation of nuclear and cytoplasmic material remains a serious challenge: the time required for cell fractionation is long compared to the time it takes some nuclear proteins to escape via diffusion [4, 8]. Furthermore, the relative quantification of protein abundance on a proteome-wide scale is only recently possible thanks to advances in mass spectrometry.

How the nuclear proteome is established during nuclear formation and subsequently maintained during interphase remains an open question. In animals and plants, the nucleus disassembles during mitosis and is rebuilt thereafter. Nuclear import plays a fundamental role in establishing nuclear composition [9, 10]. Throughout interphase, which can last many years in some somatic cells, nuclear composition has to be maintained. This is a challenge as proteins smaller than ~40 kDa in molecular weight can pass nuclear pores freely. Diffusion of larger proteins is restricted but not completely prevented. Ultimately, this would lead to intermixing of nuclear and cytoplasmic contents. Continuous nuclear export has been shown to keep cytosolic proteins out of the nucleus [11]. As an alternative but not incompatible mechanism proteins may bind large structures like DNA or assemble into large protein complexes, thereby practically preventing their diffusion through the pores. For example, antibody fragments directed against histones remain in the nucleus even though they lack a nuclear localization signal [12]. The contributions of active transport and passive retention to the maintenance of distinct nuclear and cytoplasmic proteomes have never been systematically investigated on the level of the proteome. While retention makes sense for proteins tightly bound to chromatin, it is not at all clear that the soluble contents of the nucleus (or the cytoplasm) can be maintained that way.

Our initial goal was to use a simple but reliable method of nuclear purification, the manual isolation of the large nuclei of the frog oocyte, to generate a reliable catalog of nuclear and cytosolic proteins. These could be accurately quantified using two recently developed methods of quantitative proteomics. Since the state of complex formation would be concentration dependent, we assessed the native molecular weight of proteins in undiluted cytosol and analyzed how nucleocytoplasmic protein localization is affected by inhibition of the cell's major nuclear export pathway. This allowed us to address fundamental questions of how the nuclear content is maintained.

Results

Proteome-wide quantification of nucleocytoplasmic partitioning

Among organelles of eukaryotic cells the nucleus is unique in not having a continuous membrane segregating its internal contents from the cytosol. In isolation procedures performed with tissue culture cells, soluble nuclear proteins could diffuse out through the nuclear pores, as well as through any other breaches in the membrane adventitiously generated by detergent or mechanical isolation. These problems may have contributed to

poor agreement about just what is a nuclear protein. A remarkable exception to the problems of nuclear isolation is the microdissection of the millimeter-sized oocytes of amphibians. The giant nuclei (~400 μm diameter) of *Xenopus laevis* oocytes can be isolated manually, which minimizes loss of material due to comparatively quick isolation and the much longer time (about 10,000 fold) it would take proteins to diffuse on this length scale compared to somatic nuclei (Movie S1) [8]. To quantify nucleocytoplasmic protein partitioning in a proteome-wide manner, we determined relative nucleocytoplasmic protein concentrations in biological and technical triplicates using two different methods of accurate multiplexed proteomics (MultiNotch-MS3 and TMT^C) [13, 14] (Fig. 1A, Fig. S1) along with our recently described genome-free proteomics approach [15]. To further control for protein leakage, nuclear isolation for experiment 3 was performed under mineral oil. We also demonstrated that the leakage of NLS-GFP out of the nucleus is much slower than nuclear isolation (Movie S1, Movie S2). For each quantified protein we calculated the Relative Nuclear Concentration (RNC), defined as the ratio of concentrations in the nucleus to the concentrations in nucleus plus the cytoplasm (Fig. S1 B, C). The RNC values obtained from the three replicates agree well, with an R^2 of at least 0.94 (Fig. 1B, Fig. S2A). Together, we quantified the RNCs for 9262 proteins (Fig. S2B, Table S1A).

The RNC histogram revealed a distinct trimodal distribution: most proteins are localized almost exclusively to either nucleus or cytoplasm while a smaller third subset is nearly equally distributed (Fig. 1C). When we use RNC values of 1/3 and 2/3 for discrimination, we quantified 55% of proteins as cytoplasmic, 17% as equidistributed and 27% as nuclear. To compare and integrate our measurements with available metadata, which is typically human, we mapped the frog proteins to human homologs via a bidirectional best blast hit approach [15]. In the absence of a reliable nuclear proteome resource, we first evaluated the quality of our data by comparing it to a database of proteins that are confidently predicted to be non-nuclear, the human MitoCarta database, a highquality inventory of mitochondrial proteins [16]. Indeed, of the 489 proteins labeled as confidently mitochondrial (MitoCarta's combined false discovery rate < 1%) that were observed in our study, we classified 477 (98%) as extra-nuclear (RNC < 1/3) (Fig. 1D). These results validate our interspecies mapping approach and provide an unbiased quality assurance for our subcellular protein localization data.

There are several subcellular localization databases, including Protein Atlas [17], Uniprot [18], LocDB [19], and GO [20] (Fig. 1E, Fig. S2C–G). Each gives different predictions for the composition of the nuclear proteome. We observe poor agreement between our measurements and these databases. Might this discrepancy be explained by the different nuclear composition in oocytes compared to somatic cells these databases rely on? The weak agreement between these databases for the prediction of nuclear proteins makes this unlikely to be the main explanation (Fig. S3E). Furthermore, when we identified proteins that are annotated as nuclear by all four databases and compared this subset with the measured RNC values the agreement with our databases increases drastically. More than 80% of these proteins are identified as nuclear proteins in our data (RNC > 2/3) (Fig. 1E). The strong overlap of this subset with our data suggests that the nuclear proteome of the frog oocyte is

similar to that of human somatic cells and that our resource will be valuable to evaluate and improve human subcellular localization databases.

Correlation of nucleocytoplasmic partitioning and native molecular weight

Our data set allowed us to test the importance of the mechanisms proposed to be involved in nucleocytoplasmic partitioning. Two mechanisms have been suggested: First, some proteins may be retained in the nucleus or cytoplasm by virtue of their large hydrodynamic radii which would impede movement through the nuclear pores [4, 21]. Second, continuous (energy-dependent) nuclear transport might be required to reverse the inevitable intermixing of nuclear and cytoplasmic proteins that would result in free diffusion through pores [11]. Of course, the cell employs both mechanisms to maintain nuclear and cytoplasmic composition, but their relative contribution has never been assessed. We were now in a position to evaluate these models directly at the proteome-wide level.

To test if partitioned proteins are preferentially large, while equidistributed proteins tend to be small, we first compared the polypeptide-molecular-weight for cytoplasmic, equidistributed, and nuclear proteins. We found only a modest overrepresentation of low-molecular weight proteins (< 40 kDa) in the equidistributed fraction. In fact, many such proteins are either entirely nuclear or completely cytoplasmic (Fig. 2A). Yet, polypeptide mass is not a good indicator for the capacity to diffuse through nuclear pores. Rather, the native molecular weight (native MW) of a protein, which considers if a polypeptide chains might assemble into large complexes with other proteins or nucleic acids, is the much more appropriate measure. Although a number of distinct stoichiometric complexes are now known [22], our knowledge is likely far from comprehensive, and weaker and less specific assemblies, some of which would require the high concentration found in the cytosol, are generally elusive.

To determine whether the native size of proteins offered better discrimination between equidistributed proteins and those that are localized to either nucleus or cytoplasm, we developed a proteome-wide approach for estimating native protein size. We prepared undiluted frog egg extract by centrifugal crushing of packed eggs to minimize dilution of cellular contents, as such dilution might perturb complex formation. Unlike typical cell lysates egg extract is still “alive” by many criteria: it can form metaphase spindles [23], cycle between interphase and mitosis [24], and form nuclei [25]. We then centrifuged the extract through protein filters of two molecular weight cutoffs (30 kDa & 100 kDa, respectively) and compared input and filtered material by quantitative proteomics (Fig. 2B). These filters do not give binary fractionation; rather they admit proteins to an extent that varies continuously with molecular weight like many gel filtration materials. Thus, the degree of filtration yields graded information about the native size of a protein or complex. To integrate the information from both filtration steps into a single value, we projected each data point onto a spline [26] and obtained a proxy for native size (Fig. 2C). Comparing this proxy against the known native MW of proteins and protein complexes reported in the literature (Table S1B) revealed excellent correlation (R^2 of 0.95; Fig. 2D). This allowed us to estimate the native MW for ~3500 proteins (Table S1C). This filtration-based approach

should be generally applicable to investigate the formation of protein complexes and their dynamics in cell extracts.

Many proteins exhibited a much larger native MW than predicted by their mere polypeptide MW (Fig. 2E). For example small proteins in the anaphase promoting complex, the proteasome, or the ribosome migrated with a molecular weight of more than 250 kDa, the upper size limit we could resolve with the filters used (Suppl. Table 1C). While we saw only a weak correlation of polypeptide MW and RNC, (Fig. 2A) the native MW revealed a clear pattern of subcellular localization based on size (Fig. 2F): essentially all natively small proteins are nearly equilibrated between nucleus and cytoplasm (RNC ~0.5). In contrast, most natively large proteins preferentially segregate either to the nucleus or cytoplasm, with some important exceptions (see below). The observed transition is gradual and occurs at approximately 100 kDa. This is larger than the reported size exclusion limit of NPCs (~40 kDa) [6]. We do not understand this discrepancy. It is possible that the functional size exclusion limit of NPCs is larger than the limit measured previously in short-term experiments [4]; over longer time periods, larger proteins may equilibrate. Alternatively, oocyte NPCs might be more permeable than those of somatic nuclei. Furthermore, while the literature typically reports a ~40 kDa cutoff some studies have reported significantly larger cutoffs up to ~150 kDa [27, 28]. Nevertheless, our data strongly indicate that size exclusion by the NPC could maintain nucleocytoplasmic partitioning by preventing free diffusion of proteins and protein complexes larger than 100 kDa. That we observe hardly any small but partitioned proteins, suggests that the cell does not typically spend transport receptor binding capacity and energy to maintain a nucleocytoplasmic concentration gradient for proteins that would diffuse rapidly through the nuclear pore.

Although most natively large proteins preferentially localize to one side of the nuclear membrane or the other, there is a small set of equipartitioned and natively large proteins, which can be seen in Fig. 2F as the peak at RNC ~0.5 and native MW >250 kDa. This set includes members of highly studied complexes like the Anaphase Promoting Complex (APC/C) and the proteasome (Fig. 2G). We suspect some undiscovered mechanism equi-partitions these large complexes.

Effect of Exportin 1 inhibition on nucleocytoplasmic protein partitioning

It was proposed that continuous, energy-dependent nuclear export is required to keep cytoplasmic proteins out of the nucleus [11]. The nuclear export receptor Exportin 1 (CRM1) has been suggested to play a major role in the maintenance of nuclear identity and is believed to be the exportin with the most diverse cargo range [11, 29]. To assess the contribution of Exportin 1-mediated nuclear export to the maintenance of nuclear composition, we inhibited Exportin 1 with Leptomycin B (LMB) [30] and monitored nucleocytoplasmic protein distribution over time (Fig. 3A). The vast majority of proteins quantified in both replicates (6411 out of 6639, 97%) did not change their localization significantly, even after 24 hours of LMB treatment (Fig. 3B, C, Table S1D). LMB was clearly effective as the remaining 3% of the proteins shift their RNC significantly towards the nucleus. While in our experiment Exportin 1 does not seem to be required to keep the bulk of cytoplasmic proteins out of the nucleus, it is likely that its activity establishes these

localization patterns initially, i.e. when nuclei re-assemble after mitosis. It is also likely that Exportin 1 is required to maintain cytosolic protein localization over very long time-scales.

The proteins that did re-localize following LMB application are interesting (Fig. 3D). Most subunits of the equi-partitioned APC/C moved towards the nucleus (Fig. 3E), consistent with an active role of Exportin 1 in their equilibration, presumably in conjunction with an importin. In contrast, proteasome subunits did not respond to LMB (Fig. 3E), indicating that different mechanisms operate here. Overall, we identified only 226 proteins that shifted localization significantly (1% false discovery rate (FDR)) towards the nucleus following inhibition of Exportin 1 (Fig. 3C, Table S1D). Of these candidate Exportin 1 substrates, 187 have not previously been identified as Exportin 1 substrates (Fig. S3A). We further characterized some of the LMB responders in human tissue culture cells (Table S2). Notably, we saw no sign of native size dependence in this response to LMB (Fig. S3B).

Confidently identified proteins responding to LMB might be of particular therapeutic interest because Exportin 1 inhibitors recently emerged as promising anti-cancer drugs [31–33]. How they selectively kill some cancer cells is poorly understood. Most intriguingly, we identified 14 distinct kinases as LMB responders (Fig. 3F) [34, 35]. Many of these kinases have described roles in cancer biology, so it is an attractive hypothesis that perturbations in signaling pathways involving these kinases might be important in the anti-cancer effects of Exportin 1 inhibitors.

Discussion

Frog oocytes are a widely used model system to study the structure and function of the cell nucleus. Much of the work on nuclear transport, the structure and function of nuclear pores, and the physical structure of the nuclear lamina, was carried out in frog oocytes [36–38]. Despite some unique properties of these giant cells, their nuclei perform all typical functions of somatic nuclei, including transcription and splicing.

By applying state-of-the-art quantitative proteomics to this representative and well-studied model, we generated the first quantitative, and we believe, reliable, resource for proteome-wide nucleocytoplasmic partitioning. We anticipate that this data will be a very useful resource for the development and improvement of subcellular localization databases and prediction algorithms [17–20, 39]. There have been previous attempts to quantify nucleocytoplasmic partitioning with quantitative proteomics. However, with the biochemical fractionations used it is very hard to purify nuclei faithfully. [40] For example, a recent large-scale proteomics paper mis-classified 73% of mitochondrial proteins as nuclear (Fig. S4) [41].The finding that the vast majority of partitioned proteins are natively large (>100 kDa) suggests that passive retention (rather than continuous nuclear import or export) dominates in the maintenance of nuclear and cytosolic composition (Fig. 4) [42]. This hypothesis was further supported by the observation that only ~3% of the proteome responds significantly to 24 hours of Exportin 1 inhibition. Furthermore, it is hard to imagine that there are sufficient import and export receptors to maintain this exclusive distribution by continuous active partitioning, alone. The total concentration of proteins exclusive to nucleus or cytoplasm can be estimated at ~2 mM [15]. This is ~200-fold higher than the

estimated ~10 μM of all import and export receptors found in the oocyte, as calculated from the same source (Fig. S3C). It seems to be an inescapable conclusion that a protein-autonomous mechanism such as passive retention is required to maintain nuclear composition in eukaryotic cells. We expect this to be true also for smaller somatic cells. However, it will be important to test this hypothesis experimentally.

Our conclusions do not diminish the importance of active nuclear transport in nucleocytoplasmic compartmentation: there is no doubt that import and export are required to segregate nuclear from cytoplasmic contents after mitosis, when the nucleus re-forms [9, 10, 36]. However, how much passive, merely size-dependent compartmentation mechanism contribute to the maintenance of pre-established localization was unknown. Is active nuclear transport at all required to separate nucleus and cytoplasm during interphase [11]? It surely will be vital to re-localize large complexes that can diffuse appreciably through NPCs over very long timescales, i.e. for cells with long interphases or post mitosis [43]. Large complexes might also disassemble over time, allowing their smaller components access to their non-steady-state compartment. Furthermore, normally cytosolic proteins that fail to assemble into their native complexes following their biosynthesis might enter the nucleus by diffusion or active import. This is the case for poly-basic proteins (such as RNA-binding translation factors) whose charged domains often act as cryptic nuclear import signals [11]. In fact, importins operate as chaperones for exposed basic domains [44].

In this study we did not analyze post-translational modifications, like phosphorylation. For some proteins we might have inadvertently averaged the subcellular localization of multiple distinct protein species. With quantitative phosphoproteomics [45] the role of phosphorylation on the proteome’s subcellular localization could be studied systematically.

Perhaps most surprisingly, our work revealed that the majority of the cell's small proteins is found in complexes greater than 100 kDa in molecular weight. This seems to contradict biochemical experience. However, in such experiments dilution and fractionation could easily dissociate large molecular assemblies. Furthermore, small proteins are easiest to purify while fractions found in large assemblies may be easily missed. Our results raise the question of whether protein-protein interactions at concentrations of ~100 mg/ml may enable many interactions that are simply not seen in *in-vitro* conditions. There is anecdotal evidence in many cases where concentrated extracts diluted even 2 or 3 fold fail to carry out complex processes, like spindle formation, nuclear assembly, and cell cycle progression. This aspect of the conclusion in this study, which after all assays protein distributions under native cellular conditions, warrants further study.

Finally, though we have stressed the generality of these mechanisms of nucleocytoplasmic partitioning, there will undoubtedly be differences between oocytes and somatic cells. The nuclear proteins identified in this study appear to be mostly common to all cell types but some are known to be special to the oocyte nucleus, for example, those involved in maintaining chromosomes for months in diplotene stage, or those that enable the oocyte to reprogram somatic nuclei to totipotency [46]. While the comparison among different cell types could also be done via imaging methods, this would be very labor intensive and time consuming. Recently, very quick nuclear isolation methods for somatic cells have been

developed [47]. Combining these with the quantitiative proteomics analysis described here might be a promising strategy for nuclear proteome analysis of somatic cells.

Experimental Procedures

Nuclear Isolation

Isolation of *X. laevis* oocytes was done essentially as previously described [48]. J line (National Xenopus Resource Center, Woods Hole, MA) females were anaesthetized with 0.2% Tricaine, and ovary lobes were surgically removed under sterile conditions. Oocytes were manually defolliculated and maintained in OCM, (320 mL sterile water, 480 mL Liebovitz medium (L-15) with glutamine (Sigma), 0.32 g bovine serum albumen (BSA; Sigma), 4 mL penicillin/streptomycin; pH was adjusted to 7.7 with NaOH). Ooctytes were allowed to recover overnight before experiments. Before sample collection, oocytes were washed three times with MMR (0.1 M NaCl, 2 mM KCl, 1 mM MgSO4, 2 mM CaCl2, 5 mM HEPES (pH 7.8), 0.1 mM EDTA) to remove BSA. For Experiments LMB-1 and LMB-2, nuclei were isolated in MMR, for Experiment RNC-TMT^C the nuclei were isolated under mineral oil (Sigma). For ‘Experiments LMB-1’ and ‘LMB-2’ oocytes were transferred into MMR with 200nM Leptomycin B. For each time experimental condition 40–50 oocytes were separated into nucleus and cytoplasm and immediately frozen on dry ice. The control II for LMB-2 was collected, without drug treatment, after the 24 h in LMB samples were collected to control for effects solely due to time outside the ovary. To confirm cell viability after 24 hours in LMB, their ability to respond to 3nM progesterone was assayed (not shown), [49]. Untreated cells were marked with Nile blue and co-imaged [48]. Samples were lysed with 250 mM Sucrose, 1% NP40 Substitute (Sigma), 5mM EDTA (pH 7.2), 1 Roche Complete mini tablet (EDTA free), 20 mM HEPES (pH 7.2), 10 μM Combretastatin 4A, and 10 μM Cyochalasin D [15]. Lysate was vortexed at maximum speed for ten seconds, pipetted ten times up and down with a 200 μL pipette tip, incubated on ice for 10 minutes, and again vortexed for ten seconds. Lysates were clarified by centrifugation at 7,500 g at 4° C for 4 minutes in a tabletop centrifuge. After gentle flicking to resuspend lipids, supernatant was removed and used for further analysis. For the GFP-NLS leakage experiment (Movie S2) 10 nL of 28mg/mL of NLS-GFP (kind gift of Daniel Levy) were injected into stage IV oocytes. After ~24 hours nuclei were isolated manually and one picture taken with bright field illumination under a dissection microscope (for the movie S2 this picture was replicated and shown as t = 0.0 min). After switching to fluorescent imaging the leakage of GFP-NLS out of the nucleus was followed in 10 second intervals.

Filter Percolation Experiment

Xenopus egg extract was prepared as previously described [23]. Extract was released into interphase by addition of 0.4 mM Ca and incubated for 20 min at RT. Aliquots were flash frozen for further analysis. In technical duplicates 200 μL of extract were added to Amicon Ultra-0.5ml Centrifugal Filter Units with 30 kDa nominal molecular weight cutoff (Millipore), and 90 μL were added to Amicon Ultra-0.5ml Centrifugal Filter Units with 100 kDa nominal molecular weight cutoff (Millipore). Filters were centrifuged for 30 minutes at 20°C at 5000 g. The ~65 μL of 30 kDa percolate were frozen for further analysis. The ~32

μL of 100 kDa percolate was also frozen for further analysis. 0.8 μL of crude extract, 11 μL of 100kDa filtrate, and 30 μL of 30kDa filtrate, were used for MS-analysis.

Data Analysis for Nucleocytoplasmic Partitioning Experiments

Human gene symbols were assigned to all sequences based on reciprocal best BLAST hit against human proteins available from UniProt as previously described [15]. The ratio of nuclear to cytoplasmic content that match the following gene symbols of equidistributed proteins (PFN1, ACTB, MDH1, TPI1, PGK1, GRHPR, HBZ, ALOXE3, GSTO1, TALDO1, HSPA1A, FAM115C, GSTM1, FABP4, SOD1, CFL1) were calculated for each experimental condition. To correct pipetting errors, the nuclear signal from the corresponding condition was divided by this ratio. For LMB-2, the two controls were averaged to provide the Experiment 1 RNC result. The Canonical RNC (Suppl. Table 1A) was calculated by averaging the RNC values from LMB-1, LMB-2, and RNC-TMT^{C}. If the RNC variance between replicates was larger than 0.05, the canonical RNC was not calculated. This was the case for 96 out of 9358 quantified proteins.

Data Analysis for Leptomycin B experiments

The ratio of nuclear to cytoplasmic content that match gene symbols of equidistributed proteins (PFN1, ACTB, MDH1, TPI1, PGK1, GRHPR, HBZ, ALOXE3, GSTO1, TALDO1, HSPA1A, FAM115C, GSTM1, FABP4, SOD1, CFL1) were calculated for each experimental condition. To correct pipetting errors, the nuclear signal from the corresponding condition was divided by this ratio as described above. Because slight errors in normalization would result in a large number of false positive responders to Leptomycin B, we further normalized each condition with LMB and the second control in Experiment II, so that the median signal was equivalent to the corresponding control. Note that this likely will lead to a slight underestimation for the actual movement of proteins towards the nucleus. In the LMB-2 experiment, when RNC values between biological replicates (2× Control, or 2× 24h LMB) disagreed by more than 4 average standard deviations, the protein was not quantified. Furthermore, proteins were filtered out if the RNC value of the 12h time-point was more than 4 standard deviations outside the Control or 24h time-point. For Experiment LMB-1, proteins were filtered out if the 2h time-point was more than 4 standard deviations outside the Control or 24h time-point. Importantly, for all filtering conditions, we did not make any assumptions about the directionality of the movement. For the final LMB responders, we only considered proteins which were quantified successfully in both LMB experiments.

Data Analysis for Physiological Protein Size Measurement

In technical duplicates, the ratios of flow-through over input were measured. The ratios were capped at 2×10^{-4} and 2×10^{6} for the 30 kDa Filter and 2×10−3 and 2×106 for the 100 kDa filter, which is the approximate maximum dynamic range for these measurements. Protein's ratios which differed by more than 7 average standard deviation (in log-space) between biological replicates were filtered out. Ratios from biological replicates were averaged in log-space. Theoretical protein size was estimated by multiplying the number of amino acids with 110 Da. The spline was fit through data, using the slmengine function from Mathwork File Exchange created by John D'Errico. We generated the spline out of 3rd polynomial

segments with four knots forced to be continuously increasing. To project the ratios for proteins onto the spline and measure its distance, we used the xy2sn function from Mathworks Exchange created by Juernjakob Dugge [26]. The resulting “proxy for protein size” was plotted against native protein sizes from the literature [18, 50–52] (Suppl. Table 1B). The correlation was used to estimate the physiological protein size in the experiment. We capped physiological protein sizes at minimally 2^2 kDa and maximally 2^8 kDa, which we estimate to be the maximum dynamic range for this experiment.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by NIH grants R01GM103785, R01HD073104 to MWK and R01GM39565 to TJM. MW was supported by the Charles A. King Trust Postdoctoral Fellowship Program, Bank of America, N.A., Co-Trustee. TG was supported by postdoctoral fellowships from the Human Frontier Science Program (HFSP) and the European Molecular Biology Organization (EMBO) and the Charles A. King Trust Postdoctoral Research Fellowship Program, Bank of America, N.A., Co-Trustee / Sara Elizabeth O'Brien Trust. MS was supported by NIH grant GM095450. We would like to thank the PRIDE team for proteomic data distribution. We thank Daniel Levy for the kind gift of NLS-GFP, Ian Swinburne and Sean Megason for access to their LSM, the HMS Nikon Imaging Center for access to spinning disc microscopes, and Chris Field and Fabian Romano for Xenopus antibodies. We thank Woong Kim, Robert Everley, Willi Haas, and Joao Paulo for help with mass spectrometers, the Gygi computational team for bioinformatics support. Thanks to Raphael Bruckner, Alban Ordureau, Laura Pontano Vaites for helping with the tissue culture experiment, and Tom Rapoport, Becky Ward, and Rosy Hosking for comments on the manuscript. Images of MS-instruments were used with kind permission from Thermo Fisher Scientific, the copyright owner. The illustration of the kinome tree was reproduced courtesy of Cell Signaling Technology, Inc.

References

1. Cavalier-Smith T. Cell evolution and Earth history: stasis and revolution. Philos Trans R Soc Lond B Biol Sci. 2006; 361:969–1006. [PubMed: 16754610]
2. Poon IK, Jans DA. Regulation of nuclear transport: central role in development and transformation? Traffic. 2005; 6:173–186. [PubMed: 15702986]
3. Rout MP, Aitchison JD, Suprapto A, Hjertaas K, Zhao Y, Chait BT. The yeast nuclear pore complex: composition, architecture, and transport mechanism. J Cell Biol. 2000; 148:635–651. [PubMed: 10684247]
4. Mohr D, Frey S, Fischer T, Guttler T, Gorlich D. Characterisation of the passive permeability barrier of nuclear pore complexes. The EMBO journal. 2009; 28:2541–2553. [PubMed: 19680228]
5. Gorlich D, Henklein P, Laskey RA, Hartmann E. A 41 amino acid motif in importin-alpha confers binding to importin-beta and hence transit into the nucleus. The EMBO journal. 1996; 15:1810–1817. [PubMed: 8617226]
6. Bagley S, Goldberg MW, Cronshaw JM, Rutherford S, Allen TD. The nuclear pore complex. Journal of Cell Science. 2000; 113(Pt 22):3885–3886. [PubMed: 11058073]
7. Izaurralde E, Kutay U, von Kobbe C, Mattaj IW, Gorlich D. The asymmetric distribution of the constituents of the Ran system is essential for transport into and out of the nucleus. The EMBO journal. 1997; 16:6535–6547. [PubMed: 9351834]
8. Paine PL, Austerberry CF, Desjarlais LJ, Horowitz SB. Protein loss during nuclear isolation. J Cell Biol. 1983; 97:1240–1242. [PubMed: 6619193]
9. Newport JW, Wilson KL, Dunphy WG. A lamin-independent pathway for nuclear envelope assembly. J Cell Biol. 1990; 111:2247–2259. [PubMed: 2277059]
10. D'Angelo MA, Anderson DJ, Richard E, Hetzer MW. Nuclear pores form de novo from both sides of the nuclear envelope. Science. 2006; 312:440–443. [PubMed: 16627745]

11. Bohnsack MT, Regener K, Schwappach B, Saffrich R, Paraskeva E, Hartmann E, Gorlich D. Exp5 exports eEF1A via tRNA from nuclei and synergizes with other transport pathways to confine translation to the cytoplasm. The EMBO journal. 2002; 21:6205–6215. [PubMed: 12426392]
12. Einck L, Bustin M. Functional histone antibody fragments traverse the nuclear envelope. J Cell Biol. 1984; 98:205–213. [PubMed: 6707085]
13. McAlister GC, Nusinow DP, Jedrychowski MP, Wühr M, Huttlin EL, Erickson BK, Rad R, Haas W, Gygi SP. MultiNotch MS3 Enables Accurate, Sensitive, and Multiplexed Detection of Differential Expression across Cancer Cell Line Proteomes. Analytical Chemistry. 2014; 86:7150–7158. [PubMed: 24927332]
14. Wühr M, Haas W, McAlister GC, Peshkin L, Rad R, Kirschner MW, Gygi SP. Accurate multiplexed proteomics at the MS2 level using the complement reporter ion cluster. Analytical Chemistry. 2012; 84:9214–9221. [PubMed: 23098179]
15. Wühr M, Freeman RM Jr, Presler M, Horb ME, Peshkin L, Gygi SP, Kirschner MW. Deep Proteomics of the Xenopus laevis Egg using an mRNA-Derived Reference Database. Current biology : CB. 2014
16. Pagliarini DJ, Calvo SE, Chang B, Sheth SA, Vafai SB, Ong SE, Walford GA, Sugiana C, Boneh A, Chen WK. A mitochondrial protein compendium elucidates complex I disease biology. Cell. 2008; 134:112–123. [PubMed: 18614015]
17. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, Zwahlen M, Kampf C, Wester K, Hober S. Towards a knowledge-based Human Protein Atlas. Nature Biotechnology. 2010; 28:1248–1250.
18. Update on activities at the Universal Protein Resource (UniProt) in 2013. Nucleic Acids Research. 2013; 41:D43–D47. [PubMed: 23161681]
19. Rastogi S, Rost B. LocDB: experimental annotations of localization for Homo sapiens and Arabidopsis thaliana. Nucleic Acids Research. 2011; 39:D230–D234. [PubMed: 21071420]
20. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Research. 2004; 32:D258–D261. [PubMed: 14681407]
21. Paine PL, Feldherr CM. Nucleocytoplasmic exchange of macromolecules. Exp Cell Res. 1972; 74:81–98. [PubMed: 4342186]
22. Ruepp A, Waegele B, Lechner M, Brauner B, Dunger-Kaltenbach I, Fobo G, Frishman G, Montrone C, Mewes HW. CORUM: the comprehensive resource of mammalian protein complexes--2009. Nucleic acids research. 2010; 38:D497–D501. [PubMed: 19884131]
23. Sawin KE, Mitchison TJ. Mitotic spindle assembly by two different pathways in vitro. J Cell Biol. 1991; 112:925–940. [PubMed: 1999463]
24. Murray AW, Kirschner MW. Cyclin synthesis drives the early embryonic cell cycle. Nature. 1989; 339:275–280. [PubMed: 2566917]
25. Hartl P, Olson E, Dang T, Forbes DJ. Nuclear assembly with lambda DNA in fractionated Xenopus egg extracts: an unexpected role for glycogen in formation of a higher order chromatin intermediate. J Cell Biol. 1994; 124:235–248. [PubMed: 8294509]
26. Merwade VM, Maidment DR, Hodges BR. Geospatial representation of river channels. J. Hydrol. Eng. 2005; 10:243–251.
27. Wang R, Brattain MG. The maximal size of protein to diffuse through the nuclear pore is larger than 60kDa. FEBS Lett. 2007; 581:3164–3170. [PubMed: 17588566]
28. Popken P, Ghavami A, Onck PR, Poolman B, Veenhoff LM. Size-dependent leak of soluble and membrane proteins through the yeast nuclear pore complex. Mol Biol Cell. 2015; 26:1386–1394. [PubMed: 25631821]
29. Guttler T, Gorlich D. Ran-dependent nuclear export mediators: a structural perspective. The EMBO journal. 2011; 30:3457–3474. [PubMed: 21878989]
30. Kudo N, Matsumori N, Taoka H, Fujiwara D, Schreiner EP, Wolff B, Yoshida M, Horinouchi S. Leptomycin B inactivates CRM1/exportin 1 by covalent modification at a cysteine residue in the central conserved region. Proceedings of the National Academy of Sciences of the United States of America. 1999; 96:9112–9117. [PubMed: 10430904]

31. Walker CJ, Oaks JJ, Santhanam R, Neviani P, Harb JG, Ferenchak G, Ellis JJ, Landesman Y, Eisfeld AK, Gabrail NY. Preclinical and clinical efficacy of XPO1/CRM1 inhibition by the karyopherin inhibitor KPT-330 in Ph+ leukemias. Blood. 2013; 122:3034–3044. [PubMed: 23970380]
32. Senapedis WT, Baloglu E, Landesman Y. Clinical translation of nuclear export inhibitors in cancer. Semin Cancer Biol. 2014; 27 C:74–86. [PubMed: 24755012]
33. Turner JG, Dawson J, Emmons MF, Cubitt CL, Kauffman M, Shacham S, Hazlehurst LA, Sullivan DM. CRM1 Inhibition Sensitizes Drug Resistant Human Myeloma Cells to Topoisomerase II and Proteasome Inhibitors both In Vitro and Ex Vivo. Journal of Cancer. 2013; 4:614–625. [PubMed: 24155773]
34. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. Science. 2002; 298:1912–1934. [PubMed: 12471243]
35. Chartier M, Chenard T, Barker J, Najmanovich R. Kinome Render: a stand-alone and web-accessible tool to annotate the human protein kinome tree. PeerJ. 2013; 1:e126. [PubMed: 23940838]
36. Dingwall C, Sharnick SV, Laskey RA. A polypeptide domain that specifies migration of nucleoplasmin into the nucleus. Cell. 1982; 30:449–458. [PubMed: 6814762]
37. Paine PL, Moore LC, Horowitz SB. Nuclear envelope permeability. Nature. 1975; 254:109–114. [PubMed: 1117994]
38. Aebi U, Cohn J, Buhle L, Gerace L. The nuclear lamina is a meshwork of intermediate-type filaments. Nature. 1986; 323:560–564. [PubMed: 3762708]
39. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K. WoLF PSORT: protein localization predictor. Nucleic Acids Res. 2007; 35:W585–W587. [PubMed: 17517783]
40. Mosley AL, Florens L, Wen Z, Washburn MP. A label free quantitative proteomic analysis of the Saccharomyces cerevisiae nucleus. J Proteomics. 2009; 72:110–120. [PubMed: 19038371]
41. Boisvert FM, Ahmad Y, Gierlinski M, Charriere F, Lamont D, Scott M, Barton G, Lamond AI. A quantitative spatial proteomics analysis of proteome turnover in human cells. Mol Cell Proteomics. 2012; 11 M111 011429.
42. Feldherr CM, Pomerantz J. Mechanism for the selection of nuclear polypeptides in Xenopus oocytes. J Cell Biol. 1978; 78:168–175. [PubMed: 566759]
43. Hetzer MW. The role of the nuclear pore complex in aging of post-mitotic cells. Aging (Albany NY). 2010; 2:74–75. [PubMed: 20354266]
44. Jakel S, Mingot JM, Schwarzmaier P, Hartmann E, Gorlich D. Importins fulfil a dual function as nuclear import receptors and cytoplasmic chaperones for exposed basic domains. The EMBO journal. 2002; 21:377–386. [PubMed: 11823430]
45. Erickson BK, Jedrychowski MP, McAlister GC, Everley RA, Kunz R, Gygi SP. Evaluating multiplexed quantitative phosphopeptide analysis on a hybrid quadrupole mass filter/linear ion trap/orbitrap mass spectrometer. Anal Chem. 2015; 87:1241–1249. [PubMed: 25521595]
46. Gurdon JB, Elsdale TR, Fischberg M. Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei. Nature. 1958; 182:64–65. [PubMed: 13566187]
47. Katholnig K, Poglitsch M, Hengstschlager M, Weichhart T. Lysis gradient centrifugation: a flexible method for the isolation of nuclei from primary cells. Methods Mol Biol. 2015; 1228:15–23. [PubMed: 25311118]
48. Mir A, Heasman J. How the mother can help: studying maternal Wnt signaling by antisense-mediated depletion of maternal mRNAs and the host transfer technique. Methods in molecular biology. 2008; 469:417–429. [PubMed: 19109723]
49. Gerhart J, Wu M, Kirschner M. Cell cycle dynamics of an M-phase-specific cytoplasmic factor in Xenopus laevis oocytes and eggs. J Cell Biol. 1984; 98:1247–1255. [PubMed: 6425302]
50. Fasman, GD. Practical Handbook of Biochemistry and Molecular Biology. Taylor & Francis; 1989.
51. Sober, HA. CRC Handbook of Biochemistry: Selected Data for Molecular Biology. CRC Press; 1970.
52. Squire PG, Himmel ME. Hydrodynamics and protein hydration. Arch Biochem Biophys. 1979; 196:165–177. [PubMed: 507801]

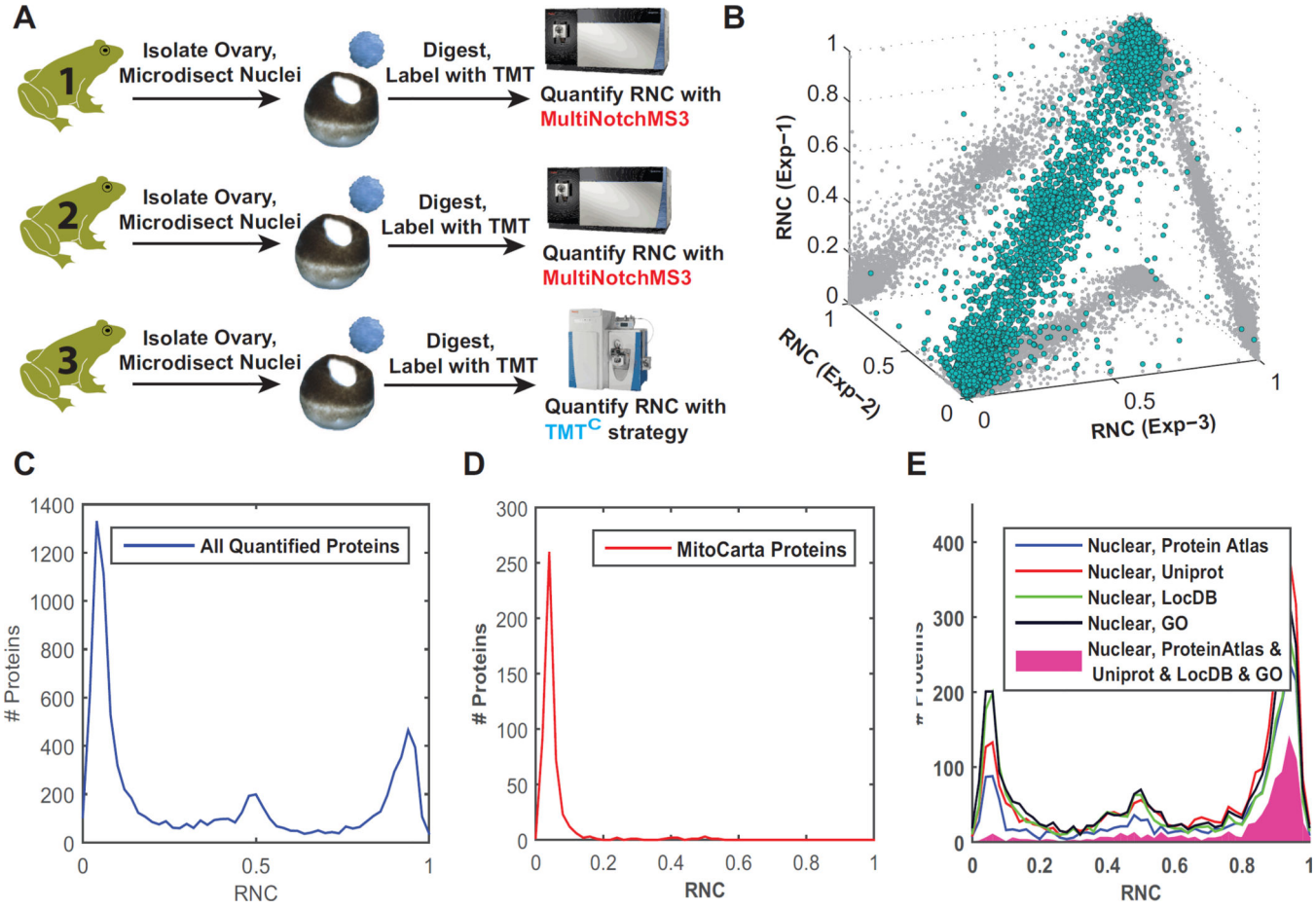

Figure 1.

Quantification of nucleocytoplasmic partitioning of the *X. laevis* oocyte proteome. **A)** Oocytes were dissected manually in three replicates, proteins digested, TMT-labeled and analyzed separately, with two different methods of accurate quantitative proteomics (MultiNotch- MS3 and TMT^C). **B)** The **R**elative **N**uclear **C**oncentration (RNC) was determined for 9262 proteins. The replicates correlated with an R^2 of at least 0.94. **C)** RNC histogram of all quantified proteins. **D)** Histogram of RNC values for proteins matched with the human MitoCarta database. **E)** RNC histogram for proteins classified as nuclear within four commonly used subcellular localization databases are highly enriched for truly nuclear proteins (purple). However, the individual databases show only moderate agreement among themselves and with our data.

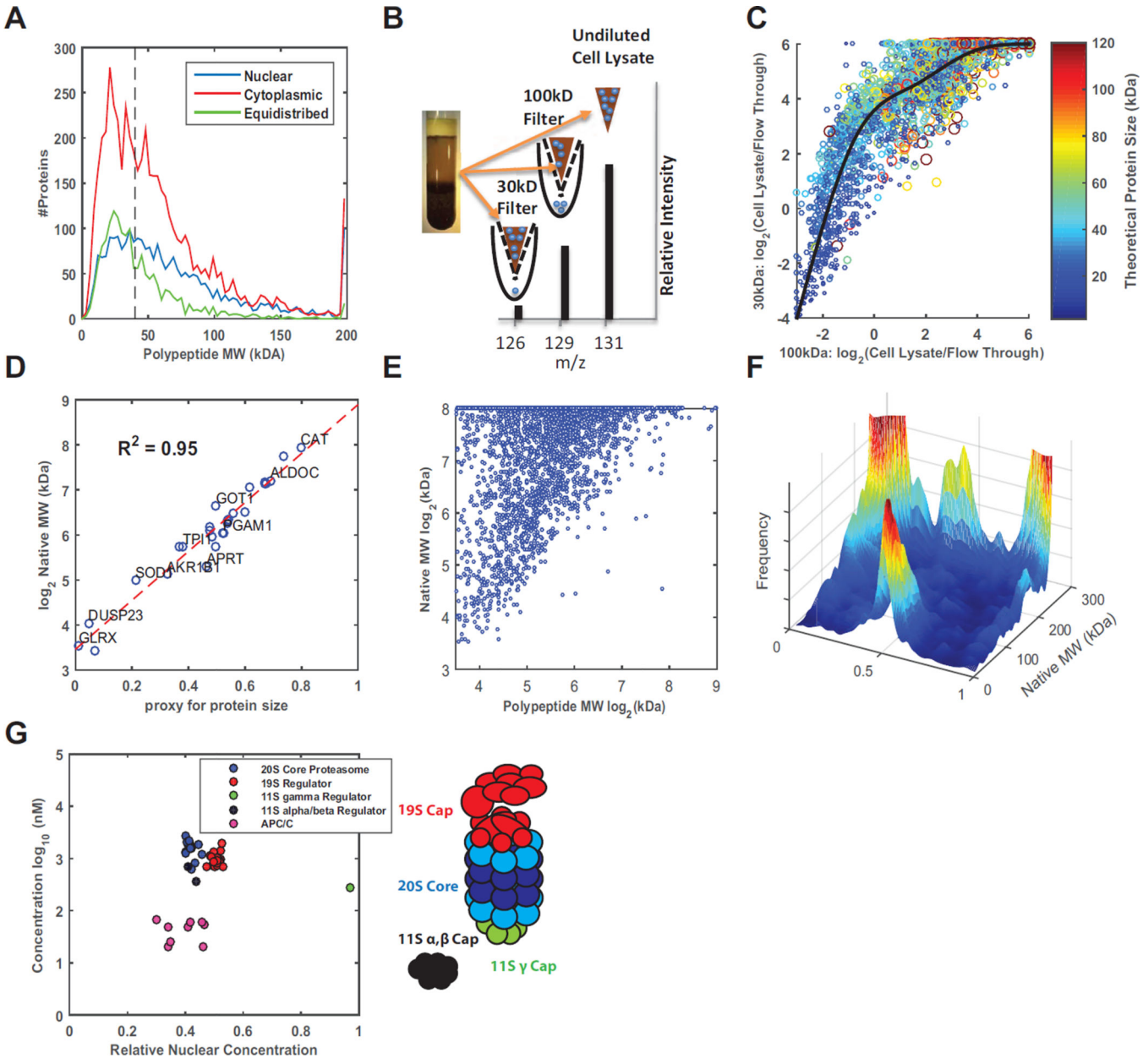

Figure 2.

Correlation of molecular weight and nucleocytoplasmic partitioning. **A)** Polypeptide MW is not a strong determinant of nucleocytoplasmic distribution. **B)** To estimate native protein sizes, cell lysate was percolated through filters of 30 or 100 kDa MW cutoff, respectively. The proteins' relative passage was quantified with the MultiNotch-MS3 approach. **C)** Ratios of input and flow-through of the indicated filters were plotted and fitted with a spline. Color code and data point size indicate polypeptide MW. Data point projection onto the spline yielded a "proxy for protein size", ranging from 0 (small – bottom left) to 1 (large – top right). **D)** This "proxy for protein size" and the experimentally determined native MW for various vertebrate proteins correlate with an R^2 of 0.95. This relationship allowed us to regress the native proteins size in a proteome-wide fashion. **E)** Plot of native MW versus

polypeptide MW indicates that many proteins behave significantly larger than their polypeptide MW suggests. The few proteins for which we measured smaller native MW than polypeptide MW likely represent measurement errors. **F)** Histogram relating native MW and RNC. Proteins smaller than ~100 kDa are preferentially equipartitioned whereas partitioned proteins are typically larger. However, a subset of natively large proteins is close to equipartitioned. Among them we found the proteasome and APC/C. **G)** Plot of estimated concentrations and RNCs for the subunits of the proteasome and the APC/C. Interestingly, the 19S and 11S α, β caps are slightly more nuclear than the core proteasome. In contrast, the 11S γ cap is exclusively nuclear.

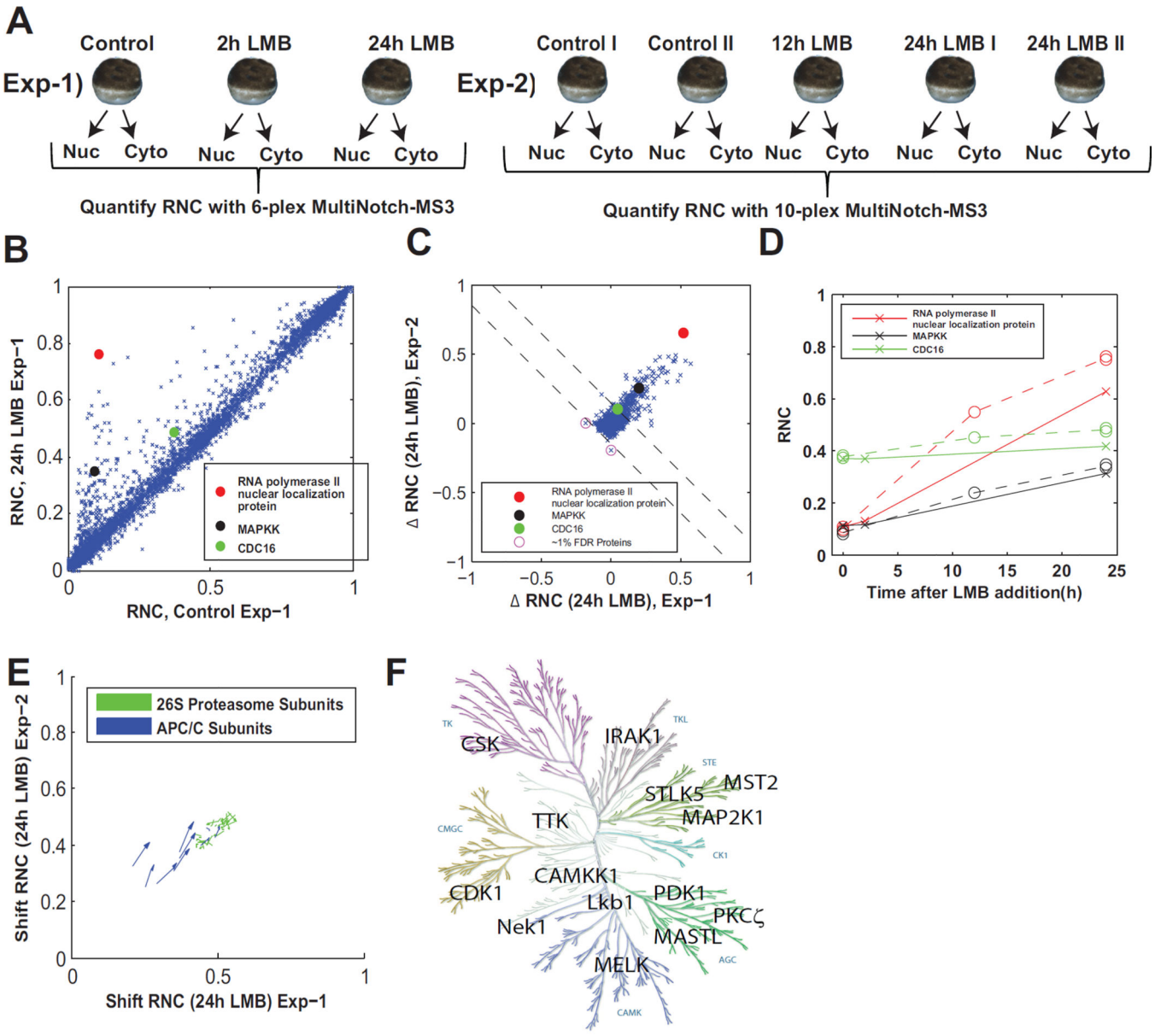

Figure 3.

Nucleocytoplasmic protein partitioning upon inhibition of Exportin 1. **A)** Experimental setup to determine the change of RNCs upon inhibition of Exportin 1 with LMB. **B)** RNCs determined for control oocytes and oocytes treated with LMB (24 h) were plotted (experiment 1). The majority of proteins did not change its localization significantly (97%). Three proteins, which re-localized to the nucleus, are highlighted for illustration. **C)** Scatter plot of RNC changes after 24h in LMB for (experiments 1 and 2). Under the assumption of noise being symmetric and LMB causing nuclear, but not cytoplasmic re-localization, we could estimate the false discovery rate (FDR) of LMB responders. With an FDR cutoff of ~1% (dotted lines) we detected 226 confident LMB responders. **D)** RNCs for all time points and replicates for the three highlighted proteins. **E)** Most subunits of the APC/C responded to LMB, suggesting that at least some large complexes present in nucleus and cytoplasm

(Fig. 2F) are equipartitioned via active bidirectional transport. We did not see any evidence for Exportin 1-dependent nuclear transport of the proteasome. **F)** Kinases are overrepresented among LMB responders (p-value = 0.002). The diagram shows these kinases.

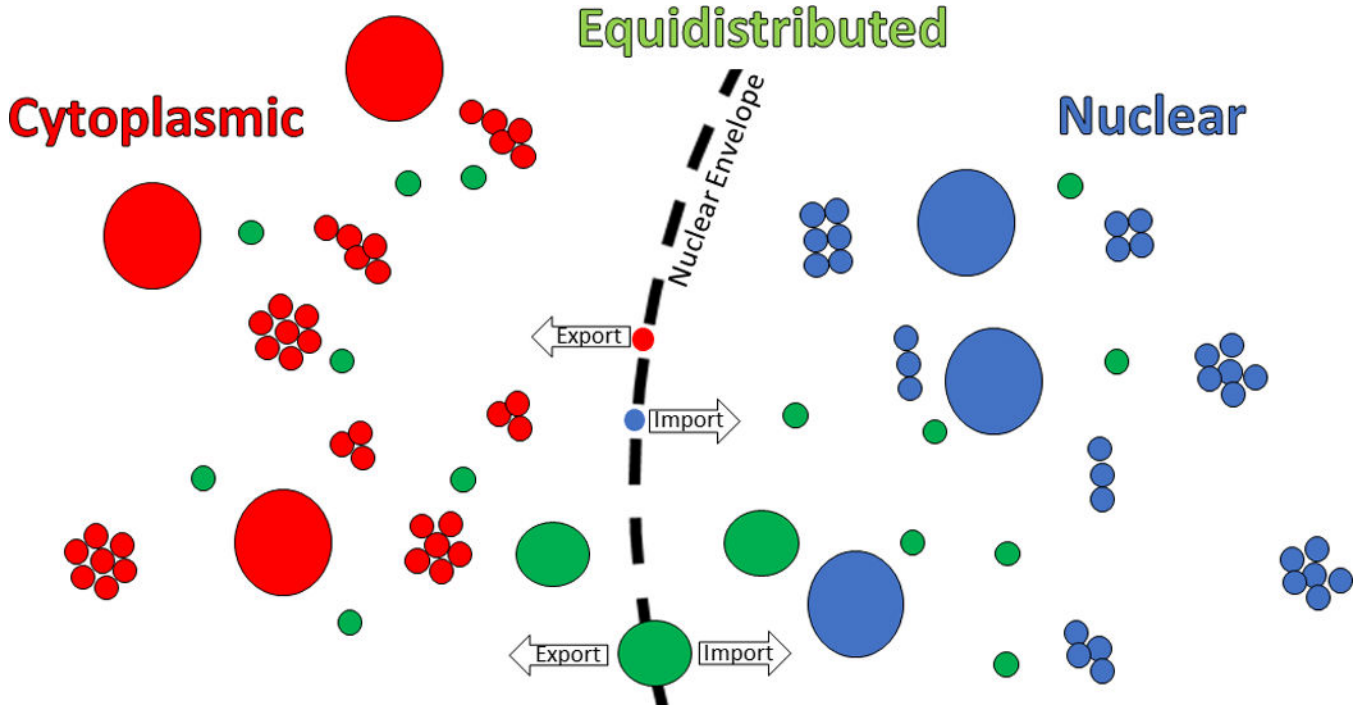

Figure 4.

The maintenance of nucleocytoplasmic partitioning is dominated by passive retention. Nuclear pores (depicted as holes in the nuclear envelope) permit the passage of small molecules but restrict that of larger ones. We observed that the vast majority of proteins smaller than ~100 kDa (small green circles) have similar concentrations in nucleus and cytoplasm. Diffusion through nuclear pores allows these proteins to equilibrate between nucleus and cytoplasm. Nearly all partitioned proteins (red or blue) have a native molecular weight larger than ~100 kDa, which prevents efficient diffusion through nuclear pores. Only very few natively small proteins are partitioned via continuous active transport. We also find a subset of natively large but equipartitioned proteins (large green circles). For some of these we provide evidence that they are equilibrated by active bidirectional transport.