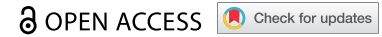Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

Evolution and diversification of the nuclear envelope

Norma E. Padilla-Mejia[a], Alexandr A. Makarov[a], Lael D. Barlow[a], Erin R. Butterfield[a], and Mark C. Field [a,b]

[a]Division of Biological Chemistry and Drug Discovery, School of Life Sciences, University of Dundee, Dundee, UK; [b]Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České, Czech Republic

ABSTRACT

Eukaryotic cells arose ~1.5 billion years ago, with the endomembrane system a central feature, facilitating evolution of intracellular compartments. Endomembranes include the nuclear envelope (NE) dividing the cytoplasm and nucleoplasm. The NE possesses universal features: a double lipid bilayer membrane, nuclear pore complexes (NPCs), and continuity with the endoplasmic reticulum, indicating common evolutionary origin. However, levels of specialization between lineages remains unclear, despite distinct mechanisms underpinning various nuclear activities. Several distinct modes of molecular evolution facilitate organellar diversification and to understand which apply to the NE, we exploited proteomic datasets of purified nuclear envelopes from model systems for comparative analysis. We find enrichment of core nuclear functions amongst the widely conserved proteins to be less numerous than lineage-specific cohorts, but enriched in core nuclear functions. This, together with consideration of additional evidence, suggests that, despite a common origin, the NE has evolved as a highly diverse organelle with significant lineage-specific functionality.

Introduction

Eukaryotes are estimated to have arisen over one and a half billion years ago – an event considered to be one of the major evolutionary transitions [1–3]. A consensus model is emerging [2], with the first eukaryotes likely evolving from within the Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota (TACK) superphylum [4]. Advances in environmental sampling and sequencing have superseded earlier rRNA-based models, which proposed three domains of eukaryotic life [5], to firmly place eukaryotes as emerging from within the Archaea and, hence, supporting a two domain paradigm [2]. At some point following the divergence from the archaeal lineage, the protoeukaryote began to acquire features typical of eukaryotic cells, including a tubulin-based flagellum, an endomembrane system, a cytoskeleton and the mitochondrion. Remarkably, this period of evolution produced a complex Last Eukaryotic Common Ancestor (LECA) that possessed a large diversity of compartments exceeding the complement of many extant unicellular organisms.

The fossil record is poor for this early period of evolution and preservation of cellular structure is of insufficient quality to draw unequivocal conclusions concerning the internal morphologies of early eukaryotic cells [6]. Consequently, molecular reconstruction of the evolutionary history of protein families is the major strategy employed to identify and, where possible, to order events during eukaryogenesis. The mitochondrion and the chloroplast are derived from endosymbiotic events and hence represent a distinctive evolutionary path [2,7]. The remaining structures/organelles in eukaryotic cells, including the nucleus, are considered to be endogenously-derived, i.e. arising through evolution and expansion in the gene complement of the proto-eukaryote; hence, the proteins defining various organelles have clear vertical descent from an archaeal ancestor. Reconstruction of the evolution of endogenously-derived organelles thus relies on the identification and analysis of marker genes/proteins that define specific organelles and their sub-structures.

Several distinct modes underpinning the evolution of compartments and organellar diversity have been proposed. These are based largely on analysis of the membrane trafficking system, which features large paralogous protein families with organelle-specific members such as GTPase and SNARE families, kinesins and others [8–10]. Some of these modes can be considered to be 'expansive', in the sense that they lead to an increase in the number of compartments. This is exemplified by Rab GTPases that specify and mediate cargo exchange between endocytic and exocytic compartments [8]. Alternatively, there are many examples of reductive processes, where lineages lose complexity; again, the Rab family is a good example – in many unicellular organisms, for example in *Saccharomyces cerevisiae*, a great many Rab genes have been lost. We termed this process 'sculpting', as the new organism is revealed only through the removal of genes. These two examples are comparatively simple to understand in evolutionary terms as they rely on changes to the numbers of paralogs within a gene family [11]; but additional mechanisms are also at play and are not always obvious in terms of simple gene counts. These include 'churning' whereby paralogs are created and deleted so the overall number remains

similar [12]; and finally 'backfilling' whereby loss of functionality through gene loss is compensated through the expansion of a different gene family [13] (Figure 1). In principle, the relevance of such modes of evolution extends to all endomembranes, including their sub-domains and functions, at all stages of eukaryotic evolution.

The nucleus is an endomembrane compartment with deep evolutionary connections to other endomembranes. In particular, nuclear pore complex (NPC) proteins and vesicle coat proteins mediating endocytosis, post-Golgi transport, ER-exit and flagellar assembly all have a common origin (reviewed in [14]). Despite significant variation in composition, the secondary structure of many NPC and coat proteins is well conserved [14], compelling evidence that the endomembrane system, and hence the great majority of the internal structure of the eukaryotic cell was well developed prior to the radiation of eukaryotes from the LECA (Figure 2). However, the details of several relevant events leading up to the LECA remain unresolved, including the relative sequential order of the origin of the nucleus and of the mitochondria [15,16]. For example, we have argued that the NPC, in its current configuration, likely emerged late in eukaryogenesis, although
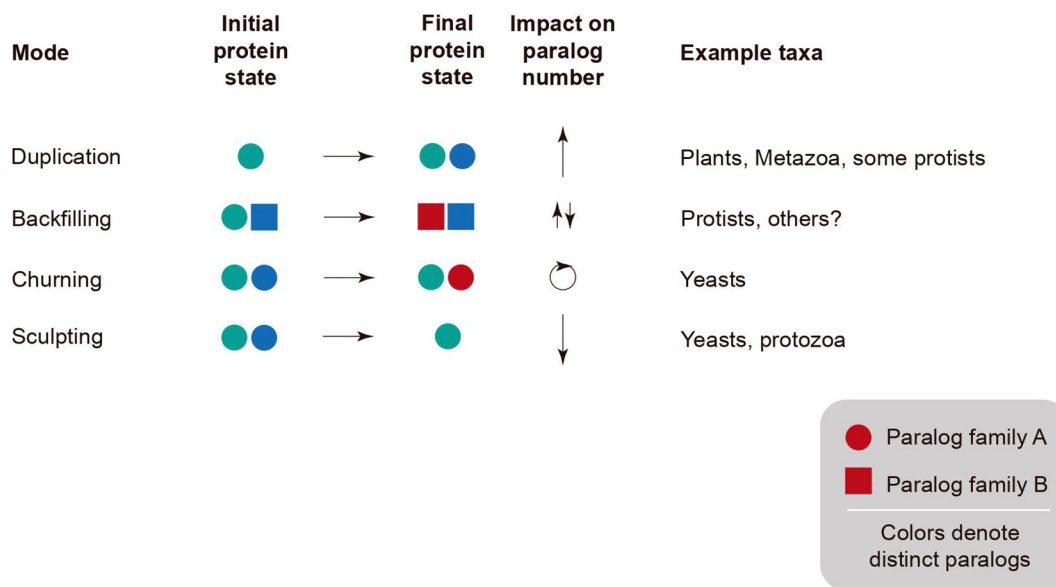
Figure 1. Modes accompanying organelle evolution. Multiple modes have been suggested for the evolution of paralogous genes involved in specifying organellar identity (see text). Gene duplication increases the number of paralogs, while sculpting decreases through gene loss. Churning can operate without an obvious or significant change to paralog number but represents the birth and death of paralogous, functionally equivalent genes while backfilling is the result of both gene loss and compensatory expansion of an unrelated gene family to suppress the impact of the initial loss.
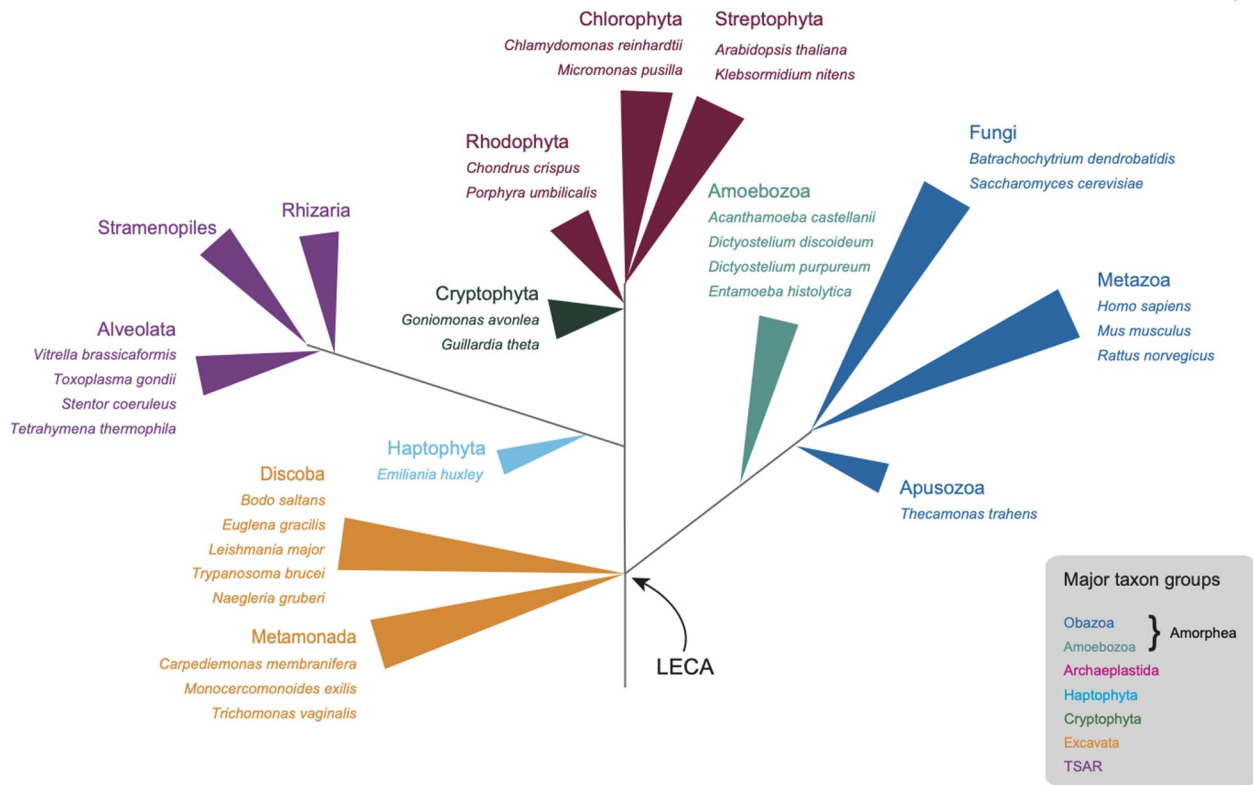
Figure 2. Eukaryotic lineages sampled and their taxonomic relationship. A schematic phylogenetic tree depicting the relationships between the major eukaryotic taxa. large groupings are color-coded as indicated in the legend at lower right and selected species included in the study are shown. LECA indicates the position of the last eukaryotic common ancestor. the tree structure is loosely based on the recent topology described in Burki et al., 2020.

this does not preclude an earlier origin for the nuclear envelope in some form [17].

The nucleus facilitates separation of transcription and mRNA processing from translation and in modern eukaryotes a considerable machinery is devoted toward securing accurate mRNA processing prior to export from the nucleus [18]. Much of the apparatus is conserved, indicating ancient origins and a central requirement to minimize accumulation of aberrant proteins. Separation of translation and transcription enabled evolution of mRNA splicing, allowing greater functional diversity to be extracted from essentially the same coding sequences and providing additional mechanisms for controlling gene expression. Further complexity of gene regulation then arises from the development of sophisticated genome packaging that allows formation of eu- and heterochromatin and, further, to chromosomal territories and complex spatial genome organization. Some nuclear functions are potentially very ancient and are likely repurposed components from preexisting systems, for example, the ESCRT system

involved in NE repair likely originated as a cytokinesis system in Archaea [19,20].

Despite these examples, variations in mechanisms of nuclear processes between organisms are well known. For example, nuclear envelope morphology during mitosis is usually either distinctly open or closed but can be semi-closed in some species such as *Schizosaccharomyces japonicus* [21]. These variations necessitate distinct mechanisms for segregation of nuclear components during division and reassembly. Open mitosis likely arose on multiple occasions, while closed mitosis is probably the ancestral as evidenced by the broader distribution of this mode [22]. Further, the mechanisms by which chromosomes are segregated are highly variable. While mitosis nearly exclusively requires a microtubule-based spindle, the location of the spindle anchor varies considerably, being at turns within the nucleus, embedded in the NE or in the cytoplasm [23]. Nuclear positioning and mechanotransduction, which in metazoa is mediated by the LINC

complex, is also likely divergent; in vascular plants, there is a range of additional systems mediating cytoskeletal connections, while some organisms lack the LINC complex altogether [24,25]. Moreover, the nuclear lamina, a fibrous network residing on the nucleoplasmic surface of the NE, is represented by several apparently dis-elected for investigation. Givenstinct systems; NMCP/CRWN proteins of plants, lamins in many lineages including metazoa and NUP-1/2 of the kinetoplastida [25–29]. So far, mammalian lamin orthologs have not been detected in yeasts, either experimentally [30] or *in silico* [25]. However, in spite of lacking canonical lamins, yeast NPC components, nucleoporins and NE-associated proteins provide a platform for anchoring chromatin to the NE and for events related to gene expression regulation, transcription, mitosis and NE stability [31–33]. Lamins are likely the ancestral lamina proteins while the more restricted plant and kinetoplastid systems appear to be lineage specific and likely later emerging [25]. Finally, the kinetochore in kinetoplastids is composed of a distinct set of proteins differentiating these organisms from all other eukaryotes [34,35].

To extend our understanding of the dynamic process of nuclear evolution beyond the NPC and lamina to the whole of the NE, we examined a set of NE and NE-associated proteins, that we collectively term 'NEA', from metazoan and kinetoplastid datasets together with published work (Figure 2). We analyzed conservation of NEA proteins across eukaryotes, finding only a small contingent to be broadly conserved with the majority of the cohort being remarkably divergent; frequently, identifiable orthologs can only be found in closely related organisms.

Methods

Genome databases. Eukaryotic species were selected for based on factors including importance as experimental organisms, phylogenetic diversity and sequence/assembly quality. This included recently published sequence data from the protozoan alga *Euglena gracilis* [36] and the freshwater sponge *Ephydatia muelleri* [37]. Details regarding all genomic and transcriptomic data sources are

provided in Table S1. Although some datasets are from genome sequencing studies and some from transcriptomic studies, in each case only predicted peptide sequences were analyzed herein. In the case of *Carpediemonas membranifera* [38] and *Hemimastix kukwesjijk* [39], nucleotide transcript sequences were translated to predicted peptide sequences using GeneMarkS-T [40].

Selection of genes for analysis. Inclusive sets of NE and NE-associated (collectively NEA) proteins were selected for investigation. Given issues concerning with purity of subcellular fractions, we sought to minimize likely contaminants within NE proteome databases and no selection based on function was made. First Seventy-five relatively high-confidence rat liver NEA protein-encoding genes were selected from a much larger cohort of potential NEA proteins identified previously [41,42]. These were not limited to nuclear envelope transmembrane (NET) proteins and both inner and outer nuclear envelope proteins were included; importantly many have been identified at the NE of mammalian cells in subsequent studies and we excluded proteins subsequently demonstrated to have no NE-association [43–45]. Second, eighty-nine *Trypanosoma brucei* NE protein-coding genes were selected based on previous studies [46,47]. In this case, proteins containing at least one predicted transmembrane domain were included. TMHMM version 2.0 c transmembrane domain predictions from TriTrypDB were used for this purpose (accessed April 2020) [48,49]. We chose this selection criterion to maximize the likelihood that the included proteins were indeed associated stably with the NE and not a peripheral interactor.

Comparative genomics. Sequence analyses were performed on the University of Dundee High Performance Computing Cluster. Predicted peptide sequences of selected genes encoding nuclear envelope proteins were used as queries for searching across eukaryotes using the Basic Local Alignment Search Tool for Proteins (BLASTp, version 2.9.0+) [50] with an E-value <0.1. The top three hits per query per organism were extracted and further filtered by an alignment coverage of greater than 10% of the query sequence. Redundant, 100% identical sequences were not included in the hit count. Reverse BLASTp of hit

sequences against the original query organism was performed and orthology predicted if the original query was within the top three hits, had an alignment coverage of greater than 10% and an E-value <0.1.

Alignment and phylogenetic analysis: For each query, sequences meeting the search criteria described above were aligned using MUSCLE [51] version 3.8.1551 with default settings. Alignments were trimmed using alncut version 1.06 [52] with gaps allowed in 25% of sequences per residue. Trimmed alignments were analyzed using FastTree version 2.1.10 [53] using default settings. In general, branch supports ≥0.95 were considered significant support for bipartitions.

Validation and binning: All data were manually inspected and met the following additional criteria beyond being included as a BLAST hit: (i) the length of the predicted protein is consistent with the query protein, (ii) the alignment demonstrates extensive regions of homology, and (iii) based on phylogeny, the putative ortholog is included in a monophyletic clade with the query and other orthologs (when paralogs were present). In many cases, the inclusive search criteria retrieved paralogs, which were excluded at this phylogenetic analysis step. Trees were viewed on Apple computers (macOS 10.15.7 operating system) using FigTree version 1.4.4 (http://tree.bio.ed.ac.uk/soft ware/figtree) or the Environment for Tree Exploration (ETE3) version 3.1.2 [54], and alignments were viewed using AliView version 1.26 (http://ormbunkar.se/aliview) [55] or JalView version 2.11.0 [56]. In cases of ambiguity, bespoke BLAST searches at NCBI were used to further validate a conclusion of found/not found. Sequence IDs for all identified orthologs are provided in Table S2.

Search results were binned according to inferred timing of evolutionary origin: Either originating in the LECA (Group A), in an amorphean ancestor of Metazoa (Group B), or much later within either Metazoa or Euglenozoa (Group C). Results in Group A included both those with orthologs nearly universally conserved among eukaryotes and those with orthologs not identified in several major taxonomic groups. Data from comparative genomics searches were converted to the Coulson plot format using CPG version 1.6.1 (https://github.com/mfield34/CPG [57]) and finalized using Illustrator (https://www.adobe.com).

Gene ontology: Gene ontology annotation for Biological Processes for mammals was obtained from Uniprot (https://www.uniprot.org) [58] and/or Quickgo-EBI (https://www.ebi.ac.uk/QuickGO/) [59] using accession IDs. Gene Ontology annotations were categorized and frequency was represented in bar charts using Microsoft Excel and finalized using Adobe Illustrator. Gene ontology annotation for trypanosome proteins were retrieved from TryTripDB release 48 beta ([49], https://tritrypdb.org/tritrypdb/app).

Results and discussion

Documenting nuclear envelope evolution and understanding eukaryogenesis. In the earliest stages of eukaryotic evolution, the nucleus may have served as a simple membranous structure enclosing the chromatin. The double lipid bilayer and composition of the NE are direct consequences of the NE being an extension of the endoplasmic reticulum (ER). With several nuclear and NE proteomes published [41,60–62], data suggest retention of functional analogs, *albeit* with divergent sequence and structure, but most information is restricted to animals and fungi. Further, protein domain shuffling [63,64] is also apparent among the NEA proteins, for example, the lamin B receptor is conserved but only due to the presence of a C_{14}-sterol reductase domain, indicating repurposing but not a pan-eukaryotic presence for an LBR [25]. Hence, more systematic analysis is warranted.

We explored NEA protein conservation by conducting searches for orthologs of 164 putatively NEA genes, using reciprocal-best-hit searching with BLAST. This cohort included 75 genes originally identified in mammals [41–45] and 89 proteins identified in *Trypanosoma brucei* [47]. The data returned from this procedure (supplementary data archive) were manually validated. All calls are conservative, with only orthologs detected with confidence deemed 'found'. Our search included members of the Amorphea (comprising organisms belonging to Opisthokonta, Amoebozoa and Apusomonada supergroups), Archaeplastida (including a diversity of plants and members of

Chlorophyta and Rhodophyta), Cryptista (represented by members of Cryptophyta), Excavata (flagellated organisms of biological and/or medical importance such as *T. brucei* and *Leishmania major*), CZAR (which encompasses several major groups of algae, protozoa and seaweeds), Haptista (comprising haptophyte algae, represented in this work by the marine microalgae *Emiliania huxleyi*) and Hemimastigophora (an early-branching lineage of free living protozoa with two rows of flagella, represented here by *Hemimastix kukwesjijk*) [39,65]. All genomes were selected for quality in terms of predicted proteome coverage completeness (see Methods for data sources and Figure 2 for taxonomic context).
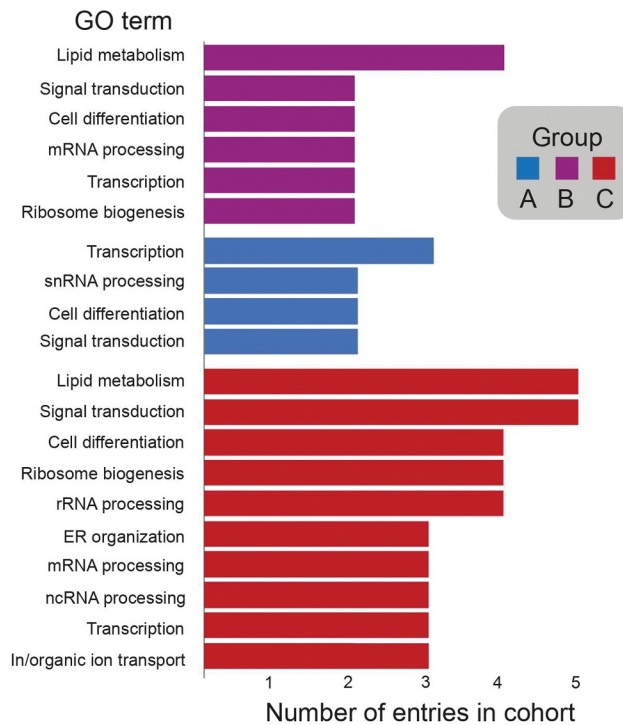
Lineage-specific proteins dominate metazoan NEA protein cohorts. Firstly, a cohort of 50 NEA proteins (Figure 3) identified initially in rodents [41] and subsequently in several human cell lines [42–45] were used as queries. We distinguished three groups (see Methods). Group A (13 proteins) are highly conserved across eukaryotes indicating an origin predating the LECA, and thus likely to support universally required functions. Gene ontogeny (GO) analyses for Biological Process revealed that these proteins are mostly involved in lipid metabolism (Figure 4), consistent with this interpretation. Group B (eight proteins) includes proteins present only in amorphean groups, i.e., Metazoa (including mammals), Fungi, Apusomonada, and Amoebozoa. GO identified these proteins as mainly related to transcription. Many of Group B may have been lost due to saprophytic or parasitic lifestyles in which the host is able to provide these requirements, as indicated on the specific losses within the Amorphea that collectively exemplify sculpting. Group C (29 proteins) comprises proteins restricted to Metazoa, suggesting an origin in a recent ancestor. NE and membrane (NE and/or ER) associated proteins in this group partake in a variety of general functions and we observe an enrichment in proteins involved in the lipid metabolism, RNA processing (including ribosomal, messenger RNAs and noncoding RNAs such as t-, sn- and piwi-RNAs) and ion transport. Additionally, several proteins stand out as being involved in complex processes characteristic to metazoans: in innate immune response, such as Tmem173/NET23/STING

[66,67], and tissue differentiation and development, such as NET37 in skeletal muscle development [68], Mospd3 in heart development [69], Nepro in neurogenesis [70]; and multiple other proteins with tissue specific functions at the nuclear envelope that appear critical for healthy tissue maintenance (Wolframin, Tmem201, Emerin and Nesprin/SYNE proteins, see below). This participation of the NEA proteins in cell/tissue differentiation is commonly accompanied by functions in signal transductions, highlighting the importance of this set of proteins in triggering signaling pathways such as MTOR, Wnt, MDA-5, Notch, kinase cascades, etc. Moreover, transcriptional regulation is a common function in the cohort, influencing regulation of RNA polymerases (I and II) such as Noc2l which acts as a transcription corepressor [71], Wdr43 which is needed for activation of promoters and favors transcription elongation [72], Int1 [73], Zmiz1/2 [74] and Rprd1b/1a which are transcriptional regulators of RNA polymerase-II transcription [75,76].

Importantly, targeting to the NE of many of these proteins has been verified experimentally as, in spite of having at least one TMD, some of these proteins do not localize to the NE but rather to the nucleoplasm or nucleolus [77]. Interestingly, some NEA proteins can be detected by microscopy at the nuclear periphery in some cell types/tissues but not in others, which has led to the concept that different tissues contain distinct pools of protein complexes that alter the localization of NEA proteins [77], making prediction of NE localization even more challenging. Although most of the proteins in our cohort reside in the NE, some of them have shown nucleoplasm localization, such as UTP15, Noc4l, Noc2l, Wdr33, Wdr75, Tmem74 [43,77]. Interestingly, some NEA proteins have been localized in different regions of the nucleus, e.g. lamin A, which typically localizes to the nuclear rim but in recent years a highly mobile nucleoplasmic pool of lamin A has been detected [78–80], although the mechanisms regulating this state of lamina-independent lamins remain obscure. Further examples are Tmem214 and Tmem70: Tmem214 is simultaneously identified as a protein [81], an ER transmembrane protein [82] and further confirmed to localize to the inner

Figure 3. Distribution of nuclear envelope proteins from metazoa across eukaryotes. Coulson plot demonstrating presence or absence of NE proteins across the eukaryotes. Filled circles indicate proteins identified, open circles indicate proteins for which orthologs were not found. Rows are predicted proteins and columns are organisms. Supergroups are colourised using the same system as in Figure 2. Three groups are recognized: Group A; highly conserved across Eukaryotic supergroups, Group B; originated in an amorphean ancestor of Metazoa.

Figure 4. Frequency of gene ontology annotation for nuclear envelope proteins from the metazoan cohort. bar chart showing frequency of GO annotation of Biological process of highly conserved proteins across supergroups (Group A), proteins present across Amorphea (Group B) and proteins restricted to Metazoa (Group C) from .Figure 3

nuclear envelope [83]; Tmem70 has multiple isoforms and is traditionally considered an inner mitochondrial membrane protein partaking in the assembly of the ATP synthase [84,85], but has also been identified as a NET [83].
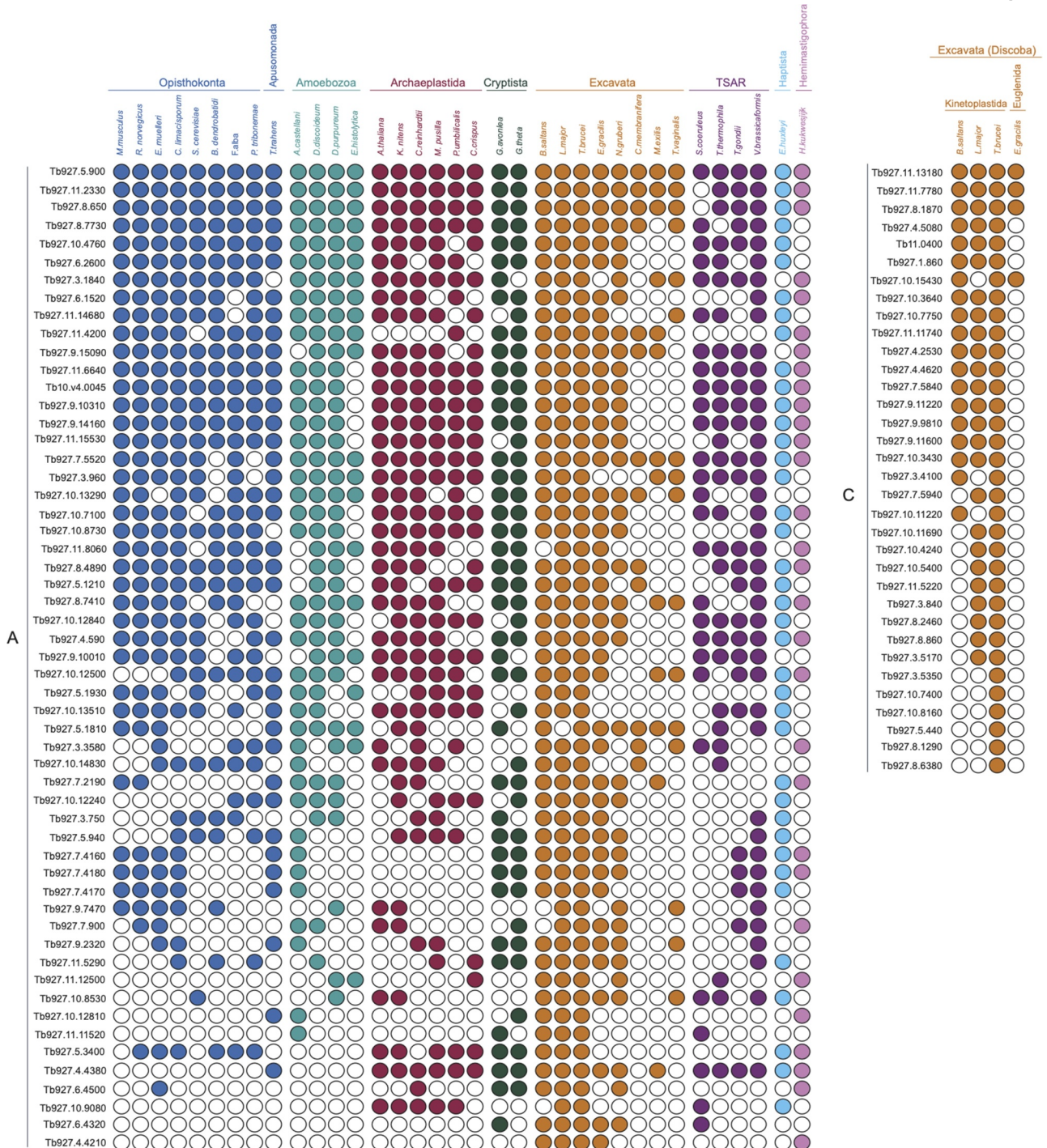
Thus, association of proteins with the NE seems to be complex to predict merely by bioinformatic tools, as their localization is influenced not only by structure but also by interacting partners and differential expression in different cell types (see below) and the orthologs of mammalian NEA proteins detected here will require experimental validation of NE association in other organisms.

Lineage-specific NEA proteins are widespread. Metazoan sequences cannot be used to identify non-orthologous lineage-specific genes in other taxa, and also have inherent bias toward metazoa-specific genes [86]. Hence, secondly, we searched using queries from the protozoan *T. brucei* [47] (Figure 5). Overall, trypanosome Group A (55 proteins) appears to have core biological functions, with GO annotation (Biological Process) indicating involvement in inorganic and organic ion transport, protein glycosylation and lipid metabolic process (Table S3). Orthologs of

some trypanosome Group A proteins were readily identified in all or nearly all other eukaryotes. For example, Tb927.5.900 is an oligosaccharyltransferase subunit, with orthologs retrieved in all genomes queried in our analysis. In other eukaryotes, these are essential for transfer of N-glycans [87] and localize at least in part to the NE [88]. Some Group A proteins, however, were less conserved, for example, the transmembrane protein Tb927.10.12810 which contains a choline phosphate cytidylyltransferase domain. In most eukaryotes, choline phosphate cytidylyltransferases localize to the NE and regulate phosphatidylcholine levels [89]. However, outwith kinetoplastids, orthologs of Tb927.10.12810 were only identified in organisms representing the unicellular taxa Apusomonada, Amoebozoa, Cryptophyta, and Hemimastigophora (Figure 5).

Group C (34 proteins) includes several proteins associated with lipid biosynthetic processes (e.g., sphingolipid, fatty acid biosynthesis) and proteins targeting to the ER; however, majority of proteins in this group are annotated as hypothetical or putative gene products. Several proteins are trypanosome-specific, represented by *T. brucei* in our analysis, including

Figure 5. Distribution of trypanosomatid nuclear envelope proteins across eukaryotes. Coulson plot demonstrating presence or absence of trypanosome NE orthologs across the eukaryotes. filled circles indicate proteins identified, open circles indicate proteins for which orthologs were not found. rows are predicted proteins and columns are organisms. supergroups are colourised using the same system as Figure 2. Two groups are recognized: Group A; scattered distribution across taxa and Group C; restricted to kinetoplastids. TriTrypDB [49] accession IDs are shown on the left.

a putative *trans*-sialidase (Tb927.5.440), a SUMO-interacting motif containing protein (Tb927.8.1290) and three hypothetical proteins (Tb927.3.5350, Tb927.8.6380 and Tb927.10.8160 with experimental evidence of localization to the NE and ER [49,62,90].

Overall, the frequency of Euglenozoa-specific proteins is comparable to the results of searches using metazoan NEA protein queries.

Moreover, with a single exception, metazoan queries (Figure 3) did not retrieve any proteins from

our trypanosoma query cohort (Figure 6) but instead others that may be important as NEA proteins, namely a beta propeller protein Tb927.8.1980, a Noc2p family protein Tb927.10.12430, CBF/Mak21 family protein Tb11.v5.0274/Tb927.10.6320/Tb927.11.2120/Tb927.4.3670. The exception to this lineage-specificity is Tm7sf2 (E9Q4M8), which appears to be an ortholog of *T. brucei* Tb927.11.15530 (Figures 3 and 6). *Frequency of lineage-specific proteins in fungi and plants*. To expand our coverage, we also examined published analysis of NE proteomes in both yeast and plants. An earlier systematic study focused on forty-five budding yeast proteins selected for NE localization from a high

throughput dataset [91]. In this case ~50% of the cohort were suggested as pan-eukaryotic with a further ~20% broadly distributed. The conserved proteins were in the main involved in core nuclear functions, including chromatin organization, with only 20% predicted as unique to Saccharomycetes. A considerable number of homologs could only be identified through a shared domain, indicating possible domain swapping as noted above for the LBR. This pattern may in part be a reflection of the extensive gene losses in *S. cerevisiae*, such that lineage-specific proteins are depleted. A high incidence of domain sharing amongst NEA proteins from different lineages, but within otherwise highly divergent
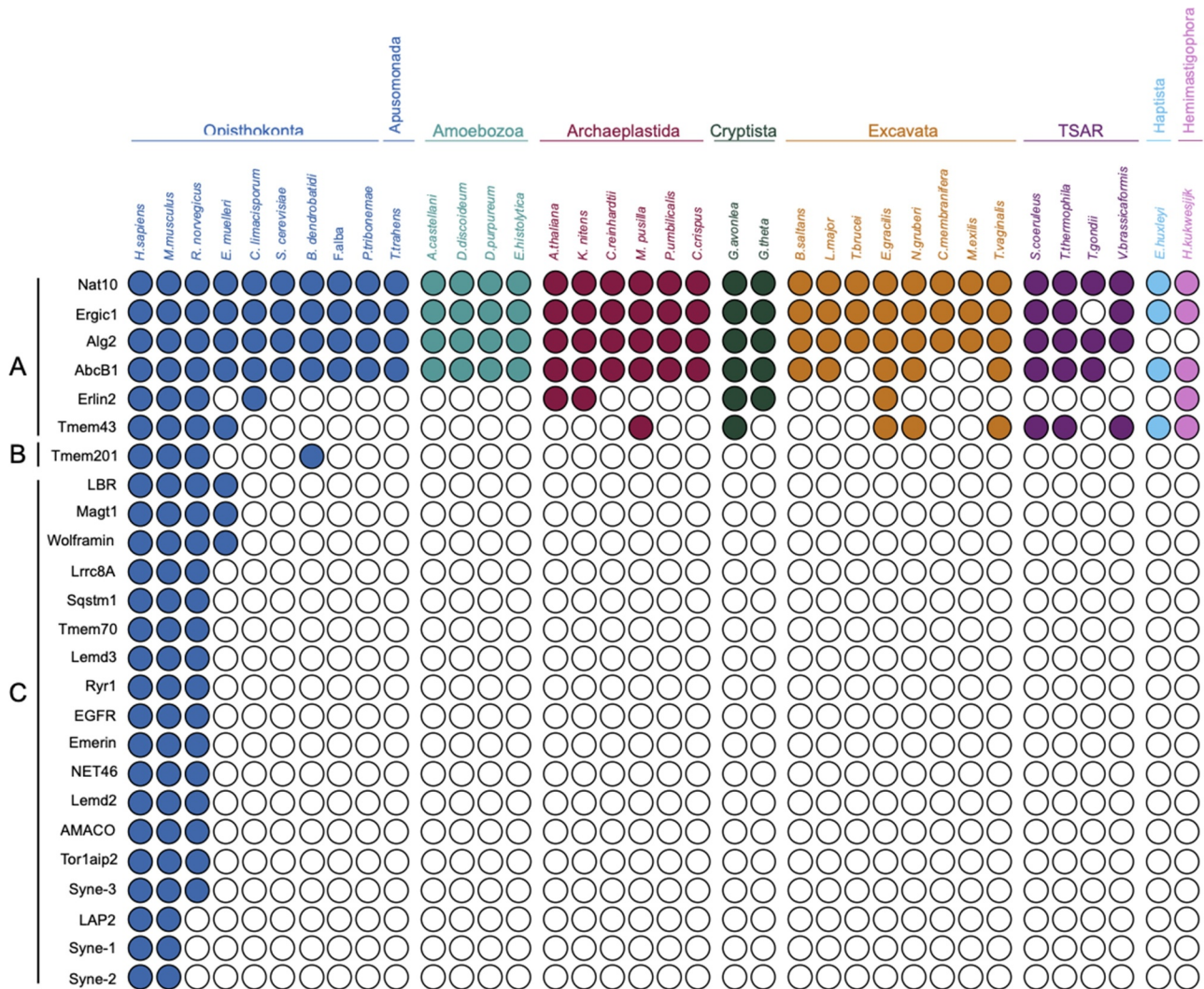
Figure 6. Distribution of disease-associated nuclear envelope proteins across eukaryotes. Coulson plot demonstrating presence or absence of NETs directly linked to human disease. Supergroups are colourised using the same system as Figure 2, Layout as in Figure 3. Three groups are recognized: Group A; highly conserved across Eukaryotic supergroups, Group B; originated in an amorphean ancestor of Metazoa, Group C; restricted to Metazoa. See text for explanation as to why LBR is excluded from Group A.

proteins, is also supported by evidence from the vascular plant *Arabidopsis thaliana*. The *A. thaliana* LINC complex for example is canonical, containing both SUN and KASH-domain proteins. However, the nesprins (see below) of metazoan cells, which also contain KASH domains and associate with the core LINC complex, are replaced by analogs that only retain a KASH domain but no other obvious homology [24,92]. Given that the KASH domain is quite short (less than 60 amino acids and in plants frequently even shorter) how this system has evolved is intriguing. SINE proteins are present across a large number of vascular plants and are components of the analogous LINC complex and likely support a link between the plant-specific lamina NMCP/CRWN proteins and the cytoskeleton [93]. There is considerable promiscuity here as all SINE proteins interact with SUN1 and SUN2; this may be related to the manner in which this family has expanded, with clear ancestral as well as lineage-restricted members that presumably evolved later. This may be, at least in part, due to tissue-specific expression [93].

NEA disease genes are dominated by lineage-specific members. The relevance of NEA proteins to a range of diseases is well documented [94–96]. To uncover whether disease-associated NEA proteins are involved in metazoan-specific or general eukaryotic functions, we examined the distributions of 25 NEA proteins previously linked to pathology (Figure 6) and encompass a broad range of conditions including myopathies, γ-globulinemia, inflammatory diseases, muscular/skeletal abnormalities, cardiomyopathies, neurological/mental conditions, glycosylation disorders and others ([42] and references in Table 1). The ability to identify an ortholog in another organism could, for example, be exploited in determining function. Nat10, Ergic1, Alg2 and Abcb1 are present broadly in all supergroups explored (Figure 6, Group A). These proteins relate to near universal functions such as maturation of SSU-rRNA, vesicle mediated transport and ATP transport, respectively. However, although related to disease, mutations in these genes do not necessarily result in detrimental phenotypes in other eukaryotes. This is reflection of context dependence, and we have noted this previously regarding some essential genes in metazoa which are absent from multiple unicellular taxa [86]. We found Alg2, Erlin2 (ER lipid raft-associated 2) and Tmem43 retained in several taxa. Mutations in

Tmem43, a four transmembrane domain protein known as LUMA in *Homo sapiens*, are linked to cardiac conduction defects [97] and Emery-Dreifuss muscular dystrophy [98,99]. Moreover, Tmem43/ LUMA upregulation is related to poorly fertile spermatozoa [100] and Tmem43/LUMA has close interactions with lamins and emerin and is involved in structural organization of the NE [99,101]. Erlin2 is a transmembrane domain protein involved in lipid homeostasis [102] and leads to motor and cognition disabilities when mutated [103]. Moreover, abnormal Erlin2 levels are found in human breast cancer [104]. Alg2, a mannosyltransferase, is involved in N-glycosylation and associated with glycosylation disorders (Table 1 and references therein). One protein seems to have originated earlier in Amorphea, Tmem201 (Figure 6, Group B), which contributes to the architecture of the NE and interacts with the LINC complex through UN1 and lamin A [105]. Mutations in this protein have been recently implicated to cause one of the variants of Emery-Dreyfuss muscular dystrophy [106].Interestingly, the eighteen remaining disease-related proteins are apparently metazoa-specific, (Figure 6, Group C, Table 1). Lamin B receptor (LBR), one of the better characterized NEA proteins residing in the inner nuclear membrane, is an important interactor of lamin B [107]. With lamina-constituting proteins being divergent in different systems [25], lamin B receptor may also constitute a case of divergence across supergroups despite its C_{14}-sterol reductase domain having a pan-eukaryotic distribution and influencing metazoan cell viability and embryogenesis in plants [108–110]. Although several proteins are associated with specific diseases, such as Nesprin 1 and 2, emerin and lamin A-associated polypeptide 2 (LAP2) (Emery-Dreifuss dystrophy and cardiomyopathy) [98,111]; others have broad pathology, such as Sqstm1 (Sequestosome 1), associated with bone, muscle and heart defects, neurodegenerative conditions and leukemia, or EGFR (Epidermal growth factor receptor) related to cancer and bowel disease (Table 1). Nesprins (Syne-1, 2 and 3) localize at the NE and are scaffolding proteins that partake in the LINC complex [112,113]. Besides nuclear structure, emerin also has roles in transcription, signaling and chromatin regulation through interactions with partner proteins [114]. It is interesting that the distributions of emerin and Nesprin-1/2/3 are substantially narrower than that of the LINC

Table 1. Diseases associated with nuclear envelope-associated genes restricted to Metazoa.

Gene	Protein	Associated Disease	Reference
Ergic1	Endoplasmic reticulum-Golgi intermediate compartment protein 1	Arthrogryposis multiplex congenita (AMC, neurophatic type), gastric cancer	[119,120]
Nat10	N-acetyltransferase 10	Hepatocellular carcinoma HGPS (Hutchinson-Gilford progeria syndrome), Muscular dystrophy	[121–124]
Alg2	α-1,3/1,6-mannosyltransferase	Congenital disorder of glycosylation (type Ii), Congenital myasthenic syndrome	[125–127]
AbcB1	ATP-binding cassette subfamily B, member 1	Inflammatory bowel disease 13, Encephalopathy	[128,129]
Tmem43	Transmembrane protein 43	Ventricular displasia 5 (ARVC5), Emery-Dreifuss muscular dystrophy 7	[99, 130–132]
Erlin2	ER lipid raft associated 2	Spastic paraplegia 18, Motor system conditions	[103,133]
Lbr	Lamin B receptor	Greenberg skeletal displasia, Pelger-Huet anomaly, Reynolds síndrome	[134,135]
Magt1	Magnesium transporter 1	Immunodeficiency with magnesium defect, Epstein-Barr virus infection and neoplasia (XMEN), Mental retardation (X-linked), Congenital disorder of glycosylation 1CC (CDG1CC)	[136,137]
Wfs1	Wolframin	Wolfram syndrome (diabetes mellitus, optic atrophy, deafness)	[138,139]
Tmem201	Transmembrane protein 201	Emery-Dreifuss muscular dystrophy	[106]
Lrrc8A	Leucine rich repeat containing 8 family, member A	Agammaglobulinemia, Glaucoma	[140,141]
Sqstm1	Sequestosome 1	Paget disease (bone turnover), Dementia, Sclerosis, Neurodegeneration, Myopathy, Alzheimer's disease, Lymphoblastic leukemia	[142–146]
Tmem70	Transmembrane protein 70	Mitochondrial disease (ATP synthase deficiency)	[147]
Lemd3	LEM domain containing 3	Buschke-Ollendorff syndrome (skin and bone disorders)	[148]
Ryr1	Ryanodine receptor 1	Malignant hyperthermia, Myopathy, Neuromuscular disorder	[149–152]
EGFR	Epidermal growth factor receptor	Lung cancer, Neonatal skin/bowel disease	[153,154]
Emd	Emerin	Emery-Dreifuss muscular dystrophy	[98,155]
NET46	Steroid transmembrane transporter	Connotruncal heart defects (CTDs) in newborns	[156]
Lemd2	LEM domain-containing protein 2	Juvenile-onset cataracts in Hutterties population	[157]
AMACO	A domain-containing protein similar to matrilin and collagen	Colon cancer	[158,159]
Tor1aip2	Torsin A-interacting protein 2	Dystonia	[160]
LAP2	Lamin A-associated polypeptide 2, Thymopoietin	Cardiomyopathy	[161]
Syne-1	Nesprin 1 (Nuclear envelope spectrine repeat protein 1)	Spinocerebellar ataxia, Emery-Dreifuss muscular dystrophy, Arthrogryposis	[98, 112, 162, 163]
Syne-2	Nesprin 2 (Nuclear envelope spectrine repeat protein 2)	Emery-Dreifuss muscular dystrophy	[98,112]
Syne-3	Nesprin 3 (Nuclear envelope spectrine repeat protein 3)	Emery-Dreifuss muscular dystrophy	[106]

complex – defined as SUN/KASH-domain protein complex crossing the nuclear lumen. As remarked above, this is likely due to replacement with functional analogs. *Tissue-specificity of nuclear envelope proteins.* Hundreds of mutations are known in genes encoding NEA proteins, including in lamins, nuclear pore components and NETs (Table 1). Although most of these proteins are ubiquitously expressed, the mutations are associated with a wide number of diverse and apparently unrelated tissue-specific diseases. Further, in many cases a single protein can harbor several mutations each causing a different disease, e.g., Lbr where mutations can cause bone disease or blood disorders (Table 1 and references therein). One prevalent hypothesis for how this occurs is that specific pathology develops from disruption of a tissue-specific function, for example, interaction with a tissue-specific partner. In Table 2, we summarize

Table 2. Differential expression in tissues of envelope-associated genes from Metazoan cohort.

Proteins from Metazoan cohort (Figure 3)			
Group in Coulson plot	Protein	Expression in tissues[1]	Reference
A	NET45, Mcat, Wdr33, NET39, Slc25a22	Enriched in liver and blood.	[43]
	NET11	Enriched in muscle and blood (B-cells).	[43]
B	Scai	Widely expressed. Higher levels in brain, spleen and thymus.	[164]
	Rprd1b	Widely expressed. Higher levels in liver, colon, prostate, uterus. Lowest levels in heart and kidney. Not detected in rectum.	[165]
	Tmem120a	Widely expressed. Higher expression in liver, heart, kidneys, colon, nociceptors and adipose tissue.	[166,167]
C	NET56	Predominantly muscle specific, with lower expression in other tissues.	[168]
	NET37	Brain, liver, spleen, skeletal muscle, heart, lung and kidney. High protein levels in astrocytes and skeletal muscle.	[68]
	Nicalin	Enriched in liver and blood.	[43]
	KDM3B	Ubiquitous, highly expressed in placenta, skeletal muscle, kidney, heart and liver.	[169]
	Wdr43	Low levels in heart, liver, lung, spleen, thymus and hippocampus.	[170]
	Lpgat1	Highly expressed in liver and placenta. Expressed in peripheral blood, lung, kidney and brain. Lower levels in colon.	[171]
	Tmem173 (Sting1)	Highly expressed in mature B cells. Present in spleen, thymus. High protein levels in dendritic cells.	[172,43]
	Torp1aip2	Widely expressed. Higher levels in non-neural cells, hippocampus and spinal cord, liver and B cells. High levels of protein in liver.	[173,43]
	Tm7sf2, Scara5, Tmem74, Tmem53	Enriched in liver.	[43]
	Tmem173, Dhrs7	Enriched in blood.	[43]
	Mospd3	Enriched in liver and muscle.	[43]
Proteins from Diseases cohort (Figure 6)			
Group in Coulson plot	Protein	Expression in tissues[1]	Reference
A	Alg2	Enriched in liver and blood.	[43]
	ABCB1	Expressed in liver, kidney, small intestine, blood and brain.	[58, 43]
	TMEM43	Highest levels in placenta. Lower levels in heart, ovary, spleen, small intestine, thymus, prostate and testis.	[101]
B	Tmem201	Widely expressed.	[43]
C	LBR	Expressed in the bone marrow, liver, heart, adrenal gland, lung, placenta and uterus. Expressed in osteoclasts and osteoblast-like cells.	[174]
	MAGT1	Ubiquitous. Low levels in brain, lung and kidney.	[175]
	Wolframin	High levels in heart, followed by brain, placenta, lung and pancreas. Low levels in liver, kidney and skeletal muscle.	[58]
	LRRC8A	High levels in bone marrow. Also found in brain, kidney, ovary, lung, liver, heart, blood cells, T-cells and B-cells.	[140]
	SQSTM1, EGFR, Lemd3, Lemd2	Ubiquitous.	[176, 177, 45]
	RYR1	Found in skeletal muscle and brain.	[178]
	Tmem70	Enriched in blood and muscle.	[43]
	Emerin	Skeletal muscle, heart, colon, testis, ovary and pancreas.	[58]
	NET46	Highly expressed in kidney, enriched in liver.	[179, 43]
	Amaco	Detected in uterus, kidney, skin, lung, intestine.	[180]
	LAP2	Ubiquitous. Abundant in adult thymus and fetal liver.	[58]
	Syne3	Ubiquitous.	[44]
	Syne1	Ubiquitous. Highly expressed in skeletal and smooth muscles, heart, spleen, leukocytes, pancreas, cerebellum, stomach, kidney and placenta.	[181, 182]
	Syne2	Ubiquitous. Higher levels in kidney, adult and fetal liver, stomach, placenta. Low levels in skeletal muscle and brain. Isoform 5 highly expressed in the pancreas, skeletal muscle and heart.	[181]

[a]Refers to levels of mRNA otherwise stated.

recent findings on the differential expression of proteins present in the cohorts in Figures 3 and 6.NE proteins can influence chromatin topology and gene expression, for example, Tmem201 (NET5), Tmem120A (NET29), NET45, NET39 and Tm7SF2 (NET47), which promote peripheral positioning of chromosomes [115–117]. Different NETs demonstrate differential impact on chromosomal positioning [116] and some, e.g. Sting1/NET23, alter chromatin through generating changes in the chromatin compaction [67]. Moreover, NEA proteins also influence transcription, e.g. Tm7SF2 during hepatic differentiation, or act as repressors as in the case for NET39, Tmem38A and WFS1 in muscle cells [115]. These activities of gene repositioning are critical for differentiation, e.g. Tmem120A in adipocytes, Tm7SF2 in hepatic cells and NET39, Tmem38A and WFS1 in myogenesis [67,115,117]. Tmem201 (Figure 6, group B), Tmem120A (Figure 3, group B) were found restricted to Opisthokonta, and Tm7SF2 (Figure 3, group C) was restricted to Metazoa, while NET45 and NET39 were identified broadly (Figure 3, group A). Whether NET45 and NET39 or other wider-distributed proteins can influence chromosome topology or induce gene regulation signaling pathways in single-cell organisms is yet to be explored.

Conclusions

Reconstructions from comparative genomics have established an overall pattern of conservation of endomembrane-system genes among all major lineages of eukaryotes including plants, protists, amoebae, animals and fungi, implying inheritance from a complex common ancestor [14]. It also suggests that a widely conserved, complex set of NEA proteins may exist, corresponding to the conservation of the NE itself. While 300–400 putative NEA proteins have been described in human and yeast cells, few orthologs of these proteins have been described across eukaryotic diversity to date. While the NEA protein datasets upon which we based our analysis are of high quality, we must emphasize the need for further experimental and in silico work to define NEA protein cohorts with greater confidence. Inherent to any proteomics investigation are sources of error, including contaminants, mislocalized proteins and proteins evading detection by mass spectrometry. In the case of the NE, isolation of pure fractions for proteomic profiling is particularly challenging due to continuity of the NE with the ER, and due to fibrous connections between the NE and both the cytoskeleton and the nucleoplasm. Further complexities arise, as many NEA proteins have dual localizations at the NE and other organelles such as ER, cytoskeleton, mitochondria and peroxisomes, and some proteins are only identified in the NE while in the early stages of their biosynthesis [60,62]. For some proteins, NE functions may be carried out by a minor fraction of the total cellular pool, as some of them can be detected in different organelles or even multiple localizations within the same organelle. Moreover, although orthologs tend to retain function [118], many widely conserved NEA proteins are characterized in only one model organism, leaving their functions in distantly related eukaryotes uncertain. This, together with the apparent prevalence of lineage-specific NEA proteins revealed here warrants parallel experimentation efforts using multiple model organisms. Our present analysis is certainly not exhaustive and is intended to provide a broad overview of modes of NEA protein evolution. We are limited not only by uncertainty in experimental identification of NEA proteins but also sources of potential error inherent to sequence similarity search methods. For example, we are aware that highly divergent orthologs will not have been retrieved by our BLAST searches, but might be retrieved through further analysis with profile searching methods. Also, we have attempted to be conservative, preferring to exclude potential orthologs when there is reason for significant doubt. Thus, further analyses may reveal somewhat more (or less) extensive diversity in nuclear composition and hence, function. Certainly, examples of churning or backfilling would not be detected here. Nevertheless, considerable divergence in NEA protein cohorts between lineages is clear, with 50% or fewer proteins detected as broadly distributed, despite low stringency searching followed by manual validation. Ribosome biogenesis, RNA processing proteins and proteins with participation in transcription are highly conserved,

unsurprisingly, followed by proteins involved in lipid biosynthesis [91]. Regardless, the cohort is dominated by lineage-specific proteins, many of which, like nesprins, possibly have analogs in other lineages. Hence, when considering the NE, variations in the NPC, lamina systems, cytoskeletal anchoring and kinetochores, the nucleus provides an example of considerable diversity despite hosting a plethora of core functions (Figure 7). Indeed, these findings suggest a rather surprising level of divergence associated with a structure that, in a very real sense, defines the eukaryotic cell.

Acknowledgments

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

ORCID

Mark C. Field 🔘 http://orcid.org/0000-0002-4866-2885

References

[1] Betts HC, Puttick MN, Clark JW, et al. Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. Nat Ecol Evol. 2018;2(10):1556–1562.

[2] Dacks JB, Field MC, Buick R, et al. The changing view of eukaryogenesis - fossils, cells, lineages and how they all come together. J Cell Sci. 2016;129 (20):3695–3703.

[3] Szathmary E, Smith JM. The major evolutionary transitions. Nature. 1995;374(6519):227–232.

[4] Williams TA, Foster PG, Nye TM, et al. A congruent phylogenomic signal places eukaryotes within the Archaea. Proc Biol Sci. 2012;279(1749):4870–4879.

[5] Winker S, Woese CR. A definition of the domains Archaea, bacteria and eucarya in terms of small sub-unit ribosomal RNA characteristics. Syst Appl Microbiol. 1991;14(4):305–310.

[6] Tang Q, Pang K, Yuan X, et al. A one-billion-year-old multicellular chlorophyte. Nat Ecol Evol. 2020;4 (4):543–549.

[7] Sagan L. On the origin of mitosing cells. J Theor Biol. 1967;14(3):255–274.

[8] Elias M, Brighouse A, Gabernet-Castello C, et al. Sculpting the endomembrane system in deep time: high resolution phylogenetics of Rab GTPases. J Cell Sci. 2012;125(10):2500–2508.

[9] Vicente JJ, Wordeman L. Mitosis, microtubule dynamics and the evolution of kinesins. Exp Cell Res. 2015;334(1):61–69.

[10] More K, Klinger CM, Barlow LD, et al. Evolution and natural history of membrane trafficking in eukaryotes. Curr Biol. 2020;30(10):R553–R64.

[11] Dacks JB, Field MC. Evolutionary origins and specialisation of membrane transport. Curr Opin Cell Biol. 2018;53:70–76.

[12] Pereira-Leal JB. The Ypt/Rab family and the evolution of trafficking in fungi. Traffic. 2008;9(1):27–38.

[13] Sparvoli D, Zoltner M, Cheng CY, et al. Diversification of CORVET tethers facilitates transport complexity in tetrahymena thermophila. J Cell Sci. 2020;133(3):3.

[14] Rout MP, Field MC. The evolution of organellar coat complexes and organization of the eukaryotic cell. Annu Rev Biochem. 2017;86(1):637–657.

[15] Martin WF, Garg S, Zimorski V. Endosymbiotic theories for eukaryote origin. Philos Trans R Soc Lond B Biol Sci. 2015;370(1678):20140330.

[16] Pittis AA, Gabaldon T. Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. Nature. 2016;531(7592):101–104.

[17] Field MC, Rout MP. Pore timing: the evolutionary origins of the nucleus and nuclear pore complex. F1000Res. 2019;8:369.

[18] Xie Y, Ren Y. Mechanisms of nuclear mRNA export: a structural perspective. Traffic. 2019;20(11):829–840.

[19] Thaller DJ, Allegretti M, Borah S, et al. LEM protein surveillance system is poised to directly monitor the nuclear envelope and nuclear transport system. Elife. 2019;8. DOI:10.7554/eLife.45284..

[20] Vietri M, Radulovic M, Stenmark H. The many functions of ESCRTs. Nat Rev Mol Cell Biol. 2020;21 (1):25–42.

[21] Aoki K, Hayashi H, Furuya K, et al. Breakage of the nuclear envelope by an extending mitotic nucleus occurs during anaphase in Schizosaccharomyces japonicus. Genes Cells. 2011;16(9):911–926.

[22] Sazer S, Lynch M, Needleman D. Deciphering the evolutionary history of open and closed mitosis. Curr Biol. 2014;24(22):R1099–103.

[23] Devos DP, Graf R, Field MC. Evolution of the nucleus. Curr Opin Cell Biol. 2014;28:8–15.

[24] Meier I, Richards EJ, Evans DE. Cell biology of the plant nucleus. Annu Rev Plant Biol. 2017;68 (1):139–172.

[25] Koreny L, Field MC. Ancient eukaryotic origin and evolutionary plasticity of nuclear lamina. Genome Biol Evol. 2016;8(9):2663–2671.

[26] Ciska M, Masuda K. Moreno Diaz de la Espina S. Lamin-like analogues in plants: the characterization of NMCP1 in allium cepa. J Exp Bot. 2013;64(6):1553–1564.

[27] DuBois KN, Alsford S, Holden JM, et al. NUP-1 is a large coiled-coil nucleoskeletal protein in trypanosomes with lamin-like functions. PLoS Biol. 2012;10 (3):e1001287.

[28] Maishman L, Obado SO, Alsford S, et al. Co-dependence between trypanosome nuclear lamina components in nuclear stability and control of gene expression. Nucleic Acids Res. 2016;44(22):10554–10570.

[29] Padilla-Mejia NE, Koreny L, Holden J, et al. A hub and spoke model of assembly for the trypanosome nuclear lamina. J Cell Sci. 2020.

[30] Strambio-de-Castillia C, Blobel G, Rout MP. Isolation and characterization of nuclear envelopes from the yeast Saccharomyces. J Cell Biol. 1995;131(1):19–31.

[31] Hirano Y, Asakawa H, Sakuno T, et al. Nuclear envelope proteins modulating the heterochromatin formation and functions in fission yeast. Cells. 2020;9(8):8.

[32] Gonzalez Y, Saito A, Sazer S. Fission yeast Lem2 and Man1 perform fundamental functions of the animal cell nuclear lamina. Nucleus. 2012;3(1):60–76.

[33] Aitchison JD, Rout MP. The yeast nuclear pore complex and transport through it. Genetics. 2012;190 (3):855–883.

[34] Akiyoshi B, Gull K. Discovery of unconventional kinetochores in kinetoplastids. Cell. 2014;156 (6):1247–1258.

[35] Field MC. The kinetochore and the origin of eukaryotic chromosome segregation. Proc Natl Acad Sci U S A. 2019;116(26):12596–12598.

[36] Ebenezer TE, Zoltner M, Burrell A, et al. Transcriptome, proteome and draft genome of Euglena gracilis. BMC Biol. 2019;17(1):11.

[37] Kenny NJ, Francis WR, Rivera-Vicens RE, et al. Tracing animal genomic evolution with the chromosomal-level assembly of the freshwater sponge Ephydatia muelleri. Nat Commun. 2020;11(1):3676.

[38] Leger MM, Kolisko M, Kamikawa R, et al. Organelles that illuminate the origins of trichomonas hydrogenosomes and giardia mitosomes. Nat Ecol Evol. 2017;1(4):0092.

[39] Lax G, Eglit Y, Eme L, et al. Hemimastigophora is a novel supra-kingdom-level lineage of eukaryotes. Nature. 2018;564(7736):410–414.

[40] Tang S, Lomsadze A, Borodovsky M. Identification of protein coding regions in RNA transcripts. Nucleic Acids Res. 2015;43(12):e78.

[41] Schirmer EC, Florens L, Guan T, et al. Nuclear membrane proteins with potential disease links found by subtractive proteomics. Science. 2003;301(5638):1380–1382.

[42] de Las Heras JI, Meinke P, Batrakou DG, et al. Tissue specificity in the nuclear envelope supports its functional complexity. Nucleus. 2013;4(6):460–477.

[43] Korfali N, Wilkie GS, Swanson SK, et al. The nuclear envelope proteome differs notably between tissues. Nucleus. 2012;3(6):552–564.

[44] Wilhelmsen K, Litjens SH, Kuikman I, et al. Nesprin-3, a novel outer nuclear membrane protein, associates with the cytoskeletal linker protein plectin. J Cell Biol. 2005;171(5):799–810.

[45] Brachner A, Reipert S, Foisner R, et al. LEM2 is a novel MAN1-related inner nuclear membrane protein associated with A-type lamins. J Cell Sci. 2005;118(24):5797–5810.

[46] Rout MP, Field MC. Isolation and characterization of subnuclear compartments from Trypanosoma brucei. identification of a major repetitive nuclear lamina component. J Biol Chem. 2001;276(41):38261–38271.

[47] DeGrasse JA, DuBois KN, Devos D, et al. Evidence for a shared nuclear pore complex architecture that is conserved from the last common eukaryotic ancestor. Mol Cell Proteomics. 2009;8(9):2119–2130.

[48] Sonnhammer EL, von Heijne G, Krogh A. A hidden markov model for predicting transmembrane helices in protein sequences. Proceedings International Conference on Intelligent Systems for Molecular Biology. 1998;6:175–182.

[49] Aslett M, Aurrecoechea C, Berriman M, et al. TriTrypDB: a functional genomic resource for the trypanosomatidae. Nucleic Acids Res. 2010;38 (suppl_1):D457–62.

[50] Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25(17):3389–3402.

[51] Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–1797.

[52] Lawrence TJ, Kauffman KT, Amrine KC, et al. FAST: FAST analysis of sequences toolbox. Front Genet. 2015;6:172.

[53] Price MN, Dehal PS, Arkin AP. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5(3):e9490.

[54] Huerta-Cepas J, Serra F, Bork P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. Mol Biol Evol. 2016;33(6):1635–1638.

[55] Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics. 2014;30(22):3276–3278.

[56] Waterhouse AM, Procter JB, Martin DM, et al. 2–a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25(9):1189–1191.

[57] Field HI, Coulson RM, Field MC. An automated graphics tool for comparative genomics: the Coulson plot generator. BMC Bioinf. 2013;14(1):141.

[58] UniProt C. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2019;47(D1):D506–D15.

[59] Huntley RP, Sawford T, Mutowo-Meullenet P, et al. The GOA database: gene ontology annotation updates for 2015. Nucleic Acids Res. 2015;43(D1): D1057–63.

[60] Tang Y, Huang A, Gu Y. Global profiling of plant nuclear membrane proteome in Arabidopsis. Nat Plants. 2020;6(7):838–847.

[61] Graumann K, Evans DE. Growing the nuclear envelope proteome. Nat Plants. 2020;6(7):740–741.

[62] Goos C, Dejung M, Janzen CJ, et al. The nuclear proteome of Trypanosoma brucei. PLoS One. 2017;12(7):e0181884.

[63] Dohmen E, Klasberg S, Bornberg-Bauer E, et al. The modular nature of protein evolution: domain rearrangement rates across eukaryotic life. BMC Evol Biol. 2020;20(1):30.

[64] Kawashima T, Kawashima S, Tanaka C, et al. Domain shuffling and the evolution of vertebrates. Genome Res. 2009;19(8):1393–1403.

[65] Burki F, Roger AJ, Brown MW, et al. The new tree of Eukaryotes. Trends Ecol Evol. 2020;35(1):43–55.

[66] Ishikawa H, Barber GN. STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling. Nature. 2008;455(7213):674–678.

[67] Malik P, Zuleger N, de Las Heras JI, et al. NET23/STING promotes chromatin compaction from the nuclear envelope. PLoS One. 2014;9(11):e111851.

[68] Datta K, Guan T, Gerace L. NET37, a nuclear envelope transmembrane protein with glycosidase homology, is involved in myoblast differentiation. J Biol Chem. 2009;284(43):29666–29676.

[69] Pall GS, Wallis J, Axton R, et al. A novel transmembrane MSP-containing protein that plays a role in right ventricle development. Genomics. 2004;84(6):1051–1059.

[70] Muroyama Y, Saito T. Identification of Nepro, a gene required for the maintenance of neocortex neural progenitor cells downstream of Notch. Development. 2009;136(23):3889–3893.

[71] Heyne K, Willnecker V, Schneider J, et al. NIR, an inhibitor of histone acetyltransferases, regulates transcription factor TAp63 and is controlled by the cell cycle. Nucleic Acids Res. 2010;38(10):3159–3171.

[72] Prieto JL, McStay B. Recruitment of factors linking transcription and processing of pre-rRNA to NOR chromatin is UBF-dependent and occurs independent of transcription in human cells. Genes Dev. 2007;21(16):2041–2054.

[73] Baillat D, Hakimi MA, Naar AM, et al. Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. Cell. 2005;123(2):265–276.

[74] Huang CY, Beliakoff J, Li X, et al. hZimp7, a novel -PIAS-like protein, enhances androgen receptor-mediated transcription and interacts with SWI/SNF-like BAF complexes. Mol Endocrinol. 2005;19(12):2915–2929.

[75] Ni Z, Olsen JB, Guo X, et al. Control of the RNA polymerase II phosphorylation state in promoter regions by CTD interaction domain-containing proteins RPRD1A and RPRD1B. Transcription. 2011;2(5):237–242.

[76] Ni Z, Xu C, Guo X, et al. RPRD1A and RPRD1B are human RNA polymerase II C-terminal domain scaffolds for Ser5 dephosphorylation. Nat Struct Mol Biol. 2014;21(8):686–695.

[77] Malik P, Korfali N, Srsen V, et al. Cell-specific and lamin-dependent targeting of novel transmembrane proteins in the nuclear envelope. Cell Mol Life Sci. 2010;67(8):1353–1369.

[78] Bronshtein I, Kepten E, Kanter I, et al. Loss of lamin A function increases chromatin dynamics in the nuclear interior. Nat Commun. 2015;6(1):8044.

[79] Moir RD, Yoon M, Khuon S, et al. Nuclear lamins A and B1: different pathways of assembly during nuclear envelope formation in living cells. J Cell Biol. 2000;151(6):1155–1168.

[80] Dechat T, Gesson K, Foisner R. Lamina-independent lamins in the nuclear interior serve important functions. Cold Spring Harbor Symposia on Quantitative Biology. 2010;75:533–543.

[81] Fong KW, Leung JW, Li Y, et al. MTR120/KIAA1383, a novel microtubule-associated protein, promotes microtubule stability and ensures cytokinesis. J Cell Sci. 2013;126(3):825–837.

[82] Li C, Wei J, Li Y, et al. Transmembrane protein 214 (TMEM214) mediates endoplasmic reticulum stress-induced caspase 4 enzyme activation and apoptosis. J Biol Chem. 2013;288(24):17908–17917.

[83] Wilkie GS, Korfali N, Swanson SK, et al. Several novel nuclear envelope transmembrane proteins identified in skeletal muscle have cytoskeletal associations. Mol Cell Proteomics. 2011;10(1):M110.003129.

[84] Houstek J, Kmoch S, Zeman J. TMEM70 protein - a novel ancillary factor of mammalian ATP synthase. Biochim Biophys Acta. 2009;1787(5):529–532.

[85] Kovalčíková J, Vrbacký M, Pecina P, et al. TMEM70 facilitates biogenesis of mammalian ATP synthase by promoting subunit c incorporation into the rotor structure of the enzyme. Faseb J. 2019;33(12):14103–14117.

[86] Dacks JB, Field MC. Eukaryotic cell evolution from a comparative genomic perspective: the endomembrane system. In: Hirt RP, Horner DS, editors. Organelles, genomes and eukaryote phylogeny: an evolutionary synthesis in the age of genomics. CRC Press; 2004;309–334.

[87] Te Heesen S, Knauer R, Lehle L, et al. Yeast Wbp1p and Swp1p form a protein complex essential for oligosaccharyl transferase activity. Embo J. 1993;12 (1):279–284.

[88] Te Heesen S, Rauhut R, Aebersold R, et al. An essential 45 kDa yeast transmembrane protein reacts with anti-nuclear pore antibodies: purification of the protein, immunolocalization and cloning of the gene. Eur J Cell Biol. 1991;56(1):8–18.

[89] Howe AG, Zaremberg V, McMaster CR. Cessation of growth to prevent cell death due to inhibition of phosphatidylcholine synthesis is impaired at 37 degrees C in Saccharomyces cerevisiae. J Biol Chem. 2002;277(46):44100–44107.

[90] Dean S, Sunter JD, Wheeler RJ. TrypTag.org: a trypanosome genome-wide protein localisation resource. Trends Parasitol. 2017;33(2):80–82.

[91] Garapati HS, Mishra K. Comparative genomics of nuclear envelope proteins. BMC Genomics. 2018;19 (1):823.

[92] Meier I. LINCing the eukaryotic tree of life - towards a broad evolutionary comparison of nucleocytoplasmic bridging complexes. J Cell Sci. 2016;129 (19):3523–3531.

[93] Zhou X, Graumann K, Wirthmueller L, et al. Identification of unique SUN-interacting nuclear envelope proteins with diverse functions in plants. J Cell Biol. 2014;205(5):677–692.

[94] Muchir A, Worman HJ. The nuclear envelope and human disease. Physiology (Bethesda). 2004;19:309–314.

[95] Dauer WT, Worman HJ. The nuclear envelope as a signaling node in development and disease. Dev Cell. 2009;17(5):626–638.

[96] Meinke P, Makarov A, Lê Thành P, et al. Nucleoskeleton dynamics and functions in health and disease. Cell Health and Cytoskeleton. 2014;7:55–69.

[97] Zheng G, Jiang C, Li Y, et al. TMEM43-S358L mutation enhances NF-kappaB-TGFbeta signal cascade in arrhythmogenic right ventricular dysplasia/cardiomyopathy. Protein Cell. 2019;10(2):104–119.

[98] Heller SA, Shih R, Kalra R, et al. Emery-Dreifuss muscular dystrophy. Muscle Nerve. 2020;61 (4):436–448.

[99] Liang WC, Mitsuhashi H, Keduka E, et al. TMEM43 mutations in Emery-Dreifuss muscular dystrophy-related myopathy. Ann Neurol. 2011;69 (6):1005–1013.

[100] Muhammad Aslam MK, Sharma VK, Pandey S, et al. Identification of biomarker candidates for fertility in spermatozoa of crossbred bulls through comparative proteomics. Theriogenology. 2018;119:43–51.

[101] Bengtsson L, Otto H. LUMA interacts with emerin and influences its distribution at the inner nuclear membrane. J Cell Sci. 2008;121(4):536–548.

[102] Browman DT, Resek ME, Zajchowski LD, et al. Erlin-1 and erlin-2 are novel members of the prohibitin family of proteins that define lipid-raft-like domains of the ER. J Cell Sci. 2006;119(15):3149–3160.

[103] Yildirim Y, Orhan EK, Iseri SA, et al. A frameshift mutation of ERLIN2 in recessive intellectual disability, motor dysfunction and multiple joint contractures. Hum Mol Genet. 2011;20(10):1886–1892.

[104] Wu H, Li J, Guo E, et al. MiR-410 acts as a tumor suppressor in estrogen receptor-positive breast cancer cells by directly targeting ERLIN2 via the ERS pathway. Cell Physiol Biochem. 2018;48 (2):461–474.

[105] Gudise S, Figueroa RA, Lindberg R, et al. 1 is functionally associated with the LINC complex and A-type lamina networks. J Cell Sci. 2011;124 (12):2077–2085.

[106] Meinke P, Kerr ARW, Czapiewski R, et al. A multistage sequencing strategy pinpoints novel candidate alleles for Emery-Dreifuss muscular dystrophy and supports gene misregulation as its pathomechanism. EBioMedicine. 2020;51:102587.

[107] Worman HJ, Yuan J, Blobel G, et al. A lamin B receptor in the nuclear envelope. Proc Natl Acad Sci U S A. 1988;85(22):8531–8534.

[108] Tsai PL, Zhao C, Turner E, et al. B receptor is essential for cholesterol synthesis and perturbed by disease-causing mutations. Elife. 2016;5:5.

[109] Subramanian G, Chaudhury P, Malu K, et al. Lamin B receptor regulates the growth and maturation of myeloid progenitors via its sterol reductase domain: implications for cholesterol biosynthesis in regulating myelopoiesis. J Immunol. 2012;188(1):85–102.

[110] Schrick K, Mayer U, Horrichs A, et al. FACKEL is a sterol C-14 reductase required for organized cell division and expansion in Arabidopsis embryogenesis. Genes Dev. 2000;14(12):1471–1484.

[111] Wheeler MA, Davies JD, Zhang Q, et al. Distinct functional domains in nesprin-1alpha and nesprin-2beta bind directly to emerin and both interactions are disrupted in X-linked Emery-Dreifuss muscular dystrophy. Exp Cell Res. 2007;313(13):2845–2857.

[112] Zhang Q, Bethmann C, Worth NF, et al. Nesprin-1 and −2 are involved in the pathogenesis of Emery Dreifuss muscular dystrophy and are critical for nuclear envelope integrity. Hum Mol Genet. 2007;16(23):2816–2833.

[113] Zhou C, Rao L, Shanahan CM, et al. Nesprin-1/2: roles in nuclear envelope organisation, myogenesis and muscle disease. Biochem Soc Trans. 2018;46 (2):311–320.

[114] Koch AJ, Holaska JM. Emerin in health and disease. Semin Cell Dev Biol. 2014;29:95–106.

[115] Robson MI, de Las Heras JI, Czapiewski R, et al. Tissue-specific gene repositioning by muscle nuclear membrane proteins enhances repression of critical developmental genes during myogenesis. Mol Cell. 2016;62(6):834–847.

[116] Zuleger N, Boyle S, Kelly DA, et al. Specific nuclear envelope transmembrane proteins can promote the location of chromosomes to and from the nuclear periphery. Genome Biol. 2013;14(2):R14.

[117] de Las Heras JI, Zuleger N, Batrakou DG, et al. Tissue-specific NETs alter genome organization and regulation even in a heterologous system. Nucleus. 2017;8(1):81–97.

[118] Klinger CM, Ramirez-Macias I, Herman EK, et al. Resolving the homology-function relationship through comparative genomics of membrane-trafficking machinery and parasite cell biology. Mol Biochem Parasitol. 2016;209(1–2):88–103.

[119] Reinstein E, Drasinover V, Lotan R, et al. Mutations in ERGIC1 cause arthrogryposis multiplex congenita, neuropathic type. Clin Genet. 2018;93(1):160–163.

[120] Wang FR, Wei YC, Han ZJ, et al. Aberrant DNA-PKcs and ERGIC1 expression may be involved in initiation of gastric cancer. World J Gastroenterol. 2017;23(33):6119–6127.

[121] Li Q, Liu X, Jin K, et al. NAT10 is upregulated in hepatocellular carcinoma and enhances mutant p53 activity. BMC Cancer. 2017;17(1):605.

[122] Balmus G, Larrieu D, Barros AC, et al. Targeting of NAT10 enhances healthspan in a mouse model of human accelerated aging syndrome. Nat Commun. 2018;9(1):1700.

[123] Larrieu D, Britton S, Demir M, et al. Chemical inhibition of NAT10 corrects defects of laminopathic cells. Science. 2014;344(6183):527–532.

[124] Barateau A, Vadrot N, Vicart P, et al. A novel lamin a mutant responsible for congenital muscular dystrophy causes distinct abnormalities of the cell nucleus. PLoS One. 2017;12(1):e0169189.

[125] Thiel C, Schwarz M, Peng J, et al. A new type of congenital disorders of glycosylation (CDG-Ii) provides new insights into the early steps of dolichol-linked oligosaccharide biosynthesis. J Biol Chem. 2003;278(25):22498–22505.

[126] Cossins J, Belaya K, Hicks D, et al. Congenital myasthenic syndromes due to mutations in ALG2 and ALG14. Brain. 2013;136(3):944–956.

[127] Belaya K, Finlayson S, Slater CR, et al. Mutations in DPAGT1 cause a limb-girdle congenital myasthenic syndrome with tubular aggregates. Am J Hum Genet. 2012;91(1):193–201.

[128] Brant SR, Panhuysen CI, Nicolae D, et al. MDR1 Ala893 polymorphism is associated with inflammatory bowel disease. Am J Hum Genet. 2003;73 (6):1282–1292.

[129] Seo J, Lee CR, Paeng JC, et al. Biallelic mutations in ABCB1 display recurrent reversible encephalopathy. Ann Clin Transl Neurol. 2020;7(8):1443–1449.

[130] Merner ND, Hodgkinson KA, Haywood AF, et al. Arrhythmogenic right ventricular cardiomyopathy type 5 is a fully penetrant, lethal arrhythmic disorder caused by a missense mutation in the TMEM43 gene. Am J Hum Genet. 2008;82(4):809–821.

[131] Dominguez F, Zorio E, Jimenez-Jaimez J, et al. Clinical characteristics and determinants of the phenotype in TMEM43 arrhythmogenic right ventricular cardiomyopathy type 5. Heart Rhythm. 2020;17 (6):945–954.

[132] Mukai T, Mori-Yoshimura M, Nishikawa A, et al. Emery-Dreifuss muscular dystrophy-related myopathy with TMEM43 mutations. Muscle Nerve. 2019;59 (2):E5–E7.

[133] Srivastava S, D'Amore A, Cohen JS, et al. Expansion of the genetic landscape of ERLIN2-related disorders. Ann Clin Transl Neurol. 2020;7(4):573–578.

[134] Thompson E, Abdalla E, Superti-Furga A, et al. Lamin B receptor-related disorder is associated with a spectrum of skeletal dysplasia phenotypes. Bone. 2019;120:354–363.

[135] Best S, Salvati F, Kallo J, et al. Lamin B-receptor mutations in Pelger-Huet anomaly. Br J Haematol. 2003;123(3):542–544.

[136] Blommaert E, Peanne R, Cherepanova NA, et al. Mutations in MAGT1 lead to a glycosylation disorder with a variable phenotype. Proc Natl Acad Sci U S A. 2019;116(20):9865–9870.

[137] Li FY, Chaigne-Delalande B, Su H, et al. XMEN disease: a new primary immunodeficiency affecting Mg2+ regulation of immunity against Epstein-Barr virus. Blood. 2014;123(14):2148–2152.

[138] Strom TM, Hortnagel K, Hofmann S, et al. Diabetes insipidus, diabetes mellitus, optic atrophy and deafness (DIDMOAD) caused by mutations in a novel gene (Wolframin) coding for a predicted transmembrane protein. Hum Mol Genet. 1998;7(13):2021–2028.

[139] Inoue H, Tanizawa Y, Wasson J, et al. A gene encoding a transmembrane protein is mutated in patients with diabetes mellitus and optic atrophy (Wolfram syndrome). Nat Genet. 1998;20(2):143–148.

[140] Sawada A, Takihara Y, Kim JY, et al. A congenital mutation of the novel gene LRRC8 causes agammaglobulinemia in humans. J Clin Invest. 2003;112 (11):1707–1713.

[141] Gasull X, Castany M, Castellanos A, et al. The LRRC8-mediated volume-regulated anion channel is altered in glaucoma. Sci Rep. 2019;9(1):5392.

[142] Laurin N, Brown JP, Morissette J, et al. Recurrent mutation of the gene encoding sequestosome 1 (SQSTM1/p62) in Paget disease of bone. Am J Hum Genet. 2002;70(6):1582–1588.

[143] Fecto F, Yan J, Vemula SP, et al. SQSTM1 mutations in familial and sporadic amyotrophic lateral sclerosis. Arch Neurol. 2011;68(11):1440–1446.

[144] Bucelli RC, Arhzaouy K, Pestronk A, et al. SQSTM1 splice site mutation in distal myopathy with rimmed vacuoles. Neurology. 2015;85(8):665–674.

[145] Salminen A, Kaarniranta K, Haapasalo A, et al. Emerging role of p62/sequestosome-1 in the pathogenesis of Alzheimer's disease. Prog Neurobiol. 2012;96(1):87–95.

[146] Gorello P, La Starza R, Di Giacomo D, et al. SQSTM1-NUP214: a new gene fusion in adult T-cell acute lymphoblastic leukemia. Haematologica. 2010;95(12):2161–2163.

[147] Cizkova A, Stranecky V, Mayr JA, et al. TMEM70 mutations cause isolated ATP synthase deficiency and neonatal mitochondrial encephalocardiomyopathy. Nat Genet. 2008;40(11):1288–1290.

[148] Hellemans J, Preobrazhenska O, Willaert A, et al. Loss-of-function mutations in LEMD3 result in osteopoikilosis, Buschke-Ollendorff syndrome and melorheostosis. Nat Genet. 2004;36(11):1213–1218.

[149] Gillard EF, Otsu K, Fujii J, et al. Polymorphisms and deduced amino acid substitutions in the coding sequence of the ryanodine receptor (RYR1) gene in individuals with malignant hyperthermia. Genomics. 1992;13(4):1247–1254.

[150] Quane KA, Healy JM, Keating KE, et al. Mutations in the ryanodine receptor gene in central core disease and malignant hyperthermia. Nat Genet. 1993;5 (1):51–55.

[151] Zhang Y, Chen HS, Khanna VK, et al. A mutation in the human ryanodine receptor gene associated with central core disease. Nat Genet. 1993;5(1):46–50.

[152] Monnier N, Ferreiro A, Marty I, et al. A homozygous splicing mutation causing a depletion of skeletal muscle RYR1 is associated with multi-minicore disease congenital myopathy with ophthalmoplegia. Hum Mol Genet. 2003;12(10):1171–1178.

[153] Paez JG, Janne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science. 2004;304(5676):1497–1500.

[154] Campbell P, Morton PE, Takeichi T, et al. Epithelial inflammation resulting from an inherited loss-of-function mutation in EGFR. J Invest Dermatol. 2014;134(10):2570–2578.

[155] Nagano A, Koga R, Ogawa M, et al. Emerin deficiency at the nuclear membrane in patients with Emery-Dreifuss muscular dystrophy. Nat Genet. 1996;12(3):254–259.

[156] Agopian AJ, Mitchell LE, Glessner J, et al. Genome-wide association study of maternal and inherited loci for conotruncal heart defects. PLoS One. 2014;9(5):e96057.

[157] Boone PM, Yuan B, Gu S, et al. Hutterite-type cataract maps to chromosome 6p21.32-p21.31, cosegregates with a homozygous mutation in LEMD2, and is associated with sudden cardiac death. Mol Genet Genomic Med. 2016;4(1):77–94.

[158] Xin B, Platzer P, Fink SP, et al. Colon cancer secreted protein-2 (CCSP-2), a novel candidate serological marker of colon neoplasia. Oncogene. 2005;24 (4):724–731.

[159] Jeun M, Lee HJ, Park S, et al. A novel blood-based colorectal cancer diagnostic technology using electrical detection of colon cancer secreted protein-2. Adv Sci (Weinh). 2019;6(11):1802115.

[160] Vander Heyden AB, Naismith TV, Snapp EL, et al. LULL1 retargets TorsinA to the nuclear envelope revealing an activity that is impaired by the DYT1 dystonia mutation. Mol Biol Cell. 2009;20 (11):2661–2672.

[161] Taylor MR, Slavov D, Gajewski A, et al. Thymopoietin (lamina-associated polypeptide 2) gene mutation associated with dilated cardiomyopathy. Hum Mutat. 2005;26(6):566–574.

[162] Gros-Louis F, Dupre N, Dion P, et al. Mutations in SYNE1 lead to a newly discovered form of autosomal recessive cerebellar ataxia. Nat Genet. 2007;39(1):80–85.

[163] Attali R, Warwar N, Israel A, et al. Mutation of SYNE-1, encoding an essential component of the nuclear lamina, is responsible for autosomal recessive arthrogryposis. Hum Mol Genet. 2009;18 (18):3462–3469.

[164] Brandt DT, Baarlink C, Kitzing TM, et al. SCAI acts as a suppressor of cancer cell invasion through the transcriptional control of beta1-integrin. Nat Cell Biol. 2009;11(5):557–568.

[165] Lu D, Wu Y, Wang Y, et al. CREPT accelerates tumorigenesis by regulating the transcription of cell-cycle-related genes. Cancer Cell. 2012;21(1):92–104.

[166] Batrakou DG, de Las Heras JI, Czapiewski R, et al. TMEM120A and B: nuclear envelope transmembrane proteins important for adipocyte differentiation. PLoS One. 2015;10(5):e0127712.

[167] Beaulieu-Laroche L, Christin M, Donoghue A, et al. TACAN is an ion channel involved in sensing mechanical pain. Cell. 2020;180(5):956–67.e17.

[168] Han S, Bahmanyar S, Zhang P, et al. Nuclear envelope phosphatase 1-regulatory subunit 1 (formerly TMEM188) is the metazoan Spo7p ortholog and functions in the lipin activation pathway. J Biol Chem. 2012;287(5):3123–3137.

[169] Hu Z, Gomes I, Horrigan SK, et al. A novel nuclear protein, 5qNCA (LOC51780) is a candidate for the myeloid leukemia tumor suppressor gene on chromosome 5 band q31. Oncogene. 2001;20(47):6946–6954.

[170] Bi X, Xu Y, Li T, et al. RNA targets ribogenesis factor WDR43 to chromatin for transcription and pluripotency control. Mol Cell. 2019;75(1):102–16.e9.

[171] Yang Y, Cao J, Shi Y. Identification and characterization of a gene encoding human LPGAT1, an endoplasmic reticulum-associated lysophosphatidylglycerol acyltransferase. J Biol Chem. 2004;279(53):55866–55874.

[172] Liu Y, Jesus AA, Marrero B, et al. Activated STING in a vascular and pulmonary syndrome. N Engl J Med. 2014;371(6):507–518.

[173] Goodchild RE, Kim CE, Dauer WT. Loss of the dystonia-associated protein TorsinA selectively disrupts the neuronal nuclear envelope. Neuron. 2005;48(6):923–932.

[174] Bennati AM, Castelli M, Della Fazia MA, et al. Sterol dependent regulation of human TM7SF2 gene expression: role of the encoded 3beta-hydroxysterol delta14-reductase in human cholesterol biosynthesis. Biochim Biophys Acta. 2006;1761(7):677–685.

[175] Zhou H, Clapham DE. Mammalian MagT1 and TUSC3 are required for cellular magnesium uptake and vertebrate embryonic development. Proc Natl Acad Sci U S A. 2009;106(37):15750–15755.

[176] Joung I, Strominger JL, Shin J. Molecular cloning of a phosphotyrosine-independent ligand of the p56lck SH2 domain. Proc Natl Acad Sci U S A. 1996;93 (12):5991–5995.

[177] Groenestege WM, Thébault S, van der Wijst J, et al. Impaired basolateral sorting of pro-EGF causes isolated recessive renal hypomagnesemia. J Clin Invest. 2007;117(8):2260–2267.

[178] Martin C, Chapman KE, Seckl JR, et al. Partial cloning and differential expression of ryanodine receptor/ calcium-release channel genes in human tissues including the hippocampus and cerebellum. Neuroscience. 1998;85(1):205–216.

[179] Yee SW, Stecula A, Chien HC, et al. Unraveling the functional role of the orphan solute carrier, SLC22A24 in the transport of steroid conjugates through metabolomic and genome-wide association studies. PLoS Genetics. 2019;15(9):e1008208.

[180] Sengle G, Kobbe B, Morgelin M, et al. Identification and characterization of AMACO, a new member of the von willebrand factor A-like domain protein superfamily with a regulated expression in the kidney. J Biol Chem. 2003;278(50):50240–50249.

[181] Zhang Q, Skepper JN, Yang F, et al. Nesprins: a novel family of spectrin-repeat-containing proteins that localize to the nuclear membrane in multiple tissues. J Cell Sci. 2001;114(Pt 24):4485–4498.

[182] Taranum S, Sur I, Müller R, et al. Cytoskeletal interactions at the nuclear envelope mediated by nesprins. International Journal of Cell Biology. 2012;2012:736524.