medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300210; this version posted December 19, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

The molecular landscape of premature aging diseases defined by multilayer network exploration

Cécile Beust[1,2,†], Alberto Valdeolivas[1,3,†], Anthony Baptista[1,4,5], Galadriel Brière[1,6], Nicolas Lévy[1,7], Ozan Ozisik[1,8], and Anaïs Baudot[1,9,10]

[1] *Aix Marseille Univ, INSERM, Marseille Medical Genetics (MMG), Marseille, France*
[2] *Present address: Univ Rennes, Inria, CNRS, IRISA - UMR 6074, F-35000 Rennes, France*
[3] *Present address: Roche Pharma Research and Early Development, Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd, Basel, Switzerland*
[4] *Present address: School of Mathematical Sciences, Queen Mary University of London, London, E1 4NS, United Kingdom*
[5] *Present address: The Alan Turing Institute, The British Library, London, NW1 2DB, United Kingdom*
[6] *Aix Marseille Univ, CNRS, I2M, Marseille, France*
[7] *Present address: Servier Research & Development, Saclay, France*
[8] *Present address: Université Paris Cité, INSERM U976, Paris, France*
[9] *CNRS, Marseille, France*
[10] *Barcelona Supercomputing Center (BSC), Barcelona, Spain*
[†] *These authors contributed equally to the work*

Abstract

Premature Aging (PA) diseases are rare genetic disorders that mimic some aspects of physiological aging at an early age. Various causative genes of PA diseases have been identified in recent years, providing insights into some dysfunctional cellular functions. However, the identification of PA genes also revealed significant genetic heterogeneity and highlighted the gaps in our understanding of PA molecular mechanisms. Furthermore, many patients remain undiagnosed. Overall, the current lack of knowledge about PA diseases hinders the development of effective diagnosis and therapies and poses significant challenges to improving patient care.

Here, we present a network-based approach to systematically unravel the cellular functions disrupted in PA diseases. Leveraging a novel community identification algorithm, we delved into a vast multilayer network of biological interactions to extract the communities of 67 PA diseases from their 132 associated genes. We found that these communities can be grouped into six distinct clusters, each reflecting specific cellular functions: DNA repair, cell cycle, transcription regulation, inflammation, cell communication, and vesicle-mediated transport. We propose that these clusters collectively represent the landscape of the molecular mechanisms that are perturbed in PA diseases, providing a framework for better understanding their pathogenesis. Intriguingly, most clusters also exhibited a significant enrichment in genes associated with physiological aging, suggesting a potential overlap between the molecular underpinnings of PA diseases and natural aging.

Keywords: Premature Aging; Network Biology; Community identification; Rare diseases

Contact: anais.baudot@univ-amu.fr

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300210; this version posted December 19, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

1 Introduction

Premature Aging (PA) diseases are rare genetic disorders that exhibit clinical features mimicking physiological aging but manifesting at an early age (Carrero et al., 2016; Schnabel et al., 2021). Aging phenotypes observed in PA diseases include, for instance, alopecia, hair graying, lipodystrophy, osteoporosis or atherosclerosis (Lessel and Kubisch, 2019; Schnabel et al., 2021). Overall, PA syndromes are clinically heterogeneous, presenting diverse sets of phenotypes, affecting different tissues and organs, and exhibiting variations in the age of onset, progression and severity (Navarro et al., 2006).

In recent years, advances in diagnosis techniques have allowed deciphering the genetic causes of various PA diseases. To date, over a hundred causative genes have been unveiled, revealing significant genetic heterogeneity (Lessel and Kubisch, 2019). Recent reviews have classified PA diseases in three major groups based on the molecular mechanisms associated with their causative genes (Grelet et al., 2019; Coppede, 2021). The first group mainly comprises diseases associated with alterations of the nuclear lamina architecture. This first group includes syndromes associated with defective Prelamin A processing, such as Hutchinson-Gilford Progeria Syndrome (HGPS) and Restrictive Dermopathy, and other laminopathies such as Mandibuloacral Dysplasia or Néstor-Guillermo Progeria Syndrome. The second group encompasses diseases resulting from alterations of DNA repair pathways, such as Werner syndrome, Cockayne syndrome, Xeroderma Pigmentosum or Bloom syndrome. Finally, the third group gathers PA diseases associated with diverse molecular defects, including connective tissue alterations (Ehlers-Danlos syndromes and Cutis Laxa), impairments in mitochondrial pathways (Fontaine progeroid syndrome), or others. For some diseases in this last group, the mechanisms underlying a premature senescence are not fully elucidated (Schnabel et al., 2021). Importantly, the genetic underpinnings of many PA diseases are still unknown, leaving numerous patients without a definitive molecular diagnosis (Coppede, 2021; Schnabel et al., 2021). Overall, gaining a deeper insight into the molecular mechanisms associated with PA diseases is crucial for enhancing diagnosis, optimizing patient care, and developing new therapeutic strategies to address the significant unmet needs in these ultra-rare, genetically-driven disorders.

The study of disease genes and proteins in their cellular network contexts provides a promising approach to understand the molecular mechanisms perturbed in diseases (Zhong et al., 2009). Indeed, proteins do not act as isolated molecules in cells but, instead, interact with each other to perform their cellular functions. Interactions are typically depicted as networks, in which the nodes represent individual genes or proteins, and the edges represent their physical or functional interactions. With the advancement of experimental techniques, biological interactions data accumulate and allow the construction of diverse biological interaction networks. These networks gather, for instance, protein-protein interactions, interactions occurring in molecular complexes, pathways, or gene/protein relationships inferred from -omics data, such as co-expression relationships inferred from transcriptomics data. This wealth of interaction data enables comprehensive studies of gene and protein cellular functions. However, given the magnitude and complexity of these interaction networks, appropriate computational methods are crucial for meaningful exploration.

Community detection is a broadly-used network analysis strategy. Biological networks tend to be structured into communities, which are groups of nodes that share more connections with one another than with the rest of the network (Hartwell et al., 1999). Communities in networks are subnetworks containing genes or proteins that are likely involved in the same processes or functions within cells. Identifying these communities enables the exploration of cellular functions using the 'guilt-by-association' principle. Importantly, by pinpointing the communities linked to disease-related genes, we gain a deeper insight into the cellular functions that are disrupted in these diseases (Furlong, 2013). A wide variety of computational methods have been proposed to extract communities from biological networks (Choobdar et al., 2019). Recently, various community identification algorithms have been extended to consider as input multiple biological networks, thereby providing improved mapping of cellular functions (MagnaniMatteo et al., 2021; Choobdar et al., 2019; Didier et al., 2015, 2018).

Our aim here is to systematically unveil the cellular functions underlying PA diseases. To this end, we conducted a comprehensive analysis of 67 PA diseases and their 132 associated genes. Leveraging a novel community identifi-

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300210; this version posted December 19, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

cation algorithm, we delved into a vast multilayer network of biological interactions, uncovering the 67 communities associated with the PA diseases. We next clustered the 67 disease communities based on their shared gene/protein nodes. This revealed a striking classification of the communities into six distinct clusters reflecting various cellular functions: DNA repair, cell cycle, transcription regulation, inflammation, cell communication, and vesicle-mediated transport. We propose that these clusters collectively represent the landscape of the dysfunctional molecular mechanisms in PA diseases. Our study thereby provides a valuable framework for decoding the pathogenesis of these disorders and guiding advancements in diagnostic and therapeutic approaches. Additionally, these clusters also exhibited a significant enrichment for genes associated with physiological aging, suggesting a potential overlap between the molecular underpinnings of PA diseases and the natural aging process.

2 Materials and methods

2.1 Premature aging disease and gene datasets

We inspected the Human Phenotype Ontology (HPO) (Köhler et al., 2017) to collect phenotypes associated with Premature Aging (PA). We selected the HPO annotation term: "HP:0007495: prematurely aged appearance" together with all its descendant terms in the HPO ontology graph (Supplementary Table S-1). In ORPHANET (Rath et al., 2012), we identified 87 diseases associated with at least one of these HPO terms. We called and hereafter refer to these 87 diseases as Premature Aging (PA) diseases. Focusing on the PA diseases having at least one associated gene in ORPHANET, we obtained 67 PA diseases associated with 132 genes (Supplementary Table S-2). Using the 132 PA genes as input and all annotated genes as background, we performed functional enrichments with the g:Profiler web server (version `e110_eg57_p18_4b54a898`). Functional enrichments were computed for Gene Ontology (The Gene Ontology Consortium et al., 2023) Biological Processes, Gene Ontology Cellular Components, and Reactome (Gillespie et al., 2022) annotations (Supplementary File S-1). P-values were corrected using the Benjamini-Hochberg procedure.

2.2 Multilayer biological network

The diverse physical and functional interactions between genes and proteins can be represented as a collection of network layers. More precisely, a multilayer networks in which the different layers share the same nodes but edges belonging to different types is called a multiplex network (Zitnik et al., 2023; Kivelä et al., 2014). We built a four-layer multiplex network of biological interactions. In all the layers, the nodes correspond to genes or proteins, which are here considered similarly and named by their approved symbols in the HUGO Gene Nomenclature Committee (HGNC) database (Seal et al., 2022). The protein-protein interactions (PPI) network layer was built by merging interactions from *Lit-BM*, *Hi-Union* (downloaded from *HuRI* (Luck et al., 2020)) and *APID* (Alonso-López et al., 2019) databases. The pathway network layer was downloaded from Reactome (Gillespie et al., 2022). Reactome provides a dataset of protein relationships inferred from pathway data. The protein relationships were inferred between reactants, polymers and all protein components of complexes (containing less than four components) involved in Reactome pathways. We built the co-expression network layer from RNA-Seq data publicly available on the Human Protein Atlas (Uhlen et al., 2015). We applied a Weighted Gene Correlation Network Analysis (WGCNA) on the Human Protein Atlas RNA-seq data using the WGCNA R package (Langfelder and Horvath, 2008). We selected a soft threshold power of 7 and exported the network with an adjacency threshold of 0.5. The molecular complex network layer was built by merging data from the *CORUM* database (Tsitsiridis et al., 2023) and the Human Protein Complex Map (Drew et al., 2021), using the R package OmniPathR (Türei et al., 2021).
The four network layers composing the multiplex network are available on the NDEx platform (Pratt et al., 2015). Network layer sizes and densities are detailed in Supplementary Table S-3. We used Cytoscape (Shannon et al., 2003) to visualize the networks.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300210; this version posted December 19, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

2.3 Community identification with iterative Random Walk with Restart (itRWR)

We implemented a novel approach, called iterative Random Walk with Restart (itRWR), to identify communities around seed node(s) in a multiplex network. We extended our previous Random Walk with Restart (RWR) approaches (Valdeolivas et al., 2018; Baptista et al., 2022). The itRWR first apply RWR to an initial set of seed(s). Then, at each iteration, the top-scored node is added to the set of seeds and the RWR is run again with this new set of seeds. The algorithm stops when a given number of iterations is reached. The itRWR implementation is based on the MultiXrank Python package (Baptista et al., 2022) and is available as a Python package on GitHub (`https://github.com/anthbapt/itRWR`).
We applied itRWR systematically to each of the 67 PA diseases. For each disease, we used its associated gene(s) as initial seed(s). The RWR parameters used to run the itRWR algorithm are described in Supplementary Table S-4. The itRWR algorithm was run with 100 iterations, leading to communities of size $100 + n$, with n being the initial number of seed(s) for a given disease (Supplementary File S-2). We also assessed the communities obtained with 50 and 150 iterations of the itRWR algorithm.
We performed a biological annotation enrichment analysis on the community obtained with 100 iterations of itRWR for each disease *g:Profiler* (version `e108_eg55_p17_0254fbf`) (Raudvere et al., 2019), using the community genes as input. We performed this enrichment analysis using the Gene Ontology (The Gene Ontology Consortium et al., 2023) Biological Processes and Cellular Components, and Reactome pathway (Gillespie et al., 2022) annotations. The background used for the statistical tests was restricted to the set of annotated genes contained in the multiplex network. P-values were corrected for multiple testing using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995) and the significance threshold was set at 0.05.

2.4 Community clustering

We computed the similarity between every pair of disease communities by computing the Jaccard index on the gene sets belonging to the two communities. The Jaccard Index is the ratio between the number of common elements between two sets (i.e., the number of genes in the intersection of the two communities) and the total number of elements in the two sets (i.e., the number of genes in the union of the two communities). We gathered the Jaccard Index values in a 67×67 similarity matrix, which we then converted into a distance matrix (by subtracting Jaccard Index similarities from 1). We next clustered the communities using hierarchical clustering of the distance matrix with the "average" method. We obtained a dendrogram of diseases. We set a cutoff at 0.7 on the dendrogram to retrieve clusters of diseases. We kept only the clusters containing at least 3 disease for further interpretation, and thereby obtained 6 clusters, each grouping three to 20 disease communities (see Supplementary File S-3 and S-4 for lists of genes and diseases in each cluster).
The clusters were generated from disease communities computed with 100 iterations of the itRWR algorithm. To assess how the choice of iteration number in the itRWR algorithm affects the clustering results, we compared the clusters generated from 100 iterations of itRWR with those from 50 and 150 iterations. We used the Rand Index for this comparison, a measure that evaluates the similarity of two clustering solutions by considering how disease pairs are grouped across different cluster sets. The Rand Index is calculated by dividing the number of disease pairs that are clustered together in two different sets of clusters by the total number of potential disease pairs. A higher Rand Index indicates greater agreement in disease clustering between the two sets, reflecting more consistent clustering outcomes.

2.5 Cluster enrichment analyses

2.5.1 Enrichment in HPO phenotypes

We used the PyHPO package (Köhler et al., 2019) to fetch HPO phenotypes and their associated ORPHANET diseases. We then performed an HPO phenotype enrichment analysis on each of the 6 community clusters. We

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300210; this version posted December 19, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

applied a hypergeometric test to assess HPO phenotype enrichments in each cluster based on the HPO phenotypes associated with all the diseases belonging to the cluster. We used the total number of ORPHANET diseases associated with at least one HPO phenotype as background (i.e, 4 262 diseases associated with 1 364 different phenotypes). P-values were corrected using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995) and the significance threshold was set at 0.05.

2.5.2 Enrichment in biological annotations

We performed biological annotation enrichment analyses on each of the six community clusters with *g:Profiler* (version `e108_eg55_p17_0254fbf`) (Raudvere et al., 2019), using the union of the genes belonging to the disease communities of each cluster as input. We performed these enrichment analyses using the Gene Ontology (The Gene Ontology Consortium et al., 2023) Biological Processes and Cellular Components, and Reactome pathway (Gillespie et al., 2022) annotations. The background used for the statistical tests was restricted to the set of annotated genes contained in the multiplex network. P-values were corrected for multiple testing using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995) and the significance threshold was set at 0.05. We filtered the enrichment results using the *orsum* Python package (Ozisik et al., 2022). *orsum* filters enrichment results by selecting more general and significant annotation terms as representatives for their less significant child terms. In addition, *orsum* allows integrating and representing enrichment results obtained in different analyses, in our case enrichments results obtained for the six community clusters. In *orsum*, we used the default parameters except the maximum representative term size (maxRepSize) that we set to 2 000 in Reactome pathway analyses.

2.5.3 Enrichment in physiological aging annotations

We performed enrichment analyses on each of the six community clusters to reveal enrichments in genes associated with physiological aging. To this goal, we used a list of 307 aging-related genes from the GenAge database (Tacutu et al., 2018a). From this list, we kept only the genes with a direct association with aging, i.e., we removed the genes presenting the association types "upstream", "downstream", "functional" and "putative". The final list of GenAge aging genes contains 107 genes. In order to analyse the enrichments of our six community clusters in GenAge aging genes, we performed a hypergeometric test. P-values were corrected using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995) and the significance threshold was set at 0,05.
We performed similar enrichment analyses using lists of genes up and down-regulated during aging in five different human tissues extracted from the study of Irizar et al. (Aramillo Irizar et al., 2018). The number of genes up-regulated and down-regulated in each tissue is detailed in the Supplementary File S-6. The lists of genes differentially expressed during aging in blood and skin were produced by the study of Irizar et al. The lists of genes differentially expressed during aging in brain, muscle and breast were extracted by Irizar et al. from three other publications (respectively Berchtold et al. (2008), Welle et al. (2003) and Pirone et al. (2012)). We mapped the Ensembl identifiers to The HUGO Gene Nomenclature Committee (HGNC) database (Seal et al., 2022). We performed hypergeometric tests for genes up-regulated and down regulated in each of our six community clusters. P-values were corrected using the Benjamini-Hochberg procedure(Benjamini and Hochberg, 1995) and the significance threshold was set at 0.05 (Supplementary File S-5).

2.6 Implementation

The itRWR algorithm is available as a Python package at `https://github.com/anthbapt/itRWR`. The pipeline developed for this study, including the identification of communities with itRWR, the clustering of communities, and the cluster enrichment analyses, is available on GitHub at `https://github.com/CecileBeust/PA_Communities`.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300210; this version posted December 19, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

3 Results

3.1 Selecting premature aging diseases and their associated genes

To systematically identify Premature Aging (PA) diseases, we selected the HPO annotation "*HP:0007495: prematurely aged appearance*" and its 12 descendant phenotype annotations in the ontology tree (Figure 1a, Supplementary Table S-1). We then selected, from ORPHANET, the diseases associated with at least one of these HPO phenotypes and associated with at least one gene. Overall, this led to the identification of 67 PA diseases associated with a total of 132 genes, ranging from one to 20 associated genes per PA disease (Supplementary Table S-2). We observed that the majority of the genes are mutated in only one disease (Figure 1b).

The selected diseases include prototypical PA diseases such as Hutchinson-Gilford Progeria Syndrome (HGPS, ORPHANET code: 740), characterized by osteoporosis, lipodystrophy and atherosclerosis (Navarro et al., 2006). HGPS is caused by specific mutations in the *LMNA* or *ZMPSTE24* genes. These mutations lead to the production of misprocessed Prelamin A isoforms and a loss of the integrity of the nuclear lamina. This loss of integrity impairs several major cellular functions, such as DNA repair, cell cycle, transcription, splicing, translation, or apoptosis (Carrero et al., 2016). Another example of selected PA disease is Ataxia-telangiectasia (ORPHANET code: 100), which is characterized by cerebellar degeneration, telangiectasia (dilatation of small surface blood vessels), immunodeficiency, cancer susceptibility and radiation sensitivity (Rothblum-Oviatt et al., 2016). It is caused by mutations in the *ATM* gene, a DNA damage sensor. Ataxia Telangiectasia belongs to the DNA repair deficient syndromes, which were among the firsts PA diseases to be deciphered. A last example is the Ehlers-Danlos syndrome (ORPHANET code: 287). It is characterized by joint hyper-mobility, skin hyper-elasticity and generalized fragility of soft tissue (Scicluna et al., 2022). It can be caused by mutations in several collagen genes (*COL1A1, COL5A1, COL5A2*) impacting the extracellular matrix organization. Hence, these three diseases described here as examples all share a PA phenotype, but are associated with the perturbation of initially distinct molecular mechanisms.

In order to study the cellular functions associated with PA diseases, we first performed a functional enrichment analysis of the 132 genes associated with the 67 PA diseases (Material and Methods). Most of the significant annotations are related with nuclear processes (Supplementary File S-1). We hypothesized that focusing solely on the genes directly linked to PA diseases does not allow revealing a comprehensive landscape of deregulated cellular functions. We hence decided to expand our analysis to study the biological network neighbourhood of the disease-associated genes.

3.2 Extracting PA disease communities from a biological multilayer network with iterative Random Walk with Restart

To elucidate the cellular functions underlying PA diseases, we systematically extracted the PA disease communities from a multilayer biological network. To this goal, we developed and applied an iterative Random Walk with Restart algorithm (Materials and Methods). The itRWR algorithm explores a multiplex network starting from nodes of interest, called the seeds. For each PA disease, we designated its associated gene(s) as the seed node(s), and executed the itRWR algorithm (Figure 1c). Following this approach, we obtained 67 disease communities (Supplementary File S-2).

Each community, gathering the multiplex network neighbourhood of the disease-associated gene(s), provides information into the cellular functions that might be disrupted in the disease. For instance, the community obtained for Hutchinson-Gilford Progeria Syndrome (HGPS) contains 102 gene nodes (two seed nodes and 100 nodes retrieved by itRWR, Supplementary File S-2). Its enrichment analyses depict a significant over-representation of genes involved in DNA repair and other nuclear processes (Material and Methods, Supplementary File S-7). The community identified for Ataxia-telangiectasia, using its mutated gene *ATM* as seed in the itRWR, contains 101 gene nodes. This disease community displays similar functional enrichments as the HGPS community (Supplementary File S-7). Finally, the community obtained for the Classical Ehlers-Danlos syndrome contains 103 gene nodes. In this case,

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300210; this version posted December 19, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

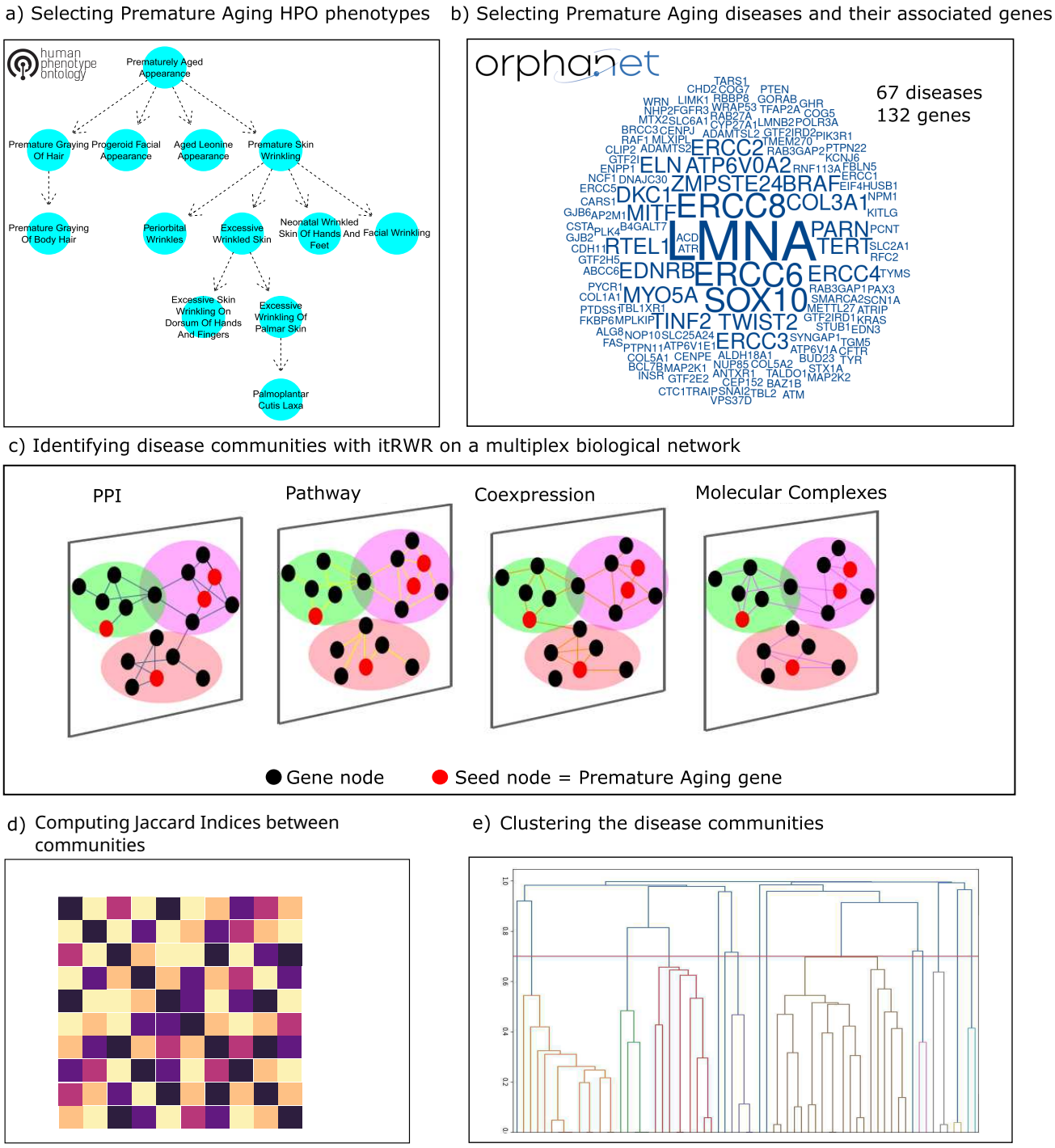

Figure 1: Overview of the workflow developed to systematically identify the cellular functions associated with premature aging (PA) diseases. a) 13 PA phenotypes are selected among HPO phenotypes (Köhler et al., 2017). b) Based on the 13 PA phenotypes, 67 PA diseases and their 132 associated genes are selected from ORPHANET. The size of the gene name in the word cloud corresponds to the number of diseases in which the gene is mutated.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300210; this version posted December 19, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Figure 1: c) Disease community identification with iterative Random Walk with Restart (itRWR). For each disease independently, the disease-associated gene(s) are used as seed(s) in the itRWR algorithm. itRWR is applied to a multilayer network composed of four layers: protein-protein interactions, pathways, molecular complexes and co-expression. In the figure, for the sake of visualisation, only three toy disease communities are represented (in pink, salmon, and green). d) Jaccard Indices are computed between the gene sets of each pair of disease communities. e) A clustering algorithm is applied on the Jaccard Index matrix to identify clusters of disease communities.

the results reveal a significant over-representation of genes involved in extracellular signaling pathways as well as cell adhesion and cell migration functions (Supplementary File S-7).

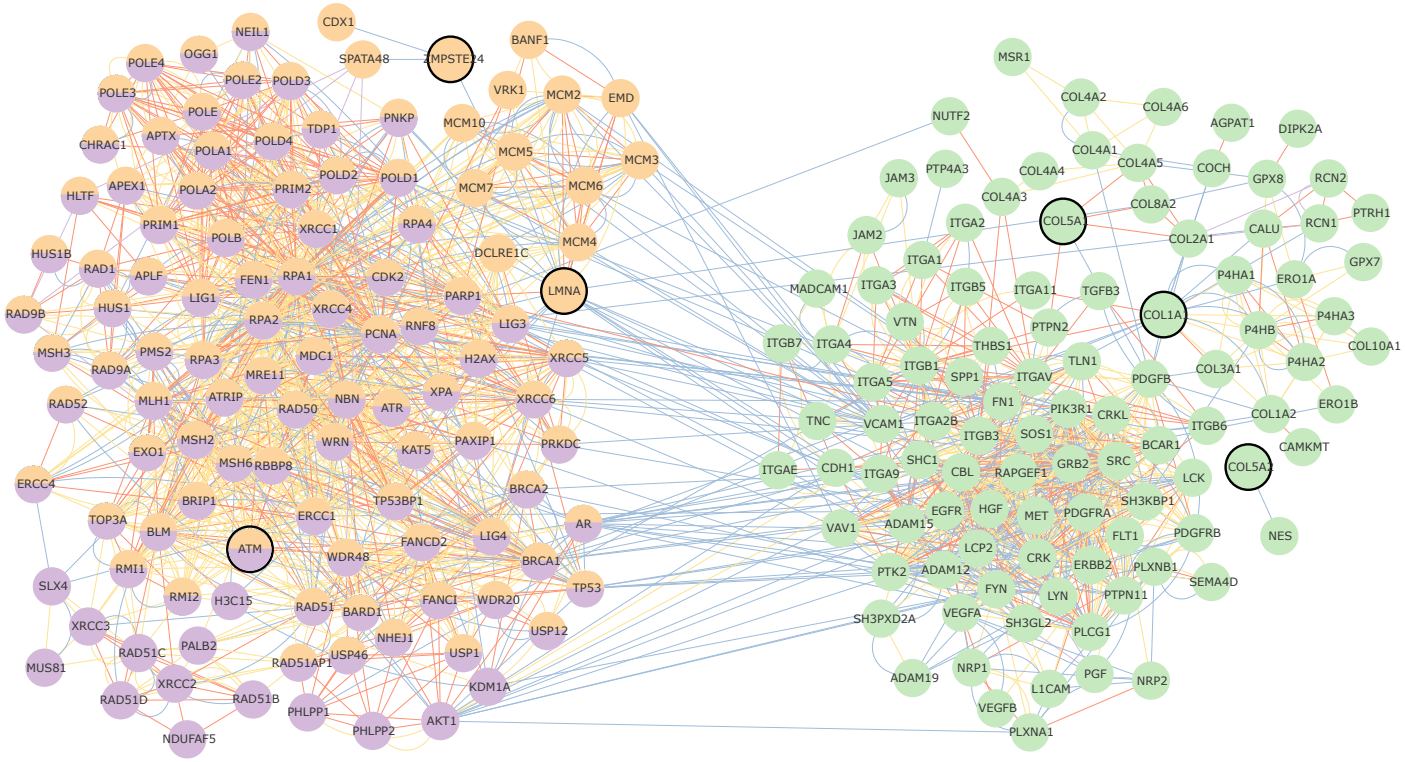

Figure 2: Disease communities of Hutchinson-Gilford Progeria Syndrome (in orange, 102 nodes, seeds: *LMNA* and *ZMPSTE24*), Ataxia-telangiectasia (in violet, 101 nodes, seed: *ATM*), and Classical Ehlers-Danlos syndrome (in green, 103 nodes, seeds: *COL1A1*, *COL5A2*, *COL5A1*). The communities were identified using itRWR on a multilayer network composed of four layers of biological interactions. The seed node(s) used for each disease are circled in black. Edge colors correspond to the interaction layer of the multilayer network: Protein-protein interactions (blue), pathways (yellow), molecular complexes (orange), co-expression (purple).

To further explore whether Ataxia-telangiectasia, HGPS, and Classical Ehlers-Danlos syndrome share perturbations of similar cellular functions, we compared the gene nodes belonging to their respective disease communities (Figure 2). We observed a high degree of overlap between the HGPS and Ataxia-telangiectasia communities, with 87 common genes, over 102 and 101 community gene nodes for HGPS and Ataxia-telangiectasia, respectively. These common gene nodes predominantly encode proteins related to DNA repair and response to DNA damage, such as

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300210; this version posted December 19, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

proteins of the DNA non-homologous end joining complex (e.g., *XRCC4-6*).

A small number of nodes are retrieved only in one of the two disease communities. For example, the phosphatases *PHLPP1, PHLPP2* are found only in the Ataxia-telangiectasia community. Contrarily, different proteins of the *MCM* family, which are involved in DNA replication, as well as proteins involved in the nuclear envelope architecture, such as *EMD*, are found only on the HGPS community (Figure 2).

Interestingly, the gene nodes of these two disease communities show no overlap with those of the Ehlers-Danlos syndrome community (Figure 2), indicating a markedly distinct node composition. The Ehlers-Danlos syndrome community mainly contains genes encoding proteins involved in tyrosine-kinase signaling pathways, e.g. *SRC* or *ERBB2*. Some of them are growth factors, like *EGFR*, which is known to be involved in cell proliferation. The community also contains several collagen proteins. Collagens are known to act as ligands of several receptors including the tyrosine-kinase receptors *DDR1* and *DDR2* (Boraschi-Diaz et al., 2017).

Overall, the three diseases used here as example seem to be associated with the perturbation of two different cellular functions, demonstrated by the grouping of the disease communities into two distinct subnetworks. In the next section, we extended the comparison to systematically assess the overlap between the disease communities extracted for the 67 PA diseases.

3.3 Clustering disease communities defines the molecular landscape of PA diseases

We hypothesized that PA disease communities with a large number of overlapping gene nodes might reveal shared disease-associated molecular mechanisms. We hence systematically compared the overlaps between the communities associated with the 67 PA diseases. To this goal, we computed the Jaccard Index between the gene sets composing each pair of disease communities (Figure 1d,e). The resulting matrix was then used to build a clustermap (Figure 3, Materials and Methods). After filtering out the clusters composed of less than three disease communities, we obtained six clusters containing between three to 20 disease communities (see Figure 3). Importantly, the obtained clusters are robust to changes in the number of iterations used in the itRWR algorithm to identify the disease communities. Indeed, the set of clusters obtained using 100 iterations in itRWR clustering is similar to the sets of clusters obtained using 50 iteration and 150 iteration (Rand indices equal to 0.74 and 0.96, respectively, Material and Methods, Supplementary Figure S1 and S2).

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300210; this version posted December 19, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

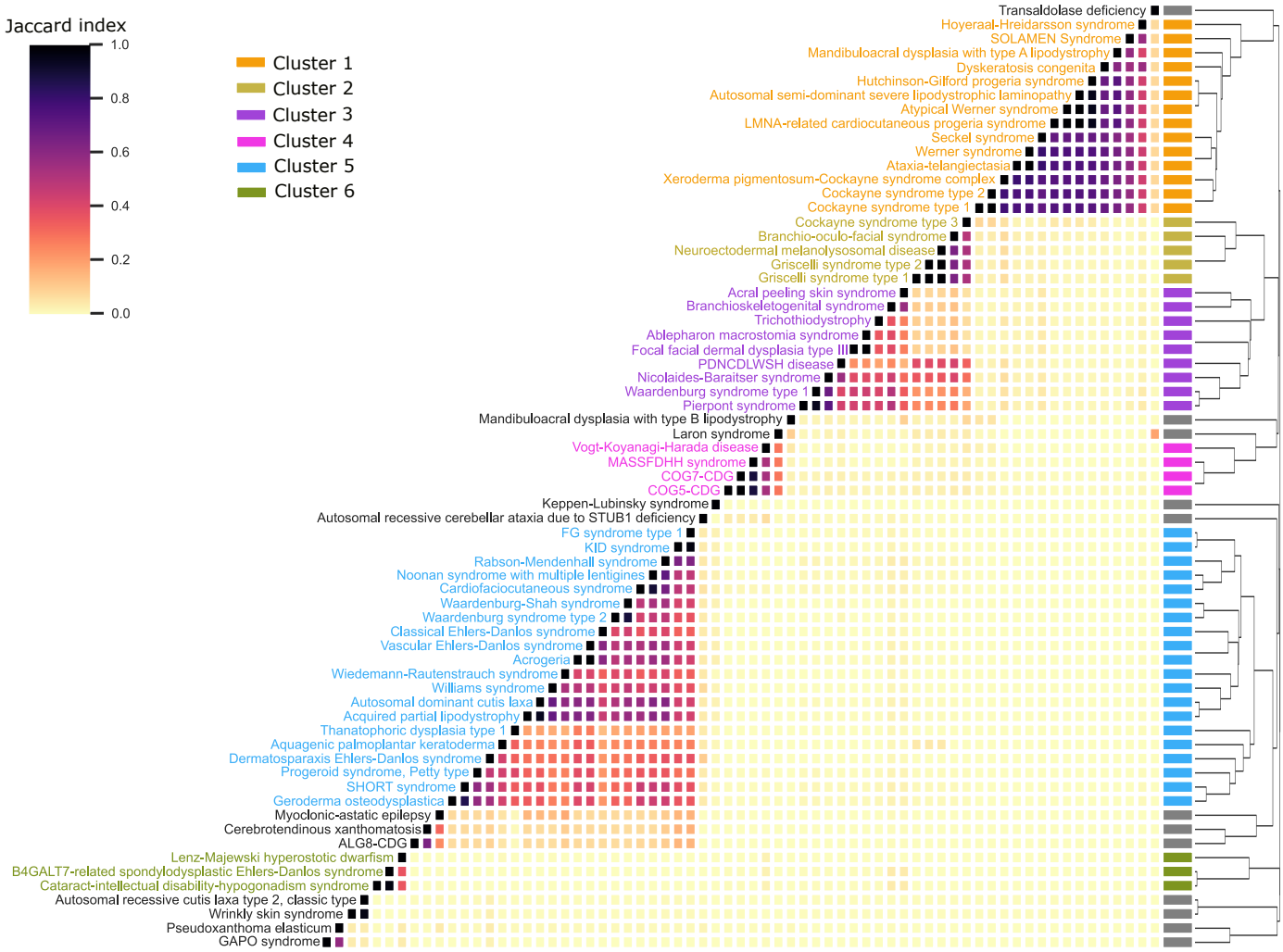

Figure 3: Clustering of the premature aging disease communities. Pairwise Jaccard indices were computed between the gene sets associated with each pair of disease communities. Then, the resulting matrix was used as input for hierarchical clustering. In the clustermap, the color gradient represents the similarities between communities (darker cells correspond to closer communities according to their Jaccard Indices). The clustering uses the average method and outputs a dendrogram. Clusters were determined by setting a cutoff value at 0.7 on the dendrogram. We kept the six clusters that contained at least three disease for downstream analyses. MASSFDHH: Moyamoya angiopathy-short stature-facial dysmorphism-hypergonadotropic hypogonadism syndrome. PDNCDLWSH: Peripheral demyelinating neuropathy-central dysmyelinating leukodystrophy-Waardenburg syndrome-Hirschsprung disease. SOLAMEN syndrome: Segmental outgrowth-lipomatosis-arteriovenous malformation-epidermal nevus syndrome

We next assessed the enrichments in cellular localizations, biological processes, and pathways, in each cluster, using the annotations of the genes within these clusters (Materials and Methods). At a coarse-grained level, we observed that the clusters 1, 2,& 3 have similar annotation profiles for cellular localization (Figure 4a). Indeed, the genes composing these three clusters are significantly enriched in nucleus, nucleoplasm and chromosome localizations (Supplementary File S-8). This nuclear localization is also reflected at the level of enriched pathways and processes, which are overall related to mitosis, cell cycle, gene expression and metabolic processes (Figure 4b, c). In contrast,

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300210; this version posted December 19, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

the clusters 4, 5, & 6 are enriched in non-nuclear localization, including membrane, cytoplasm, and/or cell periphery localization (Figure 4a). This is also reflected by the enrichments in signal transduction and cell communication identified in these clusters (Figure 4b, c, Supplementary File S-8). Furthermore, we examined the cluster enrichments in phenotypes, using the HPO annotations of the diseases belonging to each cluster (Materials and Methods, Supplementary File S-9). It should be noted that the HPO phenotype "prematurely aged appearance" is significant in all clusters. This is in line with the initial criteria used to select PA diseases. Indeed, PA diseases were selected as associated with this specific HPO term or any of its descendent terms. Our analysis did not identify any other distinct patterns in phenotype enrichments (Figure 4d).

We next focused more deeply on the annotation enrichments of each cluster (Supplementary Files 8-9). Cluster 1 gathers 14 diseases and a total of 232 genes. As stated previously, the genes belonging to this cluster have significant enrichments in nuclear and chromosome localization annotations, similar to Clusters 2 & 3. However, Cluster 1 displays an additional very strong enrichment in telomeric regions of the chromosomes. At the biological process and pathways levels, this is reflected by significant enrichments in telomere and chromosome organization processes. The most significant cellular functions of Cluster 1 are DNA metabolic process and DNA repair. The HPO phenotype enrichment analysis of the diseases belonging to Cluster 1 reveals skin and hair abnormalities, bone and cardiovascular defects, lipodystrophy, and growth delay. The second cluster contains five diseases and a total of 156 genes. At the functional level, its most over-represented annotations are related to mitotic G1 phase and G1/S transition. A significant HPO enrichments is *Hypopigmentation of hair*. Cluster 3 contains nine diseases and a total of 282 genes. It displays similar nuclear localization enrichments as Clusters 1 & 2 but with an additional significant enrichment in genes annotated for the transcriptor regulator complex. The most significant functional enrichments are related to transcription regulation. Significantly enriched Reactome pathways point to signal transduction, including signaling by nuclear receptor, Notch and TGFB. HPO enrichments exhibit phenotypes also retrieved in other Clusters, such as hair and skin abnormalities, and some specific phenotypes, for instance ocular lacrimation abnormalities. In summary, despite their shared enrichments in nuclear localization, biological processes and pathways, the genes belonging to Clusters 1, 2, & 3 exhibit subtle differences in their cellular function enrichments.

Clusters 4, 5, 6 display enrichment patterns different from the first three clusters (Supplementary File S-8-9). Cluster 4 is composed of four diseases and 138 genes. The genes belonging to Cluster 4 are significantly annotated with cytoplasmic localizations. At the level of GO Biological Processes, the enriched annotations are mainly related to signal transduction, inflammation, and immune system processes, with the strongest enrichments being in the I-κB Kinase/NF-κB signaling. The Reactome pathways enriched in this cluster are Toll-like Receptor cascades, death signaling and immune system related pathways. Cluster 5 is the largest cluster. It contains 20 diseases and 466 genes. The cluster genes are enriched in membrane and cell periphery localizations, but also displays significant enrichments in anchoring junction and extracellular region. These subcellular localizations are reflected at the level of biological process and pathway that reveal enrichments in cell adhesion and migration, extracellular matrix organization and integrin cell surface interactions, and signaling with receptor tyrosine kinases. At the phenotype level, many significant HPO terms are the same as the terms significant in Cluster 1, such as hair and skin abnormalities, and lipodystrophy. But additional phenotypes, such as joint mobility defects, are retrieved only in the Cluster 5. Finally, the last cluster, Cluster 6, contains only three diseases and 150 genes. These genes are enriched in localization annotations related to endosome and vesicle membrane, reflected in processes and pathways enrichments in vesicle-mediated transport and membrane trafficking.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300210; this version posted December 19, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

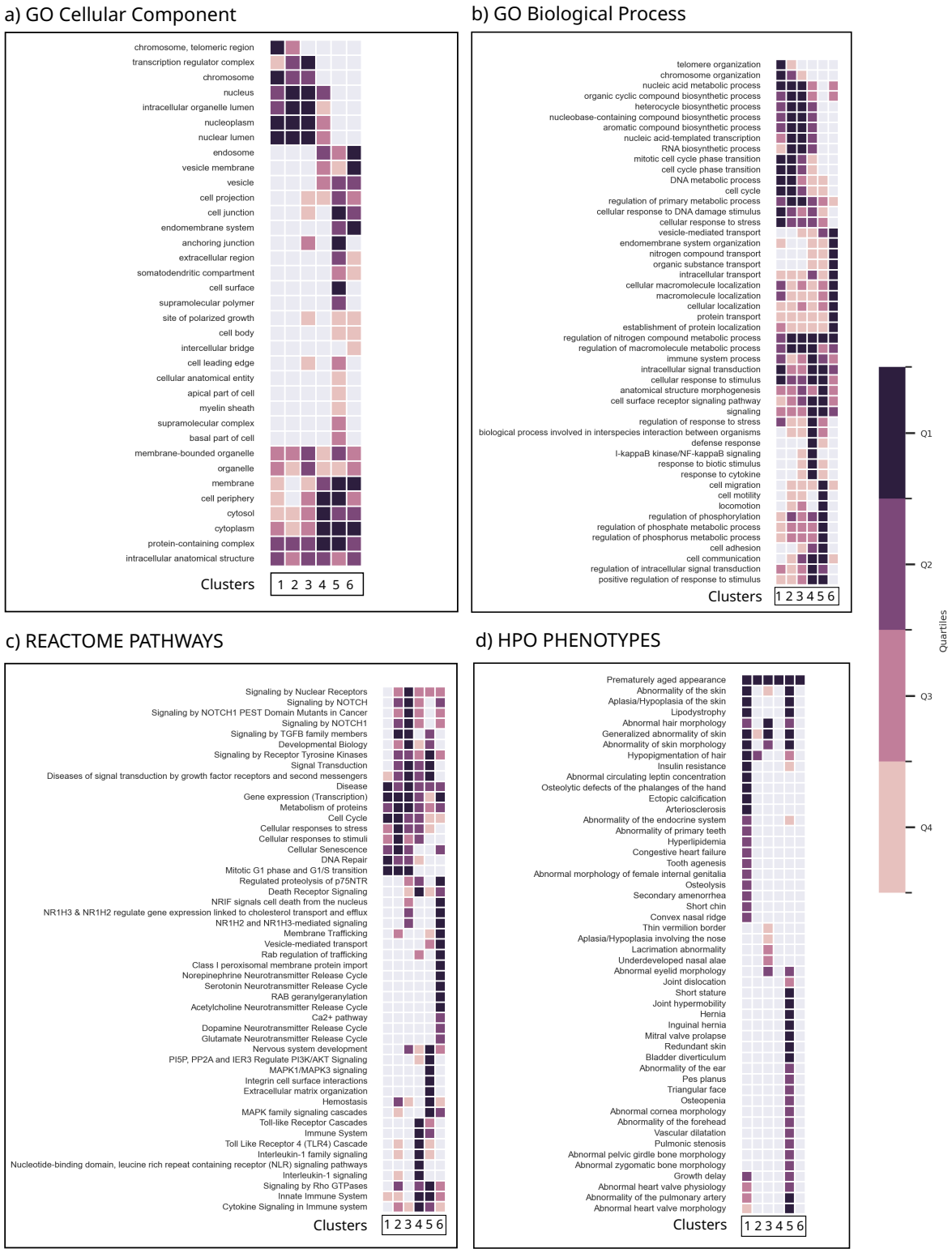

Figure 4

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300210; this version posted December 19, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Figure 4: Cluster enrichments in (a) Cellular Components , (b) GO Biological Processes, (c) Reactome pathways, and (d) HPO phenotypes. For GO Biological Process, Cellular Components and Reactome pathways, enrichments were computed with g:Profiler using the cluster gene lists as inputs and filtered using orsum. For HPO, a hypergeometric test was applied using the cluster disease lists as input. In both cases, significant annotations were selected with a threshold set at 0.05 on the corrected p-values. The color gradient reflects the quartile of the annotation significance.

In summary, we observed that the Clusters 1, 2, & 3 display similarly enriched nuclear localizations and processes and pathways related to nuclear functioning. But they also display subtle differences in their cellular functions. Indeed, Cluster 1 is more significantly associated with DNA repair mechanisms, Cluster 2 with cell cycle, in particular mitosis G1/S phases, and Cluster 3 with transcription regulation. Of note, a large majority of the PA diseases considered as classical progeroid syndrome (Kubben and Misteli, 2017), such as Hutchinson-Gilford Progeria or Werner Syndromes, belong to Cluster 1. The three last clusters (Clusters 4, 5, & 6) are associated with different and non-nuclear molecular mechanisms. Cluster 4 gathers inflammation and immune system pathways; Cluster 5 is associated with cellular communication; Cluster 6 is related with vesicle-mediated transport. Overall, the six clusters hence reveal a landscape of the molecular mechanisms associated with PA diseases.

3.4 Linking the molecular landscape of PA diseases to Physiological Aging

In order to identify potential links between premature and physiological aging, we assessed the cluster enrichments in genes related to physiological aging. We first tested whether the clusters are enriched in physiological aging genes extracted from the GenAge database (Tacutu et al., 2018b) (Material and Methods). This analysis revealed that all clusters except cluster 6 are significantly enriched in GenAge genes (corrected p-values from 10^{-14} to 10^{-03}, Figure 5, Supplementary File S-5). Next, we tested the cluster enrichments in genes up and down-regulated during physiological aging in five different human tissues (blood, skin, brain, muscle and breast, Materials and Methods). We observed that Cluster 5 is enriched in genes up-regulated in skin during physiological aging (corrected p-value = 4.9×10^{-04}). This is consistent with the set of PA diseases gathered in Cluster 5, such as different types of Ehlers-Danlos syndromes or the Autosomal dominant cutis laxa, which are mainly characterized by skin abnormalities. In Clusters 4 & 5, we noticed a strong enrichment in genes up-regulated in the brain during physiological aging. Finally, Clusters 2 & 3 display significant enrichments in genes up-regulated in muscle and down-regulated in breast during physiological aging. Cluster 6 does not present any significant enrichment in genes associated with physiological aging, even if some of the cluster genes are deregulated, in particular down-regulated in brain tissue during physiological aging (Supplementary File S-5).

Discussion

In this study, we provide a detailed overview of the molecular mechanisms involved in PA diseases through a systematic and data-driven functional analysis of genes associated with these conditions. This molecular landscape expands the scope of the molecular mechanisms previously reviewed for PA diseases. Indeed, to the best of our knowledge, these molecular mechanisms were so far only studied from a literature review perspective. Recent literature reviews broadly classify PA diseases into three groups corresponding to defects in nuclear architecture, DNA repair, and other processes (Carrero et al., 2016; Grelet et al., 2019; Coppede, 2021). The data-driven molecular landscape we propose here groups the PA diseases into six clusters primarily annotated for DNA repair, cell cycle, transcription regulation, inflammation, cell communication, and vesicle-mediated transport.

Overall, the first three clusters exhibit a close functional relationship, as they are all associated with nuclear processes. Specifically, Cluster 1 predominantly pertains to DNA repair mechanisms, Cluster 2 is more significantly linked to cell cycle regulation, and Cluster 3 primarily involves transcription regulation. These mechanisms represent

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300210; this version posted December 19, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

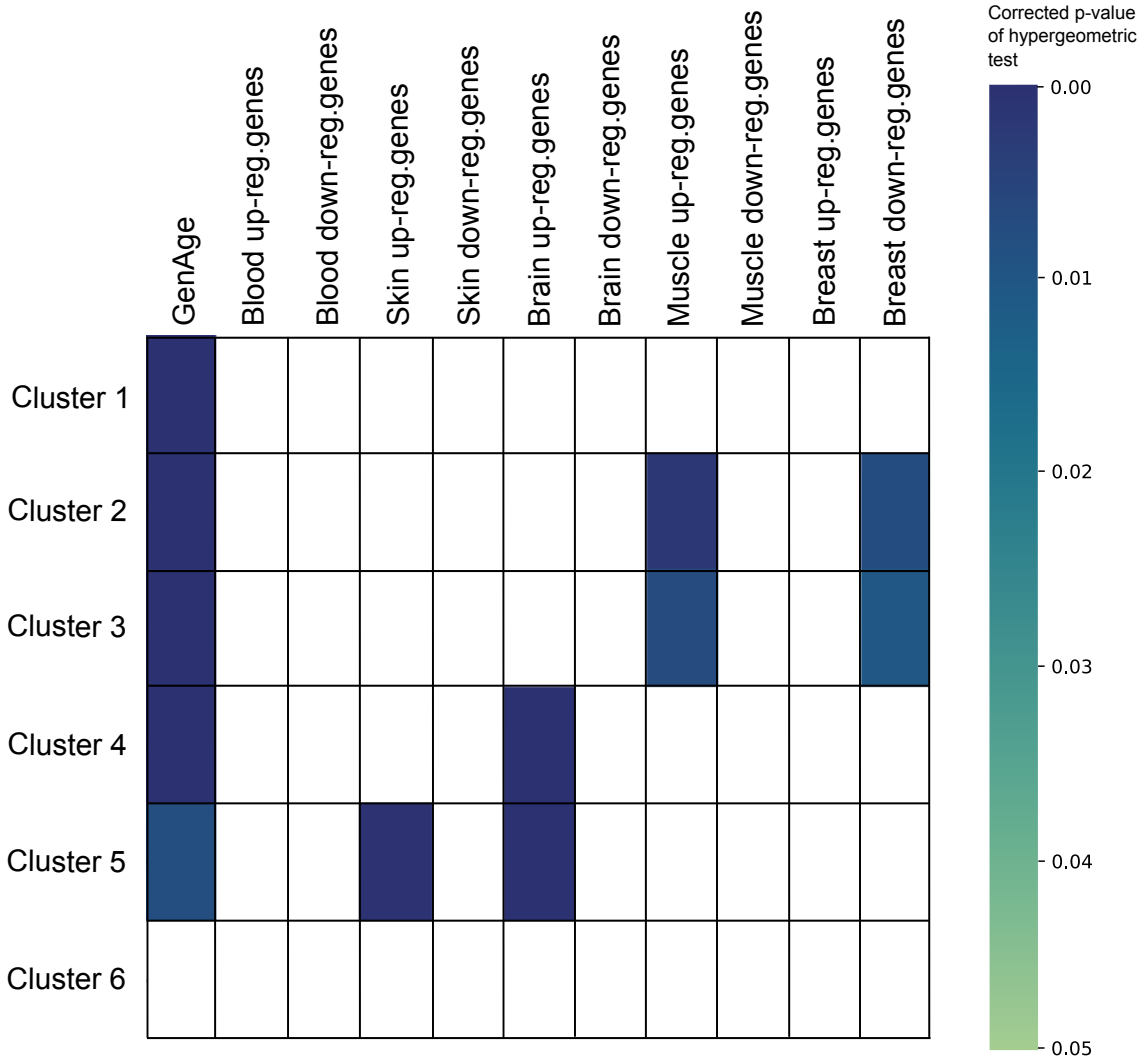

Figure 5: Cluster enrichments in physiological aging genes. The first column represents the enrichment results in physiological aging genes extracted from the GenAge database (Tacutu et al., 2018b). The other columns represent enrichment results of clusters in genes up and down-regulated during physiological aging in several tissues from the study of (Aramillo Irizar et al., 2018). The colors represent the p-value of the hypergeometric test.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300210; this version posted December 19, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

interconnected nuclear processes, which tend to become dysregulated in a cascade-like fashion when any one of them is impaired.

Regarding PA diseases associated with defects in the nuclear envelope architecture, typically caused by mutations in lamin A or proteins interacting with lamin A, our findings reveal their belonging to Cluster 1 rather than the formation of an isolated cluster. Cluster 1 most significant annotations pertain to DNA repair processes. The maintenance of nuclear envelope integrity is closely tied to DNA repair, as evidenced by the intra-nuclear accumulation of progerin (the truncated isoform of prelamin A in Hutchinson-Gilford Progeria Syndrome) leading to impaired repair of double-strand DNA breaks and delayed recruitment of ATM proteins at breakage sites (Cau et al., 2014). While existing evidence underscores such functional link between nuclear envelope architecture and DNA repair, we hypothesize that the discovery of novel interactions and the expansion of biological networks may lead to the division of Cluster 1 into two subclusters, potentially reflecting functional nuances among these PA diseases.

The remaining three clusters predominantly relate to non-nuclear cellular functions. Cluster 5 encompasses genes and proteins involved in cell communication and extracellular processes, including those implicated in Ehlers-Danlos syndromes, which affect connective tissues. Cluster 4, characterized by its association with inflammation, includes genes related to glycosylation disorders. Notably, Cluster 6, primarily annotated for vesicle-mediated transport, represents a cellular function that, to our knowledge, has not been reviewed in the context of PA diseases. This underscores a potentially under-explored area in PA research, suggesting new avenues for investigation.

Overall, our data-driven approach has allowed us to systematically classify PA diseases and broaden our understanding of their molecular perturbations. As discussed previously, the expansion of biological interaction networks is expected to further refine our clusters. Particularly, larger clusters such as Clusters 1 and 5, which currently encompass a broad range of PA diseases, may eventually be subdivided into more specific subclusters. Furthermore, we expect that certain diseases that are currently not part of any cluster due to the prerequisite of a minimum of three closely related diseases for cluster formation, could form new clusters as additional interaction data becomes available. We can also note that our methodology focused on clustering disease communities, thereby spotlighting the shared cellular functions among PA diseases. An alternative approach could be to examine disease-specific subnetworks. For instance, in Figure 2, we can pinpoint nodes unique to the HGPS disease community. The HGPS-specific nodes include *EMD*, which codes for lamin A-interacting protein Emerin, and various DNA replication factors. This kind of targeted analysis could yield deeper insights into the unique molecular pathways and interactions characteristic of each PA disease.

We systematically selected all the diseases associated with a PA phenotype using HPO (Köhler et al., 2021) and ORPHANET (Rath et al., 2012) databases. This approach gathered a wide spectrum of diseases with various clinical presentation and molecular pathogenesis. Our set of PA diseases include classical progeroid syndromes, such as Hutchinson-Gilford Progeria Syndrome, which are unambiguously considered as PA syndromes in the literature, but also other diseases which belonging to the category of progeroid syndrome or PA diseases might be debatable. For instance, the premature aging phenotype present in Griscelli syndrome patients is "premature graying of hair", and the classification of this syndrome as PA might be questioned (Hennekam, 2020). In addition, we also identified some questionable gene-disease associations. For instance, in our study, Acrogeria is grouped in Cluster 5 alongside Ehlers-Danlos PA diseases. This clustering stems from Acrogeria's association with mutations in *COL3A1* in ORPHANET. However, this indication does not consider various pieces of evidence linking Acrogeria to disease-causing variants in *LMNA* and *ZMPSTE24* (Hadj-Rabia et al., 2014; Maroofian et al., 2018). Factoring in these associations might shift Acrogeria into Cluster 1. Another example are Cockayne Syndromes that are split in clusters 1 and 2. Two main Cockayne Syndrome types exist, CSA and CSB, caused by disease-causing variants in *ERCC6* and *ERCC8*. However, databases such as ORPHANET and OMIM (OMI) have documented extremely rare phenotypes as Cockayne Syndromes related to other variants in *ERCC1* and *ERCC4*. Misdiagnoses, whether clinical or molecular, are a possibility that can introduce errors into database entries. The algorithm we developed in this study incorporates these data, although they may not be entirely relevant, potentially affecting the accuracy of our findings. Overall, an alternative to our systematic data-driven strategy could have been to select a set of diseases based on literature reviews or based on expert knowledge. However, in order to ensure unbiased and reproducible studies, we preferred

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300210; this version posted December 19, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

not to filter *a priori* the diseases.

Our data-driven classification could lead to a deeper understanding of PA diseases, with compelling implications for diagnosis and treatment. Firstly, unraveling the molecular underpinnings of these syndromes can improve diagnosis and pave the way for targeted therapies. More specifically, multilayer interaction data could aid in identifying previously unknown biological connections and crucial disease biomarkers, essential for clinical trials but often missing in PA and rare diseases. Secondly, categorizing these diseases into groups with shared cellular dysfunction can open up new avenues for treatment strategies. For instance, therapeutic molecules targeting different diseases belonging to the same cluster could be developed once the common pathophysiological mechanism has been validated. Moreover, innovative therapies, akin to those for rare cancers, might be adapted for non-oncological rare diseases. Preclinical *in vitro* and *in vivo* studies confirming the potential efficacy/safety ratio of a drug and its relevance in multiple indications, might lead to include patients being affected with different phenotypes in early clinical phases. Such a "basket trial" could expedite clinical timelines and be cost-effective (Zanello et al., 2023). Laslty, a better understanding of PA disease etiology and pathogenesis can shed light on the molecular mechanisms involved physiological aging (Hennekam, 2020; Carrero et al., 2016).

Aging is a multifaceted process governed by numerous molecular mechanisms, many of which are yet to be fully understood. PA diseases are an important source of information for studying the molecular mechanisms that could also be perturbed during physiological aging (Carrero et al., 2016). In recent years, knowledge of the links between PA diseases and physiological aging has been increasing. Our approach, which leverages different sources of information to delineate the molecular landscape of PA and connect it to the known and deregulated physiological aging genes, is interesting for hypothesis generation. For instance, our observation that Cluster 5 is enriched in genes up-regulated in physiological skin aging is relevant, as the main phenotypic pattern of PA diseases included in this cluster is a skin defect. Likewise, the association of Cluster 4 with genes that are up-regulated in brain tissue during aging may underscore the role of inflammation as a primary and secondary mechanism in neurological aging.

Our PA molecular landscape was defined based on the exploration of a multilayer interaction network. The study of biological networks in pathological contexts is at the heart of network biomedicine (Zitnik et al., 2023). Network medicine assumption is that molecular components associated with similar diseases are belonging to a close network vicinity (Vidal et al., 2011). Biological interactions have been demonstrated as important features to predict aging-related genes (Fabris et al., 2017, 2016; Freitas et al., 2011; Kerepesi et al., 2018), and network medicine approaches have been successfully used to study aging-related genes and molecular mechanisms (Li et al., 2021; Li and Milenkovic, 2022; Faisal and Milenkovifá, 2014; Newaz and Milenkovic, 2022).

In order to leverage as much as possible the interaction information available in the literature, we consider jointly multiple interaction sources gathered in a multilayer network. Considering multilayer networks is a pertinent way to reduce the incompleteness of individual interaction sources (Hammoud and Kramer, 2020). Multilayer networks represent a formidable source of information to study cellular functioning and identify disease-associated perturbations, but they are large and complex and appropriate exploration tools are mandatory. Community identification is widely use to identify subnetworks of genes or proteins involved in common cellular processes (Choobdar et al., 2019). Different algorithms exist for community identification in multiplex networks (MagnaniMatteo et al., 2021). For instance, the ACLcut algorithm is a local community identification approach based on RWR (Jeub et al., 2017). We proposed here a novel approach called iterative Random Walk with Restart. This novel approach is based on our previous work extending the Random Walk with Restart algorithm to explore multilayer networks (Valdeolivas et al., 2018; Baptista et al., 2022). We decided to set a maximal number of iterations to obtain disease communities of similar sizes. However, future version of the algorithms could be implemented with a different stopping criteria, for instance by optimizing measures such as the conductance measure in the ACLcut algorithm (Jeub et al., 2017). Importantly, our itRWR and protocol can be reused to evaluate and compare the molecular bases of any set of diseases.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300210; this version posted December 19, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Conflict of Interest

AV is employee and shareholder of F. Hoffmann-La Roche Ltd. NL is shareholder of ProGeLife and its subsidiary Calysens. NL is employee (Chief Scientist Rare Disease) at the insternational research institute Servier.

Acknowledgments

This work received support from the French National Research Agency (ANR-21-CE45-000101), the Association Française contre les Myopathies (AFM), the Excellence Initiative of Aix-Marseille University - A*Midex, a French "Investissements d'Avenir" programme - Institute MarMaRa AMX-19-IET-007, and the European Union's Horizon 2020 research and innovation programme under the EJP RD COFUND-EJP No 825575. GPT-4 was used to improve syntax and readability.

Data and code availability

- Data:
  - Supplementary Materials: `https://zenodo.org/records/10400905`
  - Networks: `https://www.ndexbio.org/index.html#/search?searchType=All&searchString=cecile.beust&searchTermExpansion=false`
- Code:
  - Community identification in multilayer networks: itRWR Python Package `https://github.com/anthbapt/itRWR`
  - Complete analysis workflow: Community identification, clustering, and enrichment analyses `https://github.com/CecileBeust/PA_Communities.git`

Supplementary Materials

Supplementary Figures:
Supplementary Figure S1: Clustering of diseases with 50 iterations of itRWR
Supplementary Figure S2: Clustering of diseases with 150 iterations of itRWR

Supplementary Tables:
Supplementary Table S1:Premature Aging HPO phenotypes
Supplementary Table S2: the 67 PA diseases and their 132 associated genes from ORPHANET
Supplementary Table S3: Number of nodes, edges, and densities of the 4 network layers composing the multiplex biological network
Supplementary Table S4: Parameters used for MultiXrank

Supplementary Files:
Supplementary File S1: Csv file containing the enrichment analysis results computed using the genes associated with physiological aging
Supplementary File S2: Excel file containing the gene nodes belonging to each of the 67 communities
Supplementary File S3: Excel file containing the gene nodes belonging to each cluster (i.e., the union of the genes belonging to the set of communities composing the cluster)

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300210; this version posted December 19, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Supplementary File S4: Csv file containing the diseases belonging to each cluster
Supplementary File S5: Excel file containing the enrichment of clusters using lists of physiological aging genes
Supplementary File S6: Csv file containing the lists of genes differentially expressed in human blood, skin, brain, muscle and breast during aging, from the study of Irizar et al.
Supplementary File S7: Excel file containing the enrichment analysis results of the 67 communities, using GO Biological Processes and Cellular Components, and Reactome pathways functional annotations
Supplementary File S8: Excel file containing the enrichment analysis results of the 6 clusters, using GO Biological Processes and Cellular Components, and Reactome pathways functional annotations. The last 2 columns contain the seed nodes of the cluster annotated for the corresponding function and their corresponding diseases
Supplementary File S9: Excel file containing enrichment analysis results of the 6 clusters, using HPO phenotypes

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300210; this version posted December 19, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

References

Dido Carrero, Clara Soria-Valles, and Carlos López-Otín. Hallmarks of progeroid syndromes: lessons from mice and reprogrammed cells. *Disease Models & Mechanisms*, 9(7):719–735, 2016. ISSN 1754-8403. doi: 10.1242/dmm.024711. URL `http://dmm.biologists.org/lookup/doi/10.1242/dmm.024711`.

Franziska Schnabel, Uwe Kornak, and Bernd Wollnik. Premature aging disorders: A clinical and genetic compendium. *Clinical Genetics*, 99(1):3–28, January 2021. ISSN 0009-9163, 1399-0004. doi: 10.1111/cge.13837. URL `https://onlinelibrary.wiley.com/doi/10.1111/cge.13837`.

Davor Lessel and Christian Kubisch. Hereditary Syndromes With Signs of Premature Aging. *Deutsches Ärzteblatt international*, July 2019. ISSN 1866-0452. doi: 10.3238/arztebl.2019.0489. URL `https://www.aerzteblatt.de/10.3238/arztebl.2019.0489`.

Claire L. Navarro, Pierre Cau, and Nicolas Lévy. Molecular bases of progeroid syndromes. *Human Molecular Genetics*, 15(SUPPL. 2):151–161, 2006. ISSN 09646906. doi: 10.1093/hmg/ddl214.

Maude Grelet, Véronique Blanck, Sabine Sigaudy, Nicole Philip, Fabienne Giuliano, Khaoula Khachnaoui, Godelieve Morel, Sarah Grotto, Julia Sophie, Céline Poirsier, James Lespinasse, Laurent Alric, Patrick Calvas, Gihane Chalhoub, Valérie Layet, Arnaud Molin, Cindy Colson, Luisa Marsili, Patrick Edery, Nicolas Lévy, and Annachiara De Sandre-Giovannoli. Outcomes of 4 years of molecular genetic diagnosis on a panel of genes involved in premature aging syndromes, including laminopathies and related disorders. *Orphanet Journal of Rare Diseases*, 14(1):288, December 2019. ISSN 1750-1172. doi: 10.1186/s13023-019-1189-z. URL `https://ojrd.biomedcentral.com/articles/10.1186/s13023-019-1189-z`.

Fabio Coppede. Mutations Involved in Premature-Ageing Syndromes. *The Application of Clinical Genetics*, Volume 14:279–295, June 2021. ISSN 1178-704X. doi: 10.2147/TACG.S273525. URL `https://www.dovepress.com/mutations-involved-in-premature-ageing-syndromes-peer-reviewed-fulltext-article-TACG`.

Quan Zhong, Nicolas Simonis, Qian-ru Li, Benoit Charloteaux, Fabien Heuze, Niels Klitgord, Stanley Tam, David Szeto, Chenwei Lin, Tong Hao, Changyu Fan, Stuart Milstein, Denis Dupuy, Robert Brasseur, David E Hill, Michael E Cusick, and Marc Vidal. Edgetic perturbation models of human inherited disorders. *Molecular Systems Biology*, 5(321), 2009. doi: 10.1038/msb.2009.80.

Leland H. Hartwell, John J. Hopfield, Stanislas Leibler, and Andrew W. Murray. From molecular to modular cell biology. *Nature*, 402(6761supp):C47–C52, 1999. ISSN 0028-0836. doi: 10.1038/35011540. URL `http://www.nature.com/doifinder/10.1038/35011540`.

Laura I. Furlong. Human diseases through the lens of network biology. *Trends in Genetics*, 29(3):150–159, Mar 2013. ISSN 0168-9525. doi: 10.1016/j.tig.2012.11.004. URL `https://doi.org/10.1016/j.tig.2012.11.004`.

Sarvenaz Choobdar, Mehmet E. Ahsen, Jake Crawford, Mattia Tomasoni, Tao Fang, David Lamparter, Junyuan Lin, Benjamin Hescott, Xiaozhe Hu, Johnathan Mercer, Ted Natoli, Rajiv Narayan, Aravind Subramanian, Jitao D. Zhang, Gustavo Stolovitzky, Zoltán Kutalik, Kasper Lage, Donna K. Slonim, Julio Saez-Rodriguez, Lenore J. Cowen, Roberto Bellotti, Sven Bergmann, and Daniel Marbach. Assessment of network module identification across complex diseases. *Nature Methods*, 16(9):843–852, September 2019. ISSN 1548-7105. doi: 10.1038/s41592-019-0509-5. URL `https://www.nature.com/articles/s41592-019-0509-5`. Number: 9 Publisher: Nature Publishing Group.

MagnaniMatteo, HanteerObaida, InterdonatoRoberto, RossiLuca, and TagarelliAndrea. Community Detection in Multiplex Networks. *ACM Computing Surveys (CSUR)*, May 2021. doi: 10.1145/3444688. URL `https://dl.acm.org/doi/10.1145/3444688`. Publisher: ACM PUB27 New York, NY, USA.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300210; this version posted December 19, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Gilles Didier, Christine Brun, and Anaïs Baudot. Identifying communities from multiplex biological networks. *PeerJ*, 3(0):e1525, 2015. ISSN 2167-8359. doi: 10.7717/peerj.1525. URL `https://peerj.com/articles/1525`.

Gilles Didier, Alberto Valdeolivas, and Anaïs Baudot. Identifying communities from multiplex biological networks by randomized optimization of modularity. *F1000Research*, 7(0):1042, 2018. ISSN 2046-1402. doi: 10.12688/f1000research.15486.1. URL `https://f1000research.com/articles/7-1042/v1`.

Sebastian Köhler, Nicole A. Vasilevsky, Mark Engelstad, Erin Foster, Julie McMurry, Ségolène Aymé, Gareth Baynam, Susan M. Bello, Cornelius F. Boerkoel, Kym M. Boycott, Michael Brudno, Orion J. Buske, Patrick F. Chinnery, Valentina Cipriani, Laureen E. Connell, Hugh J.S. Dawkins, Laura E. DeMare, Andrew D. Devereau, Bert B.A. De Vries, Helen V. Firth, Kathleen Freson, Daniel Greene, Ada Hamosh, Ingo Helbig, Courtney Hum, Johanna A. Jähn, Roger James, Roland Krause, Stanley J.F. Laulederkind, Hanns Lochmüller, Gholson J. Lyon, Soichi Ogishima, Annie Olry, Willem H. Ouwehand, Nikolas Pontikos, Ana Rath, Franz Schaefer, Richard H. Scott, Michael Segal, Panagiotis I. Sergouniotis, Richard Sever, Cynthia L. Smith, Volker Straub, Rachel Thompson, Catherine Turner, Ernest Turro, Marijcke W.M. Veltman, Tom Vulliamy, Jing Yu, Julie Von Ziegenweidt, Andreas Zankl, Stephan Züchner, Tomasz Zemojtel, Julius O.B. Jacobsen, Tudor Groza, Damian Smedley, Christopher J. Mungall, Melissa Haendel, and Peter N. Robinson. The human phenotype ontology in 2017. *Nucleic Acids Research*, 45(D1):D865–D876, 2017. ISSN 13624962. doi: 10.1093/nar/gkw1039.

Ana Rath, Annie Olry, Ferdinand Dhombres, Maja Miličić Brandt, Bruno Urbero, and Segolene Ayme. Representation of rare diseases in health information systems: The orphanet approach to serve a wide range of end users. *Human Mutation*, 33(5):803–808, 2012. ISSN 10597794. doi: 10.1002/humu.22078.

The Gene Ontology Consortium, Suzi A Aleksander, James Balhoff, Seth Carbon, J Michael Cherry, Harold J Drabkin, Dustin Ebert, Marc Feuermann, Pascale Gaudet, Nomi L Harris, David P Hill, Raymond Lee, Huaiyu Mi, Sierra Moxon, Christopher J Mungall, Anushya Muruganugan, Tremayne Mushayahama, Paul W Sternberg, Paul D Thomas, Kimberly Van Auken, Jolene Ramsey, Deborah A Siegele, Rex L Chisholm, Petra Fey, Maria Cristina Aspromonte, Maria Victoria Nugnes, Federica Quaglia, Silvio Tosatto, Michelle Giglio, Suvarna Nadendla, Giulia Antonazzo, Helen Attrill, Gil dos Santos, Steven Marygold, Victor Strelets, Christopher J Tabone, Jim Thurmond, Pinglei Zhou, Saadullah H Ahmed, Praoparn Asanitthong, Diana Luna Buitrago, Meltem N Erdol, Matthew C Gage, Mohamed Ali Kadhum, Kan Yan Chloe Li, Miao Long, Aleksandra Michalak, Angeline Pesala, Armalya Pritazahra, Shirin C C Saverimuttu, Renzhi Su, Kate E Thurlow, Ruth C Lovering, Colin Logie, Snezhana Oliferenko, Judith Blake, Karen Christie, Lori Corbani, Mary E Dolan, Harold J Drabkin, David P Hill, Li Ni, Dmitry Sitnikov, Cynthia Smith, Alayne Cuzick, James Seager, Laurel Cooper, Justin Elser, Pankaj Jaiswal, Parul Gupta, Pankaj Jaiswal, Sushma Naithani, Manuel Lera-Ramirez, Kim Rutherford, Valerie Wood, Jeffrey L De Pons, Melinda R Dwinell, G Thomas Hayman, Mary L Kaldunski, Anne E Kwitek, Stanley J F Laulederkind, Marek A Tutaj, Mahima Vedi, Shur-Jen Wang, Peter D'Eustachio, Lucila Aimo, Kristian Axelsen, Alan Bridge, Nevila Hyka-Nouspikel, Anne Morgat, Suzi A Aleksander, J Michael Cherry, Stacia R Engel, Kalpana Karra, Stuart R Miyasato, Robert S Nash, Marek S Skrzypek, Shuai Weng, Edith D Wong, Erika Bakker, Tanya Z Berardini, Leonore Reiser, Andrea Auchincloss, Kristian Axelsen, Ghislaine Argoud-Puy, Marie-Claude Blatter, Emmanuel Boutet, Lionel Breuza, Alan Bridge, Cristina Casals-Casas, Elisabeth Coudert, Anne Estreicher, Maria Livia Famiglietti, Marc Feuermann, Arnaud Gos, Nadine Gruaz-Gumowski, Chantal Hulo, Nevila Hyka-Nouspikel, Florence Jungo, Philippe Le Mercier, Damien Lieberherr, Patrick Masson, Anne Morgat, Ivo Pedruzzi, Lucille Pourcel, Sylvain Poux, Catherine Rivoire, Shyamala Sundaram, Alex Bateman, Emily Bowler-Barnett, Hema Bye-A-Jee, Paul Denny, Alexandr Ignatchenko, Rizwan Ishtiaq, Antonia Lock, Yvonne Lussi, Michele Magrane, Maria J Martin, Sandra Orchard, Pedro Raposo, Elena Speretta, Nidhi Tyagi, Kate Warner, Rossana Zaru, Alexander D Diehl, Raymond Lee, Juancarlos Chan, Stavros Diamantakis, Daniela Raciti, Magdalena Zarowiecki, Malcolm Fisher, Christina James-Zorn, Virgilio Ponferrada, Aaron Zorn, Sridhar Ramachandran, Leyla Ruzicka, and Monte Westerfield. The Gene Ontology knowledgebase in 2023. *Genetics*, 224(1):iyad031, May 2023. ISSN 1943-2631. doi: 10.1093/genetics/iyad031. URL `https://doi.org/10.1093/genetics/iyad031`.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300210; this version posted December 19, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Marc Gillespie, Bijay Jassal, Ralf Stephan, Marija Milacic, Karen Rothfels, Andrea Senff-Ribeiro, Johannes Griss, Cristoffer Sevilla, Lisa Matthews, Chuqiao Gong, Chuan Deng, Thawfeek Varusai, Eliot Ragueneau, Yusra Haider, Bruce May, Veronica Shamovsky, Joel Weiser, Timothy Brunson, Nasim Sanati, Liam Beckman, Xiang Shao, Antonio Fabregat, Konstantinos Sidiropoulos, Julieth Murillo, Guilherme Viteri, Justin Cook, Solomon Shorser, Gary Bader, Emek Demir, Chris Sander, Robin Haw, Guanming Wu, Lincoln Stein, Henning Hermjakob, and Peter D'Eustachio. The reactome pathway knowledgebase 2022. *Nucleic Acids Research*, 50(D1):D687–D692, January 2022. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkab1028. URL `https://academic.oup.com/nar/article/50/D1/D687/6426058`.

Marinka Zitnik, Michelle M. Li, Aydin Wells, Kimberly Glass, Deisy Morselli Gysi, Arjun Krishnan, T. M. Murali, Predrag Radivojac, Sushmita Roy, Ana√Øs Baudot, Serdar Bozdag, Danny Z. Chen, Lenore Cowen, Kapil Devkota, Anthony Gitter, Sara Gosline, Pengfei Gu, Pietro H. Guzzi, Heng Huang, Meng Jiang, Ziynet Nesibe Kesimoglu, Mehmet Koyuturk, Jian Ma, Alexander R. Pico, Nata °a Pr æulj, Teresa M. Przytycka, Benjamin J. Raphael, Anna Ritz, Roded Sharan, Yang Shen, Mona Singh, Donna K. Slonim, Hanghang Tong, Xinan Holly Yang, Byung-Jun Yoon, Haiyuan Yu, and Tijana Milenkovifá. Current and future directions in network biology, September 2023. URL `http://arxiv.org/abs/2309.08478`. arXiv:2309.08478 [q-bio].

Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P. Gleeson, Yamir Moreno, and Mason A. Porter. Multilayer networks. *Journal of Complex Networks*, 2(3):203–271, 2014. ISSN 20511329. doi: 10.1093/comnet/cnu016.

Ruth L Seal, Bryony Braschi, Kristian Gray, Tamsin E M Jones, Susan Tweedie, Liora Haim-Vilmovsky, and Elspeth A Bruford. Genenames.org: the HGNC resources in 2023. *Nucleic Acids Research*, 51(D1):D1003–D1009, 10 2022. ISSN 0305-1048. doi: 10.1093/nar/gkac888. URL `https://doi.org/10.1093/nar/gkac888`.

Katja Luck, Dae-Kyum Kim, Luke Lambourne, Kerstin Spirohn, Bridget E. Begg, Wenting Bian, Ruth Brignall, Tiziana Cafarelli, Francisco J. Campos-Laborie, Benoit Charloteaux, Dongsic Choi, Atina G. Coté, Meaghan Daley, Steven Deimling, Alice Desbuleux, Amélie Dricot, Marinella Gebbia, Madeleine F. Hardy, Nishka Kishore, Jennifer J. Knapp, István A. Kovács, Irma Lemmens, Miles W. Mee, Joseph C. Mellor, Carl Pollis, Carles Pons, Aaron D. Richardson, Sadie Schlabach, Bridget Teeking, Anupama Yadav, Mariana Babor, Dawit Balcha, Omer Basha, Christian Bowman-Colin, Suet-Feung Chin, Soon Gang Choi, Claudia Colabella, Georges Coppin, Cassandra D'Amata, David De Ridder, Steffi De Rouck, Miquel Duran-Frigola, Hanane Ennajdaoui, Florian Goebels, Liana Goehring, Anjali Gopal, Ghazal Haddad, Elodie Hatchi, Mohamed Helmy, Yves Jacob, Yoseph Kassa, Serena Landini, Roujia Li, Natascha van Lieshout, Andrew MacWilliams, Dylan Markey, Joseph N. Paulson, Sudharshan Rangarajan, John Rasla, Ashyad Rayhan, Thomas Rolland, Adriana San-Miguel, Yun Shen, Dayag Sheykhkarimli, Gloria M. Sheynkman, Eyal Simonovsky, Murat Taşan, Alexander Tejeda, Vincent Tropepe, Jean-Claude Twizere, Yang Wang, Robert J. Weatheritt, Jochen Weile, Yu Xia, Xinping Yang, Esti Yeger-Lotem, Quan Zhong, Patrick Aloy, Gary D. Bader, Javier De Las Rivas, Suzanne Gaudet, Tong Hao, Janusz Rak, Jan Tavernier, David E. Hill, Marc Vidal, Frederick P. Roth, and Michael A. Calderwood. A reference map of the human binary protein interactome. *Nature*, 580(7803):402–408, April 2020. ISSN 1476-4687. doi: 10.1038/s41586-020-2188-x. URL `https://www.nature.com/articles/s41586-020-2188-x`.

Diego Alonso-López, Francisco J Campos-Laborie, Miguel A Gutiérrez, Luke Lambourne, Michael A Calderwood, Marc Vidal, and Javier De Las Rivas. APID database: redefining protein–protein interaction experimental evidences and binary interactomes. *Database*, 2019, January 2019. ISSN 1758-0463. doi: 10.1093/database/baz005. URL `https://academic.oup.com/database/article/doi/10.1093/database/baz005/5304002`.

M. Uhlen, L. Fagerberg, B. M. Hallstrom, C. Lindskog, P. Oksvold, A. Mardinoglu, A. Sivertsson, C. Kampf, E. Sjostedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A.-K. Szigyarto, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P.-H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne,

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300210; this version posted December 19, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

J. Nielsen, and F. Ponten. Tissue-based map of the human proteome. *Science*, 347(6220):1260419–1260419, 2015. ISSN 0036-8075. doi: 10.1126/science.1260419. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1260419.

Peter Langfelder and Steve Horvath. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9(1):559, December 2008. ISSN 1471-2105. doi: 10.1186/1471-2105-9-559. URL https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-559.

George Tsitsiridis, Ralph Steinkamp, Madalina Giurgiu, Barbara Brauner, Gisela Fobo, Goar Frishman, Corinna Montrone, and Andreas Ruepp. CORUM: the comprehensive resource of mammalian protein complexes–2022. *Nucleic Acids Research*, 51(D1):D539–D545, January 2023. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkac1015. URL https://academic.oup.com/nar/article/51/D1/D539/6830667.

Kevin Drew, John B Wallingford, and Edward M Marcotte. hu.MAP 2.0: integration of over 15,000 proteomic experiments builds a global compendium of human multiprotein assemblies. *Molecular Systems Biology*, 17(5), May 2021. ISSN 1744-4292, 1744-4292. doi: 10.15252/msb.202010016. URL https://onlinelibrary.wiley.com/doi/10.15252/msb.202010016.

Dénes Türei, Alberto Valdeolivas, Lejla Gul, Nicolàs Palacio-Escat, Michal Klein, Olga Ivanova, Márton Ölbei, Attila Gábor, Fabian Theis, Dezső Módos, Tamás Korcsmáros, and Julio Saez-Rodriguez. Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. *Molecular Systems Biology*, 17(3), March 2021. ISSN 1744-4292, 1744-4292. doi: 10.15252/msb.20209923. URL https://onlinelibrary.wiley.com/doi/10.15252/msb.20209923.

Dexter Pratt, Jing Chen, David Welker, Ricardo Rivas, Rudolf Pillich, Vladimir Rynkov, Keiichiro Ono, Carol Miello, Lyndon Hicks, Sandor Szalma, Aleksandar Stojmirovic, Radu Dobrin, Michael Braxenthaler, Jan Kuentzer, Barry Demchak, and Trey Ideker. NDEx, the Network Data Exchange. *Cell Systems*, 1(4):302–305, October 2015. ISSN 24054712. doi: 10.1016/j.cels.2015.10.001. URL https://linkinghub.elsevier.com/retrieve/pii/S2405471215001477.

Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S Baliga, Jonathan T Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape : A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, 1(Karp 2001):2498–2504, 2003. doi: 10.1101/gr.1239303.metabolite.

Alberto Valdeolivas, Laurent Tichit, Claire Navarro, Sophie Perrin, Gaëlle Odelin, Nicolas Levy, Pierre Cau, Elisabeth Remy, and Anaïs Baudot. Random walk with restart on multiplex and heterogeneous biological networks. *Bioinformatics*, 1(August):1–9, 2018. ISSN 1367-4803. doi: 10.1093/bioinformatics/bty637. URL https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/bty637/5055408.

Anthony Baptista, Aitor Gonzalez, and Anaïs Baudot. Universal multilayer network exploration by random walk with restart. *Communications Physics*, 5(1):170, July 2022. ISSN 2399-3650. doi: 10.1038/s42005-022-00937-9. URL https://www.nature.com/articles/s42005-022-00937-9.

Uku Raudvere, Liis Kolberg, Ivan Kuzmin, Tambet Arak, Priit Adler, Hedi Peterson, and Jaak Vilo. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, 47(W1):W191–W198, July 2019. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkz369. URL https://academic.oup.com/nar/article/47/W1/W191/5486750.

Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300210; this version posted December 19, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Sebastian Köhler, Leigh Carmody, Nicole Vasilevsky, Julius O B Jacobsen, Daniel Danis, Jean-Philippe Gourdine, Michael Gargano, Nomi L Harris, Nicolas Matentzoglu, Julie A McMurry, David Osumi-Sutherland, Valentina Cipriani, James P Balhoff, Tom Conlin, Hannah Blau, Gareth Baynam, Richard Palmer, Dylan Gratian, Hugh Dawkins, Michael Segal, Anna C Jansen, Ahmed Muaz, Willie H Chang, Jenna Bergerson, Stanley J F Laulederkind, Zafer Yüksel, Sergi Beltran, Alexandra F Freeman, Panagiotis I Sergouniotis, Daniel Durkin, Andrea L Storm, Marc Hanauer, Michael Brudno, Susan M Bello, Murat Sincan, Kayli Rageth, Matthew T Wheeler, Renske Oegema, Halima Lourghi, Maria G Della Rocca, Rachel Thompson, Francisco Castellanos, James Priest, Charlotte Cunningham-Rundles, Ayushi Hegde, Ruth C Lovering, Catherine Hajek, Annie Olry, Luigi Notarangelo, Morgan Similuk, Xingmin A Zhang, David Gómez-Andrés, Hanns Lochmüller, Hélène Dollfus, Sergio Rosenzweig, Shruti Marwaha, Ana Rath, Kathleen Sullivan, Cynthia Smith, Joshua D Milner, Dorothée Leroux, Cornelius F Boerkoel, Amy Klion, Melody C Carter, Tudor Groza, Damian Smedley, Melissa A Haendel, Chris Mungall, and Peter N Robinson. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Research*, 47(D1):D1018–D1027, January 2019. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gky1105. URL `https://academic.oup.com/nar/article/47/D1/D1018/5198478`.

Ozan Ozisik, Morgane Térézol, and Anaïs Baudot. orsum: a Python package for filtering and comparing enrichment analyses using a simple principle. *BMC Bioinformatics*, 23(1):293, December 2022. ISSN 1471-2105. doi: 10.1186/s12859-022-04828-2. URL `https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-022-04828-2`.

Robi Tacutu, Daniel Thornton, Emily Johnson, Arie Budovsky, Diogo Barardo, Thomas Craig, Eugene Diana, Gilad Lehmann, Dmitri Toren, Jingwei Wang, Vadim E Fraifeld, and João P de Magalhães. Human Ageing Genomic Resources: new and updated databases. *Nucleic Acids Research*, 46(D1):D1083–D1090, January 2018a. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkx1042. URL `http://academic.oup.com/nar/article/46/D1/D1083/4599180`.

Peer Aramillo Irizar, Sascha Schäuble, Daniela Esser, Marco Groth, Christiane Frahm, Steffen Priebe, Mario Baumgart, Nils Hartmann, Shiva Marthandan, Uwe Menzel, Jule Müller, Silvio Schmidt, Volker Ast, Amke Caliebe, Rainer König, Michael Krawczak, Michael Ristow, Stefan Schuster, Alessandro Cellerino, Stephan Diekmann, Christoph Englert, Peter Hemmerich, Jürgen Sühnel, Reinhard Guthke, Otto W. Witte, Matthias Platzer, Eytan Ruppin, and Christoph Kaleta. Transcriptomic alterations during ageing reflect the shift from cancer to degenerative diseases in the elderly. *Nature Communications*, 9(1):327, January 2018. ISSN 2041-1723. doi: 10.1038/s41467-017-02395-2. URL `https://www.nature.com/articles/s41467-017-02395-2`. Number: 1 Publisher: Nature Publishing Group.

Nicole C. Berchtold, David H. Cribbs, Paul D. Coleman, Joseph Rogers, Elizabeth Head, Ronald Kim, Tom Beach, Carol Miller, Juan Troncoso, John Q. Trojanowski, H. Ronald Zielke, and Carl W. Cotman. Gene expression changes in the course of normal brain aging are sexually dimorphic. *Proceedings of the National Academy of Sciences*, 105(40):15605–15610, October 2008. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0806883105. URL `https://pnas.org/doi/full/10.1073/pnas.0806883105`.

Stephen Welle, Andrew I. Brooks, Joseph M. Delehanty, Nancy Needler, and Charles A. Thornton. Gene expression profile of aging in human muscle. *Physiological Genomics*, 14(2):149–159, July 2003. ISSN 1094-8341, 1531-2267. doi: 10.1152/physiolgenomics.00049.2003. URL `https://www.physiology.org/doi/10.1152/physiolgenomics.00049.2003`.

Jason R. Pirone, Monica D'Arcy, Delisha A. Stewart, William C. Hines, Melissa Johnson, Michael N. Gould, Paul Yaswen, D. Joseph Jerry, Sallie Smith Schneider, and Melissa A. Troester. Age-Associated Gene Expression in Normal Breast Tissue Mirrors Qualitative Age-at-Incidence Patterns for Breast Cancer. *Cancer Epidemiology, Biomarkers & Prevention*, 21(10):1735–1744, October 2012. ISSN 1055-9965, 1538-7755.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300210; this version posted December 19, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

doi: 10.1158/1055-9965.EPI-12-0451. URL https://aacrjournals.org/cebp/article/21/10/1735/68990/Age-Associated-Gene-Expression-in-Normal-Breast.

Cynthia Rothblum-Oviatt, Jennifer Wright, Maureen A. Lefton-Greif, Sharon A. McGrath-Morrow, Thomas O. Crawford, and Howard M. Lederman. Ataxia telangiectasia: a review. *Orphanet Journal of Rare Diseases*, 11(1): 159, November 2016. ISSN 1750-1172. doi: 10.1186/s13023-016-0543-7.

Kirsty Scicluna, Melissa M. Formosa, Rosienne Farrugia, and Isabella Borg. Hypermobile Ehlers-Danlos syndrome: A review and a critical appraisal of published genetic research to date. *Clinical Genetics*, 101(1):20–31, January 2022. ISSN 1399-0004. doi: 10.1111/cge.14026.

Iris Boraschi-Diaz, Jennifer Wang, John S. Mort, and Svetlana V. Komarova. Collagen Type I as a Ligand for Receptor-Mediated Signaling. *Frontiers in Physics*, 5, 2017. ISSN 2296-424X. URL https://www.frontiersin.org/articles/10.3389/fphy.2017.00012.

Nard Kubben and Tom Misteli. Shared molecular and cellular mechanisms of premature ageing and ageing-associated diseases. *Nature Reviews Molecular Cell Biology*, 18(10):595–609, October 2017. ISSN 1471-0072, 1471-0080. doi: 10.1038/nrm.2017.68. URL https://www.nature.com/articles/nrm.2017.68.

Robi Tacutu, Daniel Thornton, Emily Johnson, Arie Budovsky, DIogo Barardo, Thomas Craig, Eugene DIana, Gilad Lehmann, Dmitri Toren, Jingwei Wang, Vadim E. Fraifeld, and Joaõ P. De Magalhães. Human Ageing Genomic Resources: New and updated databases. *Nucleic Acids Research*, 46(D1):D1083–D1090, 2018b. ISSN 13624962. doi: 10.1093/nar/gkx1042.

Pierre Cau, Claire Navarro, Karim Harhouri, Patrice Roll, Sabine Sigaudy, Elise Kaspi, Sophie Perrin, Annachiara De Sandre-Giovannoli, and Nicolas L??vy. Nuclear matrix, nuclear envelope and premature aging syndromes in a translational research perspective. *Seminars in Cell and Developmental Biology*, 29:125–147, 2014. ISSN 10963634. doi: 10.1016/j.semcdb.2014.03.021. URL http://dx.doi.org/10.1016/j.semcdb.2014.03.022.

Sebastian Köhler, Michael Gargano, Nicolas Matentzoglu, Leigh C Carmody, David Lewis-Smith, Nicole A Vasilevsky, Daniel Danis, Ganna Balagura, Gareth Baynam, Amy M Brower, Tiffany J Callahan, Christopher G Chute, Johanna L Est, Peter D Galer, Shiva Ganesan, Matthias Griese, Matthias Haimel, Julia Pazmandi, Marc Hanauer, Nomi L Harris, Michael J Hartnett, Maximilian Hastreiter, Fabian Hauck, Yongqun He, Tim Jeske, Hugh Kearney, Gerhard Kindle, Christoph Klein, Katrin Knoflach, Roland Krause, David Lagorce, Julie A McMurry, Jillian A Miller, Monica C Munoz-Torres, Rebecca L Peters, Christina K Rapp, Ana M Rath, Shahmir A Rind, Avi Z Rosenberg, Michael M Segal, Markus G Seidel, Damian Smedley, Tomer Talmy, Yarlalu Thomas, Samuel A Wiafe, Julie Xian, Zafer Yüksel, Ingo Helbig, Christopher J Mungall, Melissa A Haendel, and Peter N Robinson. The Human Phenotype Ontology in 2021. *Nucleic Acids Research*, 49(D1):D1207–D1217, January 2021. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkaa1043. URL https://academic.oup.com/nar/article/49/D1/D1207/6017351.

Raoul C.M. Hennekam. Pathophysiology of premature aging characteristics in Mendelian progeroid disorders. *European Journal of Medical Genetics*, 63(11):104028, November 2020. ISSN 17697212. doi: 10.1016/j.ejmg.2020.104028. URL https://linkinghub.elsevier.com/retrieve/pii/S1769721220303177.

Smail Hadj-Rabia, Jacob Mashiah, Patrice Roll, Amandine Boyer, Patrice Bourgeois, Philippe Khau Van Kien, Nicolas Lévy, Annachiara De Sandre-Giovannoli, Christine Bodemer, and Claire Navarro. A new lamin a mutation associated with acrogeria syndrome. *The Journal of Investigative Dermatology*, 134(8):2274–2277, August 2014. ISSN 1523-1747. doi: 10.1038/jid.2014.158.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300210; this version posted December 19, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Reza Maroofian, Michela Murdocca, Hossein Rezaei-Delui, Amirhossein Nekooei, Majid Mojarad, Federica Sangiuolo, Giuseppe Novelli, Andrea Superti-Furga, and Maria Rosaria D'Apice. A novel in-frame deletion in ZMPSTE24 is associated with autosomal recessive acrogeria (Gottron type) in an extended consanguineous family. *Clinical Dysmorphology*, 27(3):88–90, July 2018. ISSN 1473-5717. doi: 10.1097/MCD.0000000000000220.

OMIM. URL `https://omim.org/`.

Galliano Zanello, Macarena Garrido,ÄêEstepa, Ana Crespo, Daniel O'Connor, Rima Nabbout, Christina Waters, Anthony Hall, Maurizio Taglialatela, Chun,ÄêHung Chan, David A Pearce, Marc Dooms, and Philip John Brooks. Targeting shared molecular etiologies to accelerate drug development for rare diseases. *EMBO Molecular Medicine*, 15(7):e17159, July 2023. ISSN 1757-4676, 1757-4684. doi: 10.15252/emmm.202217159. URL `https://www.embopress.org/doi/10.15252/emmm.202217159`.

Marc Vidal, Michael E. Cusick, and Albert-László Barabási. Interactome Networks and Human Disease Marc. *Cell*, 144(6):986–998, 2011. ISSN 03681874. doi: 10.1016/j.cell.2011.02.016.Interactome.

Fabio Fabris, João Pedro de Magalhães, and Alex A. Freitas. A review of supervised machine learning applied to ageing research. *Biogerontology*, 2017. ISSN 1389-5729. doi: 10.1007/s10522-017-9683-y. URL `http://link.springer.com/10.1007/s10522-017-9683-y`.

Fabio Fabris, Alex A. Freitas, and Jennifer M. A. Tullet. An Extensive Empirical Comparison of Probabilistic Hierarchical Classifiers in Datasets of Ageing-Related Genes. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 13(6):1045–1058, November 2016. ISSN 1545-5963, 1557-9964, 2374-0043. doi: 10.1109/TCBB.2015.2505288. URL `https://ieeexplore.ieee.org/document/7346435/`.

Alex A Freitas, Olga Vasieva, and Jo√£o Pedro de Magalh√£es. A data mining approach for classifying DNA repair genes into ageing-related or non-ageing-related. *BMC Genomics*, 12:27, January 2011. ISSN 1471-2164. doi: 10.1186/1471-2164-12-27. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3031233/`.

Csaba Kerepesi, Bálint Daróczy, Ádám Sturm, Tibor Vellai, and András Benczúr. Prediction and characterization of human ageing-related proteins by using machine learning. *Scientific Reports*, 8(1):1–13, 2018. ISSN 20452322. doi: 10.1038/s41598-018-22240-w.

Qi Li, Khalique Newaz, and Tijana Milenkovifá. Improved supervised prediction of aging-related genes via weighted dynamic network analysis. *BMC Bioinformatics*, 22:520, October 2021. ISSN 1471-2105. doi: 10.1186/s12859-021-04439-3. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8543111/`.

Qi Li and Tijana Milenkovic. Supervised Prediction of Aging-Related Genes From a Context-Specific Protein Interaction Subnetwork. *IEEE/ACM transactions on computational biology and bioinformatics*, 19(4):2484–2498, 2022. ISSN 1557-9964. doi: 10.1109/TCBB.2021.3076961.

Fazle E. Faisal and Tijana Milenkovifá. Dynamic networks reveal key players in aging. *Bioinformatics (Oxford, England)*, 30(12):1721–1729, June 2014. ISSN 1367-4811. doi: 10.1093/bioinformatics/btu089.

Khalique Newaz and Tijana Milenkovic. Inference of a Dynamic Aging-related Biological Subnetwork via Network Propagation. *IEEE/ACM transactions on computational biology and bioinformatics*, 19(2):974–988, 2022. ISSN 1557-9964. doi: 10.1109/TCBB.2020.3022767.

Zaynab Hammoud and Frank Kramer. Multilayer networks: aspects, implementations, and application in biomedicine. *Big Data Analytics*, 5(1):2, December 2020. ISSN 2058-6345. doi: 10.1186/s41044-020-00046-0. URL `https://bdataanalytics.biomedcentral.com/articles/10.1186/s41044-020-00046-0`.

Lucas G. S. Jeub, Michael W. Mahoney, Peter J. Mucha, and Mason A. Porter. A local perspective on community structure in multilayer networks. *Network Science*, 5(2):144–163, 2017. doi: 10.1017/nws.2016.22.