

**SEMMELWEIS EGYETEM
DOKTORI ISKOLA**

Ph.D. értekezések

2495.

NÁNÁSI TIBOR

Funkcionális Idegtudományok

című program

Programvezető: Dr. Vizi E. Szilveszter, professor emeritus

Témavezető: Dr. Ulbert István, egyetemi tanár

MACHINE LEARNING AND GRAPH METHODS TO STUDY AGING AND EPILEPSY

PhD Thesis

Tibor Nánási MD

Semmelweis University

János Szentágothai Doctoral School of Neurosciences



Supervisor:

István Ulbert, MD, DSc

Official reviewers:

Zsuzsanna Arányi, MD, DSc

Zoltán Gáspári, PhD

Head of the Complex Examination Committee:

Dániel Bereczki, MD, DSc

Members of the Complex Examination Committee:

Anita Kamondi, MD, DSc

Sándor Pongor, DSc

Budapest

2020

TABLE OF CONTENTS

ABBREVIATIONS	5
1. INTRODUCTION	6
1.1. Graphs in systems biology	6
1.2. Machine learning in epilepsy and molecular biology	8
1.3. Support Vector Machines	9
1.4. Epilepsy	10
1.5. Aging	11
1.6. Parabiosis	14
2. OBJECTIVES	15
3. MATERIALS AND METHODS	16
3.1. Electrocorticography and Seizure Onset Zone information	16
3.2. Selection and preprocessing of electrocorticography data samples	17
3.3. Protein-Protein Interactome data	18
3.4. Transcriptomic and proteomic data sources and preparation	19
3.5. Classification of electrocorticography recordings using Support Vector Machines	21
3.6. Predictome: integrating genome-wide molecular measurements with interactome information using Support Vector Machines	23
3.7. Evaluation of SVM predictive performance: Matthews Correlation Coefficient, Jaccard Similarity	24
3.8. Monte Carlo probability of measured Eigenvector Centrality in dataset specific Predictomes (pMC)	25
3.9. Control model excluding PPI information	26
3.10. Pathway Analysis, Sliding Enrichment Pathway Analysis (SEPA)	27
3.11. Selection of critical pathways and genes, and data visualization	28
3.12. Comparison with literature	28
3.13. Coverage of the genomic analysis	31
3.14. Implementation notes	31
4. RESULTS	32
4.1. Seizure Onset Zone detection requires information from multiple frequency bands	32
4.2. Predictome-based transcriptome analysis highlights functionally related genes	40
4.3. Interactome information enhances reproducibility of Pathway Analysis results	41
4.4. Interactome information enhances robustness and abundance of pathway enrichment in top gene sets	44

4.5. SVM unmask biological information embedded in gene product interactions.....	46
4.6. Genes influenced by Brain Aging and Parabiosis are forming a continuous network in the Interactome.....	49
4.7. Pathway Analysis reveals shared aspects of Brain Aging and Parabiosis	53
5. DISCUSSION	60
5.1. Electrophysiology and Seizure Onset Zone detection	60
5.2. Molecular biology – method development and validation.....	62
5.3. Molecular biology – brain aging and parabiosis	64
5.4. Linear and Gaussian kernels in Seizure Onset Zone detection and in transcriptomics	66
6. CONCLUSION.....	70
7. SUMMARY	72
8. ÖSSZEFOGLALÁS	73
9. BIBLIOGRAPHY	74
10. BIBLIOGRAPHY OF THE CANDIDATE’S PUBLICATIONS.....	97
10.1. Publications related to this Thesis	97
10.2. Publications not related to this Thesis.....	98
11. ACKNOWLEDGEMENTS.....	100

ABBREVIATIONS

EC – Eigenvector Centrality

ECoG – Electrocorticography

EEG – Electroencephalography

FDR – False Discovery Rate

MCC – Matthews Correlation Coefficient

PA – Pathway Analysis

pMC – Monte Carlo probability

PPI – Protein-Protein Interaction

REM – Rapid Eye Movement

RMA – Robust Multi-Array Analysis

SEPA – Sliding Enrichment Pathway Analysis

SOZ – Seizure Onset Zone

SVM – Support Vector Machine

1. INTRODUCTION

1.1. Graphs in systems biology

Complex systems of multiple interacting elements often exhibit behaviors of emergent character. Rather than isolated agents, the components are mutually influenced by the activity of each other. Such systems can be analyzed on multiple different levels and from a wide array of perspectives. As a first step, static models can be drawn by mapping the interacting elements in a binary way. Considering the system as a graph, then the corresponding adjacency matrix contains that information in a compact form. Topological structure of the whole system, its sub-regions, or local neighborhoods of individual elements can be analyzed using this representation and elements can be classified both based on their position in local or in system-level topology.

A considerable subset of biological phenomena can be represented using such tools (Barabási & Oltvai, 2004; Busiello et al., 2017; Mihalik et al., 2012). From ecology to behavioral sciences, through networks of large-scale neural activity, down to the molecular level of gene product interactions (or even approaching atomic scales in structural biology) we can find countless examples where graph theoretical considerations are useful. Even though the real systems are temporal processes, considering the stationary aspects of them could lead to a better understanding of embedded interaction dynamics and to describe factors contributing to functionality. Different graphs representing alternative states of the system, or temporal snapshots can be compared to access non-stationary properties. Furthermore, the connected elements (graph nodes) themselves are not necessarily reflecting discrete physical entities like subjects or molecules but can be outputs of preceding analytic steps like measurement values associated to the studied items. Temporal and dynamic information can be encoded in these pre-processing steps.

From the long list of possible examples, in this work we will focus on two distinct biological systems, first, the large-scale functional neural network of the human brain and second, transcriptomic landscapes delivered from the same organ. Considering functional brain networks, topological information can be delivered by studying the wiring diagram of neural elements on different scales of anatomy. Parallel to this, molecular interactions are usually curated in databases and similarly organized to functional abstractions, such as pathways. Ongoing neural activity can be measured using a wide array of tools including functional neuroimaging and electrophysiology. Among the best examples for stationary processes, we can consider non-REM sleep recordings (Ujma et al., 2019). In case of molecular biology, gene expression levels or protein concentrations can be measured and analyzed together with various other molecular factors such as genetic polymorphisms (Banlaki et al., 2015; Kovacs-Nagy et al., 2013). Comparing phenotypes can reveal information about the importance of those patterns. In both systems, reactions to environmental as well as internal changes are related to dynamic processes or phase transitions, where intermittent short-term or long-lasting chronic perturbations alter the behavior of the system. Studying those responses leads to a more profound understanding of the hidden characteristics of complex biological phenomena. Changes of functional brain activity can be investigated both in normal and pathological context. Fast changes occur as responses to sensory stimuli or reflecting mental activity and their detection can be used to develop brain-machine interface systems among myriad of other theoretical and practical applications (Lee et al., 2019). On the other hand, quick pathological phase shifts can be observed in epilepsy (Jung et al., 2011; Kim et al., 2015; Ortega et al., 2008; van Mierlo et al., 2013; Wilke et al., 2011). This domain allows us to study locally controlled activities on the micro- or mesoscopic scale when analyzing inter-ictal activities and full-scale phase transitions between normal and pathological activity as well during seizures. Functional connectivity of epileptic and supposedly healthy brain tissue can be compared by analyzing responses elicited by direct cortical stimulation in patients implanted with semi-invasive intracranial electrode arrays (ECoG grids) using graph tools (File et al., 2020).

Parallel to this, molecular biology provides us examples of dynamic processes when studying long-term reorganizations of gene expression or proteome concentration patterns relatable to aging or chronic neurodegenerative diseases (De Magalhães & Tacutu, 2015; Fernandes et al., 2016; Guebel et al., 2018; Kirkwood & Kowald, 1997; Kiss et al., 2009; Simkó et al., 2009). Interestingly, de-facto phase transitions in this rather understudied phenomenon were only considered recently in the literature (Lehallier, Gate, Schaum, Nanasi, Eun Lee, et al., 2019; Lehallier, Gate, Schaum, Nanasi, Lee, et al., 2019).

1.2. Machine learning in epilepsy and molecular biology

Applications of machine learning benefited tremendously from the exponential growth of accessible computational power in the recent years. Consequently, an expansion of comparable scale can be observed in the number of such techniques and the real-world problems they have been applied to (Ben-Hur et al., 2008; Fernández-Delgado et al., 2014). Healthy (Lee et al., 2019; Ujma et al., 2019) and epileptic (Abbasi & Goldenholz, 2019) brain activity can be analyzed with great efficiency. Machine learning can help clinical seizure detection (Fergus et al., 2016), prediction (Rasheed et al., 2020; Usman et al., 2017), and even intervention (Anderson et al., 2007).

On the molecular level, applicability of these tools is equally widespread, from cancer diagnosis (Smolander, Dehmer, & Emmert-Streib, 2019) to optimization of therapy (Tseng et al., 2018), to autoimmune (Smolander, Dehmer, & Emmert-Streib, 2019), and to age-related diseases, such as coronary artery disease (Ayatollahi et al., 2019), and Alzheimer's Disease (Dukart et al., 2011). Aging itself (Fabris et al., 2017; Kerepesi et al., 2018), and putative lifespan-extending compounds in model animals (Barardo et al., 2017) have been investigated.

Interestingly, the intersection between machine learning and interaction network studies is relatively small, compared to the popularity of the two approaches as stand-alone methods. As a notable example, hierarchical structure of the yeast transcriptomic machinery (Chen et al., 2016) have been reconstructed in a data-driven manner using machine learning tools and an integrative analysis of yeast fission have been performed, combining machine learning with network biology (Pancaldi et al., 2012). Plant-pathogen interactions have also been studied from this perspective (Mishra et al., 2018).

1.3. Support Vector Machines

Support Vector Machines (SVM) are geometrically inspired classification tools which are working by the idea of drawing an n-dimensional hyperplane, which provides a clean-cut separation of classes of observations projected into an n-dimensional space, where n equals to the number of features measured (Ben-hur et al., 2001; Ben-Hur et al., 2008). Observations themselves can be properties of the investigated objects, like frequency band power of a channel in case of EEG recordings or concentrations of individual proteins in a blood sample. Classes are sets of measurements to be separated, like cases and controls. The SVM algorithm distinguishes itself from other tools by its specialized focus on hard-to-classify observations. Hence, possible hyperplanes are not evaluated using all observations but instead, geometric distances from the nearest class exemplars are used for this purpose. The machine optimizes the separator hyperplane in order produce a hyperplane with maximum margin from those exemplars, the “support vectors”.

The technique is known to be resilient to overfitting and generally copes well with outliers which are common in biological data. SVMs can use different kernels to transform the measurement values and consequently the projected positions of the observations in the hyperspace, effectively reformulating complex problems into linearly separable patterns. Importantly, both linear and radial basis function (Gaussian) kernels proven to be effective

in binary classification of transcriptome data to an extent to be comparable favorably to state of the art deep belief networks (Smolander, Dehmer, & Emmert-Streib, 2019). Good generalization can be achieved even with moderate sample sizes, which is an important factor given the limited availability of biological data. For a wide variety of mid-scale problems, SVM algorithms were shown to outperform neural networks and various other machine learning approaches (Fernández-Delgado et al., 2014). Similarly, they have the potential to surpass Extreme Learning Machines (Chorowski et al., 2014).

1.4. Epilepsy

Pharmacologically intractable epilepsy is a neurological disorder with tremendous impact on the quality of life. Besides the direct interference of the unpredictable seizure on day to day activities it is important to consider the long-term consequences of such episodes of extreme neural activity on the organization of synaptic networks. Repeated epileptiform activities can gradually reshape brain connectivity and as such they can compromise development in affected children. Disturbances can manifest in increased likelihood of psychiatric disorders (Hoare, 1984). With the presence of neurobehavioral comorbidities, a wide range of cognitive domains, especially executive functions can be compromised (Hermann et al., 2008). Lifelong effects of childhood onset epilepsy can be devastating – prospective cohort studies revealed profound structural changes of the brain (Garcia-Ramos et al., 2017) as well as severe decline in language, semantic, and visuomotor functions (Karrasch et al., 2017). Therefore, reliable diagnosis is important and even radical forms of treatment, such as destructive surgical techniques can be favorable choices considering the costs and benefits.

On the other hand, epilepsy can also be connected to aging and neurodegeneration. Epilepsy is the third most common neurological disorder of old age (Ravdin & Katzen, 2013).

Neurodegenerative changes typically occur earlier and with greater prevalence in patients suffering from chronic epilepsy, which in turn tends to become more severe with aging (Kotloski et al., 2019). Furthermore, the well-known hallmark of various neurodegenerative diseases, the amyloid- β_{1-42} ($A\beta_{1-42}$) peptide may also increase the surface expression of dopamine D1 receptors which influences both the hippocampal epileptic threshold and synaptic plasticity. In late-onset epilepsy patients, $A\beta_{1-42}$ levels were significantly decreased in cerebrospinal fluid, suggesting cerebral deposition (Costa et al., 2016). Various clinical, electrophysiological, and molecular similarities emphasize further the relatedness of epilepsy to a wide array of diseases, including Alzheimer's Disease and schizophrenia (Cendes et al., 2019).

Animal studies revealed complex patterns in hippocampal remodeling in temporal lobe epilepsy, whereas initial increases of volume were followed by a subsequent atrophy (Roggenhofer et al., 2019). In epileptic foci, cerebrovascular angiogenesis and remodeling were observed together with increased permeability of the blood-brain barrier. Those changes have been shown both in human patients and in experimental models of seizure and involving molecular factors such as Interleukins, TGF- β , PDGF and VEGFR2 (Marchi & Lerner-Natoli, 2013). Other extracellular aspects emphasize further the intervened nature of various degenerative brain pathologies and epilepsy. Increased activity of matrix metalloproteinases (MMPs) and dysregulation of the balance between them and their inhibitors (TIMPs) has been implicated in the pathogenesis of drug dependence, Alzheimer's disease, and epilepsy (Mizoguchi et al., 2011). The peculiar balance of MMP-TIMP activity influences the cleavage of neurotransmitter receptors, growth factors that mediate cell adhesion, synaptogenesis, synaptic plasticity, and long-term potentiation, among others. Interestingly, human cord plasma treatment ameliorates cognitive deficits associable to hippocampal dysfunction in aged mice and for this effect, TIMP2 found to be necessary (Castellano et al., 2017).

1.5. Aging

Our understanding on the process of aging with special emphasis of brain aging is limited: parallel theories have been developed from concepts describing a precisely timed “aging program” to viewpoints emphasizing stochasticity and emergent nature (Partridge & Gems, 2002), including Protein-Protein Interaction (PPI) network considerations (De Magalhães & Tacutu, 2015; Kiss et al., 2009).

Gene expression changes related to synaptic plasticity, calcium signaling, glutamate and GABA receptors, vesicular transport, mitochondrial function, stress response, antioxidant defense and DNA repair were observed in the human brain (Lu et al., 2004). Rather than isolated events, tissue- and organ-level alternations can directly affect each other on multiple scales. Importance of neuron-immune crosstalk involving inhibitor of nuclear factor kappa B (I κ B kinase-b, IKK-b) nuclear factor kappa B (NF- κ B) and gonadotropin-releasing hormone (GnRH) signaling was demonstrated. The central role of non-neural elements such as microglia has been proposed, highlighting the role of Interleukin-33 in Alzheimer’s Disease (Fu et al., 2016), in development (Vainchtein et al., 2018) and recently, in neuroplasticity and aging (Nguyen et al., 2020). Furthermore, hypothalamic immunity changes are shown to be affecting whole-body aging in mice (Zhang et al., 2013).

The discussion about the cross-species conservativeness of the genomic background of aging is far from settled. Particular aging-associated gene expression changes are shown to be remarkably conserved across distant eukaryotic species, like nematodes and yeast (Smith et al., 2008) but also within the class of mammals (Jobson et al., 2010; Semeiks & Grishin, 2012). This conserved system known to be enriched in elements functionally related to genomic instability, telomere attrition, epigenetic alternations, loss of proteostasis, deregulated nutrient sensing, mitochondrial dysfunction, cellular senescence, stem cell exhaustion, and altered intercellular communication (López-Otín et al., 2013). Meta-analysis of rodent and human studies confirmed the prominent role of the immune system,

mitochondria, energy metabolism and cellular senescence while also revealing clues about the involvement of collagen-related processes, the lysosome, apoptotic and cell cycle-related processes (de Magalhães et al., 2009).

On the other hand, comparison of multi-tissue data across human and mouse yielded strikingly weak overlaps of lists of homologs with significant age-related expression changes (Swindell et al., 2012; J. Yang et al., 2015; Zahn et al., 2007) and also, even chimpanzee neocortex seems to be utterly different to its human counterpart in this regard (Fraser et al., 2005). It is plausible to think about aging as a multifactorial process of heterogeneous origin and phenomenology, which could show considerable discrepancies across individuals and also within the same individual, even on sub-organ level (Stegeman & Weake, 2017). DNA methylation based epigenetic age acceleration measurements (Horvath, 2013) showed remarkable differences when comparing 30 anatomic sites (Horvath et al., 2015a). In human brain tissue, compared to the cerebellum, age seems to have a much higher impact on neocortical gene activity (Fraser et al., 2005). Similarly, using epigenetic clock analysis, the cerebellum found to be especially resilient to aging-related changes whereas the frontal and temporal cortices were comparable (Horvath et al., 2015b).

Graph aspects can be taken into account, as aging genes shown to be highly connected (central) members of the interactome forming a continuous subnetwork (Bell et al., 2009; Kirkwood & Kowald, 1997). Furthermore, it seems that particularly central or critical network elements, such as hubs, have a higher tendency to be associated to age-related pathologies (Budovsky et al., 2007; Ferrarini et al., 2005; Promislow, 2004). The concept of “Guilt-by-association” proven to be useful in finding novel aging-related genes as selecting candidates from the interaction partners of already established aging-genes (De Magalhães & Toussaint, 2004). Taking a step further, integrating those findings with a wide array of graph-topological measurements and functional annotation data, advanced data mining tools can be employed to predict novel genes with possible influence on aging (Fabris et al., 2017;

Kerepesi et al., 2018; Li et al., 2010). Focusing on the neural tissue by comparing age-associated transcriptome changes in fruitfly ganglia and human brain, interactome regions of robust behavior have been revealed which were especially enriched in cell-cycle regulating elements (Xue et al., 2007).

1.6. Parabiosis

Parabiosis is a surgically induced state that connects the circulatory systems of multiple organisms. When the chronological age of the connected subjects differ, effects of heterochronic parabiosis can be studied. Such experiments showed that multiple tissues, including muscle, liver, heart, pancreas, kidney, bone and brain, can be functionally and structurally rejuvenated in old mice connected to young in a heterochronic parabiotic state (Baht et al., 2015; Conboy et al., 2005; Q. Huang et al., 2018; Katsimpardi et al., 2014; Loffredo et al., 2013; Salpeter et al., 2013; Sinha et al., 2014; Villeda et al., 2011, 2014).

Further research concluded that infusion of plasma (the soluble, cell-free fraction of blood) can replicate multiple phenotypic changes associable to heterochronic parabiosis. Old plasma accelerates brain aging in young mice (Villeda et al., 2011) and on the flip side, young plasma can be used to reverse aspects of brain aging in old mice (Villeda et al., 2014). Recently, similar revitalizing effects of human umbilical cord plasma have been reported (Castellano et al., 2017) indicating the presence of evolutionarily conserved mechanisms. In turn, analysis of aging in human blood plasma revealed common aging patterns shared by humans and model animals, and this conserved aging signature could be altered by heterochronic parabiosis (Lehallier et al., 2019). These results underline the relevance of parabiotic mouse experiments in the study of human brain aging.

2. OBJECTIVES

Support Vector Machines, as well as graph-based data analysis can be effectively used in various fields of neuroscience, including neurophysiology and molecular biology. My previous co-authored publications include applications of graph theory on molecular (Kiss et al., 2009; Simkó et al., 2009) as well as neural systems (File et al., 2020; Nánási et al., 2016). Statistical (Banlaki et al., 2015; File et al., 2020; Kovacs-Nagy et al., 2013) and model fitting tools (Lehallier, Gate, Schaum, Nanasi, Eun Lee, et al., 2019).

In this Thesis, I will explore the synergistic application of machine learning and graph techniques to study large-scale functional brain activity recorded using semi-invasive electrophysiological tools (ECoG) as well as transcriptomic and proteomic changes relatable to organism-level and brain aging.

In my previous works I have demonstrated the usability of network-based models (File et al., 2020) and multi-modal feature integration (Nánási et al., 2016) in Seizure Onset Zone localization from intracranial ECoG recordings. In this thesis, the feasibility of ECoG analysis using machine learning models will be investigated.

In the fields of transcriptomics and proteomics, the number of integrative studies involving both graph and machine learning tools is limited when compared to the extensive literature on both topics. Aiming to analyze omics measurement in context of a priori knowledge on gene product interactions, a novel integrative method, the Predictome approach will be introduced and validated on repeated transcriptomic measurements of human brain aging. Finally, proteomic changes elicited by heterochronic parabiosis will be explored using the validated model to investigate the relevance of parabiosis and the importance of evolutionally conserved, circulating factors in blood plasma in the context of brain aging.

3. MATERIALS AND METHODS

3.1. Electrocorticography and Seizure Onset Zone information

Electrocorticography (ECoG) data has been acquired and processed as described in preceding works (File et al., 2020; Nánási et al., 2016). In this study, previously acquired clinical data from six patients suffering from pharmacologically intractable focal epilepsy was used. Following the clinical protocols, antiepileptic drug administration was discontinued, or doses reduced to facilitate the emergence of epileptiform activities of diagnostic value. Then, implantation of flexible electrode grid arrays to subdural positions and continuous ECoG monitoring was performed in the Department of Functional Neurosurgery and Center of Neuromodulation, National Institute of Clinical Neurosciences, Budapest, Hungary. Seizure Onset Zones (SOZ) were identified by experts in the Epilepsy Centrum, Department of Neurology, National Institute of Clinical Neurosciences, Budapest, Hungary.

The diagnostic setup offered 1 cm of spatial resolution (in terms of distances between adjacent ECoG grid electrodes). Electrode positions were reconstructed by integrating information from pre-implantation structural MRI and post-implantation CT scans using the FreeSurfer software (<https://surfer.nmr.mgh.harvard.edu/>). Data acquisition rate and precision were 1024 Hz and 16 bit. For an overview of the number of ECoG channels and their relative position to SOZ, please consult with Table 1.

Table 1 – Available ECoG recording channels

# patient	ECoG electrodes		Total
	SOZ	Non-SOZ	
1	10	27	37
2	3	29	32
3	5	25	30
4	14	34	48
5	11	16	27
6	8	32	40

3.2. Selection and preprocessing of electrocortigraphy data samples

For each patient, 3-minute-long segments of deep non-REM sleep ECoG recordings were selected for analysis (n=2 for one patient, n=4 for five patients), free from any obvious epileptic activity except of spikes and with at least 1 hour of temporal separation from actual seizures. From these segments, seven frequency bands were produced by applying Butterworth band-pass filtering: slow delta (1-2 Hz), fast delta (2-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), slow gamma (30-45 Hz) and fast gamma (45-80 Hz) bands. Filters were applied in a forward-backward manner to prevent phase distortions. The range of 45-55 Hz was omitted to rule out artefacts potentially introduced by the 50 Hz power grid frequency (Table 2).

Table 2 – analyzed frequency bands of the ECoG signal.

Band name	Frequency
Slow delta	1-2 Hz
Fast delta	2-4 Hz
Theta	4-8 Hz
Alpha	8-13 Hz
Beta	13-30 Hz
Slow gamma	30-45 Hz
Fast gamma	55-85 Hz

Amplitude information was accessed in the form of absolute Hilbert transformed values of the filtered signals. Band power was defined as squared amplitude. To obtain a reasonable

number of samples to be presented to the machine learning models, we have dissected the band power time series into non-overlapping epochs of 10 second length (Table 3). This step was added after the filtering and Hilbert transformations in order to avoid edge artefacts which could be present to some extent even with the refined Butterworth technique referenced above.

Table 3 – summary of the acquired ECoG data epochs.

# patient	Obtained data	Total epochs
1	2 x 180 sec	36
2	4 x 180 sec	72
3	4 x 180 sec	72
4	4 x 180 sec	72
5	4 x 180 sec	72
6	4 x 180 sec	72

3.3. Protein-Protein Interactome data

Multiple PPI databases exist with the mutual aim but with different logics and philosophies to collect and organize the already acquired knowledge on the interaction network of proteins (Ceol et al., 2009; Gioutlakis et al., 2017; Kerrien et al., 2007; Keshava Prasad et al., 2009; Stark, 2006; Szklarczyk et al., 2017; Türei et al., 2016; Veres et al., 2015). Since reproducibility is a core issue in the field, we shifted our preferences from coverage towards reliability when selecting a reference interactome map. Several approaches exist to incorporate the probabilistic nature of the curated protein-protein interaction links however as the method of choice is characteristic to the source, merging them (which is preferable given their limited overlap) would pose another challenge and the solution would be, inevitably, heuristic and arbitrary. From the wide array of available databases therefore we have focused on two, distinct by their shared high standards of reliability protocols. OmniPath (Türei et al., 2016) and PICKLE (Gioutlakis et al., 2017; Klapa et al., 2013) are

both defining confirmation by at least two independent sources as an inclusion criteria for their manually curated protein interaction links. Merging them produced a novel deposit of highly dependable PPI data describing 195.456 interactions between 16.005 proteins which is, to our knowledge, could be regarded as the most comprehensive body of reliability-optimized interactome information currently available. When constructing the unified database, we have opted for a protein-based namespace (UniProt) natively accessible by both sources to avoid additional namespace-mapping artefacts.

3.4. Transcriptomic and proteomic data sources and preparation

To study human brain aging, genome-wide RNA sequencing data (Deluca et al., 2012; Mortazavi et al., 2008; Wang, Zhong; Gerstein, Mark; Snyder, 2009) obtained from the GTEx Consortium was analyzed (The GTEx Consortium, 2013). Notably, the GTEx dataset contains a rather unique redundancy for frontal cortex and cerebellum as in a considerable subset of cases tissue samples from the exact same individuals were re-sampled and measured by two independent laboratories. Although minor variations are expectable based on the inevitable differences in sampled tissues (about 1 mm distance between the two subsequent sampling areas), ischemic time and handling, this redundancy gives a good opportunity to enhance the reliability and to test the stability of the obtained results.

It has been shown recently (Lehallier, Gate, Schaum, Nanasi, Eun Lee, et al., 2019) that aging can be characterized by multiple, temporally separated waves of proteomic changes. Crests of the undulating proteome alternations are reached at ages of 34, 60 and 78 years of age and they are reflecting distinct biological pathways, implying profound temporal heterogeneity of the aging process. The crests were marked as points of interest of our transcriptome analysis using the GTEx data. However, the age distribution of the included patients forces us to compromise in this regard as there is insufficient data to study the last wave of aging

and sample sizes corresponding the first wave are limited. Also, as patient age is reported with the precision of decades in the GTEx database, we had to alter the definition of the first wave of aging accordingly.

We decided to analyze the first and second waves of aging by comparing age ranges of 20 to 39 with 40 to 49 for the first and 50 to 59 with 60 to 69 for the second wave. Covariate distributions found to be well-balanced across these subgroups, with the exception of early-stage cerebellar aging, where mild trends of deviation from equilibrium could be found for both sex ($p_{\text{Wilcoxon}}=0.09$) and post-mortem interval ($p_{\text{Wilcoxon}}=0.1$) covariates.

Normalized data readily available from the GTEx Portal is optimized for a subsequent eQTL-analysis with stringent filtering criteria which is not fulfilled by the RNA probes in a uniform manner, resulting discrepancies in the available gene sets to work with. As our goal was increasing robustness through repeated measurements, we opted to re-normalize the raw readout data using robust multi-array analysis (RMA, (Irizarry et al., 2003)) to obtain equal coverage of the transcriptomic landscape.

Our investigation was limited to the intersection of the set of measured genes or proteins and the set of traceable elements within the available Protein-Protein Interaction data. For the PPI framework, we have selected the SwissProt / UniProt ID system which is natively supported by both used interactome databases. Mapping from RNA to protein space was carried out using the online interface provided by UniProt (uniprot.org). In non-bijective cases when multiple RNA IDs could have been associated to a single UniProt entity, feature values for all possible ID matching combinations were calculated as described in the following sections and their averages were considered as the final results corresponding to entities curated in the PPI network. For the final useable sample sizes, please consult with Table 4.

Table 4 – Summary of analyzed transcriptomic datasets. Post-mortem interval (PMI) of the tissue samples and gender of donors were used as covariates. Uneven coverage across repeated laboratory measurements indicated by superscripts ¹ and ² ; uniform re-normalization of raw RNA sequencing counts with RMA offered a solution to this problem.

Tissue	Phenotype	Sample Size		Genes	
		Young	Old	without RMA	with RMA
Frontal Cortex (2x)	Early Aging (20-39y → 40-49y)	6	9	13619 ¹	15375
	Late Aging (50-59y → 60-69y)	29	39	13602 ²	
Cerebellum (2x)	Early Aging (20-39y → 40-49y)	11	12	13570 ¹	15375
	Late Aging (50-59y → 60-69y)	35	47	13503 ²	

Heterochronic parabiosis was explored using previously published measurements (Lehallier, Gate, Schaum, Nanasi, Eun Lee, et al., 2019) based on aptamer technology (Gold et al., 2010). Plasma protein concentration in young and old heterochronic parabiont animals were compared to iso-chronic (same age) parabiotic controls. This arrangement was selected to rule out possible side effects of the parabiotic state, such as modified locomotion or stress levels which are unrelated to the rejuvenation or provoked aging effect elicited by the shared circulatory system. From the available protein concentration readouts, only elements with associable PPI nodes were processed (Table 5).

Table 5 – Summary of analyzed proteomic datasets. All animals were males.

Parabiosis	Control (isochronic)	Treated (heterochronic)	Plasma proteins
Effect on young animal	5	6	1337
Effect on old animal	9	9	1337

3.5. Classification of electrocorticography recordings using Support Vector Machines

ECoG data was processed and segmented into epochs as described above. Based on their channel of origin, these recordings were assigned to SOZ and non-SOZ classes and presented to the SVM algorithm as observations (Table 6). The Support Vector Machines were tasked to predict SOZ / non-SOZ labels and ultimately, to reconstruct expert opinion on SOZ / non-SOZ status.

Table 6 – Number of observations presented to the SVM algorithm

# patient	Observations presented to SVM		
	SOZ	Non-SOZ	Total
1	360	972	1332
2	216	2088	2304
3	360	1800	2160
4	1008	2448	3456
5	792	1152	1944
6	576	2304	2880

The ECoG signal has been filtered to seven frequency bands and power information from each band has been extracted prior to these steps. It has been shown previously that extra information on brain activity can be acquired by combining data from different frequency bands (Canolty & Knight, 2010; Cohen et al., 2009; López-Azcárate et al., 2010, 2013; Maris et al., 2011; McGinn & Valiante, 2014; Nánási et al., 2016; Scheffzük et al., 2011; Sharott et al., 2009; Tort et al., 2009, 2008; Van Ooyen et al., 2018; von Nicolai et al., 2014). We hypothesize that presenting multiband information to our machine learning model could enhance SOZ reproduction as well. In this merit, 8 types of SVM models have been constructed: in 7 cases, observations contained features only on power measured on a singular band, like theta- or beta powers. In the 8th case however, all band power measurements were condensed into feature vectors of 7 dimension representing different characteristics of the observations to be classified.

Given the heterogeneity in terms of clinical parameters, and exact grid localization and grid dimensions, patients were analyzed separately. In order to facilitate robustness of our findings, 5-fold cross-validation was used during model training and evaluation. Kernel choice effects were investigated by comparing the output of linear and Gaussian (RBF) kernel SVM models. Predictors were standardized and default kernel parameters were used.

3.6. Predictome: integrating genome-wide molecular measurements with interactome information using Support Vector Machines

In our analysis, we aimed to combine information extracted from experimental gene expression measurements with a priori knowledge accumulated on the functional interactions of the gene products on the protein level. The nascent Protein-Protein Interaction Network (PPI) was transformed into phenotype-specific weighted graphs (Predictomes) with link weights delivered from the performance of SVM models in a process described as follows.

Each link (interaction) of the PPI network was considered as a separate model fitting task to be solved. Two continuous predictors were defined as the gene expression levels or direct concentration readouts corresponding to the two interacting proteins forming the link in question. Non-genetic covariates like sex and sample post-mortem interval (PMI) were included as additional predictors in case of human data. Predictors were standardized then the models were trained to classify the samples into younger-older or control-pathological groups in a binary manner. Performance was measured in the form of Matthews Correlation Coefficient (MCC). Samples were divided into 3 folds for cross-validation and predictive performances of the trained models were measured on the test sets (1/3). The procedure was repeated (so learning and test sets were randomized) 30 times and the resulting MCCs (3x30)

were averaged for each link to define its weight. Hence, Protein-Protein Interaction links were outfitted with SVM performance metrics on age or parabiogenic state prediction.

Finally, for human measurements, the resulting graph was subjected to thresholding: link weights were adjusted by the performance of a similarly trained and evaluated SVM model using only the non-genetic covariates without actual transcript data. So, corrected link weights in the Predictome are representing the performance gain of an SVM model outfitted with gene expression data corresponding to the linked nodes over the covariate-only null model. Negative link weights were interpreted as the inability of the model to extract excess information from genomic data on phenotype and therefore set to 0. Such links with zero weight are treated as non-existent by the subsequent algorithms employed to quantify node importance (see next chapters).

3.7. Evaluation of SVM predictive performance: Matthews Correlation Coefficient, Jaccard Similarity

In this work, SVM performance was measured in the form of Matthews Correlation Coefficient (MCC) (Matthews, 1975) between prediction and ground truth. As sample sizes are unbalanced, especially for ECoG data (where both the number of recording electrodes and the ratio of SOZ / non-SOZ classes vary across patients), MCC is a favorable choice (Powers, 2007) over more generic metrics, such as accuracy.

To quantify results when testing model stability of replicated laboratory measurements (available only for human brain aging transcriptomes), Jaccard Similarity (Jaccard Index, Jaccard similarity coefficient) was used (Jaccard, 1912). This metric is defined as the size of the intersection divided by the size of the union of the sample sets (highlighted genes or pathways using the two replicates), and much like MCC, gives a more realistic insight on the

robustness of item selection tasks than sheer overlap across the highlighted elements when set sizes are different.

3.8. Monte Carlo probability of measured Eigenvector Centrality in dataset specific Predictomes (pMC)

To evaluate the importance of individual agents, information represented by the graph (Predictome) must be projected to node (gene) level. Eigenvector Centrality (Newman, 2004) prioritizes genes participating in multiple, highly predictive interactions and genes forming highly predictive interactions with the ones falling into the former category. Its “recursive logic” practically generates high scores based on local importance and spreads that through the network via strong links, thereby extends the influence of the nodes in the topological space. The metric has been shown to be informative in various biological problems, including functional brain network studies (Gao et al., 2019; Skouras et al., 2019) and cancer genomics (Al-Aamri et al., 2019), and thought to emphasize elements of central super-regulatory role or a critical targets of regulatory pathways (Vella et al., 2017). In our case, it naturally highlights genes whose expression levels are individually correlated to group affiliation and also, it favors contributors of interactions where neither gene’s expression level, but their combined information (such as expression ratio) could be used as a successful predictor of phenotype.

However, (weighted) graph centrality measurements in general are heavily influenced by sparsity features of the (unweighted) network, such as local density and node degree. To avoid focusing on genes solely by their central position in the unweighted PPI network, Monte Carlo analysis has been conducted whereas link weights were shuffled between the existing connections, and node scores were re-calculated 100 000 times. Final node significance level was assigned based on the number of times the actual measurement-based

value surpassed the ones delivered from those surrogate networks in the form of Monte Carlo probability of measured Eigenvector Centrality, pMC (see Figure 1).

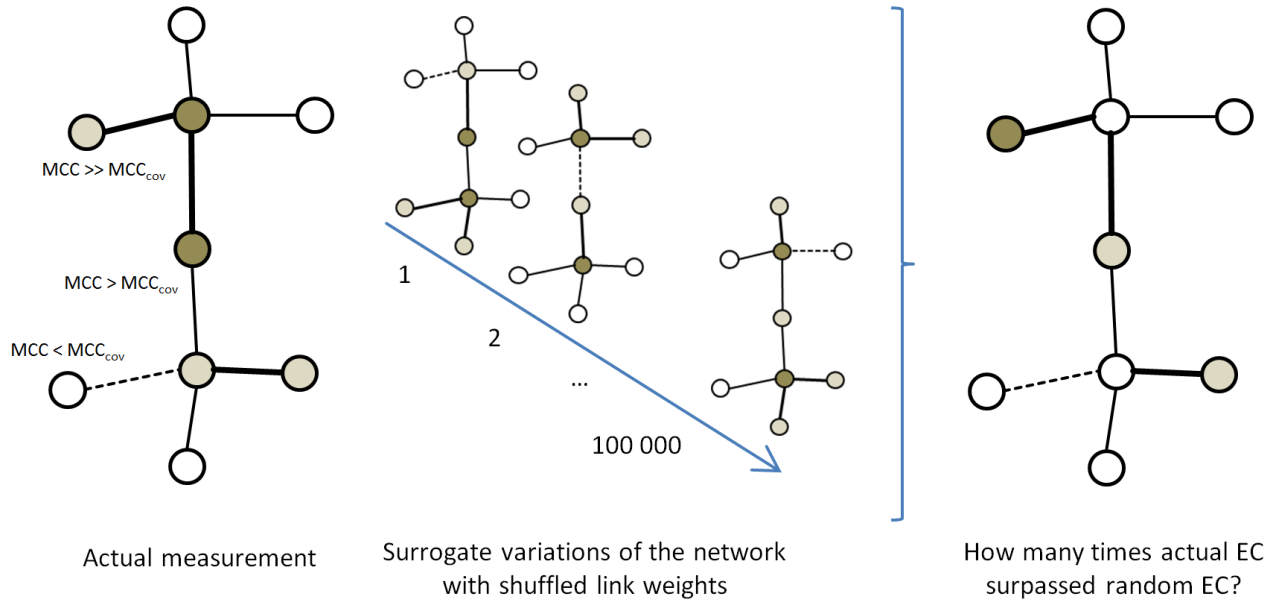


Figure 1 – Calculating Monte Carlo probability of measured Eigenvector Centrality (pMC) in the dataset specific Predictomes. Predictome link weights were defined as increase of Matthews Correlation Coefficient (MCC) in the phenotype prediction task while having access to gene expression information on the linked elements, compared to models outfitted with covariate data only. While keeping the unweighted topology of the PPI network intact, assigned link weights were shuffled randomly 100 000 times and Eigenvector Centrality (EC) values of the nodes were recalculated for these networks, and compared to EC delivered from actual measurement data. Monte Carlo probability expresses the rate on which the measured EC surpasses the surrogate values, hence, quantifies its per-node significance corrected to unweighted network topology.

3.9. Control model excluding PPI information

To evaluate the influence of the integrated PPI information on robustness and interpretability, an alternative model has been constructed where SVMs had to rely only on individual gene expression data instead of gene pairs to predict phenotype. Other steps of the workflow (i.e. inclusion of covariates, cross-validation and performance metrics) were similar. The

investigated gene space was restricted to PPI-traceable elements to facilitate comparisons with the integrative model. In cases of unambiguous transcriptome readout to protein projections, all possible solutions have been examined and results were averaged. Given the limitation that the surrogate approach cannot be used without the inclusion of the network which weights to be randomized, nodes were simply ranked by their predictive performance on group affiliation.

3.10. Pathway Analysis, Sliding Enrichment Pathway Analysis (SEPA)

Pathway Analysis (PA) is a widely used concept in the interpretation of high throughput genome-wide data. Most forms of PA are employing the calculation of statistical enrichment of biologically annotated elements among genes highlighted by a former step of the analytic workflow to quantify their functional attunement. The consensual biological processes and the corresponding lists of genes related to them are curated in various functional annotation databases (Ashburner et al., 2000; Fabregat et al., 2018; Kanehisa & Goto, 2000). Although progressive, integrative frameworks capable to handle several functional annotation systems parallelly are gaining momentum (D. W. Huang et al., 2009), serious questions have been emerged regarding to their reliability (Wadi et al., 2016b, 2016a). Sticking with self-contained platforms, we have opted to employ Reactome (Fabregat et al., 2018) as our standard testing library. With its relatively confined semantic space this annotation system allowed further robustness and dimension reduction as the number of investigated gene/protein entities exceeded the quantity of the associated terms more than sevenfold.

The exact methodology of performing the statistical calculations in PA is nontrivial and still forms a subject of heavy research and development (García-Campos et al., 2015). Overrepresentation analysis (ORA, (Dopico et al., 2015; Wang et al., 2013; Zeeberg et al., 2003)) is regarded as a conservative standard in the field (García-Campos et al., 2015). In

this work, we resorted on the closely related Sliding Enrichment Pathway Analysis (SEPA, (Lehallier, Gate, Schaum, Nanasi, Eun Lee, et al., 2019)) which enhances the process of cutoff selection by considering the ranking of selected genes of interest. Within SEPA, functional enrichment for Reactome terms was quantified using Fisher's exact test (Fisher, 1922) in an iteratively broadened list of top scoring elements, with increments of 1 until covering all genes with significant pMC (see previous chapters). FDR correction was carried out according to Benjamini and Hochberg (Benjamini & Hochberg, 1995). Phenotype-associated Reactome terms were required to be consistently significant ($q < 0.05$ for at least 20 incremental lists).

3.11. Selection of critical pathways and genes, and data visualization

Visualization and outline the most important results from analyses of genomic scale is challenging. Enriched pathways can be highly redundant because of the embedded hierarchical structure of molecular biology terms (like S phase is a subcategory of Cell cycle phases). The goal is providing a concise summary of functional changes of genome activity landscape.

To achieve this in an unbiased, data-driven way, the hierarchical network of Reactome terms (available on Reactome.org) was regarded as an unweighted network, then subjected to modularization using the Louvain algorithm (Blondel et al., 2008). The resulting modules representing groups of terms densely connected in the Reactome network and expected to share biological meaning. Then, top-scoring pathways were picked separately from each of these modules.

3.12. Comparison with literature

Highlighted gene and pathway sets were compared with literature data to verify the integrity of our results in an unbiased way. Literature mining has been performed in PubMed abstracts and co-occurrence of gene or pathway names with keywords (“*Aging*”, “*Age-related*”) have been counted. Additionally, the GenAge database (De Magalhães & Toussaint, 2004) has been queried for genes.

Overrepresentation of GenAge membership was quantified using Fisher’s exact test. Gene or pathway name frequency in aging-related PubMed abstracts was subjected to a Wilcoxon rank sum test, comparing these metrics delivered for highlighted and non-highlighted items.

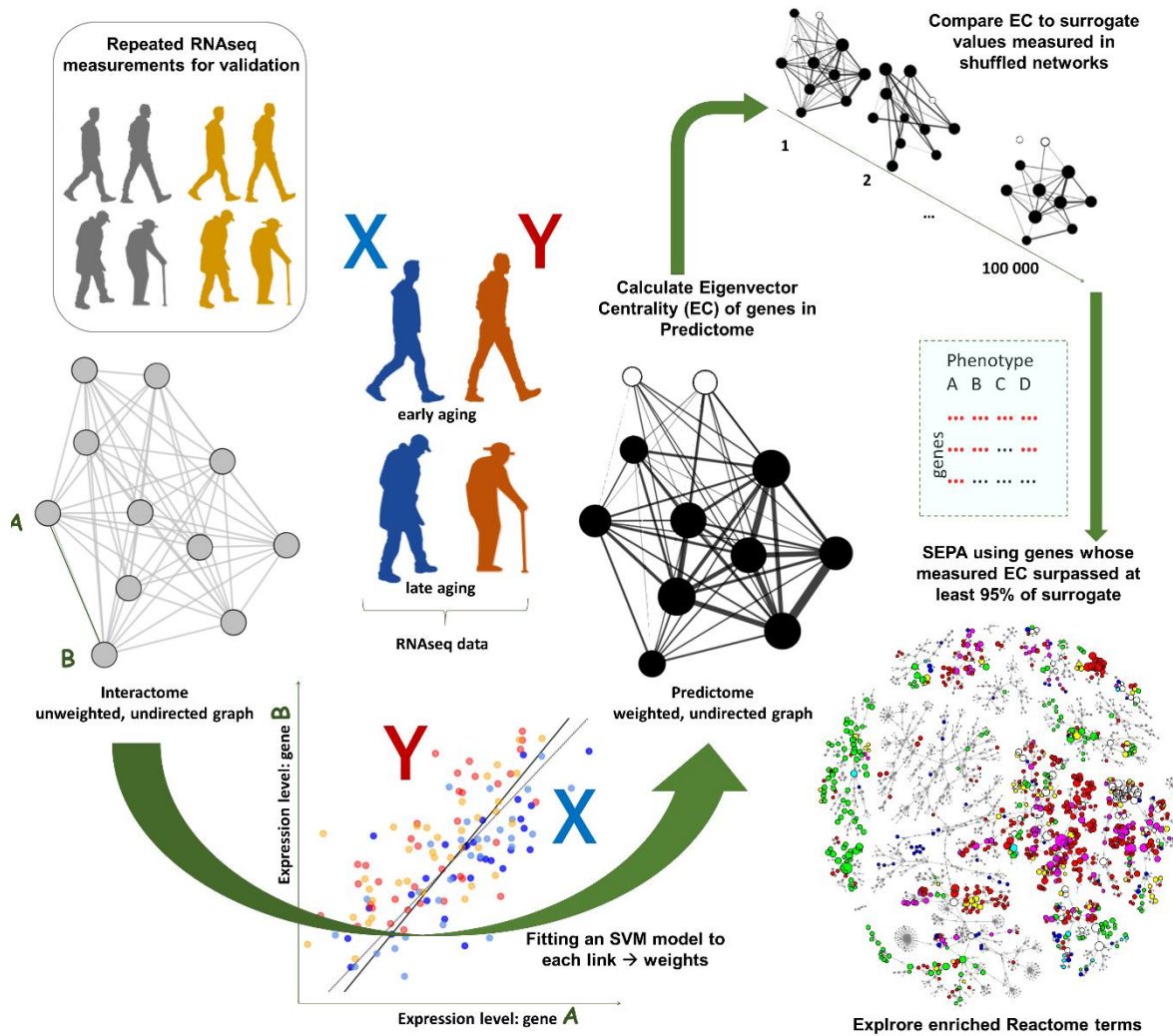


Figure 2 – Overview of the Predictome approach used in this thesis. Genome-wide RNAseq measurements were compared across younger and older subjects with small age differences, corresponding to the first and second waves of aging (see Methods, chapter 4). The network of a priori known protein interactions (Methods, chapter 3) was outfitted with weights using performance metrics of Support Vector Machine models tasked to predict phenotype from measurements associable to the linked entities. This empirically weighted graph was termed as the Predictome (Methods, chapters 6 and 7). Individual gene importance was determined by Eigenvector Centrality. To compensate for topological biases, actual measurement values were compared with results arising from surrogate networks with shuffled edge weights, then genes were ranked based on the Monte-Carlo significance of their Eigenvector Centrality (Methods, chapter 8). Finally, functional attunements of ranked gene lists were explored by Sliding Enrichment Pathway Analysis performed in the Reactome term space (Methods, chapters 10 and 11).

3.13. Coverage of the genomic analysis

Based on the available genome-wide RNAseq data, the interactome network (PPI) could be outfitted with link weights and node-level statistics with good coverage. Brain aging data readily available from the GTEx Portal could be matched to 84.4-85.1% of the curated proteins and to 80.2-80.3% of the interactions. Our RMA re-normalization of raw GTEx data enhanced both of those metrics to 96.0-96.1%. For pathway analysis, 2190 (97.8%) from the total of 2240 Reactome terms could be associated to entities enlisted in our merged PPI database. Also, from the 16005 PPI-trackable entities, 9401 (58.7%) could be directly linked to biological functions using Reactome. Direct proteomic measurements in mouse blood plasma covered only a modest 8.35% of the PPI space, which were found to be associable to 1590 Reactome terms (71%). It's noteworthy that albeit the overlap across the item sets curated in Reactome and our PPI framework is not perfect, the transcripts or proteins directly not associable to Reactome terms can still influence the enrichment scores via local and global interactome topologies when using the network-based models.

3.14. Implementation notes

RMA normalization of raw sample readouts was performed with the YARN package (Paulson J, Chen C, Lopes-Ramos C, Kuijjer M, Platig J, Sonawane A, Fagny M, Glass K, 2019) and literature mining using custom written scripts in the R software environment. Other parts of the analysis and visualization were carried out in MATLAB. Computationally intensive tasks (such as model fitting and Monte Carlo statistics) were performed on a dual Intel Xeon E5-2680v2 system outfitted with 20 physical processor cores and 56 GBytes of RAM, using custom-written, highly parallelized MATLAB scripts.

4. RESULTS

4.1. Seizure Onset Zone detection requires information from multiple frequency bands

Relying on single-band power features, regardless of the used frequency band or applied kernel transformations, SVM was unable to find separation planes reliably sorting epoch data to SOZ or non-SOZ origin. In other words, the machine learning algorithm failed to identify rules providing similar results to judgement of neurologist experts. Still, differences between distinct frequency band and kernel performances exist.

In contrast, SVM models outfitted with full spectral information provided better results for both linear and Gaussian approaches. Importantly, Gaussian kernel SVM employing multiple frequency band powers as feature vectors was able to reproduce expert decisions with great fidelity in all patients. Ability of different models to highlight epochs of SOZ origin summarized in Table 7 below, in terms of Matthews Correlation Coefficient.

Table 7 – Summary of model fitting results to reproduce expert Seizure Onset Zone selection using different SVM kernels and frequency band information: Matthews Correlation between prediction and ground truth. For all patients, combining full spectral information and Gaussian kernel transformations provided the best results (bold, black). Linear SVM achieved comparable performance only with combined spectral information for patients #1 and #2. Access to the full spectrum resulted in dramatic improvements in model fitting in case of patient #2 when using linear SVM and for patients #3, #4, and #6 with Gaussian kernel, emphasizing the importance of information encoded in cross-frequency relationships. Interestingly, for patients #1 and #5, the restricted use of the gamma frequency range produced virtually identical output with both kernels (bold, red), indicating the presence of linearly separable epileptic patterns in some cases.

patient / frequency	Linear SVM						Gaussian kernel SVM					
	#1	#2	#3	#4	#5	#6	#1	#2	#3	#4	#5	#6
slow delta	0	0	0	0	0	0	0.2	0.33	0	0	0.21	0.08
fast delta	0.26	0	0	0	0	0	0.31	0.07	0	0	0.23	0.12
theta	0.26	0	0	0	0	0	0.3	0.1	0	0	0.22	0
alpha	0	0	0	0	0	0	0	0	0	0	0.27	0
beta	0	0	0	0	0	0	0.09	0	0	0	0.32	0
slow gamma	0.45	0	0	0	0.24	0	0.45	0	0	0.03	0.24	0
fast gamma	0.37	0	0	0	0.22	0	0.37	0.05	0	0	0.23	0
COMBINED	0.58	0.39	0	0	0.27	0	0.65	0.4	0.44	0.29	0.49	0.37

Seemingly, linearly separable, and more complex cases are both present in this model fitting task. SVM was able to approximate expert SOZ detection (hence fit the model to replicate that decision) operating within the untransformed measurement space in cases of patient #1 and #2 and proven to be partially successful with patient #5. In two cases, the decisions were largely driven by information present in the gamma frequency range. However, sometimes single-band data is not enough to highlight the SOZ. For patient #2 in the linear, and for patient #3 in the Gaussian feature space, the SVM purely relies on cross-frequency constellations.

To analyze our results in a greater detail, SVM classification results were projected to the reconstructed anatomical positions of the ECoG electrodes. Also, principal component analysis (PCA) using the whole spectral information was performed on individual epochs (observations), then, SVM predictions on SOZ status were visualized in the PCA space. On the following pages we provide some examples. Defining successful models by the ability to classify epochs originating from SOZ electrodes as epileptic with greater likelihood than epochs of non-SOZ origin, slow rhythm data alone was sufficient to fit our models to clinical data for 4 out of 6 patients. For this performance, however, the SVM models had to exploit Gaussian kernel transformations (Figure 3). In contrast, fast rhythms enabled the machine learning algorithm to pinpoint the SOZ even without kernel transformations for patients #1 and #5. Solutions provided by linear and Gaussian models are remarkably similar, both in the anatomical and in the PCA space (Figures 4-5).

Results provided by models given access to the whole frequency spectrum information were generally superior. Both linear SVM and Gaussian kernel SVM could combine band power data synergistically (Figures 6-7).

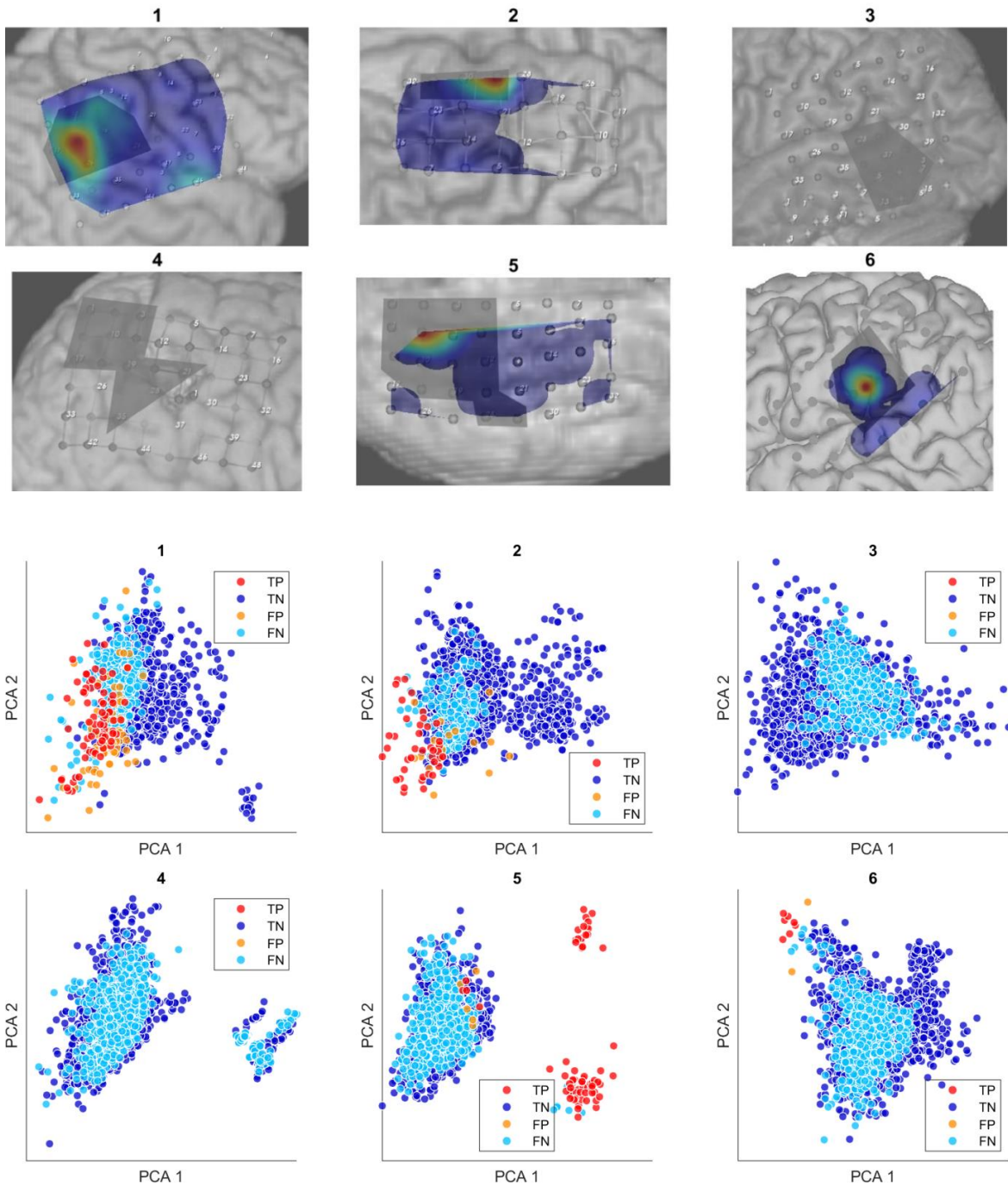


Figure 3 – visualization of **Gaussian** kernel SVM classification using the **slow delta** band. On the reconstructed anatomical images, grey shading marks the clinical SOZ. Rate of epileptic classification of epochs originating from ECoG electrodes indicated with coloring with red zones representing the highest value. Below, individual epochs visualized in PCA space, differentiating true positive (TP), true negative (TN), false positive (FP), and false negative (FN) categories.

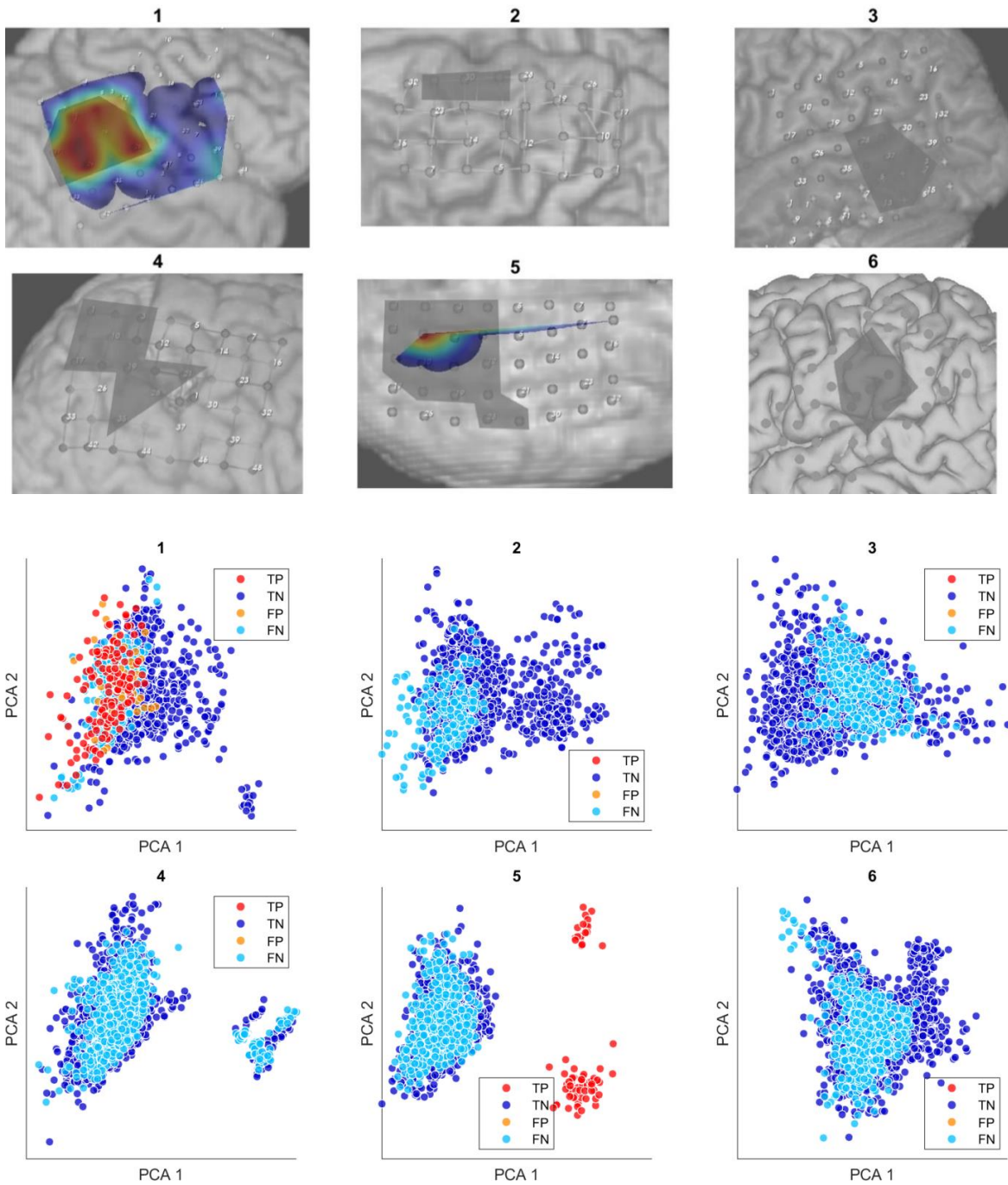


Figure 4 – visualization of **linear SVM** classification using the **slow gamma** band. On the reconstructed anatomical images, grey shading marks the clinical SOZ. Rate of epileptic classification of epochs originating from ECoG electrodes indicated with coloring with red zones representing the highest value. Below, individual epochs visualized in PCA space, differentiating true positive (TP), true negative (TN), false positive (FP), and false negative (FN) categories.

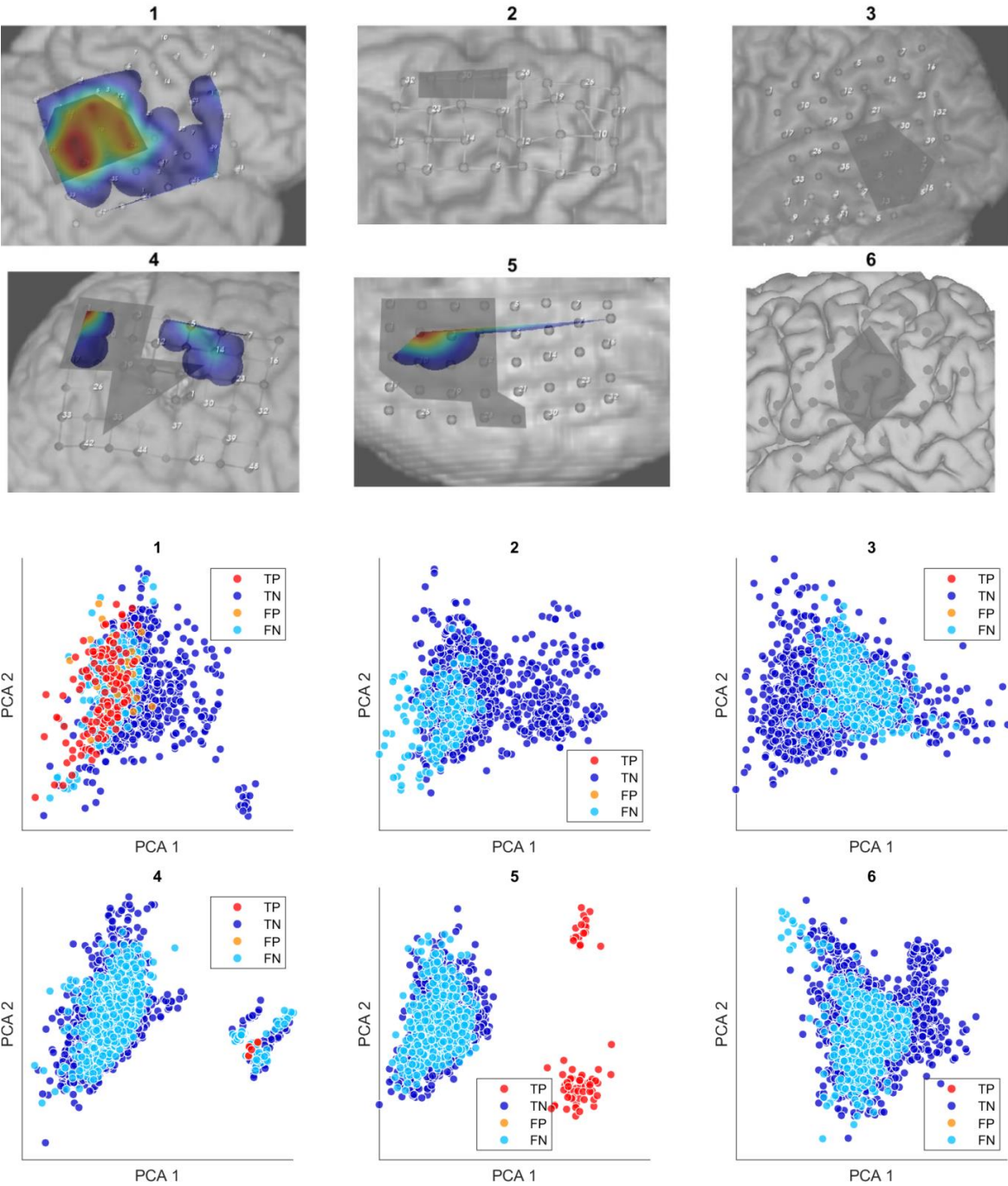


Figure 5 – visualization of **Gaussian** kernel SVM classification using the **slow gamma** band. On the reconstructed anatomical images, grey shading marks the clinical SOZ. Rate of epileptic classification of epochs originating from ECoG electrodes indicated with coloring with red zones representing the highest value. Below, individual epochs visualized in PCA space, differentiating true positive (TP), true negative (TN), false positive (FP), and false negative (FN) categories.

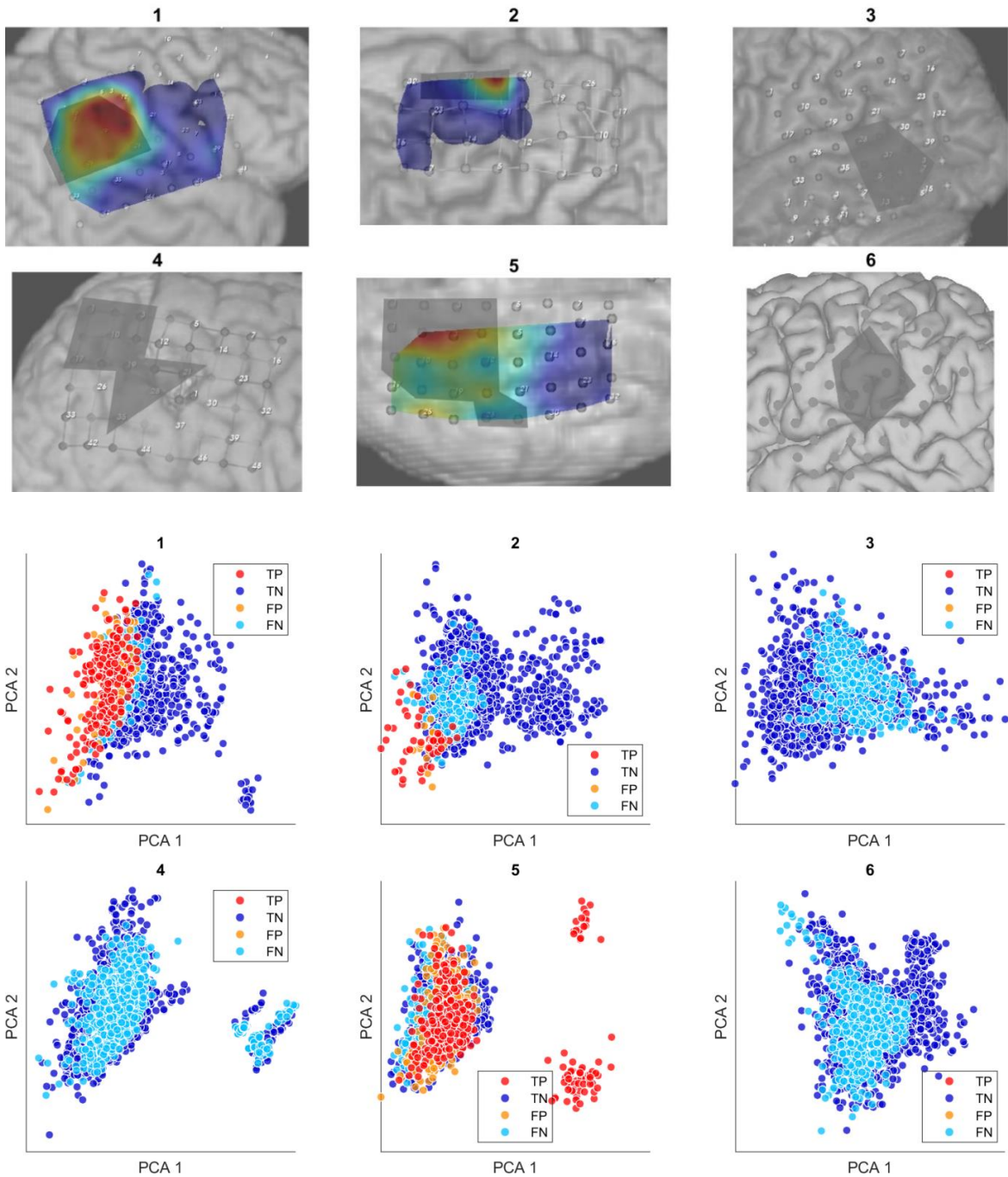


Figure 6 – visualization of **linear SVM** classification combining **all** spectral features. On the reconstructed anatomical images, grey shading marks the clinical SOZ. Rate of epileptic classification of epochs originating from ECoG electrodes indicated with coloring with red zones representing the highest value. Below, individual epochs visualized in PCA space, differentiating true positive (TP), true negative (TN), false positive (FP), and false negative (FN) categories.

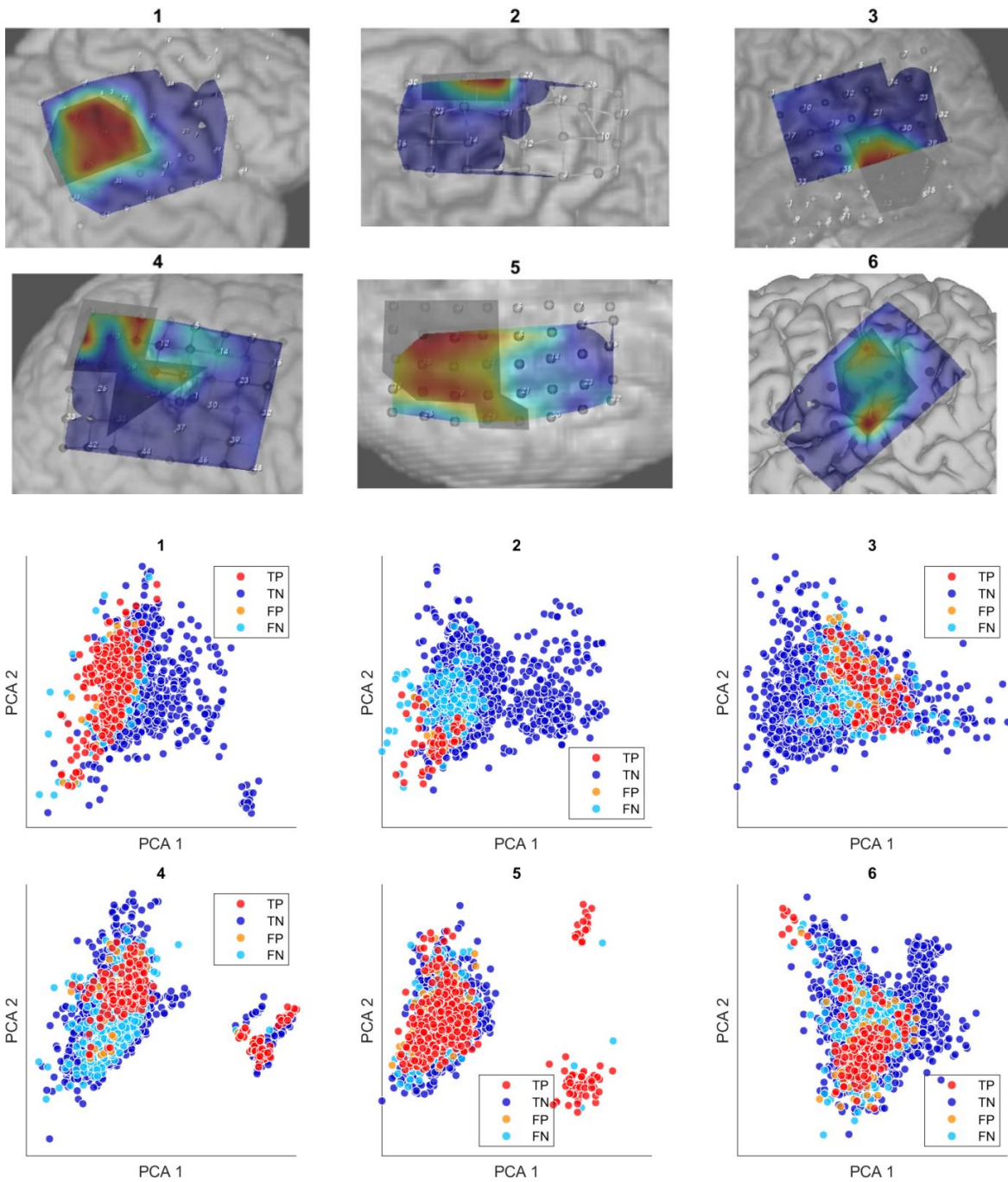


Figure 7 – visualization of **Gaussian** kernel SVM classification combining **all** spectral features. On the reconstructed anatomical images, grey shading marks the clinical SOZ. Rate of epileptic classification of epochs originating from ECoG electrodes indicated with coloring with red zones representing the highest value. Below, individual epochs visualized in PCA space, differentiating true positive (TP), true negative (TN), false positive (FP), and false negative (FN) categories.

4.2. Predictome-based transcriptome analysis highlights functionally related genes

Ranked gene lists delivered by various methods were analyzed using the SEPA method. Effects of algorithmic choices (normalization method, SVM kernel, inclusion of interactome data) on the functional interpretability of detected genomic changes were explored, quantified as the abundance of enriched pathways. We focused on data corresponding to the second wave of aging (50-59y to 60-69y) for this exploratory step due to sample size limitations (early aging) or weaker genomic coverage (blood plasma) in other datasets.

Regardless to the used SVM kernel and normalization method, extensive interactome areas could be highlighted in the prediction-weighted network. The number of proteins associable to nodes with significantly changed Eigenvector Centrality values ranged from 1411 to 1925, covering 8.82-12.03% of the 16005 PPI nodes. RMA renormalization raised those values to 1647-2530 (10.29-15.81%) indicating the increased sensitivity of the refined analysis.

Inclusion of Protein-Protein Interactome (PPI) information seemingly guided the selection process towards enhanced interpretability in context of a priori biological knowledge curated in the Reactome database. Depending on parameter settings, the integrative analysis yielded 140 to 378 significant FDR-corrected enrichment probabilities, highlighting 6.39-17.26% of the Reactome namespace (n=2190).

Considering network topology, so the information stored in the nascent PPI graph describing the co-dependent relationships of the studied genomic variables, found to be instrumental to reach that kind of semantic abundance. We have also examined the available transcriptomic data on a single-gene basis and ranked the associable proteins according to model predictive performance. Such process favors proteins with corresponding RNA levels showing straightforward correlation to group affiliation in case of linear kernel models; more

complicated relationships, such as U-shaped responses can be handled by the Gaussian kernel. Using those ranked lists, sets were selected with sizes matching the ones highlighted by the integrative method. Despite the obvious gene-level correlations with aging, the abundance of enriched Reactome terms within those sets was strikingly low, especially for the cerebellum. Interestingly, linear SVM outperformed the Gaussian kernel variant regardless of data normalization method, reaching a maximum of 46 enrichment hits (2.1% of the Reactome namespace). Results summarized in Table 8.

Table 8 – Abundance of significantly changing genes and highlighted Reactome terms within those sets. Inclusion of Protein-Protein Interactome information enhanced the Reactome-based interpretability of the highlighted items.

database	renormalized with RMA	changing genes (pMC < 0.05)		enriched Reactome terms			
		linear	Gaussian	without PPI		with PPI	
				linear	Gaussian	linear	Gaussian
Cerebellum #1	yes	2115	1708	0	0	232	197
Cerebellum #2	yes	1959	1647	2	1	316	293
Frontal Cortex #1	yes	2530	2313	46	0	351	140
Frontal Cortex #2	yes	2305	2401	4	11	378	258
Cerebellum #1	no	1929	1796	2	0	140	367
Cerebellum #2	no	1411	1925	0	0	332	368
Frontal Cortex #1	no	1773	1577	21	0	317	186
Frontal Cortex #2	no	1849	1575	43	10	358	277

4.3. Interactome information enhances reproducibility of Pathway Analysis results

Similarly, the second wave of aging has been studied to gather insights on parameter choice effects on robustness. Regardless of the inclusion of PPI information, linear kernel SVM models produced a Jaccard similarity of 0.29-0.31 across highlighted gene sets, whereas Gaussian kernels reached a value of 0.13-0.18 in this respect when using RMA normalized data. Without renormalization, those values were more modest, ranging from 0.19 to 0.24 in case of the linear and from 0.13 to 0.14 with the Gaussian kernel. Importantly, linear SVM provided a better agreement across repeated measurements in all cases.

Interpretability of gene lists derived with or without integrating the PPI information to the workflow showed stark differences. Robustly enriched terms were virtually absent with traditional analysis. On the other hand, the integrative approach showed much higher replicability both within and across normalization protocols: term set Jaccard similarities for repeated measurements always exceeded the corresponding values delivered for individual genes. Again, linear SVM surpassed the Gaussian kernel in terms of consistency. Mean Jaccard similarity across tissues reached 0.38 with and 0.28 without RMA renormalization, whereas Gaussian model performances averaged only between 0.21 to 0.23 (see Table 9).

As reproducibility poses a key challenge on the field, we decided to use this verification step to empirically select the most robust normalization and kernel combination to be employed for detailed analysis of the available genomic datasets. Therefore, in the next chapters, we are focusing on the results delivered by linear SVM models on data refined by RMA normalization.

Table 9 – Reproducibility of highlighted genes and Reactome terms across redundant measurements in the GTEx transcriptomic database. Cerebellar and cortical brain samples

were re-sampled with approx. 1 mm spatial differences and processed via similar protocols by independent laboratories. For RMA normalization, raw readouts were reprocessed. Transcriptome data was projected to nodes (proteins) in the unified Protein-Protein Interaction (PPI) network. SVM prediction of age range (50-59y versus 60-69y) was performed using information corresponding to singular gene expression or to pairs of interacting elements in the PPI network – usage of the PPI data was noted in the latter case. Gene similarity values indicating the robustly highlighted nodes of the PPI network and Reactome values are referring to enriched term sets within these sets.

tissue (2x)	renormalized with RMA	use PPI	SVM kernel	genes		Reactome terms	
				overlap	Jaccard	overlap	Jaccard
Cerebellum	yes	yes	linear	955	0,31	169	0,45
Cerebellum	yes	yes	Gaussian	389	0,13	73	0,18
Frontal Cortex	yes	yes	linear	1044	0,28	173	0,31
Frontal Cortex	yes	yes	Gaussian	648	0,16	78	0,24
Cerebellum	yes	no	linear	972	0,31	0	0,00
Cerebellum	yes	no	Gaussian	423	0,14	0	0,00
Frontal Cortex	yes	no	linear	1098	0,29	1	0,02
Frontal Cortex	yes	no	Gaussian	725	0,18	0	0,00
Cerebellum	no	yes	linear	532	0,19	89	0,23
Cerebellum	no	yes	Gaussian	470	0,14	148	0,25
Frontal Cortex	no	yes	linear	664	0,22	163	0,32
Frontal Cortex	no	yes	Gaussian	371	0,13	77	0,20
Cerebellum	no	no	linear	627	0,23	0	0,00
Cerebellum	no	no	Gaussian	436	0,13	0	0,00
Frontal Cortex	no	no	linear	702	0,24	13	0,25
Frontal Cortex	no	no	Gaussian	360	0,13	0	0,00

4.4. Interactome information enhances robustness and abundance of pathway enrichment in top gene sets

As we have seen in the previous chapter, aging results in profound changes of the transcriptomic landscape. The genes of significantly altered expression levels or significantly changed centrality in the Predictome network are numerous. This level of abundance, however, could hinder interpretation: ranking on a finer scale is needed to pinpoint the most important elements instead of highlighting roughly 10% of the genome.

In this merit, we explored the robustness of ranking across analyses stemmed from repeated laboratory measurement data. For Predictome results, primary ranking was performed according to Monte Carlo probability of Eigenvector Centrality (pMC) and ties resolved according to single-gene SVM performance.

The issue of reproducibility turned out to be especially severe when only a small number of critical genes were selected and somewhat ameliorated as the set size increased, implying considerable stochasticity of gene ranks across repeated experiments. Interestingly, Predictome approach enhanced the robustness of gene selection predominantly when working with smaller set sizes, sometimes doubling the overlap across repeated measurements compared to single-gene analysis.

Interpretability of gene lists derived with or without the Predictome showed stark differences with pathway analysis. When relying only on gene expression data to define top gene sets, significant Reactome term enrichments were scarce and heavily cutoff-dependent. The maximum number of highlighted pathways was 12 and importantly, neither of these results could be replicated. Incorporating network information yielded more voluminous and balanced interpretability with considerable reproducibility across repeated measurements (Figure 8).

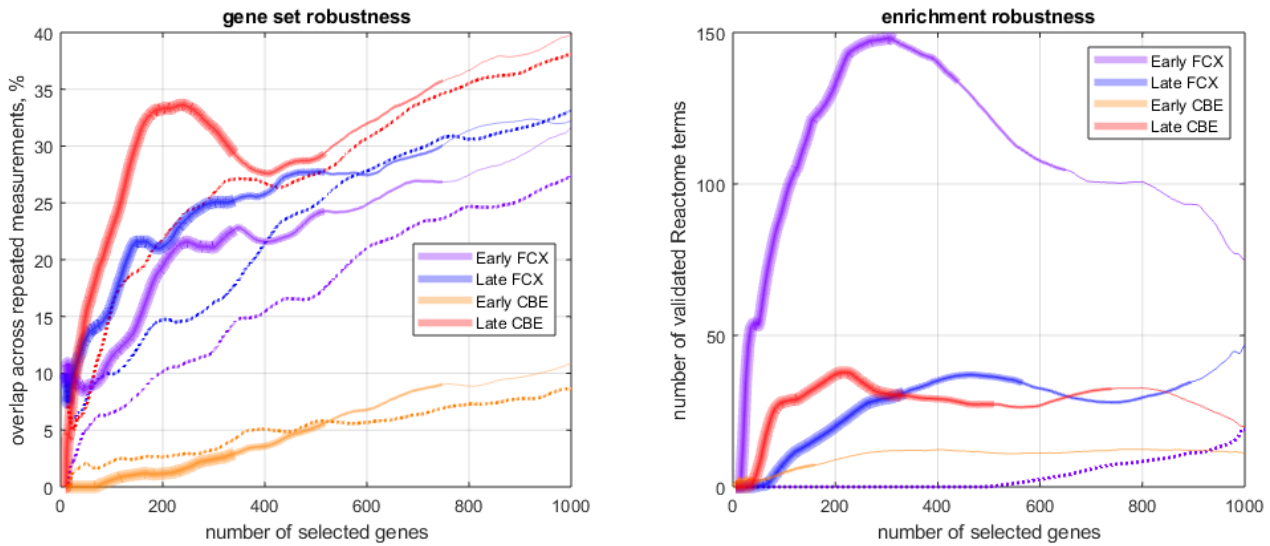


Figure 8 – Reproducibility of gene sets and enriched Reactome term sets within them as a function of selected gene set size. **Solid lines**: genes were ranked using the **Predictome** approach; **dotted lines**: only **singular gene expression** was used for ranking. Decreasing line thickness indicating the points where Monte Carlo probabilities of Eigenvector Centrality elevation (pMC) in the Predictomes are reaching 0.00001, 0.0001 and 0.001, respectively. For both samples in the frontal cortex (FCX) and for late aging in the cerebellum (CBE), the Predictome approach stabilizes gene selection across repeated experiments and enables more robust critical set definition especially when working with stringent cutoffs. However, reproducibility suffers considerably when the group balances are compromised as we see in the case of early cerebellar aging. Differences between the two approaches are becoming even more pronounced when the results are projected to the enriched term space. Without Predictome, no significant pathways could be confirmed after FDR correction. In contrast, the network-based technique yielded considerable reproducibility of biological interpretation.

Taken together, Predictome models based on linear SVM were required to study functional attunement of gene sets of reasonable size in a reproducible manner. This effect can be linked to the enhanced sensitivity of the approach and its tendency to select functionally related gene groups. Inclusion of network information apparently ameliorates stochasticity of individual gene measurements, supposedly by introducing an adaptive filtering on the detected transcriptomic changes operating along interactome connectivity. Functional regions of the interactome tend to form densely connected regions (modules). Given the interlinked nature of elements populating those regions, highly predictive genes on phenotype could elevate the score of their interacting partners, thereby effect the ranking of the whole functional module.

4.5. SVM unmask biological information embedded in gene product interactions

In the previous chapter, we have shown the enhanced effectivity of the Predictome approach to select functionally related gene sets associable to phenotype in a robust manner. This effect could be partially achieved by the spread of node importance to adjacent neighbors in the interactome network. If expression level of a gene correlates strongly with phenotype, all models involving this gene – like with all its interaction partners in the PPI – will have good predictive performance on phenotype. Here, we asked whether all the benefits of the Predictome approach could be explained by this effect, or the machine learning algorithms can extract extra information on phenotype, inaccessible to methods operating only on singular gene expression levels.

Strikingly, we found that in multiple occasions link based SVM models could predict group affiliation even if the genes associated to the link in question did not show altered expression when analyzed separately. In multiple cases, not the gene expression levels themselves, but their ratio was informative. Figure 9A illustrates how this integrated analysis can unmask

information by combining pairs of genes. For example, the expression of TUBB4B and TUBG1 which are two Tubulin genes, major constituent of microtubules, did not change during aging but their relative concentration shifted from TUBG1 to TUBB4B dominance in older individuals. In fact, the complementary information brought by connected genes improved age prediction for 43,000+ interactions (22-23% of the literature-based PPI network, Figure 9B).

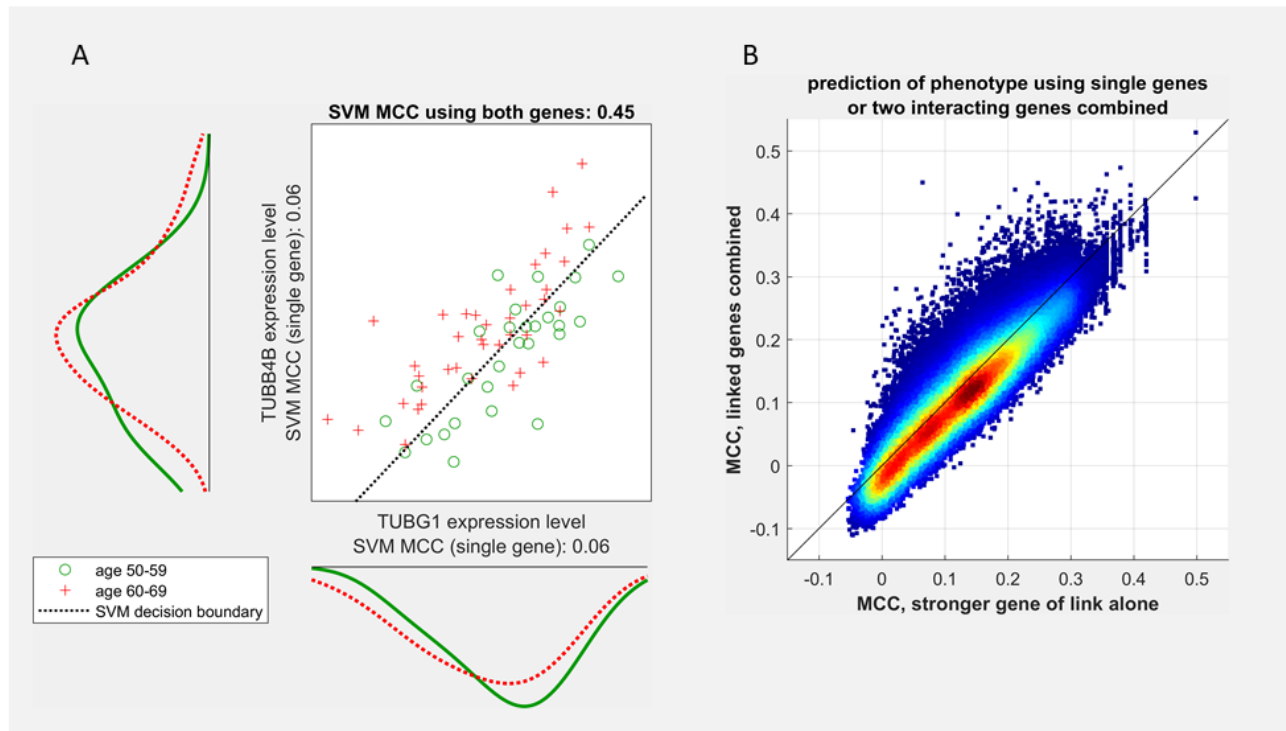


Figure 9 – **A**) PPI-driven linear SVM models reveal higher order relationships between gene products and phenotype. In our example, middle-aged and older individuals can be separated along the expression ratio of two tubulins in the neocortex. Taken separately, levels of TUBB4B and TUBG1 are not correlated with aging. However, their relative concentration shifts from TUBG1 to TUBB4B dominance in older individuals – a feature that can be exploited by the linear SVM model when given access to expression levels of both genes simultaneously. SVM performance was quantified using Matthews Correlation Coefficient (MCC) between prediction of group affiliation and ground truth. **B**) Performance of SVM models combining gene expression information corresponding to linked genes, compared to performance achievable using the linked genes separately (neocortex, later stage of aging). In case of the instances below the diagonal of equal performance, no extra information can be extracted by combining the linked genes: the increased model complexity compromises generalization and the measured performance of the model drops. However, massive performance gain upon combining expression information of the linked genes is a feature exhibited by a considerable fraction (22-23%) of all interactions.

4.6. Genes influenced by Brain Aging and Parabiosis are forming a continuous network in the Interactome

As mentioned in the previous chapters, biological functions tend to be associated to interacting genetic elements. These continuous subnetworks could represent pathways of related biological processes. Here, we wanted to see whether aging-related changes are emerging in an orchestrated way in the Interactome or are they affecting random, disjunct subnetworks in a more stochastic fashion. Also, our goal was to visualize the most important aging genes selected by our novel integrative approach.

Inclusion criteria for visualization and annotation were the following. First, the most robust aging genes were picked, reaching repeatedly significant Eigenvector Centrality change in at least two redundant database pairs. Therefore, confirmed association with both aging stages or with both brain areas was required. Then, from the two replications, the weaker pMC was considered for each gene. Early Brain Aging and Late Brain Aging scores were calculated as the means of these values across frontal cortex and cerebellum. Second, genes influenced by Parabiosis (either by provoked aging in young animals or rejuvenation in old animals) were added to this list. As changes of opposite direction are expectable in this setup, the stronger pMC value was used as a Parabiosis score. Genes annotated in detail if they have reached significant score in at least two out of the three categories (Early Aging, Late Aging, Parabiosis).

Within these constraints, 658 genes were selected for visualization and 21 genes for detailed annotation. Interestingly, 635 of the corresponding proteins are forming a densely connected, continuous subnetwork with 2751 interaction links, indicating their functional relatedness (Figure 9). Detailing the most important elements of this network, we present the pMC statistics of the annotated genes from Figure 10 in Table 10.

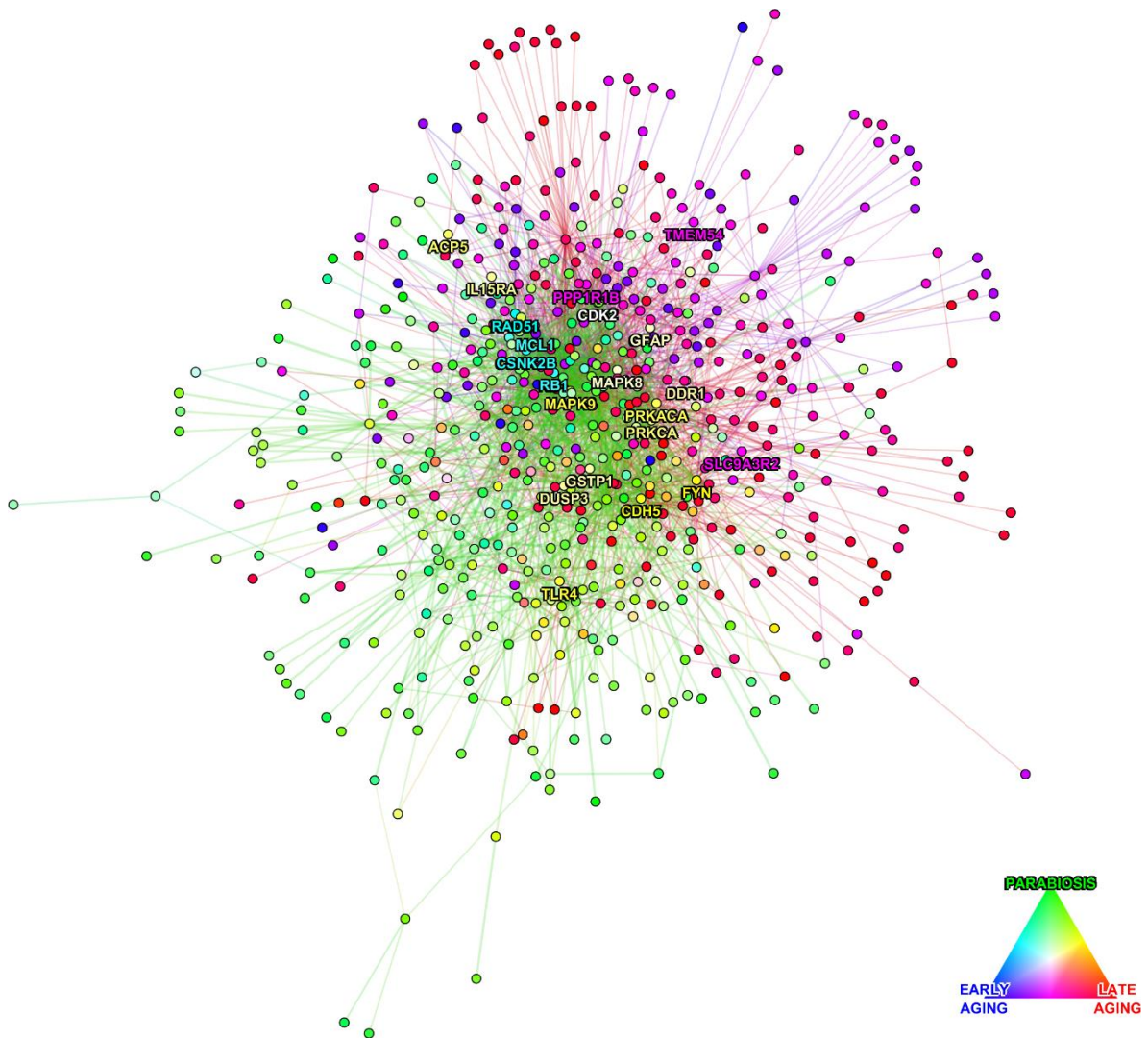


Figure 10 – Genes associable to brain aging in a robust manner and to proteins influenced by heterochronic parabiosis. Colored nodes representing genes relateable to the studied phenomena based on their elevated Eigenvector Centrality (pMC) in the phenotype specific Predictomes. Lines symbolizing curated interactions of corresponding proteins in the PPI database, with colors showing performance of linear SVM models predicting phenotypes from data corresponding to the linked elements. Aging- and parabiosis-related genes are forming a densely connected, continuous subnetwork of 635 nodes and 2751 links within the Interactome. 23 unconnected nodes were omitted for visual clarity.

Table 10 – Annotated genes (elements linked to multiple phenotypes) from the aging- and parabiosis-associated subnetwork. For aging phenotypes, the higher pMC values (weaker significance) are shown across the two repeated measurements. Zero values for aging phenotypes indicating that the Eigenvector Centrality of the given node surpassed all 100 000 surrogate networks in both sample repeats. In case of Parabiosis, pMC values delivered from the provoked Aging experiment (young animal exposed to old blood) and from the Rejuvenation experiment (old animal exposed to young blood). Co-occurrence in PubMed abstracts with words “Aging” or “Age-related” noted in the last column.

phenotype / gene	Early Aging		Late Aging		Parabiosis		PubMed references (n)
	FCX	CBE	FCX	CBE	Aging	Rejuv	
GFAP	0,04618	0,89246	0	0	0,02041	0,00629	1078
DUSP3	0,01713	0,84682	0,00573	0,00024	0,44025	0	2
GSTP1	0,00212	0,98645	0,00413	0,00011	0,6301	0	114
ACP5	0,99866	0,84459	0,01412	0,03001	0,13265	0	15
MAPK9	0,75672	0,9692	0	0,00001	0,11141	0	3
TLR4	0,9965	0,80845	0	0,00007	0,24302	0,00515	357
PRKACA	1	0,67405	0	0,14078	0,99297	0,0215	3
CDH5	0,99885	0,9027	0,14664	0,00005	0,82445	0	4
CDK2	0,00791	0,00013	0	0	0	0,99908	119
CSNK2B	0	0,02298	0,99877	0,9101	0	1	0
MCL1	0,00718	0,0138	0,92404	0,95179	0,00154	0,98634	26
RB1	0,00309	0,00196	1	0,88581	0,03022	0,95399	80
RAD51	0	0,18618	0,99863	1	0,00104	0,99714	94
MAPK8	0,00207	0,5909	0	0	0,00012	0,74381	55
DDR1	0,318	0,96329	0	0,00374	0,04086	0,64804	7
FYN	1	0,99991	0	0	0,01434	1	65
PRKCA	1	0,61468	0,00082	0	0,00173	0,20939	20
IL15RA	0,51837	1	0,08654	0,00079	0,00564	0,94887	6
TMEM54	0,02205	0,0209	0,00406	0	1	1	0
SLC9A3R	0,0277	0,14006	0,00198	0,11871	1	1	0
PPP1R1B	0,10612	0,00661	0,12745	0	1	1	15

A control analysis using the STRING database (Szklarczyk et al., 2017) confirmed the existence of the highly connected aging subnetwork. This alternative repository of PPI information has wider coverage but in general, less reliable than the sources we used as it incorporates predicted and supposed interactions as well as experimentally confirmed links. Curated connections outfitted with a heuristic reliability parameter which can be tuned. At medium confidence (standard usage of the STRING web interface) our aging- and parabiosis-related genes formed a dense network with 7036 links. Number of connections decreased when using higher confidence cutoffs, resulting in 2989 links with high and 2177 links with highest confidence. Regardless of the parameter settings, enrichment for interactions within the selected set was highly significant ($p < 10^{-16}$ in all cases) indicating the presence of the aging subnetwork.

Literature mining confirmed the relevance of both the broaden set presented in Figure 9 and the narrower list detailed in Table 10. From the 317 genes mutually listed in the GenAge and our PPI databases, 62 was part of the highlighted aging network ($p_{\text{Fisher}} = 3.3 \cdot 10^{-25}$). Mutual appearance in PubMed abstracts with keywords “*Aging*” ($p_{\text{Wilcoxon}} = 1.5 \cdot 10^{-78}$) and “*Age-related*” ($p_{\text{Wilcoxon}} = 8.7 \cdot 10^{-67}$) was more frequent compared to another genes. The annotated subset was also enriched in GenAge elements ($2.5 \cdot 10^{-6}$) and overrepresented in the literature for “*Aging*” ($p_{\text{Wilcoxon}} = 7.5 \cdot 10^{-7}$) and “*Age-related*” ($p_{\text{Wilcoxon}} = 1.3 \cdot 10^{-7}$).

Taken together, Predictome analysis of aging produced results in line with the literature. Also, our results are suggesting that the aging associated Interactome changes are occurring in a highly orchestrated manner.

4.7. Pathway Analysis reveals shared aspects of Brain Aging and Parabiosis

Comparative analysis of the single-gene and Predictome-based approaches revealed the most profound differences in terms of functional enrichment abundance and stability across repeated measurements. In fact, replicable elements on single-gene basis were scarce, implying low level of reliability. Therefore, we aimed to give a broad overview of the aging-associated biological processes suggested by our novel, presumably more robust method. In this merit, we visualized the hierarchal tree of Reactome terms as an unweighted graph using the force-directed layout (Fruchterman & Reingold, 1991) which represents functionally related or embedded biological processes as adjacent nodes. Aging-related elements were highlighted in this semantic network in Figure 11.

271 pathways were selected based on their robust significance across repeated transcriptomic measurements of human brain aging or their association to either rejuvenation or provoked aging in the parabiotic state. From these items, 5 found to show universal association to all human brain aging datasets and furthermore, influenced by the parabiotic state. “*Gene expression (Transcription)*”, “*Generic Transcription Pathway*”, “*RNA Polymerase II Transcription*” are umbrella terms denoting widespread changes of the gene expression machinery, however, “*SUMOylation*” and “*SUMO E3 ligases SUMOylate target proteins*” are more specific hits, emphasizing the importance of regulation through small ubiquitin-like modifier proteins in the aging process. Four Reactome terms found to be robustly, but exclusively enriched in early aging and all of them were related to “*mRNA Splicing*”. The exclusive set of late aging consisted of 108 elements describing various biological concepts from “*Hemostasis*”, “*Axon guidance*” and “*Integration of energy metabolism*” to widespread examples of innate immune functionality, including “*Toll-like Receptor Cascades*”, “*Signaling by Interleukins*” and “*Inflammasomes*”. The parabiosis-specific set was populated by 145 items with emphasis on “*Cell Cycle*”, “*DNA Repair*”, “*Metabolism of proteins*” and on “*Terminal pathway of complement*”, among others.

Interestingly, the later phase of human brain aging was characterized by multiple key pathways which in turn were also affected by the parabiotic state. This intersection consisted of 9 elements: “*Apoptosis*”, “*Programmed Cell Death*”, “*Cellular Senescence*”, “*Signal Transduction*”, “*Signaling by Non-Receptor Tyrosine Kinases*”, “*Transcriptional regulation by RUNX2*”, “*Signaling by PTK6*”, “*Signaling by TGF-beta family members*” and “*Signaling by TGF-beta Receptor Complex*”.

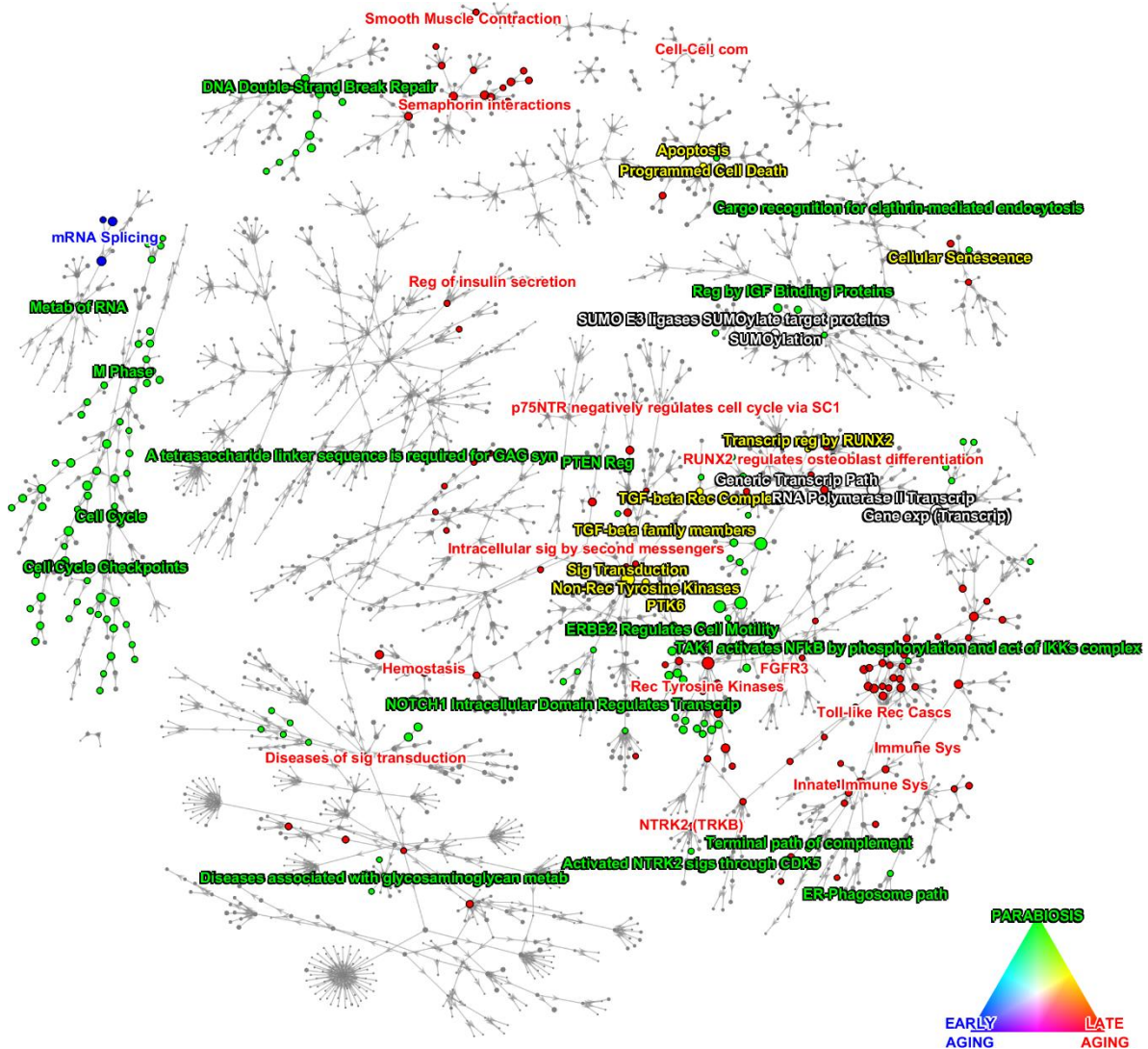


Figure 11 – Reactome terms associable to brain aging in a robust manner and to proteins influenced by heterochronic parabiosis. Colored nodes representing pathways relatable to the studied phenomena based on SEPA analysis in the phenotype specific Predictomes. Lines symbolizing hierarchal, embedded relationships within Reactome.

Table 11 – Annotated pathways in provoked Parabiotic Aging and Rejuvenation and their level of significance in various phenotypes

phenotype / Reactome	Early Aging		Late Aging		Parabiosis	
	FCX	CBE	FCX	CBE	Aging	Rejuv
Signaling by PTK6	0,308901	0,778412	0,027576	0,000171	0,228307	0,005993
Signaling by Non-Receptor Tyrosine Kinases	0,308901	0,778412	0,027576	0,000171	0,228307	0,005993
ERBB2 Regulates Cell Motility	0,999998	0,999896	0,176781	0,144783	1	0,000705
Terminal pathway of complement	1	1	0,89668	0,996631	0,760494	0,009096
Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)	0,73741	0,941316	0,739409	0,894856	0,470105	0,009681
Cargo recognition for clathrin-mediated endocytosis	0,142846	0,999926	0,56716	0,232165	0,391144	0,021765
Diseases associated with glycosaminoglycan metabolism	1	0,885389	0,64055	0,453806	1	0,043045
A tetrasaccharide linker sequence is required for GAG synthesis	0,967559	0,875458	0,6086	0,30827	1	0,049615
RNA Polymerase II Transcription	1,69E-14	5,89E-08	0,017797	0,002455	1,75E-06	0,864056
Gene expression (Transcription)	3,02E-14	5,7E-08	0,02268	0,007328	1,82E-06	0,868246
Generic Transcription Pathway	1,21E-10	5,7E-08	0,005318	0,000481	6,17E-06	0,864056
SUMOylation	9,08E-07	0,026086	0,00747	0,004153	0,000474	0,397644
SUMO E3 ligases SUMOylate target proteins	4,08E-07	0,027102	0,005593	0,003283	0,000474	0,397644
Signaling by TGF-beta Receptor Complex	0,00034	0,46144	0,01994	0,030514	0,005791	0,645971
Signaling by TGF-beta family members	0,003509	0,48529	0,00719	0,011512	0,030987	0,535967
Apoptosis	8,37E-05	0,384223	0,001968	0,003426	0,018656	0,958091
Programmed Cell Death	6,13E-05	0,387947	0,001365	0,004712	0,01915	0,917063
Cellular Senescence	0,001342	0,360332	1,58E-06	0,007685	0,019964	0,728074
Signal Transduction	0,027697	0,267301	6,12E-09	8,86E-11	0,040277	0,247336
Transcriptional regulation by RUNX2	1,95E-06	0,310832	0,017813	0,026917	0,048463	0,718793
Cell Cycle	1,08E-07	0,469739	0,009039	0,496917	0,000101	0,872208
DNA Double-Strand Break Repair	0,033965	0,804147	0,045242	0,674786	0,000309	1
PTEN Regulation	1,58E-10	0,190683	0,001311	0,179445	0,044848	1
Cell Cycle Checkpoints	0,00014	0,61761	0,091627	0,520273	6,87E-06	0,996139
M Phase	0,00011	0,522147	0,649461	0,80767	0,003906	0,914165
ER-Phagosome pathway	0,005837	0,914248	0,11331	0,533155	0,01061	0,952438
Metabolism of RNA	5,28E-23	0,232475	0,808288	0,809515	0,015474	0,997186
NOTCH1 Intracellular Domain Regulates Transcription	0,258518	0,78444	0,000424	0,360141	0,022427	1
Activated NTRK2 signals through CDK5	1	1	0,163903	0,450414	0,028401	0,700729
TAK1 activates NFkB by phosphorylation and activation of IKKs complex	0,121503	0,999247	0,067181	0,105702	0,033917	0,998621

Table 12 – Annotated pathways in Brain Aging and their level of significance in various phenotypes

phenotype / Reactome	Early Aging		Late Aging		Parabiosis	
	FCX	CBE	FCX	CBE	Aging	Rejuv
mRNA Splicing	9,37E-09	0,00016	0,884176	0,733447	0,055536	1
Diseases of signal transduction	0,016199	0,305918	0,009725	0,02808	0,280092	0,603771
Intracellular signaling by second messengers	6,78E-05	0,318885	2,67E-05	0,000237	0,18672	0,467033
Toll-like Receptor Cascades	0,031593	0,359231	2,26E-05	0,000112	0,327274	0,265921
Cell-Cell communication	0,760876	0,901478	0,003286	0,030476	0,741707	0,697839
CTLA4 inhibitory signaling	0,947923	0,897215	0,000791	0,013573	0,459076	1
Hemostasis	0,773219	0,488919	0,000259	6,16E-06	0,47967	0,537665
Innate Immune System	0,546283	0,617656	1,4E-05	3,49E-05	0,805397	0,31436
p75NTR negatively regulates cell cycle via SC1	1	0,999896	0,022963	0,025734	1	1
Regulation of insulin secretion	0,894596	0,999926	0,015595	0,007189	0,445943	0,823791
RUNX2 regulates osteoblast differentiation	0,23639	0,258471	0,002312	0,000203	0,727638	1
Semaphorin interactions	0,964065	0,848809	1,06E-05	0,000739	0,269694	1
Signaling by FGFR3	0,062128	0,579173	0,040634	0,022063	0,464291	1
Signaling by NTRK2 (TRKB)	0,482369	0,999951	0,006223	0,003039	0,161803	0,894002
Signaling by Receptor Tyrosine Kinases	0,126397	0,579985	1,07E-07	6,45E-07	0,337487	0,198806
Smooth Muscle Contraction	0,704147	0,642998	0,00216	0,002072	1	0,854852

Further investigating the intervened network of similarities across area- or phase-specific aspects of aging and pro- or anti-aging alignments of parabiosis, a structured web of overlaps could be found (Figure 12). The early wave of aging was characterized with more abundant changes in the frontal cortex compared to cerebellum; a discrepancy largely eliminated during the second wave. In line with previous findings describing human plasma proteomic changes (Lehallier, Gate, Schaum, Nanasi, Eun Lee, et al., 2019), only a minority of factors found to be shared across the two waves of aging, implying similar, non-linear alternations of the brain transcriptomic landscape.

Effects elicited by the exposure to an old circulatory system in young animals were more widespread compared to the opposite direction of heterochronic parabiosis, Rejuvenation. The changes related to Provoked Aging shown extensive similarities with early aging in the frontal cortex, and for the second wave the strong preference towards frontal cortical patterns compared to cerebellar ones remained.

Rejuvenation effects were relatively isolated, with 23 out of 29 elements being exclusive to that phenotype, including “*EGFR downregulation*”, “*Clathrin-mediated endocytosis*” and “*Terminal pathway of complement*”. Two terms, “*Metabolism of proteins*” and “*Downregulation of ERBB2 signaling*” were shared with the early wave of aging in the frontal cortex. “*Signaling by Non-Receptor Tyrosine Kinases*”, and “*Signaling by PTK6*”, pathways characteristic to the later wave of aging in both investigated brain areas, were found to be affected by Rejuvenation. Further similarities with this wave were “*Signaling by ERBB4*” in the frontal cortex and “*SHC1 events in ERBB2 signaling*” in the cerebellum.

Literature mining results were confirmative. The 271 pathways linked to aging or to parabiosis by the Predictome approach (Figure 10) were extensively referenced in previous studies on aging or age-related conditions. Co-occurrences with words “*Aging*” ($p_{\text{Wilcoxon}} = 6 \cdot 10^{-5}$) and “*Age-related*” ($p_{\text{Wilcoxon}} = 6.7 \cdot 10^{-5}$) in PubMed abstracts were both significant. Central terms picked for annotation (Tables 11 and 12) showed even more substantial attunement ($p_{\text{Wilcoxon}} = 10^{-9}$ and $p_{\text{Wilcoxon}} = 1.2 \cdot 10^{-11}$), further solidifying coherence of our results with a priori knowledge.

Comparison of highlighted Reactome term sets (Jaccard similarities and overlaps)


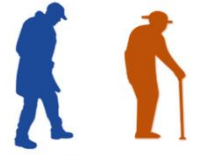
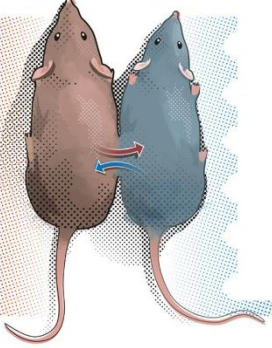
Early FCX	336	0.03	0.16	0.11	0.01	0.20	
Early CBE	9	10	0.02	0.02	0.00	0.04	
Late FCX	78	5	238	0.35	0.01	0.13	
Late CBE	57	5	122	232	0.01	0.04	
Rejuvenation	2	0	3	3	29	0.01	
Prov. Aging	79	5	41	15	1	131	
	Early FCX	Early CBE	Late FCX	Late CBE	Rejuvenation	Prov. Aging	

Figure 12 – overlaps (green) and Jaccard similarities (red) across the Reactome term sets associated to investigated phenotypes. For frontal cortex (FCX) and cerebellum (CBE) datasets, significant enrichment for both repeated measurements were required. Early and later waves of human brain aging were compared with plasma proteomic changes elicited by heterochronic parabiosis in old animals (Rejuvenation) and young animals (Provoked Aging). Images: Nik Spencer/Nature, (Scudellari, 2015).

5. DISCUSSION

5.1. Electrophysiology and Seizure Onset Zone detection

It has been shown previously that the Seizure Onset Zone (SOZ) can be characterized by distinct features of the ECoG signal. Analysis of graphs defined using various similarity or correlative metrics across signals from selected electrode pairs turned out to be useful tool to highlight SOZ. In turn, the feasibility of these approaches indicates alternations of epileptogenic foci in terms of generating and reacting to periodic signals both in the amplitude and in the phase space.

A previous analysis by the author (File et al., 2020) relied on external stimulation of brain tissue hence introduced well-controlled activation to the system. Altered reactivity to incoming signals or altered connections spreading local signals can be related to epilepsy intuitively. There, we found distinct graph topological changes specific to the SOZ and its connections to surrounding tissue in a response to direct electrical cortical stimulation.

Another study (Nánási et al., 2016) involved resting state ECoG recordings. To separate the observations from stimulus processing and seizures to an even greater extent, we opted to use data obtained during deep non-REM sleep, a technique which has been carried over to this thesis. The notable within-subject uniformity of the non-REM stage was expected to further enhance the robustness of our results and to emphasize the role of “baseline” connectivity patterns in absence of internal or external perturbations, such as sensory stimuli or mind state (Pótári et al., 2017; Ujma et al., 2019, 2020). The former analysis relied on various feature extraction methods from the ECoG signal while operating in the amplitude and phase spaces together, combining the two via previously described and novel heuristic metrics. The observed superiority of SOZ detection of approaches integrating multiple

subsets of information delivered by processing the ECoG signal emphasizes the complex nature of epileptogenic alternations. Among the tested algorithms the best performer turned out to be a heuristic method combining slow frequency information with a novel heuristic measurement delivered from a multi-band filtering approach.

In this work, our aim was to build on these previous findings and to test whether a more sophisticated machine learning approach could extract relevant information on SOZ localization using relatively straightforward and local features, such as band power measured on individual ECoG channels. Hence, integration steps fusing multi-channel information are now performed by the machine learning algorithm itself and not presented intrinsically by higher order features, like couplings, which are in turn, somewhat heuristic concepts themselves. We expected that an advanced algorithm could retrieve comparable information from more primitive features of the ECoG signal.

Data of non-REM sleep was collected for six ECoG patients during the standard diagnostic evaluation preceding curative surgery. ECoG recordings of 3-minute length were carefully selected to obtain visually uniform recordings of deep sleep. Here, we asked whether it is possible to reproduce expert SOZ-selection based on the abundance of various frequency band activities and their relationships to each other, in the non-epileptic (resting state) brain activity. We have chosen Support Vector Machines as a viable compromise between complexity and training data requirements. Also, SVM algorithms can handle problems with unbalanced classes and they cope well with outliers – both situations are characteristic to electrophysiological readouts. Effects of kernel choice and inclusion of multiband frequency information were tested. Regardless of kernel, delta and gamma frequency bands found to be the most important predictors of pathology. However, clinically defined Seizure Onset Zones were replicable with great fidelity only when using a Gaussian kernel SVM model given access to full spectral information.

Epilepsy is linked to brain aging, and brain aging can be manipulated in mouse models by heterochronic parabiosis. Furthermore, changes in deep sleep slow wave activity were linked to aging using scalp EEG recordings (Pótári et al., 2017; Ujma et al., 2019) and in turn, our analysis revealed that the same frequency band contains enough information to reconstruct clinically defined Seizure Onset Zones in 4 out of 6 cases from seizure-free deep sleep ECoG data. However, as the number of patients with intracranial EEG recordings was insufficient to analyze aging, and electrophysiology from parabiosis experiments was unavailable, we had to rely on alternative methods to study brain aging. Therefore, we opted to use publicly available datasets describing gene expression and protein concentration on a genomic scale and proceeded to develop novel bioinformatic tools inspired by the approaches proven to be successful in electrophysiology.

5.2. Molecular biology – method development and validation

To study aging in the molecular level, we applied a novel integrative approach to investigate transcriptomic changes related to two stages of human neocortical aging. Furthermore, we compared our findings with the effects of heterochronic murine parabiosis on the plasma proteome. Our method is based on the analysis of the Predictome, which is defined as an Interactome (in this case: Protein-Protein Interaction) network weighted according to the predictive performance of Support Vector Machine (SVM) models fitted to gene expression data or direct plasma protein concentration measurements corresponding to each link. The significance of Eigenvector Centrality of elements in this weighted graph was quantified using Monte Carlo statistics, summarizing both local and global topological characteristics of the resulting network on a node (protein) level.

The key idea of our PPI-Predictome framework is to transform the unweighted protein-protein interactome network into a weighted graph, where link weights are related to the extractable information on phenotype from measurements corresponding to the linked

proteins. Then, we can highlight critical items in the Predictome using statistics delivered from graph metrics. Graph theory offers an especially well adaptable platform for integrating transcriptomic or proteomic readouts with curated protein-protein interaction information. As the used facet of a priori knowledge describes the elementary building blocks (interactions) of the biological system, it is less biased by constantly evolving conceptual categories than higher level representations (like pathways). The Predictome representation also helps to deal with genes whose importance could only be assumable in context – “guilt-by-association” reasoning like this was already proven to be useful when studying the aging process (De Magalhães & Toussaint, 2004) or other pathologies, like cancer (Módos et al., 2017). Additionally, the used Eigenvector Centrality metric incorporates non-local features of the graph when quantifying gene importance.

Multiple other approaches are present in the literature to exploit a priori knowledge encoded in the Interactome when analyzing high-throughput gene expression. Examples are available from various fields, including cardiovascular diseases (Azuaje et al., 2010; Liu et al., 2016; Nair et al., 2014), cancer (Y. Guo & Xing, 2016; Nibbe et al., 2010; Y. Yu et al., 2015), and immunology (Procaccini et al., 2016). Similarly, application of protein co-expression networks have been developed (Gibbs, Baratt, et al., 2013; Gibbs, Gralinski, et al., 2013) and applied on proteomic data describing problems related to immunology (C. Guo et al., 2014; Wu et al., 2014), neurobiology (MacDonald et al., 2015) and again, cancer (Kanonidis et al., 2016; X. Yu et al., 2016). These methods are, usually, building functional maps of systems corresponding to separate phenotypes by modifying or thresholding the Interactome network on the basis of co-expression measurements, then, following this step, they are comparing the resulting maps to deduct conclusions regarding to the studied phenotypic changes. The Predictome approach is fundamentally different from such techniques as it encodes information differentiating the studied phenotypes directly into the constructed model. Instead of post-hoc comparisons deliverable from secondary measurements, Predictome links are encoding the decisive potential of the connected elements on phenotype explicitly.

Aging can be characterized by widespread, but subtle changes of transcriptomic activity. The caveat of analyzing such complex and heterogeneous data lies mostly in the reliability of the results instead of sensitivity: false positive statements could arise from the sheer disproportionateness of the variable space (genome-wide) and sample size (typically 10 to 1000 subjects). The problem is well-known in the literature and no trivial solutions are available to negate this obstacle (Bellman, 1957, 1961; Trunk, 1979). Therefore, we opted to maximize robustness of our analysis and validated the Predictome approach by employing repeated measurements on the experimental level. Top genes identified using the Predictome showed enhanced robustness when compared to traditional analysis carried out on individual genes. Pathway Analysis demonstrated even more substantial benefits of the novel approach, surpassing the single-gene technique both in abundance and in reproducibility of Reactome term enrichment results.

Furthermore, outputs of the integrative model were highly coherent with knowledge accumulated in the literature. Both on gene and pathway level, items highlighted by the Predictome approach were frequently referred together in PubMed abstracts with “Age-related” alternations or “Aging”. Furthermore, genes enlisted in the GenAge aging database were greatly enriched among critical elements of aging Predictomes.

5.3. Molecular biology – brain aging and parabiosis

Brain aging was found to be characterized by numerous robust alternations of transcription activity, signal transduction and metabolic integration, well reflecting the literature. Also, we observed the strong presence of immune- and vasculature-related pathways together with cellular death, senescence, survival and neural development. To our knowledge, the role of *SCF-KIT*, *ERBB2*, *ERBB4* and *PTK6* signaling was implicated only indirectly in aging and neurodegeneration so far. With the Predictome approach, *SCF-KIT Signaling* found to be significantly altered in the later stage of aging in the frontal cortex and in the cerebellum,

ERBB2 and *ERBB4* Signaling in both late aging sample and in early neocortical aging, whereas *PTK6* was associated to late aging in both brain areas as well as to parabiotic rejuvenation. *NTRK1* signaling (significance of which was found to be more prolific than of the already described *NTRK2*, and which was only connected previously to aging through its role in Alzheimer's Disease) have been highlighted in both stages of neocortical and later stage of cerebellar aging. Among *Axon guidance* related mechanisms, it was novel to see *Sema4D* and *EPHA*, *EPHB* to be affected by age (later stage). Analysis of condition-specific subgraphs revealed the shared role of *Toll-like receptor cascades*, *Interleukins*, *VEGF*, *TGF- β* and various other signal transduction pathways in both aging waves. Previously established findings implying the dominance of non-linear changes in the aging plasma proteome (Lehallier, Gate, Schaum, Nanasi, Eun Lee, et al., 2019) were found to be valid in case of aging brain transcriptomes as well. Despite the presence of shared patterns between the early and later aging waves, this intersection only contains the minority of wave-specific pathway perturbations.

Importantly, the relevance of parabiotic models to study human brain aging could be verified. Young parabionts exposed to old plasma, a phenotype relatable to accelerated aging shared multiple aspects with human brain aging. These features included alternations of *Cell Cycle* mechanisms, *DNA Repair* and *SUMOylation*. Rejuvenation effect elicited by young plasma on old parabionts is known to be more limited which was well reflected by the confined set of associable Reactome terms. However, similarities could be found with both stages of human brain aging including changes in *Metabolism of proteins* and *ERBB2* signaling (early-stage) or *PTK6* and *ERBB4* signaling (late-stage) in a robust manner. Together with recent findings on the intervened nature of systemic milieu and brain aging (Pluinage & Wyss-Coray, 2020; A. C. Yang et al., 2020), those results confirm the importance of plasma measurements and parabiosis to investigate brain aging and related diseases.

5.4. Linear and Gaussian kernels in Seizure Onset Zone detection and in transcriptomics

Various machine learning approaches are demonstrating strong performance when they are tasked to recognize patterns in (biological) data. The choice of Linear SVM for molecular data is based on validation results in repeated laboratory measurements and on the need of the biological interpretability of model performance. More complex methods often exploiting non-linearity, a feature posing a challenge by our relative lack of knowledge on how such complicated concentration relationships could govern the biological function of the given interaction. In the linear case, accurate predictions can typically arise in two scenarios: when the abundance of one or both of the interacting gene products are correlated to phenotype (regardless of the directionality of those changes) or in a more intricate way when the ratio of the expression levels of the two genes are changing across phenotypes. In the latter situation, the synergistic nature of combining data on interacting gene products is explicitly required to uncover embedded relations to phenotype.

The synergistic effect was widespread and found to be present in more than 20% of curated interactions in the investigated protein-protein interaction network when studying the second wave of brain aging. One of the most prominent examples when the Predictome approach could unmask otherwise hidden associations to phenotype was the changing ratio of tubulins associable with neocortical aging. Although individually not correlated with age, the TUBG1 / TUBB4B ratio soundly reflects age group affiliation, with lower relative TUBG1 levels in older individuals. TUBG1 is associated with cortical dysplasias and embryonic stem cell pluripotency, a protein required for cell cycle progression. TUBB4B is associated with Leber congenital amaurosis with early-onset deafness, development (through Slit-Robo signaling) and innate immunity. The relatedness of aging to virtually all these processes makes the two interacting tubulins relevant to the phenotype, regardless their lack of individual correlations to it.

Radial basis function (RBF, or Gaussian) kernel transformations are generally favored when working with highly complex data. Non-linear relationships and XOR logical disjunctions can be solved this way. However, although powerful they are, results achieved by Gaussian kernels are sometimes difficult to interpret. Gaussian kernels are capable to find exotic separator lines by mapping the measurement points to a higher dimensional space where a separating hyperplane can be easily found. Projected back to the (linear) measurement coordinates, these planes can form curved, even folding lines which can be a benefit over the linear approach and enables the Gaussian kernel to solve XOR-like problems. Consequently, regions assigned to a given class in the feature space can be disjunct. This can pose a problem in case of highly stochastic data when novel measurement can contain outliers. In the folded extra-dimensional space in which the Gaussian kernel operates, those previously uncharted regions of the measurement space can be adjacent to unlikely elements (Figure 13).

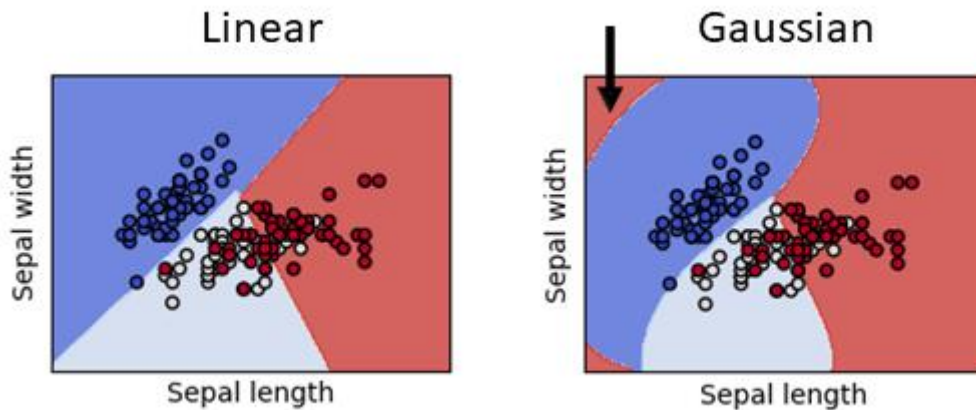


Figure 13 – visualization of Support Vector Machine kernel choice effects on classification of Fisher’s classic Iris dataset (Buitinck et al., 2011; Fisher, 1936; Varoquaux et al., 2015). Note the disjunct area on the Gaussian example (indicated by arrow). Novel, outlier measurement can be classified to the red class, a debatable decision.

This kind of kernel behavior can be both advantageous and harmful. In case of ECoG, our task was to identify hidden epileptogenic activities. Pathological alternations are expected to be somewhat uniform, given the clinical feasibility of expert SOZ detection. On the other hand, healthy brain activity is highly complex and diverse across brain regions and individuals. Electrode arrays sampled different gyri and even spanned across different lobes occasionally. Therefore, the non-SOZ class of observations expected to be non-uniform in the measurement space. From the two kernel variants, only Gaussian possesses the ability to select a circumscribed zone in the feature space when working with two-class problems (Figure 14, panel C).

In contrast, our molecular biology exercises were characterized by the presence of two distinct classes of observations. We cannot expect such dramatic differences in feature variability when comparing two age groups or the two parabiatic treatment setting. In line with those expectations, the linear SVM model outperformed Gaussian kernel SVM when working with those molecular datasets (Figure 14, panels B and D). Also, biological interpretability of the results was rendered more straightforward using the simpler model: strong prediction performance could be the consequence of either altered abundance or different concentration ratio of the linked gene products across phenotypes, as discussed above.

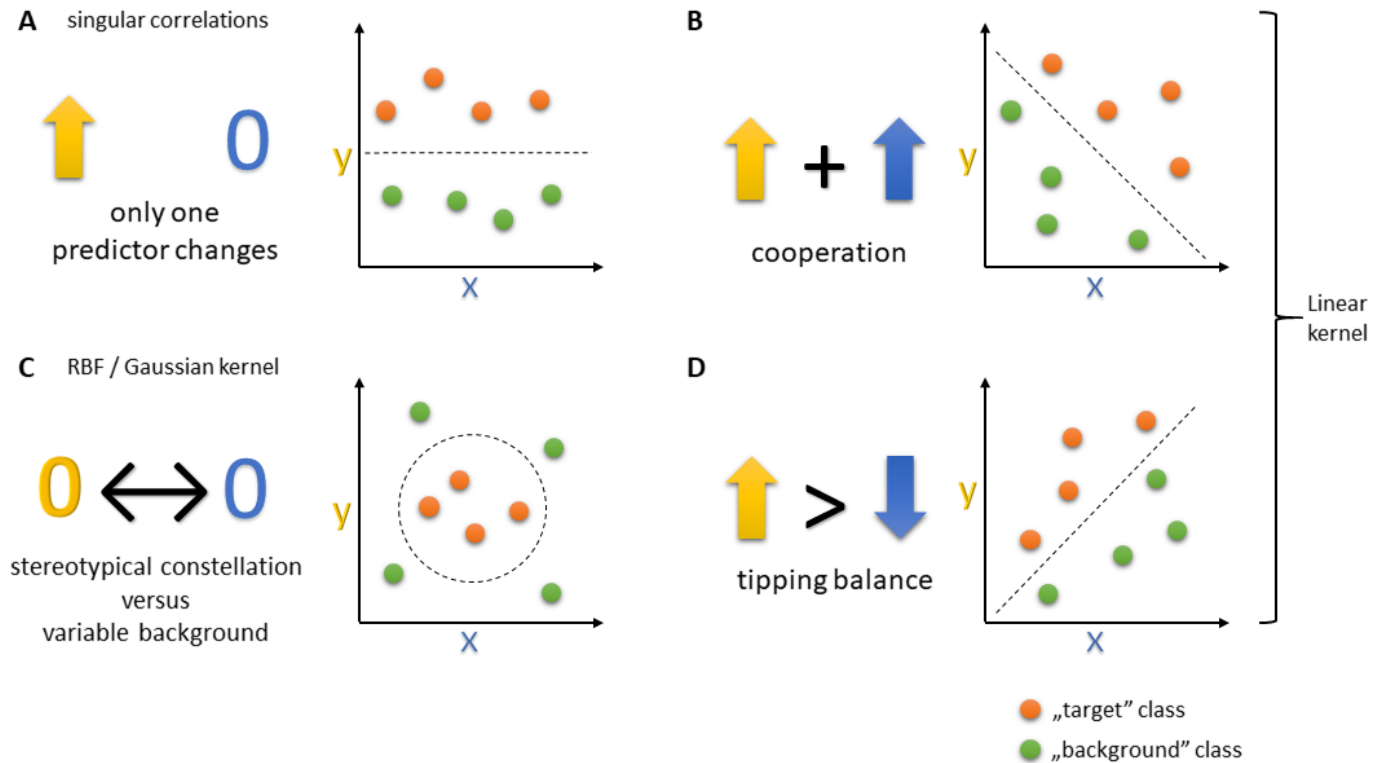


Figure 14 – Variations for feature interactions and the potential of various classification approaches to exploit them. **A)** When one predictor shows clean differences between the compared groups, correlation, Student’s t-test, or Wilcoxon rank sum test is sufficient to detect the feature with predictive potential on phenotype. **B)** and **D)** when predictors are changing in a coordinated way as a function of phenotype and the two classes are comparable in terms of variability, linear kernel SVM can be a good approximation to extract information encoded in feature interactions. In the **B)** case, the summarized abundance of the two predictors gives information on phenotype. In the **D)** case, the relationship can be described as a shift in feature ratio. On the contrary, the situation exemplified on panel **D)** cannot be solved without further transformation of the feature space by RBF or Gaussian kernels. Here, in-class variability differs substantially in the original feature space as a more stereotypical, less variable phenotype must be highlighted against a non-uniform “background”.

6. CONCLUSION

In this work, we have investigated the integrated application of machine learning and graph techniques in two problems related to brain diseases: aging and epilepsy. Despite using different types of input, namely intracranially recorded ECoG signals of deep sleep, post-mortem brain transcriptomics, and proteomic measurements in blood plasma, the combination of techniques proven to be highly synergistic.

In case of ECoG, the author investigated the usability of network-based models (File et al., 2020) and multi-modal feature integration (Nánási et al., 2016) in Seizure Onset Zone localization in previous works. In this thesis, the feasibility of ECoG analysis with Support Vector Machines was demonstrated by reconstructing expert SOZ definitions from seizure-free recordings.

Distinct waves of molecular changes at different stages of life were explored only recently, with contributions from the author (Lehallier, Gate, Schaum, Nanasi, Eun Lee, et al., 2019; Lehallier, Gate, Schaum, Nanasi, Lee, et al., 2019). In the literature, various pathological changes could be linked to aging, including epilepsy. Here, analysis of the transcriptomic changes associable to brain aging and their overlap with plasma proteomic alternations elicited by heterochronic parabiosis was performed. Inspired by the success of machine learning and graph methods in ECoG analysis, a novel framework – the Predictome analysis – was conceptualized to integrate literature data with laboratory measurements carried out on the phenotype of interest.

Predictomes greatly surpassed single-gene analysis in both robustness and interpretability. The proposed approach can be used for any kind of omics techniques when the goal is to identify differences between groups even in cases where sample sizes are limited, or phenotype differences are small. Our results are also pointing out both the possibility and the need of development of novel tools considering gene interaction relationships to achieve a more reliable and more meaningful analysis of large-scale omics readouts. Furthermore, the relevance and usability of heterochronic parabiosis models in studying human brain aging was confirmed.

7. SUMMARY

In this thesis, we will explore the synergism of machine learning and graph-based techniques in analyzing biological data of a wide variety of sources. These tools, separately, have been proven to be useful to find patterns of change in large-scale molecular datasets, but their combination is more widespread in electrophysiology studies.

From electrocorticography data recorded in epileptic patients, Support Vector Machine models were able to reconstruct expert localization of pathogenic zones. The models relied on non-trivial features combining information from multiple frequency bands in a non-linear manner and could successfully operate on seizure-free deep sleep recordings devoid of obvious epileptiform activity. Inspired by this, feasibility of models combining multiple sample features were investigated in transcriptomic and proteomic data from aging phenotypes.

When interpreting transcriptomic data, genes of interest are usually selected based on their altered expression and functional gene product relationships are considered only port-hoc on the pathway level. Here, we analyzed RNA sequencing readouts corresponding to interacting protein pairs using Support Vector Machines to build “Predictomes”, a graph with links weighted according to the performance of the models to predict phenotype. Central elements of this network showed increased replicability across redundant laboratory measurements and, better interpretability when compared to results delivered without integrating protein interaction information. Finally, blood plasma proteomic changes of provoked aging and rejuvenation in parabiosis were analyzed by the validated model and compared with human brain aging.

Our results underlying both the possibility and the need of development of novel tools considering gene interaction relationships to achieve a more reliable and more meaningful analysis of large-scale omics readouts. Furthermore, the relevance of parabiosis on the study of human brain aging was confirmed, suggesting the prominent role of evolutionally conserved, circulating factors in the blood.

8. ÖSSZEFOGLALÁS

A dolgozat témája gépi tanulás és gráfelméleti eszközök integratív alkalmazásának bemutatása különböző forrásokból származó biológiai adatokon. Bár mindkét megközelítést alkalmazzák a molekuláris biológia egyes területein, előnyös tulajdonságaik kombinálása mégis inkább az idegtudományokban terjedt el.

Epilepsziás betegekből származó elektrokortikográfiás adatokból kiindulva gépi tanulós eljárásokkal (Support Vector Machine) sikeresen rekonstruáltuk a patogén zónák szakértői lokalizációját. A modellek nem-triviális jellemzőkre, több frekvenciasávból származó információk nem-lineáris kombinációira támaszkodtak, és sikeresen alkalmazhatóak voltak görcsmentes, mélyalvásban rögzített felvételeken is, jelezve az epileptogén agyterületek nyugalmi állapotban is tetten érhető, megváltozott aktivitási mintázatát. Célunk egy hasonló elven működő, multimodális szinergizmusra támaszkodó módszer kifejlesztése volt genomi léptékű, molekuláris biológiai adatok integratív elemzéséhez. Elektrofiziológiai méréseknél a kinyert frekvencia-specifikus adatok triviálisan elektródákhoz, ezek pedig agyterületekhez rendelhetők. A molekuláris adatok esetében az együtt vizsgálandó változók csoportosítását irodalmi adatok alapján, az egymással funkcionálisan kölcsönható elemek mentén végeztük el. Az egyes fehérjepárokra megfelelő transzkripciós adatokra illesztett gépi tanulós modellek teljesítményét vizsgálva így végül egy súlyozott gráfhoz, a „Prediktómhoz” jutunk, melynek centrális elemei a vizsgált fenotípusra jellemzők.

Eljárásunkat az emberi agy öregedését leíró, redundáns transzkriptomikai méréseken validáltuk. Az integratív megközelítés reprodukálhatóság és interpretálhatóság tekintetében is felülmúlta a funkcionális géntermék-interakciókat nem használó kontroll-modellt. Végül, a vérplazma-proteóm parabiótikus változásait hasonlítottuk össze az emberi agyi öregedéssel. Eredményeink egyszerre jelzik a funkcionális géntermék-kölcsönhatások figyelembevételének lehetőségét és szükségességét genomi léptékű adatok integratív elemzésekor. Továbbá, igazolják a parabiózis, mint állatkísérletes modell relevanciáját az emberi agy öregedésének tanulmányozásában.

9. BIBLIOGRAPHY

- Abbasi, B., & Goldenholz, D. M. (2019). Machine learning applications in epilepsy. *Epilepsia*, *60*(10), 2037–2047. <https://doi.org/10.1111/epi.16333>
- Al-Aamri, A., Taha, K., Al-Hammadi, Y., Maalouf, M., & Homouz, Di. (2019). Analyzing a co-occurrence gene-interaction network to identify disease-gene association. *BMC Bioinformatics*, *20*(1), 1–15. <https://doi.org/10.1186/s12859-019-2634-7>
- Anderson, W. S., Kudela, P., Cho, J., Bergey, G. K., & Franaszczuk, P. J. (2007). Studies of stimulus parameters for seizure disruption using neural network simulations. *Biological Cybernetics*, *97*(2), 173–194. <https://doi.org/10.1007/s00422-007-0166-0>
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, *25*, 25. <https://doi.org/10.1038/75556>
- Ayatollahi, H., Gholamhosseini, L., & Salehi, M. (2019). Predicting coronary artery disease: A comparison between two data mining algorithms. *BMC Public Health*, *19*(1), 1–9. <https://doi.org/10.1186/s12889-019-6721-5>
- Azuaje, F., Devaux, Y., & Wagner, D. R. (2010). Coordinated modular functionality and prognostic potential of a heart failure biomarker-driven interaction network. *BMC Systems Biology*, *4*. <https://doi.org/10.1186/1752-0509-4-60>
- Baht, G. S., Silkstone, D., Vi, L., Nadesan, P., Amani, Y., Whetstone, H., Wei, Q., & Alman, B. A. (2015). Exposure to a youthful circulation rejuvenates bone repair through modulation of β -catenin. *Nature Communications*, *6*(May), 1–9. <https://doi.org/10.1038/ncomms8131>

- Banlaki, Z., Elek, Z., Nanasi, T., Szekely, A., Nemoda, Z., Sasvari-Szekely, M., & Ronai, Z. (2015). Polymorphism in the serotonin receptor 2a (HTR2A) gene as possible predisposal factor for aggressive traits. *PloS One*, *10*(2), e0117792. <https://doi.org/10.1371/journal.pone.0117792>
- Barabási, A. L., & Oltvai, Z. N. (2004). Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics*, *5*(2), 101–113. <https://doi.org/10.1038/nrg1272>
- Barardo, D. G., Newby, D., Thornton, D., Ghafourian, T., de Magalhães, J. P., & Freitas, A. A. (2017). Machine learning for predicting lifespan-extending chemical compounds. *Aging*, *9*(7), 1721–1737. <https://doi.org/10.18632/aging.101264>
- Bell, R., Hubbard, A., Chettier, R., Chen, D., & Miller, J. P. (2009). A Human Protein Interaction Network Shows Conservation of Aging Processes between Human and Invertebrate Species. *PLoS Genet*, *5*(3), 1000414. <https://doi.org/10.1371/journal.pgen.1000414>
- Bellman, R. E. (1957). *Dynamic programming*. Princeton University Press.
- Bellman, R. E. (1961). *Adaptive control processes: a guided tour*. Princeton University Press.
- Ben-hur, A., Horn, D., & Vapnik, V. (2001). Support Vector Clustering. *Journal of Machine Learning Research*, *2*, 125–137. <https://doi.org/10.4249/scholarpedia.5187>
- Ben-Hur, A., Soon Ong, C., ren Sonnenburg, S., Schö lkopf, B., & Rä tsch, G. (2008). Support Vector Machines and Kernels for Computational Biology. *PLoS Computational Biology*, *4*(10), e1000173. <https://doi.org/10.1371/journal.pcbi.1000173>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society*, *57*, 289–300. <https://doi.org/10.2307/2346101>
- Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10). <https://doi.org/10.1088/1742-5468/2008/10/P10008>

- Budovsky, A., Abramovich, A., Cohen, R., Chalifa-Caspi, V., & Fraifeld, V. (2007). Longevity network: Construction and implications. *Mechanisms of Ageing and Development*, *128*(1), 117–124. <https://doi.org/10.1016/j.mad.2006.11.018>
- Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Andreas, C. M., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., & Vanderplas, J. (2011). *Experiences From the Scikit-Learn Project*. 1–15.
- Busiello, D. M., Suweis, S., Hidalgo, J., & Maritan, A. (2017). Explorability and the origin of network sparsity in living systems. *Scientific Reports*, *7*(1), 1–8. <https://doi.org/10.1038/s41598-017-12521-1>
- Canolty, R. T., & Knight, R. T. (2010). The functional role of cross-frequency coupling. *Trends in Cognitive Sciences*, *14*(11), 506–515. <https://doi.org/10.1016/j.tics.2010.09.001>
- Castellano, J. M., Mosher, K. I., Abbey, R. J., McBride, A. A., James, M. L., Berdnik, D., Shen, J. C., Zou, B., Xie, X. S., Tingle, M., Hinkson, I. V., Angst, M. S., & Wyss-Coray, T. (2017). Human umbilical cord plasma proteins revitalize hippocampal function in aged mice. *Nature*, *544*(7651), 488–492. <https://doi.org/10.1038/nature22067>
- Cendes, F., Diaz-arrastia, R., Förstl, H., & Fenton, A. A. (2019). *What Are the Implications for Translational Research?* 302–312. <https://doi.org/10.1016/j.yebeh.2017.09.016>.Epilepsy
- Ceol, A., Chatr Aryamontri, A., Licata, L., Peluso, D., Briganti, L., Perfetto, L., Castagnoli, L., & Cesareni, G. (2009). MINT, the molecular interaction database: 2009 update. *Nucleic Acids Research*, *38*(SUPPL.1), 532–539. <https://doi.org/10.1093/nar/gkp983>
- Chen, L., Cai, C., Chen, V., & Lu, X. (2016). Learning a hierarchical representation of the yeast transcriptomic machinery using an autoencoder model. *BMC Bioinformatics*, *17*(1). <https://doi.org/10.1186/s12859-015-0852-1>

- Chorowski, J., Wang, J., & Zurada, J. M. (2014). *Review and performance comparison of SVM-and ELM-based classifiers*. <https://doi.org/10.1016/j.neucom.2013.08.009>
- Cohen, M. X., Elger, C. E., & Fell, J. (2009). Oscillatory activity and phase-amplitude coupling in the human medial frontal cortex during decision making. *Journal of Cognitive Neuroscience*, *21*(2), 390–402. <https://doi.org/10.1162/jocn.2008.21020>
- Conboy, I. M., Conboy, M. J., Wagers, A. J., Girma, E. R., Weissman, I. L., & Rando, T. A. (2005). Conboy, 2005, Nature, Rejuvenecimento celular e nicho.pdf. *Nature*, *433*(7027), 760–764. <http://www.nature.com/nature/journal/v433/n7027/abs/nature03260.html%0Afile:///Users/andrewmckay/Dropbox/Papers/Library/Library.papers3/Articles/2005/Conboy/NatureConboy.pdf%0Apapers3://publication/uuid/FCAE9822-AEEE-426B-83D2-FB86BFB6327D>
- Costa, C., Parnetti, L., D'Amelio, M., Tozzi, A., Tantucci, M., Romigi, A., Siliquini, S., Cavallucci, V., Di Filippo, M., Mazzocchetti, P., Liguori, C., Nobili, A., Eusebi, P., Mercuri, N. B., & Calabresi, P. (2016). Epilepsy, amyloid- β , and D1 dopamine receptors: a possible pathogenetic link? *Neurobiology of Aging*, *48*, 161–171. <https://doi.org/10.1016/j.neurobiolaging.2016.08.025>
- de Magalhães, J. P., Curado, J., & Church, G. M. (2009). Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics*, *25*(7), 875–881. <https://doi.org/10.1093/bioinformatics/btp073>
- De Magalhães, J. P., & Tacutu, R. (2015). Integrative Genomics of Aging. *Handbook of the Biology of Aging: Eighth Edition*, 263–285. <https://doi.org/10.1016/B978-0-12-411596-5.00009-5>
- De Magalhães, J. P., & Toussaint, O. (2004). GenAge: A genomic and proteomic network map of human ageing. *FEBS Letters*, *571*(1–3), 243–247. <https://doi.org/10.1016/j.febslet.2004.07.006>

- Deluca, D. S., Levin, J. Z., Sivachenko, A., Fennell, T., Nazaire, M. D., Williams, C., Reich, M., Winckler, W., & Getz, G. (2012). RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics*, 28(11), 1530–1532. <https://doi.org/10.1093/bioinformatics/bts196>
- Dopico, X. C., Evangelou, M., Ferreira, R. C., Guo, H., Pekalski, M. L., Smyth, D. J., Cooper, N., Burren, O. S., Fulford, A. J., Hennig, B. J., Prentice, A. M., Ziegler, A. G., Bonifacio, E., Wallace, C., & Todd, J. A. (2015). Widespread seasonal gene expression reveals annual differences in human immunity and physiology. *Nature Communications*, 6(May), 1–13. <https://doi.org/10.1038/ncomms8000>
- Dukart, J., Schroeter, M. L., & Mueller, K. (2011). The Alzheimer’s Disease Neuroimaging Initiative (2011) Age Correction in Dementia-Matching to a Healthy Brain. *PLoS ONE*, 6(7), 22193. <https://doi.org/10.1371/journal.pone.0022193>
- Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., Milacic, M., Roca, C. D., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Viteri, G., Weiser, J., ... D’Eustachio, P. (2018). The Reactome Pathway Knowledgebase. *Nucleic Acids Research*, 46(D1), D649–D655. <https://doi.org/10.1093/nar/gkx1132>
- Fabris, F., Magalhães, J. P. de, & Freitas, A. A. (2017). A review of supervised machine learning applied to ageing research. *Biogerontology*, 18(2), 171–188. <https://doi.org/10.1007/s10522-017-9683-y>
- Fergus, P., Hussain, A., Hignett, D., Al-Jumeily, D., Abdel-Aziz, K., & Hamdan, H. (2016). A machine learning system for automated whole-brain seizure detection. *Applied Computing and Informatics*, 12(1), 70–89. <https://doi.org/10.1016/j.aci.2015.01.001>
- Fernandes, M., Wan, C., Tacutu, R., Barardo, D., Rajput, A., Wang, J., Thoppil, H., Thornton, D., Yang, C., Freitas, A., & de Magalhães, J. P. (2016). Systematic analysis of the gerontome reveals links between aging and age-related diseases. *Human Molecular Genetics*, 25(21), 4804–4818. <https://doi.org/10.1093/hmg/ddw307>

- Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D., & Fernández-Delgado, A. (2014). Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? In *Journal of Machine Learning Research* (Vol. 15).
- Ferrarini, L., Bertelli, L., Feala, J., McCulloch, A. D., & Paternostro, G. (2005). A more efficient search strategy for aging genes based on connectivity. *Bioinformatics*, *21*(3), 338–348. <https://doi.org/10.1093/bioinformatics/bti004>
- File, B., Nánási, T., Tóth, E., Bokodi, V., Tóth, B., Hajnal, B., Kardos, Z., Entz, L., Eross, L., Ulbert, I., & Fabó, D. (2020). Reorganization of Large-Scale Functional Networks during Low-Frequency Electrical Stimulation of the Cortical Surface. *International Journal of Neural Systems*, *30*(3), 1–15. <https://doi.org/10.1142/S0129065719500229>
- Fisher, R. A. (1922). On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P. *Journal of the Royal Statistical Society*, *85*(1), 87. <https://doi.org/10.2307/2340521>
- Fisher, R. A. (1936). The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, *1*(1), 1–8.
- Fraser, H. B., Khaitovich, P., Plotkin, J. B., Pääbo, S., & Eisen, M. B. (2005). Aging and gene expression in the primate brain. *PLoS Biology*, *3*(9), 1653–1661. <https://doi.org/10.1371/journal.pbio.0030274>
- Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph Drawing by Force-directed Placement. *Software - Practice and Experience*, *21*(NOVEMBER), 1129–1164.
- Fu, A. K. Y., Hung, K. W., Yuen, M. Y. F., Zhou, X., Mak, D. S. Y., Chan, I. C. W., Cheung, T. H., Zhang, B., Fu, W. Y., Liew, F. Y., & Ip, N. Y. (2016). IL-33 ameliorates Alzheimer's disease-like pathology and cognitive decline. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(19), E2705–E2713.

<https://doi.org/10.1073/pnas.1604032113>

Gao, J., Leung, H. K., Wu, B. W. Y., Skouras, S., & Sik, H. H. (2019). The neurophysiological correlates of religious chanting. *Scientific Reports*, 9(1), 1–9. <https://doi.org/10.1038/s41598-019-40200-w>

García-Campos, M. A., Espinal-Enríquez, J., & Hernández-Lemus, E. (2015). Pathway analysis: State of the art. *Frontiers in Physiology*, 6(DEC), 1–16. <https://doi.org/10.3389/fphys.2015.00383>

Garcia-Ramos, C., Bobholz, S., Dabbs, K., Hermann, B., Joutsa, J., Rinne, J. O., Karrasch, M., Prabhakaran, V., Shinnar, S., & Sillanpää, M. (2017). Brain structure and organization five decades after childhood onset epilepsy. *Human Brain Mapping*, 38(6), 3289–3299. <https://doi.org/10.1002/hbm.23593>

Gibbs, D. L., Baratt, A., Baric, R. S., Kawaoka, Y., Smith, R. D., Orwoll, E. S., Katze, M. G., & Mcweeney, S. K. (2013). Protein co-expression network analysis (ProCoNA). *Journal of Clinical Bioinformatics*, 1–10.

Gibbs, D. L., Gralinski, L., Baric, R. S., & McWeeney, S. K. (2013). Multi-omic network signatures of disease. *Frontiers in Genetics*, 4(JAN), 1–11. <https://doi.org/10.3389/fgene.2013.00309>

Gioutlakis, A., Klapa, M. I., & Moschonas, N. K. (2017). PICKLE 2.0: A human protein-protein interaction meta-database employing data integration via genetic information ontology. *PLoS ONE*, 12(10), 1–17. <https://doi.org/10.1371/journal.pone.0186039>

Gold, L., Ayers, D., Bertino, J., Bock, C., Bock, A., Brody, E. N., Carter, J., Dalby, A. B., Eaton, B. E., Fitzwater, T., Flather, D., Forbes, A., Foreman, T., Fowler, C., Gawande, B., Goss, M., Gunn, M., Gupta, S., Halladay, D., ... Zichi, D. (2010). Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS ONE*, 5(12).

<https://doi.org/10.1371/journal.pone.0015004>

Guebel, D. V., Ertaylan, G., Vinod, P. K., Lanke, V., Moolamalla, S. T. R., & Roy, D. (2018). *Integrative Analysis of Hippocampus Gene Expression Profiles Identifies Network Alterations in Aging and Alzheimer's Disease*. <https://doi.org/10.3389/fnagi.2018.00153>

Guo, C., Liu, X. J., Cheng, Z. X., Liu, Y. J., Li, H., & Peng, X. (2014). Characterization of protein species and weighted protein co-expression network regulation of *Escherichia coli* in response to serum killing using a 2-DE based proteomics approach. *Molecular BioSystems*, *10*(3), 475–484. <https://doi.org/10.1039/c3mb70404a>

Guo, Y., & Xing, Y. (2016). Weighted gene co-expression network analysis of pneumocytes under exposure to a carcinogenic dose of chloroprene. *Life Sciences*, *151*, 339–347. <https://doi.org/10.1016/j.lfs.2016.02.074>

Hermann, B. P., Jones, J. E., Sheth, R., Koehn, M., Becker, T., Fine, J., Allen, C. A., & Seidenberg, M. (2008). Growing up with epilepsy: A two-year investigation of cognitive development in children with new onset epilepsy. *Epilepsia*, *49*(11), 1847–1858. <https://doi.org/10.1111/j.1528-1167.2008.01735.x>

Hoare, P. (1984). the Development of Psychiatric Disorder Among Schoolchildren With Epilepsy. *Developmental Medicine & Child Neurology*, *26*(1), 3–13. <https://doi.org/10.1111/j.1469-8749.1984.tb04399.x>

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology*, *14*(10), R115. <https://doi.org/10.1186/gb-2013-14-10-r115>

Horvath, S., Mah, V., Lu, A. T., Woo, J. S., Choi, O.-W., Jasinska, A. J., Riancho, J. A., Tung, S., Coles, N. S., Braun, J., Vinters, H. V., & Coles, L. S. (2015a). The cerebellum ages slowly according to the epigenetic clock. *Aging*, *7*(5), 294–306.

<http://www.ncbi.nlm.nih.gov/pubmed/26000617><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4468311>

Horvath, S., Mah, V., Lu, A. T., Woo, J. S., Choi, O.-W., Jasinska, A. J., Riancho, J. A., Tung, S., Coles, N. S., Braun, J., Vinters, H. V., & Coles, L. S. (2015b). The cerebellum ages slowly according to the epigenetic clock. *Aging*, 7(5), 294–306.

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1), 44–57. <https://doi.org/10.1038/nprot.2008.211>

Huang, Q., Ning, Y., Liu, D., Zhang, Y., Li, D., Zhang, Y., Yin, Z., Fu, B., Cai, G., Sun, X., & Chen, X. (2018). A Young Blood Environment Decreases Aging of Senile Mice Kidneys. *J Gerontol A Biol Sci Med Sci.*, 73(4), 421-428. <https://doi.org/doi:10.1093/gerona/glx183>

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-barclay, Y. D., Antonellis, K. J., & Speed, T. P. (2003). *Exploration , Normalization , and Summaries of High Density Oligonucleotide Array Probe Level Data.* June, 249–264.

Jaccard, P. (1912). the Distribution of the Flora in the Alpine Zone. *New Phytologist*, 11(2), 37–50. <https://doi.org/10.1111/j.1469-8137.1912.tb05611.x>

Jobson, R. W., Nabholz, B., & Galtier, N. (2010). An evolutionary genome scan for longevity-related natural selection in mammals. *Molecular Biology and Evolution*, 27(4), 840–847. <https://doi.org/10.1093/molbev/msp293>

Jung, Y.-J., Kang, H.-C., Choi, K.-O., Lee, J. S., Kim, D.-S., Cho, J.-H., Kim, S.-H., Im, C.-H., & Kim, H. D. (2011). Localization of ictal onset zones in Lennox-Gastaut syndrome using directional connectivity analysis of intracranial electroencephalography. *Seizure*,

20(6), 449–457. <https://doi.org/10.1016/j.seizure.2011.02.004>

Kanehisa, M., & Goto, S. (2000). *KEGG; Kyoto Encyclopedia of Genes and Genomes .pdf*. 28(1), 27–30. <https://doi.org/10.1093/nar/27.1.29>

Kanonidis, E. I., Roy, M. M., Deighton, R. F., & Le Bihan, T. (2016). Protein co-expression analysis as a strategy to complement a standard quantitative proteomics approach: Case of a glioblastoma multiforme study. *PLoS ONE*, 11(8), 1–22. <https://doi.org/10.1371/journal.pone.0161828>

Karrasch, M., Tiitta, P., Hermann, B., Joutsa, J., Shinnar, S., Rinne, J., Anttinen, A., & Sillanpää, M. (2017). Cognitive Outcome in Childhood-Onset Epilepsy: A Five-Decade Prospective Cohort Study. *Journal of the International Neuropsychological Society*, 23(4), 332–340. <https://doi.org/10.1017/S1355617716001077>

Katsimpardi, L., Litterman, N. K., Schein, P. A., Miller, C. M., Loffredo, F. S., Wojtkiewicz, G. R., Chen, J. W., Lee, R. T., Wagers, A. J., & Rubin, L. L. (2014). Vascular and neurogenic rejuvenation of the aging mouse brain by young systemic factors. *Science*, 344(6184), 630–634. <https://doi.org/10.1126/science.1251141>

Kerepesi, C., Daróczy, B., Sturm, Á., Vellai, T., & Benczúr, A. (2018). Prediction and characterization of human ageing-related proteins by using machine learning. *Scientific Reports*, 8(1), 1–13. <https://doi.org/10.1038/s41598-018-22240-w>

Kerrien, S., Alam-Faruque, Y., Aranda, B., Bancarz, I., Bridge, A., Derow, C., Dimmer, E., Feuermann, M., Friedrichsen, A., Huntley, R., Kohler, C., Khadake, J., Leroy, C., Liban, A., Lieftink, C., Montecchi-Palazzi, L., Orchard, S., Risse, J., Robbe, K., ... Hermjakob, H. (2007). IntAct - Open source resource for molecular interaction data. *Nucleic Acids Research*, 35(SUPPL. 1), 561–565. <https://doi.org/10.1093/nar/gkl958>

Keshava Prasad, T. S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., Balakrishnan, L., Marimuthu, A., Banerjee, S., Somanathan, D. S., Sebastian, A., Rani, S., Ray, S., Harrys Kishore, C. J., Kanth, S., ... Pandey, A. (2009). Human Protein Reference Database -

- 2009 update. *Nucleic Acids Research*, 37(SUPPL. 1), 767–772. <https://doi.org/10.1093/nar/gkn892>
- Kim, J.-Y., Kang, H.-C., Kim, K., Kim, H. D., & Im, C.-H. (2015). Localization of epileptogenic zones in Lennox-Gastaut syndrome (LGS) using graph theoretical analysis of ictal intracranial EEG: a preliminary investigation. *Brain & Development*, 37(1), 29–36. <https://doi.org/10.1016/j.braindev.2014.02.006>
- Kirkwood, T. B. L., & Kowald, A. (1997). Network theory of aging. *Experimental Gerontology*, 32(4–5), 395–399. [https://doi.org/10.1016/S0531-5565\(96\)00171-4](https://doi.org/10.1016/S0531-5565(96)00171-4)
- Kiss, H. J. M., Mihalik, Á., Nánási, T., Ory, B., Spiró, Z., Soti, C., & Csermely, P. (2009). Ageing as a price of cooperation and complexity: Self-organization of complex systems causes the gradual deterioration of constituent networks. *BioEssays*, 31(6), 651–664. <https://doi.org/10.1002/bies.200800224>
- Klapa, M. I., Tsafou, K., Theodoridis, E., Tsakalidis, A., & Moschonas, N. K. (2013). Reconstruction of the experimentally supported human protein interactome: What can we learn? *BMC Systems Biology*, 7(October). <https://doi.org/10.1186/1752-0509-7-96>
- Kotloski, R. J., Dowding, J., Hermann, B. P., & Sutula, T. P. (2019). Epilepsy and aging. In *Handbook of Clinical Neurology* (1st ed., Vol. 167). Elsevier B.V. <https://doi.org/10.1016/B978-0-12-804766-8.00025-X>
- Kovacs-Nagy, R., Elek, Z., Szekely, A., Nanasi, T., Sasvari-Szekely, M., & Ronai, Z. (2013). Association of aggression with a novel microRNA binding site polymorphism in the wolframin gene. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, 162(4), 404–412. <https://doi.org/10.1002/ajmg.b.32157>
- Lee, M. H., Kwon, O. Y., Kim, Y. J., Kim, H. K., Lee, Y. E., Williamson, J., Fazli, S., & Lee, S. W. (2019). EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy. *GigaScience*, 8(5), 1–16. <https://doi.org/10.1093/gigascience/giz002>

- Lehallier, B., Gate, D., Schaum, N., Nanasi, T., Eun Lee, S., Yousef, H., Moran Losada, P., Berdnik, D., Keller, A., Verghese, J., Sathyan, S., Franceschi, C., Milman, S., Barzilai, N., & Wyss-Coray, T. (2019). Undulating changes in human plasma proteome profiles across the lifespan. *Nature Medicine*. <https://doi.org/10.1038/s41591-019-0673-2>
- Lehallier, B., Gate, D., Schaum, N., Nanasi, T., Lee, S. E., Yousef, H., Losada, P. M., Berdnik, D., Keller, A., Verghese, J., Sathyan, S., Franceschi, C., Milman, S., Barzilai, N., & Wyss-Coray, T. (2019). Undulating changes in human plasma proteome across lifespan are linked to disease. *BioRxiv*. <https://doi.org/10.1101/751115>
- Li, Y.-H., Zhang, G.-G., & Guo, Z. (2010). Computational Prediction of Aging Genes in Human. *International Conference on Biomedical Engineering and Computer Science (ICBECS), 2010, 1–4 (IEEE 2010)*.
- Liu, J., Jing, L., & Tu, X. (2016). Weighted gene co-expression network analysis identifies specific modules and hub genes related to coronary artery disease. *BMC Cardiovascular Disorders, 16*(1), 1–8. <https://doi.org/10.1186/s12872-016-0217-3>
- Loffredo, F. S., Steinhauser, M. L., Jay, S. M., Gannon, J., Pancoast, J. R., Yalamanchi, P., Sinha, M., Dall’Osso, C., Khong, D., Shadrach, J. L., Miller, C. M., Singer, B. S., Stewart, A., Psychogios, N., Gerszten, R. E., Hartigan, A. J., Kim, M. J., Serwold, T., Wagers, A. J., & Lee, R. T. (2013). Growth differentiation factor 11 is a circulating factor that reverses age-related cardiac hypertrophy. *Cell, 153*(4), 828–839. <https://doi.org/10.1016/j.cell.2013.04.015>
- López-Azcárate, J., Nicolás, M. J., Cordon, I., Alegre, M., Valencia, M., & Artieda, J. (2013). Delta-mediated cross-frequency coupling organizes oscillatory activity across the rat cortico-basal ganglia network. *Frontiers in Neural Circuits, 7*, 155. <https://doi.org/10.3389/fncir.2013.00155>
- López-Azcárate, J., Tainta, M., Rodríguez-Oroz, M. C., Valencia, M., González, R., Guridi, J., Iriarte, J., Obeso, J. A., Artieda, J., & Alegre, M. (2010). Coupling between beta and high-frequency activity in the human subthalamic nucleus may be a pathophysiological mechanism in Parkinson’s disease. *The Journal of Neuroscience: The Official Journal*

of the Society for Neuroscience, 30(19), 6667–6677.
<https://doi.org/10.1523/JNEUROSCI.5459-09.2010>

López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2013). The hallmarks of aging. *Cell*, 153(6). <https://doi.org/10.1016/j.cell.2013.05.039>

Lu, T., Pan, Y., Kao, S., Li, C., Kohane, I., Chan, J., & Yankner, B. A. (2004). Gene regulation and DNA damage in the ageing human brain. *Nature*, 429(6994), 883–891.
<https://doi.org/10.1038/nature02618.1>.

MacDonald, M. L., Ding, Y., Newman, J., Hemby, S., Penzes, P., Lewis, D. A., Yates, N. A., & Sweet, R. A. (2015). Altered glutamate protein co-expression network topology linked to spine loss in the auditory cortex of schizophrenia. *Biological Psychiatry*, 77(11), 959–968. <https://doi.org/10.1016/j.biopsych.2014.09.006>

Marchi, N., & Lerner-Natoli, M. (2013). Cerebrovascular remodeling and epilepsy. *Neuroscientist*, 19(3), 304–312. <https://doi.org/10.1177/1073858412462747>

Maris, E., van Vugt, M., & Kahana, M. (2011). Spatially distributed patterns of oscillatory coupling between high-frequency amplitudes and low-frequency phases in human iEEG. *NeuroImage*, 54(2), 836–850. <https://doi.org/10.1016/j.neuroimage.2010.09.029>

Matthews, B. W. (1975). *T4 phage lysozyme*. 405, 442–451. [https://doi.org/10.1016/0005-2795\(75\)90109-9](https://doi.org/10.1016/0005-2795(75)90109-9)

McGinn, R. J., & Valiante, T. A. (2014). Phase-amplitude coupling and interlaminar synchrony are correlated in human neocortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 34(48), 15923–15930.
<https://doi.org/10.1523/JNEUROSCI.2771-14.2014>

Mihalik, Á., Kaposi, A. S., Kovács, I. A., Nánási, T., Palotai, R., Rák, Á., Szalay-Beko, M. S., & Csermely, P. (2012). How creative elements help the recovery of networks after crisis: Lessons from biology. *Networks in Social Policy Problems*, 179–188.
<https://doi.org/10.1017/CBO9780511842481.010>

- Mishra, B., Kumar, N., & Mukhtar, M. S. (2018). Systems Biology and Machine Learning in Plant–Pathogen Interactions. *Molecular Plant-Microbe Interactions*, 32(1), 45–55. <https://doi.org/10.1094/mpmi-08-18-0221-fi>
- Mizoguchi, H., Yamada, K., & Nabeshima, T. (2011). Matrix metalloproteinases contribute to neuronal dysfunction in animal models of drug dependence, Alzheimer’s disease, and epilepsy. *Biochemistry Research International*, 2011. <https://doi.org/10.1155/2011/681385>
- Módos, D., Bulusu, K. C., Fazekas, D., Kubisch, J., Brooks, J., Marczell, I., Szabó, P. M., Vellai, T., Csermely, P., Lenti, K., Bender, A., & Korcsmáros, T. (2017). Neighbours of cancer-related proteins have key influence on pathogenesis and could increase the drug target space for anticancer therapies. *Npj Systems Biology and Applications*, 3(1). <https://doi.org/10.1038/s41540-017-0003-6>
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, 5(7), 621–628. <https://doi.org/10.1038/nmeth.1226>
- Nair, J., Ghatge, M., Kakkar, V. V., & Shanker, J. (2014). Network analysis of inflammatory genes and their transcriptional regulators in coronary artery disease. *PLoS ONE*, 9(4). <https://doi.org/10.1371/journal.pone.0094328>
- Nánási, T., File, B., Tóth, E., Entz, L., Ulbert, I., Fabó, D., & ErÅ’ss, L. (2016). Synergism of spectral and coupling modalities in epileptic focus localization from iEEG recordings. *PRNI 2016 - 6th International Workshop on Pattern Recognition in Neuroimaging*, 2–5. <https://doi.org/10.1109/PRNI.2016.7552354>
- Newman, M. E. J. (2004). Analysis of weighted networks. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 70(5), 9. <https://doi.org/10.1103/PhysRevE.70.056131>
- Nguyen, P. T., Dorman, L. C., Pan, S., Molofsky, A. B., Kheirbek, M. A., Molofsky, A. V.,

- Nguyen, P. T., Dorman, L. C., Pan, S., Vainchtein, I. D., Han, R. T., & Nakao-inoue, H. (2020). Microglial Remodeling of the Extracellular Matrix Promotes Synapse Plasticity. Article Microglial Remodeling of the Extracellular Matrix Promotes Synapse Plasticity. *Cell*, 1–16. <https://doi.org/10.1016/j.cell.2020.05.050>
- Nibbe, R. K., Koyutü, M., & Chance, M. R. (2010). An integrative -omics approach to identify functional sub-networks in human colorectal cancer. *PLoS Computational Biology*, 6(1). <https://doi.org/10.1371/journal.pcbi.1000639>
- Ortega, G. J., Sola, R. G., & Pastor, J. (2008). Complex network analysis of human ECoG data. *Neuroscience Letters*, 447(2–3), 129–133. <https://doi.org/10.1016/j.neulet.2008.09.080>
- Pancaldi, V., Saraç, Ö. S., Rallis, C., Mclean, J. R., Převorovský, M., Gould, K., Beyer, A., & Bähler, J. (2012). *Predicting the Fission Yeast Protein Interaction Network*. 2, 453. <https://doi.org/10.1534/g3.111.001560>
- Partridge, L., & Gems, D. (2002). Mechanisms of ageing: Public or private? *Nature Reviews Genetics*, 3(3), 165–175. <https://doi.org/10.1038/nrg753>
- Paulson J, Chen C, Lopes-Ramos C, Kuijjer M, Platig J, Sonawane A, Fagny M, Glass K, Q. J. (2019). *YARN: Robust Multi-Condition RNA-Seq Preprocessing and Normalization*. <http://bioconductor.org/packages/release/bioc/html/yarn.html>
- Pluvinage, J. V., & Wyss-Coray, T. (2020). Systemic factors as mediators of brain homeostasis, ageing and neurodegeneration. *Nature Reviews Neuroscience*, 21(2), 93–102. <https://doi.org/10.1038/s41583-019-0255-9>
- Pótári, A., Ujma, P. P., Konrad, B. N., Genzel, L., Simor, P., Körmendi, J., Gombos, F., Steiger, A., Dresler, M., & Bódizs, R. (2017). Age-related changes in sleep EEG are attenuated in highly intelligent individuals. *NeuroImage*, 146, 554–560. <https://doi.org/10.1016/j.neuroimage.2016.09.039>

- Powers, D. M. W. (2007). Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation. *Technical Report SIE-07-001, School of Informatics and Engineering Flinders University, Adelaide, Australia, December.*
- Procaccini, C., Carbone, F., Di Silvestre, D., Brambilla, F., De Rosa, V., Galgani, M., Faicchia, D., Marone, G., Tramontano, D., Corona, M., Alviggi, C., Porcellini, A., La Cava, A., Mauri, P., & Matarese, G. (2016). The Proteomic Landscape of Human Ex Vivo Regulatory and Conventional T Cells Reveals Specific Metabolic Requirements. *Immunity, 44*(2), 406–421. <https://doi.org/10.1016/j.immuni.2016.01.028>
- Promislow, D. E. L. (2004). Protein networks, pleiotropy and the evolution of senescence. *Proceedings of the Royal Society B: Biological Sciences, 271*(1545), 1225–1234. <https://doi.org/10.1098/rspb.2004.2732>
- Rasheed, K., Qayyum, A., Qadir, J., Sivathamboo, S., Kwan, P., Kuhlmann, L., O'Brien, T., & Razi, A. (2020). *Machine Learning for Predicting Epileptic Seizures Using EEG Signals: A Review.* 1–15. <http://arxiv.org/abs/2002.01925>
- Ravdin, L. D., & Katzen, H. L. (2013). Handbook on the neuropsychology of aging and dementia. *Handbook on the Neuropsychology of Aging and Dementia*, 1–527. <https://doi.org/10.1007/978-1-4614-3106-0>
- Roggenhofer, E., Santarnecchi, E., Muller, S., Kherif, F., Wiest, R., Seeck, M., & Draganski, B. (2019). Trajectories of brain remodeling in temporal lobe epilepsy. *Journal of Neurology, 266*(12), 3150–3159. <https://doi.org/10.1007/s00415-019-09546-z>
- Salpeter, S. J., Khalaileh, A., Weinberg-Corem, N., Ziv, O., Glaser, B., & Dor, Y. (2013). Systemic regulation of the age-Related decline of pancreatic β -Cell replication. *Diabetes, 62*(8), 2843–2848. <https://doi.org/10.2337/db13-0160>
- Scheffzük, C., Kukushka, V. I., Vyssotski, A. L., Draguhn, A., Tort, A. B. L., & Brankačk,

- J. (2011). Selective coupling between theta phase and neocortical fast gamma oscillations during REM-sleep in mice. *PloS One*, 6(12), e28489. <https://doi.org/10.1371/journal.pone.0028489>
- Scudellari, M. (2015). Ageing research: Blood to blood. *Nature*, 517(7535), 426–429. <https://doi.org/10.1038/517426a>
- Semeiks, J., & Grishin, N. V. (2012). A method to find longevity-selected positions in the mammalian proteome. *PLoS ONE*, 7(6). <https://doi.org/10.1371/journal.pone.0038595>
- Sharott, A., Moll, C. K. E., Engler, G., Denker, M., Grün, S., & Engel, A. K. (2009). Different subtypes of striatal neurons are selectively modulated by cortical oscillations. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 29(14), 4571–4585. <https://doi.org/10.1523/JNEUROSCI.5097-08.2009>
- Simkó, G. I., Gyurkó, D., Veres, D. V, Nánási, T., & Csermely, P. (2009). Network strategies to understand the aging process and help age-related drug design. *Review Genome Medicine*, 1. <https://doi.org/doi: 10.1186/gm90>
- Sinha, M., Jang, Y. C., Oh, J., Khong, D., Wu, E. Y., Manohar, R., Miller, C., Regalado, S. G., Loffredo, F. S., Pancoast, J. R., Hirshman, M. F., Lebowitz, J., Shadrach, J. L., Cerletti, M., Kim, M., Serwold, T., Goodyear, L. J., Rosner, B., Lee, R. T., & Wagers, A. J. (2014). Restoring Systemic GDF11 Levels Reverses Age-Related Dysfunction in Mouse Skeletal Muscle. *Science*, May, 1–6.
- Skouras, S., Falcon, C., Tucholka, A., Rami, L., Sanchez-Valle, R., Lladó, A., Gispert, J. D., & Molinuevo, J. L. (2019). Mechanisms of functional compensation, delineated by eigenvector centrality mapping, across the pathophysiological continuum of Alzheimer's disease. *NeuroImage: Clinical*, 22(June 2018), 101777.

<https://doi.org/10.1016/j.nicl.2019.101777>

Smith, E. D., Tsuchiya, M., Fox, L. A., Dang, N., Hu, D., Kerr, E. O., Johnston, E. D., Tchao, B. N., Pak, D. N., Welton, K. L., Promislow, D. E. L., Thomas, J. H., Kaeberlein, M., & Kennedy, B. K. (2008). Quantitative evidence for conserved longevity pathways between divergent eukaryotic species *Erica*. *Genome Research*, *18*, 564–570. <https://doi.org/10.1101/gr.074724.107.1>

Smolander, J., Dehmer, M., & Emmert-Streib, F. (2019). Comparing deep belief networks with support vector machines for classifying gene expression data from complex disorders. *FEBS Open Bio*, *9*, 1232–1248. <https://doi.org/10.1002/2211-5463.12652>

Smolander, J., Dehmer, M., & Emmert-Streib, F. (2019). Comparing deep belief networks with support vector machines for classifying gene expression data from complex disorders. *FEBS Open Bio*. <https://doi.org/10.1002/2211-5463.12652>

Stark, C. (2006). BioGRID: a general repository for interaction datasets. *Nucleic Acids Research*, *34*(90001), D535–D539. <https://doi.org/10.1093/nar/gkj109>

Stegeman, R., & Weake, V. M. (2017). Transcriptional Signatures of Aging. *Journal of Molecular Biology*, *429*(16), 2427–2437. <https://doi.org/10.1016/j.jmb.2017.06.019>

Swindell, W. R., Johnston, A., Sun, L., Xing, X., Fisher, G. J., Bulyk, M. L., Elder, J. T., & Gudjonsson, J. E. (2012). Meta-profiles of gene expression during aging: Limited similarities between mouse and human and an unexpectedly decreased inflammatory signature. *PLoS ONE*, *7*(3). <https://doi.org/10.1371/journal.pone.0033204>

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N. T., Roth, A., Bork, P., Jensen, L. J., & Von Mering, C. (2017). The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Research*, *45*(D1), D362–D368.

<https://doi.org/10.1093/nar/gkw937>

The GTEx Consortium. (2013). The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*, 45(6), 580–585. <https://doi.org/10.1038/ng.2653>

Tort, A. B. L., Komorowski, R. W., Manns, J. R., Kopell, N. J., & Eichenbaum, H. (2009). Theta-gamma coupling increases during the learning of item-context associations. *Proceedings of the National Academy of Sciences of the United States of America*, 106(49), 20942–20947. <https://doi.org/10.1073/pnas.0911331106>

Tort, A. B. L., Kramer, M. A., Thorn, C., Gibson, D. J., Kubota, Y., Graybiel, A. M., & Kopell, N. J. (2008). Dynamic cross-frequency couplings of local field potential oscillations in rat striatum and hippocampus during performance of a T-maze task. *Proceedings of the National Academy of Sciences of the United States of America*, 105(51), 20517–20522. <https://doi.org/10.1073/pnas.0810524105>

Trunk, G. V. (1979). A Problem of Dimensionality: A Simple Example. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(3), 306–307. <https://doi.org/10.1109/TPAMI.1979.4766926>

Tseng, C.-D., Shieh, C.-S., Huang, Y.-J., Chao, P.-J., & Lee, T.-F. (2018). Using LASSO regression based SVM classification to improve the predictive performance of radiation-induced pneumonitis complication in breast cancer. <https://doi.org/10.1080/02533839.2018.1534616>

Türei, D., Korcsmáros, T., & Saez-Rodriguez, J. (2016). OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nature Methods*, 13(12), 966–967. <https://doi.org/10.1038/nmeth.4077>

Ujma, P. P., Baudson, T. G., Bódizs, R., & Dresler, M. (2020). The relationship between chronotype and intelligence: the importance of work timing. *Scientific Reports*, 10(1), 1–6. <https://doi.org/10.1038/s41598-020-62917-9>

- Ujma, P. P., Simor, P., Steiger, A., Dresler, M., & Bódizs, R. (2019). Individual slow-wave morphology is a marker of aging. *Neurobiology of Aging*, *80*(April), 71–82. <https://doi.org/10.1016/j.neurobiolaging.2019.04.002>
- Usman, S. M., Usman, M., & Fong, S. (2017). Epileptic Seizures Prediction Using Machine Learning Methods. *Computational and Mathematical Methods in Medicine*, *2017*. <https://doi.org/10.1155/2017/9074759>
- Vainchtein, I. D., Chin, G., Cho, F. S., Kelley, K. W., Miller, J. G., Chien, E. C., Liddel, S. A., Nguyen, P. T., Nakao-inoue, H., Dorman, L. C., Akil, O., Joshita, S., Barres, B. A., Paz, J. T., Molofsky, A. B., & Molofsky, A. V. (2018). Astrocyte-derived interleukin-33 promotes microglial synapse engulfment and neural circuit development. *Science*, *1273*(March), 1269–1273.
- van Mierlo, P., Carrette, E., Hallez, H., Raedt, R., Meurs, A., Vandenberghe, S., Van Roost, D., Boon, P., Staelens, S., & Vonck, K. (2013). Ictal-onset localization through connectivity analysis of intracranial EEG signals in patients with refractory epilepsy. *Epilepsia*, *54*(8), 1409–1418. <https://doi.org/10.1111/epi.12206>
- Van Ooyen, A., Jun, S. C., Chai, R., Dimitriadis, S. I., & Marimpis, A. D. (2018). Enhancing Performance and Bit Rates in a Brain-Computer Interface System With Phase-to-Amplitude Cross-Frequency Coupling: Evidences From Traditional c-VEP, Fast c-VEP, and SSVEP Designs. *Frontiers in Neuroinformatics | Www.Frontiersin.Org*, *1*, 19. <https://doi.org/10.3389/fninf.2018.00019>
- Varoquaux, G., Buitinck, L., Louppe, G., Grisel, O., Pedregosa, F., & Mueller, A. (2015). Scikit-learn. *GetMobile: Mobile Computing and Communications*, *19*(1), 29–33. <https://doi.org/10.1145/2786984.2786995>
- Vella, D., Zoppis, I., Mauri, G., Mauri, P., & Silvestre, D. Di. (2017). From protein-protein interactions to protein co-expression networks: a new perspective to evaluate large-scale proteomic data. *EURASIP Journal on Bioinformatics and Systems Biology*, *2017*, 6.

<https://doi.org/10.1186/s13637-017-0059-z>

- Veres, D. V., Gyurkó, D. M., Thaler, B., Szalay, K. Z., Fazekas, D., Korcsmáros, T., & Csermely, P. (2015). ComPPI: A cellular compartment-specific database for protein-protein interaction network analysis. *Nucleic Acids Research*, *43*(D1), D485–D493. <https://doi.org/10.1093/nar/gku1007>
- Villeda, S. A., Luo, J., Mosher, K. I., Zou, B., Britschgi, M., Bieri, G., Stan, T. M., Fainberg, N., Ding, Z., Eggel, A., Lucin, K. M., Czirr, E., Park, J. S., Couillard-Després, S., Aigner, L., Li, G., Peskind, E. R., Kaye, J. A., Quinn, J. F., ... Wyss-Coray, T. (2011). The ageing systemic milieu negatively regulates neurogenesis and cognitive function. *Nature*, *477*(7362), 90–96. <https://doi.org/10.1038/nature10357>
- Villeda, S. A., Plambeck, K. E., Middeldorp, J., Castellano, J. M., Mosher, K. I., Luo, J., Smith, L. K., Bieri, G., Lin, K., Berdnik, D., Wabl, R., Udeochu, J., Wheatley, E. G., Zou, B., Simmons, D. A., Xie, X. S., Longo, F. M., & Wyss-coray, T. (2014). Young blood reverses age-related impairments in cognitive function and synaptic plasticity in mice. *Nature Medicine*, *20*(6), 659–663. <https://doi.org/10.1038/nm.3569>
- von Nicolai, C., Engler, G., Sharott, A., Engel, A. K., Moll, C. K., & Siegel, M. (2014). Corticostriatal coordination through coherent phase-amplitude coupling. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(17), 5938–5948. <https://doi.org/10.1523/JNEUROSCI.5007-13.2014>
- Wadi, L., Meyer, M., Weiser, J., Stein, L. D., & Reimand, J. (2016a). Impact of knowledge accumulation on pathway enrichment analysis. *BioRxiv*, *29*(February), 049288. <https://doi.org/10.1101/049288>
- Wadi, L., Meyer, M., Weiser, J., Stein, L. D., & Reimand, J. (2016b). Impact of outdated gene annotations on pathway enrichment analysis. *Nature Methods*, *13*(9), 705–706. <https://doi.org/10.1038/nmeth.3963>
- Wang, Zhong; Gerstein, Mark; Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, *10* (1), 57–63. <https://doi.org/10.1038/nrg2484>

- Wang, J., Duncan, D., Shi, Z., & Zhang, B. (2013). WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Research*, 41(Web Server issue), 77–83. <https://doi.org/10.1093/nar/gkt439>
- Wilke, C., Worrell, G., & He, B. (2011). Graph analysis of epileptogenic networks in human partial epilepsy. *Epilepsia*, 52(1), 84–93. <https://doi.org/10.1111/j.1528-1167.2010.02785.x>
- Wu, D., Liu, X., Liu, C., Liu, Z., Xu, M., Rong, R., Qian, M., Chen, L., & Zhu, T. (2014). Network analysis reveals roles of inflammatory factors in different phenotypes of kidney transplant patients. *Journal of Theoretical Biology*, 362, 62–68. <https://doi.org/10.1016/j.jtbi.2014.03.006>
- Xue, H., Xian, B., Dong, D., Xia, K., Zhu, S., Zhang, Z., Hou, L., Zhang, Q., Zhang, Y., & Han, J. D. J. (2007). A modular network model of aging. *Molecular Systems Biology*, 3(147). <https://doi.org/10.1038/msb4100189>
- Yang, A. C., Stevens, M. Y., Chen, M. B., Lee, D. P., Stähli, D., Gate, D., Contrepois, K., Chen, W., Iram, T., Zhang, L., Vest, R. T., Chaney, A., Lehallier, B., Olsson, N., Bois, H., Hsieh, R., Cropper, H. C., Berdnik, D., Li, L., ... Quake, S. R. (2020). *Physiological blood – brain transport is impaired with age by a shift in transcytosis*. <https://doi.org/10.1038/s41586-020-2453-z>
- Yang, J., Huang, T., Petralia, F., Long, Q., Zhang, B., Argmann, C., Zhao, Y., Mobbs, C. V., Schadt, E. E., Zhu, J., & Tu, Z. (2015). *Synchronized age-related gene expression changes across multiple tissues in human and the link to complex diseases*. <https://doi.org/10.1038/srep15145>
- Yu, X., Feng, L., Liu, D., Zhang, L., Wu, B., Jiang, W., Han, Z., & Cheng, S. (2016). Quantitative proteomics reveals the novel co-expression signatures in early brain development for prognosis of glioblastoma multiforme. *Oncotarget*, 7(12), 14161–

14171. <https://doi.org/10.18632/oncotarget.7416>

- Yu, Y., Li, S., Wang, H., & Bi, L. (2015). Comprehensive network analysis of genes expressed in human oropharyngeal cancer. *American Journal of Otolaryngology - Head and Neck Medicine and Surgery*, 36(2), 235–241. <https://doi.org/10.1016/j.amjoto.2014.11.002>
- Zahn, J. M., Poosala, S., Owen, A. B., Ingram, D. K., Lustig, A., Carter, A., Weeraratna, A. T., Taub, D. D., Gorospe, M., Mazan-Mamczarz, K., Lakatta, E. G., Boheler, K. R., Xu, X., Mattson, M. P., Falco, G., Ko, M. S. H., Schlessinger, D., Firman, J., Kummerfeld, S. K., ... Becker, K. G. (2007). AGEMAP: A gene expression database for aging in mice. *PLoS Genetics*, 3(11), 2326–2337. <https://doi.org/10.1371/journal.pgen.0030201>
- Zeeberg, B. R., Feng, W., Wang, G., Wang, M. D., Fojo, A. T., Sunshine, M., Narasimhan, S., Kane, D. W., Reinhold, W. C., Lababidi, S., Bussey, K. J., Riss, J., Barrett, J. C., & Weinstein, J. N. (2003). GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol.*, 4(4), R28. <https://doi.org/10.1186/gb-2003-4-4-r28>
- Zhang, G., Li, J., Purkayastha, S., Tang, Y., Zhang, H., Yin, Y., Li, B., Liu, G., & Cai, D. (2013). Hypothalamic programming of systemic ageing involving IKK- β , NF- κ B and GnRH. *Nature*, 497(7448), 211–216. <https://doi.org/10.1038/nature12143>

10. BIBLIOGRAPHY OF THE CANDIDATE'S PUBLICATIONS

10.1. Publications related to this Thesis

File, Bálint*, Tibor Nánási*, Emília Tóth, Virág Bokodi, Brigitta Tóth, Boglárka Hajnal, Zsófia Kardos, László Entz, Loránd Eröss, István Ulbert, and Dániel Fabó. 2020. “Reorganization of Large-Scale Functional Networks during Low-Frequency Electrical Stimulation of the Cortical Surface.” *International Journal of Neural Systems* 30 (3): 1–15. <https://doi.org/10.1142/S0129065719500229>.

* *contributed equally*

Nánási, Tibor, Bálint File, Emília Tóth, László Entz, István Ulbert, Dániel Fabó, and Loránd Eröss. 2016. “Synergism of Spectral and Coupling Modalities in Epileptic Focus Localization from IEEG Recordings.” *PRNI 2016 - 6th International Workshop on Pattern Recognition in Neuroimaging*, 2–5. <https://doi.org/10.1109/PRNI.2016.7552354>.

Lehallier, Benoit, David Gate, Nicholas Schaum, Tibor Nanasi, Song Eun Lee, Hanadie Yousef, Patricia Moran Losada, Daniela Berdnik, Andreas Keller, Joe Verghese, Sanish Sathyan, Claudio Franceschi, Sofiya Milman, Nir Barzilai, Tony Wyss-Coray. 2019. “Undulating changes in human plasma proteome profiles across the lifespan.” *Nature Medicine*. <https://doi.org/10.1038/s41591-019-0673-2>.

Lehallier, Benoit, David Gate, Nicholas Schaum, Tibor Nanasi, Song Eun Lee, Hanadie Yousef, Patricia Moran Losada, Daniela Berdnik, Andreas Keller, Joe Verghese, Sanish Sathyan, Claudio Franceschi, Sofiya Milman, Nir Barzilai, Tony Wyss-Coray. 2019. “Undulating changes in human plasma proteome across lifespan are linked to disease.” *bioRxiv*.
<https://doi.org/10.1101/751115>

10.2. Publications not related to this Thesis

Richárd Fiáth, Katharina T Hofer, Vivien Csikós, Domonkos Horváth, Tibor Nánási, Kinga Tóth, Frederick Pothof, Christian Böhler, Maria Asplund, Patrick Ruther, and István Ulbert. 2018. “Long-term recording performance and biocompatibility of chronically implanted cylindrically-shaped, polymer-based neural interfaces. ” *Biomedical Engineering/Biomedizinische Technik* 63(3): 301-315.
<https://doi.org/10.1515/bmt-2017-0154>

Banlaki, Zsofia, Zsuzsanna Elek, Tibor Nanasi, Anna Szekely, Zsofia Nemoda, Maria Sasvari-Szekely, and Zsolt Ronai. 2015. “Polymorphism in the Serotonin Receptor 2a (HTR2A) Gene as Possible Predisposal Factor for Aggressive Traits.” *PloS One* 10 (2): e0117792. <https://doi.org/10.1371/journal.pone.0117792>.

Kovacs-Nagy, Reka, Zsuzsanna Elek, Anna Szekely, Tibor Nanasi, Maria Sasvari-Szekely, and Zsolt Ronai. 2013. “Association of Aggression with a Novel MicroRNA Binding Site Polymorphism in the Wolframin Gene.” *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics* 162 (4): 404–12. <https://doi.org/10.1002/ajmg.b.32157>.

Mihalik, Ágoston, Ambrus S. Kaposi, István A. Kovács, Tibor Nánási, Robin Palotai, Ádám Rák, Máté S. Szalay-Beko, and Peter Csermely. 2012. “How Creative Elements Help the Recovery of Networks after Crisis: Lessons from Biology.” *Networks in Social Policy Problems*, 179–88. <https://doi.org/10.1017/CBO9780511842481.010>.

Kiss, Huba J.M., Ágoston Mihalik, Tibor Nánási, Bálint Ory, Zoltán Spiró, Csaba Soti, and Peter Csermely. 2009. “Ageing as a Price of Cooperation and Complexity: Self-Organization of Complex Systems Causes the Gradual Deterioration of Constituent Networks.” *BioEssays* 31 (6): 651–64. <https://doi.org/10.1002/bies.200800224>.

Simkó, Gábor I, Dávid Gyurkó, Dániel V Veres, Tibor Nánási, and Peter Csermely. 2009. “Network Strategies to Understand the Aging Process and Help Age-Related Drug Design.” *Review Genome Medicine* 1 90. <http://doi.org/10.1186/gm90>

11. ACKNOWLEDGEMENTS

I would like to express my gratitude towards my former and current teachers, especially Beáta Hegedüs, Balázs Szilágyi, Péter Csermely, Mária Sasvári-Székely, István Ulbert, Tony Wyss-Coray, and Benoit Lehallier. My grateful thanks are also extended to all the members of the Ulbert and Wyss-Coray laboratories.

Finally, I would like to thank my Parents, and my dog, Gonzó, for their patience and support.