

FAIR and bias-free network modules for mechanism-based disease redefinitions

Citation for published version (APA):

Hassan, A. A. A. M. (2021). *FAIR and bias-free network modules for mechanism-based disease redefinitions*. [Doctoral Thesis, Maastricht University]. Maastricht University.
<https://doi.org/10.26481/dis.20210707ah>

Document status and date:

Published: 01/01/2021

DOI:

[10.26481/dis.20210707ah](https://doi.org/10.26481/dis.20210707ah)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

FAIR AND BIAS-FREE NETWORK MODULES FOR MECHANISM-BASED DISEASE REDEFINITIONS



AHMED A. HASSAN

Doctoral thesis

**FAIR and bias-free network modules for
mechanism-based disease redefinitions**

Ahmed A. Hassan

2021

FAIR and bias-free network modules for mechanism-based disease redefinitions

DISSERTATION

To obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus,
Prof. Dr. Rianne M. Letschert,
in accordance with the decision of the Board of Deans,
to be defended in public on Wednesday 7th of July 2021, at 10.00

by

Ahmed Adel Ahmed Mohamed Hassan
Born on 7 September 1990, Egypt

Promotor

Prof. Dr. Harald H.H.W. Schmidt

Copromotor

Prof. Dr. Michel J. Dumontier

Assessment Committee

Prof. Dr. Ir. Ilja C.W. Arts

Prof. Dr. Ir. Chris T.A. Evelo

Prof. Dr. Jörg Menche (CE-M-M Research Center for Molecular
Medicine of the Austrian Academy of Sciences, Vienna,
Austria)

Dr. Jaanika Kronberg-Guzman (Institute of Genomics,
University of Tartu, Estonia)

© Ahmed A. Hassan, Maastricht 2021.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior written permission of the author.

Cover Ahmed A. Hassan, 2021

Production Ipskamp Printing B.V.

ISBN 978-94-6421-395-9

*To my family and my lovely wife who provided me with unwavering support
on this journey*

Contents

1	Introduction	5
1.1	References	11
2	On the need to overcome inconsistent symptom-based disease classifications	15
2.1	Introduction	17
2.2	Disease classification in biomedical research, clinical practice and its limitations	19
2.3	Impact of current limitations in disease classification on medicine	22
2.4	The need for change: endotypes in asthma and inflammatory bowel diseases	25
2.5	A translational gap: incompatibilities between classifications	27
2.6	Towards a mechanism-based taxonomy and new medicine	32
2.7	Implications of mechanism-based disease classification for drug discovery	35
2.8	Conclusion	36
2.9	References	44
3	Transcription Factor NRF2 as a Therapeutic Target for Chronic Diseases: A Systems Medicine Approach	53
3.1	Introduction	55
3.2	From Nuclear Factor (Erythroid-Derived 2)– Like 2 Interactome to Nuclear Factor (Erythroid- Derived 2)–Like 2 Diseasome	56
3.3	Target Validation of Nuclear Factor (Erythroid-Derived 2)–Like 2 in Human Disease States	63
3.4	The Kelch-Like ECH-Associated Protein 1 Paradox in Cancer	85
3.5	Nuclear Factor (Erythroid-Derived 2)– Like 2 Drugome .	89
3.6	Conclusions	111
3.7	References	112

4	From single drug targets to synergistic network pharmacology in ischemic stroke	155
4.1	Introduction	157
4.2	Results	158
4.3	Discussion	167
4.4	Material and Methods	172
4.5	Acknowledgments	173
4.6	References	174
5	NOX5-induced hypertension endotype	205
5.1	Introduction	207
5.2	Results	208
5.3	Discussion	219
5.4	References	232
6	Towards FAIR protocols and workflows: The OpenPREDICT case study	259
6.1	Introduction	261
6.2	Background	262
6.3	The FAIR workflows approach	265
6.4	OpenPREDICT Case Study	272
6.5	Evaluation	278
6.6	Discussion	288
6.7	Conclusions	293
6.8	Acknowledgments	294
6.9	References	295
7	Un-biased housekeeping gene selection for high-validity expression analysis	303
7.1	Introduction	306
7.2	Results	307
7.3	Discussion	318
7.4	Materials and Methods	320
7.5	References	328
8	General Discussion	343
8.1	Disease definitions: a major roadblock in medicine	344

8.2	Network medicine and the future of disease diagnostics and therapeutics	344
8.3	Overcoming reproducibility crisis in the FAIR way	346
8.4	Conclusion	347
8.5	References	349
9	Summary	351
10	Impact	355
	Included work	359
	About the author	361
	Acknowledgments	363

1

Introduction

The ultimate goal of medicine is to understand the root causes of diseases, which then allows us to find their cures. Despite intensive research on decoding common diseases and the remarkable increase in human life expectancy, we are but slightly closer to that goal than we were a century ago. In the case of diabetes, for example, the first line of treatment drug was introduced over 60 years ago [1]. While for stroke, the second most cause of death worldwide, the situation is even direr with only one acute treatment available introduced over 20 years ago that leaves half of the patients impaired or dead [2]. In fact, none of the drugs used has been developed to address the causal factors of those conditions.

Without understanding disease mechanisms, however, the main focus of drug therapy has been to alleviate disease symptoms and to chronically manage conditions rather than curing them. Moreover, the drug discovery process has been mainly driven by hypotheses and increasingly fails to deliver patient benefit (and successful market entry). The aggregated socioeconomic effects of these failures are huge in terms of costs on the pharmaceutical industry and healthcare systems, but most importantly on the life expectancy of patients and their quality of life [3, 4].

Current definitions of diseases are inherently biased as they depend heavily on obsolete classifications of pathologies. The International Classification of Diseases, which is the de-facto clinical classification of diseases in clinical practice developed in 1893 and adopted by the World Health Organisation in 1946, distinguishes diseases either the symptoms they present or by organs they affect [5]. Such classification ignores the fact that symptoms are the end result of a long chain of biological interactions and organs are but subcomponents in an interconnected system. More importantly, this has deeply established the idea that diseases can be studied and treated in isolation. Therefore we can not undertake the task of understanding disease mechanisms without properly redefining them.

With the possibility of whole genome sequencing and genome-wide association studies, it was hoped that the elucidation of disease causes would accelerate. This has primarily been the case

only for rare, monogenic diseases. Chronic or so-called complex diseases have remained a black box despite numerous associated or correlated genes. The introduction of modern bioinformatics and network science into medicine is leading now, however, to a redefinition of diseases by abandoning organ and symptom-based taxonomies. According to these network and systems approaches [6], a disease is a continuum of causal factors and interactions over time leading to a deviation from the normal state of a living system. In practice, however, this is difficult to study and alter, therefore we rely on reductionism [7, 8], which postulates that we can fully describe a complex system given an understanding of its individual parts. Under such a reductionist framework, diseases are to be defined based on a multiplicity of factors, including symptoms, organs, genes, however, neglecting their interplay.

Network medicine provides a new perspective on how to study and explain diseases. It views diseases as modules within complex networks of genetic, proteomic, metabolic, regulatory, and environmental interactions of which pathophenotypes are emergent properties due to the perturbations in one or more of these modules [9]. Network medicine is one of the fields fueled by the concepts of network science, and it transfers its knowledge and approaches to the domain of medicine, enabling the study of biological networks topology and dynamics. Diseases can be naturally described and explored in the light of network medicine, giving rise to more mechanistic definitions in terms of the networks representing multiple levels of living systems and the interactions of their subcomponents.

Exacerbating this challenge is the crisis of reproducibility. In a recent Nature survey [10], more than 70% of participating researchers said they were unable to reproduce the results of published research, sometimes of their own. Numerous studies have identified huge numbers of risk target genes for all human common diseases, yet seldom do these targets translate to efficacious drugs across different populations [11, 12]. One major cause of this issue is the inaccessibility to research, however, open access to publications alone is not enough as it is essential to have access to the data and the exact methods used in studies. With the advent of big data and machine learning techniques in the

biomedical field, this challenge only gets more serious and requires a radical solution.

In this thesis, I focus on the demonstration and application of network medicine approaches to address the gap of non-mechanistic disease definitions, as well as the gap of irreproducibility of the majority of basic and computational research in the biomedical field. In the forthcoming chapters, **I try to address six main research questions:**

[RQ1]	Are the current disease classification systems suitable for precision medicine?	Chapter 2
[RQ2]	How can we use network medicine to understand the role of NRF2 in metabolic processes and diseases?	Chapter 3
[RQ3]	Can we predict synergistic drug targets for ischemic stroke using network medicine approaches?	Chapter 4
[RQ4]	Is network medicine able to discover unknown disease mechanisms in hypertension?	Chapter 5
[RQ5]	How can we apply FAIR principles to experimental analysis workflows?	Chapter 6
[RQ6]	Can we predict housekeeping gene candidates under different experimental conditions? How to make it accessible and reproducible?	Chapter 7

To further guide the reader, brief summaries of the respective chapters of the dissertation are given below.

[RQ1] Are the current disease classification systems suitable for precision medicine?

To appreciate the importance of a new approach to disease definitions, disease ontologies representing those definitions as structured classifications or taxonomies, are tackled first in **chapter 2**. Existing ontologies are reviewed and the different goals of each are discussed. Indeed for most ontologies, the common problem is that they ignore mechanistic factors of disease in their definitions and classification. International Classification of Diseases (ICD), Medical Subject Headings (MeSH), and Online Mendelian Inheritance in Man (OMIM) to name a few, are all examples of frequently used taxonomies, both clin-

ically and academically, that classify disease definitions to body localizations or apparent symptoms [13]. Big health data, rich omics, network medicine are replacing such understanding with definitions that capture the molecular basis of diseases and the heterogeneity of their manifestations in individuals. Only in such context, phenomena validated by strong scientific evidence such as endotypes [14–17] and single cellular mechanisms leading to multiple comorbidities [18] are justified and explained.

[RQ2] How can we use network medicine to understand the role of NRF2 in metabolic processes and diseases?

Having made the case for mechanistic disease redefinitions and the utility of systems and network medicine in such an endeavour, the next steps are applications of network medicine approaches to different therapeutic areas. In **chapter 3**, the role of NRF2 transcription factor is studied within the human interactome, using multiscale networks to find implicated pathways, diseases, and relevant drug candidates [19].

[RQ3] Can we predict synergistic drug targets for ischemic stroke using network medicine approaches?

Recognizing a highly unmet medical need in stroke [20], an integrative multi-target approach was developed in **chapter 4**, following the techniques of network medicine. The primary target, NADPH oxidase type 4 (Nox4), is chosen as a strongly validated target involved in neurotoxicity and poststroke blood-brain barrier dysfunction, and network medicine techniques are used to predict its synergistic targets.

[RQ4] Is network medicine able to discover unknown disease mechanisms in hypertension?

In **chapter 5**, network medicine approaches are applied to predict a molecular mechanism applicable to a subset of hypertensive patients, indicating the existence of an endotype that involves the target NADPH oxidase type 5 (Nox5). The NADPH oxidase enzyme family has been reported by genome-wide association studies (GWAS) stud-

ies to be involved in higher hypertension risk following the oxidative stress mechanism [21].

[RQ5] How can we apply FAIR principles to experimental analysis workflows?

With regards to the challenge of reproducibility, **chapter 6** presents a framework developed using the FAIR (Findable, Accessible, Interoperable, and Reusable) principles which can be applied to research workflows. The FAIR data principles, now widely accepted in the scientific community, offer guidelines on improving the reproducibility of published research. The GO FAIR initiative provides a set of implementations that have been mainly focused on the reproducibility of data artifacts consumed or produced by experiments, however, the process by which the data is generated, analysed, or transformed is as important for arriving at the same results.

[RQ6] Can we predict housekeeping gene candidates under different experimental conditions? How to make it accessible and reproducible?

Another challenge that persisted during the in-vitro and in-vivo validation experiments, and is significantly affecting the reproducibility of wet-lab results, is the normalization of gene expression studies [22]. A major factor is the selection of housekeeping genes for normalization, addressed in **chapter 7**. Current methods of relying on previous studies and generic reference genes are inherently biased as there is strong evidence that there are no universal housekeeping genes under all organisms, tissues, conditions [23].

Finally, **chapter 8** concludes with summarizing the answers to the research questions, and a general discussion on the contributions of the dissertation as a whole, as well as future perspectives on network medicine and disease definitions.

1.1 References

- [1] J. C. Clapham. “Sixty Years of Drug Discovery for Type 2 Diabetes: Where Are We Now?” In: *Methods in Molecular Biology*. 2019, 1–30. DOI: 10.1007/978-1-4939-9882-1_1.
- [2] D. W. Howells, E. S. Sena, and M. R. Macleod. “Bringing rigour to translational medicine”. In: *Nature Reviews Neurology* 10.1 (2013), 37–43. DOI: 10.1038/nrneuro1.2013.232.
- [3] O. J. Wouters, M. McKee, and J. Luyten. “Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018”. In: *JAMA* 323.9 (2020), p. 844. DOI: 10.1001/jama.2020.1166.
- [4] J. A. DiMasi, H. G. Grabowski, and R. W. Hansen. “Innovation in the pharmaceutical industry: New estimates of R&D costs”. In: *Journal of health economics* 47 (2016), 20–33. DOI: 10.1016/j.jhealeco.2016.01.012.
- [5] M. G. Dozmorov. “Reforming disease classification system—are we there yet?” In: *Annals of Translational Medicine* 6.S1 (2018), S30–S30. DOI: 10.21037/atm.2018.09.36.
- [6] R. Berlin, R. Gruen, and J. Best. “Systems Medicine Disease: Disease Classification and Scalability Beyond Networks and Boundary Conditions”. In: *Frontiers in Bioengineering and Biotechnology* 6 (2018). DOI: 10.3389/fbioe.2018.00112.
- [7] A. C. Ahn et al. “The Limits of Reductionism in Medicine: Could Systems Biology Offer an Alternative?” In: *PLoS Medicine* 3.6 (2006), e208. DOI: 10.1371/journal.pmed.0030208.
- [8] A.-L. Barabási. “The network takeover”. In: *Nature Physics* 8.1 (2011), 14–16. DOI: 10.1038/nphys2188.
- [9] J. Baumbach and H. H. Schmidt. “The End of Medicine as We Know It: Introduction to the New Journal, Systems Medicine”. In: *Systems Medicine* 1.1 (2018), 1–2. DOI: 10.1089/sysm.2017.28999.jba.
- [10] M. Baker. “1,500 scientists lift the lid on reproducibility”. In: *Nature* 533.7604 (2016), 452–454. DOI: 10.1038/533452a.

- [11] L. Shu, M. Blencowe, and X. Yang. “Translating GWAS Findings to Novel Therapeutic Targets for Coronary Artery Disease”. In: *Frontiers in Cardiovascular Medicine* 5 (2018). DOI: 10.3389/fcvm.2018.00056.
- [12] L. B. Scheinfeldt et al. “Challenges in Translating GWAS Results to Clinical Care”. In: *International Journal of Molecular Sciences* 17.8 (2016). DOI: 10.3390/ijms17081267.
- [13] M. Hofmann-Apitius et al. “Towards the taxonomy of human disease”. In: *Nature Reviews Drug Discovery* 14.2 (2015), 75–76. DOI: 10.1038/nrd4537.
- [14] G. P. Anderson. “Endotyping asthma: new insights into key pathogenic mechanisms in a complex, heterogeneous disease”. In: *The Lancet* 372.9643 (2008), 1107–1119. DOI: 10.1016/S0140-6736(08)61452-X.
- [15] C. A. Cooper et al. “Common variant rs356182 near SNCA defines a Parkinson’s disease endophenotype”. In: *Annals of Clinical and Translational Neurology* 4.1 (2016), 15–25. DOI: 10.1002/acn3.371.
- [16] A. Agustí, B. Celli, and R. Faner. “What does endotyping mean for treatment in chronic obstructive pulmonary disease?” In: *The Lancet* 390.10098 (2017), 980–987. DOI: 10.1016/S0140-6736(17)32136-0.
- [17] M. Jutel and P. Gajdanowicz. “Chapter 5 - Revised Disease Nomenclature Including Disease Endotypes”. In: *ScienceDirect*. Ed. by I. Agache and P. Hellings. Academic Press, 2019, 27–29.
- [18] S. D. Ghiassian et al. “Endophenotype Network Models: Common Core of Complex Diseases”. In: *Scientific Reports* 6.1 (2016), 1–13. DOI: 10.1038/srep27414.
- [19] A. Cuadrado et al. “Transcription Factor NRF2 as a Therapeutic Target for Chronic Diseases: A Systems Medicine Approach”. In: *Pharmacological Reviews* 70.2 (2018). Ed. by M. C. Michel, 348–383. DOI: 10.1124/pr.117.014753.
- [20] T. Chen et al. “Long-term unmet needs after stroke: systematic review of evidence from survey studies”. In: *BMJ Open* 9.5 (2019), e028137. DOI: 10.1136/bmjopen-2018-028137.

-
- [21] R. M. Touyz et al. "NOX5: Molecular biology and pathophysiology". In: *Experimental Physiology* 104.5 (2019), 605–616. DOI: 10.1113/ep086204.
 - [22] I. V. Vyang et al. "Within the fold: assessing differential expression measures and reproducibility in microarray assays". In: *Genome Biology* 3.11 (2002), research0062.1. DOI: 10.1186/gb-2002-3-11-research0062.
 - [23] E. M. Tunbridge, S. L. Eastwood, and P. J. Harrison. "Changed Relative to What? Housekeeping Genes and Normalization Strategies in Human Brain Gene Expression Studies". In: *Biological Psychiatry* 69.2 (2011), 173–179. DOI: 10.1016/j.biopsych.2010.05.023.

2

On the need to overcome inconsistent symptom-based disease classifications

Adapted from: Elisa Anastasi, **Ahmed A. Hassan**, Johannes Boltze, Ana I Casas, Mahmoud Elbatreek, Alexander G. B. Grønning, Zeinab Mamdouh, Cristian Nogales, Edwin K. Silverman, Anil Wipat, Harald H.H.W. Schmidt, Emre Guney. “On the need to overcome symptom-based and domain-specific disease classifications towards translational systems medicine” – *First Full Draft*.

Abstract

2

Drug therapy is imprecise. The vast majority of compounds that meet regulatory efficacy criteria require a high number of treated individuals to prove their population-wide efficacy. Most disease definitions lack a causal mechanistic understanding impeding precision diagnosis and treatment. What unites these shortcomings of the current state of medicine is their reliance on 19th/20th-century disease definitions that mainly describe symptoms within silos of organ-based disciplines. Systems and Network medicine, instead, unifies comorbidities, shared risk or causal genes, pleiotropic drugs, shared symptoms, and biomarkers for an integrated, multiscale network of diseases. Current phenotypes are endotyped to unravel the hidden causal mechanism and establish an entirely new disease taxonomy. However, a major roadblock in achieving this are the profound inconsistencies of disease ontologies used across genetic, preclinical and clinical research, healthcare and epidemiology. Here, we analyse the extent of this problem and suggest adopting an integrated uniformly used phenotype ontology for the initial research phase. Founded on several proof-of-concept studies, however, we predict to finally reach an entirely new mechanism-based disease ontology for molecular pathology, mechanism-based drug repurposing and precision medicine.

2.1 Introduction

The classification of diseases is a consequence of our desire to bring order to the complexity surrounding the biomedical universe. Conventional disease classification offers a standardised language, allowing health professionals, scientists, and patients to name diseases together with their stages and subtypes, in a consistent and concise manner. Nosology, the branch of science aiming to classify diseases, dates back to Hippocrates and the 5th century B.C. However, the first version of the International Classification of Diseases (ICD) was developed in 1893 and has been recurrently updated by the World Health Organization (WHO) every ten years as a widely used classification scheme [1]. Since ICD was established, an increasing number of other general or disease-specific classification systems have also been suggested (Table 2.1), aiming to address several shortcomings that were not catered for by others, i.e. structural organization, classification specificity, or direct applicability. Thus, as an attempt to harmonize different disease definitions, initiatives like (i) Disease Ontology (DO) [2], and (ii) Mondo Disease Ontology (Mondo) [3] emerged. These platforms aim to overcome the heterogeneity concerning disease terminology by providing a formal semantic model for integrating multiple ontologies, nomenclatures, dictionaries and vocabularies. However, a homogeneous, curated, broadly applicable, and publicly available disease classification system is currently a long-term goal rather than a current reality. Here, we first evaluated the most relevant disease classifications commonly used in biomedical research and clinical practice, focusing on difficulties associated with heterogeneous disease definitions and the impossibility to map different disease ontologies. Moreover, we propose to entirely redefine the concept of disease by a purely mechanistic approach far from the common organ- and symptom-based classifications. Additionally, we present the most recent systems medicine applications that leverage large-scale molecular and clinical data to better understand connections between diseases. Indeed, we further define how to reconcile currently existing disease misclassifications to ultimately endotype patients based on their impaired underlying pathomechanism (Figure 2.1).

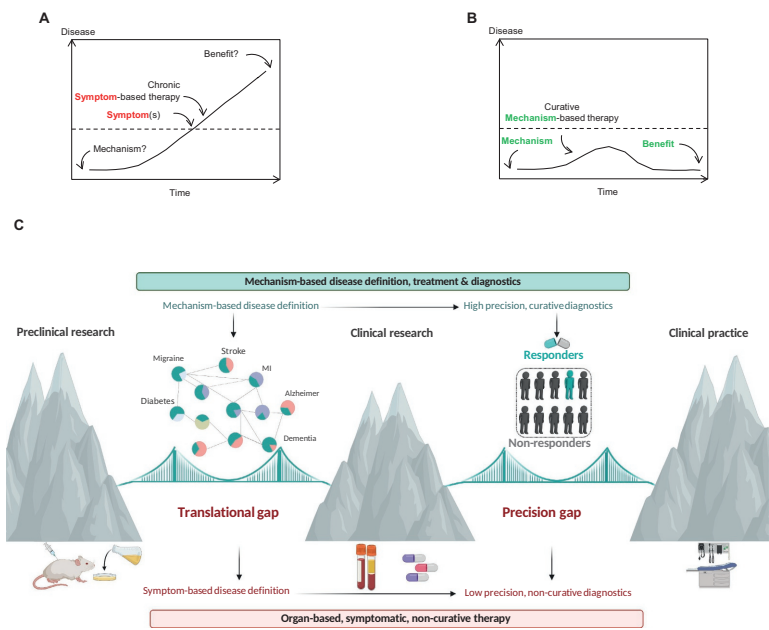


Figure 2.1: Schematic overview of disease classification from traditional and systems medicine perspectives and the therapeutic consequences. (A) The current approach for chronic diseases treatment is depicted. Diseases are not identified until symptoms start to manifest, then delayed therapy is applied targeting the alleviation of symptoms. (B) While with mechanistic-based therapy, precise biomarkers are used to identify disease well before symptoms onset. (C) The current landscape for the medical field, with large gaps between the different domains. Below, the standard low precision approach is shown. Mechanism-based disease definitions relying on multi-modal network approaches can bridge the translational gap between preclinical and clinical research. Subsequently, high precision diagnostics derived from network approaches can be used to identify endophenotype groups within populations and apply treatments curated for the underlying mechanisms.

2.2 Disease classification in biomedical research, clinical practice and its limitations

In biomedical and clinical domains, disease classifications facilitate (i) patient diagnoses; (ii) the characterisation of newly discovered disease subtypes; (iii) clinical decisions; and (iv) data integration and analysis. Due to various historical and practical considerations on how diseases are perceived in different research areas, an overwhelming number of disease classification efforts exist in the literature (Table 2.1). These classifications differ in various aspects including structure, domain coverage, richness, complexity, community acceptance, maintenance, license, and construction method.

Table 2.1: Clinical annotation and classification systems used in clinical practice and biomedical research

Ontology	Name	Hierarchical?	Size (number of terms)	Complexity	Application field
ICD9, 10, 11	International Classification of Diseases	Yes	17561 (ICD9), 91737 (ICD10), 31055 (ICD11) (as of 05/11/2018)	Simple taxonomy	Electronic Health Records, Clinical Coding, Qualitative Research
SNOMED CT	Systematized Nomenclature of Medicine - Clinical Terms	Yes	75890 (as of 18/12/2018)	Simple taxonomy	Electronic Health Records, Natural Language Processing, Medical Records Systems, Computerized Forms and Records Control
OMIM	Online Mendelian Inheritance in Man	No	8743 (as of 01/11/2018)	Flat list	Computational Biology, Genomics, Genome-Wide Association Study
MeSH	Medical Subject Headings	Yes	11572 (v2018), 11648 (v2019) (as of 01/11/2018)	Complex taxonomy	Information Storage and Retrieval, Abstracting and Indexing, Natural Language Processing
MONDO	Monarch Disease Ontology	Combines multiple disease resources	23263 (as of 12/11/2018)	Complex taxonomy. Semi-automatically constructed	Computational Biology, Whole Genome Sequencing, Translational Medical Research
DO	Disease Ontology	Yes	9095 (as of 11/11/2020)	Simple taxonomy	Computational Biology, Molecular Sequence Annotation, Gene Expression Profiling
HPO	Human Phenotype Ontology	Multiple disease resources	17387 (as of 05/11/2018)	Multiple disease resources: medical literature, Orphanet, DECIPHER, and OMIM.	Computational Biology, Gene Expression Profiling, Molecular Sequence Annotation
ORDO	Orphanet Rare Diseases Ontology	poly-hierarchical classification	13871 (as of 05/11/2018)	Complex taxonomy	Rare Diseases, Orphan Drug Production
MEDIC	CTD's merged disease vocabulary	Yes	12984 (as of 05/11/2018)	Complex taxonomy Mainly subset of MeSH, contains Multiple disease identifiers: MeSH, OMIM	Database Management Systems, Information Storage and Retrieval, Toxicogenetics
UMLS	Unified Medical Language System	N/A	N/A	Huge number of Integrated terminologies, classifications and coding standards	Natural Language Processing, Information Storage and Retrieval, Electronic Health Records

Table 2.1 continued from previous page

Ontology Name	Hierarchical?	Size (number of terms)	Complexity	Application field
EFO Experimental Factor Ontology	Yes	22126 (as of 07/11/2018)	Simple taxonomy	Gene Expression Profiling, Computational Biology, Gene Regulatory Networks
MedDRA Medical Dictionary for Drug Regulatory Activities	Yes	71208 (as of 20/02/2019)	Simple taxonomy	Adverse Drug Reaction Reporting Systems, Drug-Related Side Effects and Adverse Reactions, United States Food and Drug Administration
IDOMAL Malaria Ontology	Yes	3158 (as of 20/02/2019)	Complex taxonomy	Gene Expression Profiling, Computational Biology, Molecular Sequence Annotation
IDO Infectious Disease Ontology	Yes	518 (as of 12/11/2018)	Complex taxonomy	Gene Expression Profiling, Computational Biology, Host-Pathogen Interactions
CVDO Cardiovascular Disease Ontology	Yes	479 (as of 12/11/2018)	Complex taxonomy	Gene Expression Profiling, Gene Regulatory Networks, Cardiovascular Diseases
MFOMD Mental Disease Ontology (Mental Functioning Ontology)	Yes	358 (as of 20/02/2019)	Complex taxonomy	Gene Expression Profiling, Genome-Wide Association Study, Alzheimer Disease, Schizophrenia

The current definition of diseases is currently structured around (i) the affected organ or tissue, i.e. colon-colitis, liver-hepatitis, (ii) present symptoms, i.e. irritable bowel syndrome, depressive disorder, or (iii) linked to the physician who first identified certain patterns in a patient, i.e. Alzheimer's disease, Parkinson's disease. Therefore, such ad hoc taxonomy guides the operational structure of medicine to organ-based clinics (i.e. cardiology, neurology) and vocational specialisations (i.e. cardiologist, neurologist). Yet, the technical advances in diagnostic tools such as high-throughput genotyping and optical imaging enhance the disease identification process and how it relates to other anatomical conditions. Most human diseases involve multiple environmental and genetic factors, where their manifestations in the body vary across individuals depending on each patient's characteristics. Accordingly, the current classification of diseases reflects the existing, often incomplete, knowledge on the disease pathology where physicians use different diagnostic tests, phenotypic signs, and symptoms to assess the severity of the condition without considering the underlying mechanistic pathways.

In this review, we use the term disease classification (or disease

taxonomy) inclusively to refer to different means of disease definition and categorisation systems including ontologies, nomenclatures, dictionaries and vocabularies (see Appendix A for key differences between several disease classification efforts widely used in biomedical research).

While well-established biomarkers or clearly defined symptoms exist, not all clinical observations are quantitative, allowing qualitative interpretation of clinical observations. Furthermore, both the heterogeneity of clinical manifestation across patients and comorbid conditions accompanying the disease (multi-morbidity) are rarely accounted for currently existing disease taxonomies, therefore impeding designing and conducting clinical trials [4]. Asthma could be considered one of the most representative cases where the disease's current definition has diverged substantially from the original due to various underlying causes. Indeed, the disease descriptions that reflect the pathobiology across patient subgroups affect diagnostics and therapeutics and play a dominant role in socio-economic decisions such as insurance claim payments or clinical annotation platforms used in healthcare institutions (see Appendix B).

The landscape of medical terminologies and clinical annotation systems is far from complete. We give a compact thus a good overview of different clinical annotation efforts used across healthcare institutions worldwide (Table 2.1 and Appendix B). However, none of these efforts is fully comprehensive or sufficient to overcome all academia, clinics, and industry (Appendix A and Appendix B). Given our current inaccurate disease definition and recurrent failure of the drug discovery industry within the last years, there is a clear need to move towards a finer-grain and interoperable disease definition that captures diseases' molecular heterogeneity among different individuals.

2.3 Impact of current limitations in disease classification on medicine

Medication is the most common intervention in health care. However, for several drugs on the market and therapeutic guidelines, population-based studies failed to show patient-relevant benefits

[5, 6]. Even worse, some antiarrhythmic drugs [7] and blood pressure-lowering alpha-blockers increased mortality [8]. For indications such as stroke, only a single drug, tPA, with more than 30 contraindications, is currently available for causal therapy, leaving 85% of patients without treatment [9]. Moreover, most marketed drugs have an extremely high number needed to treat (NNT), requiring several patients to be treated before identifying any therapeutic benefit. Concerning discovering new, more precise drugs, research efficiency halves every nine years, leading to insufficient novelty, project failures, and recurrent scientific frustration [10]. In fact, within the last decade, primarily immunosuppressants, antineoplastic, and antiviral drugs were marketed, noticing the field's poor efficiency ratio. In line with this failure, massive financial loss in the pharmaceutical industry resulted in abandoned diseases, symptomatic therapies and ultimately untreatable patients.

The imprecision and failures of the drug field are closely coupled with our current disease definition almost intact since the 19th century. The drug discovery field often relies on treating symptoms, or normalising risk factors, assuming that by modulating a surrogate parameter the ultimate therapeutic goal will be reached, i.e. cholesterol reduction for heart failure prevention. However, neither symptoms nor risk factors are mechanistic definitions of a disease not leading to patient cure but just to pure symptomatic relief. In fact, most disease phenotypes are partially unknown or not mechanically understood with the exception of specific rare diseases. In that scenario, a precise (often single and severe) mutation has been identified where, in some cases, specific therapy is available.

The phenotypic definitions of the existing disease classification systems do not imply any relationship between the clinically observable characteristics and a causal underlying mechanism. In particular umbrella terms such as 'common' or 'complex' diseases or 'syndromes' artificially unify multiple underlying mechanisms into one disease only because of a similar phenotype [11–14]. These different mechanisms represent the endotypes of the visible phenotype and represent the lead towards a mechanistic understanding and redefinition of disease.

Thus our current medical taxonomy entirely focused on organ-based classifications is not sustainable. In fact, disease pathomechanisms cannot be entirely identified by simply focusing on one specific organ, without considering further biologically mechanistic connections. With the exception of specific genetic diseases, common pathomechanisms are caused by several molecular impaired mechanisms, e.g. signalling or metabolic events, which become dysfunctional. However, these mechanisms are broadly conserved and therefore relevant in different organs. Thus, an impaired biological mechanism will ultimately lead to disease in all tissues where that mechanism plays an essential physiological role, not exclusively on a single organ.

Importantly, we are currently facing a vast innovation roadblock primarily related to poor research quality and constant irreproducibility as one of the major limitations for a successful clinical translation [15, 16]. Reasons for this include positive publication bias and several statistical errors [16, 17]. In fact, accurate design and reporting of pre-clinical experimentation similar to clinical practice, i.e. multi-centre, randomization and blinding, the inclusion of aged animals, both genders, etc. are rather the exception than the rule in preclinical stroke research resulting in a lack of reproducibility. Hence, three main knowledge gaps and roadblocks in our understanding of diseases include:

- Organ/symptom- rather than mechanism-based disease definitions and underuse of systemic approaches;
- Lack of relevance of pre-clinical disease models for target validation;
- Unsuccessful translation of pre-clinical research into clinical application.

Collectively, these lead to a therapeutic innovation roadblock exemplified by the decline in successful drug developments, lack of therapeutic precision (Figure 2.1) and unmet medical needs.

2.4 The need for change: endotypes in asthma and inflammatory bowel diseases

It is possible to replace descriptive phenotypes and umbrella terms step by step with mechanistic disease definitions similar to a rare disease. This requires so-called endotyping, i.e. splitting up a descriptive clinical disease term into endotypes that lead towards a molecular mechanism. The obvious leads are associated genes or effective drugs as both of them are associated with proteins (gene products or drug targets) and several proteins will form signalling modules [18], i.e. molecular mechanisms. For some endotypes, genes may not have been discovered or no drugs are yet available. In such cases, the molecular mechanisms may be derived from triangulation through the genes or drugs of associated comorbidities etc., a process which is called multiscale modelling and network medicine. Eventually all 'common', 'complex' or 'chronic' diseases and 'syndromes' will be endotyped and replaced by rarer or rare but precise mechanistic disease definition [19].

One prominent example of the coexistence of endotypes is asthma, a condition whose name derives from ancient Greek *αστημα* (*âsthma*, "laborious breathing"). The disease term asthma intrinsically involves a symptom-based disease definition (i.e. difficulties in breathing) or, in other words, subsumes a heterogeneous 'syndrome' or a 'complex' disease of unknown origin. While clinicians have made several attempts to subclassify asthma in a descriptive manner, ranging from different origins and triggers to different cell populations in the sputum of asthmatic patients and different degrees of severity, geneticists often treat asthma as a single entity with which genome-wide associations are studied [20]. Pre-clinical research studies of asthma can be performed in vitro through 3D cell culture models (e.g., bronchospheres, [21] and in vivo through animal models in *Drosophila*, mice, rats, guinea pigs, cats, dogs, pigs, primates and equines through different routes of sensitization and challenge. Here, despite being more traditional and leading to intense inflammation, ovalbumin has been replaced by aeroallergens, such as house dust mite, to use the allergens that may trigger some forms of human asthma [22].

In detail, early clinical disease classification systems included allergic vs. non-allergic asthma and childhood vs. adult-onset asthma, while more recent nosology has focused on endotypes based on the underlying immunological abnormalities (Th2 vs. non-Th2) as well as the impact of the inflammasome, epithelial barrier, and type 17 [23]. More recently, unsupervised cluster analysis approaches based on clinical and physiological characteristics of asthma have also been applied [24]. Although both cellular (e.g., sputum and blood eosinophilia) and various multi-omic biomarkers for asthma have been proposed [25], they have neither been widely utilized in clinical research nor in care.

Current asthma treatment is based on the severity of respiratory symptoms, physiological derangements, and exacerbations [23]. For mild to moderate asthmatics, bronchodilator medications (short and/or long-acting beta-2 agonists) and/or inhaled corticosteroids are employed. Only in severe asthmatics, who have persistent and uncontrolled disease manifestations, are any elements of disease endotypes considered in treatment decisions. In severe asthmatics with elevated total IgE levels, omalizumab (anti-IgE) can be utilized, while patients with elevated circulating eosinophils can receive treatment with mepolizumab (anti-IL5), reslizumab (anti-IL5), benralizumab (anti-IL5 receptor alpha), and dupilumab (anti-IL4 receptor alpha) [26]. Principled treatment of asthma based on disease endotypes and network analysis of multi-omics data could lead to greater efficacy and fewer adverse reactions.

With respect to identifying molecular causes (and possibly new diagnostic and/or therapeutic targets) for asthma, collaborative genome-wide association studies in need of large sample sizes have typically focused on a minimum common denominator disease definition (e.g., physician-diagnosed asthma [27]). However, genetic differences likely contribute to asthma heterogeneity; for example, different genetic associations in childhood and adult-onset asthmatics have been implicated [28]. As a result, many primary disease-relevant genetic determinants of asthma endotypes may have been missed.

Another example where no clear biomarkers are distinguishing

two related conditions is the case of Crohn's disease (CD) and ulcerative colitis (UC). Collectively known as inflammatory bowel disease, CD and UC are two types of autoimmune disorders affecting the colon and small intestine. Although diet, biodiversity of the microbiota, and various genetic loci are linked to these conditions, there is no disease-specific marker that can reliably distinguish CD and UC patients [29]. The level of calprotectin, a protein released by neutrophils, when present in the human faeces drives the diagnosis of these diseases which can be only confirmed by biopsy. Due to the symptomatic similarities between the two diseases, the location and severity of the inflammation across different organs are used to classify the disease. Accordingly, in a substantial number of cases, a definitive diagnosis cannot be made reliably [30].

When looking back at the history of science, discoveries like these resemble a scientific crisis, a term coined by the philosopher of science, Thomas Kuhn. A crisis in a scientific community may occur if significant findings are deemed as anomalies or inexplicable by the current founding theories and beliefs that govern the community.

2.5 A translational gap: incompatibilities between classifications

Identifying the causal underlying molecular mechanisms of diseases is essential for diagnosis, endo-phenotyping of patients, and subsequent personalised treatment. Multi-layered network analysis is routinely used to achieve this goal by establishing links between diseases on the one hand, and their associated genes, symptoms, comorbidities, pathways, protein similarity, protein interactions, and patient clinical data on the other hand. Unfortunately, different disciplines use different disease classifications that do not overlap hampering such multiscale modelling. Clinicians, for example, might use ICD as a description of patients' data, while biomedical researchers would use DO for annotating interconnected genetic and protein networks. A common approach to connect these disciplines is to rely on ontology mapping. The distinct purposes of disease, however, often lead to incompatibilities that hinder the translation from one domain to the other. In addition, the differences between

ontologies are non-systematic. While, for example, Medical Subject Headings (MeSH) is more verbose in the definition of hypertension, it is less descriptive of different taxonomic categorizations of ischemic stroke. On the other hand, ICD is more descriptive of stroke but less so of hypertension. Hypertension in ICD can be derived from different disease classes and therefore, diseases that are unified under the same tree branch in DO or MeSH are split up in ICD. The limited overlap among these resources recommends caution while interpreting the true meaning of diagnoses across different clinical data sources, and mapping the disease annotations between various research studies.

In general, four discernible groups of disease classifications can be identified: (i) clinical such as ICD and SNOMED, (ii) biomedical research such as Online Mendelian Inheritance in Man (OMIM) and MeSH, (iii) rare diseases such as Orphanet, and (iv) consolidation initiatives such as MONDO and DO. In order to quantify the gap between ontologies, an analysis was carried out on the most used ones falling within the above defined groups: ICD-10, OMIM, MeSH, DO, Orphanet, and MONDO. The pairwise mapping ratio for each pair of those classifications was calculated and visualized in a heatmap (Figure 2.2). This ratio is a normalized measure of how many unique terms from ontology A are mapped to counterparts in ontology B. The mappings used were extracted from the EMBL-EBI Ontology Xref Service (OxO; <https://www.ebi.ac.uk/spot/oxo/>). Alas, not all the classifications offer direct mapping to one another, and manual efforts to create those mapping are very costly and prone to human error and biases. The OxO service annotates the mappings by distance, i.e. direct mappings are equivalent to distance 1, and indirect multi-hop mappings through intermediary classifications are annotated with distance 2 or more. It is clear from the heatmap that most classifications suffer huge information loss when mapping to one another. Translating OMIM to ICD-10, for example, causes a loss of 54% of OMIM terms, while the opposite direction causes the vast majority of ICD-10, around 94%, to be lost. This becomes less surprising when considering the purpose of each of the classifications, where ICD-10 includes many terms describing symptoms and apparent clinical observations, while OMIM is only including disease phenotypes with genetic origins.

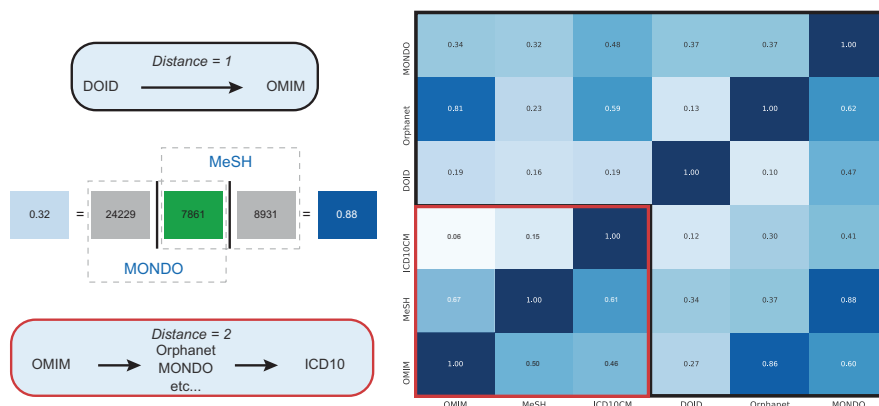


Figure 2.2: Heatmap representing pairwise mappings ratios between several ontologies. The pairwise ontology mappings were calculated between a number of representative disease ontologies (OMIM, MeSH, ICD10CM, DOID, Orphanet, MONDO) based on source data from the EMBL-EBI Ontology Xref Service (OxO). The values were then normalized by the total unique number of terms of each ontology, which serves as a proxy for assessing the coverage of terms of an ontology. The two ratios of each pair mappings should be analyzed together to be able to draw correct interpretations from the figure. The ratios are calculated based on direct mappings between the ontologies (distance = 1), however, for some ontologies (OMIM, MeSH, ICD10CM) there are no direct mappings, therefore mapping with intermediate ontologies is used (distance = 2)

Focusing on ICD-10, MONDO, and Orphanet, as classifications with very distinctive purposes, shows in more detail the source of such disparity (Figure 2.3). Out of 28,454 different disease terms in ICD-10, only 8% and 15% are mappable to Orphanet and MONDO respectively. Parkinson's disease, for example, is represented by a single term in ICD-10, whilst in Orphanet and MONDO, it has 7 and 11 subtypes respectively. MONDO, as an integrative ontology, seems to map exactly to Orphanet rare diseases. The ease of unified description of rare diseases indicates that endophenotyping of common disease towards mechanistically defined rare diseases will be a successful strategy.

To further understand the implications of such inconsistencies, we delineate various shortcomings with problematic mapping examples between ICD-10 and OMIM. In several cases, diseases are found in ICD-10 while not present in the OMIM classification, heart failure

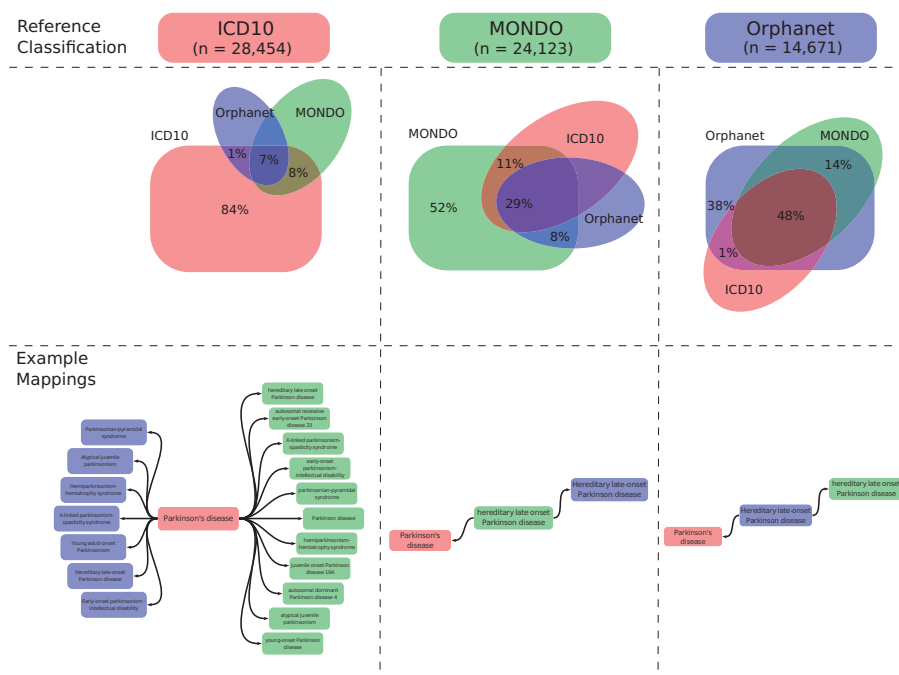


Figure 2.3: Venn diagram representations of intersection between ICD10, MONDO, and Orphanet ontologies. (A) Each of the diagrams shows the percentage of intersecting unique terms, where the intersection is equivalent to terms with mapping between ontologies. In addition, each diagram takes one of the ontologies as a base, i.e. the intersection numbers shown are the terms mapped of the base ontology. (B) Example mappings from the base ontology to the other two. To enable comparison, the chosen examples use related disease terms all describing different phenotypes of Parkinson's disease.

being one of the most prominent examples. Therefore, this results in a loss of information as many diseases cannot be mapped and thus will hamper proper diagnostics. Another cross-mapping limitation arises from the difference between disease description terms in two ontologies. In some instances, the same diseases are named differently in both classifications, e.g. shell teeth (ICD-10, K00.5), while in OMIM is considered dentinogenesis imperfecta, shields type III (125500). In addition, ICD-10 often presents broad descriptions, while OMIM is extremely specific even with distinct disease subtypes. In fact, many disease conditions are associated with only one code in ICD-10 and

multiple separate entries in OMIM. For instance, early-onset cerebellar ataxia, Friedreich ataxia (autosomal recessive), and X-linked recessive spinocerebellar ataxia have only one ICD-10 code (G11.1) while they correlate to 41 OMIM codes.

ICD-10 and OMIM use different classification criteria which can result in misleading cross-mapping. Specifically, OMIM focuses on a mutated gene/locus while ICD-10 classification is according to disease severity or location. In fact, OMIM classifies mental retardation into autosomal recessive (72 subtypes), autosomal dominant (63 subtypes), and X-linked (108 subtypes), while in ICD-10 the same disease is organized based on the severity, i.e. mild, moderate, severe, and profound (F70-F79).

ICD-10 ontology tends to overuse collective terms such as “other” or “unspecified” in its classification which hinders the homogenization process amongst disease ontologies. Surprisingly, many drug-related disorders in OMIM do not exist in the ICD-10. Poor metabolism of efavirenz (614546) exemplifies this scenario together with several rare diseases not present in ICD10. Moreover, several OMIM diseases use the terms “progression of”, “protection against”, or “susceptibility to”, while no such terminology is used in ICD-10. This can also result in wrong/imprecise mapping, or even worse, a complete loss of data. For instance, Psoriasis, susceptibility to (605606) in OMIM might be imprecisely mapped to Psoriasis (L40) in ICD-10, or even not mapped.

Some diseases/syndromes in OMIM can only be mapped to multiple codes in ICD-10. For example, nephrolithiasis/osteoporosis, hypophosphatemic (612287), can be assigned to three different sections in ICD-10: Calculus of kidney (N20.0), osteoporosis in other diseases classified elsewhere (M82.8) and disorders of phosphorus metabolism and phosphatases (E83.3). This problem is also associated with wrong mapping or loss of information. The last problem encountered is that using pre-existing mapping sources, i.e., MONDO, KEGG, Malacards, or Orphanet, has proven to be sometimes wrong, imprecise, or too general. Nijmegen breakage syndrome, for example, is mapped to Immunodeficiency associated with other specified major defects (D82.8) in KEGG. In Orphanet and MONDO, it is assigned to other specified

congenital malformation syndromes, not elsewhere classified (Q87.8). In MalaCards, it is mapped to both: other specified congenital malformation syndromes, not elsewhere classified (Q87.8), and cerebellar ataxia with defective DNA repair (G11.3)

Proper disease coding is crucial for processes where data need to be entered into and shared between clinical registries, genotype-phenotype databases, and biobanks. Registering, combining, and exchanging data is essential for both research and clinical fields [31] and can be hindered by such disease mapping-related problems. However, mapping one disease classification to another, a task frequently encountered in today's data-driven medicine [32] is not at all a trivial task and often leads to inconsistencies in the data. This task will be undertaken, but ideally and as a call to action all biomedical researchers from now on associate their data to one single ontology, even if nowadays it has to still be a descriptive organ- and symptom-based one.

2.6 Towards a mechanism-based taxonomy and new medicine

As more data from clinical research become available, our understanding of pathophysiology improves, making the discrepancy between existing disease definitions even more apparent. Indeed, some of the issues with current disease definitions in clinical trial design were already raised almost a decade ago by [4], where they called for a reform of the human disease taxonomy and asked for the incorporation of more recent scientific advances in molecular and genetic medicine. In continuation of Kola's work, [33] re-called for a movement towards disease definitions based on molecular mechanisms. There is little doubt that a reformed disease taxonomy would lead to more efficient drug discovery, better therapies and prognosis predictions. To this end, PRECISESADS and AETIONOMY, two on-going EU projects, aim to develop new taxonomies for systematic immune diseases and neurodegenerative diseases through data mining and analysis of multi-omics data for large (virtual) cohorts of patients, by discovering mechanistic differences that will allow for better diagnostics and treatments [34–36]. Results from the two projects are expected to reinforce the position of mechanistic disease classification [37–39]. While the need for updating the existing

disease definitions within the clinical annotation ecosystem involving different terminologies is clear, the practical implementation of a refined, interoperable, endotype-based and high coverage classification scheme is easier said than done. We take asthma as an example of how the new classification would materialize. It is clear that the broad term “asthma” fails to cover the complexity of the phenotypes that manifests through different mechanisms across patients. Potentially, a fine-grain definition, say referring to a group of patients with genetic risk factors and who have been smokers during a certain period of their lives, would be “chronic obstructive asthma characterised by (i) over-expansion and/or loss of elasticity of bronchial tubes, (ii) in situ/familial mutations on interleukins, and (iii) high inflammasome activity”. This definition would not only cover the clinical observations that would be identified by pulmonary diagnostic tests or the underlying genetic background of the patient but also the activity of genes within the cellular network involved in a certain endotype such as the inflammasome. Accordingly, one would need to turn to the analysis tools widely used in systems medicine, to check the levels of metabolites or abundance of proteins that belong to the subnetwork of proteins in the human interactome involved in the inflammation pathways in a given patient. Without a doubt, there are key considerations and challenges to the feasibility of such classification in practice. First and foremost, the knowledge on the factors contributing to diseases and their causality is far from complete for most complex disorders. Second, having a fine-grain, underlying cause-based classification of diseases might hinder its adoption and practical use as not only does it involve getting familiarized with new terminology, but also additional tests for identifying the specific causes of the disease in a given patient.

Though some diseases can be explained by well known genetic perturbations in a straightforward manner, pinpointing the underlying cause behind many complex diseases is not trivial. For instance, the human gut microbiota has increasingly been associated with various diseases [40, 41], but still little is known about the order of the causality; is a gut microbiome responsible for the disease or is the disease responsible for the dysbiosis of the microbiome [42, 43]? On top of this, studies have shown that host variants influence the composition

Table 2.2: Example mechanistic disease definitions for several diseases defined traditionally by symptoms rather than underlying etiology

	Disorder	ICD10 definition	Cause of disease	New mechanistic definition
Simple genetic inheritance	Cystic fibrosis	E84 Cystic fibrosis	Genetic predisposition	$\Delta F508$
Complex genetic trait	Type 1 diabetes mellitus	E10 Type 1 diabetes mellitus	Genetic predisposition	Genetic variations in IDDM1/ IDDM2/ GCK
Genetic + environmental influence	Asthma	J45 Asthma	Genetic predisposition boosted by environmental effects	Genetic variations in YKL-40 gene, high pollution levels.
Affected by microbiome	Irritable bowel syndrome	K58 Irritable bowel syndrome	Genetic predisposition and microbiome composition	Genetic variations in NOD2 and FUT2

of the gut microbiome [40], all of which makes finding disease causality more difficult. Additional layers of complexity are added to the tracking of disease mechanisms when individual lifestyles and diets are included as parameters of the search. Lifestyle and diet directly affect the composition of microbiomes in the gut, which again influence physical and even mental health [40, 41, 44]. Gut dysbiosis is linked to a range of neurological disorders, like multiple sclerosis and Parkinson [44, 45], and also mental illnesses including neuropsychiatric conditions such as autism and schizophrenia [41, 45, 46]. Mental disorders are types of multifactorial illnesses whose endogenous disease phenotypes are difficult to determine. As an attempt to ease diagnostics and treatment of diseases relating to mental health, ICD 11 has dedicated special attention to this segment of diseases [47]. Table 2.2 lists some example diseases with their prospective mechanistic definitions.

We, therefore, propose to use genetic, omic and other big data approaches to develop a multi-scale network of all human diseases with unprecedented precision and usability by virtual reality (VR) 3D interrogation. This will allow unravelling different phenotypes that cluster together due to shared gene, symptom, drug, comorbidity associations, elucidate hidden underlying mechanisms that refine our molecular definition of diseases, and capture this in a new ICD-X disease classification system. At a clinical validation level, this will provide

targeted diagnostic and precision therapeutic approaches with a substantially greater chance of achieving treatment success. Our objectives are bold but necessary and achievable.

2.7 Implications of mechanism-based disease classification for drug discovery

In the endeavour of drug repurposing, researchers apply pathway-based or network-based approaches to identify drugs or drug targets [48, 49]. An example where they use a systematic data-driven approach to identify potential drug targets is the work by [50], where they use a PPI network to identify acute ischemic stroke-related network modules and find potential drug targets as well as candidate drugs. When utilising these approaches, scientists implicitly put aside symptom-based disease definitions and try to find suitable drug targets based on relevant mechanistic pathways. In general, the network- and pathway-based strategies might be advantageous when looking for potential drug targets that e.g. are associated with GWAS-associated genes, which themselves may not be ideal drug targets [48, 51]. This point underlines a general relevance of a shift to a mechanism-based disease understanding for treatment discovery. Looking at diseases mechanistically will simplify the discovery of drugs or combinations of drugs that can reset disease-causing perturbations of signalling modules. The coupling of disease definitions based on mechanisms and drug repurposing might not only improve the conditions for many patients worldwide but also help to turn the declining drug-discovery efficacy around [52]. The earliest example and application of a mechanistic redefinition of disease and mechanistic drug approval are pembrolizumab (Keytruda®) and larotrectinib (Vitrakvi®), two cancer drugs based on gene variants rather than and organ-based or histological tumour definition [53, 54].

2.8 Conclusion

2

Despite the recent high-throughput molecular data-driven findings, the medical textbook knowledge on diseases still consists of organ- and symptom-based disease definitions. Accordingly, current disease classification initiatives –the standardised vocabularies of disease terms– fail to reflect a quantitative and evidence-based understanding of pathological mechanisms underlying diseases. Human diseases often involve molecular perturbations of various biological pathways and are substantially heterogeneous at the population level. The recent accumulation of high quality quantitative biomedical data has shed light on the etiology of complex diseases and thus has arisen the need for revising the classification of diseases. Diseases defined mainly as symptoms or phenotypes are hardly mechanistically treatable [55]. The field is moving towards mechanism-based disease definitions. A disease should no longer be addressed and treated as a symptom or phenotype, but as a biological mechanism such as a subgraph of the interactome. This will not only facilitate communication between health professionals but also allow for targeted mechanistic therapies and drug repurposing.

A mechanism-based classification of diseases requires a solid understanding of the current use of various disease classification schemes and clinical annotation systems definitions. We introduced various disease vocabularies and clinical terminologies, went over recent applications of these classifications towards gaining a better understanding of systems medicine and delineated the implications of dispersion among the existing classification efforts. Disease vocabularies and clinical terminologies address an important need in the biomedical domain: communication across peers in an interdisciplinary, interinstitutional and international manner. They provide a formal classification of disease and clinical terms and entail a hierarchy corresponding to the relationships between these terms. Therefore, these taxonomic initiatives offer a standardised vocabulary of disease terms, allowing health professionals, scientists and patients alike to name diseases (as well as their stages and subtypes) in a consistent and concise way through a common language. Yet, existing classification schemes capture a snapshot of the current medical viewpoint and knowledge

on the genetic, environmental and histological perspectives pertinent to diseases. Accordingly, a large variety of disease classification and clinical annotation systems are available in the literature, as well as many disease- and gene-annotation resources using different classifications.

Traditionally bioinformaticians/data analysts do not understand clinical medicine and the biochemistry and pharmacology of disease mechanisms and very rarely validate their hypotheses; in turn, clinicians do not have the holistic view of patient data, have over-specialised and underuse big data to incorporate the commonalities between diseases and heterogeneity across patients. With the recent advances in systems biology and abundance in omics data, the need to redefine non-traumatic, non-communicable diseases into more mechanistic endotypes and identify patient subgroups that exhibit these endotypes has become more apparent than ever. Furthermore, the increasing role of AI-based models in medicine and data-driven analysis for improved diagnosis [56] presents another pressing need for unified and both user and machine-friendly vocabularies that will pave the way for a more precise, interoperable and practical clinical annotation ecosystem, leading to new personalised therapeutic options based on underlying mechanisms.

Appendix A

2 Lost in translation: Classification, taxonomy or ontology?

It is important to distinguish between a 'classification', a 'taxonomy' and an 'ontology', three terms that are often confused and used interchangeably when discussing disease terminology and taxonomy of diseases. Ontologies, taxonomies and classifications differ by the richness of information contained in each. Specifically, a classification labels data, a taxonomy provides information where a concept has a "is a kind of" relationship with a broader term, and an ontology enriches data with information on the relationships of each term with each other using a subclass hierarchy. All provide a structure of concepts, but an ontology provides the user with further information about the concepts and their relationships [57]. In this review, to avoid confusion between terms, we will use the terms disease classification and disease taxonomy inclusively to refer to different means of disease definition and categorisation systems including ontologies, nomenclatures, dictionaries and vocabularies.

To demonstrate the diverse use of classifications and the key differences between them, we briefly introduce three popular disease classification initiatives: ICD, Medical Subject Headings (MeSH [58]) and Disease Ontology (DO [59]). ICD is currently the standard method for diagnostics and inpatient hospital coding in the United States and some European countries and is mandated by the Health Insurance Portability and Accountability Act of 1996 (HIPAA) [60]. The ICD codes help unify the electronic health records (EHR) for reporting to facilitate medical claims nationally and internationally. In addition, it enables data aggregation and specific studies on diseases and drugs [61]. Examples of EHR systems using ICD are: Vanderbilt University Medical Center Starpanel and EDW systems [62], the Stanford University Medical Center STRIDE CDW CDW [63], and the Danish HER system [64]. While ICD definitions are made easy and intuitive for clinical use, it lacks specificity induced by variation of code assignment and biases of methods of the healthcare facilities, which might hinder later data extraction and analysis [65–68].

The MeSH vocabulary, on the other hand, was created to index and annotate medical literature. MeSH terms are standardized terms that refer to defined concepts with a specific hierarchy. These terms are frequently used in systematic reviews, information retrieval, and biomedical literature mining fields. For instance, popular resources such as MEDLINE [69] and ClinicalTrials.gov [70] employ MeSH terms to annotate signs, symptoms, diagnostic procedures, diseases, and drugs in publications and clinical trials. [71–73]. Some drawbacks to MeSH vocabulary are the high cost for maintenance and update, and possible inadequacies in coverage. It takes on average 2 – 3 months to annotate new PubMed articles, and the annotation process is ultimately subjective and can introduce inconsistencies [74, 75].

In contrast to broad medical vocabularies such as MeSH, Disease Ontology (DO) was originally developed to focus only on disease concepts. It aims to harmonise disease concepts from various taxonomies under one umbrella, to enable unified disease annotation and systematic analysis of diseases. Consequently, DO has attracted interest in the areas of data integration and annotation, disease mapping, and computational analysis of diseases and their associations [76–79]. Although DO embarks on encapsulating “a comprehensive theory of disease” [2], it still largely relies on non-mechanistic disease definitions based on symptomatic or unknown causes, such as essential hypertension (DOID:10825), basal ganglia calcification (DOID:0060230), or intrinsic asthma (DOID:9360).

Appendix B

Various software platforms and clinical annotation systems used in practice

The categorization of manifestation of diseases across individuals and related clinical and phenotypic observations from patients (such as symptoms, exam results and laboratory tests) only makes sense in the context of the infrastructure supporting them. Disease classification is an integral part of clinical annotation, monitoring, decision, and management systems and software platforms used at health-care institutions. Therefore, it is essential to understand how differ-

ent clinical management software such as EPIC (www.epic.com) and FHIR (www.hl7.org/fhir/) are used, how medical terminologies such as ICD, SNOMED-CT, CHOP, MedDRA, and LOINC affect them and vice versa.

Medical vocabularies play an important role in how diseases are currently defined in clinical practice. Yet, many medical terminologies are constructed taking practical reasons into account, such as providing a reference for medical billing. Few of these clinical annotation and management systems are targeted towards the correct assignment of observed patient conditions and the medical decisions made within a clinic or a hospital. Such systems aim to simplify and standardise the medical claims process by utilising different clinical annotations, including disease classification, to encode diagnoses, treatments, and other clinical activities. While ICD10 is employed for encoding the diagnosis of diseases, the Swiss Classification of Operations (CHOP) records examinations and treatments. Treatments are arranged in chapters corresponding to the anatomical organ systems (Information Magazine of the Swiss Federal Statistical Office 1/2012, [80]). Similar to other classification efforts, CHOP follows a tree structure further subdividing the chapters into categories, sub-categories, and finally independent procedures. Despite the benefits associated with having a separate encoding for examinations and treatments, CHOP still suffers from many of the same shortcomings as current disease classification systems. Examples include hierarchical organisation following anatomical organ systems and definitions of procedures that address a symptom rather than a mechanism.

SNOMED-CT is a terminology designed to capture all concepts related to the medical field. The concepts are coded to be machine-readable, however, they are attached to a human-understandable meaning as well. SNOMED-CT concepts include clinical findings, symptoms, diagnoses, procedures, procedures, body structures, organisms and other etiologies, substances, pharmaceuticals, devices and specimens. It also contains relationships between concepts, which are represented as concepts themselves, enabling analysis and knowledge inference. Therefore, it is not a surprise that SNOMED-CT

terminology has a variety of uses in the medical domain such as: Electronics Health Records systems, diagnostics, clinical decision support systems, laboratory reporting, genetic databases, and many more.

Diagnosing diseases in clinics often involves testing in medical laboratories. LOINC (Logical Observation Identifiers Names and Codes) was developed in 1994 to provide an electronic standard for identifying laboratory and clinical test results (Official LOINC web page <https://loinc.org/faq/basics/>). The LOINC database stores codes of 3-7 characters indicating tests, measurements and observations and a hierarchical network of the LOINC objects can be calculated using RELMA (Regenstrief LOINC Mapping Assistant), which is a tool developed for searching the LOINC database and mapping local codes to LOINC codes (<https://loinc.org/faq/structure-of-loinc-codes-and-names/>).

Other systems, such as Disease Related Groups (DRG) and Spitalplanungs-Leistungsgruppen (SPLG), focus on creating high-level versions of the above systems, to enable the grouping of patients for control and budgeting of hospitals and medical services. DRG is a meta classification of ICD-9. It is currently used worldwide to pay hospitals for medical services provided to patients. It is seen as a middle ground between paying hospitals a fee per service, which encourages unnecessary and extra services that patients might not need, and a global fixed budget, which focuses on limiting costs at the expense of patients' health. While the motivation behind DRG is mainly concerned with financial aspects, SPLG is oriented more towards medical criteria. These criteria include: that the groups form medically meaningful units, the compatibility with hospital organization, clarity and conciseness, and grouping of similarly complex services. SPLG groups are based on ICD-10 for diagnostic codes and CHOP for treatment codes. The result is around 145 service groups, which is a manageable number for daily care in contrast to thousands of codes in either ICD or CHOP.

Due to the existence of a large number of clinical systems with different purposes, the medical field has put some efforts

to try to collect and standardize these systems and allow them to work together. Among such efforts is the Health Level Seven International (HL7) Fast Healthcare Interoperability Resources (FHIR) framework. The idea to standardize health care data and promote straightforward exchange, integration and retrieval of electronic health information can be dated back to 1987 where HL7 was founded. FHIR is intended to be usable worldwide in a wide variety of contexts including clinical, healthcare-related administrative, public health and research data (Online HL7 documentation, <https://www.hl7.org/fhir/overview-clinical.html>). The target user-groups of FHIR are individuals and organizations developing software and architecting interoperable solutions. This means that FHIR is a standard that third parties can use for tool or app development (Online HL7 documentation, <https://www.hl7.org/fhir/overview-clinical.html>). FHIR makes it possible to use one or more disease ontologies or terminologies, collectively called “code systems”, to describe the health status of patients (Online HL7 documentation, <https://www.hl7.org/fhir/codesystem.html>). It is possible to use international ontologies such as ICD-10 and SNOMED-CT as well as local clinical dictionaries.

Acknowledgements

Funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 777111 (REPO-TRIAL) is gratefully acknowledged. This reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains.

2.9 References

- [1] R. A. Israel. "The history of the International Classification of Diseases". en. In: *Health Bull (Edinb)* 49.1 (Jan. 1991), pp. 62–66.
- [2] *Disease ontology*. <https://disease-ontology.org/>. Accessed: 2021-3-30.
- [3] T. M. Initiative. <http://mondo.monarchinitiative.org/>.
- [4] I. Kola and J. Bell. "A call to reform the taxonomy of human disease". en. In: *Nat. Rev. Drug Discov.* 10.9 (Aug. 2011), pp. 641–642.
- [5] J. Loscalzo. "Personalized cardiovascular medicine and drug development: time for a new paradigm". en. In: *Circulation* 125.4 (Jan. 2012), pp. 638–645.
- [6] N. Nosengo. "Can you teach old drugs new tricks?" en. In: *Nature* 534.7607 (June 2016), pp. 314–316.
- [7] D. S. Echt et al. *Mortality and Morbidity in Patients Receiving Encainide, Flecainide, or Placebo*. 1991.
- [8] "Major cardiovascular events in hypertensive patients randomized to doxazosin vs chlorthalidone: the antihypertensive and lipid-lowering treatment to prevent heart attack trial (ALLHAT). ALLHAT Collaborative Research Group". en. In: *JAMA* 283.15 (Apr. 2000), pp. 1967–1975.
- [9] A. I. Casas et al. "From single drug targets to synergistic network pharmacology in ischemic stroke". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 116.14 (Apr. 2019), pp. 7129–7136.
- [10] J. W. Scannell et al. "Diagnosing the decline in pharmaceutical R&D efficiency". en. In: *Nat. Rev. Drug Discov.* 11.3 (Mar. 2012), pp. 191–200.
- [11] I. Agache et al. "Untangling asthma phenotypes and endotypes". en. In: *Allergy* 67.7 (July 2012), pp. 835–846.
- [12] P. Tomassen et al. "Inflammatory endotypes of chronic rhinosinusitis based on cluster analysis of biomarkers". en. In: *J. Allergy Clin. Immunol.* 137.5 (May 2016), 1449–1456.e4.

-
- [13] S. D. Ghiassian et al. "Endophenotype Network Models: Common Core of Complex Diseases". en. In: *Sci. Rep.* 6 (June 2016), p. 27414.
- [14] M. Hinchcliff and J. M. Mahoney. "Towards a new classification of systemic sclerosis". en. In: *Nat. Rev. Rheumatol.* 15.8 (Aug. 2019), pp. 456–457.
- [15] F. Prinz, T. Schlange, and K. Asadullah. "Believe it or not: how much can we rely on published data on potential drug targets?" en. In: *Nat. Rev. Drug Discov.* 10.9 (Aug. 2011), p. 712.
- [16] C. Kleinschnitz et al. "NOS knockout or inhibition but not disrupting PSD-95-NOS interaction protect against ischemic brain damage". en. In: *J. Cereb. Blood Flow Metab.* 36.9 (Sept. 2016), pp. 1508–1512.
- [17] G. Llovera et al. "Results of a preclinical randomized controlled multicenter trial (pRCT): Anti-CD49d treatment for acute brain ischemia". en. In: *Sci. Transl. Med.* 7.299 (Aug. 2015), 299ra121.
- [18] J. Menche et al. "Disease networks. Uncovering disease-disease relationships through the incomplete interactome". en. In: *Science* 347.6224 (Feb. 2015), p. 1257601.
- [19] P. Vallance. "An audience with Patrick Vallance". en. In: *Nat. Rev. Drug Discov.* 9.11 (Nov. 2010), p. 834.
- [20] K. W. Kim and C. Ober. "Lessons Learned From GWAS of Asthma". In: *Allergy, Asthma Immunology Research* 11.2 (2018), 170–187. DOI: 10.4168/aaair.2019.11.2.170.
- [21] T. G. Guney et al. "Characterisation of a novel 3D model of the airways". en. In: *Eur. Respir. J.* 46.suppl 59 (Sept. 2015).
- [22] M. V. Aun et al. "Animal models of asthma: utility and limitations". en. In: *J. Asthma Allergy* 10 (Nov. 2017), pp. 293–301.
- [23] I. Agache and C. A. Akdis. "Precision medicine and phenotypes, endotypes, genotypes, regiotypes, and theratypes of allergic diseases". en. In: *J. Clin. Invest.* 129.4 (Mar. 2019), pp. 1493–1503.
- [24] W. C. Moore et al. "Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program". en. In: *Am. J. Respir. Crit. Care Med.* 181.4 (Feb. 2010), pp. 315–323.

- [25] A. Custovic, J. Henderson, and A. Simpson. “Does understanding endotypes translate to better asthma management options for all?” en. In: *J. Allergy Clin. Immunol.* 144.1 (July 2019), pp. 25–33.
- [26] L. Colas, D. Hassoun, and A. Magnan. “Needs for Systems Approaches to Better Treat Individuals With Severe Asthma: Predicting Phenotypes and Responses to Treatments”. en. In: *Front. Med.* 7 (Mar. 2020), p. 98.
- [27] F. Demenais et al. “Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks”. en. In: *Nat. Genet.* 50.1 (Jan. 2018), pp. 42–53.
- [28] M. A. R. Ferreira et al. “Genetic Architectures of Childhood and Adult-Onset Asthma Are Partly Distinct”. en. In: *Am. J. Hum. Genet.* 104.4 (Apr. 2019), pp. 665–684.
- [29] T. Bennike et al. “Biomarkers in inflammatory bowel diseases: current status and proteomics identification strategies”. en. In: *World J. Gastroenterol.* 20.12 (Mar. 2014), pp. 3231–3244.
- [30] M. Guindi and R. H. Riddell. “Indeterminate colitis”. en. In: *J. Clin. Pathol.* 57.12 (Dec. 2004), pp. 1233–1244.
- [31] A. Sollie et al. “A New Coding System for Metabolic Disorders Demonstrates Gaps in the International Disease Classifications ICD-10 and SNOMED-CT, Which Can Be Barriers to Genotype–Phenotype Data Sharing”. In: *Hum. Mutat.* 34.7 (2013), pp. 967–973.
- [32] M. Lehne et al. “Why digital medicine depends on interoperability”. en. In: *NPJ Digit Med* 2 (Aug. 2019), p. 79.
- [33] M. Hofmann-Apitius et al. “Towards the taxonomy of human disease”. en. In: *Nat. Rev. Drug Discov.* 14.2 (Feb. 2015), pp. 75–76.
- [34] M. Teruel, C. Chamberlain, and M. E. Alarcón-Riquelme. “Omics studies: their use in diagnosis and reclassification of SLE and other systemic autoimmune diseases”. en. In: *Rheumatology* 56.suppl_1 (Apr. 2017), pp. i78–i87.

-
- [35] T. Charlon et al. "Single Nucleotide Polymorphism Clustering in Systemic Autoimmune Diseases". en. In: *PLoS One* 11.8 (Aug. 2016), e0160270.
 - [36] M. Naz, E. Younesi, and M. Hofmann-Apitius. "Systematic analysis of GWAS data reveals genomic hotspots for shared mechanisms between neurodegenerative diseases". In: *J. Alzheimers Dis. Parkinsonism* 7.5 (2017).
 - [37] E. Younesi et al. "PDON: Parkinson's disease ontology for representation and modeling of the Parkinson's disease knowledge domain". en. In: *Theor. Biol. Med. Model.* 12 (Sept. 2015), p. 20.
 - [38] D. Domingo-Fernández et al. "Multimodal mechanistic signatures for neurodegenerative diseases (NeuroMMSig): a web server for mechanism enrichment". en. In: *Bioinformatics* 33.22 (Nov. 2017), pp. 3679–3681.
 - [39] D. Domingo-Fernández et al. "PathMe: merging and exploring mechanistic pathway knowledge". en. In: *BMC Bioinformatics* 20.1 (May 2019), p. 243.
 - [40] A. B. Hall, A. C. Tolonen, and R. J. Xavier. "Human genetic variation and the gut microbiome in disease". en. In: *Nat. Rev. Genet.* 18.11 (Nov. 2017), pp. 690–699.
 - [41] J. A. Griffiths and S. K. Mazmanian. "Emerging evidence linking the gut microbiome to neurologic disorders". en. In: *Genome Med.* 10.1 (Dec. 2018), p. 98.
 - [42] M. J. B. Ilseung Cho. "The Human Microbiome: at the interface of health and disease". en. In: *Nat. Rev. Genet.* 13.4 (), p. 260.
 - [43] A. B. Hill. "The Environment and Disease: Association or Causation?" en. In: *Proc. R. Soc. Med.* 58.5 (May 1965), p. 295.
 - [44] H. Tremlett et al. "The gut microbiome in human neurological disease: A review". en. In: *Ann. Neurol.* 81.3 (Mar. 2017), pp. 369–382.
 - [45] W. H. Moos et al. "Microbiota and Neurological Disorders: A Gut Feeling". en. In: *Biores. Open Access* 5.1 (2016), p. 137.

- [46] G. Fond et al. "The "psychomicrobiotic": Targeting microbiota in major psychiatric disorders: A systematic review". In: *Pathol. Biol.* 63.1 (Feb. 2015).
- [47] The Lancet. "ICD-11". In: *Lancet* 393.10188 (June 2019).
- [48] S. Pushpakom et al. "Drug repurposing: progress, challenges and recommendations". en. In: *Nat. Rev. Drug Discov.* 18.1 (Oct. 2018), pp. 41–58.
- [49] T. T. Ashburn and K. B. Thor. "Drug repositioning: identifying and developing new uses for existing drugs". en. In: *Nat. Rev. Drug Discov.* 3.8 (Aug. 2004), pp. 673–683.
- [50] Y. Wang et al. "Network-Based Approach to Identify Potential Targets and Drugs that Promote Neuroprotection and Neurorepair in Acute Ischemic Stroke". en. In: *Sci. Rep.* 7.1 (Jan. 2017), pp. 1–12.
- [51] C. S. Greene and B. F. Voight. "Pathway and network-based strategies to translate genetic discoveries into effective therapies". en. In: *Hum. Mol. Genet.* 25.R2 (June 2016), R94–R98.
- [52] N. J. Schork. *Personalized medicine: Time for one-person trials*. <http://www.nature.com/news/personalized-medicine-time-for-one-person-trials-1.17411>. Accessed: 2021-3-30. Apr. 2015.
- [53] R. F. Sweis and J. J. Luke. "Mechanistic and pharmacologic insights on immune checkpoint inhibitors". en. In: *Pharmacol. Res.* 120 (June 2017), pp. 1–9.
- [54] M. Pollack et al. "Transforming approaches to treating TRK fusion cancer: historical comparison of larotrectinib and histology-specific therapies". en. In: *Curr. Med. Res. Opin.* 37.1 (Jan. 2021), pp. 59–70.
- [55] S. Bakchine. "Treating Alzheimer's Disease : Combine or Fail?" In: *The Journal Of Prevention of Alzheimer's Disease* (2019), 1–3. DOI: 10.14283/jpad.2019.16.
- [56] E. J. Topol. "High-performance medicine: the convergence of human and artificial intelligence". en. In: *Nat. Med.* 25.1 (Jan. 2019), pp. 44–56.

-
- [57] R. Van Rees. "Clarity in the usage of the terms ontology, taxonomy and classification". In: *Civil Engineering -New York then Reston-* 20.432 (July 2008).
 - [58] F. B. Rogers. "Medical subject headings". en. In: *Bull. Med. Libr. Assoc.* 51 (Jan. 1963), pp. 114–116.
 - [59] E. Schriml Lynn M. and Mitraka et al. "Human Disease Ontology 2018 update: classification, content and workflow expansion". en. In: *Nucleic Acids Res.* 47.D1 (Jan. 2019), pp. D955–D962.
 - [60] *Health Information Privacy*. <http://www.hhs.gov/ocr/hipaa/>. Accessed: 2021-3-30. Aug. 2015.
 - [61] S. E. Drösler et al. "International comparability of patient safety indicators in 15 OECD member countries: a methodological approach of adjustment by secondary diagnoses". en. In: *Health Serv. Res.* 47.1 Pt 1 (Feb. 2012), pp. 275–292.
 - [62] I. Danciu et al. "Secondary use of clinical data: the Vanderbilt approach". en. In: *J. Biomed. Inform.* 52 (Dec. 2014), pp. 28–35.
 - [63] H. J. Lowe et al. "STRIDE—An integrated standards-based translational research informatics platform". en. In: *AMIA Annu. Symp. Proc.* 2009 (Nov. 2009), pp. 391–395.
 - [64] K. Bernstein et al. "Modelling and implementing electronic health records in Denmark". en. In: *Int. J. Med. Inform.* 74.2-4 (Mar. 2005), pp. 213–220.
 - [65] B. Khokhar et al. "Systematic review of validated case definitions for diabetes in ICD-9-coded and ICD-10-coded data in adult populations". en. In: *BMJ Open* 6.8 (Aug. 2016), e009952.
 - [66] N. McCormick et al. "Validity of Diagnostic Codes for Acute Stroke in Administrative Databases: A Systematic Review". en. In: *PLoS One* 10.8 (Aug. 2015), e0135834.
 - [67] M. Peng et al. "Coding reliability and agreement of International Classification of Disease, 10th revision (ICD-10) codes in emergency department data". en. In: *Int J Popul Data Sci* 3.1 (July 2018), p. 445.

- [68] A. Y. Xiao et al. "The Use of International Classification of Diseases Codes to Identify Patients with Pancreatitis: A Systematic Review and Meta-analysis of Diagnostic Accuracy Studies". en. In: *Clin. Transl. Gastroenterol.* 9.10 (Oct. 2018), p. 191.
- [69] MEDLINE. https://www.nlm.nih.gov/medline/medline_overview.html. Accessed: 2021-3-30.
- [70] *ClinicalTrials.gov*. <http://clinicaltrials.gov/>. Accessed: 2021-3-30.
- [71] K. Schwarze et al. "Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature". en. In: *Genet. Med.* 20.10 (Oct. 2018), pp. 1122–1130.
- [72] X. Zhou et al. "Human symptoms-disease network". en. In: *Nat. Commun.* 5 (June 2014), p. 4212.
- [73] S. Kim, L. Yeganova, and W. J. Wilbur. "Meshable: searching PubMed abstracts by utilizing MeSH and MeSH-derived topical terms". en. In: *Bioinformatics* 32.19 (Oct. 2016), pp. 3044–3046.
- [74] M. Huang, A. Névél, and Z. Lu. "Recommending MeSH terms for annotating biomedical articles". en. In: *J. Am. Med. Inform. Assoc.* 18.5 (Sept. 2011), pp. 660–667.
- [75] M. E. Funk and C. A. Reid. "Indexing consistency in MEDLINE". en. In: *Bull. Med. Libr. Assoc.* 71.2 (Apr. 1983), pp. 176–183.
- [76] D. Croft et al. "The Reactome pathway knowledgebase". en. In: *Nucleic Acids Res.* 42.Database issue (Jan. 2014), pp. D472–7.
- [77] S. Köhler et al. "The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data". en. In: *Nucleic Acids Res.* 42.Database issue (Jan. 2014), pp. D966–74.
- [78] M. Žitnik et al. "Discovering disease-disease associations by fusing systems-level molecular data". en. In: *Sci. Rep.* 3 (Nov. 2013), p. 3202.
- [79] L. Cheng et al. "DisSim: an online system for exploring significant similar diseases and exhibiting potential therapeutic drugs". en. In: *Sci. Rep.* 6 (July 2016), p. 30024.

-
- [80] Federal Statistical Office. *Schweizerische Operationsklassifikation (CHOP)*. <https://www.bfs.admin.ch/bfs/en/home/statistics/catalogues-databases/publications.assetdetail.483959.html>. Accessed: 2021-3-30. July 2016.

3

Transcription Factor NRF2 as a Therapeutic Target for Chronic Diseases: A Systems Medicine Approach

Adapted from: A. Cuadrado et al. "Transcription Factor NRF2 as a Therapeutic Target for Chronic Diseases: A Systems Medicine Approach". In: *Pharmacological Reviews* 70.2 (2018). Ed. by M. C. Michel, pp. 348–383. DOI: 10.1124/pr.117.014753.

Abstract

3

Systems medicine has a mechanism-based rather than a symptom- or organ-based approach to disease and identifies therapeutic targets in a non hypothesis driven manner. In this work, we apply this to transcription factor nuclear factor (erythroid-derived 2)-like 2 (NRF2) by cross-validating its position in a protein–protein interaction network (the NRF2 interactome) functionally linked to cytoprotection in low-grade stress, chronic inflammation, metabolic alterations, and reactive oxygen species formation. Multiscale network analysis of these molecular profiles suggests alterations of NRF2 expression and activity as a common mechanism in a subnetwork of diseases (the NRF2 diseasome). This network joins apparently heterogeneous phenotypes such as autoimmune, respiratory, digestive, cardiovascular, metabolic, and neurodegenerative diseases, along with cancer. Importantly, this approach matches and confirms *in silico* several applications for NRF2-modulating drugs validated *in vivo* at different phases of clinical development. Pharmacologically, their profile is as diverse as electrophilic dimethyl fumarate, synthetic triterpenoids like bardoxolone methyl and sulforaphane, protein–protein or DNA–protein interaction inhibitors, and even registered drugs such as metformin and statins, which activate NRF2 and may be repurposed for indications within the NRF2 cluster of disease phenotypes. Thus, NRF2 represents one of the first targets fully embraced by classic and systems medicine approaches to facilitate both drug development and drug repurposing by focusing on a set of disease phenotypes that appear to be mechanistically linked. The resulting NRF2 drugome may therefore rapidly advance several surprising clinical options for this subset of chronic diseases.

3.1 Introduction

Lifespan has almost doubled in the last century, and aging-specific diseases are now becoming prevalent. However, the pathologic mechanisms underlying most of them are poorly understood and treated rather by correcting symptoms or risk factors. Moreover, contrary to a hitherto linear approach that considered one disease, one medicine, chronic diseases demonstrate a high degree of connectedness and a need for more precise, mechanism-based disease definitions rather than the current organ- and symptom-based. After the human genome sequencing and the development of molecular networks, a new concept of disease is thus emerging, in which diseases are diagnosed not only by clinical symptoms, but mainly by the underlying molecular signatures [1]. The fact that different pathophenotypes have a shared molecular mechanism provides also a rationale toward a new concept of therapy summarized as “several diseases, one medicine” and drug repurposing. Network medicine, i.e., the application of network concepts to the analysis of dynamic connections among diseases and drugs, provides a new opportunity to develop this new approach. Chronic diseases in the elderly are most likely characterized by the loss of homeostasis during aging or as a result of environmental factors, all of them leading to low-grade stress by pathologic formation of reactive oxygen species (ROS), chronic inflammation, and metabolic unbalance. Based on a network medicine approach, in this review we will present extensive evidence indicating that the nuclear factor (erythroid-derived 2)–like 2 (NRF2), as the master regulator of multiple cytoprotective responses and a key molecular node within a particular cluster of diseases, provides a new strategy for drug development and repurposing.

3.2 From Nuclear Factor (Erythroid-Derived 2)–Like 2 Interactome to Nuclear Factor (Erythroid-Derived 2)–Like 2 Diseaseome

3.2.1 A. Nuclear Factor (Erythroid-Derived 2)–Like 2 as a Master Regulator of Cellular Homeostasis

NRF2 is a basic region-leucine zipper (bZip) transcription factor (Fig. 3.1) that forms heterodimers with small musculoaponeurotic fibrosarcoma protein (MAF) K, G, and F in the nucleus. The heterodimer recognizes an enhancer sequence termed antioxidant response element (ARE) that is present in the regulatory regions of over 250 genes (ARE genes) [2, 3]. These ARE genes encode a network of cooperating enzymes involved in phase I, II, and III biotransformation reactions and antioxidant mechanisms that generate NADPH, glutathione (GSH), and thioredoxin reactions; lipid and iron catabolism; and interaction with other transcription factors, etc. [3]. Recently, NRF2 was also found to regulate the expression of several proteasome subunits and autophagy genes, providing additional interest for its control of proteostasis [4–7].

The great significance of NRF2 from a clinical perspective is that it might be targeted pharmacologically with patient benefit. The main mechanism regulating the transcriptional activity of NRF2 is the control of protein stabilization by the E3 ligase adapter Kelch-like erythroid cell–derived protein with Cap’n’collar homology (ECH)-associated protein 1 (KEAP1) (Fig. 3.2). KEAP1 is a homodimeric protein that bridges NRF2 with the E3 ligase complex formed by Cullin 3 and RING-box protein 1 (CUL3/RBX1). Under homeostatic conditions, the N-terminal domain of the KEAP1 homodimer binds one molecule of NRF2 at two amino acid sequences with low (aspartate, leucine, and glycine; DLG) and high (glutamate, threonine, glycine, and glutamate; ETGE) affinity, and hence presents NRF2 to ubiquitination by CUL3/RBX1 [8] and subsequent degradation by the proteasome. KEAP1 is a redox and electrophile sensor that upon modification of critical cysteines loses its ability to repress NRF2 (Fig. 3.2; Biomarkers as Nuclear Factor (Erythroid-Derived 2)–Like 2 Signature and for Monitoring Target Engagement). An alternative

mechanism of regulation of NRF2 stability is the phosphorylation mediated by glycogen synthase kinase 3 (GSK-3) (Fig. 3.2). This kinase phosphorylates a domain of NRF2 (aspartate, serine, glycine, isoleucine, serine;DSGIS) and hence creates a recognition motif for the E3 ligase adapter b-transducin repeat containing E3 ubiquitin protein ligase (b-TrCP) that presents NRF2 to a CUL1/RBX1 complex, leading to an alternative pathway for ubiquitin-dependent proteasome degradation of NRF2. Therefore, KEAP1 and GSK-3/ b-TrCP tightly control NRF2 protein levels in the context of redox homeostasis and cell signaling, respectively [9]. Other mechanisms of NRF2 regulation at protein, mRNA, or gene level have been reported [3], but at least these two are amenable to pharmacological regulation.

3.2.2 Positioning Nuclear Factor (Erythroid-Derived 2)– Like 2 and Its Regulatory Pathway in the Human Interactome and Diseasome

The gene encoding NRF2, termed NFE2L2, is highly polymorphic and presents amutagenic frequency of 1 per every 72 bp. An excellent review on this topic reported in 2015 up to 18 single-nucleotide polymorphisms (SNPs), most of them in the 59 regulatory region and in intron 1 [13]. Several of these SNPs might constitute functional haplotypes that are associated with risk at onset or progression of chronic diseases. Variations in functional haplotypes may have a subtle impact on a proportion of individual swho exhibit clinical symptoms of specific diseases, yet they may have a profound effect at the population level and may define specific strategies to target this gene in precision medicine.

Recent advances in network medicine have provided quantitative tools to characterize how the interplay between genes and their interactions (interactome) is related to pathology [14–16], how dysregulated molecular networks are common to various diseases [17], and how diseases manifest in particular tissues [18]. To understand the relevance of NRF2 in pathology from the systems medicine perspective, first we have generated the human interactome map. We have integrated and curated information on physical interactions among proteins involved in the

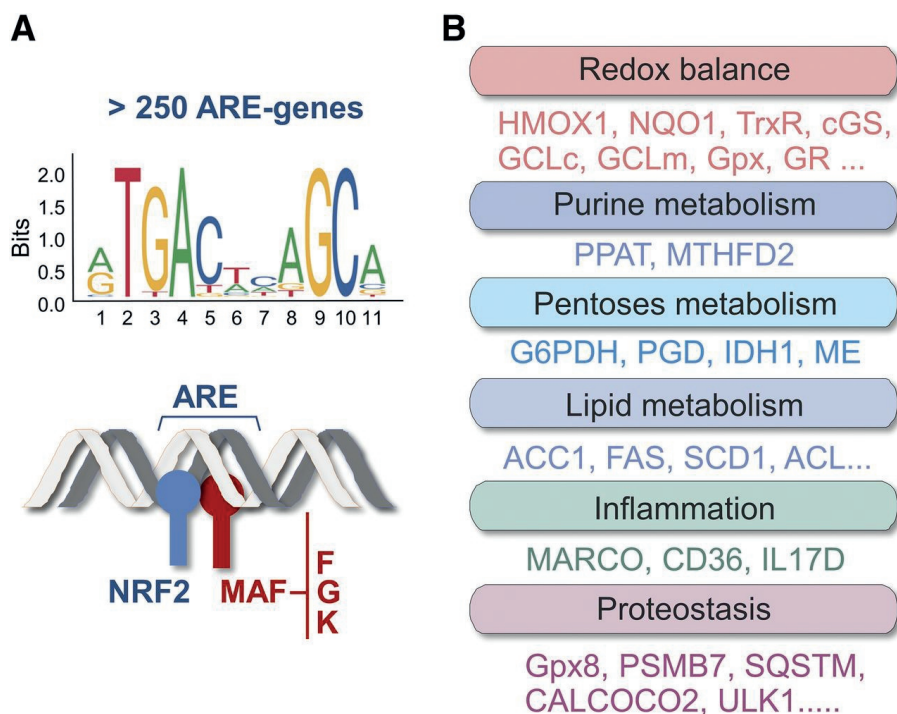


Figure 3.1: NRF2 as a master regulator of cytoprotective responses. (A) NRF2 heterodimerizes with the members of MAF through their bZip domain. The heterodimer binds to an enhancer sequence termed ARE that is present in the regulatory regions of over 250 genes (ARE genes). (B) These genes participate in the control of redox metabolism, inflammation, and proteostasis balance, as indicated. The existence of susceptibility SNPs in NFE2L2, elevated levels of its target genes in brain necropsies, and positive data from preclinical studies suggests that the imbalance in proteostasis, redox, and inflammatory control may be counterbalanced by NRF2 activation. ACl, ATP citrate lyase; ACC1, acetyl-coenzyme A carboxylase 1; CALCOCO2, calcium binding and coiled-coil domain 2; cGS, c-glutamate cysteine synthetase; FAS, fatty acid synthase; G6PDH, glucose-6-phosphate dehydrogenase; Gpx, glutathione peroxidase; Gpx8, glutathione peroxidase 8; GR, glutathione reductase; HMOX1, heme oxygenase-1; IDH1, isocitrate dehydrogenase 1; ME, malic enzyme; MTHFD2, methylenetetrahydrofolate dehydrogenase 2; PGD, phosphogluconate dehydrogenase; PPAT, phosphoribosyl pyrophosphate amidotransferase; PSMB7, proteasome subunit b type-7; SCD1, stearoyl-CoA desaturase; TrxR, thioredoxin reductase; ULK1, unc-51 like autophagy activating kinase 1.

NRF2-regulating pathway [3, 9]. The interaction data have been taken

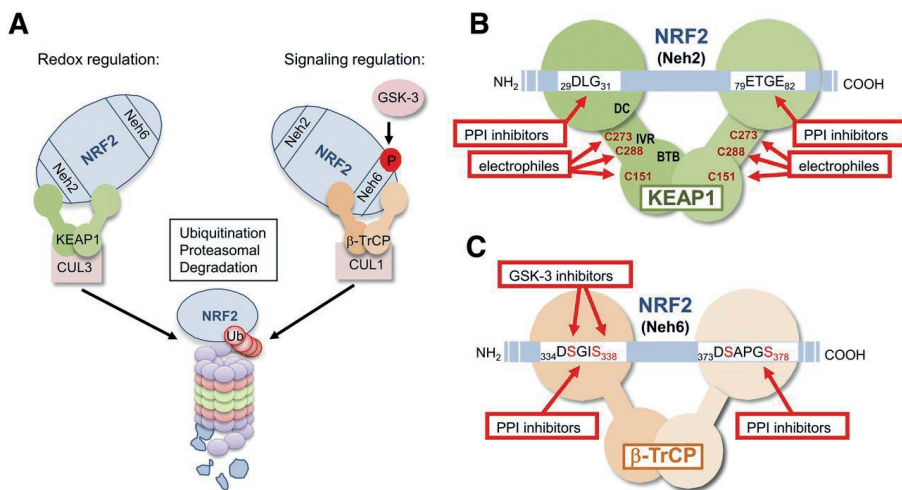


Figure 3.2: Regulation of NRF2 stability by KEAP1 and b-TrCP and its pharmacological targeting. (A) According to the dual regulation model [10], two domains of NRF2, termed Neh2 and Neh6, participate in NRF2 degradation in response to redox and electrophile changes (KEAP1) and to signaling kinases (b-TrCP), respectively. The Neh2 domain binds the E3 ligase adapter KEAP1 that presents NRF2 for ubiquitination to a CUL3/RBX1 complex. The Neh6 domain requires previous phosphorylation by GSK-3 to bind the E3 ligase adapter b-TrCP and subsequent ubiquitination by a CUL1/RBX1 complex (see text for details). (B) Detail of the binding between NRF2 and KEAP1 and current strategies to target this interaction. The KEAP1 homodimer binds NRF2 at two motifs of the Neh2 domain: the low-affinity (29-DLG-31) and the high-affinity (79-ETGE-82) binding sites. Current strategies to disrupt this interaction include the following: electrophiles that alter sulfhydryl groups of cysteines C151, 273, and 288; PPI inhibitors that alter the docking of NRF2 to the DC domain of KEAP1. (C) Hypothetical binding of NRF2 and b-TrCP and suggested strategies to target this interaction. The b-TrCP homodimer binds to the Nhe6 domain at the phospho-motif 334-DSGIS-338 when it is phosphorylated by GSK-3 [10, 11] and at the phospho-motif 373-DSAPGS-378 independently of GSK-3 [12]. In this figure we postulate that, by analogy with KEAP1, one b-TrCP homodimer interacts with one molecule of NRF2 at the two phospho-motifs, but experimental evidence is still lacking. Two possible strategies to disrupt this interaction include the use of GSK-3 inhibitors and PPI inhibitors.

from the recently published human interactome that compiles data across several protein–protein interaction (PPI) resources [17, 19]. However, the currently available information for development of the NRF2 interactome is limited by the fact that, because NRF2 is a

very short half-life protein, some meaningful interactions may be undetected. Nevertheless, some well-known proteins that physically interact with NRF2 are found in the interactome, including KEAP1, b-TrCP, and MAFs. Another group of NRF2-interacting proteins corresponds to nuclear proteins with functions in regulation of gene expression. These include proteins related with bZip transcription factors, nuclear receptors or coactivators, or proteins involved in histone acetylation. Therefore, the NRF2 interactome evidences additional mechanisms of gene regulation beyond those directly connected to this transcription factor. NRF2 is phosphorylated at several residues, and therefore it is expected to interact with several kinases that include GSK-3 and several protein kinase C isoforms. These kinases are downstream of some membrane receptors and adapter/scaffold proteins. In addition to this physical interaction, we have identified several biologic functions that are enriched in the NRF2 neighborhood, including metabolic processes, such as biosynthesis of pentose, tetrapyrrole, heme, glucose 6-phosphate, cysteine, GSH, glyceraldehyde-3-phosphate, and NADPH. Most of these regulatory proteins do not interact directly with each other but are connected through proteins acting as mediators (Fig. 3.3; a more detailed description of specific interacting molecules can be found at [20]).

Perturbations of the NRF2 interactome have been reported in several diseases. We have developed a diseasome map by curating a list of 37 NRF2-related diseases based on DisGeNET [21] and GeneCards [22] databases, as well as knowledge from some studies in animal models. We have also retrieved disease–gene associations for these pathologies using DisGeNET, OMIM, and GWAS databases [17] (Table 3.1a). The interactome-based proximity [23] of NRF2 to known disease genes for each of the NRF2-related disease phenotypes is shown in Fig. 3.4. NRF2 is significantly closer to the known disease genes of the digestive system and cancers, such as prostate, liver, and lung neoplasms, compared with randomly selected proteins, highlighting the key functional role of NRF2 in these pathologies. Moreover, NRF2 was found to be proximal to various proteins related to metabolic and cardiovascular diseases, such as diabetes, hyperglycemia, ischemia, middle cerebral artery infarction, and atherosclerosis. Protein interactions of NRF2 also

connect it to genes associated to respiratory disorders such as asthma, pulmonary fibrosis, and pulmonary emphysema, as well as to neurodegenerative conditions such as Alzheimer disease (AD), Parkinson disease (PD), and amyotrophic lateral sclerosis (ALS).

The interactome-based proximity therefore offers a perspective on how NRF2 can be linked to various pathologic conditions. The relationships among diseases have been previously summarized as a network, termed *diseasome*, that is connecting them based on genetic [1] and clinical [24, 25] commonalities. In Fig. 3.5, we have defined a network of diseases, the NRF2 *diseasome*, based on shared genes, symptom similarity, and comorbidities. NRF2 appears to connect diseases governed substantially by inflammatory processes, such as acute kidney injury, liver cirrhosis, and atherosclerosis. Furthermore, neurodegenerative diseases such as AD, PD, Huntington's disease, and ALS constitute a cluster consistent with recent studies implicating also NRF2 in neuroinflammatory processes [26–30]. The regulation of NRF2 by kinase-signaling cascades [29] could explain the well-connected cluster of cancers, particularly supporting the involvement of NRF2 in pathologic ROS formation underlining colon [31] and breast [32] tumors.

Table 3.1: Cluster of diseases with evidence of NRF2 association

Disease phenotypes with evidence of genetic association with NRF2 were selected from the DisGeNet database. DisGeNet integrates disease– gene association information from various resources, such as UniProt, ClinVar, GWAS Catalog, and Comparative Toxicogenomics Database and scores disease–gene associations according to the number of resources and publications supporting these associations. The curation of the list was based on the following criteria: 1) only pathophenotypes with more than one citation in Pubmed were selected; 2) the score of reliability was set to a threshold of 0.001; and 3) disease entries with very similar names or overlapping terms were simplified to one single entry.

Pathophenotype	Reliability Score	Pathophenotype	Reliability Score
Diabetic nephropathy	0.2016	Diabetic cardiomyopathy	0.0803
Liver cirrhosis	0.2005	Middle cerebral artery infraction	0.0800
Nonalcoholic steatohepatitis	0.2005	Breast neoplasms	0.0087
Acute kidney injury	0.2000	Vitiligo	0.0076
Pulmonary fibrosis	0.2000	Atherosclerosis	0.0067
Nonsmall cell lung carcinoma	0.1252	Asthma	0.0043
Squamous cell carcinoma	0.1243	Leukemia	0.0038
Liver neoplasms	0.1238	Colon neoplasm	0.0038
Hyperglycemia	0.1208	Gastrointestinal diseases	0.0029
Drug-induced liver injury	0.1200	Parkinson disease	0.0026
Prostatic neoplasms	0.1200	Systemic lupus erythematosus nephritis	0.0026
Chronic obstructive pulmonary disease	0.0899	Glioma	0.0024
Colorectal neoplasms	0.0847	Amyotrophic lateral sclerosis	0.0022
Alzheimer disease	0.0837	Ischemia	0.0016
Type 2 diabetes mellitus	0.0814	Pulmonary emphysema	0.0013
Chronic kidney disease	0.0808	Pancreatic neoplasms	0.0013
Diabetic retinopathy	0.0805	Vascular diseases	0.0013
Huntington's disease	0.0805	Sepsis	0.0013

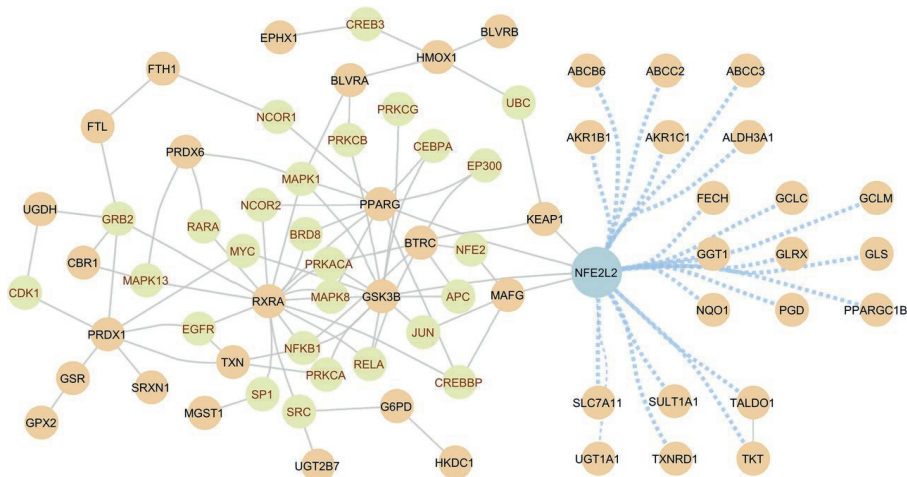


Figure 3.3: Locating the NRF2 regulatory pathway in the human interactome. NRF2 plays a key role in pathologic ROS formation as well as inflammatory and metabolic responses by coordinating the activity of various proteins. These interactions constitute a molecular interaction network, the NRF2 interactome. The known physical interactions between proteins involved in the NRF2 regulatory pathway (i.e., regulator proteins, brown circles) and the proteins through which they are connected (i.e., mediator proteins, green circles) are shown with a gray link. The regulator proteins that involve more than one mediator protein to connect to NRF2 are shown with a blue link. The regulator proteins are curated from literature, and their interactions are retrieved from various resources, including IntAct, MINT, BioGRID, HPRD, KEGG, and PhosphoSite. The list of all the proteins that interact with NRF2 in the human interactome is available online at <http://sbi.imim.es/data/nrf2/>.

3.3 Target Validation of Nuclear Factor (Erythroid-Derived 2)–Like 2 in Human Disease States

3.3.1 Key Role of Nuclear Factor (Erythroid-Derived 2)–Like 2 in Resolution of Inflammation

Persistent inflammation is a hallmark of all pathophenotypes found in the NRF2 diseasome. This is most likely because inflammation is associated with increased local and systemic pathologic formation of reactive oxygen species (ROS). In fact, ROS and reactive nitrogen species (RNS) stimulate and aggravate inflammatory responses that are mechanistically related to the

activation of transcription factor, p65 subunit of nuclear factor κ -light-chain enhancer of activated B cells (NF- κ B) [33]. Very simplified, in resting immune cells, NF- κ B is retained in the cytosol through interaction with the nuclear κ -B inhibitor (I κ Ba). Pathogen-associated molecular pattern molecules derived from microorganisms as well as damage-associated molecular pattern molecules released in response to tissue damage stimulate cognate receptors expressed by immune cells that lead to activation of I κ B kinase (IKK) β . This kinase phosphorylates I κ Ba, targeting it for degradation and allowing nuclear translocation and activation of NF- κ B (Napetschnig and Wu, 2013). These events are submitted to redox control through several modes of regulation of I κ Ba (Bowie and O'Neill, 2000; Morgan and Liu, 2011; Siomek, 2012), but one that has been described recently involves the regulation of IKK β stability by KEAP1. Just like NRF2, IKK β possesses an ETGE motif that enables its binding to KEAP1 for ubiquitination and proteasomal degradation. Therefore, under basal redox conditions, active KEAP1 targets IKK β for degradation and then I κ Ba inhibits NF- κ B. By contrast, in the presence of ROS, KEAP1 is inhibited and IKK β is stabilized, phosphorylating I κ Ba and leading to its degradation and therefore to upregulation of NF- κ B [34].

Because NRF2 is a master regulator of redox homeostasis, it exerts an indirect control on NF- κ B activity. Lipopolysaccharide (LPS) activates simultaneously a fast, proinflammatory NF- κ B response and a slow NRF2 response. The NF- κ B response is subsequently inhibited when NRF2 is maximally active [35]. For instance, Ras-related C3 botulinum toxin substrate 1, a small G protein of the Rho family, activated the NF- κ B pathway and NRF2 overexpression blocked, whereas NRF2 knockdown enhanced NF- κ B-dependent transcription [35]. Consistently, in NRF2-deficient (Nrf2^{-/-}) mice challenged with LPS or tumor necrosis factor (TNF)- α , the activity of IKK was exacerbated and led to increased phosphorylation and degradation of I κ B [36].

NRF2 also induces an anti-inflammatory phenotype that modulates the functions of CD8⁺ T cells [37] as well as in macrophages and microglia ([26], 2014a; [38]). This is because NRF2 increases cysteine and GSH levels in macrophages through regulation

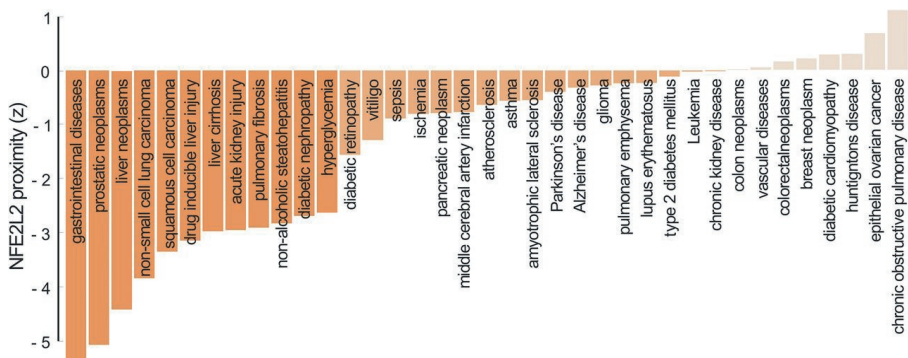


Figure 3.4: Systems medicine view of NRF2 from the perspective of the human interactome. Diseases are triggered by mutations that perturb proteins and their interactions, affecting a certain neighborhood in the NRF2 interactome. Interactome-based proximity measures the distance of genes to those disease neighborhoods. The bars show the proximity of NRF2 gene (NFE2L2) to known disease genes that participate in NRF2-related pathophenotypes. These bars highlight the role of NRF2 in digestive system diseases and cancers. For a given disease, the proximity first calculates the distance from NRF2 to the closest known disease gene and then compares that distance to a random expectation that is estimated using average distances between randomly selected proteins in the interactome. The z-score reported by interactome-based proximity corresponds to the significance of the observed distance between NRF2 and the disease genes. A negative value indicates that the observed distance is lower than what would be expected by chance. The bars are colored in varying tones of orange based on the z-score: significantly proximal (dark orange), proximal (orange), not proximal (light orange), respectively. The known disease genes are taken from DisGeNet, OMIM, and GWAS Catalog.

of the cystine/glutamate transporter and the GSH-synthesizing enzyme g-glutamyl cysteine ligase modulator and catalytic subunits [g-glutamyl cysteine ligase modulator subunit (GCLM) and g-glutamyl cysteine ligase catalytic subunit (GCLC)]. Conversely, GSH depletion sensitizes macrophages to NRF2 activation by LPS [39]. All of these studies point to NRF2 as an anti-inflammatory factor, crucial in controlling the intensity and duration of inflammatory responses (Fig. 3.6).

NRF2 and NF- κ B crosstalk through feed forward and feedback mechanisms (Fig. 3.7). At the transcriptional level, NF- κ B activates NRF2 expression due to the existence of several functional binding

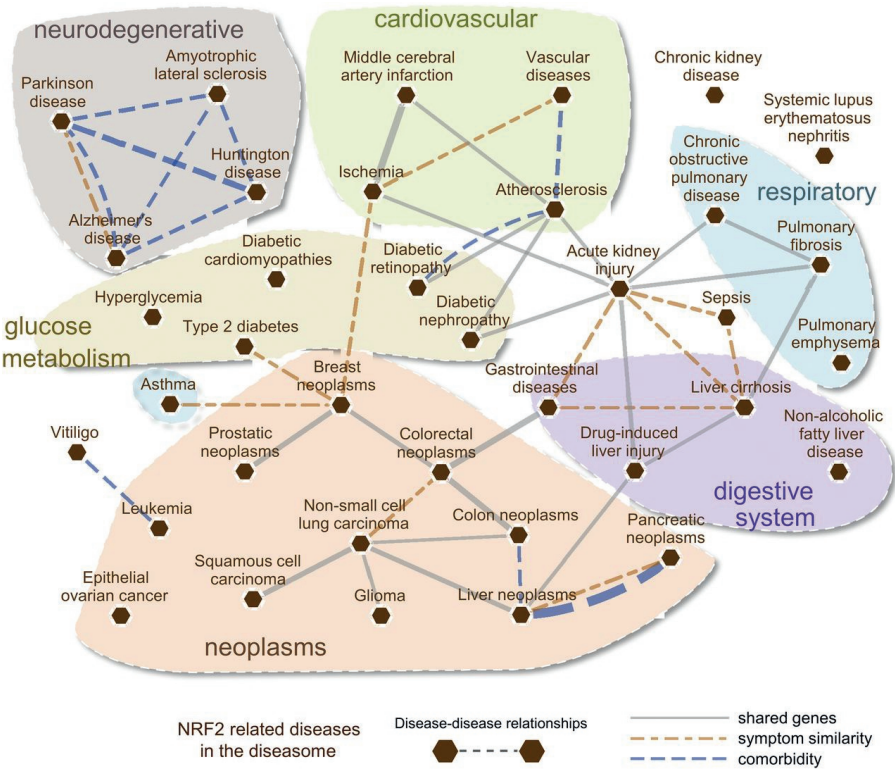


Figure 3.5: Current status of the NRF2 diseasesome. The relationships between diseases are represented as a network in which pathophenotypes are linked by common genetic and clinical descriptors. In the figure, nodes (red hexagons) represent diseases, and the edges are similarities among them based on shared genes, common symptoms, and comorbidities (gray, orange, and blue lines, respectively). The genes and symptoms associated with the diseases are used to identify disease pairs that have a significant genetic and symptomatic overlap calculated using the Jaccard index. Among significant disease–disease connections ($P < 0.05$, assessed by Fisher’s exact test based on the observed gene or symptom overlap), only the links that have an elevated overlap and comorbidity are shown to eliminate potentially spurious connections. Accordingly, the diseases that share at least 10% of the disease-associated genes and more than half of the associated symptoms are included in the figure. The comorbidity information is extracted from medical insurance claims, representing disease pairs that tend to occur together in the population (relative risk .2).

sites in the promoter region of the NFE2L2 gene, thus inducing a negative feedback loop [40]. Moreover, both NF- κ B and NRF2 transcription

factors require the coactivator CBP/p300, which is a histone acetyltransferase that acetylates and increases the DNA-binding capacity. As such, NF- κ B overexpression hampers the availability of CBP/p300 for NRF2, hence reducing its transcriptional capacity, whereas NF- κ B knockdown shows the opposite effect [41]. Additionally, NF- κ B may promote interaction of histone deacetylase-3 with MAF proteins, therefore preventing their dimerization with NRF2 [41]. NF- κ B binds and translocates KEAP1 to the nucleus, thus favoring NRF2 ubiquitination and degradation in this cellular compartment [42]. The E3 ligase adapter b-TrCP tags both I κ Ba [43] and NRF2 [9–11] for proteasomal degradation, and therefore it can lead to increased NF- κ B activity.

The anti-inflammatory activity of NRF2 was thought to rely only on modulation of redox metabolism or crosstalk with NF- κ B. However, NRF2 can also directly block the transcription of the proinflammatory genes interleukin (IL)-6 and IL-1b in macrophages upon exposure to LPS [44]. LPS exposure or pharmacological activation of NRF2 leads to its binding to the proximal promoters of these proinflammatory genes and blocks the recruitment of RNA pol II. The mechanism appears to be independent of the binding of NRF2 to its well-established ARE enhancer. In other studies, NRF2 could directly regulate the expression of several other macrophage-specific genes, such as macrophage receptor with collagenous structure, a receptor required for bacterial phagocytosis, or CD36, a scavenger receptor for oxidized low-density lipoprotein [45, 46]. Similarly, the gene encoding the proinflammatory cytokine, IL-17D, contains AREs, and this NRF2–T helper (Th)17 axis seems to confer protection against tumorigenesis and viral infections [47].

Chronic inflammatory processes involve adhesion of leukocytes to the vascular endothelium and infiltration into the damaged tissue. Both processes appear to be modulated by NRF2 together with at least one of its target genes HMOX1, coding heme oxygenase-1 (HO-1). The NRF2/HO-1 axis inhibits adhesion of inflammatory cells to the endothelium by modulating the expression of several cell adhesion molecules such as vascular cell adhesion molecule 1 (Banning and Brigelius-Flohe, 2005; [48]). Additionally, NRF2/HO-1 inhibits metalloproteinase-9 in macrophages that are necessary for migration

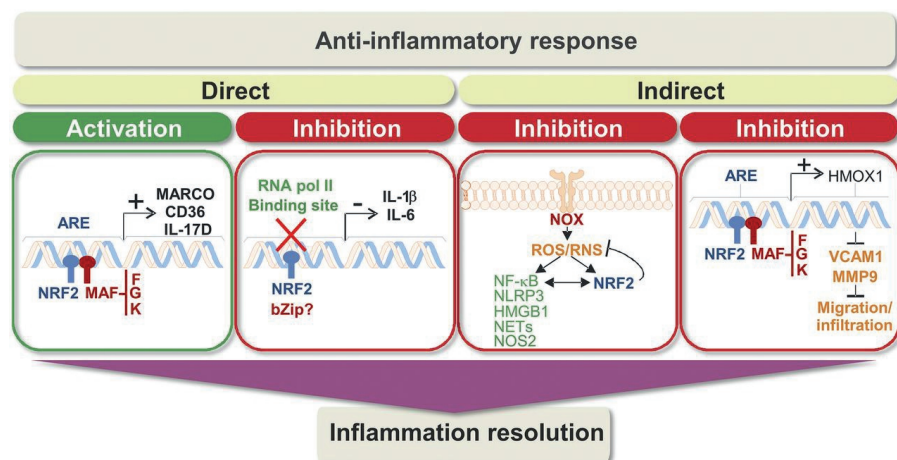


Figure 3.6: Direct and indirect regulation of inflammation by NRF2. Direct mechanisms of action include transcriptional induction of anti-inflammatory genes as well as transcriptional repression of proinflammatory genes. In the second case, the quotation mark indicates that further work is required to identify the bZip partner of NRF2 in this function, if any. Indirect mechanisms to counteract inflammation involve ROS/RNS modulation and inhibition of migration/infiltration of immune cells. Overall, these pathways lead to an anti-inflammatory response that helps to properly resolve inflammation. The existence of polymorphisms in NFE2L2 associated with reduced transcriptional activity, the altered levels of target genes in patients, and promising data from preclinical studies support a relevant role of NRF2 in inflammation resolution.

of immune cells within tissues [49].

Numerous preclinical studies have reported that activation of NRF2 by natural compounds [50] or by disrupting its negative regulator KEAP1 leads to potent anti-inflammatory effects in myeloid leukocytes [51] and macrophages [52]. In observational studies, a polymorphism in NFE2L2 was associated with reduced transcriptional activity correlating with increased risk of inflammatory bowel disease [53] and chronic gastritis [54]. One example of the immune modulatory action of NRF2 is provided in the central nervous system. Injured neurons release fractalkine, a chemokine that specifically activates the phosphatidylinositol 3 kinase/AKT (PI3K/AKT) pathway in microglia, resulting in inhibition of GSK-3 β and upregulation of NRF2 [28]. In this study, necropsies from AD and

progressive supranuclear palsy patients exhibited a compensatory increase in fractalkine levels together with upregulated NRF2 protein, suggesting that this pathway contributes to limiting the inflammatory response in the diseased brain.

3.3.2 Nuclear Factor (Erythroid-Derived 2)–Like 2 in Autoimmune Diseases

At the periphery of the NRF2 diseasome cluster, we found several autoimmune disease phenotypes, such as vitiligo, asthma, multiple sclerosis (MS), and systemic lupus erythematosus (SLE). Indeed, extensive work in animal models of experimental autoimmune encephalomyelitis (EAE) and rheumatoid arthritis (RA), as well as clinical evidence in MS and psoriasis further points to a role of NRF2 in autoimmune diseases. Oxidative tissue damage and apoptosis may increase the generation of autoantigens, leading to activation of T cells and production of autoantibodies by B cells, e.g., as observed for 3-nitrotyrosine–positive proteins [55]. In addition, loss of phase II detoxification enzymes, many of which are transcriptionally regulated by NRF2, results in increased production of reactive intermediates that contribute to formation of haptens or damaged macromolecules that sometimes become immunogenic, increasing hence the pool of autoantigens that trigger autoimmune reactions. Because NRF2-regulated enzymes play a critical role in the detoxification of many chemicals, it is conceivable that NRF2 may be a protective mechanism against the environmental contribution to autoimmune pathogenesis (Ma, 2013). Potential mechanisms for NRF2-mediated regulation of autoimmunity also involve suppression of proinflammatory Th1 and Th17 responses and activation of immunosuppressive T regulatory (Treg) and Th2 ones. There is also increasing evidence that NRF2 may control the differentiation and function of dendritic cells (DCs) and macrophages involved in antigen presentation and regulation of adaptive immune responses. In fact, NRF2 deficiency was shown to alter the function and phenotype of DCs by increasing the expression of costimulatory molecules and consequently the antigen-specific T cell reactivity [56].

MS is a chronic inflammatory disease characterized by

infiltration of autoreactive immune cells into the central nervous system. The absence of NRF2 exacerbates the development of EAE, which is a mouse model of MS [57]. Part of the effects associated with NRF2 deficiency may be related to the reduced levels of HO-1. Thus, mice with a myeloid-specific HO-1 deficiency exhibited higher incidence of lesions, accompanied by activation of antigen-presenting cells and infiltration of inflammatory Th17 and myelin-specific T cells [58]. Knockdown of KEAP1 [44] or treatment with a wide range of small molecules that activate NRF2 [59] inhibited the development and severity of disease. NRF2 is strongly upregulated in active MS lesions, and the expression of NRF2-responsive genes is predominantly found in areas of initial myelin destruction [60]. In MS brains, NRF2 and its targets NADPH:quinone oxidoreductase (NQO1) and HO-1 are mainly expressed in infiltrating macrophages and to a lesser extent in astrocytes, most likely as a compensatory response against pathologic ROS formation. In contrast, there is a lack of NRF2 and antioxidant gene expression in oligodendrocytes, and this may underlie their damage and loss in MS [61]. As a consequence of the altered immune and redox homeostasis, HO-1 expression is reduced in peripheral blood mononuclear cells of MS patients and is downregulated during exacerbation of the disease [62]. Of note, gene expression profiling in interferon- β -treated patients identified NRF2 as a potential mediator of long-term antioxidant response and neuronal preservation [63].

SLE is underlined by high oxidative environment, deregulated cell death, and defects in removal of dead cells, which leads to cell necrosis as source of autoantigens. Female mice deficient in NRF2 develop with age a multiorgan autoimmune disorder similar to SLE, characterized by increased DNA oxidation, lipid peroxidation, splenocyte apoptosis, presence of antibodies against double-strand DNA and the Smith antigen, along with important tissue damage (vasculitis, glomerulonephritis, hepatitis, and myocarditis) [64]. The fact that only female mice show progression to SLE suggests that female-specific factors may contribute to break immune tolerance to self-antigens [64]. NRF2 deficiency also results in enhanced proliferative responses of CD4⁺ T cells, altered CD4⁺/CD8⁺ ratio, and promotion of proinflammatory Th17 in SLE [65, 66]. In fact,

NRF2 depletion has been associated with Th17 differentiation and function during development of lupus nephritis, which seems to be mediated by regulation of the suppressor of cytokine signaling 3/phosphorylated signal transducer and activator of transcription (STAT)3 pathway and IL-1b signaling [66]. In addition, the salivary glands of Nrf2^{-/-} mice show intense lymphocyte infiltration, reminiscent of the Sjögren syndrome, which is often associated with SLE [65]. SLE patients exhibit alterations in repair mechanisms of oxidative DNA damage [67], high serum levels of oxidized proteins, apolipoprotein C3 [68], oxidized phospholipids, and autoantibodies against oxidatively modified lipoproteins [69]. NRF2 polymorphisms have not been associated with SLE susceptibility, although the SNP rs35652124 was related to increased risk of nephritis in childhood onset of Mexican females [70].

Rheumatoid arthritis (RA) is a systemic inflammatory disease with a complex but still elusive autoimmune profile, in which neutrophils, macrophages, and lymphocytes are actively recruited and activated in the inflamed joints. This results in increased secretion of proinflammatory mediators such as ROS/RNS, eicosanoids, cytokines (IL-17, TNF- α , interferon- γ , IL-6, and IL-1b), and catabolic enzymes that trigger hyperproliferation of synovial fibroblasts, joint swelling, and progressive destruction of cartilage and bone [71]. Deletion of the NRF2 gene increases vulnerability to joint alterations in experimental RA models. For instance, in mice expressing the T cell receptorKRN and the major histocompatibility complex class II molecule A(g7) (the K/BNarthritis model), and in antibody-induced arthritis, NRF2 deficiency accelerates the incidence and aggravates the disease course [72, 73]. NRF2 deficiency dramatically upregulates migration of inflammatory cells, expression of cyclooxygenase-2 and inducible nitric oxide synthase, production of ROS and RNS, and release of proinflammatory cytokines and chemokines. Moreover, NRF2 may be a protective factor for bone metabolism in arthritis [72], and NRF2/HO-1 activation exerts anti-inflammatory and antioxidant effects in animal models of RA and in human synovial fibroblasts [73]b. It is interesting that antirheumatic gold(I)-containing compounds that stimulate the antioxidant response through activation of NRF2 and upregulation of HO-1 and GCLC proved

clinical efficacy in RA [44]. Moreover, NRF2/HO-1 activation mediates the induction of synovial cell apoptosis and inhibition of proinflammatory cytokine production by cilostazol [74] as well as the anti-inflammatory effects of H₂S and related compounds, which are able to modify by sulfhydration the cysteine residues of KEAP1 [73]. Other drugs that induce NRF2 and HO-1 signaling, such as rebamipide, can divert the differentiation of human and murine CD4⁺ T cells toward an immunosuppressive Treg phenotype and inhibit the differentiation of TCD4⁺ cells toward inflammatory Th17 cells through specific inhibition of STAT3 [75]. Excessive ROS generation within the inflamed synovium appears to contribute to the pathogenesis of RA because patients show a marked increase in ROS formation, lipid peroxidation, protein oxidation, DNA damage, and decrease in the activity of the antioxidant defense mechanisms, all of these contributing to tissue damage and disease progression [76]. In response to pathologic ROS formation, the NRF2 pathway is activated in synovial cells of RA patients and in joints of antibody-induced arthritic mice, but this response is apparently insufficient to counteract the disease progression [73].

Vitiligo is a skin inflammatory disorder characterized by the accumulation of ROS in the epidermis, which participates in the death of melanocytes. These molecules modify DNA and melanosomal proteins with formation of autoantigens and activation of an autoimmune response against melanocytes [77]. Genetic studies have revealed associations of NRF2 promoter SNPs with susceptibility to develop vitiligo, such as the SNP at 2650 position [78], whereas the C allele of rs35652124 was shown to be associated with protective effects in a Han Chinese population [79]. NRF2 and its downstream detoxification target genes NQO1, GCLC, and GCLM are upregulated in the epidermis of vitiligo patients, suggesting insufficient activation of this defensive mechanism [80].

3.3.3 Nuclear Factor (Erythroid-Derived 2)-Like 2 in Chronic Respiratory Diseases

The relevance of NRF2 in respiratory diseases was reviewed in 2010 (Cho and Kleeberger, 2010), and in this work we will highlight

only the most relevant findings (Fig. 3.8). Cigarette smoke is a main risk factor for chronic obstructive pulmonary disease (COPD). COPD patients have dysfunctional alveolar macrophages that lead to uncontrolled ROS production, proinflammatory mediators, defective phagocytosis, and an array of metalloproteinases that participate in tissue damage. In fact, the emphysematous lung tissue shows a direct correlation between alveolar macrophage density in the parenchyma and severity of lung destruction [81]. Impaired phagocytic activity of alveolar macrophages is a major cause of recurrent bacterial and viral infections that cause acute exacerbations of COPD and are a major source of morbidity and mortality. Nrf2^{-/-} mice exhibit enhanced susceptibility to cigarette smoke-induced emphysema [82]. Importantly, activation of NRF2 with the isothiocyanate sulforaphane (SFN) restores bacteria recognition and phagocytosis, enhances pulmonary bacterial clearance by alveolar macrophages, and reduces inflammation in wild-type mice, but not in Nrf2^{-/-} mice exposed to cigarette smoke [45]. In humans, the transcriptional signature of NRF2 is decreased in alveolar macrophages from patients with smoking-related lung emphysema as compared with smoking and nonsmoking patients without emphysema [83]. Decreased NRF2 expression is associated with increased macrophage expression of the lipid peroxidation product 4-hydroxynonenal. In the NFE2L2 promoter, a functional haplotype constituted by three SNPs and one triplet repeat, polymorphism that produces low to medium NRF2 expression [84], was associated with increased risk to develop COPD [85]. The low-expressing haplotypes were significant predictors for developing respiratory failure. Thus, the 2617A allele of SNP rs6721961 had a significantly higher risk for developing acute lung injury [86].

Pathologic ROS formation may play a role in the pathogenesis of chronic lung fibrosis. An early study demonstrated that pulmonary fibrosis induced by bleomycin is more severe in Nrf2^{-/-} than in wild-type mice [87]. In fact, wild-type mice induced an antioxidant and anti-inflammatory response by upregulating NRF2, and this could not be achieved in the Nrf2^{-/-} mice. Later, it was verified that patients with idiopathic pulmonary fibrosis or chronic sarcoidosis/hypersensitivity pneumonitis exhibit increased expression of NRF2 and aug-

mented levels of low-mol. wt. antioxidants in bronchoalveolar lavage fluids, such as uric acid, ascorbic acid, retinol, and α -tocopherol, suggesting an unsuccessful adaptive response to the ROS challenge [88]. Mechanistically, NRF2 deficiency increases myofibroblast differentiation, whereas pharmacological induction of NRF2 with SFN results in lower number of myofibroblasts and attenuation of the profibrotic effects of transforming growth factor- β (TGF- β) [89].

3.3.4 Nuclear Factor (Erythroid-Derived 2)–Like 2 in the Digestive System

The prominent position of pathophenotypes of the digestive system in the NRF2 diseasome highlights the relevance of the transcriptional signature of NRF2 as a potent adaptive mechanism to chronic oxidative damage and inflammatory stress triggered by xenobiotics. In the gastrointestinal tract (GI), chronic exposure to xenobiotics triggers dysfunctional interactions between the microbiota of the intestinal lumen and the immune system (Aviello and Knaus, 2017). This can lead to chronic diseases of the GI tract, like the inflammatory bowel disease (IBD) phenotypes comprising Crohn's disease and ulcerative colitis, in which there is evidence of activation of a protective NRF2 response. For instance, in colonic tissues from IBD patients, colon epithelial cells responded to inflammatory signals through a NRF2-dependent adaptation that was associated with increased proteasome protein expression [90]. Monocyte-derived macrophages from IBD patients evidenced a specific NRF2-dependent gene expression profile that was exacerbated in response to LPS, further suggesting an attempt to attenuate the inflammatory challenge [91]. At the genetic level, a particular genotype of the NFE2L2 gene (26862684) was associated with the development of ulcerative colitis, especially in females, in a Japanese cohort [92]. In fact, pathologic processes in the GI are highly dependent on the genetic background of the host in relation with dysfunctional interaction between the microbiota of the intestinal lumen and the immune system (Aviello and Knaus, 2017).

One of the symbiotic effects of the microbiota of the GI is to release moderate amounts of ROS that elicit a cytoprotective response mediated by NRF2 in epithelial colonocytes and infiltrating immune cells

[93]. Moreover, cytoprotective molecules that are under the transcriptional control of NRF2 in eukaryotes can also be produced by commensal bacteria. For instance, the HO-1 homologs in the microbiota may greatly contribute to GI homeostasis, and this can be therapeutically exploited for local delivery of carbonmonoxide to the intestine [94].

The proven involvement of NRF2 in maintaining GI homeostasis makes this transcription factor a promising therapeutic target in IBD. Thus, several chemical compounds and dietary supplements might exhibit beneficial effects, like melatonin, 3-(3-pyridylmethylidene)-2-indolinone, butyrate, *Lactobacillus casei*, L-carnitine, 4-vinyl-2,6-dimethoxyphenol (canolol), lacto-wolfberry (formulated product of wolfberries in skimmed milk), etc. [95]. Therefore, it is of utmost importance to define the involvement of NRF2 in chronic and acute diseases of the GI tract for better guidance on the therapeutic approach for modulating the NRF2 pathway.

The liver is also a first line of defense against food xenobiotics. Therefore, it is not surprising that the NRF2 diseasome highlights the relevance of this transcription factor in pathophenotypes associated with liver damage. Early work with the Nrf2^{-/-} mouse model demonstrated its protective effect against acetaminophen-induced hepatocellular injury, benzo[a]pyrene-induced tumor formation, and Fas- and TNF- α -mediated hepatocellular apoptosis (Aleksunes and Manautou, 2007). The higher sensitivity of Nrf2^{-/-} mice to chemical toxicity correlated with reduced basal and inducible expression of detoxification enzymes. In humans, the functional haplotype of three NRF2 promoter SNPs that result in reduced NRF2 expression was significantly associated with development of gastric mucosal inflammation, either independently or by interacting with *Helicobacter pylori* infection [54]. Analysis of the transcriptional signature of NRF2 in patients with primary biliary cholangitis indicated that these patients exhibit reduced NRF2 expression together with low levels of HO-1 and GCLC proteins, and these impairments are more advanced in patients with cirrhosis [96].

Pathologic ROS formation is a key mechanism of hepatocellular injury and disease progression in patients with nonalcoholic steato-

hepatitis (NASH) (Fig. 3.8). This disease evolves in two phases, one of progressive accumulation of fatty acids in hepatocytes and a second that involves liver injury and inflammatory pathologic ROS formation [97]. Accordingly, mice fed with a high-fat diet (HFD) developed a simple steatosis, characterized by increased hepatic fat deposition without inflammation or fibrosis, but *Nrf2*^{2/2} mice presented exacerbated hepatic steatosis and substantial inflammation, consistent with NASH [98]. It is interesting, however, that the hepatocyte-specific KEAP1 deletion, while reducing liver steatosis, did not alter inflammation during development of NASH, suggesting a compensatory mechanism [99]. At least in the rat model of NASH, dietary NRF2 activators attenuate the progression of liver fibrosis [100]. Markers of pathologic ROS formation were increased in liver biopsies of NASH patients, and the NRF2 signature was increased, suggesting an attempt to reduce the oxidant and inflammatory burden [101].

3.3.5 Nuclear Factor (Erythroid-Derived 2)–Like 2 in the Cardiovascular System

The NRF2 diseasome cluster points to the high susceptibility of the cardiovascular system to changes in the cellular redox balance and the development of well-known comorbidities like atherosclerosis, hypertension, and diabetes (Griendling and FitzGerald, 2003a,b; [102]; [103]) (Fig. 3.8). A role of NRF2 in preventing these pathophenotypes has been demonstrated in *Nrf2*^{2/2} mice, which exhibit impaired cardiac structure (more remodeling events) and function (less fractional shortening) in response to chronic endurance exercise [104]. They are also more susceptible to develop heart failure after myocardial infarction [105]. In contrast, constitutive activation of NRF2 creates a reductive state, characterized by increased cardiac GSH/glutathione disulfide ratio and decreased ROS formation and malondialdehyde levels [106]. In humans, microarray analysis in Tako–Tsubo cardiomyopathy indicated an increase in pathologic ROS levels and a compensatory upregulation of NRF2 during the acute phase of this contractile dysfunction [107]. Recently, systemic inflammation and pathologic ROS formation in hemodialysis patients were associated with downregulation of NRF2 [108], and two promoter polymorphisms (rs35652124 and rs6721961) were associated with increased risk of mortality in these pa-

tients [109].

One of the most relevant targets of NRF2 in endothelial homeostasis is HO-1, which is usually paralleled by upregulation of ferritin, hence decreasing free iron levels and preventing Fenton-type reactions. Bilirubin, which is generated from the combined activity of HO-1 and biliverdin reductase, is one of the most powerful endogenous antioxidants that scavenges ROS/RNS [110], and is highly efficient in preventing lipid peroxidation in vitro [111]. Hmox12/2 mice show increased pulmonary hypertension in response to chronic hypoxia [112], and pharmacological HO-1 induction improves diabetic complications [113], as well as nitroglycerin-induced vascular dysfunction (nitrate tolerance) [114]. Recently, Hmox12/2 mice were shown to display upregulated NADPH oxidase-2, vascular pathologic ROS formation, markers of inflammation, endothelial dysfunction, and higher blood pressure in response to angiotensin-II [48]. In fact, high serum levels of bilirubin are inversely correlated with the incidence of coronary artery disease [115]. Bilirubin prevents the activation of the vascular NADPH oxidase [116], involved in the development of cardiovascular diseases (Griendling and FitzGerald, 2003a,b; [102]; [103]). Patients suffering from peripheral artery disease, which is a common manifestation of atherosclerosis, present reduced levels of HO-1 [117].

3.3.6 Nuclear Factor (Erythroid-Derived 2)–Like 2 in Metabolic Diseases

Type 2 diabetes mellitus (T2DM) is one of the most common chronic metabolic diseases and is highly underlined in the NRF2 diseasome. Pathologic ROS formation in insulin-sensitive tissues, as well as in pancreas, has been found in T2DM patients, resulting in severe impairment of both insulin secretion by pancreatic β cells and insulin action in peripheral tissues [118] (Fig. 3.8). Likewise, pathologic ROS levels also contribute to the pathogenesis of diabetic complications due to nonenzymatic glycation of proteins. This has been evidenced in diabetic nephropathy, in which the glomeruli exhibit pathologic ROS levels and a compensatory elevation of NRF2 [119]. Since 2007, several studies have addressed the role of NRF2 in

T2DM and its complications using animal models and cell lines. In vitro studies in human cells reported that NRF2 activation is achieved with acute exposure to high glucose, whereas longer incubation times or oscillating glucose concentration failed to activate NRF2 [120, 121]. Accordingly, these studies pointed out that NRF2 activation is dependent on glucose concentration and dynamics. In contrast, NRF2 is downregulated in peripheral blood mononuclear cells of prediabetic and diabetic patients, suggesting that NRF2 could be an important therapeutic target [122].

The impact of NRF2 deficiency on hyperglycemia was first shown in *Nrf2*^{2/2} mice, where oxidative and nitrosative alterations were enhanced and led to early-stage renal injury [123]. In a subsequent study, streptozotocin-induced diabetic *Nrf2*^{2/2} mice exhibited exacerbated glomerular injury, together with high ROS production and increased expression of the profibrotic markers TGF- β and fibronectin [119]. In this diabetic model, NRF2 protected against dysfunction of the blood–retina barrier and the progression of diabetic retinopathy [124]. Likewise, HFD-induced increase in vascular ROS levels was significantly exacerbated in *Nrf2*^{2/2} mice and was accompanied by a severe endothelial dysfunction, as shown by diminished acetylcholine-induced relaxation of aorta and increased expression of intercellular adhesion molecule-1 and TNF- α [120].

NRF2 plays a complex role in tissue-specific insulin resistance. Thus, HFD-fed *Nrf2*^{2/2} mice displayed better insulin sensitivity due to enhanced insulin signaling in liver and skeletal muscle than their wildtype counterparts, but conversely, these mice developed a severe NASH due to excessive hepatic lipotoxicity linked to pathologic ROS formation [125]. Accordingly, this study dissociated hepatic insulin resistance from the development of NASH. In light of these data, a subsequent study demonstrated that the livers of HFD-fed *Nrf2*^{2/2} mice exhibited higher pathologic ROS formation by a significant depletion of GSH due to attenuated expression of the CYP2A5 enzyme [126]. The knockdown of NRF2 in hepatocytes enhanced the apoptosis induced by palmitate, a fatty acid that is highly elevated in insulin-resistant obese patients. This effect was correlated with increased production of pathologic ROS, again reinforcing the key role of NRF2 in the progres-

sion of NASH [127].

To further examine the role of NRF2 in the metabolic syndrome, NRF2 was ablated in leptin-deficient (ob/ob) mice, a model with an extremely positive energy balance [128]. Interestingly, global ob/ob/Nrf2^{-/-} mice or adipocyte-specific ob/ob/Nrf2^{-/-} mice displayed reduced white fat mass, revealing NRF2 as a key player in adipogenesis. These mice had an even more severe metabolic syndrome that was characterized by hyperlipidemia, aggravated insulin resistance, and hyperglycemia, suggesting a mechanistic linkage between the metabolic syndrome and pathologic ROS formation.

Another subset of studies has evaluated the effects of persistent induction of NRF2 in glucose metabolism. Genetic NRF2 induction making use of a hypomorphic allele of Keap1 (Keap1^{flox/-} mutant) decreased blood glucose in the obese diabetic db/db mice by suppressing hepatic glucose 6 phosphatase through the repression of cAMP-CREB signaling in hepatocytes, as well as other gluconeogenic genes, such as peroxisome proliferator-activated receptor coactivator-1α [129]. Additionally, enhancement of NRF2 activity in Keap1^{-/-} knocked down mice increased the phosphorylation of AMP-activated protein kinase (AMPK) in the liver, as well as insulin signaling in skeletal muscle, resulting in a substantial improvement of glucose tolerance [130]. Due to the pleiotropic activities of NRF2 in the context of T2DM, the results of all these and other studies have evidenced the need to design multiple genetic and pharmacological strategies to elucidate the full array of NRF2 functions in tissues involved in the control of whole-body glucose homeostasis.

In addition to diabetic factors like age, body weight, and blood glucose, genetic factors that are linked to NRF2 have been poorly studied in humans. In a Chinese population, the SNP rs6721961 has been associated with pathologic ROS formation and risk of newly diagnosed T2DM and may also contribute to impaired insulin secretory capacity and increased insulin resistance [131]. The same SNP was associated with diabetes in Mexican mestizo men [132]. In a case-control study performed with Han volunteers, a significant

difference in genotypic and allelic frequencies of four SNPs of the NFE2L2 gene was found between T2DM patients with and without complications, including peripheral neuropathy, nephropathy, retinopathy, foot ulcers, and microangiopathy [133].

3.3.7 Nuclear Factor (Erythroid-Derived 2)–Like 2 in Neurodegenerative Diseases

The NRF2 diseasesome provides evidence of NRF2 involvement in several neurodegenerative diseases, including AD and PD, which represent the most prevalent cognitive and motor disorders of the elderly. In neurodegenerative diseases, the connection between low-grade pathologic ROS formation and proteostasis is particularly relevant because most of these pathophenotypes are characterized by abnormal aggregation of specific proteins (Fig. 3.8). Evidence pointing to pathologic ROS formation in proteinopathy, as well as NRF2 as regulator of proteasome and autophagy was provided in cellular and animal models [6]. Initially, it was reported that the autophagy cargo protein sequestosome 1 (SQSTM1) competes with NRF2 for binding to KEAP1. SQSTM1 takes KEAP1 to the autophagosome degradative pathway, therefore upregulating NRF2 [134]. More recently, it was found that NRF2 regulates the expression of autophagy genes involved in autophagy initiation, cargo recognition, elongation, and autolysosome clearance [5]. In this study, amyloidopathy and tauopathy induced by transgene overexpression of human mutant amyloid precursor protein and tau were aggravated in Nrf2^{-/-} mice. A connection between NRF2 deficiency and neurodegeneration is supported by a growing body of evidence in animal models (Johnson and Johnson, 2015). The general point of view is that damaged neurons try to activate NRF2-dependent transcription, presumably to increase their own survival. Additionally, upregulation of NRF2 in astrocytes participates in metabolic compensations, including increased supply of GSH for augmenting their proliferative capacity (Bolanos, 2016), whereas NRF2 upregulation in microglia returns this immune cell to a resting state [135].

Ramsey et al. [136] evidenced the nuclear localization of NRF2 in dopaminergic neurons of patients with PD. Other studies found

that amyloid precursor protein- and tau-injured neurons expressed increased levels of NRF2 and its target SQSTM1, probably as compensatory mechanism to clear these toxic proteins through autophagy [5, 28]. In agreement with these results, the levels of HO-1, NQO1, GCLM, and SQSTM1 are increased in AD and PD brains [137]; [138–140]. However, there is some controversy in the field, as Ramsey's study described the accumulation of NRF2 in the cytosol of AD-injured neurons, suggesting an impaired capacity of these neurons to upregulate NRF2 transcriptional activity. In addition, the cytoprotective proteins associated with NRF2 expression, such as NQO1 and SQSTM1, were partly sequestered in Lewy bodies, suggesting impaired neuroprotective capacity of the NRF2 signature in PD patients [140]. One possible explanation for this discrepancy could be that the levels of NRF2 and its target genes might change during ageing and disease progression.

Some SNP haplotypes of NFE2L2 were associated with decreased risk or delayed onset of ALS, AD, or PD. The onset of ALS was analyzed in two studies regarding three functional promoter SNPs that were previously linked to high gene expression. Interestingly, this haplotype was associated with a 4-year delay in onset of ALS [141], but another study did not find a clear association [142]. Regarding AD, one haplotype allele was associated with 2-year earlier age onset of AD, suggesting that variants of the NFE2L2 gene may affect AD progression (von [143]). Genetic association of NFE2L2 with PD has been analyzed in more detail. Three SNPs in the NFE2L2 promoter (rs6721961, rs6706649, and rs35652124) were evidenced as protective haplotype in a case-control study [144]. Such haplotype delayed the onset of disease in a Swedish cohort or even reduced the risk of PD in a Polish cohort. These results were supported by four new independent European case-control studies [145], but were not replicated in a Taiwanese population [146], suggesting disparity in ethnicities and environmental factors. As an alternative approach, PD cells derived from olfactory mucosa were exposed to smoke extract or pesticide to assess gene–environment interaction, and several SNPs were identified that affect the susceptibility to these toxins [147]. Altogether, it is possible that a slight activation of NRF2, such as that found for some functional haplotypes of the NFE2L2 gene, should be enough to trigger protective

mechanisms in the brain.

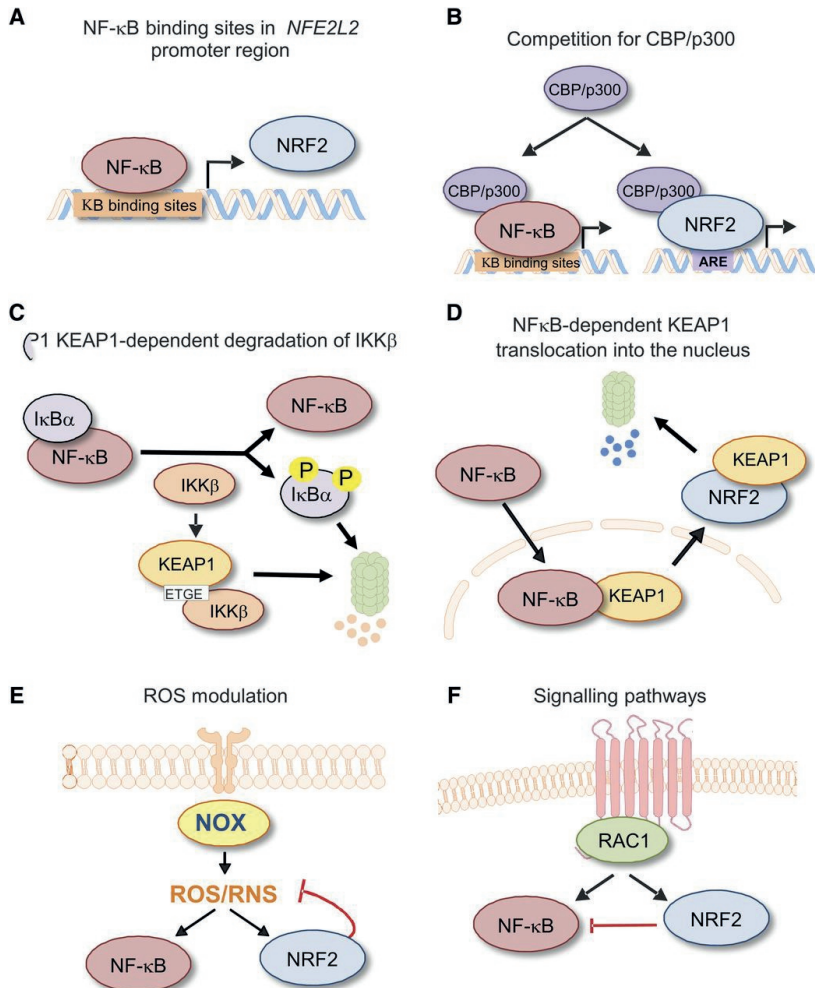


Figure 3.7: Crosstalk between NF- κ B and NRF2 occurs at different levels. (A) Responsive elements have been identified in the promoter region of *NFE2L2*. (B) Both NRF2 and NF- κ B transcription factors compete for binding to the transcriptional coactivator CREB-binding protein (CBP/p300). (C) The NF- κ B activating kinase IKK β contains an ETGE motif allowing KEAP1 binding and subsequent ubiquitin–proteasome degradation. (D) NF- κ B was reported to bind and translocate KEAP1 to the nucleus, thereby promoting NRF2 degradation. (E) ROS produced during inflammation activate NF- κ B and NRF2; finally, NRF2 attenuates ROS and consequently NF- κ B activity. (F) Different proinflammatory signals activate the Rho GTPase RAC1, which leads to NF- κ B and NRF2 activation. Then NRF2 inhibits RAC1-mediated activation of NF- κ B.

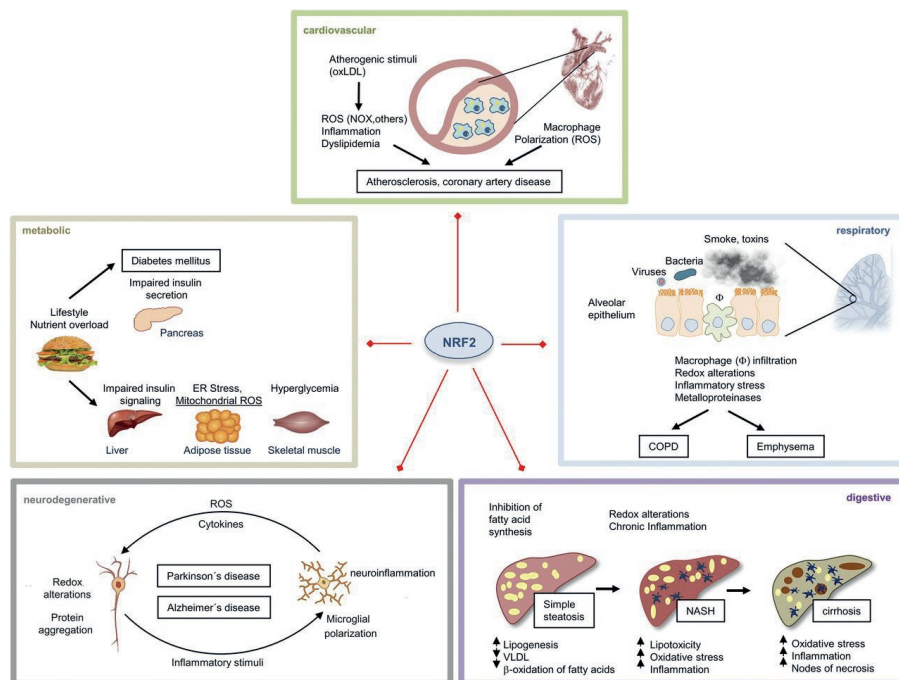


Figure 3.8: Role of NRF2 in common mechanisms and pathophenotypes of chronic diseases. The picture provides some examples extracted from the NRF2 diseasome cluster of Fig. 3.5. Common pathomechanisms of these diseases include abnormally high ROS levels and low-grade chronic inflammation that participate in tissue damage. NRF2 provides a cytoprotective signature against these and other tissue-specific alterations by regulating the expression of numerous cytoprotective genes.

3.4 The Kelch-Like ECH-Associated Protein 1 Paradox in Cancer

An apparent dichotomy appears to exist in the role of NRF2 in tumorigenesis and further tumor progression. On one hand, by activating biotransformation reactions, NRF2 protects against chemically induced carcinogenesis. Preclinical studies have demonstrated complete protection against aflatoxin B(1)-induced liver cancer after pharmacological activation of NRF2 in rats [148]. In contrast, the protective responses elicited by NRF2 provide a growth advantage in established cancers, and this will be the focus of this section.

Constantly increased levels of ROS can sustain tumorigenesis through alteration of genomic stability, along with activation of specific redox signaling circuits and inflammatory processes that favor survival and proliferation of tumor cells [149]. Therefore, upregulation of NRF2 represents a mechanism of adaptation of cancer cells to tolerate high ROS levels that propel tumor progression (Schumacker, 2006) as well as to maintain cancer stem cells that are responsible for tumor relapse and formation of distant metastases [150]. For instance, the NRF2 signature in cancer stem cells from human colorectal tumors pointed out protective mechanisms mediated by high levels of GCLC, glutathione peroxidase, and thioredoxin reductase-1 that underlie the ability of these cells to counteract stressors and chemotherapeutics [151]. From this perspective, NRF2 behaves in cancer cells like an oncogene that by inducing chronic activation of ARE-mediated cytoprotective responses affords adaptation to their oxidative environment (Panieri and Santoro, 2016). Several mechanisms of malignant activation of NRF2 have been reported, including somatic mutations, epigenetics, and oncogenic signaling alterations.

Close to 600 somatic mutations have been reported in cancer along the coding sequence of NFE2L2 [152]. In Fig. 3.9, we show the results from a dataset of 10,000 cancer patients [153]. In most cases, these mutations alter the interaction of the DLG and ETGE motifs of NRF2 with KEAP1, hence inducing hyperactivation of NRF2 in several solid tumors, including esophagus, skin, lung, and

larynx carcinomas [154, 155]. For instance, in advanced esophageal squamous cancer, gain-of-function mutations of NRF2 were associated with tumor recurrence and poor prognosis due to increased proliferation, attachment-independent survival, and resistance to chemo- and radiotherapy [156]. Loss-of-function mutations in the KEAP1 gene are also frequent in some solid tumors such as lung cancer [157]. Based on the strong evidence that this pathway regulates b-TrCP/NRF2, it is strange that disrupting somatic mutations have not been found at the interface between NRF2 and b-TrCP. This fact suggests that such mutations are not viable for unknown reasons or that the increase in NRF2 levels that would result from the escape of b-TrCP is not sufficient to drive oncogenicity.

Nonetheless, somatic mutations account for chronic NRF2 activation only in a fraction of cancer patients. At the level of gene expression, it is interesting that an allele of the SNP rs6721961 (2617C . A) located at the ARE enhancer of the human NRF2 gene abolished self-induction of NRF2, and this correlated with remarkable survival of these cancer patients [158]. Epigenetic changes due to promoter hypermethylation of three CpG sites of KEAP1 have been described in lung tumors, resulting in consequent NRF2 activation that could be reversed by 5-aza-29-deoxycytidine treatment [159]. The role of miRNAs in the posttranscriptional regulation of NRF2 levels has been reviewed recently (Kurinna and Werner, 2015). Briefly, miR200a targets the KEAP1 mRNA in human breast cancer cells, leading to its degradation and consequent activation of NRF2 [160]. In turn, miR28 facilitates the degradation of the NRF2 mRNA [161].

Oncogenes or mutated tumor suppressors may enhance the activation of NRF2 in cancer. Endogenous oncogenic alleles of KRAS, BRAF, or c-MYC upregulate NRF2, presumably through oncogene-mediated ROS generation and consequent chronic inactivation of KEAP1 [162]. The mutated form of the tumor suppressor p53, which sustains the growth of cancer cells by enhancing nutrient uptake and synthesis of building blocks, can also upregulate NRF2, possibly through the crosstalk with the Sp1 transcription factor that binds to the NRF2 promoter [163].

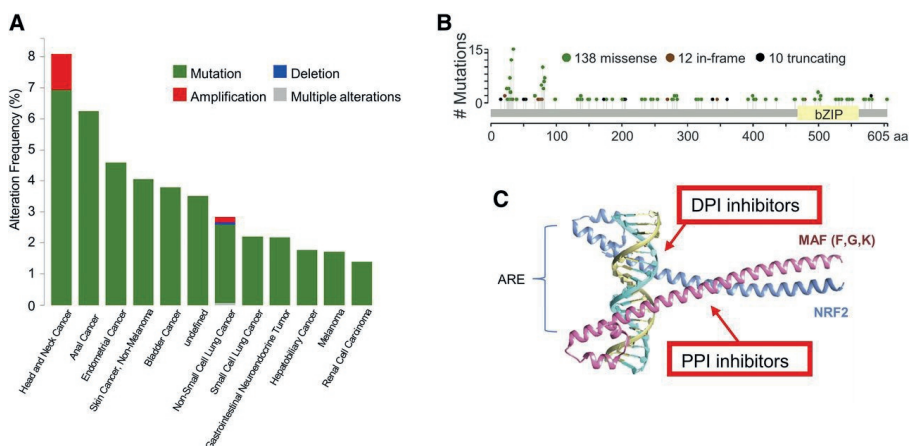


Figure 3.9: Somatic mutations found in tumors of the MSK-IMPACT Clinical Sequencing Cohort (MSKCC) study [153] and pharmacologic strategies to inhibit NRF2. (A) Percentage of tumors with NRF2 mutations. (B) Distribution of mutations along the NRF2 polypeptide. (C) PyMOL representation of the interaction between the NRF2/MAFF heterodimer and the ARE element. Blue, NRF2; pink, MAFF. The red arrows indicate possible mechanisms of inhibition of NRF2 by small molecules that could target the bZip domains of interaction between NRF2 and MAF proteins (PPI inhibitors) or the interface of interaction of the NRF2-MAF heterodimer with the ARE [DNA-protein interaction (DPI)].

The phosphorylation of NRF2 (phospho-NRF2) by various protein kinases is a potential mechanism of activation in 107 hepatocellular carcinomas. Increased levels of phospho-NRF2 were associated with reduced KEAP1 expression and poor 5-year overall survival of patients exhibiting this distinctive phenotype [164]. Additionally, mutations in the phosphatase and tensin homolog (PTEN) tumor suppressor sustain hyperactive and oncogenic phosphatidylinositol-3-kinase (PI3K)-AKT signaling and consequent increase in NRF2 activity due to the downregulation of the PTEN/GSK-3/ b-TrCP pathway for the proteasomal degradation of NRF2 [9–11]. Therapeutic interventions targeting the PTEN/GSK-3/b-TrCP pathway should take into consideration that GSK-3 may act both as tumor suppressor and tumor promoter, and is also implicated in the generation of cancer stem cells [165].

Some stress-induced proteins interact with KEAP1 and thus com-

pete with its NRF2 binding in cancer cells. Consequently, NRF2 escapes KEAP1-mediated degradation. One of the NRF2 competitors for KEAP1 binding is the phosphorylated form of the autophagy-adaptor protein SQSTM1 that occurs during selective autophagy used by cancer cells for sustaining their own growth [166]. The cyclin-dependent kinase inhibitor p21 that promotes cell cycle arrest in cancer stem cells was also demonstrated to inhibit the association of NRF2 with KEAP1 through the interaction of its KRR motif with DLG and ETGE motifs in NRF2 [167]. Recently, it was shown that dipeptidyl-peptidase 3, which bears an ETGE motif, may compete with NRF2 for binding to KEAP1 [168]. Overexpression of dipeptidyl-peptidase 3, possibly induced by chronic alteration of the redox status, correlates with increased expression of ARE genes and poor prognosis, particularly in estrogen receptor-positive breast cancer [32].

NRF2 induces metabolic changes that contribute to cancer progression. For instance, a multiplatform nontargeted metabolomics study identified patterns of metabolite changes in breast tumor samples [169]. They found that GSH and 3-(4-hydroxyphenyl)lactate were positively correlated with the involvement of BRCA1 in redox homeostasis through interaction with NRF2. Metabolomics studies also indicate that NRF2 can increase aerobic glycolysis in cancer cells to support their high-energy requirements. This occurs through NRF2-mediated induction of Mn-superoxide dismutase expression, leading to elevated mitochondrial production of hydrogen peroxide and to activation of AMPK. The process is regulated by caveolin-1, which binds directly to both NRF2 and KEAP1, and impedes on NRF2 activation and hence on the glycolytic shift. This is apparently one explanation why glycolytic tumors, which are generally more aggressive, have a caveolin-1^{low}/Mn-superoxide dismutase^{high} phenotype [170]. NRF2 can also drive glucose and glutamine toward anabolic pathways required for tumor cell proliferation [171]. In the presence of active PI3K–AKT signaling and loss of KEAP1 activity, NRF2 was shown to induce the shift of glucose metabolism from glycolysis toward anabolic pathways (purine synthesis) in cancer [171, 172]. This is underlined by NRF2-mediated transcription of genes involved in the pentose phosphate pathway and generation of NADPH (glucose 6 phosphate dehydrogenase; phosphogluconate

dehydrogenase, malic enzyme 1, isocitrate dehydrogenase 1, transketolase, and transaldolase), along with genes involved in purine nucleotide synthesis (phosphoribosyl pyrophosphate amidotransferase, methylenetetrahydrofolate dehydrogenase 2).

The fact that activation of NRF2 confers a growth advantage to cancer cells might argue that its pharmacologic activation in chronic diseases presented in this work might imply a high risk of developing cancer. However, it must be considered that the oncogenic activity of NRF2 requires mutations in its gene or in KEAP1, which results in very high and persistent induction of NRF2 signaling. This is not the case in pharmacological therapy, in which it is possible to modulate drug dosing and NRF2 activity. Moreover, empirical evidence indicates that subjects enrolled in clinical trials with NRF2 activators do not exhibit increased cancer risk. This is best exemplified in the case of patients with MS, who have been taking the NRF2 activator dimethyl fumarate for several years since it was approved by the regulatory agencies in 2013. Conversely, the use of NRF2 inhibitors in cancer patients might lead to manifestation of the pathophenotypes described in the NRF2 diseasome. This is a possibility that will need further investigation when NRF2 inhibitors reach the clinic.

3.5 Nuclear Factor (Erythroid-Derived 2)– Like 2 Drugome

This section attempts to develop a NRF2 drugome that might be useful for future clinical directions to target therapeutically NRF2 centered on the pathophenotypes of the NRF2 diseasome. As stated in Fig. 3.2B, the pharmacological activation of NRF2 is being pursued for increasing its stability by targeting KEAP1. These strategies are based on the discovery of either electrophile compounds that alter the KEAP1 structure or small molecules that prevent the docking of NRF2 to KEAP1. Although not yet demonstrated empirically, the GSK-3 inhibitors should prevent the recognition of NRF2 by b-TrCP and some compounds could be discovered to prevent binding of NRF2 to b-TrCP. The inhibition of NRF2 is being analyzed with compounds that target the bZip dimerization domain to prevent formation of the active NRF2/MAF heterodimer. By comparison with other transcription factors, it might be possible to find small molecules that impede binding

of the NRF2/MAF heterodimer to the ARE (Fig. 3.9C). This section summarizes the most important findings from a translational point of view in both de novo drug discovery and repurposing.

3.5.1 Electrophilic Nuclear Factor (Erythroid-Derived 2)-Like 2 Inducers

3

The majority of known physiologic or pharmacological NRF2 inducers are electrophilic molecules that covalently modify, by oxidation or alkylation, cysteine residues present in the thiol-rich KEAP1 protein (Hur and Gray, 2011; [50]). KEAP1 is one of the best-suited proteins to act as electrophilic or redox sensor, as it contains 27 cysteine residues in humans and functions as an electrophile trap. The cysteines C151, C273, and C288 of KEAP1 appear to be the most prone to electrophile reaction (Fig. 3.2B), although there are some specificities [173, 174]. Electrophile adducts inhibit KEAP1 in two different ways. One is induction of a conformational change in KEAP1 that will result in loss of its binding capacity to NRF2. The other is the blockade of the interaction between KEAP1 and CUL3/RBX1, resulting in sequestration of KEAP1 with NRF2 and further stabilization of newly synthesized NRF2 [174–177].

At least 30 recent patents for NRF2 modulators are indexed in the World International Property Organization. These patents are protecting chalcone derivatives, novel amide triterpenoid derivatives, deuteriumsubstituted fumarate derivatives, 3-alkylamino-1Hindolyl acrylate derivatives, withanolide, a benzyl derivative containing an activated vinyl group, andrographolide or [S]+apomorphine, and sesquiterpene lactone derivative [178]. Although most of these compounds proved to be useful to some degree from a preclinical proof-of-concept perspective, their clinical value is to date generally very limited. Only a few of them have entered clinical trials, and regulatory bodies, such as Food and Drug Administration or European Medicines Agency, have approved even fewer. We in this study discuss the most developed NRF2 activators along the translational pipeline.

Fumaric acid esters are the most prominent example of a KEAP1

modifier, and dimethyl fumarate (DMF) is to date the only Food and Drug Administration- and European Medicines Agency-approved drug registered as NRF2 activator. The monoester form of DMF, monomethyl fumarate (MMF), was described as its active metabolite. DMF and MMF are Michael acceptors that directly react with cysteine residues present in KEAP1 [179].

DMF and other fumaric acid esters have been used for treating psoriasis for over 50 years, starting at a time when the function of NRF2 was still unknown. This compound was licensed in Europe under the commercial name of Fumaderm. Clinical trials showed a decrease in the psoriasis area and severity index to 50%–80% after 12–16 weeks of DMF therapy [180, 181]. More recently, DMF has demonstrated its efficacy in the treatment of adults with moderate to severe chronic plaque psoriasis in a phase III trial (BRIDGE) [182]. The mechanism of action underlying fumarates in remission of psoriatic lesions includes the decrease in the number of peripheral T cells along with a shift from a Th1 toward a Th2 immune response [183, 184]. In another autoimmune disease, SLE, fumaric acid esters have been used as systemic combination therapy in the treatment of severe, extensive, and refractory cutaneous manifestations (Saracino and Orteu, 2017).

DMF was approved in 2013 for the treatment of MS under the commercial name Tecfidera [185]. The use of DMF in MS patients was propelled by positive results obtained in the MS mouse model of EAE. Significant therapeutic effects on the disease course and histology were associated with a markedly reduced macrophage-mediated inflammation in the spinal cord. Multiplex cytokine analysis in blood evidenced an increase of the anti-inflammatory cytokine IL-10 in DMF-treated animals [186]. Moreover, DMF also improved preservation of myelin, axons, and neurons in wild-type, but not in *Nrf2*^{2/2} mice [187]. In humans, DMF demonstrated a significant reduction of lesions and annualized relapse rate in MS [188]. Two phase III clinical trials, DEFINE and CONFIRM, substantiated these results [189, 190]. Therefore, DMF is currently used as the first line of treatment of relapsing-remitting MS that cannot be treated by traditional therapies. New formulations of DMF are being tested and patented to improve drug bioavailability and efficacy [178]. For instance, MMF has been used to develop a

second generation of NRF2 inducers as prodrugs [191]. The lead compound ALKS-8700, a 2-(2,5-dioxo-1-pyrrolidinyl)ethyl ester derivative of MMF, is rapidly converted into MMF in the body, hence increasing its bioavailability and reducing gastrointestinal side effects. ALKS-8700 is currently under phase III clinical trial (EVOLVE MS).

DMF and MMF modulate the immune response. For example, they inhibit the maturation of DCs by reducing the release of inflammatory cytokines and hence the ability of DCs to process antigens. Moreover, DMF and MMF activate natural killer cells to lyse DCs and enhance apoptosis of both DCs and T cells [183, 192]. As such, DMF and MMF impede T cell-mediated autoreactivity. Some studies indicate that DMF also induces type II DCs by triggering GSH depletion, which results in enhanced HO-1 activity and suppression of STAT1 phosphorylation. These classic type II DCs suppress Th1- and Th17-mediated responses in favor of Th2 ones. Furthermore, the increased production of IL-10 by DCs favors the differentiation of CD4⁺ T cells toward a suppressive Treg phenotype [183, 193]. DMF also inhibits the nuclear translocation of NF- κ B [194] and consequently the production of inflammatory mediators, such as TNF- α , IL-1b, IL-6, chemokines, adhesion molecules, and nitric oxide in microglia and astrocytes [195], as well as in peripheral blood mononuclear cells [196]. In addition, DMF exerts antiangiogenic effects that are dependent on the downregulation of vascular endothelial growth factor receptor-2 expression in endothelial cells [197]. Recent findings indicated that DMF reduced the number of CD4⁺, CD8⁺, Th1, and Th17 cells, whereas the CD4⁺/CD8⁺ ratio and the Th2 subset were increased in the blood of these patients. Interestingly, the inhibitory effects of DMF/MMF on T cell activation were confined mainly to memory T cells [198]. These immunomodulatory activities of DMF or MMF are important for the protection of oligodendrocytes against ROS-induced cytotoxicity [199].

Additional mechanisms might explain the inhibition of NF- κ B independently of NRF2 activation. Thus, DMF may interact with cysteine residues in several proteins that regulate NF- κ B signaling [200]. In addition, DMF inhibits ubiquitin-conjugating enzymes and thus prevents the degradation of the I κ B repressor of NF- κ B in response to IL-1b or Toll-like receptor agonists [201]. Moreover, DMF binds directly to

specific cysteine residues in protein kinase C- α , a key kinase involved in signaling by the T cell receptor [200]. In addition, MMF and DMF activate the hydroxycarboxylic acid receptor-2, resulting in inhibition of NF- κ B and downregulation of proinflammatory cytokines and adhesion molecules [202, 203] and leading to decreased neutrophil infiltration [202]. Although these NRF2-independent effects would be relevant in the acute inflammatory phase of EAE, the neuroprotective efficacy of DMF during chronic autoimmune demyelination depends on NRF2 activation [204]. The clinical benefit of DMF treatment in both Nrf2^{-/-} and wild-type mice was associated with a reduction of inflammatory Th1 and Th17 cells, as well as with induction of anti-inflammatory M2 monocytes. At the same time, decreased expression of CD80 and CD86 costimulatory molecules was observed in wild-type, but not in Nrf2^{-/-} mice, indicating that at least these effects were NRF2-dependent [205].

The success of DMF for the treatment of autoimmune diseases indicates that other diseases that share common pathomechanisms underlined by chronic, lowgrade inflammation and pathologic ROS formation might benefit from the repositioning of this drug. In a mouse model of Huntington's disease, the survival rate, muscle function, and body weight were preserved with DMF treatment, and this was associated with an increased number of intact neurons [187]. Also, in a recent preclinical study of PD, using the α -synucleinopathy model of this disease, DMF was neuroprotective in wild-type, but not in Nrf2^{-/-} mice due to impaired autophagy induction ([140]).

DMF was shown to prevent endothelial dysfunction and cardiovascular pathologic ROS formation and inflammation in diabetic mice [206], and decreased atherosclerosis, kidney dysfunction, and other diabetic complications were reported in apolipoprotein E-deficient mice after streptozotocin injection [207]. Additionally, several studies indicated that DMF might exert antitumor activity by inhibiting the NF- κ B pathway, hence adding therapeutic value in the treatment of aggressive cancers [208]. DMF is a relevant example of the drug-repurposing concept within the network pharmacology approach.

Synthetic triterpenoids are derivatives of 2-cyano- 3,12-dioxo-

oleana-1,9(11)-dien-28-oate (CDDO; bardoxolone, RTA401) that resemble the natural product oleanolic acid. They exhibit Michael acceptor activity through its α - β unsaturated scaffold and represent the most potent inducers of NRF2 [178]. They interact with C151 of KEAP1 and impede its interaction with CUL3, hence leading to NRF2 activation [177]. Proof-of-principle studies strongly support the use of synthetic triterpenoids for degenerative diseases and are being the focus of intensive research as antioxidant modulators of inflammation by Reata/ Abbott. For instance, CDDO-imidazole (CDDO-Im RTA403) induced in peritoneal neutrophils of wildtype but not Nrf2^{-/-} mice the expression of various antioxidant genes (Hmox1, Gclc, Gclm, and Nqo1) and attenuated LPS-induced ROS generation and production of proinflammatory cytokines, consequently decreasing mortality [209]. CDDO-ethyl amide (RTA405) and CDDO-CDDOtrifluoethyl amide (RTA 404) had significant effects across all endpoints measured in a toxin-induced PD model (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine) [210]. In the EAE model of MS, CDDO-CDDO- trifluoethyl amide suppressed inflammation, pathologic ROS formation, and myelin degeneration [211].

CDDO-methyl ester (CDDO-Me, RTA 402) was the first CDDO that reached in clinical trials for the treatment of diabetic nephropathy [212]. Although the results of the phase II were very encouraging, CDDO-Me was later withdrawn at phase III (BEACON trial) due to cardiovascular safety issues (Zhang, 2013) that were not related to NRF2 but most likely to an off-target alteration of endothelin signaling [213, 214]. Currently, CDDO-Me is under clinical study as potential treatment of Alport syndrome and pulmonary hypertension (Table 3.2). In an effort to improve its safety profile, further studies have led to the development of CDDOdifluoropropionamide (RTA-408, omaveloxone), which is currently in phase II trial for the treatment of Friedreich's ataxia, ocular inflammation, and pain after ocular surgery.

Oltipraz is an organosulfur compound that is used as an antischistosomal agent and is currently in phase III trial for the treatment of nonalcoholic steatohepatitis. Advanced clinical trials for the treat-

ment of Huntington's disease are under development with minocycline, an antibiotic that has demonstrated neuroprotective properties due to NRF2 activation [215]. Another NRF2 inducer in phase I clinical study for the treatment of acute kidney disease is CXA-10, a nitro fatty acid with anti-inflammatory properties through the activation of NRF2 (Batthyany and Lopez, 2015). Many other NRF2 inducers with the same mechanism of action have been described in the last years [59, 216, 217], and some are in preclinical studies, such as the compound VEDA-1209, a chalcone derivative with a good anti-inflammatory profile for the treatment of ulcerative colitis.

SFN is an isothiocyanate produced from enzymatic cleavage of the organosulfur compound glucoraphanin, which is present in sprouts of broccoli, cabbage, and other Brassicacea plants. The catalytic reaction is driven by the enzyme myrosinase that is found in plants and microbiota of the GI tract [218]. More recently, SFN has been obtained by chemical synthesis [219]. Translation of SFN to the clinic has been achieved by administration of SFN-containing broccoli sprout powder to patients with T2DM [220]. Broccoli powder decreased plasma malondialdehyde and oxidized lowdensity lipoprotein (LDL) and increased the total antioxidant capacity. Cardiovascular risk factors such as serum triglycerides, oxidized LDL/LDL ratio, and atherogenic index of plasma (log of triglycerides/highdensity lipoprotein ratio) were also reduced. Furthermore, proinflammatory markers such as C-reactive protein and IL-6 were decreased. In a more recent study, SFN administered as concentrated broccoli sprout extract suppressed glucose production from hepatocytes by nuclear translocation of NRF2 and decreased expression of key enzymes involved in gluconeogenesis. Moreover, SFN reduced fasting blood glucose and glycated hemoglobin in obese patients with T2DM [221]. SFN-induced activation of NRF2 protected renal cells against lupus nephritis by reducing the ROS burden and by inhibiting the NF- κ B and TGF- β 1 signaling pathways [222].

In regard to neurodegenerative diseases, it has been shown that SFN crosses the blood brain barrier and provides sufficient cerebral bioavailability to activate the NRF2 signature and to reduce LPS-elicited neuroinflammation, as reflected in the reduction of

proinflammatory markers (inducible nitric oxide synthase, IL-6, TNF- α) and microgliosis in the hippocampus [223]. SFN also safeguarded dopaminergic neurons against the parkinsonian toxin 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine and attenuated astrogliosis and microgliosis [224]. In line with these findings, SFN reduced the levels of phosphorylated tau and increased Beclin-1 and LC3-II, suggesting that NRF2 activation might facilitate degradation of this toxic protein through autophagy in the brain [225]. SFN-treated rats subjected to spinal cord injury had significantly decreased levels of inflammatory cytokines, reduced contusion volume, and improved coordination [226]. This drug also ameliorated EAE by preserving the blood-brain barrier and by reducing pathologic ROS formation and the number of inflammatory cells [227]. SFN has been used to date in at least 32 clinical studies addressing chronic diseases such as cancer, asthma, chronic kidney disease, T2DM, cystic fibrosis, autism, and schizophrenia [228, 229] (Table 3.2).

Altogether, these observations paved the way for the development of other SFN-derived compounds exhibiting an improved pharmacokinetic profile. SFN is an oily substance with low stability in hydrophilic media. Its physicochemical profile prompted Evgen Pharma (Wilmslow, Cheshire, England) to develop a cyclodextrin complex formulation, Sulforadex, which is under phase II clinical trial for the treatment of subarachnoid hemorrhage. SFN was also hybridized with melatonin to generate the ITH12674, a compound that was designed to have a dual drug-prodrug mechanism of action for treatment of brain ischemia [230].

Curcumin is the main curcuminoid found in turmeric and has been used for the treatment of obesity, metabolic syndrome, and pre-diabetes. A nontargeted metabolomics study to investigate the effects of curcumin on rat liver was conducted by means of gas chromatography with electron impact mass spectrometry. The intermittent intake of curcumin upregulated NRF2 and displayed antioxidant and anti-inflammatory roles in the protection against liver damage [231]. Oral consumption of curcumin is effective in lowering serum triglycerides, IL-1b, IL-4, and vascular endothelial growth factor, and in increasing adiponectin levels in blood. In T2DM patients, curcumin decreases the

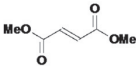
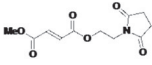
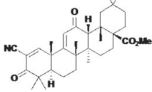
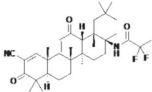
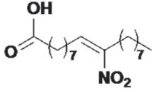
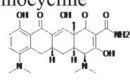
levels of fasting blood glucose, glycated hemoglobin, serum free fatty acids, triglycerides, and uric acid, and increases the levels of lipoprotein lipase [232, 233].

Resveratrol is a polyphenol that protects plants against fungal infection and is found in the skin of grapes, red wine, berries, and many other plants. Resveratrol exerts antioxidant properties through activation of NRF2 signaling by downregulating KEAP1 expression and by activating the protein deacetylase sirtuin-1 [234]. In healthy subjects, the dietary administration of resveratrol prevented the elevation in plasma of cholesterol, endotoxins, prooxidants, and inflammatory markers (p47phox, KEAP1, IL-1b, and TNF- α). These events correlated with the elevation of NRF2 activity as determined by enhanced expression of its targets NQO1 and glutathione S-transferase [235]. In T2DM patients, insulin sensitivity was improved after 4 weeks of treatment, as determined by enhanced insulin signaling via AKT, decreased pathologic ROS formation, and reduced levels of glycated hemoglobin [236, 237]. Overall, resveratrol was reported to prevent major cardiovascular, inflammatory, oxidative, and metabolic complications in hypertension, hypercholesterolemia, atherosclerosis, ischemic heart disease, diabetes, and metabolic syndrome in animal models and patients [238].

A problem that is frequently overlooked is the lack of selectivity of electrophilic KEAP1 inhibitors. Electrophiles react with different nucleophiles present in the cell, thus exhibiting off-target and undesired side effects. For instance, CDDO-Im can interact with more than 500 different targets [239]. In general, several protein phosphatases contain redox-sensitive cysteines in their catalytic center, and some KEAP1 inhibitors may modify and inactivate these phosphatases, hence disturbing signaling networks. One of these phosphatases is PTEN [240–242]. The catalytic C124 residue of PTEN can be modified through adduct formation with strong electrophiles such as CDDO-Im [243] and tertbutylhydroquinone [244]. Then, the increased activation of the PI3K/AKT pathway involves inhibition of GSK-3 and subsequent stabilization of NRF2 (Fig. 3.2C) [10, 11]. Moreover, KEAP1 interacts with other proteins that also contain the high-affinity binding motif ETGE [168], such as Bcl-2 and IKK β [245, 246]. Therefore, some results ob-

tained from KEAP1-deficient cells may not necessarily be related to NRF2 activation.

Table 3.2: Selected NRF2 inducers acting as electrophilic modifiers of KEAP1
The reference corresponds to the code in ClinicalTrials.gov.

Compound	Disease	Clinical Trial	Reference
Dimethyl fumarate			
	Multiple sclerosis Psoriasis Rheumatoid arthritis Adult brain glioblastoma Cutaneous T cell lymphoma Obstructive sleep apnea Chronic lymphocytic leukemia small lymphocytic lymphoma	Phase II Phase I Phase II Phase II Phase I	Approved Approved NCT00810836 NCT02337426 NCT02546440 NCT02438137 NCT02784834
ALKS-8700 (MMF-derivate)			
	Multiple sclerosis	Phase III	NCT02634307
CDDO-Me (RTA402)			
	Diabetic nephropathy Chronic kidney disease, T2DM, diabetic nephropathy Liver disease Hepatic impairment Advanced solid tumors Lymphoid malignancies Alport syndrome Pulmonary hypertension Pulmonary arterial hyperten- sion Renal insufficiency, T2DM	Phase II Phase III Phase I/II Phase I Phase I Phase II/III Phase III Phase III Phase II	NCT00811889 NCT01351675 NCT00550849 NCT01563562 NCT00529438 NCT00508807 NCT03019185 NCT03068130 NCT02657356 NCT01053936
CDDO-DFPA (RTA-408)			
	Mitochondrial myopathy Friedreich's ataxia Inflammation and pain follow- ing ocular surgery Corneal endothelial cell loss, ocular pain and inflammation, cataract surgery Phase II Melanoma Breast cancer Non-small cell lung cancer, melanoma	Phase II Phase II Phase II NCT02128113 Phase I/II Phase II Phase I	NCT02255422 NCT02255435 NCT02065375 NCT02259231 NCT02142959 NCT02029729
CXA-10			
	Acute kidney injury	Phase I	NCT02248051
Minocycline			
	HD Intracerebral hemorrhage Retinitis pigmentosa Intracerebral hemorrhage	Phase II/III Phase I/II Phase I/II Phase I/II	NCT00277355 NCT01805895 NCT02140164 NCT03040128

Chapter 3. Transcription Factor NRF2 as a Therapeutic Target for Chronic Diseases: A Systems Medicine Approach

Table 3.2 continued from previous page


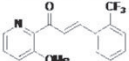
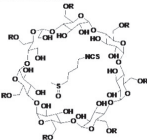
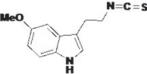
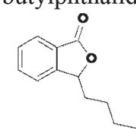
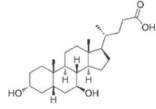
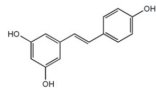
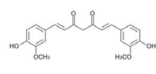
Compound	Disease	Clinical Trial	Reference
<div>Oltipraz</div> 	NASH Schistosomiasis Lung cancer	Phase III Phase I	 Approved NCT00006457
<div>VEDA-1209</div> 	Ulcerative colitis	Preclinical PK	—
	Cystic fibrosis	Phase II	NCT01315665
	Schizophrenia	Phase II/III	NCT02880462
		Phase II	NCT02810964
		Phase II	NCT01716858
	COPD	Phase II	NCT01335971
	Atopic asthmatics	Phase I	NCT01845493
	Autism	Phase II	NCT01474993
		Phase II	NCT02909959
		Phase II	NCT02677051
		Phase II	NCT02654743
		Phase I/II	NCT02561481
	Healthy	Phase I	NCT01008826
		Phase I	NCT02023931
	Melanoma	Phase I	NCT01568996
	Asthma	Phase I	NCT01845493
		Phase I/II	NCT01183923
	Prostate cancer	Phase II	NCT01228084
	Breast cancer	Phase II	NCT00843167
	Lung cancer	Phase II	NCT03232138
	Environmental carcinogenesis	Phase II	NCT01437501
	Alcohol sensitivity	Phase II	NCT01845220
	Aging	Phase II	NCT03126539
	Allergic rhinitis	Phase II	NCT02885025
	Helicobacter pylori infection	Phase IV	NCT03220542
	T2DM	Phase II	NCT02801448
	Head and neck squamous cell carcinoma, head and neck cancer, tobacco-related carcinoma	Phase I	NCT03182959
<div>Sulforadex</div> 	Subarachnoid hemorrhage Breast neoplasm Prostate cancer	Phase II Phase I/II Phase I	NCT02614742 NCT02970682 NCT02055716 NCT01948362
<div>ITH12674</div> 	Brain ischemia	Preclinical PK	—

Table 3.2 continued from previous page

Compound	Disease	Clinical Trial	Reference
DL-3-n-butylphthalide 	Alzheimer disease Acute ischemic stroke Vascular cognitive impairment Cerebrovascular occlusion, collateral blood circulation, anterior cerebral circulation infarction	Phase II Phase I/II Phase II/III Phase IV	NCT02711683 NCT02149875 NCT02993367 NCT02594995
Ursodiol 	Cholestasis HD Barrett esophagus, low-grade dysplasia Chronic hepatitis C T2DM	Phase II/III Phase I Phase II Phase III Phase II	NCT00846963 NCT00514774 NCT01097304 NCT00200343 NCT02033876
Resveratrol 	T2DM Colon cancer COPD Friedreich ataxia NASH Non-ischemic cardiomyopathy Endometriosis Chronic renal insufficiency Metabolic syndrome X Chronic subclinical inflammation AD HD Colorectal cancer	Phase I Phase I Phase III Phase I/II Phase II/III Phase III Phase IV Phase III Phase II Phase III Phase II Phase III Phase III Phase I	NCT01677611 NCT00256334 NCT02245932 NCT01339884 NCT02030977 NCT01914081 NCT02475564 NCT02433925 NCT02114892 NCT01492114 NCT01504854 NCT00743743 NCT02336633 NCT00433576
Curcumin 	T2DM, cardiovascular risk Schizophrenia, cognition, psychosis Acute kidney injury, abdominal aortic aneurysm Chronic kidney diseases, T2DM AD Neoplasms Crohn's disease Chronic schizophrenia Mild cognitive impairment Prostate cancer Major depression	Phase IV Phase I/II Phase II/III Phase II/III Phase I/II Phase II Phase III Phase IV Phase II Phase III Phase IV	NCT01052025 NCT02104752 NCT01225094 NCT03262363 NCT00164749 NCT02944578 NCT02255370 NCT02298985 NCT01811381 NCT02064673 NCT01750359

3.5.2 Protein–Protein Interaction Inhibitors for Nuclear Factor (Erythroid-Derived 2)–Like 2 Activation

To overcome the pitfall of selectivity, a new class of NRF2 inducers that prevent the docking of NRF2 to KEAP1 has emerged [247]. The use of PPI inhibitors has been achieved by the prior elucidation of the X-ray crystal structure of KEAP1 [248] bound to a peptide containing the high-affinity binding ETGE motif of NRF2 [249]. KEAP1

contains a six-bladed b-propeller with specific hydrophobic and hydrophilic residues that participate in the docking of the ETGE motif that adopts a b-hairpin structure. Docking is mainly favored by electrostatic interactions between several arginines of KEAP1 and the two glutamates in the ETGE motif [248, 249]. The docking to KEAP1 of the low-affinity DLG motif of NRF2 has also been characterized [8]. Based on these interactions, peptidomimetic compounds were the first example of PPI inhibitors with significantly improved selectivity over electrophiles [250–252]. These inhibitors show weak activity in cells, and a new provocative strategy has now been reported based on the use of cyclic peptides. One of these peptides exhibited high-binding affinity for KEAP1 and activation of NRF2 and elicited anti-inflammatory effects in mouse macrophages [253].

The discovery of new peptides and small-molecule inhibitors of the KEAP1/NRF2 interaction has been reviewed recently [254, 255]. Briefly, a series of truncated NRF2 peptides was initially evaluated as direct inhibitors of PPI using surface plasmon resonance and fluorescence polarization assays [250]. The minimal peptide sequence with inhibitory capacity was the 9-mer sequence of LDE-ETGE-FL [256, 257]. In parallel, Wells and collaborators [258] searched for new putative peptide ligands using a phage display library combined with high-throughput fluorescence polarization assay. They found that hybrid peptides based upon the ETGE motif of NRF2 and SQSTM1 have superior binding activity to KEAP1 compared with either native peptide alone. To facilitate cellular uptake, a peptide was designed with the ETGE motif fused to the cell transduction domain of the HIV-Tat protein and the cleavage sequence of calpain (DEETGE-Cal-Tat). This peptide showed neuroprotective and cognitive-preserving effects in a mouse model of cerebral ischemia [259].

Table 3.3: Selected NRF2 inducers acting as NRF2–KEAP1 protein–protein interaction inhibitors

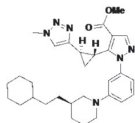
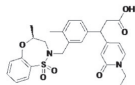
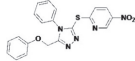
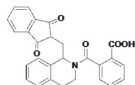
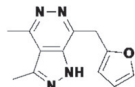
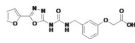
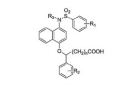
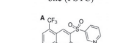
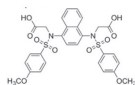
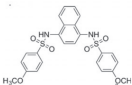
Compound	Patent	Title	Applicant
Compound 7	WO2017060855	Arylcyclohexyl pyrazoles as NRF2 regulators	GlaxoSmithKline Astex Therapeutics
			
Compound 15	WO2016/202253	NRF2 regulators	GlaxoSmithKline Astex Therapeutics, GlaxoSmithKline (China), R&D
			
BM19	WO2014/197818	Small-molecule activators of NRF2 pathway	General Hospital, Regents of the University of California
			
LH601	WO2013/067036	Direct inhibitors of KEAP1–NRF2 interaction as antioxidant inflammation modulators	Rutgers, The State University of New Jersey, Broad Institute
			
	WO2011/156889	Novel modulators of NRF2 and uses thereof	TRT Pharma Gerald Batist, Jian Hui Wu
AN-465/144580	JP2011/0167537	KEAP1 protein-binding compound, crystal of complex between the same and KEAP1 protein, and method for producing the same	Toray Industries
			
3-(2-Oxo-2-phenylethylidene)-2,3,6,7-tetrahydro-1H-pyrazino[2,1-a]isoquinolin-4(11H)-one (Compound I)	WO2017124835	1-sulfonamido-4-aryloxy compound, and preparation method and medicinal application thereof	China Pharmaceutical University
			
3-(pyridin-3-ylmethyl)-5-(trifluoromethyl)-2H-chromen-2-one (PSTC)	WO2015/092713	NRF2 regulators	GlaxoSmithKline Astex Therapeutics Limited
			
Compound 2	WO2011156889A1	Novel regulators of NRF2 and uses thereof	TRT Pharma Gerald Batist, Jian Hui Wu
			

Table 3.3 continued from previous page

Compound	Patent	Title	Applicant
Compound 16 	WO2016/202253	NRF2 regulators	GlaxoSmithKline Astex Therapeutics, GlaxoSmithKline (China), R&D CO., LTD

Five families of PPI inhibitors have been described: tetrahydroisoquinoline [247, 260], thiopyrimidine [251], naphthalene [261], carbazone [262], and urea derivatives [263]. Table 3.3 compiles recent patents addressing these small molecules. Although these compounds are very promising, it is still needed to demonstrate that they are selective for the KEAP1/NRF2 interaction, because KEAP1 also targets at least Bcl2 and IKK [168, 245, 246].

From the large number of compounds indexed in the available libraries, the compounds LH601, benzenesulfonylpyrimidone 2, N-phenyl-benzenesulfonamide, and a series of 1,4-diphenyl-1,2,3-triazoles might be very well-suited candidates to inhibit the PPI with KEAP1 [Wen_et_al_2015 Nasiri_et_al_2016, 250, 260, 264]. These studies described in detail the atomic interaction with KEAP1, the affinity, and the thermodynamics parameters of binding. The therapeutic efficacy of these compounds is to be analyzed in future work in which safety, potency, and blood brain barrier permeability should be addressed.

3.5.3 Drug Targets Other Than Kelch-Like ECH-Associated Protein 1 for Nuclear Factor (Erythroid-Derived 2)-Like 2 Activation

Protein kinase GSK-3 phosphorylates the two serine residues in the DSGIS sequence of NRF2 to generate a phosphorylation-dependent degradation motif or phosphodegron (Fig. 3.2). This phosphodegron is recognized by the E3 ligase adapter b-TrCP, leading to ubiquitindependent proteasomal degradation of NRF2. Therefore, GSK-3 inhibitors should stop NRF2 degradation by preventing the generation of this phosphodegron. GSK-3 is an important kinase in AD and other pathophenotypes. It phosphorylates the cytoskeletal protein tau, facilitating the formation of neurofibrillary

tangles, which are pathologic intracellular aggregates that disturb axonal transport and lead to neuronal death [265]. Therefore, it has been speculated that GSK-3 inhibition might have the double benefit of preventing neurofibrillary tangle formation and NRF2 degradation. Unfortunately, most pipelines for the development of GSK-3 inhibitors have been discontinued due to futility, although in most cases there was not good evidence of target modulation (Palomo and Martinez, 2017).

Conceptually, inhibitors of the b-TrCP-phosphoNRF2 interaction should also lead to NRF2 activation as they should disrupt this branch of NRF2 degradation (Fig. 3.2). The molecular interactions between the b-propeller of b-TrCP and a peptide containing the NRF2 phosphodegron have been resolved by NMR [11]. As it happens for KEAP1/EGTE, the most relevant amino acids appear to be several arginine residues of b-TrCP that interact with the two phosphoserines of the DpSGIpS motif. However, the discovery of small molecules that could inhibit the b-TrCP-phosphoNRF2 interaction is still to come.

Additional strategies have been developed to inhibit the NRF2 repressor BACH1, a bZip protein that makes heterodimers with MAF proteins and blocks expression of ARE genes. Efficient inhibition of BACH1 by the HPP-4382 compound has been described *in vitro* [266], but, prior to a full clinical trial, the safety and efficacy profile of HPP-4382 will have to be demonstrated *in vivo*. Considering that other pathways may also influence NRF2 activity, it is reasonable to speculate that a combinatorial approach will be the best way to activate this transcription factor.

3.5.4 Nuclear Factor (Erythroid-Derived 2) Like 2 Inhibitors

NRF2 has a “dark side” related to its oncogenic activity when constitutively and highly overexpressed. Therefore, NRF2 inhibition has been proposed as a mechanism to sensitize cancer cells to chemotherapeutic drugs or radiotherapy [267]. Two strategies can be envisioned to inhibit NRF2 with small molecules: PPI inhibitors that disrupt the bZip interaction between NRF2 and MAFs, and DNA–protein interaction inhibitors that block binding of NRF2-MAF to the ARE (Fig.

3.9C). Both strategies are hampered by the need by such drugs to overcome the large free energy of association between protein–protein and, to a lesser extent, protein–DNA interfaces. Nevertheless, such drugs have been found for other bZip transcription factors such as STAT3–STAT3, MYC–MAX, and JUNFOS [268], and new small molecules are being described for NRF2–MAF. For instance, malabaricone-A is a pro-oxidant compound that overcomes leukemia resistance by targeting NRF2 [269]. Ascorbic acid (vitamin C), a well-known ROS scavenger, was found to sensitize imatinib-resistant cancer cells by decreasing the levels of the NRF2/ARE complex, reducing the expression of the GCLC gene and dropping GSH levels [270]. All-transretinoic acid is another example of NRF2 inhibitor that significantly decreases NRF2 activation by potent electrophilic NRF2 inducers in vitro and in vivo. It activates the retinoic acid receptor α , which forms a complex with NRF2, hence impeding the binding of the transcription factor to ARE genes [271].

Natural products such as brusatol [272, 273], ochratoxin A [270, 274], and trigonelline [275] have also been found to inhibit NRF2. However, their mechanism of action is not fully understood. In fact, a significant issue related to currently available compounds is the profound off-target effect that they might have. For instance, promising results with brusatol were recently discouraged by the finding that this drug exerts a general and unspecific inhibition of protein synthesis, resulting in the drop of NRF2 levels, but also of many other proteins with rapid turnover [276]. Similarly, the antiprotozoal agent halofuginone, used in veterinary practice, enhances the chemosensitivity of cancer cells by suppressing NRF2 accumulation, but this effect appears to be indirect by inhibiting prolyl-transfer RNA synthesis that is strongly required for ribosomal translation of NRF2 as well as many other proline-containing proteins [277].

A novel approach to identify selective NRF2 inhibitors has been reported recently by the use of quantitative highthroughput screen of small-molecule inhibitors [278]. The authors identified a first-in class compound, termed ML385, which most likely prevented the binding of NRF2 to other bZip coactivators. This compound blocked NRF2 tran-

scriptional activity and sensitized KEAP1-deficient cells to carboplatin and other chemotherapeutic drugs. Additional studies are needed to confirm whether ML385 is selective for NRF2 or if it also inhibits other bZip transcription factors.

In light of the highly favorable systemic effects of NRF2 in various tumor pathophenotypes, a specific targeting of NRF2 with small-molecule inhibitors seems to provide an excellent clinical approach. However, it is necessary to determine whether cancer treatment with NRF2 inhibitors increases the risk of other pathophenotypes of the NRF2 diseasome.

3.5.5 Repurposing Instead of De Novo Drug Discovery and Development

As previously discussed, numerous compounds are under development to provide a benefit for the pathophenotypes associated with the NRF2 diseasome. An alternative approach is to give drugs that are already in clinical use for a certain pathomechanism a new use for the treatment of other pathomechanisms that are connected to NRF2. This section provides the basis for repositioning some commonly used drugs based on their role in NRF2 regulation.

Metformin is the first-line monotherapy for the T2DM. According to Fig. 3.6, it provides therapeutic benefit to the NRF2 subcluster of pathophenotypes related to glucose metabolism. In fact, SFN reduces hepatic glucose production and improves glucose control in patients with T2DM [221]. Interestingly, some evidence suggests that metformin may be effective in preventing other nonglycemic pathophenotypes of the NRF2 diseasome, including cardiovascular (Nesti and Natali, 2017), respiratory [279], digestive (Bauer and Duca, 2016), neurodegenerative [280], autoimmune [281], and neoplastic [282] disorders. The mechanism of action of metformin is not completely clear, but it involves inhibition of mitochondrial complex I, thus increasing the AMP/ATP ratio [283, 284] and leading to activation of the energy sensor AMPK [285, 286]. Importantly, AMPK activates NRF2 [287, 288], and pharmacological targeting of this axis attenuates inflammation after stroke [97] or

endotoxin exposure [289, 290]. Indeed, metformin activates NRF2 in an AMPK-dependent manner, resulting in inhibition of inflammatory responses in preclinical rodent models of transient global cerebral ischemia [291, 292]. Glucose metabolism and inflammation may not be the only pathomechanisms affected by metformin/NRF2. In fact, other salutary effects have been described for redox [293, 294] and protein homeostasis [295].

Statins prevent and reduce cardiovascular pathophenotypes. In addition to a lipid-lowering effect, statins appear to protect against pathomechanisms associated with the NRF2 network such as inflammatory [296–298] and pathologic ROS formation [299]. They are competitive inhibitors of 3-hydroxy-3-methyl-glutaryl-CoA reductase, which catalyzes the rate-limiting reaction in cholesterol synthesis. Other pleiotropic effects include the upregulation of transcription factor Krüppel-like factor 2, which is induced early during progression of cirrhosis and lessens the development of hepatic vascular dysfunction [300]. Recent evidence indicates that at least some statins activate NRF2. In a proteomic study conducted in isolated hepatocytes, high concentrations of simvastatin activated NRF2, probably as a defensive mechanism [301]. The pretreatment of neural stem cells with lovastatin activated the NRF2 pathway and elicited protection against hydrogen peroxide-induced cell death [299]. In liver cirrhosis, simvastatin activates an axis formed by Krüppel-like factor 2 and NRF2 to reduce the oxidative burden and inflammatory response of stellate cells, improving liver fibrosis, endothelial dysfunction, and portal hypertension. The mechanism of activation of NRF2 by simvastatin is not completely clear, but it appears to involve elements found in the NRF2 interactome, such as mitogen-activated protein kinase, PI3K/AKT pathways [302], and GSK-3 [303].

Other cases for drug repurposing can be inferred from the NRF2 interactome of Fig. 3.4, in particular with signaling kinases. As indicated in Fig. 3.2C, GSK-3 phosphorylates the Neh6 domain of NRF2, leading to the recognition by b-TrCP and further ubiquitindependent proteasomal degradation. GSK-3 is active in the absence of stimuli and inactive when signaling cascades that activate AKT and other ki-

nases lead to phosphorylation of GSK-3 at its N-terminal pseudosubstrate domain. It follows that medications known to target signaling kinases may be used to upregulate (GSK-3 inhibitors) or downregulate (PI3K/AKT inhibitors) the NRF2 signature.

GSK-3 participates in at least some pathophenotypes found in the NRF2 diseasome such as diabetes and neurodegeneration [304, 305]. A broad spectrum of GSK-3 inhibitors has been discovered from natural and synthetic origins [306], but probably the best evidence for repurposing a GSK-3 inhibitor to increase NRF2 activity stems from the clinical use of lithium as mood stabilizer [307]. Although bipolar disorder and depression are not found at this time in the NRF2 diseasome, it is becoming evident that they exhibit neuroinflammatory and degenerative pathophenotypes that at least in mouse models imply deregulation of NRF2 [308–310].

The NRF2 interactome also provides a justification for the inhibition of NRF2 by cancer drugs that block signaling kinases, thus activating GSK-3. For example, the epidermal growth factor receptor inhibitor erlotinib leads to NRF2 inhibition, participating in tumor cell sensation in nonsmall cell lung cancer [311]. The kinase cascade inhibitor sorafenib, used in therapy of hepatocellular carcinoma, also leads to inhibition of NRF2 and its downstream targets metallothionein-1 [312] and methylenetetrahydrofolate dehydrogenase 1 [313].

Finally, search for repurposing drugs that might impinge on NRF2 regulation has been done to date in two relevant studies. Using a fluorescence correlation spectroscopy-based screening system, two of 1633 drugs significantly increased NRF2 protein levels in HepG2 cells: chlorophyllin and bonaphton [314]. In another study, the connectivity map database that comprises gene expression profiles for human cell lines treated with 1309 agents [315, 316] was analyzed [317] in search for potential redox regulators through activation of NRF2 [318]. This study found astemizole, a potent antihistamine drug, used in allergic conditions, as a novel NRF2 activator.

3.5.6 Biomarkers as Nuclear Factor (Erythroid-Derived 2)–Like 2 Signature and for Monitoring Target Engagement

3 The evaluation of the redox status in patients or population studies is hampered by the short half-life of ROS, in the range of milliseconds or nanoseconds [319]. Therefore, biomarkers of pathologic ROS formation are based on measuring the traces left by ROS, which are normally terminal oxidation products of cellular molecules, many of them being nonspecific [320]. On the contrary, activation of NRF2 and subsequent expression of its target genes is an indirect but reliable estimation of the total exposure of the organism to pathologic ROS formation. Because NRF2 activation is a well-established cellular response to environmental stressors, it has been considered as biomarker of exposure to xenobiotics. In lung, a data mining of several transcription studies followed by Ingenuity pathway analysis reported that the NRF2 signature is upregulated in healthy smokers, therefore suggesting that NRF2-regulated antioxidant genes play a central role in protection against toxic effects of tobacco smoke [321]. Similarly, the levels of NQO1, an enzyme regulated by NRF2, were 15-fold higher in liver tissue obtained from acetaminophenoverdosed patients [322]. The association between disease and nutrition is frequently based on unreliable self-reporting [323]. Measuring biomarkers of response to nutrients, supposedly having beneficial effects by activating NRF2, could provide a reliable method to validate nutritional studies. However, this possibility is still unexplored.

The changes associated with NRF2 transcription could be useful as biomarker for monitoring the efficacy of drugs aimed at reducing pathologic ROS formation by inhibitors of xanthine oxidase and NADPH oxidase. Similarly, exposure to environmental chemicals could be detected and monitored by defining a global protein and gene expression profile [324]. This approach is similar to the use of phase I drug metabolizing enzymes, in which cytochrome P450, which is induced by various xenobiotics through the Ah receptor, can be used as indicator of marine pollution [325]. Daily oral administration of fumaric acid esters over 12 weeks was associated with the increased expression of NRF2 target genes in the skin of patients with psoriasis [326]. Similarly, a fivefold increase in the mRNA level of NQO1 has

been reported in peripheral blood mononuclear cells obtained from cancer patients that received a daily dose of CDDO-Me for 3 weeks [327].

The use of the transcriptional signature of NRF2 as a biomarker requires a good knowledge of the mechanisms involved in activation of ARE genes, as most of the NRF2 targets are regulated by additional transcription factors. It is therefore important to analyze the expression of several ARE genes. For instance, a study using NRF2 as predictor for response to treatment in lung squamous cell carcinoma has proposed the use of 28 genes to define a NRF2 activation profile [328].

3.6 Conclusions

Systems medicine together with network pharmacology highlights a cluster of chronic disease pathophenotypes in which NRF2 plays a fundamental role. These diseases share common mechanisms, including oxidative, inflammatory, and metabolic alterations. The NRF2 interactome, the NRF2 diseasome, and the NRF2 drugome presented in this work are still in an early stage of development, but they represent a first attempt to structure NRF2 as a common therapeutic and systems medicine approach. The forthcoming refinement of current databases and upcoming clinical outcome data will further improve the accuracy of this new approach to pharmacology and mechanism-based drug repurposing. This paper provides a road map for a comprehensive strategy for drug discovery to activate or inhibit NRF2 and highlights the need of translational efforts toward the development of de novo drugs or the repurposing of drugs that target NRF2 as a common element in chronic diseases.

3.7 References

- [1] K.-I. Goh et al. "The human disease network". In: *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA* 104.21 (2007), 8685–8690. DOI: {10.1073/pnas.0701361104}.
- [2] Q. Ma. "Role of Nrf2 in Oxidative Stress and Toxicity". In: *ANNUAL REVIEW OF PHARMACOLOGY AND TOXICOLOGY, VOL 53, 2013*. Ed. by Insel, PA. Vol. 53. Annual Review of Pharmacology and Toxicology. 2013, 401+. DOI: {10.1146/annurev-pharmtox-011112-140320}.
- [3] J. D. Hayes and A. T. Dinkova-Kostova. "The Nrf2 regulatory network provides an interface between redox and intermediary metabolism". In: *TRENDS IN BIOCHEMICAL SCIENCES* 39.4 (2014), 199–218. DOI: {10.1016/j.tibs.2014.02.002}.
- [4] M. Pajares et al. "Redox control of protein degradation". In: *REDOX BIOLOGY* 6 (2015), 409–420. DOI: {10.1016/j.redox.2015.07.003}.
- [5] M. Pajares et al. "Transcription factor NFE2L2/NRF2 is a regulator of macroautophagy genes". In: *AUTOPHAGY* 12.10 (2016), 1902–1916. DOI: {10.1080/15548627.2016.1208889}.
- [6] M. Pajares, A. Cuadrado, and A. I. Rojo. "Modulation of proteostasis by transcription factor NRF2 and impact in neurodegenerative diseases". In: *REDOX BIOLOGY* 11 (2017), 543–553. DOI: {10.1016/j.redox.2017.01.006}.
- [7] M. R. D. L. Vega et al. "Role of Nrf2 and Autophagy in Acute Lung Injury". In: *Current Pharmacology Reports* 2.2 (2016), pp. 91–101. DOI: 10.1007/s40495-016-005302.
- [8] K. I. Tong et al. "Different electrostatic Potentials define ETGE and DLG motifs as hinge and latch in oxidative stress response". In: *MOLECULAR AND CELLULAR BIOLOGY* 27.21 (2007), 7511–7521. DOI: {10.1128/MCB.00753-07}.

-
- [9] A. Cuadrado. "Structural and functional characterization of Nrf2 degradation by glycogen synthase kinase 3/beta-TrCP". In: *FREE RADICAL BIOLOGY AND MEDICINE* 88.B (2015), 147–157. DOI: {10.1016/j.freeradbiomed.2015.04.029}.
 - [10] P. Rada et al. "SCF/beta-TrCP Promotes Glycogen Synthase Kinase 3-Dependent Degradation of the Nrf2 Transcription Factor in a Keap1-Independent Manner". In: *MOLECULAR AND CELLULAR BIOLOGY* 31.6 (2011), 1121–1133. DOI: {10.1128/ MCB.01204-10}.
 - [11] P. Rada et al. "Structural and Functional Characterization of Nrf2 Degradation by the Glycogen Synthase Kinase 3/beta-TrCP Axis". In: *MOLECULAR AND CELLULAR BIOLOGY* 32.17 (2012), 3486–3499. DOI: {10.1128/MCB.00180-12}.
 - [12] S. Chowdhry et al. "Nrf2 is controlled by two distinct beta-TrCP recognition motifs in its Neh6 domain, one of which can be modulated by GSK-3 activity". In: *ONCOGENE* 32.32 (2013), 3765–3781. DOI: {10.1038/onc.2012.388}.
 - [13] H.-Y. Cho, J. Marzec, and S. R. Kleeberger. "Functional polymorphisms in Nrf2: implications for human disease". In: *FREE RADICAL BIOLOGY AND MEDICINE* 88.B (2015), 362–372. DOI: {10.1016/j.freeradbiomed.2015.06.012}.
 - [14] A.-L. Barabasi, N. Gulbahce, and J. Loscalzo. "Network medicine: a network-based approach to human disease". In: *NATURE REVIEWS GENETICS* 12.1 (2011), 56–68. DOI: {10.1038/nrg2918}.
 - [15] M. Vidal, M. E. Cusick, and A.-L. Barabasi. "Interactome Networks and Human Disease". In: *CELL* 144.6 (2011), 986–998. DOI: {10.1016/j.cell.2011.02.016}.
 - [16] E. Guney et al. "Network-based in silico drug efficacy screening". In: *NATURE COMMUNICATIONS* 7 (2016). DOI: {10.1038/ncomms10331}.
 - [17] J. Menche et al. "Uncovering disease-disease relationships through the incomplete interactome". In: *SCIENCE* 347.6224 (2015). DOI: {10.1126/science.1257601}.

- [18] M. Kitsak et al. "Tissue Specificity of Human Disease Module". In: *SCIENTIFIC REPORTS* 6 (2016). DOI: {10.1038/srep35241}.
- [19] D. Tuerei et al. "NRF2-ome: An Integrated Web Resource to Discover Protein Interaction and Regulatory Networks of NRF2". In: *OXIDATIVE MEDICINE AND CELLULAR LONGEVITY* (2013). DOI: {10.1155/2013/737591}.
- [20] *Neighborhood of NFE2L2 in the human interactome*. <http://sbi.imim.es/data/nrf2/>. Accessed: 2021-3-30.
- [21] J. Pinero et al. "DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants". In: *NUCLEIC ACIDS RESEARCH* 45.D1 (2017), D833–D839. DOI: {10.1093/nar/gkw943}.
- [22] G. Stelzer et al. "The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses". In: *Current Protocols in Bioinformatics* (2016). DOI: 10.1002/cpbi.5.
- [23] E. Guney and B. Oliva. "Analysis of the Robustness of Network-Based Disease-Gene Prioritization Methods Reveals Redundancy in the Human Interactome and Functional Diversity of Disease-Genes". In: *PLOS ONE* 9.4 (2014). DOI: {10.1371/journal.pone.0094686}.
- [24] C. A. Hidalgo et al. "A Dynamic Network Approach for the Study of Human Phenotypes". In: *PLOS COMPUTATIONAL BIOLOGY* 5.4 (2009). DOI: {10.1371/journal.pcbi.1000353}.
- [25] X. Zhou et al. "Human symptoms-disease network". In: *NATURE COMMUNICATIONS* 5 (2014). DOI: {10.1038/ncomms5212}.
- [26] A. I. Rojo et al. "Nrf2 Regulates Microglial Dynamics and Neuroinflammation in Experimental Parkinson's Disease". In: *GLIA* 58.5 (2010), 588–598. DOI: {10.1002/glia.20947}.

-
- [27] I. Lastres-Becker et al. "alpha-Synuclein expression and Nrf2 deficiency cooperate to aggravate protein aggregation, neuronal death and inflammation in early-stage Parkinson's disease". In: *HUMAN MOLECULAR GENETICS* 21.14 (2012), 3173–3192. DOI: {10.1093/hmg/dds143}.
 - [28] I. Lastres-Becker et al. "Fractalkine activates NRF2/NFE2L2 and heme oxygenase 1 to restrain tauopathy-induced microgliosis". In: *BRAIN* 137.1 (2014), 78–91. DOI: {10.1093/brain/awt323}.
 - [29] K.-A. Jung, S. Lee, and M.-K. Kwak. "NFE2L2/NRF2 Activity Is Linked to Mitochondria and AMP-Activated Protein Kinase Signaling in Cancers Through miR-181c/Mitochondria-Encoded Cytochrome c Oxidase Regulation". In: *ANTIOXIDANTS & REDOX SIGNALING* 27.13 (2017), 945–961. DOI: {10.1089/ars.2016.6797}.
 - [30] X. Wang et al. "Tenuigenin inhibits LPS-induced inflammatory responses in microglia via activating the Nrf2-mediated HO-1 signaling pathway". In: *EUROPEAN JOURNAL OF PHARMACOLOGY* 809 (2017), 196–202. DOI: {10.1016/j.ejphar.2017.05.004}.
 - [31] C. Gonzalez-Donquiles et al. "The NRF2 transcription factor plays a dual role in colorectal cancer: A systematic review". In: *PLOS ONE* 12.5 (2017). DOI: {10.1371/journal.pone.0177549}.
 - [32] K. Lu et al. "NRF2 Induction Supporting Breast Cancer Cell Survival Is Enabled by Oxidative Stress-Induced DPP3-KEAP1 Interaction". In: *CANCER RESEARCH* 77.11 (2017), 2881–2892. DOI: {10.1158/0008-5472.CAN-16-2204}.
 - [33] P. Wenzel et al. "Redox regulation of cardiovascular inflammation - Immunomodulatory function of mitochondrial and Nox-derived reactive oxygen and nitrogen species". In: *FREE RADICAL BIOLOGY AND MEDICINE* 109.SI (2017), 48–60. DOI: {10.1016/j.freeradbiomed.2017.01.027}.

- [34] D.-F. Lee et al. "KEAP1 E3 Ligase-Mediated Downregulation of NF-kappa B Signaling by Targeting IKK beta". In: *MOLECULAR CELL* 36.1 (2009), 131–140. DOI: {10.1016/j.molcel.2009.07.025}.
- [35] A. Cuadrado et al. "Transcription Factors NRF2 and NF-kappa B Are Coordinated Effectors of the Rho Family, GTP-binding Protein RAC1 during Inflammation". In: *JOURNAL OF BIOLOGICAL CHEMISTRY* 289.22 (2014), 15244–15258. DOI: {10.1074/jbc.M113.540633}.
- [36] R. K. Thimmulappa et al. "Nrf2-dependent protection from LPS induced inflammatory response and mortality by CDDO-imidazolide". In: *BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS* 351.4 (2006), 883–889. DOI: {10.1016/j.bbrc.2006.10.102}.
- [37] L. K. Sha et al. "Loss of Nrf2 in bone marrow-derived macrophages impairs antigen-driven CD8(+) T cell function by limiting GSH and Cys availability". In: *FREE RADICAL BIOLOGY AND MEDICINE* 83 (2015), 77–88. DOI: {10.1016/j.freeradbiomed.2015.02.004}.
- [38] B. Bruene et al. "Redox Control of Inflammation in Macrophages". In: *ANTIOXIDANTS & REDOX SIGNALING* 19.6 (2013), 595–637. DOI: {10.1089/ars.2012.4785}.
- [39] M. Diotallevi et al. "Glutathione Fine-Tunes the Innate Immune Response toward Antiviral Pathways in a Macrophage Cell Line Independently of Its Antioxidant Properties". In: *FRONTIERS IN IMMUNOLOGY* 8 (2017). DOI: {10.3389/fimmu.2017.01239}.
- [40] S. A. Rushworth et al. "The high Nrf2 expression in human acute myeloid leukemia is driven by NF-kappa B and underlies its chemo-resistance". In: *BLOOD* 120.26 (2012), 5188–5198. DOI: {10.1182/blood-2012-04-422121}.
- [41] G.-H. Liu, J. Qu, and X. Shen. "NF-kappa B/p65 antagonizes Nrf2-ARE pathway by depriving CBP from Nrf2 and facilitating recruitment of HDAC3 to MafK". In: *BIOCHIMICA ET BIOPHYSICA ACTA-MOLECULAR CELL RESEARCH* 1783.5 (2008), 713–727. DOI: {10.1016/j.bbamcr.2008.01.002}.

-
- [42] M. Yu et al. "Nuclear factor p65 interacts with Keap1 to repress the Nrf2-ARE pathway". In: *CELLULAR SIGNALLING* 23.5 (2011), 883–892. DOI: {10.1016/j.cellsig.2011.01.014}.
 - [43] J. T. Winston et al. "The SCF beta-TRCP-ubiquitin ligase complex associates specifically with phosphorylated destruction motifs in I kappa B alpha and beta-catenin and stimulates I kappa B alpha ubiquitination in vitro". In: *GENES & DEVELOPMENT* 13.3 (1999), 270–283. DOI: {10.1101/gad.13.3.270}.
 - [44] E. H. Kobayashi et al. "Nrf2 suppresses macrophage inflammatory response by blocking proinflammatory cytokine transcription". In: *NATURE COMMUNICATIONS* 7 (2016). DOI: {10.1038/ncomms11624}.
 - [45] C. J. Harvey et al. "Targeting Nrf2 Signaling Improves Bacterial Clearance by Alveolar Macrophages in Patients with COPD and in a Mouse Model". In: *SCIENCE TRANSLATIONAL MEDICINE* 3.78 (2011). DOI: {10.1126/scitranslmed.3002042}.
 - [46] T. Ishii and G. E. Mann. "Redox status in mammalian cells and stem cells during culture in vitro: Critical roles of Nrf2 and cystine transporter activity in the maintenance of redox balance". In: *REDOX BIOLOGY* 2 (2014), 786–794. DOI: {10.1016/j.redox.2014.04.008}.
 - [47] R. Saddawi-Konefka et al. "Nrf2 Induces IL-17D to Mediate Tumor and Virus Surveillance". In: *CELL REPORTS* 16.9 (2016), 2348–2358. DOI: {10.1016/j.celrep.2016.07.075}.
 - [48] P. Wenzel et al. "Heme oxygenase-1 suppresses a pro-inflammatory phenotype in monocytes and determines endothelial function and arterial hypertension in mice and humans". In: *EUROPEAN HEART JOURNAL* 36.48 (2015), 3437–3446. DOI: {10.1093/eurheartj/ehv544}.
 - [49] E. Bourdonnay et al. "Redox-Sensitive Regulation of Gene Expression in Human Primary Macrophages Exposed to Inorganic Arsenic". In: *JOURNAL OF CELLULAR BIOCHEMISTRY* 107.3 (2009), 537–547. DOI: {10.1002/jcb.22155}.

- [50] T. Satoh, S. R. McKercher, and S. A. Lipton. “Nrf2/ARE-mediated antioxidant actions of pro-electrophilic drugs”. In: *FREE RADICAL BIOLOGY AND MEDICINE* 65 (2013), 645–657. DOI: {10.1016/j.freeradbiomed.2013.07.022}.
- [51] X. Kong et al. “Enhancing Nrf2 Pathway by Disruption of Keap1 in Myeloid Leukocytes Protects against Sepsis”. In: *AMERICAN JOURNAL OF RESPIRATORY AND CRITICAL CARE MEDICINE* 184.8 (2011), 928–938. DOI: {10.1164/rccm.201102-0271OC}.
- [52] W. Lin et al. “Sulforaphane suppressed LPS-induced inflammation in mouse peritoneal macrophages through Nrf2 dependent pathway”. In: *BIOCHEMICAL PHARMACOLOGY* 76.8 (2008), 967–973. DOI: {10.1016/j.bcp.2008.07.036}.
- [53] T. Arisawa et al. “Nrf2 gene promoter polymorphism and gastric carcinogenesis”. In: *HEPATO-GASTROENTEROLOGY* 55.82-83 (2008), 750–754.
- [54] T. Arisawa et al. “The relationship between Helicobacter pylori infection and promoter polymorphism of the Nrf2 gene in chronic gastritis”. In: *INTERNATIONAL JOURNAL OF MOLECULAR MEDICINE* 19.1 (2007), 143–148.
- [55] L. Thomson et al. “Immunoglobulins Against Tyrosine-Nitrated Epitopes in Coronary Artery Disease”. In: *CIRCULATION* 126.20 (2012), 2392+. DOI: {10.1161/CIRCULATIONAHA.112.103796}.
- [56] L. M. A. Al-Huseini et al. “Nuclear Factor-erythroid 2 (NF-E2) p45-related Factor-2 (Nrf2) Modulates Dendritic Cell Immune Function through Regulation of p38 MAPK-cAMP-responsive Element Binding Protein/Activating Transcription Factor 1 Signaling”. In: *JOURNAL OF BIOLOGICAL CHEMISTRY* 288.31 (2013), 22281–22288. DOI: {10.1074/jbc.M113.483420}.
- [57] D. A. Johnson et al. “The Absence of the Pro-antioxidant Transcription Factor Nrf2 Exacerbates Experimental Autoimmune Encephalomyelitis”. In: *TOXICOLOGICAL SCIENCES* 114.2 (2010), 237–246. DOI: {10.1093/toxsci/kfp274}.

-
- [58] S. Tzima et al. "Myeloid heme oxygenase-1 regulates innate immunity and autoimmunity by modulating IFN-beta production". In: *JOURNAL OF EXPERIMENTAL MEDICINE* 206.5 (2009), 1167–1179. DOI: {10.1084/jem.20081582}.
 - [59] I. Buendia et al. "Nrf2-ARE pathway: An emerging target against oxidative stress and neuroinflammation in neurodegenerative diseases". In: *PHARMACOLOGY & THERAPEUTICS* 157 (2016), 84–104. DOI: {10.1016/j.pharmthera.2015.11.003}.
 - [60] S. Licht-Mayer et al. "Cell type-specific Nrf2 expression in multiple sclerosis lesions". In: *ACTA NEUROPATHOLOGICA* 130.2 (2015), 263–277. DOI: {10.1007/s00401-015-1452-x}.
 - [61] J. van Horssen et al. "Nrf2 and DJ1 are consistently upregulated in inflammatory multiple sclerosis lesions". In: *FREE RADICAL BIOLOGY AND MEDICINE* 49.8 (2010), 1283–1289. DOI: {10.1016/j.freeradbiomed.2010.07.013}.
 - [62] P. Fagone et al. "Heme oxygenase-1 expression in peripheral blood mononuclear cells correlates with disease activity in multiple sclerosis". In: *JOURNAL OF NEUROIMMUNOLOGY* 261.1-2 (2013), 82–86. DOI: {10.1016/j.jneuroim.2013.04.013}.
 - [63] E. Croze et al. "Interferon-beta-1b-induced short- and long-term signatures of treatment activity in multiple sclerosis". In: *PHARMACOGENOMICS JOURNAL* 13.5 (2013), 443–451. DOI: {10.1038/tpj.2012.27}.
 - [64] J. Li, T. D. Stein, and J. A. Johnson. "Genetic dissection of systemic autoimmune disease in Nrf2-deficient mice". In: *PHYSIOLOGICAL GENOMICS* 18.3 (2004), 261–272. DOI: {10.1152/physiolgenomics.00209.2003}.
 - [65] Q. Ma, L. Battelli, and A. F. Hubbs. "Multiorgan autoimmune inflammation, enhanced lymphoproliferation, and impaired homeostasis of reactive oxygen species in mice lacking the antioxidant-activated transcription factor Nrf2". In: *AMERICAN JOURNAL OF PATHOLOGY* 168.6 (2006), 1960–1974. DOI: {10.2353/ajpath.2006.051113}.

- [66] M. Zhao et al. "Nuclear Factor Erythroid 2-related Factor 2 Deficiency Exacerbates Lupus Nephritis in B6/lpr mice by Regulating Th17 Cell Function". In: *SCIENTIFIC REPORTS* 6 (2016). DOI: {10.1038/srep38619}.
- [67] M. D. Evans et al. "Aberrant processing of oxidative DNA damage in systemic lupus erythematosus". In: *BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS* 273.3 (2000), 894–898. DOI: {10.1006/bbrc.2000.3078}.
- [68] P. E. Morgan et al. "Serum protein oxidation and apolipoprotein CIII levels in people with systemic lupus erythematosus with and without nephritis". In: *FREE RADICAL RESEARCH* 41.12 (2007), 1301–1312. DOI: {10.1080/10715760701684809}.
- [69] J. Frostegard et al. "Lipid peroxidation is enhanced in patients with systemic lupus erythematosus and is associated with arterial and renal disease manifestations". In: *ARTHRITIS AND RHEUMATISM* 52.1 (2005), 192–200. DOI: {10.1002/art.20780}.
- [70] E. J. Cordova et al. "The NRF2 gene variant,-653G/A, is associated with nephritis in childhood-onset systemic lupus erythematosus". In: *LUPUS* 19.10 (2010), 1237–1242. DOI: {10.1177/0961203310367917}.
- [71] C. A. Roberts, A. K. Dickinson, and L. S. Taams. "The Interplay Between Monocytes/Macrophages and CD4(+) T Cell Subsets in Rheumatoid Arthritis". In: *FRONTIERS IN IMMUNOLOGY* 6 (2015), 1–19. DOI: {10.3389/fimmu.2015.00571}.
- [72] N. Maicas et al. "Deficiency of Nrf2 Accelerates the Effector Phase of Arthritis and Aggravates Joint Disease". In: *ANTIOXIDANTS & REDOX SIGNALING* 15.4 (2011), 889–901. DOI: {10.1089/ars.2010.3835}.
- [73] W. Wu et al. "Inflammatory Stress Sensitizes the Liver to Atorvastatin-Induced Injury in ApoE-/- Mice". In: *PLOS ONE* 11.7 (2016). DOI: {10.1371/journal.pone.0159512}.

-
- [74] S. Y. Park et al. "Cilostazol Enhances Apoptosis of Synovial Cells From Rheumatoid Arthritis Patients With Inhibition of Cytokine Formation via Nrf2-Linked Heme Oxygenase 1 Induction". In: *ARTHRITIS AND RHEUMATISM* 62.3 (2010), 732–741. DOI: {10.1002/art.27291}.
 - [75] S.-J. Moon et al. "Rebamipide Suppresses Collagen-Induced Arthritis Through Reciprocal Regulation of Th17/Treg Cell Differentiation and Heme Oxygenase 1 Induction". In: *ARTHRITIS & RHEUMATOLOGY* 66.4 (2014), 874–885. DOI: {10.1002/art.38310}.
 - [76] S. Datta et al. "Correlation of oxidant status with oxidative tissue damage in patients with rheumatoid arthritis". In: *CLINICAL RHEUMATOLOGY* 33.11 (2014), 1557–1564. DOI: {10.1007/s10067-014-2597-z}.
 - [77] H. Xie et al. "Vitiligo: How do oxidative stress-induced autoantigens trigger autoimmunity?" In: *JOURNAL OF DERMATOLOGICAL SCIENCE* 81.1 (2016), 3–9. DOI: {10.1016/j.jdermsci.2015.09.003}.
 - [78] C.-P. Guan et al. "The susceptibility to vitiligo is associated with NF-E2-related factor2 (Nrf2) gene polymorphisms: a study on Chinese Han population". In: *EXPERIMENTAL DERMATOLOGY* 17.12 (2008), 1059–1062. DOI: {10.1111/j.1600-0625.2008.00752.x}.
 - [79] P. Song et al. "Genetic polymorphism of the Nrf2 promoter region is associated with vitiligo risk in Han Chinese populations". In: *JOURNAL OF CELLULAR AND MOLECULAR MEDICINE* 20.10 (2016), 1840–1850. DOI: {10.1111/jcmm.12874}.
 - [80] V. T. Natarajan et al. "Transcriptional Upregulation of Nrf2-Dependent Phase II Detoxification Genes in the Involved Epidermis of Vitiligo Vulgaris". In: *JOURNAL OF INVESTIGATIVE DERMATOLOGY* 130.12 (2010), 2781–2789. DOI: {10.1038/jid.2010.201}.

- [81] R. Finkelstein et al. "ALVEOLAR INFLAMMATION AND ITS RELATION TO EMPHYSEMA IN SMOKERS". In: *AMERICAN JOURNAL OF RESPIRATORY AND CRITICAL CARE MEDICINE* 152.5 (1995), 1666–1672. DOI: {10.1164/ajrccm.152.5.7582312}.
- [82] T. Rangasamy et al. "Genetic ablation of Nrf2 enhances susceptibility to cigarette smoke-induced emphysema in mice". In: *JOURNAL OF CLINICAL INVESTIGATION* 114.9 (2004), 1248–1259. DOI: {10.1172/JCI200421146}.
- [83] D. Goven et al. "Altered Nrf2/Keap1-Bach1 equilibrium in pulmonary emphysema". In: *THORAX* 63.10 (2008), 916–924. DOI: {10.1136/thx.2007.091181}.
- [84] T. Yamamoto et al. "Identification of polymorphisms in the promoter region of the human NRF2 gene". In: *BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS* 321.1 (2004), 72–79. DOI: {10.1016/j.bbrc.2004.06.112}.
- [85] C.-C. Hua et al. "Functional haplotypes in the promoter region of transcription factor Nrf2 in chronic obstructive pulmonary disease". In: *DISEASE MARKERS* 28.3 (2010), 185–193. DOI: {10.1155/2010/143142}.
- [86] J. M. Marzec et al. "Functional polymorphisms in the transcription factor NRF2 in humans increase the risk of acute lung injury". In: *FASEB JOURNAL* 21.9 (2007), 2237–2246. DOI: {10.1096/fj.06-7759com}.
- [87] H.-Y. CHO et al. "The transcription factor NRF2 protects against pulmonary fibrosis". In: *The FASEB Journal* 18.11 (2004), 1258–1260. DOI: 10.1096/fj.03-1127fje.
- [88] P. Markart et al. "Alveolar Oxidative Stress is Associated with Elevated Levels of Nonenzymatic Low-Molecular-Weight Antioxidants in Patients with Different Forms of Chronic Fibrosing Interstitial Lung Diseases". In: *ANTIOXIDANTS & REDOX SIGNALING* 11.2 (2009), 227–240. DOI: {10.1089/ars.2008.2105}.

-
- [89] E. Artaud-Macari et al. "Nuclear Factor Erythroid 2-Related Factor 2 Nuclear Translocation Induces Myofibroblastic Dedifferentiation in Idiopathic Pulmonary Fibrosis". In: *ANTIOXIDANTS & REDOX SIGNALING* 18.1 (2013), 66–79. DOI: {10.1089/ars.2011.4240}.
 - [90] M.-L. Kruse et al. "Colonic Lamina Propria Inflammatory Cells from Patients with IBD Induce the Nuclear Factor-E2 Related Factor-2 Thereby Leading to Greater Proteasome Activity and Apoptosis Protection in Human Colonocytes". In: *INFLAMMATORY BOWEL DISEASES* 22.11 (2016), 2593–2606. DOI: {10.1097/MIB.0000000000000925}.
 - [91] J. K. Baillie et al. "Analysis of the human monocyte-derived macrophage transcriptome and response to lipopolysaccharide provides new insights into genetic aetiology of inflammatory bowel disease". In: *PLOS GENETICS* 13.3 (2017). DOI: {10.1371/journal.pgen.1006641}.
 - [92] T. Arisawa et al. "Nrf2 gene promoter polymorphism is associated with ulcerative colitis in a Japanese population". In: *HEPATO-GASTROENTEROLOGY* 55.82-83 (2008), 394–397.
 - [93] R. M. Jones et al. "Lactobacilli Modulate Epithelial Cytoprotection through the Nrf2 Pathway". In: *CELL REPORTS* 12.8 (2015), 1217–1225. DOI: {10.1016/j.celrep.2015.07.042}.
 - [94] J. C. Onyiah et al. "Herne oxygenase-1 and carbon monoxide regulate intestinal homeostasis and mucosal immune responses to the enteric microbiota". In: *GUT MICROBES* 5.2 (2014), 220–224. DOI: {10.4161/gmic.27290}.
 - [95] S. Orena et al. "Extracts of Fruits and Vegetables Activate the Antioxidant Response Element in IMR-32 Cells". In: *JOURNAL OF NUTRITION* 145.9 (2015), 2006–2011. DOI: {10.3945/jn.115.216705}.
 - [96] U. Wasik et al. "Protection against oxidative stress mediated by the Nrf2/Keap1 axis is impaired in Primary Biliary Cholangitis". In: *SCIENTIFIC REPORTS* 7 (2017). DOI: {10.1038/srep44769}.

- [97] X. Wang et al. "Gliptins Suppress Inflammatory Macrophage Activation to Mitigate Inflammation, Fibrosis, Oxidative Stress, and Vascular Dysfunction in Models of Nonalcoholic Steatohepatitis and Liver Fibrosis". In: *ANTIOXIDANTS & REDOX SIGNALING* 28.2 (2018), 87–109. DOI: {10.1089/ars.2016.6953}.
- [98] I. Reccia et al. "Non-alcoholic fatty liver disease: A sign of systemic disease". In: *METABOLISM-CLINICAL AND EXPERIMENTAL* 72 (2017), 94–108. DOI: {10.1016/j.metabol.2017.04.011}.
- [99] P. Ramadori et al. "Hepatocyte-specific Keap1 deletion reduces liver steatosis but not inflammation during non-alcoholic steatohepatitis development". In: *FREE RADICAL BIOLOGY AND MEDICINE* 91 (2016), 114–126. DOI: {10.1016/j.freeradbiomed.2015.12.014}.
- [100] R. Shimozone et al. "Nrf2 Activators Attenuate the Progression of Nonalcoholic Steatohepatitis-Related Fibrosis in a Dietary Rat Model". In: *MOLECULAR PHARMACOLOGY* 84.1 (2013), 62–70. DOI: {10.1124/mol.112.084269}.
- [101] Y. Takahashi et al. "Does Hepatic Oxidative Stress Enhance Activation of Nuclear Factor-E2-Related Factor in Patients with Nonalcoholic Steatohepatitis?" In: *ANTIOXIDANTS & REDOX SIGNALING* 20.3 (2014), 538–543. DOI: {10.1089/ars.2013.5470}.
- [102] D. Harrison et al. "Role of oxidative stress in atherosclerosis". In: *AMERICAN JOURNAL OF CARDIOLOGY* 91.3, S (2003). Symposium on Biology of Atherosclerosis and Its Implications for Treatment Strategies, ATLANTA, GEORGIA, MAR, 2002, 7A–11A.
- [103] D. Jay, H. Hitomi, and K. K. Griendling. "Oxidative stress and diabetic cardiovascular complications". In: *FREE RADICAL BIOLOGY AND MEDICINE* 40.2 (2006), 183–192. DOI: {10.1016/j.freeradbiomed.2005.06.018}.

-
- [104] G. Shanmugam et al. "Constitutive activation of Nrf2 induces a stable reductive state in the mouse myocardium". In: *REDOX BIOLOGY* 12 (2017), 937–945. DOI: {10.1016/j.redox.2017.04.038}.
 - [105] J. Strom and Q. M. Chen. "Loss of Nrf2 promotes rapid progression to heart failure following myocardial infarction". In: *TOXICOLOGY AND APPLIED PHARMACOLOGY* 327 (2017), 52–58. DOI: {10.1016/j.taap.2017.03.025}.
 - [106] G. Shanmugam et al. "Chronic Endurance Exercise Impairs Cardiac Structure and Function in Middle-Aged Mice with Impaired Nrf2 Signaling". In: *FRONTIERS IN PHYSIOLOGY* 8 (2017). DOI: {10.3389/fphys.2017.00268}.
 - [107] H. M. Nef et al. "Expression profiling of cardiac genes in Tako-Tsubo cardiomyopathy: Insight into a new cardiac entity". In: *JOURNAL OF MOLECULAR AND CELLULAR CARDIOLOGY* 44.2 (2008), 395–404. DOI: {10.1016/j.yjmcc.2007.10.015}.
 - [108] L. M. Pedruzzi et al. "Systemic inflammation and oxidative stress in hemodialysis patients are associated with down-regulation of Nrf2". In: *JOURNAL OF NEPHROLOGY* 28.4 (2015), 495–501. DOI: {10.1007/s40620-014-0162-0}.
 - [109] Y. Shimoyama et al. "Polymorphism of Nrf2, an Antioxidative Gene, is Associated with Blood Pressure and Cardiovascular Mortality in Hemodialysis Patients". In: *INTERNATIONAL JOURNAL OF MEDICAL SCIENCES* 11.7 (2014), 726–731. DOI: {10.7150/ijms.8590}.
 - [110] T. Jansen et al. "Conversion of biliverdin to bilirubin by biliverdin reductase contributes to endothelial cell protection by heme oxygenase-1-evidence for direct and indirect antioxidant actions of bilirubin". In: *JOURNAL OF MOLECULAR AND CELLULAR CARDIOLOGY* 49.2 (2010), 186–195. DOI: {10.1016/j.yjmcc.2010.04.011}.
 - [111] R. Stocker et al. "Bilirubin is an antioxidant of possible physiological importance". In: *Science* 235.4792 (1987), 1043–1046. DOI: 10.1126/science.3029864.

- [112] H. Christou et al. "Prevention of hypoxia-induced pulmonary hypertension by enhancement of endogenous heme oxygenase-1 in the rat". In: *CIRCULATION RESEARCH* 86.12 (2000), 1224–1229. DOI: {10.1161/01.RES.86.12.1224}.
- [113] A. L. Kruger et al. "Up-regulation of heme oxygenase provides vascular protection in an animal model of diabetes through its antioxidant and antiapoptotic effects". In: *JOURNAL OF PHARMACOLOGY AND EXPERIMENTAL THERAPEUTICS* 319.3 (2006), 1144–1152. DOI: {10.1124/jpet.106.107482}.
- [114] P. Wenzel et al. "Heme oxygenase-1 - A novel key player in the development of tolerance in response to organic nitrates". In: *ARTERIOSCLEROSIS THROMBOSIS AND VASCULAR BIOLOGY* 27.8 (2007), 1729–1735. DOI: {10.1161/ATVBAHA.107.143909}.
- [115] P. Hopkins et al. "Higher serum bilirubin is associated with decreased risk for early familial coronary artery disease". In: *ARTERIOSCLEROSIS THROMBOSIS AND VASCULAR BIOLOGY* 16.2 (1996), 250–255. DOI: {10.1161/01.ATV.16.2.250}.
- [116] J. Y. Kwak et al. "BILIRUBIN INHIBITS THE ACTIVATION OF SUPEROXIDE-PRODUCING NADPH OXIDASE IN A NEUTROPHIL CELL-FREE SYSTEM". In: *BIOCHIMICA ET BIOPHYSICA ACTA* 1076.3 (1991), 369–373. DOI: {10.1016/0167-4838(91)90478-I}.
- [117] S. S. Signorelli et al. "Plasma heme oxygenase-1 is decreased in peripheral artery disease patients". In: *MOLECULAR MEDICINE REPORTS* 14.4 (2016), 3459–3463. DOI: {10.3892/mmr.2016.5644}.
- [118] A. Uruno, Y. Yagishita, and M. Yamamoto. "The Keap1-Nrf2 system and diabetes mellitus". In: *ARCHIVES OF BIOCHEMISTRY AND BIOPHYSICS* 566 (2015), 76–84. DOI: {10.1016/j.abb.2014.12.012}.
- [119] T. Jiang et al. "The Protective Role of Nrf2 in Streptozotocin-Induced Diabetic Nephropathy". In: *DIABETES* 59.4 (2010), 850–860. DOI: {10.2337/db09-1342}.

-
- [120] Z. Ungvari et al. "Adaptive induction of NF-E2-related factor-2-driven antioxidant genes in endothelial cells in response to hyperglycemia". In: *AMERICAN JOURNAL OF PHYSIOLOGY-HEART AND CIRCULATORY PHYSIOLOGY* 300.4 (2011), H1133–H1140. DOI: {10.1152/ajpheart.00402.2010}.
 - [121] T.-S. Liu et al. "Oscillating high glucose enhances oxidative stress and apoptosis in human coronary artery endothelial cells". In: *JOURNAL OF ENDOCRINOLOGICAL INVESTIGATION* 37.7 (2014), 645–651. DOI: {10.1007/s40618-014-0086-5}.
 - [122] A. S. Jimenez-Osorio et al. "Nrf2 and Redox Status in Prediabetic and Diabetic Patients". In: *INTERNATIONAL JOURNAL OF MOLECULAR SCIENCES* 15.11 (2014), 20290–20305. DOI: {10.3390/ijms151120290}.
 - [123] K. Yoh et al. "Hyperglycemia induces oxidative and nitrosative stress and increases renal functional impairment in Nrf2-deficient mice". In: *GENES TO CELLS* 13.11 (2008), 1159–1170. DOI: {10.1111/j.1365-2443.2008.01234.x}.
 - [124] Z. Xu et al. "NRF2 plays a protective role in diabetic retinopathy in mice". In: *DIABETOLOGIA* 57.1 (2014), 204–213. DOI: {10.1007/s00125-013-3093-8}.
 - [125] P. J. Meakin et al. "Susceptibility of Nrf2-Null Mice to Steatohepatitis and Cirrhosis upon Consumption of a High-Fat Diet Is Associated with Oxidative Stress, Perturbation of the Unfolded Protein Response, and Disturbance in the Expression of Metabolic Enzymes but Not with Insulin Resistance". In: *MOLECULAR AND CELLULAR BIOLOGY* 34.17 (2014), 3305–3320. DOI: {10.1128/MCB.00677-1}.
 - [126] Y. Cui et al. "Experimental nonalcoholic fatty liver disease in mice leads to cytochrome p450 2a5 upregulation through nuclear factor erythroid 2-like 2 translocation". In: *REDOX BIOLOGY* 1.1 (2013), 433–440. DOI: {10.1016/j.redox.2013.08.003}.

- [127] M. Pilar Vaidecantos et al. "Essential role of Nrf2 in the protective effect of lipoic acid against lipoapoptosis in hepatocytes". In: *FREE RADICAL BIOLOGY AND MEDICINE* 84 (2015), 263–278. DOI: {10.1016/j.freeradbiomed.2015.03.019}.
- [128] P. Xue et al. "Adipose Deficiency of Nrf2 in ob/ob Mice Results in Severe Metabolic Syndrome". In: *DIABETES* 62.3 (2013), 845–854. DOI: {10.2337/db12-0584}.
- [129] A. Uruno et al. "The Keap1-Nrf2 System Prevents Onset of Diabetes Mellitus". In: *MOLECULAR AND CELLULAR BIOLOGY* 33.15 (2013), 2996–3010. DOI: {10.1128/MCB.00225-13}.
- [130] J. Xu et al. "Energy absorption ability of buckyball C-720 at low impact speed: a numerical study based on molecular dynamics". In: *NANOSCALE RESEARCH LETTERS* 8 (2013), 1–10. DOI: {10.1186/1556-276X-8-54}.
- [131] X. Wang et al. "Association between the NF-E2 Related Factor 2 Gene Polymorphism and Oxidative Stress, Anti-Oxidative Status, and Newly-Diagnosed Type 2 Diabetes Mellitus in a Chinese Population". In: *INTERNATIONAL JOURNAL OF MOLECULAR SCIENCES* 16.7 (2015), 16483–16496. DOI: {10.3390/ijms160716483}.
- [132] A. Sarai Jimenez-Osorio et al. "Association of Nuclear Factor-Erythroid 2-Related Factor 2, Thioredoxin Interacting Protein, and Heme Oxygenase-1 Gene Polymorphisms with Diabetes and Obesity in Mexican Patients (vol 2016, pg 8, 2016)". In: *OXIDATIVE MEDICINE AND CELLULAR LONGEVITY* (2017). DOI: {10.1155/2017/7543194}.
- [133] I. M.-J. Xu et al. "Transketolase counteracts oxidative stress to drive cancer development". In: *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA* 113.6 (2016), E725–E734. DOI: {10.1073/pnas.1508779113}.
- [134] M. Komatsu et al. "The selective autophagy substrate p62 activates the stress responsive transcription factor Nrf2 through inactivation of Keap1". In: *NATURE CELL BIOLOGY* 12.3 (2010), 213–U17. DOI: {10.1038/ncb2021}.

-
- [135] A. I. Rojo et al. "The PTEN/NRF2 Axis Promotes Human Carcinogenesis". In: *ANTIOXIDANTS & REDOX SIGNALING* 21.18 (2014), 2498–2514. DOI: {10.1089/ars.2014.5843}.
 - [136] C. P. Ramsey et al. "Expression of Nrf2 in neurodegenerative diseases". In: *JOURNAL OF NEUROPATHOLOGY AND EXPERIMENTAL NEUROLOGY* 66.1 (2007), 75–85. DOI: {10.1097/nen.0b013e31802d6da9}.
 - [137] F. van Muiswinkel et al. "Expression of NAD(P)H: quinone oxidoreductase in the normal and Parkinsonian substantia nigra". In: *NEUROBIOLOGY OF AGING* 25.9 (2004), 1253–1262. DOI: {10.1016/j.neurobiolaging.2003.12.010}.
 - [138] A. Cuadrado, P. Moreno-Murciano, and J. Pedraza-Chaverri. "The transcription factor Nrf2 as a new therapeutic target in Parkinson's disease". In: *EXPERT OPINION ON THERAPEUTIC TARGETS* 13.3 (2009), 319–329. DOI: {10.1517/13543780802716501}.
 - [139] H. M. Schipper et al. "Heme oxygenase-1 and neurodegeneration: expanding frontiers of engagement". In: *JOURNAL OF NEUROCHEMISTRY* 110.2 (2009), 469–485. DOI: {10.1111/j.1471-4159.2009.06160.x}.
 - [140] I. Lastres-Becker et al. "Repurposing the NRF2 Activator Dimethyl Fumarate as Therapy Against Synucleinopathy in Parkinson's Disease". In: *ANTIOXIDANTS & REDOX SIGNALING* 25.2 (2016), 61–77. DOI: {10.1089/ars.2015.6549}.
 - [141] P. Bergstrom et al. "Association of NFE2L2 and KEAP1 haplotypes with amyotrophic lateral sclerosis". In: *AMYOTROPHIC LATERAL SCLEROSIS AND FRONTOTEMPORAL DEGENERATION* 15.1-2 (2014), 130–137. DOI: {10.3109/21678421.2013.839708}.
 - [142] A. LoGerfo et al. "Lack of Association between Nuclear Factor Erythroid-Derived 2-Like 2 Promoter Gene Polymorphisms and Oxidative Stress Biomarkers in Amyotrophic Lateral Sclerosis Patients". In: *OXIDATIVE MEDICINE AND CELLULAR LONGEVITY* (2014). DOI: {10.1155/2014/432626}.

- [143] M. von Otter et al. "Nrf2-encoding NFE2L2 haplotypes influence disease progression but not risk in Alzheimer's disease and age-related cataract". In: *MECHANISMS OF AGEING AND DEVELOPMENT* 131.2 (2010), 105–110. DOI: {10.1016/j.mad.2009.12.007}.
- [144] M. von Otter et al. "Association of Nrf2-encoding NFE2L2 haplotypes with Parkinson's disease". In: *BMC MEDICAL GENETICS* 11 (2010). DOI: {10.1186/1471-2350-11-36}.
- [145] M. von Otter et al. "Genetic associations of Nrf2-encoding NFE2L2 variants with Parkinson's disease a multicenter study". In: *BMC MEDICAL GENETICS* 15 (2014). DOI: {10.1186/s12881-014-0131-4}.
- [146] Y.-C. Chen et al. "Genetic analysis of NFE2L2 promoter variation in Taiwanese Parkinson's disease". In: *PARKINSONISM & RELATED DISORDERS* 19.2 (2013), 247–250. DOI: {10.1016/j.parkreldis.2012.10.018}.
- [147] M. Todorovic et al. "Comprehensive Assessment of Genetic Sequence Variants in the Antioxidant 'Master Regulator' Nrf2 in Idiopathic Parkinson's Disease". In: *PLOS ONE* 10.5 (2015). DOI: {10.1371/journal.pone.0128030}.
- [148] N. M. Johnson et al. "Complete Protection against Aflatoxin B-1-Induced Liver Cancer with a Triterpenoid: DNA Adduct Dosimetry, Molecular Signature, and Genotoxicity Threshold". In: *CANCER PREVENTION RESEARCH* 7.7 (2014), 658–665. DOI: {10.1158/1940-6207.CAPR-13-0430}.
- [149] S. Reuter et al. "Oxidative stress, inflammation, and cancer: How are they linked?" In: *Free radical biology medicine* 49.11 (2010), 1603–1616. DOI: 10.1016/j.freeradbiomed.2010.09.006.
- [150] I.-g. Ryoo, S.-h. Lee, and M.-K. Kwak. "Redox Modulating NRF2: A Potential Mediator of Cancer Stem Cell Resistance". In: *OXIDATIVE MEDICINE AND CELLULAR LONGEVITY* (2016). DOI: {10.1155/2016/2428153}.

-
- [151] B. L. Emmink et al. "The secretome of colon cancer stem cells contains drug-metabolizing enzymes". In: *JOURNAL OF PROTEOMICS* 91 (2013), 84–96. DOI: {10.1016/j.jprot.2013.06.027}.
 - [152] J. Gao et al. "3D clusters of somatic mutations in cancer reveal numerous rare mutations as functional targets". In: *GENOME MEDICINE* 9 (2017). DOI: {10.1186/s13073-016-0393-x}.
 - [153] A. Zehir et al. "Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients". In: *NATURE MEDICINE* 23.6 (2017), 703+. DOI: {10.1038/nm.4333}.
 - [154] Y. R. Kim et al. "Oncogenic NRF2 mutations in squamous cell carcinomas of oesophagus and skin". In: *JOURNAL OF PATHOLOGY* 220.4 (2010), 446–451. DOI: {10.1002/path.2653}.
 - [155] K. Taguchi and M. Yamamoto. "The KeAP1-NRF2 System in Cancer". In: *FRONTIERS IN ONCOLOGY* 7 (2017). DOI: {10.3389/fonc.2017.00085}.
 - [156] T. Shibata et al. "NRF2 Mutation Confers Malignant Potential and Resistance to Chemoradiation Therapy in Advanced Esophageal Squamous Cancer". In: *NEOPLASIA* 13.9 (2011), 864–U133. DOI: {10.1593/neo.11750}.
 - [157] A. Singh et al. "Dysfunctional KEAP1-NRF2 interaction in non-small-cell lung cancer". In: *PLOS MEDICINE* 3.10 (2006), 1865–1876. DOI: {10.1371/journal.pmed.0030420}.
 - [158] Y. Okano et al. "SNP (-617C > A) in ARE-Like Loci of the NRF2 Gene: A New Biomarker for Prognosis of Lung Adenocarcinoma in Japanese Non-Smoking Women". In: *PLOS ONE* 8.9 (2013). DOI: {10.1371/journal.pone.0073794}.
 - [159] R. Wang et al. "Hypermethylation of the Keap1 gene in human lung cancer cell lines and lung cancer tissues". In: *BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS* 373.1 (2008), 151–154. DOI: {10.1016/j.bbrc.2008.06.004}.

- [160] G. Eades et al. "miR-200a Regulates Nrf2 Activation by Targeting Keap1 mRNA in Breast Cancer Cells". In: *JOURNAL OF BIOLOGICAL CHEMISTRY* 286.47 (2011), 40725–40733. DOI: {10.1074/jbc.M111.275495}.
- [161] M. Yang et al. "MiR-28 regulates Nrf2 expression through a Keap1-independent mechanism". In: *BREAST CANCER RESEARCH AND TREATMENT* 129.3 (2011), 983–991. DOI: {10.1007/s10549-011-1604-1}.
- [162] G. M. De Nicola et al. "Oncogene-induced Nrf2 transcription promotes ROS detoxification and tumorigenesis". In: *NATURE* 475.7354 (2011), 106–U128. DOI: {10.1038/nature10189}.
- [163] M.-C. Tung et al. "Mutant p53 confers chemoresistance in non-small cell lung cancer by upregulating Nrf2". In: *ONCOTARGET* 6.39 (2015), 41692–41705. DOI: {10.18632/oncotarget.6150}.
- [164] J. Chen et al. "Clinical implication of Keap1 and phosphorylated Nrf2 expression in hepatocellular carcinoma". In: *CANCER MEDICINE* 5.10 (2016), 2678–2687. DOI: {10.1002/cam4.788}.
- [165] J. A. McCubrey et al. "GSK-3 as potential target for therapeutic intervention in cancer". In: *ONCOTARGET* 5.10 (2014), 2881–2911. DOI: {10.18632/oncotarget.2037}.
- [166] T. Shimizu et al. "Accumulation of phosphorylated p62 is associated with NF-E2-related factor 2 activation in hepatocellular carcinoma". In: *JOURNAL OF HEPATO-BILIARY-PANCREATIC SCIENCES* 23.8 (2016), 467–471. DOI: {10.1002/jhbp.364}.
- [167] W. Chen et al. "Direct Interaction between Nrf2 and p21(Cip1/WAF1) Upregulates the Nrf2-Mediated Antioxidant Response". In: *MOLECULAR CELL* 34.6 (2009), 663–673. DOI: {10.1016/j.molcel.2009.04.029}.
- [168] B. E. Hast et al. "Proteomic Analysis of Ubiquitin Ligase KEAP1 Reveals Associated Proteins That Inhibit NRF2 Ubiquitination". In: *CANCER RESEARCH* 73.7 (2013), 2199–2210. DOI: {10.1158/0008-5472.CAN-12-4400}.

-
- [169] X. Tang et al. "A joint analysis of metabolomics and genetics of breast cancer". In: *BREAST CANCER RESEARCH* 16.4 (2014). DOI: {10.1186/s13058-014-0415-9}.
 - [170] P. C. Hart et al. "Caveolin-1 regulates cancer cell metabolism via scavenging Nrf2 and suppressing MnSOD-driven glycolysis". In: *ONCOTARGET* 7.1 (2016), 308–322. DOI: {10.18632/oncotarget.5687}.
 - [171] Y. Mitsuishi et al. "Nrf2 Redirects Glucose and Glutamine into Anabolic Pathways in Metabolic Reprogramming". In: *CANCER CELL* 22.1 (2012), 66–79. DOI: {10.1016/j.ccr.2012.05.016}.
 - [172] X. Xu et al. "Genetic variants of nuclear factor erythroid-derived 2-like 2 associated with the complications in Han descents with type 2 diabetes mellitus of Northeast China". In: *JOURNAL OF CELLULAR AND MOLECULAR MEDICINE* 20.11 (2016), 2078–2088. DOI: {10.1111/jcmm.12900}.
 - [173] T. Yamamoto et al. "Physiological significance of reactive cysteine residues of keap1 in determining Nrf2 activity". In: *MOLECULAR AND CELLULAR BIOLOGY* 28.8 (2008), 2758–2770. DOI: {10.1128/MCB.01704-07}.
 - [174] R. Saito et al. "Characterizations of Three Major Cysteine Sensors of Keap1 in Stress Response". In: *Molecular and Cellular Biology* (2015), MCB.00868–15. DOI: 10.1128/mcb.00868-15.
 - [175] G. Rachakonda et al. "Covalent modification at Cys151 dissociates the electrophile sensor Keap1 from the ubiquitin ligase CUL3". In: *CHEMICAL RESEARCH IN TOXICOLOGY* 21.3 (2008), 705–710. DOI: {10.1021/tx700302s}.
 - [176] L. Baird and A. T. Dinkova-Kostova. "Diffusion dynamics of the Keap1-Cullin3 interaction in single live cells". In: *BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS* 433.1 (2013), 58–65. DOI: {10.1016/j.bbrc.2013.02.065}.
 - [177] A. Cleasby et al. "Structure of the BTB Domain of Keap1 and Its Interaction with the Triterpenoid Antagonist CDDO". In: *PLOS ONE* 9.6 (2014). DOI: {10.1371/journal.pone.0098896}.

- [178] H. Sun et al. "Recent progress in the development of small molecule Nrf2 modulators: a patent review (2012-2016)". In: *EXPERT OPINION ON THERAPEUTIC PATENTS* 27.7 (2017), 763–785. DOI: {10.1080/13543776.2017.1325464}.
- [179] S. X. Lin et al. "The anti-inflammatory effects of dimethyl fumarate in astrocytes involve glutathione and haem oxygenase-1". In: *ASN NEURO* 3.2 (2011). DOI: {10.1042/AN20100033}.
- [180] P. J. Altmeyer et al. "ANTIPSORIATIC EFFECT OF FUMARIC-ACID DERIVATIVES - RESULTS OF A MULTICENTER DOUBLE-BLIND-STUDY IN 100 PATIENTS". In: *JOURNAL OF THE AMERICAN ACADEMY OF DERMATOLOGY* 30.6 (1994), 977–981. DOI: {10.1016/S0190-9622(94)70121-0}.
- [181] U. Mrowietz, E. Christophers, and P. Altmeyer. "Treatment of psoriasis with fumaric acid esters: results of a prospective multicentre study". In: *BRITISH JOURNAL OF DERMATOLOGY* 138.3 (1998), 456–460.
- [182] U. Mrowietz et al. "Efficacy and safety of LAS41008 (dimethyl fumarate) in adults with moderate-to-severe chronic plaque psoriasis: a randomized, double-blind, Fumaderm (R) - and placebo-controlled trial (BRIDGE)". In: *BRITISH JOURNAL OF DERMATOLOGY* 176.3 (2017), 615–623. DOI: {10.1111/bjd.14947}.
- [183] K. Ghoreschi et al. "Fumarates improve psoriasis and multiple sclerosis by inducing type II dendritic cells". In: *JOURNAL OF EXPERIMENTAL MEDICINE* 208.11 (2011), 2291–2303. DOI: {10.1084/jem.20100977}.
- [184] S. Tahvili, B. Zandieh, and Z. Amirghofran. "The effect of dimethyl fumarate on gene expression and the level of cytokines related to different Thelper cell subsets in peripheral blood mononuclear cells of patients with psoriasis". In: *INTERNATIONAL JOURNAL OF DERMATOLOGY* 54.7 (2015), E254–E260. DOI: {10.1111/ijd.12834}.
- [185] Z. Xu et al. "Dimethyl fumarate for multiple sclerosis". In: *Cochrane Database of Systematic Reviews* (2015). DOI: 10.1002/14651858.cd011076.pub2.

-
- [186] S. Schilling et al. "Fumaric acid esters are effective in chronic experimental autoimmune encephalomyelitis and suppress macrophage infiltration". In: *CLINICAL AND EXPERIMENTAL IMMUNOLOGY* 145.1 (2006), 101–107. DOI: {10.1111/j.1365-2249.2006.03094.x}.
 - [187] G. Ellrichmann et al. "Efficacy of Fumaric Acid Esters in the R6/2 and YAC128 Models of Huntington's Disease". In: *PLOS ONE* 6.1 (2011). DOI: {10.1371/journal.pone.0016172}.
 - [188] S. Schimrigk et al. "Oral fumaric acid esters for the treatment of active multiple sclerosis: an open-label, baseline-controlled pilot study". In: *EUROPEAN JOURNAL OF NEUROLOGY* 13.6 (2006), 604–610. DOI: {10.1111/j.1468-1331.2006.01292.x}.
 - [189] R. J. Fox et al. "Placebo-Controlled Phase 3 Study of Oral BG-12 or Glatiramer in Multiple Sclerosis". In: *NEW ENGLAND JOURNAL OF MEDICINE* 367.12 (2012), 1087–1097. DOI: {10.1056/NEJMoa1206328}.
 - [190] R. Gold et al. "Placebo-Controlled Phase 3 Study of Oral BG-12 for Relapsing Multiple Sclerosis". In: *NEW ENGLAND JOURNAL OF MEDICINE* 367.12 (2012), 1098–1107. DOI: {10.1056/NEJMoa1114287}.
 - [191] T. A. Zeidan et al. *Prodrugs of Fumarates and Their Use in Treating Various Diseases*. 2014.
 - [192] Z. Al-Jaderi and A. A. Maghazachi. "Vitamin D-3 and Monomethyl Fumarate Enhance Natural Killer Cell Lysis of Dendritic Cells and Ameliorate the Clinical Score in Mice Suffering from Experimental Autoimmune Encephalomyelitis". In: *TOXINS* 7.11 (2015), 4730–4744. DOI: {10.3390/toxins7114730}.
 - [193] H. M. Ockenfels et al. "The antipsoriatic agent dimethyl-fumarate immunomodulates T-cell cytokine secretion and inhibits cytokines of the psoriatic cytokine network". In: *BRITISH JOURNAL OF DERMATOLOGY* 139.3 (1998). Conference on Psoriasis - from Gene to Clinic, LONDON, ENGLAND, DEC 05-07, 1996, 390–395.

- [194] H. Peng et al. "Dimethyl Fumarate Inhibits Dendritic Cell Maturation via Nuclear Factor kappa B (NF-kappa B) and Extracellular Signal-regulated Kinase 1 and 2 (ERK1/2) and Mitogen Stress-activated Kinase 1 (MSK1) Signaling". In: *JOURNAL OF BIOLOGICAL CHEMISTRY* 287.33 (2012), 28017–28026. DOI: {10.1074/jbc.M112.383380}.
- [195] M. S. Brennan et al. "The NRF2 transcriptional target, OSGIN1, contributes to monomethyl fumarate-mediated cytoprotection in human astrocytes". In: *SCIENTIFIC REPORTS* 7 (2017). DOI: {10.1038/srep42054}.
- [196] S. Eminel et al. "Dimethyl- and monomethylfumarate regulate indoleamine 2,3-dioxygenase (IDO) activity in human immune cells". In: *EXPERIMENTAL DERMATOLOGY* 26.8 (2017), 685–690. DOI: {10.1111/exd.13138}.
- [197] M. Meissner et al. "Suppression of VEGFR2 Expression in Human Endothelial Cells by Dimethylfumarate Treatment: Evidence for Anti-Angiogenic Action". In: *JOURNAL OF INVESTIGATIVE DERMATOLOGY* 131.6 (2011), 1356–1364. DOI: {10.1038/jid.2011.46}.
- [198] Q. Wu et al. "Dimethyl Fumarate Selectively Reduces Memory T Cells and Shifts the Balance between Th1/Th17 and Th2 in Multiple Sclerosis Patients". In: *JOURNAL OF IMMUNOLOGY* 198.8 (2017), 3069–3080. DOI: {10.4049/jimmunol.1601532}.
- [199] R. H. Scannevin et al. "Fumarates Promote Cytoprotection of Central Nervous System Cells against Oxidative Stress via the Nuclear Factor (Erythroid-Derived 2)-Like 2 Pathway". In: *JOURNAL OF PHARMACOLOGY AND EXPERIMENTAL THERAPEUTICS* 341.1 (2012), 274–284. DOI: {10.1124/jpet.111.190132}.
- [200] M. M. Blewett et al. "Chemical proteomic map of dimethyl fumarate-sensitive cysteines in primary human T cells". In: *SCIENCE SIGNALING* 9.445 (2016). DOI: {10.1126/scisignal.aaf7694}.

-
- [201] V. A. McGuire et al. "Dimethyl fumarate blocks proinflammatory cytokine production via inhibition of TLR induced M1 and K63 ubiquitin chain formation". In: *SCIENTIFIC REPORTS* 6 (2016). DOI: {10.1038/srep31159}.
 - [202] H. Chen et al. "Hydroxycarboxylic acid receptor 2 mediates dimethyl fumarate's protective effect in EAE". In: *JOURNAL OF CLINICAL INVESTIGATION* 124.5 (2014), 2188–2192. DOI: {10.1172/JCI72151}.
 - [203] G. O. Gillard et al. "DMF, but not other fumarates, inhibits NF-kappa B activity in vitro in an Nrf2-independent manner". In: *JOURNAL OF NEUROIMMUNOLOGY* 283 (2015), 74–85. DOI: {10.1016/j.jneuroim.2015.04.006}.
 - [204] R. A. Linker et al. "Fumaric acid esters exert neuroprotective effects in neuroinflammation via activation of the Nrf2 antioxidant pathway". In: *BRAIN* 134.3 (2011), 678–692. DOI: {10.1093/brain/awq386}.
 - [205] U. Schulze-Topphoff et al. "Dimethyl fumarate treatment induces adaptive and innate immune modulation independent of Nrf2". In: *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA* 113.17 (2016), 4777–4782. DOI: {10.1073/pnas.1603907113}.
 - [206] A. Sharma et al. "The nuclear factor (erythroid-derived 2)-like 2 (Nrf2) activator dh404 protects against diabetes-induced endothelial dysfunction". In: *CARDIOVASCULAR DIABETOLOGY* 16 (2017). DOI: {10.1186/s12933-017-0513-y}.
 - [207] S. M. Tan et al. "Derivative of Bardoxolone Methyl, dh404, in an Inverse Dose-Dependent Manner Lessens Diabetes-Associated Atherosclerosis and Improves Diabetic Kidney Disease". In: *DIABETES* 63.9 (2014), 3091–3103. DOI: {10.2337/db13-1743}.
 - [208] I. Kastrati et al. "Dimethyl Fumarate Inhibits the Nuclear Factor B Pathway in Breast Cancer Cells by Covalent Modification of p65 Protein". In: *JOURNAL OF BIOLOGICAL CHEMISTRY* 291.7 (2016), 3639–3647. DOI: {10.1074/jbc.M115.679704}.

- [209] R. Thimmulappa et al. "Nrf2 is a critical regulator of the innate immune response and survival during experimental sepsis". In: *JOURNAL OF CLINICAL INVESTIGATION* 116.4 (2006), 984–995. DOI: {10.1172/JCI25790}.
- [210] N. A. Kaidery et al. "Targeting Nrf2-Mediated Gene Transcription by Extremely Potent Synthetic Triterpenoids Attenuate Dopaminergic Neurotoxicity in the MPTP Mouse Model of Parkinson's Disease". In: *ANTIOXIDANTS & REDOX SIGNALING* 18.2 (2013), 139–157. DOI: {10.1089/ars.2011.4491}.
- [211] T. K. Pareek et al. "Triterpenoid modulation of IL-17 and Nrf-2 expression ameliorates neuroinflammation and promotes remyelination in autoimmune encephalomyelitis". In: *SCIENTIFIC REPORTS* 1 (2011). DOI: {10.1038/srep00201}.
- [212] P. E. Pergola et al. "Bardoxolone Methyl and Kidney Function in CKD with Type 2 Diabetes". In: *NEW ENGLAND JOURNAL OF MEDICINE* 365.4 (2011), 327–336. DOI: {10.1056/NEJMoal105351}.
- [213] D. de Zeeuw et al. "Bardoxolone Methyl in Type 2 Diabetes and Stage 4 Chronic Kidney Disease". In: *NEW ENGLAND JOURNAL OF MEDICINE* 369.26 (2013), 2492–2503. DOI: {10.1056/NEJMoal1306033}.
- [214] M. P. Chin et al. "Mechanisms Contributing to Adverse Cardiovascular Events in Patients with Type 2 Diabetes Mellitus and Stage 4 Chronic Kidney Disease Treated with Bardoxolone Methyl". In: *AMERICAN JOURNAL OF NEPHROLOGY* 39.6 (2014), 499–508. DOI: {10.1159/000362906}.
- [215] X. Kuang et al. "Attenuation of oxidative stress, inflammation and apoptosis by minocycline prevents retrovirus-induced neurodegeneration in mice". In: *BRAIN RESEARCH* 1286 (2009), 174–184. DOI: {10.1016/j.brainres.2009.06.007}.
- [216] I. Buendia et al. "Neuroprotective mechanism of the novel melatonin derivative Neu-P11 in brain ischemia related models". In: *NEUROPHARMACOLOGY* 99 (2015), 187–195. DOI: {10.1016/j.neuropharm.2015.07.014}.

-
- [217] I. Buendia et al. "New melatonin-cinnamate hybrids as multi-target drugs for neurodegenerative diseases: Nrf2-induction, antioxidant effect and neuroprotection". In: *FUTURE MEDICINAL CHEMISTRY* 7.15 (2015), 1961–1969. DOI: {10.4155/fmc.15.99}.
- [218] T. W. Kensler et al. "Keap1-Nrf2 Signaling: A Target for Cancer Prevention by Sulforaphane". In: *NATURAL PRODUCTS IN CANCER PREVENTION AND THERAPY*. Ed. by Pezzuto, JM and Suh, N. Vol. 329. Topics in Current Chemistry-Series. 2013, 163–177. DOI: {10.1007/128_2012_339}.
- [219] T. Kim et al. "The synthesis of sulforaphane analogues and their protection effect against cisplatin induced cytotoxicity in kidney cells". In: *BIOORGANIC & MEDICINAL CHEMISTRY LETTERS* 25.1 (2015), 62–66. DOI: {10.1016/j.bmcl.2014.11.014}.
- [220] Z. Bahadoran et al. "Broccoli sprouts powder could improve serum triglyceride and oxidized LDL/LDL-cholesterol ratio in type 2 diabetic patients: A randomized double-blind placebo-controlled clinical trial". In: *DIABETES RESEARCH AND CLINICAL PRACTICE* 96.3 (2012), 348–354. DOI: {10.1016/j.diabres.2012.01.009}.
- [221] A. S. Axelsson et al. "Sulforaphane reduces hepatic glucose production and improves glucose control in patients with type 2 diabetes". In: *SCIENCE TRANSLATIONAL MEDICINE* 9.394 (2017). DOI: {10.1126/scitranslmed.aah4477}.
- [222] Z.-Y. Jiang et al. "Discovery of Potent Keap1-Nrf2 Protein-Protein Interaction Inhibitor Based on Molecular Binding Determinants Analysis". In: *JOURNAL OF MEDICINAL CHEMISTRY* 57.6 (2014), 2736–2745. DOI: {10.1021/jm5000529}.
- [223] N. G. Innamorato et al. "The transcription factor Nrf2 is a therapeutic target against brain inflammation". In: *JOURNAL OF IMMUNOLOGY* 181.1 (2008), 680–689. DOI: {10.4049/jimmunol.181.1.680}.

- [224] A. Jazwa et al. "Pharmacological Targeting of the Transcription Factor Nrf2 at the Basal Ganglia Provides Disease Modifying Therapy for Experimental Parkinsonism". In: *ANTIOXIDANTS & REDOX SIGNALING* 14.12 (2011), 2347–2360. DOI: {10.1089/ars.2010.3731}.
- [225] C. Jo et al. "Nrf2 reduces levels of phosphorylated tau protein by inducing autophagy adaptor protein NDP52". In: *NATURE COMMUNICATIONS* 5 (2014). DOI: {10.1038/ncomms4496}.
- [226] X. Wang et al. "Activation of the Nuclear Factor E2-Related Factor 2/Antioxidant Response Element Pathway Is Neuroprotective after Spinal Cord Injury". In: *JOURNAL OF NEUROTRAUMA* 29.5 (2012), 936+. DOI: {10.1089/neu.2011.1922}.
- [227] B. Li et al. "Sulforaphane ameliorates the development of experimental autoimmune encephalomyelitis by antagonizing oxidative stress and Th17-related inflammation in mice". In: *EXPERIMENTAL NEUROLOGY* 250 (2013), 239–249. DOI: {10.1016/j.expneurol.2013.10.002}.
- [228] C. G. Duran et al. "A proof-of-concept clinical study examining the NRF2 activator sulforaphane against neutrophilic airway inflammation". In: *RESPIRATORY RESEARCH* 17 (2016). DOI: {10.1186/s12931-016-0406-8}.
- [229] C. A. Houghton, R. G. Fassett, and J. S. Coombes. "Sulforaphane and Other Nutrigenomic Nrf2 Activators: Can the Clinician's Expectation Be Matched by the Reality?" In: *OXIDATIVE MEDICINE AND CELLULAR LONGEVITY* (2016). DOI: {10.1155/2016/7857186}.
- [230] J. Egea et al. "Melatonin-sulforaphane hybrid ITH12674 induces neuroprotection in oxidative stress conditions by a 'drug-prodrug' mechanism of action". In: *BRITISH JOURNAL OF PHARMACOLOGY* 172.7, SI (2015), 1807–1821. DOI: {10.1111/bph.13025}.

-
- [231] P. Qiu et al. "Overdose Intake of Curcumin Initiates the Unbalanced State of Bodies". In: *JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY* 64.13 (2016), 2765–2771. DOI: {10.1021/acs.jafc.6b00053}.
 - [232] L.-X. Na et al. "Curcuminoids exert glucose-lowering effect in type 2 diabetes by decreasing serum free fatty acids: a double-blind, placebo-controlled trial". In: *MOLECULAR NUTRITION & FOOD RESEARCH* 57.9, SI (2013), 1569–1577. DOI: {10.1002/mnfr.201200131}.
 - [233] S. Chuengsamarn et al. "Reduction of atherogenic risk in patients with type 2 diabetes by curcuminoid extract: a randomized controlled trial". In: *JOURNAL OF NUTRITIONAL BIOCHEMISTRY* 25.2 (2014), 144–150. DOI: {10.1016/j.jnutbio.2013.09.013}.
 - [234] Z. Ungvari et al. "Resveratrol confers endothelial protection via activation of the antioxidant transcription factor Nrf2". In: *AMERICAN JOURNAL OF PHYSIOLOGY-HEART AND CIRCULATORY PHYSIOLOGY* 299.1 (2010), H18–H24. DOI: {10.1152/ajpheart.00260.2010}.
 - [235] H. Ghanim et al. "A Resveratrol and Polyphenol Preparation Suppresses Oxidative and Inflammatory Stress Response to a High-Fat, High-Carbohydrate Meal". In: *JOURNAL OF CLINICAL ENDOCRINOLOGY & METABOLISM* 96.5 (2011), 1409–1414. DOI: {10.1210/jc.2010-1812}.
 - [236] P. Brasnyo et al. "Resveratrol improves insulin sensitivity, reduces oxidative stress and activates the Akt pathway in type 2 diabetic patients". In: *BRITISH JOURNAL OF NUTRITION* 106.3 (2011), 383–389. DOI: {10.1017/S0007114511000316}.
 - [237] J. K. Bhatt, S. Thomas, and M. J. Nanjan. "Resveratrol supplementation improves glycemic control in type 2 diabetes mellitus". In: *NUTRITION RESEARCH* 32.7 (2012), 537–541. DOI: {10.1016/j.nutres.2012.06.003}.
 - [238] N. Xia et al. "Antioxidant effects of resveratrol in the cardiovascular system". In: *BRITISH JOURNAL OF PHARMACOLOGY* 174.12, SI (2017), 1633–1646. DOI: {10.1111/bph.13492}.

- [239] M. M. Yore et al. "Proteomic Analysis Shows Synthetic Oleanane Triterpenoid Binds to mTOR". In: *PLOS ONE* 6.7 (2011). DOI: {10.1371/journal.pone.0022862}.
- [240] S. R. Lee et al. "Reversible inactivation of the tumor suppressor PTEN by H₂O₂". In: *JOURNAL OF BIOLOGICAL CHEMISTRY* 277.23 (2002), 20336–20342. DOI: {10.1074/jbc.M111899200}.
- [241] Y. Kitagishi and S. Matsuda. "Redox regulation of tumor suppressor PTEN in cancer and aging (Review)". In: *INTERNATIONAL JOURNAL OF MOLECULAR MEDICINE* 31.3 (2013), 511–515. DOI: {10.3892/ijmm.2013.1235}.
- [242] S.-J. Han et al. "Assay of the redox state of the tumor suppressor PTEN by mobility shift". In: *METHODS* 77-78 (2015), 58–62. DOI: {10.1016/j.ymeth.2015.01.007}.
- [243] I. Pitha-Rowe et al. "Synthetic Triterpenoids Attenuate Cytotoxic Retinal Injury: Cross-talk between Nrf2 and PI3K/AKT Signaling through Inhibition of the Lipid Phosphatase PTEN". In: *INVESTIGATIVE OPHTHALMOLOGY & VISUAL SCIENCE* 50.11 (2009), 5339–5347. DOI: {10.1167/iovs.09-3648}.
- [244] A. I. Rojo et al. "Redox Control of Microglial Function: Molecular Mechanisms and Functional Significance". In: *ANTIOXIDANTS & REDOX SIGNALING* 21.12 (2014), 1766–1801. DOI: {10.1089/ars.2013.5745}.
- [245] J.-E. Kim et al. "Suppression of NF-kappa B signaling by KEAP1 regulation of IKK beta activity through autophagic degradation and inhibition of phosphorylation". In: *CELLULAR SIGNALLING* 22.11 (2010), 1645–1654. DOI: {10.1016/j.cellsig.2010.06.004}.
- [246] S. C. Cazanave et al. "Degradation of Keap1 activates BH3-only proteins Bim and PUMA during hepatocyte lipoapoptosis". In: *CELL DEATH AND DIFFERENTIATION* 21.8 (2014), 1303–1312. DOI: {10.1038/cdd.2014.49}.

- [247] B. G. Richardson et al. "Non-electrophilic modulators of the canonical Keap1/Nrf2 pathway". In: *BIOORGANIC & MEDICINAL CHEMISTRY LETTERS* 25.11 (2015), 2261–2268. DOI: {10.1016/j.bmcl.2015.04.019}.
- [248] B. Padmanabhan et al. "Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer". In: *MOLECULAR CELL* 21.5 (2006), 689–700. DOI: {10.1016/j.molcel.2006.01.013}.
- [249] S.-C. Lo et al. "Structure of the Keap1 : Nrf2 interface provides mechanistic insight into Nrf2 signaling". In: *EMBO JOURNAL* 25.15 (2006), 3605–3617. DOI: {10.1038/sj.emboj.7601243}.
- [250] L. Hu et al. "Discovery of a small-molecule inhibitor and cellular probe of Keap1-Nrf2 protein-protein interaction". In: *BIOORGANIC & MEDICINAL CHEMISTRY LETTERS* 23.10 (2013), 3039–3043. DOI: {10.1016/j.bmcl.2013.03.013}.
- [251] D. Marcotte et al. "Small molecules inhibit the interaction of Nrf2 and the Keap1 Kelch domain through a non-covalent mechanism". In: *BIOORGANIC & MEDICINAL CHEMISTRY* 21.14 (2013), 4011–4019. DOI: {10.1016/j.bmc.2013.04.019}.
- [252] A. F. Winkel et al. "Characterization of RA839, a Noncovalent Small Molecule Binder to Keap1 and Selective Activator of Nrf2 Signaling". In: *JOURNAL OF BIOLOGICAL CHEMISTRY* 290.47 (2015), 28446–28455. DOI: {10.1074/jbc.M115.678136}.
- [253] M.-C. Lu et al. "Discovery of a head-to-tail cyclic peptide as the Keap1-Nrf2 protein-protein interaction inhibitor with high cell potency". In: *EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY* 143 (2018), 1578–1589. DOI: {10.1016/j.ejmech.2017.10.052}.
- [254] D. A. Abed et al. "Discovery of direct inhibitors of Keap1-Nrf2 protein-protein interaction as potential therapeutic and preventive agents". In: *ACTA PHARMACEUTICA SINICA B* 5.4 (2015), 285–299. DOI: {10.1016/j.apsb.2015.05.008}.

- [255] Z.-Y. Jiang, M.-C. Lu, and Q.-D. You. "Discovery and Development of Kelch-like ECH-Associated Protein 1. Nuclear Factor Erythroid 2-Related Factor 2 (KEAP1:NRF2) Protein-Protein Interaction Inhibitors: Achievements, Challenges, and Future Directions". In: *JOURNAL OF MEDICINAL CHEMISTRY* 59.24 (2016), 10837–10858. DOI: {10.1021/acs.jmedchem.6b00586}.
- [256] Y. Chen et al. "Kinetic Analyses of Keap1-Nrf2 Interaction and Determination of the Minimal Nrf2 Peptide Sequence Required for Keap1 Binding Using Surface Plasmon Resonance". In: *CHEMICAL BIOLOGY & DRUG DESIGN* 78.6 (2011), 1014–1021. DOI: {10.1111/j.1747-0285.2011.01240.x}.
- [257] D. Inoyama et al. "Optimization of Fluorescently Labeled Nrf2 Peptide Probes and the Development of a Fluorescence Polarization Assay for the Discovery of Inhibitors of Keap1-Nrf2 Interaction". In: *JOURNAL OF BIOMOLECULAR SCREENING* 17.4 (2012), 435–447. DOI: {10.1177/1087057111430124}.
- [258] R. Hancock et al. "Peptide inhibitors of the Keap1-Nrf2 protein-protein interaction with improved binding and cellular activity". In: *ORGANIC & BIOMOLECULAR CHEMISTRY* 11.21 (2013), 3553–3557. DOI: {10.1039/c3ob40249e}.
- [259] J. Tu et al. "Cell-Permeable Peptide Targeting the Nrf2-Keap1 Interaction: A Potential Novel Therapy for Global Cerebral Ischemia". In: *JOURNAL OF NEUROSCIENCE* 35.44 (2015), 14727–14739. DOI: {10.1523/JNEUROSCI.1304-15.2015}.
- [260] E. Jnoff et al. "Binding Mode and Structure-Activity Relationships around Direct Inhibitors of the Nrf2-Keap1 Complex". In: *CHEMMEDCHEM* 9.4 (2014), 699–705. DOI: {10.1002/cmdc.201300525}.
- [261] T. Jiang et al. "Nrf2 suppresses lupus nephritis through inhibition of oxidative injury and the NF-kappa B-mediated inflammatory response". In: *KIDNEY INTERNATIONAL* 85.2 (2014), 333–343. DOI: {10.1038/ki.2013.343}.
- [262] N. Ranjan et al. "Selective inhibition of bacterial topoisomerase I by alkynyl-bisbenzimidazoles". In: *MEDCHEMCOMM* 5.6 (2014), 816–825. DOI: {10.1039/c4md00140k}.

-
- [263] M. SATO et al. *KEAP1 PROTEIN BINDING COMPOUND, CRYSTAL OF COMPLEX BETWEEN THE SAME AND KEAP1 PROTEIN, AND METHOD FOR PRODUCING THE SAME*. 2013.
- [264] H. C. Bertrand et al. "Design, Synthesis, and Evaluation of Triazole Derivatives That Induce Nrf2 Dependent Gene Products and Inhibit the Keap1-Nrf2 Protein-Protein Interaction". In: *JOURNAL OF MEDICINAL CHEMISTRY* 58.18 (2015), 7186–7194. DOI: {10.1021/acs.jmedchem.5b00602}.
- [265] T. Silva et al. "Alzheimer's disease, enzyme targets and drug discovery struggles: From natural products to drug prototypes". In: *AGEING RESEARCH REVIEWS* 15 (2014), 116–145. DOI: {10.1016/j.arr.2014.03.008}.
- [266] O. C. Attucks et al. "Induction of Heme Oxygenase I (HMOX1) by HPP-4382: A Novel Modulator of Bach1 Activity". In: *PLOS ONE* 9.7 (2014). DOI: {10.1371/journal.pone.0101044}.
- [267] L. Milkovic, N. Zarkovic, and L. Saso. "Controversy about pharmacological modulation of Nrf2 for cancer therapy". In: *REDOX BIOLOGY* 12 (2017), 727–732. DOI: {10.1016/j.redox.2017.04.013}.
- [268] J. L. Yap et al. "Small-Molecule Inhibitors of the ERK Signaling Pathway: Towards Novel Anticancer Therapeutics". In: *CHEMMEDCHEM* 6.1 (2011), 38–48. DOI: {10.1002/cmdc.201000354}.
- [269] A. Manna et al. "The variable chemotherapeutic response of Malabaricone-A in leukemic and solid tumor cell lines depends on the degree of redox imbalance". In: *PHYTOMEDICINE* 22.7-8 (2015), 713–723. DOI: {10.1016/j.phymed.2015.05.007}.
- [270] T. Tarumoto et al. "Ascorbic acid restores sensitivity to imatinib via suppression of Nrf2-dependent gene expression in the imatinib-resistant cell line". In: *EXPERIMENTAL HEMATOLOGY* 32.4 (2004), 375–381. DOI: {10.1016/j.exphem.2004.01.007}.

- [271] X. J. Wang et al. "Identification of retinoic acid as an inhibitor of transcription factor Nrf2 through activation of retinoic acid receptor alpha". In: *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA* 104.49 (2007), 19589–19594. DOI: {10.1073/pnas.0709483104}.
- [272] D. Ren et al. "Brusatol enhances the efficacy of chemotherapy by inhibiting the Nrf2-mediated defense mechanism". In: *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA* 108.4 (2011), 1433–1438. DOI: {10.1073/pnas.1014275108}.
- [273] A. Olayanju et al. "Brusatol provokes a rapid and transient inhibition of Nrf2 signaling and sensitizes mammalian cells to chemical toxicity-implications for therapeutic targeting of Nrf2". In: *FREE RADICAL BIOLOGY AND MEDICINE* 78 (2015), 202–212. DOI: {10.1016/j.freeradbiomed.2014.11.003}.
- [274] A. Limonciel and P. Jennings. "A Review of the Evidence that Ochratoxin A Is an Nrf2 Inhibitor: Implications for Nephrotoxicity and Renal Carcinogenicity". In: *TOXINS* 6.1 (2014), 371–379. DOI: {10.3390/toxins6010371}.
- [275] A. Arlt et al. "Inhibition of the Nrf2 transcription factor by the alkaloid trigonelline renders pancreatic cancer cells more susceptible to apoptosis through decreased proteasomal gene expression and proteasome activity". In: *ONCOGENE* 32.40 (2013), 4825–4835. DOI: {10.1038/onc.2012.493}.
- [276] B. Harder et al. "Brusatol overcomes chemoresistance through inhibition of protein translation". In: *MOLECULAR CARCINOGENESIS* 56.5 (2017), 1493–1500. DOI: {10.1002/mc.22609}.
- [277] K. Tsuchida et al. "Halofuginone enhances the chemo-sensitivity of cancer cells by suppressing NRF2 accumulation". In: *FREE RADICAL BIOLOGY AND MEDICINE* 103 (2017), 236–247. DOI: {10.1016/j.freeradbiomed.2016.12.041}.

-
- [278] A. Singh et al. "Small Molecule Inhibitor of NRF2 Selectively Intervenes Therapeutic Resistance in KEAP1-Deficient NSCLC Tumors". In: *ACS CHEMICAL BIOLOGY* 11.11 (2016), 3214–3225. DOI: {10.1021/acscchembio.6b00651}.
- [279] N. Sato et al. "Metformin attenuates lung fibrosis development via NOX4 suppression". In: *RESPIRATORY RESEARCH* 17 (2016). DOI: {10.1186/s12931-016-0420-x}.
- [280] M. Markowicz-Piasecka et al. "Metformin - a Future Therapy for Neurodegenerative Diseases". In: *PHARMACEUTICAL RESEARCH* 34.12 (2017), 2614–2627. DOI: {10.1007/s11095-017-2199-y}.
- [281] M. Schuiveling et al. "Metformin, A New Era for an Old Drug in the Treatment of Immune Mediated Disease?" In: *Current Drug Targets* 19.8 (2018), 945–959. DOI: 10.2174/1389450118666170613081730.
- [282] B. M. Heckman-Stoddard et al. "Repurposing metformin for the prevention of cancer and cancer recurrence". In: *DIABETOLOGIA* 60.9 (2017), 1639–1647. DOI: {10.1007/s00125-017-4372-6}.
- [283] M. Y. El-Mir et al. "Dimethylbiguanide inhibits cell respiration via an indirect effect targeted on the respiratory chain complex I". In: *JOURNAL OF BIOLOGICAL CHEMISTRY* 275.1 (2000), 223–228. DOI: {10.1074/jbc.275.1.223}.
- [284] M. R. Owen, E. Doran, and A. P. Halestrap. "Evidence that metformin exerts its anti-diabetic effects through inhibition of complex 1 of the mitochondrial respiratory chain". In: *BIOCHEMICAL JOURNAL* 348.3 (2000), 607–614. DOI: {10.1042/0264-6021:3480607}.
- [285] D. Hardie. "The AMP-activated protein kinase pathway - new players upstream and downstream". In: *JOURNAL OF CELL SCIENCE* 117.23 (2004), 5479–5487. DOI: {10.1242/jcs.01540}.
- [286] G. Rena, D. G. Hardie, and E. R. Pearson. "The mechanisms of action of metformin". In: *DIABETOLOGIA* 60.9 (2017), 1577–1585. DOI: {10.1007/s00125-017-4342-z}.

- [287] L. Wang et al. "MD2 Blockage Protects Obesity-Induced Vascular Remodeling via Activating AMPK/Nrf2". In: *OBESITY* 25.9 (2017), 1532–1539. DOI: {10.1002/oby.21916}.
- [288] C. Zhao et al. "Fortunellin protects against high fructose-induced diabetic heart injury in mice by suppressing inflammation and oxidative stress via AMPK/Nrf-2 pathway regulation". In: *BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS* 490.2 (2017), 552–559. DOI: {10.1016/j.bbrc.2017.06.076}.
- [289] X. Ci et al. "Betulin exhibits anti-inflammatory activity in LPS-stimulated macrophages and endotoxin-shocked mice through an AMPK/AKT/Nrf2-dependent mechanism". In: *CELL DEATH & DISEASE* 8 (2017). DOI: {10.1038/cddis.2017.39}.
- [290] H. Lv et al. "Xanthohumol ameliorates lipopolysaccharide (LPS)-induced acute lung injury via induction of AMPK/GSK3 beta-Nrf2 signal axis". In: *REDOX BIOLOGY* 12 (2017), 311–324. DOI: {10.1016/j.redox.2017.03.001}.
- [291] G. Ashabi et al. "Pre-treatment with metformin activates Nrf2 antioxidant pathways and inhibits inflammatory responses through induction of AMPK after transient global cerebral ischemia". In: *METABOLIC BRAIN DISEASE* 30.3 (2015), 747–754. DOI: {10.1007/s11011-014-9632-2}.
- [292] M. A. Kaisar et al. "Offsetting the impact of smoking and e-cigarette vaping on the cerebrovascular system and stroke injury: Is Metformin a viable countermeasure?" In: *REDOX BIOLOGY* 13 (2017), 353–362. DOI: {10.1016/j.redox.2017.06.006}.
- [293] D. Kocer, F. Bayram, and H. Diri. "The effects of metformin on endothelial dysfunction, lipid metabolism and oxidative stress in women with polycystic ovary syndrome". In: *GYNECOLOGICAL ENDOCRINOLOGY* 30.5 (2014), 367–371. DOI: {10.3109/09513590.2014.887063}.

-
- [294] M. T. Kelleni, E. F. Amin, and A. M. Abdelrahman. "Effect of Metformin and Sitagliptin on Doxorubicin-Induced Cardiotoxicity in Rats: Impact of Oxidative Stress, Inflammation, and Apoptosis". In: *JOURNAL OF TOXICOLOGY* (2015). DOI: {10.1155/2015/424813}.
 - [295] H.-H. Tsai et al. "Metformin promotes apoptosis in hepatocellular carcinoma through the CEBPD-induced autophagy pathway". In: *ONCOTARGET* 8.8 (2017), 13832–13845. DOI: {10.18632/oncotarget.14640}.
 - [296] R. Pantan et al. "Synergistic effect of atorvastatin and Cyanidin-3-glucoside on angiotensin II-induced inflammation in vascular smooth muscle cells". In: *EXPERIMENTAL CELL RESEARCH* 342.2 (2016), 104–112. DOI: {10.1016/j.yexcr.2016.02.017}.
 - [297] W.-J. Wu et al. "S-propargyl-cysteine attenuates inflammatory response in rheumatoid arthritis by modulating the Nrf2-ARE signaling pathway". In: *REDOX BIOLOGY* 10 (2016), 157–167. DOI: {10.1016/j.redox.2016.08.011}.
 - [298] A.-R. Hwang et al. "Fluvastatin inhibits AGE-induced cell proliferation and migration via an ERK5-dependent Nrf2 pathway in vascular smooth muscle cells". In: *PLOS ONE* 12.5 (2017). DOI: {10.1371/journal.pone.0178278}.
 - [299] A. Abdanipour et al. "Evaluation of Lovastatin Effects on Expression of Anti-apoptotic Nrf2 and PGC-1 alpha Genes in Neural Stem Cells Treated with Hydrogen Peroxide". In: *MOLECULAR NEUROBIOLOGY* 49.3 (2014), 1364–1372. DOI: {10.1007/s12035-013-8613-5}.
 - [300] G. Marrone et al. "KLF2 exerts antifibrotic and vasoprotective effects in cirrhotic rat livers: behind the molecular mechanisms of statins". In: *GUT* 64.9 (2015), 1434–1443. DOI: {10.1136/gutjnl-2014-308338}.
 - [301] Y.-E. Cho et al. "Integrative analysis of proteomic and transcriptomic data for identification of pathways related to simvastatin-induced hepatotoxicity". In: *PROTEOMICS* 13.8, SI (2013), 1257–1275. DOI: {10.1002/pmic.201200368}.

- [302] H. J. Jang et al. "Simvastatin induces heme oxygenase-1 via NF-E2-related factor 2 (Nrf2) activation through ERK and PI3K/Akt pathway in colon cancer". In: *ONCOTARGET* 7.29 (2016), 46219–46229. DOI: {10.18632/oncotarget.10078}.
- [303] C.-H. Lin et al. "Lovastatin protects neurite degeneration in LRRK2-G2019S parkinsonism through activating the Akt/Nrf pathway and inhibiting GSK3 beta activity". In: *HUMAN MOLECULAR GENETICS* 25.10 (2016), 1965–1978. DOI: {10.1093/hmg/ddw068}.
- [304] E. Beurel, S. F. Grieco, and R. S. Jope. "Glycogen synthase kinase-3 (GSK3): Regulation, actions, and diseases". In: *PHARMACOLOGY & THERAPEUTICS* 148 (2015), 114–131. DOI: {10.1016/j.pharmthera.2014.11.016}.
- [305] M. Maqbool and N. Hoda. "GSK3 Inhibitors in the Therapeutic Development of Diabetes, Cancer and Neurodegeneration: Past, Present and Future". In: *CURRENT PHARMACEUTICAL DESIGN* 23.29 (2017), 4332–4350. DOI: {10.2174/1381612823666170714141450}.
- [306] I. Khan et al. "Natural and synthetic bioactive inhibitors of glycogen synthase kinase". In: *EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY* 125 (2017), 464–477. DOI: {10.1016/j.ejmech.2016.09.058}.
- [307] C.-T. Chiu et al. "Therapeutic Potential of Mood Stabilizers Lithium and Valproic Acid: Beyond Bipolar Disorder". In: *PHARMACOLOGICAL REVIEWS* 65.1 (2013), 105–142. DOI: {10.1124/pr.111.005512}.
- [308] M. Dolores Martin-de Saavedra et al. "Nrf2 participates in depressive disorders through an anti-inflammatory mechanism". In: *PSYCHONEUROENDOCRINOLOGY* 38.10 (2013), 2010–2022. DOI: {10.1016/j.psyneuen.2013.03.020}.
- [309] A. E. Freitas et al. "Agmatine, by Improving Neuroplasticity Markers and Inducing Nrf2, Prevents Corticosterone-Induced Depressive-Like Behavior in Mice". In: *MOLECULAR NEUROBIOLOGY* 53.5 (2016), 3030–3045. DOI: {10.1007/s12035-015-9182-6}.

-
- [310] W. Yao et al. "Role of Keap1-Nrf2 signaling in depression and dietary intake of glucoraphanin confers stress resilience in mice". In: *SCIENTIFIC REPORTS* 6 (2016). DOI: {10.1038/srep30659}.
- [311] C. Xiaobo et al. "TUSC2(FUS1)-erlotinib Induced Vulnerabilities in Epidermal Growth Factor Receptor(EGFR) Wildtype Non-small Cell Lung Cancer(NSCLC) Targeted by the Repurposed Drug Auranofin". In: *Scientific Reports* 6.1 (2016). DOI: 10.1038/srep35741.
- [312] A. Houessinon et al. "Metallothionein-1 as a biomarker of altered redox metabolism in hepatocellular carcinoma cells exposed to sorafenib". In: *MOLECULAR CANCER* 15 (2016). DOI: {10.1186/s12943-016-0526-2}.
- [313] D. Lee et al. "Folate cycle enzyme MTHFD1L confers metabolic advantages in hepatocellular carcinoma". In: *JOURNAL OF CLINICAL INVESTIGATION* 127.5 (2017), 1856–1872. DOI: {10.1172/JCI90253}.
- [314] Y. Yoshizaki et al. "Drug-Repositioning Screening for Keap1-Nrf2 Binding Inhibitors using Fluorescence Correlation Spectroscopy". In: *SCIENTIFIC REPORTS* 7 (2017). DOI: {10.1038/s41598-017-04233-3}.
- [315] J. Lamb et al. "The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease". In: *SCIENCE* 313.5795 (2006), 1929–1935. DOI: {10.1126/science.1132939}.
- [316] F. Iorio et al. "Identification of small molecules enhancing autophagic function from drug network analysis". In: *AUTOPHAGY* 6.8 (2010), 1204+. DOI: {10.4161/auto.6.8.13551}.
- [317] Q.-Y. Zhang et al. "Identification of Non-Electrophilic Nrf2 Activators from Approved Drugs". In: *Molecules* 22.6 (2017), p. 883. DOI: 10.3390/molecules22060883.

- [318] M. Xiong et al. "Identification of transcription factors for drug-associated gene modules and biomedical implications". In: *BIOINFORMATICS* 30.3 (2014), 305–309. DOI: {10.1093/bioinformatics/btt683}.
- [319] P. Ghezzi et al. "The oxidative stress theory of disease: levels of evidence and epistemological aspects". In: *British Journal of Pharmacology* 174.12 (2016), 1784–1796. DOI: 10.1111/bph.13544.
- [320] J. Frijhoff et al. "Clinical Relevance of Biomarkers of Oxidative Stress". In: *ANTIOXIDANTS & REDOX SIGNALING* 23.14 (2015), 1144–1170. DOI: {10.1089/ars.2015.6317}.
- [321] A. Comandini et al. "Markers of anti-oxidant response in tobacco smoke exposed subjects: A data-mining review". In: *PULMONARY PHARMACOLOGY & THERAPEUTICS* 23.6, SI (2010), 482–492. DOI: {10.1016/j.pupt.2010.05.006}.
- [322] L. M. Aleksunes, M. Goedken, and J. E. Manautou. "Up-regulation of NAD(P)H quinone oxidoreductase 1 during human liver injury". In: *WORLD JOURNAL OF GASTROENTEROLOGY* 12.12 (2006), 1937–1940. DOI: {10.3748/wjg.v12.i12.1937}.
- [323] E. Archer, G. Pavea, and C. J. Lavie. "The Inadmissibility of What We Eat in America and NHANES Dietary Data in Nutrition and Obesity Research and the Scientific Formulation of National Dietary Guidelines". In: *MAYO CLINIC PROCEEDINGS* 90.7 (2015), 911–926. DOI: {10.1016/j.mayocp.2015.04.009}.
- [324] P. Ghezzi et al. "The oxidative stress theory of disease: levels of evidence and epistemological aspects". In: *BRITISH JOURNAL OF PHARMACOLOGY* 174.12, SI (2017), 1784–1796. DOI: {10.1111/bph.13544}.
- [325] M. P. Cajaraville et al. "The use of biomarkers to assess the impact of pollution in coastal environments of the Iberian Peninsula: a practical approach". In: *SCIENCE OF THE TOTAL ENVIRONMENT* 247.2-3 (2000). 2nd Iberian Congress on Environmental Contamination and Toxicology, BASQUE COUNTRY,

-
- SPAIN, JUN 08-12, 1998, 295–311. DOI: {10.1016/S0048-9697(99)00499-4}.
- [326] A. J. Onderdijk et al. “Regulated genes in psoriatic skin during treatment with fumaric acid esters”. In: *BRITISH JOURNAL OF DERMATOLOGY* 171.4 (2014), 732–741. DOI: {10.1111/bjd.13128}.
- [327] D. S. Hong et al. “A Phase I First-in-Human Trial of Bardoxolone Methyl in Patients with Advanced Solid Tumors and Lymphomas”. In: *CLINICAL CANCER RESEARCH* 18.12 (2012), 3396–3406. DOI: {10.1158/1078-0432.CCR-11-2703}.
- [328] D. W. Cescon et al. “NRF2 Pathway Activation and Adjuvant Chemotherapy Benefit in Lung Squamous Cell Carcinoma”. In: *CLINICAL CANCER RESEARCH* 21.11 (2015), 2499–2505. DOI: {10.1158/1078-0432.CCR-14-2206}.

From single drug targets to synergistic network pharmacology in ischemic stroke

Adapted from: Ana I. Casas, **Ahmed A. Hassan**, Simon J. Larsen, Vanessa Gomez-Rangel, Mahmoud Elbatreek, Pamela W. M. Kleikers, Emre Guney, Javier Egea, Manuela G. López, Jan Baumbach, and Harald H. H. W. Schmidt. "From single drug targets to synergistic network pharmacology in ischemic stroke". In: *Proceedings of the National Academy of Sciences*.

Abstract

4 Drug discovery faces an efficacy crisis to which ineffective mainly single-target and symptom-based rather than mechanistic approaches have contributed. We here explore a mechanism-based disease definition for network pharmacology. Beginning with a primary causal target, we extend this to a second using guilt-by-association analysis. We then validate our prediction and explore synergy using both cellular in vitro and mouse in vivo models. As a disease model we chose ischemic stroke, one of the highest unmet medical need indications in medicine, and reactive oxygen species forming NADPH oxidase type 4 (Nox4) as a primary causal therapeutic target. For network analysis, we use classical protein–protein interactions but also metabolite-dependent interactions. Based on this protein–metabolite network, we conduct a gene ontology-based semantic similarity ranking to find suitable synergistic cotargets for network pharmacology. We identify the nitric oxide synthase (Nos1 to 3) gene family as the closest target to Nox4. Indeed, when combining a NOS and a NOX inhibitor at subthreshold concentrations, we observe pharmacological synergy as evidenced by reduced cell death, reduced infarct size, stabilized blood–brain barrier, reduced reoxygenation-induced leakage, and preserved neuromotor function, all in a supraadditive manner. Thus, protein–metabolite network analysis, for example guilt by association, can predict and pair synergistic mechanistic disease targets for systems medicine-driven network pharmacology. Such approaches may in the future reduce the risk of failure in single-target and symptom-based drug discovery and therapy.

Significance

Current one drug–one target–one disease approaches in drug discovery have become increasingly inefficient. Network pharmacology defines disease mechanisms as networks best targeted by multiple, synergistic drugs. Using the high unmet medical need indication stroke, we here develop an integrative in silico approach based on a primary target, NADPH oxidase type 4, to identify a mechanistically related cotarget, NO synthase, for network pharmacology. Indeed, we validate both in vivo and in vitro, including humans, that both NOX4 and NOS inhibition is highly synergistic, leading to a significant reduction of infarct volume, direct neuroprotection, and blood–brain-barrier stabilization. This systems medicine approach provides a ground plan to decrease current failure in the field by being implemented in other complex indications.

4.1 Introduction

In drug discovery, a “one disease–one target–one drug” approach is common practice, primarily to simplify compound screening, reduce unwanted side effects, and simplify registration [1]. This approach, however, oversimplifies disease mechanisms, which are in fact complex subnetworks within the interactome [2, 3]. Moreover, disease definitions are mostly symptom- rather than mechanism-based, and hence the therapeutics are likewise. Not surprisingly, drug discovery has thus become increasingly inefficient [4]. Conversely, systems medicine and network pharmacology define diseases according to causal mechanisms [5, 6]. Moreover, network pharmacology aims to further enhance this by targeting not only a single component within such a network but by combining drugs within these networks with the aim of achieving synergy and dose reduction [7]. However, most network databases are curated [8]; the de novo identification of such networks is only in its beginning. De novo network enrichment from a single primary validated target toward at least one secondary target holds high promise for systems medicine [9] but is currently not possible

To address this challenge, we designed an approach that (i) is integrative, (ii) is based on the network pharmacology paradigm, (iii) predicts targets instead of drugs, (iv) is validated through experiment, and (v) is readily applicable by a broad range of biomedical scientists. In fact, our approach can be established as a powerful tool and therefore implemented in novel, complex, and frequently unexpected indications where already-marketed drugs can be repurposed, leading to new therapies. Our strategy amends the limitations of previous approaches, for example, simple pairwise combination of drugs as opposed to targeting networks [10, 11], or combining drugs, which may have different off-target effects, rather than drug targets [10–12]. Furthermore, most proposed computational methods have not been validated experimentally for de novo predictions [10–14]. Moreover, most of these methods rely on drug similarity signatures extracted from chemical structures, targets, and side effect profiles, introducing a potential bias toward the pharmacological classes currently represented in knowledge bases [15] and limiting their applicability to de novo candidate discovery [16].

We therefore here develop a simple and integrative in silico approach to pair an existing validated primary causal therapeutic target with a synergistic cotarget within a network pharmacology strategy. We importantly validate our prediction both in vitro and in vivo, including a suitable in vitro human model. As a disease model for this we chose ischemic stroke, a multifactorial high unmet medical need indication for which no neuroprotective therapy is currently available. As a mechanistic starting point (node), we selected the reactive oxygen-forming enzyme NADPH oxidase type 4 (NOX4), a preclinically highly validated target directly involved in neurotoxicity and post stroke blood–brain barrier dysfunction [17, 18].

4.2 Results

4.2.1 Guilt-by-Association Analysis and Network Construction

To identify synergistic and mechanistically related cotargets for NOX4, we employed a guilt-by-association analysis on a multilayered molecular interaction network. Since many signaling events are gov-

erned by intermediate metabolites rather than protein–protein interactions [19], we considered this approach alone as insufficient to search for secondary targets. We therefore combined protein–protein interactions with protein–metabolite interactions to overcome such a potential bias or limitation.

We adopted a bottom-up approach consisting of three interacting computational modules starting from a well-known clinical target in stroke, NOX4 [17], as our primary target protein and seed node (Fig. 4.1). In module 1, we expanded from this seed node to obtain a network of candidate targets and related metabolites, resulting in five metabolites which were extended to their interactors, yielding 537 proteins. The main product of Nox4 [20], hydrogen peroxide (H₂O₂), and the substrate, oxygen (O₂), were manually added due to their absence from the Human Metabolome Database (HMDB), and a curation request was sent to the database. As a filtering step for narrowing down the interactions search space, drug–target interactions were used, resulting in 166 potential druggable target proteins. In module 2, a protein–protein interaction network was constructed based on the previously obtained druggable target proteins (Fig. 4.1). Subsequently, networks from modules 1 and 2 were combined to obtain a two-layered network determining the closest interaction partners of our primary target by means of guilt by association [21]. Hence, several levels of connectedness to NOX4 were observed, via direct protein interactions or indirect metabolic interactions (Fig. 4.2A), of which we consider the highest level, including nine proteins, as suitable NOX4-synergistic targets (Table 4.1). The full list of protein connectedness is reported in Table 4.S2.

4.2.2 Semantic Similarity of Gene Ontology Terms Affirms Network Analysis Results

Semantic similarity quantifies the closeness or relatedness of two strings or terms, in our case the different gene ontology (GO) molecular function annotations [22]. In module 3 of our approach (Fig. 4.1), we computed a single score measuring the similarity of each GO term pair, which was then employed to compare the functional relatedness of two proteins. In brief, the functional relatedness score of two pro-

Table 4.1: Network proteins ranked according to their connectedness to NOX4 through its metabolites

Protein symbol	Protein name	UniProt ID	Connectedness to NOX4
NOS1	Nitric oxide synthase, brain	P29475	4
NOS2	Nitric oxide synthase, inducible	P35228	4
NOS3	Nitric oxide synthase, endothelial	P29474	4
HMOX1	Heme oxygenase 1	P09601	4
HMOX2	Heme oxygenase 2	P30519	4
DUOX1	Dual oxidase 1	Q9NRD9	4
DUOX2	Dual oxidase 2	Q9NRD8	4
PPOX	Protoporphyrinogen oxidase	P50336	4
AOX1	Aldehyde oxidase	Q06278	4

teins was calculated by combining the similarity scores of every possible pair of GO terms annotating the two proteins. For scoring term pairs, we used the Wang et al. method [23] due to its ability to infer similarity according to the GO hierarchy, and not only the immediate terms in comparison. To combine these scores into the functional relatedness score of two proteins, we used the best average match strategy, as it accounts for both similar and dissimilar terms and is less affected by the number of terms available for comparison [22]. Based on the assumption that functions of proteins act as a proxy for structural and biological similarity, we ranked the proteins according to their functional relatedness to NOX4 calculated based on GO similarity scores. Finally, the candidate proteins were filtered to extract the top 10 targets functionally most similar to NOX4 (Fig. 4.2B).

The intersection of the outcome of the semantic analysis with the list of the most connected targets from the network analysis narrowed down the candidate list of targets to only four: CYBB, NOS2, NOS3, and NOS1, which ranked as the topmost functionally similar drug targets, with similarity scores based on the molecular functions from the GO annotations of 0.87, 0.70, 0.67, and 0.67, respectively (Fig. 4.2B and C). NOX1 also showed an equivalent score to NOS1; however, previous studies using a combined preclinical metaanalysis described that NOX1 plays no role in brain ischemia [24]. Moreover, NOX4KO mice treated with a NOX inhibitor showed no additional effect, suggesting no additional NOX1-dependent mechanism in stroke [17]. Having predicted a close connection between NOX4 and the NOS enzyme

family by in silico hybrid protein–metabolic network analysis, we next wanted to validate our finding stepwise, first in vitro, then in vivo, with respect to mechanistic synergy and thus applicability for network pharmacology.

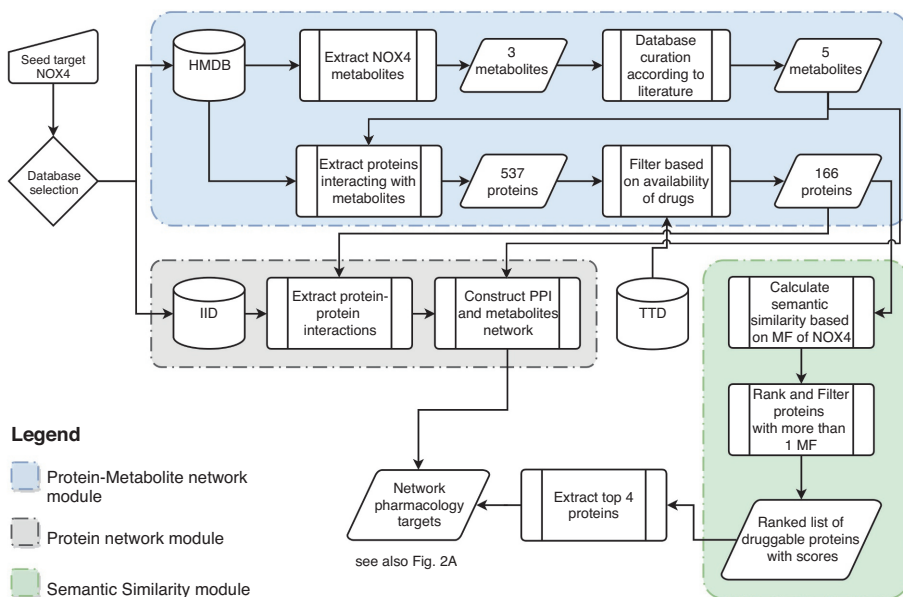


Figure 4.1: Computational workflow for target prioritization via network pharmacology. The computational target prioritization pipeline consists of three interdependent modules. The blue module extracts the metabolites interacting with the protein NOX4 from the Human Metabolome Database, performs curation of the metabolites, extracts the proteins interacting with them, and filters them based on the availability of drugs from the Therapeutic Target Database (TTD). The gray module uses the Integrated Interaction Database (IID) to extract protein–protein interactions of the proteins yielded by the blue module and constructs a network out of them. The green module calculates gene ontology-based semantic similarity scores of the output of the blue module compared with NOX4 using molecular function (MF) annotations, ranks the proteins based on their similarity scores, and excludes proteins with less than one molecular function. The output of the green module is used to annotate the network with the top four proteins.

4.2.3 In Vitro Cotarget Validation and Drug Identification

For in vitro validation, we used two models: an organotypic hippocampal culture (OHC) and human brain microvascular endothelial

cells as a blood–brain barrier model. In the OHC model, oxygen and glucose deprivation (OGD) followed by reoxygenation (Fig. 4.3A) resulted in the increased expression of our primary target, NOX4 [17, 18] (Fig. 4.3B), and all different NOS isoforms (Fig. 4.3C) within 2, 4, 8, 12, and 24 h post-OGD. Combinatory treatment with subthreshold concentrations of the NOX4 inhibitor GKT136901 (0.1 μ M) and the NOS inhibitor L-NAME (0.3 μ M) significantly reduced cell death (Fig. 4.3D) and formation of reactive oxygen and nitrogen species 24 h post-OGD, while individual treatment with these subthreshold concentrations had no effect (Fig. 4.3E). Likewise, early kinetics (15-min) assessment of reactive oxygen species (ROS) formation postcotreatment reflected a significant reduction compared with single therapies. Similarly, in the human blood–brain barrier model, cotreatment with the same subthreshold concentrations of GKT136901 (0.1 μ M) and L-NAME (0.3 μ M) reduced cell death (Fig. 4.3F) and prevented the increase in permeability after hypoxia (Fig. 4.3G). These data validated both the mechanism-based nature of NOX4 and NOS as a target and their *in silico* predicted synergistic interaction, since chosen monotherapies were not significantly effective.

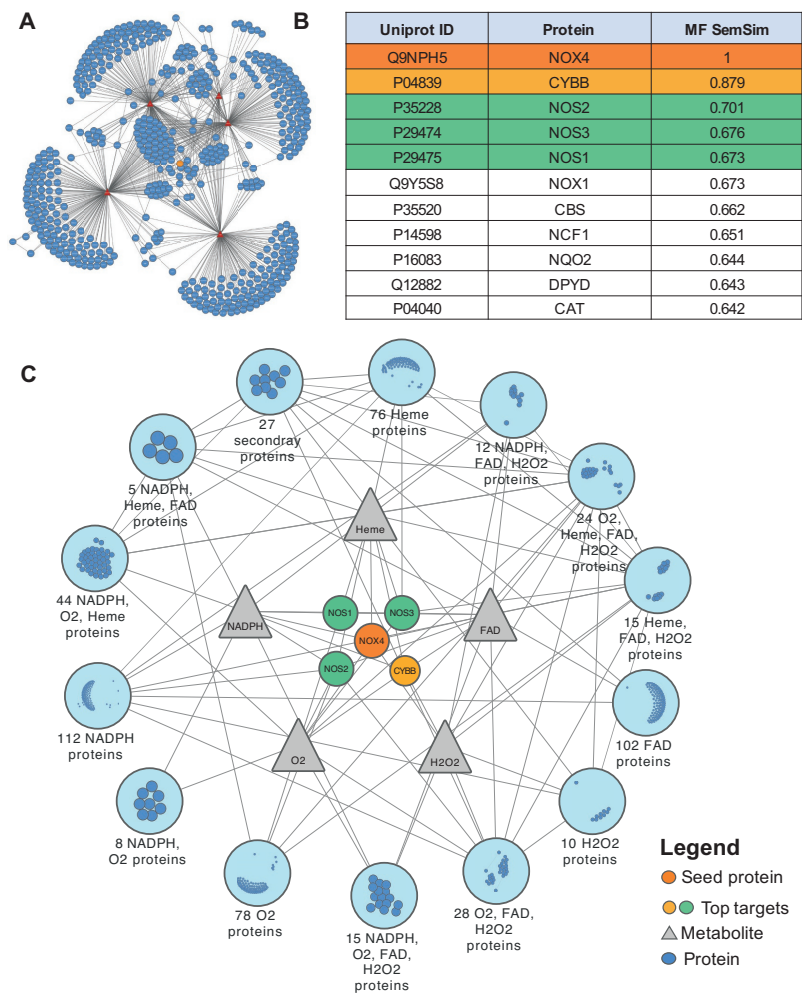


Figure 4.2: Integrated NOX4-extended multilayer network of biomolecular interactions used for candidate extraction and the involved protein semantic similarity ranking. (A) The full network constructed using the primary protein, NOX4 (orange node), connected to its direct metabolic interactors (red nodes), which have been linked to the proteins (blue nodes) interacting with them. We also show all protein–protein and metabolite–protein interactions (gray edges). (B) The semantic similarity ranking based on molecular functions (MF SemSim) of proteins with the top four similar proteins is highlighted. (C) The simplified network with only the primary protein, and the top four similar proteins and metabolites shown individually, while the rest of the proteins are grouped as modules and their interactions are merged.

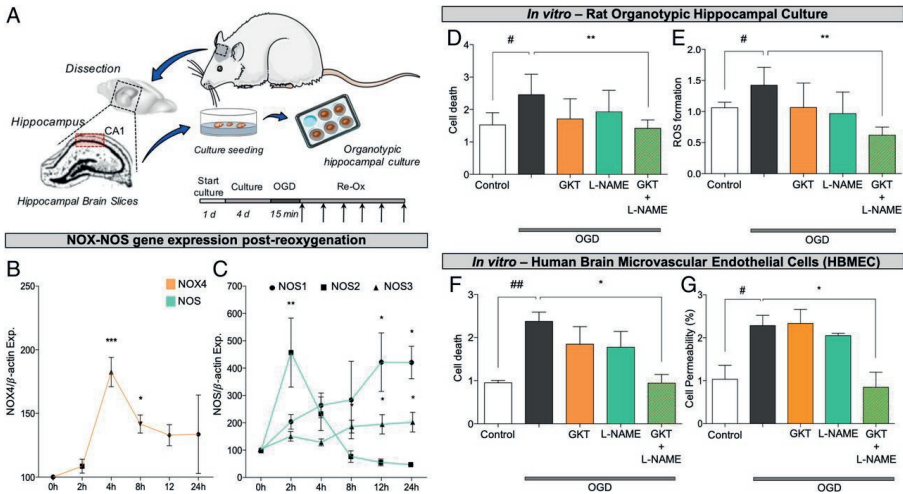


Figure 4.3: In vitro cotarget validation and drug identification of NOX4 and NOS inhibitors as a combinatory treatment. (A) Organotypic hippocampal cultures prepared from hippocampal slices were cultured for 4 d and subsequently subjected to 15 min of OGD period followed by 24-h treatment. Samples for gene expression detection were collected at 0, 2, 4, 8, 12, and 24 h post-OGD. (B) NOX4 expression was up-regulated at 4 and 8 h in comparison with the beginning of the ischemia period (* $P < 0.05$, *** $P < 0.001$; $n = 3$). (C) Inducible NOS (NOS2; square) was up-regulated only in the first 2 h post-OGD, while neuronal NOS (NOS1; circle) was up-regulated in the final 12 to 24 h after the OGD period. Similarly, endothelial NOS (NOS3; triangle) was also significantly up-regulated at 8, 12, and 24 h post-OGD (* $P < 0.05$, ** $P < 0.01$; $n = 4$). Gene expression was normalized using β -actin as housekeeping gene. (D) Cell death was significantly reduced in OHCs treated with GKT136901 (0.01 μ M) and L-NAME (0.3 μ M) in combination (** $P < 0.01$; $n = 8$; green slashed bar) in comparison with control slices (# $P < 0.05$ with respect to basal; $n = 8$; gray bar). Individual treatments show no effect. (E) ROS formation was also significantly decreased in OHCs treated with the combination of GKT136901 (0.01 μ M) and L-NAME (0.3 μ M) in comparison with nontreated slices. Again, individual treatments show no effect on cell death. # $P < 0.05$ compared with basal conditions (gray bar; $n = 5$); ** $P < 0.01$ with respect to nontreated slices (gray bar; $n = 5$). (F) Combinatory treatment of GKT136901 and L-NAME increases cell viability in human brain microvascular endothelial cells subjected to hypoxia/reoxygenation (Re-Ox). ## $P < 0.01$ with respect to basal conditions ($n = 4$; gray bar); * $P < 0.05$ with respect to nontreated cells ($n = 4$; green slashed bar). (G) Cell permeability was assessed by measuring Evans blue fluorescence post-OGD. Evans blue diffusion was significantly reduced in cells treated with GKT136901 (0.01 μ M) and L-NAME (0.3 μ M) in combination (# $P < 0.05$; $n = 4$; gray bar) in comparison with nontreated cells (* $P < 0.05$; $n = 4$; green slashed bar). Error bars are mean \pm SD.

4.2.4 In Vivo Validation of Network Pharmacology for Clinical Translation

To validate our network pharmacology approach in an in vivo model relevant for clinical translation, we used the mouse occlusion of the middle cerebral artery (MCAO) model in the absence or presence of GKT136901 (10 mg/kg) or L-NAME (3 mg/kg). Due to the many translational failures in stroke [25], the Stroke Treatment Academic Industry Roundtable (STAIR) established a set of guidelines to improve the success rate. Following these STAIR criteria, we assessed both a transient and permanent model, male and female, old and young mice. First, in transient MCAO, single subthreshold treatments showed no neuroprotection (Fig. 4.4A) but combinatory treatment significantly reduced brain infarctions compared with controls (Fig. 4.4A), both at 1 h and, importantly [26], 3 h poststroke (Fig. 4.4A), suggesting a wide therapeutic time window in agreement with our in vitro expression kinetics. Similar effects in a permanent MCAO model suggested therapeutic effect independent of reperfusion (Fig. 4.4A), and thus promise for patients where thrombolysis or/and thrombectomy is not recommended. In humans, stroke mostly occurs in the elderly population, and patient prognosis is directly influenced by age [27]. Thus, we confirmed these effects in aged female and male mice (Fig. 4.4A). Although smaller infarct sizes poststroke is an important read-out, neurofunctional outcome and quality of life postischemia are the main clinical parameters. Hence, we additionally assessed three independent neuromotor functioning tests in the adult mice treated 1 and 3 h poststroke together with the aged mice model: the Bederson score (Fig. 4.4B), elevated body swing test (Fig. 4.4C), and four-limb hanging wire test (Fig. 4.4D), which all were significantly improved 1 h postoperation (PO), and Bederson and the four-limb test also 3 h PO (Fig. 4.4 C–E). Monotherapies were only assessed 1 h poststroke treatment in adult mice due to ethical restrictions (Material and Methods). Thus, dual NOX/NOS inhibition poststroke leads to a potent synergistic, mechanism-based, and neuroprotective effect, further confirming that both targets are causally linked in a clinically translatable manner.

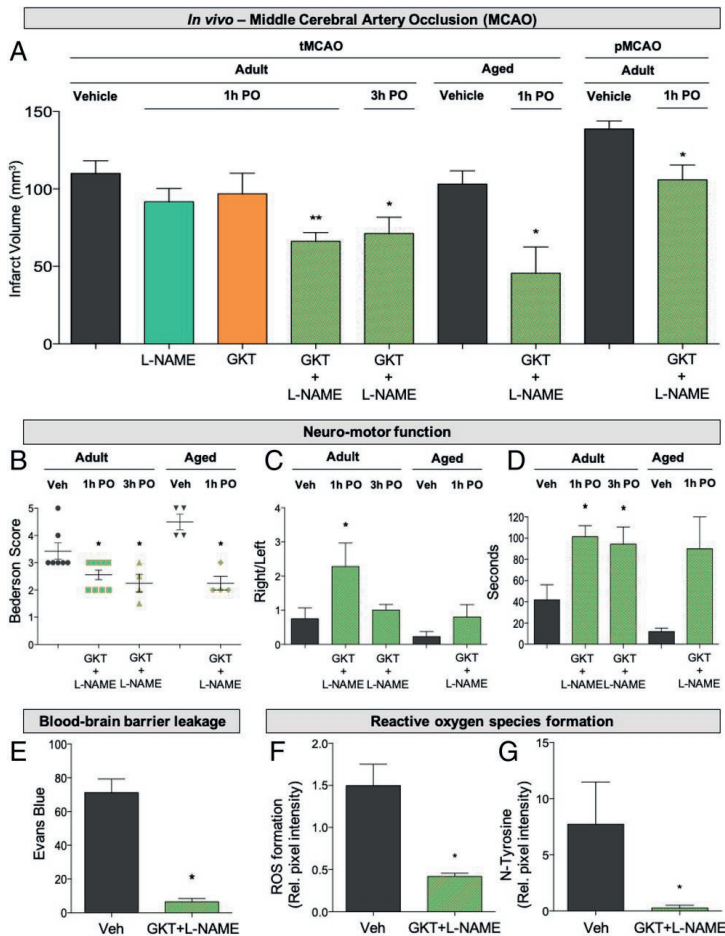


Figure 4.4: In vivo validation of network pharmacology for clinical translation. (A) Twenty-four hours after tMCAO infarct size was reduced in mice treated with GKT136901 (10 mg/kg) and L-NAME (3 mg/kg) in combination 1 h (**P < 0.01; n = 6) and 3 h poststroke (*P < 0.05; n = 5), while individual treatment showed no effect in reduction of infarct size. Infarct volume was also significantly reduced in aged animals treated with the combination (GKT+L-NAME) 1 h poststroke (**P < 0.01; n = 5). Similarly, combinatory treatment decreased infarct volume after permanent occlusion of the MCA in adult mice (*P < 0.05; n = 5). (B) With respect to the neurological outcome of the combinatory treatment in surviving mice, neurological outcome (Bederson score) was improved in the adult mice treated 1 h poststroke (*P < 0.05; n = 9), 3 h poststroke (*P < 0.05; n = 5), and the aged model (*P < 0.05; n = 4).

Figure 4.4 (cont.): (C) Likewise, the elevated body swing test indicated a significant increase for the right swing number/total swing number ratio in adult mice treated 1 h PO (* $P < 0.05$; $n = 9$) but not in the other groups. (D) Significantly improved motor outcome was detected after four-limb hanging test in all groups: 1 h PO (* $P < 0.05$; $n = 9$), 3 h PO (* $P < 0.05$; $n = 5$), and aged animals (* $P < 0.05$; $n = 4$). (E) Blood–brain barrier integrity assessed by Evans blue extravasation was preserved in treated animals compared with nontreated mice at day 1 after 1 h of tMCAO (* $P < 0.05$; $n = 4$). (F) Treated mice showed decreased ROS formation compared with their respective nontreated animals (* $P < 0.05$; $n = 4$). (G) N-Tyr–positive cells were significantly reduced with the combinatory therapy compared with nontreated mice. (* $P < 0.05$; $n = 4$). Error bars are mean \pm SD.

4.2.5 Prevention of Blood–Brain Barrier Disruption and ROS Formation upon Stroke Treatment

The cerebral vasculature, which is critical for the maintenance of the blood–brain barrier (BBB), is particularly susceptible to oxidative stress [28, 29]. To test whether dual inhibition of NOX/NOS leads to the blood–brain barrier phenotype, we assessed the integrity of the blood–brain barrier after ischemic stroke. In line with previous findings, combinatory treatment significantly reduced blood–brain barrier disruption upon stroke compared with nontreated mice (Fig. 4.4E).

To link the neuroprotective effect on the enzymatic activity of both ROS sources, we measured oxidative stress and N-Tyr formation in brain tissue cryosections. ROS generation and N-Tyr formation were dramatically reduced in treated mice (Fig. 4.4F) after 24 h of cotreatment (Fig. 4.4G), demonstrating a direct link in ROS reduction and a broad neuroprotective effect (Figs. 4.S2 and 4.S3).

4.3 Discussion

We here report a proof of concept for an *in silico* discovery approach to pair a single validated therapeutic drug target with another mechanistically related one for synergistic network pharmacology. Our multilayered interactome analysis including metabolites coupled with semantic similarity ranking detects pathomechanistically related proteins which can be cotargeted. Using this approach, we extend NOX4 to the closest functional neighbor

gene, NOS.

In search for a secondary, synergistic, and causal network pharmacology target prediction, data-driven or modeling-based techniques have been developed. A data-driven approach integrated multiple sources of data on drugs such as target proteins and their pathways, medical indications, therapeutic effects, and side effects [30]. DrugComboRanker prioritizes synergistic drug combinations [31] by constructing a functional drug network, although restricted to cancer drug–gene profiles. Here, community detection is performed via Bayesian nonnegative matrix factorization and, finally, similar drugs are inferred based on an adjacency matrix built from the drug–target network.

With respect to modeling, a network-based approach ranked combinations of proteins using a topological score calculated from an integrated protein–protein interaction (PPI) network constructed and enriched with gene expression data from a singular disease phenotype [13]. Random forests were used to predict drug combinations by exploiting network features generated from PPI data, and drug chemical and pharmacogenomic features from drug-induced expression profiles [11]. MASCOT is a model-driven machine-learning algorithm that leverages curated dynamic models of signaling networks and their disease states to predict synergistic targets of a desired therapeutic effect [14]. In comparison with all these previous reports, our two-step approach utilized experimental databases on multilevel molecular networks in a rather simple and generic manner, and our predictions were experimentally validated both in vitro and in vivo.

PPI networks or interactomes have been commonly used to understand complex disease mechanisms [32–36]. However, PPI networks are just one level of molecular interaction networks. Most signaling events are not due to wild-type PPIs but involve metabolites or metabolic protein modifications. We therefore used protein–metabolite networks in conjunction with PPI networks. In addition, this solved another issue of molecular interaction databases. Current databases suffer from selection and detection biases, high rates of false positives,

and low rates of coverage [5, 37].

However, metabolome databases still have severe shortcomings. When querying the most comprehensive HMDB with seven additional key signaling enzymes, key metabolites were consistently missing, namely for soluble guanylate cyclase, the substrate, GTP, and the product, 2',5'-cGMP, as well as GMP, cAMP, and 2',3'-cGMP being wrongly listed. Thus, our approach will improve considerably once these or other metabolic databases are intensely curated and become more complete.

Although our method is generic and can be applied to other cases in concept, there are restrictions. For instance, the current computational pipeline supports one seed target; however, it is generally applicable to more than one seed target. For multiple seed targets, one would have to look for druggable candidates with the shortest average distance in the graph representation of the integrated network to all seeds.

In the network analysis step, the patchiness of the data sources may affect predictions, while for the semantic similarity, the availability and accuracy of GO annotations may impact on the ranking. In addition, we report only a single validated application of our method, and further use cases will be needed. In fact, NOX4-related targets, namely NCF1, NQO2, and DPYD, might also show potential. Targeting NCF1, also known as p47 (NOX subunit), may lead to indirect NOX4 activation. However, modulating NCF1 is so far not possible, since protein-protein interaction inhibitors proved noneffective and no further network pharmacology strategy could be achieved [38]. Moreover, ribosyl dihydronicotinamide dehydrogenase (NQO2), a ROS-generating enzyme, shows a direct acetaminophen side effect, while this drug has been shown as protective in stroke, demonstrating a direct link [39]. However, when weighing NQO2 and NOS as a cotarget of NOX4, we would still prefer NOS, because with NOX4 and NQO2 we would both target ROS formation with possibly no synergy but rather additive effects.

We thus validated the therapeutic applicability of our in silico net-

work pharmacology hypothesis both in vitro and in vivo by coadministering both a NOX inhibitor and NOS inhibitor, respectively, in three different species including a human BBB model. Of high translational relevance, combining a NOX and NOS inhibitor conveyed in direct neuroprotection in three different brain ischemia models, rat organotypic hippocampal culture, transient and permanent MCAO in mice, and human brain microvascular cells as a BBB model. Importantly, this was achieved at concentrations and doses, respectively, that on their own were ineffective. This will allow extension of the clinical translation of NOX4 inhibition in stroke to be enhanced in efficacy and safety by lowering in risk of any potential side effects, increasing mechanistic-based synergy and reducing the number needed to treat. Thus, our multitargeted approach therefore focuses on NOX4 inhibition coadministered with a NOS inhibitor while, due to synergy, reducing the doses/concentrations of both drugs to individual sub-threshold levels.

On a mechanistic level, interaction between reactive oxygen species and NO, for example, to toxify NO via intermediate peroxynitrite formation, has been shown before [40]. The source of ROS, however, has not always been identified [41–43]. Also, the signaling networks of NOS/NO on the one side and ROS formation and ROS targets on the other have been annotated as rather independent. Here, we show that, at least in disease, both networks represent subnetworks of the same common mechanism that involves both NOS and NOX4. At least NOX4 is one relevant ROS source interacting with NO, or NO's downstream signaling and online pathways have been suggested for curation, as far as possible. Importantly, this does not imply that all ROS sources will interact as well with NOS. Such assumptions represent a shortcoming of the current curated NOS and ROS pathways, as they combine, for example, all ROS sources and all ROS targets into one scheme. With respect to the relevant NOS isoform, the best characterized and validated is NOS1 [44], whereas NOS3 is rather protective [45] and NOS2 expression commences only 12 h after the onset of the stroke in an in vivo rat model [46], while our findings suggest different expression in mice primarily due to model and species differences.

With respect to clinical application there are two other NOS inhibitors worth being considered apart from L-NAME, which has been tested for the longest time. First, Vasopharm is developing VAS203 as a NOS common inhibitor currently entering phase III clinical trials for traumatic brain injury. Since it is a NOS common inhibitor like L-NAME it has a similar spectrum, but concerns have arisen with respect to covalent off-target effects [47, 48] and depression of kidney function [49]. Second, S-methyl-L-thiocitrulline has also been tested in humans but appears to have only a limited spectrum, possibly limited to NOS1 [1], while our *in silico* prediction ranked NOS2 > NOS3 > NOS1. Such isoform-selective inhibitors will certainly be of interest in future studies aiming at deciphering the contribution of individual NOS isoforms in ROS–NO interaction.

Clearly, more *de novo* generated pathways as subgraphs of the interactome are necessary to eliminate such assumptions. Also, NOX4 generates H₂O₂, whereas typically superoxide, O₂⁻, is considered the key interfering molecule with respect to NO. Moreover, NOX4 has also been identified as a positive indirect transcriptional regulator of the major H₂S-producing enzyme, cystathionine β -synthase [50], which plays a key role in the central nervous system and circulation linked to worse poststroke outcome [51]. Thus, poststroke NOX4-dependent inhibition of the cystathionine β -synthase pathway may also result in at least additive effective effects in stroke within the same mechanistic network.

With respect to NOS1 to 3, the possibility exists that one of the isoforms may actually be protective. Importantly, the here-presented network approaches are by definition undirected, namely whether a cotarget needs to be inhibited or activated is not always immediately obvious. Chronically, NOS3 inhibition is certainly not of benefit [11]; however, in an acute setting, even endothelial NOS-derived NO may for a time window where it interacts with NOX4-derived ROS be detrimental. Pan-NOS inhibition is almost as effective as NOS1 knockout. Thus, pharmacological validation will in many cases remain an essential component when interpreting and curating network pharmacology discovery results.

Thus, from a chemical point of view, the NOX4–NOS interaction that we predicted and validated was surprising, and may involve a hitherto underappreciated interaction of H₂O₂ with transition metal centers to form singlet oxygen [40, 52].

In conclusion, our present and other network pharmacology approaches [1, 53] provide a roadmap to reduce the risk of failure in single drug target development by moving toward multiple targeting of de novo causal networks to increase therapeutic efficacy and reduce individual drug dosing and possible side effects due to mechanism-based synergy [53, 54]. We suggest extending our approach to other unmet medical need indications, where currently only single drug- or symptom-based approaches are available.

4.4 Material and Methods

Detailed experimental procedures are provided in Supplemental Material and Methods.

4.4.1 Study Design

All animal experiments were performed after approval of the protocol by the Institutional Ethics Committee of the Autonomous University of Madrid according to European guidelines for the use and care of animals for research. The dropout rates were four mice in the vehicle groups [transient (t)MCAO, aged, permanent (p)MCAO, and Evans blue] versus three mice in different treatment groups (tMCAO and pMCAO) (Table 4.S2). Post hoc power analysis for adult mice is included in Table 4.S3.

4.4.2 Transient Occlusion of the Middle Cerebral Artery

The model was previously described in [17].

4.4.3 Human Brain Microvascular Endothelial Cells Subjected to Hypoxia

Human brain microvascular endothelial cells (HBMECs) (Cell Systems) were cultured to ~90% confluence. Cell medium was

replaced by non-FBS-containing medium (2 mL per well) following 6 h of hypoxia (94.8% N₂, 0.2% O₂, and 5% CO₂) and 24 h of reperfusion in the presence or absence of the pharmacological treatment (see supplements section for details).

4.4.4 Statistical Analysis

All results obtained from the *in vitro* (hippocampal brain slices, OHCs, HBMECs) and *in vivo* (tMCAO) ischemia models were analyzed using Prism 5.0 software (GraphPad Software). Data were expressed as the means \pm SEM of separate experiments. Statistical comparisons between groups were performed using one-way ANOVA, followed by a Newman–Keuls multiple-comparison test. Differences between two groups were considered significant at $P < 0.05$. Numbers of animals necessary to detect a standardized effect size on infarct volumes ≥ 0.2 (vehicle-treated control mice vs. treated mice) were determined via a priori sample size calculation with the following assumptions: $\alpha = 0.05$; $\beta = 0.2$; 20% SD of the mean. In each case, when only two groups were compared, the unpaired two-tailed Student's *t* test was applied followed by a Mann–Whitney *U* test, where significance was considered at $P < 0.05$.

4.5 Acknowledgments

This study was supported by the ERC (Advanced Investigator Grant 294683/RadMed and Proof-of-Concept Grant 737586/SAVE-BRAIN, both to H.H.H.W.S.), Spanish Ministry of Economy and Competence (SAF2015-63935R to M.G.L.), Fondo de Investigaciones Sanitarias (Instituto de Salud Carlos III/Fondo Europeo de Desarrollo Regional) (Programa Miguel Servet: CP14/00008, PI16/00735), Fundación Mutua Madrileña (J.E.), short-term scientific missions by the COST Actions EU-ROS and OpenMultiMed, and Kootstra Talented Fellowship (UM) (to A.I.C.). J.B.'s work was financially supported by VILLUM Young Investigator Grant 13154. J.B. and H.H.H.W.S. also received support from H2020 Project 777111-REPO-TRIAL.

4.6 References

- [1] M. A. Yildirim et al. “Drug-target network”. In: *Nature Biotechnology* 25.10 (2007), pp. 1119–1126.
- [2] J. Aguirre-Plans et al. “Proximal Pathway Enrichment Analysis for Targeting Comorbid Diseases via Network Endopharmacology”. In: *Pharmaceuticals* 11.3 (2018), UNSP 61.
- [3] A. L. Hopkins. “Network pharmacology”. In: *Nature Biotechnology* 25.10 (2007), pp. 1110–1111.
- [4] J. W. Scannell et al. “Diagnosing the decline in pharmaceutical R&D efficiency”. In: *Nature Rev. Drug Discovery* 11.3 (2012), pp. 191–200.
- [5] J. Menche et al. “Uncovering disease-disease relationships through the incomplete interactome”. In: *Science* 347.6224 (2015), p. 1257601.
- [6] F. Langhauser et al. “A diseasome cluster-based drug repurposing of soluble guanylate cyclase activators from smooth muscle relaxation to direct neuroprotection”. In: *Npj Systems Biol. Applications* 4 (2018), UNSP 8.
- [7] A. L. Barabasi and Z. N. Oltvai. “Network biology: Understanding the cell’s functional organization”. In: *Nature Rev. Genetics* 5.2 (2004), 101–U15.
- [8] R. Batra et al. “On the performance of de novo pathway enrichment”. In: *Npj Systems Biol. Applications* 3 (2017), UNSP 6.
- [9] N. Alcaraz et al. “De novo pathway-based biomarker identification”. In: *Nucleic Acids Research* 45.16 (2017), e151.
- [10] P. Koduru and R. Chaganti. “Congenital chromosome breakage clusters within giemsa-light bands and identifies sites of chromatin instability”. In: *Cytogenetics Cell Genetics* 49.4 (1988), pp. 269–274.
- [11] X. Li et al. “Prediction of synergistic anti-cancer drug combinations based on drug target network and drug induced gene expression profiles”. In: *Artificial Intelligence In Medicine* 83 (2017), pp. 35–43.

-
- [12] M. V. G. "Time characteristics of local cerebral blood flow autoregulation in hypotension and hypertension". In: *Patologicheskaya Fiziologiya I Eksperimental'naya Terapiya* 4 (1985), pp. 29–32.
- [13] F. Vitali et al. "Network-based target ranking for polypharmacological therapies". In: *J. Biomedical Informatics* 46.5 (2013), pp. 876–881.
- [14] H. E. Chua, S. S. Bhowmick, and L. Tucker-Kellogg. "Synergistic target combination prediction from curated signaling networks: Machine learning meets systems biology and pharmacology". In: *Methods* 129 (2017), pp. 60–80.
- [15] S. Vilar and G. Hripcsak. "The role of drug profiles as similarity metrics: applications to repurposing, adverse effects detection and drug-drug interactions". In: *Briefings In Bioinformatics* 18.4 (2017), pp. 670–681.
- [16] E. Guney. "Reproducible drug repurposing: when similarity does not suffice". In: *Pacific Symposium On Biocomputing 2017* (2017), pp. 132–143.
- [17] A. I. Casas et al. "NOX4-dependent neuronal autotoxicity and BBB breakdown explain the superior sensitivity of the brain to ischemic damage". In: *Proc. National Acad. Sciences United States Am.* 114.46 (2017), pp. 12315–12320.
- [18] C. Kleinschnitz et al. "Post-Stroke Inhibition of Induced NADPH Oxidase Type 4 Prevents Oxidative Stress and Neurodegeneration". In: *Plos Biol.* 8.9 (2010), e1000479.
- [19] F. Li, W. Xu, and S. Zhao. "Regulatory Roles of Metabolites in Cell Signaling Networks". In: *J. Genetics Genomics* 40.7 (2013), pp. 367–374.
- [20] Y. Nisimoto et al. "Nox4: A Hydrogen Peroxide-Generating Oxygen Sensor". In: *Biochem.* 53.31 (2014), pp. 5111–5120.
- [21] D. Piovesan et al. "Protein function prediction using guilty by association from interaction networks". In: *Amino Acids* 47.12 (2015), pp. 2583–2592.

- [22] C. Pesquita et al. "Metrics for GO based protein semantic similarity: a systematic evaluation". In: *Bmc Bioinformatics* 9 (2008), S4.
- [23] J. Z. Wang et al. "A new method to measure the semantic similarity of GO terms". In: *Bioinformatics* 23.10 (2007), pp. 1274–1281.
- [24] P. W. M. Kleikers et al. "A combined pre-clinical meta-analysis and randomized confirmatory trial approach to improve data validity for therapeutic target validation". In: *Scientific Reports* 5 (2015), p. 13428.
- [25] V. E. O'Collins et al. "1,026 experimental treatments in acute stroke". In: *Annals Neurology* 59.3 (2006), pp. 467–477.
- [26] I. D. Pena et al. "Strategies to Extend Thrombolytic Time Window for Ischemic Stroke Treatment: An Unmet Clinical Need." In: *Journal Of Stroke* 19.1 (2017), pp. 50–60.
- [27] R.-L. Chen et al. "Ischemic stroke in the elderly: an overview of evidence". In: *Nature Rev. Neurology* 6.5 (2010), pp. 256–265.
- [28] C. Lehner et al. "Oxidative Stress and Blood-Brain Barrier Dysfunction Under Particular Consideration of Matrix Metalloproteinases". In: *Antioxidants & Redox Signaling* 15.5 (2011), pp. 1305–1323.
- [29] A.-M. Enciu, M. Gherghiceanu, and B. O. Popescu. "Triggers and Effectors of Oxidative Stress at Blood-Brain Barrier Level: Relevance for Brain Ageing and Neurodegeneration". In: *Oxidative Medicine Cellular Longevity* 2013 (2013), p. 297512.
- [30] X.-M. Zhao et al. "Prediction of Drug Combinations by Integrating Molecular and Pharmacological Data". In: *Plos Computational Biol.* 7.12 (2011), e1002323.
- [31] L. Huang et al. "DrugComboRanker: drug combination discovery based on target network analysis". In: *Bioinformatics* 30.12 (2014), pp. 228–236.
- [32] M. Vidal, M. E. Cusick, and A.-L. aszlo Barabasi. "Interactome Networks and Human Disease". In: *Cell* 144.6 (2011), pp. 986–998.

-
- [33] M. Atthias E. . Futschik, G. autam Chaurasia, and H. anspeter Herzel. "Comparison of human protein-protein interaction maps". In: *Bioinformatics* 23.5 (2007), pp. 605–611.
 - [34] B. Lehne and T. Schlitt. "Protein-protein interaction databases: Keeping up with growing interactomes". In: *Human Genomics* 3.3 (2009), pp. 291–297.
 - [35] T. Rolland et al. "A Proteome-Scale Map of the Human Interactome Network". In: *Cell* 159.5 (2014), pp. 1212–1226.
 - [36] V. Mehta and L. Trinkle-Mulcahy. "Recent advances in large-scale protein interactome mapping." In: *F1000research* 5 (2016).
 - [37] L. A. N. Amaral. "A truer measure of our ignorance". In: *Proc. National Acad. Sciences United States Am.* 105.19 (2008), pp. 6795–6796.
 - [38] S. Altenhofer et al. "Evolution of NADPH Oxidase Inhibitors: Selectivity and Mechanisms for Target Engagement". In: *Antioxidants & Redox Signaling* 23.5 (2015), pp. 406–427.
 - [39] T. P. Miettinen and M. Bjoerklund. "NQO2 Is a Reactive Oxygen Species Generating Off-Target for Acetaminophen". In: *Mol. Pharmaceutics* 11.12 (2014), pp. 4395–4404.
 - [40] R. Radi. "Oxygen radicals, nitric oxide, and peroxynitrite: Redox pathways in molecular medicine". In: *Proc. National Acad. Sciences United States Am.* 115.23 (2018), pp. 5839–5848.
 - [41] M. Geiszt. "NADPH oxidases: New kids on the block". In: *Cardiovascular Research* 71.2 (2006), pp. 289–299.
 - [42] M. G. Espey et al. "A chemical perspective on the interplay between NO, reactive oxygen species, and reactive nitrogen oxide species". In: *Nitric Oxide: Novel Actions, Deleterious Effects Clinical Potential* 962 (2002), pp. 195–206.
 - [43] P. Ghezzi et al. "The oxidative stress theory of disease: levels of evidence and epistemological aspects". In: *British J. Pharmacology* 174.12 (2017), pp. 1784–1796.
 - [44] C. Kleinschnitz et al. "NOS knockout or inhibition but not disrupting PSD-95-NOS interaction protect against ischemic brain damage". In: *J. Cerebral Blood Flow Metabolism* 36.9 (2016), pp. 1508–1512.

- [45] M. Endres et al. "Targeting eNOS for stroke protection". In: *Trends In Neurosciences* 27.5 (2004), pp. 283–289.
- [46] M. Niwa et al. "Time course of expression of three nitric oxide synthase isoforms after transient middle cerebral artery occlusion in rats". In: *Neurologia Medico-chirurgica* 41.2 (2001), pp. 63–72.
- [47] V. T.-V. Dao et al. "Pharmacology and Clinical Drug Candidates in Redox Medicine". In: *Antioxidants & Redox Signaling* 23.14 (2015), pp. 1113–1129.
- [48] Q.-A. Sun et al. "Off-target thiol alkylation by the NADPH oxidase inhibitor 3-benzyl-7-(2-benzoxazolyl)thio-1,2,3-triazolo[4,5-d]pyrimidine (VAS2870)". In: *Free Radical Biol. Medicine* 52.9 (2012), pp. 1897–1902.
- [49] C. Ott et al. "Effects of the nitric oxide synthase inhibitor ronopterin (VAS203) on renal function in healthy volunteers". In: *British J. Clinical Pharmacology* 85.5 (2019), pp. 900–907.
- [50] R. K. Mistry et al. "Transcriptional Regulation of Cystathionine-gamma-Lyase in Endothelial Cells by NADPH Oxidase 4-Dependent Signaling". In: *J. Biological Chem.* 291.4 (2016), pp. 1774–1788.
- [51] S. J. Chan et al. "Cystathionine beta-Synthase Inhibition Is a Potential Therapeutic Approach to Treatment of Ischemic Injury". In: *Asn Neuro* 7.2 (2015).
- [52] A. A. Noronhadutra, M. M. Epperlein, and N. Woolf. "Reaction of nitric-oxide with hydrogen-peroxide to produce potentially cytotoxic singlet oxygen as a model for nitric oxide-mediated killing". In: *Febs Lett.* 321.1 (1993), pp. 59–62.
- [53] A. L. Hopkins. "Network pharmacology: the next paradigm in drug discovery". In: *Nature Chem. Biol.* 4.11 (2008), pp. 682–690.
- [54] J. Tang and T. Aittokallio. "Network Pharmacology Strategies Toward Multi-Target Anticancer Therapies: From Computational Models to Experimental Design Principles". In: *Current Pharmaceutical Design* 20.1 (2014), pp. 23–36.

Supplemental Material and Methods

Network analysis and semantic ranking

NOX4 was considered as initial stroke related target and primary protein. The goal is to identify more related targets in silico using publicly available molecular interaction databases. Since associated proteins can interact directly (physical protein-protein interactions) or indirectly (through downstream reactions through metabolites), we opted to cover the two data sources. Therefore, we integrated two databases: i) the Integrated Interactions Database (IID) [1] for protein-protein interactions, and ii) the Human Metabolome Database (HMDB) [2] for protein-metabolite interactions (See Table 4.S4 for details on database versions, access dates, interactions, and evidence levels). Since IID is an integrated PPI database with experimentally detected interactions from several other PPI databases, predicted interactions and orthologous interactions; we chose to focus only on the experimentally detected interactions to limit the noise in our data. Integrating IID and HMDB gave rise to a two-layered network, which we constructed as follows: We first extracted the metabolites interacting with our primary target, NOX4, and using expert curation, we added important metabolites interactions that were missing from the HMDB database, namely oxygen (O₂) and hydrogen peroxide (H₂O₂). We subsequently extracted all proteins interacting with those metabolites yielding a two-layered NOX4-centric protein/metabolite-network. Since we are interested in identifying immediately clinically relevant targets, we filtered all proteins in the network to identify those annotated as targets of known drugs. To this end we utilized the Therapeutic Target Database (TTD) (See Table 4.S4) to obtain a list of druggable proteins, and we extracted the direct molecular interactions between them and thereby constructed the final, filtered network of potential targets. However, this network provides only a means of assessing the connectedness of proteins to our primary protein and not a quantitative measure to assess the functional similarity of candidate synergistic targets. Thus, we used a semantic analysis pipeline to rank the connected proteins. We utilized a strategy of scoring the similarity of molecular functions in the Gene Ontology terms of each of the proteins as opposed to

those of NOX4 using the semantic the GOSim R software package [3]. We used the Wang [4] method to compute the semantic similarity which yields a score in the range of 0 to 1, with 1 being identical to the primary protein. The Wang method is an ontology-based approach where the semantic similarities of the GO terms are calculated based on their locations in the directed acyclic graph (DAG) representing the GO ontology. It defines the semantic value of each term based on the sum of the weighted contributions of all its ancestors, where closer ancestor terms have higher contribution, and which is used to calculate the semantic similarity between GO term pairs. The similarity between two proteins is then derived from the semantic similarities of all GO terms annotating both proteins. Finally, we used Cytoscape to visualize both the expanded and the reduced network with top targets annotated.

Study design

All animal experiments were performed after approval of the protocol by the institutional Ethics Committee of Autonomous University of Madrid (Madrid, Spain) according to the European Guidelines for the use and care of animals for research. All efforts were made to minimize animal suffering and to reduce the number of animals used in the experiments. Animals were housed under controlled conditions ($22 \pm 1^\circ\text{C}$, 55–65% humidity, 12h light-dark cycle), and have been given free access to water and standard laboratory chow. 8-16 week adult male and female mice and 1-year old mice (aged) were used for the in vivo animal study. Animals were excluded from end-point analyses if death occurred within 24h after tMCAO or pMCAO and if intracerebral haemorrhage occurred as macroscopically assessed during brain sampling. The drop-out rates were 4 mice in the vehicle groups (tMCAO, aged, pMCAO and Evans Blue) versus 3 mice in different treatment groups (tMCAO and pMCAO) (Table 4.S2). Post-hoc power analysis for adult mice is included in Table 4.S3. 2-3 months old Sprague-Dawley rats were used for the in vitro studies.

Ex-vivo acute model: Preparation of hippocampal brain slices and induction of oxygen and glucose deprivation

Experiments were performed using hippocampal brain slices (HBS) from adult Sprague-Dawley rats (2-3 months) as previously described in [5, 6]. Rats were decapitated and forebrains were rapidly removed from the skull and placed into ice-cold Krebs bicarbonate dissection buffer (pH 7.4), containing: NaCl 120 mM, KCl 2 mM, CaCl₂ 0.5 mM, NaHCO₃ 26 mM, MgSO₄ 10 mM, KH₂PO₄ 1.18 mM, glucose 11 mM and sucrose 200 mM. The hippocampus was quickly dissected and subsequently cut into transverse slices 300 μ m-thick using a Tissue Chopper McIlwain. To help the tissue to recover from the slicing trauma, slices were incubated in Krebs buffer for 45 min at 34°C (stabilization period). Then, oxygen and glucose deprivation (OGD) was induced by incubating the slices during 15 min in a Krebs-bicarbonate solution where glucose was replaced by 2-deoxyglucose (OGD solution). In parallel, control slices were incubated during 15 min in a Krebs-bicarbonate solution without sucrose (control solution). At least 15 minutes before starting the experiment, both control and OGD solutions were bubbled with either 95% O₂/5% CO₂ or 95% N₂/5% CO₂ gas mixture to ensure O₂ and N₂ saturation respectively. After the OGD period, slices were returned to an oxygenated Krebs-bicarbonate solution containing glucose for 120 min (Re-oxygenation period). During this period, the slices were treated with 0.1, 0.3, 1 and 3 μ M GKT136901 and 0.3, 1 and 3 μ M L-NAME. Non-treated slices were used as 100% viability control. Synergy between compounds was considered since mono-therapies using sub-threshold concentrations/doses of both compounds were not therapeutically effective and effect size can be provided.

Cell viability of hippocampal brain slices

Hippocampal slices were collected immediately after the re-oxygenation period and subsequently incubated with 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT, Sigma-Aldrich, Spain) in Krebs bicarbonate solution (0.5 mg/ml) for 30 min at 37°C. The tetrazolium ring of MTT is cleaved by active dehydrogenases in viable cells, producing a precipitated

formazan derivative [7]. This formazan derivative was solubilized by adding 200 μ M DMSO. The optical density was measured spectrophotometrically at 540 nm using a micro plate reader. Absorbance values obtained in control slices were set to 100% viability and experimental variables were normalized with respect to this value.

Preparation of Organotypic Hippocampal Slices and induction of oxygen and glucose deprivation

4

Hippocampal brain slices for organotypic cultures (OHC) were obtained from brains of 7-10-days-old Sprague Dawley rats. Organotypic cultures were prepared based on the methods previously described in [8, 9]. Pups were quickly decapitated and forebrains removed from the skull and dissected. Hippocampus was cut into 300 μ m-thick slices using a Tissue Chopper McIlwain. Slices were separated in sterile ice-cold Hank's balanced salt solution (HBSS, Biowest, Madrid, Spain). Six slices were placed on each Millicell-0.4 μ m culture inserts (Millipore, Madrid, Spain). Each Millicell culture insert was placed in a well within a six-well culture plate. Neurobasal medium (Invitrogen, Madrid, Spain) enriched with 10% of fetal bovine serum (Sigma-Aldrich, Madrid, Spain) was used for the next 24h (1 ml/well). After 24h, B27 supplement and antioxidants were added to the culture medium. Slices were in culture for 4 days before inducing the OGD period. On day 6, neurobasal medium was replaced by 1 ml/well of OGD solution composed of (in mM): NaCl 137.93 mM, KCl 5.36 mM, CaCl₂ 2 mM, MgSO₄ 1.19 mM, NaHCO₃ 26 mM, KH₂PO₄ 1.18 mM, and 2-deoxyglucose 11 mM (Sigma-Aldrich, Madrid, Spain). The cultures were then placed in a partially open airtight chamber (Billups and Rothenberg) exposed within 3 min to 95% N₂/5% CO₂ gas flow to ensure oxygen removal. After that, the chamber was sealed for 15 min at 37°C (OGD period). At the same time, control cultures were maintained under normoxic conditions in a solution with the same composition as previously described but containing normal glucose 15 mM (Sigma-Aldrich, Madrid, Spain) instead of 2-deoxyglucose. Afterwards, slice cultures were pharmacologically treated with 0.01 μ M GKT136901 and 0.3 μ M L-NAME separately or in combination before returning them to

normal oxygen and glucose concentrations for 24h (re-oxygenation period). Non-treated slices were used as internal control.

Quantification of cell death in organotypic hippocampal cultures

Cell death was determined in the CA1 region by staining the OHCs with Propidium iodide (PI) (Invitrogen, Madrid, Spain). Slices were incubated with PI (1 μ g/ml) and Hoechst (5 μ g/ml) during 30 min. Fluorescence was measured in a fluorescence-inverted NIKON Eclipse T2000-U microscope at 10X magnification. Wavelengths of excitation and emission for PI and Hoechst were 530/350, and 580/460 nm, respectively. Fluorescence analysis was performed using the Metamorph software version 7.0. Hoechst staining was used to normalize PI fluorescence with respect to the number of nuclei as described in [10]. Data were normalized with respect to the control values that were considered as 1.

ROS measurement in Organotypic Hippocampal Culture

To determine ROS/RNS production in OHCs, we used the fluorescent probe H2DCFDA as has been previously described in (29). Organotypic hippocampal slices were loaded with 10 μ M H2DCFDA, which diffuses through the cell membrane and is hydrolyzed by intracellular esterases to the non-fluorescent form dichlorofluorescein (DCFH). DCFH reacts with intracellular H₂O₂ to form the green fluorescent signal. Fluorescence was measured in a fluorescence-inverted NIKON Eclipse T2000-U microscope. Wavelengths of excitation and emission were 485 and 520 nm, respectively. All variables were normalized with respect to the control values (considered as 1).

ROS kinetics: DHE

Ex-vivo acute model: Hippocampal brain slices: ROS production was evaluated in real-time by the fluorescence dye DHE method (10, 11). Stock solution of DHE (3.2 mM) was dissolved in previously oxygenated Krebs solution and added to the hippocampal brain slices. The average fluorescence intensity was monitored using a 10X objective in the CA1 region of the hippocampus every 30 sec during the first

10 min of the re-oxygenation period. Excitation and emission wavelength were 485 nm, and >580 nm, respectively. Fluorescence analysis was performed using the Metamorph software version 7.0. Each value was divided by the initial fluorescence value for normalization.

Ex-vivo chronic model: Organotypic hippocampal culture: Similarly, a stock solution of DHE (3.2 mM) was dissolved in Krebs solution and added to the culture insert. Fluorescence measurements were performed at 0, 2, 4, 8 and 12h after the OGD period using a 10X objective in the CA1 region of the hippocampus. Same emission and excitation wavelength were used. Fluorescence analysis was performed using the Metamorph software version 7.0.

RNA extraction, quantification and reverse transcription

Hippocampal brain slices from organotypic culture were crushed and homogenized using TRI Reagent ®(Sigma-Aldrich, The Netherlands). 100 μ l of chloroform was added to the samples followed by 15 min centrifugation at 11.000 rpm and 4°C. After centrifugation, the upper phase was preserved (mRNA). 250 μ l isopropanol was added to the upper phase and incubated during 1h at -20°C or overnight at -80°C. After incubation, samples were centrifuged during 10 min at 13.000 rpm and 4°C. 200 μ l ethanol 80% was added to the supernatant (pellet discarded) followed by 10 min centrifugation at 13.000 rpm. After removal of the ethanol, mRNA was dissolved in 10 μ l RNase free water. mRNA was quantified spectrophotometrically using the Nanodrop 2000 device. Later, 0.08 μ g of total mRNA was reverse transcribed to cDNA with the High Capacity Reverse Transcription Kit (Applied Biosystems, The Netherlands) according to manufacturer's protocol.

Real-time PCR

mRNA levels of Nox4, neuronal NOS (NOS1), inducible NOS (NOS2) and endothelial NOS (NOS3), were quantified using the fluorescent Taqman® technology. We used TaqMan® gene expression arrays (TaqMan® Universal PCR Master Mix, ThermoFisher Scientific, The Netherlands) specific for rat: Nox4

(Rn01506793_m1, ThermoFisher Scientific, The Netherlands), NOS1 (Rn00694747_m1, ThermoFisher Scientific, The Netherlands), NOS2 (Rn00561646_m1, ThermoFisher Scientific, The Netherlands), NOS3 (Rn02132634_s1, ThermoFisher Scientific, The Netherlands). β -actin (Rn00667869_m1, ThermoFisher Scientific, The Netherlands) was used as a house-keeping gene. Water controls were included to ensure specificity and the comparative $2^{-\Delta\Delta C_t}$ method was used for relative quantification of gene expression.

In vivo MCAO ischemia model

The model was conducted as described in [11]. C57Bl6/J mice were anesthetized with isoflurane (0.6% in oxygen). The animal was placed on a heating-pad, and rectal temperature was maintained at 37.0°C using a servo-controlled rectal probe-heating pad (Cibertec, Spain). Transient cerebral ischemia was induced using an intraluminal filament technique. Using a surgical microscope (Tecnoscopia OPMI pico, Carl Zeiss, Meditec Iberia SA, Spain), a midline neck incision was made and the right common and external carotid arteries were isolated and permanently ligated. A microvascular temporarily ligature was placed on the internal carotid artery to temporarily stop the blood flow. A silicon rubber-coated monofilament (6023910PK10, Doccol Corporation, Sharon, MA, USA) was inserted through a small incision into the common carotid artery and advanced into the internal carotid artery until a resistance is felt. The tip of the monofilament is then precisely located at the origin of the right middle cerebral artery and thus interrupting blood flow. The filament was held in place by a tourniquet suture on the common carotid artery to prevent filament relocation during the ischemia period. Animals were maintained under anaesthesia during 1h occlusion period followed by the reperfusion period just started when the monofilament is removed. After the surgery, wounds were carefully sutured and animals could recover from surgery in a temperature-controlled cupboard. For permanent MCAO (pMCAO) the occluding filament was left in situ until sacrificing the animals [12]. Operation time per animal did not exceed 20 minutes.

Treatment with NO synthase and NADPH oxidase inhibitors

L-NAME was dissolved in saline. GKT136901 was dissolved in a mixture of DMSO/saline in a ratio of 1/99. L-NAME (3 mg/kg), GKT136901 (10 mg/kg) or vehicle (DMSO/saline in a ratio of 1/99) was injected i.p. 1h and 3h after reperfusion, i.e. 1h or 3h after removal of the filament. Stroke outcome in mono therapies were only assessed 1h post-reperfusion due to ethical limitations. Our current ethical proposal is focused on go/no-go decisions to further reduce number of animals as much as possible. Therefore, testing already validated non-effective therapeutic strategy in different animal groups was explicitly not allowed. Therefore, only the combination therapy (effective) was validated in long-term treatment, aged animals, permanent occlusion brain ischemia model and different post-stroke outcomes.

Determination of infarct size

After sacrificing the mice, brains were quickly removed and cut in four 2-mm thick coronal sections using a mouse brain slice matrix (Harvard Apparatus, Spain). Brain slices were stained for 15 min at room temperature with 2% 2,3,5-triphenyltetrazolium chloride (TTC; Sigma-Aldrich, The Netherlands) in PBS to visualize the infarctions [13]. Indirect infarct volumes were calculated by volumetry (ImageJ software, National Institutes of Health, USA) according to the following equation:

$$V_{indirect}(mm^3) = V_{infarct} * (1 - (V_{ih} - V_{ch})/V_{ch})$$

, where the term $(V_{ih} - V_{ch})$ represents the volume difference between the ischemic hemisphere and the control hemisphere and $(V_{ih} - V_{ch})/V_{ch}$ expresses this difference as a percentage of the control hemisphere.

Assessment of neuro-motor functional outcomes

Three different neuro-motor functional tests were assessed in all mice groups (adult male, female and aged) treated 1h after filament removal. The Bederson Score [13] is defined by: Score 0, no apparent neurological deficits; 1, body torsion and forelimb flexion; 2, right

side weakness and thus decreased resistance to lateral push; 3, unidirectional circling behavior; 4, longitudinal spinning; 5, no movement. Within the elevated body swing test, the mice were held ~1 cm from the base of its tail and then elevated above the surface in the vertical axis around 20 cm. We considered a swing whenever the animal moved its head out of the vertical axis to either the left or the right side (more than 10 degrees). Ratio of right/left swings were subsequently analyzed. Finally, to directly evaluate strength, the four-limbs hanging wire test was performed. The mouse was placed on the center of the wire with a diameter of 10 cm. Later, the wire was slowly inverted and placed at ~40 cm above a paper towel bedding. The time until the mouse fell from the wire was recorded considering as maximum time 120s.

Determination of blood-brain-barrier leakage and brain oedema

To determine the permeability of the cerebral vasculature and brain oedema formation after 1h occlusion of the middle cerebral artery (tMCAO), 2% Evans blue tracer (Sigma Aldrich, The Netherlands) was diluted in 0.9% NaCl and then injected intraperitoneally just after removing the filament. Measurement of Evans Blue extravasation was performed as described in [11].

Oxidative stress: DHE staining

ROS production formation in vivo brain tissue was determined using the florescence dye dihydroethidium (Thermo Scientific Technology, The Netherlands). Frozen brain cryosections were fixated in 4% paraformaldehyde (PFA) in PBS and then incubated with 2 μ M DHE (2mM stock solution) for 30 minutes at 37°C. After three washing steps with PBS, slices were incubated with Hoechst (Hoechst 33342, Sigma-Aldrich, The Netherlands) 2 ng/ml for 10 min at 37°C. The relative pixel intensity was measured in identical regions with (ImageJ software, National Institutes of Health, USA).

Immunohistochemistry: N-Tyr staining

Brain tissue cryosections (10 μm) were fixed with 4% PFA in PBS. After fixation, sections were incubated for 1h at room temperature with a rabbit polyclonal anti-nitrotyrosine antibody (1:100; ThermoFisher Scientific, A-21285) in blocking buffer. After three washes in PBS, sections were incubated with the secondary antibody, Alexa Fluor 488 donkey anti-rabbit (1:100; ThermoFisher Scientific, A-21206) for 45 min at room temperature. Then, the fluorescent Hoechst33342 dye (2 ng/ml; ThermoFisher Scientific, The Netherlands) was added for 10 min at room temperature. Sections were washed in PBS and then mounted using a Dako Fluorescence Mounting Medium (S3023, Agilent Technologies). Immunofluorescent signals were viewed using a Leica DMI3000 B fluorescence microscope. Quantitative analysis of nitro-tyrosine fluorescence was performed with (ImageJ software, National Institute of Health, USA).

Human brain microvascular endothelial cell (HBMEC) cultures subjected to hypoxia

HBMEC (Cell systems, USA) between passage 3 and 9 were cultured to approximately 95% confluence using specialized cell medium (EGM-2 MV BulletKit, Lonza, The Netherlands) enriched with 5% fetal bovine serum (FBS; Sigma-Aldrich, The Netherlands) before starting the hypoxia period. For hypoxia studies, HBMECs were seeded at specific density (6×10^4 cells/ml) in 12 wells-plate and incubated during 24h at 37°C. Then, cell medium was replaced with non-FBS containing medium (2 ml/well) following by 6h of hypoxia (94,8% N₂, 0.2% O₂ and 5% CO₂) at 37°C using hypoxia workstations (Ruskin Invivo2 400 station, The Netherlands). The hypoxia period was followed by 24h of reperfusion in the presence or absence of 0,1 μM GKT136901 (Genkyotech, Switzerland) and 0,3 μM L-NAME, incubated separately or in combination. Control cells were exposed to normoxia (75% N₂, 20% O₂ and 5% CO₂) and enriched medium during the hypoxia period. All flasks and well plates were pre-treated with fibronectin (Sigma-Aldrich, The Netherlands) solution (1:100 in PBS).

Assessment of cell viability in HBMEC

After 24h of re-oxygenation period, cell viability was assessed using the colorimetric MTT assay [14]. MTT solution (5 mg/ml) was added to each well (100 μ l/ml) and incubated for 2h at 37°C. The formazan salt formed was solubilized by adding 350 μ l/well DMSO. The optical density was measured spectrophotometrically at 540 nm using a micro plate reader. Absorbance values obtained in control cells were set to 100% viability.

Assessment of cell permeability in HBMEC

2 x 10⁴ HBMECs were grown to confluence on membranes of Transwell inserts (collagen-coated Transwell Pore Polyester Membrane Insert; pore size = 3.0 μ m, Corning, The Netherlands) 24h before inducing 6h of ischemic conditions followed by 24h re-oxygenation period where cells were treated with 1 μ M GKT136901. Cell permeability was assessed using the Evans Blue dye (Sigma-Aldrich, The Netherlands). Before the diffusion experiment, medium was removed and cells were washed once with assay buffer. The same buffer (1.5 ml) was added to the abluminal side of the insert. Permeability buffer (0.5 ml) containing 4% bovine serum albumin (Sigma-Aldrich, The Netherlands) and 0.67 mg/ml Evans blue dye was loaded on the luminal side of the insert followed by 15 min incubation at 37°C. The concentration of Evans Blue in the abluminal chamber was measured by determining the absorbance of 150 μ l buffer at 630 nm using a microplate reader.

Transient occlusion of the middle cerebral artery (tMCAO). Model previously described in [15].

Table 4.S1: Network proteins and their connectedness to NOX4

Protein Symbol	Uniprot ID	Connectedness to NOX4
NAXD	Q8IW45	1
DHFR2	Q86XF0	1
AKR1E2	Q96JD6	1
AKR1B10	O60218	1
AASDHPPT	Q9NRN7	1
TECR	Q9NZ01	1
SRD5A3	Q9H8P0	1
NCF1C	A8MVU1	1
NCF1B	A6NI72	1
HTATIP2	Q9BUP3	1
DHRS3	O75911	1
PTGR2	Q8N8N7	1
HSD17B12	Q53GQ0	1
NUDT7	P0C024	1
IYD	Q6PHW0	1
A9Z1X1	A9Z1X1	1
AASS	A4D0W4	1
AASDH	Q4L235	1
MTHFD1L	Q6UB35	1
DHFR	B0YJ76	1
FAR2	Q96K12	1
FAR1	Q8WVX9	1
ALDH1L2	Q3SY69	1
RDH10	Q8IZV5	1
RDH11	Q8TC12	1
RDH14	Q9HBH5	1
RDH8	Q9NYR8	1
HSD3B2	P26439	1
AKR7A2	O43488	1
KDSR	Q06136	1
RDH12	Q96NR8	1
PYCR3	Q53H96	1
HSD17B6	O14756	1
DHRS9	Q9BPW9	1
NCF1	P14598	1
RAC2	P15153	1
GLRX	P35754	1
NCF4	Q15080	1
NCF2	P19878	1
DECR2	Q9NUI1	1
CRYZ	Q08257	1
DECR1	Q16698	1
PGD	P52209	1
PECR	Q9BY49	1
MECR	Q9BV79	1
ALDH1L1	O75891	1
TSTA3	Q13630	1
MTHFD1	P11586	1
NNT	Q13423	1

Table 4.S1: Network proteins and their connectedness to NOX4

Protein Symbol	Uniprot ID	Connectedness to NOX4
HMGCR	P04035	1
IDH1	O75874	1
IDH2	P48735	1
ME2	P23368	1
ME3	Q16798	1
ME1	P48163	1
ALDH18A1	P54886	1
AASS	Q9UDR5	1
HSD17B1	P14061	1
HSD17B3	P37058	1
DHFR	P00374	1
ADH1C	P00326	1
ADH6	P28332	1
ADH1A	P07327	1
ADH7	P40394	1
ADH1B	P00325	1
ADH5	P11766	1
HSD11B1	P28845	1
HSD11B2	P80365	1
SRD5A1	P18405	1
SRD5A2	P31213	1
GAPDH	P04406	1
GLUD1	P00367	1
GAPDHS	O14556	1
GMPR	P36959	1
GMPR2	Q9P2T1	1
GLUD2	P49448	1
AKR1D1	P51857	1
DHCR7	Q9UBM7	1
QDPR	P09417	1
ALDH3B1	P43353	1
ALDH3B2	P48448	1
AKR1A1	P14550	1
DHDH	Q9UQ10	1
HSD17B8	Q92506	1
HSD17B2	P37059	1
SPR	P35270	1
GRHPR	Q9UBQ7	1
AKR1C2	P52895	1
AKR1C1	Q04828	1
AKR1C3	P42330	1
H6PD	O95479	1
G6PD	P11413	1
DCXR	Q7Z4W1	1
PYCR2	Q96C36	1
PYCR1	P32322	1
ALDH1B1	P30837	1
ALDH3A2	P51648	1
ALDH2	P05091	1
ALDH1A3	P47895	1

Table 4.S1: Network proteins and their connectedness to NOX4

Protein Symbol	Uniprot ID	Connectedness to NOX4
ALDH7A1	P49419	1
ALDH3A1	P30838	1
ALDH9A1	P49189	1
TM7SF2	O76062	1
FDFT1	P37268	1
ALDH4A1	P30038	1
AKR1B1	P15121	1
CBR3	O75828	1
DHRS4	Q9BTZ2	1
CBR1	P16152	1
AKR1C4	P17516	1
ALDH6A1	Q02252	1
FASN	P49327	1
RSAD2	Q8WXXG1	1
PGRMC2	O15173	1
PGRMC1	O00264	1
NRF1	Q16656	1
NFE2L1	Q14494	1
NENF	Q9UMX5	1
SLC25A28	Q96A46	1
SLC25A37	Q9NYZ2	1
SLC48A1	Q6P1K1	1
HERC2	O95714	1
HEBP2	Q9Y5Z4	1
HEBP1	Q9NRV9	1
HBM	Q6B0K9	1
HBE1	P02100	1
HBZ	P02008	1
HBQ1	P09105	1
FRRS1	Q6ZNA5	1
FADS3	Q9Y5Q0	1
FADS2P1	A8MWK0	1
FA2H	Q7L5A8	1
EIF2AK1	Q9BQI3	1
CYB5B	O43169	1
CYP4V2	Q6ZWL3	1
CYP2W1	Q8TAV3	1
CYP26C1	Q6V0L0	1
CYP26B1	Q9NR63	1
COX15	Q7KZN9	1
CDC37	Q16543	1
CYB5D2	Q8WUJ1	1
CYB5D1	Q6P9G0	1
CYB561D1	Q8N8Q1	1
CYP27C1	Q4G0S4	1
ABCB10	Q9NRK6	1
ABCB7	O75027	1
ABCB6	Q9NP58	1
CD163	Q86VB7	1
FADS2	O95864	1

Table 4.S1: Network proteins and their connectedness to NOX4

Protein Symbol	Uniprot ID	Connectedness to NOX4
COX5A	P20674	1
CYBRD1	Q53TN4	1
SLC46A1	Q96NT5	1
CYB5A	P00167	1
HRG	P04196	1
TSPO	P30536	1
FADS1	O60427	1
CYBA	P13498	2
NFE2	Q16621	1
FLVCR1	Q9Y5Y0	1
TF	P02787	1
CYP26A1	O43174	1
PRDM2	Q13029	1
HCCS	P53701	1
CYCS	P99999	1
CYB561	P49447	1
CYB561D2	O14569	1
KCNMA1	Q12791	1
AMBP	P02760	1
HFE	Q30201	1
COX10	Q12887	1
HPX	P02790	1
HBG1	P69891	1
TBXAS1	P24557	1
PTGIS	Q16647	1
GUCY1B2	O75343	1
GUCY1B3	Q02153	1
ALAD	P13716	1
FECH	P22830	1
UQCRH	P07919	1
CYC1	P08574	1
MT-CYB	P00156	1
UQCR10	Q9UDW1	1
SUCLA2	Q9P2R7	1
CBS	P35520	1
BLVRA	P53004	2
HMBS	P08397	1
UROS	P10746	1
UROD	P06132	1
ALAS1	P13196	1
ALAS2	P22557	1
TET3	O43151	1
TET2	Q6N021	1
TET1	Q8NFU7	1
PIR	O00625	1
PHF8	Q9UPP1	1
RIOX1	Q9H6W3	1
RIOX2	Q8IUF8	1
KDM8	Q8N371	1
FAXDC2	Q96IV6	2

Table 4.S1: Network proteins and their connectedness to NOX4

Protein Symbol	Uniprot ID	Connectedness to NOX4
CYP21A2	P08686	2
AGMO	Q6ZNB7	1
ALKBH1	Q13686	1
ADI1	Q9BV57	1
NOS1	B3VK56	2
CP	A5PL27	1
CYP2D6	Q6NWU0	3
CYP21A2	Q08AG9	3
P4HTM	Q9NXG6	1
EGLN3	Q9H6Z9	1
EGLN2	Q96KS0	1
EGLN1	Q9GZT9	1
DOHH	Q9BU89	1
KDM2B	Q8NHM5	1
KDM2A	Q9Y2K7	1
P4HA3	Q7Z4N8	1
P3H3	Q8IVL6	1
P3H2	Q8IVL5	1
P3H1	Q32P28	1
FTH1	P02794	1
BCO2	Q9BYV7	1
ALKBH2	Q6NS38	1
ADO	Q96SZ5	1
NGB	Q9NPG2	2
HBD	P02042	2
CYGB	Q8WWM9	2
CYP4F22	Q6NT55	3
CYP2U1	Q7Z449	2
CYP20A1	Q6UW02	2
CYP1A1	A0N0X8	3
IDO2	Q6ZQW0	2
SCD5	Q86SK9	1
CYP4A22	Q5TCH4	3
COX4I2	Q96KJ9	1
COX6A1	P12074	1
MT-CO2	P00403	1
CYP8B1	Q9UNU6	3
CYP39A1	Q9NYL5	3
TDO2	P48775	1
CYP11B2	P19099	1
CYP46A1	Q9Y6A2	3
TMLHE	Q9NVH6	1
COX8C	Q7Z4L0	1
COX7A2L	O14548	1
COX6C	P09669	1
COX8A	P10176	1
COX6B2	Q6YFQ2	1
COX4I1	P13073	1
COX7B2	Q8TF08	1
COX6B1	P14854	1

Table 4.S1: Network proteins and their connectedness to NOX4

Protein Symbol	Uniprot ID	Connectedness to NOX4
COX5B	P10606	1
COX7C	P15954	1
COX7B	P24311	1
COX6A2	Q02221	1
CH25H	O95992	1
COX7A2	P14406	1
FXN	Q16595	2
HBG2	P69892	2
CYP2R1	Q6VVX0	3
MB	P02144	2
COX7A1	P24310	1
CYP24A1	Q07973	3
COQ7	Q99807	2
GGCX	P38435	1
MT-CO1	P00395	2
CYP7B1	O75881	3
HBA1	P69905	2
HBB	P68871	2
BCO1	Q9HAY6	1
CYP2A6	P11509	3
CYP2A7	P20853	3
CYP2J2	P51589	3
CYP4F8	P98187	3
CYP2S1	Q96SQ9	3
CYP2C8	P10632	3
CYP19A1	P11511	3
CYP1A2	P05177	3
CYP4Z1	Q86W10	3
CYP4B1	P13584	3
CYP3A7	P24462	3
CYP2A13	Q16696	3
CYP1A1	P04798	3
CYP3A5	P20815	3
CYP2B6	P20813	3
CYP4X1	Q8N118	3
CYP2F1	P24903	3
CYP4F12	Q9HCS2	2
CYP2C18	P33260	3
CYP2D6	P10635	2
CYP1B1	Q16678	3
CYP3A43	Q9HB55	3
CYP4F11	Q9HBI6	3
CYP2E1	P05181	3
HAAO	P46952	1
CYP17A1	P05093	3
CYP27B1	O15528	3
CYP2C19	P33261	3
CYP27A1	Q02318	3
CYP2C9	P11712	3
HIF1AN	Q9NWT6	1

Table 4.S1: Network proteins and their connectedness to NOX4

Protein Symbol	Uniprot ID	Connectedness to NOX4
ASPH	Q12797	1
CYP3A4	P08684	3
CYP4F3	Q08477	3
CYP4F2	P78329	3
ALOX15	P16050	1
PTGS1	P23219	2
PTGS2	P35354	2
ALOX12	P18054	1
ALOX15B	O15296	1
ALOX5	P09917	1
CDO1	Q16878	2
HSD17B7	P56937	2
HGD	Q93099	1
HPD	P32754	1
MIOX	Q9UGB7	1
CYP11A1	P05108	2
CYP7A1	P22680	3
CP	P00450	1
TPH2	Q8IWU9	1
FTMT	Q8N4E7	1
TPH1	P17752	1
PLOD3	O60568	1
PLOD2	O00469	1
PLOD1	Q02809	1
CYP11B1	P15538	2
IDO1	P14902	2
CYP4A11	Q02928	3
BBOX1	O75936	1
SCD	O00767	1
P4HA1	P13674	1
P4HA2	O15460	1
PAH	P00439	1
TH	P07101	1
PHYH	O14832	1
CYP51A1	Q16850	3
NSDHL	Q15738	2
MSMO1	Q15800	2
DBH	P09172	1
PAM	P19021	1
CPOX	P36551	2
TYR	P14679	1
GPX8	Q8TED1	1
NOXO1	Q8NFA2	1
NOXA1	Q86UR1	2
LOXL2	Q9Y4K0	2
PXDNL	A1KZ92	2
PXDN	Q92626	2
SOD3	P08294	2
PRDX5	P30044	1
SOD1	P00441	2

Table 4.S1: Network proteins and their connectedness to NOX4

Protein Symbol	Uniprot ID	Connectedness to NOX4
SOD2	P04179	2
CAT	P04040	3
SUOX	P51687	3
GPX2	P18283	1
GPX3	P22352	1
GPX4	P36969	1
GPX1	P07203	1
GPX6	P59796	1
GPX5	O75715	1
GPX7	Q96SL4	1
HAO2	Q9NYQ3	2
HAO1	Q9UJM8	2
LOX	P28300	2
PNPO	Q9NVS9	2
AOC2	O75106	2
AOC3	Q16853	2
AOC1	P19801	2
EPX	P11678	2
MPO	P05164	2
PRDX6	P30041	2
LPO	P22079	2
TPO	P07202	2
NDUFB2	A4D1T5	1
STEAP4	Q687X5	2
STEAP2	Q8NFT2	2
STEAP1	Q9UHE8	2
SQOR	Q9Y6N5	1
RETSAT	Q6NUM9	1
RNLS	Q5VYX0	1
QSOX2	Q6ZRP7	3
QSOX1	O00391	3
PYROXD1	Q8WU10	1
PDP2	Q9P2J9	1
PCYOX1L	Q8NBM8	1
PCYOX1	Q9UHG3	3
OXNAD1	Q96HP4	1
IL4I1	Q96RQ9	3
NOX4	Q9NPH5	5
NOX1	Q9Y5S8	2
NDOR1	Q9UHB4	1
MTO1	Q9Y2Z2	1
MICAL3	Q7RTP6	3
MICAL2	O94851	3
MICAL1	Q8TDZ2	3
KDM1A	O60341	1
FOXRED2	Q8IWF2	1
FOXRED1	Q96CU9	1
ERO1B	Q86YB8	1
ERO1A	Q96HE7	1
DUS4L	O95620	1

Table 4.S1: Network proteins and their connectedness to NOX4

Protein Symbol	Uniprot ID	Connectedness to NOX4
DUS3L	Q96G46	1
DUS1L	Q6P1R4	1
CRY2	Q49AN0	1
KDM1B	Q8NB78	1
GFER	P55789	3
AIFM3	Q96NN9	1
AIFM2	Q9BRQ8	1
ACOXL	Q9NUZ1	1
ACADVL	P49748	1
STEAP3	Q658P3	2
NOX3	Q9HBY0	2
NOX5	Q96PH1	3
COQ6	Q9Y2Z9	2
FLAD1	Q8NFF5	1
Q5HYI4	Q5HYI4	3
CYB5RL	Q6IPT4	1
CYB5R4	Q7L1T6	2
CYB5R2	Q6BCY4	1
CYB5R1	Q9UHQ9	1
TXNRD3	Q86VQ6	2
DUOX1	Q9NRD9	4
PRODH2	Q9UF12	1
ACAD9	Q9H845	1
ACAD8	Q9UKU7	1
LDHD	Q86WU2	1
DUOX2	Q9NRD8	4
CRY1	Q16526	1
AIFM1	O95831	1
SMOX	Q9NWM0	3
KMO	O15229	3
TKFC	Q3LXA3	1
L2HGDH	Q9H9P8	1
D2HGDH	Q8N465	1
PRODH	O43272	1
ETFB	P38117	1
APP	P05067	2
DHCR24	Q15392	2
FMO6P	O60774	3
ETFA	P13804	1
CYBB	P04839	5
NQO1	P15559	2
ACADSB	P45954	1
AGPS	O00116	1
NDUFB11	Q9NX14	1
PPOX	P50336	4
NOS3	P29474	4
NOS1	P29475	4
NOS2	P35228	4
SARDH	Q9UL12	1
DPYD	Q12882	2

Table 4.S1: Network proteins and their connectedness to NOX4

Protein Symbol	Uniprot ID	Connectedness to NOX4
BLVRB	P30043	3
DUS2	Q9NX74	1
MTRR	Q9UBK8	2
DAO	P14920	3
ALDH1A2	O94788	1
FDXR	P22570	2
DMGDH	Q9UI17	1
GCDH	Q92947	1
CHDH	Q8NE62	1
ALDH1A1	P00352	1
HMOX1	P09601	4
GPD2	P43304	1
GSR	P00390	2
ENPP3	O14638	1
SDHB	P21912	1
ACOX3	O15254	3
NDUFS6	O75380	1
NDUFA13	Q9P0J0	1
MT-ND6	P03923	1
MTHFR	P42898	2
NDUFB9	Q9Y6M9	1
NDUFV1	P49821	1
FMO1	Q01740	3
DHODH	Q02127	3
POR	P16435	3
XDH	P47989	3
NDUFS7	O75251	1
SDHD	O14521	2
MT-ND4	P03905	1
NDUFA9	Q16795	1
ETFDH	Q16134	1
NDUFA5	Q16718	1
NDUFAB1	O14561	1
NDUFB7	P17568	1
NDUFB3	O43676	1
FMO3	P31513	3
NDUFC2	O95298	1
MAOA	P21397	3
IVD	P26440	1
FMO4	P31512	3
ACOX2	Q99424	1
NDUFB5	O43674	1
NDUFS1	P28331	1
MAOB	P27338	3
MTR	Q99707	1
NDUFA2	O43678	1
TXNRD2	Q9NNW7	2
NDUFA7	O95182	1
FMO2	Q99518	3
MT-ND2	P03891	1

Table 4.S1: Network proteins and their connectedness to NOX4

Protein Symbol	Uniprot ID	Connectedness to NOX4
SQLE	Q14534	3
NDUFS8	O00217	1
HMOX2	P30519	4
SDHA	P31040	1
NDUFB4	O95168	1
DDO	Q99489	3
ACOX1	Q15067	3
NDUFA6	P56556	1
NDUFA10	O95299	1
FMO5	P49326	3
NDUFV2	P19404	1
NDUFS5	O43920	1
NDUFS3	O75489	1
NQO2	P16083	1
NDUFS4	O43181	1
SC5D	O75845	3
TXNRD1	Q16881	2
NDUFA11	Q86Y39	1
MT-ND4L	P03901	1
NDUFA4	O00483	1
NDUFA8	P51970	1
NDUFA3	O95167	1
NDUFA4L2	Q9NRX3	1
NDUFA1	O15239	1
AOX1	Q06278	4
CYB5R3	P00387	1
NDUFA12	Q9UI09	1
PAOX	Q6QHF9	3
PIPOX	Q9P0Z9	3
SDHC	Q99643	2
NDUFS2	O75306	1
ENPP1	P22413	1
NDUFB1	O75438	1
MT-ND1	P03886	1
ACADM	P11310	1
ACADS	P16219	1
ACADL	P28330	1
DLD	P09622	1

Table 4.S2: Animals excluded from the statistical analysis after tMCAO

Animals	Ischemia model	Duration	Excluded Animals	Reason of exclusion
C57/Bl6J	Control (TTC)	24h	1 of 8	Dead within the first 24h
C57/Bl6J	L-NAME (TTC)	24h	1 of 6	Dead within the first 24h
C57/Bl6J	GKT136901 (TTC)	24h	1 of 5	Dead within the first 24h
C57/Bl6J	C. ther. 1h PO (TTC)	24h	0 of 6	-
C57/Bl6J	C. ther. 3h PO (TTC)	24h	0 of 6	-
C57/Bl6J	Control Aged (TTC)	24h	1 of 6	Dead during the surgery
C57/Bl6J	C. ther. Aged (TTC)	24h	0 of 5	-
C57/Bl6J	pMCAO Control (TTC)	24h	1 of 7	Dead during the surgery
C57/Bl6J	pMCAO C. ther. 1h PO (TTC)	24h	1 of 6	Dead within the first 24h
C57/Bl6J	Control (EB)	24h	1 of 6	Dead during the surgery
C57/Bl6J	C. ther. (EB)	24h	0 of 5	-
C57/Bl6J	Control (DHE)	24h	0 of 4	-
C57/Bl6J	C. ther. (DHE)	24h	0 of 4	-
C57/Bl6J	Control (N-Tyr)	24h	0 of 4	-
C57/Bl6J	C. ther. (N-Tyr)	24h	0 of 4	-

tMCAO: transient occlusion of the middle cerebral artery; TTC: 2,3,5- triphenyltetrazolium hydrochloride; C. ther.: combinatory therapy; EB: Evans Blue. Animal exclusion procedures are described in the respective methods parts.

Table 4.S3: Power analysis

Adult mice - tMCAO – GKT+L-NAME (10 mg/kg, 3 mg/kg)				
Infarct size (mm3)				
	Mean	SD	N	Power for measured difference (%)
Vehicle	110,1	21,5	7	89,7
Treatment	66,2	13,9	6	
Four limb hanging				
	Mean	SD	N	Power for measured difference (%)
Vehicle	41,7	38,2	7	98,2
Treatment	101	31,1	5	
Elevated body swing				
	Mean	SD	N	Power for measured difference (%)
Vehicle	0,8	0,8	7	96,1
Treatment	2,3	2,1	9	
Aged mice - tMCAO – GKT+L-NAME (10 mg/kg, 3 mg/kg)				
Infarct size (mm3)				
	Mean	SD	N	Power for measured difference (%)
Vehicle	103,19	19,6	5	25,4
Treatment	45,61	37,85	5	

tMCAO: transient middle cerebral artery occlusion; SD: standard deviation; N: number of animals. We conducted a post hoc analysis of power in the different animal groups. For each animal treatment group, a pooled variance of the vehicle and treatment groups was calculated from mean, SD and n-number with n the size of the group and CV the coefficient of variation (SD/Mean) of the group. Power was calculated for the measured difference using Russ Lenth's power software with an alpha of 0.05, the measured effect (%) and the calculated pooled variances [(Lenth, R.V 2006-9, java Applets for Power and Sample Size [Computer Software], Retrieved 02-17-2014, from <http://www.stat.uiowa.edu/~rlenth/Power>].

Table 4.S4: Databases used and their details

Database Name	Version	Access date	Number of nodes/edges	Evidence level used
Human Metabolite Database (HMDB)	4.0	1/24/18	114,100/ 866,004	All
Integrated Interactions Database (IID)	Human 2017-04	2017-12	18,627/ 923,913	exp
Therapeutic Database (TTD)	Target Sep 15th, 2017	2018-02	2957/-	-

References

- [1] M. Kotlyar et al. "Integrated interactions database: tissue-specific view of the human and model organism interactomes". In: *Nucleic Acids Research* 44 (Oct. 2015), pp. D536–D541. DOI: 10.1093/nar/gkv1115.
- [2] D. S. Wishart et al. "HMDB 4.0: the human metabolome database for 2018". In: *Nucleic Acids Research* 46 (Jan. 2018), D608–D617. DOI: 10.1093/nar/gkx1089.
- [3] H. Fröhlich et al. "GOSim – an R-package for computation of information theoretic GO similarities between terms and gene products". In: *BMC Bioinformatics* 8 (May 2007). DOI: 10.1186/1471-2105-8-166.
- [4] J. Z. Wang et al. "A new method to measure the semantic similarity of GO terms". In: *Bioinformatics* 23.10 (2007), pp. 1274–1281.
- [5] J. Egea et al. "Neuroprotection afforded by nicotine against oxygen and glucose deprivation in hippocampal slices is lost in 7 nicotinic receptor knockout mice". In: *Neuroscience* 145 (Mar. 2007), pp. 866–872. DOI: 10.1016/j.neuroscience.2006.12.036.
- [6] J. Egea et al. "Small synthetic hyaluronan disaccharides afford neuroprotection in brain ischemia-related models". In: *Neuroscience* 265 (Apr. 2014), pp. 313–322. DOI: 10.1016/j.neuroscience.2014.01.032.
- [7] F. Denizot and R. Lang. "Rapid colorimetric assay for cell growth and survival". In: *Journal of Immunological Methods* 89 (May 1986), pp. 271–277. DOI: 10.1016/0022-1759(86)90368-6.
- [8] J. Huuskonen et al. "A refined in vitro model to study inflammatory responses in organotypic membrane culture of postnatal rat hippocampal slices". In: *Journal of Neuroinflammation* 2 (2005), p. 25. DOI: 10.1186/1742-2094-2-25.

- [9] J. Egea et al. "Melatonin-sulforaphane hybrid ITH12674 induces neuroprotection in oxidative stress conditions by a 'drug-prodrug' mechanism of action". In: *British Journal of Pharmacology* 172 (Feb. 2015), pp. 1807–1821. DOI: 10.1111/bph.13025.
- [10] J. Egea et al. "Galantamine elicits neuroprotection by inhibiting iNOS, NADPH oxidase and ROS in hippocampal slices stressed with anoxia/reoxygenation". In: *Neuropharmacology* 62 (Feb. 2012), pp. 1082–1090. DOI: 10.1016/j.neuropharm.2011.10.022.
- [11] C. Kleinschnitz et al. "Post-Stroke Inhibition of Induced NADPH Oxidase Type 4 Prevents Oxidative Stress and Neurodegeneration". In: *PLoS Biology* 8 (Sept. 2010). Ed. by M. McLeod, e1000479. DOI: 10.1371/journal.pbio.1000479.
- [12] V. Bindokas et al. "Superoxide production in rat hippocampal neurons: selective imaging with hydroethidine". In: *The Journal of Neuroscience* 16 (Feb. 1996), pp. 1324–1336. DOI: 10.1523/jneurosci.16-04-01324.1996.
- [13] J. B. Bederson et al. "Evaluation of 2,3,5-triphenyltetrazolium chloride as a stain for detection and quantification of experimental cerebral infarction in rats." In: *Stroke* 17 (Nov. 1986), pp. 1304–1308. DOI: 10.1161/01.str.17.6.1304.
- [14] H. Li et al. "Associations of NADPH oxidase-related genes with blood pressure changes and incident hypertension: The Gen-Salt Study". In: *Journal of Human Hypertension* 32 (Feb. 2018), pp. 287–293. DOI: 10.1038/s41371-018-0041-6.
- [15] A. I. Casas et al. "NOX4-dependent neuronal autotoxicity and BBB breakdown explain the superior sensitivity of the brain to ischemic damage". In: *Proceedings of the National Academy of Sciences* 114 (Oct. 2017), pp. 12315–12320. DOI: 10.1073/pnas.1705034114.

5

5

NOX5-induced uncoupling of endothelial NO synthase is a causal mechanism and theragnostic target of an age-related hypertension endotype

Adapted from: M. H. Elbatreek et al. "NOX5-induced uncoupling of endothelial NO synthase is a causal mechanism and theragnostic target of an age-related hypertension endotype". In: *PLOS Biology* 18 (Nov. 2020). Ed. by C. W. Lo, e3000885. DOI: 10.1371/journal.pbio.3000885.

Abstract

Hypertension is the most important cause of death and disability in the elderly. In 9 out of 10 cases, the molecular cause, however, is unknown. One mechanistic hypothesis involves impaired endothelium-dependent vasodilation through reactive oxygen species (ROS) formation. Indeed, ROS forming NADPH oxidase (Nox) genes associate with hypertension, yet target validation has been negative. We re-investigate this association by molecular network analysis and identify Nox5, not present in rodents, as a sole neighbor to human vasodilatory endothelial nitric oxide (NO) signaling. In hypertensive patients, endothelial microparticles indeed contained higher levels of NOX5—but not NOX1, NOX2, or NOX4—with a bimodal distribution correlating with disease severity. Mechanistically, mice expressing human Nox5 in endothelial cells developed—upon aging—severe systolic hypertension and impaired endothelium-dependent vasodilation due to uncoupled NO synthase. We conclude that NOX5-induced uncoupling of endothelial NO synthase is a causal mechanism and theragnostic target of an age-related hypertension endotype. Nox5 knock-in (KI) mice represent the first mechanism-based animal model of hypertension.

5.1 Introduction

Hypertension is of major medical relevance as a risk factor for myocardial infarction, stroke, and other chronic conditions and death [1]. Except for 5% of patients with secondary hypertension (due to renal artery stenosis, adrenal adenomas, pheochromocytomas, and numerous single-gene mutations involving renal transporters [2]), the cause of hypertension remains unknown in the remaining 95% of all cases. In these cases of so-called ‘essential hypertension,’ treatments have to focus on symptomatic vasodilatory drug therapy and lifestyle management. Even this symptomatic antihypertensive therapy is sometimes ineffective, i.e., in treatment-resistant hypertension, and requires a high number to treat, with many patients still experiencing adverse outcomes such as stroke and myocardial infarction [3].

One molecular mechanism of hypertension that has been suggested for decades is oxidative stress, i.e., an unphysiological production of reactive oxygen species (ROS), which in blood vessels interferes with vasodilation by the endothelium-derived relaxing factor, nitric oxide (NO) [4]. No hypertension-relevant cellular source of ROS, however, has been identified to either prove this hypothesis or exploit it for a mechanism-based or even curative clinical therapy.

Recent genome-wide association studies (GWAS) [5] in search of hypertension risk genes point towards NADPH oxidases (Nox), the only known enzyme family dedicated to ROS formation, in particular the genes Nox4 and Nox5. This matches preclinical studies, excluding other vascular NOX isoforms, i.e., NOX1 and NOX2, for causing hypertension, unless animals are infused with pro-hypertensive agents [6–8]. Mice overexpressing p22phox (an essential component of NOX 1–4) or lacking the antioxidant extracellular superoxide dismutase (ec-SOD) in the vasculature develop hypertension upon ageing [9]. Thus, the data regarding the role of NOX1 and NOX2 regarding preclinical hypertension are controversial and suggest that these isoforms may contribute to basal blood pressure regulation but not hypertension [6, 8, 10, 11]. With respect to NOX4, this isoform is widely expressed, but it appears irrelevant for blood pressure or hypertension [12, 13] and is rather vasoprotective [13]. Regarding NOX5, this enzyme is physio-

logically expressed in vascular endothelial cells of human blood vessels and may be associated with diabetic nephropathy [14–16]; mice expressing human Nox5 in vascular smooth muscle cells are, however, normotensive [17]. Thus, the role of NOX5 in hypertension is unclear.

Network medicine [18] predicts that, for most particularly complex diseases, not a single protein but protein modules, i.e., sub-graphs of the interactome, are in fact relevant [19–21]. Therefore, we set out to re-investigate the association of NOX with hypertension and NO-dependent vasodilation using 3 complementing unbiased in silico approaches and to validate any prediction both in mice and—if possible—also human patient samples.

5

5.2 Results

5.2.1 NOX5 is the only direct neighbor of endothelial NO-cyclic GMP signaling

To explore the possible link between NOX isoforms and hypertension and NO-dependent vasodilation, we constructed a pruned molecular subnetwork from the first neighbors of NOX family members and nitric oxide-cyclic GMP related proteins as seed nodes in the experimentally validated interactome obtained from the IID [22] interactome database. These included NOX1, NOX3, NOX4, NOX5, NOS1, NOS3, GUCYA1, GUCYA2, GUCYB1, PDE5A, PDE9A, and PRKG1, but not NOX2 and NOS2. The resulting subnetwork was further pruned according to the subnetwork-participation-degree (SPD) to correct for hub nodes (i.e. proteins that occurred mainly because of their high number of interactions in the whole network). This resulted in a disease module consisting of several connected components, which revealed that all NOX isoforms but NOX5 were excluded as a close neighbor of endothelial NO-cyclic GMP signaling. NOX5 fell into the same connected component with the genes encoding the NO receptor, GUCYA1, GUCYA2 and GUCYB1, and with endothelial NOS (NOS3) (Fig 5.1A). From IID, this connection is based on a physical interaction suggested by high-throughput affinity chromatography [23].

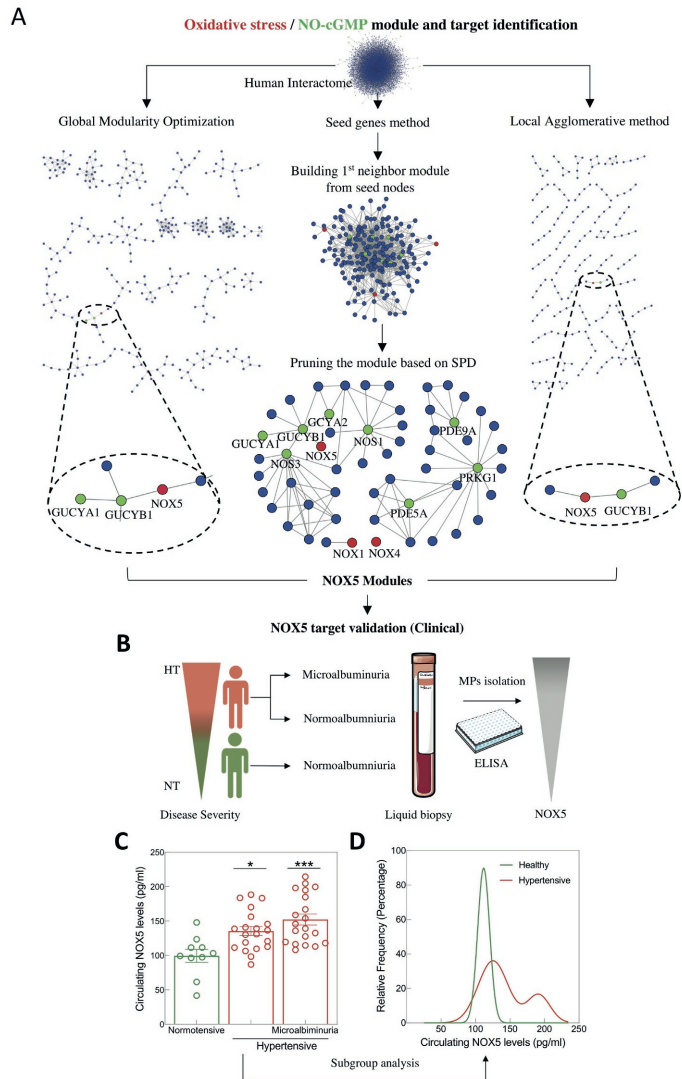


Figure 5.1: Identification of NOX5 as direct neighbor of endothelial NO-cGMP signaling and clinical validation in hypertension.

To cross-check our in-silico findings of a NOX disease module, we

Figure 5.1 (cont.): A. NOX module was constructed by 1st neighbor subnetwork pruned based on subnetwork-participation-degree (SPD) (the middle panel) where NOX isoforms (red nodes) and NO-cGMP related proteins (green nodes) were used as seed nodes. The resulting NOX module was confirmed with two disease module identification methods, global modularity optimization (the left panel) and the agglomerative local method (the right panel). All the methods identified NOX5 as the closest link to NO-cGMP signaling and excluded NOX1-4 and all other known ROS sources (not shown). B. NOX5 levels in endothelial microparticles (MPs) isolated from plasma of normotensive (NT), normoalbuminuric subjects and hypertensive (HT) normoalbuminuric and microalbuminuric patients were measured by ELISA. C. NOX5 levels were increased in hypertensive patients with normoalbuminuria (n=20) compared to normotensive subjects (n=10). NOX5 levels were even higher in hypertensive patients with microalbuminuria (n=20). Comparison between groups was done by one-way ANOVA followed by Tukey's multiple comparisons test. D. Subgroup analysis of all hypertensive patients shows bimodal distribution ($p=0.0007$, two-tailed F-test, adjusted $r\text{-squared}=0.9973$). All data are represented as mean \pm S.E.M. of n independent experiments * $P < 0.05$, *** $P < 0.001$.

employed two additional computational network module identification methods, global modularity optimization, and agglomerative local search, both of which have been top-performers in the recent Module Identification DREAM Challenge [24]. In brief, the global modularity optimization approach combines multiple module detection algorithms to avoid suboptimal partitions resulting from individual algorithms [25]. The agglomerative local method uses the SPICi algorithm [26] to optimize the local density of modules around seed nodes. With all three in-silico methods, we reached the same conclusion: Exclusion of all NOX genes, except NOX5, as a direct neighbor of endothelial nitric oxide-cyclic GMP signaling (Fig 5.1A).

5.2.2 NOX5 protein levels are increased in hypertensive patients

To test this causal endothelial NOX5 hypothesis for human hypertension, we enrolled consecutive outpatients with essential hypertension and a baseline estimated glomerular filtration rate (eGFR) ≥ 30 mL/min/1.73 m^2 . Study subjects were divided into 3 groups, healthy (n=10), hypertensive patients with normoalbuminuria (n=20), and hypertensive patients with moderately increased albuminuria (previously termed microalbuminuria) (n=20). The baseline characteristics of the patients are listed in Table 5.1. To

measure NOX5 protein levels, circulating endothelial microparticles, i.e. membrane vesicles that are released from endothelial cells upon cellular activation or cell death and carry endothelial proteins [27], were isolated from plasma of the participants (Fig 5.1B). We observed higher NOX5 protein levels in endothelial microparticles of hypertensive versus normotensive subjects and within hypertensive subjects, patients with microalbuminuria showed even higher NOX5 protein levels (Fig 5.1C). These data suggest that NOX5 levels are associated with hypertension and correlate with disease severity. Hypertension is rather an umbrella term that may involve different molecular mechanisms all resulting in a similar phenotype, i.e. elevated blood pressure. NOX5-dependent hypertension may be such an endotype but apply only to a subset of patients [28–30]. We, therefore, performed a subgroup analysis of all hypertensive patients and, indeed, NOX5 levels showed a bimodal distribution (Fig 5.1D). Based on this, approximately every fourth hypertensive patient would fall into a high NOX5 mechanotype, which according to the PPI interaction would cause NO-cGMP signaling dysfunction. With respect to other vascular NOX isoforms i.e., NOX1, NOX2 and NOX4, their levels did not differ significantly between normotensive and hypertensive subjects. However, patients with microalbuminuria showed higher NOX1, but lower NOX2 levels compared to normotensive subjects (Fig 5.S1A–C). The higher NOX1 levels in those patients is in agreement with previous preclinical data showing that NOX1 is not associated with elevation of blood pressure, but with renal oxidative stress in a model of chronic hypertension [6] and renal pathology [31–33]. In addition, we determined plasma asymmetric dimethylarginine (ADMA) levels, a biomarker of NOS uncoupling and endothelial dysfunction [34]. We found that ADMA levels were significantly increased in hypertensive patients compared to healthy individuals (Fig 5.S1D), which is in agreement with previous findings [35, 36].

5.2.3 Endothelial NOX5 induces age-related hypertension

In the absence of NOX5 specific inhibitors, we tested the possible role of NOX5 in endothelial NO-cGMP signaling dysfunction and hypertension in mice. Mice, however, lack the *Nox5* gene. We, there-

Table 5.1: Baseline characteristics in healthy individuals and hypertensive patients.

	Healthy (n = 10)	HTN no albuminuria (n = 20)	HTN with microalbuminuria (n = 20)	p-Value
Age (y)	44 ± 4	56 ± 3	60 ± 3	0.012
Men	7 (70%)	12 (60%)	13 (65%)	0.859
Diabetes	0 (0%)	3 (15%)	6 (30%)	0.118
BMI	24.1 ± 0.94	26.1 ± 0.76	26.9 ± 0.73	0.096
Smoking	0 (0%)	7 (35%)	9 (45%)	0.042
T. Chol	178 ± 11	193 ± 9	199 ± 8	0.358
Triglyceride	158 ± 42	146 ± 19	214 ± 27	0.139
HDL	50 ± 3	47 ± 3	41 ± 2	0.081
LDL	97 ± 11	117 ± 8	115 ± 9	0.36
Fasting glucose	96 ± 6	104 ± 8	120 ± 7	0.107
Serum Cr	0.81 ± 0.07	0.91 ± 0.05	1.01 ± 0.07	0.144
Uric acid	6.2 ± 0.53	6.1 ± 0.31	6.5 ± 0.29	0.656
GFR	87.0 ± 5.69	83.2 ± 3.57	79.3 ± 5.52	0.606
FRS	5.2 ± 1.73	8.2 ± 1.78	10.9 ± 1.9	0.168
ACR	0.008 ± 0.0009	0.010 ± 0.0001	0.059 ± 0.0008	<0.001
hsCRP	0.56 ± 0.33	0.25 ± 0.03	0.52 ± 0.09	0.188
Adiponectin	16.0 ± 2.75	20.4 ± 3.08	18.5 ± 2.99	0.671
NT-pro-BNP	75.1 ± 11.95	80.9 ± 8.76	95.8 ± 16.9	0.572
Medications				
ACE-I	0 (0%)	4 (20%)	1 (5%)	0.143
ARB	0 (0%)	14 (70%)	16 (80%)	<0.001
CCB	0 (0%)	13 (65%)	14 (70%)	<0.001
Beta-blocker	0 (0%)	2 (10%)	9 (45%)	0.005
Thiazides	0 (0%)	6 (30%)	8 (40%)	0.069
Statin	1 (10%)	3 (15%)	7 (35%)	0.185

fore, analyzed a knock-in mouse model expressing human Nox5 in its physiological endothelial cell location [37] (Fig 5.2A). In young (9 – 15 weeks old) NOX5 KI mice of both genders (n = 19-20), systolic blood pressure, diastolic blood pressure and mean arterial pressure were, however, not different from age- and sex-matched wild type (WT) mice (Fig 5.2B and 5.2C and Fig 5.S2A). Upon aging (68 – 87 weeks), though, systolic blood pressure and mean arterial pressure were significantly elevated throughout the day in KI (n = 33) compared to age- and sex-matched WT mice (n = 31) (Fig 5.2D and Fig 5.S2B). Diastolic blood pressure remained unmodified (Fig 5.2E) as was heart to body weight ratio (Fig 5.S3A) indicating that there was no cardiac hypertrophy in KI mice, which is in line with the late (age-dependent) development of hypertension. In addition, cardiac hypertrophy does not necessarily exist with hypertension. For example, other hypertensive animal models such as eNOS KO mice do not have cardiac hypertrophy [38, 39]. This is also in agreement with clinical observations that not all hypertensive patients (subgroups) have cardiac hypertrophy [40–42]. Notably, there was no difference in blood pressure between aged KI male (72-hour MAP, 120 ± 3.4 , n = 15, p = 0.13) and female (114 ± 2.1 , n = 18) mice.

Taken together, our observations indicate that in mice, expression of NOX5 in the endothelium leads, upon aging, to a selective elevation of systolic arterial blood pressure. Having established the potential of NOX5 to induce a hypertensive phenotype, we wanted to test the mechanistic link to vascular NO-cGMP signaling as suggested from the in-silico network analysis Fig. 5.2. Preclinical validation of NOX5 in hypertension. A. Both young WT and NOX5 KI mice were normotensive however only KI mice become hypertensive upon ageing. B and C. There was no significant difference in systolic (B) and diastolic (C) blood pressure between young WT (n=19) and KI (n=20). D and E. Aged KI mice (n=33) had higher systolic (D) but similar diastolic (E) blood pressure compared to WT (n=31). Telemetry data were analyzed by two-way repeated measures ANOVA followed by Sidak's multiple comparisons test. All data are represented as mean \pm S.E.M. of n individual animals **P < 0.01.

5.2.4 Endothelial NOX5 causes endothelial dysfunction by uncoupling NOS

5 In the thoracic aorta, femoral artery and saphenous artery isolated from aged KI and WT mice of both genders ($n = 9$), we analyzed the structural, smooth muscle and endothelial vasomotor properties. Collectively, these blood vessels cover the entire range of large elastic conduit, muscular conduit and small muscular resistance-sized arteries, respectively. In the thoracic aorta, femoral artery and saphenous artery of the aged animals, the relation between resting wall tension and arterial lumen diameter did not differ between KI and WT mice (Fig 5.S4A-C). It is therefore unlikely that the blood pressure phenotype of the KI mice resulted from stiffening or inward remodeling of the conduit or resistance arteries. This is in agreement with previous studies that structural stiffening of arteries is not a general observation in rodent models of (essential) hypertension [43–46]. Also, clinical data, especially from aged hypertensive patients, is in line with this finding [47–50]. To test the effect of endothelial NOX5 on endothelium-dependent, NO-cGMP mediated relaxation, arterial segments were pre-contracted by either depolarization (K^+), $\alpha 1$ -adrenergic activation (phenylephrine), or endothelin-1; vasorelaxation was then induced by acetylcholine (ACh) the classical endothelium-derived relaxing factor stimulant [51]. In femoral arteries, irrespective whether pre-contracted with K^+ , phenylephrine or endothelin-1, the amplitudes of ACh-induced relaxing responses were significantly smaller in KI compared to WT mice (Fig 5.3A) (Fig 5.S5A and 5.S5D). Conversely, in the saphenous artery (Fig 5.3B) (Fig 5.S5B and 5.S5E) and thoracic aorta (Fig 5.S5C, 5.S5F and 5.S5G), ACh-induced relaxing responses did not differ between KI and WT mice. When comparing ours to previous studies, the ACh-induced relaxation of the saphenous arteries of both mice groups seemed attenuated [52–54]. There are two possible explanations for this discrepancy. First, we have used very old mice and other studies used young mice [52–54]. Second, we have used mice with a mixed genetic background (80% 129/Sv and 20% C57Bl6) and previous data showed that endothelium-dependent relaxation is smaller in 129/Sv compared with C57Bl6 mice [55].

In segments of the thoracic aorta and femoral artery contracted with 256 nM endothelin-1, the relaxing effects of Ach were reversed by 100 μ M L-NAME (a pharmacological inhibitor of NO synthases) and this did not differ significantly between preparations of KI and WT mice (Fig 5.S5H and 5.S5I). In the saphenous arteries of both types of mice, L-NAME did not modify the relaxing effects of Ach (Fig 5.S5J). This is in line with our earlier findings indicating that in these resistance-sized arteries Ach-induced relaxation is mediated by endothelium-dependent hyperpolarization but not NO [52–54]. To check whether this caliber-specific effect on vasomotor function was due to differential expression of NOX5 along the systemic arterial tree, we measured NOX5 gene expression by quantitative PCR. There was, however, no difference in NOX5 gene expression between thoracic aorta, femoral arteries and saphenous arteries of NOX5 KI mice (Fig 5.S6). These data suggest an aging-dependent dysfunction of endothelial NO-cGMP signaling due to NOX5 and that this effect is not uniformly distributed along the systemic arterial tree. We next tested whether, alternatively, chronic changes in the underlying arterial smooth muscle layer could have contributed to the observed blood pressure and vasomotor phenotypes in aged NOX5 KI mice. Contractile responses to K⁺ and the sensitivity and maximal responsiveness to phenylephrine and endothelin-1 did, however, not differ between KI and WT mice in all arterial segments (Fig 5.S7A-I). Also, the blunting of agonist-induced contractile responses by indomethacin was similar in the thoracic aorta of the KI and WT mice (Fig 5.S7G-I). To test which component of NO-cGMP signaling was most likely affected, we tested an uncoupling effect on endothelial NOS [56] or oxidative damage of the NO receptor, soluble guanylate cyclase yielding oxidized or heme-free apo-sGC [57, 58]. To test for sGC/apo-sGC, relaxing responses to the NO donor compound and sGC stimulator, PAPA/NO (0.01-10 μ M) [59], and to the apo-sGC activator, BAY 60-2770 (0.01-10 μ M) [57], were analyzed. Neither the PAPA/NO (Fig 5.S8A-C) nor the BAY 60-2770 response (Fig 5.S8D) differed between WT and KI mice. These observations suggested that sGC was not dysfunctional in aged NOX5 KI mice. Uncoupling of endothelial NOS is considered a major cause of endothelial dysfunction characterized by decreased NO formation and increased superoxide production and occurs

mainly when ROS oxidize the NOS cofactor tetrahydrobiopterin (H₄Bip) [60]. When we incubated femoral arteries of aged NOX5 KI mice, precontracted with phenylephrine, with the H₄Bip precursor sepiapterin (100 μ M), Ach-induced relaxations were greatly improved and became indistinguishable from those in WT (Fig 5.3A). Also, femoral arteries of aged NOX5 KI mice showed higher superoxide formation (as indicated by DHE staining) than WT and this increase was inhibited by pretreatment with the NOS inhibitor, L-NAME (Fig 5.S9). We also tested whether the impaired Ach-induced relaxation in the femoral arteries of NOX5KI mice could be reversed by adding antioxidants to the bath medium. Neither 10 μ M N-acetylcysteine nor 100 μ M tempol, however, were effective (Fig 5.S10). Collectively, these data suggest that endothelial NOX5 induces endothelial dysfunction by uncoupling of endothelial NOS leading to impaired endothelium-dependent relaxation of muscular conduit arteries and thus systolic hypertension (Fig 5.3C).

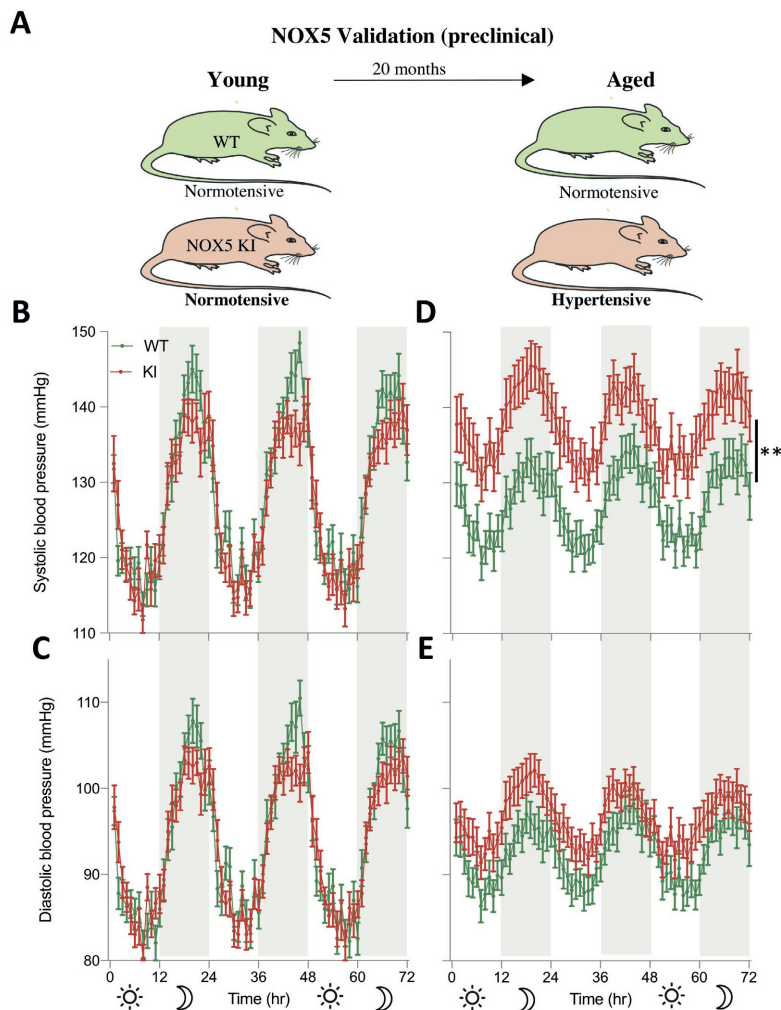


Figure 5.2: Preclinical validation of NOX5 in hypertension. (A) Both young WT and NOX5 KI mice were normotensive; however, only KI mice become hypertensive upon ageing. (B, C) There was no significant difference in systolic (B) and diastolic (C) blood pressure between young WT (n = 19) and KI (n = 20). (D, E) Aged KI mice (n = 33) had higher systolic (D) but similar diastolic (E) blood pressure compared to WT (n = 31). Telemetry data were analyzed by two-way repeated measures ANOVA followed by Sidak’s multiple comparisons test. All data are represented as mean \pm SEM of n individual animals $**p < 0.01$. All raw data are included in the S1 Data file. KI, knock-in; NOX5, NADPH oxidase 5; WT, wild type.

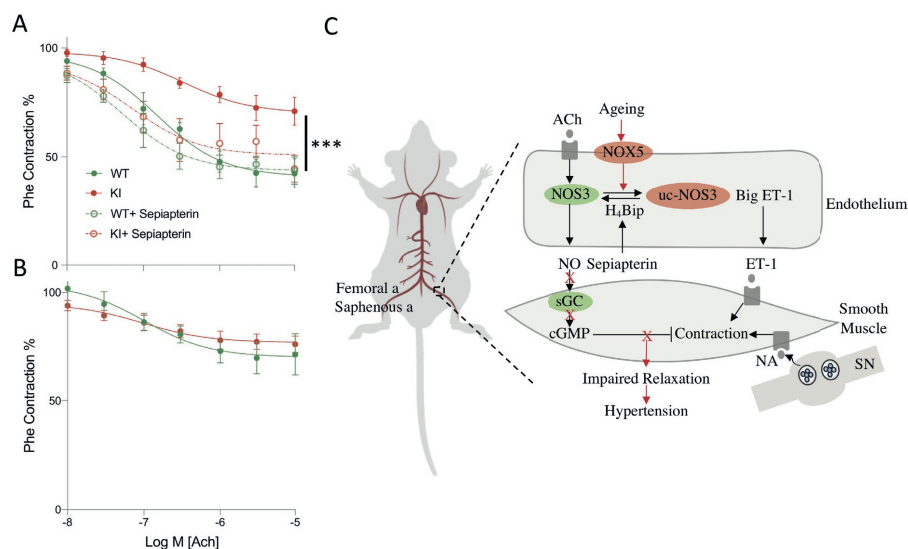


Figure 5.3: Endothelial NOX5 induces endothelial dysfunction and hypertension by uncoupling NOS. (A) Femoral arteries of aged KI mice ($n = 9$) precontracted with Phe had less responsiveness to ACh-induced relaxation compared to WT ($n = 9$), while pretreatment with sepiapterin ($100 \mu\text{M}$) improved relaxation in KI ($n = 4$), which was not different from WT ($n = 3$). (B) Saphenous arteries of aged KI mice ($n = 9$) precontracted with Phe showed no difference in ACh-induced relaxation compared to WT ($n = 8$). Myograph data were analyzed by two-way ANOVA followed by Sidak's multiple comparisons test. (C) Schematic representation of NOX5-induced age-dependent hypertension. In ageing, endothelial NOX5 is activated and interferes with normal NO-cGMP signaling, which results in impaired vascular smooth muscle relaxations and raised blood pressure. All data are represented as mean \pm SEM of n individual animals *** $p < 0.001$. All raw data are included in the S1 Data file. ACh, acetylcholine; cGMP, cyclic GMP; ET-1, endothelin-1; H₄Bip, tetrahydrobiopterin; KI, knock-in; NA, noradrenaline; NO, nitric oxide; NOS3, endothelial nitric oxide synthase; NOX5, NADPH oxidase; Phe, phenylephrine; sGC, soluble guanylate cyclase; SN, sympathetic nerve; uc-NOS3, uncoupled NOS3; WT, wild type.

5.3 Discussion

Based on human genetic, human clinical, and genetic pre-clinical mechanistic validation, we here report the first identified causal molecular mechanism of human systolic hypertension associated with aging. This endotype affects approximately one in four patients and molecularly consists of a NOX5-induced uncoupling of endothelial NO synthase followed by impaired endothelium-dependent vasodilation in muscular conduit arteries. Detection of elevated levels of NOX5 in circulating microparticles appears to serve as a mechanism-based liquid biopsy marker to stratify patients for therapeutic intervention. Based on our in-vivo validation such an intervention may include the H4Bpi precursor and NOS recoupling agent, sepiapterin and, once such compounds should become clinically available [61–64], a NOX5 inhibitor. Here we used a KI mouse model that expresses human NOX5 in endothelial cells and white blood cells mimicking, to a large extent, the physiological pattern of NOX5 expression in humans [37]. While we cannot exclude a potential contribution of leukocytic NOX5 on hypertension in aged animals in our in vivo experiments, a contribution of adhering leukocytes is unlikely in our ex vivo experiments. Expression of human NOX5 in mice led with aging to severe systolic hypertension. This was not due to stiffening, structural remodeling, or increased sensitivity to vasoconstrictor stimuli in the systemic arterial tree. It rather resulted from a regionally selective and specific attenuation of NO-mediated endothelium-dependent relaxation in medium-sized muscular conduit arteries via uncoupling of NOS. Collectively, our data warrant clinical proof-of-concept trials aiming at a theragnostic [65] strategy in carefully stratified hypertensive patients based on the detection of elevated NOX5, ROS overproduction, e.g. by oxidatively modified proteins, and mechanism-based functional repair by a network pharmacology approach that inhibits NOX5 and recouples NOS. If successful, this will be the first case of a molecular redefinition of the phenotypic disease definition, essential hypertension, and the first step towards precision medicine in a currently high number-needed-to-treat indication. This may be the case for about a fourth of all hypertensives. Our estimate, though significant, is limited by the relatively small number of patients in the present study and needs to be validated in larger and international hypertension cohort. The fact

that NOX5 is also involved in a worse outcome in stroke [37], for which hypertension is a major risk factor [66], and correlates with atherosclerosis [67], this approach may not only lower blood pressure, but also two major consequences of hypertension, stroke and myocardial infarction. On a broader scale, our present three-fold interactome-based approach for disease module discovery, pre-clinical and clinical validation appears to be applicable to a wide range of common or complex diseases. We here identify NOX5 as the missing link between ROS and impaired NO signaling. The full module consists of NOX5, NOS3, the different subunits of the NO receptor sGC and the phosphodiesterases, PDE5 and PDE9, as well as the cGMP dependent protein kinase, PKG1. The dysregulation of such a module may be best treated by multiple drugs targeting different protein components. In the present case, the WT mice mimic pharmacological NOX5 inhibition (not yet available) and the treatment with sepiapterin, NOS recoupling. Disease module construction is a young research field at the interface of biomedicine and bioinformatics. We began with a seed gene-based approach, based on clinically validated proteins. NOXs were suggested by GWAS in old [5] but not in younger patients [68] and are the only known enzyme family solely dedicated to ROS formation [69]. The NO-cGMP pathway is important for blood pressure regulation and its dysfunction a hallmark of hypertension [70]. Our approach yielded a module containing a sufficient set of NO-cGMP signaling components and, importantly, with NOX5 as sole ROS source. We confirmed our findings independently by two complementing in-silico network module detection approaches. As a potential limitation of our in silico findings, the interaction between NOX5 and cGMP-related proteins has been shown experimentally so far in only one study [23]. Our clinical data are yet still correlative. Endothelial microparticles are, however, well established surrogate biomarkers associated with hypertension and its progression [71, 72]. They produce ROS, contain NOX, induce endothelial dysfunction and impair endothelium-dependent relaxation [73]. In human endothelial cells, angiotensin II, the target of clinically used angiotensin type 1 receptor blockers and angiotensin converting enzyme inhibitors, and the pro-hypertensive autacoid, endothelin-1, increase NOX5 expression (gene and protein) and activity [74]. In addition to its physiological vascular expression in endothelial cells, NOX5 is also increased (gene, protein, and activity) to higher levels in human

renal proximal tubule cells of hypertensive patients [75], which may contribute to the observed correlation of NOX5, blood pressure and microalbuminuria. Induction of NOX5 in smooth muscle cells does not cause hypertension per se but correlates with advanced atherosclerotic lesions and diseased coronary arteries show high NOX5 expression and activity [67]. Hypertension is rather an umbrella term for different blood pressure-elevating mechanisms. Some of these may be related to worse clinical outcomes, some not. NOX5-dependent hypertension, however, appears to be disease-relevant as this molecular mechanism also leads to or aggravates hypertension-associated clinical outcomes, stroke [37], myocardial infarction [76] and renal failure [75, 77].

GWAS had delivered two NOX candidate genes, Nox4 and Nox5. Knocking out Nox4 is without a blood pressure phenotype [12, 78]; in many models, Nox4 was rather a vasoprotective [13, 79, 80]. The enzymatic product of NOX4, H₂O₂, activates NOS [81] and is an alternative endothelium-derived relaxing factor in its own right [82–84]. This left Nox5 as a candidate gene, which was confirmed by first neighbor analysis linked to NO-cGMP signaling. Unlike NOX4, NOX5 produces superoxide which can uncouple endothelial NO synthase in hypertension by oxidation of H₄Bpi [85, 86] in a sepiapterin-reversible manner [87] or scavenge NO to form peroxynitrite, which may also act as a cytotoxic agent or vasodilator [88, 89]. The mechanistic validation of this role of NOX5 in hypertension was performed in a pre-clinical mouse KI model expressing NOX5, not present in the mouse genome, in the physiological cell type, endothelial cells where NOX5 is important for endothelial migration and angiogenesis [90]. In mice expressing NOX5 non-physiologically in smooth muscle cells, blood pressure is normal and angiotensin II-induced pressure effects are not augmented [74]. Of note, in other animal models of hypertension—e.g., spontaneously hypertensive (SHR) and Dahl salt-sensitive rats, in which NOX5 is not present—antioxidants can reduce blood pressure [91–93]. This may suggest that, if relevant also to humans, other endotypes of hypertension may involve other ROS sources (mitochondria, xanthine oxidase, or myeloperoxidase).

Of particular interest is the regionally selective effect of NOX5, which was reminiscent of the vascular heterogeneity in age-related en-

dothelial dysfunction [94, 95]. It was present in muscular femoral conduit arteries but not in the small resistance-sized saphenous arteries (lumen diameter $<250\ \mu\text{m}$) of the KI mice and could thus selectively result in systolic hypertension with accompanying elevation of arterial pulse pressure. However, we have not considered other small resistance arteries, including renal arteries, which control sodium handling and blood pressure and are more relevant in secondary renovascular hypertension [96]. This effect of NOX5 was not reversed in our ex vivo experiments by antioxidants, presumably because NOS uncoupling was already established chronically in vivo by NOX5-derived superoxide or NO production was reduced as part of age-related endothelial dysfunction. Preclinically, sepiapterin or H_4Bip lower elevated blood pressure induced by NOS uncoupling [97, 98]. As a potential limitation to our study, we did not re-examine in vivo in NOX5 KI mice the efficacy of sepiapterin to reverse NOS uncoupling and hypertension. Another limitation and general pitfall of all animal models (particularly the KI mouse model) are differences between human and mouse physiology and that expressing a human protein in mice may not wholly reflect its role in human pathology [99].

With respect to the clinical dimension of our findings, our observation that plasma ADMA levels were elevated in hypertensive patients speaks in favor of NOS uncoupling [35, 36]; yet a similar mechanistic link to NOX5 as we obtained pre-clinically in NOX5 KI mice would require the ex vivo analysis of isolated human blood vessels or an interventional trial with sepiapterin. Indeed, sepiapterin analogs, i.e., folic acid and H_4Bip , are clinically effective to reduce elevated blood pressure by improving endothelial function [100, 101]. Folic acid, either alone [102] or in combination with antihypertensives [103, 104], reduces the risk of cardiovascular and cerebrovascular events in hypertensive patients.

Overall, our findings using in silico network approaches as well as further clinical and preclinical validation explain the long-observed correlation between oxidative stress, endothelial dysfunction, and systolic hypertension. Humanized endothelial cell NOX5 KI mice represent the first mechanism-based animal model of human age-related hypertension and endothelial dysfunction. Therapeutically, NOX5 in-

hibition and NOS recoupling, ideally in combination with mechanistic biomarker stratification, e.g., based on endothelial microparticle liquid biopsies, represents a first-in-class mechanism-based approach for curative antihypertensive therapy obviating the need for symptomatic vasodilators.

Materials and Methods

Ethics Statement

This human subjects' study was approved by the research ethics committee of Taipei Veterans General Hospital (Taipei Veterans General Hospital Institutional Review Board, No.: 96-12-42A), and all participants provided their written informed consent. The study complied with the Declaration of Helsinki. It was approved by the appropriate health authorities, an independent ethics committees, and the Institutional Review Board of Taipei Veterans General Hospital. All animal experimental research was approved by the Animal Welfare Body (IvD) of Maastricht University, the Animal Ethics Committee (Dier Experimenten Commissie [DEC]), and the Central Authority for Scientific Procedures on Animals (Centrale Commissie Dierproeven [CCD]). The DEC project numbers are 2012-123 and PV 2017-029, and the license number is AVD1070020174486. The animal care and use protocol adhere to the ARRIVE guidelines ("Animal Research: Reporting of In Vivo Experiments"; <https://arriveguidelines.org/>), Wet op dierproeven (WoD; 2014, <https://wetten.overheid.nl/BWBR0003081/2019-01-01>), the Dierproevenregeling (DpR; 2014, <https://zoek.officielebekendmakingen.nl/stcrt-2014-34746.html>), the Dierproevenbesluit (DpB; 2014, <https://wetten.overheid.nl/BWBR0035866/2014-12-18>), and the EU Directive 2010/63 (2010, <https://eur-lex.europa.eu/eli/dir/2010/63/2019-06-26>).

Study design

The human subjects sample size was determined by G*Power software. For mice, we used a power analysis according to the

formula $n=2x \ s2x \ (Za/2+Zb)2/D2$ (L. Sachs, *Angewandte Statistik*, Springer, 1983, Berlin, Springer Verlag). Human subjects with history or clinical evidence of angina, myocardial infarction, congestive heart failure, peripheral vascular disease, inflammatory disease, or any disease predisposing to vasculitis were excluded. Causes of secondary hypertension were excluded by appropriate investigations. Patients with stage 4 and 5 chronic kidney disease (GFR <30 mL/min/1.73 m^2) were also excluded. Human samples were allocated to different groups based on blood pressure and albuminuria values. Mice were allocated to experimental groups according to genotypes. Investigators were blinded to the experimental groups. Replicate experiments were successful. All experiments were reproduced at least three times with independent biological samples.

5

In-silico methods

We extracted a molecular subnetwork from experimentally validated protein-protein interactions from the IID [22] database (interactome) version 2018-11 using NOX family members and nitric oxide-cyclic GMP related proteins as seed nodes. This set of seeds comprises NOX1, NOX3, NOX4, NOX5, NOS1, NOS3, GUCYA1, GUCYA2, GUCYB1, PDE5A, PDE9A and PRKG1. We obtained the subnetwork induced by all first neighbors of the seed genes from the interactome. The induced subnetwork was then pruned according to the subnetwork-participation-degree (SPD), defined as the degree of the node (protein) inside the subnetwork normalized by the degree of the node in the full interactome. The SPD quantifies how enriched the interactions of a given protein are in a given subnetwork. This way, we emerged a weighted disease module, which is represented by a set of connected components and some single nodes. We selected an SPD cutoff corresponding to 80% of the cumulative sum of the percentage of the nodes as a cutoff value in the pruning step, as this includes most module-specific interactions while excluding most non-specific nodes. The final subnetwork consisted of 56 proteins and 83 protein-protein interactions. Moreover, we applied the two top-ranked disease network module identification methods from the Module Identification DREAM Challenge [24] to the interactome. We selected these methods from two complementary categories

of methods, global and local. The main difference between these two methods is that global methods exploit the global structure information of PPI networks, while the local method considers only local neighbor information. The global modularity optimization method (M1) of the DREAM challenge bundled in the MONET tool and the agglomerative local method (L1) from SPICi tool [26] have been selected (as best-performers in their categories) and applied to find disease modules in the interactome. Note that M1 is an ensemble approach combining multiple module detection algorithms to avoid suboptimal partitions resulting from individual algorithms [25], which notably works without any seed nodes in this tool. The agglomerative L1 method clusters the network greedily, starting from automatically selected local seeds with a high weighted degree. This algorithm improves the local density of modules in the neighboring region of seed nodes.

Human study participants

We designed the current study based on a previous study in which we had enrolled consecutive outpatients with essential hypertension and a baseline estimated GFR ≥ 30 mL/min/1.73 m^2 at Taipei Veterans General Hospital between April 2008 and December 2008 [105, 106]. Hypertension was defined as systolic blood pressure ≥ 140 mmHg, diastolic blood pressure ≥ 90 mmHg, or use of antihypertensive drugs. Subjects with history or clinical evidence of angina, myocardial infarction, congestive heart failure, peripheral vascular disease, inflammatory disease, or any disease predisposing to vasculitis were excluded. Causes of secondary hypertension were excluded by appropriate investigations. Patients with stage 4 and 5 chronic kidney disease (GFR < 30 mL/min/1.73 m^2) were also excluded. Medical history, including cardiovascular risk factors, previous and present cardiovascular events, and current medication regimen, was obtained during a personal interview and from medical files. Weight, height, and waist circumference were measured and body mass index (BMI) was calculated. Brachial blood pressure was measured by a physician with a mercury sphygmomanometer after patients sat for 15 minutes or longer. The average of 3 measurements was used for the analysis.

Blood and urine measurements

5 Venous blood samples were collected from all patients after 8 hours of overnight fasting for measurement. The blood samples were centrifuged at 3000 rpm for 10 minutes immediately after collection, and the plasma samples were kept frozen at -70°C until analysis. Each standard and plasma sample was analyzed twice, and the mean value was used in all subsequent analyses. The plasma high-sensitivity C-reactive protein (hs-CRP) level was determined using a latex-enhanced immunonephelometric assay (Dade Behring, Marburg, Germany). Plasma N-terminal pro b-type natriuretic peptide (NT-proBNP) was determined by a sandwich immunoassay (EIMA) with two antibodies (Cortez Diagnostics, Calabasas, CA, USA). Plasma ADMA levels were measured by ADMA Fast ELISA kit (DLD Diagnostika GMBH, Hamburg, Germany). Overnight urine samples were obtained for measurement of the albumin excretion rate. Normoalbuminuria was defined as albumin excretion rate of less than 20mg/min, moderately increased albuminuria (previously known as microalbuminuria) was defined as albumin excretion rate of 20–200mg/min, and severely increased albuminuria (previously known as macroalbuminuria) was defined as albumin excretion rate more than 200 mg/min.

Endothelial microparticles extraction and measurement of NOX5

CD144+ microparticles were isolated as described with modifications [107, 108]. Briefly, Dynabeads G (Invitrogen, Carlsbad, CA) were washed with PBS containing 0.1% BSA and then reconstituted with PBS. Anti-CD144 antibody (Santa Cruz Biotechnology, Dallas, TX) which specifically targets endothelial cells was mixed with prewashed Dynabeads G for 2 hours and then incubated with plasma samples at 1:200 dilution overnight at 4°C. After precipitation, Dynabeads G were washed with PBS and 1% Tween-20 three times. The purity of CD144+ MPs, determined by FACS analysis, was 70%±5.6%. With the use of FITC-conjugated beads as size references, the size of such particles was assessed to be <0.5 μ m in diameter. Human NADPH oxidase 5 (NOX5) levels were measured using commercially available enzyme-linked

immunosorbent assay (ELISA, Cusabio Technology LLC, Houston, Texas) kit according to the manufacturer's instructions. Samples were stored at -70°C from the date of collection in 2008 until testing for NOX5 in 2014 (a totally 50 samples were available). The intra-assay and inter-assay variation coefficients of the tests were <8% and <10%, respectively. NOX1, NOX2 and NOX4 were also measured by ELISA kits (Cloud-Clone, Katy, TX, USA) in another cohort of patients. The baseline characteristics of those patients are listed in Supplementary Table 5.S1.

Animals

Mice naturally do not express the NOX5 gene, therefore, we have generated and validated humanized NOX5 KI mice as previously described [37]. Briefly, the model was developed using the hypoxanthine phospho-ribosyl-transferase (Hprt) targeted transgenic approach under the control of the Tie2 promoter. Therefore, our NOX5 KI mice express the NOX5 in endothelial and white blood cells which mimic the physiological human expression of NOX5. The expression of NOX5 in the KI mice tissues was previously validated by quantitative real-time PCR and compared to Wild Type (WT) mice [37]. Age- and gender-matched groups of male and female mice (9-15 weeks old, n = 19-20) and (68-87 weeks old, n = 31-33) were used. All mice were allowed free access to water and food in a temperature-regulated room (22°C) and placed in a 12 h light-dark cycle.

Blood Pressure Recording (telemetry)

NOX5 KI and WT Mice were anesthetized with isoflurane (induction, 3-4%; maintenance, 1.5-2.5%) and echocardiography (ultrasound) was performed (Fig 5.S11). To implant the telemetry transmitters, 5 days after the ultrasound, mice were anesthetized with the same protocol and preoperative analgesia was done by subcutaneous injection of 0.05 mg/kg buprenorphine repeated every 12 hours. Each mouse was placed on a heating pad (UNO temperature control unit, UNO Roestvaststaal BV) and body temperature was monitored using a rectal probe and maintained at 37.0°C using a feedback-controlled infrared light. An incision in the skin overlying the carotid artery was made.

Via this incision, in the subcutaneous space of the flank a pocket was created for inserting the telemetry transmitter (TA11PA-C10; Data Sciences, Inc., St. Paul, MN) to monitor blood pressure, heart rate, and motor activity. The left carotid artery was dissected free and 3 ligatures (5-0, silk) were placed: at the bifurcation of the internal and external carotid to close the vessel, at the heart to temporarily close the vessel and one in between to fixate the catheter. Via a small hole cut in the artery, the catheter was introduced and advanced into the aortic arch. Then, the pocket in the flank was filled with 3mL prewarmed saline and the transmitter was placed in the pocket. The wound was then closed using a polysorb 5-0 suture. All surgical procedures were performed under aseptic conditions. Post-operative analgesia was done by a subcutaneous injection of 0.05 mg/kg buprenorphine after 6 hr. and 5 mg/kg carprofen after 24 and 48 hr. Mice were allowed to recover for 7-14 days before starting the measurement. Mice were housed individually in a quiet room. Blood pressure was measured over a 72-hour period [109–111], with 10 cycles of 75 sec. per hour. Radio signals from the transmitter were continuously monitored with a fully automated data-acquisition system (Dataquest A.R.T.; Data Sciences, Inc.). Mice were sacrificed by CO₂/O₂ inhalation and organs were taken out for further analysis. Organ and body weights are presented in Fig 5.S3.

Myograph

After mice were sacrificed, thoracic aortae, femoral and saphenous arteries were dissected free from perivascular adipose tissue and mounted in a wire myograph (DMT, Aarhus, DK). The organ chamber was filled with Krebs-Ringer bicarbonate-buffered salt solution (KRB) that was continuously aerated with 95% O₂/5% CO₂ and maintained at 37°C. The passive stretch procedure was performed to mimic the physiologically relevant internal lumen diameter as previously described [112]. Arterial contractile and relaxing responses were recorded at lumen diameters corresponding to a distending pressure of 100 mmHg in the thoracic aorta and femoral artery and at 90% of this diameter in the resistance-sized saphenous artery. This is justified because diastolic arterial blood pressure did not differ significantly between the aged KI and WT mice. In view of

the comparable diameter-tension relationships, these diameters did not differ significantly between the two mouse strains (Fig 5.S12). The diameter tension relationships were constructed according to the law of Laplace for a cylindrical tube: $P = T/R$ (with P for transmural pressure, T for wall tension and R for the lumen radius of the tube). Consequently, wall stiffness can be determined equally by recording 1) changes in tension in response to imposed changes in radius (wire myography) or 2) changes in radius in response to changes in transmural pressure (pressure myography) [113, 114]. Here we used the former approach because it has a higher throughput than the latter. Stress-strain relationships and even better “incremental elastic (Young’s) modulus” (which require additional recording of wall thickness) only come into play to discriminate between structural and material property contributions to changes in arterial stiffness. Part of the isolated thoracic aorta was studied in the absence and part in the continuous presence of 10 μM indomethacin to inhibit the production of prostaglandins that can act as endothelium-derived vasoactive factors in this vessel. The vessels were tested for their contractile response to 40mM K^+ , concentration-response curves of phenylephrine (0.01 to 100 μM) and endothelin-1 (1 – 256nM) followed by acetylcholine (ACh) (0.01 to 100 μM), PAPA/NO (0.01-10 μM) or Bay60-2770 (0.01-10 μM)-induced relaxations. The wall tension of the vessel segment was continuously recorded with LabChart Pro (ADInstruments, Oxford, UK).

Measurement of superoxide formation: DHE staining

Superoxide was measured in femoral arteries using the fluorescence dye dihydroethidium (DHE) (Thermo Scientific Technology, The Netherlands). Frozen femoral arteries cryosections were fixated in 4% paraformaldehyde (PFA) in PBS and then incubated with 2 μM DHE for 30 minutes at 37°C. After three washing steps with PBS, slices were incubated with DAPI (Sigma-Aldrich, The Netherlands) 2 $\mu\text{g}/\text{ml}$ for 10 min. Sections were washed in PBS and then mounted using a Dako Fluorescence Mounting Medium (S3023, Agilent Technologies). Immunofluorescent signals were viewed using a Leica DMI3000 B fluorescence microscope. Some arteries were pretreated with 500 μM L-NAME for 30 min at 37°C before performing the DHE staining.

RNA Extraction, cDNA Synthesis and Quantitative Real-Time PCR

Thoracic aortae, femoral arteries and saphenous arteries were isolated from mice and immediately submerged in RNAlater solution (Thermo Fisher Scientific). RNA was extracted using RNeasy® Micro Kit (Qiagen) according to the manufacturer's protocol. cDNA was synthesized from 1 µg total RNA in 20 µl reactions using High Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific). After synthesis, the cDNA was stored at -20°C. RT-qPCR was performed on CFX96™ Real-Time PCR Detection System (Bio-Rad). All reactions were performed in triplicates in a total volume of 20 µl each using TaqMan® Universal PCR Master Mix (Applied Biosystems- Life Technologies) according to manufacturer's instructions. 3 µl cDNA was used as a template and pre-designed TaqMan® primers of β -actin and Nox5 were used. The specific assay IDs for the primers used are shown in Supplementary Table 5.S2. The standard PCR conditions were as follows: 10 min at 95°C, followed by 15 s at 95°C and 1 min at 60 °C, 59 repeats. The amount of mRNA was normalized to the measured expression of β -actin mRNA.

Statistical analysis

All human and animal data are expressed as the mean \pm SEM for numeric variables and as the number (percentage) for categorical variables. Comparisons of continuous variables between the two mice groups were performed by unpaired two-tailed Student's t-test and among the three human groups by one-way analysis of variance (ANOVA) followed by Tukey's multiple comparisons test (post-hoc test). Comparisons of categorical variables among the human groups were assessed by a χ^2 (chi-square). Comparisons between the two mice groups, in telemetry data were done by two-way repeated measures ANOVA and in myograph by ordinary two-way ANOVA, followed by Sidak's multiple comparisons test. For the subgroup analysis of the NOX5 levels in human subjects, a frequency analysis was carried out, where the bin width is calculated using Sturges' rule [115]. To assess the modality of the data, the output frequencies were fitted with a single Gaussian and sum of two Gaussian distributions, and a two-tailed F-test with a null hypothesis as Gaussian and alternative

hypothesis as a sum of two-Gaussians was performed. Additionally, the adjusted r-squared values were compared to select the best fitting distribution for the sample. Given the bimodal nature of the sample, the area under each Gaussian distribution was calculated using the formula “Amplitude*SD/0.3989” and subsequently, the proportion of NOX5 mechanotype was reported as the ratio between the two distributions. Data were analyzed using GraphPad Prism Version 8.2 (GraphPad Software Inc., San Diego, CA). A p-value of less than 0.05 was considered to indicate statistical significance after multiple testing correction.

5.4 References

- [1] M. H. Olsen et al. "A call to action and a lifecourse strategy to address the global burden of raised blood pressure on current and future generations: the Lancet Commission on hypertension". In: *The Lancet* 388 (Nov. 2016), pp. 2665–2712. DOI: 10.1016/s0140-6736(16)31134-5.
- [2] S. Oparil et al. "Hypertension". In: *Nature Reviews Disease Primers* 4 (Mar. 2018), p. 18014. DOI: 10.1038/nrdp.2018.14.
- [3] L. G. Ogden et al. "Long-Term Absolute Benefit of Lowering Blood Pressure in Hypertensive Patients According to the JNC VI Risk Stratification". In: *Hypertension* 35 (Feb. 2000), pp. 539–543. DOI: 10.1161/01.hyp.35.2.539.
- [4] R. J. Gryglewski, R. M. J. Palmer, and S. Moncada. "Superoxide anion is involved in the breakdown of endothelium-derived vascular relaxing factor". In: *Nature* 320 (Apr. 1986), pp. 454–456. DOI: 10.1038/320454a0.
- [5] A. T. Kraja et al. "New Blood Pressure–Associated Loci Identified in Meta-Analyses of 475000 Individuals". In: *Circulation: Cardiovascular Genetics* 10 (Oct. 2017). DOI: 10.1161/circgenetics.117.001778.
- [6] A. Yogi et al. "Renal Redox-Sensitive Signaling, but Not Blood Pressure, Is Attenuated by Nox1 Knockout in Angiotensin II-Dependent Chronic Hypertension". In: *Hypertension* 51 (Feb. 2008), pp. 500–506. DOI: 10.1161/hypertensionaha.107.103192.
- [7] C. E. Murdoch et al. "Role of endothelial Nox2 NADPH oxidase in angiotensin II-induced hypertension and vasomotor dysfunction". In: *Basic Research in Cardiology* 106 (Apr. 2011), pp. 527–538. DOI: 10.1007/s00395-011-0179-7.
- [8] C. M. Sag et al. "Distinct Regulatory Effects of Myeloid Cell and Endothelial Cell NADPH Oxidase 2 on Blood Pressure". In: *Circulation* 135 (May 2017), pp. 2163–2177. DOI: 10.1161/circulationaha.116.023877.

-
- [9] J. Wu et al. "Immune activation caused by vascular oxidation promotes fibrosis and hypertension". In: *Journal of Clinical Investigation* 126 (Nov. 2015), pp. 50–67. DOI: 10.1172/jci80761.
 - [10] S. Dikalov et al. "Distinct roles of Nox1 and Nox4 in basal and angiotensin II-stimulated superoxide and hydrogen peroxide production". In: *Free Radical Biology and Medicine* 45 (Nov. 2008), pp. 1340–1351. DOI: 10.1016/j.freeradbiomed.2008.08.013.
 - [11] G. Gavazzi et al. "Decreased blood pressure in NOX1-deficient mice". In: *FEBS Letters* 580 (Dec. 2005), pp. 497–504. DOI: 10.1016/j.febslet.2005.12.049.
 - [12] C. Kleinschnitz et al. "Post-Stroke Inhibition of Induced NADPH Oxidase Type 4 Prevents Oxidative Stress and Neurodegeneration". In: *PLoS Biology* 8 (Sept. 2010). Ed. by M. McLeod, e1000479. DOI: 10.1371/journal.pbio.1000479.
 - [13] R. Ray et al. "Endothelial Nox4 NADPH Oxidase Enhances Vasodilatation and Reduces Blood Pressure In Vivo". In: *Arteriosclerosis, Thrombosis, and Vascular Biology* 31 (June 2011), pp. 1368–1376. DOI: 10.1161/atvbaha.110.219238.
 - [14] C. Holterman, N. Read, and C. J. Kennedy. "Nox and renal disease". In: *Clinical Science* 128 (Dec. 2014), pp. 465–481. DOI: 10.1042/cs20140361.
 - [15] J. C. Jha et al. "NADPH Oxidase Nox5 Accelerates Renal Injury in Diabetic Nephropathy". In: *Diabetes* 66 (July 2017), pp. 2691–2703. DOI: 10.2337/db16-1585.
 - [16] J. C. Jha et al. "Endothelial or vascular smooth muscle cell-specific expression of human NOX5 exacerbates renal inflammation, fibrosis and albuminuria in the Akita mouse". In: *Diabetologia* 62 (June 2019), pp. 1712–1726. DOI: 10.1007/s00125-019-4924-z.
 - [17] A. C. Montezano et al. "NADPH Oxidase 5 Is a Pro-Contractile Nox Isoform and a Point of Cross-Talk for Calcium and Redox Signaling-Implications in Vascular Function". In: *Journal of the*

- American Heart Association* 7 (June 2018). DOI: 10.1161/jaha.118.009388.
- [18] A.-L. Barabási, N. Gulbahce, and J. Loscalzo. “Network medicine: a network-based approach to human disease”. In: *Nature Reviews. Genetics* 12 (Jan. 2011), 56–68. DOI: 10.1038/nrg2918.
 - [19] N. Alcaraz et al. “De novo pathway-based biomarker identification”. In: *Nucleic Acids Research* 45 (July 2017), e151–e151. DOI: 10.1093/nar/gkx642.
 - [20] R. Batra et al. “On the performance of de novo pathway enrichment”. In: *npj Systems Biology and Applications* 3 (Mar. 2017). DOI: 10.1038/s41540-017-0007-2.
 - [21] J. Menche et al. “Uncovering disease-disease relationships through the incomplete interactome”. In: *Science* 347 (Feb. 2015). DOI: 10.1126/science.1257601.
 - [22] M. Kotlyar et al. “IID 2018 update: context-specific physical protein–protein interactions in human, model organisms and domesticated species”. In: *Nucleic Acids Research* 47 (Nov. 2018), pp. D581–D589. DOI: 10.1093/nar/gky1037.
 - [23] E. L. Huttlín et al. “Architecture of the human interactome defines protein communities and disease networks”. In: *Nature* 545 (May 2017), pp. 505–509. DOI: 10.1038/nature22366.
 - [24] S. Choobdar et al. “Assessment of network module identification across complex diseases”. In: *Nature Methods* 16 (Aug. 2019), pp. 843–852. DOI: 10.1038/s41592-019-0509-5.
 - [25] A. Arenas, A. Fernández, and S. Gómez. “Analysis of the structure of complex networks at different resolution levels”. In: *New Journal of Physics* 10 (May 2008), p. 053039. DOI: 10.1088/1367-2630/10/5/053039.
 - [26] P. Jiang and M. Singh. “SPICi: a fast clustering algorithm for large biological networks”. In: *Bioinformatics* 26 (Feb. 2010), pp. 1105–1111. DOI: 10.1093/bioinformatics/btq078.

-
- [27] F. Dignat-George and C. M. Boulanger. “The Many Faces of Endothelial Microparticles”. In: *Arteriosclerosis, Thrombosis, and Vascular Biology* 31 (Jan. 2011), pp. 27–33. DOI: 10.1161/atvbaha.110.218123.
 - [28] J. Aguirre-Plans et al. “Proximal Pathway Enrichment Analysis for Targeting Comorbid Diseases via Network Endopharmacology”. In: *Pharmaceuticals* 11 (June 2018), p. 61. DOI: 10.3390/ph11030061.
 - [29] A. Mazein et al. “Systems medicine disease maps: community-driven comprehensive representation of disease mechanisms”. In: *npj Systems Biology and Applications* 4 (June 2018). DOI: 10.1038/s41540-018-0059-y.
 - [30] N. S. Burns and P. W. Miller. “Learning What We Didn’t Know — The SPRINT Data Analysis Challenge”. In: *New England Journal of Medicine* 376 (June 2017), pp. 2205–2207. DOI: 10.1056/nejmp1705323.
 - [31] C. E. Holterman et al. “Nephropathy and Elevated BP in Mice with Podocyte-Specific NADPH Oxidase 5 Expression”. In: *Journal of the American Society of Nephrology* 25 (Nov. 2013), pp. 784–797. DOI: 10.1681/asn.2013040371.
 - [32] J. Sáenz-Medina et al. “Nox1-derived oxidative stress as a common pathogenic link between obesity and hyperoxaluria-related kidney injury”. In: *Urolithiasis* 48 (Nov. 2019), pp. 481–492. DOI: 10.1007/s00240-019-01170-w.
 - [33] M. Muñoz et al. “Differential contribution of Nox1, Nox2 and Nox4 to kidney vascular oxidative stress and endothelial dysfunction in obesity”. In: *Redox Biology* 28 (Jan. 2020), p. 101330. DOI: 10.1016/j.redox.2019.101330.
 - [34] U. Förstermann and T. Münzel. “Endothelial Nitric Oxide Synthase in Vascular Disease”. In: *Circulation* 113 (Apr. 2006), pp. 1708–1714. DOI: 10.1161/circulationaha.105.602532.

- [35] F. Perticone et al. "Asymmetric Dimethylarginine, L-Arginine, and Endothelial Dysfunction in Essential Hypertension". In: *Journal of the American College of Cardiology* 46 (Aug. 2005), pp. 518–523. DOI: 10.1016/j.jacc.2005.04.040.
- [36] A. Sonmez et al. "Plasma Apelin and ADMA Levels in Patients with Essential Hypertension". In: *Clinical and Experimental Hypertension* 32 (May 2010), pp. 179–183. DOI: 10.3109/10641960903254505.
- [37] A. I. Casas et al. "Calcium-dependent blood-brain barrier breakdown by NOX5 limits postreperfusion benefit in stroke". In: *Journal of Clinical Investigation* 129 (Mar. 2019), pp. 1772–1778. DOI: 10.1172/jci124283.
- [38] A. Bubikat et al. "Local Atrial Natriuretic Peptide Signaling Prevents Hypertensive Cardiac Hypertrophy in Endothelial Nitric-oxide Synthase-deficient Mice". In: *Journal of Biological Chemistry* 280 (June 2005), pp. 21594–21599. DOI: 10.1074/jbc.m501103200.
- [39] A. Gödecke et al. "Coronary Hemodynamics in Endothelial NO Synthase Knockout Mice". In: *Circulation Research* 82 (Feb. 1998), pp. 186–194. DOI: 10.1161/01.res.82.2.186.
- [40] R. B. Devereux et al. "Left ventricular hypertrophy in hypertension. Prevalence and relationship to pathophysiologic variables." In: *Hypertension* 9 (Feb. 1987). DOI: 10.1161/01.hyp.9.2.pt_2.ii53.
- [41] J. B. Park and E. L. Schiffrin. "Small artery remodeling is the most prevalent (earliest?) form of target organ damage in mild essential hypertension". In: *Journal of Hypertension* 19 (May 2001), pp. 921–930. DOI: 10.1097/00004872-200105000-00013.
- [42] C. Cuspidi et al. "Prevalence of left-ventricular hypertrophy in hypertension: an updated review of echocardiographic studies". In: *Journal of Human Hypertension* 26 (Nov. 2011), pp. 343–349. DOI: 10.1038/jhh.2011.104.

-
- [43] Y. Bézie et al. "Fibronectin Expression and Aortic Wall Elastic Modulus in Spontaneously Hypertensive Rats". In: *Arteriosclerosis, Thrombosis, and Vascular Biology* 18 (July 1998), pp. 1027–1034. DOI: 10.1161/01.atv.18.7.1027.
 - [44] D. Hayoz et al. "Conduit artery compliance and distensibility are not necessarily reduced in hypertension." In: *Hypertension* 20 (July 1992), pp. 1–6. DOI: 10.1161/01.hyp.20.1.1.
 - [45] P. Lacolley et al. "Influence of graded changes in vasomotor tone on the carotid arterial mechanics in live spontaneously hypertensive rats". In: *British Journal of Pharmacology* 115 (Aug. 1995), pp. 1235–1244. DOI: 10.1111/j.1476-5381.1995.tb15031.x.
 - [46] H. D. Intengan and E. L. Schiffrin. "Structure and Mechanical Properties of Resistance Arteries in Hypertension". In: *Hypertension* 36 (Sept. 2000), pp. 312–318. DOI: 10.1161/01.hyp.36.3.312.
 - [47] C. Bussy et al. "Intrinsic Stiffness of the Carotid Arterial Wall Material in Essential Hypertensives". In: *Hypertension* 35 (May 2000), pp. 1049–1054. DOI: 10.1161/01.hyp.35.5.1049.
 - [48] S. Laurent et al. "Isobaric compliance of the radial artery is increased in patients with essential hypertension". In: *Journal of Hypertension* 11 (Jan. 1993), pp. 89–98. DOI: 10.1097/00004872-199301000-00013.
 - [49] S. Laurent et al. "Elastic modulus of the radial artery wall material is not increased in patients with essential hypertension." In: *Arteriosclerosis and Thrombosis: A Journal of Vascular Biology* 14 (July 1994), pp. 1223–1231. DOI: 10.1161/01.atv.14.7.1223.
 - [50] S. Laurent. "Arterial Wall Hypertrophy and Stiffness in Essential Hypertensive Patients". In: *Hypertension* 26 (Aug. 1995), pp. 355–362. DOI: 10.1161/01.hyp.26.2.355.
 - [51] R. F. Furchgott and J. V. Zawadzki. "The obligatory role of endothelial cells in the relaxation of arterial smooth muscle by acetylcholine". In: *Nature* 288 (Nov. 1980), pp. 373–376. DOI: 10.1038/288373a0.

- [52] R. Chennupati et al. "Endothelium-dependent hyperpolarization-related relaxations diminish with age in murine saphenous arteries of both sexes". In: *British Journal of Pharmacology* 169 (July 2013), pp. 1486–1499. DOI: 10.1111/bph.12175.
- [53] R. Chennupati et al. "Endothelial Arginine Resynthesis Contributes to the Maintenance of Vasomotor Function in Male Diabetic Mice". In: *PLoS ONE* 9 (July 2014). Ed. by J. Isenberg, e102264. DOI: 10.1371/journal.pone.0102264.
- [54] R. Chennupati et al. "Deletion of endothelial arginase 1 does not improve vasomotor function in diabetic mice". In: *Physiological Reports* 6 (June 2018), e13717. DOI: 10.14814/phy2.13717.
- [55] M. J. Ryan et al. "Endothelial Dysfunction and Blood Pressure Variability in Selected Inbred Mouse Strains". In: *Arteriosclerosis, Thrombosis, and Vascular Biology* 22 (Jan. 2002), pp. 42–48. DOI: 10.1161/hq0102.101098.
- [56] V. Gebhart et al. "Site and mechanism of uncoupling of nitric-oxide synthase: Uncoupling by monomerization and other misconceptions". In: *Nitric Oxide* 89 (Aug. 2019), pp. 14–21. DOI: 10.1016/j.niox.2019.04.007.
- [57] C. B. Mendes-Silverio et al. "Activation of Haem-Oxidized Soluble Guanylyl Cyclase with BAY 60-2770 in Human Platelets Lead to Overstimulation of the Cyclic GMP Signaling Pathway". In: *PLoS ONE* 7 (Nov. 2012). Ed. by H. H. H. W. Schmidt, e47223. DOI: 10.1371/journal.pone.0047223.
- [58] J.-P. Stasch et al. "Targeting the heme-oxidized nitric oxide receptor for selective vasodilatation of diseased blood vessels". In: *Journal of Clinical Investigation* 116 (Sept. 2006), pp. 2552–2561. DOI: 10.1172/jci28371.
- [59] J. A. Hrabie et al. "New nitric oxide-releasing zwitterions derived from polyamines". In: *The Journal of Organic Chemistry* 58 (Mar. 1993), pp. 1472–1476. DOI: 10.1021/jo00058a030.

-
- [60] R. Kietadisorn, R. P. Juni, and A. L. Moens. "Tackling endothelial dysfunction by modulating NOS uncoupling: new insights into its pathogenesis and therapeutic possibilities". In: *American Journal of Physiology-Endocrinology and Metabolism* 302 (Mar. 2012), E481–E495. DOI: 10.1152/ajpendo.00540.2011.
 - [61] S. Altenhöfer et al. "The NOX toolbox: validating the role of NADPH oxidases in physiology and disease". In: *Cellular and Molecular Life Sciences* 69 (May 2012), pp. 2327–2343. DOI: 10.1007/s00018-012-1010-9.
 - [62] S. Altenhöfer et al. "Evolution of NADPH Oxidase Inhibitors: Selectivity and Mechanisms for Target Engagement". In: *Antioxidants & Redox Signaling* 23 (Aug. 2015), pp. 406–427. DOI: 10.1089/ars.2013.5814.
 - [63] F. Augsburger et al. "Pharmacological characterization of the seven human NOX isoforms and their inhibitors". In: *Redox Biology* 26 (Sept. 2019), p. 101272. DOI: 10.1016/j.redox.2019.101272.
 - [64] V. T.-V. Dao et al. "Isoform-selective NADPH oxidase inhibitor panel for pharmacological target validation". In: *Free Radical Biology and Medicine* 148 (Feb. 2020), pp. 60–69. DOI: 10.1016/j.freeradbiomed.2019.12.038.
 - [65] S. Frangos and J. R. Buscombe. "Why should we be concerned about a "g"?" In: *European Journal of Nuclear Medicine and Molecular Imaging* 46 (Nov. 2018), pp. 519–519. DOI: 10.1007/s00259-018-4204-z.
 - [66] C. Hörnsten et al. "High blood pressure as a risk factor for incident stroke among very old people". In: *Journal of Hypertension* 34 (Oct. 2016), pp. 2059–2065. DOI: 10.1097/hjh.0000000000001048.
 - [67] T. J. Guzik et al. "Calcium-Dependent NOX5 Nicotinamide Adenine Dinucleotide Phosphate Oxidase Contributes to Vascular Oxidative Stress in Human Coronary Artery Disease". In: *Journal of the American College of Cardiology* 52 (Nov. 2008), pp. 1803–1809. DOI: 10.1016/j.jacc.2008.07.063.

- [68] H. Li et al. "Associations of NADPH oxidase-related genes with blood pressure changes and incident hypertension: The Gen-Salt Study". In: *Journal of Human Hypertension* 32 (Feb. 2018), pp. 287–293. DOI: 10.1038/s41371-018-0041-6.
- [69] M. H. Elbatreek et al. "Reactive Oxygen Comes of Age: Mechanism-Based Therapy of Diabetic End-Organ Damage". In: *Trends in Endocrinology & Metabolism* 30 (May 2019), pp. 312–327. DOI: 10.1016/j.tem.2019.02.006.
- [70] M. Hermann, A. Flammer, and T. F. Lüscher. "Nitric oxide in hypertension". In: *Journal of Clinical Hypertension (Greenwich, Conn.)* 8 (Dec. 2006), 17–29. DOI: 10.1111/j.1524-6175.2006.06032.x.
- [71] T. Helbing. "Role of microparticles in endothelial dysfunction and arterial hypertension". In: *World Journal of Cardiology* 6 (2014), p. 1135. DOI: 10.4330/wjc.v6.i11.1135.
- [72] m. p. t. E. d. j. u. v. y. j. J. Shantsila E. In: ().
- [73] D. Burger et al. "Endothelial Microparticle-Derived Reactive Oxygen Species: Role in Endothelial Signaling and Vascular Function". In: *Oxidative Medicine and Cellular Longevity* 2016 (2016), pp. 1–10. DOI: 10.1155/2016/5047954.
- [74] A. C. Montezano et al. "Nicotinamide Adenine Dinucleotide Phosphate Reduced Oxidase 5 (Nox5) Regulation by Angiotensin II and Endothelin-1 Is Mediated via Calcium/Calmodulin-Dependent, Rac-1-Independent Pathways in Human Endothelial Cells". In: *Circulation Research* 106 (Apr. 2010), pp. 1363–1373. DOI: 10.1161/circresaha.109.216036.
- [75] P. Yu et al. "Unique role of NADPH oxidase 5 in oxidative stress in human renal proximal tubule cells". In: *Redox Biology* 2 (2014), pp. 570–579. DOI: 10.1016/j.redox.2014.01.020.
- [76] N. E. Hahn et al. "NOX5 Expression Is Increased in Intramyocardial Blood Vessels and Cardiomyocytes after Acute Myocardial Infarction in Humans". In: *The American Journal of Pathology* 180 (June 2012), pp. 2222–2229. DOI: 10.1016/j.ajpath.2012.02.018.

-
- [77] C. E. Holterman, J. F. Thibodeau, and C. R. Kennedy. "NADPH oxidase 5 and renal disease". In: *Current Opinion in Nephrology and Hypertension* 24 (Jan. 2015), pp. 81–87. DOI: 10.1097/mnh.0000000000000081.
 - [78] G. Bouabou et al. "Nox4 genetic inhibition in experimental hypertension and metabolic syndrome". In: *Archives of Cardiovascular Diseases* 111 (Jan. 2018), pp. 41–52. DOI: 10.1016/j.acvd.2017.03.011.
 - [79] K. Schröder et al. "Nox4 Is a Protective Reactive Oxygen Species Generating Vascular NADPH Oxidase". In: *Circulation Research* 110 (Apr. 2012), pp. 1217–1225. DOI: 10.1161/circresaha.112.267054.
 - [80] C. Veith et al. "NADPH Oxidase 4 is Not Involved in Hypoxia-Induced Pulmonary Hypertension". In: *Pulmonary Circulation* 6 (Sept. 2016), pp. 397–400. DOI: 10.1086/687756.
 - [81] R. P. Brandes, I. Takac, and K. Schröder. "No Superoxide—No Stress?" In: *Arteriosclerosis, Thrombosis, and Vascular Biology* 31 (June 2011), pp. 1255–1257. DOI: 10.1161/atvbaha.111.226894.
 - [82] H. Miura et al. "Role for Hydrogen Peroxide in Flow-Induced Dilation of Human Coronary Arterioles". In: *Circulation Research* 92 (Feb. 2003). DOI: 10.1161/01.res.0000054200.44505.ab.
 - [83] T. M. Leurgans et al. "Endothelin-1 shifts the mediator of bradykinin-induced relaxation from NO to H₂O₂ in resistance arteries from patients with cardiovascular disease". In: *British Journal of Pharmacology* 173 (Apr. 2016), pp. 1653–1664. DOI: 10.1111/bph.13467.
 - [84] H. Shimokawa. "Hydrogen peroxide as an endothelium-derived hyperpolarizing factor". In: *Pflügers Archiv - European Journal of Physiology* 459 (Feb. 2010), pp. 915–922. DOI: 10.1007/s00424-010-0790-8.

- [85] U. Landmesser et al. "Oxidation of tetrahydrobiopterin leads to uncoupling of endothelial cell nitric oxide synthase in hypertension". In: *Journal of Clinical Investigation* 111 (Apr. 2003), pp. 1201–1209. DOI: 10.1172/jci200314172.
- [86] C. Dumitrescu et al. "Myocardial ischemia results in tetrahydrobiopterin (BH4) oxidation with impaired endothelial function ameliorated by BH4". In: *Proceedings of the National Academy of Sciences* 104 (Sept. 2007), pp. 15081–15086. DOI: 10.1073/pnas.0702986104.
- [87] B. M. Mitchell, A. M. Dorrance, and R. C. Webb. "GTP cyclohydrolase 1 inhibition attenuates vasodilation and increases blood pressure in rats". In: *American Journal of Physiology-Heart and Circulatory Physiology* 285 (Nov. 2003), H2165–H2170. DOI: 10.1152/ajpheart.00253.2003.
- [88] M. Moro et al. "The formation of nitric oxide donors from peroxynitrite". In: *British Journal of Pharmacology* 116 (Oct. 1995), pp. 1999–2004. DOI: 10.1111/j.1476-5381.1995.tb16404.x.
- [89] M. Wu et al. "Involvement of nitric oxide and nitrosothiols in relaxation of pulmonary arteries to peroxynitrite". In: *American Journal of Physiology-Heart and Circulatory Physiology* 266 (May 1994), H2108–H2113. DOI: 10.1152/ajpheart.1994.266.5.h2108.
- [90] X. Pi et al. "NADPH Oxidase–Generated Reactive Oxygen Species Are Required for Stromal Cell–Derived Factor-1 α –Stimulated Angiogenesis". In: *Arteriosclerosis, Thrombosis, and Vascular Biology* 34 (Sept. 2014), pp. 2023–2032. DOI: 10.1161/atvbaha.114.303733.
- [91] B. Rodriguez-Iturbe et al. "Antioxidant-Rich Diet Relieves Hypertension and Reduces Renal Immune Infiltration in Spontaneously Hypertensive Rats". In: *Hypertension* 41 (Feb. 2003), pp. 341–346. DOI: 10.1161/01.hyp.0000052833.20759.64.

-
- [92] M. Akpaffiong and A. Taylor. "Antihypertensive and vasodilator actions of antioxidants in spontaneously hypertensive rats". In: *American Journal of Hypertension* 11 (Dec. 1998), pp. 1450–1460. DOI: 10.1016/s0895-7061(98)00183-6.
 - [93] S. Meng et al. "Oxidative Stress in Dahl Salt-Sensitive Hypertension". In: *Hypertension* 41 (June 2003), pp. 1346–1352. DOI: 10.1161/01.hyp.0000070028.99408.e8.
 - [94] M. Barton et al. "Anatomic heterogeneity of vascular aging: role of nitric oxide and endothelin". In: *Hypertension (Dallas, Tex.: 1979)* 30 (Oct. 1997), 817–824. DOI: 10.1161/01.hyp.30.4.817.
 - [95] R. L. Matz et al. "Vascular bed heterogeneity in age-related endothelial dysfunction with respect to NO and eicosanoids". In: *British Journal of Pharmacology* 131 (Sept. 2000), pp. 303–311. DOI: 10.1038/sj.bjp.0703568.
 - [96] L. W. Miksche, U. Miksche, and F. Gross. "Effect of Sodium Restriction on Renal Hypertension and on Renin Activity in the Rat". In: *Circulation Research* 27 (Dec. 1970), pp. 973–984. DOI: 10.1161/01.res.27.6.973.
 - [97] S. Wang et al. "Acute Inhibition of Guanosine Triphosphate Cyclohydrolase 1 Uncouples Endothelial Nitric Oxide Synthase and Elevates Blood Pressure". In: *Hypertension* 52 (Sept. 2008), pp. 484–490. DOI: 10.1161/hypertensionaha.108.112094.
 - [98] E. Podjarny et al. "Effect of chronic tetrahydrobiopterin supplementation on blood pressure and proteinuria in 5/6 nephrectomized rats". In: *Nephrology Dialysis Transplantation* 19 (July 2004), pp. 2223–2227. DOI: 10.1093/ndt/gfh383.
 - [99] J. P. Sundberg et al. "The mouse as a model for understanding chronic diseases of aging: the histopathologic basis of aging in inbred mice". In: *Pathobiology of Aging & Age-related Diseases* 1 (Jan. 2011), p. 7179. DOI: 10.3402/pba.v1i0.7179.
 - [100] M. Porkert et al. "Tetrahydrobiopterin: a novel antihypertensive therapy". In: *Journal of Human Hypertension* 22 (Mar. 2008), pp. 401–407. DOI: 10.1038/sj.jhh.1002329.

- [101] M. P. McRae. "High-dose folic acid supplementation effects on endothelial function and blood pressure in hypertensive patients: a meta-analysis of randomized controlled clinical trials". In: *Journal of Chiropractic Medicine* 8 (Mar. 2009), pp. 15–24. DOI: 10.1016/j.jcm.2008.09.001.
- [102] X. Kong et al. "Platelet Count Affects Efficacy of Folic Acid in Preventing First Stroke". In: *Journal of the American College of Cardiology* 71 (May 2018), pp. 2136–2146. DOI: 10.1016/j.jacc.2018.02.072.
- [103] Y. Huo et al. "Efficacy of folic acid therapy in primary prevention of stroke among adults with hypertension in China: the CSPPT randomized clinical trial". In: *JAMA* 313 (2015), pp. 1325–35. DOI: 10.1001/jama.2015.2274.
- [104] W.-W. Wang et al. "A Meta-Analysis of Folic Acid in Combination with Anti-Hypertension Drugs in Patients with Hypertension and Hyperhomocysteinemia". In: *Frontiers in Pharmacology* 8 (Aug. 2017). DOI: 10.3389/fphar.2017.00585.
- [105] C.-Y. Hsu et al. "Increased Circulating Endothelial Apoptotic Microparticle to Endothelial Progenitor Cell Ratio Is Associated with Subsequent Decline in Glomerular Filtration Rate in Hypertensive Patients". In: *PLoS ONE* 8 (July 2013). Ed. by G. P. Fadini, e68644. DOI: 10.1371/journal.pone.0068644.
- [106] P.-H. Huang et al. "Increased circulating CD31+/annexin V+ apoptotic microparticles and decreased circulating endothelial progenitor cell levels in hypertensive patients with microalbuminuria". In: *Journal of Hypertension* 28 (Aug. 2010), pp. 1655–1665. DOI: 10.1097/hjh.0b013e32833a4d0a.
- [107] F. Shang et al. "MicroRNA-92a Mediates Endothelial Dysfunction in CKD". In: *Journal of the American Society of Nephrology* 28 (July 2017), pp. 3251–3261. DOI: 10.1681/asn.2016111215.
- [108] Z. Chen et al. "Oxidative Stress Activates Endothelial Innate Immunity via Sterol Regulatory Element Binding Protein 2 (SREBP2) Transactivation of MicroRNA-92a". In: *Circulation* 131 (Mar. 2015), pp. 805–814. DOI: 10.1161/circulationaha.114.013675.

-
- [109] Y. Wang et al. "EPHB4 Protein Expression in Vascular Smooth Muscle Cells Regulates Their Contractility, and EPHB4 Deletion Leads to Hypotension in Mice". In: *Journal of Biological Chemistry* 290 (May 2015), pp. 14235–14244. DOI: 10.1074/jbc.M114.621615.
 - [110] P. Xu et al. "Endothelial Dysfunction and Elevated Blood Pressure in Mas Gene-Deleted Mice". In: *Hypertension* 51 (Feb. 2008), pp. 574–580. DOI: 10.1161/hypertensionaha.107.102764.
 - [111] J. K. Shirey-Rice et al. "Norepinephrine transporter variant A457P knock-in mice display key features of human postural orthostatic tachycardia syndrome". In: *Disease Models & Mechanisms* 6 (Apr. 2013), pp. 1001–1011. DOI: 10.1242/dmm.012203.
 - [112] R. Lazor et al. "Endothelin-1 Does Not Mediate the Endothelium-Dependent Hypoxic Contractions of Small Pulmonary Arteries in Rats". In: *Chest* 110 (July 1996), pp. 189–197. DOI: 10.1378/chest.110.1.189.
 - [113] M. Bloksgaard et al. "Imaging and modeling of acute pressure-induced changes of collagen and elastin microarchitectures in pig and human resistance arteries". In: *American Journal of Physiology-Heart and Circulatory Physiology* 313 (July 2017), H164–H178. DOI: 10.1152/ajpheart.00110.2017.
 - [114] F. Pourageaud and J. G. R. De Mey. "Structural properties of rat mesenteric small arteries after 4-wk exposure to elevated or reduced blood flow". In: *American Journal of Physiology-Heart and Circulatory Physiology* 273 (Oct. 1997), H1699–H1706. DOI: 10.1152/ajpheart.1997.273.4.h1699.
 - [115] D. W. Scott. "Sturges' rule". In: *Wiley Interdisciplinary Reviews: Computational Statistics* 1 (Nov. 2009), pp. 303–306. DOI: 10.1002/wics.35.

Supporting information

Table 5.S1: Baseline characteristics in healthy individuals and hypertensive patients of the second cohort. Values are mean \pm SD or number (%). BMI, body mass index; LDL, low density lipoprotein.

	Normotensive no proteinuria (n = 20)	Hypertension no proteinuria (n = 20)	Hypertension with proteinuria (n = 20)	p value
Age	63 \pm 7	69 \pm 12	67 \pm 10	0.168
Gender (male)	13 (65%)	16 (80%)	14 (70%)	0.563
BMI	24 \pm 3	26 \pm 4	26 \pm 4	0.413
Smoking	8 (40%)	9 (45%)	9 (45%)	0.71
Total Cholesterol	164 \pm 36	146 \pm 31	170 \pm 42	0.126
LDL-cholesterol	102 \pm 39	89 \pm 27	97 \pm 32	0.513
Uric acid	5.6 \pm 1.8	6.4 \pm 1.2	6.8 \pm 2.1	0.118

Table 5.S2: qPCR assays.

Gene	Assay ID
β -actin	Mm02619580_g1
Nox5	Hs00225846_m1

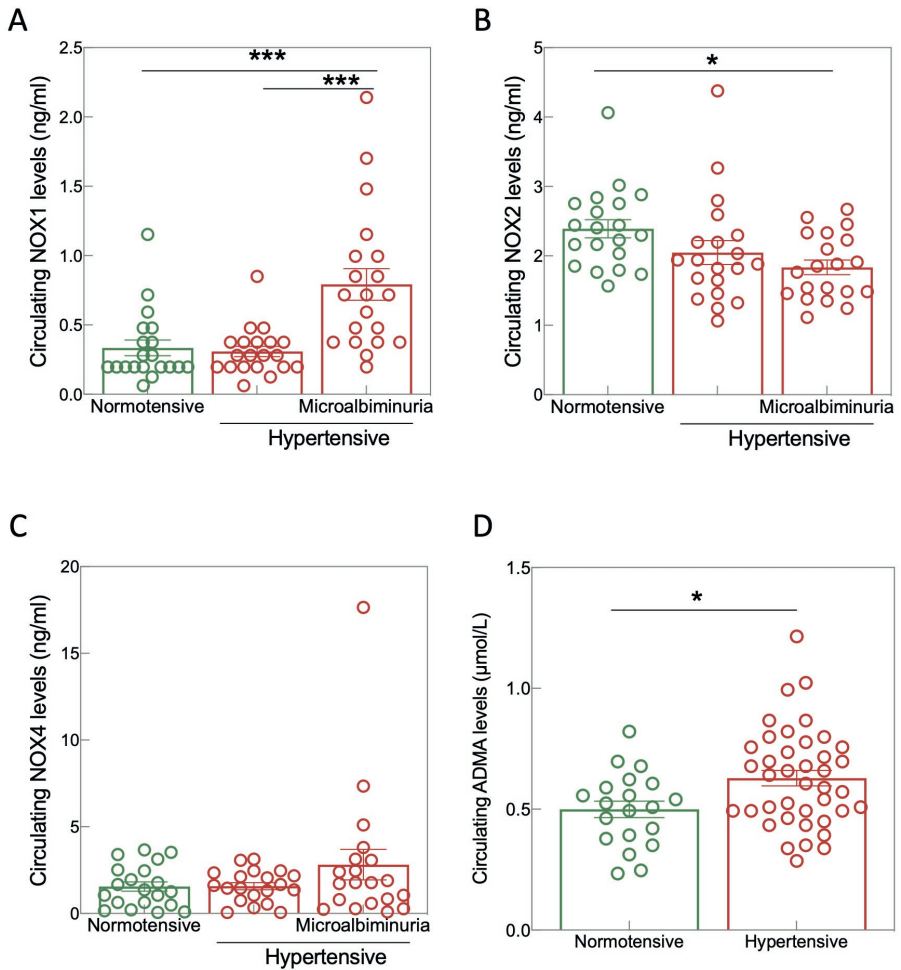


Figure 5.S1: Levels of NOX1, NOX2, NOX4, and ADMA in normotensive and hypertensive individuals. (A–C) There were no differences in NOX1, NOX2, or NOX4 levels between hypertensive patients with normoalbuminuria ($n = 20$) compared to normotensive individuals ($n = 20$). NOX1 levels were higher (A), but NOX2 were lower (B) in hypertensive patients with microalbuminuria ($n = 20$) compared to normotensive individuals. Comparison between groups was done by one-way ANOVA followed by Tukey's multiple comparisons test. (D) ADMA levels were significantly higher in hypertensive patients ($n = 40$) compared to healthy individuals ($n = 20$). Comparison between the two groups was done by two-tailed unpaired t test. All data are represented as mean \pm SEM, * $p < 0.05$, *** $p < 0.001$.

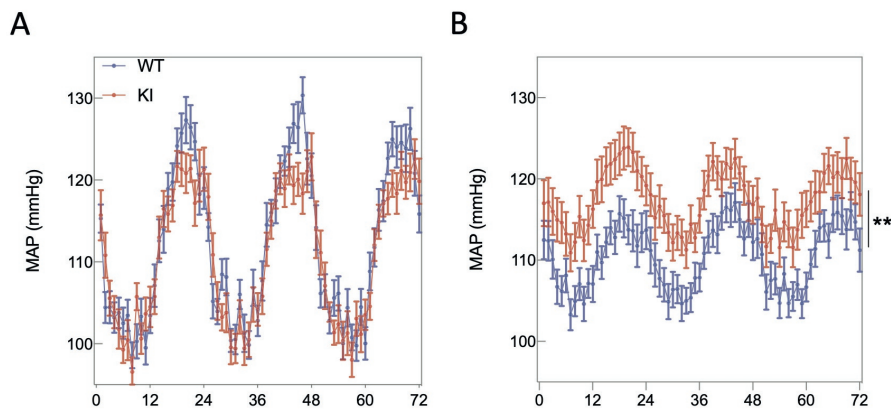


Figure 5.S2: MAP in young and aged WT and KI mice. (A) There was no significant difference in MAP between young WT (n = 19) and KI (n = 20). (B) Aged KI mice (n = 33) had higher MAP compared to WT (n = 31). Telemetry data were analyzed by two-way repeated measures ANOVA followed by Sidak's multiple comparisons test. All data are represented as mean \pm SEM of n individual animals **p < 0.01.

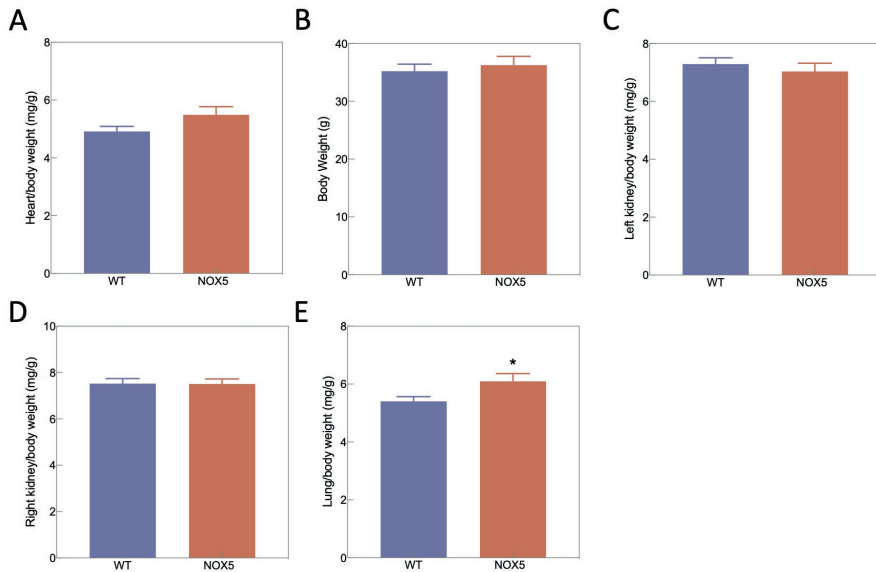


Figure 5.S3: Body and organ weights in aged WT and KI mice. (A–E) There was no difference in body, heart, and kidney weights between WT ($n = 24$) and KI mice ($n = 20$); however, lung/body weight ratio was higher in KI mice. Comparison between groups were done by two-tailed unpaired t test. All data are represented as mean \pm SEM of n individual animals * $p < 0.05$.

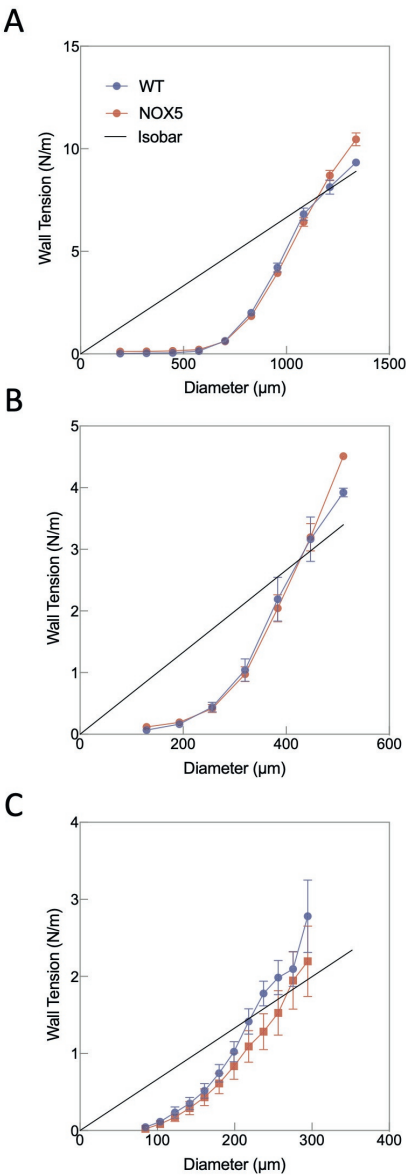


Figure 5.S4: Arterial stiffness in aged WT and KI mice. (A–C) The relation between resting wall tension and arterial lumen diameter did not differ between KI ($n = 9$) and WT ($n = 9$) mice in thoracic aortae (A), femoral arteries (B), and saphenous arteries (C). All data are represented as mean \pm SEM of n individual animals.

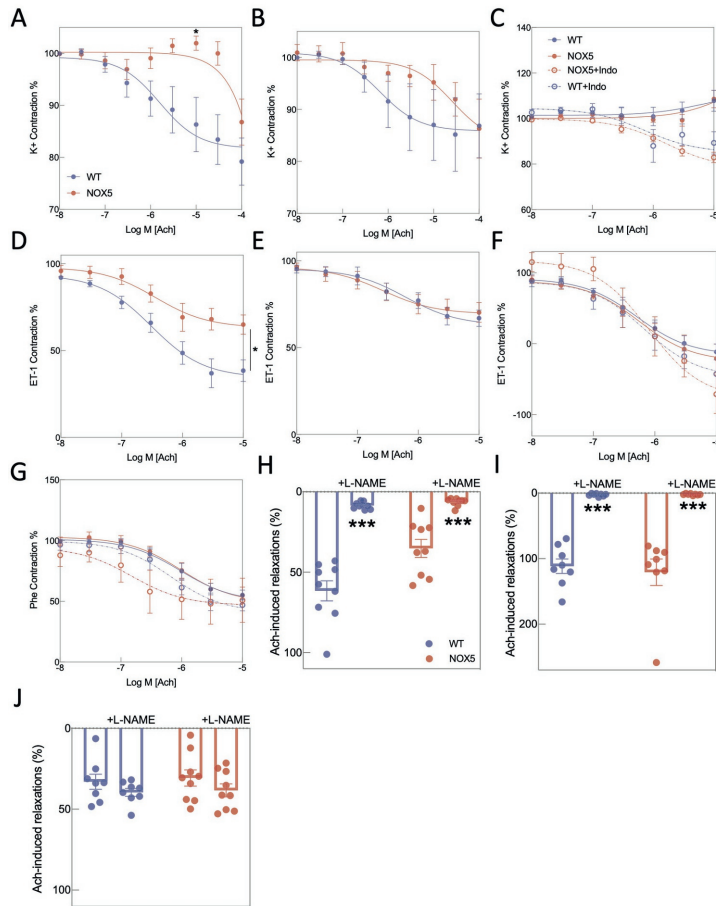


Figure 5.S5: Ach-induced relaxations in arteries of aged WT and KI mice. (A–C) Ach-induced relaxations were impaired in femoral arteries (A) of aged KI mice ($n = 9$) compared to WT ($n = 8–9$) but not in saphenous arteries (B) and thoracic aortae (with/without indomethacin) (C) precontracted with K^+ . (D–F) Ach-induced relaxations were impaired in femoral arteries (D) of aged KI mice ($n = 9$) compared to WT ($n = 8–9$) but not in saphenous arteries (E) and thoracic aortae with/without indomethacin (F) precontracted with endothelin-1. (G) There was no difference in Ach-induced relaxations in thoracic aortae (with/without indomethacin) precontracted with phenylephrine between WT ($n = 9$) and KI mice ($n = 9$). (H–J) Ach-induced relaxations in arteries made to contract with endothelin-1 were reversed by $100 \mu\text{M}$ L-NAME in femoral arteries (H) and thoracic aortae (I), but not saphenous arteries (J) of both aged KI mice ($n = 8–9$) and WT ($n = 8–9$). Myograph data were analyzed by two-way ANOVA followed by Sidak's multiple comparisons test. All data are represented as mean \pm SEM of n individual animals, * $p < 0.05$.

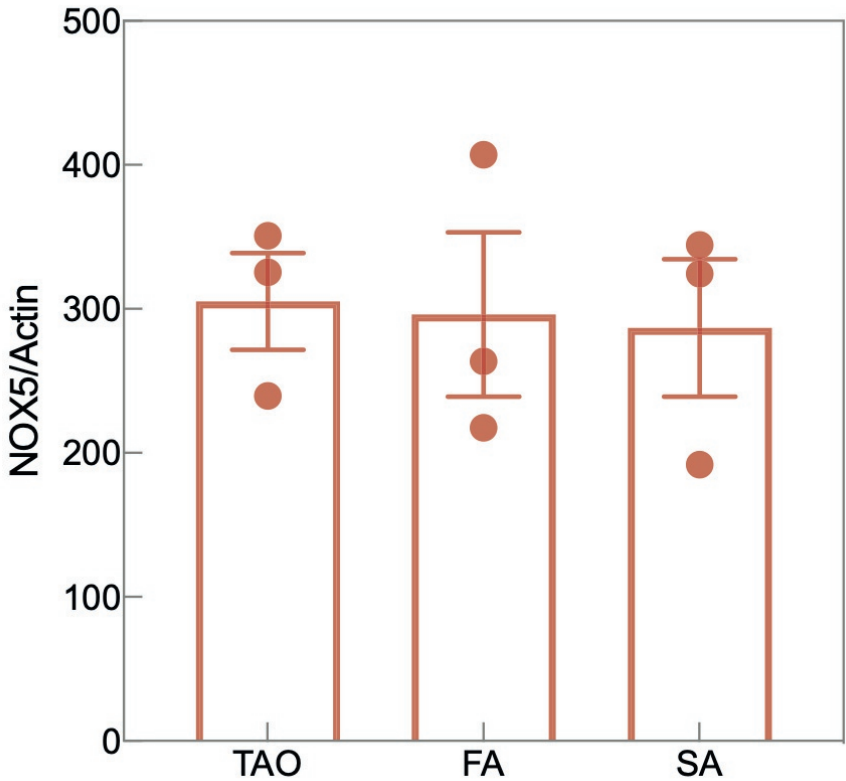


Figure 5.S6: qPCR of NOX5 in TAO, FAs, and SAs of aged KI mice. There was no difference in NOX5 gene expression between the 3 vessel types (n = 3, each in duplicates). Comparison were done by one-way ANOVA. All data are represented as mean \pm SEM of n individual animals. All raw data are included in the S1 Data file. FA, femoral artery; SA, saphenous artery; TAO, thoracic aortae.

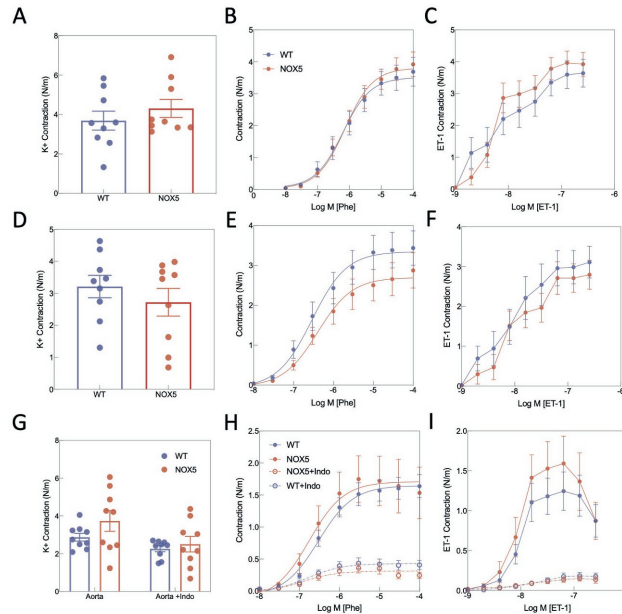


Figure 5.S7: Contractile responses in arteries of aged WT and KI mice. There was no difference in contractile responses to K⁺, phenylephrine, and endothelin-1 in femoral arteries (A–C), saphenous arteries (D–F), and thoracic aortae (with/without indomethacin) (G–I) between WT (n = 8–9) and KI mice (n = 9). Comparison between 2 groups in contractile responses to K⁺ was done by two-tailed t test. Other myograph data were analyzed by two-way ANOVA followed by Sidak's multiple comparisons test. All data are represented as mean ± SEM of n individual animals.

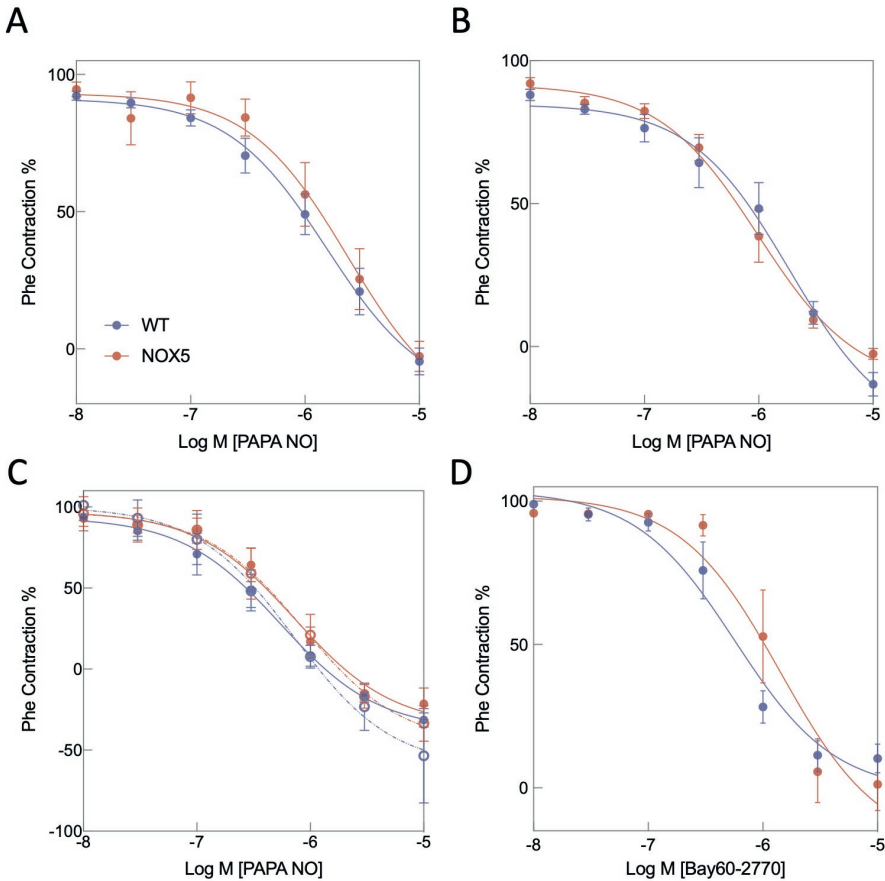


Figure 5.S8: Endothelium-independent relaxations in arteries of aged WT and KI mice. (A–C) Relaxations induced by the NO donor, PAPA NO (0.01–10 μ M), in femoral arteries (A), saphenous arteries (B), and thoracic aortae (with/without indomethacin) (C) did not differ between WT (n = 8–9) and KI mice (n = 9). (D) Relaxations induced by the apo-sGC activator, Bay60-2770 (0.01–10 μ M), in femoral arteries did not differ between WT (n = 4) and KI mice (n = 4). Myograph data were analyzed by two-way ANOVA followed by Sidak’s multiple comparisons test. All data are represented as mean \pm SEM of n individual animals.

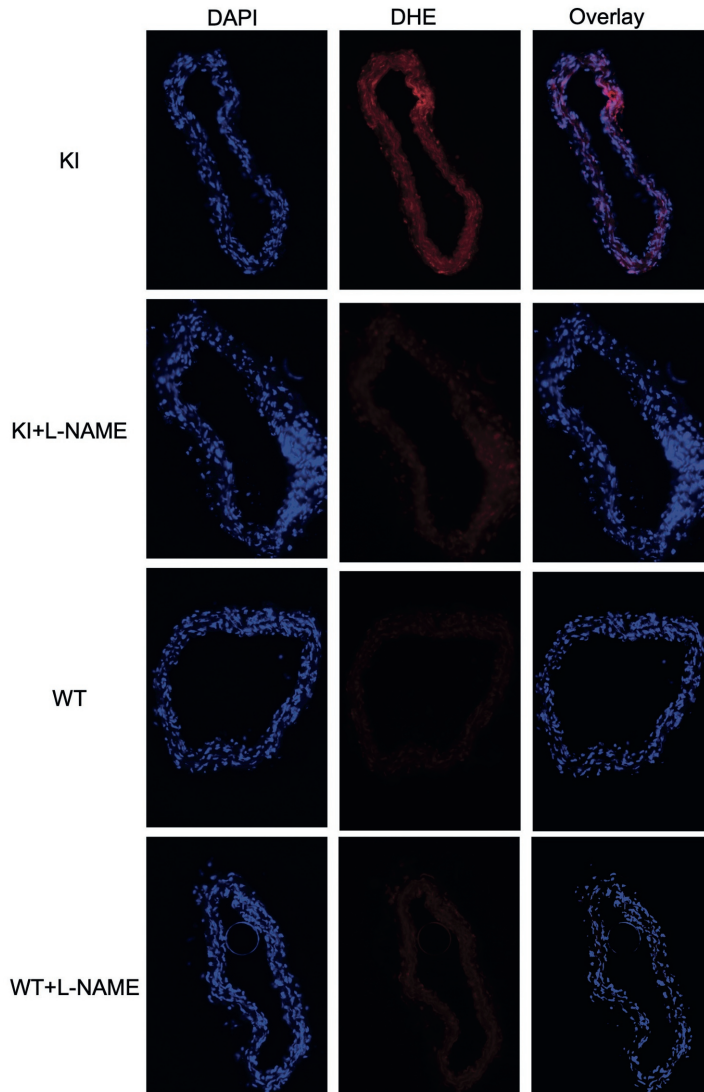


Figure 5.S9: Representative images of DHE staining of femoral arteries of aged WT and KI mice.

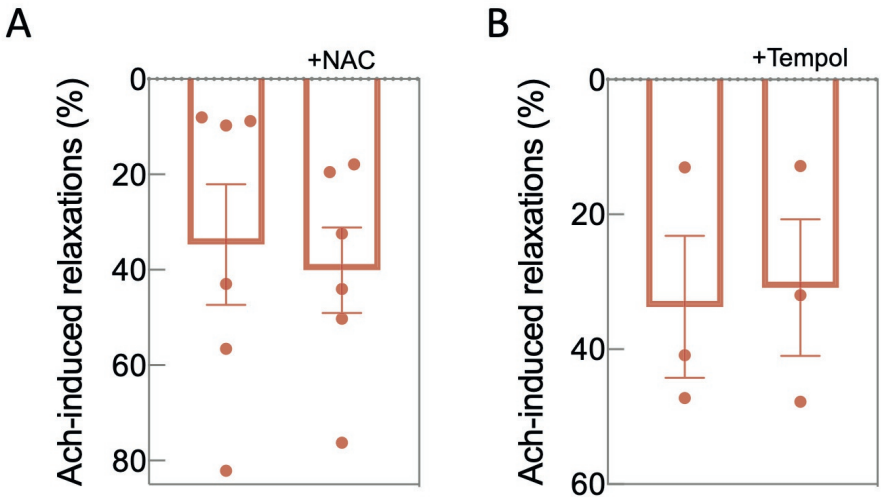


Figure 5.S10: Ach-induced relaxations in femoral arteries of aged KI mice treated with antioxidants. In segments of femoral artery ($n = 3-6$) made to contract with $10 \mu\text{M}$ phenylephrine, relaxing effects of Ach ($10 \mu\text{M}$) were not reversed by $10 \mu\text{M}$ N-acetylcysteine (NAC) (A) or $100 \mu\text{M}$ tempol (B). Comparison between groups were done by two-tailed unpaired t test. All data are represented as mean \pm SEM of n individual animals.

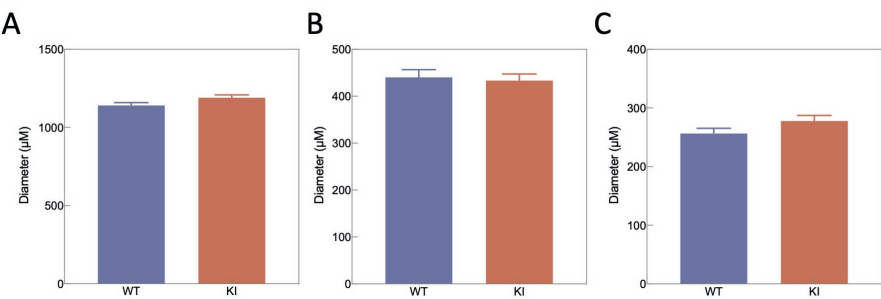


Figure 5.S11: Echocardiography in aged WT and KI mice. There were no differences in all parameters between WT ($n = 28$) and KI mice ($n = 29$). Comparison between groups were done by two-tailed unpaired t test. All data are represented as mean \pm SEM of n individual animals.

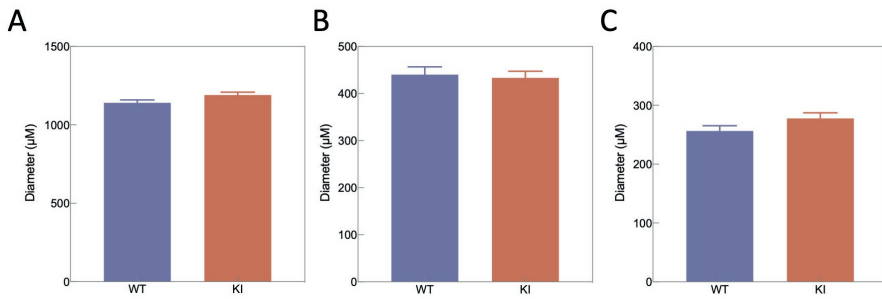


Figure 5.S12: Arteries diameter in aged WT and KI mice. There were no differences in diameter of thoracic aortae (A), femoral arteries (B), and saphenous arteries (with/without indomethacin) between WT ($n = 9$) and KI mice ($n = 9$). Comparison between groups were done by two-tailed unpaired t test. All data are represented as mean \pm SEM of n individual animals.

6

Towards FAIR protocols and workflows: The OpenPREDICT case study

6

Adapted from: Remzi Celebi, Joao Rebelo Moreira, **Ahmed A. Hassan**, Sandeep Ayyar, Lars Ridder, Tobias Kuhn and Michel Dumontier “Towards FAIR protocols and workflows: the OpenPREDICT use case”. In: *PeerJ Computer Science*.

Abstract

It is essential for the advancement of science that researchers share, reuse and reproduce each other's workflows and protocols. The FAIR principles are a set of guidelines that aim to maximize the value and usefulness of research data and emphasize the importance of making digital objects findable and reusable by others. The question of how to apply these principles not just to data but also to the workflows and protocols that consume and produce them is still under debate and poses a number of challenges. In this paper, we describe a two-fold approach of simultaneously applying the FAIR principles to scientific workflows as well as the involved data. We apply and evaluate our approach on the case of the PREDICT workflow, a highly cited drug repurposing workflow. This includes FAIRification of the involved datasets, as well as applying semantic technologies to represent and store data about the detailed versions of the general protocol, the concrete workflow instructions, and their execution traces. We propose a semantic model to address these specific requirements and evaluate it based on a set of competency questions. This semantic model consists of classes and relations from a number of existing ontologies, including Workflow4ever, PROV, EDAM and BPMN. Our evaluation shows the high degree to which our FAIRified OpenPREDICT workflow now adheres to the FAIR principles and the practicality and usefulness of our model to answer relevant competency questions.

6.1 Introduction

Reproducible results are one of the main goals of science. A recent survey, however, showed that more than 70% of researchers have been unsuccessful in reproducing another research experiment and more than 50% failed to reproduce their own research studies [1]. The rate of non-reproducibility for pharmacological studies is particularly worrying. Together with their high costs and their high rate of failure (around 90%), this highlights the need for new approaches in drug discovery [2]. For these reasons, we chose pharmacology as the field to apply and test the approach we will introduce below. Specifically, we will be looking into drug repositioning, where small molecules approved for one indication are repurposed for a new indication. Drug repositioning is gaining recognition as a safe, effective and lower-cost approach to uncover new drug uses [3, 4]. The availability of public data, both in the form of literature curated knowledge and ‘omics data has created exciting opportunities for computational drug repositioning. For instance, gene expression data in repositories such as the Gene Expression Omnibus (GEO) enable the analysis of correlations between drug and gene expression — termed the Connectivity Map approach — to find chemicals that may counter cellular disorders [5], including Alzheimer’s, and small cell lung cancer [6, 7]. More sophisticated approaches use network analysis and machine learning to efficiently combine drug and disease data [8–12].

The ability to reproduce original research results is contingent on the availability of the original data, methods and results. The FAIR principles [13], describe a set of requirements for data management and stewardship to make research data Findable, Accessible, Interoperable and Reusable. Ongoing efforts on FAIR cover data policies, data management plans, identifier mechanisms, standards, and data repositories [14]. Highly diverse communities, from the biomedical sciences to the social sciences and humanities, are now working towards defining standards for publication and sharing of data. In anticipation, new methods and infrastructure are needed to facilitate the generation of FAIR data and workflows.

Here, we describe a methodology to publish scientific workflows

as FAIR data. We are using the term workflow here to include computational steps implemented in software but also manual steps, such as manual data cleaning steps or wet-lab activities. We evaluate our method by applying it to the PREDICT drug repositioning workflow. Based on this example, we will try to answer our research question of how we can use existing vocabularies and techniques to make scientific workflows more open and FAIR, with a particular focus on the interoperability aspect. The main contributions of this paper are (a) general guidelines to make scientific workflows open and FAIR, focusing on the interoperability aspect, (b) the OpenPREDICT use case, demonstrating the open and FAIR version of the PREDICT workflow, (c) new competency questions for previously unaddressed reproducibility requirements, and (d) evaluation results on the practicality and usefulness of our approach.

6.2 Background

6

Below, we refer to the most relevant background with respect to reproducibility, workflow systems and applying FAIR to workflows.

6.2.1 Scientific Workflows and Reproducibility

According to the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [15], a workflow is a “plan that defines role(s), task(s), and a specific structure for tasks to be executed, usually supporting the work of an organization”, and a plan is a description of instructions with an explicit goal. A scientific workflow, therefore, is such a plan that implements scientific methods to work towards the general goal of scientific knowledge gathering and organization. Certain scientific workflows can be automated through workflow systems, which are software systems that enable the representation and execution of structured tasks.

The lack of relevant details in the published descriptions of scientific workflows [16] is an important factor contributing to the non-reproducibility rates of 64% in pharmacology [17, 18], 89% in cancer research [19], and in 66% psychology [20]. A recent analysis of over 1.4 million Jupyter notebooks (available in GitHub) found that only

24.11% of the notebooks could be executed without errors and only 4.03% produced the same results [21]. As a consequence, it has been reported that data scientists spend 19% of their time finding, understanding and accessing datasets, and 60% of their time cleaning and organizing these datasets to use in their studies [22]. Thus, only 20% of the time is left for data scientists to spend on their core activities, such as mining data, refining algorithms, building training sets and analyzing the results.

6.2.2 Workflow Systems

To tackle the workflow decay phenomenon [23], a number of recent initiatives are targeting the improvement of the reproducibility of computational workflows, for example the Common Workflow Language (CWL) [24], which has become the de facto standard for syntactic interoperability of computational workflow management systems. In order to improve semantic interoperability and connect workflows to real-world entities in a systematic way, additions of semantic models and methods have been proposed, for example the Workflow4ever project with its Research Objects method [25].

Provenance is an important aspect of workflows, which can be classified into prospective provenance, retrospective provenance, and workflow evolution provenance. Prospective provenance refers to the specifications or “recipes” that describe the workflow steps, their execution order, and their inputs and outputs [26]. Retrospective provenance refers to the information about actual workflow executions that happened in the past, including the concrete activities that consumed inputs and produced outputs, as well as information about the execution environment [27]. Workflow evolution provenance refers to tracking the versions of workflows specifications and the respective data, as the workflow specification is changed and improved over time.

A number of models and methods have been developed to capture these different kinds of provenance. The PROV ontology [28] provides the vocabulary and model for provenance in general, which can be used in conjunction with top-level ontologies such as DOLCE and other general vocabularies such as Dublin Core and schema.org.

Several approaches have been proposed to apply PROV to workflows, such as the Open Provenance Model for Workflows (OPMW) [29], P-PLAN [30], and CWLProv [31]. Other notable examples include ProvBook and the Reproduce-me ontology [32, 33] for workflows in Jupyter notebooks, the ML-Schema ontology for machine learning workflows [34], the Publishing Workflow Ontology (PWO) for workflows in scientific publications [35], and the Business Process Modelling Notation (BPMN) to specify business processes [36]. Other approaches, such as SMART protocols [37] and protocols.io, target the description of laboratory protocols.

6.2.3 Applying FAIR to Workflows

The FAIR principles have received significant attention, but we currently lack overarching approaches to align them with scientific protocols and workflows in a broad sense. Making a workflow FAIR-compliant entails that general-purpose software can interpret it and understand its context. The application of FAIR in healthcare, for example, has shown that these principles can boost data-driven applications that require the integration of data coming from different sources, achieving “interoperability without the need to all speak exactly the same language” [38]. Recent initiatives have outlined how FAIR can be applied to software [39, 40], contributing towards the goal of applying FAIR not just to input and output data, but to the entire process in between, in order to solve the current problem that even human experts are often unable to reconstruct the specific steps and parameters of a workflow from what is published in scientific articles [16].

FAIRification [41] consists of a number of steps required to transform an existing data element to its FAIR version, typically leveraging RDF technology [42] for the interoperability aspect. RDF is a broadly applicable formal language to achieve the semantic interoperability principle I1. FAIRification starts by retrieving the non-FAIR data from its sources. Subsequently, these datasets are analyzed to understand the data structures and how they are mapped to concepts of the domain. The next step — semantic modelling — is a major activity comprising semantic harmonisation and integration, requiring the reuse and/or creation of models compliant with the FAIR principles. Once

the dataset is aligned with semantic definitions, it can be expressed in RDF and augmented with metadata. The last step is to store the FAIRified data into a findable and accessible manner.

6.3 The FAIR workflows approach

In this section we describe our workflow representation requirements, with a special focus on the coverage of manual steps, different workflow abstraction levels, and versioning on all these levels. We formulate these requirements as competency questions and present a configuration of elements from existing semantic models as a unified model to answer these competency questions.

6.3.1 Requirements and Competency Questions

With the help of structured interviews with data scientists and a gap analysis of the literature, we formulated user requirements for the reproducibility of workflows (the details of the interviews are given in Appendix 1). The interviewees stated that they experience many challenges in reproducing their or others' work, due to the lack of details of workflow steps, data cleaning and filtering. Also, essential information, such as processing parameters or design details needed to reproduce the results, is often missing. Some of these requirements are already covered by existing approaches while others have not been addressed so far. The interviews indicated that the definitions of manual processes of workflows are usually missing or incomplete, which is a requirement poorly addressed by computational workflow approaches. Often, software libraries, packages and versions of tools used are not explicitly recorded. The interviewees suggested making metadata of the datasets accessible, adding richer prospective and retrospective provenance and allowing for fine-grained workflow versioning linked to outputs produced during distinct executions. A unanimous recommendation was to allow for the separate input of relevant workflow parameters so that one can run the same workflow multiple times with different processing options without having to change the workflow itself.

The representation of software environment details (e.g., the

used libraries and packages) is already addressed by some of the surveyed semantic models, like Workflow4ever, CWLProv and Reproduce-me. We checked the capabilities of the existing semantic approaches to address the needs collected from the interviews. We concluded that none of the related work could completely address all the requirements together. The missing parts can be put in three main categories: (CQ1) Manual steps description and executions; (CQ2) abstraction levels of workflows; and (CQ3) versioning of executed workflows. Therefore, we propose the following additional sets of competency questions (CQ) to cover these missing parts:

The first group of questions (CQ1) is about manual steps:

- **CQ1.1:** Which steps are meant to be executed manually and which to be executed computationally?
- **CQ1.2:** For the manual steps, who are the agents responsible to execute them (individuals and roles)?
- **CQ1.3:** Which datasets were manually handled and their respective formats?
- **CQ1.4:** What are the types of manual steps involved?

The second group (CQ2) is about the instantiation of general workflows by more specific ones:

- **CQ2.1:** What are the main steps of a general workflow?
- **CQ2.2:** What are the steps of a specific workflow and how are they described?
- **CQ2.3:** What higher-level description instantiated a certain workflow step?
- **CQ2.4:** Who or what method made the instantiation of a semantic/meta level description of a step into an executable workflow step?

The third group (CQ3) are questions about versioning of workflows and their executions:

- **CQ3.1:** What are the existing versions of a workflow and what are their provenances?
- **CQ3.2:** Which instructions were removed/changed/added from one version to another?
- **CQ3.3:** Which steps were automatized from one version to another?
- **CQ3.4:** Which datasets were removed/changed/added for the different versions?
- **CQ3.5:** Which workflow version was used in each execution and what was generated?

To the best of our knowledge, none of the previous research on semantic modelling of workflows (or protocols/processes) addresses all these requirements together. In few cases some semantic models only partially cover some questions, as explained in the prior section.

6.3.2 Unified Model

From the study of the diverse existing semantic models for workflows and protocols, we compiled a unified conceptual model covering the elements required to answer our competency questions. For this, we applied the ontology-driven conceptual modelling approach [43], which is based on the Unified Foundational Ontology (UFO) and its ontological language OntoUML [44].

Figure 6.1 illustrates the main elements of our unified model¹, which is primarily based on DOLCE Ultra Lite (DUL), PROV, P-PLAN and BPMN 2.0. The most relevant ontology used is P-PLAN, which provides an abstract terminology of the main building blocks to describe plans.

¹<https://w3id.org/fair/plex>

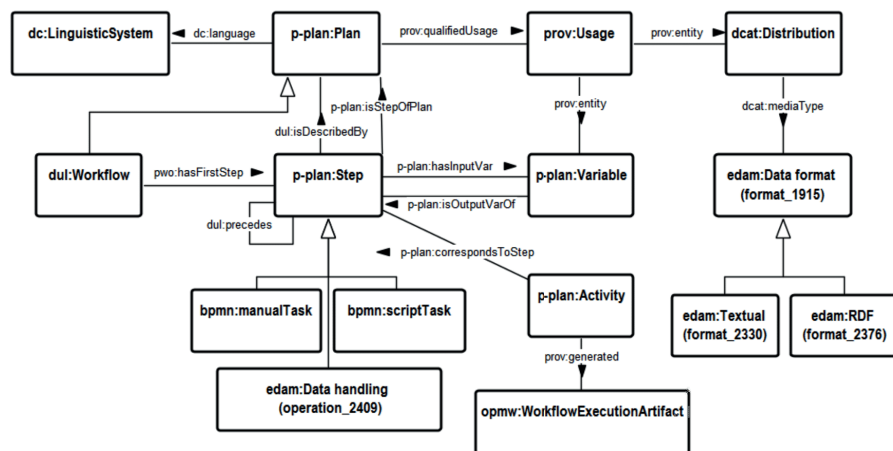


Figure 6.1: Unified semantic model for workflows

6

The *p-plan:Plan* category is the core element of our unified model and is the class used to classify any type of instruction. It allows for the composition of instruction by means of smaller steps (*p-plan:Step*) that have input and output variables (*p-plan:Variable*). With the *pwo:hasFirstStep* property, we can indicate the first step of a plan, and with *dul:precedes*, we can indicate whenever a step precedes another, thereby enabling the representation of sequential and parallel steps.

We decouple a particular step within a workflow from its instruction with the pattern *p-plan:Step dul:isDescribedBy p-plan:Plan*, where each step always points to one plan. This approach allows us to separate the workflow steps, enabling the reuse of instructions by different workflows. Therefore, in our approach, a step is a lightweight object (like a pointer) that serves only for ordering of instructions without coupling them to the specific workflow. Besides that, we use the *dul:isDescribedBy* property as a self-relationship of *p-plan:Plan*, to represent that an instruction describes another instruction in a different abstraction level. With this approach, we can represent anything from high-level abstract protocols to concrete and executable workflow steps, and the links between these levels. This can be used to first represent the general protocol and then move to

the definition of the executable steps akin to the common software engineering phases of specification and implementation. Our model can however also be used in the other direction to extract a new common protocol from similar existing concrete workflows. At the more abstract levels, instructions are written in a natural language like English (or possibly pseudo-code), whereas at the lowest level, we find the executable specifications, which can be written in a programming language and thereby automatically executable. Alternatively, at the lowest level instructions can be in natural language, such as for wet-lab instructions, which can naturally only be executed in a manual fashion. For example, the first general step of a specification of a machine learning pipeline like OpenPREDICT (to which we will come back to shortly) might be to “load all features and gold standard” (a *p-plan:Plan*). The concrete execution of this general step is described by four concrete and executable steps (written in a language such as Python), each having a link (*dul:isDescribedBy*) to the general description of the step.

We use the BPMN 2.0 ontology for the representation of manual and computational activities with *bpmn:ManualTask* and *bpmn:ScriptTask*, which we both define as subclasses of *p-plan:Step*. With this approach, the modeller can therefore include manual and automated steps in the same workflow. More specific classes can be used for particular workflow systems, such as *reprod:Cell* as a kind of *bpmn:ScriptTask* describing a code cell in a Jupyter Notebook.

We follow the FAIR Data Point specification² for the representation of datasets (input and output) through the *dcat:Dataset* element, which should be linked to the available distributions (*dcat:Distribution*) through the *dcat:distribution* property, and the URL to download the distribution is represented with *dcat:downloadURL*. We improved this approach with data formats from the EDAM ontology through the *dcat:mediaType* property. We use *prov:qualifiedUsage* for variable bindings. For example, the instruction (*p-plan:Plan*) to “download a dataset and save it in the local environment” has a link (*prov:qualifiedUsage*) to the “binding the

²<https://github.com/FAIRDataTeam/FAIRDataPoint-Spec>

online dataset to a local variable” (*prov:Usage*), which represents the connection between the dataset distribution (*dcat:Distribution*) and the local variable (*p-plan:Variable*) through instances of the *prov:entity* properties. For the representation of retrospective provenance, i.e., information about prior executions of a workflow, we follow the P-PLAN approach by using *p-plan:Activity* and linking it to the steps with *p-plan:correspondsToStep*.

To represent the roles of the different involved agents (such as people and software), we use the agent associations as defined in PROV. For example, the Jupyter Notebook (*prov:SoftwareAgent*) was used as execution environment (*prov:Role*) for all computational steps of the OpenPREDICT workflow. Furthermore, as a practical design decision, we extended the notion of *prov:Association* for endurants, so the modeller can apply the association pattern similarly to the perdurant way, i.e., use the property *prov:hadPlan* from *p-plan:Association* to *p-plan:Plan* instead of the relation from *prov:Activity* through *prov:qualifiedAssociation*. Therefore, this approach allows the modeller to represent the association of agent roles to an instruction. For example, Remzi is the OpenPREDICT main developer, so the “Remzi as developer of OpenPREDICT” (*prov:Association*) links to (a) the “Developer” (*prov:Role*) through *prov:hadRole* property, (b) the Remzi object (a *prov:Agent*) through *prov:agent*; and (c) all OpenPREDICT instructions (*p-plan:Plan*), through *prov:hadPlan*. Notice that, although the terminology of these properties targeted the perdurant aspect (*prov:Activity*), these properties are also useful for the endurant aspect. Ideally, they should have the adequate endurant terminology, so instead of *prov:hadPlan*, it should be “*prov:hasPlan*” (similarly for *prov:hadRole* too).

One of the most important links is the one between a workflow execution and its created outputs. For this, we specialized the PROV approach by using *prov:generated* to link a workflow activity (*p-plan:Activity*) to an output artefact (*opmw:WorkflowExecutionArtifact*). Therefore, each step execution can generate workflow execution artefacts. To represent the specifics of machine learning workflows, we moreover use the ML-Schema ontology (*mls*), such as to specify the trained model and its evaluation measures (via *mls:ModelEvaluation*

and *mls:EvaluationMeasure*). For example, it can be used to specify the accuracy of the models that were trained during different executions. For the representation of versioning, finally, we use *dc:hasVersion* to assign version identifiers and *prov:wasRevisionOf* to link to the previous versions, and apply this to all relevant elements, including workflows, instructions, software and datasets.

6.3.3 Case Study Topic

We evaluate our approach below with a case study of a computational drug repurposing method based on machine learning, called PREDICT [9]. PREDICT is one of the most frequently cited drug repurposing methods and provides a ranking of drug-disease associations on the basis of their similarity to a set of known associations. PREDICT has reported a high AUC (0.90) for predicting drug indications, though neither the original data nor the software to produce the results are available.

The features for the drug prediction classifier included five drug-drug similarity measures and two disease-disease similarity measures. The similarities between drugs were calculated based on molecular fingerprints, common side effects of drugs, target protein sequence alignment, semantic similarity of target genes of drugs in the Gene Ontology, and closeness of target proteins in human protein-protein interaction network. For the disease aspect, two disease-disease similarities were calculated based on medical description of diseases and semantic similarity of disease terms in the Human Phenotype Ontology. The method transforms drug-drug and disease-disease similarities into integrated features to be used for logistic regression training.

For evaluating the performance of the logistic regression, 10-fold cross-validation was used in two different ways: One in which 10% of drugs are hidden and one in which 10% of associations are hidden. In the first strategy, 10% randomly selected drugs in the gold standard and the known indications associated with them were removed. The positive training set consisted of the remaining 90% of drugs and the indications associated with them. The negative training set consisted of randomly generated drug-disease associations which were not in

the positive set. For the second strategy, the known associations were divided into 90% positive training and 10% positive test sets, while negative training and test sets were built using randomly generated drug-disease associations from respective sets. In the next section, we report on the application of our approach to this use case.

6.4 OpenPREDICT Case Study

As a case study, we took the original PREDICT workflow, as introduced above, and transformed it with our approach to make it open and FAIR. We therefore call the resulting workflow OpenPREDICT. It implements the same steps of the original PREDICT, i.e., five drug-drug similarity and two disease-disease similarity measures were used to train a logistic regression classifier to predict potential drug-disease association (see Figure 6.2). Therefore, we follow the same general protocol of these four steps:

6

1. **Data preparation:** In this step, the necessary dataset is collected and preprocessed.
2. **Feature Generation:** In this step, we generate features from the collected data sets. Drug-drug and disease-disease similarity scores were combined by computing the weighted geometric mean. Thus, we combine 5 drug-drug similarity measures and 2 disease-disease similarity measures, resulting in 10 features.
3. **Model Training:** In this step, the generated features from the previous step are to be used to train in a simple logistic classifier.
4. **Model Evaluation:** This step uses 2 different cross-validation approaches: One where 10 % of drugs is hidden and one where 10 % of associations is hidden for testing. ROC AUC, AUPR, accuracy, precision and F-score of the classifier on test data are reported.

Below we explain how we made a FAIR version of PREDICT's input data and then show how we used our approach to model the OpenPREDICT workflow that is consuming this data. The implemen-

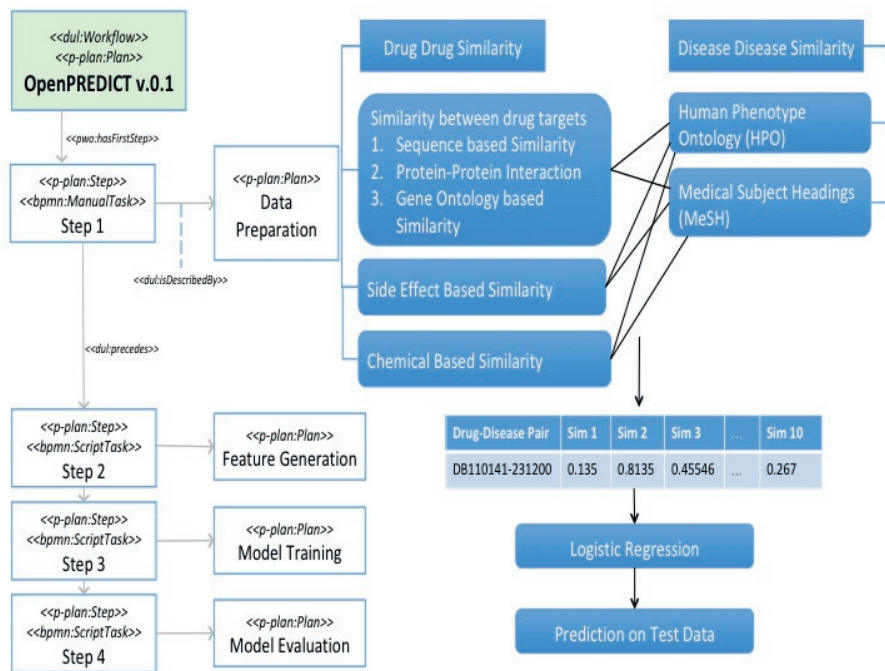


Figure 6.2: OpenPREDICT Workflow (version 0.1) with manual and computational steps .

tation and the workflow description of OpenPREDICT are available on GitHub³.

6.4.1 FAIRified Data Collection

Since the original data used in PREDICT is not publicly available, we collected data from open sources and made it FAIR with Linked Data [45] representations. We obtained data about drugs, drug targets, drug chemical structure, and drug target sequence from Drug-Bank [46], and additional drug targets from the KEGG dataset [47]. The SIDER dataset [48] was used for drug side effects and the HGNC and the GOA datasets [49, 50] were used for gene identifier mapping and gene ontology (GO) annotation respectively. We used Linked Data versions of the above-mentioned datasets from Bio2RDF [51], which is

³<https://github.com/fair-workflows/openpredict>

an influential resource for the biomedical sciences, providing a network of data collected from several major biological databases. On top of that, we used the supplementary file provided by [52] for protein-protein interactions and disease phenotype annotations that link HPO terms to OMIM diseases⁴. MeSH annotations were collected from [53]⁵ and annotations were also obtained by NCBO annotator API [54] using the OMIM disease description.

The data that was not yet in a Linked Data format were converted to RDF with a FAIRification process [41]. We kept the copies of the retrieved non-RDF datasets in our GitHub repository to prevent data access issues that may arise if data sources are unavailable. We also stored the collected datasets in a triplestore and created SPARQL queries to access the triplestore in order to produce the features for PREDICT's method.

Our OpenPREDICT workflow has two versions (0.1 and 0.2). In the first, we experimented with the FAIRifier tool with the two inputs that are provided as text files, i.e., (protein-protein interactions in human interactome) and (disease phenotypic descriptions). Besides the formalization of the manual steps through our approach, we also provide guidelines for the manual steps. In the second, we wrote Python scripts for FAIRification process of these datasets, evolving most of the manual steps to computational ones. Table 6.1 summarizes the list of all datasets used in version v0.1 and v0.2. OpenPREDICT v0.2 also uses a different MESH annotation dataset for disease similarity (indicated with † in Table 6.1).

For this FAIRification process, the data needs to be mapped to formal semantic models. In our case, important concepts included “protein-protein interaction” from Bioportal, “protein interactions” from EDAM (*edam:topic_0128*), the Bio2RDF properties *bio2rdf:interactor_a* and *bio2rdf:interactor_b* for the gene interactors representing the role of a gene in a protein-protein interaction, and “disease” and “has phenotype” from SIO (*SIO_010299* and *SIO_001279*).

⁴<https://hpo.jax.org/app/download/annotation>

⁵https://paccanarolab.org/disease_similarity

Table 6.1: All datasets used in OpenPREDICT version v0.1 and v0.2.

Dataset file	Retrieved	Data format	Download URL
Bio2RDF r4 datasets (Drugbank, KEGG, HGNC, SIDER and GOA)	2019-08-15	.nq (RDF) compressed as .gz	https://download.bio2rdf.org/#/release/4/
PREDICT drug indication gold standard	2019-08-15	.tab with tabular separator	https://www.ncbi.nlm.nih.gov/pmc/article/PMC3159979/bin/msb201126-s4.xls
Pubchem-Drugbank mappings	2019-08-15	.tab with tabular separator	https://raw.githubusercontent.com/dhimmel/drugbank/gh-pages/data/mapping/pubchem.tsv
Protein-protein interactions	2019-08-15	.txt with tabular separator	https://media.nature.com/full/nature-assets/srep/2016/161017/srep35241/extref/srep35241-s3.txt
HPO Phenotype annotations	2019-08-15	.tab with tabular separator	http://compbio.charite.de/jenkins/job/hpo.annotations/lastSuccessfulBuild/artifact/misc/phenotype_annotation_hpoteam.tab
†MESH Phenotype annotations	2019-08-15	.tab with tabular separator	http://www.paccanarolab.org/static_content/disease_similarity/mim2mesh.tsv
MESH Phenotype annotations (BioPortal)	2019-08-15	.txt file	https://github.com/fair-workflows/openpredict/blob/master/data/external/meshAnnotationsFrom/BioPorttalUsingOMIMDesc.txt

6.4.2 OpenPREDICT workflow representation

Figure 6.2 illustrates the main steps of the OpenPREDICT workflow, in which the main protocol is represented as a *dul:Workflow* and a *p-plan:Plan*, with version set through the *dc:hasVersion* property. The workflow consists of four steps: data preparation, feature generation, model training and evaluation, and presentation of results. Each one is defined by (*dul:isDescribedBy*) its own *p-plan:Plan*. In the first version of OpenPREDICT (0.1) all steps within the data preparation were manual (*bpmn:ManualTask*), as the FAIRification process and the preparation steps on data that were already provided as RDF. The second version of OpenPREDICT (0.2) automated most of these manual steps, requiring less human intervention. We will now go through some of the most important aspects of this representation.

Prospective provenance. We decoupled the workflow steps from the instructions, linking a *p-plan:Step* to a *p-plan:Variable* through *p-plan:hasInputVar* and *p-plan:hasOutputVar*, while the *p-plan:Plan* links

to the *prov:Usage* through the *prov:qualifiedUsage* property, describing how to bind the variable to other resources. This is an example:

```
opredict:Step_Download_Drugbank_dataset
  rdf:type bpmn:ManualTask ;
  rdf:type edam:operation_2409 ;
  rdf:type p-plan:Step ;
  p-plan:hasOutputVar opredict:
    Variable_Drugbank_dataset_online ;
  p-plan:isStepOfPlan opredict:Plan_Main_Protocol_v01 ;
  dul:isDescribedBy opredict:Plan_Download_Drugbank_dataset ;
  dul:precedes opredict:Step_Save_Drugbank_dataset ;
  rdfs:label "Download Drugbank dataset" ;
.

opredict:Plan_Download_Drugbank_dataset
  rdf:type p-plan:Plan ;
  dc:description "Download Drugbank dataset" ;
  dc:language :LinguisticSystem_xsd_language_English ;
  rdfs:label "Download Drugbank dataset" ;
  prov:qualifiedUsage opredict:
    Usage_Fetch_download_Drugbank_dataset_to_variable ;
.

opredict:Usage_Fetch_download_Drugbank_dataset_to_variable
  rdf:type prov:Usage ;
  rdfs:label "Link variable to download Drugbank dataset" ;
  prov:entity opredict:Distribution_release-4-drugbank-
    drugbank.nq.gz;
  prov:entity opredict:Variable_Drugbank_dataset_online ;
.

opredict:Distribution_release-4-drugbank-drugbank.nq.gz
  rdf:type dcat:Distribution ;
  rdfs:label "release/4/drugbank/drugbank.nq.gz" ;
  dcat:downloadURL "http://download.bio2rdf.org/files/release
    /4/drugbank/drugbank.nq.gz" ;
  dcat:mediaType opredict:DataFormat_nq_compressed_gz ;
.

opredict:Variable_Drugbank_dataset_online
  rdf:type p-plan:Variable ;
  rdfs:label "Drugbank dataset online" ;
.
```

Retrospective provenance. We represent the concrete executions that happened and the concrete output that was generated with a *p-plan:Activity* that is linked to a *p-plan:Step* through the *p-plan:correspondsToStep* property and to the outputs (*opmw:WorkflowExecutionArtifact*) through *prov:generated*. Each output has a value (e.g., accuracy rate) and is linked to *prov:Generation* through the *prov:qualifiedGeneration* property, which specifies when the generation occurred with (*prov:atTime*). This is an example:

```

opredict:
  Activity_Model_preparation_train_and_evaluation_Execution
  rdf:type p-plan:Activity ;
  p-plan:correspondsToStep opredict:
    Step_Model_preparation_train_and_evaluation ;
  prov:generated opredict:
    ModelEvaluation_Accuracy_Execution_1546302862 ;
  prov:generated opredict:
    ModelEvaluation_AveragePrecision_Execution_1546302862 ;
  prov:generated opredict:
    ModelEvaluation_F1_Execution_1546302862 ;
  prov:generated opredict:ModelEvaluation_Precision_1546302862
    ;
  prov:generated opredict:
    ModelEvaluation_Recall_Execution_1546302862 ;
  prov:generated opredict:
    ModelEvaluation_RocAuc_Execution_1546302862 ;
.

opredict:ModelEvaluation_Accuracy_Execution_1546302862
  rdf:type mls:ModelEvaluation ;
  dc:description "0.833336" ;
  mls:specifiedBy opredict:
    EvaluationMeasure_PredictiveAccuracy ;
  prov:qualifiedGeneration opredict:
    Generation_Execution_1546302862 ;
.

opredict:Generation_Execution_1546302862
  rdf:type prov:Generation ;
  prov:atTime "2019-01-01T00:02:31.011"^^xsd:dateTime ;

```

Versioning of workflows. We track the modification across the two versions with the *dc:hasVersion* property on the level of a *dul:Workflow*, *p-plan:Plan*, *dc:LinguisticSystem*, and *prov:SoftwareAgent*. Furthermore, we use the *prov:wasRevisionOf* property to link to the previous version. This is an example:

```
opredict:Plan_Main_Protocol_v02
  rdf:type p-plan:Plan ;
  rdf:type dul:Workflow ;
  dc:created "2019-05-15" ;
  dc:creator opredict:Agent_Remzi ;
  dc:description "OpenPREDICT Main Protocol v.0.2" ;
  dc:hasVersion "0.2" ;
  dc:language :LinguisticSystem_xsd_language_English ;
  dc:modified "2019-07-03" ;
  pwo:hasFirstStep opredict:Step_Prepare_Input_Data_Files_v02
    ;
  rdfs:label "Main Protocol v.0.2" ;
  prov:wasAttributedTo opredict:Agent_Remzi ;
  prov:wasRevisionOf opredict:Plan_Main_Protocol_v01 ;
.

opredict:Plan_Main_Protocol_v01
  rdf:type p-plan:Plan ;
  rdf:type dul:Workflow ;
  dc:created "2018-11-27" ;
  dc:creator opredict:Agent_Remzi ;
  dc:description "OpenPREDICT Main Protocol v.0.1" ;
  dc:hasVersion "0.1" ;
  dc:language :LinguisticSystem_xsd_language_English ;
  dc:modified "2019-05-15" ;
  pwo:hasFirstStep opredict:Step_Prepare_Input_Data_Files ;
  rdfs:label "Main Protocol v.0.1" ;
  prov:wasAttributedTo opredict:Agent_Remzi ;
.
```

6.5 Evaluation

In this section, we describe the evaluation of our approach, consisting of two parts: First, we revisit each FAIR principle and explain

how the principle is addressed. Second, we applied the traditional ontology validation methodology by answering the competency questions through the execution of SPARQL queries (the concrete queries are available in GitHub repo).

6.5.1 Addressing the FAIR principles

In order for our workflow to comply with FAIR principles, we checked each FAIR criterion defined in [13], as identified between parentheses below. First, global and persistent identifiers were assigned to resources defined in the workflow and associated data. Rich metadata for workflow and input and output data were created using HCLS⁶ and FAIR data point specification (F2). In addition, the metadata we generated contains an explicit global and persistent identifier of the data they describe (F3). In order to enable the workflow and the data used to be searched, they were uploaded in a triple-store as a FAIR Data Point⁷. Data can be queried through SPARQL over HTTP(S) protocol (A1.1). Since the data is not private or protected, we don't require authentication and authorisation mechanism (A1.2). All data and metadata are permanently available at Zenodo⁸ to make the metadata accessible even the data is no longer available (A2). We used RDF and OWL with commonly used controlled vocabularies and ontologies such as Bio2RDF vocabulary, SIO and PROV to model input data and workflows (I1). HCLS dataset specification and FAIR Data Point specification were used to define the metadata and provenance of data (I2). Meaningful links between (meta)data such as Bio2RDF links and data and workflow were created (I3). To increase reusability of the workflow, we describe the workflow and its data with community standards such as ML-Schema and P-PLAN (R1). We provide the license (R1.1) and provenance information in the metadata using FAIR data point specification (R1.2), and HCLS specification (R1.3) and PROV.

⁶<https://www.w3.org/TR/hcls-dataset/>

⁷<https://graphdb.dumontierlab.com/repositories/openpredict>

⁸<https://doi.org/10.5281/zenodo.3770918>

6.5.2 Answering competency questions

Besides evaluating whether each FAIR principle was addressed, we also assessed the unified model using the common semantic validation approach, which is based on SPARQL queries used to answer the competency questions. All questions listed in Section 3 could be answered by running the SPARQL queries over the OpenPREDICT use case. The complete queries and results can be found online⁹. Therefore, the reproduction of this validation can be performed by re-executing the queries on the RDF representation of the OpenPREDICT workflow. Below we explain the result for each competency question.

CQ1 - Questions about manual steps.

6

CQ1.1: Which steps are meant to be executed manually and which to be executed computationally? The SPARQL query we built to answer this question first filters all steps within the first version of OpenPREDICT workflow (*opredict:Plan_Main_Protocol_v01*). The results show each step and its type - manual (*bpmn:ManualTask*) or computational (*bpmn:ScriptTask*) - as well as the respective instructions (*p-plan:Plan*) that describe the steps. In summary, OpenPREDICT v0.1 has 28 manual steps and 14 computational steps (42 in total), while v0.2 has 9 manual steps and 9 computational steps (18 in total). This difference reflects the automatization of most of the manual steps within data preparation (evolving from manual to computational) and the simplification of the computational steps described in fewer Jupyter Notebook cells.

CQ1.2: For the manual steps, who are the agents responsible to execute them? To answer this question we filtered the results for only manual steps through the statement:

```
values ?stepType { bpmn:ManualTask }
```

⁹<https://github.com/fair-workflows/openpredict>

The result is a list of all steps and roles related to each one, such as executor, creator, developer and publisher. For example, Remzi is creator, developer and executor of all instructions, while Ahmed is a developer of some computational steps and Joao is the executor of the entire OpenPREDICT workflow. This approach allows for the representation of multiple roles played by different agents within each step.

As in related approaches such as Workflow4ever and Reproduce-me, we use the PROV ontology to address the different types of agents and roles through the *prov:wasAttributedTo* property, and apply the *dc:creator* and *dc:publisher* properties for the direct relation from an instruction to an agent.

CQ1.3: Which datasets were manually handled and what are their formats?

OpenPREDICT's computational steps use datasets, as explained in Section 4.1, that required manual pre-processing. The difference between v0.1 and v0.2 is that we automated the manual pre-processing of two datasets in v0.2; MESH Phenotype annotations and protein-protein inter-actions. The main elements of the query reflect the FAIR data point specification with DCAT elements (*dcat:Distribution*, *dcat:downloadURL* and *dcat:mediaType*), PROV (*prov:Usage* and *prov:qualifiedUsage*) and EDAM classification for data handling steps (*edam:operation_2409*) and data formats (media types).

CQ1.4: What are the types of manual steps involved, and what are their inputs and outputs?

Similar to the Reproduce-me approach, our ontology leverages on the P-PLAN ontology to address the variables used as input and output of the manual steps, mostly during data preparation in OpenPREDICT v0.1, such as downloading and saving the datasets listed in the results of CQ1.3. For example, the input of *opredict:Step_Save_files_in_triplestore* are variables that indicate the local file of each dataset (serialized as RDF) and the output variable indicating the endpoint to upload all datasets (*opredict:Variable_Triplestore_endpoint_for_input_data*).

When changing the filter from manual steps to computational steps, the pattern followed was to classify the output variables of a step (a Jupyter Notebook cell) according to the data saved in files. For example, in feature generation, the *opredict:Step_Feature_generation_01_Pipeline_Source_Cell11* has an output variable for drug fingerprint similarity, indicating the generation of the file “drugs-fingerprint-sim.csv”

CC2 - Questions about instantiation of general workflows by more specific ones.

CC2.1: What are the main steps of a general workflow?

OpenPREDICT workflow follows the common machine learning pipeline process of: data preparation, feature generation, model training, model evaluation and presentation of results. The query returns these steps by looking for the first step of the workflow (through *pwo:hasFirstStep*) and following the preceding path in a recursive way, e.g.,

```
?step1 dul:precedes ?step2.  
?step2 dul:precedes ?step3.  
?step3 dul:precedes ?step4. (until there is no preceding steps  
)
```

The classification of the step is given by the EDAM specializations of the Operation concept (*operation_0004*), such as *Data Handling* for data preparation (*edam:operation_2409*). For the sake of simplicity, model training and evaluation were performed within the same step. The main steps are listed below:

```
opredict:Step_Prepate_Input_Data_Files  
opredict:Step_Feature_generation_Pipeline_OpenPREDICT_ipynb  
opredict:Step_Model_preparation_train_and_evaluation_Workflow_  
-OpenPREDCIT_ML_ipynb
```

```
opredict:Step_Format_results_for_presentation
```

CQ2.2: What are the steps of a specific workflow? Similar to the previous question, the SPARQL query uses the properties that allow for the ordering of steps' execution (*pwo:hasFirstStep* and *dul:precedes*). The pattern *p-plan:Step dul:isDescribedBy p-plan:Plan* allows us to answer this question by representing how a step is described by an instruction. This pattern resembles the one used by Workflow4ever, which applies the *wfdesc:hasWorkflowDefinition (dul:isDescribed)* to link a *wfdesc:Workflow (p-plan:Step)* to a *wfdesc:WorkflowDefinition (p-plan:Plan)*, aiming at representing the instructions (e.g., a Python script) that are natively understood by the *wfdesc:WorkflowEngine (prov:SoftwareAgent)*. However, different from this approach, we classify the instruction language (*p-plan:Plan dc:language dc:LinguisticSystem*), allowing for the representation of instructions that follow computer language or natural language, which includes pseudo-code — commonly used to specify algorithms before implementing in a particular computer language.

The results show that OpenPREDICT has 78 steps in total, where 60 steps belong to v0.1 and 18 belong to v0.2, each step linked to an instruction. 9 instructions were reused from v0.1 to v0.2 regarding data preparation, thus, v0.2 presents 9 new instructions that are used to automate the data preparation phase. These instructions are written as either English (natural language) or Python 3.5 (computer language), where most of the Python ones refer to the Jupyter notebook cells for feature generation and model training and evaluation.

CQ2.3: What higher-level description does a certain workflow step instantiate? The SPARQL query to answer this question includes the pattern *p-plan:Plan dul:isDescribedBy p-plan:Plan*, which extends the capability described in the previous question, i.e. decoupling steps from instructions, enabling the representation of different abstraction levels of instructions and their relations.

This pattern resembles the links between specification artefacts (e.g., conceptual model, activity diagrams and use cases) and implementation artefacts (e.g., software code, deployment procedures and automated tests) in software engineering. Usually, a specification artefact aims at describing the instructions necessary to enable a programmer to create the software code, sometimes automatically generated as in model-driven engineering. For example, a pseudo-code within an activity diagram (*p-plan:Plan*) may describe the behaviour expected (*dul:isDescribed*) for the algorithm behind a service endpoint, which may be implemented as a Python script (*p-plan:Plan*).

OpenPREDICT did not formally follow the specification phase of software engineering since it is a research project, having the code developed from the data scientist interpretation perspective about publications related to PREDICT. In research-oriented data science, this type of approach is common. However, we created some examples of the pattern that represent the specification of OpenPREDICT workflow. Therefore, the results of this query include 10 Jupyter Notebook cell instructions (*p-plan:Plan*), representing implementation artefacts, that were specified (*p-plan:isDescribedBy*) by 3 specification instructions (*p-plan:Plan*). The level of abstraction can be derived from the properties of the instruction. For example, the 10 Jupyter Notebook cell instructions were written (*dc:language*) in Python 3.5 (*schema:ComputerLanguage*), while the 3 specification instructions were written in English (*en* value of *xsd:language*). Furthermore, this approach enables links of *s* (specification artefacts) *x* *i* (implementation artefacts), where *i* > *s*, i.e., a specification artefact usually describes several software code lines (instructions). In OpenPREDICT, the first specification instruction guides the load of input datasets, which is linked to cells 1-5 of the feature generation step, while the second guides the calculation of scores between pairs of drugs and compute similarity feature, which is linked to cells 6-9.

CQ3 - Questions about versioning of workflows and their executions

CQ3.1: What are the existing versions of a workflow and what are their provenance? The collective workflow (the whole) is represented as a *dul:Workflow* and a *p-plan:Plan*. Similar to other approaches (Workflow4ever, Reproduce-me, CWLProv, among others) the query to answer this question makes use of DC properties (e.g., *dc:creator*, *dc:created*, *dc:modified*) and PROV (e.g. *prov:wasAttributedTo*) for prospective provenance. It also covers workflow versioning through *dc:hasVersion* and *prov:wasRevisionOf*, where the former is responsible for version of *dul:Workflow* and the latter to link an instruction to another (*p-plan:Plan* *prov:wasRevisionOf* *p-plan:Plan* pattern). The retrospective (executions) provenance is supported by the link from an execution (a *p-plan:Activity*) to the correspondent step (*p-plan:correspondsToStep* property), which is a pattern that resembles most of the aforementioned semantic models. The main difference here is the assumption that any instruction (*p-plan:Plan*) should be versionable, thus, all executions link to a versioned instruction. Differently from Workflow4ever approach, here we do not introduce any elements regarding the specification of the changes (e.g., *roevo:ChangeSpecification*). The results for OpenPREDICT show 2 workflows (v0.1 and v0.2), both created by and attributed to Remzi, where v0.2 links to the prior version (v0.1).

CQ3.2: Which instructions were removed/changed/added from one version to another? Three SPARQL queries were written to answer whether the instructions of OpenPREDICT v0.1 were removed or changed or added in v0.2. Each SPARQL uses the identifier of the workflow versions (retrieved in CQ3.1) as an input parameter to perform the comparison from one version to another. For the query for removed instructions, it considers all instructions used in v0.1 that are not used in v0.2 and excludes the instructions that were changed. For the query for changed instructions, it considers the instructions with the *prov:wasRevisionOf* property. For the query for added instructions, the SPARQL query uses the reverse logic from the

removed.

Forty-seven instructions were removed from v0.1 to v0.2 due to the refactoring of the code of feature generation, model training and model evaluation, and the elimination of several manual steps in data preparation. Three instructions were changed, reflecting the porting of the FAIRification manual steps to computational steps in data preparation, i.e., download and save human interactome Barabási, and phenotype annotations. Seven instructions were added in v0.2, where 3 of them represent the new Python scripts for data preparation of the new data sources, other 3 represent the new scripts for feature generation and the remaining for model training.

CQ3.3: Which steps were automatized from one version to another?

This query is quite similar to the one used for changed instructions (CQ3.2), but it makes explicit that the old version of the instruction used as manual step (*bpmn:ManualTask*) was modified to an instruction used as computational step (*bpmn:ScriptTask*) in the new version. The results confirm the findings from the previous query regarding the 3 instructions that were ported from manual steps to computational steps, namely the data preparation top-level instruction, the FAIRification instructions (download and save human interactome Barabási, and phenotype annotations). Although our approach covers change management, we face the same challenges regarding the dependency of the developer practices for code versioning. This means that, for example, a developer is free to choose whether to remove files from an old version of the software implementation and add files to the new version, even though these files refer to the same capability or routines. Most of the version controls track the changes when the files (old and new) have the same name and path (i.e., the step identifier), which is a similar approach used here.

CQ3.4: Which datasets were removed, changed, or added from one version to the next?

This question can be answered by mixing the same query of CQ1.3 (datasets manually used) with the logic used

in query CQ3.2, i.e., one SPARQL query to the datasets removed, one for the changed and one for the added. The query results over OpenPREDICT (v0.1 and v0.2) confirm the findings of CQ1.3, where none datasets were removed from the old version to the new, none changed and 2 were added.

CQ3.5: Which workflow version was used in each execution and what was generated?

This question is answered by using the pattern *p-plan:Activity p-plan:correspondsToStep p-plan:Step*, where the step is part of the *dul:Workflow* that provides the workflow version. The OpenPREDICT workflow had 14 executions represented with our unified model, exemplifying the execution of some computational steps, i.e., each one a particular Jupyter Notebook cell. Therefore, this approach allows for the representation of multiple executions of each step according to the version of the corresponding instruction. Each execution inherits the properties of *p-plan:Activity*, e.g., the event start and end time points. Furthermore, each execution is associated to the correspondent generated artefacts through the *p-plan:Activity prov:generated opmw:WorkflowExecutionArtifact* pattern, a similar approach of Workflow4ever, which applied the inverse property *prov:wasGeneratedBy*. An artefact generated by an execution can be an evaluation measure of the trained model, such as the model accuracy and recall for that particular execution, i.e., a *mls:ModelEvaluation*. Therefore, OpenPREDICT executions generated the values about the model evaluation measures of accuracy, average precision, F1, precision, recall and ROC AUC. For example, the results show that the model accuracy of v0.1 is 0.83, while v0.2 is 0.85. This query can be further extended by considering the particular version of each instruction that the executed step implements. In addition, ideally, each output of a Jupyter Notebook cell should be represented as a *opmw:WorkflowExecutionArtifact*, so all generated outputs are stored (similar to ProvBook/Reproduce-me approach). This query can be easily changed to provide aggregations for related analytical questions, such as how often each workflow version was executed.

6.6 Discussion

Before we move on to discuss the encountered reproducibility challenges and other issues, we would like to first highlight the two FAIR perspectives that our approach embodies and demonstrates. Firstly, FAIR applies to the datasets that a scientific workflow consumes and produces, e.g., the protein–protein interactions dataset used by OpenPREDICT, and we need FAIRification approaches to raise existing datasets to this standard. This is the aspect the FAIR principles originally focused on. On top of that, we have here proposed and exemplified a second perspective. This additional perspective regards the workflows’ own FAIRification, i.e., the process of aligning them with the FAIR principles, which relies on a semantic modelling approach such as the one described in this paper. Our work therefore expands the notions of FAIR and FAIRification from the relatively static artifacts of datasets to the dynamic processes of workflows.

6

6.6.1 Reproducibility Challenges

It was expected that our study would be unable to fully reproduce the accuracy of the method reported in the PREDICT paper due to use of different input datasets. The performance results of this study are lower than originally reported. The PREDICT paper reported an AUC of 0.90 in cross-validation but using the same gold standard, we could only achieve a AUC of 0.83.

We were also able to obtain the drug and disease similarity matrices used in PREDICT from the authors via email request. Given 5 drug–drug similarity measures for 593 drugs and 2 disease–disease similarity measures for 313 diseases, there are resulting 10 features of combined drug–disease similarities. The logistic classifiers were trained with these pre-computed similarity scores and an average AUC of 0.85 was obtained from 10 repetitions in a 10-fold cross-validation scheme. This is still a significant difference from the AUC of 0.90 what the authors reported in PREDICT study. This indicates that there was more likely an error in the design or implementation of evaluation, and not the aggregation of data nor the calculation of drug–drug and disease–

disease similarity scores.

While attempting to reproduce the PREDICT study, we faced the following issues, which we have turned into generic recommendations (highlighted in italics).

1. **Insufficient documentation** Essential details concerning the calculation of features were not clearly defined, nor was the software code to perform the calculations provided. *Many details of an experiment, including data sets, processing parameters and applied software and algorithms, need to be specified in order to facilitate the replication of the results. A methods section in a scientific article may not be the best place to provide all this information as it is usually limited by size constraints and different organization styles of journals and conference proceedings, leading to a lack of required detail.*
2. **Inaccessible or missing data** Since no data except the gold standard data (drug-disease associations) were given, the features for the PREDICT workflow were reconstructed using the publicly accessible databases DrugBank and KEGG and SIDER. However, we could not check if this resulted in exactly the same datasets. *The original data that were available to the authors could be absent or no longer accessible to others for many reasons. Sufficient data should be published to enable reproducing a study.*
3. **Versioning and change of data** In PREDICT, publicly accessible datasets have been used to construct models and validate hypotheses for prediction of drug indications and drugs were identified by their Drugbank IDs. However, Drugbank IDs are subject to change over time. For example, two drugs (DB00510, DB00313) in the original dataset were merged to the same drug within the current version of the Drugbank. *Results or hypotheses may change as a result of updated input data. In order to reconstruct the original conclusion, it is important to record the version or the date of the data that were used in a study. This is especially important as publicly accessible datasets are increasingly used to construct models and validate hypotheses for prediction of drug indications.*

4. **Execution environment and third-party dependencies** In the PREDICT study, the versions of some software tools, such as the library for semantic similarity calculation, were not specified. *The versions of software libraries, packages and tools used in a workflow should be explicitly mentioned, and an effort must be made to maintain the access to those releases used in the original workflow.*

6.6.2 Further Issues Encountered

While the execution of the FAIRification process in the OpenPREDICT was straightforward, the semantic modelling of the unified workflow model was challenging. The reuse of existing semantic vocabularies for the representation of our unified model proved to be an extensive task. There are several existing semantic approaches to represent workflows that present reproducibility issues and different conceptualizations, sometimes overlapping in their terminology. CWL and WDL, two most commonly used computational workflow languages, are designed to separate the workflow description from its execution. These languages are aimed to exchange and run computational workflows reproducibly in different environments. In contrast, we aimed to define workflows that involve both manual and computational steps, using semantic web standards and enriching workflows with sufficient metadata to make them FAIR. The prospective part of Workflow4ever implementation (wfdesc¹⁰) has consistency issues such as missing disjointness and licensing elements, besides not conforming to the documentation (e.g., for all elements related to workflow templates). On the other hand, semantic models like DUL, PROV and P-PLAN presented higher quality and common foundations (in DOLCE), while being easier to reuse and extend. Although CWLProv also provides an ontology-based on W4ever semantic models, it is oriented only to retrospective provenance of computational steps, reusing most of the predicates that P-PLAN extends (from PROV). Furthermore, the URI (<https://w3id.org/cwl/prov>) does not provide a concrete description of the new predicates (e.g., cwlprov:image) and neither resolves to the RDF model (TBox).

¹⁰<https://raw.githubusercontent.com/wf4ever/ro/0.1/wfdesc.owl>

A question that may arise is whether it would be better to create a new ontology from scratch rather than creating a unified model based on the existing ontologies. We believe that high quality semantic models should be reused, taking benefit from the lessons learned. Furthermore, we consider that reusing existing semantic workflow models actually improve semantic interoperability, while creating a new ontology may impede interoperability if it is not accompanied with alignments to the existing semantic models. Therefore, our approach appears to lead to improved semantic interoperability. Because we reused several semantic models, the competency questions that they target are potentially addressed by our approach. For example, the gap in our approach regarding the representation of change management for versioning can be addressed by reusing some elements from the versioning approach of Workflow4ever, e.g., *roevo:Change*, *roevo:ChangeSpecification* and *roevo:VersionableResource*.

Deciding which type of approach should be used for role representation should be based on the needs for either a fine-grained definition of the role/relator pattern (a reified relationship), such as the *prov:Association* approach, or a simple property, such as *dc:creator*. While the former (1) enriches the definition of the role (an improved representation capability), the latter (2) is less verbose:

```
(1)
?Association prov:agent opredict:Remzi;
prov:hadRole opredict:Creator;
prov:hadPlan ?plan.

(2)
?plan dc:creator 'Remzi' .
```

One of the main challenges is to understand the different terminology used for similar conceptualizations. Although the definitions of terms like plan, process, protocol, procedure, workflow, plan specification and standard operating procedure seem to be the same (or quite overlapping), their meanings become notoriously ambiguous across varied communities. How to grasp these semantic differences is a cru-

cial question that needs further exploration. For example, in the bioinformatics community, the term *Workflow* usually refers to an implemented (computational) piece of software, i.e., a set of programming language instructions, usually developed with a *Workflow Management System* as a *Workflow Application* [55]. Meanwhile, in software engineering, the *Workflow* term is usually referred to as a detailed business process within the Business Process Modelling (BPM) research. Usually, the BPM languages conform to graphical notations (e.g., BPMN, EPC, ARIS), targeted to human comprehension rather than computational ends (process design/modelling). Additionally, some BPM languages focus on representing, at a lower level of abstraction, the process execution details, e.g., BPEL (process implementation) [56]. This is a topic extensively covered by Service Oriented Architecture (SOA) initiatives. Several related works target the gap that exists between business process models and workflow specifications and implementations, such as service composition schemes [57] and formal provenance of process executions and versioning [58]. Furthermore, some of these languages provide predicates for forks and conditionals, which were intentionally not included in the unified model since they have a high complexity. This is still a topic under discussion in the CWL community, for example.

In future work, we will improve the modelling of manual steps by studying and possibly incorporating predicates from the SMART protocols ontology. We will characterize the abstraction levels of workflows based on multi-level process modelling approaches, such as the widespread adopted APQC's Process Classification Framework (PCF). The PCF provides 5 abstraction levels for process specification, from a high abstraction level to detailed workflow specification: category (level 1), process group (level 2), process (level 3), activity (level 4) and task (level 5). Although this framework aims at providing a methodological approach for business process specification, we should investigate whether the minimal information elements of each level require proper representation in the ontology. We should also consider the challenges of process refinement ("process description in a more fine-grained representation") [59]. A process refinement mechanism maps and/or derives models from a higher-level specification to a detailed level, equivalent to vertical and exogenous model transformations in

model-driven engineering. Typical refinement categories will be investigated, such as activity decomposition principles about event delivery and execution condition transference [60, 61]. The representation of intentionality of the activities within business processes will also be addressed in future work through goal-oriented semantic process modeling [62], linking goals to activities and roles.

Industry-oriented approaches are also being investigated, such as Extract, Transform and Loading (ETL/ELT) for data warehousing and SQL Server Integration Services, which considers a workflow as a control flow, while a `dataflow` transforms data from a source to a destination. Furthermore, Product Line Management (PLM) tools should be investigated, especially the ones that cover Laboratory Information Management System (LIMS), which provides important concepts such as Bill-of-Materials (BoM), specifications and their certifications. For example, in PLM a *specification* is a description of raw materials and packaging materials and semi-finished and finished products. This description may contain product characteristics (e.g., chemical compounds), recipes (e.g., BoM), production methods, quality norms and methods, artwork, documents and others.

Ultimately, initiatives like CWL, the Center for Expanded Data Annotation and Retrieval (CEDAR) (for metadata management) [63] and FAIRsharing.org (for indexing FAIR standards) may be used as building blocks for the envisioned FAIR workbench tool, which can be a reference implementation over a workflow system such as Jupyter Notebook (e.g., a plug-in). Finally, the validation of the reproducibility level of a workflow should consider specific FAIR metrics that take in consideration specific recommendations (e.g., from CWLProv approach) and the practices for higher reproducibility of Jupyter notebooks [21].

6.7 Conclusions

In this work, we examined how FAIR principles can be applied to scientific workflows. We adopted the FAIR principles to make the PREDICT workflow, a drug repurposing workflow based on machine learning, open, reproducible and interoperable. From this stems the

main contribution of this paper, the OpenPREDICT case study, which demonstrates how to make a machine learning workflow FAIR and open. To do this, we created a unified model that reuses several semantic models to show how a workflow can be semantically modeled. We published the workflow representation, data and meta-data in a triple store which was used as a FAIR data point. In addition, new competency questions have been defined for FAIR workflows and how these questions can be answered through SPARQL queries. Among the main lessons learned, we highlight how the main existing workflow modelling approaches can be reused and enhanced by our unified model. However, reusing these semantic models showed to be a challenging task, once they present reproducibility issues and different conceptualizations, sometimes overlapping in their terminology.

In the future, we envision that the intensive human effort that we had to perform in order to make a workflow FAIR will be taken care of by smart and intuitive workflow tools. As a prototype of such a tool, we are currently developing the FAIR workbench as a general tool that allows users to deal with workflows and protocols in a semantic and FAIR form.

6.8 Acknowledgments

This research was funded by the NWO (grant number 628.011.011) and the Netherlands eScience Center (grant number P 17.0201). Support for this work was provided by NCATS, through the Biomedical Data Translator program (NIH award OT2TR003434)

6.9 References

- [1] M. Baker. “1,500 scientists lift the lid on reproducibility”. In: *Nature* 533.7604 (2016), pp. 452–454. DOI: 10.1038/533452a.
- [2] J. W. Scannell et al. “Diagnosing the decline in pharmaceutical R&D efficiency”. In: *Nature Reviews Drug Discovery* 11.3 (2012), pp. 191–200. DOI: 10.1038/nrd3681.
- [3] T. T. Ashburn and K. B. Thor. “Drug repositioning: identifying and developing new uses for existing drugs”. In: *Nature Reviews Drug Discovery* 3.8 (2004), pp. 673–683. DOI: 10.1038/nrd1468.
- [4] S. H. Sleight and C. L. Barton. “Repurposing Strategies for Therapeutics”. In: *Pharmaceutical Medicine* 24.3 (2010), pp. 151–159. DOI: 10.1007/bf03256811.
- [5] T. Barrett and R. Edgar. “[19] Gene Expression Omnibus: Microarray Data Storage, Submission, Retrieval, and Analysis”. In: *Methods in Enzymology*. Vol. 411. Academic Press, 2006, pp. 352–369. DOI: [https://doi.org/10.1016/S0076-6879\(06\)11019-8](https://doi.org/10.1016/S0076-6879(06)11019-8).
- [6] J. Lamb et al. “The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease”. In: *Science* 313.5795 (2006), pp. 1929–1935. DOI: 10.1126/science.1132939.
- [7] M. Sirota et al. “Discovery and Preclinical Validation of Drug Indications Using Compendia of Public Gene Expression Data”. In: *Science Translational Medicine* 3.96 (2011), 96ra77–96ra77. DOI: 10.1126/scitranslmed.3001318.
- [8] F. Cheng et al. “Prediction of Drug-Target Interactions and Drug Repositioning via Network-Based Inference”. In: *PLOS Computational Biology* 8.5 (2012), e1002503. DOI: 10.1371/journal.pcbi.1002503.
- [9] A. Gottlieb et al. “PREDICT: a method for inferring novel drug indications with application to personalized medicine”. In: *Molecular Systems Biology* 7.1 (2011), p. 496. DOI: 10.1038/msb.2011.26.

- [10] R. Hoehndorf et al. "Mouse model phenotypes provide information about human drug targets". In: *Bioinformatics* 30.5 (2013), pp. 719–725. DOI: 10.1093/bioinformatics/btt613.
- [11] C. Wu et al. "Computational drug repositioning through heterogeneous network clustering". In: *BMC systems biology* 7 Suppl 5.Suppl 5 (2013), S6–S6. DOI: 10.1186/1752-0509-7-S5-S6.
- [12] H. Bisgin et al. "A phenome-guided drug repositioning through a latent variable model". In: *BMC bioinformatics* 15.1 (2014), pp. 267–267. DOI: 10.1186/1471-2105-15-267.
- [13] M. D. Wilkinson et al. "The FAIR Guiding Principles for scientific data management and stewardship". In: *Nature* 3 (2016), p. 160018. DOI: 10.1038/sdata.2016.18.
- [14] S. Collins et al. "Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data". In: (2018). DOI: 10.2777/15242.
- [15] S. Borgo and C. Masolo. "Ontological Foundations of dolce". In: *Theory and Applications of Ontology: Computer Applications*. Ed. by R. Poli, M. Healy, and A. Kameas. Dordrecht: Springer Netherlands, 2010, pp. 279–295. DOI: 10.1007/978-90-481-8847-5_13.
- [16] N. A. Vasilevsky et al. "On the reproducibility of science: unique identification of research resources in the biomedical literature". In: *PeerJ* 1 (2013), e148–e148. DOI: 10.7717/peerj.148.
- [17] J. P. A. Ioannidis. "Contradicted and Initially Stronger Effects in Highly Cited Clinical Research". In: *JAMA* 294.2 (2005), pp. 218–228. DOI: 10.1001/jama.294.2.218.
- [18] F. Prinz, T. Schlange, and K. Asadullah. "Believe it or not: how much can we rely on published data on potential drug targets?" In: *Nature Reviews Drug Discovery* 10.9 (2011), pp. 712–712. DOI: 10.1038/nrd3439-c1.

-
- [19] C. G. Begley and L. M. Ellis. “Drug development: Raise standards for preclinical cancer research”. In: *Nature* 483.7391 (2012), pp. 531–533. DOI: 10.1038/483531a.
 - [20] R. A. Klein et al. “Investigating variation in replicability: A “many labs” replication project”. In: *Social Psychology* 45.3 (2014), pp. 142–152. DOI: 10.1027/1864-9335/a000178.
 - [21] J. F. Pimentel et al. “A large-scale study about quality and reproducibility of jupyter notebooks”. In: *Proceedings of the 16th International Conference on Mining Software Repositories*. IEEE Press. 2019, pp. 507–517.
 - [22] CrowdFlower. *Data Science Report*. Report. accessed on 09 October 2019. 2016.
 - [23] K. Hettne et al. “Best practices for workflow design: How to prevent workflow decay”. In: *CEUR Workshop Proceedings* 952 (2012).
 - [24] C. W. Language (CWL). *Common Workflow Language*.
 - [25] K. Belhajjame et al. “Using a suite of ontologies for preserving workflow-centric research objects”. In: *Journal of Web Semantics* 32 (2015), pp. 16–42. DOI: <https://doi.org/10.1016/j.websem.2015.01.003>.
 - [26] S. Cohen-Boulakia et al. “Scientific workflows for computational reproducibility in the life sciences: Status, challenges and opportunities”. In: *Future Generation Computer Systems* 75 (2017), pp. 284–298. DOI: <https://doi.org/10.1016/j.future.2017.01.012>.
 - [27] F. Z. Khan et al. “Sharing interoperable workflow provenance : A review of best practices and their practical application in CWLProv”. In: *GigaScience* (2019), pp. 1–26. DOI: 10.5281/zenodo.1966881.
 - [28] T. Lebo et al. “Prov-o: The prov ontology”. In: *W3C recommendation* 30 (2013).
 - [29] L. Moreau et al. “The open provenance model: An overview”. In: *International Provenance and Annotation Workshop*. Springer. 2008, pp. 323–326.

- [30] D. Garijo and Y. Gil. “Augmenting PROV with Plans in P-PLAN: Scientific Processes as Linked Data”. In: *LISC@ISWC*. 2012.
- [31] S. Soiland-Reyes et al. “Capturing interoperable reproducible workflows”. In: *Workshop on Research Objects: Workshop at IEEE eScience 2018*. 2018.
- [32] S. Samuel and B. König-Ries. “Combining P-Plan and the REPRODUCE-ME ontology to achieve semantic enrichment of scientific experiments using interactive notebooks”. In: *European Semantic Web Conference*. Springer. 2018, pp. 126–130.
- [33] S. Samuel and B. König-Ries. “ProvBook: Provenance-based Semantic Enrichment of Interactive Notebooks for Reproducibility.” In: *In Proceedings of the ISWC 2018 Posters & Demonstrations, Industry and Blue Sky Ideas Tracks co-located with ISWC*. 2018.
- [34] G. Correa Publio et al. “ML-Schema: Exposing the Semantics of Machine Learning with Schemas and Ontologies”. In: *Reproducibility in Machine Learning Workshop, ICML*. 2018.
- [35] H. A. Hartanto, R. Sarno, and N. F. Ariyani. “Warning criterion ontology for measuring of compliance in standard operating procedure implementation”. In: 2017.
- [36] M. Rospocher, C. Ghidini, and L. Serafini. “An ontology for the Business Process Modelling Notation”. In: *FOIS*. 2014.
- [37] G. A. L. F. C. O. Giraldo Olga. “Using semantics for representing experimental protocols”. In: *Journal of Biomedical Semantics* (2017).
- [38] M. Imming et al. *FAIR Data Advanced Use Cases: from principles to practice in the Netherlands*. 2018.
- [39] C. H. Neil and K. Daniel S. *FAIR enough? Can we (already) benefit from applying the FAIR data principles to software?* 2018. DOI: 10.6084/m9.figshare.7449239.v2.
- [40] A.-L. Lamprecht et al. “Towards FAIR principles for research software”. In: *Data Science* (Nov. 2019). Ed. by P. Groth, pp. 1–23. DOI: 10.3233/DS-190026.
- [41] A. Jacobsen et al. “A generic workflow for the data FAIRification process”. In: *Data Intelligence* (2019), pp. 56–65.

-
- [42] O. Lassila, R. R. Swick, et al. "Resource description framework (RDF) model and syntax specification". In: (1998).
 - [43] G. Guizzardi et al. "Towards ontological foundations for conceptual modeling: The unified foundational ontology (UFO) story". In: *Applied Ontology* 10.3-4 (2015), pp. 259–271. DOI: 10.3233/AO-150157.
 - [44] J. L. R. Moreira et al. "Menthor Editor: An Ontology-Driven Conceptual Modeling Platform". In: *JOWO@FOIS*. 2016.
 - [45] C. Bizer, T. Heath, and T. Berners-Lee. "Linked data-the story so far". In: *International journal on semantic web and information systems* 5.3 (2009), pp. 1–22.
 - [46] D. S. Wishart et al. "DrugBank: a knowledgebase for drugs, drug actions and drug targets". In: *Nucleic acids research* 36.suppl_1 (2008), pp. D901–D906.
 - [47] M. Kanehisa et al. "KEGG for linking genomes to life and the environment". en. In: *Nucleic Acids Research* 36.Database (Dec. 2007), pp. D480–D484. DOI: 10.1093/nar/gkm882.
 - [48] M. Kuhn et al. "A side effect resource to capture phenotypic effects of drugs". In: *Molecular systems biology* 6.1 (2010).
 - [49] K. A. Gray et al. "Genenames. org: the HGNC resources in 2015". In: *Nucleic acids research* 43.D1 (2015), pp. D1079–D1085.
 - [50] D. Barrell et al. "The GOA database in 2009—an integrated Gene Ontology Annotation resource". In: *Nucleic acids research* 37.suppl_1 (2009), pp. D396–D403.
 - [51] A. Callahan, J. Cruz-Toledo, and M. Dumontier. "Ontology-based querying with Bio2RDF's linked open data". In: *Journal of biomedical semantics*. Vol. 4. 1. BioMed Central. 2013, S1.
 - [52] J. Menche et al. "Uncovering disease-disease relationships through the incomplete interactome". In: *Science* 347.6224 (2015), p. 1257601.
 - [53] H. Caniza, A. E. Romero, and A. Paccanaro. "A network medicine approach to quantify distance between hereditary disease modules on the interactome". In: *Scientific reports* 5 (2015), p. 17658.

- [54] N. F. Noy et al. "BioPortal: ontologies and integrated data resources at the click of a mouse". en. In: *Nucleic Acids Research* 37.Web Server (July 2009), W170–W173. DOI: 10.1093/nar/gkp440.
- [55] S. M. S. Da Cruz, M. L. M. Campos, and M. Mattoso. "A foundational ontology to support scientific experiments". In: *CEUR Workshop Proceedings* 938.2001 (2012), pp. 144–155.
- [56] M. Rosemann and J. vom Brocke. "The Six Core Elements of Business Process Management". In: *Handbook on Business Process Management 1: Introduction, Methods, and Information Systems*. Ed. by J. vom Brocke and M. Rosemann. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015, pp. 105–122. DOI: 10.1007/978-3-642-45100-3_5.
- [57] B. Stephan, B. Thomas, and R. Manfred. "Bridging the Gap between Business Process Models and Service Composition Specifications". In: *Service Life Cycle Tools and Technologies: Methods, Trends and Advances*. Ed. by L. Jonathan, M. Shang-Pin, and L. Alan. Hershey, PA, USA: IGI Global, 2012, pp. 124–153. DOI: 10.4018/978-1-61350-159-7.ch007.
- [58] A. Krishna, P. Poizat, and G. Salaün. "Checking business process evolution". In: *Science of Computer Programming* 170 (2019), pp. 1–26. DOI: <https://doi.org/10.1016/j.scico.2018.09.007>.
- [59] Y. Ren et al. "Process Refinement Validation and Explanation with Ontology Reasoning". In: *Service-Oriented Computing*. Springer Berlin Heidelberg, 2013, pp. 515–523.
- [60] Y. Jiang et al. "A novel flexible activity refinement approach for improving workflow process flexibility". In: *Computers in Industry* 80 (2016), pp. 1–15. DOI: <https://doi.org/10.1016/j.compind.2016.03.002>.
- [61] M. z. Muehlen and M. Rosemann. "Multi-Paradigm Process Management". In: *CAiSE Workshops*. 2004, pp. 169–175.

-
- [62] J. Horkoff et al. "Goal-oriented requirements engineering: an extended systematic mapping study". In: *Requirements Engineering* 24.2 (2019), pp. 133–160. DOI: 10.1007/s00766-017-0280-z.
- [63] R. S. Gonçalves et al. "The CEDAR Workbench: An Ontology-Assisted Environment for Authoring Metadata that Describe Scientific Experiments". In: *The Semantic Web – ISWC 2017* (2017), 103–110. DOI: 10.1007/978-3-319-68204-4_10.

7

Un-biased housekeeping gene selection for high-validity expression analysis

7

Adapted from: Ana I. Casas, **Ahmed A. Hassan**, Quirin Manz, Christian Wiwie, Pamela Kleikers, Javier Egea, Manuela G. López, Markus List, Jan Baumbach, Harald H.H.W. Schmidt. “Un-biased housekeeping gene selection for high-validity expression analysis” – *Submitted to PLOS Biology*.

Abstract

Differential gene expression, normalized to a single housekeeping (HK) gene, is widely used to identify (patho)physiological mechanisms. HK gene selection is, however, often biased, based on previous literature, thereby introducing risks of systematic errors and discordant results. Here we examine these risks in a brain hypoxia model reproducing the established upregulation of the *Nox4* gene and pharmacologic target. To identify *published* HK genes in models of brain hypoxia, with and without pharmacological intervention, we conducted a systematic literature review. The top eight published HK genes were validated by real-time quantitative PCR, but their majority showed unacceptable expression variance and were rejected. The remaining HK genes were ranked according to statistical stability. When normalizing *Nox4* expression with either the most stable or any of the rejected HK genes we obtained even qualitatively opposing results. To overcome literature-based bias and the need for systematic reviews and stability analysis for every single research question, we developed the HouseKeepR algorithm for *de novo* HK gene identification. Here, all publicly available gene expression datasets relevant to a given disease model are merged and automatically ranked based on stability. Many of the resulting top ten genes were hitherto not routinely used in brain hypoxia analysis. Importantly, they all reproduced the relative *Nox4* induction. For broad applicability by the research community, HouseKeepR is now implemented as an open-access interactive web application to select HK genes in the future in an automated, unbiased and tissue/disease-specific manner for high validity analysis of differential expression.

Author Summary

In drug discovery, disease genes are often identified by differential expression, e.g. in comparison to healthy conditions. In this process, data normalization to a stable, so-called housekeeping gene is, of course, essential but surprisingly arbitrary or biased. We show for an important target in ischemic stroke that this can yield variable or even invalid results. Instead, we validate and offer to the research community a widely applicable, open-access solution. HouseKeepR is a user-friendly web application that selects suitable and stable housekeeping genes from in an automated, unbiased as well as tissue and disease-specific manner.

7.1 Introduction

Gene expression analysis is a fundamental tool in biomedical sciences to understand cellular regulation [1], define diagnostics strategies, and identify targets for drug discovery and therapeutic intervention [2, 3]. Due to its high sensitivity and specificity, real-time quantitative PCR (RT-qPCR) is the gold standard and most widely used technology to assess mRNA levels [4]. However, to be able to make any statement regarding the differential expression of a gene (control vs. treatment), normalization using a stable housekeeping gene (HK) is necessary. Such reference genes are conventionally defined as those which are stable and consistently expressed across all cells and conditions [5]. To this end, it is common practice to select genes related to basal cell metabolism as their expression is thought to fulfill these stability requirements [5]. Yet, the expression levels of several standard HK genes vary strongly in different tissues, cell types, disease models and therapeutic conditions [5, 6]. Since the validity of gene expression analysis and hypotheses on disease mechanisms entirely depend on the selection of appropriate HK genes [7], diametrically opposing results can be obtained if data normalization is performed using non-stable HK genes [8, 9]. In the worst scenario, such results could mislead subsequent drug discovery. Moreover, drugs are able to modulate HK gene expression [10] and thereby affect the analysis of potential therapeutic effects.

To ensure a suitable HK gene selection process, several approaches have been suggested. To rank stably expressed genes based on pair-wise comparisons, the most basic approach is the comparative delta-Ct method [11]. Other strategies use expression variance of either individual, a subset of genes, or all genes represented as a co-variance matrix [12–14]. These are based on the assumption of a constant distribution of all samples which, however, is not always the case and may lead to the removal of desired variations. Several studies have identified HK genes under different conditions using microarray expression data [15–17]. The huge number of publicly available data sets may indeed increase the confidence in HK gene selection and resolve the preselection bias problem. Unfortunately, only a handful of these studies

have developed tools to automate their approaches and make them available for future access [13, 18], where RefGenes [13] and NormFinder [14] are two clear examples. Both tools have significant limitations as they do not scale well for usage in genome-wide search for de novo HK gene candidates. We here address these unmet needs in existing HK gene selection by developing and validating a microarray-based, bias-free approach in a user-friendly manner.

7.2 Results

7.2.1 Systematic review identifying HK in models of brain hypoxia

Housekeeping gene selection for gene expression normalization remains arbitrary and is primarily focused on literature review or previously performed experiments [19]. In a systematic literature review, we identified the most frequently used HK genes in ischemic stroke in two broadly used ex vivo brain ischemia models: (i) acute Hippocampal Brain Slices (HBS), and (ii) Organotypic Hippocampal Culture (OHCs). Literature screening led to 38 and 16 original articles respectively after considering our exclusion criteria (Fig. 7.S1). We identified 6 potential HK genes considering β -actin and *Gapdh* as the preferred ones for the selected models (Table 7.1). We further included two additional HK genes previously suggested as the most suitable for *in vivo* stroke models, i.e. *Sdha* and *Ywhaz* [20].

Table 7.1: Housekeeping genes used for in vitro gene expression determination in brain ischemia models

Hippocampal Brain Slices – Acute model								
Published	Excluded	Included	Housekeeping genes					
364	353	11	Hprt 1	β -actin 4	18S 1	Gapdh 5		
Organotypic Hippocampal Culture – Chronic model								
Published	Excluded	Included	Housekeeping genes					
249	222	27	β -actin 9	Hprt 1	Gapdh 12	18S 2	β 2-mg 2	Rpl13 1

7.2.2 The majority of previously identified HK genes show unacceptable expression variance

Gene expression normalization plays a key role in target identification, therapy design through RT-qPCR, microarray and RNA-seq analysis [21], besides being essential for obtaining reliable conclusions [22]. Therefore, we validated previously selected HK genes using three widely used brain ischemia models in two different species. First, as an acute *ex-vivo* ischemia model, rat hippocampal brain slices (HBS) were subjected to oxygen and glucose deprivation (OGD) for 15 min followed by 2h of the re-oxygenation period (Re-Ox) including different pharmacological interventions (Fig. 7.1A). Upon RT-qPCR analysis (Table 7.S1), 5 out of 8 previously defined candidate genes significantly varied under these experimental conditions and were considered unsuitable for this model. In contrast, *Gapdh*, *Hprt* and *Rpl13* remained stable after RT-qPCR validation (Fig. 7.1B). Ideally, HK genes are consistently expressed without being influenced by any condition or pharmacological intervention [23]. To assess whether previously detected gene instability is reproducible, we used an additional chronic ischemia model. Organotypic Hippocampal Cultures (OHCs) were subjected to 15min of OGD followed by 24h of re-oxygenation in presence of different pharmacological treatments (Fig. 7.1C). *Ywhaz* and *18S* remained unstable but surprisingly, the broadly used HK gene *Gapdh*, was also discarded for this model (Fig. 7.1D).

To further corroborate our findings, we compared the *ex vivo* models with the most physiologically relevant in vivo stroke model, the occlusion of the middle cerebral artery (MCAO). Adult mice were subjected to a 1h-transient MCAO subsequently treated post-stroke (Fig. 7.1E). In line with previous findings, *Ywhaz* and *18S* remained statistically variable while β -actin, one of the preferred HK genes in the biomedical field, was unstable (Fig. 7.1F). We demonstrated that even in similar experimental conditions, HK genes are extremely variable across different species, pre-clinical models, and pharmacological interventions. Therefore, HK genes validated under specific conditions can indeed not be directly translated to similar experimental scenarios.

7.2.3 Contradicting results comparing stable and rejected HK genes

All scientific statements derived from poor normalization could lead to wrong conclusions [9]. To experimentally validate our literature-based HK gene selection strategy, we considered the established upregulation of *NADPH oxidase 4 (Nox4)* upon stroke. *Nox4* is one of the most relevant therapeutic targets for translational drug therapy in brain ischemia, currently under clinical testing (REPO-STROKE Phase I). Upon ischemia, *Nox4* is broadly induced in different cell types with a unique role in the brain [24]. Hence, neurovascular deletion of *Nox4* leads to direct neuroprotection, blood-brain barrier stability, and improved neuro-motor functioning post-stroke [25, 26]. We normalized *Nox4* gene expression using three HK gene candidates that were previously rejected in all experimental models, i.e. (i) *Yhawz*, (ii) β -actin, and (iii) *18S*, as well as the most suitable candidates already identified for this model, i.e. (i) *Hprt*, (ii) *Gapdh*, and (iii) *Sdha*. As expected, *Sdha* normalization showed a significant *Nox4* upregulation post-ischemia which was completely prevented through direct pharmacological intervention. In contrast, *Yhawz* normalization resulted in opposite results contradicting previous literature (Fig 7.2). Hence, identical therapeutic conditions could be interpreted as both, clearly neuroprotective or absolutely ineffective based on the selected normalization process, and therefore potentially lead to poor therapy identification.

7.2.4 HouseKeepR allows for robust *de novo* housekeeping gene identification

Our literature-based HK identification approach shows that all potential HK genes need to be experimentally validated per model, tissue and condition, leading to an extreme experimental burden. To overcome the lack of stability in gene expression normalization we developed HouseKeepR, a web tool that robustly ranks HK gene candidates across gene expression data sets (Fig 7.3). In contrast to existing approaches, genes are not only ranked by variance but also for consistently high average expression for each data set individually. To ensure that results are robust towards sampling bias, HouseKeepR uses a bootstrapping strategy to obtain a distribution, expected (mean) rank,

and rank variance for each gene. HouseKeepR can be applied to user-specified gene expression data sets which are automatically retrieved from the public Gene Expression Omnibus (GEO) database [27].

7.2.5 HouseKeepR is easily accessible through a user-friendly web interface

For broad applicability, we have developed a user-friendly web interface with integrated search and sample annotation functions (<https://exbio.wzw.tum.de/housekeepr>). HouseKeepR requires the following parameters as input: tissue, condition and organism. Users can choose a few optional parameters such as the number of final HK gene candidates, and the number of bootstrap replications (defaults to 10,000 repetitions). HouseKeepR searches GEO and reports all data sets matching the selected parameters [27]. Users subsequently need to choose at least two data sets and annotate condition and control samples to perform the analysis. After starting the analysis, the server will download the data sets, preprocess them, and run the HouseKeepR algorithm. Finally, the results are displayed as a ranking table (Fig 7.S2), and a rank distribution across data sets (Fig 7.S3). The run time depends on the number of the data set and on the response time of the GEO database. Data are retrieved from GEO and Ensembl, which are subjected to changes and may affect the reproducibility of results. To address this issue, HouseKeepR allows users to save sessions, which persists results and the parameters used in the analysis for later retrieval. Furthermore, users can optionally select older Ensembl versions from the Ensembl archive.

HouseKeepR enables researchers to quickly generate hypotheses about candidate housekeeping genes for their experiments and allows for meta-analyses for any tissue, condition and organism with a few clicks, grasping back to a large and continuously growing number of data sets available in GEO.

7.2.6 HK genes candidates are robust to data set sampling and across non-overlapping data sets

Through bootstrapping, HouseKeepR allows studying the stability of the results under different random samplings from the included data sets. As additional validation of consistency, we executed HouseKeepR 10 times with the same data sets and bootstrap parameters but using different random seeds. In this context, we computed the reproducibility measure (R) as the average of all pairwise overlap coefficients “A comparative analysis of similarity measures akin to the Jaccard index in collaborative recommendations: empirical and theoretical perspective” calculated between each pair of results lists (with the top 50 HK genes) produced by each run. The ten runs achieved an R value >0.99 , which demonstrates the robustness of our algorithm.

Another important measure for HouseKeepR is the reproducibility of HK gene candidates when comparing different non-overlapping data sets for the same tissue, organism and condition. We selected two non-overlapping sets of 12 and two data sets, respectively. The top 20 HK gene candidates of both groups were compared, yielding an overlap coefficient of 0.5 (p-value = $9.75e-29$) (Fig 7.S4) [29]. The significant overlap indicates that HouseKeepR successfully predicts stable HK genes across different expression platforms and experimental setups. As a negative control, we ran HouseKeepR on a group of two data sets of different organisms and conditions. As expected, the overlap coefficient of the top 20 HK genes to the previously selected ones was 0 (Fig 7.S4), confirming that predictions are condition- and organism-dependent.

7.2.7 NormFinder confirms the stability of HouseKeepR candidate genes

We next evaluated the results of HouseKeepR with a complementary computational approach. NormFinder employs a more elaborate statistical model to test gene stability which is computationally expensive and thus not suited for systematic screening across several data sets (Fig 7.S5). Stable genes were defined using NormFinder’s stability score with a cut-off value of 0.15 [30]). 80% of the top 10 HK genes

(Fig 7.S6) generated by HouseKeepR were confirmed to be stable by NormFinder, highlighting that HouseKeepR reported genes are also considered robust by independent statistical evaluation.

7.2.8 HouseKeepR suggested HK genes are consistent in *Nox4* target validation

For an in vitro validation of HouseKeepR, we selected 12 data sets (Table 7.S3) specific to ischemic stroke in rats and mice for identifying condition-specific de-novo HK gene candidates (see Methods for details). Using previously selected data sets, HouseKeepR calculated an overall ranking of 19,878 genes over 10,000 bootstrap samples. We then assessed *Nox4* expression normalized against a panel of the top 10 HK candidates, after excluding mitochondrial genes (Fig 7.4A-B). Through individual normalization using each of the top 10 in silico predicted candidates, *Nox4* gene expression remained unchanged under different experimental conditions (Fig 7.4C). Similarly, while normalizing against the top 2 genes from the panel, *Nox4* expression was increased upon hypoxia although no effect was detected under pharmacological treatment. However, using normalization panels of 4, 6, 8 and 10 previously predicted genes showed stable and reliable gene expression data (Fig 7.4D). Hence, we demonstrate that HouseKeepR provides tissue- and condition-specific HK gene panels that lead to accurate, reliable and stable results under different experimental conditions.

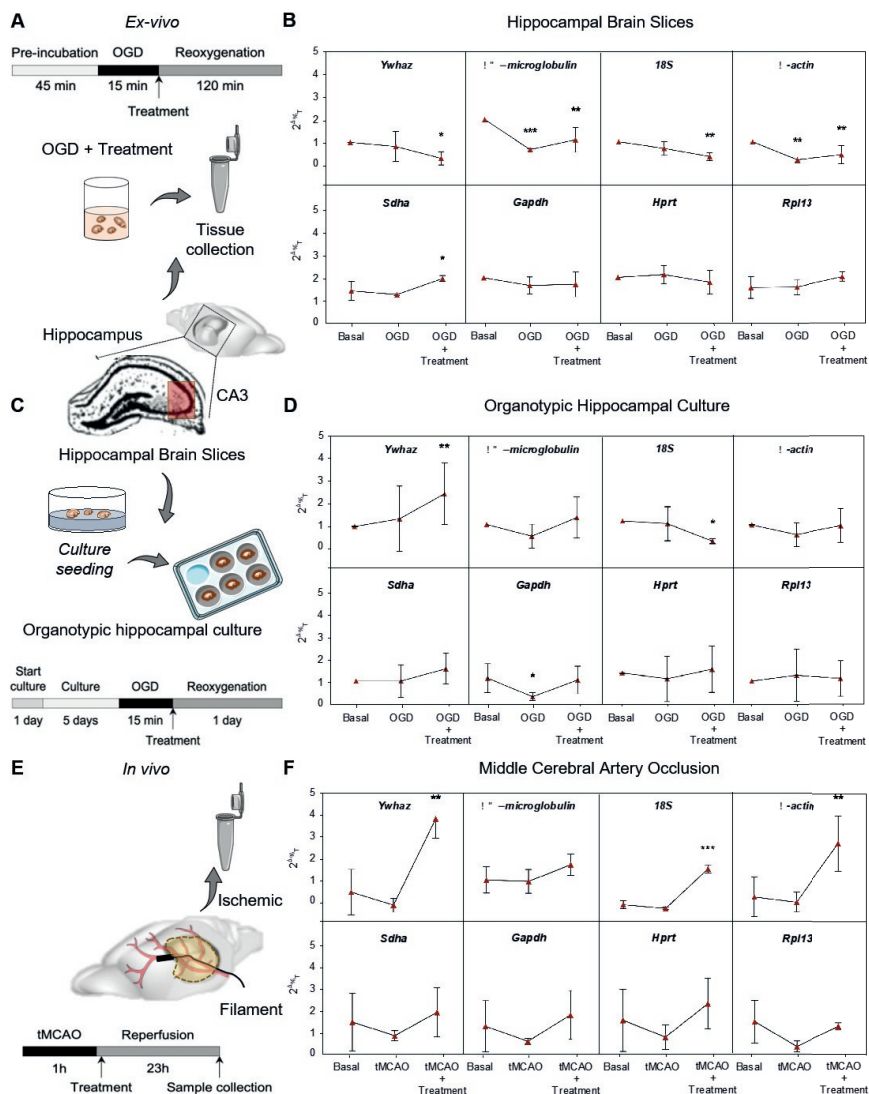


Figure 7.1: Experimental validation of eight housekeeping genes in three different brain ischemia models. (A) Rat hippocampal brain slices were stabilized for 45 min followed by 15 min of oxygen and glucose deprivation (OGD) and 2h of re-oxygenation (Re-Ox). (B) Gapdh, Hprt and Rpl13 gene expression remain stable while Ywhaz, 18S and Sdha are significantly up- or down-regulated after pharmacological intervention. β 2-microglobulin and β -actin showed instability in all hypoxic conditions (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, $n = 5$).

Figure 7.1 (cont.): (C) After 5 days of culture stabilization, organotypic hippocampal cultures (OHCs) were exposed to 15 min of OGD followed by 24h of Re-Ox. (D) While β 2-microglobulin, β -actin, Sdha, Hprt and Rpl13 expression remains stable under all experimental conditions, Ywhaz, Gapdh and 18S resulted unsuitable housekeeping genes for this model. (* $p < 0.05$, ** $p < 0.01$, $n = 5$). (E) Adult WT mice were subjected to a 1h-transient occlusion of the middle cerebral artery (tMCAO) followed by 23h of reperfusion. (F) β 2-microglobulin, Sdha, Gapdh, Hprt and Rpl13 gene expression remained stable under all interventions while Ywhaz, 18S and β -actin are significantly up-regulated due to treatment (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, $n = 5$).

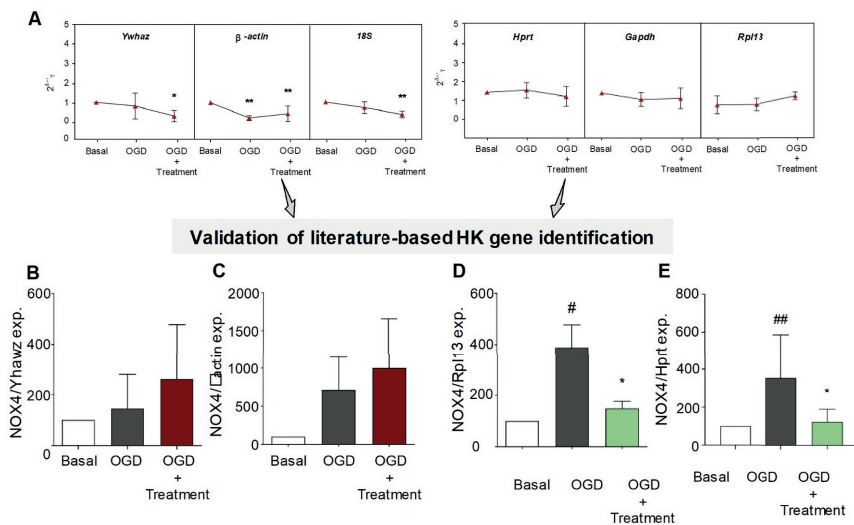


Figure 7.2: Opposing gene expression results based on specific housekeeping gene selection. (A) Using the OHCs in vitro model, RT-qPCR validation resulted in 3 unstable genes (Yhawz, 18S and Gapdh) and 5 promising housekeeping candidates (Sdha, β 2-microglobulin, Rpl3, β -actin and Hprt). Based on a NormFinder gene stability analysis, Sdha resulted the best candidate for this model. (B) Tissue from OHCs subjected to 3 different experimental conditions, i.e. Basal, OGD and Treatment was collected for further NADPH oxidase 4 (Nox4) expression assessment. Nox4 gene expression was normalized against (i) previously detected unstable genes (Yhawz), and (ii) most stable candidate (Sdha). Opposing results were identified depending on the selected gene. (# $p < 0.05$, $n = 5$) later prevented by 2h of pharmacological treatment (* $p < 0.05$, $n = 5$).

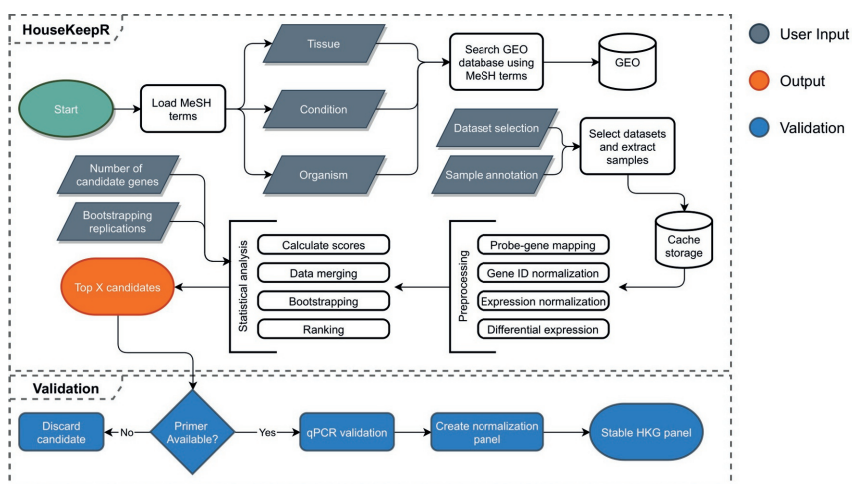


Figure 7.3: HouseKeepR algorithm flow chart. First, MeSH terms are parsed and preloaded to be used as standardized search terms for expressions datasets. User inputs for tissue, condition and organism, encoded in MeSH terms, are used to search and download datasets from Gene Expression Omnibus (GEO). The datasets are then pre-processed and statistically analyzed using a score formula engineered to find housekeeping genes. The top ranking genes, based on user choice, are selected to be validated, where the availability of the primers for performing quantitative polymerase chain reaction (RT-qPCR) plays a role in the final selection. Gray colored elements show user required input, while the orange elements are the outputs the user, and the blue elements are external validation steps conducted via wet lab experiments.

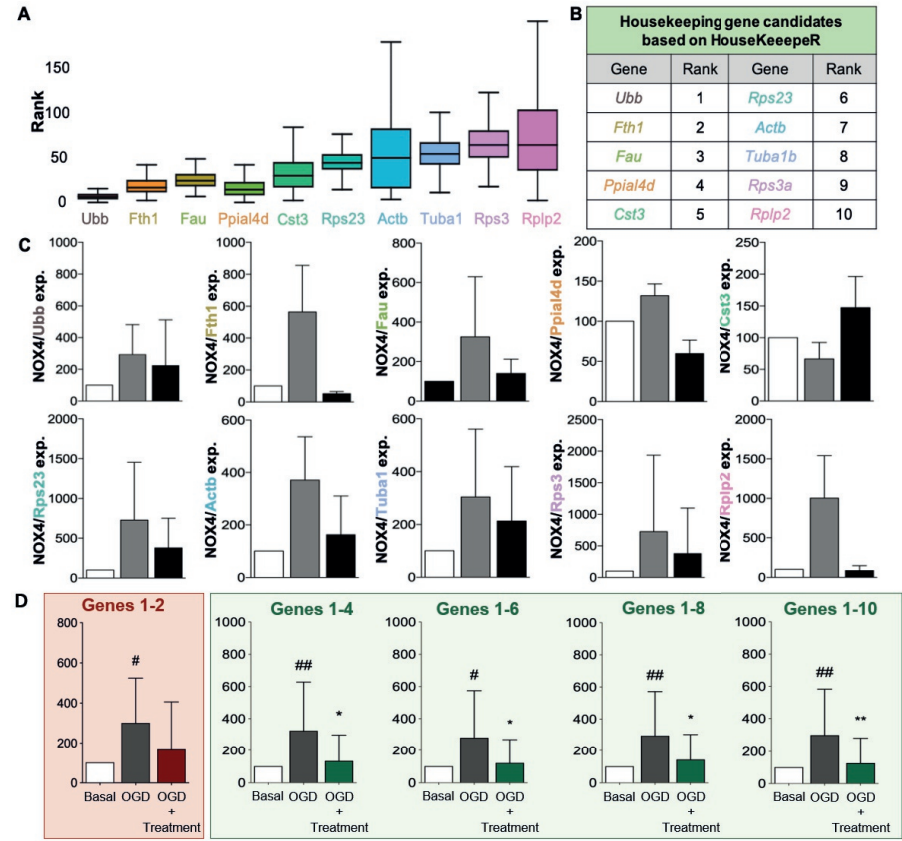


Figure 7.4: Un-biased de novo housekeeping gene panel generation and validation. (A) Ranking distribution of the top 10 housekeeping gene candidates generated over 10,000 bootstrap runs over all the samples in the 12 datasets selected for analysis. The order of the plot follows the final ranking in B. (B) The final ranking of the top 10 housekeeping gene candidates: Ubb, Fth1, Fau, Ppia14d, Cst3, Rps23, Actb, Tuba1, Rps3, Rplp2. This ranking is calculated by taking the mean bootstrap rank from A and penalizing it for the absence of genes from each of the 12 datasets selected for analysis.

Figure 7.4 (cont.): (C) Nox4 gene expression was normalized against the top 10 genes of the ranking under 3 different experimental conditions, i.e. Basal, oxygen and glucose deprivation (OGD) and pharmacologically treated hippocampal tissue post-OGD. (D) 5 housekeeping genes panels were generated and validated: (i) Panel 1; 2 genes (Ubb and Fth1); (ii) Panel 2, 4 genes (Ubb, Fth1, Fau and Ppial4d); (iii) Panel 3, 6 genes (Ubb, Fth1, Fau, Ppial4d, Cst3 and Rps23); (iv) Panel 4, 8 genes (Ubb, Fth1, Fau, Ppial4d, Cst3, Rps23, Actb and Tuba 1); (v) Panel 5, 10 genes (Ubb, Fth1, Fau, Ppial4d, Cst3, Rps23, Actb, Tuba 1, Rps3 and Rplp2). Panel 2, 3, 4 and 5 showed stable Nox4 gene expression results while Panel 1 remained statistically not significant. (#p < 0.05, n = 6; ##p < 0.01, n = 6; *p < 0.05, n = 6; **p < 0.01, n = 6).

7.3 Discussion

The choice of a HK gene is critical for the interpretation of gene expression studies and can lead to false conclusions [31, 32]. Here we demonstrate for both ex-vivo and in-vivo models of brain ischemia that literature-derived HK genes display considerable variance that would result in opposing conclusions whether to consider Nox4 as a drug target [24, 26] or not. Our results suggest that the most widely used methods for HK gene selection, i.e., literature-based and hence biased, may lead to reproducibility issues as part of the quality crisis in biomedical research [33]. Thus, a systematic, unbiased method for HK gene selection is evidently needed. Moreover, for practical reasons this method should be tissue- and condition-specific and user-friendly so that it will be widely adopted by biomedical scientists without affording programming and coding skills. We have examined methods that select HK based on gene expression data and found limitations that hinder their adoption. RefGenes, for example, which is part of the free version of the Genevestigator (<https://genevestigator.com>) platform, relies on the standard deviation of HK genes to define housekeeping gene candidates. However, by focusing on a single gene expression data set, it neglects variations between conditions, which can be misleading in cases of low signal-to-noise ratio. While RefGenes offers a graphical user interface, it does not support exporting of results for downstream analysis. NormFinder which is available as Excel and R script, is another widely adopted method for validating HK genes in RT-qPCR data. In contrast to RefGenes, it considers not only the variance within a group of samples but also between groups, allowing for analysis across data sets, organisms, or conditions. While NormFinder is in principle applicable to microarray expression data, it is typically used for the validation of a set of candidate genes and not suited for a genome-wide search for *de novo* HK gene candidates, where the runtime of the NormFinder algorithm becomes the limiting factor. Since existing tools do not satisfy one or more of the aforementioned criteria, we developed HouseKeepR, a web application for *de novo* HK gene discovery that fulfills, to our knowledge for the first time, all these criteria.

HouseKeepR suggests HK gene candidates that are highly repro-

ducible both in silico and in vitro. Using our model system and target gene, Nox4, expression changes could be detected with single HK genes; the more subtle response of Nox4 expression to pharmacological treatment was only detectable when using a panel of HK genes for normalization. Thus, unbiased panel-based normalization with condition-specific and cross-data set validated HK genes should become the future gold standard.

HouseKeepR is currently limited to microarray data through the GEO database. As RNA-seq data will become more prevalent, we plan to incorporate additional data sources such as ARCHS4 [34] or GEMMA [35]. RNA-seq may in fact eventually supersede qRT-PCR based gene expression analysis as the gold standard. However, HK gene-based normalization will also need to be applied in this context [36, 37] and the need for unbiased and robust HK gene panels remains just as relevant.

In conclusion, with HouseKeepR we enable the wider biomedical research community to discover and use condition-specific HK genes leveraging a large set of publicly available data. This will help to avoid extensive and costly HK gene validation experiments, which would otherwise be necessary for each tissue, condition, model and type of intervention, and ultimately improve the reproducibility of gene expression studies for disease gene and therapeutic target identification.

7.4 Materials and Methods

7.4.1 Systematic review

A literature review focused on two in vitro models of brain ischemia: (i) hippocampal brain slices (HBS), and (ii) organotypic hippocampal culture (OHC) was performed. PubMed was searched for original papers and conference abstracts where these *ex-vivo* models appeared. No terms for RT-qPCR were included since PubMed only screens abstracts, titles and keywords, and RT-qPCR details are frequently mentioned only in the methods section. No language restriction was used. Our search strategy for the HBS model identified 364 records. First, these hits were screened based on title, abstract and results excluding other, non-related *ex-vivo* models. Publications without RT-qPCR experiments or inaccessible full-text, i.e., only title/abstract published were excluded. Finally, 16 articles were included for full text screening. In 1 article, no housekeeping gene was used and in 4, RT-qPCR was conducted using other species. After a full-text assessment, 11 articles were considered. Similarly, our search strategy for model 2 (OHCs) identified 249 records in PubMed. Following first screening, 38 articles were included for full-text screening. 2 of these did not use housekeeping genes and 9 conducted RT-qPCR experiments using tissue from other species. Therefore, in total, 27 articles were considered for the OHC model. Studies were included if (i) the specific *ex-vivo* ischemia model was used; (ii) specific conditions were considered within the experiment; (iii) RT-qPCR experiments were conducted.

7.4.2 Animals

Rats used for *ex-vivo* experiments were handled following the Guide for the Care and Use of Laboratory Animals and were previously approved by the Institutional Ethics Committee of Universidad Autónoma de Madrid, Spain, according to the European guidelines for the use and care of research animals in accordance with the European Union Directive of 22 September 2010 (2010/63/UE) and the Spanish Royal Decree of 1 February 2013 (53/2013). Similarly, *in vivo* experiments in mice strictly followed the Dutch law on animal

experiments and were approved by the local animal experimental committee (Maastricht University DEC2011-106). Both mice and rats were housed under controlled conditions of temperature (22°C), humidity (55–65%), light (12h light-dark cycles) and with free access to water and standard laboratory chow. Male and female C57/Bl6 mice aged 8-12 weeks, Sprague-Dawley adult rats (8-12 weeks) and pups (7-10 days) were used.

7.4.3 Ex-vivo acute model: Preparation of hippocampal brain slices and induction of oxygen and glucose deprivation

Experiments were performed using hippocampal brain slices from adult male Sprague-Dawley rats (8-12 weeks) as previously described in [38, 39]. Briefly, rats were quickly decapitated, and forebrains were rapidly removed from the skull and placed into ice-cold Krebs bicarbonate dissection buffer (pH 7.4), containing (in mM): NaCl 120, KCl 2, CaCl₂ 0.5, NaHCO₃ 26, MgSO₄ 10, KH₂PO₄ 1.18, glucose 11 and sucrose 200. At least 20 minutes before starting the experiment, chamber solutions were bubbled with either 95% O₂/5% CO₂ or 95% N₂/5% CO₂ gas mixtures to ensure O₂ and N₂ saturation respectively. The hippocampus was quickly dissected and subsequently cut into transverse slices of 300 μm thick using a Tissue Chopper Mcllwain. To recover from slicing trauma, slices were incubated in Krebs buffer for 45 min at 34°C (stabilization period). Then, control slices were incubated for a 15 min period in a Krebs-bicarbonate solution without sucrose (control solution). Oxygen and glucose deprivation was induced by incubating the slices during 15 min in a glucose-free Krebs-bicarbonate solution in which glucose was replaced by 2-deoxyglucose (OGD solution). Both solutions were pre-bubbled for 30 minutes with a 95% O₂/5% CO₂ or 95% N₂/5% CO₂ respectively. All experiments were performed at 37°C. Following the OGD period slices were returned back to an oxygenated Krebs-bicarbonate solution containing glucose for 120 min (Re-Ox period). During the Re-Ox period either a pharmacological treatment (OGD + Treatment), or no extra measures (OGD) were taken. After the Re-Ox period, slices were collected and quickly shock-frozen.

7.4.4 Ex-vivo chronic model: Preparation of Organotypic Hippocampal Slices and induction of oxygen and glucose deprivation

Hippocampal brain slices for cultures were obtained from brains of 7- to 10-days-old Sprague-Dawley rats. Organotypic cultures were prepared based on the methods previously described in [25]. Briefly, pups were quickly decapitated, and brains removed from the skull and dissected. The hippocampus was cut into 300 μm -thick slices using a Tissue Chopper Mcllwain. Then, they were separated in sterile ice-cold Hank's balanced salt solution (HBSS, Biowest, Madrid, Spain) containing (in mM): glucose 15, CaCl_2 1.3, KCl 5.36, NaCl 137.93, KH_2PO_4 0.44, Na_2HPO_4 0.34, MgCl_2 0.49, MgSO_4 0.44, NaHCO_3 4.1, HEPES 25, 100 U/ml penicillin, and 0.100 mg/ml gentamicin. Six slices were placed on each Millicell-0.4 μm culture inserts (Millipore, Madrid, Spain) within each well of a six-well culture plate. Specific neurobasal medium (Invitrogen, Madrid, Spain) enriched with 10% of fetal bovine serum (Sigma-Aldrich, Madrid, Spain) was used for the next 24 h (1ml/well). 24 h later, B27 supplement and antioxidants were added to the culture medium. Slices were in culture for 4 d before inducing the OGD period. On day 6, inserts were placed in 1 ml of OGD solution composed of (in mM): NaCl 137.93, KCl 5.36, CaCl_2 2, MgSO_4 1.19, NaHCO_3 26, KH_2PO_4 1.18, and 2-deoxyglucose 11 (Sigma-Aldrich, Madrid, Spain). The cultures were then placed in an airtight chamber (Billups-Rothenberg Inc., USA) and exposed during 5 min to 95% N_2 /5% CO_2 gas flow to ensure oxygen removal. Then, the chamber was sealed for 15 min and placed at 37°C (OGD period). At the same time, control cultures were maintained under a normoxic atmosphere in a solution with the same composition as previously described but containing glucose (15 mM) instead of 2-deoxyglucose. Pharmacological treatment was added to the cultures before returning them to normal oxygen and glucose concentrations for 24 h (reoxygenation period). After the Re-Ox period, slices were collected and quickly shock-frozen.

7.4.5 In-vivo model: Transient Occlusion of the Middle Cerebral Artery (tMCAO) in mice

Stroke surgery was conducted as previously described in [40]. After administration of a painkiller, animals were anesthetized with isoflurane (induction 4-5% in air, maintenance 2-2.5% in air) and placed on a heating-pad which maintains rectal temperature at 37.0°C using a feedback-controlled infrared lamp. Using a surgical microscope (Wild M5A, Wild Heerbrugg, Gais, CH), a midline neck incision was made and both the right common and external carotid arteries were isolated and permanently ligated while a microvascular clip was temporarily placed on the internal carotid artery. A small incision into the common carotid artery was performed where a silicon rubber-coated 6.0 nylon monofilament (602312PK10, Docol Corporation, Sharon, MA, USA) was inserted until a resistance is felt. The tip of the monofilament should be specifically located intracranially at the origin of the right middle cerebral artery and thereby interrupting blood flow. The filament was fixed with a tourniquet suture to prevent dislocation. 1 h after occlusion of the middle cerebral artery, reperfusion was initiated by monofilament removal. Wounds were carefully sutured and animals could recover in a temperature-controlled cupboard. Pharmacological treatment was given via i.p. injections, 2 and 12 h after start of ischemia. Animals were sacrificed 24 h after induction of ischemia by cervical dislocation. Brains were quickly removed and shock-frozen.

7.4.6 RNA extraction, quantification and reverse transcription

Hippocampal brain slices and brain tissue from *ex-vivo* and *in vivo* models were crushed and homogenized using TRI Reagent® (Sigma-Aldrich, The Netherlands). 100 µl of chloroform was added to the samples followed by a 15 min centrifugation at 11.000 rpm and 4°C. After centrifugation, 250 µl of isopropanol was added to the upper phase (mRNA) and then kept during 1 h at -20°C. After incubation, samples were centrifuged during 10 min at 13.000 rpm and 4°C. 200 µl ethanol 80% was added to the supernatant followed by 10 min centrifugation at 13.000 rpm. After ethanol removal, the mRNA was dissolved in RNase free water. mRNA was quantified spectrophotometrically us-

ing the Nanodrop 2000 device. 0.08 μ g of total mRNA was reverse transcribed to cDNA with the High Capacity Reverse Transcription Kit (Applied Biosystems, The Netherlands) according to the manufacturer's protocol.

7.4.7 Real-time PCR

mRNA levels of studied genes were quantified using the fluorescent Taqman[®] technology. We used TaqMan[®] gene expression arrays (TaqMan[®] Universal PCR Master Mix, ThermoFisher Scientific, The Netherlands) for all species. 1) For rat: *β 2-microglobulin* (Rn00560865_1, ThermoFisher Scientific, The Netherlands), *β -actin* (Rn00667869_m1, ThermoFisher Scientific, The Netherlands), *Rpl13* (Rn00821946_m1, ThermoFisher Scientific, The Netherlands), *18S* (Hs99999901_s1, ThermoFisher Scientific, The Netherlands), *Hprt* (Rn01527840_m1, ThermoFisher Scientific, The Netherlands), *Sdha* (Rn00590475_m1, ThermoFisher Scientific, The Netherlands), *Ywhaz* (Rn00755072_m1, ThermoFisher Scientific, The Netherlands), *Gapdh* (Rn01775763_g1, ThermoFisher Scientific, The Netherlands), *Nox4* (Rn01506793_m1, ThermoFisher Scientific, The Netherlands) 2) For mice: *β 2-microglobulin* (Mm00437762_m1, ThermoFisher Scientific, The Netherlands), *β -actin* (Mm02619580_g1, ThermoFisher Scientific, The Netherlands), *Rpl13* (Mm02526700_g1, ThermoFisher Scientific, The Netherlands), *18S* (Hs99999901_s1, ThermoFisher Scientific, The Netherlands), *Hprt* (Mm03024075_m1, ThermoFisher Scientific, The Netherlands), *Sdha* (Mm01352366_m1, ThermoFisher Scientific, The Netherlands), *Ywhaz* (Mm03950126_s1, ThermoFisher Scientific, The Netherlands), *Gapdh* (Mm99999915_g1, ThermoFisher Scientific, The Netherlands), *Ppia4d* (Mm01191872_g1, ThermoFisher Scientific, The Netherlands), *Fth1* (Mm00850707_g1, ThermoFisher Scientific, The Netherlands), *Cst3* (Mm00438347_m1, ThermoFisher Scientific, The Netherlands), *Rps23* (Mm03019701_g1, ThermoFisher Scientific, The Netherlands), *Rplp2* (Mm00782638_s1, ThermoFisher Scientific, The Netherlands), *Ubb* (Mm01622233_g3, ThermoFisher Scientific, The Netherlands), *Fau* (Mm02601595_u1, ThermoFisher Scientific, The Netherlands), *Tuba1* (Mm00846967_g1, ThermoFisher Scientific, The Netherlands), *Rps3* (Mm00656272_m1, ThermoFisher Scientific, The Netherlands) (Supplementary Table 2). Water controls were included

to ensure specificity and the comparative $2^{-\Delta\Delta C_t}$ method was used for relative quantification of gene expression.

7.4.8 NormFinder stability score

The NormFinder algorithm [14] can be used to assess the stability of HK genes across groups based on RT-qPCR measurements. Briefly, NormFinder models the variation within and between sample groups to estimate a HK gene candidate's expression stability in the form of a distribution

$$f_{ig} = z_{ig} - \theta_g - \alpha_i$$

where z_{ig} is the mean, log transformed expression level of the gene i for all samples in group g , θ_g is the average amount of mRNA for group g , and α_i is the mean expression level for gene i over all groups. Hence, the resulting distribution is an additive measure of variance within and between groups. NormFinder transforms the resulting stability distribution to an easier-to-interpret stability value ρ_{ig} by taking the absolute value of the mean + 1 standard deviation:

$$\rho_{ig} = |\text{mean}(f_{ig}) + \text{sd}(f_{ig})|$$

7.4.9 HouseKeepR method

HK genes should also show stable expression levels across tissues and conditions, i.e., a low variance and log fold change are desirable. These characteristics can be captured using the following scoring function:

$$\text{Score} = \frac{\text{MeanExpression}}{\text{FoldChange} \times \text{Variance}}$$

This simple score prefers genes with a high expression which show little difference between conditions and low variance over all samples. An ideal HK gene candidate achieves a high score across various data sets. To compare scores across data sets in spite of batch effects and differences in technical platforms, we thus rank genes within a single dataset and prefer HK genes that show little rank variance across diverse data sets. To avoid that the selected HK genes exhibit high rank stability by chance, we use bootstrapping [41], i.e. sampling with

replacement. As the bootstrap samples will always have a different composition, we can judge the effect of sampling bias on rank stability. Genes with stable ranks over a large number of bootstrap repetitions, e.g. 10,000, are selected as the top performing HK gene candidates.

7.4.10 HouseKeepR R shiny web interface

Existing tools like NormFinder do not offer a user-friendly interface, do not afford genome-wide coverage and do not support the integrative analysis across publicly available gene expression data. This motivated us to make the HouseKeepR method easily accessible to the community via an R shiny web interface which leads users through the process of de-novo detection of suitable HK genes for user-selected tissues and conditions. The basis for this analysis is the Gene Expression Omnibus (GEO) [42], a widely used repository currently offering access to more than a hundred thousand gene expression profiles of more than three million samples. HouseKeepR allows users to directly query GEO for datasets related to specific organisms, conditions and tissues. HouseKeepR shows the result of the query and provides additional information about available studies. Next, users can select studies relevant for de-novo HK gene discovery and assign condition and control labels for each sample. Once an analysis was started by the user, HouseKeepR will (i) download expression data using the 'GEOquery' R package [43], (ii) map microarray probe identifiers to gene identifiers such as Entrez or Ensembl, (iii) normalize expression values per dataset to reduce potential bias, and (iv) compute the log fold change between condition and control samples using the 'limma' Bioconductor package [44]. Once the analysis is complete, HouseKeepR reports the top performing HK genes together with informative visualizations of their ranks across data sets. Figure 3 provides an overview of these steps. HouseKeepR is released under an open-access license (<https://github.com/biomedbigdata/housekeepr>) and available for local use as a docker container or an online web application at <https://exbio.wzw.tum.de/housekeepr>.

7.4.11 HouseKeepR application to mouse and rat models of ischemic stroke

We demonstrate the practical application of HouseKeepR by validating de-novo identified HK genes with RT-qPCR. To this end, we selected rat and mouse as organisms, ischemia, ischemic or stroke as condition and brain as tissue. This query returned 163 data sets, from which we excluded those that did not satisfy our model criteria or that showed special model attributes that could affect predictions of housekeeping genes, such as genetically modified organisms. Finally, 12 high-quality data sets (Table 7.S3) were selected for further analysis.

7.4.12 Statistical analysis

Experimental results were presented as means \pm SEM. Differences between groups were determined by applying a one-way ANOVA followed by two-way ANOVA followed by Dunnett's Multiple Comparison test or by Student's two-tailed t-test and Mann-Whitney test experiments when appropriate. For repeated measurements, a two-way ANOVA was used. Statistical analysis was conducted using GraphPad Prism version 5.00. The level of statistical significance was set at $p < 0.05$.

7.5 References

- [1] T. J. Atkinson and M. S. Halfon. "Regulation of gene expression in the genomic context". In: *Computational and Structural Biotechnology Journal* 9 (Jan. 2014), e201401001. DOI: 10.5936/csbj.201401001.
- [2] F. Xie, T. P. Frazier, and B. Zhang. "Identification, characterization and expression analysis of MicroRNAs and their targets in the potato (*Solanum tuberosum*)". In: *Gene* 473 (Feb. 2011), pp. 8–22. DOI: 10.1016/j.gene.2010.09.007.
- [3] M. R. Fannon. "Gene expression in normal and disease states — identification of therapeutic targets". In: *Trends in Biotechnology* 14 (Aug. 1996), pp. 294–298. DOI: 10.1016/0167-7799(96)10041-x.
- [4] M. Kubista et al. "The real-time polymerase chain reaction". In: *Molecular Aspects of Medicine* 27 (Apr. 2006), pp. 95–125. DOI: 10.1016/j.mam.2005.12.007.
- [5] O. Thellin et al. "Housekeeping genes as internal standards: use and limits". In: *Journal of Biotechnology* 75 (Oct. 1999), pp. 291–295. DOI: 10.1016/s0168-1656(99)00163-7.
- [6] J. Huggett et al. "Real-time RT-PCR normalisation; strategies and considerations". In: *Genes Immunity* 6 (Apr. 2005), pp. 279–284. DOI: 10.1038/sj.gene.6364190.
- [7] N. Jain, S. Vergish, and J. P. Khurana. "Validation of housekeeping genes for normalization of gene expression data during diurnal/circadian studies in rice by RT-qPCR". In: *Scientific Reports* 8 (Feb. 2018). DOI: 10.1038/s41598-018-21374-1.
- [8] C. Tricarico et al. "Quantitative real-time reverse transcription polymerase chain reaction: normalization to rRNA or single housekeeping genes is inappropriate for human tissue biopsies". In: *Analytical Biochemistry* 309 (Oct. 2002), pp. 293–300. DOI: 10.1016/s0003-2697(02)00311-1.
- [9] K. Dheda et al. "The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization". In: *Analytical Biochemistry* 344 (Sept. 2005), pp. 141–143. DOI: 10.1016/j.ab.2005.05.022.

-
- [10] A. L. Schroder, K. E. Pelch, and S. C. Nagel. "Estrogen modulates expression of putative housekeeping genes in the mouse uterus". In: *Endocrine* 35 (Feb. 2009), pp. 211–219. DOI: 10.1007/s12020-009-9154-6.
 - [11] J. Vandesompele et al. "Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes". In: *Genome biology* 3 (2002), RESEARCH0034. DOI: 10.1186/gb-2002-3-7-research0034.
 - [12] M. W. Pfaffl et al. "Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper – Excel-based tool using pair-wise correlations". In: *Biotechnology Letters* 26 (Mar. 2004), pp. 509–515. DOI: 10.1023/b:bile.0000019559.84305.47.
 - [13] T. Hruz et al. "RefGenes: identification of reliable and condition specific reference genes for RT-qPCR data normalization". In: *BMC Genomics* 12 (Mar. 2011). DOI: 10.1186/1471-2164-12-156.
 - [14] C. L. Andersen, J. L. Jensen, and T. F. Ørntoft. "Normalization of Real-Time Quantitative Reverse Transcription-PCR Data: A Model-Based Variance Estimation Approach to Identify Genes Suited for Normalization, Applied to Bladder and Colon Cancer Data Sets". In: *Cancer Research* 64 (Aug. 2004), pp. 5245–5250. DOI: 10.1158/0008-5472.can-04-0496.
 - [15] C.-Y. Chia et al. "High expression stability of microtubule affinity regulating kinase 3 (MARK3) makes it a reliable reference gene". In: *IUBMB Life* 62 (Jan. 2010), pp. 200–203. DOI: 10.1002/iub.295.
 - [16] B. M. H. Keng et al. "Transcriptome Analysis of *Spermophilus lateralis* and *Spermophilus tridecemlineatus* Liver Does Not Suggest the Presence of *Spermophilus*-Liver-Specific Reference Genes". In: *ISRN Bioinformatics* 2013 (May 2013), pp. 1–8. DOI: 10.1155/2013/361321.

- [17] I. H. K. Too and M. H. T. Ling. "Signal Peptidase Complex Subunit 1 and Hydroxyacyl-CoA Dehydrogenase Beta Subunit Are Suitable Reference Genes in Human Lungs". In: *ISRN Bioinformatics* 2012 (2012), pp. 1–7. DOI: 10.5402/2012/790452.
- [18] F. Xie et al. "miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs". In: *Plant Molecular Biology* 80 (Jan. 2012), pp. 75–84. DOI: 10.1007/s11103-012-9885-2.
- [19] V. P. Mane et al. "Systematic method for determining an ideal housekeeping gene for real-time PCR analysis". In: *Journal of biomolecular techniques: JBT* 19 (Dec. 2008), 342–347.
- [20] C. Gubern et al. "Validation of housekeeping genes for quantitative real-time PCR in in-vivo and in-vitro models of cerebral ischaemia". In: *BMC Molecular Biology* 10 (2009), p. 57. DOI: 10.1186/1471-2199-10-57.
- [21] L. Jacob, J. A. Gagnon-Bartsch, and T. P. Speed. "Correcting gene expression data when neither the unwanted variation nor the factor of interest are observed". In: *Biostatistics* 17 (Aug. 2015), pp. 16–28. DOI: 10.1093/biostatistics/kxv026.
- [22] H. Noh and R. Gunawan. "Inferring gene targets of drugs and chemical compounds from gene expression profiles". In: *Bioinformatics* 32 (Mar. 2016), pp. 2120–2127. DOI: 10.1093/bioinformatics/btw148.
- [23] J. Ye et al. "Selection of suitable reference genes for qRT-PCR normalisation under different experimental conditions in *Eucommia ulmoides* Oliv". In: *Scientific Reports* 8 (Oct. 2018). DOI: 10.1038/s41598-018-33342-w.
- [24] C. Kleinschnitz et al. "Post-Stroke Inhibition of Induced NADPH Oxidase Type 4 Prevents Oxidative Stress and Neurodegeneration". In: *PLoS Biology* 8 (Sept. 2010). Ed. by M. McLeod, e1000479. DOI: 10.1371/journal.pbio.1000479.
- [25] A. I. Casas et al. "From single drug targets to synergistic network pharmacology in ischemic stroke". In: *Proceedings of the National Academy of Sciences* 116 (Apr. 2019), 7129–7136. DOI: 10.1073/pnas.1820799116.

-
- [26] A. I. Casas et al. "NOX4-dependent neuronal autotoxicity and BBB breakdown explain the superior sensitivity of the brain to ischemic damage". In: *Proceedings of the National Academy of Sciences* 114 (Oct. 2017), pp. 12315–12320. DOI: 10.1073/pnas.1705034114.
 - [27] R. Edgar. "Gene Expression Omnibus: NCBI gene expression and hybridization array data repository". In: *Nucleic Acids Research* 30 (Jan. 2002), pp. 207–210. DOI: 10.1093/nar/30.1.207.
 - [28] V. Verma and R. K. Aggarwal. "A comparative analysis of similarity measures akin to the Jaccard index in collaborative recommendations: empirical and theoretical perspective". In: *Social Network Analysis and Mining* 10 (June 2020). DOI: 10.1007/s13278-020-00660-9.
 - [29] I. Rivals et al. "Enrichment or depletion of a GO category within a class of genes: which test?" In: *Bioinformatics* 23 (Dec. 2006), pp. 401–407. DOI: 10.1093/bioinformatics/btl633.
 - [30] M. Köhler et al. "Validation of reference genes for the normalization of RT-qPCR gene expression in *Acanthamoeba* spp." In: *Scientific Reports* 10 (June 2020). DOI: 10.1038/s41598-020-67035-0.
 - [31] N. Silver et al. "Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR". In: *BMC Molecular Biology* 7 (2006), p. 33. DOI: 10.1186/1471-2199-7-33.
 - [32] S. Waxman and E. Wurmbach. "De-regulation of common housekeeping genes in hepatocellular carcinoma". In: *BMC Genomics* 8 (2007), p. 243. DOI: 10.1186/1471-2164-8-243.
 - [33] S. A. Iqbal et al. "Reproducible Research Practices and Transparency across the Biomedical Literature". In: *PLOS Biology* 14 (Jan. 2016). Ed. by D. L. Vaux, e1002333. DOI: 10.1371/journal.pbio.1002333.
 - [34] A. Lachmann et al. "Massive mining of publicly available RNA-seq data from human and mouse". In: *Nature Communications* 9 (Apr. 2018). DOI: 10.1038/s41467-018-03751-6.

- [35] A. Zoubarev et al. "Gemma: a resource for the reuse, sharing and meta-analysis of expression profiling data". In: *Bioinformatics* 28 (July 2012), pp. 2272–2273. DOI: 10.1093/bioinformatics/bts430.
- [36] C. Evans, J. Hardin, and D. M. Stoebe. "Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions". In: *Briefings in Bioinformatics* 19 (Feb. 2017), pp. 776–792. DOI: 10.1093/bib/bbx008.
- [37] C. Antoniades et al. "Association of plasma asymmetrical dimethylarginine (ADMA) with elevated vascular superoxide production and endothelial nitric oxide synthase uncoupling: implications for endothelial function in human atherosclerosis". In: *European Heart Journal* 30 (Mar. 2009), pp. 1142–1150. DOI: 10.1093/eurheartj/ehp061.
- [38] J. Egea et al. "Neuroprotection afforded by nicotine against oxygen and glucose deprivation in hippocampal slices is lost in $\alpha 7$ nicotinic receptor knockout mice". In: *Neuroscience* 145 (Mar. 2007), pp. 866–872. DOI: 10.1016/j.neuroscience.2006.12.036.
- [39] I. Buendia et al. "Neuroprotective mechanism of the novel melatonin derivative Neu-P11 in brain ischemia related models". In: *Neuropharmacology* 99 (Dec. 2015), pp. 187–195. DOI: 10.1016/j.neuropharm.2015.07.014.
- [40] E. Göb et al. "Blocking of plasma kallikrein ameliorates stroke by reducing thromboinflammation". In: *Annals of Neurology* 77 (Mar. 2015), pp. 784–803. DOI: 10.1002/ana.24380.
- [41] S. M. S. Lee and M. C. Pun. "On m out of n Bootstrapping for Nonstandard M-Estimation With Nuisance Parameters". In: *Journal of the American Statistical Association* 101 (Sept. 2006), pp. 1185–1197. DOI: 10.1198/016214506000000014.
- [42] T. Barrett et al. "NCBI GEO: archive for functional genomics data sets—update". In: *Nucleic Acids Research* 41 (Nov. 2012), pp. D991–D995. DOI: 10.1093/nar/gks1193.

-
- [43] S. Davis and P. S. Meltzer. "GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor". In: *Bioinformatics* 23 (May 2007), pp. 1846–1847. DOI: 10.1093/bioinformatics/btm254.
- [44] M. E. Ritchie et al. "limma powers differential expression analyses for RNA-sequencing and microarray studies". In: *Nucleic Acids Research* 43 (Jan. 2015), e47–e47. DOI: 10.1093/nar/gkv007.

Supplemental information

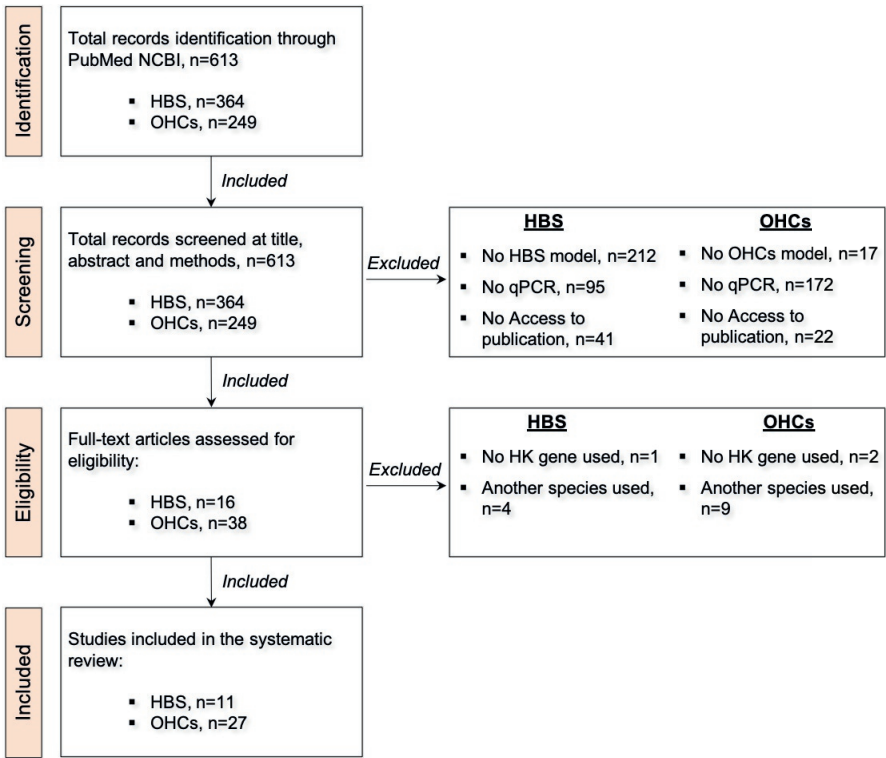


Figure 7.S1: Flow diagram of the housekeeping based systematic review. First, we proceed with suitable literature identification followed by partial screening (title and abstract), full text screening (results and methods sections), and final implementation of previously decided exclusion criteria (listed in the right boxes above).

Gene Symbol(s)	Ensembl ID(s)	#Data sets	Rank Mean	Rank Variance	Rank *
Mt-co1	ENSRNOG00000034234	11/12	5.3131	2.9023	1
Ubb	ENSRNOG00000042271	9/12	9.1297	3.272	2
Mt-atp6	ENSRNOG00000031979	7/12	9.5614	178.4173	3
Mt-co2	ENSRNOG00000030371	11/12	15.4169	5.6141	4
Fth1	ENSRNOG00000022619	11/12	20.1958	9.1929	5
LOC687780, Fau, LOC100360647	ENSRNOG00000020982, ENSRNOG00000046393, ENSRNOG00000047365	11/12	26.5895	8.3883	6
Ppia14d	ENSRNOG000000055068	6/12	27.4902	215.5905	7
Mt-co3	ENSRNOG00000030700	7/12	28.4778	178.9081	8
Mt-nd3	ENSRNOG00000033615	8/12	30.2172	10.7486	9
Cst3	ENSRNOG00000005195	12/12	30.6617	17.5866	10
Mt-nd1	ENSRNOG00000030644	10/12	30.7543	23.1536	11
Rps23	ENSRNOG00000016580	12/12	43.1639	12.6347	12
Mt-cyb	ENSRNOG00000031766	11/12	49.8794	26.3948	13
Actb	ENSRNOG00000034254	12/12	51.8344	39.2029	14
Tuba1b	ENSRNOG00000053468	10/12	60.3932	16.589	15
Rps3a	ENSRNOG00000011893	11/12	66.8536	19.6935	16
LOC108352650, LOC108350501, LOC108351482, Rps29	ENSRNOG00000004196, ENSRNOG00000028939, ENSRNOG00000029443, ENSRNOG00000032542	12/12	68.0881	46.319	17
Rplp2, LOC100911575	ENSRNOG00000002116, ENSRNOG00000037607	12/12	68.2227	19.1185	18
LOC108351137	ENSRNOG00000018630	12/12	72.3901	25.4942	19
Tuba1a	ENSRNOG000000060728	10/12	75.1609	32.0659	20

Figure 7.S2: HousekeepR candidate genes ranking. The final ranking as shown in the HousekeepR application after the analysis has finished. This ranking shows the HGNC gene symbols as well as their corresponding Ensembl ID. Multiple splice variants are mapped to the same row. Indicated is also the number of datasets where this gene was found among the selected expression datasets, the mean and variance of ranks derived from the bootstrapping of all samples, and finally, the ranking of the HK genes within the current analysis.

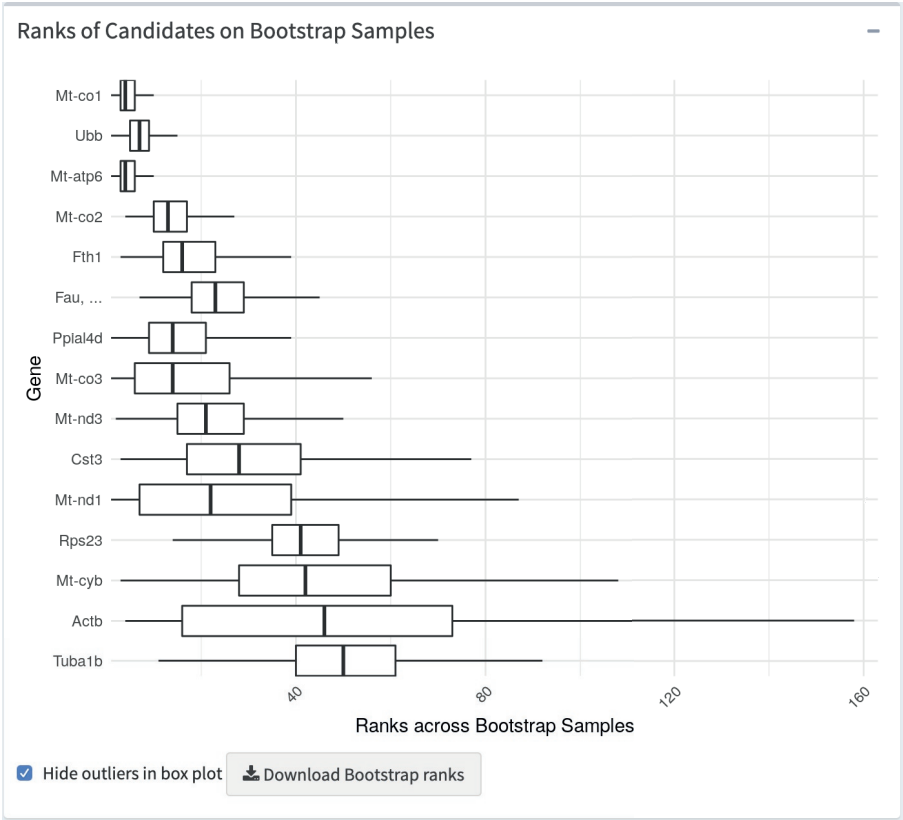


Figure 7.S3: HousekeepR gene ranking distribution over samples. To help with the explicability of results, HousekeepR shows a box plot presenting the distribution of HK candidate genes ranking over all the samples of the chosen datasets. The plot shows the median, first and third quartiles of the ranking for each candidate gene, while not taking into account the penalization based on missing samples. Since outliers could occur often when bootstrapping, a checkbox is added to hide them.

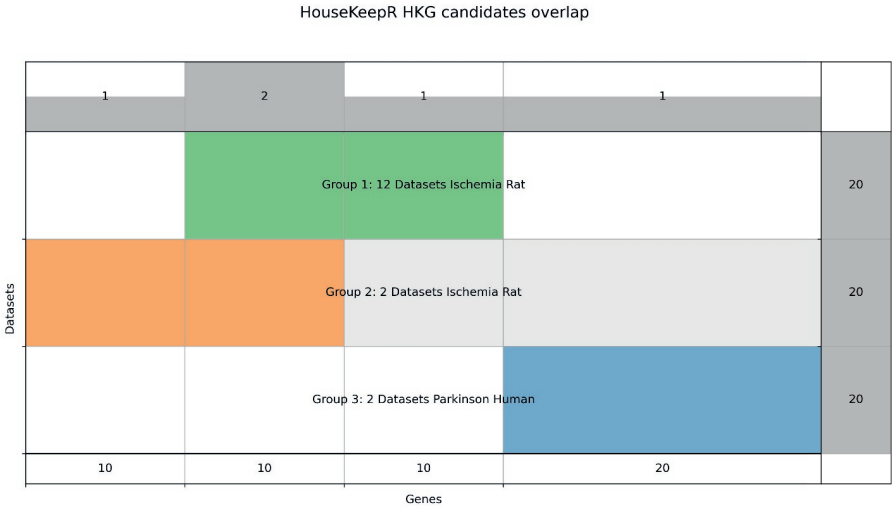


Figure 7.S4: Overlap between the top 20 gene candidates generated by HouseKeepR on 3 different groups of datasets. The 3 different groups illustrate the stability and reliability of HouseKeepR predictions. Groups 1 and 2 share the same tissue, condition and organism, i.e., brain, ischemia and rat, while group 3 shares only the tissue with the other, i.e. brain, Parkinson and human. Group 1 and 2 overlap in 10 out of 20 genes, while there is no overlap with group 3.

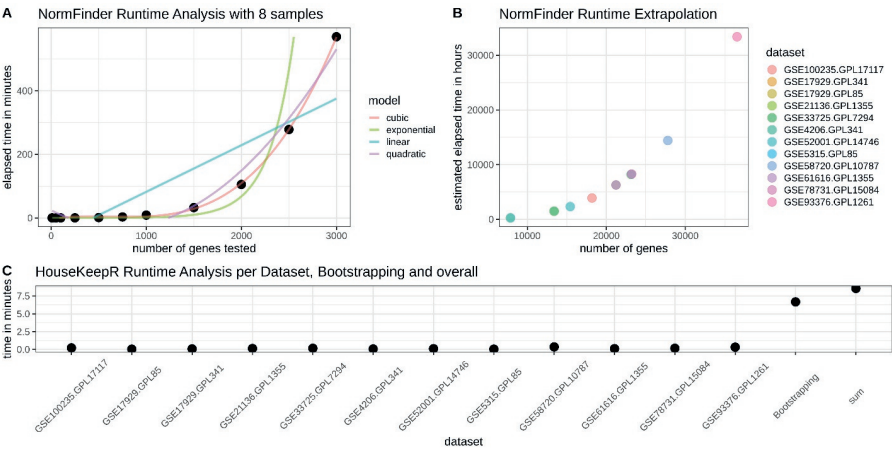


Figure 7.S5: Runtime comparison between NormFinder and HouseKeepR. (A) NormFinder analysis on a small dataset with 8 samples and up to 3,000 genes revealing a cubic increase in runtime. (B) Extrapolation of the runtime for the datasets selected in this study. (C) Runtime of the HouseKeepR approach including bootstrapping for the same dataset as well as the overall runtime.

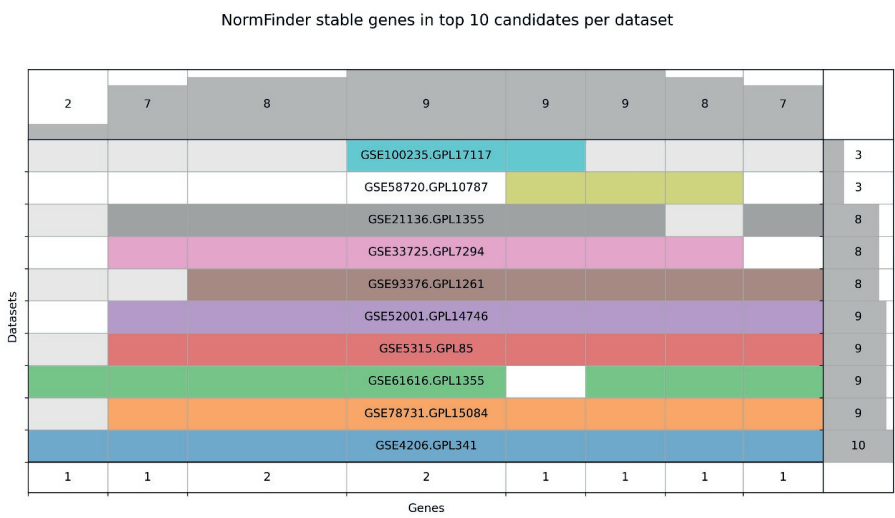


Figure 7.S6: Stability of top 10 HouseKeepR HKG candidates evaluated by NormFinder. For validating the accuracy of HouseKeepR housekeeping genes, the stability of genes was calculated within each dataset using NormFinder algorithm. Stable genes were considered to be with a stability score of 0.20 or lower. The figure shows how many genes out of the top 10 HouseKeepR candidates are stable (x-axis subtotals) as well as the overlap of stability across the different dataset used in the analysis (y-axis subtotals).

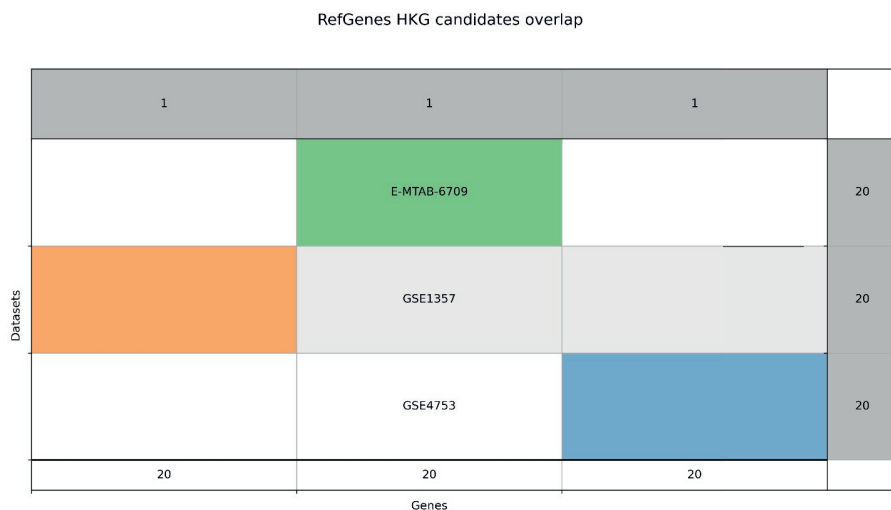


Figure 7.S7: Overlap between the top 20 gene candidates generated by RefGenes on three different datasets. The figure shows three different datasets with different underlying platforms (E-MTAB-6709 is an RNA-seq platform, while GSE1357 and GSE4753 are two different Affymetrix microarray platforms). There is no overlap between the gene candidates for each dataset since RefGenes relies only on standard deviation of expression levels, which are require normalization between platforms.

Table 7.S1: Search strategy systematic review

Model	Search term
HBS	(OGD OR oxygen-glucose deprivation OR oxygen glucose deprivation OR combined oxygen and glucose deprivation OR oxygen-glucose deprivation and reoxygenation OR oxygen and glucose deprivation) AND (hippocampal brain slices OR rat hippocampal brain slices OR rodent hippocampal brain slices OR acute hippocampal slices OR acute hippocampal brain slices)
OHCs	(OGD OR oxygen-glucose deprivation OR oxygen glucose deprivation OR combined oxygen and glucose deprivation OR oxygen-glucose deprivation and reoxygenation OR oxygen and glucose deprivation) AND (organotypic brain slices OR organotypic brain slice culture OR rodent hippocampal slice cultures OR rodent organotypic hippocampal slice cultures OR OHSCs OR organotypic hippocampal slice culture OR organotypic hippocampal slices OR organotypic hippocampal culture)

Table 7.S2: Candidate Reference genes

Gene symbol	Gene name	TaqMan assay number - Rat	TaqMan assay number - Mice
<i>β2-microglobulin</i>	<i>β2-microglobulin</i>	Rn00560865	Mm00437762
<i>β-actin</i>	<i>β-actin</i>	Rn00667869	Mm02619580
<i>Rpl13</i>	<i>Ribosomal Protein L13A</i>	Rn00821946	Mm02526700
<i>18S</i>	<i>Eukaryotic 18S rRNA</i>	Hs99999901	Hs99999901
<i>Hprt</i>	<i>Hypoxanthine phosphoribosyl-transferase</i>	Rn01527840	Mm03024075
<i>Sdha</i>	<i>Succinate dehydrogenase complex, subunit A</i>	Rn00590475	Mm01352366
<i>Ywhaz</i>	<i>Tyrosine 3-monooxygenase</i>	Rn00755072	Mm03950126
<i>Gadph</i>	<i>Glyceraldehyde-3-phosphate dehydrogenase</i>	Rn01775763	Mm99999915
<i>Ubb</i>	<i>ubiquitin B</i>	-	Mm01622233
<i>Fau</i>	<i>FAU ubiquitin like and ribosomal protein S30 fusion</i>	-	Mm02601595
<i>Fth1</i>	<i>Ferritin heavy chain 1</i>	-	Mm00850707
<i>Cst3</i>	<i>Cystatin 3</i>	-	Mm00438341
<i>Pp1a4d</i>	<i>Peptidylprolyl isomerase (cyclophilin)-like 4</i>	-	Mm01191872
<i>Rps23</i>	<i>Ribosomal protein 23</i>	-	Mm03019701
<i>Tuba1b</i>	<i>tubulin alpha 1b</i>	-	Mm00846967
<i>Rplp2</i>	<i>Ribosomal protein lateral stalk subunit P2</i>	-	Mm00782638
<i>Rps3</i>	<i>Ribosomal protein S3</i>	-	Mm00656272

Table 7.S3: Selected data sets for HouseKeepR ischemia analysis

GEO Data set Accession	Platform ID
GSE100235	GPL17117
GSE17929	GPL341
GSE17929	GPL85
GSE21136	GPL1355
GSE33725	GPL7294
GSE4206	GPL341
GSE52001	GPL14746
GSE5315	GPL85
GSE58720	GPL10787
GSE61616	GPL1355
GSE78731	GPL15084
GSE93376	GPL1261

8

General Discussion

8.1 Disease definitions: a major roadblock in medicine

In this thesis, I attempted to tackle two major roadblocks towards finding cures for non-communicable diseases. The first is our outdated definitions of diseases. Current clinical practice posits that diseases can be sufficiently described by the symptoms they present confined within body organs, leading to uniform clinical diagnosis and therapy for varied underlying molecular mechanisms. Contrarily, systems approach in medicine evidentiates the non-organ specific nature of diseases better described as perturbations in a multiscale network of molecular interactions [1]. However, achieving such a change is not a small feat, since the organ-based definitions are enshrined in the operational structure of the medical practice and its specializations. Hence, a broad analysis is conducted on the existing diseases taxonomies used in clinical and academic domains to measure the size of the problem. The difference between existing disease taxonomies are described and the extent of their suitability for use in precision medicine is studied. The conclusion drawn is that all of the disease classifications in use are not fully suitable for such purpose, and only by including omics data can disease classifications move towards more mechanistic definitions that take into account individual genetic and epigenetic makeup as well as environmental factors. A strategy is then outlined to reconcile current disease classifications with emergent concepts from network and systems medicine such as disease modules, endophenotyping, and mechanism-based drug therapies, where successful examples of this model are provided such as Asthma endophenotypes and cancer drugs based on mutations.

8.2 Network medicine and the future of disease diagnostics and therapeutics

Borrowing from the success of network science in understanding complex phenomena, several case studies are presented for the use of network medicine to discover disease mechanisms and predict drug targets for effective biomarkers and treatments. With regards to the role of NRF2 in metabolic processes and diseases, a protein-protein interaction (PPI) network was constructed and enriched with biological functions, then analyzed using an interactome-based proximity mea-

sure. Furthermore, the NRF2 diseasome, a multilayer disease-gene interactions network, is generated from shared genes, symptom similarity and disease comorbidity. Both networks are at different scales however complementing each other following a systems perspective, and yielding a wide set of diseases that NRF2 is related to either functionally or through interactions with neighbouring genes, such as digestive system disorders, cancers, metabolic and cardiovascular diseases, respiratory disorders, inflammatory-related conditions, and neurodegenerative diseases. Finally, the NRF2 interactome is used to generate hypotheses of drugs targeting one or multiple NRF2 implicated diseases and more importantly generation of drug repurposing candidates. This result is indeed in line with current and past studies, as well as approved drugs and ongoing clinical trials.

Regarding synergistic drug targets for ischemic stroke, a multilayer functional module was constructed from PPI, protein-metabolite, and drug-target interactions using a seeded approach, starting with the primary target NADPH oxidase type 4 (NOX4). Analysis of constructed network module using guilt-by-association, enrichment by Gene Ontology (GO) molecular function, and semantic similarity of the enriched labels, revealed nitric oxide synthase (NOS) enzyme family to be a promising candidate for synergistic drug targeting to NOX4. Further in-vitro and in-vivo validation showed that the network medicine approach was successful in predicting synergistic co-targets and hence reducing the risk of failure of single target and symptom-based drug discovery [2].

On whether network medicine is able to discover unknown disease mechanisms in hypertension, a seed-based PPI network is constructed and analyzed using multiple computational module identification methods. First neighbours analysis method is used to create a NOX5 subnetwork from the initial network derived from IID interactome database. Subnetwork participation degree is used to exclude hub nodes generating a further pruned topological module. Alternatively, two other module identification methods were used, global modularity optimization and agglomerative local search, arriving at a consensus result that among the NOX family only NOX5

is implicated in the nitric oxide GMP signaling. Further statistical analysis of a cohort of essential hypertensive patients, suggested that one-quarter of essential hypertensive patients fall under the newly discovered mechanism. Once more, this demonstrates the success of network medicine uncovering molecular mechanisms and redefining diseases accordingly enabling precise intervention.

8.3 Overcoming reproducibility crisis in the FAIR way

The second roadblock is the reproducibility of biomedical research which is a well-known issue with several studies raising flags on how cumbersome it is to replicate published research [3–7]. The problem is complex due to the intricate nature of manual steps and processes conducted through the course of studies and experiments. Additionally, the goal of reproducibility should not be restricted to the validation of results, however, reusing and extending previous methods is one of the core tenets of science. This thesis presents two solutions within the scope of the reproducibility problem. First, regarding the reproducibility of workflows, FAIR principles are sought to be applied to both computation and manual workflows. A new framework complementing existing approaches and tools is described that allows not only the input and output data of experiments to be FAIRified but also the workflow methods, implementation, provenance, and evolution to be captured and integrated into a standard RDF format that is findable, reusable, and extendable. To address the workflow decay phenomenon [8], a new semantic ontology was developed by reusing existing models such as DOLCE, W3C PROV, P-PLAN and BPMN 2.0. The new ontology gives researchers the ability to represent different parts of the workflow including manual steps that are essential to most scientific studies yet unappreciated by other approaches. To exemplify the new semantic model, PREDICT [9], a computational drug repositioning workflow based on machine learning, is FAIRified into OpenPREDICT then represented by a multistep plan, including both automated and manual steps, in the new ontology, which was published in a triple store used as a FAIR data point [10].

Regarding the question of predictability and reproducibility of

housekeeping genes for normalizing gene expression experiments, a new data-driven computational algorithm, HouseKeepR, was developed to overcome this issue via integrating and statistically analysing multiple gene expression studies with the same organism, tissue, and condition. Both in-vitro and in-vivo validations were carried out via an already established target, NOX4, for brain ischemia, and the predictions of the algorithm were shown to be superior to the commonly used methods, namely preselected gene subsets and systematic reviews. Repeatedly analyses of the algorithm produced the same housekeeping gene candidates given the same parameters. To bolster shareability and reuse, the HouseKeepR algorithm was made available through a user-friendly web application that can be used by experimental and computational researchers alike to find the optimal housekeeping genes for their experiments.

8.4 Conclusion

The medical practice is still relying on the Oslerian tradition where the patient is viewed as a collection of symptoms that when analyzed by the corresponding organ specialist will provide a diagnosis and a way of treatment [11]. With “the network takeover” [12] of medicine, however, network medicine is challenging this ideology, and trying to balance a deep understanding of the cell components and their disease correlations alongside a holistic view of the cell molecular interactions and their manifestations. While identifying the genetic associations of disease, as in GWAS studies, is an important step, it alone doesn’t provide us with a full picture of how disease phenotypes emerge from genetic variations. Network medicine has promised a different perspective on how we define and study diseases by introducing several organizing principles or hypotheses on which biological networks can be analyzed [13]. Several of these principles were put into application in this thesis to perform different tasks of in-silico discovery and hypothesis generation. The disease module hypothesis, for example, is used one time to identify a disease module for hypertension, and another time in reverse to find the different disease modules a gene product is involved in. Similarly, the local hypothesis is used to discover

synergistically interacting proteins involved in a stroke disease module.

Correct identification of disease modules is, therefore, an essential step to describe diseases within the interactome. Even though the human interactome is still incomplete, as not all positive and negative relations [14] between proteins have been mapped yet, we still can use it to construct and dissect disease maps to help us understand their mechanisms [15]. To address the incompleteness problem of the interactome, multi-layer networks can be used as a complementary solution, such as combining protein-protein interaction networks with protein-metabolite networks which was employed in one of the studies in this thesis [2]. Another accessible approach is to use machine learning algorithms to predict missing interactome links based on PPI networks with several features such as co-expression, co-localization, structural similarity, and functional similarity [16].

Network medicine approaches are creating a revolution in the way we understand diseases and discover drugs. However, it is still an enormous undertaking to successfully transform all our investigations into diseases using networks, and to propagate and adopt this change in the clinic. New frameworks and systems need to be in place in order to facilitate such change, and generic computational tools need to be developed to streamline the application of network methods to all diseases. Only then can we make significant progress towards decoding pathomechanisms, solving the diseaseome puzzle, and ensuring the development of effective precise cures.

8.5 References

- [1] R. Berlin, R. Gruen, and J. Best. “Systems Medicine Disease: Disease Classification and Scalability Beyond Networks and Boundary Conditions”. In: *Frontiers in Bioengineering and Biotechnology* 6 (2018). DOI: 10.3389/fbioe.2018.00112.
- [2] A. I. Casas et al. “From single drug targets to synergistic network pharmacology in ischemic stroke”. In: *Proceedings of the National Academy of Sciences* 116.14 (2019), 7129–7136. DOI: 10.1073/pnas.1820799116.
- [3] A. Mullard. “Reliability of “new drug target” claims called into question”. In: *Nature Reviews Drug Discovery* 10.9 (2011), 643–644. DOI: 10.1038/nrd3545.
- [4] C. G. Begley and L. M. Ellis. “Raise standards for preclinical cancer research”. In: *Nature* 483.7391 (2012), 531–533. DOI: 10.1038/483531a.
- [5] M. Baker. “1,500 scientists lift the lid on reproducibility”. In: *Nature* 533.7604 (2016), 452–454. DOI: 10.1038/533452a.
- [6] K. Mullane, M. J. Curtis, and M. Williams. “Chapter 1 - Reproducibility in Biomedical Research”. In: *ScienceDirect*. Ed. by M. Williams, M. J. Curtis, and K. Mullane. Academic Press, 2018, 1–66.
- [7] A. Stuppel, D. Singerman, and L. A. Celi. “The reproducibility crisis in the age of digital medicine”. In: *npj Digital Medicine* 2.1 (2019). DOI: 10.1038/s41746-019-0079-z.
- [8] J. Zhao et al. *Why workflows break — Understanding and combating decay in Taverna workflows*. 2012. DOI: 10.1109/eScience.2012.6404482.
- [9] A. Gottlieb et al. “PREDICT: a method for inferring novel drug indications with application to personalized medicine”. In: *Molecular Systems Biology* 7 (2011), p. 496. DOI: 10.1038/msb.2011.26.
- [10] R. Celebi et al. “Towards FAIR protocols and workflows: The OpenPREDICT case study”. In: *arXiv:1911.09531 [cs]* (2019).

- [11] D. H. Roukos. “Networks medicine: from reductionism to evidence of complex dynamic biomolecular interactions”. In: *Pharmacogenomics* 12.5 (2011), 695–698. DOI: 10.2217/pgs.11.28.
- [12] A.-L. Barabási. “The network takeover”. In: *Nature Physics* 8.1 (2011), 14–16. DOI: 10.1038/nphys2188.
- [13] A.-L. Barabási, N. Gulbahce, and J. Loscalzo. “Network medicine: a network-based approach to human disease”. In: *Nature Reviews Genetics* 12.1 (2011), 56–68. DOI: 10.1038/nrg2918.
- [14] P. Blohm et al. “Negatome 2.0: a database of non-interacting proteins derived by literature mining, manual annotation and protein structure analysis”. In: *Nucleic Acids Research* 42.Database issue (2014), D396–D400. DOI: 10.1093/nar/gkt1079.
- [15] J. Menche et al. “Uncovering disease-disease relationships through the incomplete interactome”. In: *Science* 347.6224 (2015). DOI: 10.1126/science.1257601.
- [16] J. Zahiri, J. Bozorgmehr, and A. Masoudi-Nejad. “Computational Prediction of Protein–Protein Interaction Networks: Algorithms and Resources”. In: *Current Genomics* 14.6 (2013), 397–414. DOI: 10.2174/1389202911314060004.

9

Summary

Most diseases are mechanistically not understood. Therapies thus largely focus on mitigating symptoms rather than cure. If the current crisis with COVID-19 has taught us one thing, it is that we must change our approach to speed up drug discovery. Reasons for this include our organ- and single-target-based focus on diseases and taxonomy in medicine in general, which extends into basic research. Basic research, moreover, is hardly able to translate its findings into applications as over 50 per cent of published biomedical research data are irreproducible.

Network medicine moves away from the reductionist view of diseases and defines them as emergent properties caused by the perturbations of a complex system that can be described as a network of interactions of its components. Here I present the application of network medicine approaches to answer multiple research questions (RQ) [refer to Chapter 1] pertaining to diseases mechanistic definitions. Additionally, I address the reproducibility problem by applying FAIR principles to scientific workflows, and tackle the gene expression normalization via arbitrary housekeeping genes problem that contributes to non efficacious drugs.

In regards to RQ1, current disease classifications are reviewed and their limitations in supporting disease mechanistic definitions, endophenotyping, and drug repurposing are highlighted. A data-driven strategy is then outlined where disease classifications rely on molecular and clinical data and recommendations provided on how it can be integrated in practice for diagnosis and treatment. Implications on the drug discovery process are discussed and examples of successful partial implementation are provided.

Network medicine techniques are then employed to understand the role of NRF2 in metabolic processes and diseases, as part of the answer to RQ2. Multiple networks are constructed with different underlying data, an interactome based on protein-protein interactions, and a diseasome based on disease-gene interactions. The output is a set of diseases that NRF2 involved in, including digestive system disorders, cancers, metabolic and cardiovascular diseases, respiratory disorders, inflammatory-related conditions, and neurodegenerative dis-

eases, and their shared common mechanisms, highlighting the importance of NRF2 in chronic diseases.

RQ3 questions the capacity to predict synergistic drug targets using network medicine. To that end, a multilayered network was constructed based on the validated primary target NADPH oxidase 4 (NOX4), and a functional network module was created using guilt-by-association and semantic similarity methods. Nitric Oxide Synthase (NOS) family is (NOS1-3) is shown to be the most synergistic targets to NOX4, which is then validated both in vitro and in vivo.

Concerning the ability to discover mechanisms of hypertension posed by RQ4, a PPI network is constructed to answer that question. Multiple topological analysis methods such as pruning by subnetwork-participation-degree, global modularity optimization, and agglomerative local search were used to identify a network module pertaining to the association of NOX with hypertension and NO-dependent vasodilation. The module showed that only NOX5 is a direct neighbour of endothelial nitric oxide-cyclic GMP signalling. On clinical validation, the levels of NOX5 were indeed observed to be directly correlated with the severity of hypertension in at least 25% of enrolled patients, suggesting the mechanism represented by the network approach as an endotype of hypertension opening the door for precision diagnostics and therapeutics.

RQ5 examines the possibility of applying FAIR principles, which are a set of guidelines describing how to make data easily findable, accessible, interoperable, and reusable, to scientific workflows and experimental protocols. The reproducibility crisis and especially in biomedical research has gained a lot of attention in the past few years as it challenges the advancement of medicine and threatens the drug development industry. Although the adoption of open access publications is growing, this is not a definite solution to the problem as reproducibility concerns all aspects of studies including the input and output data, exact method steps, and any parameters involved. In answer to this, a framework is developed to FAIRify workflows, as well as the data they consume and produce, where workflows are represented by a semantic model created within the approach.

Design and evaluation of the semantic model were done according to competency questions that are formulated from interviews with biomedical researchers. Moreover, the applicability of the framework was tested on a well-known drug repurposing workflow, converting it from an obscure one to a FAIR counterpart.

Under the same theme of reproducibility, RQ6 is concerned with the challenge of replicating the results of gene expression experiments. Normalization using housekeeping genes is addressed as a major issue, as experimental evidence asserts that there are no universally stable genes under all conditions, tissues, and in different organisms. HouseKeepR is an algorithm developed to solve this problem, by dynamically extracting a set of housekeeping genes for each specific combination of organisms, conditions, and tissues, hence, researchers can make unbiased selections. To ensure the reproducibility of the analyses of HouseKeepR itself, the algorithm is presented to the public as a free highly accessible web application.

Under such a scheme, easily reproducible network medicine approaches, it is believed that current diseases will thus become in the future rather symptoms of underlying causal pathomechanisms. And diseases will consequently be defined and named after that mechanism, representing a hot spot or vulnerable area within the human interactome. Several of my findings are now in clinical trialling, in stroke, heart failure and hypertension. Only then it can be said if these approaches represent proofs-of-concept for an entirely new outlook on disease and medicine in general. Based on the substantial pre-clinical validation it is, however, predicted that network and systems medicine will fundamentally change our healthcare system, from symptoms to mechanisms, from single targets to modules, from low precision to high precision, and eventually from treating to curing.

10

Impact

The costs of chronic diseases

Chronic diseases are the cause of 60% of all deaths around the world. This has a huge effect on societies and their economies. With 41 million deaths per year, and \$47 trillion projected global costs by 2030 as the world's population ages, the impact of finding innovative and effective therapies for these diseases is enormous.

The value behind network medicine approaches

In this thesis, I present the end-to-end value of network medicine approaches as a driver for medical innovation applied to two challenging areas responsible for almost one-third of the world's deaths: ischemic stroke and hypertension. Starting with complex networks modelling of the diseases and their underlying genetic modules, the causal mechanisms are analysed and then used to redefine appropriate endotypes. Utilizing our mechanistic understanding, multiple synergistic drugs are used to correct the dysfunctions of the networks, making use of drug repurposing to greatly slash the time to market. Both examples provide IP that benefits a large base and helps in alleviating social and economical hardship from struggling healthcare providers and their patients.

Stroke: the long-awaited therapy

In ischemic stroke, the analysis of the ROCG network has led to the first safe and widely applicable neuroprotective combination therapy. Three low dosage affordable compounds, namely NADPH oxidase (NOX) inhibitor, Nitric Oxide Synthase (NOS) inhibitor, and Soluble Guanylate Cyclase (sGC) stimulator, are used which are either already approved or in late clinical trials. This delivers a huge benefit in contrast to the only currently available drug, rt-PA, which has a very short window of administration of about 4.5 hours, costs a few thousand dollars, and has side effects such as bleeding. The combination therapy has been patented and was shown effective in both small and large animals trials. With clinical trials still to be carried out, collaborations with the industry or developing the patent into a spin-off company are two approaches on how to bring the end value to the

patients.

Hypertension: the unknown disease

Whilst stroke is the direct cause of many deaths and disabilities, hypertension is one of the main culprits. Clinically, hypertension is treated uniformly among the population wherein 90% of the cases it manifests with unknown molecular causes. Patients are therefore grouped under one generic disease definition called essential hypertension which is treated through lowering of blood pressure. In this thesis, the first causal mechanism of hypertension is discovered by investigating the human interactome network pruned for studying the role of NOX5 in hypertension. The presented mechanism identifies an endotype of hypertension which accounts for around 25% of hypertensive patients, where NOX5 induces the uncoupling of NOS causing impairment of endothelium-dependent vasodilation. Therefore the proposed treatment is to administer NOX5 specific inhibitors plus NOS recoupling agents to the affected subgroup. The end-value for patients is two-fold, a precise biomarker to stratify seemingly equivalent phenotypes and a mechanistic therapy targetted at the main cause with readily available compounds.

Repeatable success

An important gap that might hinder the generation of scientific and social value is the reproducibility issue. No reproducible research, even if it follows better paradigms, causes not only public distrust in the value of research but also wasting months of research and millions in funding. Therefore, contributing to solving this issue is another aspect of the value of this thesis. A framework based on FAIR principles, which are well-accepted guidelines for reproducibility of research, is developed to allow the reproduction of scientific workflows. The requirements of this new framework were collected from researchers in different subfields of the biomedical domain. A major added value is the inclusion of workflows with manual steps, which are very common within the field. Additionally, competency questions were formulated to be used in the evaluation of other frameworks as to whether they satisfy FAIR principles and the requirements of the scientific com-

munity. The next step is to develop a set of tools that will automate the procedures of the framework for multiple platforms, allowing researchers to find, import, reuse others' workflows as well as publish their own. This, in turn, will allow new approaches such as network medicine ones to be easily accessible by scientists, lower the need for multiple validation studies, and focus scientific attention on the most promising results.

The end of all chronic diseases

The presented network medicine approach is not exclusive to the examples demonstrated in this thesis, but can and should be generalized to all chronic diseases. Only through understanding the molecular mechanisms of diseases and redefining them can we start finding precise and affordable cures, that would drive down costs of healthcare and have a positive social and economic impact. This approach is only successful due to the collaboration of biomedical, clinician, and data scientists, and has the potential to engage the pharmaceutical industry, healthcare providers, and insurance companies. Ultimately, the goal is to make our healthcare systems fair and effective, gain back the trust of the public that medicine can not only treat but cure all chronic diseases.

Included work

Chapter 2

Elisa Anastasi, **Ahmed A. Hassan**, Johannes Boltze, Ana I. Casas, Mahmoud Elbatreek, Alexander G. B. Grønning, Zeinab Mamdouh, Cristian Nogales, Edwin K. Silverman, Anil Wipat, Harald H.H.W. Schmidt, Emre Guney. "On the need to overcome symptom-based and domain-specific disease classifications towards translational systems medicine" – *First Full Draft*

Chapter 3

Antonio Cuadrado, Gina Manda, **Ahmed A. Hassan**, et al. "Transcription Factor NRF2 as a Therapeutic Target for Chronic Diseases: A Systems Medicine Approach". Published in: *Pharmacological Reviews*. DOI: 10.1124/pr.117.014753

Chapter 4

Ana I. Casas, **Ahmed A. Hassan**, Simon J. Larsen, Vanessa Gomez-Rangel, Mahmoud Elbatreek, Pamela W. M. Kleikers, Emre Guney, Javier Egea, Manuela G. López, Jan Baumbach, and Harald H. H. W. Schmidt. "From single drug targets to synergistic network pharmacology in ischemic stroke". Published in: *Proceedings of the National Academy of Sciences*. DOI: 10.1073/pnas.1820799116

Chapter 5

Mahmoud H. Elbatreek, Sepideh Sadegh, Elisa Anastasi, Emre Guney, Cristian Nogales, Tim Kacprowski, **Ahmed A. Hassan**, et al. "NOX5-induced uncoupling of endothelial NO synthase is a causal mechanism and theragnostic target of an age-related hypertension endotype". Published in: *PLOS Biology*. DOI: 10.1371/journal.pbio.3000885

Chapter 6

Remzi Celebi, Joao Rebelo Moreira, **Ahmed A. Hassan**, Sandeep Ayyar, Lars Ridder, Tobias Kuhn and Michel Dumontier “Towards FAIR protocols and workflows: the OpenPRE-DICT use case”. Published in: *PeerJ Computer Science*. DOI: 10.7717/peerj-cs.281

Chapter 7

Ana I. Casas*, **Ahmed A. Hassan***, Quirin Manz, Christian Wiwie, Pamela Kleikers, Javier Egea, Manuela G. López, Markus List, Jan Baumbach, Harald H.H.W. Schmidt. “Un-biased housekeeping gene selection for high-validity expression analysis” – *Submitted to BMC Biology*

*Have contributed equally to this work

About the author

Ahmed A. Hassan is born and raised in the beautiful coastal city of Alexandria in Egypt. He obtained his bachelor's degree in Communications and Electronics engineering from Alexandria University in 2012, and his master's degree in Machine Learning and Data Mining from Université Jean Monnet in Saint-Etienne, France in 2016.

Ahmed is a data and machine learning enthusiast who likes to research and reshape data to derive new meanings and teach machines how to better represent it. In 2016, he moved to Germany to start a career in data science, however, his affinity for research led him to Maastricht in 2017 to start a PhD journey in the new and exciting field of network medicine. In the course of the next 4 years, he worked within the brilliant research department of Pharmacology and Personalized Medicine under the supervision of Prof. Harald H.H.W. Schmidt, where he worked in an interdisciplinary environment contributing to the computational output of the group.

He also worked in close collaboration with other computational research groups in larger EU projects on biomedical data homogenization, precision medicine and drug repurposing. He attended several conferences, workshops, and mentored undergraduate students.

Maastricht has been a great influence on Ahmed with its nice weather, friendly people and social warmth, but most importantly, as he found the love of his life in the city, whom he married in 2019. On the personal side, Ahmed enjoys playing football, stargazing, and reading about astronomy, philosophy and religion.

Acknowledgments

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ
وَعَلَّمَكَ مَا لَمْ تَكُنْ تَعْلَمُ وَكَانَ فَضْلُ اللَّهِ عَلَيْكَ عَظِيمًا. (57:21)
ذَلِكَ فَضْلُ اللَّهِ يُؤْتِيهِ مَنْ يَشَاءُ وَاللَّهُ ذُو الْفَضْلِ الْعَظِيمِ. (4:113)

I start by praising and thanking Allah, the most merciful, the most wise and the most generous, for giving me strength and steadfastness every step along the way towards the PhD.

My deepest and utmost gratitude goes to Harald, my supervisor and mentor, and a positive force in my PhD years until the last moment. I am first and foremost thankful for the opportunity you have given me to work with such a great team with a significant vision. I have learnt a lot from you, and I am forever indebted for all the help you have given me over the years, both professionally and personally.

I would like to extend my deepest appreciation to my colleagues at PPM lab, Ana, Mahmoud, Mayra, Cristian, Alexandra, Zeinab, Ivo. You have made this journey worthwhile and I have learnt a lot from all of you. Thank you for the meaningful work on a common research vision, and thank you for all the good moments we shared along the road. Special thanks owed to Marlou, who went to great lengths to support me in all administrative aspects of my PhD, and made sure everything is done in a timely fashion.

I wish to thank Michel Dumontier, Jan Baumbach, Emre Guney, Markus List, and Christian Wiwie whose support was instrumental in completing my PhD.

I am also grateful to all my friends who offered me their encouragement and support during the PhD, Omar Samir, Maged Aldehri, Hesham Elnakhlawy, Mohamed Hanafy, Karim Zaghloul.

Acknowledgments

Finally, I cannot begin to express my gratitude to my family. My parents who have always motivated me and pushed me towards success with all the mental and emotional support, I am forever grateful, and may Allah help me repay but a small part of all that you have selflessly given me throughout the years. My sisters, Nourhan and Sarah, thank you for your moral support and know that I am very proud of you both. My dear wife, Lydia, thank you for being beside me every step along the way. You have made this road much easier with your never-ending care, support and motivation.

Ahmed A. Hassan
Maastricht
2021-07-07

